Biological warfare spectroscopy: Development of rapid reagentless, *in situ* detection of biological threat agents and assessment and correction of the environmental conditioning on the vibrational spectrum.

Claire Louisa Pickering, BSc.

Department of Pure and Applied Chemistry

PhD Thesis 2018

Volume I

‘This thesis is the result of the author’s original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.’

Signed: Date:

For my Nan, my inspiration and driving force in everything I do.

Acknowledgements

Firstly, I would like to thank my family. Without their continuing love and support I would never have considered starting a PhD, let alone been able finish one! Thank you for everything you have done and continue to do for me.

I would like to thank my supervisor, Dr Matthew Baker, and co-supervisor Prof Roy Goodacre, without whose advice and guidance I would not have been able to submit this thesis. Thanks also to Elaine Perkins and Will Sellors at DSTL for their support.

I would also like to thank all of my friends, particularly Lizzie and Alan for helping to keep me sane and also going a bit insane with me throughout my PhD.

To my research groups, thank you for supporting me and being every bit as crazy as me especially Shawn, Katie, Kim, Holly, Lila, Angela, Kat and Mekhela. I have definitely made some friends for life.

A very special thank you goes to my partner James who has given me the strength and determination to produce this thesis. You have listened to me rant and rave about the highs and lows of this project, most of the time just nodding along waiting for me to shut up! Thank you for all of your love and support and also your frequent trips to the shop for coffee and cake!

List of Abbreviations

3-D -	3-Dimensional
Abs -	Absorbance
ALS -	Asymmetric Least Square
ANAA -	Advanced Nucleic Acid Analyser
ATCC -	American Type Culture Collection
ATR -	Attenuated Total Reflection
ATR FTIR -	Attenuated Total Reflection Fourier Transform Infrared
B. atrop -	*Bacillus atrophaeus*
BC -	Before Christ
BT -	*Bacillus thuringiensis*
BTK -	*Bacillus thuringiensis var. kurstaki*
BTKl -	*Bacillus thuringiensis var. kurstaki HD-1 cry-*
BTKx -	*Bacillus thuringiensis var. kurstaki HD-1 cry+*
BTWC -	The Biological and Toxin Weapons Convention
BW -	Biological Warfare
BWA(s) -	Biological Warfare Agent(s)
CaDPA -	Calcium Dipicolinic Acid
CaF_2 -	Calcium Fluoride
CARS -	Confocal Anti-Stokes Raman Spectroscopy
CART -	Classification and Regression Trees

CBRNE -	The National Strategy for Chemical, Biological, Radiological, Nuclear and Explosives
CCA -	Canonical Correlation Analysis
CCD -	Charge Coupled Device
CDC -	Centre for Disease Control
CFU -	Colony Forming Units
CO_2 -	Carbon Dioxide
COW -	Correlation Optimised Warping
CWA -	Chemical Warfare Agent
DF -	Discriminant Function
DFA -	Discriminant Function Analysis
DI -	Deionised
Di ATR -	Diamond Attenuated Total Reflection
DiZnSe -	Diamond Zinc Selenide
DNA -	Deoxyribonucleic Acid
DPA -	Dipicolinic Acid
DRIFTS -	Diffuse Reflectance Infrared Fourier Transform Spectroscopy
DSTL -	Defence Science and Technology Laboratory
DTGS -	Deuterated Triglycine Sulfate
E. coli -	*Eschericha coli MRE 162*
ESI-MS -	Electrospray Ionisation- Mass Spectrometry
FBI(s) -	Federal Bureau of Investigation(s)

FLAPS -	Fluorescence Aerodynamic Particle Sizer
FT -	Fourier Transform
FTIR -	Fourier Transform Infrared
GC -	Gas Chromatography
Ge ATR -	Germanium Attenuated Total Reflection
HCA -	Hierarchical Cluster Analysis
HT-FTIR -	High Throughput - Fourier Transform Infrared
HT-IR -	High Throughput – Infrared
HTS-	High Throughput Spectroscopy
ICPMS -	Inductively Coupled Plasma Mass Spectrometry
IR -	Infrared
IRE(s) -	Internal Reflection Elements(s)
KPLS -	Kernel Partial Least Squares
LDA -	Linear Discriminant Analysis
LPS -	Lipopolysaccharide
LWD -	Long Working Distance
MALDI-ToF-MS -	Matrix Assisted Laser Desorption/Ionisation - Time of Flight - Mass Spectrometry
MCT -	Mercury Cadmium Telluride
MoD -	Ministry of Defence
MS -	Mass Spectrometry
MVA -	Multivariate Analysis
NaCl -	Sodium Chloride

NR -	Noise Reduction
OD -	Optical Density
PC -	Principal Component
PCA -	Principal Component Analysis
PC-DFA -	Principal Component – Discriminant Function Analysis
PC-LDA -	Principal Component – Linear Discriminant Analysis
PCR -	Polymerase Chain Reaction
PDL -	Peptidoglycan
PE -	Potential Energy
PF -	*Pseudomonas fluorescens*
PLS -	Partial Least Squares
PLS-DA -	Partial Least Squares for Discriminant Function Analysis
PyMS -	Pyrolysis Mass Spectrometry
Py-GC-DMS -	Pyrolysis - Gas Chromatography- Differential Mobility Spectrometry
RNA -	Ribonucleic Acid
SERS -	Surface Enhanced Raman Spectroscopy
S/N -	Signal to Noise
SVM -	Support Vector Machine
TEM -	Transmission Electron Microscopy
TGS -	Triglycine Sulfate
ToF-SIMS -	Time of Flight - Secondary Ion Mass Spectrometry
UCLan -	University of Central Lancashire

UoM -	University of Manchester
UoS -	University of Strathclyde
UK -	United Kingdom
US -	United States
USA -	United States of America
UTI -	Urinary Tract Infection
UV -	Ultraviolet
UVRR -	Ultraviolet Resonance Raman
UV-Visible -	Ultraviolet-Visible
VCC -	Viable Cell Count
VN -	Vector Normalisation

List of Nomenclature

α-	Alpha
β-	Beta
°-	Degrees
°C-	Degrees Centigrade
gl^{-1}-	Grams per Litre
g-	Gravitational Acceleration
L-	Litres
µL-	Microlitres
µm-	Micrometres
ml-	Millilitres
mm-	Millimetres
M-	Molar
nm-	Nanometres
%-	Percent
cm^{-1}-	Reciprocal Centimetres
mm^{-1}-	Reciprocal Millimetres
s-	Seconds
Spores/L-	Spores per Litre

Table of Contents

Chapter 2: Bacterial Preparation, Construction of Substrate Library and Method Development for Substrate Removal using Vibrational Spectroscopy

Chapter 3: Downselection of Spectroscopic Technique for Bacterial Identification: FTIR vs Raman

Appendix D

Table of Figures

Chapter 1: A Review of Bacterial Biological Warfare and the Use of Vibrational Spectroscopy for Agent Identification

Chapter 2: Bacterial Preparation, Construction of Substrate Library and Method Development for Substrate Removal using Vibrational Spectroscopy

Chapter 3: Downselection of Spectroscopic Technique for Bacterial Identification: FTIR vs Raman

Chapter 4: Method Development for in situ Analysis of Surface Deposited Bacteria Using DRIFTS and the Effects of Environmental Conditioning

Chapter 5: Assessment of Effects of Environmental Conditioning on FTIR Spectra from Surface Deposited Bacteria using High Throughput Infrared Spectroscopy

Appendix A

Appendix B

Appendix C

Appendix D

Table of Tables

Chapter 1: A Review of Bacterial Biological Warfare and the Use of Vibrational Spectroscopy for Agent Identification

Chapter 2: Bacterial Preparation, Construction of Substrate Library and Method Development for Substrate Removal using Vibrational Spectroscopy

Chapter 4: Method Development for in situ Analysis of Surface Deposited Bacteria Using DRIFTS and the Effects of Environmental Conditioning

Chapter 5: Assessment of Effects of Environmental Conditioning on FTIR Spectra from Surface Deposited Bacteria using High Throughput Infrared Spectroscopy

Appendix A

Appendix B

Abstract

Current established bacterial identification techniques (cell culture and genetic analysis), are often costly and time-consuming processes. The ability to rapidly identify bacteria offers utility in a number of important areas, especially where pathogens could be left in a natural environment for a prolonged period of time, on various different backgrounds after their initial release. The 2015 UK National Security Strategy and Strategic Defence and Security Review (NSS SD) lists an attack on the UK or its Overseas Territories by another state or proxy using chemical, biological, radiological or nuclear (CBRN) weapons as a tier 2 threat; therefore, methods to identify biological warfare agents (BWAs) are a major priority for bio-defence.

Vibrational spectroscopy is a rapid, cheap and non-destructive technique that has previously been used to identify bacteria. This project uses varying temperature and humidity levels to represent a hot dry climate and assess the impact upon the bacteria using vibrational spectroscopy. The main focus of the project looks at the effect that substrate and environmental conditions have on the spectral signature of bacteria. The bacteria chosen for this study include surrogates of BWAs and bacteria that are commonly found in the environment.

The results of the project demonstrate that Fourier Transform Infrared spectroscopy is the optimal method for bacterial identification when compared with Raman for identifying samples found on complex matrices. Supervised and unsupervised multivariate analysis (MVA) of the data was performed using principle

component analysis (PCA) and discriminant function analysis (DFA) to show separation of Gram-type, bacterial strain and time point. This project shows the development of a methodology that can be used on a handheld spectrometer where the spectral contribution from a complex matrix is removed to provide a bacterial spectrum. This methodology has great promise for rapid, *in situ* identification of 'real world' BWA samples.

Chapter 1

A Review of Bacterial Biological Warfare and the Use of Vibrational Spectroscopy for Agent Identification

Claire L. Pickering[1,3], Holly J. Butler[1], Cerys A. Jenkins[1], Elaine Perkins[2], William Sellors[2], Royston Goodacre[3] and Matthew J. Baker[1] *

[1]WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK

[2] DSTL, CB Detection Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK

[3]School of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7ND, UK

* Corresponding Author: Dr Matthew J Baker email: matthew.baker@strath.ac.uk Tel No: 0141 548 4700. Twitter: @ChemistryBaker

Keywords: Biological warfare, Vibrational spectroscopy, Surface deposited, Environmental

1.1 Abstract

The use of biological warfare agents (BWAs) is a threat that is often addressed in national strategies, including the UK national security strategy, used to assess the current risks faced by a nation. Current methods used to identify these agents are

often laborious, destructive to the fungus, bacterium or virus, located centrally rather than in the field and are time consuming. Ideally, the situation requires a technique that can rapidly detect and identify the pathogenic agent allowing the implementation of the most appropriate treatment and hazard management procedures more quickly. Moreover, *in situ* detection and rapid identification of bacteria reduces the effect of terrorism and subsequent consequences of such an attack. Analytical techniques, such as mass spectrometry (MS) and vibrational spectroscopy (infrared (IR) and Raman), have been explored as potential approaches for the rapid field-deployable identification of BWAs. This review provides a critical insight into current techniques used for bacterial identification using historical evidence and discusses the potential of novel vibrational spectroscopic approaches in combatting the threat of BWAs.

1.2 Introduction

The use of bacteria in warfare is not a new concept and examples date back to 1400 BC where infected animals were used to spread disease in enemy camps. [2] The use of animals, decaying bodies and blood to cause disease within enemy camps was common in Europe as recently as the early 20th century. Figure 1 illustrates, in chronological order, some of the most significant incidents where biological warfare agents (BWAs) have been employed and can be split into three prominent epochs: 1400 BC-1914, 1914-1945 and post World War II.

	Pre World War I
14th Century BC	• Hitties used infected animals to spread tularaemia to thier enemies
800-146 BC	• Greek, Persian and Roman empires polluted water supplies with dead animals
1155	• Human bodies were used to pollute water in Italy by Barbossa
1346	• Mongals used dead bodies to spread plague during the siege on the city of Caffa
1422	• Lithuanian armies used plague infected bodies to infect the town of Carolstein
1495	• Leprocy infected blood is mixed with wine by Spaniards and sold to the French
1650	• Saliva from rabies infected dogs was used by Polish armies
1763	• French and English troops used the smallpox virus against Native Americans
1797	• The area surrounding Mantua was flooded by Napoleonic armies to aid the spread of Malaria
1863	• Smallpox and yellow fever clothing is sold to spread the diseases during the American Civil war
1899 and 1904	• The Hague conventions outlaw 'poison or poisoned arms' use but not bacterial weapons
	World War I and II
1914-1918	• Diseases caused in humans and animals include cholera, plague, anthrax and glanders
1925	• Creation of the Geneva Protocol bans the use of chemical and biological agents in war
1932	• Unit 731 (Japan) tests BWAs on human and animal subjects resulting in thousands of deaths
1937	• Development of the Pingfan Institute (Japan), the first large scale biological warfare complex
1939-1945	• Diseases during the second world war include anthrax, plague, glanders and typhus
1940	• Research into BW begins in UK government facilities at Porton Down, now part of DSTL
1940/41	• Japan uses grains, wheat and plague infected fleas to spread the disease within China
1942	• BWA research service was created at Fort Detrick (US) including production and testing facilities • Experiments using *B. anthracis* begin on Gruinard Island by British scientists
	Post World War II
1950-1980	• Vietnam, Korea and Afghanistan were accused of using BWAs, but claims were unsubstantiated
1972	• Creation of the BTWC ordered countries to cease the production of BWAs
Post 1974	• Iraq begin to develop a wide-ranging collection of BWAs, under the control of Saddam Hussein
1979	• Outbreak of anthrax in Sverdlovsk, Russia as a result of an accidental leak at a secret facility
1984	• Followers of Bhagwan Shree Rajneesh cult contaminate food in two Dallas restaurants
1995	• Aum Shrinkyo carry out a numer of chemical and biological attacks in Japan
2001	• "Anthrax letters attacks" carried out in America resulting in 5 deaths
2001	• Documents are found at Tanark farms, Afghanistan, showing plans for mass production of BWAs
2003	• Letters sent to the Department of Transportation and the White House contained traces of ricin
2009	• Anthrax found in heroin causing 47 confimed cases of infection with 13 deaths
2010	• Infected meat causes outbreak of anthrax in Bangledesh resulting in over 300 cases of infection
2015	• *B. anthracis* accidentally sent from Utah to Australia, South Korea, Canada and 17 states

Figure 1 Timeline detailing cases involving bacteria that can commonly be used as biological warfare agents, some of which are the offensive use of the agents. Information compiled from [1-13]

Prior to the commencement of the World War I and II, the use of BWAs remained fairly primitive in methodology; however, with the growth in knowledge of microbial pathogenicity and transmission as well as subsequent advancements within the field, specific virulent microbial agents have been exploited for offensive use. The authors would like to highlight [1-3], for more detailed information on the finer details of case studies and the history of BWAs. Also of particular interest is 'A Higher Form of Killing' [11,12] as it covers the historical use of both chemical and biological weapons compiling MoD documentation and information from Dr Rex Watson, the then Director of Porton Down. The UK national security strategy is a system that classifies risks such as biological and chemical attacks based on a three-tier system where the risks are categorised based on likelihood and impact. [18] The threat level for biological warfare currently within the UK is classed as a tier two threat. Current tier one threats include the likes of a cyber-attack or a natural disaster such as severe flooding while tier three threats include public disorder and severe heatwaves.

The modern concept of using bacteria as BWAs paradoxically came about as a reaction to their inclusion in the Geneva protocol, 1925, which highlighted their potential application for use in warfare. [19] The Japanese biological warfare offensive, led by Shirō Ishii initiated a research centre in 1935 where work to begin setting up the institute, referred to as the Pingfan Institute, was authorised in 1937 and was the first biological warfare complex in the world. At the Pingfan Institute mice, guinea pigs, sheep, horse and humans were all used as test subjects for the weapons being produced there. American investigators examined human remains and found that a number of different agents had been used causing diseases such as: anthrax (*Bacillus*

anthracis), plague (*Yersinia pestis*) and glanders (*Burkholderia mallei*) to name a few. The institute, and all documents held there, were deliberately destroyed prior to Russian armies arriving in 1945. [11] Although the scale of production at the site has never been confirmed, in 1949 Russia estimated production levels for the institute being around 8 tonnes of bacteria per month which in comparison to the scale of the American's efforts seems minimal as the site at Vigo in Indiana was capable of producing around 100 tonnes a month. [20]

During the early days of BWA development, Japanese production and testing was far superior to the Anglo-American efforts. Sir Maurice Hankey spearheaded the British biological program which was largely developed at the specialist laboratory at Porton Down. He began his work consulting with the Medical Research Council in 1934 with initial work focusing on "the practicality of biological warfare and to make recommendation as to the countermeasures which should be taken to deal with such an eventuality". [21] However, this soon changed and by 1938 emphasis had turned to offensive weapon development once German research into BWAs, particularly the work surrounding anthrax, began to surface. The main focus of the British biological warfare unit surround *B. anthracis* and its potential use as a BWA, this was due to its ability to form spores. [22] Spores are able to survive in harsh, nutrient deficient conditions for prolonged periods of time, a very useful property for BWAs to have, and hence their prevalence in current bioterrorism. Prolonged survival within nature can lead to outbreaks of disease after the initial attack once it is reintroduced to a nutrient rich environment, such as a human or animal host. [23] In 1948, the UK Ministry of defence (MoD) began trails on Gruinard Island, an island off the west

coast of Scotland, to test the persistence of *B. anthracis* spores and its potential for use as a BWA. [24] The study involved collecting soil samples annually between 1948 and 1968 and testing the effect of harsh environmental conditions over time on the virulence of the bacteria. The results of the study found that despite these parameters, the bacteria collected from the island compared to a laboratory strain of *B. anthracis* showed no loss of virulence. This Scottish island remained off limits to public access and has only relatively recently (1990) been decontaminated; however, due to the history of the island the population of Gruinard Island is still zero.

In 1969, it was suggested by Britain that a biological weapons convention should be created to prohibit the stockpiling and production of these agents. [19] The convention was finally implemented in 1972, when countries including Britain, the United States and the Soviet Union eventually signed. [4] A source of controversy arose due to evidence that the Soviet Union continued with their large-scale BW research program due to an accidental leak of *B. anthracis* in 1979 in Sverdlovsk, that resulted in 68 fatalities and hundreds of casualties. [25]

1.3 Microbiology theory

There are two main types of cells, prokaryotic and eukaryotic. Bacteria are prokaryotic cells while animals, plants and fungi are eukaryotic cells. Prokaryotic cells are defined as “simple organisms with no nucleus or organelles” and “the DNA present in the cell is not organized into chromosomes”. [26] Eukaryotic cells are

defined as “organisms that have a nuclear compartment that contains its genetic information”. [27] Figure 2 shows the structural differences in prokaryote and eukaryote cells. [28]

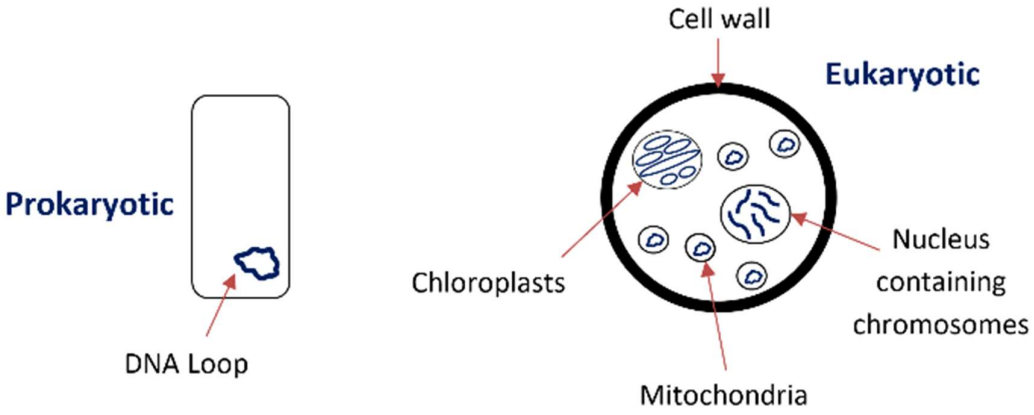

Figure 2: Adapted figure showing examples of prokaryotic and eukaryotic cells. The cells shown are not drown to scale. [26]

There are two main types of cells, prokaryotic and eukaryotic. Bacteria are prokaryotic cells while animals, plants and fungi are eukaryotic cells. Prokaryotic cells are defined as “simple organisms with no nucleus or organelles” and “the DNA present in the cell is not organized into chromosomes”. [26] Eukaryotic cells are defined as “organisms that have a nuclear compartment that contains its genetic information”. [27] Figure 2 shows the structural differences in prokaryote and eukaryote cells. [28]

There are a number of other features that differ depending on cell type, including size, generation time and the ratio of the components of the cell. Typically, prokaryotic cells (bacteria) are smaller cells, at around 1 µm, with a generation time of between 20 mins and 10 hours, while eukaryotic cells (yeast and animal) are much

larger, measuring between 10 and 100 μm with a generation time of 2-20 hours. The weight percentage of DNA, RNA, proteins, lipids and carbohydrates differs not only between pro and eukaryotes but also within the cell type, with variations seen between yeast and animal cells. Generally animal cells have higher weight percentages of DNA, RNA, proteins and lipids, with lower levels of carbohydrates compared to yeasts and bacteria. [29] A table with the proportion of macromolecules found within an Escherichia coli (E. coli) cell are shown below in Table I. This project focuses solely on the use and interrogation of prokaryotic cells, and therefore eukaryotic cells will not be discussed.

Table I: Macromolecules commonly found within E. coli represented as percentages within a dry cell. [29]

CELL COMPONENT	PERCENTAGE OF DRY CELL MASS
PROTIENS	55
RNA	21
LIPIDS	9
POLYSACCHARIDES	5
DNA	3
LIPOPOLYSACCHARIDES	3

Bacterial classification is primarily carried out using a differential staining technique known as Grams. Named from the German microbiologist, Dr Hans Christian Gram, the technique exploits key structural differences in the cell wall. Gram hypothesised that differing concentrations of fats, proteins and sugars within the cell wall. Gram-positive bacteria, which have a thick cell wall packed with PDL, produce a blue-to-purple colour as the crystal violet-iodide adheres very well to the peptidoglycan. [30] Gram-positive bacteria include Bacillus, Clostridium, Staphylococcus and Streptococcus. [31] Gram-negative bacteria have a thinner cell

wall containing less PDL and an outer membrane covered in LPS, making it difficult for the crystal violet-iodine to adhere to the cell. Thus, the crystal violet-iodine is easily washed away prior to the application of a Safranin, a red dye, which gives gram-negative bacteria a red appearance. This is a result of the lipids in the cell wall being dissolved by the crystal violet-iodine. [30] Gram-negative bacteria include Escherichia, Salmonella and Pseudomonas. [31] Gram staining is a very simple method that can be used to easily narrow down the potential identity of a bacterium, but does not provided a conclusive result as to its exact identity.

Another preliminary method of identification is to look at a bacterium's shape. Bacteria are single cell microbes that can vary in shape and are generally classified into one of five different groups which are shown in Figure 3. [32] Commonly encountered warfare agents such as Bacillus anthracis (B. anthracis) and Clostridium botulinum (C. botulinum), which cause the diseases anthrax and botulism respectively, are found to have a rod shape. [31]

Figure 3: The five shapes that bacteria can take from left to right: Rod (Bacillus), Spherical (Streptococcus), Spiral (Borrelia), Comma (Vibro) and Corkscrew (Campylobacter).

Gram staining and shape are features of a bacterium that can be used for preliminary indication but again cannot be used for the conclusive classification of

bacteria. [33] However, both Gram staining and shape can be used to rule out suspected bacteria in the case of an attack. If a particular bacterium is alleged to have been used, these techniques can immediately rule out the agent if the characteristics do not correlate to the known profile of that bacterium.

An interesting feature of some bacteria is their ability to form endospores which changes the spectral profile of a bacterium. Endospores enable a bacterium to survive in harsh, nutrient-deficient environments that what would otherwise be difficult for vegetative bacteria to survive. Endospores have an adjusted structure that has two membranes with a thin layer of spore-specific peptidoglycan that differs to the forms that can be found in the bacteria's vegetative state. [34] This layer of peptidoglycan can be identified by analytical techniques to indicate which form of bacteria you are working with, giving an indication as to what kind of environment a bacterium has been exposed to. It is the ability of B. anthracis to produce these endospores that makes it possible to find viable spores on Gruinard island some 24 years after they were deposited [24], making it a good biological warfare agent due to its persistence and robustness combined with high levels of pathogenicity.

B. anthracis is ideal to use as a BWA due to its ability to produce endospores that can remain inactive for many years, making them able to survive in extreme conditions. The bacterial spores only become active again when they return to a nutrient-rich environment, such as that of a human or animal host via ingestion, inhalation or cutaneous contact. As the spores enter the host their structure changes from the inactive to the active form of the bacteria. The bacteria rapidly replicate,

increasing the toxicity of the bacterium, causing severe illness or, in cases of inhalation, death, with fatality rates above 90%. When the body or carcass of a victim of anthrax poisoning is left to decompose naturally, the Bacillus anthracis form spores that can become airborne and infect a new host or can be ingested by other animals, continuing the cycle. [35]

1.4 Notable Anthrax Case Studies

B. anthracis is commonly the agent of choice when it comes to BW and was one of the agents chosen by Japanese cult Aum Shinrikyo. [36] Aum Shinrikyo is a Japanese apocalyptic cult that carried out a series of chemical and biological attacks, mainly taking place in Japan, between 1990 and 1995. During this time a total of seven biological attacks took place, three using botulinum toxin with the other four using *B. anthracis.* [36] All of the attacks used a spray mechanism for the dissemination of agents, which was considered one of the primary reasons the attack failed; although the failure was also associated with the use of vaccine or inactive strains of the agents. Samples collected from the scene in 1993, were only then analysed six years later, demonstrating the common delay associated with analysis of BWAs. [37] None of the biological attacks carried out by Aum Shinrikyo caused any causalities, however had the attacks been more successful slow detection and identification rates in currents techniques could have resulted in higher casualty and fatality rates due to longer exposure, delays and ill-informed treatment procedures.

Less than a decade later, an unrelated attack was carried out *via* the postal system in America on a number of US media and government offices using *B. anthracis* (causative agent of anthrax). [38] The attacks resulted in 22 casualties, including 5 fatalities. Letters were sent to the media and state senators containing solid and aerosolised forms of *B. anthracis* respectively. [38] Identification of the substance was confirmed as *B. anthracis* by employing standard cell culture techniques, using samples collected from the letters and envelopes. Blood samples, bronchial washings and pleural fluid were collected from patients were also analysed using polymerase chain reaction (PCR) and gamma phage lysis [39] Part of this methodology requires a three day incubation period, which resulted in a delay between the attack and identification of the bacterial agent. [40] Prolonged identification of BWAs has many serious implications, including delays in effective treatment, which highlights the need for a rapid and reliable technique to allow appropriate treatment to be administered and precautions to be implemented to avoid the spread of infection. Note that detection and identification are quite different: detection is used to establish that a threat has happened and to direct the most appropriate protective action to be taken on site; by contrast, identification is unequivocal and used for epidemiological purposes so that the most appropriate political response can be made (we note that George Bush Jr was President of the USA at the time). During the 2001 USA bioterrorism attack, DNA typing using amplified fragment length polymorphism was used to establish that a derivative of the virulent Ames strain (a laboratory strain for research, developing vaccines and tests) was the BWA employed. [41]

1.5 Issues with BWA Identification

For effective detection of BWAs, bacteria found in the field that require analysis, are likely to have been exposed to different environmental conditions and will be found on a variety of varying surfaces. All of these factors will affect the integrity of the bacterial samples, in turn affecting our ability to correctly identify and classify the bacterial species. This is one important factor that needs to be considered when developing a new technique for identification and classification of bacteria, specifically when considering rapid in situ characterisation of BWAs. The following section looks at current issues that exist in BWA detection and classification techniques.

There are a number of factors that govern whether a person will be infected by a BWA during an attack including the amount of time a person is exposed to the agent and the concentration of the organism they are exposed to. It is said that on average a human adult will inhale 10^{-1}00 L of air per minute, for *B. anthracis* to cause infection, 8000^{-1}0000 spores are needed. [17] So exposure at a concentration of 100 spores/L would take 10 minutes to inhale enough spores to cause infection. At a concentration of 1000 spores/L it would take just one minute for someone to be exposed to the same bacterial load. The previous two exposure times are calculated using the minimum inhalation volume and an infectious dose of 10000 spores and assuming no removal of spores. [17] In reality, an average healthy human may be infected more rapidly due to a larger air intake per minute; however, this provides a useful guide to the environmentally relevant detection limits that an identification

technique should be able to attain. High sensitivity (low limit of detection) and the ability to detect and identify bacteria rapidly are the two main features required to combat the low concentrations and short exposure times that cause infection. Furthermore, as bacterial warfare may occur in a variety of environments and settings, it is crucial that the method of detection is robust enough to be applied in a variety of scenarios, at a rate that is sufficient to minimise the effect caused during an attack. This will depend on the agent used, its concentration and the location where the attack takes place, however there is a desire to develop a method that allows for real-time analysis of samples. [42]

It is also important that the technique is able to discriminate between bacteria of similar genotypes accurately, and crucially between pathogenic and non-pathogenic strains. A review carried out in 2012 discusses various methods used to identify *B. anthracis* in environmental samples, including biochemical and nucleic acid based methods, and showed that identification with existing methods is particularly difficult due to the genetic similarity of *B. anthracis* to non-pathogenic environmental *Bacillus* spp. [12] The literature agrees that there is a desperate need for a novel rapid identification technique that is able to provide discriminatory information. [13,15] Some of the specific traits highlighted for a new technique include high sensitivity and specificity, small and portable instrumentation, a long shelf-life and rapid measurement times to name a few, with others shown in Figure 4. The latter of these attributes could be considered the most important, with 15 minutes considered a desirable analysis time. [14] High sensitivity and specificity allows for accurate and precise identification of the agents used to allow the optimum

treatment and hazard management procedures to be implemented, in turn reducing the effect of an attack. Rapid detection times are important for ensuring that the number of people exposed to the agent can be minimised. The ability to identify bacteria on any complex matrix would allow the technique to recognise areas that are contaminated and what type of bacteria are there.

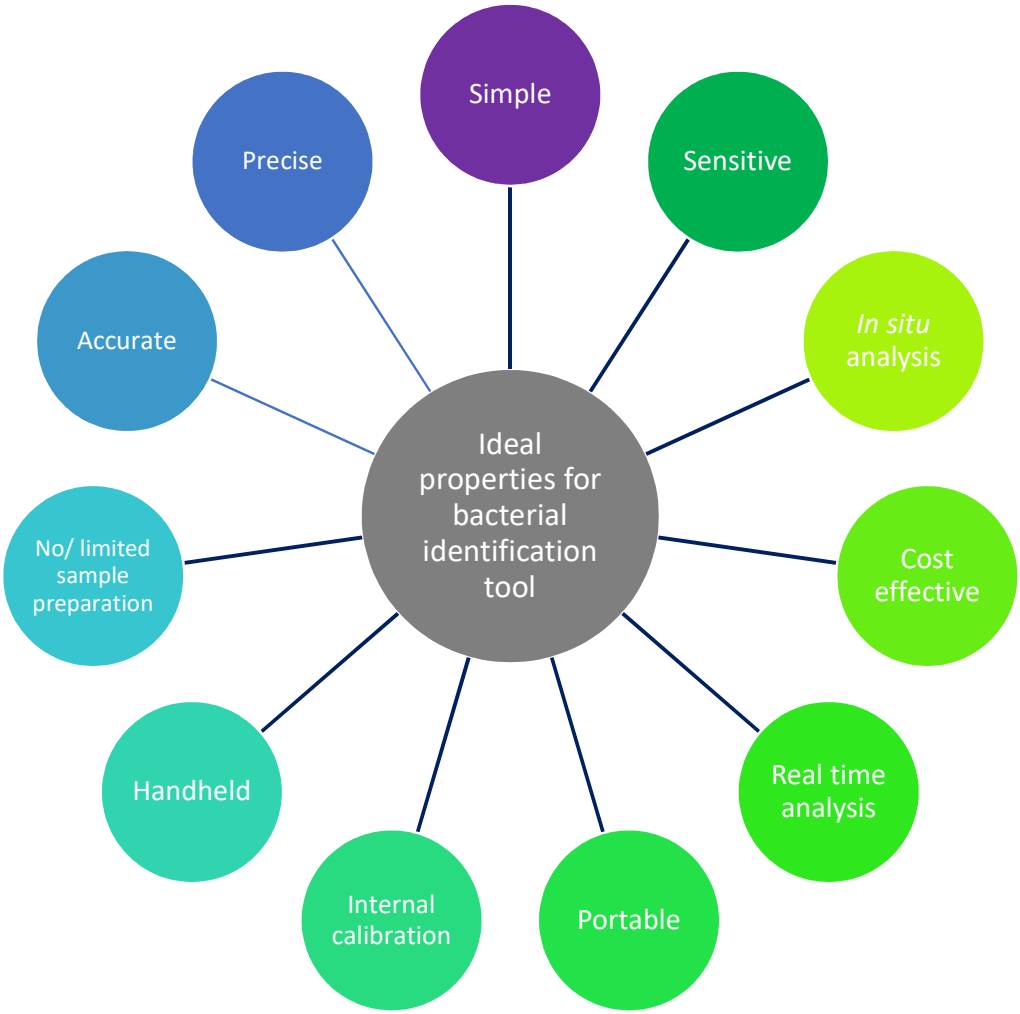

Figure 4 Diagram showing the ideal properties for a new biodetector/ identification method for use on BWAs. Compiled from [12,13, 14,15, 16,17]

1.6 Techniques Currently used for BWA Identification

Traditional methods used for bacterial identification, specifically cell culture and genetic analysis, are both costly and time-consuming processes largely associated with their extensive sample preparation. [14] Enzyme based tests, such as the catalase test, can also be used to classify a bacterium at a very high taxonomic level by establishing how the bacterial culture interacts with oxygen, and thus may infer whether the bacteria are aerobic, anaerobic, facultative anaerobic or microaerophilic organisms. During the catalase test, if the culture produces a foam when added to hydrogen peroxide, then the catalase enzyme is present. Tests such as this will take place during cell culture, which will be reliant upon a successful and sufficient period of culture, which in some instance may take more than seven days to show any sign of growth. [43] This will significantly impact the time taken to identify the bacterial species. However, tests like this lack sufficient discriminatory ability to separate closely related species unless many biochemical tests are performed in parallel. [44]

Polymerase chain reaction (PCR) technologies are commonly used for analysis of bacteria including BWAs. [45] When using PCR combined with a microarray-based assay, very high sensitivity and specificity values (up to 96 and 98% respectively) were produced within 3 hours from a set of 186 blood culture samples containing a variety of different genus. [46] Further development of PCR technologies has been focused upon reducing the time taken for identification. Research using real-time PCR has shown it is possible using pag and cap genes to detect a single *B. anthracis* spore in

100 L air within an hour. [47] Real-time PCR differs from conventional PCR as it monitors the fluorescent signal produced by the nucleic acids as the PCR process is occurring as opposed to after the PCR process has ceased. [48] Although this is faster than the conventional PCR method, it was initially far from the optimal detection time of under 15 minutes. [14] Other research groups that have been using real-time PCR brought identification times to just seven minutes with *Erwinia herbicola* cells by using an Advanced Nucleic Acid Analyser (ANAA). [49] Moving forward from identification of bacterial cells, the detection of bacterial spores is a more challenging process due to their nucleic acid being encased in a resistant shell (spore coat) meaning that the spores need to be disrupted prior to analysis, elongating the analysis time. In order to address this the development of a minisonicator allows for the disruption of bacterial spores in 30 seconds, dramatically reducing the time for the overall process bringing detection down to 15 minutes. [50] Although this reduces analysis times, this advancement does not address issues with instrument portability and benefit of little or no sample preparation.

Despite the advancements that have been shown, it is apparent from reviewing current techniques for bacterial identification, that there are a number of limitations in the currently available methods, particularly in terms of the need for optimum laboratory-based working environments (which for PCR need to be very clean in order to reduce contamination), speed of identification, cost and lack of portability. A method of identification would require improvements in these attributes in order to address sufficiently the requirements of a technique appropriate for BWA identification in the field. Equally as important are high

sensitivity and specificity to ensure that there is a balance between ease of use and quality of results. Crucially, for any new detection method we need to ensure that the number of false alarms is low but maintains high detection rates high to minimise potentially lethal exposure. Although there have been significant improvements in the 'gold standard' techniques (such as PCR), it is important to consider alternative approaches that may provide the attributes of an ideal technique for bacterial identification. Table II details the current and potential techniques that can be used for bacterial identification and compares the number of ideal properties each technique possesses, including FTIR and Raman spectroscopy.

The cost effectiveness of a technique depends on the specificity of a result and the cost of instrumentation and reagents, where in this case specificity relates to strain level identification. FTIR, Raman and mass spectrometry all have portable instrumentation that are cheaper than their benchtop equivalents, making them suitable *in situ* analysis. By having these cheaper, portable instruments the techniques become more cost effective, particularly for IR and Raman analysis where there is very little cost associated with an individual test following the initial outlay for the instrumentation. The only costs associated with these techniques are the regents used to clean them between samples. Mass spectrometry requires solvents for dissolving samples in, however these tend not to be too costly. Cell culture and enzyme tests often require multiple tests to be performed in order to get a result as specific as the instrument-based techniques, therefore they are not particularly cost effective as they require multiple reagents and equipment. Ease of use relates to how simple a method or technique is to use, from preparing a sample to collecting the

data, or visually inspecting the results of the tests. Mass spectrometry and PCR require a sample to be prepared prior to analysis using solvents or primers, making them more complex than a technique such as Raman or FTIR.

Table II Table of the ideal properties of a technique to identify of BWAs and how these correlate with the different techniques used for identification.

		Technique					
		Mass spectrometry	Cell culture/ enzyme tests	PCR	Real-time PCR	FTIR	Raman
Ideal properties	Sensitive	✓	✓	✓	✓	✓	✓
	Specific	✓	✓	✓	✓	✓	✓
	Portable	✓	X	X	✓	✓	✓
	Ease of use	X	X	X	✓	✓	✓
	No/little sample preparation	X	X	X	X	✓	✓
	Cost effective	✓	X	X	✓	✓	✓
	Real time analysis	✓	✓	X	✓	✓	✓

1.7 Spectroscopy fundamentals

Vibrational spectroscopy is an analytical technique widely implemented in the field of biological sciences [85,87,90], which is able to rapidly determine the chemical composition of a given sample. It has been extensively applied to identify the presence of bacteria, and also discriminate between bacterial species and strains at high rates of sensitivity and specificity. Since a number of proof-of-concept studies in

the 1980's, the number of papers published in this field has significantly increased (Figure 5), largely due to Dieter Naumann's pioneering work on Fourier transform infrared (FTIR) spectroscopy for bacterial characterisations at the Robert Koch Institute, Berlin, Germany. [82]

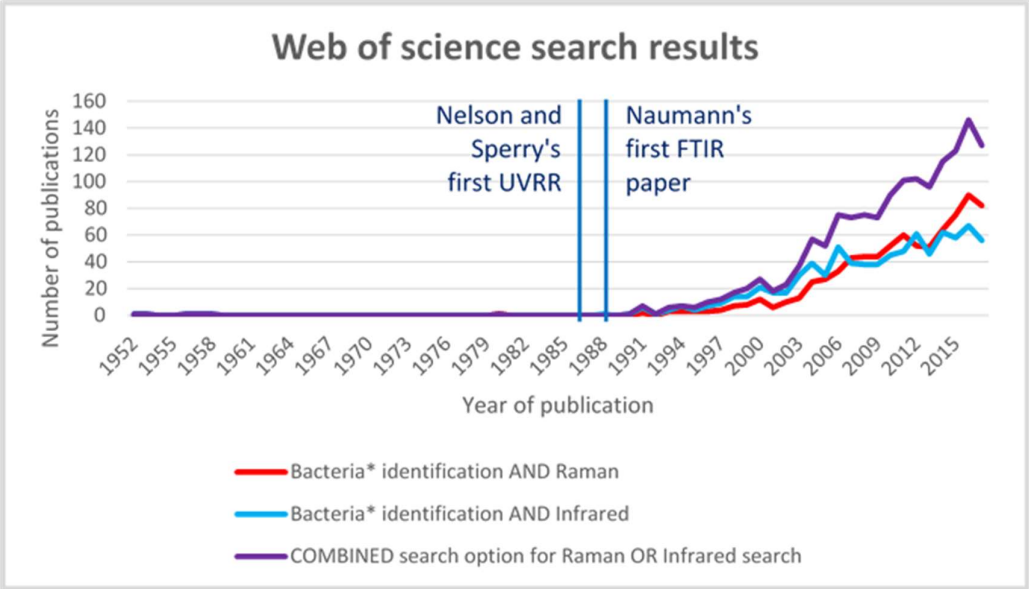

Figure 5 Web of science results for the number of papers published when conducting a search for 'Bacteria identification' AND either 'Raman' or 'Infrared' in the topic [51]*

Vibrational spectroscopy is a non-destructive technique that can derive highly specific chemical information with minimal sample preparation and fast analysis times and thus may be well suited towards bacterial identification. Moreover, the development of handheld instrumentation allows the possibility of portable vibrational spectroscopic analysis which further contributes towards this technique being ideal and translatable for use in the identification of BWAs in situ. [52]

Vibrational spectroscopy is a term widely used to describe the techniques used to measure the interaction of electromagnetic radiation with matter, and encompasses

the techniques of Raman and infrared (IR) spectroscopy. [46] This is based on the principle that this interaction causes the molecular bonds present in the sample to vibrate at discrete frequencies. Every compound produces a unique spectrum based on the composition and structure of the sample, with a spectrum often acting as a 'fingerprint' for a sample. [44,53] Although both techniques present similar information regarding the sample, IR and Raman are considered to be complementary techniques as some bonds which are observed in an IR spectrum will often not appear in a Raman spectrum, and vice versa. This is based on the mutual exclusion principle [45] and is due to the fact that both techniques rely on distinctly different principles: for a vibrational mode to be detected by Raman it will exhibit a change in molecular polarizability, while those modes considered to be IR active will demonstrate a change in dipole moment.

One issue that initially hindered the application of infrared spectroscopy for any biological analysis was spectral collection time, which heated the sample and caused degradation. It was only after the end of the Second World War that collection times were reduced down from hours to minutes [54,55], with now a days Fourier transform (FT) systems achieving excellent infrared spectral quality in seconds by using a Michelson interferometer and Triglycine Sulfate (TGS) detector. [56] Since then Deuterium Triglycine Sulfate (DTGS) has become the detectors of choice as it is cheap, good sensitivity and doesn't require any cooling before use unlike a Mercury Cadmium Telluride (MCT) detector. [57] Spectra can be acquired using a number of different sampling modes, that can be tailored the sample type, allowing flexible and optimised sampling. The sampling modes that have previously been used in the

analysis of bacteria are reflectance, transmission and attenuated total reflectance (ATR) and for a period diffuse reflectance absorbance approaches. [58] Transmission FTIR analysis works with wet samples that often require samples to undergo drying onto an IR transparent substrate during sample preparation as water absorbs strongly in the mid-IR (which is common in FTIR), which is a very easy process but adds time to the process. By contrast, an advantage of ATR is that limited drying of the sample is needed and an intimate contact between the sample and the ATR crystal (a highly refractive IR substrate), so measurements are quicker. Limited sample preparation is a key property to consider when trying to translate a technique for in situ analysis of samples.

Raman spectroscopy was first illustrated in 1928 by Sir Chandrasekhara Venkata Raman [59] and this built on previous predictions by Smekal in 1923. [60] Raman spectroscopy is based on light scattering and there are two general types of scattering types. Elastic scattering, also referred to as Rayleigh scattering, involves the excitation of a photon that is promoted to a virtual state before dropping back to the ground state with no overall energy loss or gain; this method usually results in the redirection of the light and the efficiency of Rayleigh scattering approximates to$1/\lambda 4$. By contrast, Raman scattering involves a change in energy of the incident light and is often referred to as inelastic light scattering. Stokes and anti-Stokes are examples of inelastic scattering that occur equidistantly from the Rayleigh line; in a Raman spectrum the x-axis refers to the wavenumber shift (units cm^{-1}) from the laser or Rayleigh line and is related to the vibrational frequency of the band, the y-axis is a measure of the energy of the vibration. Stokes scattering is found at a lower energy

than the Rayleigh line with anti-Stokes found at a higher energy level. [66] Stokes and anti-Stokes represent Raman scattering as the molecule gains or loses energy, respectively and a particular vibrational mode will have the same vibrational frequency and so will be equidistant from the Rayleigh line as Stokes would be (e.g. for a benzene ring) 1000 cm^{-1} and anti Stokes this occurs at 1000 cm^{-1}. Both phenomena are considerably much rarer than elastic scattering, with as few as 1 in 10^{6-8} photons being inelastically scattered. Stokes scattering is on the whole more common (ca. 10-fold), as most molecules naturally exist in the ground state at room temperature. [75] Raman spectroscopy differs from infrared in that IR promotes photons into a vibrational state, whereas Raman promotes the molecule to a virtual state before the molecule then returns to either a vibrational (Stokes Raman) or ground state (anti-Stokes Raman) with the release of a photon.

1.7.1 Electromagnetic spectrum

The electromagnetic spectrum represents all types of radiation, ranging from low frequency radio waves up to high frequency gamma rays (γ-rays). Figure 6 shows the different region of the electromagnetic spectrum with the corresponding frequencies, highlighting the mid infrared region. Rotational spectroscopy occurs in the far infrared and microwave region of the spectrum, as little energy is required for rotations to occur. [61] Vibrational spectroscopy uses the principles of the interaction of light with matter to identify a compound as the light causes the molecular bonds to vibrate. Vibrational spectroscopy is a term widely used to describe the techniques

of Raman and Infrared (IR) spectroscopy, which are primarily concerned with molecular/bond vibrations in the mid and near regions of the IR spectrum. The wavenumber range commonly interrogated in IR spectroscopy is 4000-400 $^{cm-1}$ where fundamental vibrations are detected. The near-infrared region is home to combination bands and overtones due to the higher frequency and higher energy radiation. Overtones are secondary bond vibrations that occur due to anharmonicity, allowing transitions over multiple electronic states from the ground state. These transitions cause weaker vibrations that occur at a higher wavelength than the fundamental bond vibration. The far-infrared region comprises low energy vibrations and rotations.

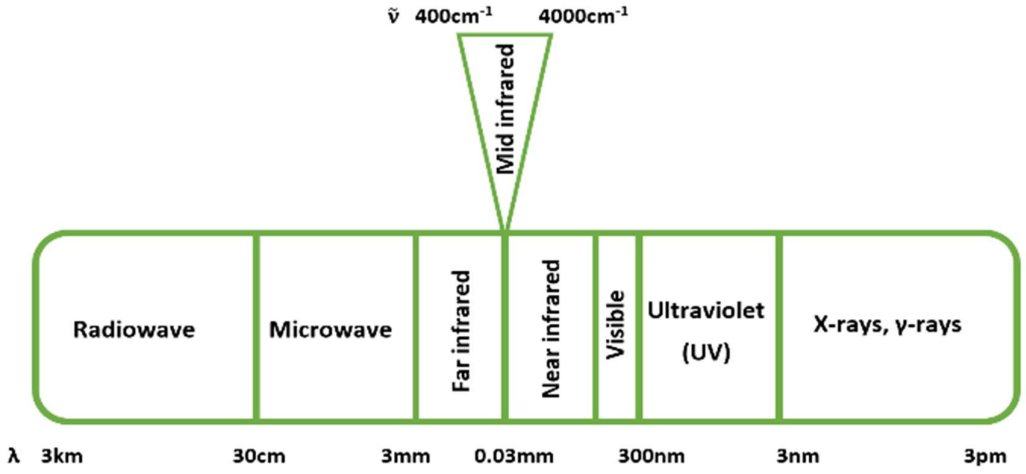

Figure 6: Electromagnetic spectrum with the key Mid-infrared region highlighted with the corresponding wavenumber range.

Frequency is a measure of the number of cycles per unit time, while wavelength is the distance between two peaks within a wave, shown in Figure 7. [62] When considering spectroscopy, it is important to consider the relationship between a number of different variables including frequency and wavelength.

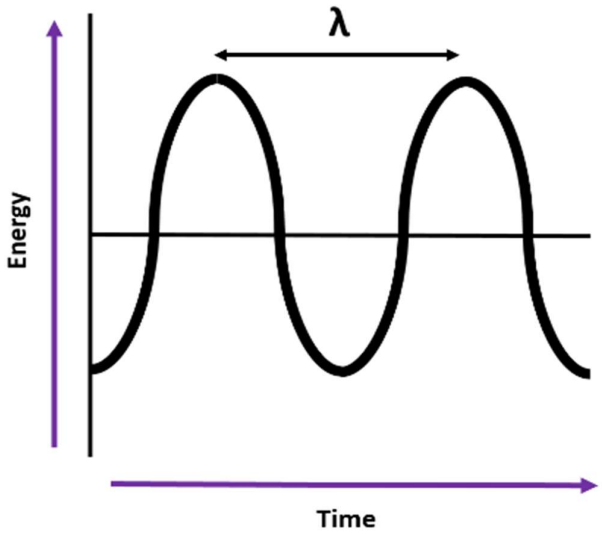

Figure 7: Schematic depicting a wave, defining the measure of wavelength by showing the distance between two peaks.

One key equation to consider for vibrational spectroscopy explores the energy emitted or absorbed by considering the frequency of the radiation (ν).

$$h\nu = |E_1 - E_2| \tag{2.1}$$

E_1 and E_2 are the energies that the transition has occurred between, ν is the frequency of the radiation and h is Planck's constant $= 6.63\ x10^{-34}$. Taking this into accounts, the relationship between wavelength and frequency can be expressed as

$$\lambda = \frac{c}{\nu} \tag{2.2}$$

where λ is the wavelength, ν is the frequency of the radiation and c is the speed of light. Wavenumber can be converted into wavelength using

$$\tilde{\nu} = \frac{1}{\lambda} = \frac{\nu}{c} \tag{2.3}$$

where $\tilde{\nu}$ is the wavenumber, ν is the frequency, λ is the wavelength and c is the speed of light.

1.7.2 Vibrational spectroscopy fundamentals

Every compound produces a unique spectrum based on the composition of a sample and different structural arrays, with a spectrum often acting as a 'fingerprint' for a sample. [63] This is due to the fact that different bonds vibrate at different wavelengths and the number of vibrational modes a molecule will exhibit is down to the linearity of the molecule. As such the following two simple equations are used

$$3N - 5 \qquad Linear\ molecules \tag{2.4}$$

$$3N - 6 \qquad Non - linear\ molecules \tag{2.5}$$

where N is the number of atoms with linear molecules using Eq. 2.4 and non-linear molecules using Eq. 2.5. [64] For example, considering a CO_2 molecule which has a centre of inversion and four types of vibrational modes (asymmetric stretching, symmetric stretching and two bending modes), the asymmetric stretch and both bends will be IR-active as a change in the dipole moment occurs and it therefore cannot be Raman active. The symmetric stretch is Raman-active as there is a change in the polarizability of the molecule; the symmetric stretch will therefore be absent from an IR spectrum. [65,66]

IR and Raman are often considered to be complementary techniques, as bonds that are not seen in the former will often appear in the latter and vice versa, so long as there is a centre of symmetry; this is the mutual exclusion principle. [67] However, this is not the case for all molecules, as for non-symmetric molecules multiple vibrational modes will be detected by the different techniques with some being detected by both IR and Raman. Raman-active molecules exhibit a change in the polarizability of the molecule [68] while IR works on the basis that if there is a change in the dipole moment the molecules will be IR-active. [64] Figure 8 below shows the fundamental differences between IR and Raman spectroscopy, which is that IR is an absorbance technique while Raman is a scattering technique. [69]

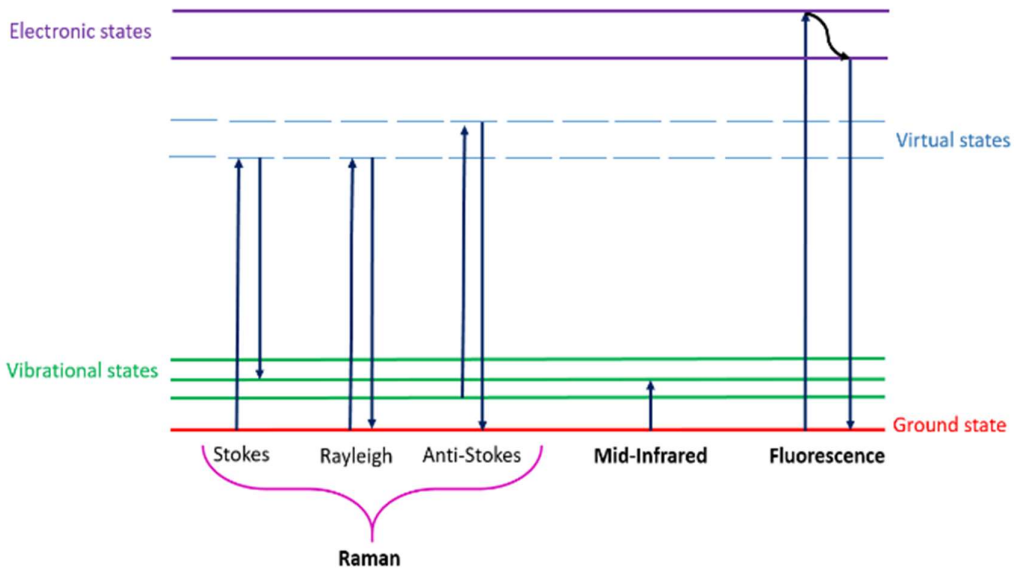

Figure 8: Adapted Jablonski diagram showing electronic transitions for Stokes, Anti-Stokes and Rayleigh scattering in Raman spectroscopy.

For a diatomic molecule, the potential energy (PE) of a harmonic vibration is calculated using

$$PE = \frac{1}{2} KX^2 \qquad (2.6)$$

Plotting the PE generated using Eq. 2.6 for increasing nuclear distances produces a symmetrical parabola. A molecule transition from one energy state to another occurs at discrete energy levels. These transitions are measured using vibrational spectroscopy. In reality, an anharmonic oscillator is a more accurate model to represent the potential energy required for a molecular transition. Figure 9 shows visually the difference between harmonic and anharmonic oscillations when considering the PE and internuclear distance. [70]

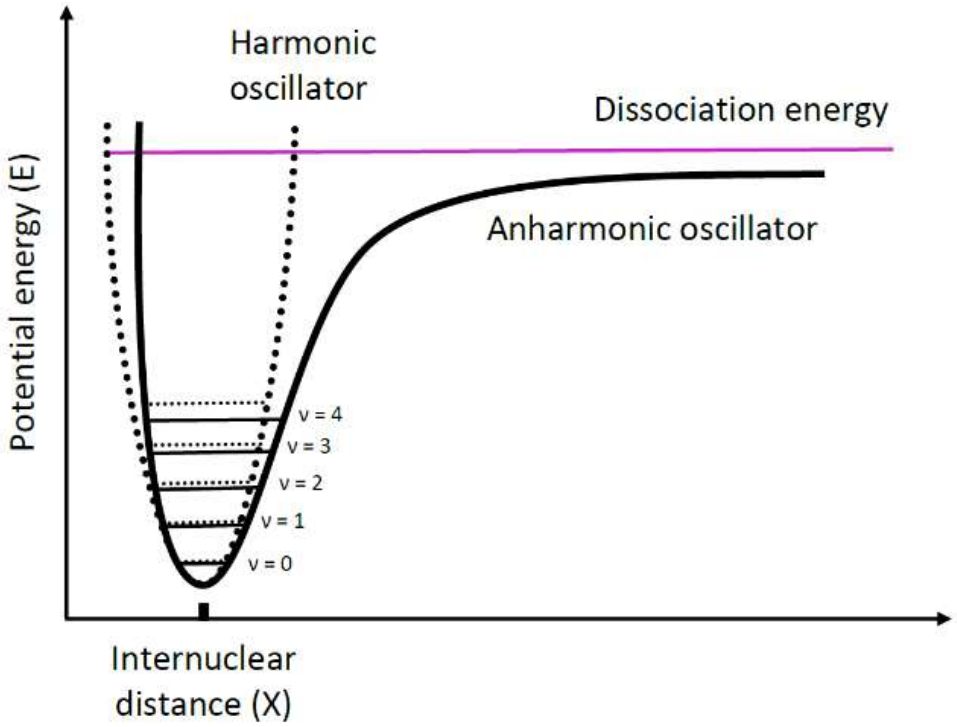

Figure 9: Potential energy (PE) level diagram illustration the difference between a harmonic and an anharmonic oscillator including the dissociation energy. v represents the energy levels where v = 0 is the energy level where fundamental transitions occur from.

As the energy levels of a harmonic oscillator are equidistant the energies are calculated using

$$E_i = \left(v_i + \frac{1}{2}\right) h\nu \tag{2.7}$$

Where v_i = 0, 1, 2….. For anharmonic transitions the PE is not linearly proportional to the increasing internuclear distance. With this, as the energy levels increase the energy between the levels required for transition to occur reduces, meaning the energy levels are not equidistant as they are in a harmonic oscillator. As the energy levels increase the energy level tends towards that required to cause dissociation of a molecule. Due to these differences, the energy equation is modified for this, with the $\mathrm{x}_e\nu_e$ term of Eq. 2.8 accounting for the anharmonicity of the molecule. The modified energy equation to incorporate anharmonicity is as follows

$$E_\upsilon = h\nu_e\left(v + \frac{1}{2}\right) - h\mathrm{x}_e\nu_e\left(v + \frac{1}{2}\right)^2 \tag{2.8}$$

Anharmonicity accounts for presence of overtones and combination bands, which are forbidden using the harmonic model where only fundamental transitions are allowed. Fundamental transitions ($\Delta\nu = \pm 1$) represent the majority of peaks observed within a spectrum. Combination bands and overtones ($\Delta\nu = \pm 2$) can also be present within a spectrum and are the result of a molecular transition through multiple quantised energy levels. As these types of transitions are less likely, especially when considered with fundamental transitions, the intensity of the peaks in a spectrum are far less intense than fundamental transitions. Figure 10 shows the difference between fundamental bands, overtones and hot bands. [70] Overtones occur when a molecule is excited by more than one energy level, for example from v=0 to v=2 which would be the first overtone, with the second overtone being a transition from v=0 to v=3 and so on. Combination bands are the result of multiple

fundamental vibrations occurring at the same time. Another type of band that can contribute to a spectrum is termed a hot band [17]. Hot bands occur when an already excited molecule transitions further to a higher energy level, for example, if a molecular transition from v=1 to v=2 occurs.

1.7.3 IR history

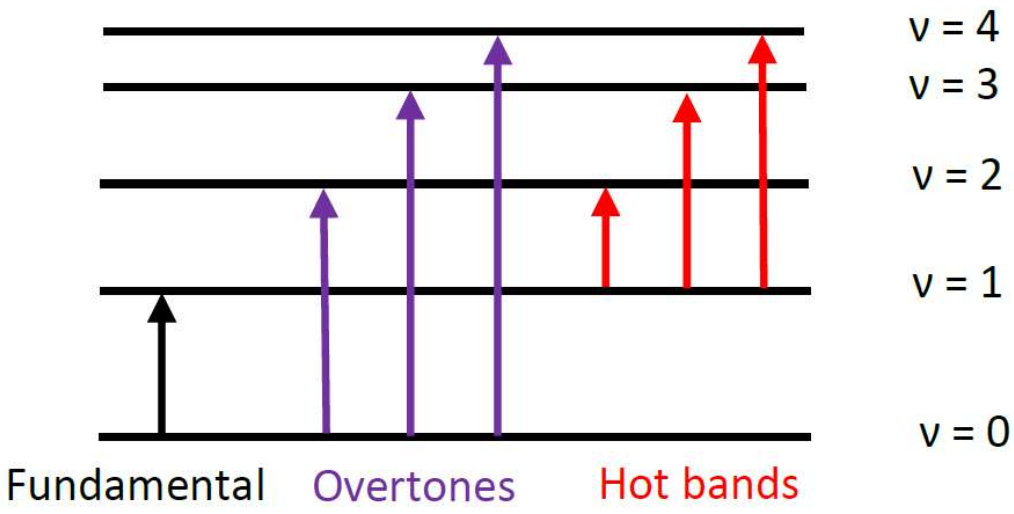

Figure 10: Examples of the differences between fundamental transitions, overtones and hot bands.

Infrared radiation was first discovered in 1800 by Sir William Herschel who was conducting research observing the solar system. [63] Prior to Herschel's work, heat and light were thought of as separate phenomena, but it has since become common to consider light and heat simultaneously. Herschel's son became the first person to make infrared radiation visible and reported the first multichannel spectrometer. One issue that hindered the development of infrared spectrometers was collection

time; it was only after the end of the Second World War that collection times were reduced down to 20 minutes. Modern Fourier transform systems have further reduced collection time down to seconds using the Michelson interferometer, shown in Figure 11, to produce high quality spectra. [71] One of the first commercially produced double beam spectrometers was made by Perkin Elmer in 1950.

1.7.4 IR instrumentation setup and sampling modalities

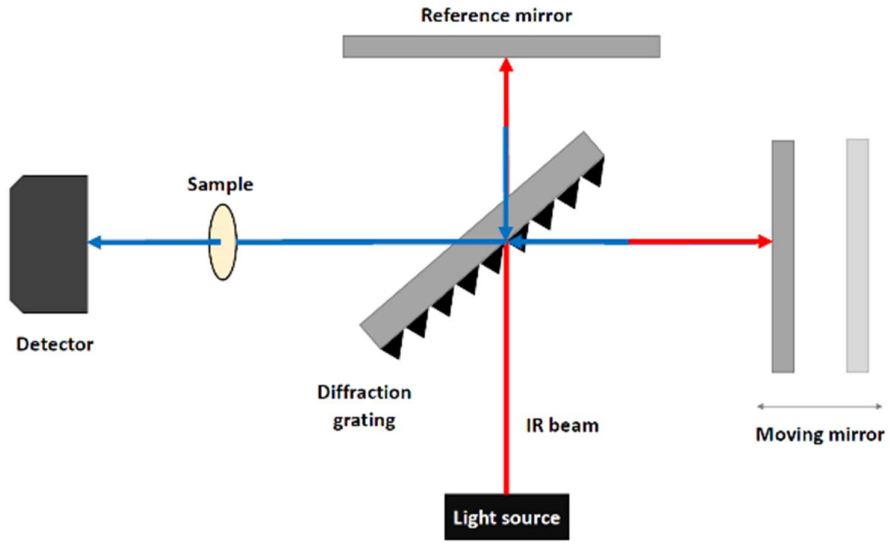

Figure 11: Schematic of a Michelson interferometer

A typical FTIR setup will produce spectra in either transmittance or absorbance. There are also other techniques that can be used, such as diffuse and specular reflection. However, the modalities that will mainly be used in this project are attenuated total reflection (ATR) and diffuse reflectance (DRIFTS). The main difference between transmission and ATR are shown in Figure 12. [72]

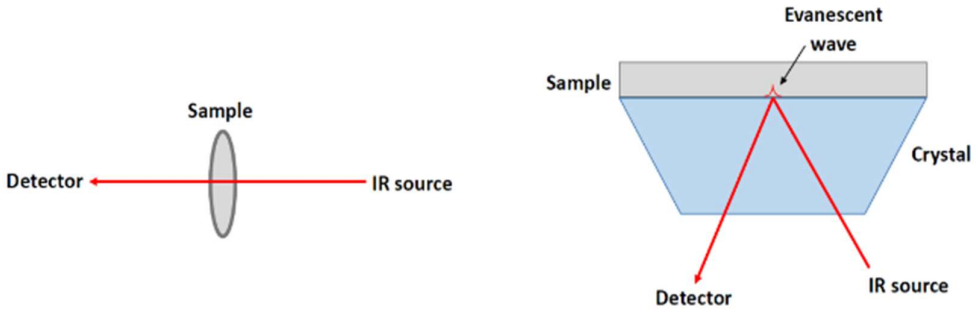

Figure 12: Adapted schematic showing the difference in techniques between transmission/absorbance and ATR.[23]

For transmission the IR beam passes through the sample, in ATR the IR beam is reflected internally and only penetrates the surface of the sample by typically $1^{-1}0$ µm, depending on the wavelength and incident angles. [73] For this reason, ATR calls for an intimate contact between the crystal and the sample. For powdered samples, it sometimes requires the use of a clamp to ensure there is sufficient contact. For solid samples that require a higher depth of penetration than can be achieved with ATR, diffuse reflectance can be a suitable alternative. Diffuse reflectance allows for bulk analysis of a sample, rather than simply surface analysis that is achieved with ATR, shown in Figure 13. [74]

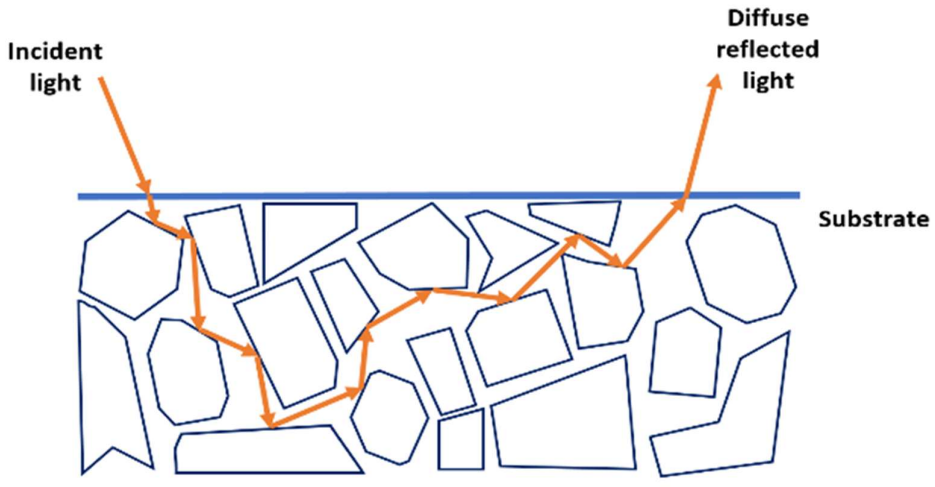

Figure 13: Schematic of diffuse reflectance spectroscopy when analysing a substrate.

1.7.5 Raman history

Raman spectroscopy was first discovered in 1928 by Sir Chandrasekhara Venkata Raman. [75] Elastic scattering, also referred to as Rayleigh scattering, involves the excitation of a photon that is promoted to a virtual state before dropping back to the ground state with no overall energy loss or gain. Stokes and anti-Stokes are examples of inelastic scattering that occur equidistantly from the Rayleigh line, with Stokes scattering found at a lower energy than the Rayleigh line and anti-Stokes found at a higher energy level. [66] Stokes and anti-Stokes are both inelastically scattered, with Stokes scattering producing more intense peaks as the electron is promoted from the ground state, at which most molecules naturally exist. Stokes is therefore the scattering often used for analysis. [75] Raman spectroscopy differs from infrared in that IR promotes electrons into a vibrational state, whereas Raman promotes it to virtual state before it falls back to either a vibrational or ground state, with the release of a photon. Figure 8 shows the electronic transitions that take place in Raman spectroscopy, representing the three different scattering types previously mentioned. [69]

Figure 14 shows a schematic of a standard Raman spectrometer fitted with a microscope and a 532 nm laser. [77] Instruments can be fitted with one or multiple lasers depending on the individual requirements of a research group. Different lasers are used depending on the samples that are being analysed, with lasers ranging from 244 nm in the UV-Visible region up to 1064 nm in the Near-infrared region.

Fluorescence is often one of the major considerations when it comes to which laser will be used. Visible lasers are often swamped by a fluorescent background, masking the already weak Raman peaks. UV-visible and near-infrared lasers such as 244, 785 and 1064 nm lasers are less affected by this. Raman instruments also have a number of other parameters that can be altered to change the spectra that are acquired. The grating that an instrument is fitted with will affect the resolution of the spectra. A 300 mm^{-1} grating will produce spectra with a lower resolution than those acquired using a 2400 mm^{-1} grating. Filters affect the amount of light reaching the surface of the sample with a range of different filters used in this project. Changing the amount of light that reaches the sample is important for a number of reasons. Firstly, certain types of sample can be degraded or burned by using a high laser power. Also, the quality of the spectra is affected as some samples produce small peaks that can also be masked when high laser powers are used. On the contrary certain sample types will require more intense radiation to produce molecular bond vibrations which are observed in the Raman spectra. With the Renishaw InVia systems the pinhole can be in or out which alters the laser beam from a line to a spot which can help with sample that are suffering from damage cause by too much laser power as the Raman laser is dispersed over a larger area. [78]

Discussed in the subsequent chapter, in section 3.2.3, no standard set of parameters could be used for spectral acquisition across all of the substrates chosen for use within this project. Therefore, all of the parameters discussed in the previous paragraph can be found in the figure legend for all of the spectra collected during this

project, these spectra are discussed within Chapters 3 and 4 as well as Appendix A and B.

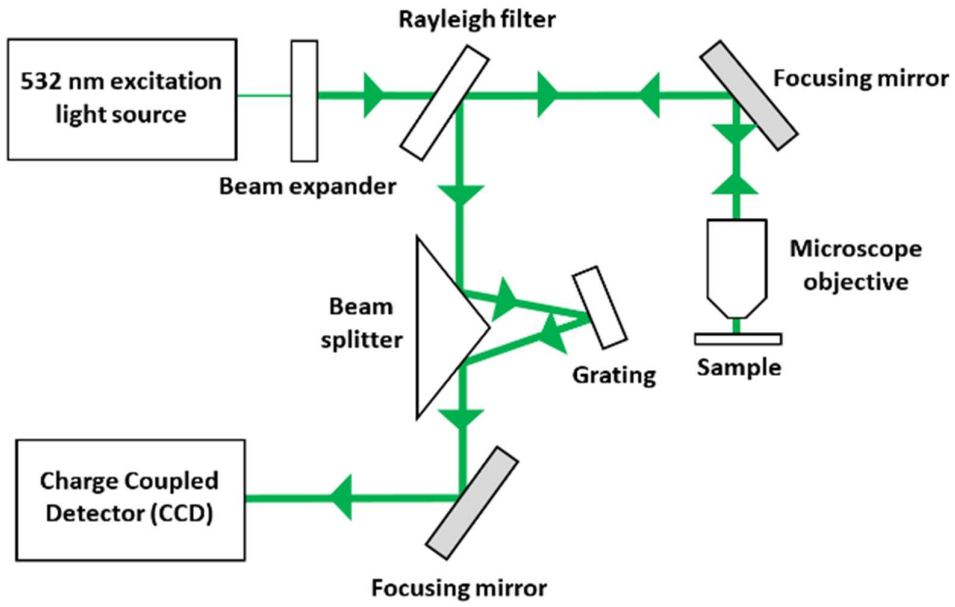

Figure 14: Schematic of a standard Raman spectrometer fitted with a microscope and a 532 nm laser.

1.8 Instrumentation

The following section discusses all of the instrumentation used throughout the course of this project including microbiology, IR and Raman instrumentation. The section has annotated photographs of the individual instruments and discusses the purpose of the instrument and, in the IR and Raman sections, the reasoning behind why that instrument was used for each particular study.

1.8.1 Microbiology instrumentation

Discussed within this section is the Bioscreen C instrumentation used for growth curve analysis to give an indication of the way that the bacterial strains used within this PhD project grow.

1.8.1.1 Bioscreen C instrumentation

For growth curve analysis Bioscreen C can be used for the automated analysis of up to 100 wells, for any desired length of time. Bioscreen C analyses bacterial growth rate in liquid media. While the instrument is automated, there are a number of variables that can be altered to allow experiments to be individually tailored. Some of these parameters include temperature, length of experiment, frequency of data collection and if the samples are to be shaken. Optical density (OD) measurements are collected from the plate and plotted against time. This allows standard growth in broth to be mimicked using small amount of sample to understand more about how a bacterium grows in a given environment. The instrumentation used for this is shown in Figure 15.

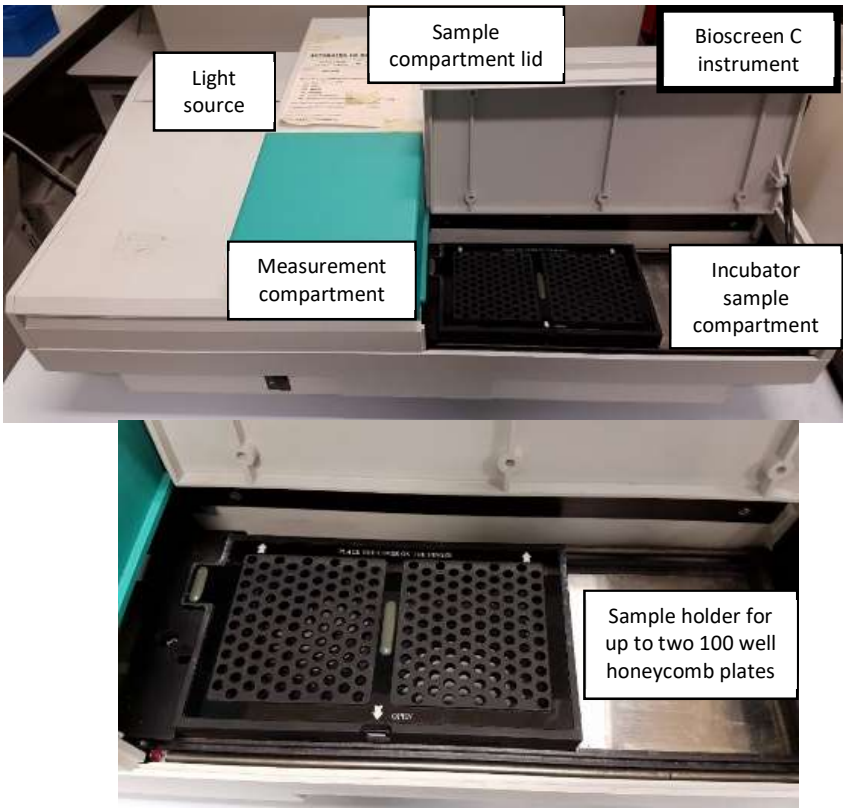

Figure 15: Photographs with labelling of the Bioscreen C instrumentation with a close up of the sample holder/ analysis area.

1.8.2 FTIR Instrumentation

Within this project four IR spectrometers, three based at the university of Strathclyde (UoS) and one at the university of Manchester (UoM), have been used for spectral acquisition. The systems include two benchtop FTIR spectrometers with an ATR accessory, one handheld FTIR using a diffuse reflectance attachment and one HT-IR system. It will be stated within the relevant research chapters which instrument has been used for the acquisition of the data.

1.8.2.1 Shimadzu 8400S benchtop spectrometer

The Shimadzu 8400S benchtop FTIR spectrometer, based at the University of Strathclyde, was initially used in this project to acquire data from the prechosen set of substrates. The Shimadzu is a research grade instrument capable of achieving a spectral resolution of 0.85 cm^{-1} with a peak-to-peak signal-to-noise (S/N) ratio of 20 000. [79]

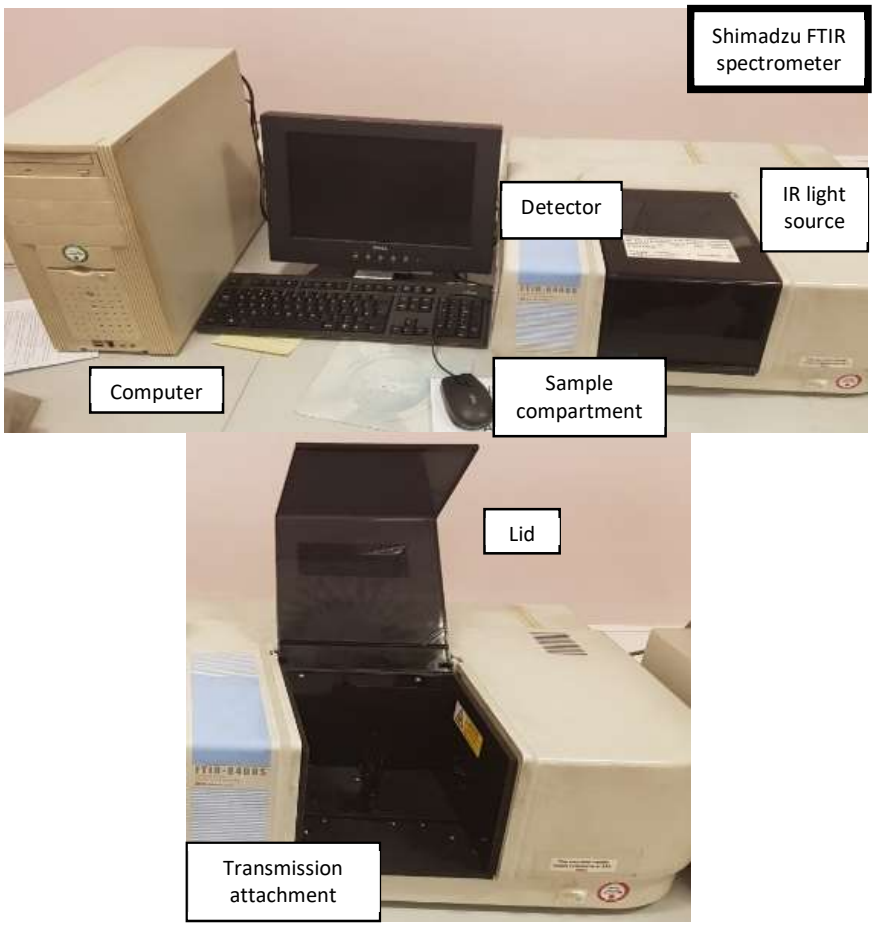

Figure 16: Photographs with labelling of the Shimadzu 8400S benchtop spectrometer with a close up of the sample analysis compartment setup for transmission analysis.

1.8.2.2 Agilent Cary 660 benchtop spectrometer

The Agilent Cary 660 benchtop FTIR spectrometer, based at the University of Strathclyde, has been used in this project to acquire the majority of the data in all studies. These studies include building a substrate library, developing a methodology to remove substrate contribution and bacterial identification. The Agilent is a research grade instrument capable of achieving a spectral resolution of 0.06 cm^{-1} with a peak-to-peak S/N ratio of 16 000. In terms of the S/N ratio the Agilent and Shimadzu are fairly comparable, however the Agilent has a significantly better spectral resolution. [80]

Figure 17: Photographs with labelling for the Agilent Cary 660 benchtop spectrometer and Pike ATR accessory

1.8.2.3 Bruker Equinox 55 with a HTS XT microplate reader attachment

A benchtop, research grade FTIR spectrometer, based at the university of Manchester, was used only within the environmental conditioning study as a method of rapid analysis to assess the effect that the environment has on the bacterial spectrum. A 96-well Silicon (Si) plate, that can hold 1-20 µl of sample, was used. [81] In comparison to the other benchtop instruments, the equinox has a spectral resolution of 0.2 cm^{-1}. [82]

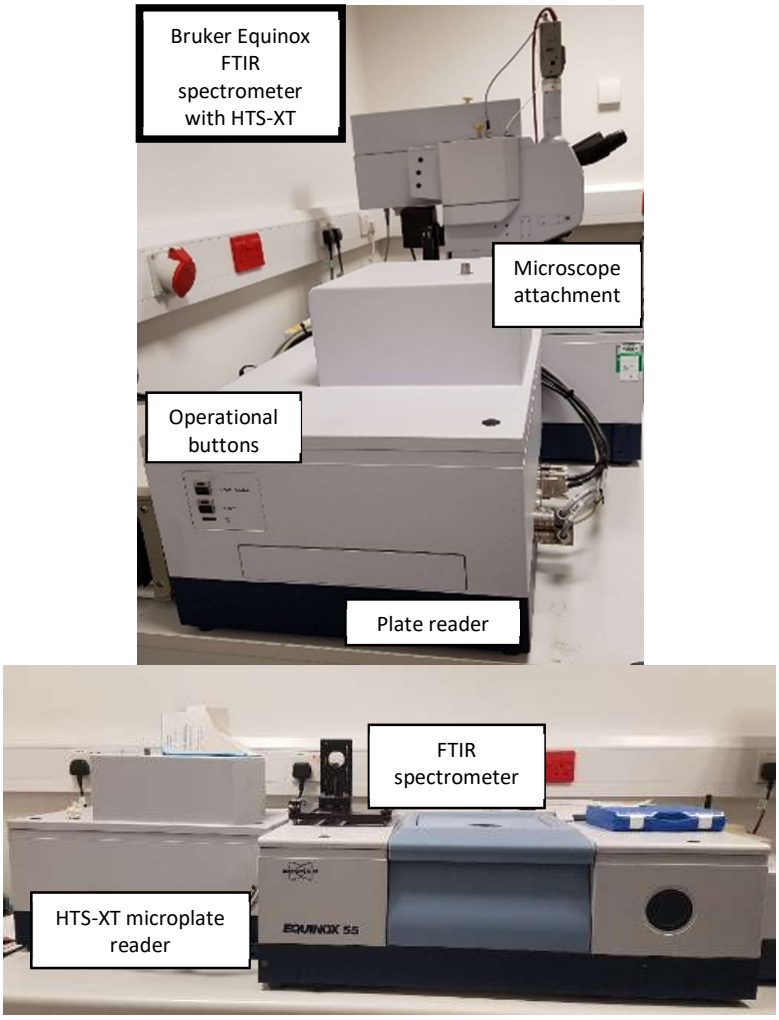

Figure 18: Photographs and labels for the Bruker Equinox 55 spectrometer and HTS-XT microplate reader attachment.

1.8.2.4 Agilent 4300 handheld spectrometer

A handheld FTIR spectrometer, based at the University of Strathclyde, was predominantly used with a diffuse reflectance attachment to acquire data for the environmental conditioning study. A diamond attenuated total reflection (Di ATR) attachment was used for comparing between the spectra produced with each accessory. For this reason, the diffuse reflectance attachment is shown in Figure 19 with a gold reflectance reference cap attached.

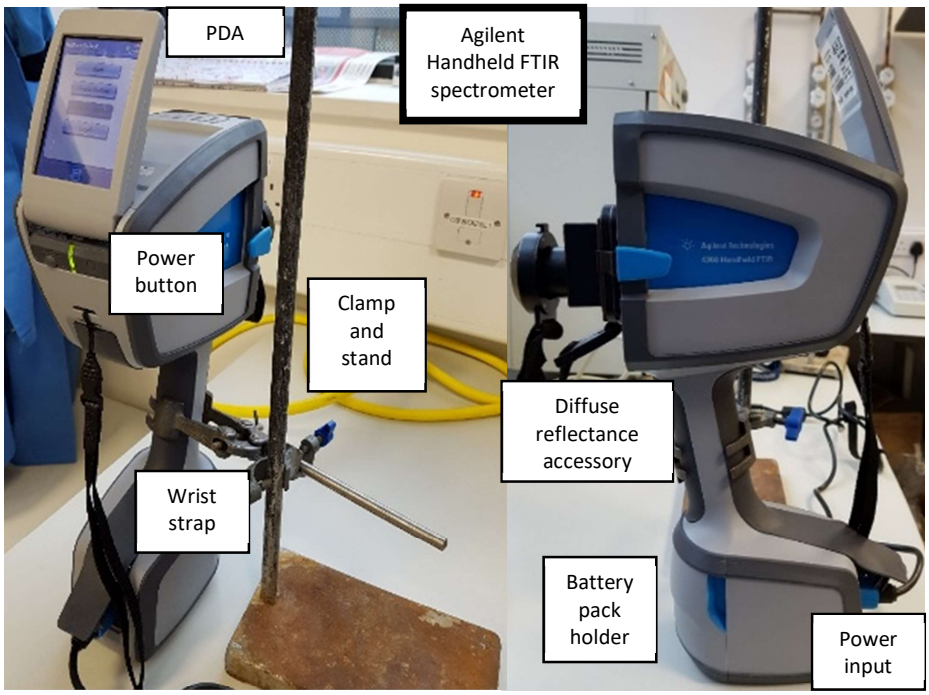

Figure 19: Photographs and labels for the Agilent 4300 handheld spectrometer

1.8.3 Raman instrumentation

Within this project three Raman spectrometers, one based at The University of Central Lancashire and two at The University of Manchester, have been used for spectral acquisition. The systems use a range of lasers including 532, 633 and 785 nm, with some systems fitted with multiple laser types. It will be defined within the following chapters the instrument and laser used for data collection.

1.8.3.1 Horiba LabRam spectrometer

The Horiba LabRam system, based at the University of Central Lancashire (UCLan) was fitted with two lasers - 532 nm and 785 nm - both of which were used for spectral acquisition of data from the substrates. For spectral acquisition of data from the bacterial samples, only the 785 nm laser was used as this was the only laser that could be used with all substrates explored within this project. This aspect of the project is discussed in section 3.2.3. Figure 20 below shows the external and internal set-up of the LabRam system indicating where the light source, laser sources, charge coupled device (CCD) detector, microscope and stage can be found. All data from this set up was acquired using a x50 long working distance (LWD) objective and therefore is the objective shown. The x50 LWD objective was used as this is the highest magnification optic available for use within this project.

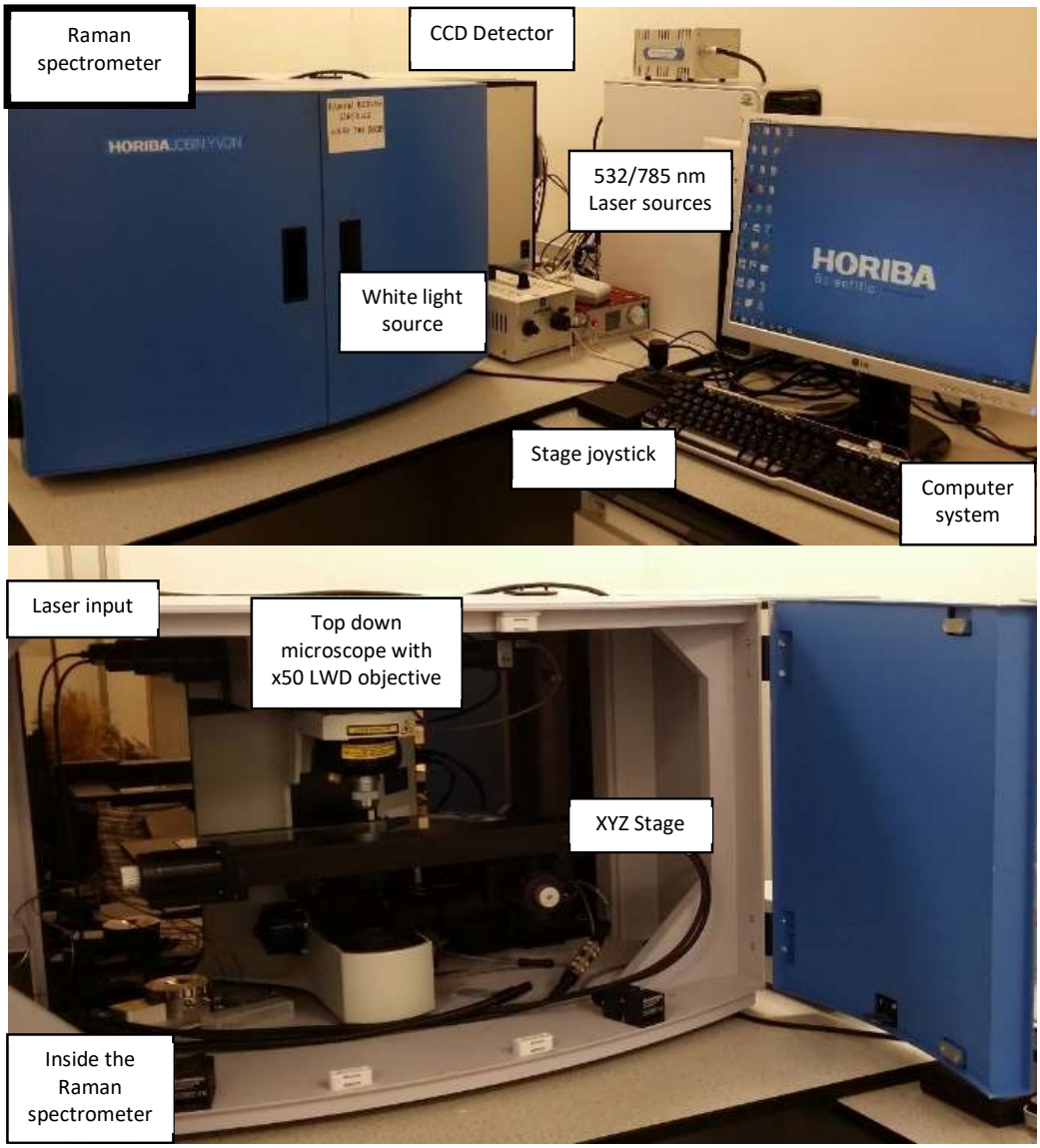

Figure 20: Photographs and labels for the Horiba LabRam Raman spectrometer

1.8.3.2 Renishaw InVia spectrometers

There were two Renishaw InVia systems used within this project, both based at the University of Manchester. One of the systems was fitted with both a 532 and a 633 nm laser, while the other system had a 785 nm laser. Figure 21 and Figure 22 show the outside and inside set up of the InVia systems, with the only difference between the two being the laser attachments. The figures indicate where the laser sources, microscope and stage can be found. All data from these set ups were acquired using a x50 LWD objective and therefore this is the objective shown in both of the figures. Figure 22, fitted with a 785 nm laser, shows a calcium fluoride (CaF_2) disk with a bacterial sample prepared on it.

Figure 21: Photographs with labels of the Renishaw InVia Raman microscope fitted with 532 and 633 nm lasers

Figure 22: Photographs with labels of the Renishaw InVia Raman microscope fitted with a 785 nm laser

1.9 Bacterial Identification using Vibrational Spectroscopy

Research published in Science in 1952 highlighted the power of IR spectroscopy for the successful discrimination between different species of bacteria, including the ability to distinguish between different strains of *Bacterium tularense.* [83] Since this point, research has expanded in this field, moving towards investigating a plethora of pathogens, some with specific interests, aiming to push the capabilities of new identification techniques towards strain level identification. [84,85] One key study that pushes this further looked at quantifying metabolism in single *E. coli* cells based on the organism's composition of carbon and nitrogen components using stable isotopic labelling. [86] Continuing research into bacterial identification has led to the creation of a number of new research fields, particularly metabolomics which looks to identify the key individual components that are responsible for differentiation, with IR and Raman being characterised as metabolic fingerprinting techniques. [87,88]

Pioneering research published by Dieter Naumann and his research group at the Robert Koch Institute, Berlin, spearheaded FTIR spectroscopy as an approach that could be applied to BWAs. This work was able to first typify a bacterial IR spectrum, and assign tentative molecular assignments to the spectrum; as illustrated in Figure 23. Naumann and colleagues concluded that a typical bacterial spectrum can be split into five main regions. [89] A typical FT-IR spectrum from bacteria (and indeed any biological samples) is described as having:

(A) The higher wavenumber region of 3000-2800 cm^{-1} which contains C-H stretches from CH_3 and CH_2 bonds found in fatty acids.

The fingerprint region of the spectrum from 1800-700 cm^{-1} is then split into 4 regions:

(B) $1800^{-1}500$ cm^{-1} is often heavily dominated by Amide I and II stretches but also covers C=O stretches and DNA/RNA related information

(C) The region of $1500^{-1}200$ cm^{-1} contains a mixture of features resulting from fatty acid components, protein and phosphates bands. A subsection of this region from $1500^{-1}400cm^{-1}$, looks again at fatty acids stretches including CH_2 and CH_3 bonds.

(D) The lower wavenumber region with vibrations at 1200-900cm^{-1} mainly contains information from carbohydrates found with the cell wall.

(E) The 900-700cm^{-1} region generally consists of many overlapping weak signals that are very characteristic of a sample with this region commonly being referred to as the "true fingerprint region".

The research also showed that FTIR, coupled with cluster analysis, was able to successfully classify bacterial strains from different genera, and also identify unknown clinical isolates (*Staphylococcus aureus* and *Streptococcus faecalis*) at similarity rates as high as 98.4% ± 0.9% when using FTIR. [89]

Following on from the work carried out by Helm *et al.* in 1991 it is now understood that an infrared spectrum from biological material can be characterised further into 7 key regions [84] : The region of 4000-3100 cm^{-1} represents broad OH/NH

stretches from proteins which are very strong in wet samples, as water will dominate this region of the spectrum. The other key region is around 1230 cm^{-1}, which focuses on the double bond asymmetric stretch of a phosphate band. This band become very important when considering the differences in Gram-positive and Gram-negative bacteria as large amounts of phosphate groups can be found in Gram-positive strains, as these contain large amounts of teichoic acids and lipoteichoic acids which are rich in phosphates.

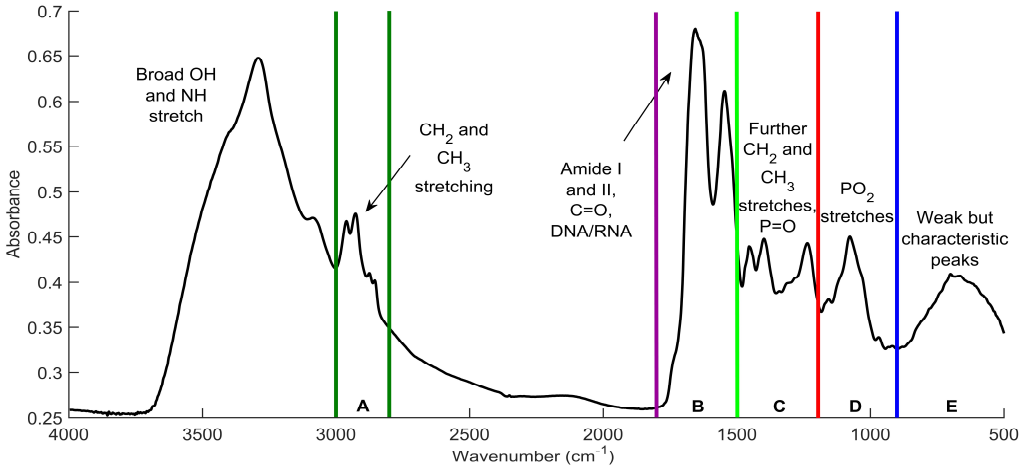

Figure 23 Figure showing a typical spectrum collected from a bacterium, split into the different regions as identified by Helm, et al. 1991: 69-79. [63] (A) has C-H stretches from CH_3 and CH_2 bonds, (B) has Amide I and II, C=O stretches and DNA/RNA information, (C) has fatty acid components, protein and phosphates bands, (D) has carbohydrates found within the cell wall, (E) is the "true fingerprint region"

The spectral regions in figure 23 are present in all bacteria and any differences are usually subtle quantitative levels of the various molecular vibrations, representing the different levels of nucleic acids, proteins, lipids and metabolites in cells. This means that simple visible inspection of spectra is not possible. FTIR in combination with multivariate analysis (MVA) is a powerful technique which is used

to analyse complex infrared spectra and has been employed to differentiate and identify bacteria. [90] Early studies by Naumann and colleagues [91] established that using second derivative data from only the $1425^{-1}485cm^{-1}$ range, attributed to CH, CH_2 and CH_3 vibrations, resulted in clear differentiation between Gram-positive and Gram-negative bacteria. These authors suggested that the grouping was due to differing structure and cell wall composition of Gram-negative bacteria causing spectral difference in the fatty acid region of the spectra. At the other end of the taxonomic scale Naumann and co-workers established that highly specific fingerprint-like patterns can be derived from IR spectra of bacteria, that in turn can be used to characterise the microorganism under investigation to subspecies level, and even down to strain level. [91]

In a later review written by Maquelin, it was also suggested that as IR and Raman are complementary techniques that the fingerprint produced by combining the data from the two types of vibrational spectroscopy would create a highly specific fingerprint that would further enhance the diagnostic capabilities of these analytical techniques. [92] Similar to Naumann's breakdown of a bacterial FTIR spectrum the review highlighted key regions of a Raman spectrum. It was shown that 2700-3100cm^{-1} and 1450cm^{-1} was characteristic of CH stretching and deformation while 1660cm^{-1} and 1250cm^{-1} correspond to amide I and amide III respectively. The Raman aspect of this study also showed well resolved peaks that correspond to DNA/RNA base ring vibrations, such as tryptophan, tyrosine and phenylalanine, that are not as commonly seen in FTIR data. FTIR spectra are often heavily dominated by broad amide peaks, masking smaller bands.

Work by Timmins *et al.* [93] used both pyrolysis mass spectrometry (PyMS; for a review see "Pyrolysis mass spectrometry and its applications in biotechnology") [94] and FTIR spectroscopy to differentiate multiple strains from 3 different species of *Candida* (*C. albicans, C. dubinienis* and *C. stellatoidea*). Both PyMS and FTIR are rapid techniques (~2 min and ~10 s respectively) and analysis of these 29 clinical isolates of *Candida* resulted in classification of these yeast into three clusters based on the three different species.

As discussed in the introduction to this review, and also highlighted in a review by Clemens *et al.*, there is an essential need for a new technique for BWA identification. We believe that vibrational spectroscopy possesses the attributes of an ideal analytical technique for BWA identification including, rapid acquisition times, robustness and portability, as well as the ability to identify samples found on complex matrices. [15] The effect of environmental conditioning on bacteria has also been investigated using FTIR spectroscopy on BWA surrogates, as well as environmental bacteria in a variety of environmental conditions. [95] Temperature and humidity were cycled from 30-44°C with a change in humidity from 40^{-1}4% to recreate conditions of a hot dry climate, derived from MoD standards. MVA of the data was able to distinguish six different bacterial species with sensitivities and specificities as high as 88% and 97.6%, respectively, when using support vector machines (SVM). Using the same processing method on the environmentally conditioned data, sensitivities and specificities as high as 100% were achieved for three of these bacterial strains. Principal component analysis linked to linear discriminant analysis (PC-LDA) of the environmentally conditioned dataset shows visually that the conditioned bacteria

were projected next to the pre-conditioned data for each strain, as well as showing good separation between the three bacterial strains. High sensitivities and specificities are key when developing an alternative technique, more so when developing an alternative for BWA identification as (i) false negatives could led to the spread of harmful bacteria, potentially infecting others and cause an outbreak of disease, whilst (ii) false positives could led to inappropriate precautionary action being taken.

The work above has illustrated how FTIR and Raman, along with other physicochemical techniques that are used for whole organism fingerprinting (e.g. PyMS), can be used to classify and identify bacteria and fungi with high sensitivity and specificity. PyMS has been used by Goodacre *et al.* to detect *Bacillus* spores from these spore-forming bacteria where the marker was a pyridine ketonium ion which was a pyrolysis product from dipicolinic acid (DPA; pyridine-2,6-dicarboxylic acid). [85] The study found characteristic markers using both PyMS and diffuse reflectance absorbance FTIR (two vibrational modes from the constrained pyridine ring vibration between $1447^{-1}439$ cm^{-1} were found to be discriminatory) that can be used to discriminate between spores and vegetative cells based on the presence or absence of DPA; a compound that is unique to the spore coat of *Bacillus* and *Clostridium* species. Combining information from different techniques provides a better overall profile of the sample, which can help when trying to differentiate between closely related strains.

1.9.1 The Detection of Bacterial Spores using Spectroscopy

The ability for BWAs to form spores has resulted in a large proportion of research using spectroscopy to identify BWA that are based on spore forming bacteria; the most obvious one being *B. anthracis* as the causal agent of anthrax, although other members of this genus including *B. cereus* and *B. subtilis* are common food-poisoning agents and so could be used to deliberately contaminate food. Research by Foster *et al.* has shown that identification of spores is possible using FTIR when combined with exploratory analysis achieving a 100% success rate of identifying their physiological state (vegetative or sporulated) with a 67% rate for the correct identification of the 9 different bacterial strain. [96] Other research groups have also examined the identification of spore forming bacteria with an extensive study carried out using IR on various strains of *Bacillus.* [97-99] In one of these studies, differentiation of bacteria based on vegetative or sporulated was investigated and was able to show that there are differences in the spectra produced from sporulated bacteria between different spore production batches despite experimental consistency. [97] Alterations were particularly evident between 1200-900 cm^{-1}, with significantly less batch-to-batch variation observed in vegetative cells. The study also found clear peak differences between spores and vegetative cells with a quartet of peaks in the region of 700 cm^{-1} found exclusively in spores and not in vegetative cells, thought to be associated with calcium dipicolinate trihydrate ($CaDP{\cdot}3H_2O$). The spectra produced from sporulated bacteria also exhibited minor shoulders and peaks at 1232, 1307 and ~1735 cm^{-1} that the authors attributed to a very small

concentration of vegetative cells/debris. The main additional peak found in vegetative cells that is often greatly reduced in intensity or completely lacking in spores is the peak at around 1739 cm^{-1} of the ester C=O stretch from lipids, which again are lacking in the spore coat compared to the cell wall of vegetative *Bacillus* species. [87]

Further research using infrared spectroscopy identified the presence of $CaDP{\cdot}3H_2O$ in the spore form of the bacteria. [98] The work reaffirmed that the peak around 1739 cm^{-1} is found in both spores and vegetative samples. The peak is much smaller on spectra collected from spores and is often only present as a result of very low concentrations of vegetative cells being found in the samples. Further analysis of the quartet of peaks, thought to be as a result of $CaDP{\cdot}3H_2O$, located at 659, 701, 725 and 766 cm^{-1} showed the specific vibrational modes responsible for them. 701 and 766 cm^{-1} were both found to relate to the CaOH bend and water CaO torsions. [88] The peaks at 725 cm^{-1} was linked with OCO bend and a ring CCC bend while 659 cm^{-1} is linked with out of plane ring torsions and C-C torsions. [88] It was also shown that strain level discrimination of *Bacillus* endospores using chemometrics was possible using principal component analysis (PCA). [99] Using PCA, it was possible to discriminate down to the strain level of *Bacillus* spores with an 85% success rate. This research not only showed the discriminatory power of IR spectroscopy for bacterial spores, but also identified the peaks that correspond to calcium dipicolinate trihydrate found in sporulated bacteria.

Raman spectroscopy has also been used to identify and discriminate various bacterial spores, identifying the differences between vegetative and sporulated

bacteria. [100] The characteristics peaks used for differentiation were found at 661, 822, 1018, 1396, 1448 and 1575 cm^{-1} and like for IR can also be attributed to the presence of CaDPA in sporulated bacteria. Also, the work was able to show how the peaks in a Raman spectrum changed depending on the length of time the bacteria were left to sporulated, especially as DPA is released upon the transformation from spore to vegetative biomass. Peaks associated with CaDPA, two phenylalanine peaks at 622 and 1004 cm^{-1} and three tyrosine peaks at 643, 827 and 853cm^{-1}, can be used to monitor incubation time and were found to change during sporulation. [100] This study also investigated the effect that the conditions used to culture the samples could have on the spectra produced and classification of bacterial strains. The results show that sample culture methods do have an effect on the spectra. This is perhaps not surprising as Raman and indeed IR are phenotypic characterisation methods and so measure the result of the response of the bacterial genome to its changing environment.

1.9.1.1 Enhancing Raman Scattering for Low-Level Detection of the DPA Spore Biomarker

Prior to normal Raman scattering on bacteria Nelson and Sperry in the late 1980s developed deep Ultraviolet Resonance Raman (UVRR) spectroscopy for bacterial characterisation. These authors published a series of papers that illustrated that UVRR can be used to gain information about Gram type, [101] as well establishing

that using excitation at 242 nm reproducible UVRR spectra could be obtained that can be used to identify bacteria. [102] UVRR spectroscopy has also been used to characterise endospore forming *Bacillus* and *Brevibacillus.* [103] Studies showed that UVRR spectroscopy was a useful tool for characterising bacteria, with a number of benefits over conventional Raman spectroscopy. This included less interference from background fluorescence which in turn increases the sensitivity of the spectra and leads to fewer stages of pre-processing, as well as a resonance enhancement (*ca.* 10^2) from aromatic molecules found in bacteria.

Surface enhanced Raman scattering (SERS) is a further approach that can be employed to aid in identifying bacterial species (for a review see "Characterisation and identification of bacteria using SERS" [104]). By exploiting the enhancement effect of metallic nanoparticles with nanoscale roughness, a significant increase in the scattering intensity can be observed, which allows for more rapid detection at much lower bacterial levels that traditional spontaneous Raman spectroscopy. [105] It has been shown that the addition of a silver colloid solutions to pure DPA allowed unequivocal detection of this spore biomarker at a concentration as low as 1 ppm. [106] Further work showed that a combination of SERS and MVA based on discriminant analysis was able to identify DPA directly in *Bacillus* spores. [107] As this method did not selectively extract DPA from spore the uniformity of the cell wall in Gram-positive bacteria was seen to be a major contributing factor to spectral reproducibility; however, overall the research presented a technique that could be used in the next generation of BWA analysis. An alternative approach presented by the same group developed SERS for identifying DPA from *Bacillus* spores: this

involved using a portable Raman spectrometer, nitric acid to extract DPA from spores, and when this extract was coupled with silver colloids the acidic conditions also enabled suitable aggregating which sped up sampling and enabled results to be obtained in seconds rather than minutes. [108] These authors reported excellent reproducibility and a very low limit of detection (LoD) down to 5 ppb for DPA which at the time was significantly lower than the limits obtained from previous studies using SERS and was equivalent to *ca*. 1,000 spores which is below the infective dose for anthrax. In a very recent collaboration between Bell's group and Goodacre's group, SERS on meso-droplets supported on superhydrophobic wires was shown to allow for exquisitely sensitive detection of DPA down to levels of 10^{-6} mol dm^{-3} which is equivalent to just 18 spores and 2 orders of magnitude better than any previous measurements for detecting this spore biomarker. [109]

As an alternative to SERS another area of promising research has looked at the use of coherent anti-stokes Raman scattering (CARS). CARS is a non-linear Raman method than can be used to tune into and enhance specific vibrational frequencies. CARS is at least two orders of magnitude more efficient than standard Raman spectroscopy and can lead to 100 times faster sensing and detection of BWA ideal for rapid *in situ* analysis. [110] Bacterial spores can be detected using a single femtosecond laser shot and it is possible to discriminate 10^4 spores showing the sensitivity of the technique. [111]

It is evident that enhanced spectroscopic techniques are versatile approaches to bacterial identification and discrimination with high levels of sensitivity and

specificity, and it is perhaps not surprising that this field has concentrated on the detection of bacterial spores. Although some enhanced spectroscopic techniques, such as SERS, are portable, like conventional spectroscopic techniques they often require additional consumables (metal nanostructures) and sample preparation prior to analysis. More conventional spectroscopic methods do not require preparation of samples so direct analysis can be carried out allowing true *in situ* analysis, which could provide the ideal technique for BWA classification and identification, but these may lack the signal boost that SERS, CARS and UVRR offer.

Dipicolonic acid (DPA) is a key characteristic marker of spores from both *Bacillus* and *Clostridium* species. It is unique to bacterial spore and plays an important role in the controlling water content of the spore coat and thereby helps the spore survive harsh environments. [112] However, as this is a general mechanism DPA is not species specific and so is found in all bacterial spores and not just *B. anthracis*. Therefore, developing a technique that relies solely on DPA for spore identification for BWA detection would be very problematic as there would be many false positives from environmental spore forming bacteria. However, within a military context any indication that a cloud contains large levels of DPA is a good indication of potentially threating actions.

1.10 Data processing

Once the data has been collected it has to undergo a number of data processing steps. Within the biospectroscopy community there is no set

methodology for data processing. The protocol used will heavily depend on the types of samples that are to be analysed, the technique used to collect the data and will also depend on the laboratory conducting the analysis. The steps used to process the data, the order in which they are performed, the importance of sharing this information is an area often discussed within the field of biophotonics. [113,114,115,116] However, the importance of developing a universal method for data processing, particularly a set order for pre-processing, has been acknowledged. This is starting to be addressed as some research groups, such as Prof. Roy Goodacre's, have made their codes used for processing in the cluster-toolbox-v2.0 which is freely available and can be found at: https://github.com/biospec. The protocols chosen for use in this project were dependant on the data being analysed. All processing of the data collected during this study was processed using MATLAB unless otherwise stated as there are instances where instrument software has been used rather than MATLAB for ease and consistency of analysis. The codes used for all data analysis performed as part of this project were a combination of in-house written code, the IRootLab toolbox and the cluster 2 toolbox created by Prof. Roy Goodacre's research group.

1.10.1 Pre-processing

Before any form of multivariate analysis can be performed it is necessary for the data to undergo a number of quality control measures due to the spectral variability caused by variation in sample thickness, atmospheric changes in water and CO_2 and instrumental variation that causes changes in the baseline. Between every

stage of analysis, and prior to any form of processing or analysis, a visual inspection is performed to ensure that there are no obvious outliers that are vastly different from the bulk of the data. This ensures that there are no spectra built into a discriminatory model that have obvious spectral deformations that are not due to biochemical changes. Visual inspection is not used to remove spectra that show slight changes that are due to biochemical changes, as these spectra are important for developing a model with a high level of robustness. The following section discusses the processes used with the data collected during the course of this project, including noise reduction, vector normalisation and extended multiplicative signal correction, before moving on to discuss the multivariate analysis methods that have been used on the processed data.

1.10.1.1 Noise Reduction (NR)

Noise reduction (NR) is key for removing noise that is present in the spectra to ensure that this does not contribute to the separation of the data when it comes to the multivariate analysis stage. Principal component analysis (PCA)-based noise reduction is the style of NR used within this study. PCA-based noise reduction is an orthogonal linear transformation that creates a principal component (PC) for each of the data points. [113] These data points are then organised based on the most variance explained down to the PCs that explain the least, which will generally contain only noise and no biochemical information. When the information is reconstructed, the PCs containing noise are not included, thus removing noise from each individual

spectrum. Caution must be used when selecting the number of PCs to use for the smoothing process, as the greater the number of PCs removed the higher the amount of smoothing seen in the data. Therefore, using too many PCs for smoothing may see small but significant peaks lost due to over-smoothing of the data. [117]

1.10.1.2 Vector Normalisation (VN)

Vector normalisation (VN) is a very important part of spectral processing, especially when working with samples that may vary in thickness, such as dried samples and tissue biopsies. [71] This is because variations in thickness can lead to a variation in peak intensities between similar spectra. The reason VN is used is to reduce the amount of variation between the different spectra. The principle behind VN creates an average absorbance by squaring the intensities of every data point, adding them together before performing the square root of the value. This value becomes known as the vector length and is equal to 1. Each spectrum is then scaled using the vector length. [64]

1.10.1.3 Extended Multiplicative Signal Correction (EMSC)

Extended Multiplicative Signal Correction (EMSC) is another type of normalisation that can be used on spectral data and is commonly used with biochemical datasets. This method of normalisation uses polynomial smoothing and

is often the favoured method used to process datasets that have a non-linear background. [118] As well as issues with the baseline, the method can also be used for processing interference and scaling issues. [119] Therefore, EMSC was used as a processing method for the data collected using the microplate reader to overcome the variation in baseline caused by differing path lengths. [120] VN is a filtering method which removes unwanted noise and other variations from the spectrum while EMSC is a model-based method where the information is filtered but not lost. [119]

1.10.1.4 Derivative

The bands observed in a typical FTIR spectrum will have a series of broad bands rather than sharp, defined peaks as a number of vibrations are represented within the same region of the spectrum. In an attempt to overcome this problem, derivatisation is commonly used to deconvolute the broad bands and produce a number of individual peaks. [113] For converting to the derivative to be successful, a high S/N ratio is required. Spectra with a low S/N ratio will contain higher levels of noise which, when derivatised, will be enhanced along with the rest of the spectrum. Performing a first derivative shows the gradient of the spectrum while a second derivative shows the change in gradient. As a result too much noise will mask important biological information as noise has sharper peaks and more sudden changes than that of a biological peak. For this reason, derivative analysis is more likely to be used with IR spectra than with Raman spectra, as Raman by nature is a much noisier technique. When it comes to interpreting the spectra produced after

derivatisation, first derivate spectra find the peak maximum now on the baseline, which makes interpreting them more difficult and therefore, second derivative is generally the more favoured method to use with IR spectra. [64]

1.10.2 Multivariate Analysis (MVA)

Once the data has been processed using one or more of the methods discussed above, the data is ready to be interrogated using multivariate analysis (MVA). MVA can be classed as either supervised of unsupervised. Unsupervised analysis simply looks for natural variations within the data and splits it into groups based upon this information. Supervised analysis requires the number of groups to be known, with the data generally being assigned to one of these groups. As with pre-processing, there is a vast array of analysis methods, therefore the only ones that are discussed below are the ones that have been used within this project.

1.10.2.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised form of MVA, meaning that no information or identifying knowledge about the sample is provided prior to the analysis being performed.

When spectra are collected there is a large amount of data points per spectrum. PCA has four main goals: to extract the most important information from

the data, reduce the dimensionality of the dataset while retaining all important information, simplify the dataset and finally produce visual outputs that relate to the differences in the data. [121] A simple way to understand this principle is by comparing a group of people, who are a collection of three-dimensional (3-D) data points, to a photograph of them, which is two-dimensional (2-D). A photograph retains all of the key information about the people such as variances in gender, hair colour and eye colour, while reducing and simplifying the amount of information contained in that dataset. Scores plots and loadings plot would also be available alongside this as a visual descriptive of the data. The scores plot shows how each data point differs from the others data points while the loadings plot shows what is responsible for the separation. Figure 24 below shows a photograph that has a group of people in it, between whom there are similarities and differences, with a mocked up PCA plot.

The most discriminating PC in this case would be gender and would therefore be explained by PC1. The second most discriminating factor may then be hair colour or eye colour and would be explained in PC2. As the PC number increases the significance of the information gained decreases. This leads to irrelevant background information to be brought into the model. Thinking back in spectral terms, this would be the likes of noise and spectral artefacts. For this reason, it is important to restrict the number of PCs used when building a model to ensure the model is unaffected by unwanted background information.

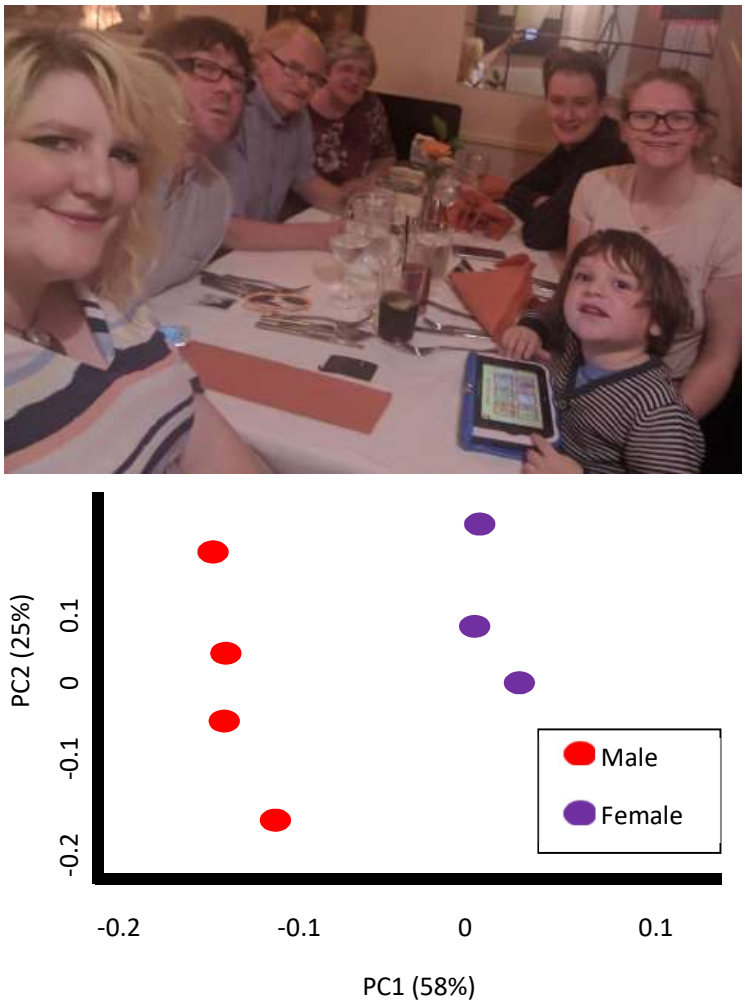

Figure 24: Photograph representing a 3-D system (top) with a mocked up PCA plot showing example separation when using PC1 and PC2 considering the data produced from the

PCA looks for the natural variance in a set of data and produces a number of outputs. The outputs produced from PCA include three variables: scores, loadings and percentage explained values. All three variables can be plotted to provide visual information; plotting of both the scores and loadings variables have been discussed previously. Scores plots describe the contribution of the chosen PCs to each of the individual spectra within a given dataset. The first PC represents the largest amount

of variance with the second PC being orthogonal to the first. Loadings represent the correlation between each individual datapoint across a wavenumber range and the chosen principal component. [121]

Plotting the percentage explained variable, commonly referred to as a scree plot, allows you to visualise the relation between the PC and the percentage variance explained. This is commonly used as a guide to judge how many PCs are required to explain 95% of the total variance of your dataset. The number of PCs needed to explain 95% variance is heavily dependent upon the dataset however, with Raman datasets usually requiring more PCs as there is generally a higher level of noise within the data. When working with biological datasets, scores plots and loadings plot are usually considered together, as the loadings are used to explain the biochemical changes behind the differences observed in the scores plot.

1.10.2.2 Multiblock Principle Component Analysis (Multiblock PCA)

Due to the complexity of some datasets, standard PCA alone is not sufficient to fully interrogate the data and highlight the underlying variance within the dataset. There are three main types of multiblock PCA: hierarchical PCA (HPCA), consensus PCA (CPCA) and generalized PCA (GPCA). [122] Within this research project both HPCA and CPCA were used to compare which showed the best split of the data. Within all forms of multiblock PCA a number of different variables are generated in addition to those produced with PCA. The data is first split into blocks which are used to generate

both block loadings and blocks scores. The block scores are then combined to create a new super block. PCA is performed on the super block to produce a super weight, which is normalised to one, as well as a new super score. This process is repeated until there is convergence of the super score. HPCA works in the same way as CPCA, however, HPCA differs in the normalisation step of the analysis where this time it is the super score that is normalised to one rather than the super weight. [123] Multiblock PCA allows you to group data into different blocks and look for the variances between these new groups. By combining data from different groups into new blocks, some variance can be removed to allow for subtler differences to be detected.

To give an example of this, consider a dataset comprised of spectra collected from five bacterial strains at four different time points. By creating blocks for each time point, combining all of the bacterial strains, the separation caused by the difference in bacterial strain is removed and allows for the effect of aging to be shown. As part of this PhD project, an environmental conditioning study will be carried out that will look at both of these factors and will assess how much of an impact aging has on the bacterial spectrum.

1.10.2.3 Discriminant Function Analysis (DFA)

Discriminant Function Analysis (DFA) is classed as a supervised technique as, unlike with PCA, some sample knowledge is used in the analysis. Group classifications

are provided to allow the algorithm to know how many groups need to be classified, making the clustering of groups more defined in comparison to PCA data. DFA uses the optimal hyperplane to explain the variation between different data points. These differences are what become discriminant functions (DFs). [124] The algorithms used for the PCA and DFA analysis were from the cluster 2 toolbox mentioned in section 0. Once significance between the DFs is identified, classification will follow and the DFA will use the prior knowledge of the number of groups to classify each of the data points. When carrying out DFA, there are a number of assumptions that are made about the dataset. These include assuming a normal distribution, that there are no obvious outliers and that the number of independent groups is at most n-2, where n is sample size. The maximum number of DFs that can be used for analysis is X^{-1}, where X is the number of groups as this is the number of degrees of freedom. [125] Importantly, the number of samples per group can be unbalanced without affecting the analysis, while other types of analysis such as Multiblock PCA requires equal sized groups. As with PCA, DFA produces scores plots and loadings plots as visual outputs to explain the variance observed within the dataset.

PC-DFA allows for data to be split into testing sets and training sets, whereby the training set is used as the model and the test set is blindly projected into it. This allows for unknown spectra to be classified. Error ellipses can be plotted around the data and used for calculating sensitivity and specificity values associated with the different groups. Sensitivity and specificity values are calculated using

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positivies+ False negatives}} \quad (2.8)$$

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives+ False positives}} \quad (2.9)$$

Where:

True positives are the number of samples correctly identified as their species.

True negatives are the number of samples correctly identified as being not their species.

False positives are the number of samples that the model has incorrectly identified as being a species which they are not.

False negatives are the number of samples that the model has incorrectly identified as not being a member of their species. [64]

1.11 Conclusions and Future Work

There are a number of key qualities a technique used for BWA identification should have including: portability to allow for *in situ* use, rapid data acquisition, high sensitivities and specificities, a low limit of detection of the pathogenic organism as well as the ability to detect and identify bacteria found on complex matrices.

Research into alternative techniques for identification and classification of bacteria has shown that the process can be improved by using MS techniques and traditional microscopy techniques. However, one of the downfalls found in these techniques is that currently the instrumentation required is not portable, limiting the application of the technique. Vibrational spectroscopy has the potential to be an

all-round more powerful technique due its ability to make the technique portable, as well as additional attributes such as speed of acquisition. The technique is able to identify Gram type and can differentiate different types of bacteria down to the specific strain.

Before any alternative techniques can be considered as a replacement for current techniques, or a complementary tool to enhance the process of bacterial identification, there is a need for additional research. Specifically, bacterial identification using both real world samples recovered from warfare situations, or from laboratory cultures but in non-ideal conditions, needs to be explored. Using non-ideal growth conditions would provide the opportunity to assess the effect of environmental stress on bacteria, and the subsequent effect this has upon the spectral fingerprint of the sample. This will more closely mimic the physical state of the bacteria that are found in BW settings.

1.12 Acknowledgements

The authors would like to acknowledge Elaine Perkins, William Sellors, and Simon Smith. We also thank the DSTL National PhD Scheme for funding this research and Claire L Pickering.

1.13 References

1 Martin, J. W., Christopher, G. W. & Eitzen, E. M. (2007).

2 Barras, V. & Greub, G. History of biological warfare and bioterrorism. *Clinical microbiology and infection* **20**, 497-502 (2014).

3 Klietmann, W. F. & Ruoff, K. L. Bioterrorism: implications for the clinical microbiologist. *Clinical microbiology reviews* **14**, 364-381 (2001).

4 Archy, W. *The biological weapons convention (BWC) at a glance*, <http://www.armscontrol.org/factsheet/bwc> (2012).

5 Centres for disease control and prevention. *A history of anthrax*, <http://www.cdc.gov/anthrax/resources/history/index.html> (2015).

6 Shea, D. A. & Gottron, F. *Ricin: Technical background and potential role in terrorism*, <http://fas.org/sgp/crs/terror/RS21383.pdf> (2013).

7 Forest, J. J. F. & Howard, R. D. *Weapons of mass destruction and terrorism*, <http://www.smi-online.co.uk/_Media/docs/Chemical_Biological_Threats_Against_Public_Water_Systems.pdf> (2012).

8 BBC news. *Bangladesh anthrax cases spread*, <http://www.bbc.co.uk/news/world-south-asia-11194760> (2010).

9 BBC news, investigation, F. a. & health, C.-S. F. d. o. p. *US anthrax scare widens to 51 labs in 17 states*, 2015).

10 Ramsay, C. N. *et al. An Outbreak of Anthrax among Heroin Users in Scotland, December 2009 to December 2010*, <http://www.documents.hps.scot.nhs.uk/posters/2011/anthrax-heroin-scotland-2010.pdf> (2011).

11 Harris, R. & Paxman, J. *A higher form of killing*. 4th edn, 53-67 (Arrow Books, 2010).

12 Irenge, L. M. & Gala, J. L. Rapid detection methods for *Bacillus anthracis* in environmental samples: a review. *Applied microbiology biotechnology* **93**, 1411-1422 (2012).

13 Ivnitski, D. *et al.* Nucleic acid approaches for detection and identification of biological warfare and infectious disease agents. *BioTechnique* **35**, 862-869 (2003).

14 Bala, C. in *Portable Chemical Sensors, Weapons Against Bioterrorism* (ed Dimitrios. P. Nikolelis) Ch. 2, 344 (Springer, 2012).

15 Clemens, G. N., Hands, J. R., Dorling, K. & Baker, M. J. Vibrational Spectroscopic Methods for Cytology and Cellular Research. *Analyst* **139**, 4411-4444 (2014).

16 Government office for science. Blackett review on wide-area biological detection. 1-33 (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/279345/14-590-blackett-review-biological-detection.pdf 2014).

17 Walt, D. R. & Franz, D. R. Biological warfare. *Analytical chemistry* **72**, 738 A - 746 A (2000).

18 HM Government. National Security Strategy and Strategic Defence and Security Review 2015 A Secure and Prosperous United Kingdom. (2015).

19 Mauroni, A. J. *Chemical and Biological Warfare: A Reference Handbook*. 2nd edn, (ABC-CLIO, Inc., 2007).

20 Müller , R. D. *Germany and the Second World War*. Vol. V: Organisation and Mobilization of the German Sphere of Power 783 (Oxford University Press, 2003).

21 Hammond, P. & Carter, G. *From Biological Warfare to Healthcare: Porton Down 1940-2000*. 7 (PAGRAVE MACMILLIAN, 2001).
22 Balmer, B. *Britain and Biological Warefare: Expert Advice and Science Policy, 1930-65*. 45 (PALGRAVE, 2001).
23 Money, N. P. *Microbiology A Very Short Introduction*. 1st edn, (Oxford University Press, 2014).
24 Manchee, R. J., Broster, M. G., Melling, J., Henstridge, R. M. & Stagg, A. J. *Bacillus anthracis* on Gruinard island. *Nature* **294**, 254-255 (1981).
25 Johnston, W. R. *Review of fall 2001 anthrax bioattacks*, <http://www.cdc.gov/niosh/nas/rdrp/appendices/chapter6/a6-45.pdf> (2005).
26 Ford, M. in *Medical Microbiology* (ed M. Ford) Ch. Introduction to Microbiology, 1-11 (Oxford University Press, 2014).
27 Craig, N. L. *et al.* in *Molecular Biology: Principles of Genome Function* Ch. 1:Genomes and the flow of biological information, 1-29 (Oxford University Press, 2014).
28 Levin, H. L. in *The Earth Through Time* (ed Ryan Flahive) Ch. 8: Earliest Earth: 2,100,000,000 Years of the Archean Eon, 208-242 (John Wiley & Sons, 2006).
29 Naumann, D. FT-Infrared and FT-Raman spectroscopy in biomedical research. *Applied Spectroscopy Reviews* **36**, 239-298, doi:https://doi.org/10.1081/ASR-100106157 (2001).
30 Engelkirk, P. G., Duben-Engelkirk, J. L. & Wilson, G. R. in *Burton's Microbiology for the Health Sciences* (ed D.B. Troy) Ch. 4: Microbial Diversity, 40-68 (Lippincott Williams & Wilkins, 2011).
31 Sullivan, J. A. *CELLS alive!*, <https://www.cellsalive.com/cells/bactcell.htm> (2017).
32 Symonds, J. *Bacteria*, <http://microbiologyonline.org/about-microbiology/introducing-microbes/bacteria> (2018).
33 England, P. H. in *Introduction to the preliminary identification of medically important bacteria and fungi from culture* 32 (Public Health England, Standards Unit, Public Health England, 2017).
34 Naumann, D. *Infrared Spectroscopy in Microbiology*. 1-18 (2006).
35 Amyes, S. G. B. *Bacteria: A very short introduction*. 1st edn, (Oxford University Press, 2013).
36 Monterey institute of international studies. *Chronology of Aum Shinrikyo's CBW activities (2001)*, <http://www.nonproliferation.org/wp-content/uploads/2016/06/aum_chrn.pdf> (2001).
37 Takahashi, H. *et al. Bacillus anthracis* Incident, Kameido, Tokyo, 1993. *Emerging infectious disease* **10**, 117-120 (2004).
38 Jernigan, D. B. *et al.* Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiological findings. *Emerging infectious diseases* **8**, 1019-1028 (2002).
39 Jernigan, J. A. *et al.* Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States. *Emerging Infectious Diseases* **7**, 933-944 (2001).
40 Committee on Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Bacillus anthracis Mailings *et al. Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Anthrax Letters*. (The National Academies Press, 2011).
41 Dalton, R. Genetic sleuths rush to identify anthrax strains in mail attacks. *Nature* **413**, 657-658, doi:10.1038/35099687 (2001).
42 Mogilevsky, G., Borland, L., Brickhouse, M. & Fountain III, A. W. Raman Spectroscopy for Homeland Security Applications. *International Journal of Spectroscopy* **2012**, 1-12, doi:http://dx.doi.org/10.1155/2012/808079 (2012).

43 Cullimore, D. R. Ch. 1, 1-15 (CRC Press, 2000).
44 Logan, N. A. & Berkeley, R. C. W. Identification of Bacillus strains using the API system. *Journal of General Microbiology* **130**, 1871-1882 (1984).
45 Thavaselvam, D. & Vijayaraghavan, R. Biological Warfare Agents. *Journal of Pharmacy and Bioallied Sciences* **2**, 179-188 (2010).
46 Järvinen, A.-K. *et al.* Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. *BMC Microbiology* **9** (2009).
47 Makino, S.-i. & Cheun, H.-i. Application of the real-time PCR for the detection of airborne microbial pathogens in reference to the anthrax spores. *Journal of Microbiological Methods* **53**, 141-147 (2003).
48 Shiply, G. L. in *Real-Time PCR* (ed M. T. Dorak) Ch. 1, 1-38 (Taylor & Francis, 2006).
49 Belgrader , P. *et al.* PCR detection of bacteria in seven minutes. *Science* **284**, 449-450 (1999).
50 Belgrader, P. *et al.* A Minisonicator To Rapidly Disrupt Bacterial Spores for DNA Analysis *Analytical chemistry* **71**, 4232-4236 (1999).
51 Web of science. in *Web of science* (2018).
52 Ellis, D. I., Muhamadali, H., Haughey, S. A., Elliott, C. T. & Goodacre, R. Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis – moving out of the laboratory and into the food supply chain. *Analytical Methods* **7**, 9401-9414, doi:10.1039/c5ay02048d (2015).
53 Goodacre, R. *et al.* Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks *Microbiology* **144**, 1157-1170 (1998).
54 Miller, F. A. Reminiscences of Pioneers and Early Commercial IR Instruments. *Analytical Chemistry* **64**, 824A-831A (1992).
55 Kawahara, F. K. Vol. 3 Ch. 7, 468 (Academic Press, 1984).
56 Griffiths, P. R. Fourier Transform Infrared Spectrometry. *Science* **222**, 297-302, doi:10.1126/science.6623077 (1983).
57 Fredricks, P. in *Analytical Instrumentation Handbook* (ed J. Cazes) Ch. 7, 163-239 (CRC Press, 2004).
58 Winson, M. K. *et al.* Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR-based approach to rapid screening for metabolite overproduction. *Analytica Chimica Acta* **348**, 273-282, doi:http://dx.doi.org/10.1016/S0003-2670(97)00237-7 (1997).
59 Raman, C. V. A New Radiation. *Indian Journal of Physics*, 387-398 (1928).
60 Ellis, D. I., Cowcher, D. P., Ashton, L., O'Hagan, S. & Goodacre, R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *Analyst* **138**, 3871-3884, doi:10.1039/c3an00698k (2013).
61 Atkins, P. & de Paula, J. in *Elements of Physical Chemistry* Ch. Foundations: Building on the Foundations, 1-17 (Oxford University Press, 2013).
62 Larkin, P. in *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation* Ch. 2: Basic Principles, 7-25 (Elsevier, 2011).
63 Barth, A. & Harris, P. I. *Biological and Biomedical Infrared Spectroscopy*. 1st edn, 1-52 (IOS Press BV, 2009).
64 Baker, M. J., Hughes, C. S. & Hollywood, K. A. *Biophotonics: Vibrational Spectroscopic Diagnostics*. (Morgan & Claypool, 2016).
65 Atkins, P., Overton, T., Rourke, J., Weller, M. & Armstrong, F. *Shriver and Atkins' Inorganic Chemistry*. 5th edn, (OUP Press).
66 Ferraro, J. R., Nakamoto, K. & Brown, C. W. in *Introductory Raman Spectroscopy* Ch. 1: Basic Theory, 1-94 (Elsevier, 2003).

67 Sathyanarayana, D. N. in *Vibrational Spectroscopy: Theory and Applications* Ch. 1: General Introduction, 1-43 (New Age International (P) Limited Publishers, 2004).
68 Sathyanarayana, D. N. in *Vibrational Spectroscopy: Theory and Applications* Ch. 7: Raman Spectroscopy, 294-347 (New Age International (P) Limited Publishers, 2004).
69 Horiba. *Jablonski Energy Diagram for Raman Scattering*, <http://www.horiba.com/uk/scientific/products/raman-spectroscopy/raman-academy/raman-tutorial/raman-scattering/> (
70 Holmes, A. *Combination Bands, Overtones and Fermi Resonances*, <https://chem.libretexts.org/Core/Physical_and_Theoretical_Chemistry/Spectroscopy/Vibrational_Spectroscopy/Vibrational_Modes/Combination_Bands%2C_Overtones_and_Fermi_Resonances> (2016).
71 Baker, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological material. *Nature Protocols* **9**, 1771-1791, doi:10.1038/nprot.2014.110 (2014).
72 Cammarata, L., Kazarian, S. G., Salter, R. A. & Welton, T. Molecular states of water in room temperature ionic liquids. *Physical Chemistry Chemical Physics* **3**, 5192-5200 (2001).
73 Banwell, C. N. in *Fundamentals of Molecular Spectroscopy* Ch. 3:Infra-red Spectroscopy, 55-99 (McGraw-Hill Book Company, 1983).
74 Shimadzu. *ABC's of the Diffuse Reflection Method*, <https://www.shimadzu.com/an/ftir/support/ftirtalk/talk1/intro.html> (2018).
75 Cid, M. M. & Bravo, J. *Structure Elucidation in Organic Chemistry: The Search for the Right Tools*. 1st edn, 145-172 (Wiley, 2015).
76 Ferraro, J. R., Nakamoto, K. & Brown, C. W. *Introductory Raman Spectroscopy*. 2nd edn, (Elsevier, 2003).
77 Butler, H. J. *et al.* Using Raman spectroscopy to characterize biological materials. *Nature Protocols* **11**, 664-687, doi:10.1038/nprot.2016.036 (2016).
78 Science, C. C. f. M. TM002-02-A Introduction to WiRE and System start-up. 28. <http://www.ccmr.cornell.edu/wp-content/uploads/sites/2/2016/08/Wire4_Training_Modules_Compilation-Part-2.pdf>.
79 Shimadzu. *FTIR-8400S: Shimadzu Fourier Transform Infrared Spectrophotometer*, <http://photos.labwrench.com/equipmentManuals/7783-2914.pdf> (
80 Agilent. *Agilent Cary 600 Series FTIR*, <https://www.agilent.com/cs/library/brochures/Agilent_FTIR_660-670-680_Brochure.pdf> (2011).
81 Bruker. *HTS-XT FT-IR High Throughput Screening Extension*, <https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/OpticalSpectrospcopy/FT-IR/HTS-XT/Flyers/HTS-XT_Flyer.pdf> (2013).
82 Siesler, H. W. *Bruker Equinox 55 FTIR/FTNIR Spectrometer*, <http://www.nir-spektroskopie.de/geraete/equinoxe.htm> (
83 Stevenson, H. J. R. & Bolduan, O. E. A. Infrared spectrophotometry as a means for identification of Bacteria. *Science* **116**, 111-113 (1952).
84 Naumann, D. Infrared Spectroscopy in Microbiology, 1-29 (2006).
85 Goodacre, R. *et al.* Detection of the Dipicolinic Acid Biomarker in Bacillus Spores Using Curie-Point Pyrolysis Mass Spectrometry and Fourier Transform Infrared Spectroscopy *Analytical Chemistry* **72**, 119-127 (2000).
86 Muhamadali, H., Chisanga, M., Subaihi, A. & Goodacre, R. Combining Raman and FT-IR Spectroscopy with Quantitative Isotopic Labeling for Differentiation of E. coli Cells at Community and Single Cell Levels *Analytical Chemistry* **87**, 4578-4586, doi:10.1021/acs.analchem.5b00892 (2015).

87 Goodacre, R. I spy with my little eye something beginning with … 'H'. *Metabolomics* **11**, 6-8, doi:10.1007/s11306-014-0762-7 (2015).
88 Goodacre *et al.* Metabolomics by numbers - acquiring and understanding global metabolite data. *Trends in Biotechnology* **22**, 245-252, doi:10.1016/j.tibtech.2004.03.007 (2004).
89 Helm, D., Labischinski, H., G., S. & Naumann, D. Classification and Identification of bacteria by Fourier-Transform Infrared Spectroscopy. *Microbiology* **137**, 69-79 (1991).
90 Naumann, D., Fijala, V. & Labischinski, H. The Differentiation and Identification of Pathogenic bacteria using FTIR and Multivariate Statistical Analysis. *Mikrochimica Acta* **94**, 373-377 (1988).
91 Naumann, D., Helm, D. & Labischinski, H. Microbiological characterizations by FT-IR spectroscopy. *Nature* **351**, 81-82 (1991).
92 Maquelin, K. *et al.* Identification of medically relevant microorganisms by vibrational spectroscopy. *Journal of Microbiological Methods* **51**, 255-271, doi:http://dx.doi.org/10.1016/S0167-7012(02)00127-6 (2002).
93 Timmins , E. M., Howell , S. A., Alsberg , B. K., Noble , W. C. & Goodacre, R. Rapid Differentiation of Closely Related Candida Species and Strains by Pyrolysis-Mass Spectrometry and Fourier Transform-Infrared Spectroscopy. *Journal of Clinical Microbiology* **36**, 367-374 (1998).
94 Goodacre, R. & Kell, D. B. Pyrolysis mass spectrometry and its applications in biotechnology. *Current Opinion in Biotechnology* **7**, 20-28, doi:http://dx.doi.org/10.1016/S0958-1669(96)80090-5 (1996).
95 McIntosh, A. J. S. *et al.* Spectroscopic analysis of bacterial biological warfare simulants and the effects of environmental conditioning on a bacterial spectrum. *Journal of Analytical Bioanalytical Chemistry* **404**, 2307-2315 (2012).
96 Foster, N. S., Thompson, S. E., Valentine, N. B., Amonette, J. E. & Johnson, T. J. Identification of Sporulated and Vegetative Bacteria Using Statistical Analysis of Fourier Transform Mid-infrared Transmission Data. *Applied Spectroscopy* **58**, 203-211 (2004).
97 Johnson, T. J. *et al.* The Infrared Spectra of Bacillus Bacteria Part I: Vegetative Bacillus versus Sporulated Cells and the Contributions of Phospholipids to Vegetative Infrared Spectra. *Applied Spectroscopy* **63**, 899-907 (2009).
98 Johnson, T. J., Williams, S. D., Valentine, N. B. & Su, Y. F. The Infrared Spectra of Bacillus Bacteria Part II: Sporulated Bacillus—The Effect of Vegetative Cells and Contributions of Calcium Dipicolinate Trihydrate,CaDP·3H2O. *Applied Spectroscopy* **63**, 908-915 (2009).
99 Forrester, J. B., Valentine, N. B., Su, Y. F. & Johnson, T. J. Chemometric analysis of multiple species of Bacillus bacterial endospores using infrared spectroscopy: Discrimination to the strain level. *Analytica Chimica Acta* **651**, 24-30 (2009).
100 De Gelder, J. *et al.* Raman spectroscopic study of bacterial endospores. *Journal of Analytical Bioanalytical Chemistry* **389**, 2143-2151 (2007).
101 Manoharan, R. *et al.* UV resonance Raman spectra of bacteria, bacterial spores, protoplasts and calcium dipicolinate. *Journal of Microbiological Methods* **11**, 1-15 (1990).
102 Dalterio, R. A., Nelson, W. H., Britt, D. & Sperry, J. F. An Ultraviolet (242 nm Excitation) Resonance Raman Study of Live Bacteria and Bacterial Components. *Applied Spectroscopy* **41**, 417-422 (1987).

103 Dıez, E. C. L. & Goodacre, R. Characterization of Microorganisms Using UV Resonance Raman Spectroscopy and Chemometrics. *Analytical Chemisty* **76**, 585-591 (2004).
104 Jarvis, R.M. & Goodacre, R. Characterisation and identification of bacteria using SERS. *Chemical Society Reviews* **37**, 931-936, doi:10.1039/b705973f (2008).
105 Sauer-Budge, A. F., Ziegler, L. D., Premasiri, W. R., Klapperich, C. M. & Lee, J. C. Rapid Bacterial Diagnostics Via Surface-Enhanced Raman Microscopy. *Spectroscopy* **27** (2012).
106 Bell, S. E., Mackle, J. N. & Sirimuthu, N. M. Quantitative surface-enhanced Raman spectroscopy of dipicolinic acid—towards rapid anthrax endospore detection. *The Analyst* **130**, 545-549 (2005).
107 Jarvis, R. M., Brooker, A. & Goodacre, R. Surface-enhanced Raman scattering for the rapid discrimination of bacteria. *Faraday Discussion* **132**, 281-292 (2006).
108 Cowcher, D. P., Xu, Y. & Goodacre, R. Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering. *Analytical chemistry* **85**, 3297-3302 (2013).
109 Cheung, M., Lee, W. W. Y., Cowcher, D. P., Goodacre, R. & Bell, S. E. J. SERS of meso-droplets supported on superhydrophobic wires allows exquisitely sensitive detection of dipicolinic acid, an anthrax biomarker, considerably below the infective dose. *Chemical Communications* **52**, 9925-9928, doi:10.1039/c6cc03521c (2016).
110 Petrov, G. I. *et al.* Comparison of coherent and spontaneous Raman microspectroscopies for non-invasive detection of single bacterial endospores. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7776-7779 (2007).
111 Pestov, D. *et al.* Single-shot detection of bacterial endospores via coherent Raman spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 422-427, doi:10.1073/pnas.0710427105 (2008).
112 Setlow, B., Atluri, S., Kitchel, R., Koziol-Dube, K. & Setlow, P. Role of Dipicolinic Acid in Resistance and Stability of Spores of Bacillus subtilis with or without DNA-Protective α/β-Type Small Acid-Soluble Proteins. *Journal of Bacteriology* **188**, 3740-3747 (2006).
113 Lasch, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging *Chemometrics and Intelligent Laboratory Systems* **117**, 100-114, doi:10.1016/j.chemolab.2012.03.011 (2012).
114 Byrne, H., Knief, P., Keating, M. & Bonnier, F. Spectral Pre and Post Processing for Infrared and Raman Spectroscopy of Biological Tissues and Cells. *Chem. Soc. Rev* **45**, 1865-1878, doi:doi:10.1039/c5cs00440c (2016).
115 Diem, M. in *Modern Vibrational Spectroscopy and Micro-Spectroscopy: Theory, Instrumentation and Biomedical Applications* Ch. 12, 251-282 (Wiley, 2015).
116 Henderson, A. Minimum Reporting Requirement for Clinical IR and Raman Spectroscopy. (2015). <https://docs.google.com/document/d/1d4KTDWQBRuwfNNiNp4Xuunz_dtLSqw2TLoq9K4J9gig/edit>.
117 Antonelli, P. *et al.* A principal component noise filter for high spectral resolution infrared measurements. *Journal of Geophysical Research* **109**, 1-22, doi:10.1029/2004JD004862 (2004).
118 Wharfe, E. S., Jarvis, R. M., Winder, C. L., Whiteley, A. S. & Goodacre, R. Fourier transform infrared spectroscopy as a metabolite fingerprinting tool for monitoring the phenotypic changes in complex bacterial communities capable of degrading

phenol. *Environmental Microbiology* **12**, 3253-3263, doi:10.1111/j.1462-2920.2010.02300.x (2010).

119 Afseth, N. K. & Kohler, A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems* **117**, 92-99, doi:https://doi.org/10.1016/j.chemolab.2012.03.004 (2012).

120 Martens, H., Nielsen, J. P. & Engelsen, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. *Analytical Chemistry* **75**, 394-404, doi:10.1021/ac020194w (2003).

121 Abdi, H. & Williams, L. J. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 433-459, doi:DOI: 10.1002/wics.101 (2010).

122 Smilde, A. K., Westerhuis, J. A. & de Jong, S. A framework for sequential multiblock component methods. *Journal of Chemometrics* **17**, 323-337, doi:10.1002/ cem.811 (2003).

123 Westerhuis, J. A., Kourti, T. & Macgregor, J. F. Analysis of Multiblock and Hierarchical PCA and PLS models. *Journal of Chemometrics* **12**, 301-321 (1998).

124 Ami, D., Mereghetti, P. & Doglia, S. M. in *Multivariate Analysis in Management, Engineering and the Sciences* (eds L. V. de Freitas & A. P. B. R. de Freitas) Ch. 10: Multivariate Analysis for Fourier Transform Infrared Spectra of Complex Biological Systems and Processes, 189-220 (Intech, 2012).

125 Poulsen, J. & French, A. *Discriminant Function Analysis*, <https://pdfs.semanticscholar.org/8e10/85b33d84d7015174e30e3e911272b6e1c2fe.pdf>

Chapter 2

Bacterial Preparation, Construction of Substrate Library and Method Development for Substrate Removal using Vibrational Spectroscopy

2.1 Introduction and Aims

This chapter discusses research that has focused on assessing the most appropriate parameters (such as spectral range, resolution and number of co-added scans) to use with each technique, infrared and Raman, by investigating the set of substrates selected for this project in collaboration with technical experts at the Defence Science and Technology Laboratories (DSTL).

The first section of this chapter discusses the development of a database of substrate spectra using infrared and Raman. It is important to have these reference spectra available to fundamentally understand the characteristics of each of the substrates to ensure that appropriate compensations can be made for any spectral contributions that arise. A range of spectral parameters were trialled such as internal reflection element (IRE) type for Attenuated Total Reflection – Fourier Transform

Infrared (ATR-FTIR) Spectroscopy and laser wavelength for Raman in order to quantify which parameters were ideal for analysis conducted in further studies.

Small amounts of sample and low sample thicknesses can lead to the substrate contributing to the overall spectrum produced by a sample, when considering the penetration depth associated with an ATR-FTIR IRE. This means identification of surface deposited samples may be skewed or not possible. Developing a method to remove the infrared spectral contribution from substrates would eliminate the need for samples to be recovered from a surface to then undergo complex sample preparation prior to analysis, allowing *in situ* analysis of samples to take place. Removing substrate contribution would allow for samples at low concentrations or with sparse coverage to still be identified *in situ*. The second section of this chapter investigates two different methods that can be used to remove the spectral contribution relating to a substrate to produce absorbance spectra of either glucose or ovalbumin. The use of glucose and ovalbumin as surrogates is important when considering the end goal of using the method for *in-situ* identification of surface deposited bacteria.

2.2 Materials and Methods

2.2.1 Substrate Study

The eight different substrates used in this study: lino, jacket, boot, clean filter, dirty filter, metal tile, concrete and a MirrIR slide are shown in Figure 1. The

substrates were chosen by Dstl as surfaces of interest, with MirrIR added as a standard substrate used commonly with ATR-FTIR.

Spectra were collected from five different locations, three spectra per location, across the surface of the substrates (see figure 1) to ensure the substrate as a whole was represented in the spectrum as some of the substrates are not completely homogenous, a factor which is key to be considered. Spectral libraries were built for both ATR-FTIR and Raman spectroscopy. The ATR-FTIR study used different Internal Reflection Elements (IREs) while the Raman study used lasers of different wavelengths to optimise the ideal parameters for analysis. As such, an overall average spectrum to represent each substrate has been produced for each different parameter and is shown in the respective sub-sections within this chapter, depending on if ATR-FTIR or Raman was used.

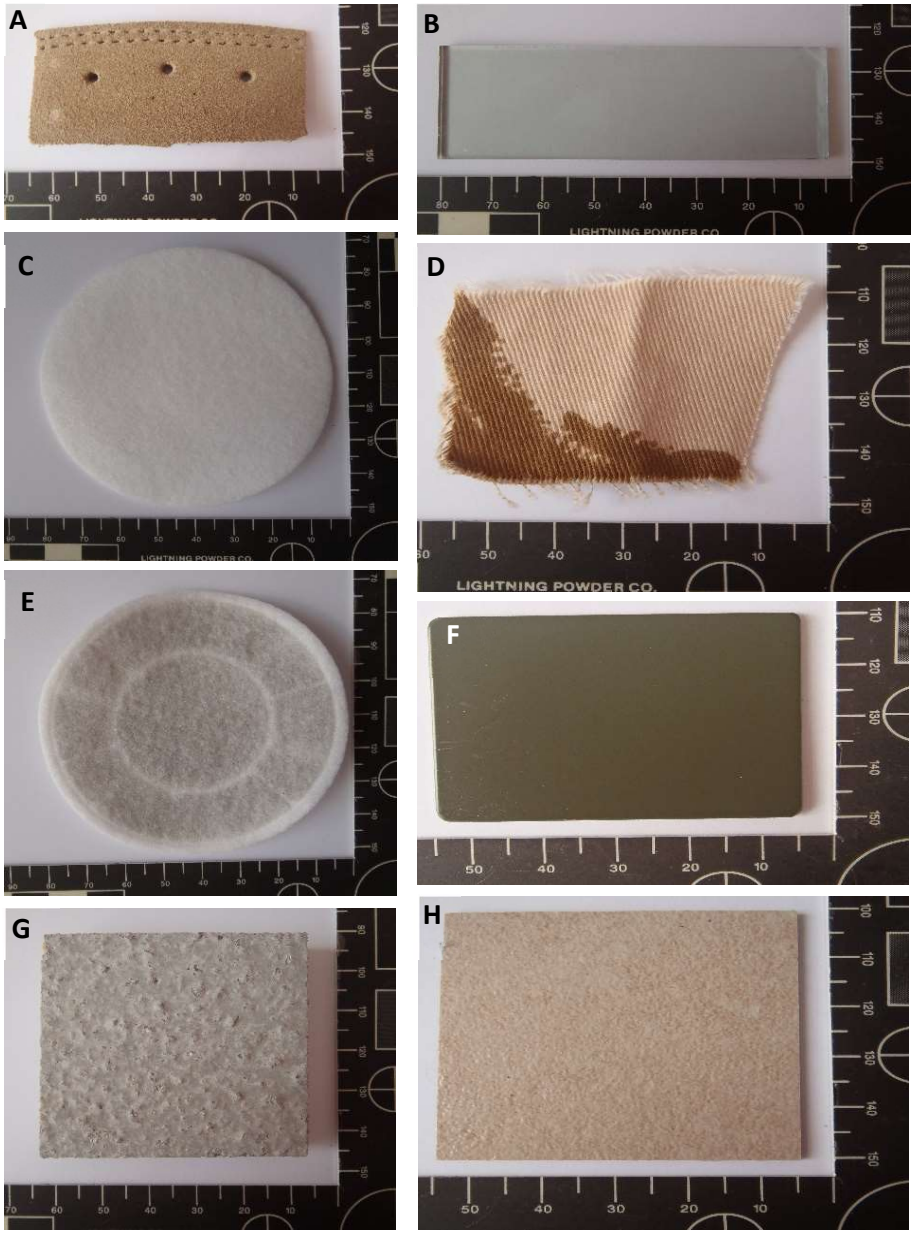

Figure 1: Photograph's of the substrates chosen for use in this study L-R: (A) boot, (B) MirrIR slide, (C) clean filter, (D) jacket, (E) dirty filter, (F) metal tile, (G) concrete and (H) lino.

2.2.2 ATR-FTIR

ATR-FTIR spectra were initially collected from a Shimadzu FTIR 8400 S in absorbance mode, using a DiZnSe IRE accessory, over a range of 4000-750 cm^{-1} with a resolution of 4 cm^{-1} and 32 co-added scans. Further ATR-FTIR analysis was carried out at the University of Strathclyde using a newly installed Agilent Cary 600 series spectrometer with a Pike MIRacle ATR DiZnSe IRE accessory attached. Spectra were collected using 32 co-added scans, 4 cm^{-1} resolution over a range of 4000-600 cm^{-1}. Both the DiZnSe and the Ge IRE were used to allow further comparative analysis to observe which accessory was most suited for future studies. Prior to spectral collection the IRE was cleaned using Virkon and ethanol (Sigma Aldrich, UK) with a background collected between each section of the substrate analysed. Background scans were collected with a resolution of 4 cm^{-1} and 32 co-added scans. The area analysed was dependent on the size of the IRE in the ATR accessory which has a ~2 mm diameter. To produce each spectrum the substrate was clamped to ensure there was an intimate contact between the substrate and the IRE, see Figure 2. For each spot three spectra were collected to act as technical replicates before the slide was removed so that the IRE could be cleaned and a new background collected before the next spot was analysed.

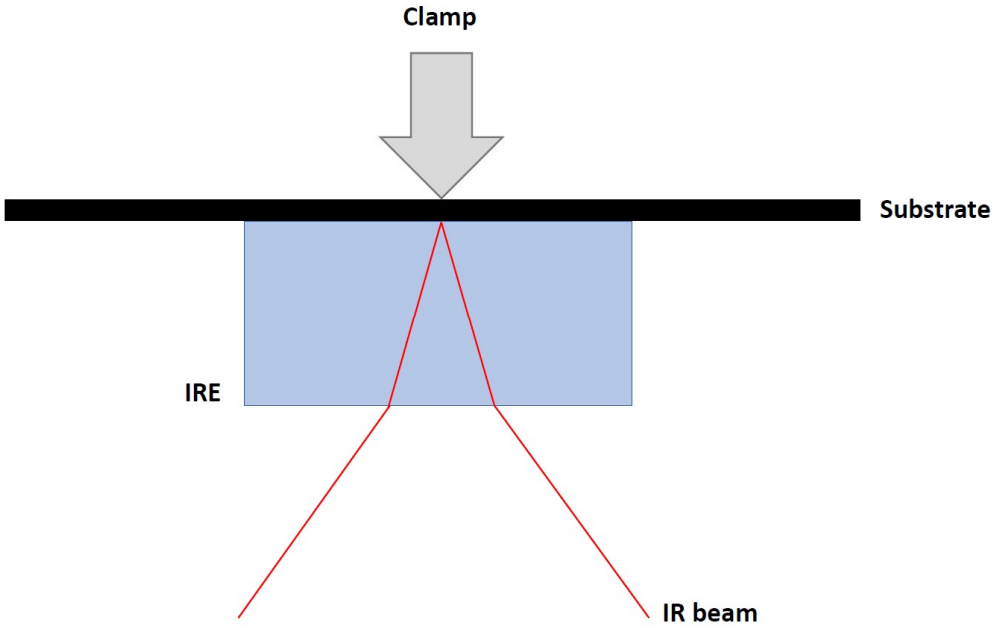

Figure 2: Schematic showing the set up for spectral collection using ATR-FTIR for building the substrate library.

2.2.3 Raman

Raman spectra from the substrates were acquired initially at the University of Central Lancashire (UCLan) using a Horiba Jobin-Yorn LabRAM HR800 confocal Raman system fitted with two lasers; 785 nm (400 mW at source) and 532 nm (300 mW at source). Different combinations of parameters, such as filter, accumulations and acquisition time, were explored in order to optimise a set of parameters that could be used for spectral acquisition across all substrates. Two different gratings were available for use with this instrument, 300 mm^{-1} and 600 mm^{-1} with the spectral resolution increasing respectively. [1] The parameters used for acquiring these spectra were a 20 s acquisition time with 2 accumulations, utilising a 300 mm^{-1} grating and

x50LWD objective as a static scan. The filter used was dependent on the substrate that was being examined with the optimum filter being either 0.1, 1, 10 or 25%. The spectral range examined when using 785 nm for clean filter, concrete, dirty filter and lino spanned from 70-1600 cm^{-1}, for boot it was 70-1700 cm^{-1} and for jacket, metal tile and a MirrIR slide ranged from 200-1600 cm^{-1}. The spectral range examined when using 532 nm for clean filter, concrete and dirty filter spanned from 70-1600 cm^{-1}, for boot it was 70-1700 cm^{-1} and for a MirrIR slide the range was 200-1600 cm^{-1}. Prior to any spectral collection all Raman instruments were calibrated using a Silicon standard.

Raman spectra were also collected at the University of Manchester (UoM) using a Renishaw InVia system with a 633 nm laser (17mW at source) fitted. This system also had two gratings, 600 and 1200 mm^{-1}, available for use. A x50LWD objective was used with a 1200 mm^{-1} grating. As with previous spectral collection there was no optimal set of parameters with different acquisition times, number of accumulations, amount of light filtered depending on the substrate being examined. This information is included in the caption for each figure which can be found in Appendix A. A multipoint map with 10 µm space between spots and depth profile with 1 µm between each acquisition were acquired for each substrate to try and understand more about the structure, homogeneity and consistency of each substrate. Outliers from the dirty filter spectral map were removed before mean and standard deviation (SD) spectra were generated. As previously mentioned, all lasers were calibrated using a silicon standard. The spectral range examined when using 633 nm for boot, clean filter and dirty filter and spanned from 70-1600 cm^{-1}.

2.2.4 Development of Substrate Removal Methodology

2.2.4.1. Glucose Study

A proof-of-principle study was carried out to develop an easy to use method of substrate removal that can be applied to different samples types such as simple sugars and protein rich bacterial samples. D-glucose (glucose) was used in this study which is a simple sugar. The principle behind the substrate removal method used for this study is to use the Beer-Lambert Equation

$$\text{Abs} = -log\frac{I_S}{I_R} \tag{3.1}$$

where, I_S the intensity of the light before going through the sample and I_R is the intensity of the light after going through the sample. Eq. 3.1 is used to manually remove the spectral components that belong to the substrate and generate an absorbance spectrum for the material on the surface, in this particular case glucose, based on the differing intensities.

A single channel environmental spectrum was collected using the DiZnSe IRE before to a substrate was clamped to the IRE. The spectrum collected from the substrate was also collected as a single channel spectrum. The substrate was unclamped and moved before a new sample spectrum was acquired. This was repeated until ten sample spectra had been acquired. The same protocol was used to collect the data required for a substrate that had been contaminated with glucose. This provided a dataset of 20 spectra per substrate, ten as standard substrate spectra and ten as contaminated substrate spectra where the contaminant was glucose.

Spectra were then exported from the IRsolution software as .txt files and opened within MATLAB to allow for the Beer-Lambert substrate removal to be performed.

To fully assess how well the method performed, the individual spectra were ranked using the following grading system: 'Very good', 'Good', 'Adequate', 'Poor' and 'Very poor'. The basis for the judgement was based on the vibration seen in the spectra produced when the substrate has been removed. Ibrahim *et al.,* discussed the key vibrations produced by glucose. [2] Key vibrations highlighted in a glucose spectrum include C-C, C=O, O-H and C-H. The definitions of each category of this grading system are based on these vibrations and can be seen in Table I. This included making sure that the key vibrations were present in the spectra produced for glucose in this study.

Table I: Four class grading system devised for ranking absorbance spectra generated after substrate removal using the Beer-Lambert method with glucose, including a description of each grade

Grade	Description
Very good	Key vibrations present with no peak inversions
Good	Some vibrations missing and/or inversions in the lipid region.
Adequate	Some vibrations missing and/or inversions in the fingerprint region where the amide, nucleic acids and carbohydrates are found.
Poor	Some vibrations missing and/or inversions in the lipid region **and** the amide region.
Very poor	Some vibrations missing and/or inversions in any region of spectrum (excluding CO_2 peaks).

For this grading system distortion of the spectra due to the baseline is not considered as none of the spectra have been corrected in this manner. It is widely accepted that the key region for identifying a bacterial spectrum is the 'fingerprint region' which ranges from 1800-800 cm^{-1}. For this reason, any spectra showing

inversion within this region along with inversions in the high wavenumber region are immediately classed as 'Very poor'. To assess if all glucose vibrations are present in the spectrum, Ibrahim *et al.,* peak table was used. [1]

During the drying of the glucose in this study there were a number of problems with the viscosity of the glucose. There were also problems with the glucose solution not being dry enough to withstand compression for analysis using the ATR-FTIR, particularly with non-porous surface such as lino. It was found that around four hours in an oven at 37°C was enough to dry the glucose so that it was not damaged during analysis and also to avoid the sample crystallising as shown in Figure 3. Single channel spectra were collected using 32 co-added scans, 4 cm^{-1} resolution over a range of 4000-750 cm^{-1}. The spectra were collected from 10 different spots across each of the eight different substrates. The ten spectra were then averaged to create a representative spectrum for the substrate. A solution of 0.15 M glucose was then applied to each of the substrates before they were placed in a 37°C oven to dry for around four hours.

Figure 3: Photograph showing crystallised glucose on a lino section of dimensions shown in Figure 1H.

Once the samples had dried individual spectra were collected from ten different spots using the Shimadzu FTIR 8400 S spectrometer in single channel mode was used using an ATR accessory. Prior to spectral collection and between spots the ATR accessory was cleaned using Virkon/Trigene and ethanol (Sigma Aldrich, UK) and a background collected, using the same settings as previously discussed. For each substrate, the area where the glucose had been spotted was clamped down to ensure there was an intimate contact between the glucose and the IRE, see Figure 4.

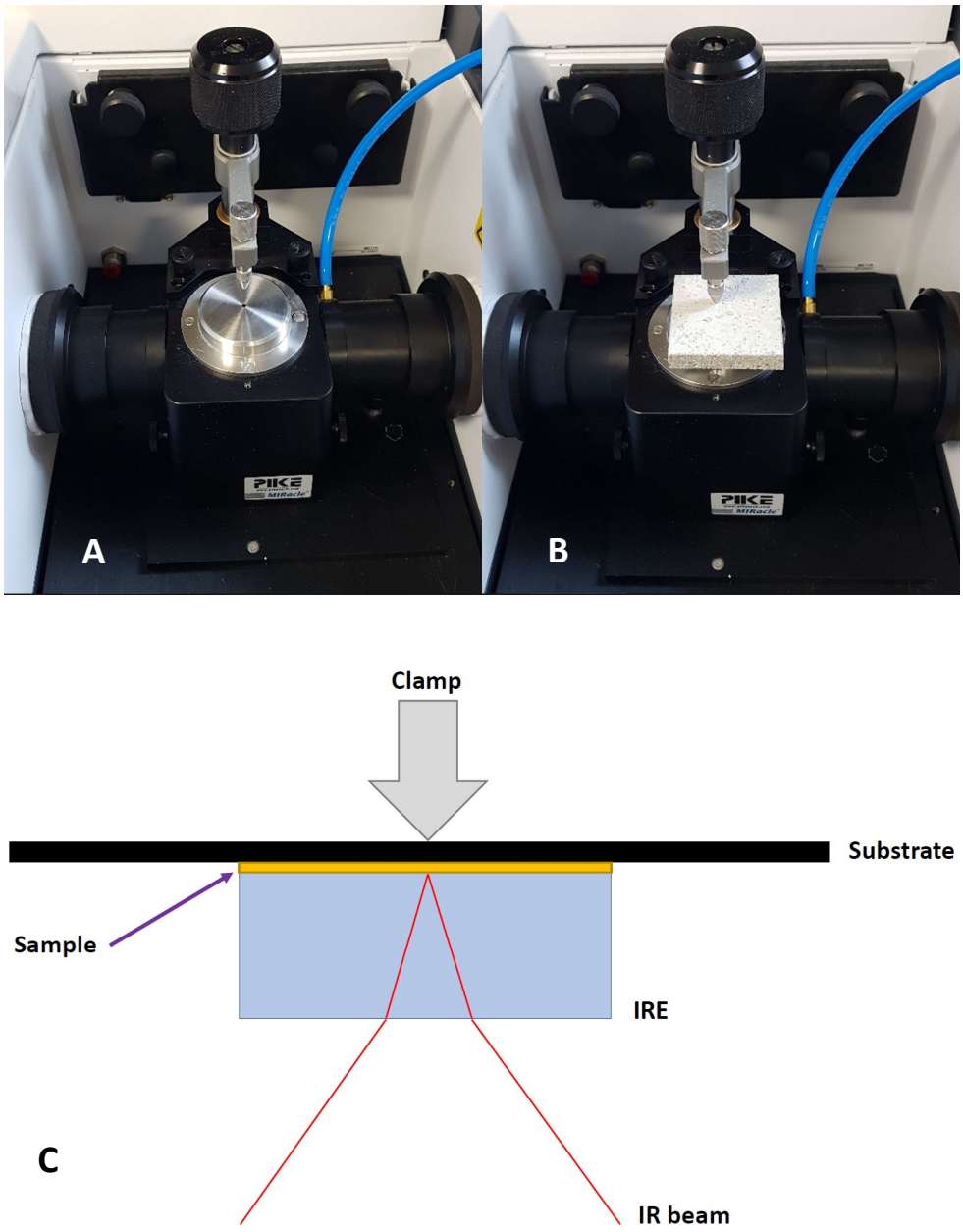

Figure 4: Photograph A shows the FTIR setup with the ATR attachment installed. Photograph B shows a piece of concrete clamped to the ATR attachment. Below these is a schematic (C) showing the setup for ATR-FTIR spectral collection from glucose and ovalbumin samples prepared on a substrate.

Once the spectra were exported from the Shimadzu software (IRsolution) and opened within MATLAB, as mentioned previously, the Beer-Lambert substrate removal stage was performed. For this, both standard substrate and substrate with

glucose spectra were opened in MATLAB. Spectra from ten spots across the substrate surface were collected, opened in MATLAB and averaged to create a single representative spectrum to later be removed. The averaged substrate spectrum was then removed individually from the spectra obtained from the substrate contaminated with glucose, using Eq. 3.1, to produce ten glucose spectra per substrate. For an example of this using spot 1, the averaged substrate spectrum was first divided by the spectra acquired from spot 1 of the glucose contaminated sample. This creates a new variable which is the single channel data for spot 1 with the substrate contribution removed, theoretically pure glucose. This variable then had the function -log performed on it to generate another new variable, which when plotted represented the absorbance spectrum for spot 1. This method was also used with ovalbumin, a more complex compound also found within bacteria. The spectra produced during both of these studies using this method can be seen in the sections 2.3.3.1, 2.3.3.2 and Appendix A.

2.2.4.2. Ovalbumin Study

A study using ovalbumin (Sigma Aldrich) rather than glucose was carried out to compare the ability of the Beer-Lambert removal method developed in section 2.2.4.1. The study used ovalbumin, a more complex compound than glucose, to first test the Beer-lambert method before trailing an alternative method where the substrate was removed within the software. This would reduce the complexity of the

methodology as the Beer-lambert method requires the data to be exported from the software and manipulated in MATLAB. A comparison of the processes used in the Beer-Lambert method and the software-based method can be seen in Figure 5.

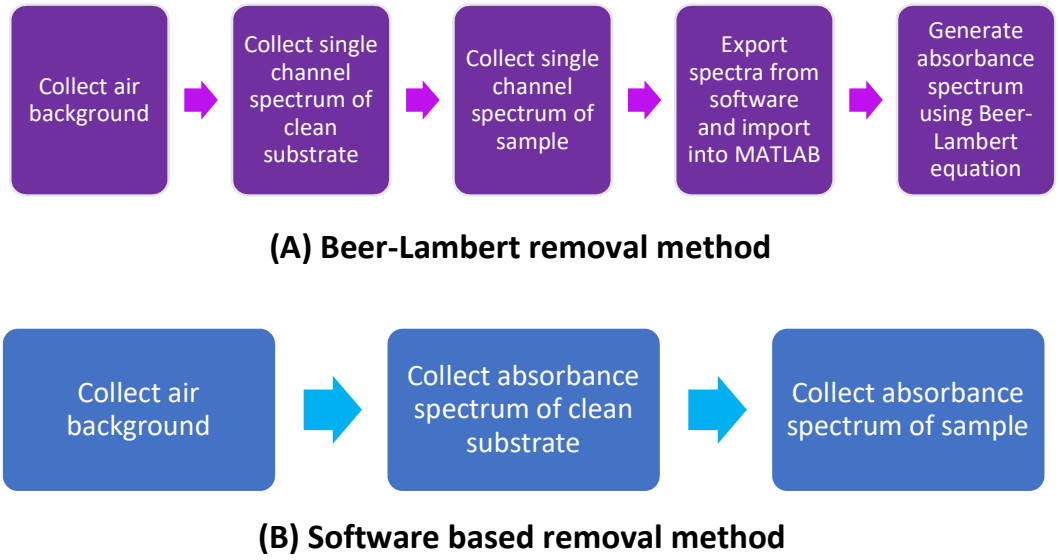

Figure 5: Flow charts showing the process followed for the Beer-Lambert removal method (A) and the software-based removal method (B).

Similarly, with the glucose study, to assess how well each of the methods performed, the individual spectra were ranked again with grading system: 'Very good', 'Good', 'Adequate', 'Poor' and 'Very poor'. The definitions of each category of this grading system are slightly different to those for the glucose study as the key regions for ovalbumin are different. Velusamy highlights that the key area of an ovalbumin spectrum is the Amide peaks found within the fingerprint region. [3] For this reason, the Amide I and II peaks (1800-1500 cm^{-1}) are key for determining the

spectral quality of the spectrum. The definitions for the ovalbumin grading system are shown in Table II.

Table II: Four class grading system devised for ranking the absorbance spectra generated after substrate removal in the ovalbumin study, for both methods, including a description of each grade

Grade	Description
Very good	All vibrations present with no peak inversions
Good	Some vibrations missing and/or inversions in the region lipid region **or** the nucleic acid/carbohydrate regions
Adequate	Some vibrations missing and/or inversions in the Amide region.
Poor	Some vibrations missing and/or inversions in the lipid region **and** the nucleic acid/carbohydrate regions. **No inversions in the Amide region.**
Very poor	Some vibrations missing and/or inversions in any region of spectrum (excluding CO_2 peaks).

This second method again used ATR-FTIR analysis with a DiZnSe IRE but rather than collecting single channel data and using the Beer-Lambert equation to manually generate absorbance spectra, absorbance spectra we collected after the instrument had been backgrounded to a clean section of substrate. This way the Agilent software (Resolutions Pro) automatically performed the removal of the substrate to produce an absorbance spectrum of the compound on the surface, in this case ovalbumin.

Two methodologies were used in this study, the first of which was the same as that used in the glucose study, however this time the Agilent Cary 660 system and Resolutions Pro software were used instead of the Shimadzu 8400S system and IRsolution software. The second method used for this study is slightly different. Two sets of substrates were required for the method, one set of standard substrates (the

uncontaminated set) and a set that had a solution of ovalbumin applied (the contaminated set). A solution of 50-100 mg/ml ovalbumin was applied to each of the substrates before they were placed in a 37°C oven to dry for around three and a half hours. As described within section 2.2.4.1, an air background was taken with the DiZnSe IRE attached before the uncontaminated substrate was clamped to the IRE and another spectrum taken. This is where the method differs from the glucose method. Within the method discussed in section 2.2.4.1 this spectrum is taken as an absorbance spectrum while here it remains in single channel mode. This became the reference background within the software. The uncontaminated substrate was removed and replaced with the one contaminated with ovalbumin. A single spectrum was obtained before the substrate was moved to a new spot. This spectrum was collected in absorbance mode and is a representative spectrum of ovalbumin with the background removed. This method was repeated until 10 absorbance spectra had been collected. Absorbance spectra were collected using 32 co-added scans, 4 cm^{-1} resolution over a range of 4000-750 cm^{-1}. The spectra were collected from 10 different spots across each of the eight different substrates.

2.3 Results

2.3.1 ATR-FTIR Substrate Study Results

A spectral database has been compiled from each of the eight different substrates chosen for this study. All of the spectra generated and the corresponding

tentative peak assignments, for this dataset can be seen in Appendix A: Figures 1-14 and Tables 1-7. Therefore, shown within this chapter are a representative selection of spectra for the ATR-FTIR study. Figure 6 and Figure 7 show a 5-spot overlay (3 replicates per spot), the mean and mean with SD cloud for clean filter using both the DiZnSe and Ge IRE. The corresponding tentative vibrations that have been assigned to the peaks can be found in Table III.

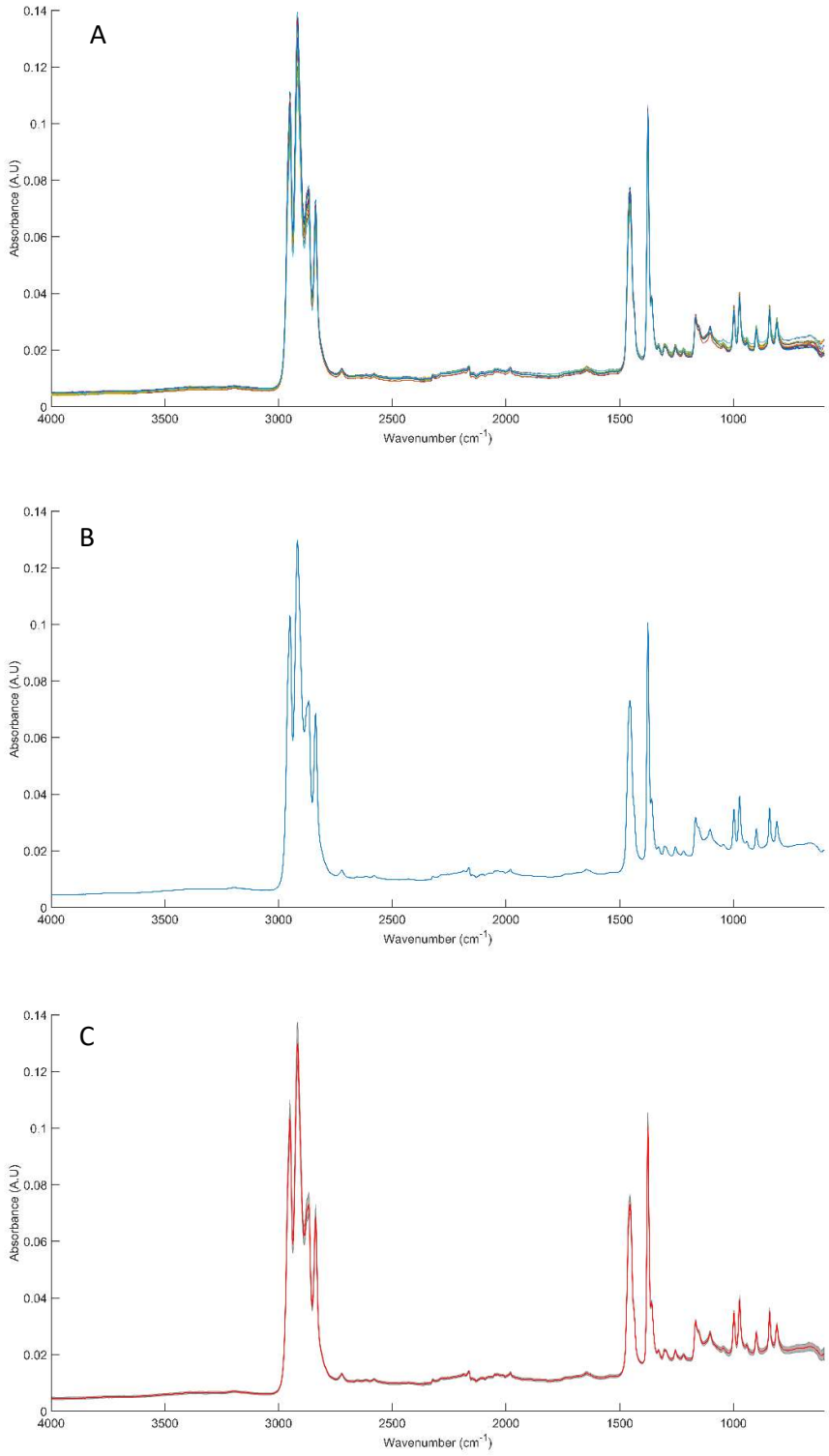

Figure 6: Collection of spectra collected from clean filter acquired using a DiZnSe IRE starting with an overlay of 15 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C)

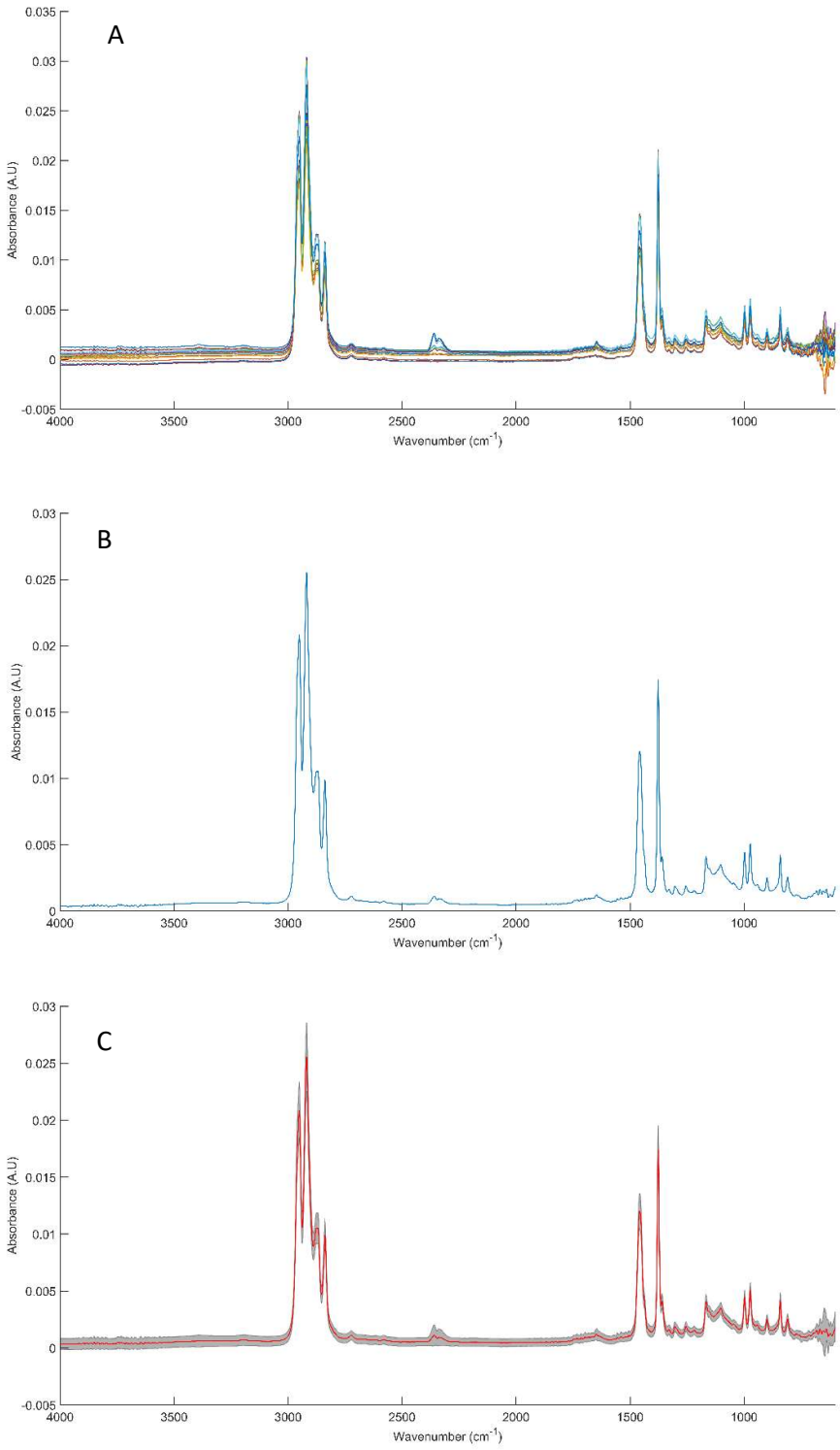

Figure 7: Collection of spectra collected from clean filter acquired using a Ge IRE starting with an overlay of 15 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

Table III: Wavenumbers and corresponding peak assignment for the average clean filter spectrum collected using ATR-FTIR with Ge and DiZnse IRE. [3,4,5,6]

Wavenumber cm^{-1}		Tentative peak assignment
Ge	DiZnSe	
808, 841, 899, 972	808, 841, 899, 972	vC-O-C, vCC alicyclic, aliphatic chains
997, 1302	997, 1302	vCC alicyclic, aliphatic chains
1103	1101	vC-S aromatic
1167	1165	vC=S
1255	1255	Amide III
1360	1360	vC-(NO_2)
1377	1375	δCH_3
1458	1454	vCC aromatic ring chain vibrations
2839, 2870, 2920, 2951	2837, 2868, 2916, 2951	vC-H

The spectra collected during this study, shown here and in Appendix A, demonstrate the complexity of the spectral signature for each substrate, with some substrates exhibiting more spectral features in their fingerprints than others due to differences in their structure. Of particular interest are the clear differences that have been detected using ATR-FTIR between the clean and dirty filter, simply by visually inspecting the spectra. An overlay of the averaged spectra (Figure 8) show a visible difference between the substrates at 668, 1647 and 3371 cm^{-1}. These differences may represent O-H bonds a result of moisture being trapped within the dirty filter as it has been used and exposed to the environment as opposed to the clean filter which is a new and unused filter. However, these peaks can also relate to amide stretches, which would be present in environmental bacteria found on the used filter. It is likely that these peaks are a combination of both O-H and amide stretches. The fact that ATR-FTIR can detect these differences highlight its potential use for the identification of bacteria found on complex substrates.

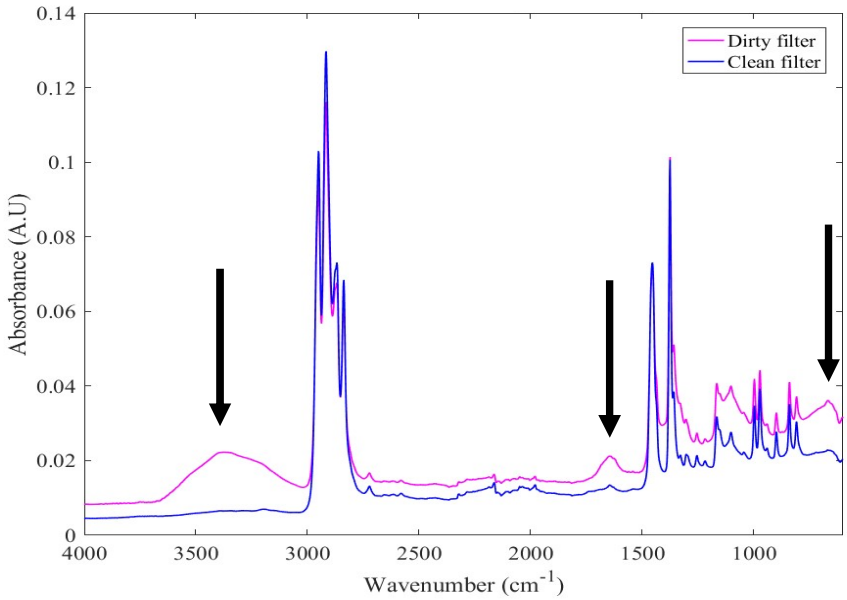

Figure 8: Averaged clean filter and dirty filter spectra acquired with a DiZnSe attachment to show the differing spectral features

This substrate study was key for understanding the spectral feature present within the substrates alone without the added complexity of a sample. A method to remove the spectral contribution can be developed but only once the structure of the substrates is first understood. Two different IREs (Ge and DiZnSe) were trialled in this study to assess the most optimal IRE by comparing the spectra produced by the same substrate from the different IREs. By changing the IRE from Ge to DiZnSe there was an observable change in the intensity of the peaks produced and a slight wavenumber shift. However, generally the same bonds were observed irrespective of IRE. The DiZnSe IRE produced the more intense spectra which were at least an order of magnitude more intense in terms of absorbance, making smaller peaks within the spectrum more prominent. This increased intensity, which highlight the less intense

peaks makes it ideal for future analysis where there are going to be contributing factors that may affect cause small but significant vibrations to be missed. Based on this factor, DiZnSe was chosen as the IRE to use in further analysis.

2.3.2 Raman

As with ATR-FTIR a spectral database has been compiled from each of the different substrates chosen for this study. However, one major difference is the spectra collected using Raman had to be cut prior to baseline corrected to remove regions containing no chemical vibrations which in turn enhance the peaks found within the spectra. The range which the spectra were cut to depended on the substrate. For clean filter, dirty filter, concrete and lino were cut to have the range 70-1600 cm^{-1}. Boot spectra are shown over the range 70-1700 cm^{-1} while jacket metal tile and MirrIR have the range 200-1600 cm^{-1}. An overlay of the spectra collected from 10 different points across a substrate has been produced for all eight substrates. The spectra were then averaged, and the corresponding peak assignments complied into tables per substrate. To give an idea of intra-substrate variability the SD was plotted around the mean for each substrate. For each substrate, an overlay, a mean and a mean with SD cloud have been grouped together with a new one produced for each laser type. As with the ATR-FTIR data a representative selection of spectra and peak assignments presented within this chapter (Figure 9, Figure 10 and Table IV) with the remaining Raman data located in Appendix A: Figures 15-28 and Tables 8-17.

Due to issues with fluorescence, stray light, burning of samples etc. a full spectral library for all substrates using the 532nm laser was not possible. In an attempt to overcome some of these issues a 633nm laser was tested as a comparison to see if a full library could be compiled using a second laser. However, similar problems were encountered with this laser also. For the acquisition, a multipoint map and depth profile were performed for each substrate. Once the Raman spectra had been acquired (using Renishaw InVia system) the spectra were exported from WiRE software into .txt format. The data was then cut and baseline corrected in Labspec software to ensure consistency between the data collected from the different lasers. Averaged spectra and peak assignments for 633nm data collected from boot, clean filter and dirty filter can be seen below in Appendix A: Figure 30-32. Similar fluorescence and interference problems that were encountered with the 532 nm laser were experienced with the 633 nm laser, therefore there are no spectra for the remaining five substrates.

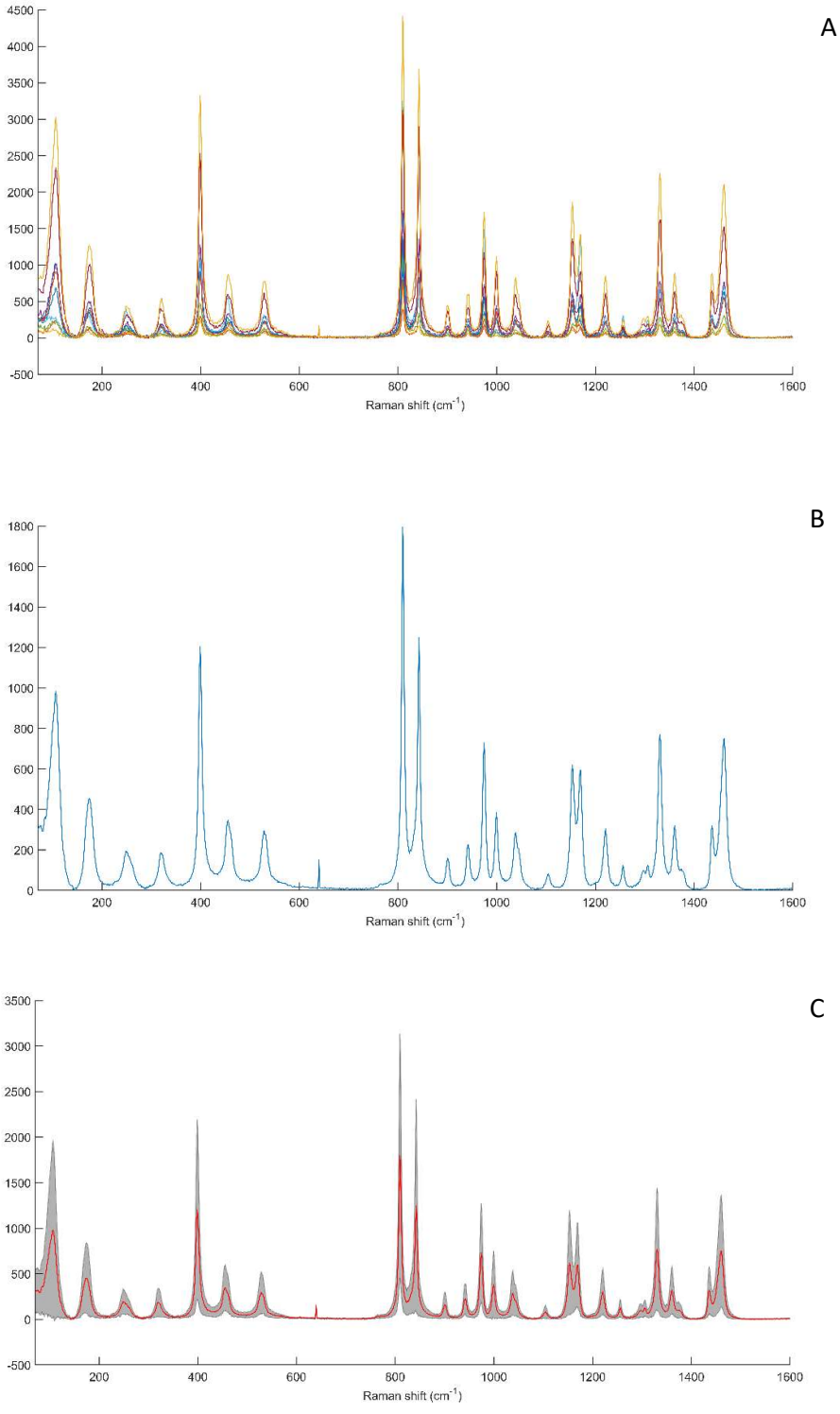

Figure 9: Collection of spectra collected from clean filter acquired using a 785nm laser starting with an overlay of 15 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

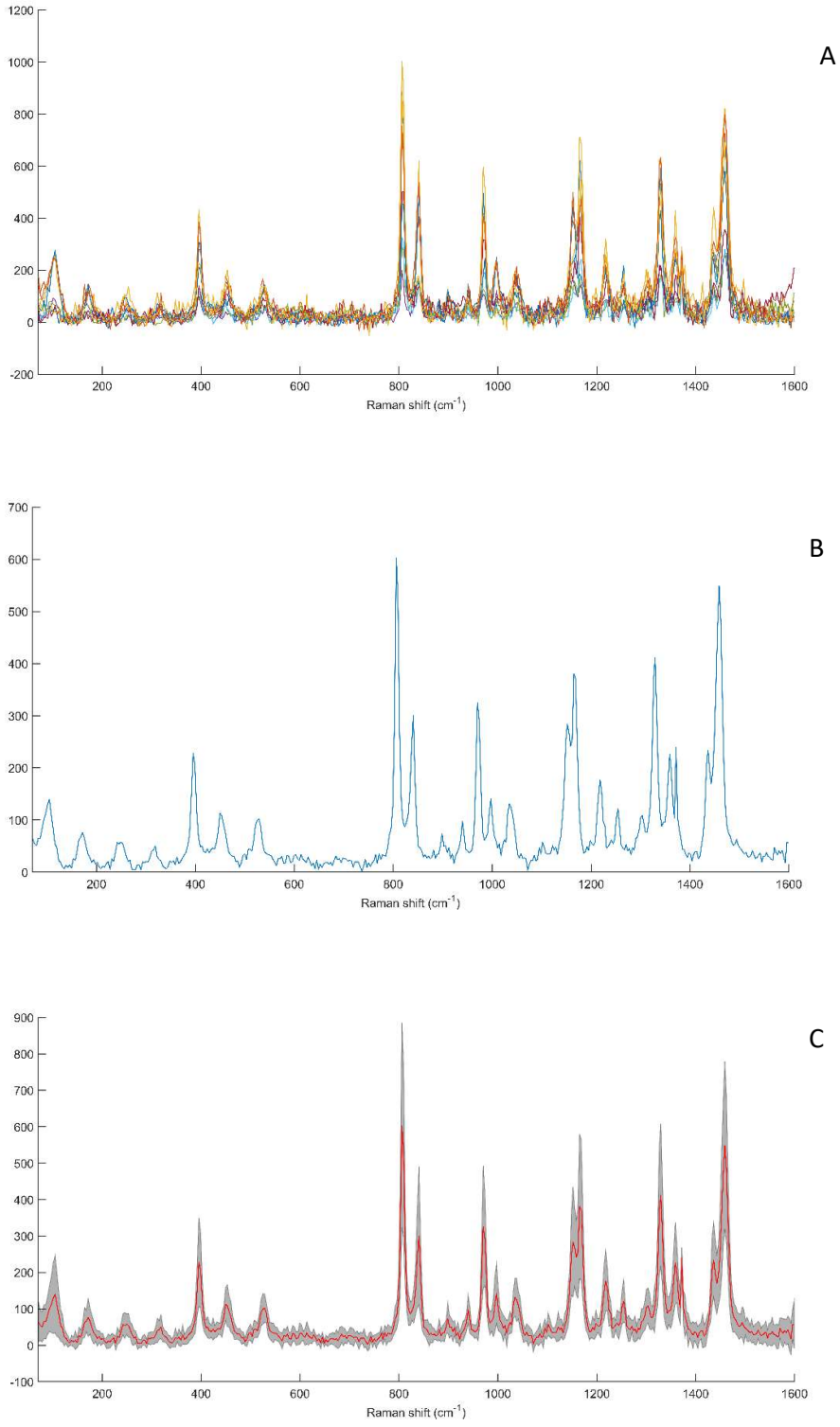

Figure 10: Collection of spectra collected from clean filter acquired using a 532nm laser starting with an overlay of 15 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

Table IV: Wavenumbers and corresponding peak assignment for the average clean filter spectrum collected using a 532 and 785nm laser. [3,4,5,6]

Wavenumber cm^{-1}		Tentative peak assignment
532nm	**785nm**	
105, 172	106,174	Lattice vibrations
249, 319, 395	249, 319, 399	δCC aliphatic chains
450	455	vS-S, vSi-O-Si
528	529	vS-S, vSi-O-Si, vC-I, vC-Br, vC-Cl
806, 899, 940, 970	809, 901, 942, 974	vC-O-C, vCC alicyclic, aliphatic chains
840	843	vC-O-C, vO-O, vCC alicyclic, aliphatic chains
998, 1035	999, 1038, 1044	vC-O-C, vCC alicyclic, aliphatic chains
1152	1105, 1153	vC-O-C asymmetric
1165, 1304	1170, 1306	vCC alicyclic, aliphatic chains
1218, 1228, 1254,	1221, 1256, 1297	Amide III
1330, 1359, 1372	1330, 1361, 1374	$vC\text{-}NO_2$
1436, 1459	1437, 1461	$\delta CH_2/\delta CH_3$ asymmetric

When considering which of the three laser types used within this study is best suited to the substrate set, the ability to compile a full spectral library was taken into consideration. Both the 532nm and 633nm lasers failed to produce spectra for all eight of the substrates used within the study as they were affected by a number of issues such as florescence. However the 785 nm laser was able to produced a full spectral library.

As the 532 and 633nm lasers suffered significantly from florescence, and other interference, it was difficult to get detailed spectra from all of the substrates. However, using a 785nm allowed for a full spectral library to be built for the substrates proving to be the optimal laser. Due to this it was necessary to optimise the parameters for spectral collection to get the most information possible, as a result showing that there is no standard set of parameters that are optimal for use

across all of the substrates. As mentioned in section 2.2.3, the parameters used for each substrate are included in the caption for the corresponding figures found within Appendix A. However, with ATR-FTIR a standard set of parameters can be used across all substrates, for all sample types, to produce detailed and reproducible spectra. This factor makes the technique well suited for this kind of analysis, especially when considering the end goal of the project is to develop a robust and universal method that can be used on any sample/substrate combination.

2.3.3 Substrate Removal Results

Due to the ability to adjust which section and depth of a sample needs to be analysed in Raman spectroscopy it is possible to focus only on the sample onto a substrate therefore removing the need to generate a method of substrate removal. While this property of Raman in theory appears to give it an advantage over FTIR, practically it was found that trying to focus the instrument correctly for each substrate type provided an additional challenge. As ATR-FTIR has a fixed penetration depth there needs to be a way to remove the spectral contribution from a substrate as it is likely this will also contribute to the spectral profile of a sample. As part of the substrate removal study both glucose and ovalbumin were used as surface deposited samples. Glucose and ovalbumin were used as they have relatively simple structures and can both be found within bacteria. Comparatively, ovalbumin has a more complex structure than glucose, therefore glucose was used in the preliminary study with ovalbumin being used in the latter developmental studies.

2.3.3.1.Glucose Study Results

The key vibrations that should be observed when interpreting a glucose spectrum are discussed by Ibrahim *et al*., where the majority of peaks found within the fingerprint region with only lipid vibrations and OH stretches found outside of this region. [1] The vibrations which are discussed by Ibrahim *et al*., are used to assess the quality of spectra produced after the substrate contribution has been removed. A single representative spectrum for each substrate was removed from the spectra collected from each glucose spot using the Beer-Lambert Equation, discussed in the previous section, to produce an absorbance spectrum, totalling ten absorbance spectra per substrate. This was repeated for all substrates with the overlays of the resulting spectra, and corresponding mean spectra shown in Appendix A: Figures 29-35. Figure 11 shows a representative example for absorbance spectra generated from lino. The tentative peak assignments for all substrates can be found in Table V.

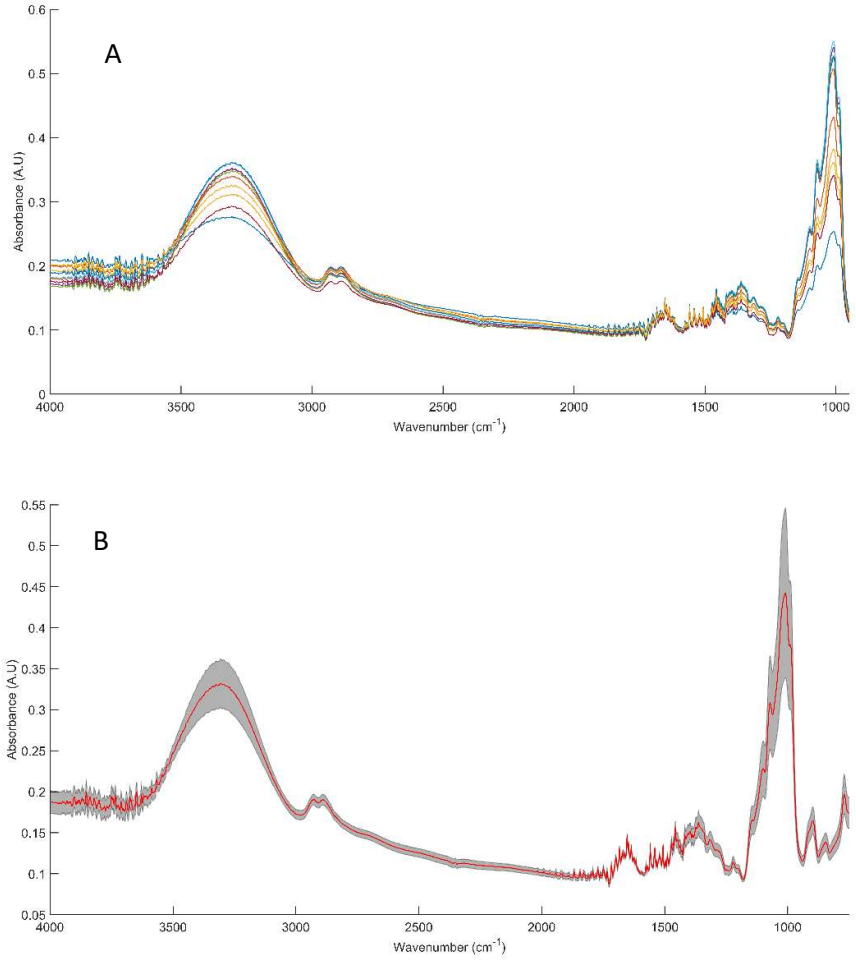

Figure 11: Spectra generated using the Beer-Lambert equation on single channel spectra collected from lino with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

Table V: Wavenumbers and corresponding peak assignment for the glucose spectra generated from all of the substrates focusing on the fingerprint region only [1]

Wavenumber (cm^{-1})								Tentative peak assignment
Boot	**Clean filter**	**Concrete**	**Dirty filter**	**Jacket**	**Lino**	**Metal tile**	**MirrIR slide**	
770	772	772	774	772	770	770	770	δCCO+, δCCH
845, 897	841, 899	868		841, 897	845, 897	843, 897	839	δCH
914				914	914	914		vCO+, vCCH+, v_{asy} ring of pyranose
988	991	990	991	999	988	988	988	vCO+, δCCO
1099	1101	1020, 1102	1028, 1101	1019, 1096	1101	1022, 1099	1103	vCO
1009, 1071, 1144	1032, 1078, 1144	1072, 1144	1074, 1148	1007, 1040, 1074, 1144	1009, 1071, 1144	1009, 1071, 1146	1011, 1074, 1148	vCO+, vCC
1198, 1221		1198	1206	1223	1202, 1221	1202, 1230	1200	vCH+, δOH in plane
1276								δOH in plane, δCH
1318 1341	1318				1314 1341	1320	1317 1339	δCCH, δOCH
1362, 1418		1373	1375		1362, 1418	1362, 1418	1362, 1418	δOCH+, δCOH+, δCCH+
1456				1456	1456	1456	1456	δCH_2+, δOCH+, δCCH
		1539		1539, 1559	1539, 1559			Amide II

This study has shown that it is possible to remove spectral contributions from the substrate from a spectrum to produce an absorbance spectrum that has vibrations characteristic of glucose. However, the spectra show there are problems with how well the method works depending on the substrate that is being removed, with generally the more porous of the substrates performing worse. This is likely to

be as a result of the sample no longer solely sitting on the surface, with the bulk of the sample found outside of the penetration depth of the IR beam.

The following grading process, based on Table I, was performed on each of the eight substrates used in the study. The ten spectra collected per substrate were individually graded with an overall grade attributed to the substrate, the results of this can be seen in Table VI.

Table VI: Assessment of the performance of the Beer-Lambert substrate removal technique for analysing glucose on substrates. The table provides the number of spectra achieving each grade for that substrate before giving the substrate an overall rating.

	Very good	Good	Adequate	Poor	Very poor	Overall grade
Boot	0	0	10	0	0	Adequate
Clean filter	0	0	4	0	6	Very poor
Concrete	4	0	0	0	6	Very Poor
Dirty filter	0	0	0	0	10	Very poor
Jacket	10	0	0	0	0	Very good
Lino	10	0	0	0	0	Very good
Metal tile	0	0	10	0	0	Adequate
MirrIR	10	0	0	0	0	Very good

Based on this grading system for glucose found on the substrate collection, three substrates scored a 'Very good' rating, two scored an 'Adequate' rating with the final three scoring a 'Very poor' rating. Overall the Beer-Lambert method of

substrate removal worked well with 72.5% of the individual spectra scoring a rating of 'Adequate' or higher. Despite this the spectra for the more porous substrates like clean filter, concrete and dirty filter (Appendix A: Figures 30, 31 and 32) show that there is some intra-substrate variation.

When considering the relative standard deviation (RSD) across the glucose spectra generated using this method, it can be seen that the RSD is significant in some of the substrates. The large RSD seen in the spectra from these substrates is likely to be down to distortions within the baseline. The areas of highest and lowest RSD across all of the substrates is shown in Table VII.

Table VII Table detailing the highest and lowest relative standard deviations (RSDs), with the correlating wavenumbers, for the glucose spectra generated from all of the substrates

Highest RSD		Substrate	Lowest RSD	
Wavenumber	RSD (%)		Wavenumber	RSD (%)
1631.83	103.56	**Boot**	1751.42	2.59
1406.15	72202.20	**Clean filter**	1558.54	20.42
1381.08	41.82	**Concrete**	1560.46	1.86
1217.12	92170.00	**Dirty filter**	1026.16	35.28
962.51	20.28	**Jacket**	1220.98	2.62
1010.73	23.35	**Lino**	1506.46	1.32
981.80	12162.20	**MirrIR**	1541.18	4.14
952.87	45.29	**Tile**	3088.14	9.10

2.3.3.2. Ovalbumin Study

In an attempt to overcome some of the issues discussed previously with regards to the removal of the substrate contribution an alternative method was trailed. The Beer-Lambert method requires the unprocessed spectra to be converted and exported from the instrument software before being imported into MATLAB for processing. The second method trailed kept the processing in the software. This method was chosen to keep the methodology as simple as possible. This is important when considering the overall end goal of the project in developing a methodology that can be used for *in situ* analysis of samples on handheld instrumentation. This would make the method more user friendly, reduce the equipment required and speed up the overall process of analysis. The performance of this method is evaluated in Table XII in the same way that the spectra generated using the Beer-Lambert method are evaluated in Table IX, both based on the quality of the spectra produced. To quantitatively assess the absorbance spectra generated for the substrate removal for each method mean and standard deviation were calculated and are shown with the overlay for each substrate for each method. The key region of the spectrum assessed for the ovalbumin data is the fingerprint region (900-1800cm^{-1}) as identified by Velusamy. [2]

While only a representative set of spectra for both the Beer-Lambert (Figure 12) and software-based removal method (Figure 13) are shown within this chapter, spectra from all substrates across both studies are discussed. All other spectra

generated using the Beer-Lambert method (method 1) can be seen in Appendix A: Figures 36-42 with the corresponding peak assignments shown in Table VIII.

The spectra generated from the software-based removal method (method 2) used in this study can be seen here in Appendix A: Figures 43-49 with the corresponding peak assignments shown in Table XI.

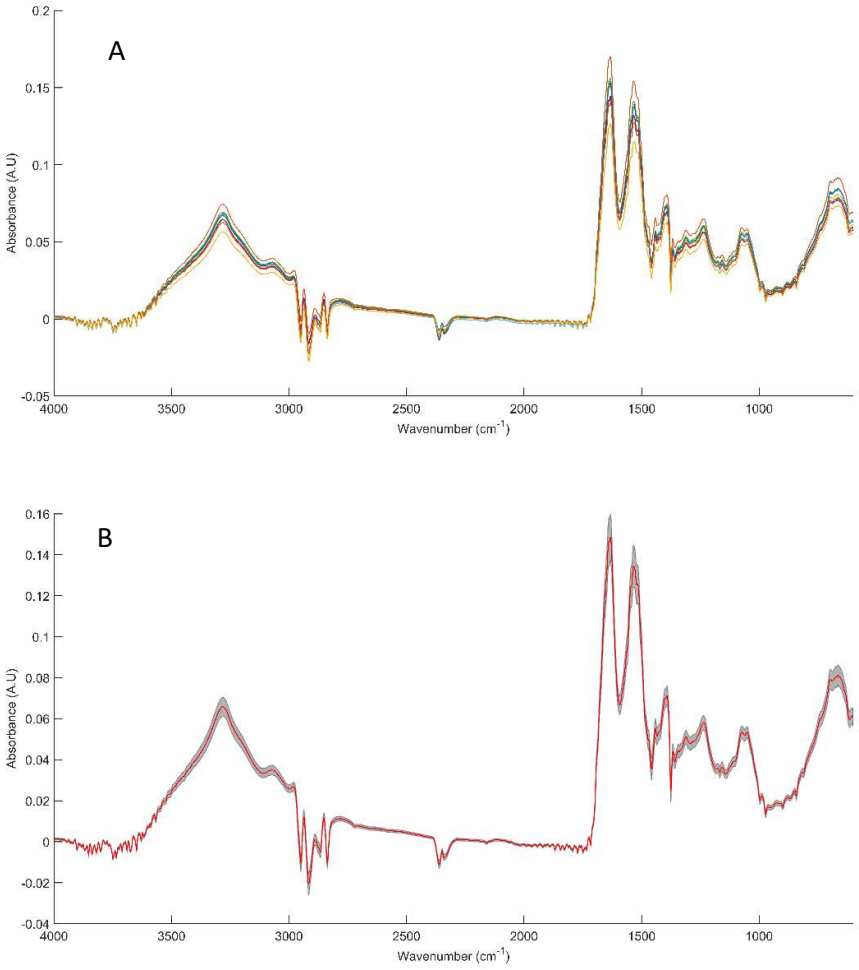

Figure 12 Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from clean filter with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

Table VIII: Wavenumbers and corresponding peak assignment for the ovalbumin spectra generated from all of the substrates focusing on the fingerprint region only for the Beer-Lambert method of substrate removal. [3,4,5,6]

Wavenumber (cm^{-1})								Tentative peak assignment
Boot	**Clean filter**	**Concrete**	**Dirty filter**	**Jacket**	**Lino**	**Metal tile**	**MirrIR slide**	
700, 743, 779	663, 698	604, 667	648, 669, 698, 744		631, 663, 716, 738	621, 700	646	Out of plane NH, OCN bend
887					885			C-O-C ring
1055, 1070	989, 1051, 1074, 1111	1074	1049, 1074,		1053, 1072, 1113	1080		CO, CC, COH and COC deformation
	1157	1155	1159		1126, 1155	1155		CN and CC stretch
	1236	1232	1236		1221, 1288	1232	1243	Amide III
1309	1311		1309		1306	1313	1309	Amide II
	1344		1338		1348			CN, CO
	1367		1365					C-H bend
1390	1392		1396		1394	1392	1390	C=O stretch of COO^-
	1441		1439, 1473		1451	1444	1444	CH deformation
1529	1535	1535	1541	1549	1531	1531	1529	Amide II
1631	1632	1631	1635	1641	1633	1631	1631	Amide I β sheet structure

Table IX: Assessment of the performance of the Beer-Lambert based substrate removal technique for analysing ovalbumin on substrates. The table provides the number of spectra achieving each grade for that substrate before giving the substrate an overall rating.

	Very good	Good	Adequate	Poor	Very poor	Overall grade
Boot	0	0	0	0	10	Very poor
Clean filter	0	10	0	0	0	Good
Concrete	0	10	0	0	0	Good
Dirty filter	0	10	0	0	0	Good
Jacket	0	0	0	0	10	Very poor
Lino	0	0	10	0	0	Adequate
Metal tile	0	0	0	0	10	Very poor
MirrIR	0	10	0	0	0	Good

Based on the grading system for ovalbumin found on the substrate collection, four substrates scored a 'Good' rating, one scored an 'Adequate' rating with the final three scoring a 'Very poor' rating, as shown in Table IX. Overall the Beer-Lambert method of substrate removal worked well with 62.5% of the individual spectra scored a rating of 'Adequate' or higher. In comparison to the glucose study that used this method it performed better with the more porous substrates that previously showed a large amount of intra-substrate variation, such as the concrete, clean filter and dirty filter all achieving a good rating.

When considering the relative standard deviation (RSD) across the ovalbumin spectra generated using the manual Beer-Lambert removal method, it can be seen that the largest RSD for all of the substrates, is over 100%. In most cases the highest amount of RSD at a given wavenumber is in the tens of thousands. These vast

numbers are caused as a result of very low absorbances within the mean spectrum, often close to zero, with large SDs. The cause of such low absorbances may be as a result of a changing baseline across the 10 ovalbumin spectra that also gets averaged when generating a mean spectrum. When considering the lowest amount of RSD, no substrate has higher than 1.5% RSD (boot), with three substrates having less than 1% RSD at a given wavenumber. The wavenumber with the lowest RSD for ovalbumin spectra produced with Method 1 is the same in six out of the eight different substrates. Both the significant reduction in the lowest RSD values, and the increased consistency of the corresponding wavenumbers, show an improvement on the data generated using this method on glucose spectra. However, the areas that have the highest RSD show a significantly higher amount of RSD when compared to glucose. The areas of highest and lowest RSD across all of the substrates is shown in Table X.

Table X Table detailing the highest and lowest relative standard deviations (RSDs), with the correlating wavenumbers, for the ovalbumin spectra generated from all of the substrates using the manual Beer-lambert method (Method 1)

Highest RSD		Substrate	Lowest RSD	
Wavenumber	RSD (%)		Wavenumber	RSD (%)
3027.68	29351.55	Boot	3748.92	1.50
2967.89	34195.40	Clean filter	3748.92	0.67
1504.12	8754.06	Concrete	3748.92	1.13
3268.73	38898.01	Dirty filter	3748.92	1.32
2829.05	3473.36	Jacket	1698.97	1.40
1907.24	4168.90	Lino	3748.92	0.74
3511.72	9956.64	MirrIR	3735.42	0.68
1475.27	12265.78	Tile	3748.92	1.21

The first method to manually remove the spectral features caused by the substrate used the Beer-Lambert equation to manually subtract these features from the absorbance spectra. This method was used with both glucose and ovalbumin. The spectra produced using this method were graded using the system detailed in Table I and Table II. From this for glucose 72.5% of spectra scored an 'Adequate' or better rating with ovalbumin scoring slightly lower at 62.5%. This showed that the method worked well with most of the spectra containing the key spectral features. The less absorbent substrates, specifically boot, lino and metal tile, tend to produce spectra with a higher SNR. However, the Beer-Lambert method appears to be less reproducible when working with absorbent substrates such as the filters, concrete and jacket, producing less intense spectra that are more difficult to interpret and spectra also being heavily affected by water vapour. MirrIR, a non-absorbent slide generally produces smooth spectra but has a large interference below 1000cm^{-1}.

A substrate-based method of removal was used to assess if the methodology could be made simpler with the overall end goal of the project in mind. Removing the step of having to transfer data from the instrument software to MATLAB to manually perform a ratio step increases the time taken for sample identification and increases the amount of computer power needed. By performing the removal in the instrument software, the processing speed increases with the spectra generated able to be immediately projected into a pre-built model allowing for the identification of the sample.

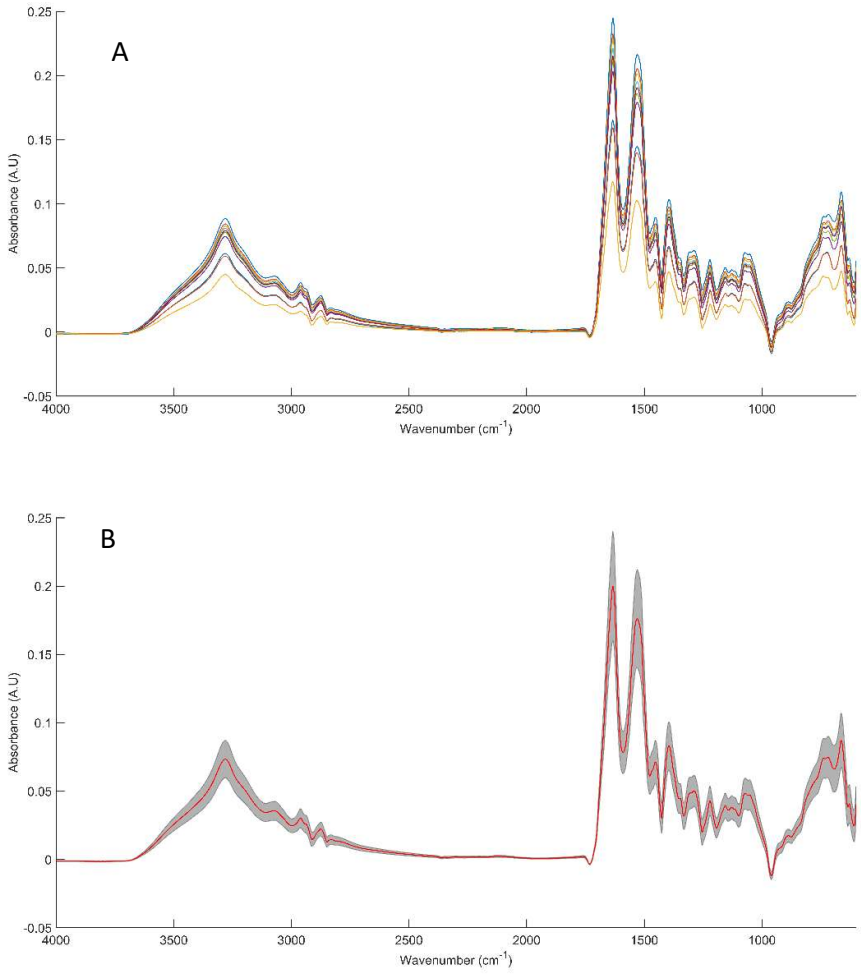

Figure 13: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from lino with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

For each substrate only the positive peaks, within the fingerprint region, have been assigned wavenumbers and a tentative peak assignment shown in Table XI below. Wavenumbers and peaks have been assigned to the single MirrIR spectrum, yellow line in Appendix A: Figure 45A, that shows positive peaks.

Table XI: Wavenumbers and corresponding peak assignment for the ovalbumin spectra generated from all of the substrates focusing on the fingerprint region only for the software-based method of substrate removal. [3,4,5,6]

Wavenumber (cm^{-1})								Tentative peak assignment
Boot	**Clean filter**	**Concrete**	**Dirty filter**	**Jacket**	**Lino**	**Metal tile**	**MirrIR slide**	
	669	660	669, 698		663, 738	662, 679, 700	656	Out of plane NH, OCN bend
					889			C-O-C ring
	1054, 1111	1078	1073, 1109		931, 1053, 1072, 1113	1080	1074	CO, CC, COH and COC deformation
	1163		1163		1126, 1155	1155	1155	CN and CC stretch
	1236	1232	1236		1223, 1288	1232	1244	Amide III
	1315		1311		1306	1313	1313	Amide II
	1340		1336		1348			N-C, C-O
	1365		1365					C-H bend
	1395		1394		1394	1394	1398	C=O stretch of COO^-
	1439, 1473		1441, 1473		1452	1448	1450	CH deformation
1516	1539, 1572	1545	1539	1543	1529	1519, 1533	1533	Amide II
1630	1639	1631	1639	1649	1633	1633	1636	Amide I β sheet structure

Table XII: Assessment of the performance of the software-based substrate removal technique for analysing ovalbumin on substrates. The table provides the number of spectra achieving each grade for that substrate before giving the substrate an overall rating.

	Very good	Good	Adequate	Poor	Very poor	Overall grade
Boot	0	0	0	0	10	Very poor
Clean filter	0	0	0	10	0	Poor
Concrete	1	0	0	0	9	Very Poor
Dirty filter	0	0	0	10	0	Poor
Jacket	0	0	0	10	0	Poor
Lino	0	0	10	0	0	Adequate
Metal tile	0	0	0	0	10	Very Poor
MirrIR	1	0	0	0	9	Very Poor

Based on the grading system for ovalbumin found on the substrate collection, one substrate scored an 'Adequate' rating, three scored a 'Poor' rating with the remaining four substrates scoring a 'Very Poor' rating, as shown in Table XII. Overall the software-based method of substrate removal worked very poorly as only 15% of the individual spectra scored a rating of 'Adequate' or higher with almost half of individual spectra scoring a 'Very poor' rating. In comparison with the Beer-Lambert based method of removal, the quality of the absorbance spectra produced using this method are of a poorer quality. This could be due to the fact that the reference background spectrum for the non-contaminated substrate is not an average of multiple spots across the substrate as with the Beer-lambert method.

When considering the relative standard deviation (RSD) across the ovalbumin spectra generated using the software-based method, it can again be seen that with all of the substrates, there is a significant RSD when considering the wavenumber with the highest RSD. When considering the areas of lowest RSD in comparison to the manual Beer-Lambert based method (Method 1), the RSD is much larger than found in Method 1. The areas of highest and lowest RSD for the absorbance spectra generated with Method 2 can be seen in Table XIIITable VII.

Table XIII Table detailing the highest and lowest relative standard deviations (RSDs), with the correlating wavenumbers, for the ovalbumin spectra generated from all of the substrates using the software-based remoaval method (Method 2)

Highest RSD		Substrate	Lowest RSD	
Wavenumber	RSD (%)		Wavenumber	RSD (%)
1500.34	**1519.86**	**Boot**	2856.04	4.45
2973.68	**11476.40**	**Clean filter**	1700.90	21.21
2163.73	**1201.47**	**Concrete**	1816.61	53.39
993.15	**8355.09**	**Dirty filter**	1637.26	43.68
1616.05	**1871.46**	**Jacket**	1477.20	3.51
3671.78	**1724.60**	**Lino**	3808.70	15.86
2316.08	**730.58**	**MirrIR**	1706.68	39.53
997.01	**6185.21**	**Tile**	1617.97	18.86

From the results presented in both Table IX and Table XII, this it can be seen that the software-based method did not perform as well as the Beer-Lambert method did when generating absorbance spectra for ovalbumin with only 15% of the individual spectra scoring an ‘Adequate’ rating or higher. A comparison of both of the methods used to generate absorbance spectra for ovalbumin can be seen in Table XIV. This shows that the Beer-Lambert method outperformed the software-based method for half of the eight substrates (clean filter, concrete, dirty filter and a MirrIR slide). The software-based method of removal only outperformed the Beer-Lambert

method with one of the eight substrates (jacket), with the methods performing to the same level for the boot, lino and metal tile substrates.

2.3.3.3.Comparison of Substrate Removal Methods

Two methods of substrate removal have been explored, both using ovalbumin, however, only one of these methods can be taken forward for use in further studies. As shown in Table VI, Table IX and Table XII the individual spectra have been graded to assess how well the method of removal worked. Table VI and Table IX show the results from the Beer-Lambert method for glucose and ovalbumin respectively while Table XII details the software-based method results with ovalbumin. Following this a direct comparison of the results generated from ovalbumin deposited on the substrates can be seen in Table XIV. The table compares the results from the Beer-Lambert method and the software-based method before concluding the optimal method based on the quality of the absorbance spectra produced.

Table XIV: Assessment of the performance of the Beer-Lambert and Software based substrate removal techniques for ovalbumin. The table provides the overall grade for each substrate with the final column detailing the best method for each substrate.

	Overall substrate grade		Overall method performance
	Beer-Lambert method	**Software based method**	
Boot	Very poor	Very poor	Equal
Clean filter	Good	Poor	Beer-Lambert method
Concrete	Good	Very Poor	Beer-Lambert method
Dirty filter	Good	Poor	Beer-Lambert method
Jacket	Very poor	Poor	Software based method
Lino	Adequate	Adequate	Equal
Metal tile	Very poor	Very Poor	Equal
MirrIR	Good	Very Poor	Beer-Lambert method

Based on the results shown in Table XIV, along with the spectra presented within this chapter and Appendix A, it is clear that using the Beer-Lambert based method is the more most optimal. Therefore, the Beer-Lambert method was the method chosen to remove unwanted spectral features resulting from the substrate. Despite the fact that the Beer-Lambert method has a slightly longer methodology due to the multi-step analysis it is still a fast method of substrate removal, producing absorbance spectra, of a good quality within 10 minutes of beginning analysis.

2.4 Conclusions

The substrate study was an important building block in understanding the spectral signature that is representative of each substrate prior to further analysis that will involve analysing surface deposited bacteria on these surfaces.

From the ATR-FTIR analysis, the type of IRE used in the ATR accessory (Ge or DiZnSe) only seemed to affect the intensity of the spectral features with the DiZnSe IRE often producing spectra that are more than twice as intense as those acquired using the Ge IRE. For this reason, DiZnSe was chosen as the accessory to carry forward for future analysis to ensure that minor peaks and changes that could be used for discrimination are detected.

Raman analysis of the substrate set showed significant changes to the spectral fingerprint when using different lasers. Using the 532 nm and 633 nm lasers fluorescence often dominated the spectra, masking peaks produced by the substrates, making it difficult to compile complete substrate libraries for these lasers. Due to the existing fluorescence profile, spectra acquired with a 532 nm or a 633 nm laser are heavily affected while spectra acquired with a 785nm laser are only marginally affected by fluorescence. [4] As spectra produced using the 785 nm often less affected by fluorescence, more detailed spectra are produced and therefore a complete set of reference substrate spectra were obtained for a substrate library. Due to the issues with fluorescence the 532 nm and 633 nm lasers were ruled out for

further analysis, leaving the 785 nm laser the optimal laser to be used for bacterial identification studies.

The proof of principle work that has been carried to assess the optimal substrate removal procedure for use with ATR-FTIR has considered two different approaches. The study allowed for the development of a robust and adaptable methodology that has the ability to be transferred to handheld instrumentation to analyse complex sample types. The study has shown that the optimal method for removing substrate contribution was to use manually remove the substrate using the Beer-Lambert equation.

2.5 References

1. Horiba, What factors affect spectral resolution in a Raman spectrometer?, http://www.horiba.com/us/en/scientific/products/raman-spectroscopy/raman-academy/raman-faqs/what-factors-affect-spectral-resolution-in-a-raman-spectrometer/, (accessed 10/05/2018, 2018).
2. M. Ibrahim, M. Alaam, H. El-Haes, A. F. Jalbout and A. de Leon, *Ecletica Quimica*, 2006, **31**, 15-21.
3. V. Velusamy, PhD, Manonmaniam Sundaranar University, 2012.
4. H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone and F. L. Martin *Nature Protocols*, 2016, **11**, 664-687.

Chapter 3

Downselection of Spectroscopic Technique for Bacterial Identification: FTIR vs Raman

3.1. Introduction and Aims

The main focus of this study was to conclude the optimal technique for the spectroscopic analysis of bacteria. This chapter discusses the analysis that has been conducted using vibrational spectroscopy and the ability for it to be used as a tool for bacterial identification and classification. This study aims to build on previous knowledge of the optimal modality for bacterial classification of surface deposited samples to identify the optimal vibrational technique. By evolving the methodology to identify more chemically complex samples compared to glucose and ovalbumin the work is developed towards the ultimate end goal of developing a method for use with handheld instruments for *in-situ* analysis of samples.

This chapter is split into three main sections with the first discussing bacterial identification of samples prepared on a MirrIR slide (Kevley Technologies, Ohio, USA) and analysed using ATR-FTIR and Raman spectroscopy. Understanding the characteristics and behaviour of the bacterium chosen for use in this project is important for developing an appropriate experimental procedure that can be applied to each individual study. The procedure tries to replicate the non-ideal setting that would be used by an untrained microbiologist preparing samples to use in warfare.

This chapter looks at how each bacterium grows and discusses the standard sample preparation method developed for use within the project, for example growing samples on a general nutrient agar rather than specialist agars. The chapter then moves on to answer one of the main aims of the overall project by assessing which spectroscopic technique performs best. Based on the findings from the previous chapter the DiZnSe IRE was used with ATR-FTIR and a 785 nm laser chosen for Raman analysis, section 3.5.

The chapter then moves on to discuss the results that were obtained during this study. Discussion follows of the processing method used on the data. The order of pre-processing was assessed before deciding which order of derivatisation is optimal for the dataset. The results generated during this process are discussed prior to concluding the optimal order of processing.

The chapter concludes by discussing the findings from the main and secondary studies detailing the optimum technique and methodology that was carried forward for future analysis of environmentally conditioned samples.

3.2. Materials and Methods

3.2.1. Bacterial Culture

Seven different bacterial strains, supplied by the Defence Science and Technology Laboratories (DSTL), were used in this study. The strains are shown in Table I and represent a combination of Gram-positive and Gram-negative bacteria

that are often used to represent Biological Warfare Agents (BWAs). The *Bacillus thuringiensis var. kurstaki* (BTK) provided was isolated from Dipel insecticide. *Bacillus thuringiensis var. kurstaki* HD-1 cry- (BTK-) is a new strain produced at DSTL in 2014 which has lost the insecticidal genes. *Bacillus thuringiensis var. kurstaki* HD-1 cry+ (BTK+) is the parent strain used to produce the cry- strain. The bacteria were cultured on Nutrient Agar plates (Agar pre-mixed nutrient agar powder from Oxoid, 'Lab-Lemco' powder 1 gl^{-1}; Yeast extract 2 gl^{-1}; Peptone 5 gl^{-1}; Sodium chloride 5 gl^{-1}; Agar 15 gl^{-1}, Petri dished from BRAND (TM)) filled with 25 ml and incubated for 24 hours at either 30 or 37°C (see Table I). No specialist agars or broths were used for growth during this project.

Table I: The bacterial strains used during this project with the American Type Culture Collection (ATCC) notation, where available. The table also includes properties such as Gram type, if the bacterium has the ability to form spores, incubation temperature and the abbreviations they are referred to as throughout this thesis

Bacterial Strain	Gram Type	Forms Spores	Incubation Temperature (°C)
Bacillus atrophaeus **(B.Atrop)**	+	Yes	37
Bacillus thuringiensis **(BT) ATCC 29730**	+	Yes	37
Bacillus thuringiensis var. kurstaki **(BTK)**	+	Yes	37
Bacillus thuringiensis var. kurstaki HD-1 cry- **(BTK-)**	+	Yes	37
Bacillus thuringiensis var. kurstaki HD-1 cry+ **(BTK+)**	+	Yes	37
Eschericha coli MRE 162 **(*E. coli*)**	-	No	37
Pseudomonas fluorescens **(PF) ATCC 13525**	-	No	30

3.2.2. Streak Plating

Streak plating [1] was used for growing the bacteria and yields enough biomass to create the concentrations desired for use within this project. A single 10µL

inoculation loop (Greiner) of bacteria from the stock solution was added to an agar plate and spread using the streak plating method shown in Figure 1.

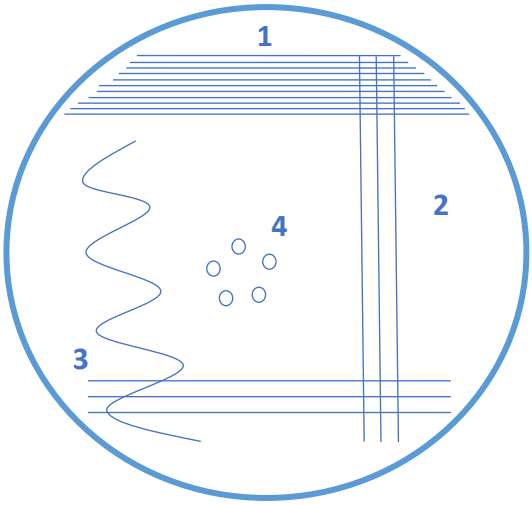

Figure 1: Showing the four different stages involved in the 'streak plating' technique. The first stage is the application of the bacteria to the agar plate spreading it across the top section (1), the bacteria then has three streaks going across it using a loop (2). The loop is then slipped before creating more streaks through the bacteria (3) before the final step of creating a wave through the bacteria along with some individual spots (4).

To begin preparing samples a loop containing a small amount of bacterial biomass is spread across the top ¼ of the plate, using a side to side motion, to spread the biomass out evenly across the agar. The plates were then rotated 90° anti-clockwise before three streaks were made horizontally across the plate starting from the left-hand side of the plate where the bacteria had been streaked. The plates were then rotated a further 90° anticlockwise. At this point the plastic loop was flipped so that the opposite side can be used to spread the bacteria. Flipping the loop between the second and third stage enhances the chance of forming single cell colonies as there is no biomass on that side of the loop. A further three streaks were made starting again at the left-hand side of the plate finishing at the right. The plates were rotated a final 90° anti-clockwise before a squiggle is made from the left to the

right of the plate through the bacteria. A few dots were made in the centre of the plates to enhance the chance of creating a single cell colony.

The plates are then placed down on the top of the lid so that the agar was upside down to avoid contamination from condensation found on the plate lid. The plates are then taped with micropore tape and are ready to be stored in an incubator set to the desired temperature for bacterial growth. A control plate was also incubated along with the prepared plates, made from the same batch of agar, containing no bacteria, as a further check for contamination as no growth should be observed on the control plate. The plates were incubated for 24 hours to allow for growth of the bacteria to ensure that there was sufficient biomass that could be collected and prepared for analysis.

3.2.3. Optical Density Calculations

Typically for most bacterial strains, harvesting all the biomass from a single plate that had been incubated for 24 hours provided a starting concentration that was too concentrated for detection with the spectrophotometer, as ideally the concentration should fall within the linear range of 0.3-0.9 absorbance units (abs). This meant that the bacterial solution was diluted by a factor of ×10 or ×20 to give a concentration that was detectable. Measurements from the spectrophometer were collected at 600 nm as the light is scattered by the cells, not absorbed.[2] The optical density (OD) of a solution correlates to the total concentration of the sample, taking into account both living and dead cells. To produce a concentration within the ideal

range for analysis with the spectrophotometer a ×20 dilution was required. To produce a 1 ml sample, 50 µl cell solution and 950 µl of 0.9% NaCl (Acros Organics) were combined and mixed in an Eppendorf. The sample was vortexed to ensure that it was homogenous, transferred to a 1ml cuvette and analysed with a spectrophotometer. Prior to the sample being analysed, 1ml of 0.9% NaCl was analysed with the spectrophotometer to give a reading of 0, demonstrating the instrument was correctly calibrated.

A ×20 dilution produced a reading that is 1/20th of the actual concentration of the solution. So, for example an OD reading of 0.1 collected from a ×20 dilution means that the starting OD of the cell solution was 2. Once this information was acquired the samples were made up to the desired concentration or OD. Eq 4.1, Eq 4.2 and Eq 4.3 show the process that was used to make up these solutions.

$$\boldsymbol{Dilution\ factor} = \frac{\boldsymbol{Actual\ OD\ (Abs)}}{\boldsymbol{Desired\ OD\ (Abs)}} \quad (4.1)$$

$$\boldsymbol{Volume\ of\ sample\ required}\ (\mu\boldsymbol{l}) = \frac{\mathbf{1000}}{\boldsymbol{Dilution\ factor}} \quad (4.2)$$

$$\boldsymbol{Volume\ of}\ \mathbf{0.9\%}\ \boldsymbol{NaCl\ required}\ (\mu\boldsymbol{l}) = \mathbf{1000}\mu\boldsymbol{l} - \boldsymbol{Volume\ of\ sample\ required}(\mu\boldsymbol{l}) \quad (4.3)$$

3.2.4. Growth Curve Analysis

Growth curve calculations were carried out on all bacterial strains used within this project to understand how each individual bacterium grows. For this analysis, a

100 well honeycomb plate (Steri) is used with samples pipetted into each of the wells (Figure 2). For each new bacterial strain, the first well of each column is filled with 200 µl of nutrient broth to act as a control. 200 µl of bacterial sample, which has been prepared to a standard OD of 0.1, was then pipetted into each of the 5 wells below the control on the honeycomb plate. The other wells below this were filled with the sample bacterial strain. For each bacterium investigated a total of six wells were filled, one with nutrient broth and five with the bacterium. The remaining four wells of the column were left empty. Figure 2 shows a typical Bioscreen C plate set up where seven bacterial strains are being examined.

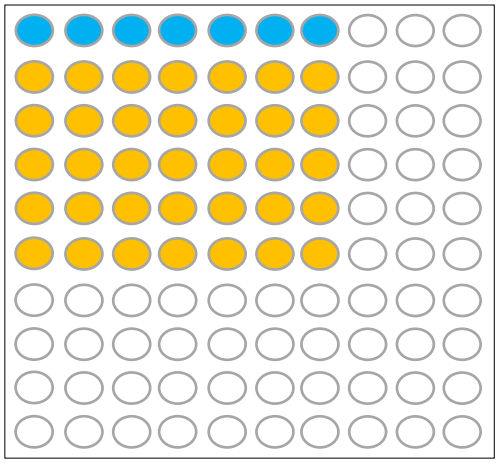

Figure 2: Set up for Bioscreen C honeycomb plate where filled orange circles indicate wells filled with bacterial samples while filled blue circles indicate wells filled with nutrient broth, white circles indicate empty wells

The parameters used with the Bioscreen C allowed for growth data to be collected over a 48-hour period, after an initial five minutes of preheating with OD values recorded every ten minutes. There was constant shaking of the plate, set up on the Bioscreen C instrument as part of the methodology, to avoid the formation of a biofilm on the samples. Once the data were acquired, the five samples were

averaged to provide one growth curve per bacterial strain to indicate the length of time the strains need to complete all stages of development.

3.2.5. Sample Preparation

The seven bacterial strains supplied by DSTL shown in Table I were all prepared in the same manner once they had been grown at their optimal temperatures. Three separate plates were cultured per bacterial species to show the level of reproducibility between culture cycles. As such each plate is referred to as a 'batch'. Preparing samples in this way allows for an assessment of the batch-to-batch variation. The following method was applied to each plate for preparing the samples: Bacteria were collected using plastic loops and suspended in 1ml of 0.9% NaCl. The cells were then 'washed' by spinning the samples in a centrifuge set to 5000g for five minutes at 4°C. The supernatant was then removed leaving a pellet of bacterial biomass, before the cells were re-suspended in 1ml of 0.9% NaCl solution. Once the NaCl was added the samples were vortexed to ensure the samples were homogenous prior to spotting. Ten 2µL spots per bacterium were spotted onto a low-e slide per batch (Figure 3) and air dried before being stored in a petri dish in the fridge at 2-8°C prior to analysis. Samples were prepared using the same method for both Raman and IR analysis. For Raman analysis data was only collected from two of the three batches due to limited access to the Raman instrumentation. However, based on the analysis of the ATR-FTIR dataset there is no evidence of batch-to-batch variation observed across the three batches in the PCA scores plots shown in Figure 15A. Therefore,

analysis of all three batches simply for batch variance was not needed with Raman. Also due to the limited access to the Raman instrument based at UCLan, analysis of samples was performed over a number of months.

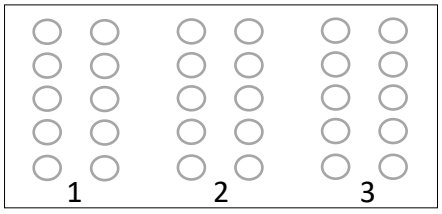

Figure 3: Schematic of bacterial samples deposited onto a MirrIR slide. The number denote the different batch that the sample belongs to.

3.2.6. Data Pre-Processing and Analysis

After the acquisition of spectral data, the data undergo several pre-processing steps before analysis. Pre-processing the data allowed for background correction and the removal of any artefacts, such as cosmic rays in Raman, or distortions caused by water vapour in FTIR. This ensures that the differences detected using multivariate analysis (MVA) were genuine changes arising from biochemical differences between samples. The following section discusses the pre-processing steps and MVA that were performed on the data. The section first discusses the pre-processing and analysis methods for ATR-FTIR data and is followed by Raman pre-processing and analysis.

3.2.6.1. FTIR Pre-Processing and Analysis

For all FTIR data collected during the bacterial identification study the same set of pre-processing parameters were used with the data. Following this, data were analysed using both unsupervised and supervised MVA methods in the form of PCA and DFA respectively. The set of steps carried out for processing FTIR data once it has been acquired can be seen in Figure 4.

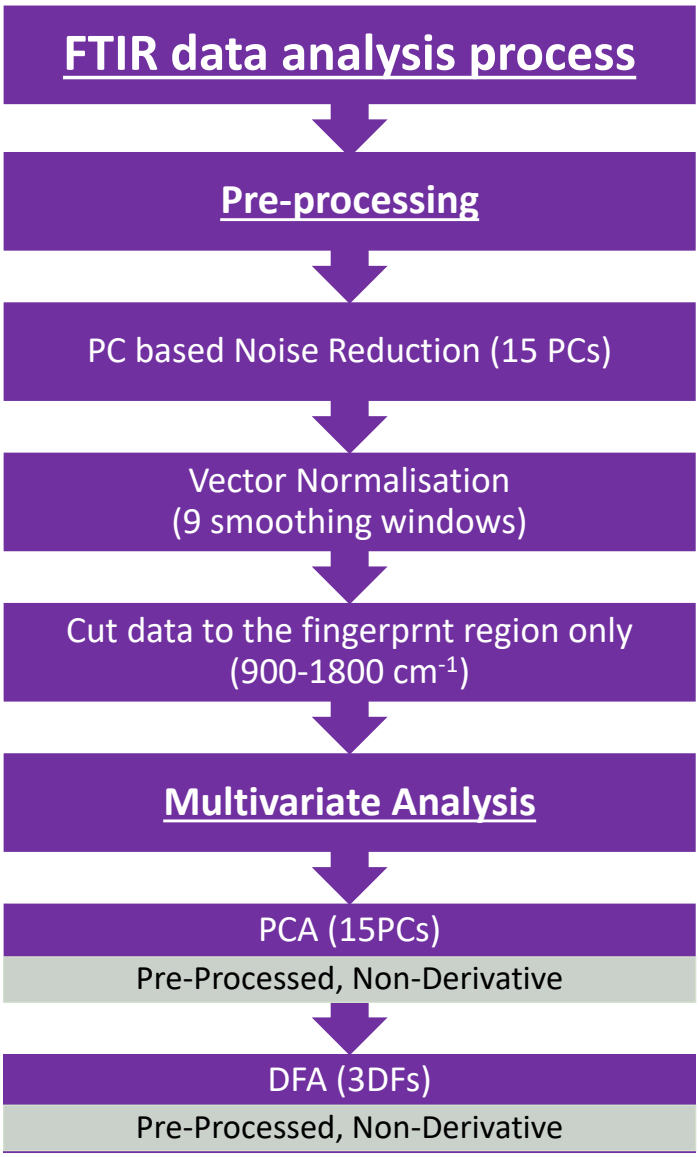

Figure 4: Data processing flowchart for processing the data collected during FTIR analysis in the bacterial identification study.

Pre-processing of the data was performed using in-house written/open source code that was used as a standard protocol for data analysis within the research group. The order in which the processing is performed is discussed later in this chapter. However, once the pre-processing had been performed the dataset was cut to the fingerprint region. Following this MVA, PCA and DFA, was performed. When performing PCA, the number of PCs retained will always differ depending on

the dataset, however, for this project enough PCs were retained to explain at least 95% of the total variance of the dataset. As mentioned previously it is important before any MVA is carried out that the data is pre-processed. Data that remains unprocessed before MVA will lead to results that can be skewed. Unprocessed data often affected by noise and instrument error as well as distorted baselines. For this reason, noise reduction followed by baseline correction is performed on the data to ensure that any separation seen after MVA is performed is not due these distortions. Assigning peaks to the loadings produced from PCA and DFA is affected by the order of derivatisation. Data that is non-derivative and 2nd derivative data is straightforward for assigning peaks, as where the peak appears in the loadings plot correlates to where a peak in the original spectrum. This means that a tentative peak assignment for the bands causing the vibration can be made. However, assigning peaks within 1st order derivative data is more difficult as the process of derivatisation causes the original maximum point of a peak to become located along the base line at zero. Therefore, the tentative assignments are assigned to where the value is zero between peaks, rather than the maximum of a peak, in the loadings plot. This way the cause of the separation seen in the PCA and DFA plots, and shown in the respective loadings, can be linked back to biological information within the sample. For a visual representation see Figure 5. [3]

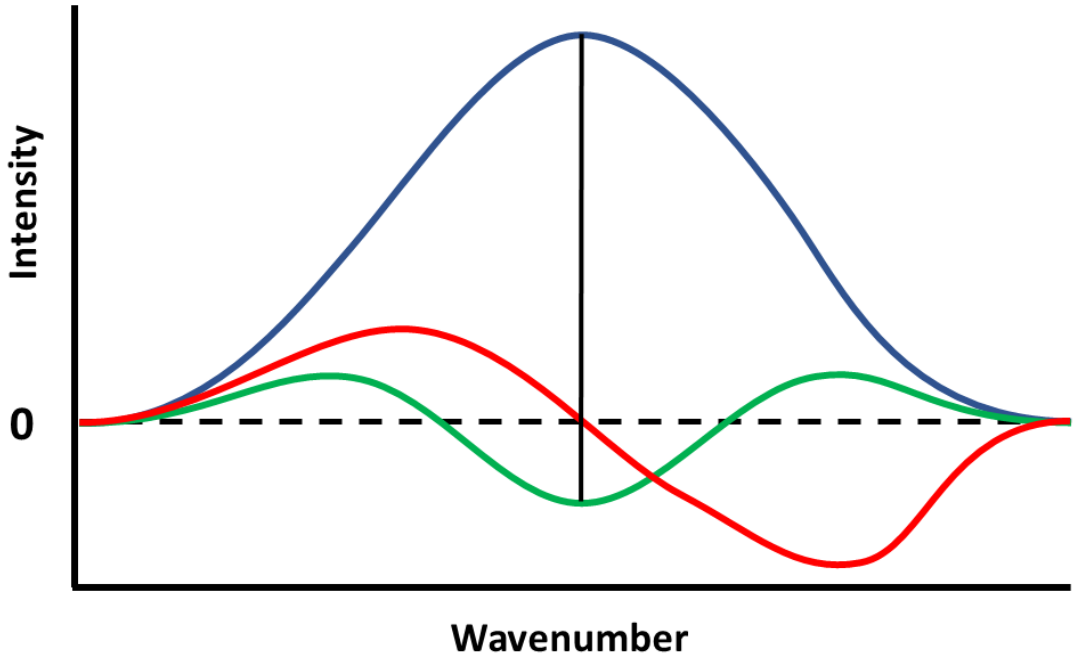

Figure 5: Schematic showing an example of an absorbance spectrum (blue) and the corresponding 1st order (red) and 2nd order (green) derivatives.

3.2.6.2. Multivariate Analysis (MVA)

Once data has been processed to correct for differences in noise, background and other non-biochemical differences, multivariate analysis (MVA) can be used to identify inter and intra-group differences that are caused by biochemical differences. Principal Component Analysis (PCA) and Linear Discriminant Function Analysis (DFA) are the two methods of MVA that have been used across all the data collected within this project. As well as the method and order of pre-processing, the number of Principal Components (PCs) and Discriminant Functions (DFs) used will also affect the separation of the data. Using too many PCs will then affect the separation seen in the PCA and DFA plot and can lead to overfitting of the DFA model. This is because more

noise gets introduced into the model when more PCs are used. Generally, the noisier a dataset is will mean that a higher number of PCs will be required to explain 95% of the variance. As such Raman datasets will require a higher number of PCs than the FTIR datasets.

3.2.6.3. Raman Pre-processing and Analysis

The Raman data collected during the bacterial identification study also had to undergo a set pre-processing steps prior to any MVA. These pre-processing steps were different to those used with FTIR but are used to address similar distortions and artefacts within the baseline and those that are caused by noise. Wavelet denoising was used as this method of noise removal is more suited to Raman data due to its ability to deal with low frequency data points caused by fluorescence. [4] Following this, the data was then analysed using PCA and DFA, the same as the whole process is shown in Figure 6. Raman, by nature, produces significantly noisier spectra in comparison to FTIR as Raman is a scattering technique that causes inherently weaker vibrations in comparison to FTIR vibrations. Derivatisation, and in turn the deconvolution of peaks causes, an enhancement of all peaks in a spectrum, both biological peaks and background noise. [3] As Raman spectroscopy produces spectra with a lower signal to noise ratio (S/R) derivatisation of the data was not performed.

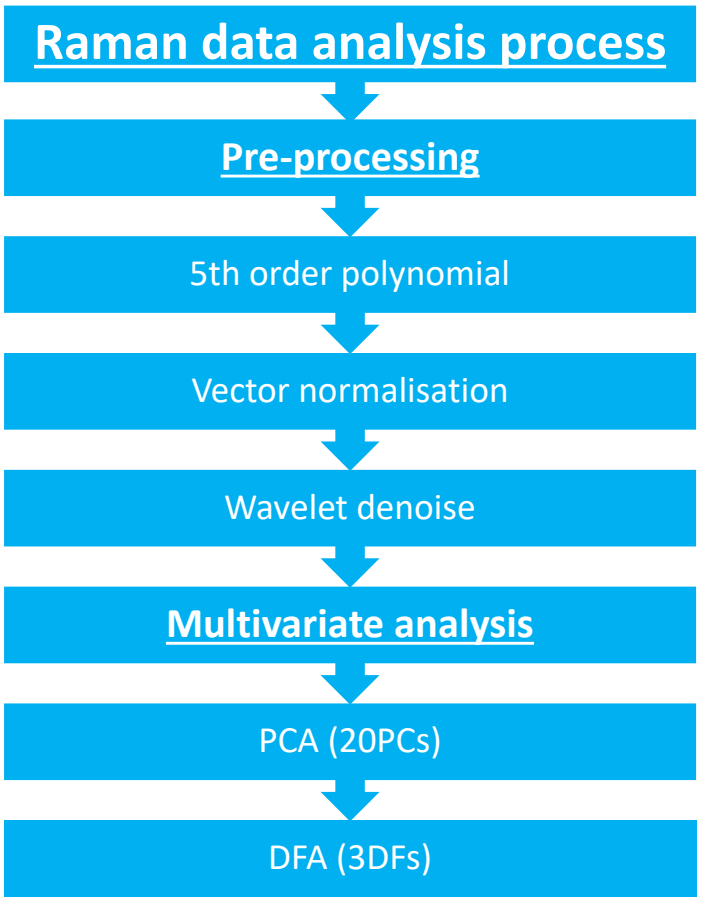

Figure 6: Data processing flowchart for processing the data collected during Raman analysis in the bacterial identification study.

The pre-processing steps were performed using IRootLab toolbox [5] as part of the standard processing procedure established within our research group. As with the FTIR data analysis, 20 PCs were used for PCA analysis to try and account for around 95% of the variance.

3.3. Results and Discussion

The following section of the chapter discusses the results that have been obtained using the Bioscreen C instrument and will discuss the optimal growth temperatures for the bacteria. The results section then moves on to discuss the results acquired using ATR-FTIR and Raman. The ATR-FTIR results section first discusses the results from the pre-processing study and the optimal order for this. The chapter then discusses the optimal derivative state for the data which has already been pre-processed. This section of the chapter works to develop the optimal data processing order for ATR-FTIR data which is concluded at the end of the section with the process shown visually in Figure 4.

The Raman section discusses the pre-processing methods used prior to the analysis of the data, discussing the issues experienced when using MirrIR with Raman before moving on to discuss the MVA results. A comparison of both techniques to see which was the most optimal for bacterial identification can be found at the end of this chapter.

3.3.1. Growth Curve Analysis

Growth curve calculations were carried out on all bacterial stains used within this project to understand how each individual bacterium grows, the results of which can be seen in Figure 8 to Figure 13.

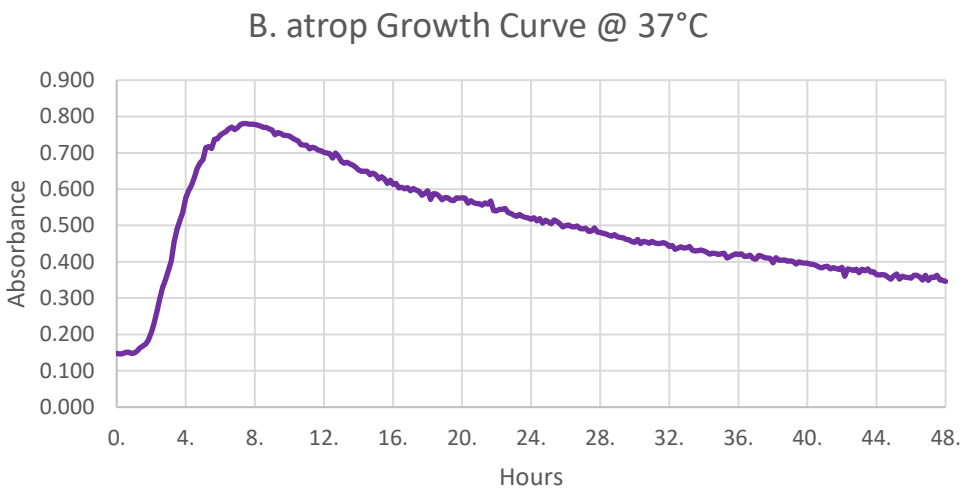

Figure 8: Growth curve graph for B. atrop grown at 37°C for 48 hours.

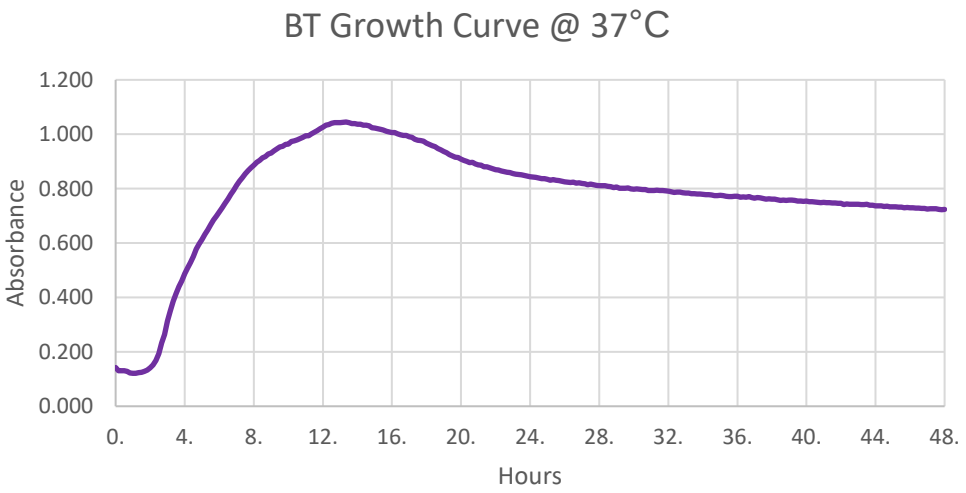

Figure 9: Growth curve graph for BT grown at 37°C for 48 hours.

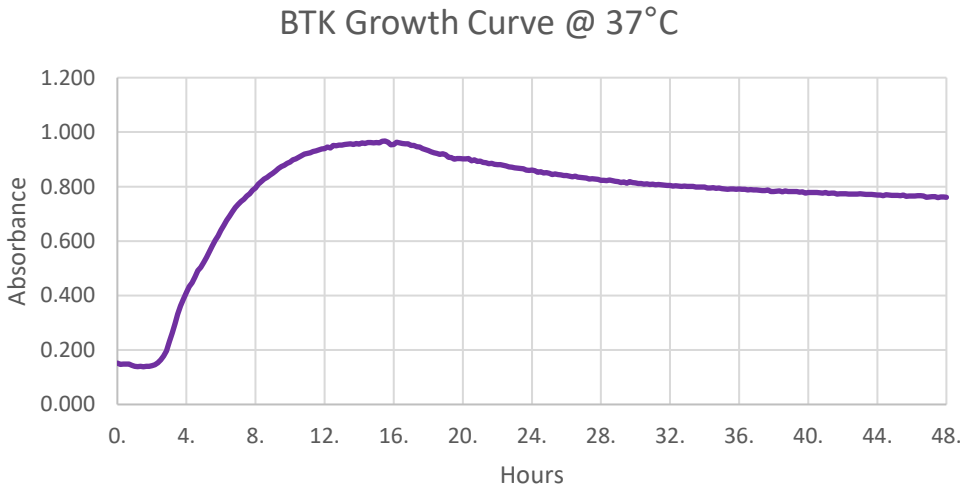

Figure 7: Growth curve graph for BTK grown at 37°C for 48 hours.

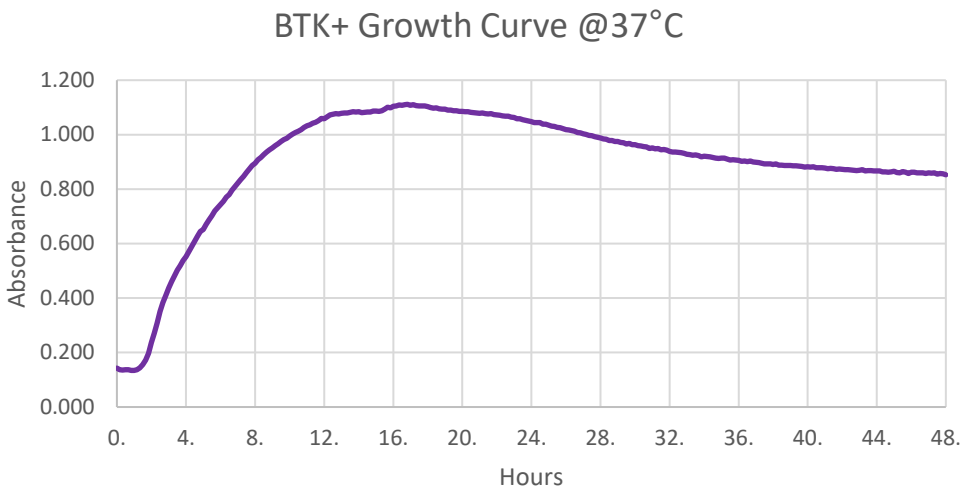

Figure 12: Growth curve graph for BT+ grown at 37°C for 48 hours.

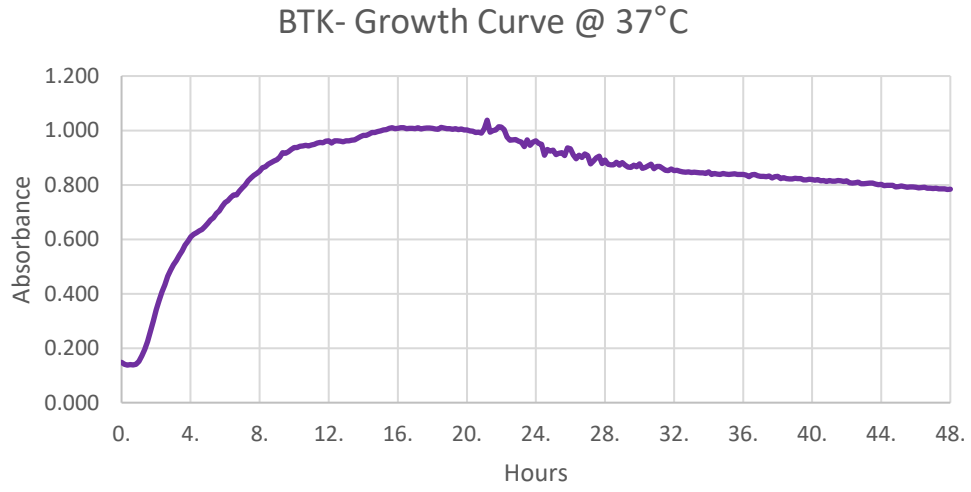

Figure 11: Growth curve graph for BTK- grown at 37°C for 48 hours.

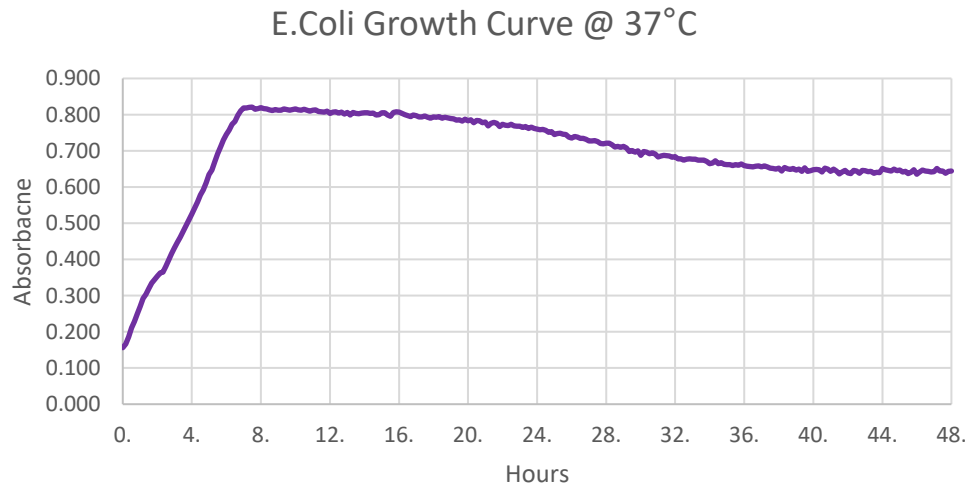

Figure 10: Growth curve graph for E.coli grown at 37°C for 48 hours.

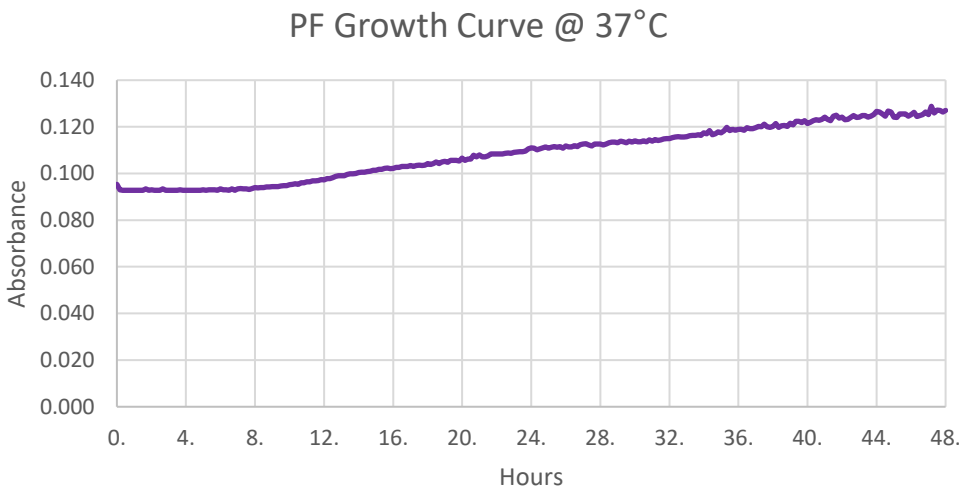

Figure 14: Growth curve graph for PF grown at 37°C for 48 hours.

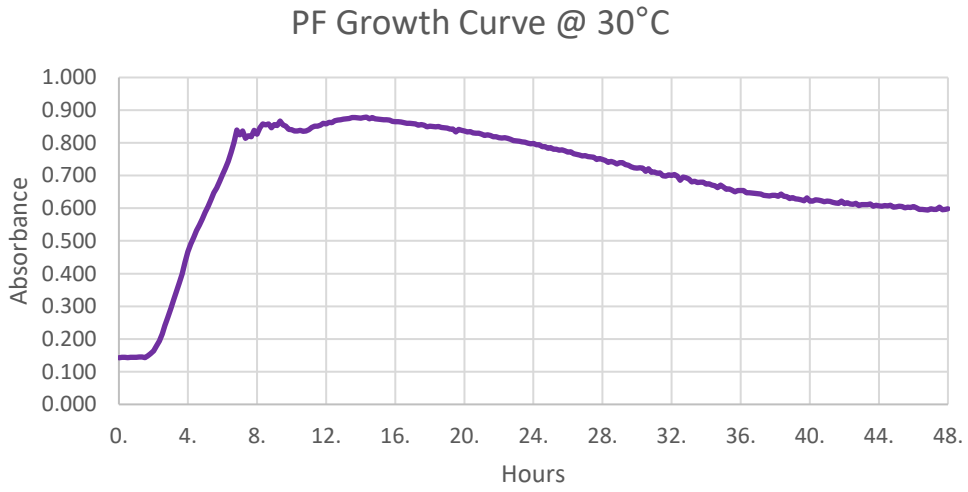

Figure 13: Growth curve graph for PF grown at 30°C for 48 hours.

From the growth curve analysis carried out over a 48-hour period it was found that all strains went through the four stages of growth within this period. Within the growth curves shown from Figure 8 to Figure 13 all bacteria have reached their peak level of growth between 8-20 hours of growth depending on the bacterial strain. However, as these curves are based on liquid samples that are continuously shaking growth may be faster due to easier access to nutrients than bacteria grown on agar plates. [6] As agar plates were used within this project all plates for all bacterial strains

were grown for 24 hours to ensure there was sufficient biomass to carry out the experiments and to be consistent with standard protocol. [7] Figure 14 and Figure 13 show the growth curves for *P. fluorescens* (PF) showing there was no growth at 37°C but there was growth at 30°C. It was also found that all except one of the bacterial strains used within this study could grow at 37°C. *P. fluorescens* (PF) was found to need a lower temperature to grow, therefore PF was grown at 30°C throughout this project. Bacteria were grown for 24 hours to ensure that each strain progressed through the lag and exponential growth phases and were in either the stationary or death phase. [8] The time that is required to reach each of these different stages of growth differs depending on the bacterium, however 24 hours was enough for each of the strains to be at this phase of its cycle.

3.3.2. ATR-FTIR

Data collected using ATR-FTIR was used to assess which order of pre-processing was the optimal method for bacterial identification, these results are discussed first. Following this, once the data had been pre-processed, the data were converted to the 1st and 2nd derivative. The three sets of data (non-derivative, 1st order derivative and 2nd order derivative) were analysed using both unsupervised (PCA) and a supervised (DFA) forms of MVA. Four different combination of pre-processing methods were trialled to see which was the optimal method for

maximising the information that can be gained from the data. The combinations are shown below in Table II, and are referred to by their Method number.

Table II: Table showing the four different methods trialled for which order is optimal for pre-processing.

	Method 1	**Method 2**	**Method 3**	**Method 4**
Step 1	CUT	CUT	Noise Reduction	Vector Normalisation
Step 2	Noise Reduction	Vector Normalisation	Vector Normalisation	Noise Reduction
Step 3	Vector Normalisation	Noise Reduction	CUT	CUT

The order in which the pre-processing methods were performed affected the variance shown in the data. The PCA and DFA plots, with the corresponding loadings and peak assignments, shown here are for Method 3 as this was chosen as the optimal method. The plots produced from Method 1, 2 and 4 can be found in Appendix B. It should be noted that the PCA and DFA plots shown in this chapter and Appendix B are chosen based on the clearest separation and clustering of the data.

3.3.2.1. Pre-Processed Data (Method 3)

The separation and clustering shown in the PCA and DFA plots is affected by the order in which different processing steps are performed. Cutting the data last resulted in more defined boundaries between the different strains, except for the BTK strains as there is a large amount of overlap between these groups regardless of processing order. However, cutting the Discriminant Function Analysis data to the

fingerprint region first resulted in tighter clustering of the bacterial strains. This may be due to less biological information to consider as the higher wavenumber range has been removed, as well as removing noisier regions of the spectrum near to the detector cut off. Cutting the data to the fingerprint region prior to pre-processing allows for more noise to be removed from the dataset, by discounting the variance caused by the higher wavenumber vibrations. The PCs and DFs responsible for showing the best separation across the Methods also changed. PC1 vs PC3 and DF1 vs DF2, shown in Appendix B: Figure 1 and Figure 4, demonstrate the best separation in Method 1 and 2 where the data was cut before the pre-processing. However, when the data is cut last as in Method 3 and 4 it is PC2 vs PC3, shown in Figure 15A and Appendix B: Figure 7A, where the best separation is seen compared to PC1 vs PC3 from Method 1 and 2 where the data has been cut first. The PCs used to plot the data will have an influence on separation seen in both the PCA and DFA plots. With Methods 3 and 4, PC1 shows the non-biochemical changes from within the dataset, whereas with Methods 1 and 2 biochemical changes are shown but are combined with the remaining noise that was not removed during pre-processing causing more overlap between the bacterial clusters, especially within the PCA plots. Thus, Methods 3 and 4 are more optimal. This is likely to be as a result of the data being cut to the fingerprint region after it has been pre-processed, so more noise within the dataset has been removed compared with Methods 1 and 2.

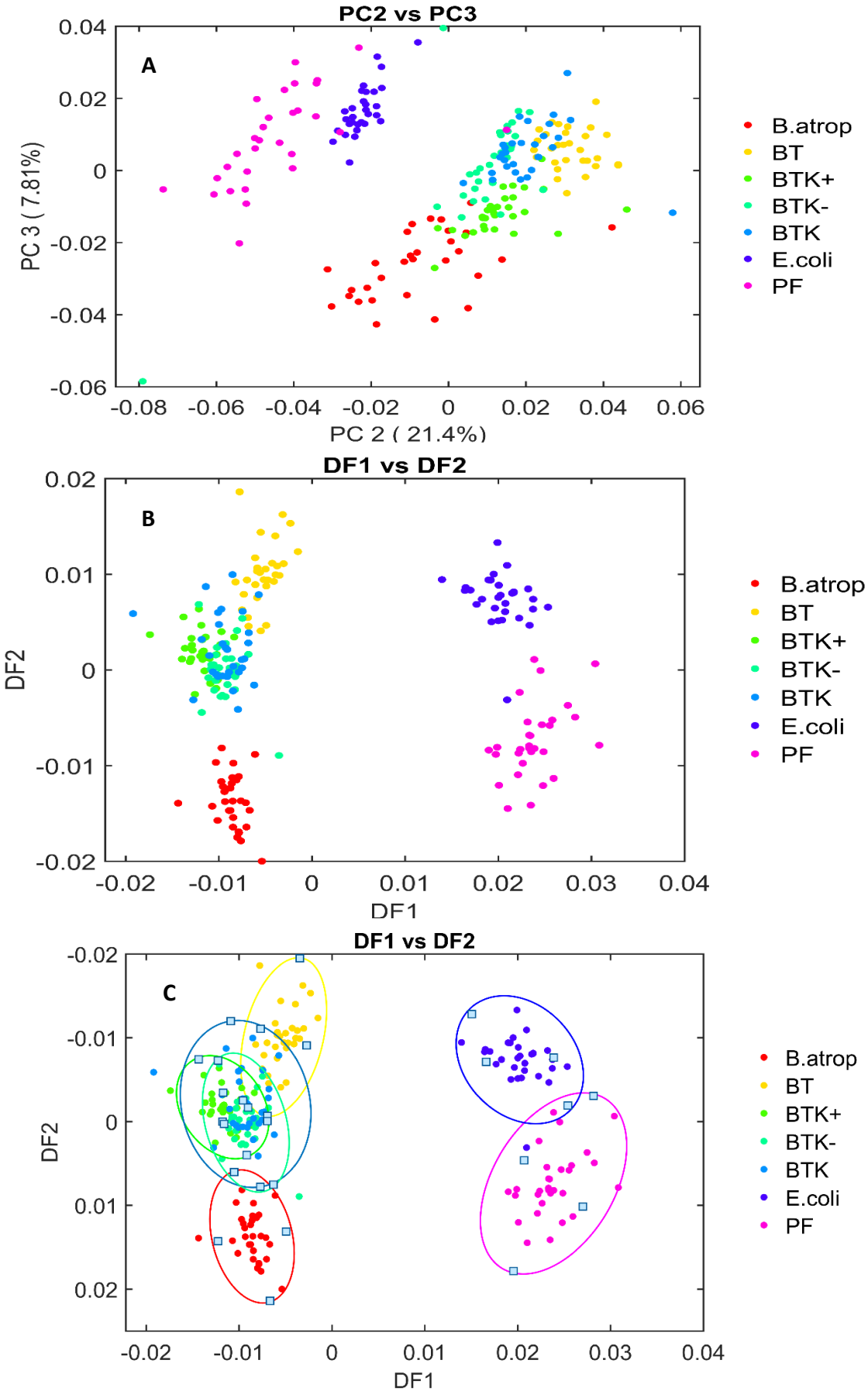

Figure 15: Figure showing a PCA plot using PC1 and PC3 (A) a DFA plot using DF1 and DF2 (B) and a the same DFA plot with a 95% confidence ellipse (C) from data pre-processed using Method 3.

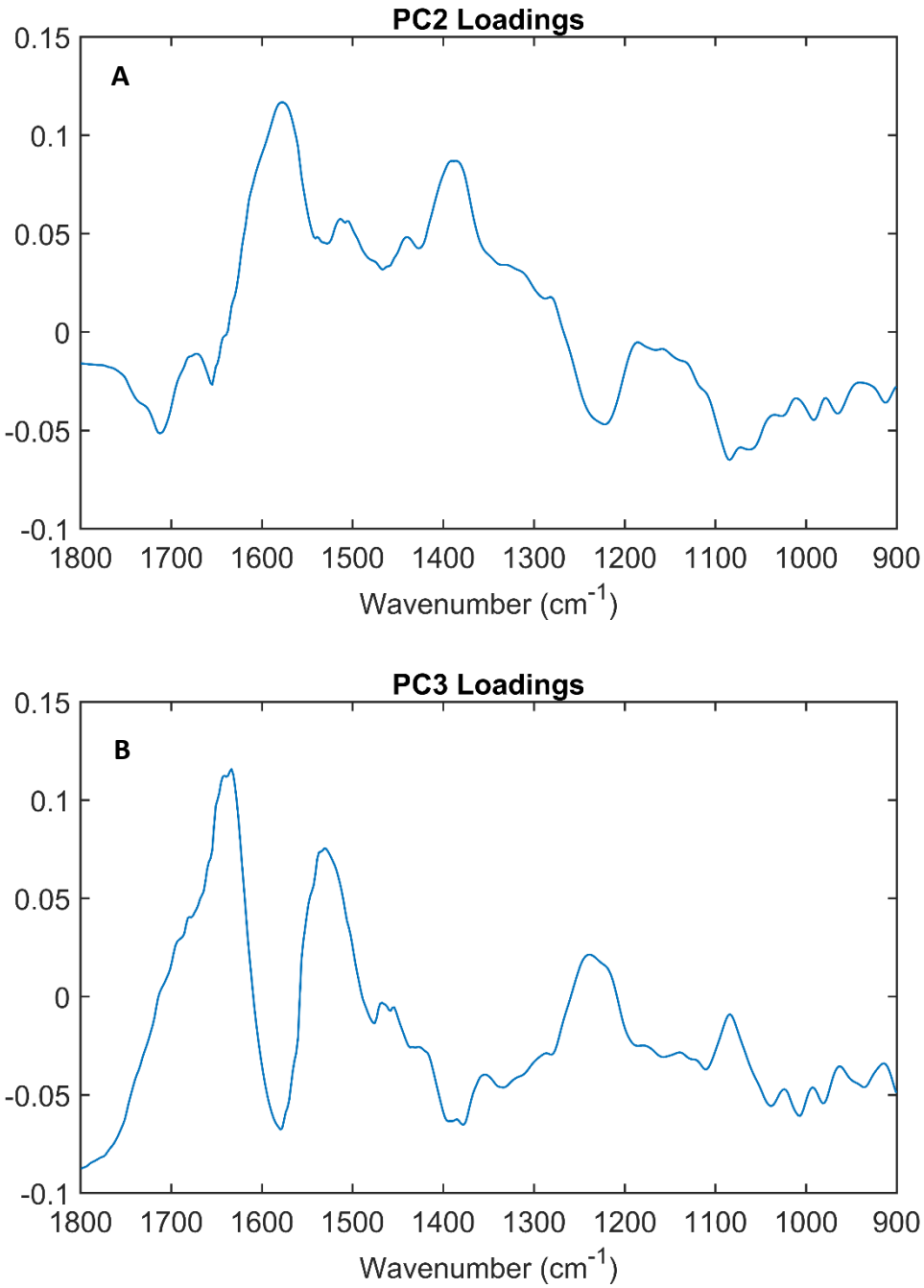

Figure 16: Figure showing a PC1 loadings plot (A) and a PC3 loadings plot (B) from data pre-processed using Method 3.

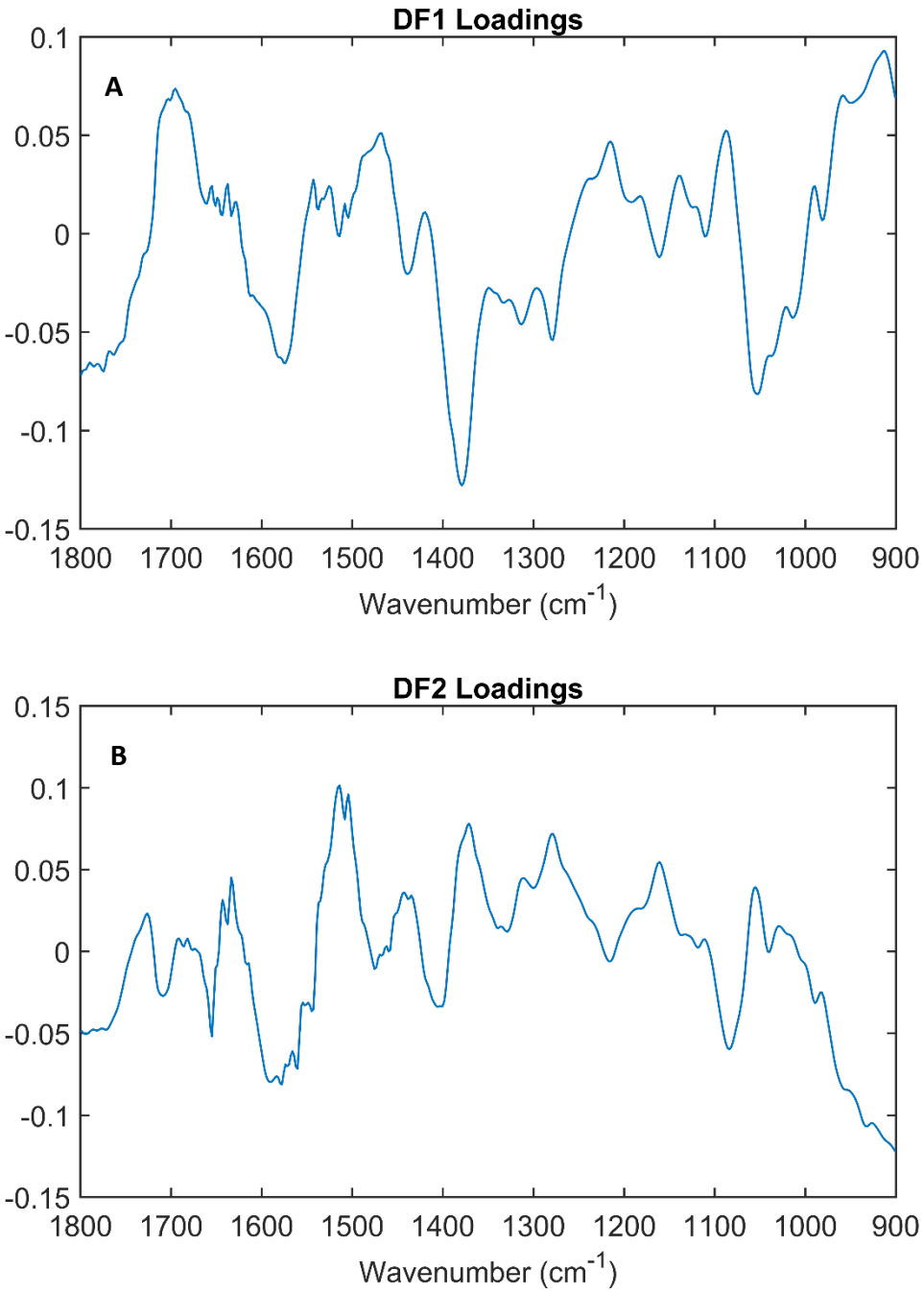

Figure 17: Figure showing a DF1 loadings plot (A) and a DF2 loadings plot (B) from data pre-processed using Method 3.

Table III: Wavenumbers and corresponding peak assignments for PC1, PC3, DF1 and DF2 loadings from Method 3. [9,10,11,3,12,]

Wavenumber (cm^{-1})				Tentative peak assignment
PC2	PC3	DF 1	DF 2	
	980	912, 958,	933, 957, 1161	vC-O, vC-C, vC-O-H, vC-O-C
	1007			Phenylalanine
	1039	1038		Phenylalanine C-H in-plane
1063				vC-N, vC-C
1084		1053, 1088	1084	vP=O (symmetric) PO^-_2
		1279	1279	Amide III band of proteins
1388	1379, 1394	1379, 1468	1371	δCH_2 and δCH_3
1514			1514	Tyrosine
1506, 1539	1531, 1537		1504	Amide II
1577	1579	1574, 1581	1560, 1572, 1577, 1591	Aspartate and glutamate carboxylate stretching
	1633, 1641			Amide I β-pleated sheet structures
			1655	Amide I α-helix
		1682, 1695		Amide I band components resulting from antiparallel Pleated sheets from β-turns of proteins
1712		1703		vC=O of esters

The PCA and DFA plots that are produced using Method 1 and 2 show the same separation and clustering of the data. PC1 separates out the Gram-negative bacteria and *B. atrop* from the BT species and sub species. The PC1 loadings show the region of the spectrum that is responsible for the separation is the wavenumber range from 900-1250 cm^{-1}. The peaks in this region are because of C-C, C-O, C-O-H vibrations and PO_2 stretches (Appendix B: Table I and Table II). These peaks are found in Gram-negative bacteria PF and E. coli, due to their different cell structure. PC3 is responsible for the separation of PF and BTK from *B. atrop, E. coli*, BT, BTK+ and BTK- are all found in the centre of the plot with respect to PC3. A similar region of the PC2 loadings for Method 3 and 4 are responsible for the separation as to those seen in the PC1 loadings for Methods 1 and 2. Again the loadings mainly show vibrations caused by PO_2 stretches around 1050-1100 cm^{-1}. PC3 is used across all four of the

methods, however the separation caused by it is different. While Methods 1 and 2 separate the Gram-negative bacteria, this is not the case in Methods 3 and 4, where both Gram-negative bacteria are found in the positive region of the plot. BT and BTK are also found in the positive region of Figure 15 and Appendix B: Figure 7B. Meanwhile, B. atrop, BTK+ and BTK- are all located in the negative region of the plot. Unlike the PCA scores plots form Method 1 and 2, no groups are separated across the centre of the PCA scores plots generated in Method 3 and 4. PCA plots using PC1 are not shown within this chapter for Methods 3 and 4 as the separation shown within these plots is not as clear as when considering PC2 vs PC3. While PC1 explains the most variance within the dataset, when considering Methods 3 and 4 this is not the PC responsible for the separation of the bacterial strains and differing Gram types. Therefore, in this case PC1 may explain any remaining noise that was not removed from the dataset during pre-processing along with similarities between the different strains. This means that PC2 and 3 contain the biological information that differs between the bacterial strains and differing Gram types. As well as the visual separation changing depending on the processing order, the percentage variance explained within the first three PC's also changes with around a 1.5% increase when the data is cut after the pre-processing has been performed, shown in Table IV.

Table IV: Table showing the percentage variance explained for the first three PC's for each of the four methods.

	PC	Method 1	Method 2	Method 3	Method 4
Percentage explained (%)	1	54.1	54	63.1	63
	2	20.3	20.2	21.4	21.3
	3	16.4	16.4	7.81	8.04
	1, 2 and 3	90.8	90.6	92.31	92.34

With a change in the percentage variance from the first three PCs also comes a change in the number of PCs required to explain 95% of the total variance. When the data is cut after processing, the number of PCs needed is reduced from five to four. This highlights that when the data is cut, to the fingerprint region in MATLAB, before performing the pre-processing steps more noise is retained in the dataset, therefore leading to a higher number of PCs needed to explain 95% of the variance within the data. A summary of this information is presented in Table V.

Table V: Table showing the number of PC's required to explain at least 95% of the total variance with the percentage explained by those PC's also stated.

	Method 1	Method 2	Method 3	Method 4
Number of PCs	5	5	4	4
Percentage explained (%)	95.67	95.43	96.47	96.45

When visually comparing DFA analysis to the PCA analysis, DFA shows more defined clustering of the bacterial groups with DF1 showing a very clear separation of Gram-positive and Gram-negative bacteria across all four methods (Figure 15B and Appendix B: Figure 1B, 4B and 7B). However, the separation shown by DF2 differs slightly depending on the order in which the pre-processing was carried out; not only if the data was cut first or not but also the order in which vector normalisation and noise reduction was performed. For Method 1, DF2 separates PF from *E. coli* as well as separating *B. atrop* and BTK+ from the other BT species and sub species (Appendix B: Figure 1B). None of these clusters were separated across the DF2 axis. However, Method 2 again separates the two Gram-negatives as well as *B. atrop* and BTK+ from the remaining BTK species and sub-species but BT was located in the centre of the

plot separates across the DF2 axis (Appendix B: Figure 4B). In Method 3 and 4, DF2 also separates PF from *E. coli* (Figure 15 and Appendix B: Figure 7B). However, they also separate *B. atrop* and BT, while other BTK species and sub-species are separated across the centre of DF2. Similar peaks can be seen in the loadings plots showing the wavenumbers that are responsible for the separation seen. For DF1, the wavenumber range from 900-1250 cm^{-1} is important as seen in PC1. The peaks in this region arise mainly due to C-C, C-O, C-O-H vibrations and PO_2 stretches (Appendix B: Table I). For Method 3, in terms of DF2 loadings, the key vibrations responsible for separating PF and *B. atrop* form the other strains are also C-C, C-O, C-O-H vibrations and PO_2 stretches. Additionally, amide I, C=O ester stretches and aspartate and glutamate stretches are responsible (Table III).

Considering the phylogenetic information, discounting the Gram-negative strains, you would expect BT and *B. atrop* to separate out from BTK and its sub-species. This is observed in DFA plots produced when using Method 3 and 4 but not Methods 1 and 2. By cutting the data to the fingerprint region after performing vector normalisation and noise reduction increases the percentage variance observed in the first three PC's. By performing the cut after the processing, it increases the variance explained by around 1.5% depending on the order in which the normalisation and noise reduction is performed, shown previously in Table IV. This, coupled with the separation shown in the DFA plots has led to Method 3 being the optimal method chosen for pre-processing the data. Figure 18 shows the effect that pre-processing has on the bacterial dataset, using the order of Method 3.

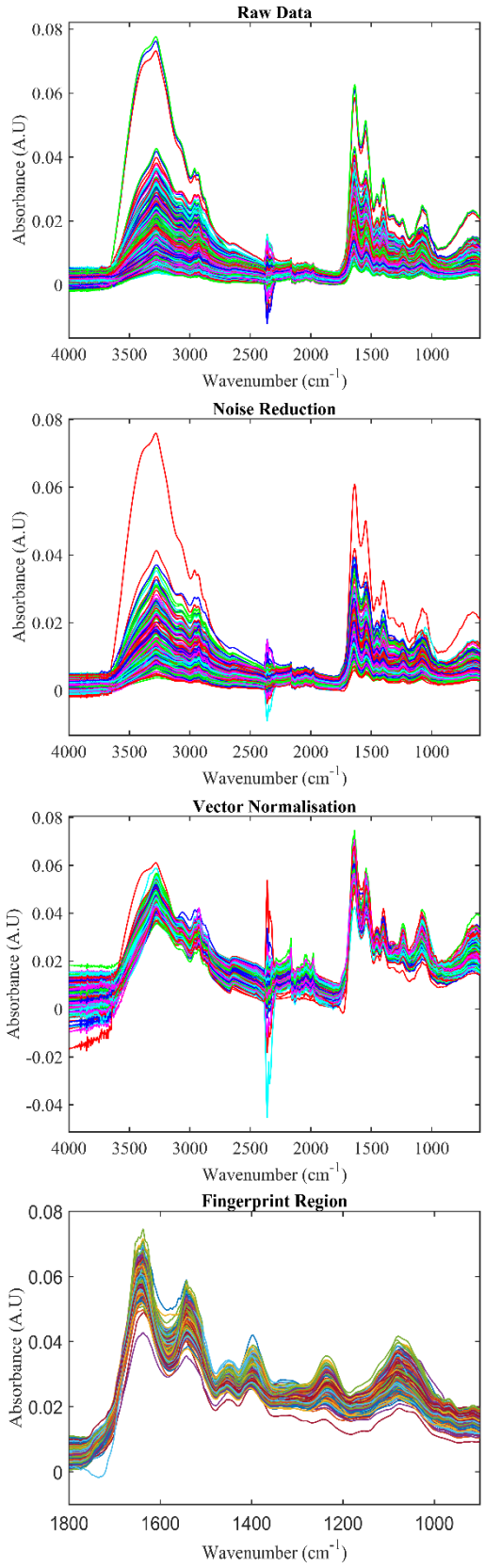

Figure 18: Figure showing the effect that the pre-processing of the data has starting with the raw data (A) followed by noise reduction (B) then vector normalisation (C) and finally the cutting the data to the fingerprint region (D).

To test the ability of the DFA model a new dataset was created by splitting the existing data into a ~$\frac{2}{3}$ training set and a ~$\frac{1}{3}$ test set with the same number of samples from each bacterial strain. This data was projected into the model with a 95% confidence ellipse shown in Figure 19. Table VI shows the sensitivity and specificity values that were calculated using equations which are shown in Chapter 2 section 8.2.3.

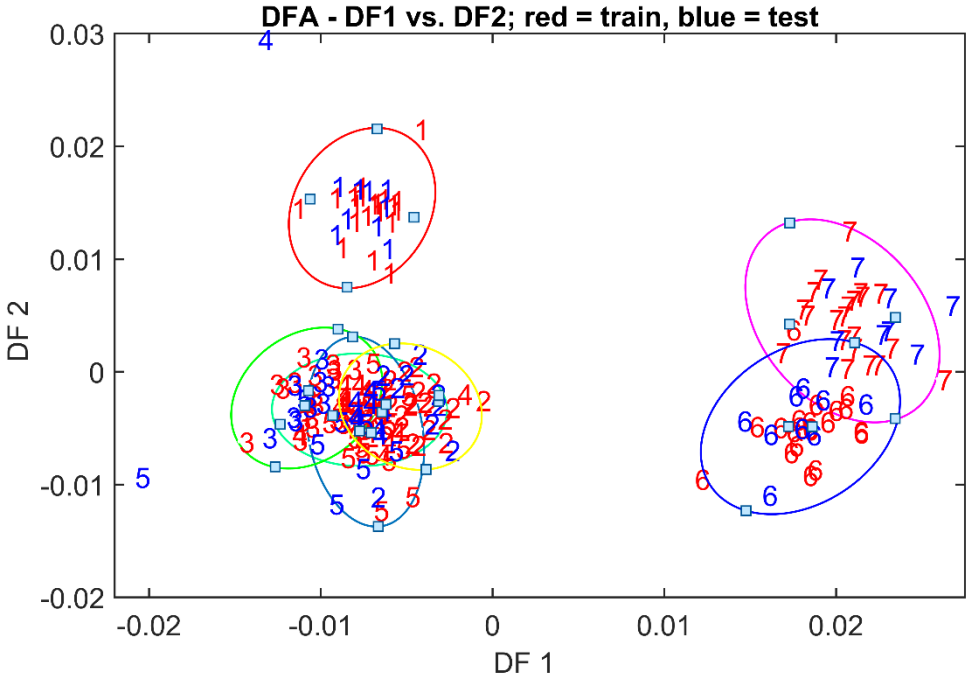

Figure 19 PCDFA plot showing DF1 vsDF2 for Method 3 data split into a $\frac{2}{3}$ training $\frac{1}{3}$ test set.

Table VI: Table showing the sensitivity and specificity values calculated for the Method 3 dataset when using a 95% confidence interval

	B. atrop (1)	BT (2)	BTK+ (3)	BTK- (4)	BTK (5)	*E. coli* (6)	PF (7)
Sensitivity (%)	100	80	100	100	87.5	75	80
Specificity (%)	100	78.4	80	**61.5**	**71.1**	96.1	94.2

While the separation seen visually in Figure 19 is not as clear as that in Figure 17, the sensitivity and specificity values shown in Table VI highlight that the model works well for the majority of the strains examined in this study. These figure show that the model has a good level of classification across all of the bacterial stains used within the study and that the model only begins to struggle with the sub strains of BT, particularly BTK- and BTK, with these values highlighted in bold within Table VI.

3.3.2.2. Derivatisation of Data

The final stage of pre-processing that is required before MVA is derivatisation. The following section shows the results produced from data that was processed using Method 3. The data was then categorised by having a 1st order or a 2nd order derivatisation performed or remaining non-derivative. The PCA plots, with corresponding loadings, for both the 1st and 2nd order derivatisation datasets can be seen in Appendix B: Figures 10, 11 and 13. The DFA plots for 1st and 2nd order derivatisation data can be seen in Figure 20. The corresponding loadings plots for this data is shown in Appendix B: Figure 12 and 14. The tentative assignments given to the peaks identified within the loadings plots for both 1st and 2nd order data can be seen in Appendix B: Table IV, V and VI.

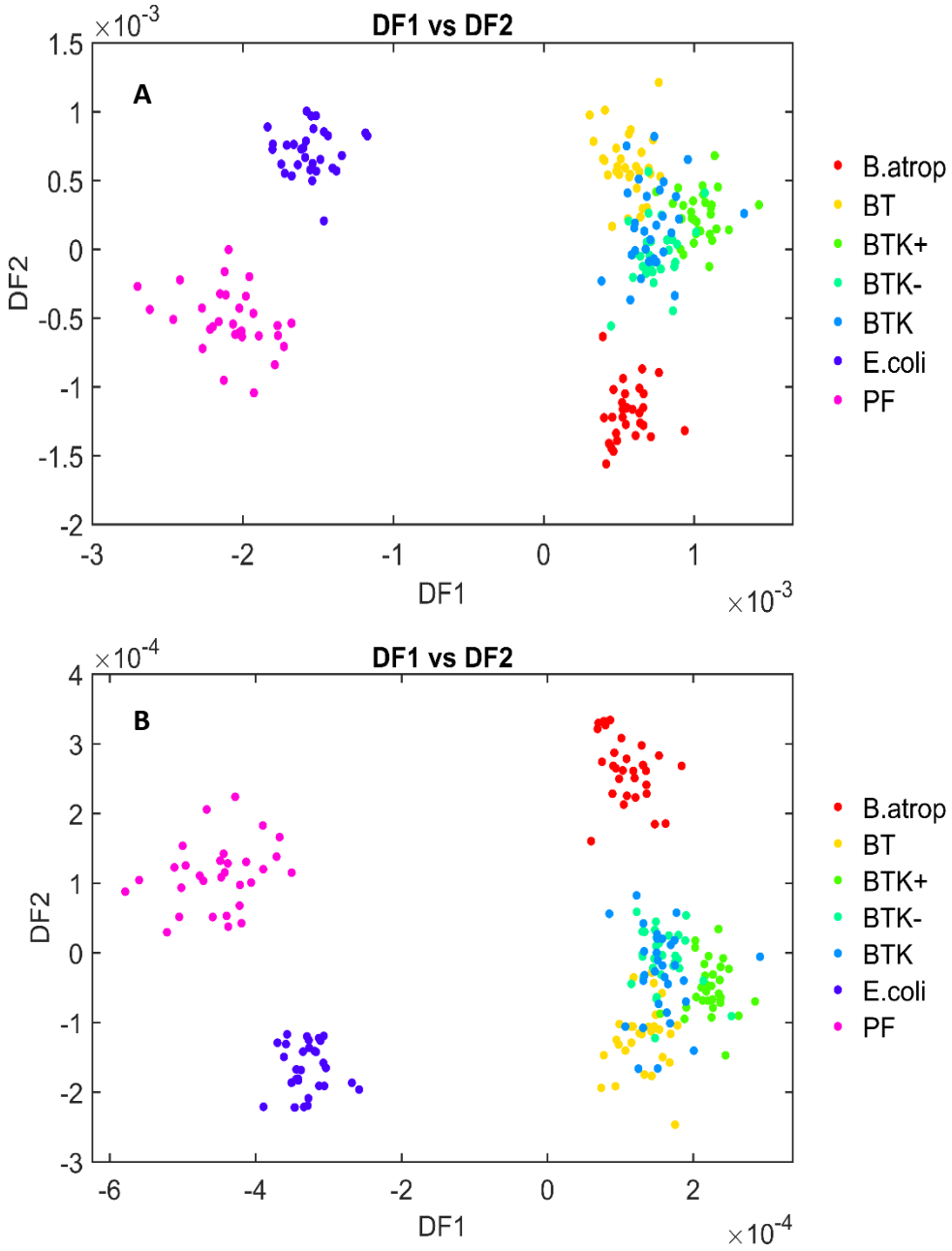

Figure 20 Figure showing a DFA plot using DF1 and DF2 for 1st order data (A) and a plot using DF1 and DF2 for 2nd order data (B).

The percentage variance explained by the first three PCs for the non-derivative, 1st order and 2nd order derivative data is shown in Table VII. From Table VII non-derivative data has 5-6% more variance explained in just three PC's compared to either 1st or 2nd order derivatives of the same dataset.

Table VII: Table showing the percentage variance explained for the first three PC's for pre-processed, 1st order derivative and 2nd order derivative data.

	PC number	Pre-processed (Method 3)	1st Derivative	2nd Derivative
Percentage explained (%)	1	63.1	63.9	55.1
	2	21.4	15	23.7
	3	7.81	7.77	7.2
	1,2 and 3	92.31	86.67	86

Prior to MVA, the final step of data pre-processing was optimised. Following on from the pre-processing carried out using Method 3, 1st and 2nd order derivatisation was performed on the data. A comparison of the percentage variance explained from the first three PC's was used as an indication of the optimal order along with the visual separation and clustering observed in the PCA and DFA plots. Table VII shows that performing either a 1st or 2nd order derivatisation on the data cause the percentage variance explained by the first three PC's to drop from 92.31% to 86.67% and 86% respectively. The drop observed in the percentage variance explained can be explained by the fact that as derivatisation is performed the noise within the dataset becomes exaggerated along with the key biological peaks. Additional to this, when interrogating the PCA plots produced by the data for 1st and 2nd order derivatives (Appendix B: Figure 10 and 13A), along with the DFA plots shown in Figure 20, the separation and the clustering of the bacterial strains and Gram type observed has no improvement on that shown in the pre-processed, non-derivative dataset from Method 3 (Figure 15).

As with previous plots, the PCA and DFA plots for both 1st and 2nd order derivatisation shows separation of the Gram-negative strains from the Gram-positive

ones. However, both the PCA and DFA models struggle to clearly separate BT from BTK and its sub species. As discussed previously it is the C-C, C-O, C-O-H vibrations and PO_2 stretches as well as Amide I, C=O ester stretches and Aspartate and Glutamate stretches are responsible for most of the separation observed in the plots (Appendix B: Table IV, V and VI). By interrogating the BTK species and sub species independently it may be possible to interpret more biological information about the subtle differences between the strains. This could lead to a two-step approach in the identification of samples where all BTK and their sub-strains are grouped in the first discriminatory model. If identified as BTK than the data would then be projected into a second discriminatory model containing only BTK species for a more accurate classification.

Gram-positive bacteria have a thick cell wall consisting of mainly peptidoglycan, while Gram-negative bacteria has multiple different layers to the cell wall including an outer membrane. It is these differences in cell structure that are most often detected in ATR-FTIR spectra by the first few PC's. An interesting property of *B. thuringiensis* (BT) and their subspecies is the presence of parasporal crystals. [13] These parasporal crystals are δ-endotoxins which can be formed during the stationary growth phase. [14] This is a property that is not present in other Bacillus species, therefore *B. atrophaeus* will not have these. Within this project four species and sub-species of BT are used. However, one of the strains has lost an insecticidal gene (BTK-), with BTK+ being the parent strain used to create BTK-. The presence of these parasporal crystals is another component that may also be detected within the first few PC's as it is a major difference between the different cell types. The

differences observed in the data shown within this chapter are shown to correlate to the changes in cell structure between species and Gram-type. This shows that the technique picks out significant biological changes making the technique ideal for future bacterial classification.

The aim of this project was to develop a methodology that was optimal for bacterial identification with an overall end goal of making the methodology into one that is suited to in situ analysis of samples. A key feature of this is to keep the methodology as simple as possible with as few steps as possible between the collection of the data to the identification of a sample. Adding in an extra step of pre-processing before moving on to MVA of the data therefore is not ideal. This coupled with the lower percentage explained and the fact that there is no improvement in the separation and clustering of the data illustrate why it was chosen not to perform a derivatisation on any future data acquired.

3.3.3. Raman

Once the Raman data had been acquired it was analysed using PCA. However, as you can see from Figure 21 there is no separation or clustering of the data based on bacterial strain. Figure 21 shows this is not possible using any combination of the first three PCs. As there is no separation or clustering when using bacterial strain as the legend, two alternative labelling systems were used to investigate the data. The labelling systems chosen were to look at the batch from which the data was collected

and the month that the data was acquired in. The PCA plots from both systems can be seen in Figure 22 and Figure 23 respectively. The loading that relate to these figures can be seen in Figure 24.

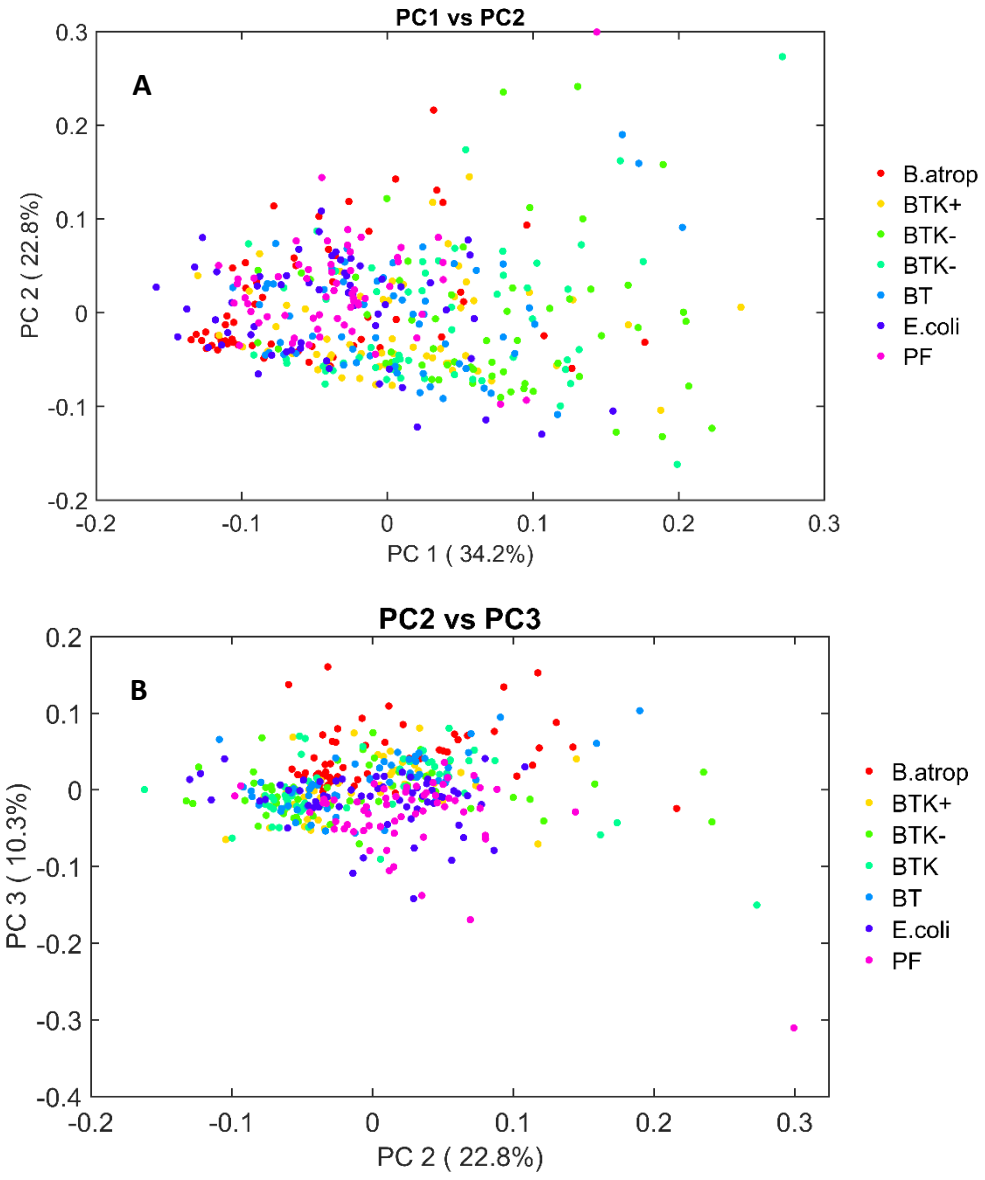

Figure 21: Figure showing a PCA plot using PC1 and PC2 (A) and a PCA plot using PC2 and PC3 (B) from Raman data collected with a 785 nm laser.

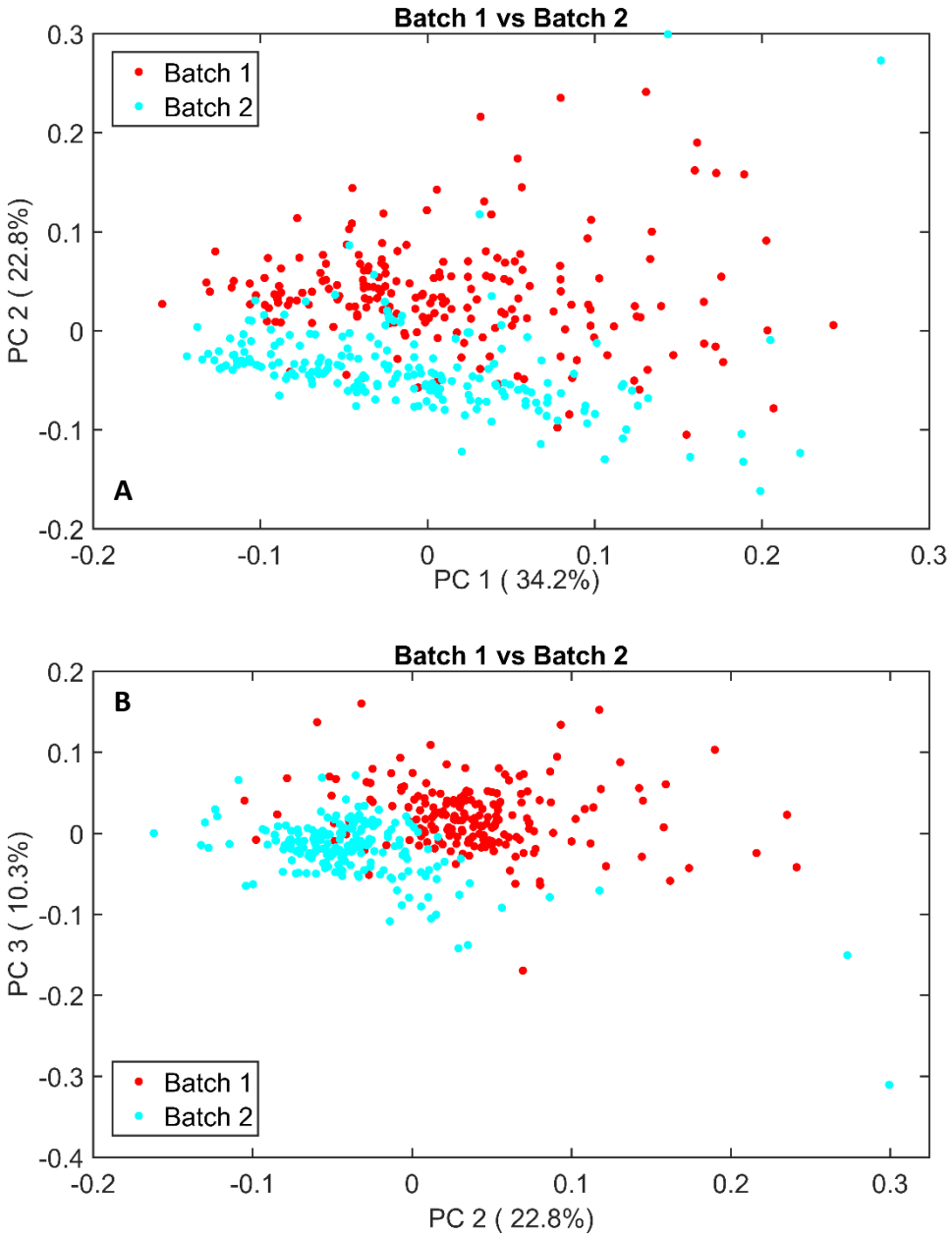

Figure 22: Figure showing a PCA plot using PC1 and PC2 (A) a PCA plot using PC1 and PC3 (B) and a PCA using PC2 and PC3 from Raman data collected with a 785 nm laser. The grouping system used in these data plots looks at the effect that the batch that the bacteria was prepared from has on the clustering of the data.

PCA analysis with batch as the legends showed that there was some separation and clustering of the data. This shows that the most significant factor for the separation in this case is not due to bacterial strain. Using PC2 vs PC3 produced the best level of separation and clustering, showing separation of Batch 1 from Batch 2. All three of the batches were prepared and spotted onto the slide at the same time. However, due to the Raman data being collected using instrumentation at UCLan where access to the instrument was limited, analysis of the samples was carried out over a period of several months. Therefore, it is highly possible that some sample degradation may have occurred between measurements as the samples aged as all spots for each bacterium were analysed for batch one before moving onto batch two samples. Although the separation seen here in the plots appears to be based on the batch from which the samples were produced, what is much more likely to be causing the separation of the data is the month in which they were analysed. ATR-FTIR analysis of three batches showed that there is no variation between them. PCA analysis was performed where the month of analysis became the legend with the plots shown in Figure 23. Figure 23 also shows that there is a higher spread of the samples in April compared to June or July. The cause of this is likely to be due to the fact that larger changes are likely to occur as the freshly prepared samples begin to dry out. Figure 24 shows the loadings produced for the Raman dataset, showing that there are clear biological peaks that are causing the separation of the data. Therefore, the separation seen within PCA plots cannot be put down to instrumental changes caused by analysing samples on different days.

Raman has successfully been used by a number of different research groups for bacterial identification and speciation. Research has used spontaneous confocal Raman, ultraviolet resonance Raman (UVRR), tip-enhanced Raman spectroscopy (TERS) [15] and surface enhanced Raman spectroscopy (SERS) [16] for bacterial identification. This work has involved preparing samples on calcium fluoride (CaF_2) windows for confocal Raman, suspension samples deposited onto a polystyrene well plate for UVRR, glass slides for TERS and an aluminium scanning electron microscopy (SEM) stub for SERS, all of which are considered ideal for Raman analysis. For all of these studies there is little to no influence on the spectral profile resulting from the substrate. A protocol paper published by Butler *et al*., discusses the different substrates that can be used for Raman, along with many other parameters that can affect a Raman spectrum, so that this information can be taken into account prior to the start of an experiment. [17]

The most likely cause of the poor separation observed within this study when considering the Raman data is that samples were prepared on MirrIR slides. MirrIR is not classed as on optimal substrate for use with Raman analysis. However, the overall objective for this project is to provide a technique that can identify samples found on any surface. The fact that Raman is significantly affected by the MirrIR substrate, as well as struggling to provide spectra for some of the substrates discussed in Chapter 3 section 3.2, indicates that Raman may not be the optimal technique to use for this project.

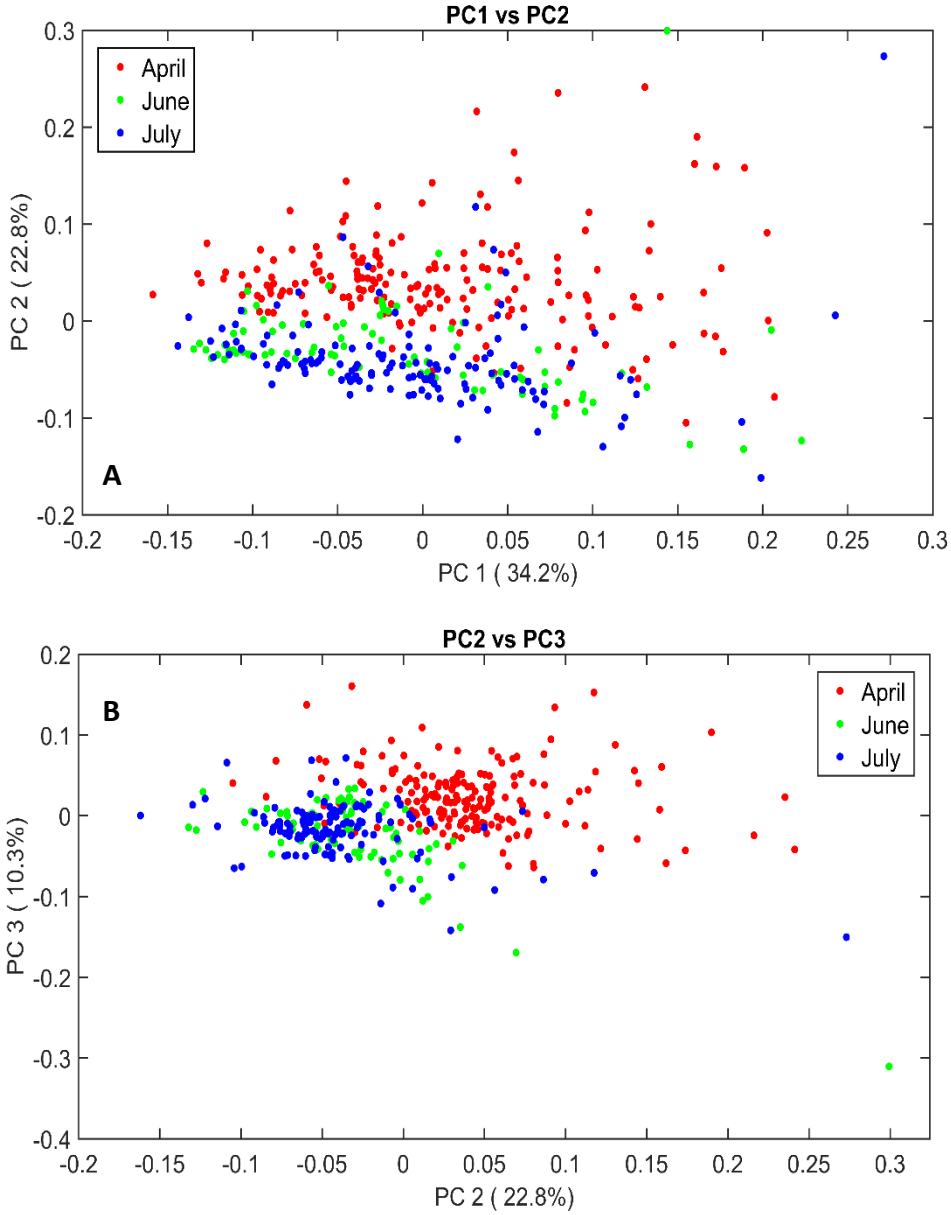

Figure 23: Figure showing a PCA plot using PC1 and PC2 (A) a PCA plot using PC2 and PC3 (B) from Raman data collected with a 785 nm laser. The grouping system used in these data plots looks at the effect that the month in which the data was acquired has on the clustering of the data.

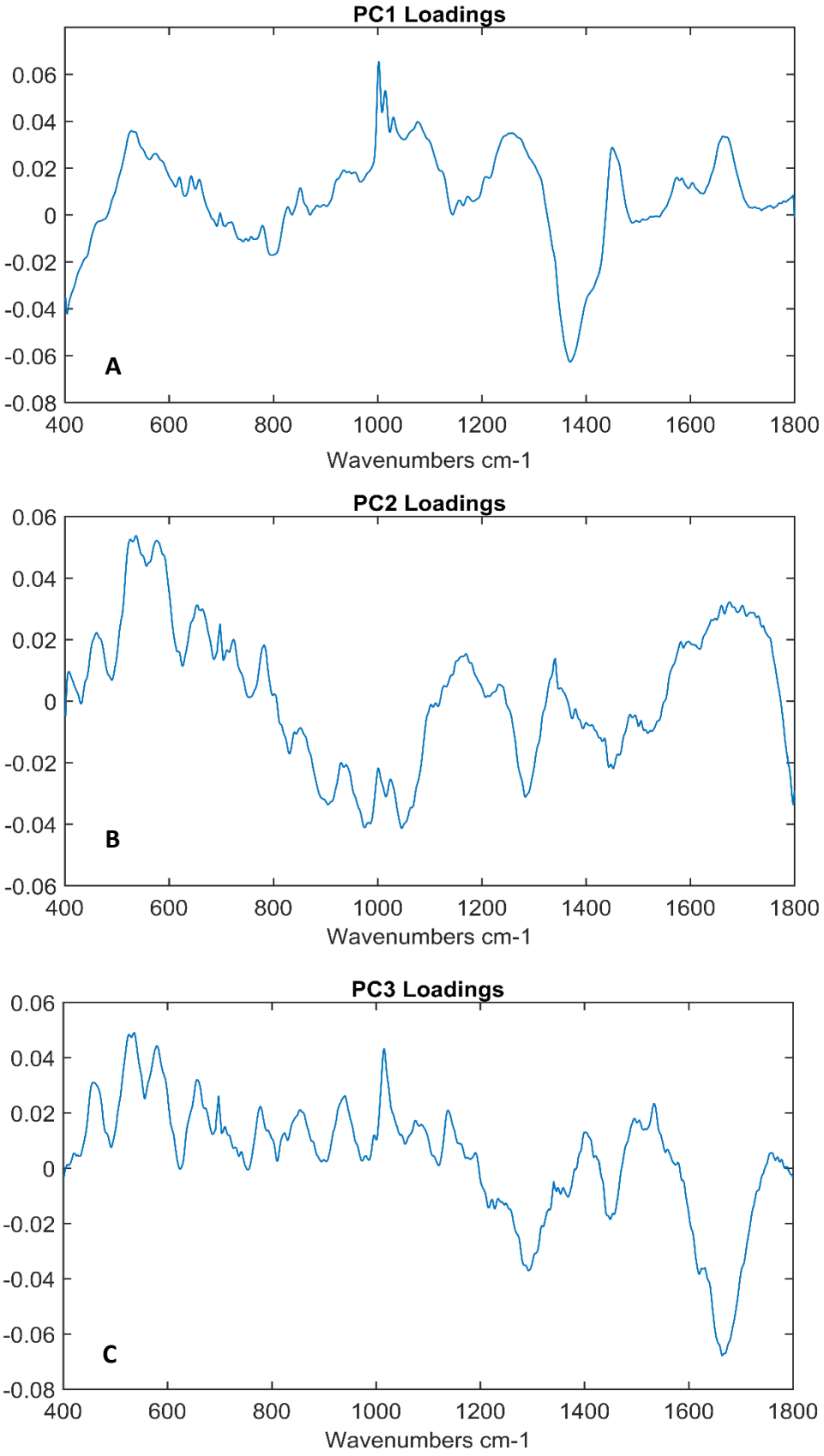

Figure 24: Figure showing a PC1 loadings plot (A) a PC2 loadings plot (B) and a PC3 loadings plot (C) from Raman data.

PCA analysis based on batch number rather than bacterium type shows that there is also some separation of the samples when considering PC2 (see Figure 22). The clustering seen in is more likely to be caused by the fact that there was several months between analysing the batch 1 and 2 samples, therefore the batch 2 samples can be considered as aged. The June and July groups are likely to overlap as there is less time between the analysis time for these samples, therefore the changes caused by sample degradation are going to be less than those observed between April and June or July.

The loadings shown for the PCA plots highlight a number of key vibrations that are responsible for the separation and clustering of the data. Based on the plots it is PC2 that is the most important loadings plot to consider as this is responsible for the separation of April from June and July. The key peaks found within the April data are the C=C torsion and ring torsion of phenyl (461, 537 and 561cm^{-1}) and C-I/C-Br/C-Cl stretches (575 and 589 cm^{-1}). June and July have many vibrations that are present and cause the separation from April. The most significant of these vibrations are peaks within the phosphodiester region (905, 976 and 987 cm^{-1}), C-O/C-C/C-O-H/C-O-C/P=O (symmetric) stretches (1046 and 1055 cm^{-1}), C-N/C-C stretches (1068 cm^{-1}) and CH_2/CH_3 vibrations (1341 cm^{-1}).

When considering the total amount of variance explained after PCA, a total of just over 67% of the variance is explained using the first three PCs for Raman analysis. When compared to FTIR, this is around 19% less than is explained by 1st and 2nd order FTIR data and 25% less than is explained by pre-processed FTIR data. This significant

drop in percentage explained is likely to be cause by both interference from the substrate and a higher level of noise. A summary of the percentage variance for FTIR (including 1st and 2nd order) and Raman is shown in Table VIII.

Table VIII: Table detailing the method of data acquisition, the order of derivatisation and the percentage variance explained by the first three PC's both individually and cumulatively.

	FTIR			**Raman**
Order of derivatisation	None	1st	2nd	None
PC1 (%)	63.1	63.9	55.1	34.2
PC2 (%)	21.4	15	23.7	22.8
PC3 (%)	7.81	7.77	7.2	10.3
Variance using first 3 PC's (%)	92.31	86.67	86	67.30

DFA was performed using the bacterial strain naming system to see if a supervised form of MVA could detect the changes between the bacterial strains that are masked in PCA. The DFA plot with its corresponding loadings are shown in Figure 25 with the tentative peak assignments compiled in Table VII.

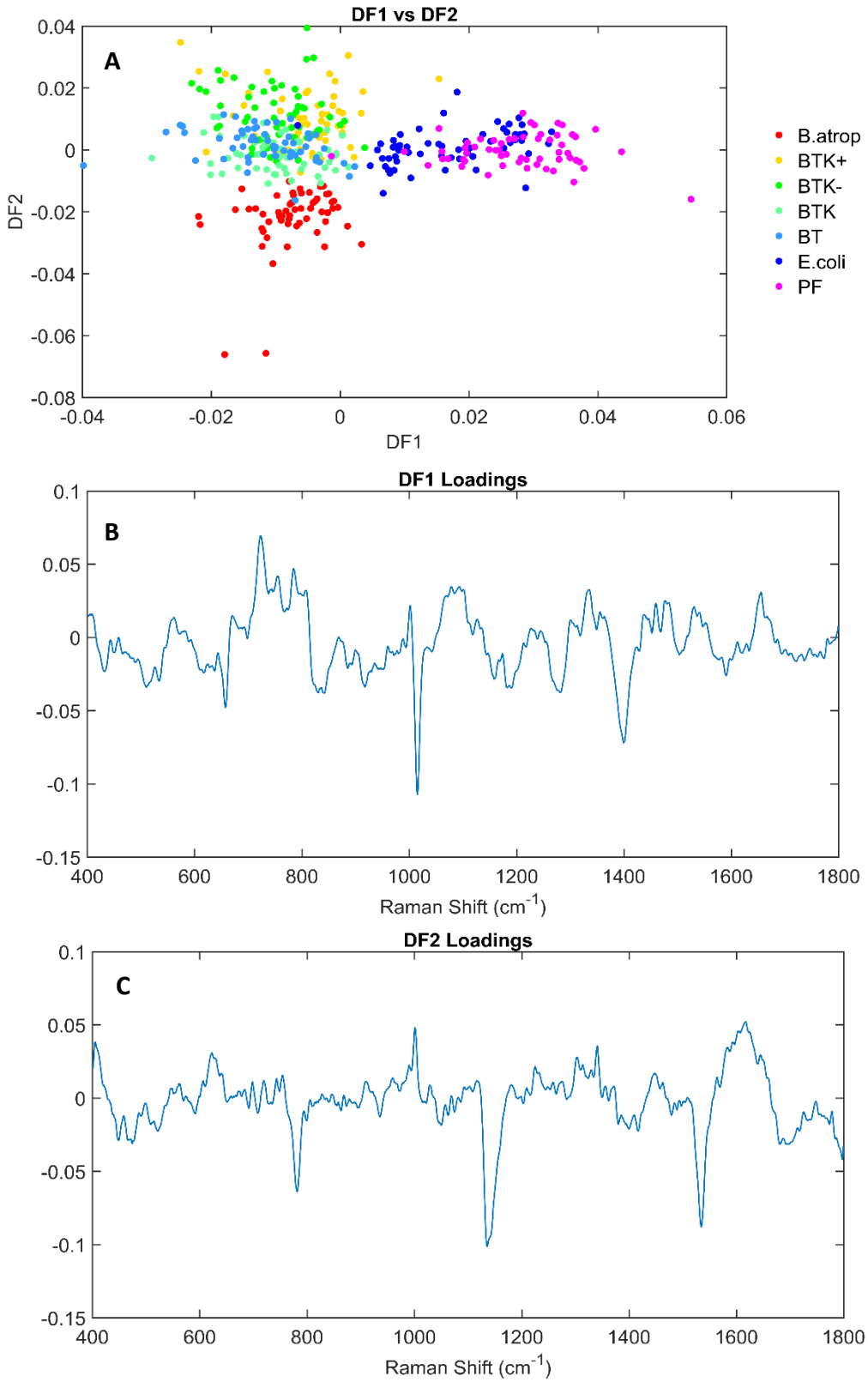

Figure 25: Figure showing a DFA plot using DF1 and DF3 (A) DF loadings plot using DF1 (B) and DF loadings plot using DF2 (C) from Raman data.

Table IX: Wavenumbers and corresponding peak assignments for PC1, PC2, PC3, DF1 and DF2 loadings from Raman PCA and DFA plots. [9,10,11,3,12,]

Wavenumber (cm^{-1})					Tentative peak assignment
PC1	PC2	PC3	DF1	DF2	
	461, 537, 561	458, 536			C=C torsion and ring torsion of phenyl
528	523	526			vS-S
	575, 589	580			vC-I, vC-Br, vC-Cl
1002			1002	623, 1001	Phenylalanine
	698	698	808, 841		δCH out of plane
642	830	855	830		Tyrosine
658	655, 664	657	657		Guanine/Thymine
			723		Adenine
			755		Tryptophan
		778		781	Cytosine/Uracil/Thymine
	905, 976, 987	940			Phosphodiester
1015, 1031	1046	1016	1015		vC-O, vC-C, vC-O-H, vC-O-C vP=O (symmetric) of PO^-_2
1077	1055		1078		vP=O (symmetric) of PO^-_2
	1068	1138	1068, 1158	1135	vC-N, vC-C
			1093		vC-C, vC-O-C glycosidic link
1253	1284	1284, 1293	1268, 1275, 1282		Amide III band of proteins
1370, 1450	1341	1400, 1495	1336, 1400, 1459, 1477 1484	1341	δCH_2 and δCH_3
		1503			δCH of phenyl ring
		1533	1531	1535	Amide II
		1620		1584, 1618	vC-C of phenyl ring
				1593, 1603,	vC=N, vNH_2 adenine
	1660, 1675	1657, 1664, 1670	1656	1606, 1625, 1646, 1652, 1660, 1696	Amide I
1663					vC=O Cytosine, uracyl
				1681	vC=O Guanine deformation N-H in plane
	1701				vC=O Guanine

DFA of the samples shows that despite the MirrIR slide having a significant impact on the spectral profile for each of the samples it is possible to see both separation and clustering of the seven bacterial strains used within the study. DF1

clearly separates the Gram-negative from the Gram-positive strains. From the DF1 loading (Figure 25B), it is the out of plane C-H stretches found at 808 and 841 cm^{-1}, Adenine (755 cm^{-1}) and Tryptophan (723 cm^{-1}) that cause the Gram-negative *E. coli* and PF to separate from the Gram-positive *Bacillus* strains. DF2 is responsible for separating *B. atrop* out in the negative region with respect to DF2 with BT, BTK+ and BTK- separating in the positive region with respect to DF2. The remaining bacterial strains can be found within the centre of the plot with respect to DF2. The vibrations that are key in separating *B. atrop* from the other bacterial strains are cytosine/uracil/thymine (781 cm^{-1}), C-N/C-C stretches (1135 cm^{-1}) and Amide II (1535 cm^{-1}).

The DFA results show that when a supervised form of MVA analysis is used the underlying factors that affect the dataset and thus impact the PCA analysis have less influence on the DFA separation and clustering. That is not to say that these factors, such as sample degradation and substrate influence, don't have an impact on the separation of the data. If you were to compare the DFA plots from the ATR-FTIR and Raman datasets you can see that there is clearer separation of the different groups with the ATR-FTIR data. If these factors were to be removed from the Raman dataset it may be possible to see enhanced separation between the bacterial strains, particularly with respect to the *Bacillus* strains.

3.4.Conclusions

Growth curve analysis of the seven bacterial strains chosen for use in this study showed that 24 hours is the optimal time for the samples to be grown at. The optimal temperature used for growth is 37°C for all strains except PF for which a lower temperature of 30°C is needed.

Four different methods were trialled to see how the order that the pre-processing is performed effects the separation and percentage variance explained seen data from the same FTIR instrument. It was found that Method 3, where the data was noise-reduced followed by vector normalisation and cutting to the fingerprint region, was most optimal for this FTIR dataset as the PCA plot has the best separation seen across the four methods. A study to assess the impact that analysing data in their derivative form was performed. This study found that MVA performed on pre-processed but non-derivative data performed the best as separation of the groups was clearer. Also, this methodology is more optimal as it has one less analysis step and produces loadings plots that are easier to interpret. This method of pre-processing FTIR data is summarised in Figure 4.

The results from the technique downselection study show that species-based separation can be observed using both FTIR and Raman spectroscopy using the semi-supervised technique of DFA. FTIR shows that separation can be seen between the groups regardless of derivative state, with a higher level of separation seen using the unsupervised method of PCA when compared with Raman.

Over 85% of the total variance can be explained using just the first three PC's with any of the derivative forms of the FTIR data with this being over 90% for pre-processed, non-derivative data. However, with the Raman dataset the first three PC's explain just over 67% of the variance, comparison is shown in Table VIII. Using a higher number of PCs for PCA and DFA analysis may improve the separation observed in the plots, however, by using more PCs the amount of noise included in the model will increase. As mentioned previously, the consequence of this is the loadings will be noisier with a higher chance that peaks will relate to non-biological features.

Raman data shows the technique is significantly affected by non-ideal substrates, in this case a MirrIR slide. PCA shows no separation of the data based on bacterial strain. DFA shows clustering of bacterial strains, however the separation seen is less than in DFA analysis of FTIR data. When looking forward to future work and the importance for the technique to work well on non-ideal substrates, FTIR appears to be the more favourable technique. This is due to both its ability to work well across a wide range of substrates and a greater penetration depth which is useful for the more porous substrates. This, coupled with the higher percentage variance explained and better visual separation of the data, has led to FTIR being the technique chosen to focus future analysis on.

3.5.References

1 Madigan, M. T., Martinko, J. M., Stahl, D. A. & Clark, D. P. *Brock Biology of Microorganisms*. 13th edn, (Pearson Education Inc., 2012).

2 Burton , Z. F. & Kaguni, J. M. Ch. 3. Bacteria: Composition, Organization and Growth, 21-28 (Academic Press, 1997).

3 Baker, M. J., Hughes, C. S. & Hollywood, K. A. *Biophotonics: Vibrational Spectroscopic Diagnostics*. (Morgan & Claypool, 2016).

4 Lasch, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging *Chemometrics and Intelligent Laboratory Systems* **117**, 100-114, doi:10.1016/j.chemolab.2012.03.011 (2012).

5 Trevisan, J., Angelov, P. P., Scott, A. D., Carmichael, P. L. & Martin, F. L. IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis. *Bioinformatics* **29**, 1095-1097, doi:https://doi.org/10.1093/bioinformatics/btt084 (2013).

6 Hedge, M. *Microbiology: Do doubling rates differ between E. coli grown on LB agar vs. liquid media?*, <https://www.quora.com/Microbiology-Do-doubling-rates-differ-between-E-coli-grown-on-LB-agar-vs-liquid-media> (2013).

7 Naumann, D. FT-Infrared and FT-Raman spectroscopy in biomedical research. *Applied Spectroscopy Reviews* **36**, 239-298, doi:https://doi.org/10.1081/ASR-100106157 (2001).

8 Pommerville, J. C. Ch. 15. Microbial Growth: Analysis of a Bacterial Growth Curve, 145-152 (Jones & Bartlett Learning, 2007).

9 Horiba. *Raman Data and Analysis*, <http://www.horiba.com/fileadmin/uploads/Scientific/Documents/Raman/bands.pdf> (

10 Clemens, G. N., Hands, J. R., Dorling, K. & Baker, M. J. Vibrational Spectroscopic Methods for Cytology and Cellular Research. *Analyst* **139**, 4411-4444 (2014).

11 Movasaghi, Z., Rehman, S. & ur Rehman, I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews* **43**, 134-179 (2008).

12 Naumann, D. *Infrared Spectroscopy in Microbiology*. 1-18 (2006).

13 Schünemann, R., Knaak, N. & Fiuza, L. M. Mode of Action and Specificity of Bacillus thuringiensis Toxins in the Control of Caterpillars and Stink Bugs in Soybean Culture. *International Scholarly Research Notices Microbiology* **2014**, 1-12, doi:dx.doi.org/10.1155/2014/135675 (2014).

14 Palma, L., Muñoz, D., Berry, C., Murillo, J. & Caballero, P. Bacillus thuringiensis Toxins: An Overview of Their Biocidal Activity. *Toxins* **6**, 3296-3325, doi:10.3390/toxins6123296 (2014).

15 Ashton, L., Lau, K., Winder, C. L. & Goodacre, R. Raman spectroscopy: lighting up the future of microbial identification. *Future Microbiology* **6**, 991-997 (2011).

16 Jarvis, R.M. & Goodacre, R. Characterisation and identification of bacteria using SERS. *Chemical Society Reviews* **37**, 931-936, doi:10.1039/b705973f (2008).

17 Butler, H. J. *et al.* Using Raman spectroscopy to characterize biological materials. *Nature Protocols* **11**, 664-687, doi:10.1038/nprot.2016.036 (2016).

Chapter 4

Method Development for *in situ* Analysis of Surface Deposited Bacteria Using DRIFTS and the Effects of Environmental Conditioning

4.1 Introduction and Aims

This study looks at surface deposited bacteria applied to a variety of surfaces that were exposed to varying temperatures. This study looks to further investigate the ability of vibrational spectroscopy to identify bacteria by considering the surface a bacterium is found on and the environment it has been exposed to. Ultimately this study aims to move the research away from the laboratory and tackle some of the issues that samples from a 'real world' setting might present.

To do this, the first section of this chapter discusses the process of developing a suitable methodology for *in situ* analysis of samples using a handheld spectrometer. It is important to have reference spectra for the substrates available to fundamentally understand the characteristics of each of the substrates for every modality. For this reason, a set of spectra were acquired using Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) for the five substrates used within this study. The

substrates used in this study were: boot, concrete, jacket, lino and metal tile. Both the clean and the dirty filter were removed from this study as they were not part of the original five chosen as the core substrates. MirrIR was also removed from the study as it was previously included simply for purpose of having a standard substrate for use with FTIR and Raman. These reference spectra are subtracted from spectra collected from surface deposited bacteria which is discussed later in this chapter. Development of a suitable method for use on a handheld spectrometer that can be used to identify surface deposited bacteria is the key aim of the overall project, the development of this method is discussed within this chapter.

The main study discussed in this chapter looks at the effect that different temperature has on the spectral signature of a bacterium. An additional study using the same seven bacterial strains and five substrates used in the temperature study was carried out to assess the impact that an alternative environmental condition would have on the spectral profile. This study used solar radiation, at standard room temperature as an example of an alternative environmental factor that can have an impact on the profile of a bacterium.

4.2 Methods and Materials

4.2.1 Substrate Preparation

Substrates were not cleaned prior to analysis to ensure that they were representative of a 'real world' scenario, and were substrates that have the potential to be contaminated if a BWA were used. The substrates used in this study were supplied by DSTL with the exception of the metal tile which is commercially available. The substrates that would be coming into contact with the agent would not have been cleaned prior to their contamination and as such would have a level of background contamination already present. For this reason, it was important to not clean the substrates prior to analysis to assess how the method could work with this background contamination as well as the surface deposited bacteria. The seven bacterial strains were spotted onto different pieces of boot, jacket and lino. For concrete and tile, seven bacterial strains were again spotted onto the substrates however, only data from two bacterial strains is shown. Issues with the way the samples dried onto these surfaces meant that there was no biological information shown in the spectra obtained. The same method of preparation was used with all of the substrates and strains. For this, two different strains were applied to the same piece of substrate but importantly in different areas with the areas clearly labelled and no contact between strains.

4.2.2 DRIFTS

As with Chapter 2 (section 2.1), it was important to compile a spectral database for the substrates to ensure that reference spectra are available for the removal. The first section of this chapter discusses the building of the spectral database for DRIFTS. The overall aim of this PhD project is to develop a methodology that is suitable for *in situ* analysis of surface deposited samples. With this in mind the methodology aimed to be robust, simple and easy to perform. This chapter aims to build a robust methodology, developed on a handheld instrument that was used to identify surface deposited bacteria found on complex matrices.

Spectra were collected using 32 co-added scans with 4 cm^{-1} resolution over a range of 4000-600 cm^{-1} using the DRIFTS accessory fitted with a coarse Gold reflectance cap for background measurements. A background spectrum was collected prior to the collection of spectra from the substrate. Background scans were acquired using the same set of parameters as for sample measurements. With ATR the bacterial sample is in direct contact with the attachment therefore, a background is needed either between every individual spectrum or new area of the substrate. However, as the bacterial sample does not come into contact with the attachment when using DRIFTS, backgrounds can be collected less frequently. Consequently, a background spectrum was acquired every 10 minutes. Background spectra were collected to ensure that any changes occurring in the atmosphere during the course of analysis were taken into account and corrected for regularly. Spectra were acquired from four different areas of the substrate, with three technical replicates

collected per area, to create an average representative spectrum for each substrate. Spectra were then averaged to generate an average diffuse reflectance spectrum per substrate. To ensure that the best possible spectral quality and reproducibility was achieved the handheld spectrometer used for analysis was clamped in an upright position so that both hands were free, one to hold the sample in place against the accessory and the other to work the instrument. This setup for this can be seen in Figure 1. In a 'real world' scenario the instrument would not be held up with a clamp. In this situation the surfaces that are being examined are likely to be bulky and unlikely to move. Smaller items that may move while being analysed can be held in place by a colleague.

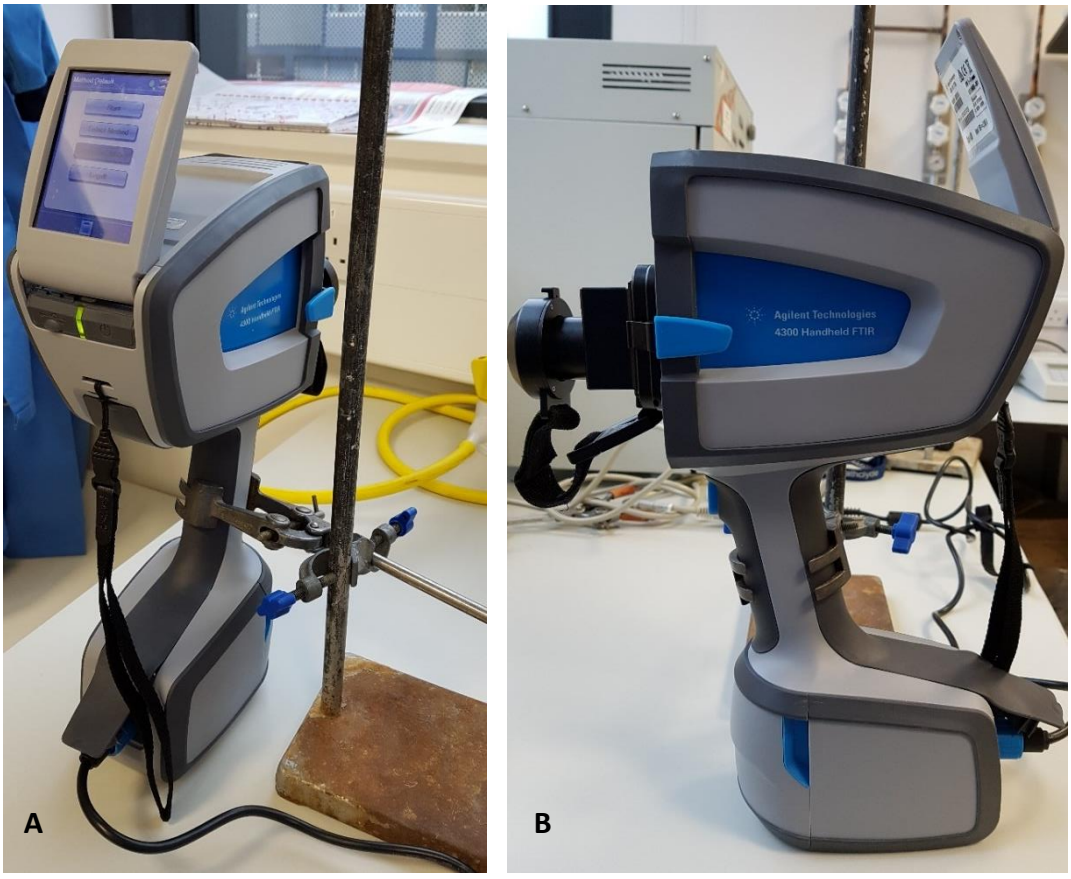

Figure 1: Figure showing the handheld spectrometer setup using a clamp to hold the spectrometer in place and the PDA raised for ease of use (A) along with a side view showing the attachment of the diffuse reflectance accessory with a coarse Gold reflectance cap fitted (B).

All seven of the bacterial samples provided by the Defence Science and Technology Laboratories (Dstl) were prepared to an Optical Density (OD) of 20, using the same method discussed previously in Chapter 3 (section 2.3). From these samples four spots, each of 2µl, were pipetted onto each of the substrates (boot, concrete, jacket, lino, tile). Two sets of samples were created, one of which acted as a control set. The control set were analysed at the end of the experiment with the non-control set analysed at the start and end as well as two intermediate time-points. These time-points were just after deposition, within the first 5 hours (T0), 24 hours after deposition (T1), 7 days after initial deposition (T2) and 14 days after initial deposition (T3), which was the final time point. Before the analysis of any samples, sets were placed into an incubator set to the temperature being used for the study (either 30, 37 or 49°C) to dry ~1 hour. This allowed the bacterial deposits to dry enough for the substrates to be held vertically and the spots not move or drip into the accessory. Once the samples were dried the non-control samples were analysed while the control samples remained in the oven. Once the samples had been analysed they were stored in clear petri dishes before then being transferred back into the incubator.

The control set of samples were only analysed at the end (T3) so that the samples are left intact, unaffected by any damage that may be caused to the samples by analysis. The samples also remained in the environment for a full 14-day period, ensuring that the samples are not exposed to any other environment, including solar radiation. Having a control set of samples ensures that any differences observed in the spectra are not as a result of the samples being repeatedly removed from a given

environment and exposed to a different one for the period of analysis time. It also reduces the chance of the samples getting contaminated or damaged, both of which would affect the spectra produced.

Once the spectra have been collected they have the substrate background manually removed using the Beer-Lambert equation. This uses the methodology developed with samples of glucose and ovalbumin in Chapter 2 (section 2.4), this time using DRIFTS rather than ATR. Although the modality is different, the principle is the same as there has still been no compensation for the substrate background.

Each environmental condition used five substrates, seven bacterial strains and four time-points. The substrate was prepared with four spots where three spectra were collected per spot. Each of the individual spectra were processed, using the Beer-Lambert equation, to produce an absorbance spectrum for every DRIFTS spectrum.

4.3 Results and Discussion

This section is split into clearly defined sections where the first half of the section discusses the data collected from the DRIFTS substrate study before moving on to discuss the environmental dataset.

4.3.1 DRIFTS

This part of this DRIFTS section covers the substrate study that was essential for generating standard substrate spectra that were to be used in the DRIFTS environmental study. This involves discussing the key vibrations observed from the five substrates downselected for use within the environmental study.

4.3.1.1 DRIFTS Substrate Library

As with ATR-FTIR and Raman spectroscopy a spectral database has been compiled from each of the five different substrates down-selected for use within this study. An overlay of the spectra collected for boot and tile can be seen in Figure 2 as an example of the quality of spectra that can be achieved using the DRIFTS accessory on the handheld FTIR spectrometer with the wavenumbers and corresponding peak assignments for these substrates compiled into Table I. The spectra, and peak assignments, for the remaining substrates can be seen in Appendix C.

Table I discusses the key vibrations present in the spectra acquired from boot and metal tile. The tile spectrum in the regions of 4000-3500 cm^{-1} and 1800-1400 cm^{-1} is clearly affected by noise and water vapour. The peaks seen in the region of 2500-3000 cm^{-1} of the tile spectra are more likely to be residues found on the surface of the substrate rather than being a part of the substrate itself, due to the incredibly low absorbance levels observed in Figure 2B. The vibrations that correlate to this range are caused by CH, CH_2 and CH_3 vibrations found in lipids. The residues may have

been deposited by the protective cover that is found on the tile to protect them from damage during transportation, or during the handling of the tile at any stage during the lifetime of the substrate.

Table I: Table with the wavenumbers and corresponding tentative peak assignments for the average spectrum collected from boot and tile, acquired using DRIFTS [1,2,3]

Wavenumber (cm^{-1})	Tentative peak assignments
976.6	vCC alicyclic, aliphatic chains
1006, 1034, 1167, 1204, 1241, 1247	vC=S, vCC alicyclic, aliphatic chains
1083, 1094, 1101, 1114, 1120, 1126, 1134, 1140, 1146	vC-O-C asymmetric
1258, 1267, 1286	vCC alicyclic, aliphatic chains
1336, 1387	vC-NO_2
1393, 1409, 1417, 1456,1465	vCH_2, δCH_3 asymmetric
1549, 1554, 1565, 1582	Amide II
1638, 1647, 1668, 1674, 1680, 1688	Amide I
2851, 2857, 2900, 2920, 2930, 2956	vC-H
3083, 3332	vO-H

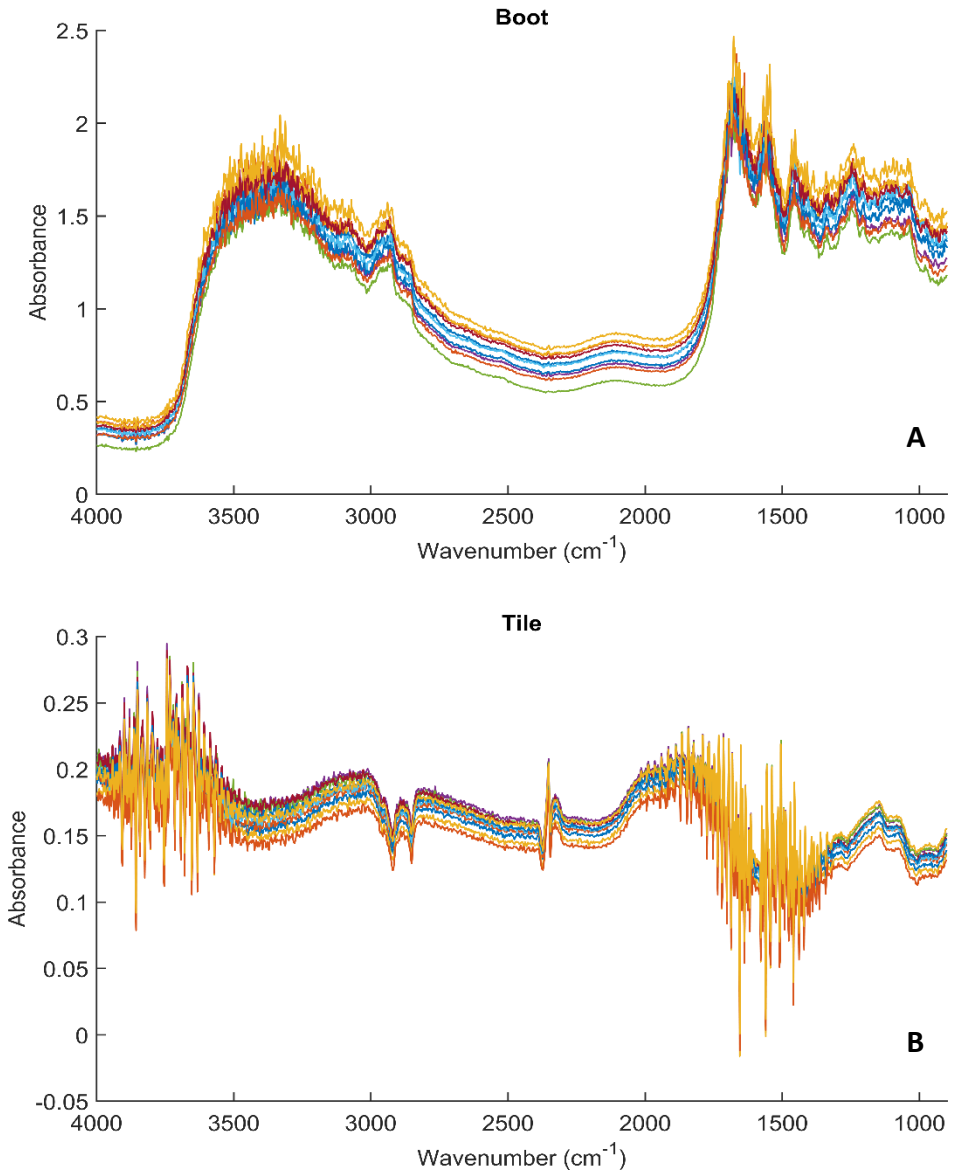

Figure 2: Spectra collected using DRIFTS showing an overlay of 10 spectra for Boot (A) and Tile (B).

4.3.1.2 DRIFTS Environmental Conditioning Study

This DRIFTS study looks at the combined effect of the environment and the substrate that the bacteria has been deposited on. As with the glucose and ovalbumin studies discussed in Chapter 2 (section 2.4), the quality and variability of the absorbance spectra generated when using the Beer-Lambert equation vary with the substrate used. The more porous substrates such as boot, concrete and jacket produce less intense, noisier spectra than the non-porous substrates such as lino and tile. A general trend seen across all of the temperature studies, as well as all of the substrates, is that there is overcompensation. Overcompensation when using the standard substrate spectra causes features of the absorbance spectra to be inverted. Despite the features of the spectrum being inverted the biological features of a bacterial spectrum are still present. Although spectra were collected across the full spectral region (4000-600 cm^{-1}), only the fingerprint region (1800-900 cm^{-1}) was interrogated during MVA. When performing the Beer-Lambert substrate removal with the tile the region where the Amide I and II bands are found became distorted due to the effect of the noise/water vapour in this region. Due to the very limited spectral contribution from the substrate (Figure 2B), a representative bacterial spectrum can be seen without having to perform Beer-Lambert removal (Figure 3). For this reason, the analysis of data collected from tile was performed on the raw data cut to the fingerprint region.

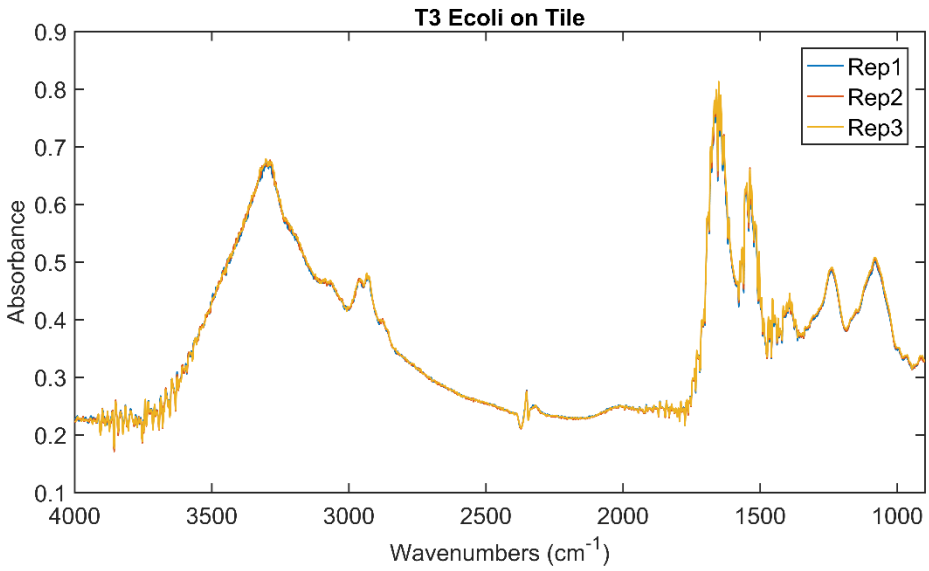

Figure 3: Example of a spot of E. coli on tile, collected at T3 during the 49C study, showing a typical bacterial spectrum.

One issue encountered during the analysis of bacterial samples prepared on some of the substrates used in the study is that the spots were easily damaged or struggled to stay attached to the surface of the substrate for the duration of the experiment. This was a common trend across the non-porous substrates of lino and tile. Another trend also seen is that with the higher temperatures more of the spots were damaged or became detached from the surface, as a consequence fewer bacterial strains were available to be examined during MVA. Despite this, there was always a minimum of four out of the seven strains with a mix of Gram-positive and Gram-negative bacteria which still allows for a good comparison.

Once the data had been processed and absorbance spectra produced for each DRIFTS spectrum, MVA could be performed. As with the glucose and ovalbumin spectra they were not pre-processed in any way before PCA and DFA were performed. It was chosen not to pre-process the spectra in the interest of developing

a methodology that can be translated for the analysis of *in situ* samples. To do this, the number of steps between collecting a spectrum and performing MVA where an identification can be made were kept to a minimum. This allowed the model's ability to separate different bacterial strains to be tested. While pre-processing has previously been optimised for spectra collected using the benchtop instrument, there are a number of parameters for each of the processing steps which can be modified by the user, based on the quality of the spectra acquired. In the interest of developing the simplest methodology for non-specialist users, pre-processing was kept to a minimum for the data acquired using DRIFTS.

Prior to PCA and DFA being performed, the data were clustered into new datasets where all seven strains could be examined at the same time. For this the spectra were grouped based on the substrate that the bacteria had been deposited on. Data was not grouped by strain to look at the substrate for analysis as in a 'real world' scenario the substrate would be a known variable. Therefore, a database can be chosen based on substrate to enhance classification. PCA was carried out on all bacterial strains and time points for a single substrate. Plots for the first three PCs and DFs were generated to try and identify which PCs/DFs represents the most discriminatory information. A selection of key figures are shown within this chapter, further PCA and DFA figures can be found in Appendix C. Due to the amount of data being examined the clearest separation between the data is not always seen between the first two PCs/DFs. The plots chosen for inclusion in this chapter use different PCs/DFs depending on which show the clearest separation however, only data from the first three PCs/DFS are used as this represents the majority of the variance seen

within the datasets. PCA and DFA of the concrete dataset shows a varying amount of separation across the temperatures. Figure 4 shows PCA plots produced from spectra collected from concrete in the 37°C study where some separation and clustering can be seen based on strain (A) and a PCA plot produced from concrete in the 30°C study where the is no clear separation of data based on strain (B).

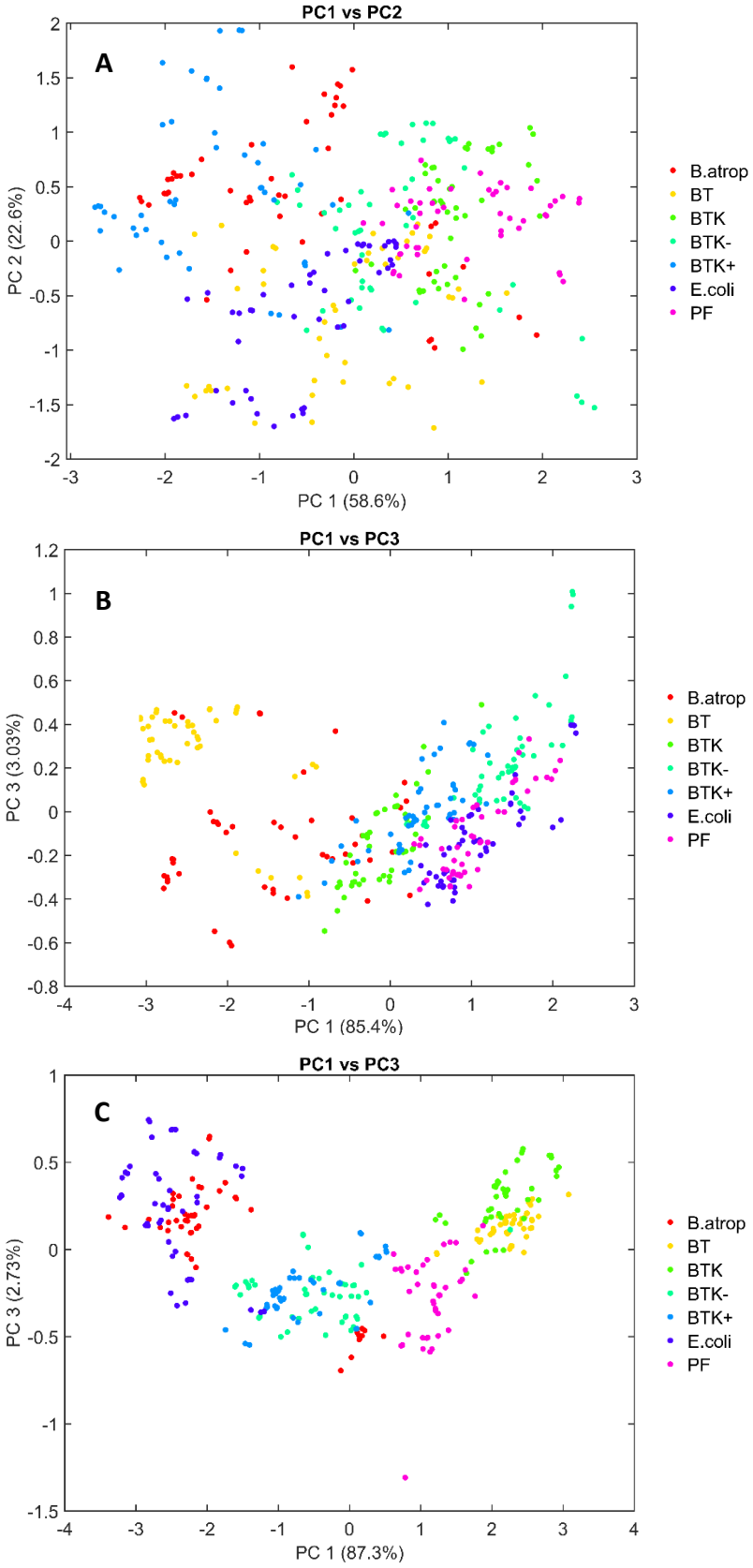

Figure 4: PCA plots showing separation of bacterial strain for spectra collected from concrete during the 30°C study (A), from concrete during the 37°C study (B) and from concrete during the 49°C study (C).

From the plots shown in Figure 4 there is a varying level of both separation and clustering of the seven different bacterial strains across all three temperature environments. In Figure 4A there is very little separation, however there is some loose clustering of the strains with some overlap between groups of bacteria. Therefore, the plot can be split into three sets of clustering: *E. coli* and BT, *B. atrop* and BTK+ and BTK, BTK- and PF.

When considering the data presented in Figure 4B and Figure 4C for the 37°C and 49°C studies, it can be seen that there is a slight horseshoe formation to the spread of the data. [4] This spread is caused by non-linearity between the groups of data so caution must be used when interpreting the results. Figure 4B sees *B. atrop* and BT separating from the remaining five bacterial strains. BTK seems to separate out best from the other strains. However, in this study there is an overlap of the Gram-negative bacteria which isn't seen in the 30°C study. Figure 4C shows the best combined separation and clustering across the three studies. The plot can generally be split into four sections with PC1 and PC3 splitting the groups. The clustering shows overlap of: BT and BTK, BTK+ and BTK-, *B. atrop* and *E. coli* with PF not overlapping with any other strain. As with the 30°C study there is no separation of the Gram-negative from the Gram-positive bacteria.

This PCA analysis was then followed up by DFA to try to improve the split of the data seen. Figure 5 shows the corresponding DFA plots to those shown in Figure 4, where there is a significant improvement in the data from both studies showing more clustering and separation based on bacterial strain.

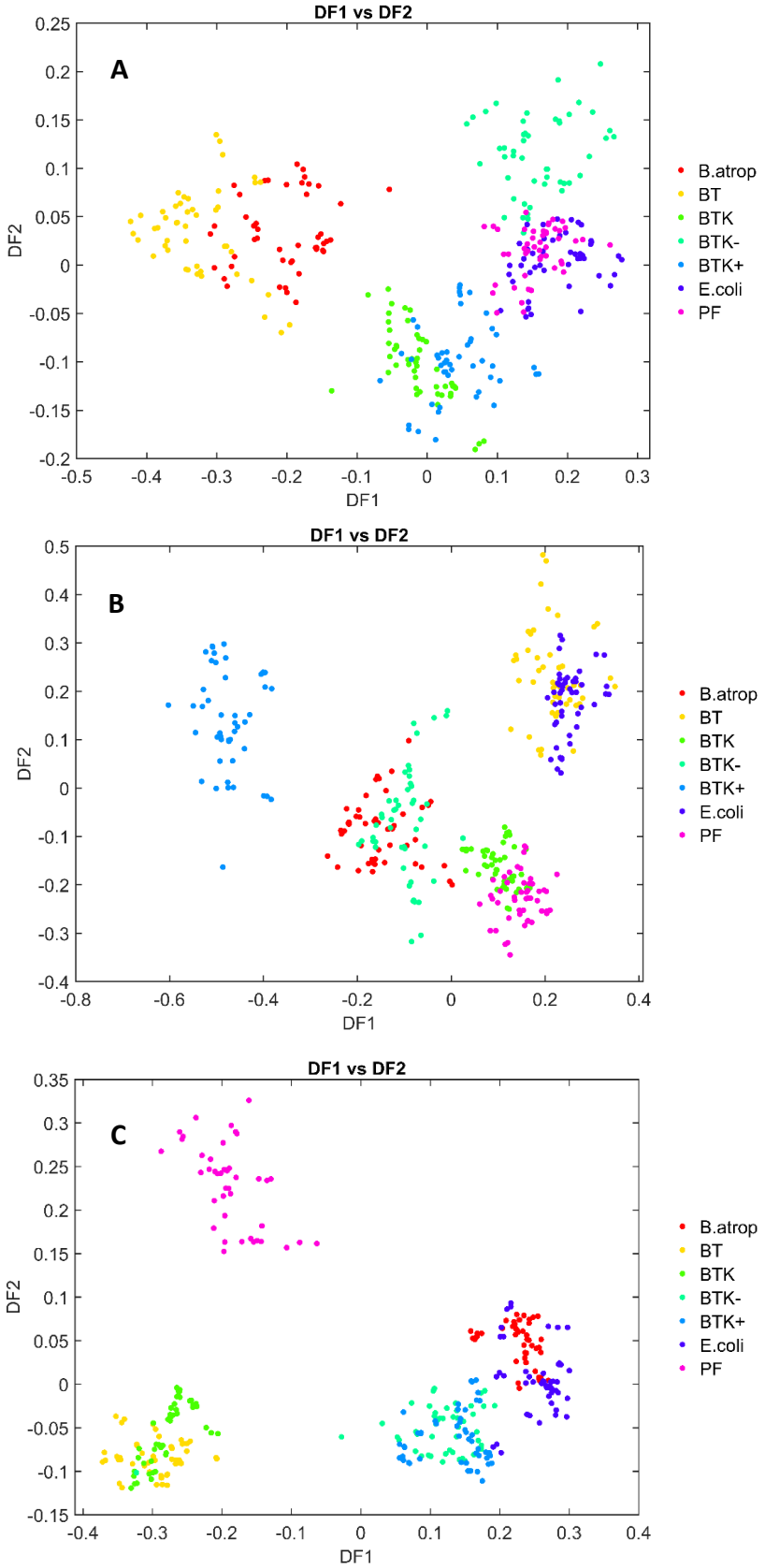

Figure 5: DFA plots showing separation of bacterial strain for spectra collected from concrete during the 30°C study (A), from concrete during the 37°C study (B) and from concrete during the 49°C study (C).

The DFA plots across all three of the temperature studies shows clearer separation and much more defined clustering of the bacterial strains in comparison to the equivalent PCA plots. DF1 in Figure 5A separates BT and *B. atrop* from the other bacteria strains. Also separated is BTK-, PF and *E. coli* with the Gram-negative strains overlapping with each other which was not seen in the PCA plot shown in Figure 4A. BTK and BTK+ are in the centre of the plot with regards to DF1. DF2 then separates BTK and BTK+ from the other five strains. DF2 also separates BTK- from the Gram-negative strains on the right of the plot.

In Figure 5B and C there is no overlap or separation of the Gram-negative bacteria. Figure 5B uses DF1 and DF2 to separate the bacteria into four distinct clusters with overlap of some of the strains. PF and BTK, *E. coli* and BT, BTK- and *B. atrop* and BTK+ found on its own. DF1 separates BTK+, BTK- and *B. atrop* from the others strains with DF2 separating BTK+, *E. coli* and BT from the other four strains.

As with Figure 5B, there are four defined clusters in Figure 5C, however this time it is PF that is the strain that does not overlap with the other bacterial strains. The three overlapping clusters are: BT and BTK, BTK- and BTK+ and *E. coli* and *B. atrop*. DF1 clearly separates PF, BT and BTK from the other four strains with DF2 separating PF from all of the other strains. DF2 can also be seen to be starting to split *B. atrop* from *E. coli.* The DF1 and DF2 loadings from each of the DFA plots shown in Figure 5 can be seen in Figure 6, Figure 7 and Figure 8.

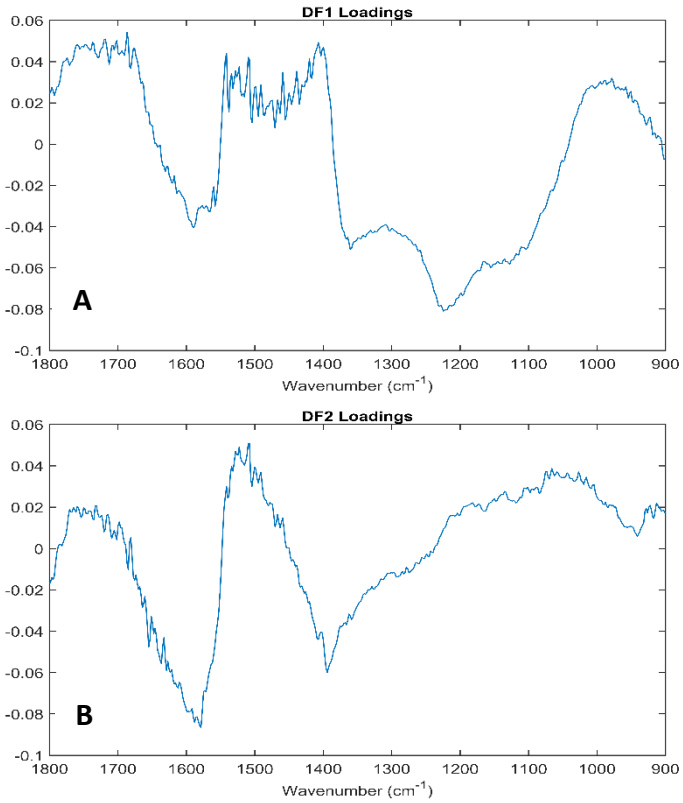

Figure 6: DFA loadings for the 30°C study showing DF1 loadings (A) and DF2 loadings (B).

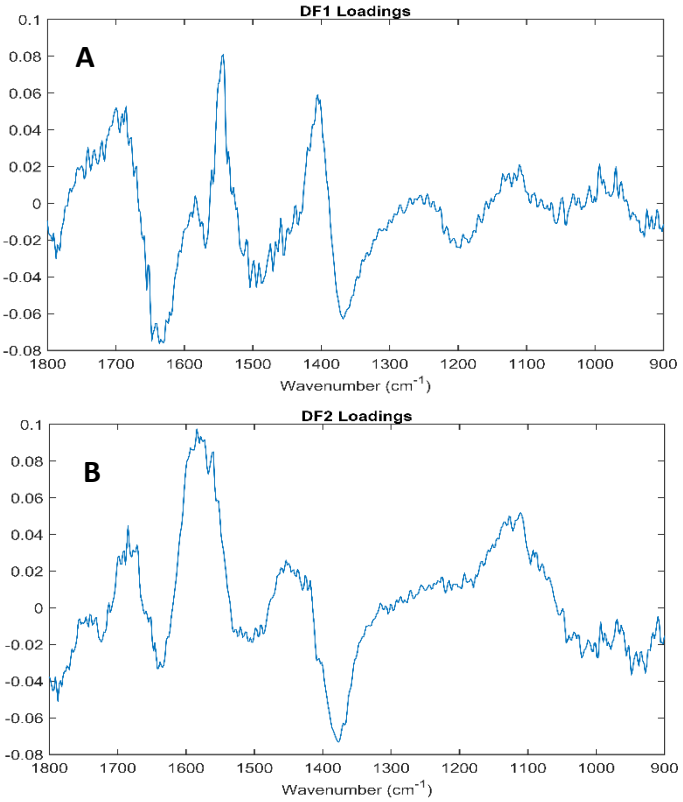

Figure 7: DFA loadings for the 37°C study showing DF1 loadings (A) and DF2 loadings (B).

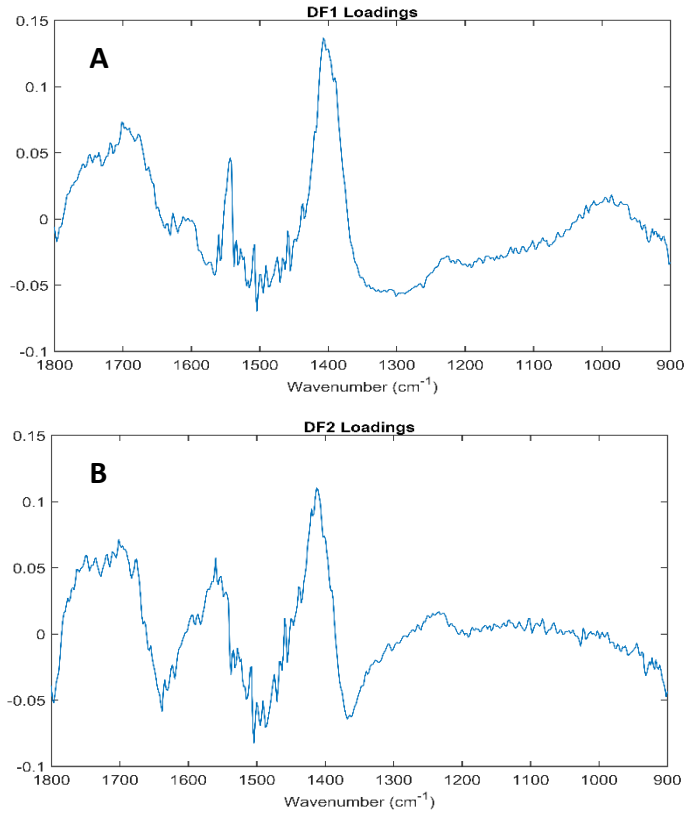

Figure 8: DFA loadings for the 49°C study showing DF1 loadings (A) and DF2 loadings (B).

The loadings for the species based study using DRIFTS can be seen in Figure 6, Figure 7 and Figure 8 with a common trend observed across the three different temperature studies is that the loadings spectra generated to describe the separation seen in the DFA plots are significantly affected by noise with very few biological features protruding through. Tentative peak assignments were provided for the peaks that are classed as significant, based on intensity, and therefore should represent physical features rather than simply identifying noise. All peak assignments have been collected into a single table that represents peaks present in the spectra collected across the different temperature studies, see

Table II.

Based on Figure 6 there are a number of different peaks represented by DF1 that are not represented by DF2. These include C=O/C-O, Amide I, phosphate vibrations and peaks within the phosphodiester region. The key vibrations within these are the ones located in the low wavenumber region representing the differences in the phosphate groups which are causing the Gram-negative bacteria, and BTK-, to separate from the other strains. The one unique vibration detected in DF2 is a series of peaks found within the Amide II region. As previously discussed changes in the Amide II region relate to the outer membrane and LPS layer found within Gram-negative bacteria.

As with the loadings plots for the 30°C (Figure 6) the loadings for the 37°C study show there are several key vibrations represented in DF1 that are not present in DF2 (Figure 7), as they are separating based on different bond vibrations. The bulk of the peaks fall within the Amide I region with C=O and CH vibrations also present. DF2 has fatty acid ester groups (1787 cm^{-1}) and phosphodiesters vibrations (1111 cm^{-1}) that are not seen in DF1.

The loading plots for the 49°C study, shown in Figure 8, have the most similarities between the groups being represented in the DF1 and DF2 loadings across the three temperature studies. Peaks attributable to Amide I, II and C=O are present in both DF1 and DF2 loadings plot as well as CH ring stretches and CH_2 wagging. Figure 8A has peaks at 1301 and 1262 cm^{-1} which are characteristic of Amide III vibrations. DF2 (Figure 8B) has fatty acid esters at 1797 and 1750 cm^{-1} as well as a CH_2 bend of a methylene chain found in lipids at 1470 cm^{-1}. No vibrations consistent with phosphate

groups are present in either of the loadings plots as there is no clear separation of the Gram-negative strains from the Gram-positive ones. Concrete is a non-homogenous, porous substrate that allows the samples absorb into the surface. Due to the penetration depth of the technique, and the strong spectral profile of the concrete (Appendix C), it is possible that not all of the bacterial signatures will be detected following removal of the spectral contribution from the substrate.

Table II: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots from the data collected from all species deposited on concrete during the 30°C, 37°C and 49°C studies [1,2,3]

Wavenumber (cm^{-1})	Tentative peak assignments
1111, 1128	Phosphodiester region, vCO, vCC, vCOH and vCOC
1155	vC-O
1224	Phosphate II (stretching PO_2^- asymmetric vibrations) in β-form DNA
1230	Asymmetric PO_2^- stretching vibrations in DNA and RNA
1262, 1301	Amide III
1353, 1359, 1360, 1362, 1368, 1377, 1390, 1394	CH_2 wagging
1400, 1401, 1405, 1407, 1413, 1420	Ring stretch vibrations mixed with CH in-plane bending
1470	CH_2 bend of methylene chain in lipids
1485, 1487, 1495, 1504, 1508, 1515, 1519, 1543, 1551, 1554, 1560, 1573, 1579, 1584, 1588 1593	Amide II (NH bend coupled to CN stretch)
1620, 1623, 1631, 1636, 1638, 1647, 1675, 1677, 1685, 1687, 1690, 1696	Amide I
1700, 1702, 1711, 1718, 1735	vC=O
1750, 1787, 1797	Fatty acid esters

After the DFA analysis had been performed the dataset was split down into smaller datasets to allow for more exploratory analysis. The data was split into individual strains on each of the substrates meaning that bacterial strain, temperature and substrate are fixed variables with length of exposure being the only

variable examined. This allowed for the examination of the affect that the prolonged exposure to a given environment has on the spectral profile of a bacterium. The splitting of the dataset was carried out across all substrates as well as all four environments, three different temperatures and a solar study.

PCA and DFA have been carried out with all of the datasets; however, only the key figures are shown within the chapter. The figures show plots from a single strain on a given substrate, across all three of the temperature studies to show a comparison of the data as the temperature increases. PCA plots showing different separation abilities will be shown first, followed by the corresponding DFA that show an improvement in the clustering and separation ability. All other PCA and DFA plots from the remaining strains and environments are shown in Appendix C.

Figure 9 shows a comparison of PCA plots produced from spectra acquired across the temperatures for BTK deposited on boot. Figure 10 shows a comparison of DFA plots produced from spectra acquired from BTK deposited on boot across the temperatures. There is a significant improvement in the separation of the data from the PCA to DFA plots, with consistent clustering for the four time-points across the three environmental conditions, with particularly distinct clustering for the 37°C dataset shown in Figure 10.

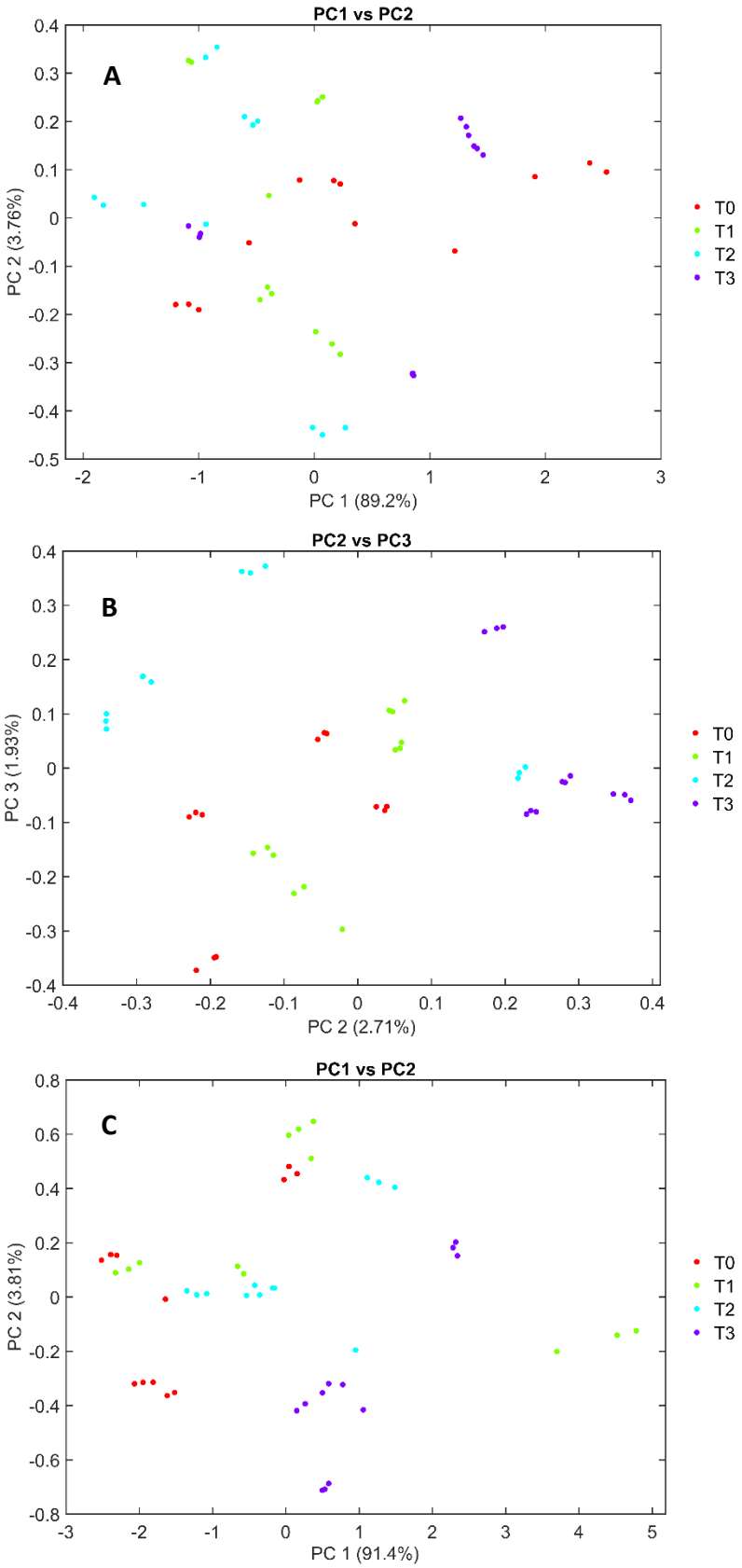

Figure 9: PCA plots showing separation of time point for spectra collected from BTK deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

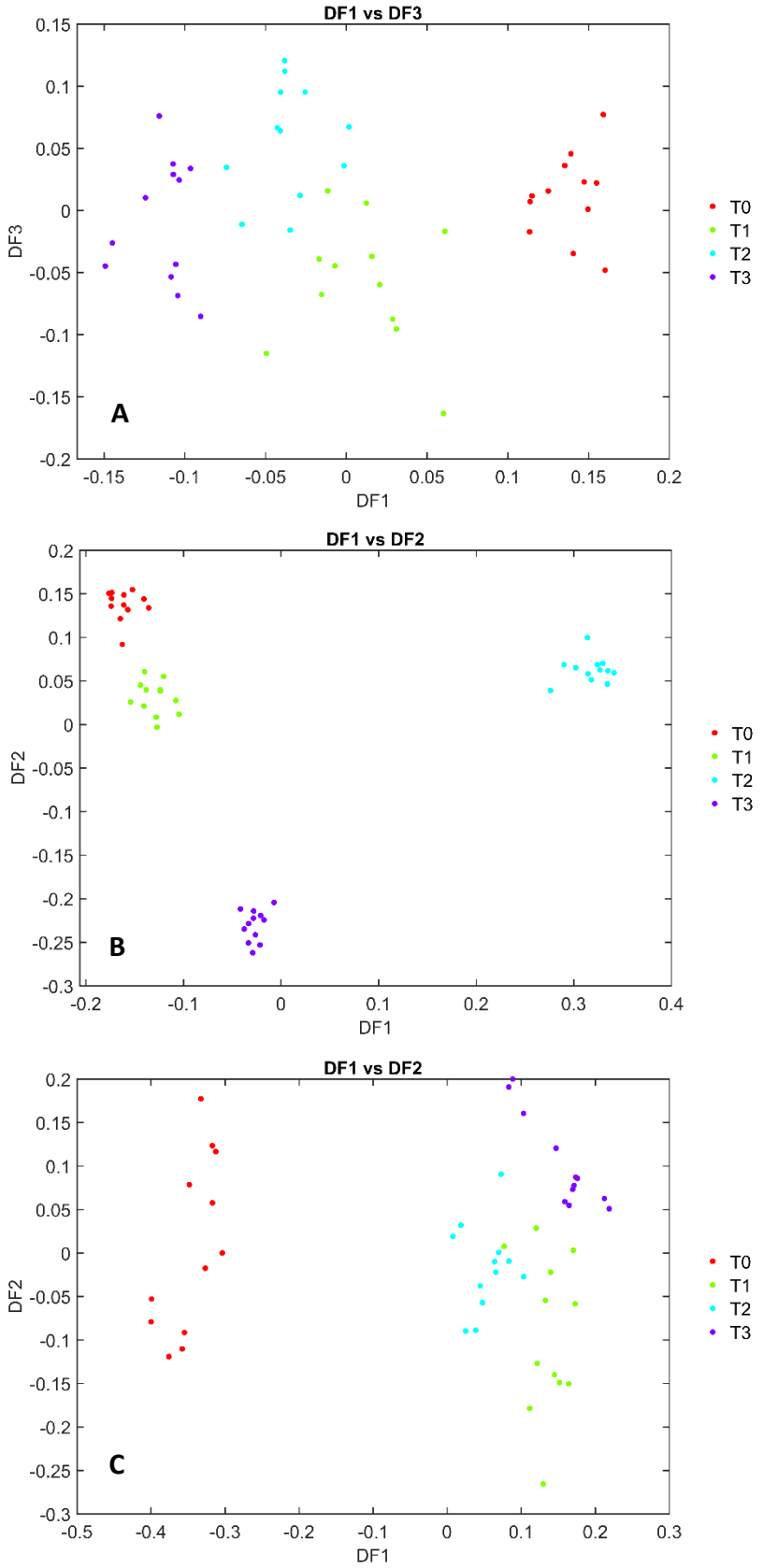

Figure 10: DFA plots showing separation of time point for spectra collected from BTK deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

A general trend seen across all of environments in the PCA plots in Figure 9 is that there is no separation or clustering of the four time-points seen in any of the BTK datasets. However, when examining the DFA plots for the same datasets, shown in Figure 10, there is significantly enhanced separation of the time-points. For all three of the DFA plots shown in Figure 10 there is both clear separation and clear clustering of the time-points, highlighting the improved quality of separation achieved when performing DFA after PCA, ultimately highlighting the need for an extra stage of MVA to provide a classification result.

DF1 is responsible for some of the separation seen in the DFA plots across all three temperatures. DF1 is responsible for the separation of T0 from the other time-points in Figure 10C, while in Figure 10B it is T2 that is separated out by DF1. Figure 10A shows that T0 and T3 are separated by DF1, with T1 and T2 falling in the centre of the plot with respect to DF1. DF2 separates T3 from the remaining time-points in both Figure 10B and C. Along with this, T0 and T1 are also separated with respect to DF2 and the general separation of the groups is much clearer in Figure 10B, in comparison to the other plots in Figure 10. When analysing Figure 10A, DF3 is mainly responsible for separating T1 and T2 from each other, with the remaining two time-points spread across the plot with respect to this DF.

When interrogating the loadings plots that relate to the DFA plots of Figure 10, it can be seen that the spectra are heavily dominated by noise, as with the plots shown in Figure 6, Figure 7 and Figure 8. Tentative assignments have been given to the most significant peaks that are standing out from the noise.

Due to the degree of noise present in the loadings, a threshold was applied to limit the number of peaks identified for assignment. Only absorbances greater than 0.05 were considered to be genuine biological variations, as opposed to noise contributions. Biological variations may exist in other regions of the spectrum, but are difficult to extract from the underlying noise. Performing noise reduction may help to enhance these underlying biological peaks, however, in this particular study it was key to keep the processing as simple as possible.

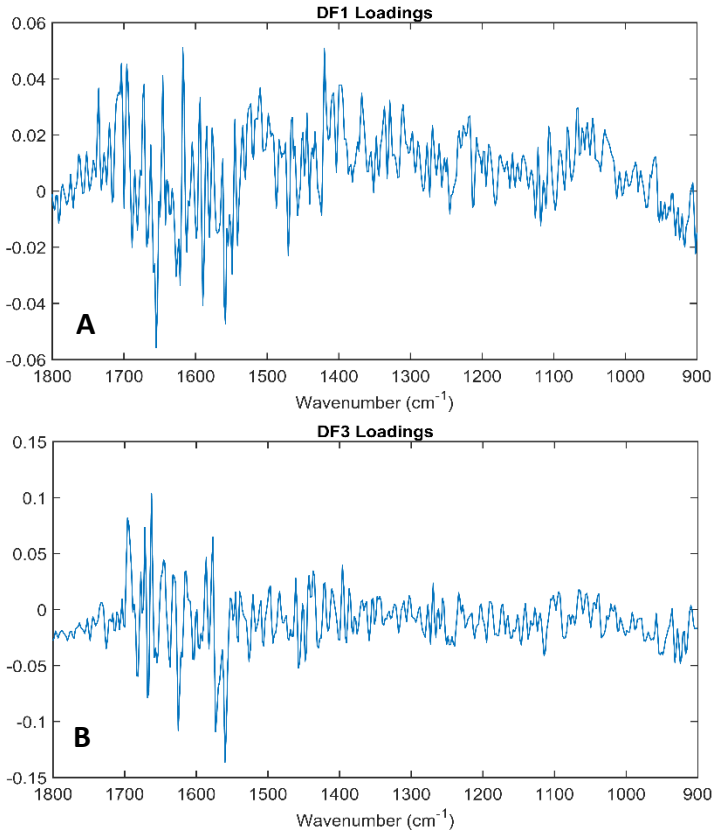

Figure 11: DFA loadings plots from the 30°C BTK study showing DF1 loadings (A) and DF3

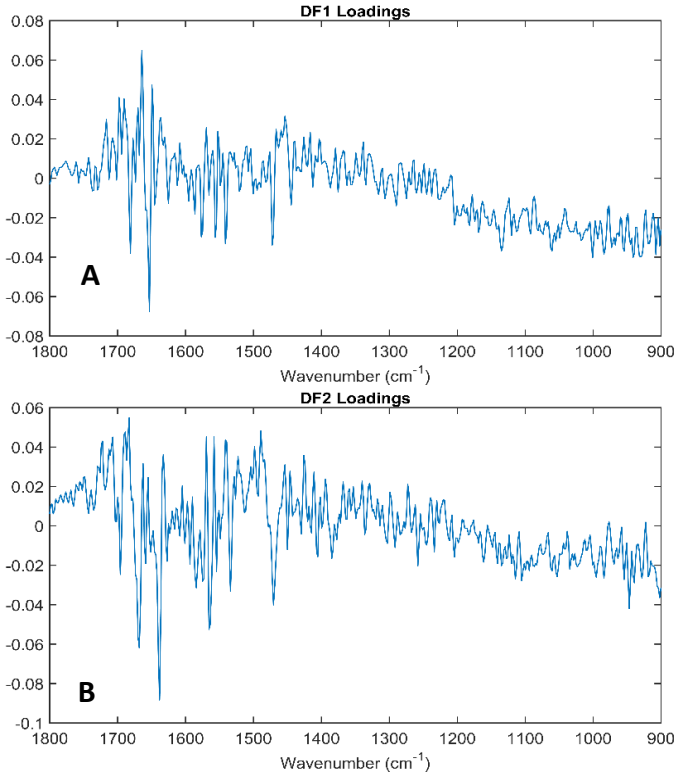

Figure 12: DFA loadings plots from the 37°C BTK study showing DF1 loadings (A) and DF2 loadings(B).

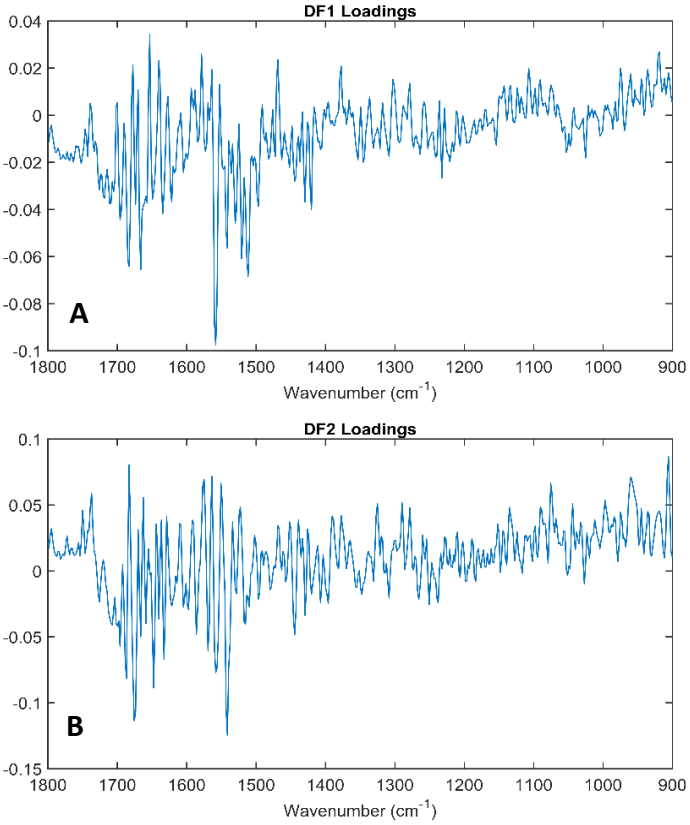

Figure 13: DFA loadings plots from the 49°C BTK study showing DF1 loadings (A) and DF3 loadings(B).

In the 30°C study, DF1 (Figure 11A) separates out T0 to the right of the plot with T3 separating out to the left of the plot, and the remaining two time-points (T1 and T2) are left in the centre of the plot with respect to DF1. DF3 (Figure 11B) splits T1 and T2 from each other, with T0 and T3 in the centre of the plot with respect to DF3. The only vibration that both of the DF loadings have in common is the Amide I vibration. DF3 has a number of peaks within the Amide II and a singular peak at 1457 cm^{-1} representing an asymmetric CH_3 vibration. The only unique peak found within the DF1 plot is found at 1420 cm^{-1} and represents C-H vibrations and ring stretching. Combined, DF1 and 2 (Figure 12) separate and cluster the data into the four individual time points. DF1 causes T2 to separate clearly from the other time points while DF2 separates T3 out from the other time points. When considering the peaks represented in the 37°C DF loadings, only peaks within the Amide I region are shown in the DF1 loadings, with Amide I and II vibrations represented in DF2. For the 49°C study, DF1 (Figure 13A) separates out T0 from the other time points while DF2 (Figure 13B) separates out T3. Both DF1 and DF2 have peaks within the Amide I and II region - these are the only peaks present in the DF1 loadings. Within the DF2 loadings there are a number of other peaks represented. These include peaks in the Amide III and Phosphodiester regions as well as C=O and symmetric PO_2^- vibrations.

Table III: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots generated from the data collected from BTK deposited on boot during the 30°C, 37°C and 49°C studies [1,2,3]

Wavenumber (cm^{-1})	Tentative peak assignments
905.7, 959.8, 997.1, 1044	Phosphodiester region, vCO, vCC, vCOH and vCOC
1075	vPO_2^- symmetric
1290, 1325	Amide III
1420	Ring stretch vibrations mixed with CH in-plane bending
1457	δCH_3 asymmetric
1511, 1521, 1541, 1551, 1553, 1558, 1560, 1564, 1565, 1569, 1573, 1575, 1577	Amide II (NH bend coupled to CN stretch)
1618, 1625, 1633, 1638, 1647, 1655, 1662, 1666, 1668, 1672, 1675, 1681, 1683, 1687, 1696	Amide I
1737	vC=O

Figure 14 shows a comparison of PCA plots produced from spectra acquired from *B. atrop* deposited onto boot in the temperature studies, showing some good separation of the time-points apart from the 30°C data (A). The ideal growth temperature for *B. atrop* is 37°C which may explain why there is poorer separation of the time points at 30°C and better at 37 and 49°C. This would be due to bacteria not growing or growing very slowly which will result in smaller, less noticeable changes over time at 30°C. At 37 and 49°C the bacteria are likely to be growing therefore, a biological response is expected and would be more noticeable within the spectral profile. Figure 15 shows the accompanying DFA plots produced from spectra acquired from *B. atrop* deposited onto boot in the temperature studies. As with other studies, with DFA analysis there is an improvement in the separation of the data with tighter clustering of data points for each time-points compared with the PCA plots produced for the same dataset.

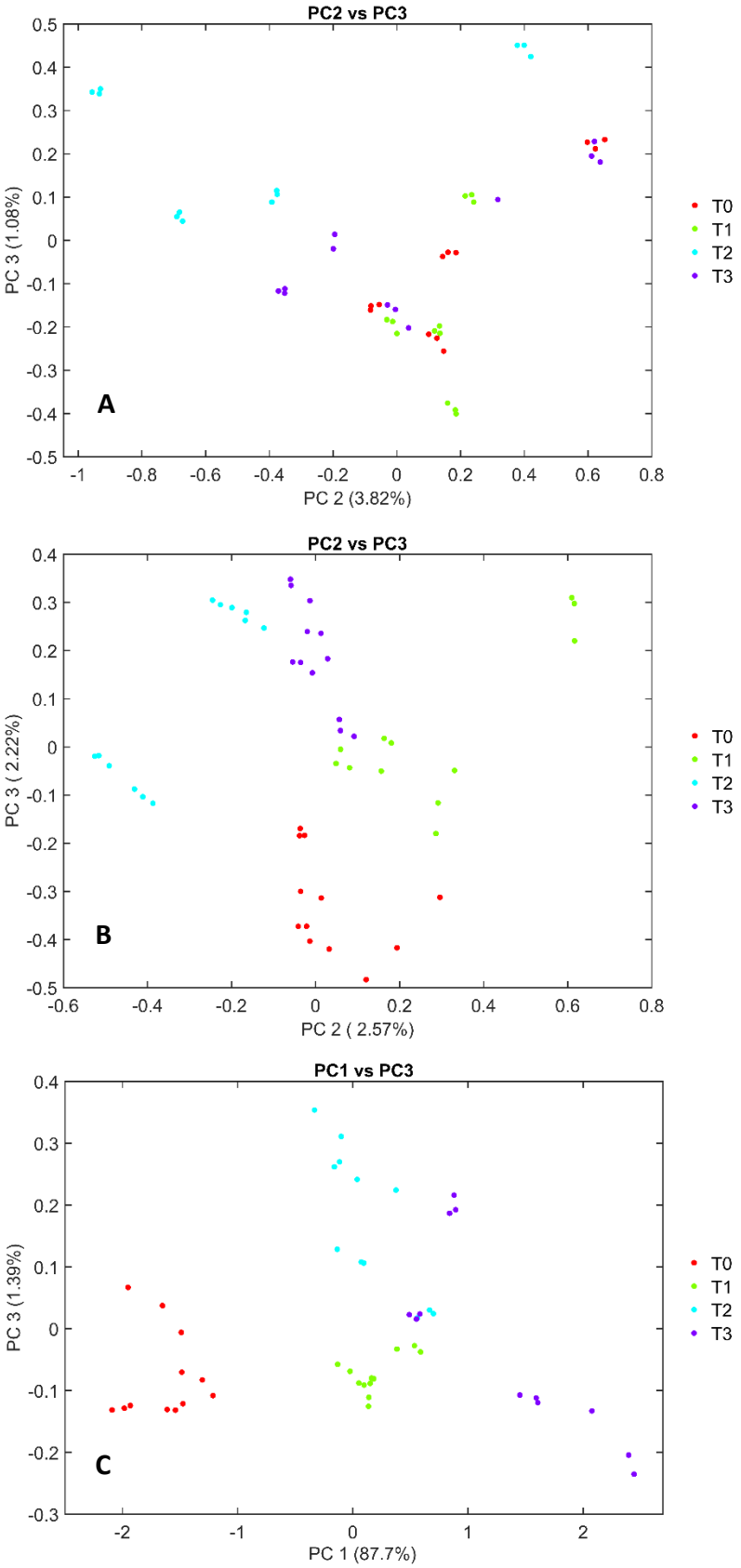

Figure 14: PCA plots showing separation of time point for spectra collected from B. atrop deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

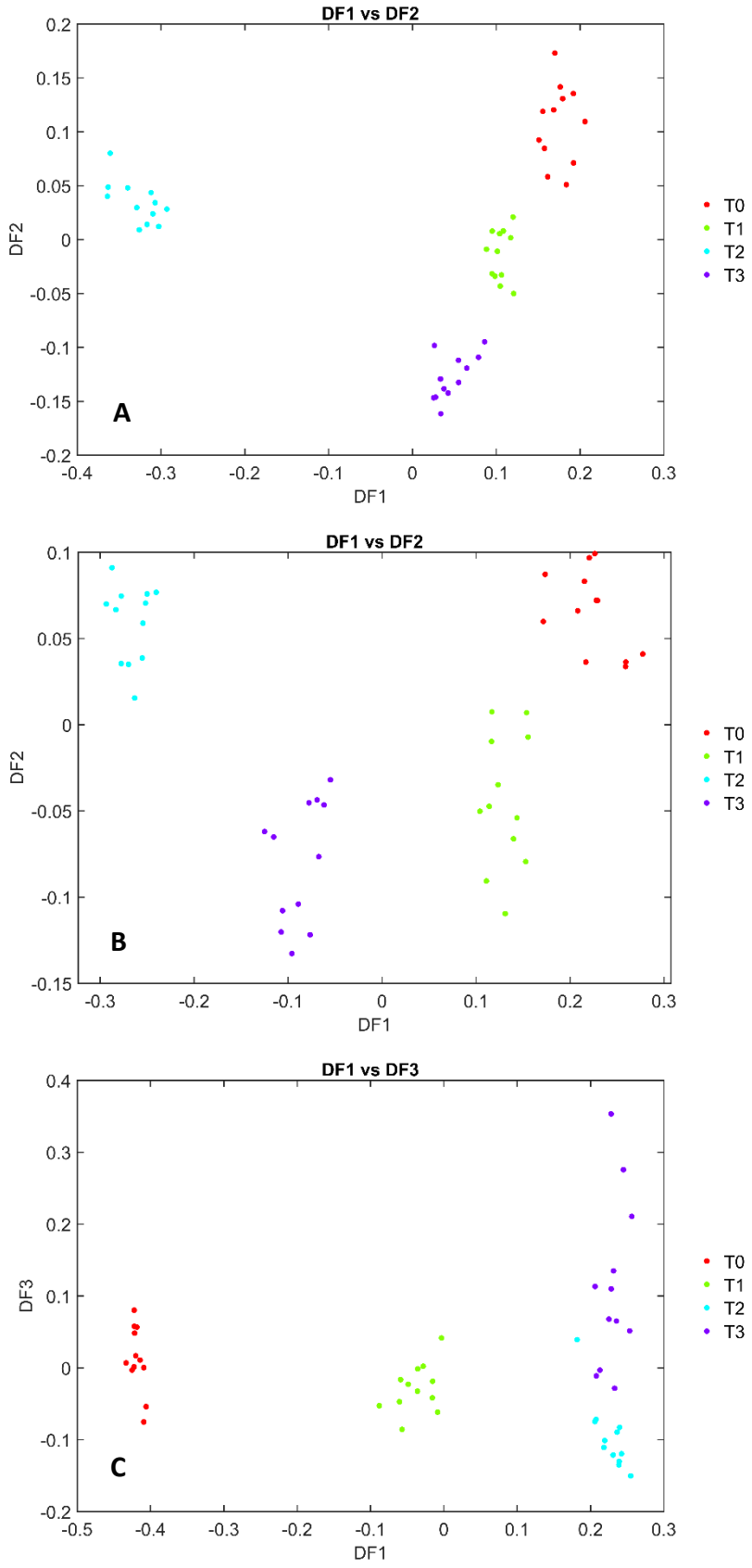

Figure 15: DFA plots showing separation of time point for spectra collected from B. atrop deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

Unlike with the PCA plots produced from the BTK datasets, there is a general trend seen across all of environments for the *B. atrop* data shown in the PCA in plots in Figure 14 that there is a good level of separation with some clustering of the four time-points. In Figure 14 all of the time-points separate out, rather than a single time splitting to leave the other three clustered together. The exception to this is in Figure 14A, where T2 clearly separates out from the other three time-points, while the other three time-points do not clearly separate. Figure 14B and C split all four of the time-points into individual clusters. In Figure 14B, PC2 causes T1 and T2 to separate from the other two time-points which are located in the centre of the plot with respect to that PC. PC3 separates out T0 and T3, as well as splitting the T2 cluster in half. For Figure 14C, PC1 separates out T0 and T3 with the other two time-points in the centre of the plot with respect to PC1. PC3 is mainly responsible for separating T2 from the other time-points. The DFA plots shown in Figure 15 show enhanced separation of the time-points across the three different temperatures, with tighter clustering. Generally, DF1 is responsible for the separation of the early time-points (T0 and T1) from the later time-points (T2 and T3), except for Figure 15A where DF1 is responsible for splitting T2 from the rest of the time-points. DF2 appears to separate T0 and T2 from T1 and T3 in Figure 15A and B. In Figure 15C, DF3 separates T2 and T3 from each other, with T0 and T1 generally sitting in the centre of the plot with respect to DF3. Again, the spectra are again heavily dominated by noise with some biochemical features protruding through the noise to illustrate what physical features are causing the separation. These have been tentatively assigned peaks and collated into Table IV.

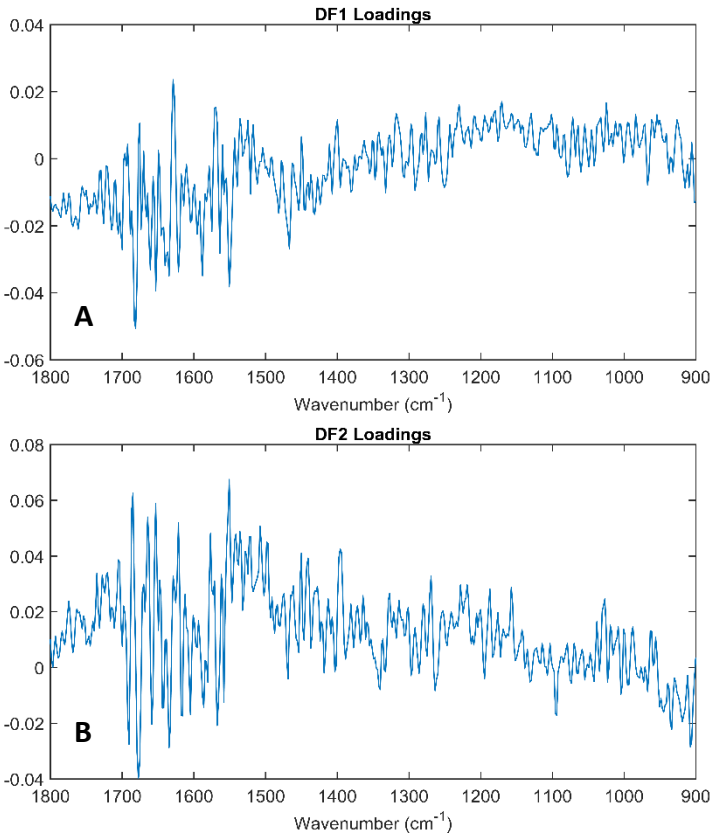

Figure 16: DFA loadings plots from the 30°C B. atrop study showing DF1 loadings (A) and DF2 loadings(B).

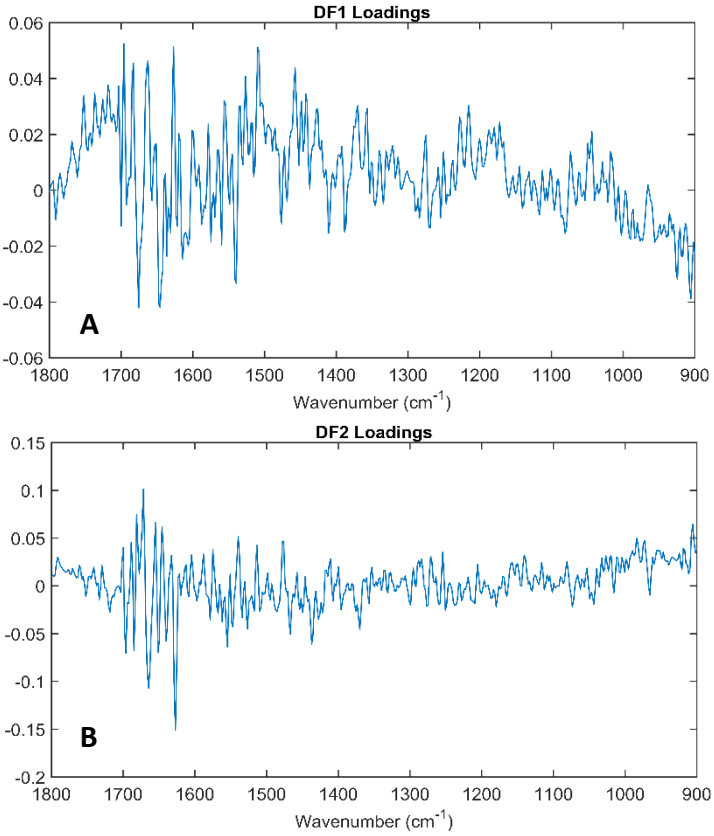

Figure 17: DFA loadings plots from the 37°C B. atrop study showing DF1 loadings (A) and DF2 loadings(B).

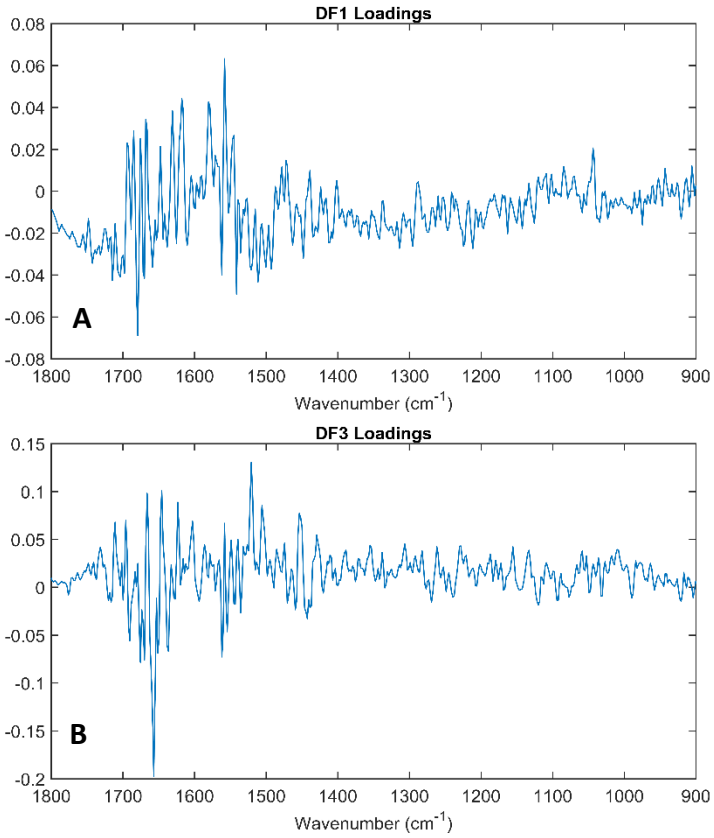

Figure 18: DFA loadings plots from the 49°C B. atrop study showing DF1 loadings (A) and DF3 loadings (B).

Table IV: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots generated from the data collected from B. atrop deposited on boot during the 30°C, 37°C and 49°C studies [1,2,3]

Wavenumber (cm^{-1})	Tentative peak assignments
905.7, 984	Phosphodiester region, vCO, vCC, vCOH and vCOC
1429, 1437, 1454, 1467,	Ring stretch vibrations mixed with CH in-plane bending
1506, 1508, 1510, 1521, 1539, 1551, 1554, 1558, 1562	Amide II (NH bend coupled to CN stretch)
1621, 1623, 1627, 1636, 1640, 1646, 1651, 1653, 1655, 1657, 1664, 1666, 1670, 1672, 1675, 1679, 1681, 1685, 1690, 1696	Amide I
1711	vC=O

For the 30°C study, DF1 (Figure 16A) separates T2 from the other time points while DF2 (Figure 16B) separates T0, T1 and T3 from each other. In the DF1 loading the only significant peaks are within the Amide I region, while in DF2 peaks are also

found within the Amide II region with peaks at 1551 and 1508 cm^{-1}. DF1 (Figure 17A) separates the early time points (T0 and T1) from the later time points (T2 and T3). DF2 (Figure 17B) separates T0 and T2 from T1 and T3 for the data collected during the 37°C study. DF1 houses only peaks within the Amide I and II regions, however DF2 also has peaks within the phosphodiester region and ring stretch vibrations. As with DF1 in the 37°C study, DF1 (Figure 18A) separates the early time points (T0 and T1) from the later time points (T2 and T3). DF3 (Figure 18B) separates T2 from T3, with T0 and T1 sitting in the centre of the plot with respect to this DF. As with the DF loadings from the 37°C, DF1 has only peaks within the Amide I and II regions. This suggests that as the bacteria ages it is the Amide region that is most affected at 49°C as proteins are known to denature at higher temperatures. The DF3 loading shows that peaks are also present within the Amide region, as well as at 1711 to represent the C=O vibration and peaks at 1454 and 1429 cm^{-1} representing C-H vibrations and ring stretches. This suggests that at the final time point (T3) there is more C=O and C-H present in the spectral profile of *B. atrop* than at the remaining three time points.

4.4 Conclusions

During the DRIFTS environmental conditioning study, it was shown that absorbance spectra from bacteria can be generated using the Beer-Lambert equation, however it can be seen that there is some overcompensation when removing the substrate contribution. Despite the overcompensation, key biological features are present and are detectable. The performance of this method is variable depending on the type of substrate used, with non-porous substrates generally performing better than porous ones. Due to the nature of the DFIRTS technique porous substrates will produce less intense spectra due to the loss of energy to the substrate. These results are also consistent with the conclusions made, with respect to the effect of the substrate, during the glucose and ovalbumin studies. The ability for the Gram-type to be determined varied depending on the environment the bacteria were left in. On the occasions where there was no clear separation based on Gram-type there was a good level of species-based clustering showing that the technique can identify differences between the species when collected from a complex substrate. Previous work shown in Chapter 3 (section 3), shows that there is no detectable difference between different batches of the same bacteria. While it is difficult to be certain where all of the biological information within the loadings plots lies, there are a number of peaks clearly protruding through the noise that have been tentatively assigned biological vibrations. As mentioned previously, performing some pre-processing on the data mainly in the form of noise reduction may enhance the biological feature within the spectrum. However, the key aspect to this study was to

develop a methodology that is rapid and user friendly, which required removing the pre-processing steps. The method then simply involved removing the substrate contribution from the bacterial spectrum before performing data analysis. With a model database set up containing multiple bacterial strains and substrates, this could lead to an identification of a sample within a number of minutes, perfectly suited to *in situ* analysis.

The overall aim of this study was to develop a methodology that can be used to identify samples *in situ* so that there was no sample preparation involved, as well as being able to provide a rapid identification of a sample. This is particularly important when handling suspected biological warfare agents (BWAs) as quicker identification of a sample would allow for the correct treatment of patients to be implemented, in turn potentially reducing the number of fatalities caused by such an event. DRIFTS has shown that separation and clustering can be seen in data that has not been pre-processed. For this reason, DRIFTS is a suitable technique that can be used for the identification of surface-deposited bacteria.

4.5 References

1 Hayashi, H. & Mukamel, S. Vibrational-Exciton Couplings for the Amide I, II, III, and A Modes of Peptides. *Journal of Physical Chemistry* **111**, 11032-11046 (2007).

2 Naumann, D. *Infrared Spectroscopy in Microbiology*. 1-18 (2006).

3 Davis, R. & Mauer, L. J. in *Current research, technology and education topics in Applied Microbiology and Microbial Biotechnology Volume II.* Vol. 2 (ed A. Mendez-Vilas) Ch. Fourier transform infrared (FT-IR) spectroscopy: A rapid tool for detection and analysis of foodborne pathogenic bacteria, 1582-1594 (Formatex Research Center, 2010).

4 Lewis, P. D. & Menzies, G. E. Vibrational spectra, principal components analysis and the horseshoe effect. *Vibrational Spectroscopy* **81**, 62-67, doi:https://doi.org/10.1016/j.vibspec.2015.10.002 (2015).

Chapter 5

Assessment of Effects of Environmental Conditioning on FTIR Spectra from Surface Deposited Bacteria using High Throughput Infrared Spectroscopy

5.1 Introduction and Aims

This chapter discusses research that has focused on assessing the effect of different environments on surface deposited bacteria. The temperatures used in the study were in the range 30°C to 49°C which are used to simulate an extremely hot dry climate as classified by the Ministry of Defence (MoD). [1] For this study the lower limit of 30°C was used for the environment as opposed to 32°C to ensure that the temperature range encompassed the optimal growth temperatures for all of the bacterial strains used within the study. An extreme hot dry climate is typical of regions such as the Middle East and Southwestern United States of America (USA), areas that have had cases involving the use of Biological Warfare Agents (BWAs) during the late 20th and early 21st Century and therefore are of particular concern and interest.

The chapter begins by outlining the setup for the experiment before moving on to discuss the results obtained and assess the effect of exposure to different

environments and how these effects can be used to aid identification of samples. This study analysed bacteria prepared on a 96-well Silicon (Si) plate with data collected using a High Throughput - Fourier Transform Infrared Spectrometer (HT-FTIR). As part of the environmental study Multiblock PCA was performed on the dataset to analyse which of the different factors has the biggest impact on the separation of the data.

After discussing the results of the HT-FTIR study a final section discusses the general trends that are common between both of the environmental studies considering the data discussed in Chapter 4 (section 3.1.2) and Chapter 5 (section 5).

5.2 Sample Preparation

For HT-FTIR analysis, carried out at the University of Manchester (UoM) using the Bruker Equinox 55 with a HTS XT microplate reader attachment, bacterial samples were prepared to an OD 20. From these samples, 12 per bacterial strain, were spotted in duplicate across two 96-well Si plates, with each well containing 20 µl of sample. As with the study carried out with DRIFTS, there was a control and non-control set of samples. Due to the nature of the instrument, this consisted of one 96-well Si plate of samples being a control plate, with a second used as a non-control plate. Bacteria were spotted onto the plate in a random order, with the order then remaining consistent across both plates and across all temperatures as well as in the solar study.

The same protocol was used across all of the temperature and solar studies. Before the analysis of any samples, both plates were dried in an oven set to 50°C for

~20 mins, until all spots were dry. It is important for HT-FTIR that the samples are fully dry before analysis takes place as the water within the samples will completely dominate the spectrum acquired from a wet spot.

5.3 Data Collection

Once the samples were dried one plate was transferred to a desiccator for storage while the other plate was analysed for the T0 time point. Once the analysis of the first plate was completed the plates were swapped to allow for analysis of the second plate, again storing the previously analysed plate in the desiccator. Once both plates had been analysed they were stored in clear petri dishes before then being placed into the environment.

For the study assessing the effect of temperature using HT-FTIR, the environment was an incubator set to the chosen temperature for the 30°C and 37°C studies, with an oven used for the 49°C study. As a consequence of using an incubator or an oven for the storage of samples, the plates were not exposed to any solar radiation other than when being transferred for analysis.

For the solar study a slightly different set up was used as the temperature was not controlled. A steel, high sided rectangular box was used to store the plates before the LED lamp was attached to the side of the box which was then covered with foil, see Figure 1. The steel box was stored in a fume hood with the front closed, only being opened on the days where a plate was removed for analysis. This set up allowed

for the temperature around the samples to be kept as consistent as possible. The setup also ensured that the exposure to other solar radiation was very minimal while ensuring any stray radiation from the lamp was directed towards the samples as the steel sides and foil reflect the light onto the samples. A temperature and humidity monitor was stored alongside the samples within the steel box to track the conditions that the samples were exposed to alongside the solar radiation. Prior to a study being carried out with bacterial samples, the box was set up with the lamp and humidity monitor to assess any temperature change caused by the lamp. The study found that there was no significant temperature change caused by the use of the lamp, therefore no compensation or extra considerations were required during this study.

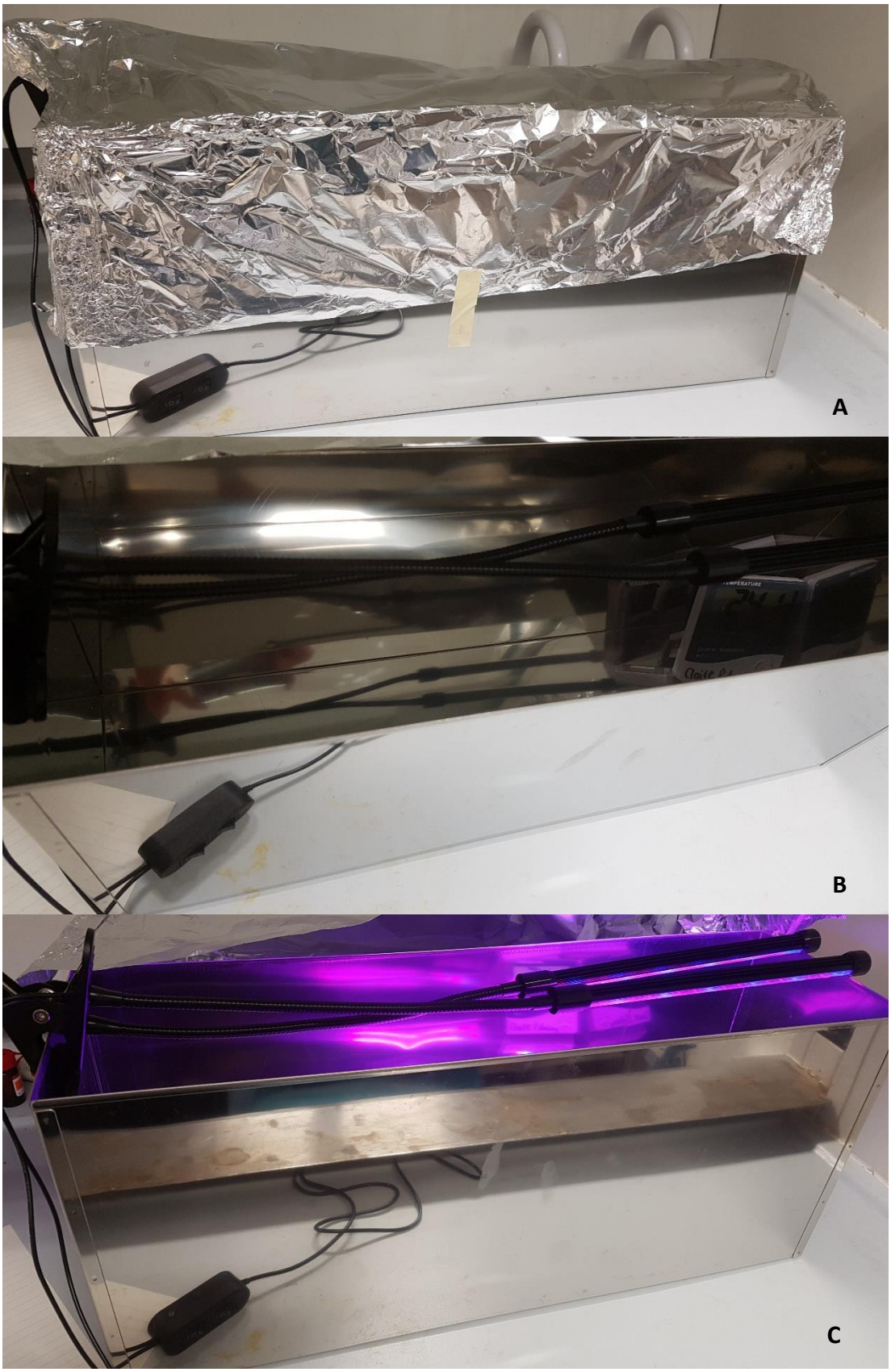

Figure 1: Figure showing the set up used for the Solar study where the box is covered with foil (A), the box is uncovered with lamp attached to the side of it and a temperature and humidity monitor (B) and the box is uncovered with lamp attached to the side of it and switched on (C).

5.4 Data Analysis

During the processing of the data, after all of the spectral acquisition had been completed, it was noted that during spotting of the bacterial samples, an extra sample of BTK had been spotted (13), and one less sample of BT (11). This was consistent across both Si plates, across all three temperatures and for the solar study. This did not affect the PCA and DFA analysis. However, for multiblock PCA to be performed, each of the bacterial strains require the same number of samples per group. To balance the groups, a 12^{th} spot for BT was generated by averaging the other 11 BT samples. A single spot was removed from BTK to leave 12 spots. As discussed in Chapter 4 (section 3.1.2) PCA/DFA plots shown within this chapter compare different PCs/DFs due to the fact that different datasets are being compared and the most discriminatory component/function may not be the same but will be found within the first three.

5.4.1 Data Pre-Processing and Multivariate Analysis

The processing method developed within Chapter 3 (section 2.6) for the pre-processing and multivariate analysis (MVA) of FTIR data was used within this chapter for the processing of the HT-FTIR dataset. The pre-processing of the data involves vector normalising (VN) the data, performing PCA-based noise reduction (NR) before cutting the data to the fingerprint region (1800-900 cm^{-1}). Once the pre-processing had been performed on this dataset MVA could be performed. This was

in the form of both PCA and DFA analysis as well as Multiblock PCA. Multiblock PCA was performed on the HT-FTIR dataset to try and explore further the effect that aging has on a bacteria sample when exposed to differing environmental conditions. As with other datasets, NR and VN and a cut to the fingerprint region were performed on the dataset prior to multiblock PCA. Additional to this, the data was also processed using extended multiplicative scatter correction (EMSC) and a fingerprint cut rather than NR and VN as a comparison of pre-processing techniques. The additional study where the data was pre-processed using EMSC was included as HT-FTIR data is more prone to suffer from scattering.

For the DRIFTS dataset the data were handled in a similar manner to that of the glucose and ovalbumin dataset discussed within Chapter 2 (section 2.4). This involved manually removing the spectral contribution of the substrate from the bacterial spectrum to produce an absorbance spectrum of the bacterial data. Once this had been obtained it was cut to the fingerprint region and used for MVA in the form of PCA and DFA. Spectra were not pre-processed prior to MVA in the interest of keeping the methodology for identification as simple as possible. If the methodology were to be used in a 'real world' scenario the data would not need to be cut to the fingerprint region prior to MVA as the region of interest can be selected prior to spectral collection. The whole spectral range (4000-600 cm^{-1}) was collected during this study for the purposes of having a complete spectral database. It should be noted that no outliers were removed from any of the models to test the robustness of the model.

5.5 Environmental Conditioning Study

For the HT-FTIR study there are four different sections to consider and analyse the interactions and relationships both individually and between them. The four sections are made up of three different temperatures and a single solar radiation study. Data from each of the temperature studies and the solar study will be shown in individual sections where they will also be discussed. Following this, there will be a discussion of each of the temperatures individually before comparing and contrasting the results seen in the three environments. The HT-FTIR section will finish with an overall comparison of results collected from the four independent studies, including the solar study.

HT-FTIR data was processed using the method discussed in Chapter 3 (section 2.6) using vector normalisation and noise reduction, before cutting the data to the fingerprint region (1800-900 cm^{-1}). Figure 2 shows a multigroup PCA, with all seven bacteria, for each of the three temperatures. The multigroup PCA plots show both the bacterial strain and time point which make it clear which is the overriding factor causing the separation seen in the plot, without having to plot them separately. The plots shown were chosen based on the PCs that show the clearest separation and clustering of the data. Following this, DFA was performed on the same datasets with plots that show the best separation shown in Figure 3 for each of the temperatures. Figure 5 to Figure 7 goes on to show the loadings plots for the most discriminatory DFs shown in Figure 3.

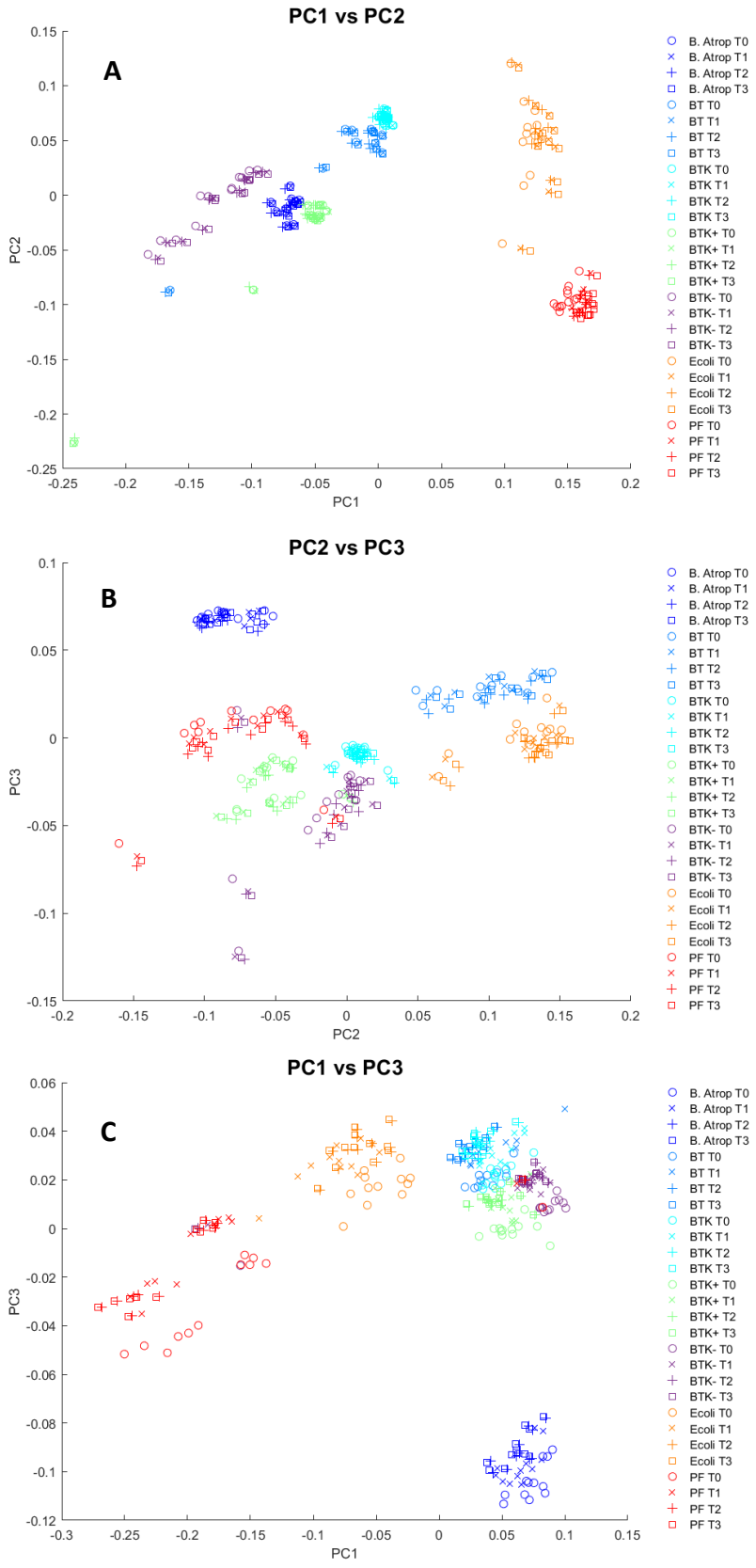

Figure 2: Multigroup PCA plots from HT-FTIR data showing separation based on both the species and the ageing of bacteria, across all three temperature studies. (A) shows PC1 vs PC2 from the 30°C study, (B) shows the PC2 vs PC3 plot from the 37°C study and (C) shows PC1 vs PC3 from the 49°C

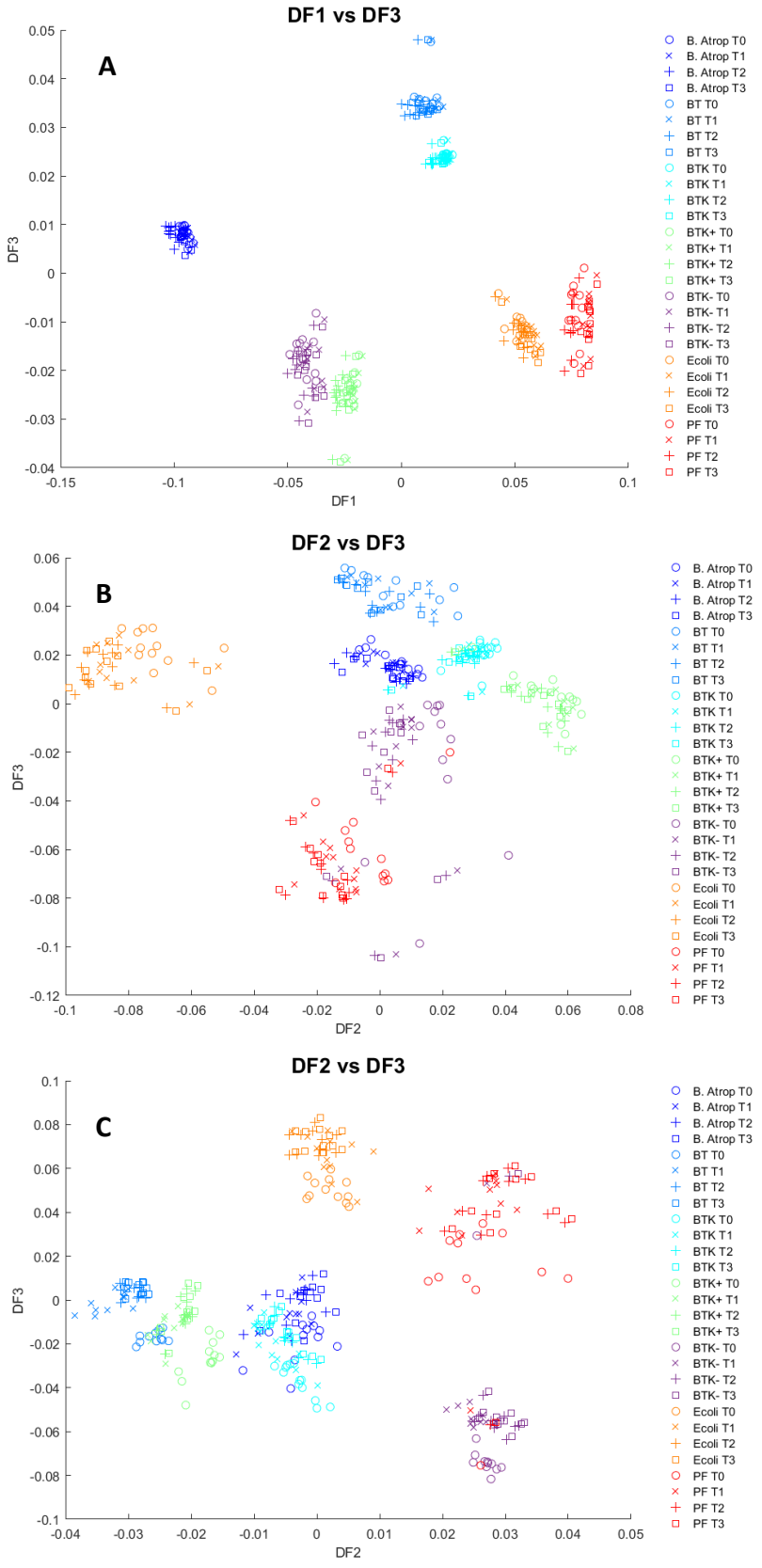

Figure 3: Multigroup DFA plots from HT-FTIR data showing separation based on both the species and the ageing of bacteria, across all three temperature studies. (A) shows PC1 vs PC3 from the 30°C study, (B) shows the PC2 vs PC3 plot from the 37°C study and (C) shows PC2 vs PC3 from the 49°C

In Figure 2A and C, PC1 is responsible for the split in the data caused by Gram-type. PC2 is responsible for the separation of *E. coli* from PF as well as BT and BTK from the other *Bacillus* species shown in Figure 2B. As PC1 is responsible for separation of Gram type, the loadings for the 30 and 49°C studies are included in this chapter (Figure 4). Figure 5 shows that the majority of the separation is caused by vibrations within the Amide I and II regions (1659, 1657, 1612, 1609 1584 and 1582 cm^{-1}) that are responsible for the separation of Gram type in both the 30 and 49°C studies. In Figure 2A, PC2 mainly separates BT and *E. coli* from the other bacterial strains. PC3 clearly separates *B. atrop* from all other bacterial strains including the other *Bacillus* strains which can be seen in Figure 2B and C. Also shown in Figure 2C is the separation between BT, BTK and its sub-species with respect to PC3. The combination of PC1 and PC2 in Figure 2A and PC2 and PC3 in Figure 2B shows that there are seven individual clusters, one for each bacterial species, however in Figure 2C there is some overlap of the *B. thuringiensis* strains. The *B. thuringiensis* strains overlap as they are phylogenetically closer to each other than they do any of the other bacterial strains. When moving on to consider the separation observed in the DFA plots, shown in Figure 3, all three temperature studies show separation and clustering of the bacterial strains. The DFs responsible for the separation of bacteria based on Gram-type vary across the three temperature ranges. DF1 clearly separates Gram-negative from Gram-positive bacteria in Figure 3A, while in Figure 3C it is DF3 that is used for the separation of Gram-type. DF2 shows this separation in Figure 3B, however the separation is less definitive than the split shown by the data collected from the other temperatures.

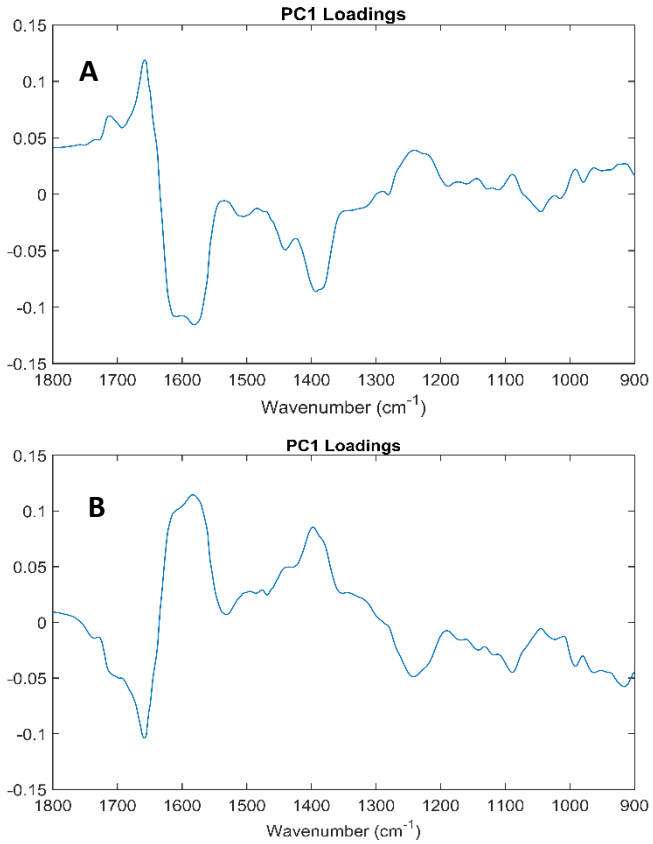

Figure 4: PC1 loadings for the 30°C study (A) and the 49°C study (B).

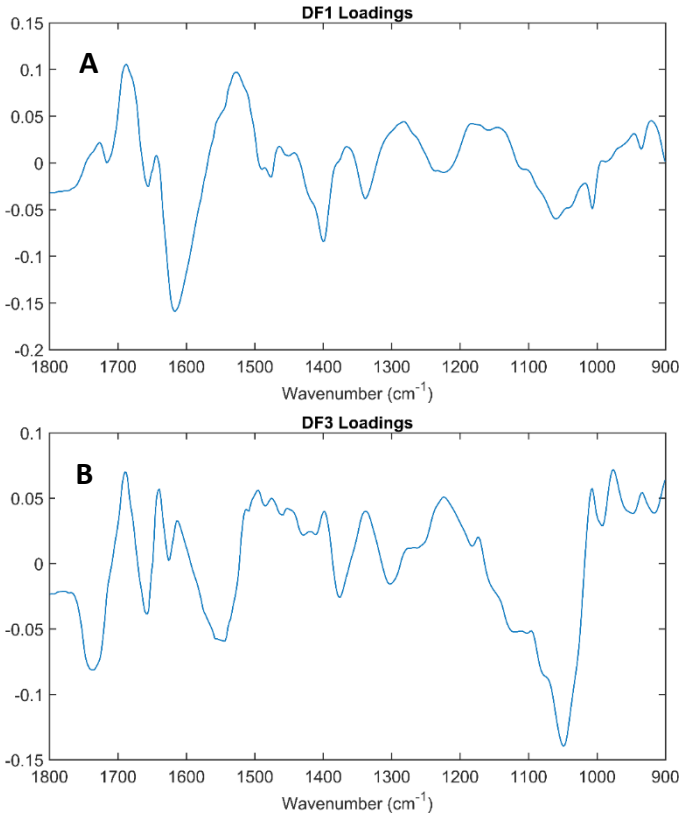

Figure 5: DFA loadings for the 30°C study showing DF1 loadings (A) and DF3 loadings (B).

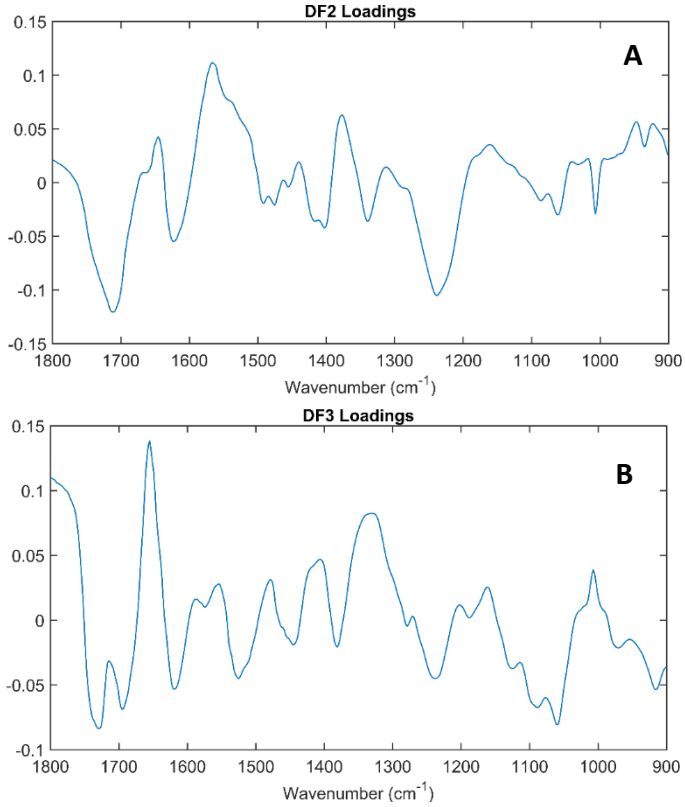

Figure 6: DFA loadings for the 37°C study showing DF2 loadings (A) and DF3 loadings (B).

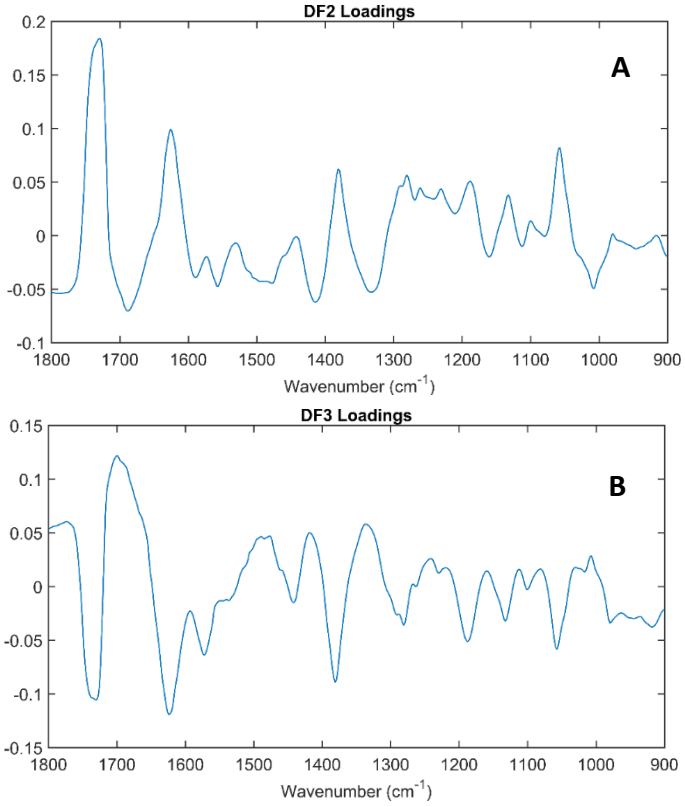

Figure 7: DFA loadings for the 49°C study showing DF2 loadings (A) and DF3 loadings (B).

In Figure 5A, showing DF1, the key peaks that cause the separation are found at the high wavenumbers relating to Amide I and II vibrations (1650-1545 cm^{-1}). These peaks are the result of vibrations from the C=O for Amide I and a combination of C-N-H bending and C-N stretching for Amide II. [2] Other vibrations that are key for the DF separation are the peaks at 1398 and 1061 cm^{-1} correlating to CH_3 vibrations and the vibrations of the phosphodiester region. Figure 5B shows DF3 which represents more biological features. Amide I and II and the phosphodiester region are still represented in DF3, however, the distribution of peaks is much more dispersed across the fingerprint region compared to DF1. Other key features used in the separation are C=O, Amide III and C=C and C-H vibrations. A key structural difference seen between Gram-positive and Gram-negative bacteria is the phosphate region of the spectrum encompassing the peaks at 1225, 1121 and 1103 cm^{-1} which causes the Gram-negative bacteria, as well as BTK- and BTK+, to separate out from the other *Bacillus* strains. The peak at 1225 cm^{-1} represents Phosphate II which covers asymmetric PO_2^- stretching vibrations, while peaks at 1121 and 1103 cm^{-1} cover the symmetric P-O-C and PO_2^- stretching.

When interrogating the loadings shown in Figure 6 there are similarities between DF2 and 3. Key vibrations highlighted in both loadings plots are Amide I (1713, 1692 and 1657 cm^{-1}) and the phosphodiester region (972, 947, 922 and 916 cm^{-1}) which can be linked to the separation of the bacterial strains. The CH_2 wagging is found exclusively with the bacteria found within the positive region of the DFA plot with respect to DF3. Different peaks within the Amide I and phosphodiester region point to possible shifts in peak position and structural changes. This can

explain why peaks relating to a particular vibration show up in both the positive and negative regions of a loadings spectrum. Amide I peaks are found in the negative region of the DF2 loading and in both positive and negative regions of DF3. The peaks found in the phosphodiester region can be found in both regions of the DF2 and the negative region of the DF3 loadings plot. The peaks seen within the Amide I and phosphodiester region are likely to be a result of structural differences found within the cell walls of the bacteria. DF2 highlights C=O, Amide II, Phosphate II and DNA/RNA bases as the key vibrations responsible for separation. Of particular interest are the peaks at 1711 and 1238 cm^{-1} which identify C=O and Phosphate II as the key vibrations for the split of the Gram-negative bacteria from the Gram-positive *Bacillus.* The phosphate bands are of particular interest when it comes to the separation shown as these relate to a fundamental difference within the cell wall. For this reason, they are often used to distinguish between Gram-positive and Gram-negative bacteria when using FTIR. [3]

For the 49°C study the DFA loadings shown in Figure 7 highlight that the key vibrations causing the separation are in the same regions for both DF2 and 3. These vibrations include C=O, Amide I and III, CH, CH_2 and CH_3 as well as vibrations in the phosphodiester region. The key vibrations that differ between the two discriminant functions is shown in Figure 7B with the peak at 1572 cm^{-1} correlating to an Amide II vibration. Changes in the Amide II region link to the overall level of protein content within the cell. For Gram-positive bacteria this is fairly low compared to that of Gram-negative bacteria which is high due to the presence of an outer membrane and a Lipopolysaccharide (LPS layer). [4] As with the data from the previous temperature

studies, the separation seen is likely to be due to the variation in the structure of the cell wall and the protein content of the different bacterial strains that are both Gram-positive and Gram-negative.

Along with a set of three temperature studies that span across 30-49°C, a solar study was carried out to evaluate the effect that another environmental variable has on a bacterial spectrum. Figure 8 and Figure 9 show the multigroup plots produced using PCA and DFA, along with the loadings plots that show the key peaks responsible for the separation seen in the DFA plot.

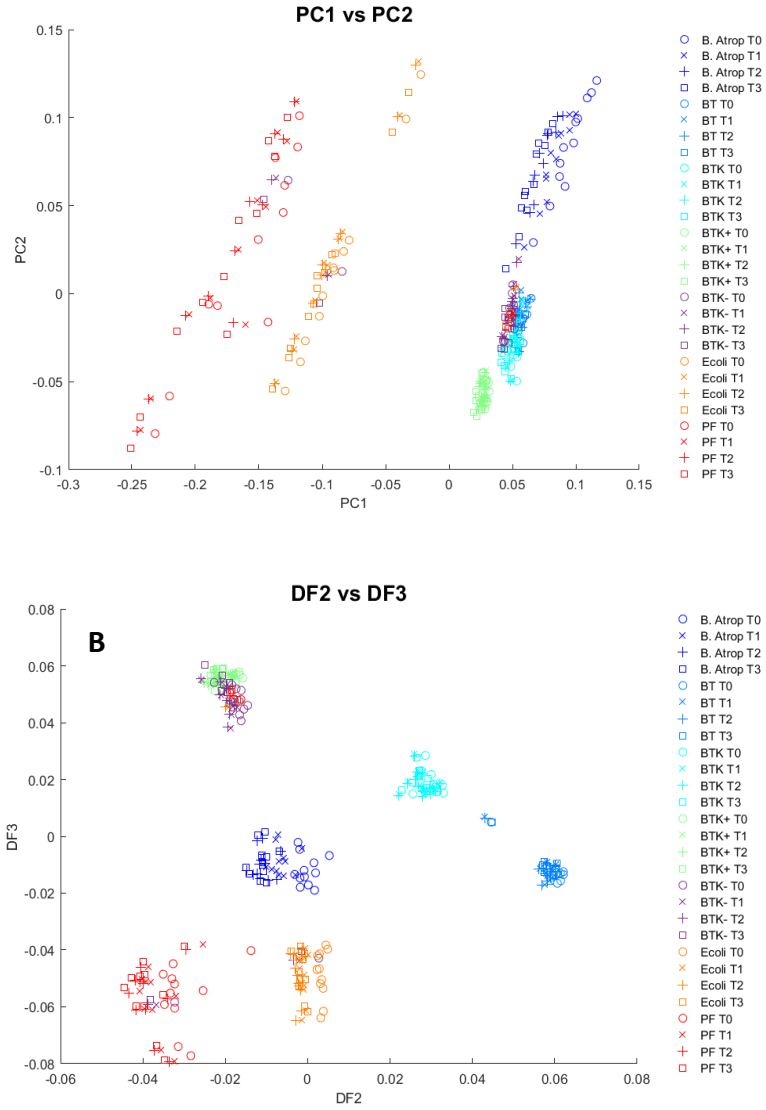

Figure 8: Multigroup MVA plots from HT-FTIR data acquired during the solar study showing PC1 vs PC2 (A) and DF2 vs DF3 (B).

From Figure 8 it can be seen that the separation is again caused by Gram-type rather than the effect of the solar radiation. However, there are signs within the bacterial clusters that the time points are beginning to separate out particularly with *B. atrop*, *E. coli* and PF. Both of the PCA and DFA plots show clear separation and clustering of the bacterial strains with PC1 in Figure 8 showing separation of the

Gram-negative strains (*E. coli* and PF) from the Gram-positive strains. Figure 8 shows that it cannot only separate the Gram-negative strains from the Gram-positive strains but also splits *E. coli* from PF. The PCA model then struggles with the separation within the Gram-positive strains. Figure 9A shows that there is a clearer separation of all strains regardless of Gram-type, however, there is still an overlap between BTKl and BTKx. A trend seen in the temperature study that is also observed in this study is that T0 can be seen to separate out the most out of all of the time-points. One possible cause for this separation could be differing levels of hydration when analysing samples at the T0 time point compared with T1, T2 and T3. For T0, the samples have only just been prepared, therefore have not had sufficient time to fully dry. This will change the biological peaks that can be seen within the loadings for the dataset, therefore causing the T0 data to separate out from the other three time-points. In Figure 8 this trend is most prominently seen in *B. atrop*, in Figure 9A this trend is also seen in *E. coli* and PF.

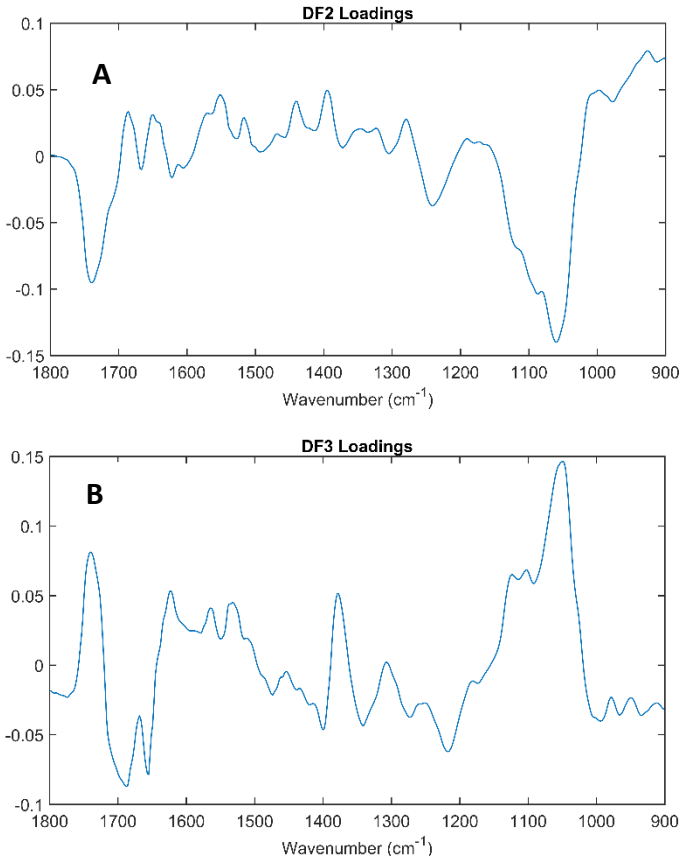

Figure 9: Loadings plots correlating to the DFA plot produced from the 49°C study showing DF2 loadings plot (A) and a DF3 loadings plot (B).

When it comes to interrogating the loadings plots for the Solar study, the peaks highlighted by DF2 and DF3, shown in Figure 9, have some similarities. The peaks highlighted include C=O, Phosphate I and peaks within the phosphodiester region. However, there are some key differences found within the DF3 plot where Amide I, the symmetric stretch of P-O-C and CH_3 are all identified. These phosphate vibrations may relate to the Gram-negative bacteria that is being separated out at the bottom of the plot with respect to DF3. The only unique vibration found in the DF2 loadings is the peak at 1121 cm^{-1} which represents symmetric stretching of a

phosphodiester band found in RNA. As discussed previously, the phosphodiester region is key in Gram-typing due to the differing levels of phosphate in Gram-positive and Gram-negative bacteria. The wavenumber and corresponding tentative peak assignments, from across the three temperature studies as well as the solar study, are shown collectively in Table I.

Table I: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots generated from all species in the 30°C, 37°C, 49°C and solar study datasets acquired using HT-FTIR [5,6,7]

Wavenumber (cm^{-1})	Tentative peak assignments
900.7, 916.2, 922, 925.8, 933.5, 947, 972, 976, 1007, 1049, 1057, 1059, 1060, 1061, 1124	Phosphodiester region, vCO, vCC, vCOH and vCOC
1076	vPO_2^- symmetric
1088, 1217	Phosphate I (stretching PO_2^- symmetric vibrations) in *B*-form DNA
1103	vP-O-C symmetric
1121	Symmetric phosphodiester stretching band RNA
1188, 1281	Amide III
1225,1238	Phosphate II (stretching PO_2^- asymmetric vibrations) in β-form DNA
1325, 1329, 1333, 1337	CH_2 wagging
1377	Adenine, Thymine, Guanine, Cytosine (DNA/RNA bases)
1379, 1381	δCH_3
1398	δCH_3 symmetric
1414, 1418	Ring stretch vibrations mixed with CH in-plane bending
1495	vC=C, CH
1528, 1537, 1545, 1558, 1566, 1572	Amide II (NH bend coupled to CN stretch)
1616, 1618, 1622, 1624, 1626, 1639, 1655, 1657, 1688, 1690, 1692, 1695, 1699	Amide I
1711, 1728, 1730, 1736, 1738, 1740	vC=O

As it is clear from the PCA and DFA plots shown in Figure 2, Figure 3, Figure 8 and Figure 9A, the overriding factor for the separation is the Gram-type of the bacteria and not the effect of the environmental conditioning. This is a trend that is not exclusive to the temperature studies as it also seen in the solar study. Despite this, there are signs that the environment does have an effect on the spectra collected. This effect is more prominent in certain bacterial strains and in certain environments. From the plots it can be seen that the tightest clustering and best separation is seen at the lowest of these temperatures (30°C). As the temperature increases there is a trend that the Gram-negative bacteria (*E. coli* and PF) appear to be more affected by the changing condition than the Gram-positive strains. This may relate to the fact that Gram-negative bacteria are unable to form endospores to help them handle a non-ideal environment. Endospores occur in some Gram-positive bacteria and are restructured, inactive forms that are caused as a response to being exposed to a non-ideal condition. [8] For spore-forming bacteria, as expected the spectral profile changes. The main differences present in a vegetative spectrum is the Amide I peak at ~1735cm^{-1}. [9] A spectrum collected from a spore lacks this shoulder but contains a unique and reproducible "quartet" of peaks between 600-800 cm^{-1} that correlate to $CaDP \cdot 3H_2O$. [10] When the endospore is reintroduced to a nutrient rich environment will revert back to its original structure, becoming active once again. It is for this reason identifying spore forming bacteria is important to prevent further infection occurring.

To observe the full effect of the environmental conditioning, MVA of the individual bacterial strains was required. As with the data presented from the dataset

as a whole, only a selection of figures will be presented within this chapter. All other relevant figures can be found in Appendix D. This will include plots from all bacterial strains as this section discusses only two of the original seven strains used in the study. Figure 10 shows PCA plots from BT across all three temperature studies, showing a mix in the separation ability. Figure 11 shows the DFA plots that correspond to the PCA plots shown in Figure 10. It can be seen that there is a slight improvement in the separation ability with some clustering of time-points in some, but not all, of the studies.

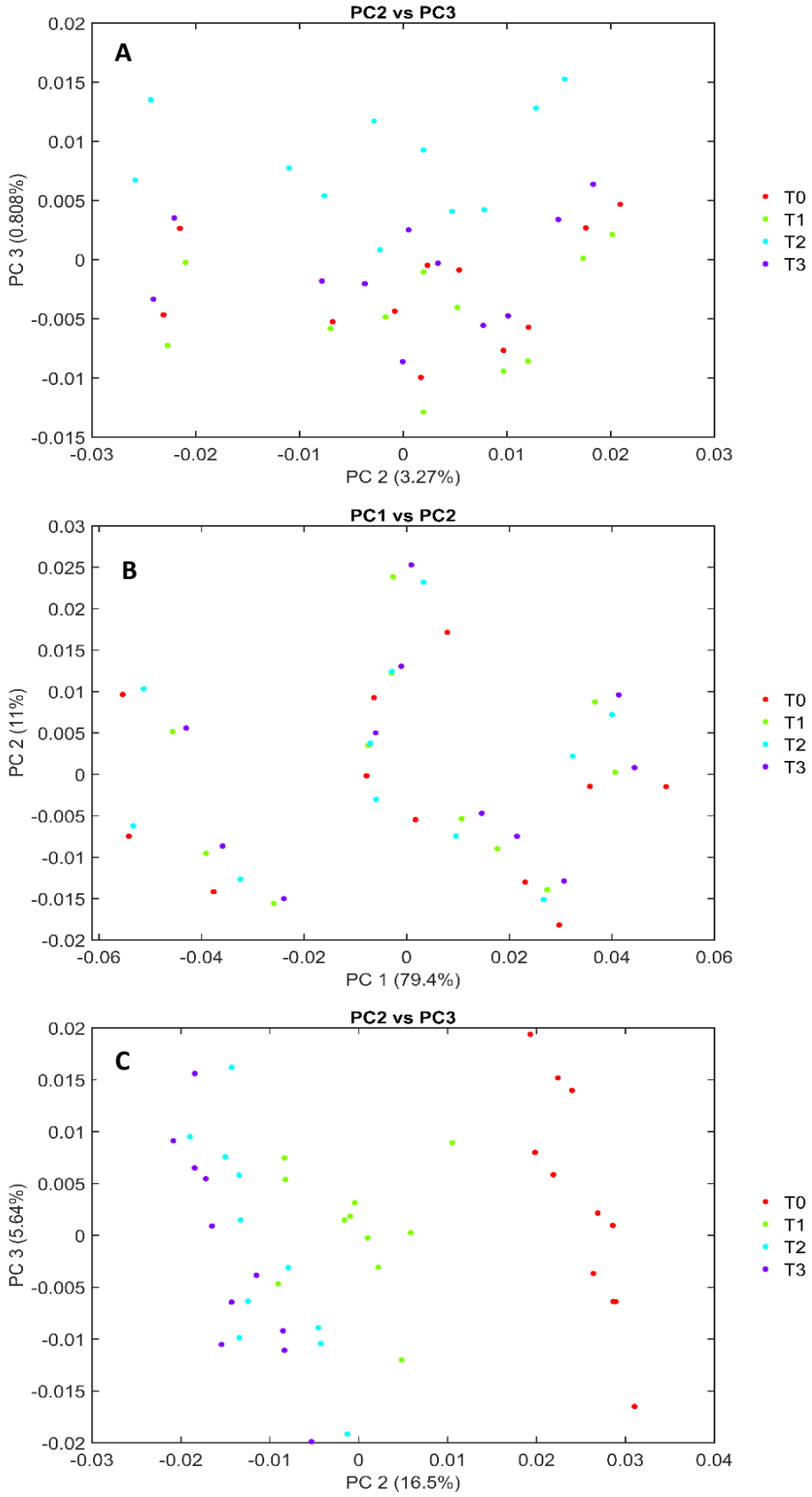

Figure 10: PCA plots showing separation based on the ageing of BT across the three temperature studies showing T2 separates out slightly in the 30°C study (A), no separation in the 37°C study (B) and some separation of T0 and T1 in the 49°C study (C).

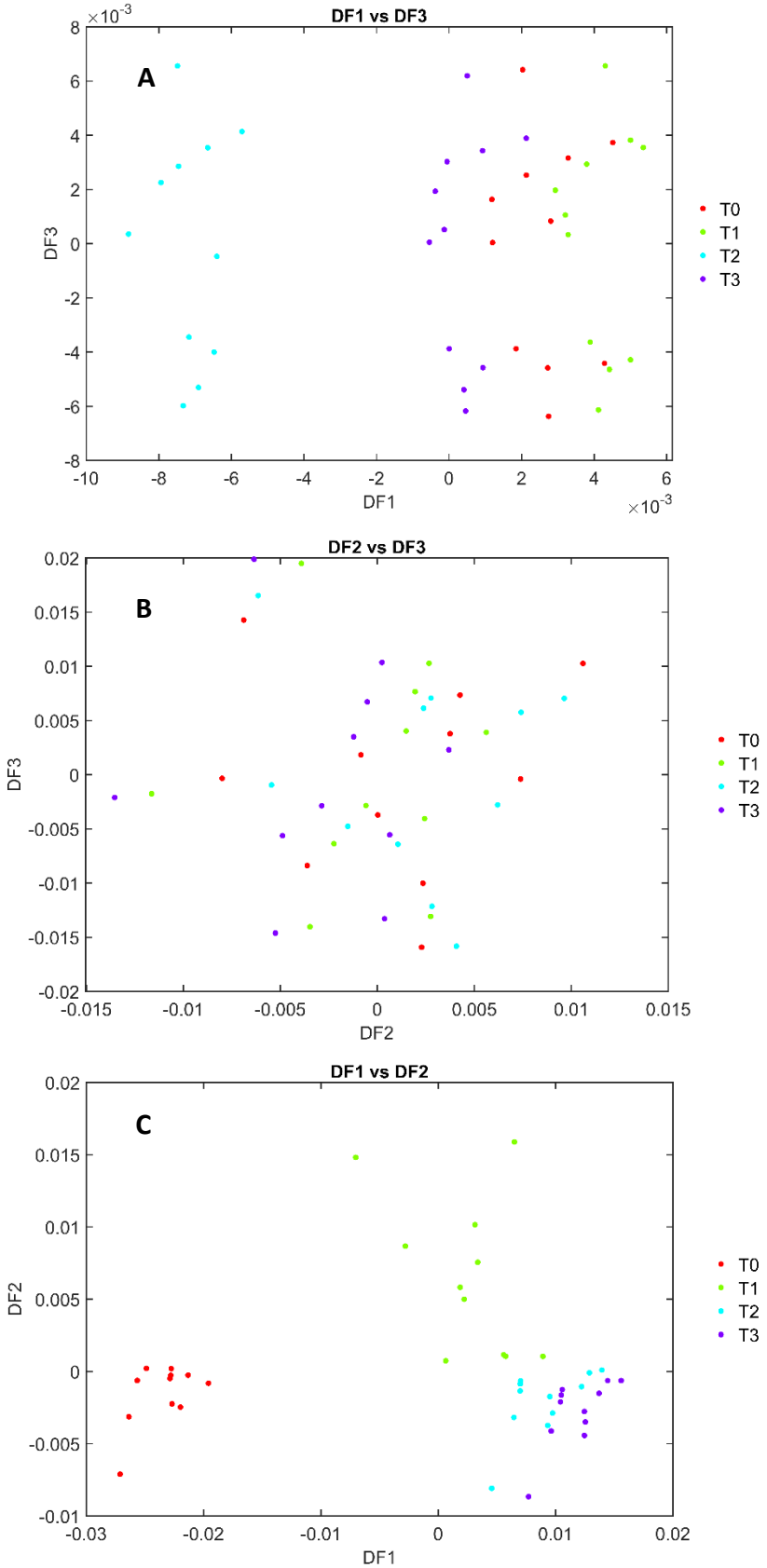

Figure 11: DFA plots showing separation based on the ageing of BT across the three temperature studies showing some separation of the time-points, with no clustering, in the 30°C study (A), no separation or clustering in the 37°C study (B) and clustering of the time-points and some separation of T0 and T1 in the 49°C study (C).

PCA and DFA were performed for all bacterial strains with two bacterial strains (BT and BTKl) shown here. The separation seen in the PCA and DFA plot produced varied for the different strains. The effect that aging or exposure to a given environment has on a bacterium may not be uniform across the seven bacterial strains. Responses to aging or a non-ideal environment are likely to differ most significantly between Gram-positive and Gram-negative strains. However, within these two groups, it is also possible that bacteria will show signs of different mechanisms designed to respond to challenging conditions. Generally, PCA plots produced from the BT dataset shows poor separation of the time points, suggesting there is clear effect on the bacterial profile caused by either aging or temperature. However, the data from the 49°C study shows that the initial time point (T0) separates slightly from the other time points. One particular feature of BT that may affect the way it handles long term exposure to non-ideal/ harsh environments is the presence of parasporal bodies, discussed in Chapter 3 (section 3.2.2). The presence of these bodies may affect the efficacy of the endospores formed which in turn would affect how the exposure to an environment and the aging of a sample could impact its spectral profile. A further, more in-depth study of BT with more regular time points and other environmental conditions may provide more information about why this particular bacterial strain differs from the others. When considering the DFA analysis Figure 11A and C show that there is separation based on time point. Figure 11A shows that it is T2 that separates out the most from the other time-points while it is T0 and T1 that separate in Figure 11C.

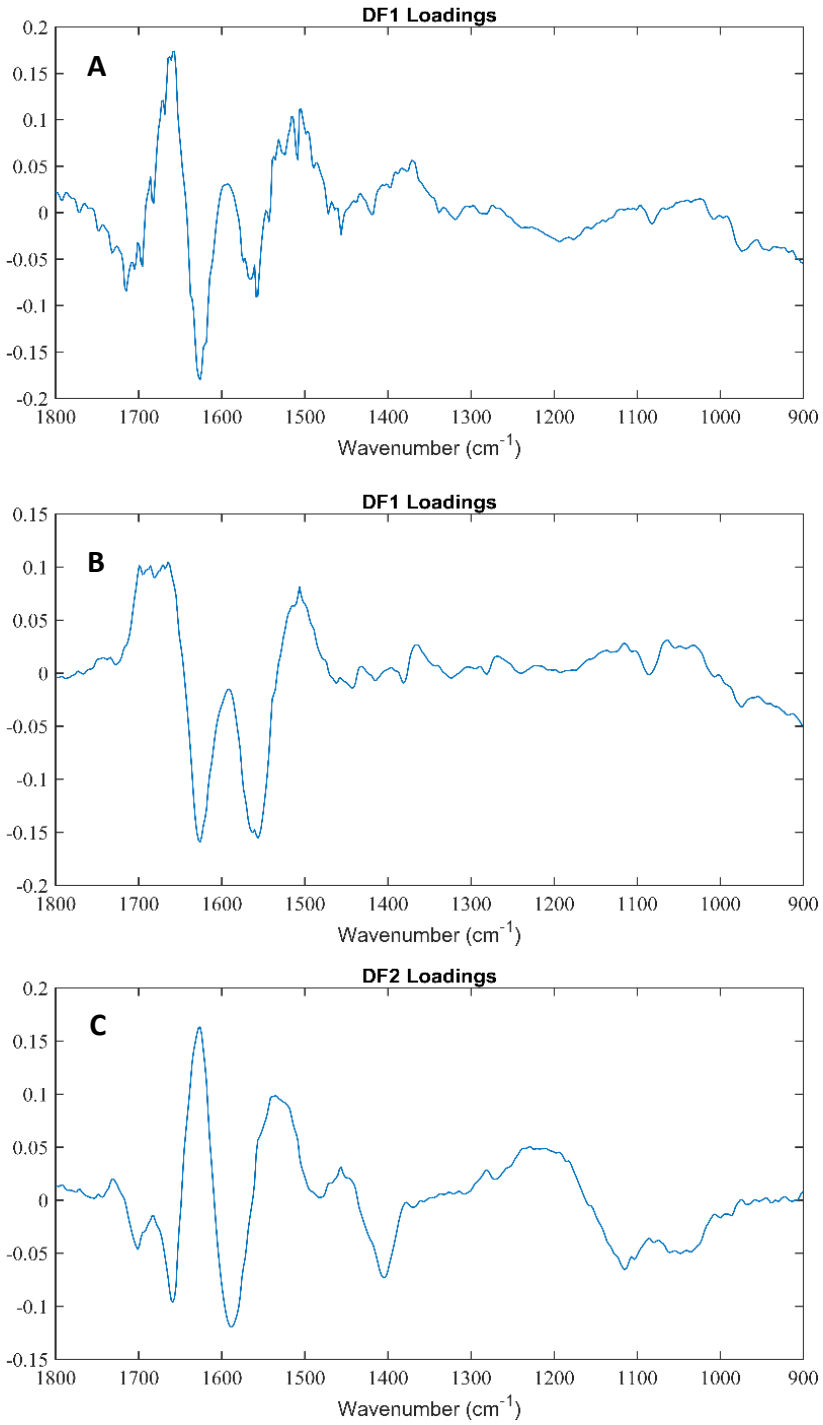

Figure 12: DFA loadings plots from the most discriminatory DF/DFs for each study for BT. (A) shows DF1 for 30°C study, (B) shows DF1 for 49°C study and (C) shows DF2 for 49°C study. Note that no DF loadings are shown for the 37°C as there is no separation seen in the DFA plot for this dataset.

The most important of the loadings plots to consider when looking at Figure 11A is DF1 as this shows the separation of T2 from the other three timepoints. For this the key vibrations identified are all located at the higher end of the wavenumber range covering vibrations from C=O and CH/NH vibrations with most of the peaks being in the Amide I and II region (Figure 12A). This may be as a specific biological response within BT that occurs when aged for seven days within a 30°C environment. Again to fully explore exactly why the data shows this change further repeat experiments would be required to confirm if a specific structural change occurs within BT under these environmental conditions.

For the 49°C study both DFs are responsible for separation, and therefore have had tentative vibrations assigned to the peaks. In Figure 12B most of the peaks again are found at the higher end of the wavenumber range including the Amide I and II regions (1480-1720 cm^{-1}). Figure 12C deals again with Amide I and II as well as the symmetric stretch of P-O-C and the CH_3 stretch.

Table II: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots generated from the BT dataset during the 30°C, 37°C and 49°C studies, acquired using HT-FTIR [2,3,4]

Wavenumber (cm^{-1})	Tentative peak assignments
1103, 1115	νP-O-C symmetric
1371	δCH and NH
1404	δCH_3 asymmetric
1485, 1497, 1499, 1504, 1506, 1516, 1520, 1531, 1535, 1539, 1557, 1558, 1562, 1566, 1574, 1587	Amide II (NH bend coupled to CN stretch)
1618, 1626, 1628, 1635, 1638, 1657, 1663, 1665, 1670, 1686, 1695, 1699	Amide I
1705, 1715	νC=O

In contrast to the plots shown in Figure 10, Figure 11 and Figure 12 for the BT dataset, Figure 13 and Figure 16 show PCA, DFA and loadings plots from BTKl where there is generally a good separation based on temperature. The separation can be seen across all three temperatures, with Figure 14, Figure 15 and Figure 16 showing that the DFA plots for this dataset have an improved separation with clustering of the time-points, particularly in Figure 14A.

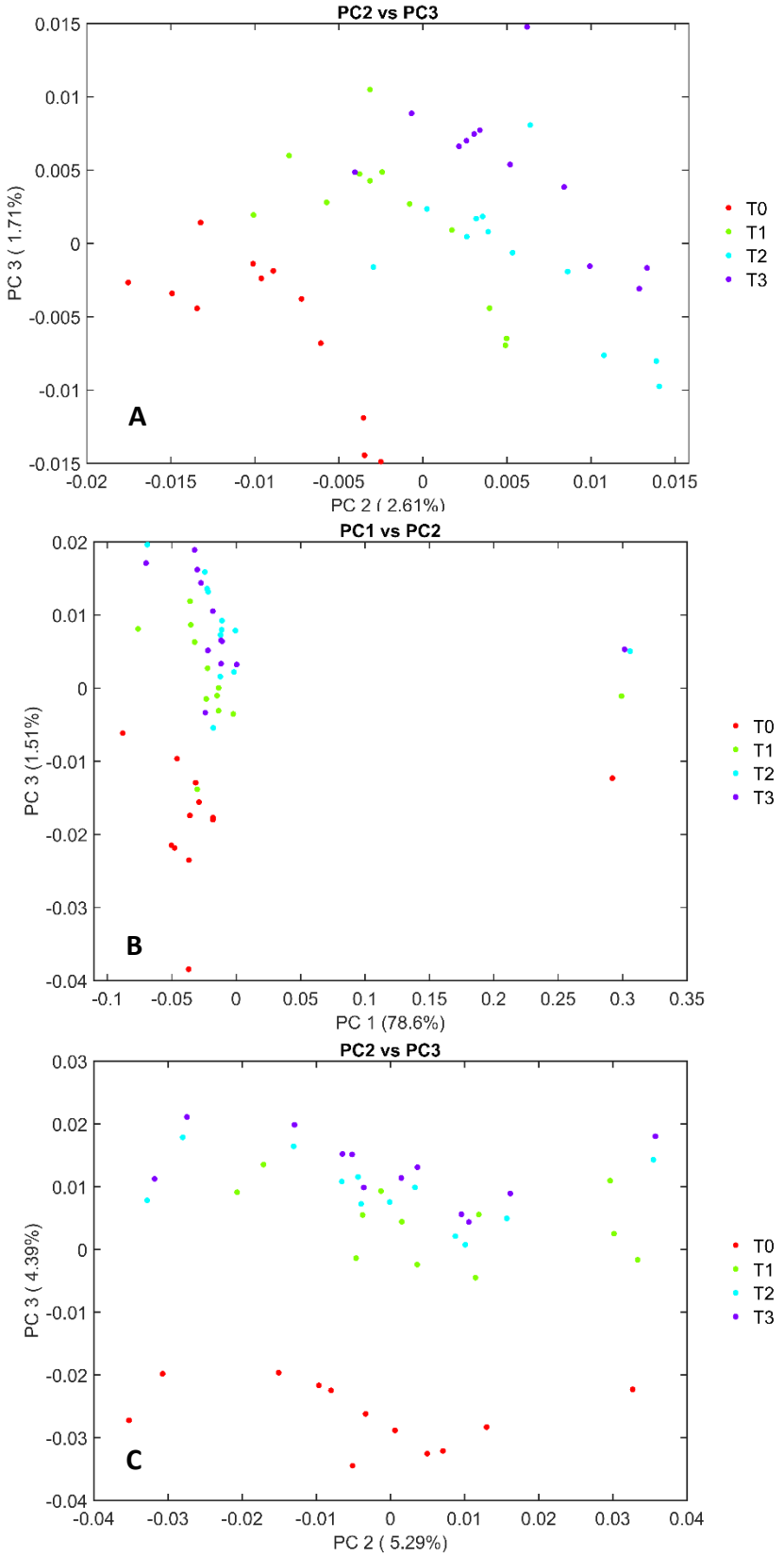

Figure 13: PCA plots showing separation based on the ageing of BTKI across the three temperature studies showing some separation of the time-points, with no clustering, in the 30°C study (A), separation of the T0 time point from the other time-points with no clustering in the 37°C study (B) and clear separation of the T0 time-points from the other time-points in the 49°C study (C).

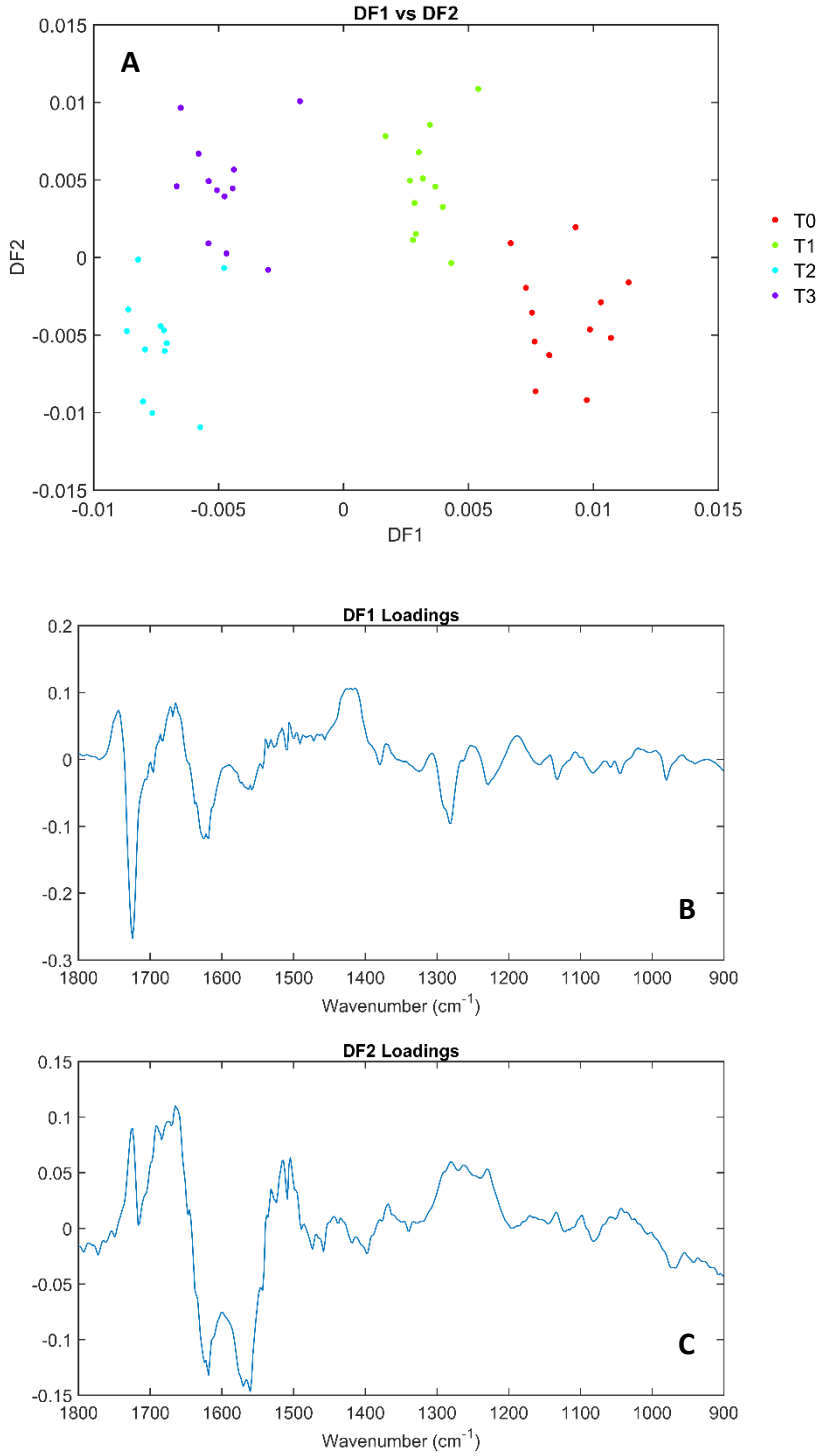

Figure 14: DFA plot showing very clear separation and clustering of the time point for BTKI in the 30°C study (A), followed by a DF1 loadings plot (B) and a DF2 loadings plot (C).

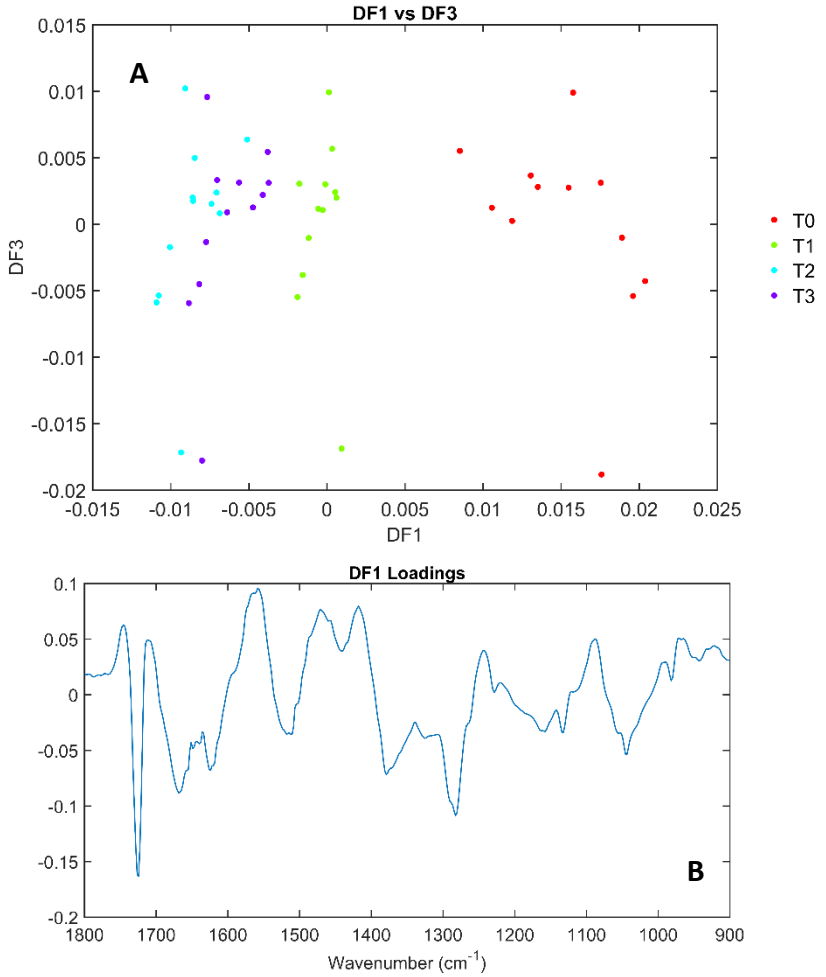

Figure 15: DFA plot showing separation of the T0 time point from the other time-points for BTKI in the 37°C study (A) followed by the DF1 loadings plot (B).

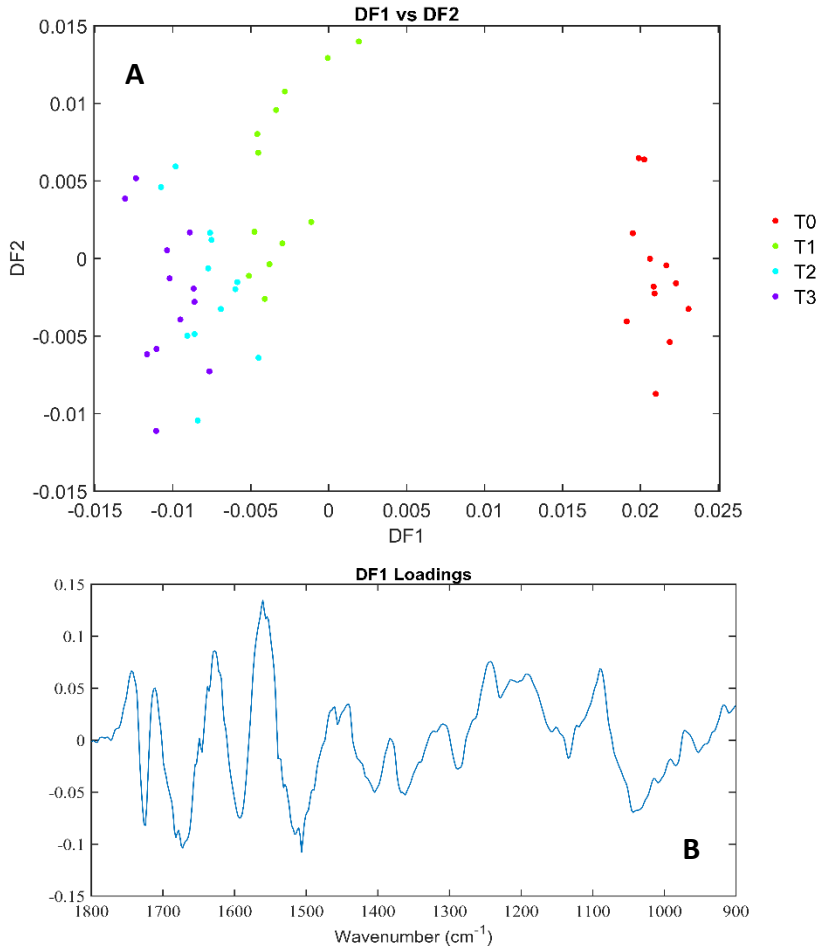

Figure 16: DFA plot showing very clear separation and clustering of the time point for BTKI in the 49°C study (A), followed by the DF1 loadings plot (B).

A general trend seen across the PCA and DFA plots produced from the BTKI dataset is that there is a better level of separation of the timepoints, showing that there is an effect on the bacterial profile produced from BTKI bacteria caused by either aging or temperature. When looking at the PCA plots produced by this dataset, shown in Figure 13, T0 separates out in all three of the temperature studies. This is likely to be based on the moisture content of the samples as the spectra are collected at the initial timepoint where they have had only a limited amount of time to dry. When considering the DFA analysis, Figure 14A shows clear separation and clustering

of the four time-points, showing that the aging of BTKl has a clear effect on the bacterial spectrum. Figure 14B and C show that T0 and T1 separates out from the remaining two time-points with some clustering.

In the 30°C study DF1 is responsible for separation of the earlier time points (T0 and T1) from the later time points (T2 and T3). DF2 then separates T0 and T2 from T1 and T3 (Figure 14A). For both the DF1 and DF2 loadings (Figure 14B and C) in this study the key peaks are found at the higher end of the wavenumber range within the fingerprint region. As such, the majority of the vibrations represented within the loadings are the same across DF1 and 2. These vibrations include C=O and Amide I, II and III stretches. The differing vibrations are the peaks at 1414 and 1425 cm^{-1} in the DF1 loading plot which represent C-H vibrations and ring stretches. The unique peak found within DF2 is found at 1231 cm^{-1} and represents the PO_2^- vibration. Phosphate bands are important to consider as they are found within the cell wall of a bacterium. A change in phosphates may indicate a change within the cell wall and structure, which could be as a survival response to being exposed to a non-ideal, nutrient deficient environment by forming an endospore. [11] For the 37°C study, only DF1 is responsible for any of the separation shown, separating T0 from the remaining time points (Figure 15A). DF1 (Figure 15B) shows the majority of the same peaks that are shown within the DF1 and DF2 loadings from the 30°C study with the exception of PO_2^-. However, the loading has some additional peaks, as the peaks present are spread across the whole of the fingerprint region. The same vibrations that are seen include C=O vibrations and Amide I, II and III stretches as well as ring stretches, while the additional vibrations include CH_2 wagging and vibrations from the Phosphate I

band and the phosphodiester region. As with the 37°C, only one DF shows separation in the 49°C study. Again, DF1 separates T0 from the remaining time points (Figure 16A). Previous vibrations such as C=O, Amide I and II, vibrations in the phosphodiester region and ring stretches are seen. A number of other vibrations are also represented in DF1 including PO_2^- and asymmetric PO_2^- vibrations, Phosphate II and COC, CC and COH vibrations (Figure 16B).

Table III: Table with the wavenumbers and corresponding tentative peak assignments for the DF loadings plots generated from the BTK- dataset during the 30°C, 37°C and 49°C studies, acquired using HT-FTIR [2,3,4]

Wavenumber (cm^{-1})	Tentative peak assignments
964.4, 972.1, 1032, 1043, 1046	Phosphodiester region, vCO, vCC, vCOH and vCOC
1072, 1088	Phosphate I (stretching PO_2^- symmetric vibrations) in *B*-form DNA
1090	Phosphate II (stretching PO_2^- asymmetric vibrations) in β-form DNA
1192	vC-O-C, C-C, C-O-H, C-O-C
1214	vPO_2^- asymmetric
1231, 1242	vPO_2^-
1261, 1263, 1281, 1283, 1290	Amide III
1362, 1369, 1379	CH_2 wagging
1404, 1414, 1418, 1425, 1458, 1470, 1472	vC-H
1487, 1489, 1495, 1504, 1506, 1514, 1516, 1543 1555, 1558, 1560, 1564, 1570, 1593	Amide II (NH bend coupled to CN stretch)
1611, 1618, 1620, 1624, 1626, 1638, 1657, 1665, 1666, 1668, 1672, 1674, 1682, 1692, 1699	Amide I
1711, 1724, 1726, 1740, 1744, 1746	vC=O

5.6 Analysis of data using Multiblock PCA

As there are a large number of variables to be considered and the separation of the data is not particularly clear with regards to the strain and age of the bacteria using standard PCA, Multiblock PCA was used to try to improve the separation caused by creating different groups referred to as 'blocks'. Figure 17 and Figure 18 show the multiblock PCA score plots from blocking the datasets by time point. This disregards the differences caused by different bacterial strains and looks simply at the separation that is cause by the stage of aging of the bacteria. Successful separation of the time points here would allow for a spectrum to be aged regardless of the bacterial strain.

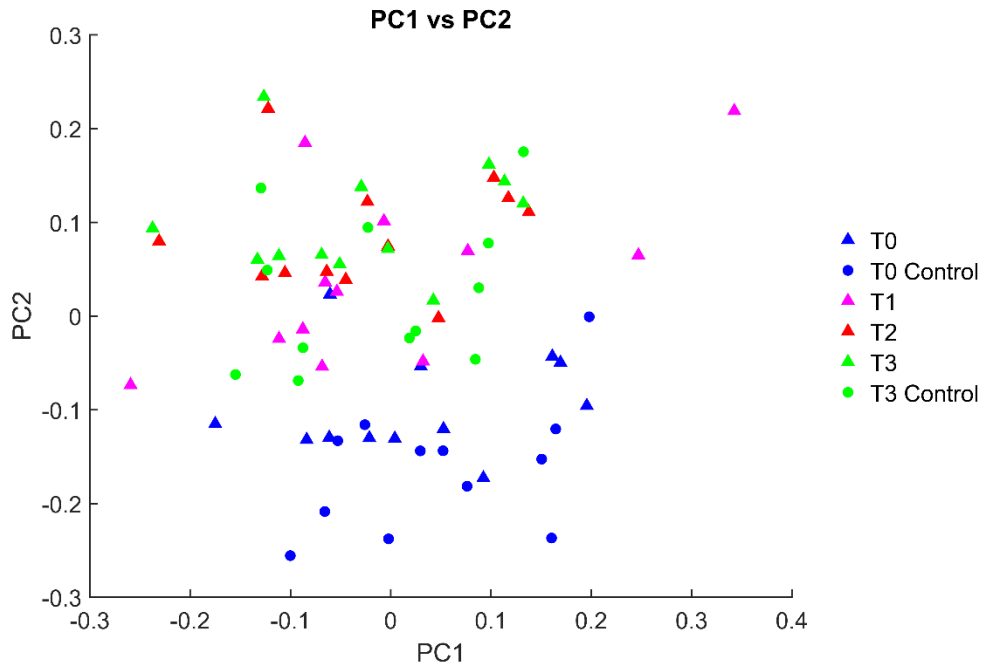

Figure 17: Multiblock PCA plot showing separation of the 49°C dataset, based on time point not bacterial strain, showing the plot of PC1 vs PC2.Data in this plot has been pre-processed using noise reduction (NR) and vector normalisation (VN).

Figure 17 shows a multiblock PCA produced from the 49°C data that has been noise reduced, vector normalised and cut to the fingerprint region. It can be seen from Figure 17 that T0 and T0 control data is separating out from the remaining three time-points. However, the remaining data are severely overlapped. From Figure 18, as with Figure 17, the T0 and T0 control clearly separates out from the other time-points. While the other time-points still overlap, in the EMSC processed data the T3 and T3 control samples are beginning to separate out from the intermediate time-points. Bases on the PC1 loadings plots (Figure 19) the separation seen from PC1 in both of the PCA plots is caused by the Amide I and II vibrations. The lower wavenumber Amide II and the higher wavenumber Amide I vibrations are found within the T0 samples while the higher wavenumber Amide II vibrations and the lower wavenumber Amide I vibrations are found within T1, T2 and T3 samples.

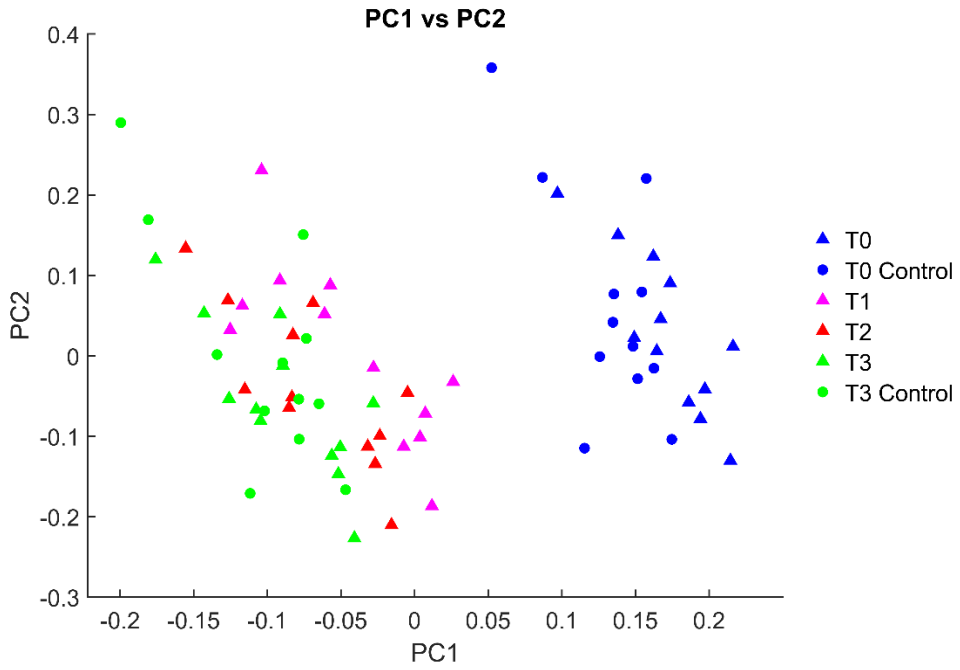

Figure 18: Multiblock PCA plot showing separation of the 49°C dataset, based on time point not bacterial strain, showing the plot of PC1 vs PC2.Data in this plot has been pre-processed using extended multiplicative scatter correction (EMSC).

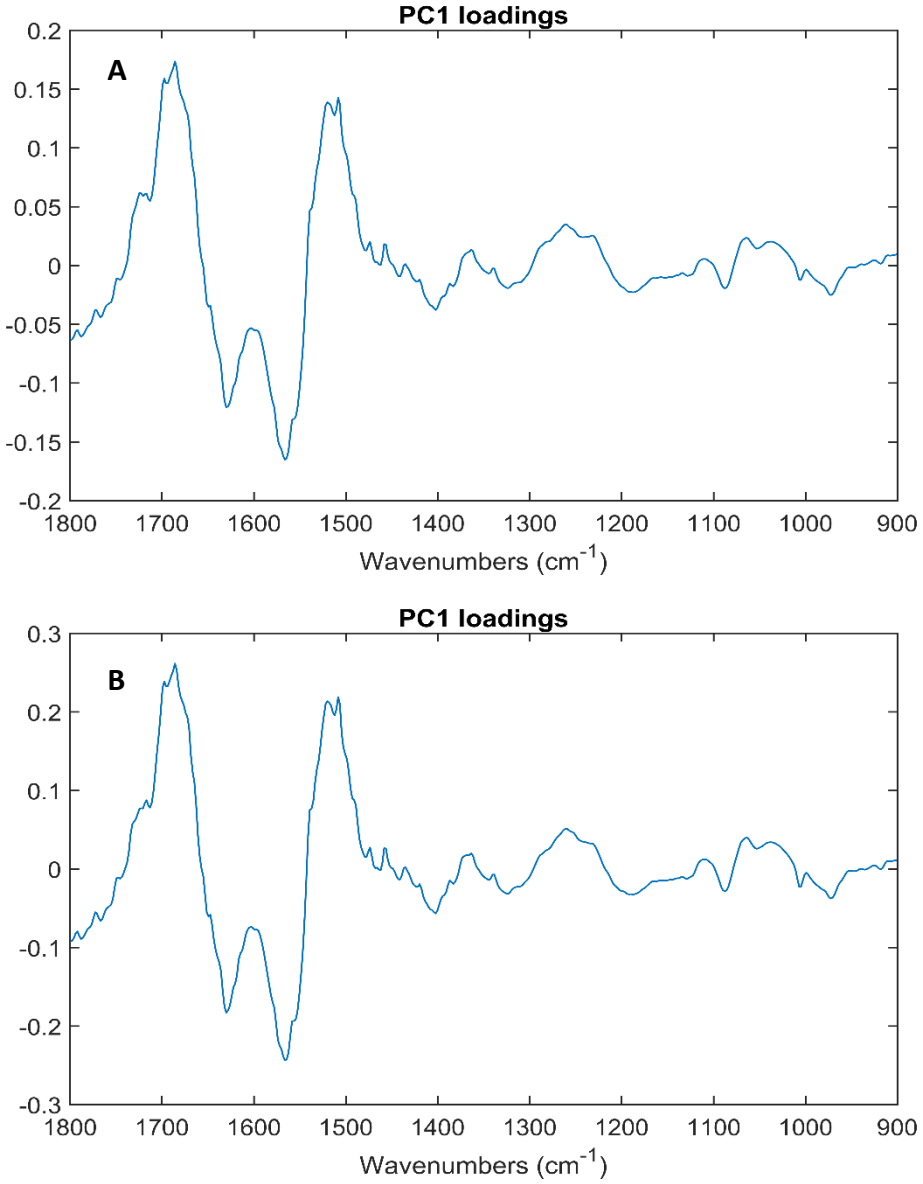

Figure 19: PC1 loadings plot for the 49°C multiblock PCA plot where data has been pre-processed using VN/NR (A) and EMSC (B).

Table IV: Table with the wavenumbers and corresponding tentative peak assignments for the PC1 loadings plot for multiblock PCA generated from the 49°C dataset processed with both VN/NR and EMSC [2,3,4]

Wavenumber (cm^{-1})	Tentative peak assignments
1500, 1508, 1520, 1557, 1566, 1572, 1580	Amide II (NH bend coupled to CN stretch)
1622, 1630, 1674, 1686, 1697	Amide I

This analysis shows that changes caused by aging has an effect of the bacterial profile, changes which can be detected regardless of bacterial strain. To take this further and test the extent to which the multiblock PCA can split based on time point, a further study with more bacterial strains and time points could be conducted. Due to the time and samples restrictions this study was not conducted within this PhD thesis.

HT-FTIR is able to differentiate between spectra collected at various stages of aging within a range of environmental temperatures. The process of identification would require a two-step approach, using PCA first to identify bacterial strain and then using a strain-specific dataset to age the bacteria. However, the ability for the bacteria to be correctly aged is dependent on the strain identified, as the quality of the separation of the various time points is dependent on the bacterial strain. When performing multiblock PCA analysis it can be seen that the initial time points begin to separate from the other time points when grouping all bacterial strains together. This shows that the bacteria are affected by aging and the changes can be detected in a spectral profile regardless of bacterial strain. However, when considering the overall aim of the project HT-FTIR is not practical as it is currently only a laboratory-based technique.

5.7 Conclusions

PCA and DFA were performed on datasets collected using both HT-FTIR and DRIFTS. The HT-FTIR study housed four individual studies (three different temperatures and a solar study) with DRIFTS having three different temperature studies. Multiblock PCA has also been performed on the four HT-FTIR datasets. This in turn has generated a large amount of visual outputs in the form of scores plots loadings plots and peak assignment tables. Shown within this chapter are a key selection of the results obtained during the course of the project with the remaining key figures and tables found in Appendix D.

5.7.1 Comparison of HT-FTIR and DRIFTS Results

When considering the effect of aging and exposure to an environment, the separation and clustering of data from different time points is generally more defined with DRIFTS than with HT-FTIR despite the corresponding loadings for DRIFTS containing a significant level of noise. This may be due to the different instruments and modalities that have been used to acquire the data. The differences in separation between DRIFTS and HT-FTIR are not thought to be the results of biological variations between the samples as they were grown and prepared in the same manner. Compared with other FTIR modalities, DRIFTS is generally a noisier technique which is seen in the standard absorbance spectra, and is also carried through to the absorbance spectra that have been produced. From both the HT-FTIR and the DRIFTS

studies, it is clear that there is separation and clustering of the bacterial strains when using either technique. The quality of the clustering and separation of the bacteria is generally enhanced by performing DFA analysis as well as PCA, rather than just using PCA alone to investigate the separation. Both techniques show that a good level of separation can be seen after performing individual strain analysis with respect to analysing the effect that aging has on the bacterial spectrum. A similarity seen across both of the techniques is that the effect of the aging differs depending on the bacterial strain that is being examined. This may be as a result of the Gram-type of the bacteria and their ability to form endospores as Gram-positive bacteria will from them while Gram-negative strains will not. The ability to form these spores will affect how a bacterium responds to being in a non-ideal environment with bacteria that do not form spores being more susceptible to damage and degradation. Another possibility, when considering BT, is the presence of parasporal bodies within the cells. These will also affect the cells ability to form spores. The variation in the separation seen within these studies highlights the importance of understanding how a bacterium's spectral profile changes both over time and when exposed to differing environmental conditions. The ability for both of these techniques to identify the changes caused by environment and aging shows promise for the DRIFTS technique to be used for the *in-situ* analysis of surface deposited bacterial samples.

Generally, when examining all bacterial strains together, HT-FTIR performs better at separating the Gram-positive from the Gram-negative strains. However, when examining the effect of aging on the bacteria for individual bacterial strains, it is DRIFTS that performs better. One of the key differences between the two

techniques is that more biochemical information can be interpreted from the loadings produced when using HT-FTIR compared to DRIFTS. However, DRIFTS spectra are unprocessed and collected from complex backgrounds, both of which will contribute to the quality of the spectra and loadings plots produced.

Both modalities have shown that Gram-type and species are the overriding factors for separation rather than the effect of aging or environment. The DRIFTS study went one step further and also considered the effect that the surface from which a spectrum was collected may have on the resulting absorbance spectrum. It was found that there was a level of overcompensation with the more porous substrates, while the non-porous substrates encountered issues with the samples getting damaged more easily. Despite this there were good levels of separation, particularly based on time point when interrogating an individual bacterial strain.

Across all studies, MVA of the datasets showed that with unsupervised analysis there was limited separation and clustering, while with supervised DFA analysis clustering and separation is much clearer. This is due to the fact that more information is initially given to the DFA model to support the analysis. This shows promise that there could one day be a methodology developed, using vibrational spectroscopy, to allow for *in situ* analysis of surface deposited bacterial samples found on complex matrices.

When taking into consideration the overall aim of the project which is the development of a technique that can be used to identify surface deposited bacteria in situ, DRIFTS is the technique that has shown that it can address this. While HT-FTIR

has shown that it is capable of identifying strain and age for a variety of environmental conditions, the technique is currently only laboratory based and therefore cannot address *in situ* analysis, as well as not being able to collect data from different substrates. The DRIFTS modality used on a handheld instrument allows for the analysis of samples in any location and on any given surface. While the preliminary data shows that DRIFTS potentially has the capability to tackle the overall aim of the project it is acknowledged that the method is far from refined and therefore more research is required to make the technique and methodology fully translatable to allow *in situ* analysis.

5.8 References

1 MoD. Environmental Handbook for Defence Materiel. Part 4: Natural Environments., 332 (Ministry of Defence, 2000).

2 Hayashi, H. & Mukamel, S. Vibrational-Exciton Couplings for the Amide I, II, III, and A Modes of Peptides. *Journal of Physical Chemistry* **111**, 11032-11046 (2007).

3 Naumann, D. *Infrared Spectroscopy in Microbiology*. 1-18 (2006).

4 Davis, R. & Mauer, L. J. in *Current research, technology and education topics in Applied Microbiology and Microbial Biotechnology Volume II.* Vol. 2 (ed A. Mendez-Vilas) Ch. Fourier transform infrared (FT-IR) spectroscopy: A rapid tool for detection and analysis of foodborne pathogenic bacteria, 1582-1594 (Formatex Research Center, 2010).

5 Baker, M. J., Hughes, C. S. & Hollywood, K. A. *Biophotonics: Vibrational Spectroscopic Diagnostics*. (Morgan & Claypool, 2016).

6 Movasaghi, Z., Rehman, S. & ur Rehman, I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews* **43**, 134-179 (2008).

7 Horiba. Raman Data and Analysis: Raman Spectroscopy for Analysis and Monitoring. <http://www.horiba.com/fileadmin/uploads/Scientific/Documents/Raman/bands.pdf>.

8 Amyes, S. G. B. *Bacteria: A very short introduction*. 1st edn, (Oxford University Press, 2013).

9 Johnson, T. J. *et al.* The Infrared Spectra of Bacillus Bacteria Part I: Vegetative Bacillus versus Sporulated Cells and the Contributions of Phospholipids to Vegetative Infrared Spectra. *Applied Spectroscopy* **63**, 899-907 (2009).

10 Johnson, T. J., Williams, S. D., Valentine, N. B. & Su, Y. F. The Infrared Spectra of Bacillus Bacteria Part II: Sporulated Bacillus—The Effect of Vegetative Cells and Contributions of Calcium Dipicolinate Trihydrate,CaDP·3H2O. *Applied Spectroscopy* **63**, 908-915 (2009).

11 Naumann, D. FT-Infrared and FT-Raman spectroscopy in biomedical research. *Applied Spectroscopy Reviews* **36**, 239-298, doi:https://doi.org/10.1081/ASR-100106157 (2001).

Chapter 6

Overall Conclusions and Scope for Future Work

This chapter discusses the overall findings from the research that has been conducted during this project and the conclusions that can be drawn from them. The chapter looks at the information that has been concluded from each of the studies and an outlook to the future of *in situ* bacterial identification.

The main objective for this research was to develop a method for in situ identification of surface deposited bacteria. Collectively the data from chapters 2,3 and 4 show that buy using a handheld spectrometer using DRIFTS it is possible to collect spectra from surface deposited bacteria. To reach this overall conclusion four separate studies were conducted with the results of the previous study leading into the next to help direct the project towards the best method.

The overall aim of the project is to develop a method of *in situ* identification of surface deposited bacteria on a given set of substrates. The data presented in chapter 2 shows the importance of understanding the composition of the substrates. To be able to develop a method where the spectral contribution from the substrate can be removed a database was compiled of reference spectra. In order to make the method as robust as possible a more extensive database of reference spectra would need to be developed. This database should include spectra from a variety of different compositions, ages and conditions. For the method to be translatable to

identify bacteria in a '*real-world scenario'* spectra will be collected from any given surface. These substrates will have been subjected to a variety of different environmental conditions which will not only have an impact on the spectral profile of the bacterium but also on the profile of the substrate.

One of the aspects to be considered when developing this method is the most suitable spectroscopic technique to use. During this study, it showed that ability to obtain Raman spectra was often dependant on the laser used. The 785 nm laser performed better than the 532 nm and 633 nm lasers for collecting reference spectra as it was less affected by fluorescence. However, the quality of spectra acquired when using the 785 nm laser was variable depending on the substrate being interrogated. To develop this study further, a wider range of laser lengths, including from outside of the fluorescence region, such as 233 or 1064 nm, could be tested. An advantage of Raman is that the surface of a sample can be examined. Therefore, depending on the thickness of a sample, the substrate my not contribute to the spectrum at all. Despite this, due to the wide variety of parameters that can be used to obtain the most optimal Raman spectrum, the technique is not ideal for novice users that need to be able to collect the best spectrum at the first attempt given the scenario that he technique is likely to be used in.

FTIR proved to be a more user-friendly technique that produces clear spectra from each of the eight substrates. For the FTIR aspect of the study Ge and DiZnSe IRE were tested to see which of them was optimal. It was found that DiZnSe produced much more intense spectra and was taken forward for further investigation. A strong

spectroscopic signature from the substrate is often considered problematic, however, it is important to have the most intense spectral profile when considering the possibility of a contaminant being found on the surface.

To remove the spectral contribution from the substrate it was the Beer-Lambert method that proved to be optimal. The method would have benefits over other methods by accessing a virtual database stocked with reference spectra. Glucose and ovalbumin were used as the samples for this study, however, developing the study using more complex sample types or even combinations of various compounds would further test the methodology to assess the areas in which the method could be improved. This would also give an indication of which sample types are most suited to this type of analysis. The main area in which this study needs to be developed is by adding more substrates to develop the database of surfaces that can be analysed, including a mix of porous and non-porous substrates.

One of the fundamental aspects to this research was the successful identification of bacteria. Part of this considered the pre-processing and MVA that was to be performed on the data, specifically the order in which the data was processed. By not including the derivatisation step it makes the process easier and more user-friendly as it removes a step from the process as well as keeping the results easier to interpret. Non-derivatised loadings are easier to provide tentative assignments for, which in turn is then related to the structure of the sample being analysed. This allows conclusions about what is causing the separation of the data to be assigned to the structural differences within the samples.

To expand on this study more bacterial strains could be added. This would allow the model to be tested further to ensure that it can classify the bacteria correctly if more closely related strains were added.

To build the most robust model possible for this technique, the study would benefit from additional bacterial strains, substrates and other environmental conditions. This study was based on the temperature ranges observed for a hot dry climate as defined by the MoD, as this is the environment where BWA samples are most likely to be found. However, that does not mean that they will exclusively be found in these environments. Databases of each of the different environments, as classified by the MoD, would need to be compiled to provide the most accurate classification possible. Other ways in which this study can be developed is to look at multiple factors that combine to define an environment, such as the effect of temperature and solar radiation. One final aspect of analysis that is important to consider in future studies is how well the instrument works in the given environments. The samples were stored within these environments, however, for analysis the samples and instrument were used at standard room temperature. Testing the equipment's ability to work in different environments and the quality of spectra produced in these environments will give a true indication of suitability of the technique for translation to identify surface deposited samples, found on complex matrices, *in situ* in any given environment.

This project has many different avenues which can be further developed to support and enhance the work carried out here. There are significant developments

to be made in this line of analysis and the development of a method that analyses surface deposited bacteria on complex matrices is the perfect starting point. There is a significant drive towards ensuring that both analysis of samples and the interpretation of results is both rapid and user-friendly. To make this possible, this project has involved moving away from bench top instruments to test handheld technology that can be taken to the sample, rather than the sample having to be brought into the lab to be analysed. Adding in other bacterial strains to the study begins the development of a reference database that will provide more accurate identification of the samples. Along with this, adding in additional substrates, as mentioned previously, will also aid with the identification of samples. To make both the bacterial and the substrate databases robust, spectra need to be incorporated from different grades of sample. This means that non-ideal samples need to be included to add extra information about the sample that is being identified. Including spectra from samples of varying ages and conditions, as well as manufacturers, will allow the database to be used to answer multiple questions about a given sample.

Ultimately this research project has shown that classification of data obtained from surface deposited bacteria found on complex matrices using vibrational spectroscopy is possible. The optimal technique has been downselected and a methodology developed to allow for the *in-situ* analysis of samples. This shows great promise for the future analysis of all sample types, and is of particular interest for the safe analysis of BWAs in the field.

Appendix A

ATR-FTIR Substrate Study Results

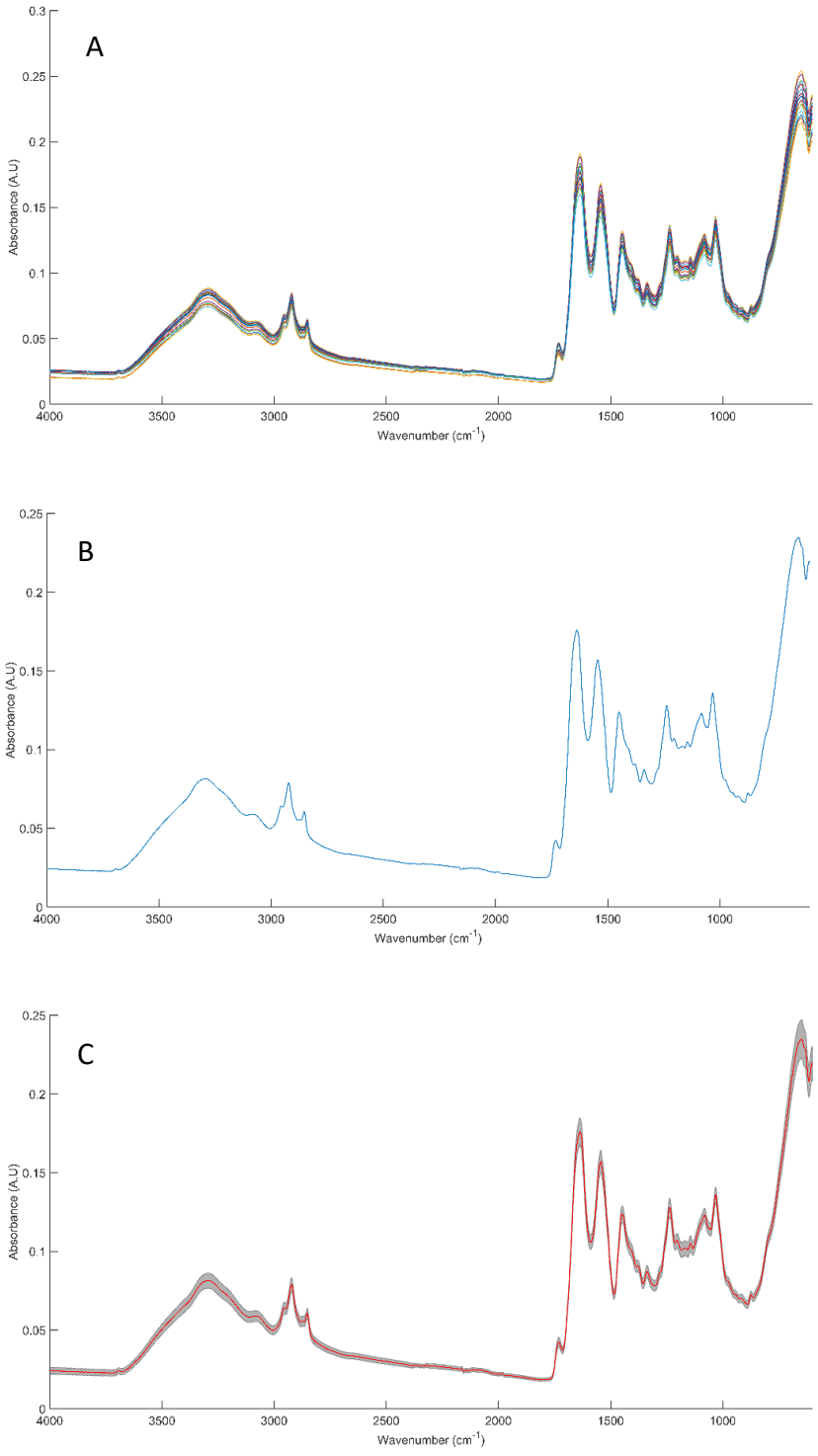

Figure 1: Collection of spectra collected from Boot acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C)

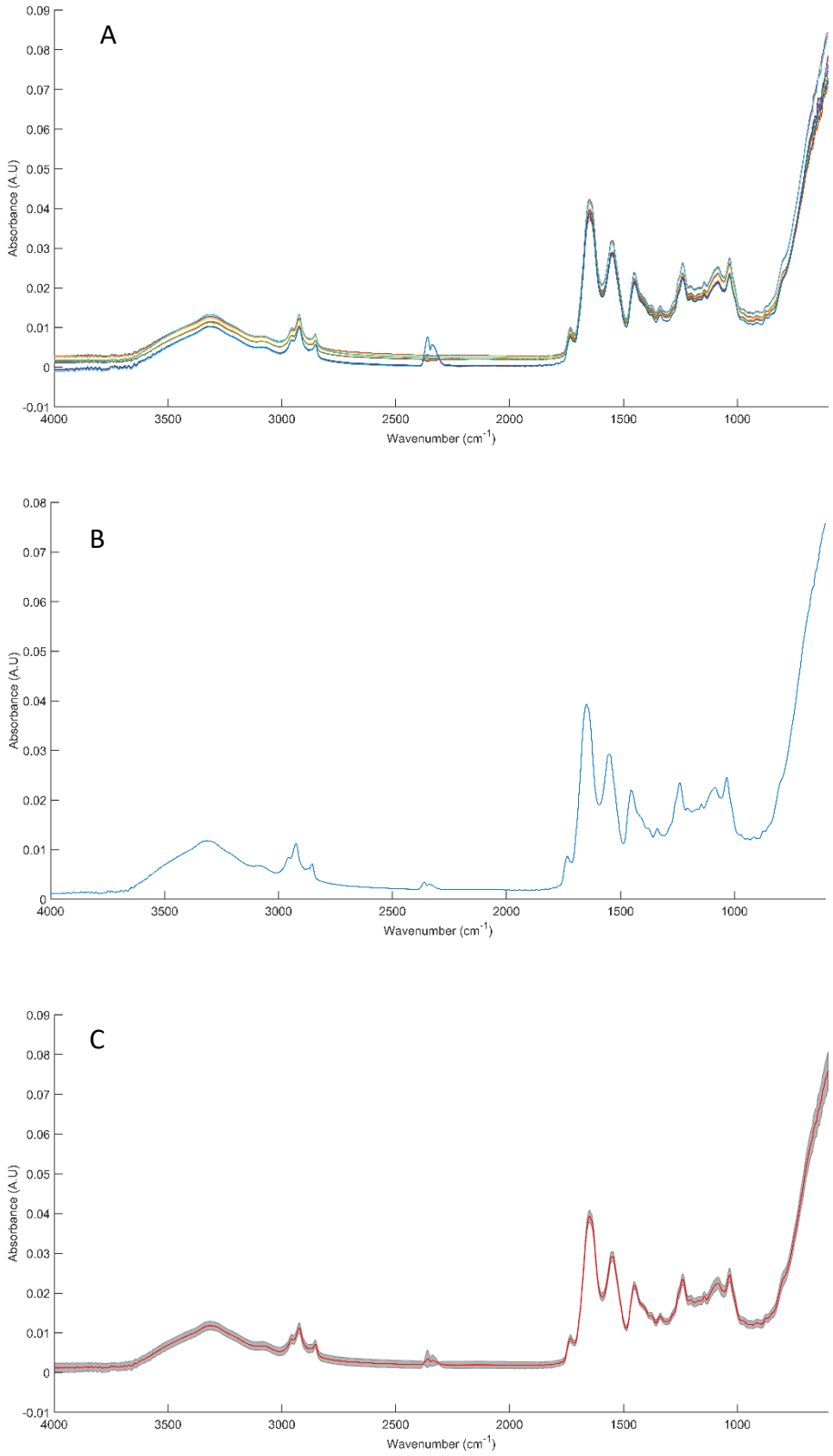

Figure 2: Collection of spectra collected from Boot acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C)

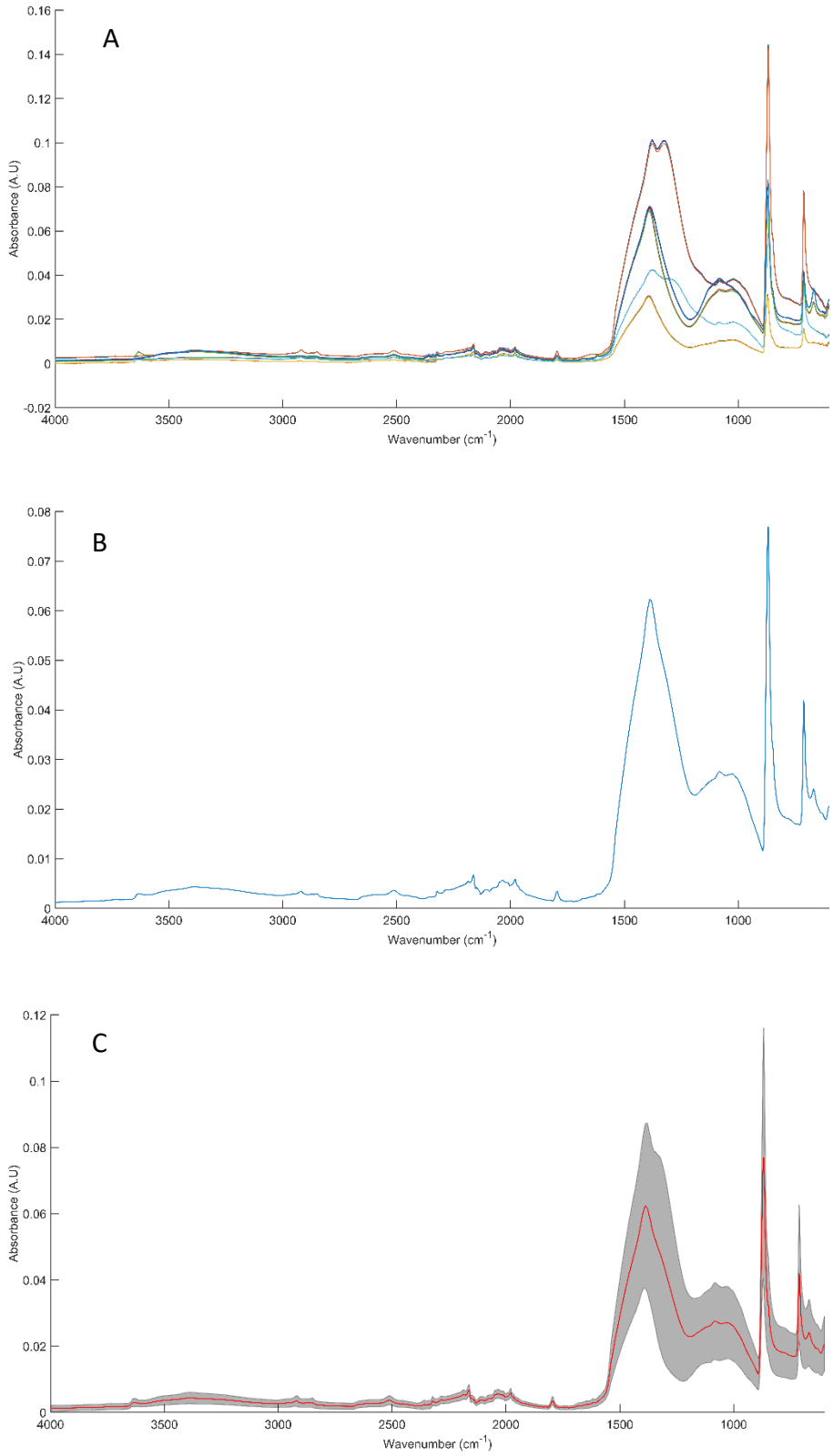

Figure 3: Collection of spectra collected from Concrete acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

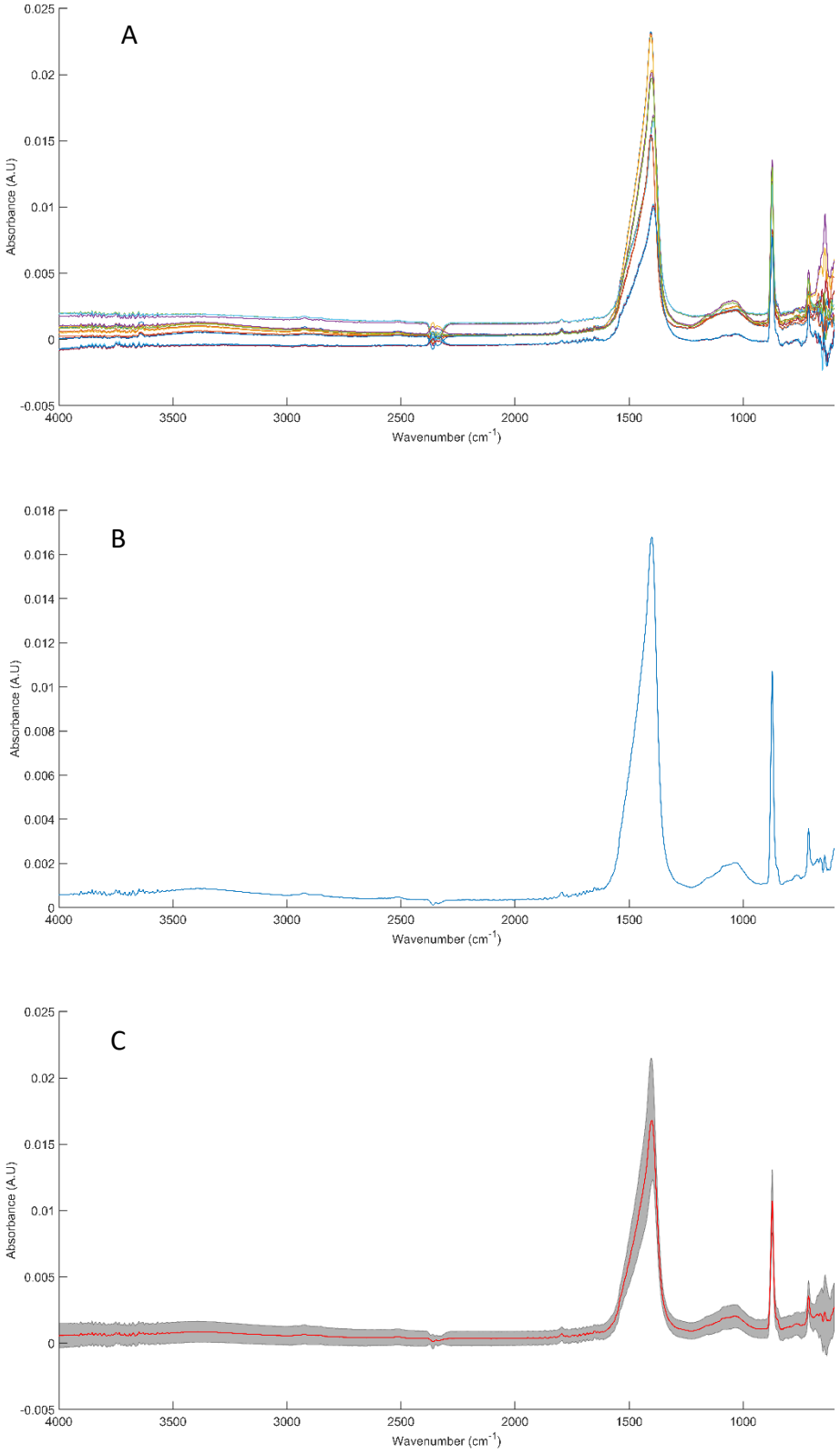

Figure 4: Collection of spectra collected from Concrete acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C)

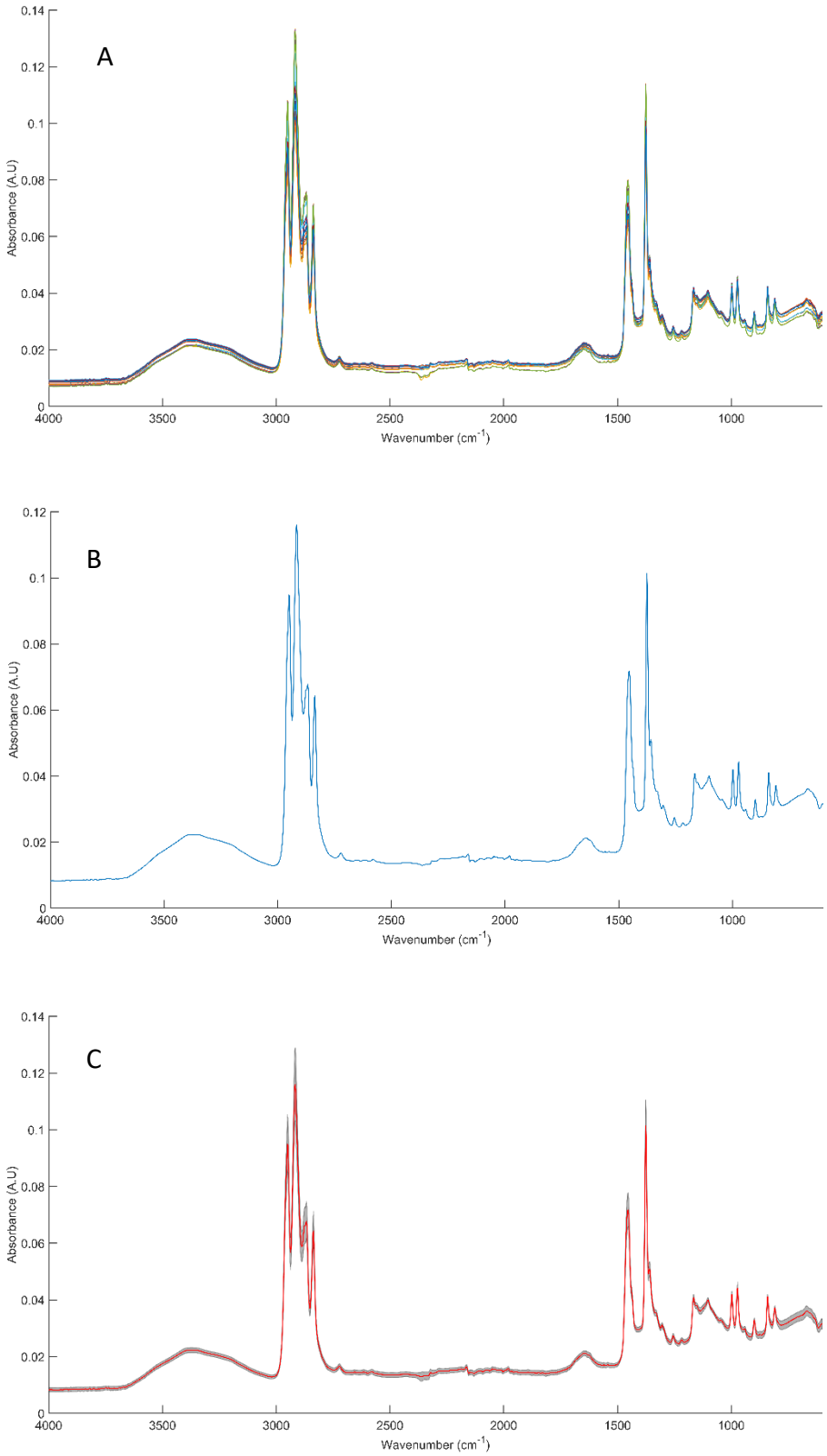

Figure 5: Collection of spectra collected from Dirty filter acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C)

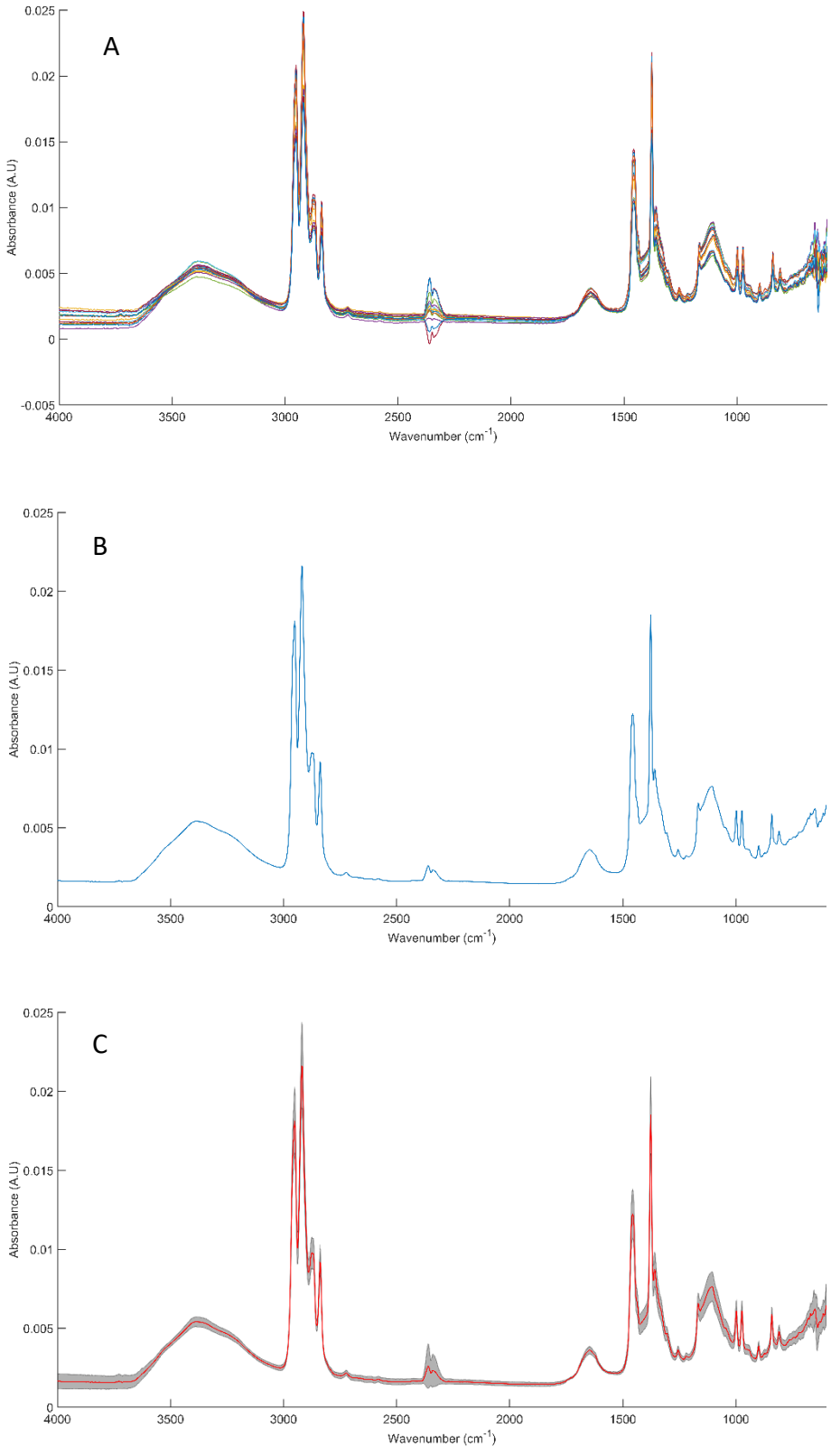

Figure 6: Collection of spectra collected from Dirty filter acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

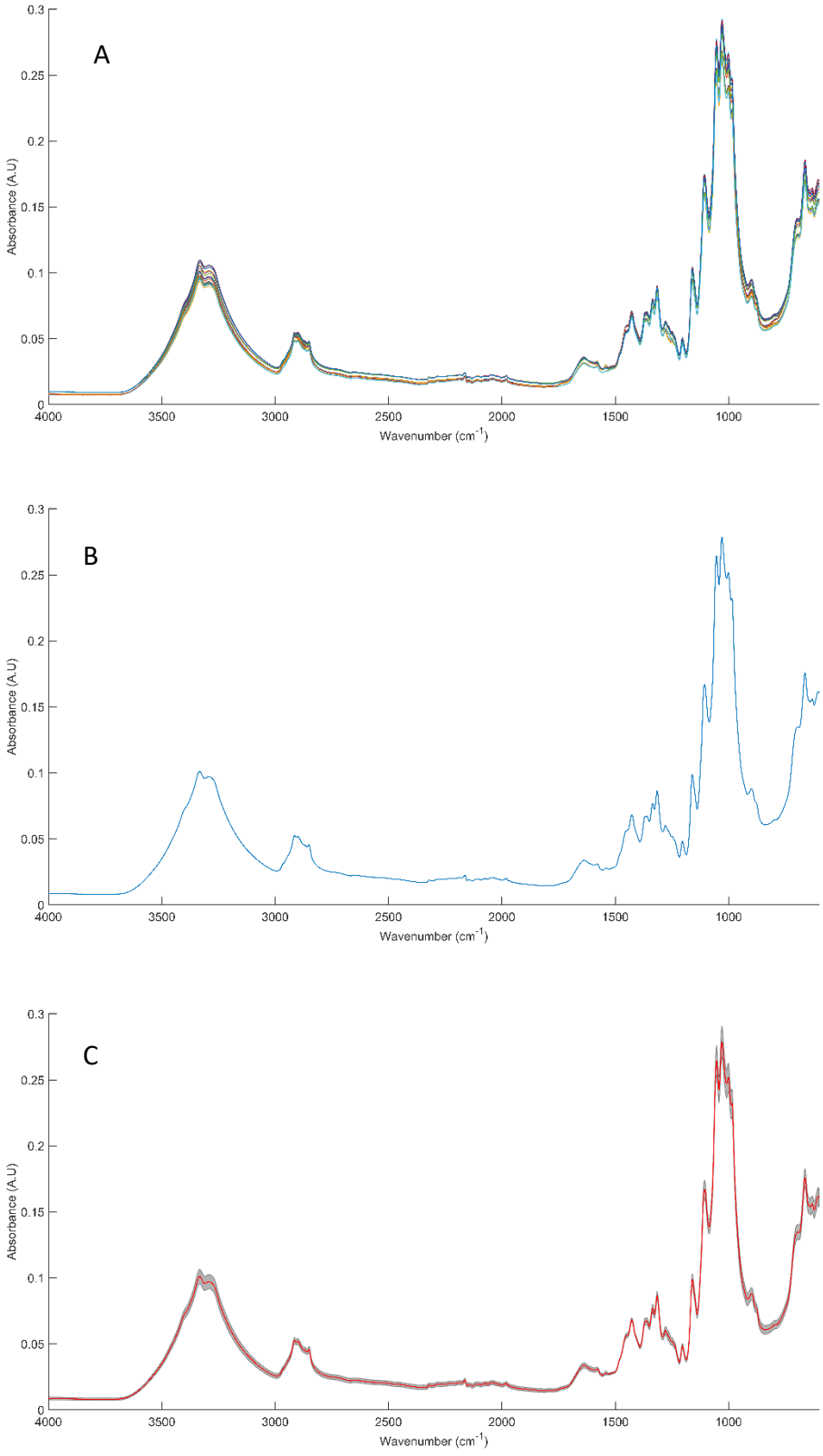

Figure 7: Collection of spectra collected from Jacket acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

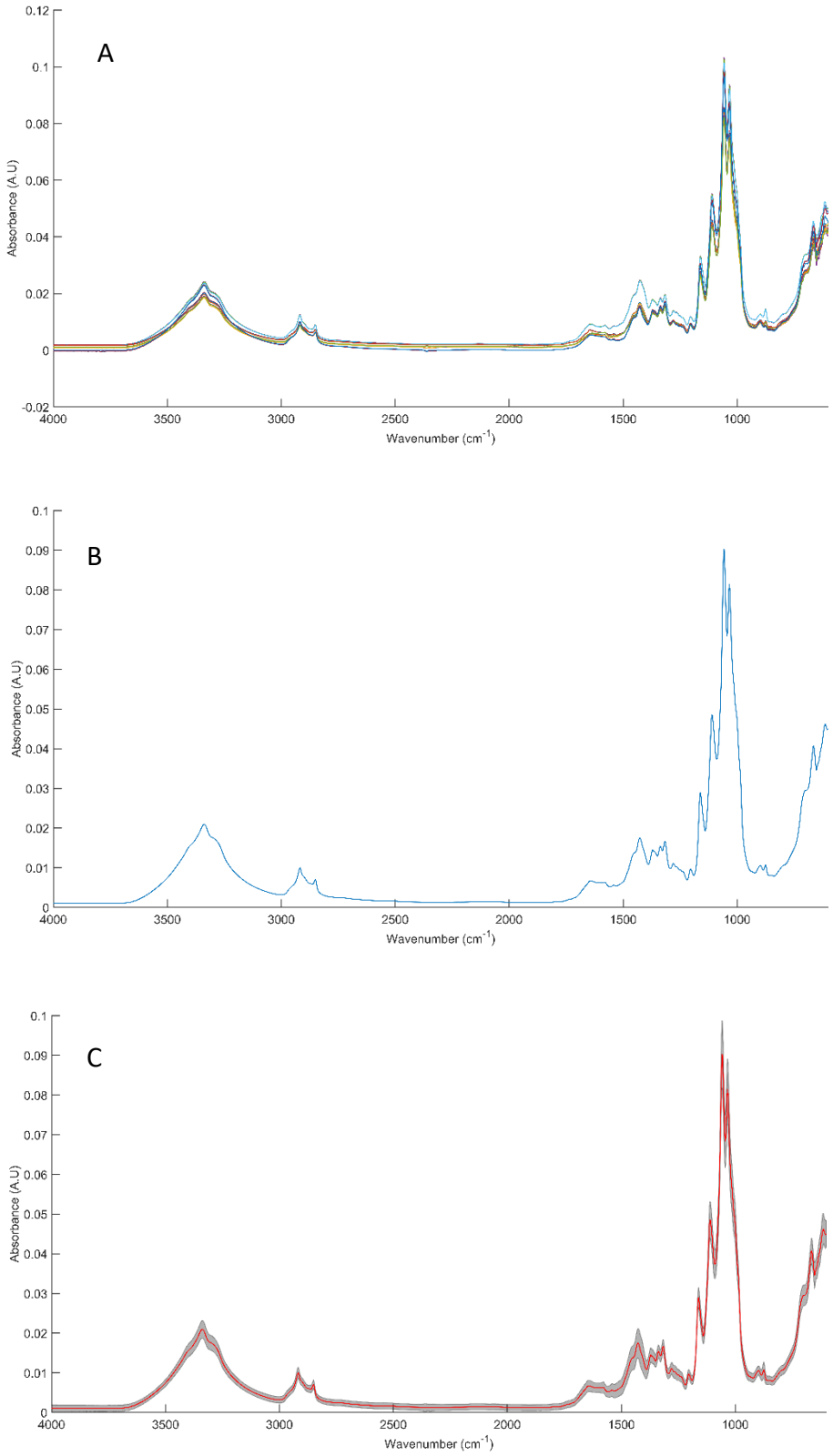

Figure 8: Collection of spectra collected from Jacket acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

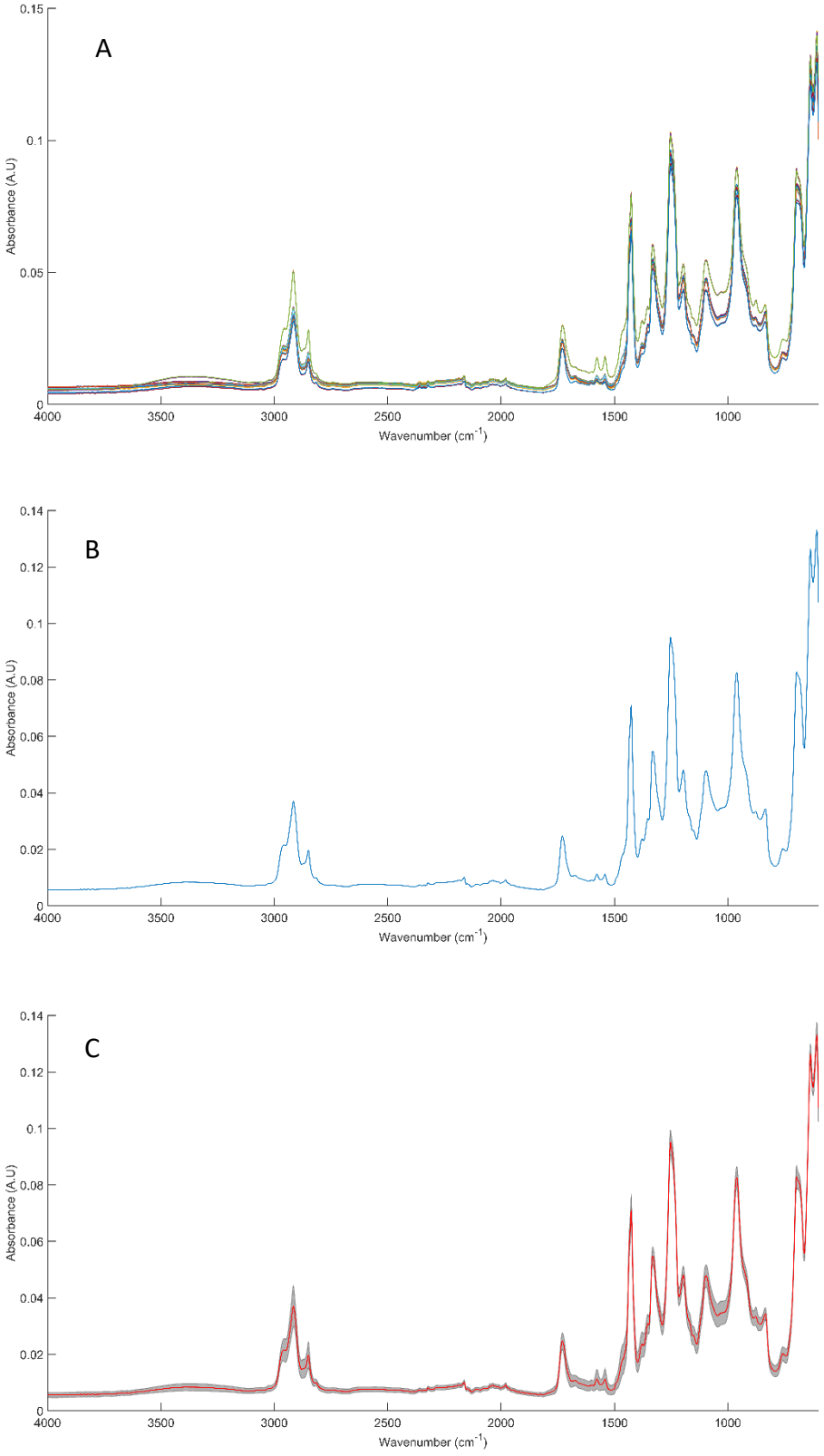

Figure 9: Collection of spectra collected from Lino acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

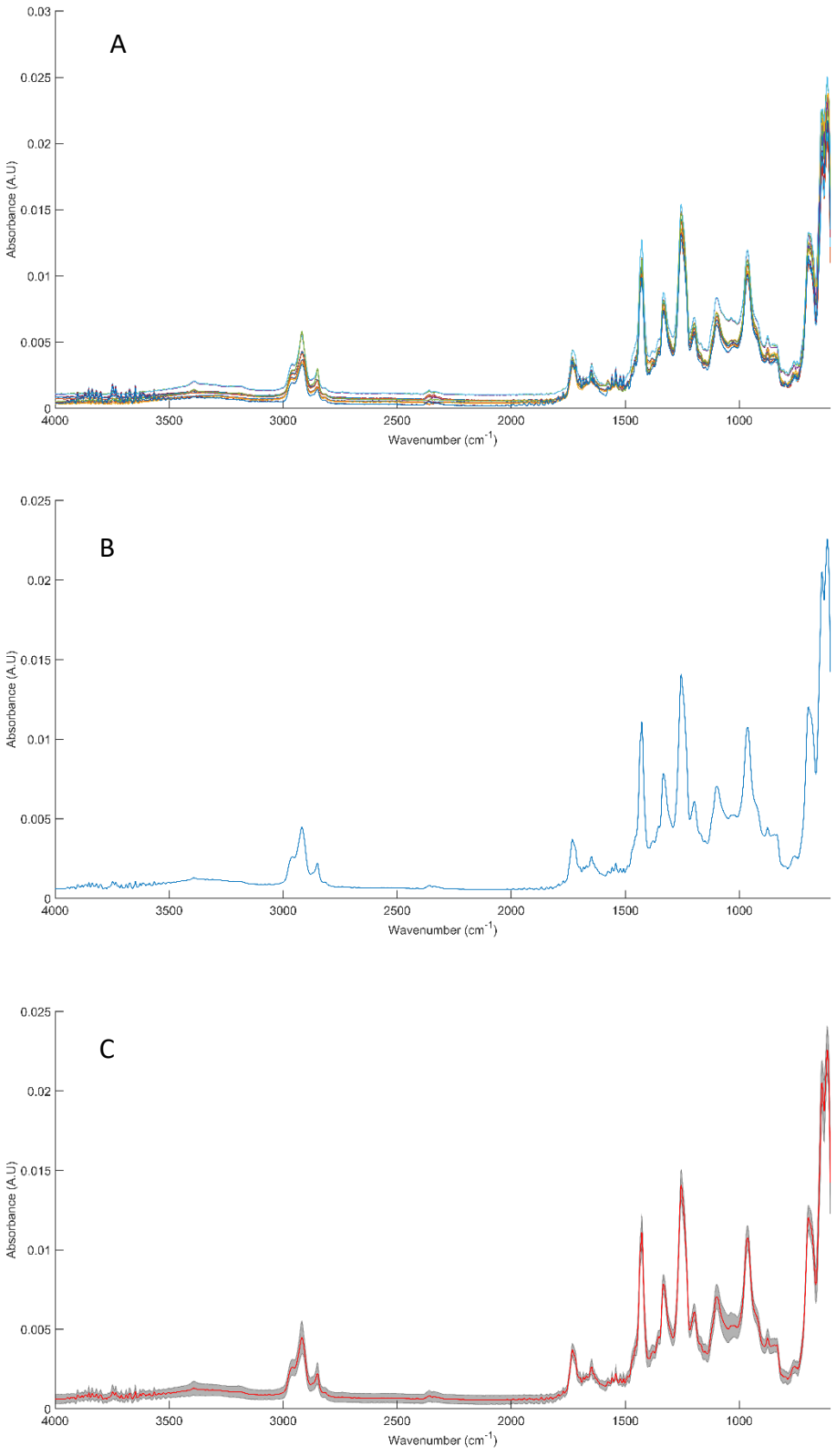

Figure 10: Collection of spectra collected from Lino acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

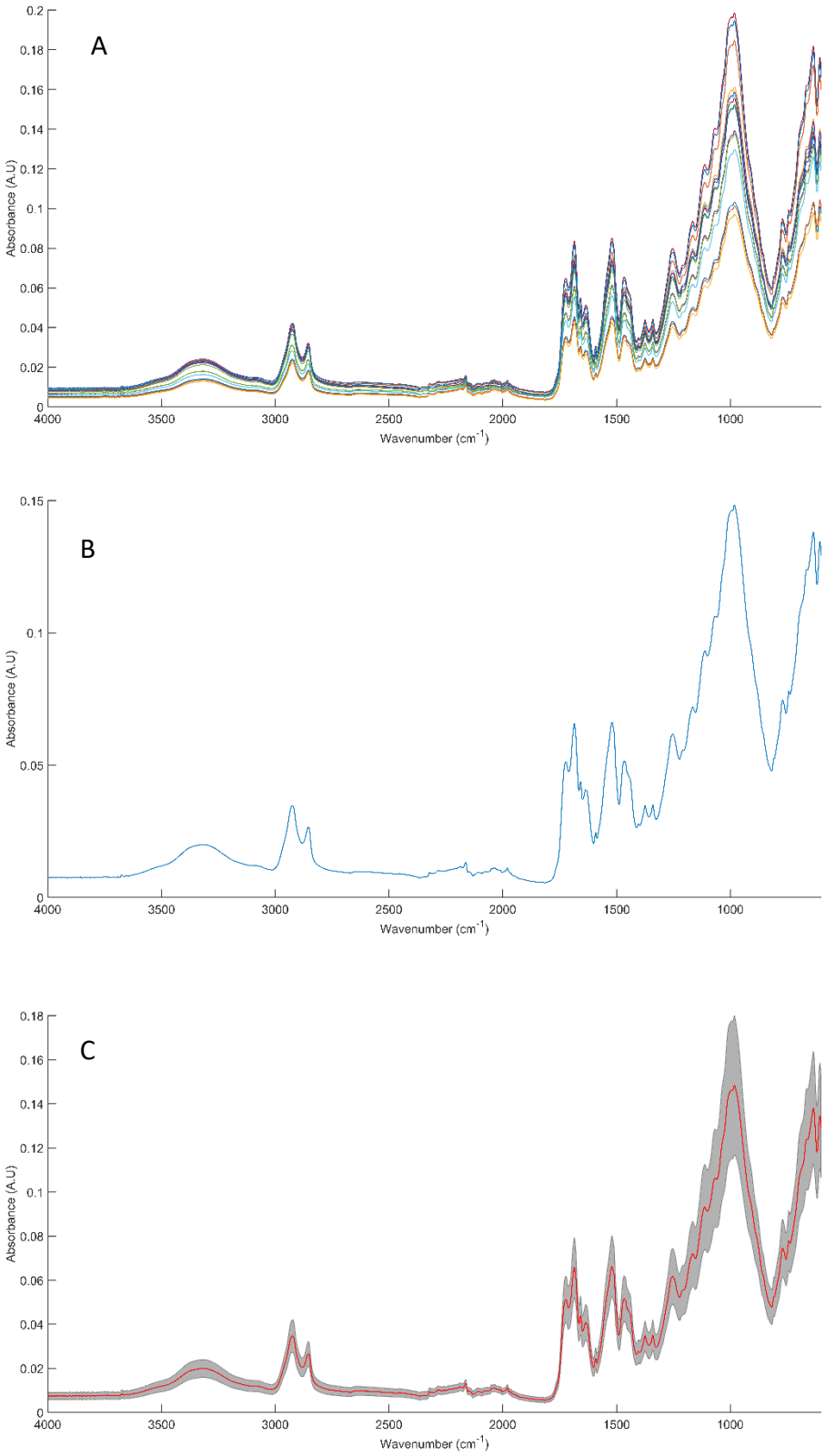

Figure 11: Collection of spectra collected from Metal tile acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

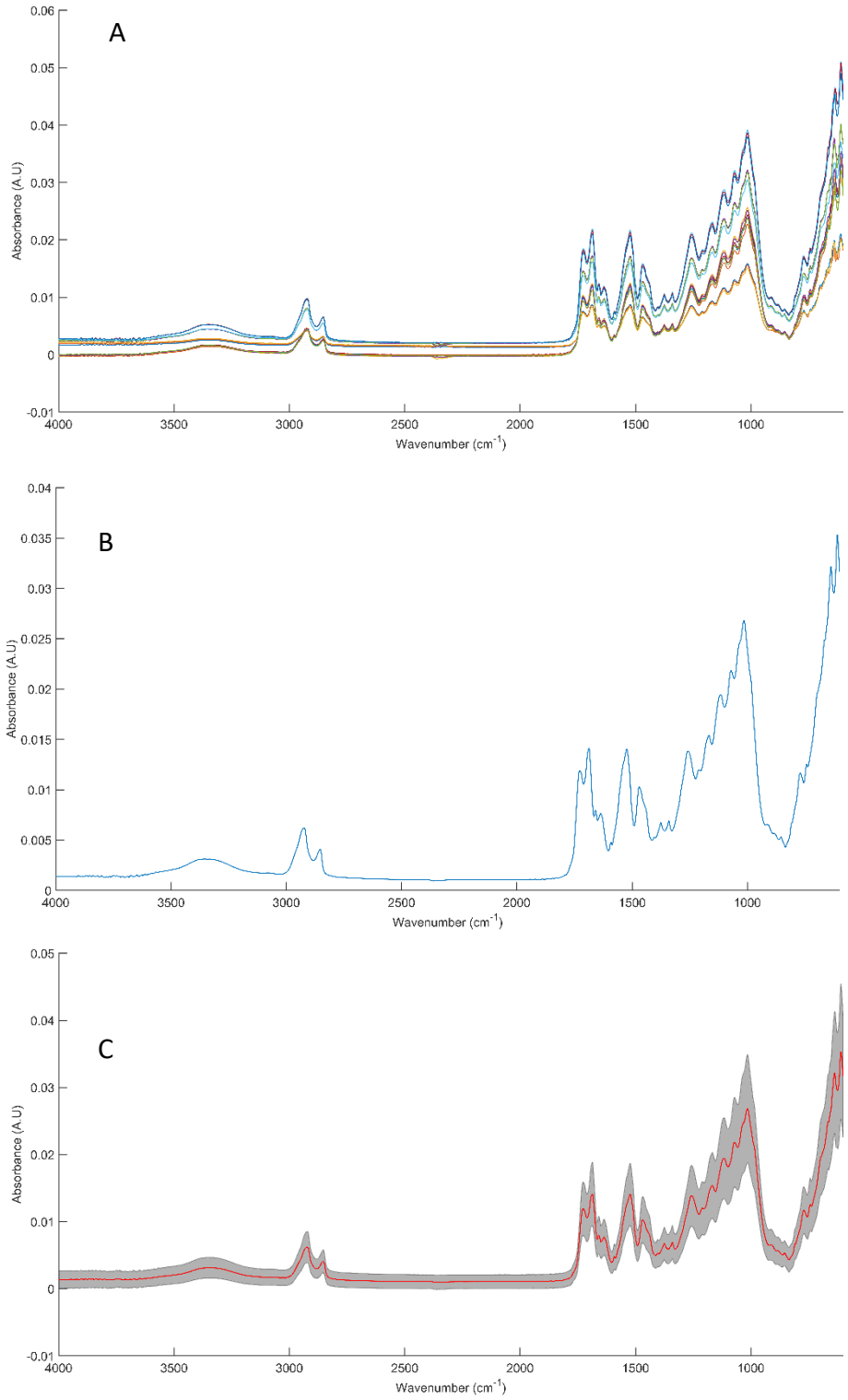

Figure 12: Collection of spectra collected from Metal tile acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

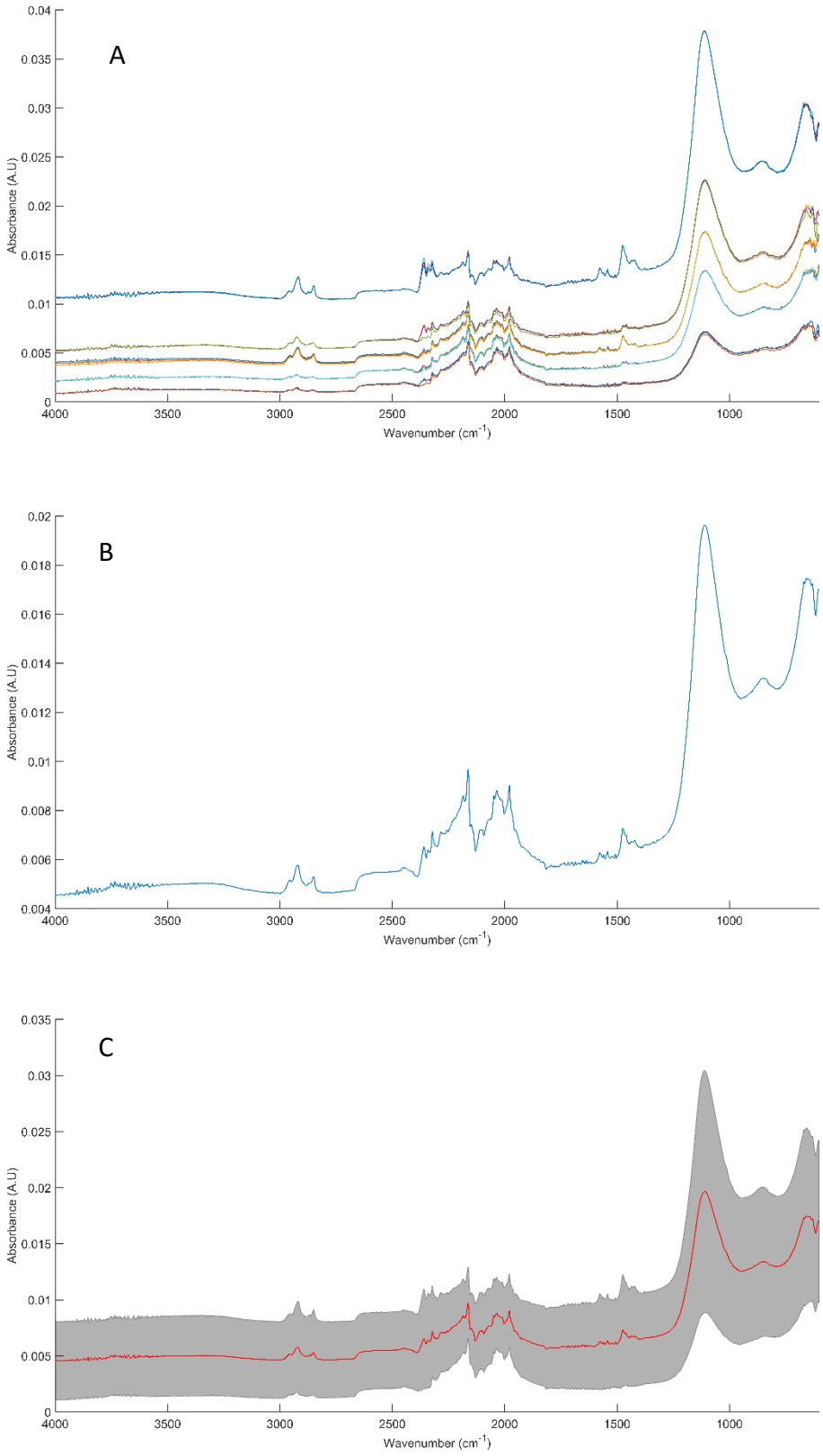

Figure 13: Collection of spectra collected from MirrIR acquired using a DiZnSe IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

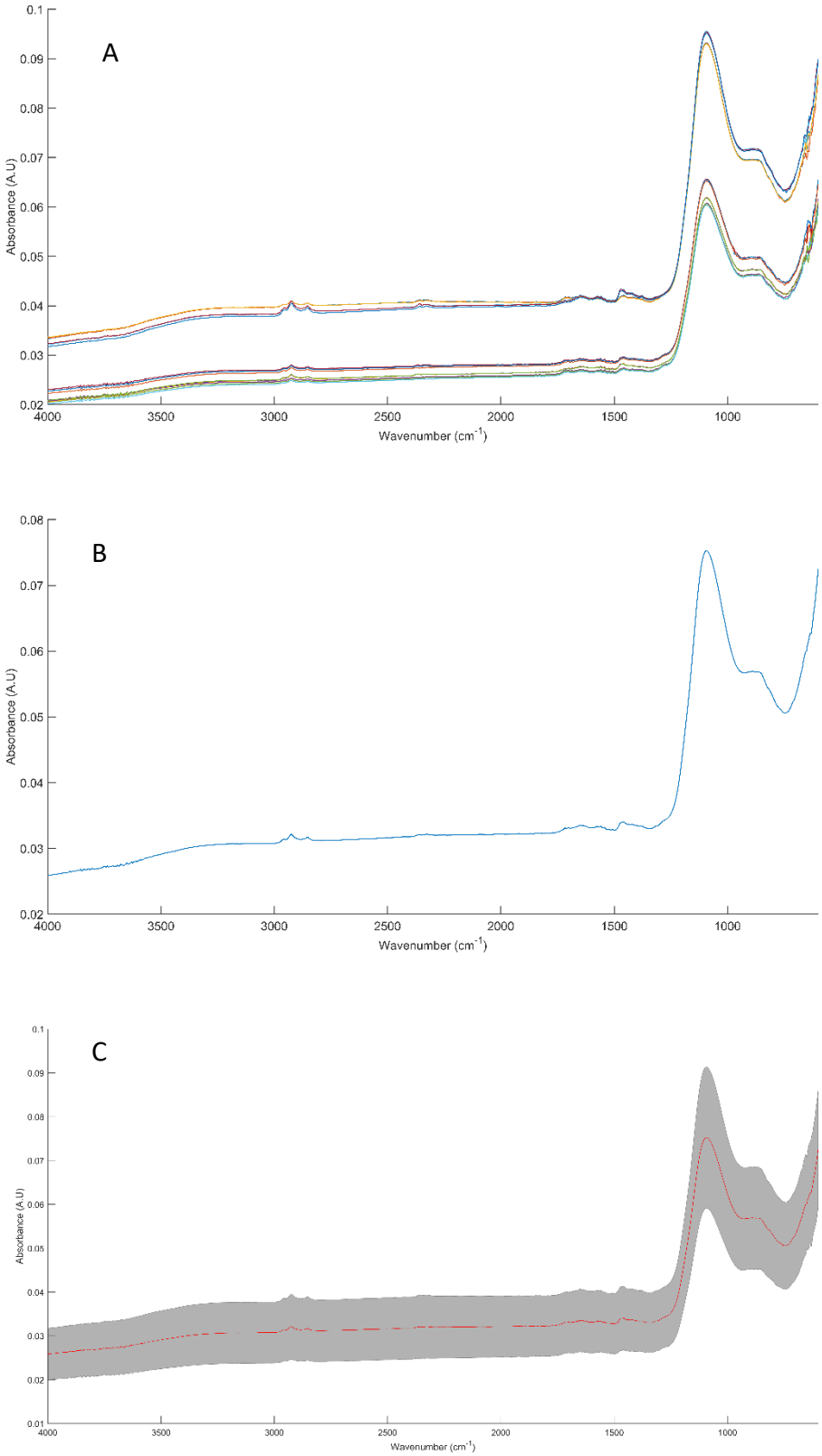

Figure 14: Collection of spectra collected from MirrIR acquired using a Ge IRE starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C).

Table 1: Wavenumbers and corresponding peak assignment for the average spectra collected from each substrate using ATR-FTIR with both Ge and DiZnse IREs [1,2,3,4]

Wavenumber (cm^{-1})							Tentative Peak Assignments
Boot	Concrete	Dirty filter	Jacket	Lino	MirrIR	Metal tile	
648	669, 712, 758		663, 700, 1001, 1030, 1034, 1279, 1280	696, 697, 835, 960, 962,	658	982, 1014, 1066, 1070,	vC-S aliphatic, vCC alicyclic, aliphatic chains
		808, 972					vC-O-C, vCC alicyclic, aliphatic chains
	868, 872	841, 899	876, 899		849,864		vC-O-C, vO-O, vCC alicyclic, aliphatic chains
		997	1105, 1111		1093, 1111	1111, 1115	vCC alicyclic, aliphatic chains
1032, 1034, 1203, 1236, 1240	1027, 1036	1165, 1167	1159,1161, 1203			1165, 1167	vC=S, vCC alicyclic, aliphatic chains
			1053, 1057				vC-O-C, vCC alicyclic, aliphatic chains
1082, 1084	1082	1103		1095, 1099			vC-S aromatic, vC-O-C asymmetric
				1196, 1198			vC=S
1144, 1146							vC-O-C asymmetric
		1255		1253		1253, 1257	Amide III
1336, 1338		1358, 1359	1315, 1317, 1334, 1336, 1361,1371	1331, 1333			vC-(NO_2)
1377	1388	1375, 1377				1375	δCH_3

	1400		1427	1425, 1427	1473	1466, 1469	δCH_2, δCH_3 asymmetric
1448, 1452		1454, 1456					vCC aromatic ring chain vibrations
						1520, 1523	vC=C, Amide II
1543, 1549			1579				vC-(NO_2) asymmetric, Amide II
1637, 1649		1647	1637, 1645	1647		1635, 1658, 1685, 1687	vC=N, Amide I
1732				1730, 1732		1722, 1727	vC=C, vC=O
2852, 2922, 2924, 2954, 2956		2837, 2839, 2868, 2875, 2917, 2920, 2951	2850, 2916, 2918	2850, 2916, 2918, 2958, 2960	2850, 2854, 2922, 2925, 2956, 2957	2852, 2854, 2922, 2925	vC-H
3082, 3093, 3294,			3294,				vO-H
3316		3371, 3387	3333, 3338			3316, 3355	vO-H, vN-H

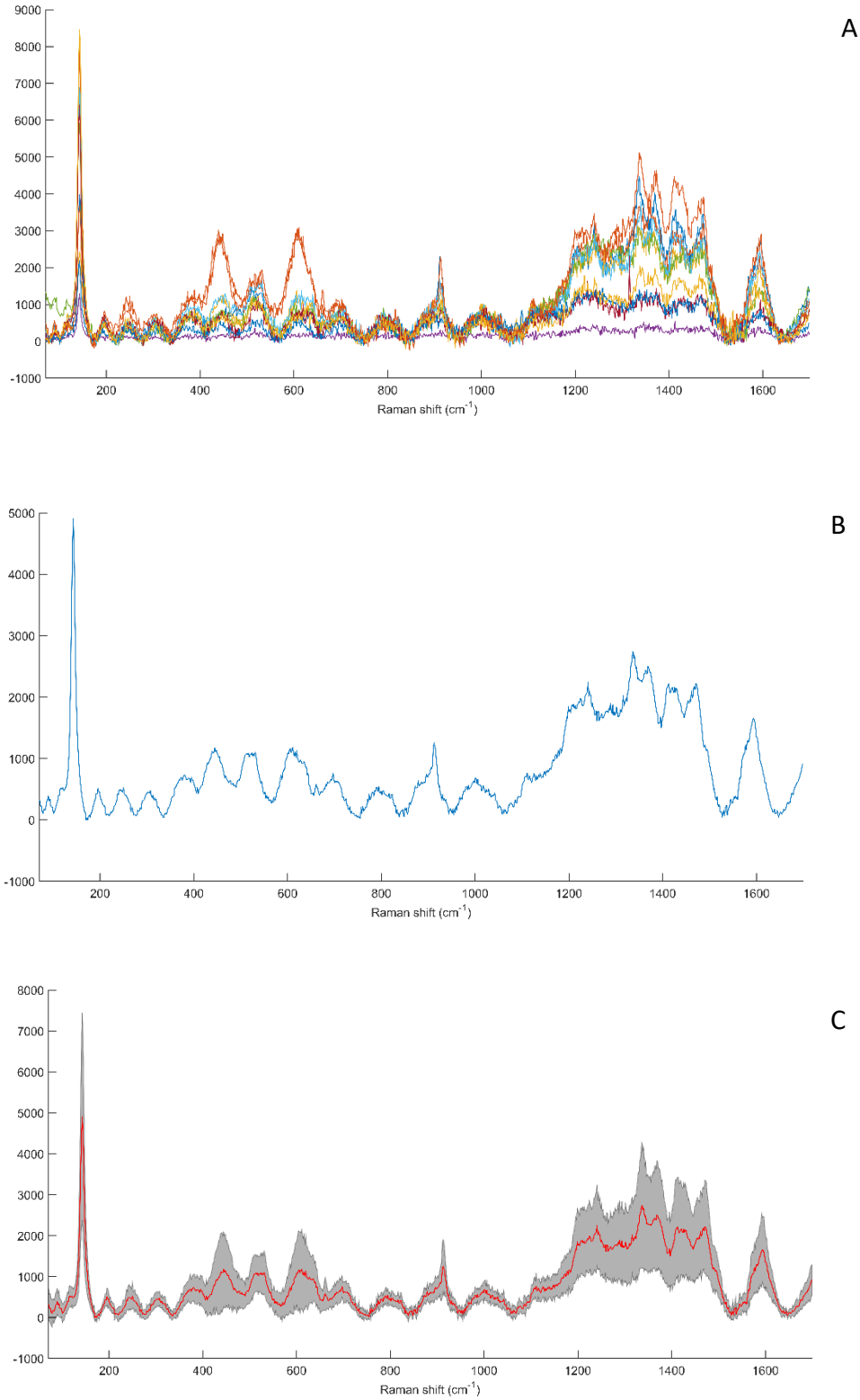

Figure 15: Collection of spectra collected from Boot acquired using a 532nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

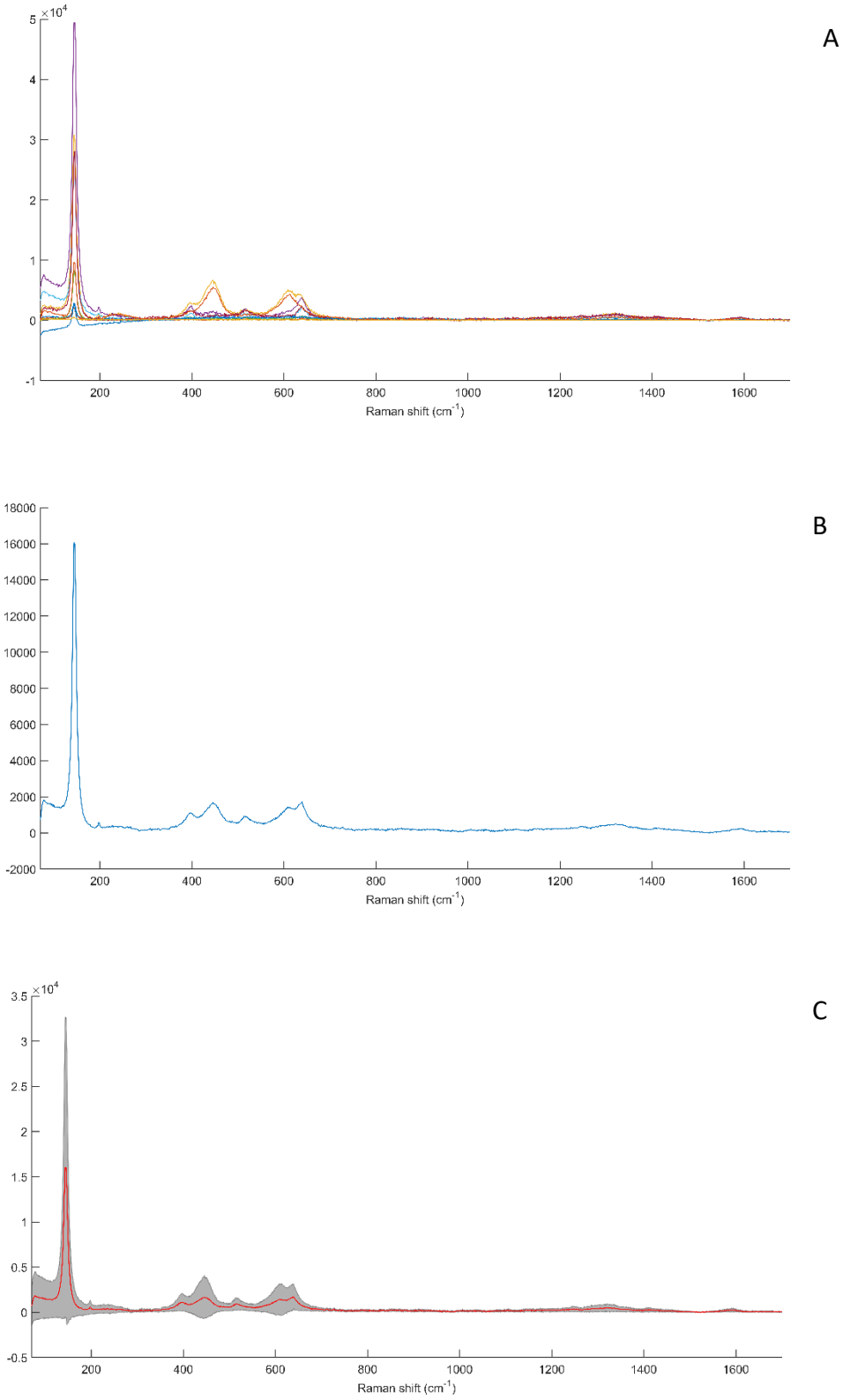

Figure 16: Collection of spectra collected from Boot acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 25% filter.

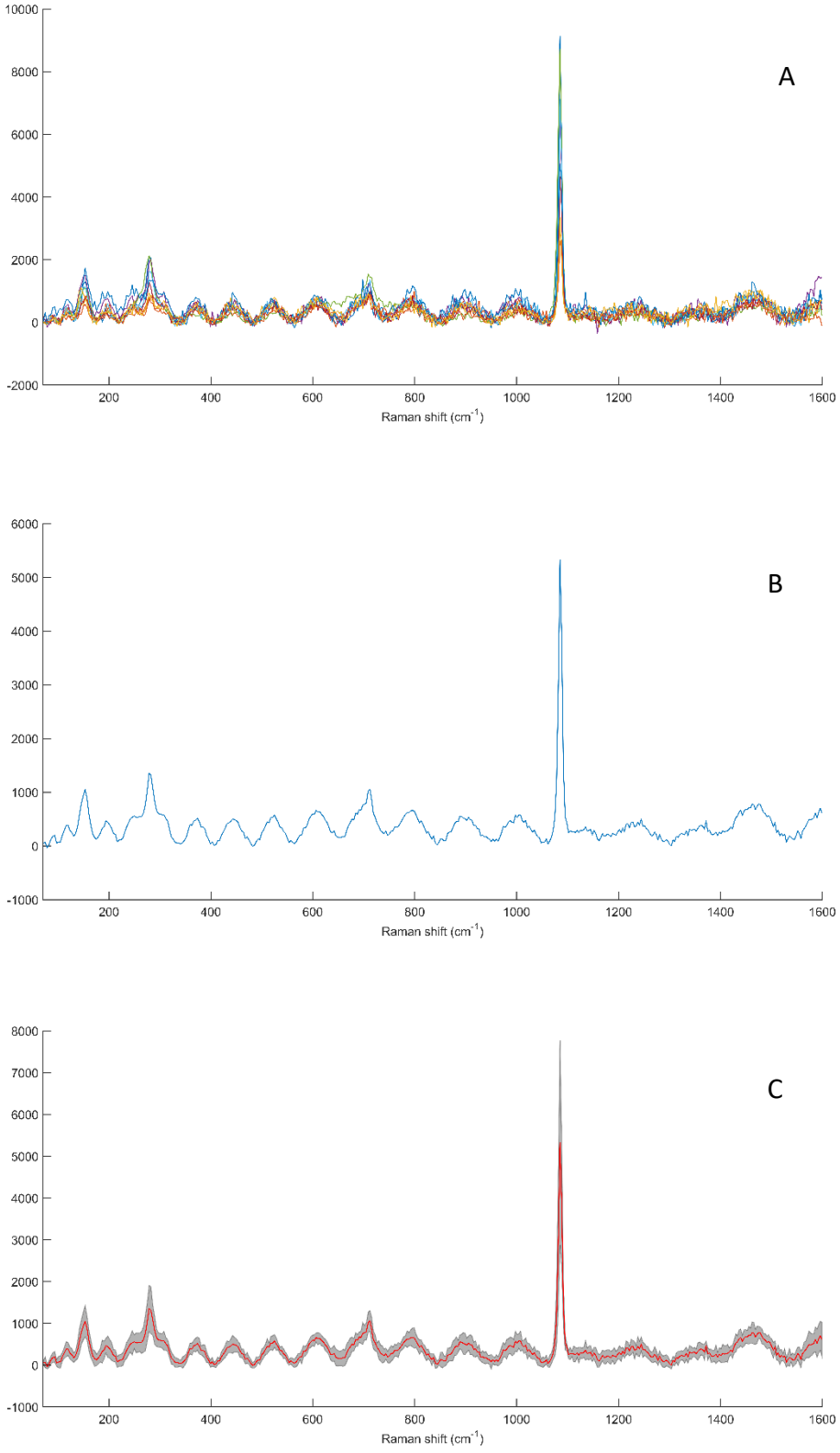

Figure 17: Collection of spectra collected from Concrete acquired using a 532nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 1% filter.

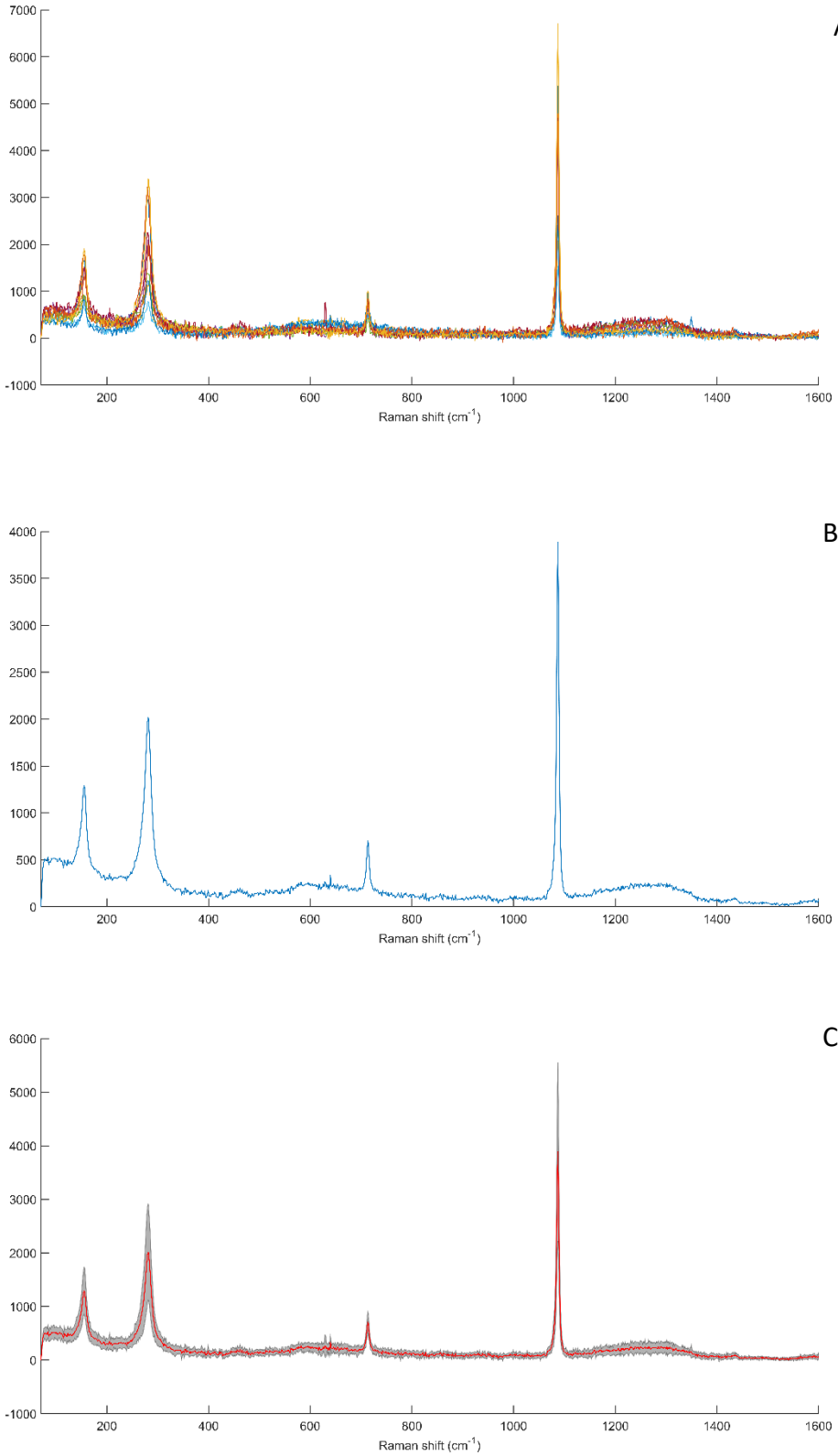

Figure 18: Collection of spectra collected from Concrete acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

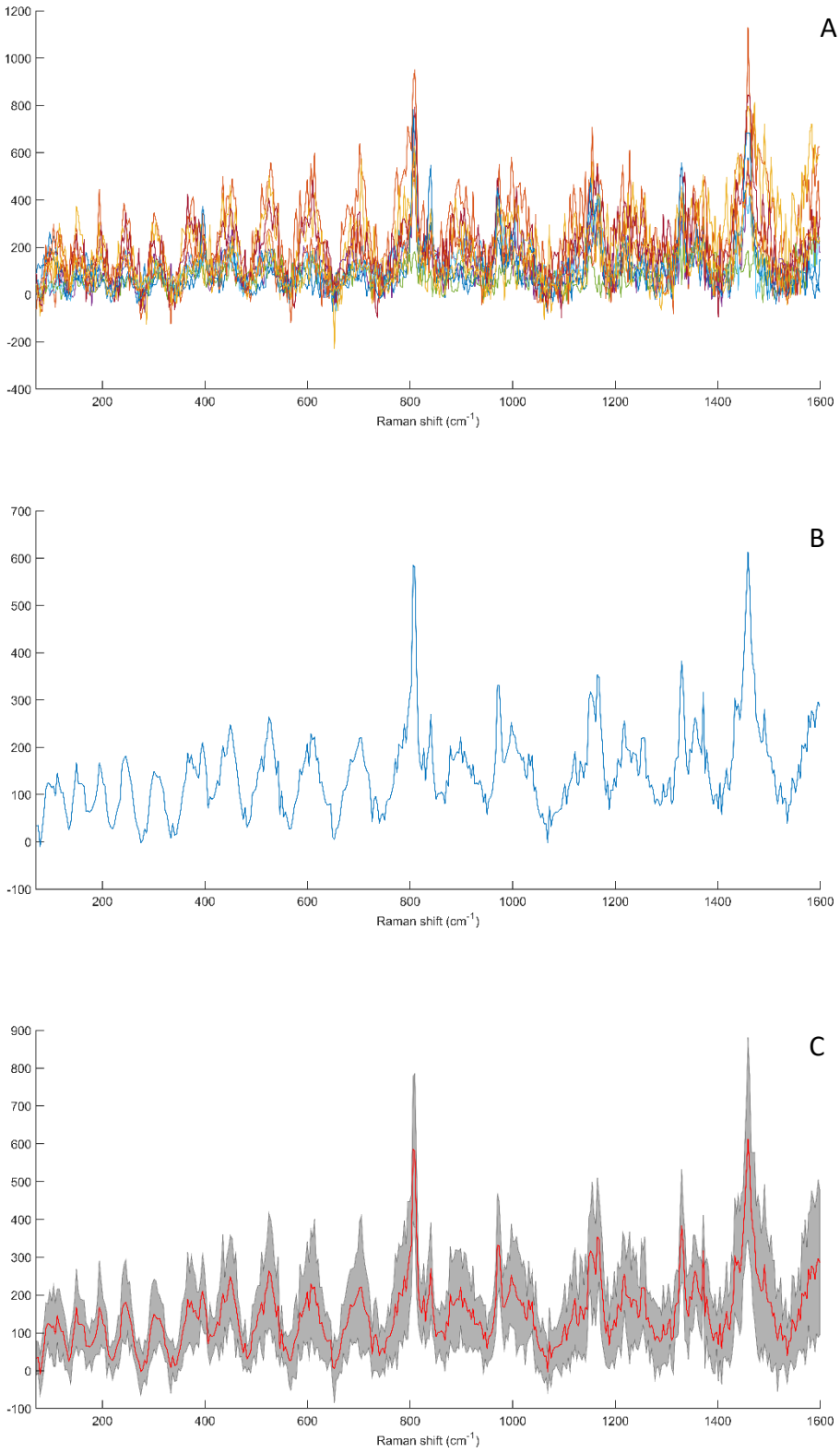

Figure 19: Collection of spectra collected from Dirty filter acquired using a 532nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 1% filter.

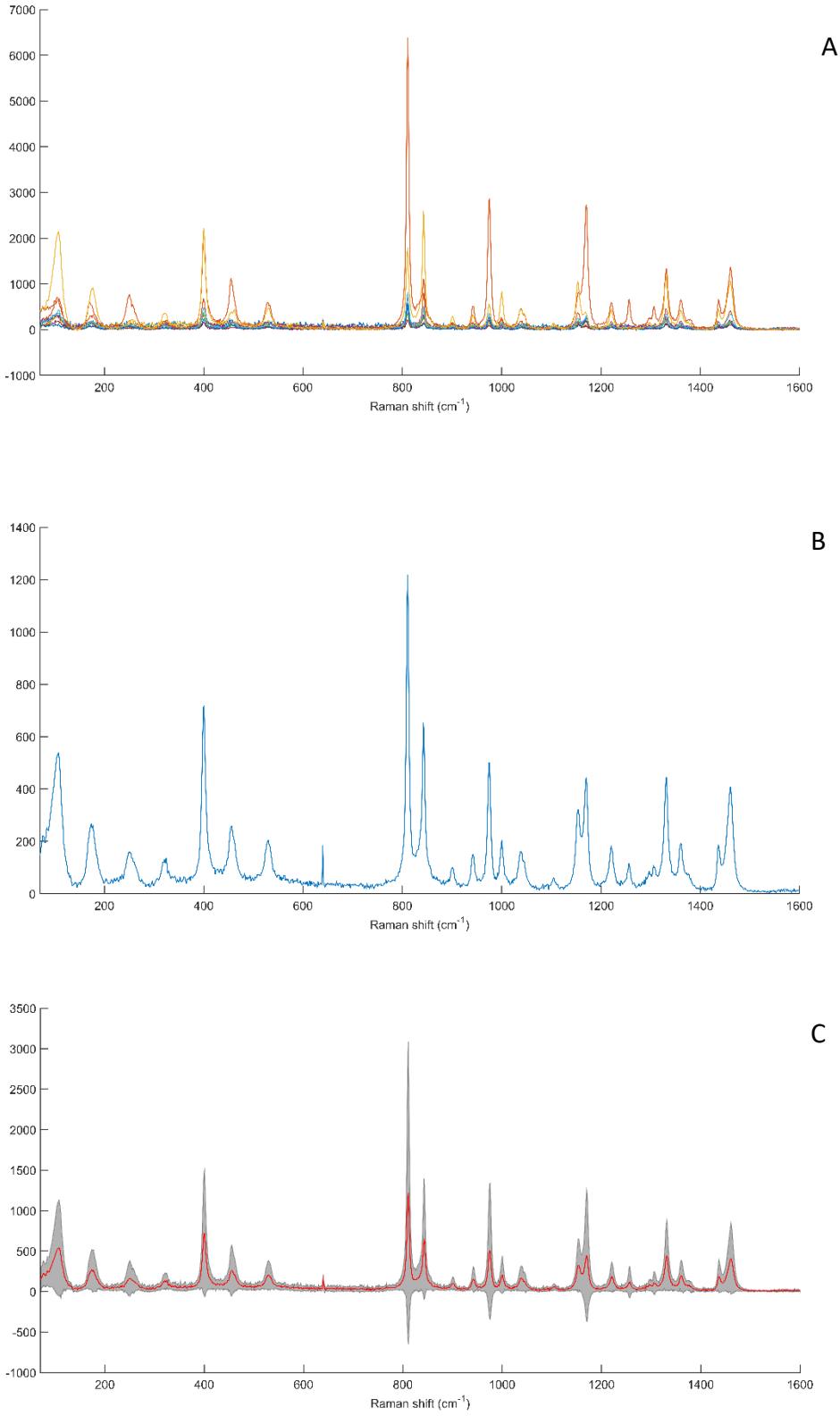

Figure 20: Collection of spectra collected from Dirty filter acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

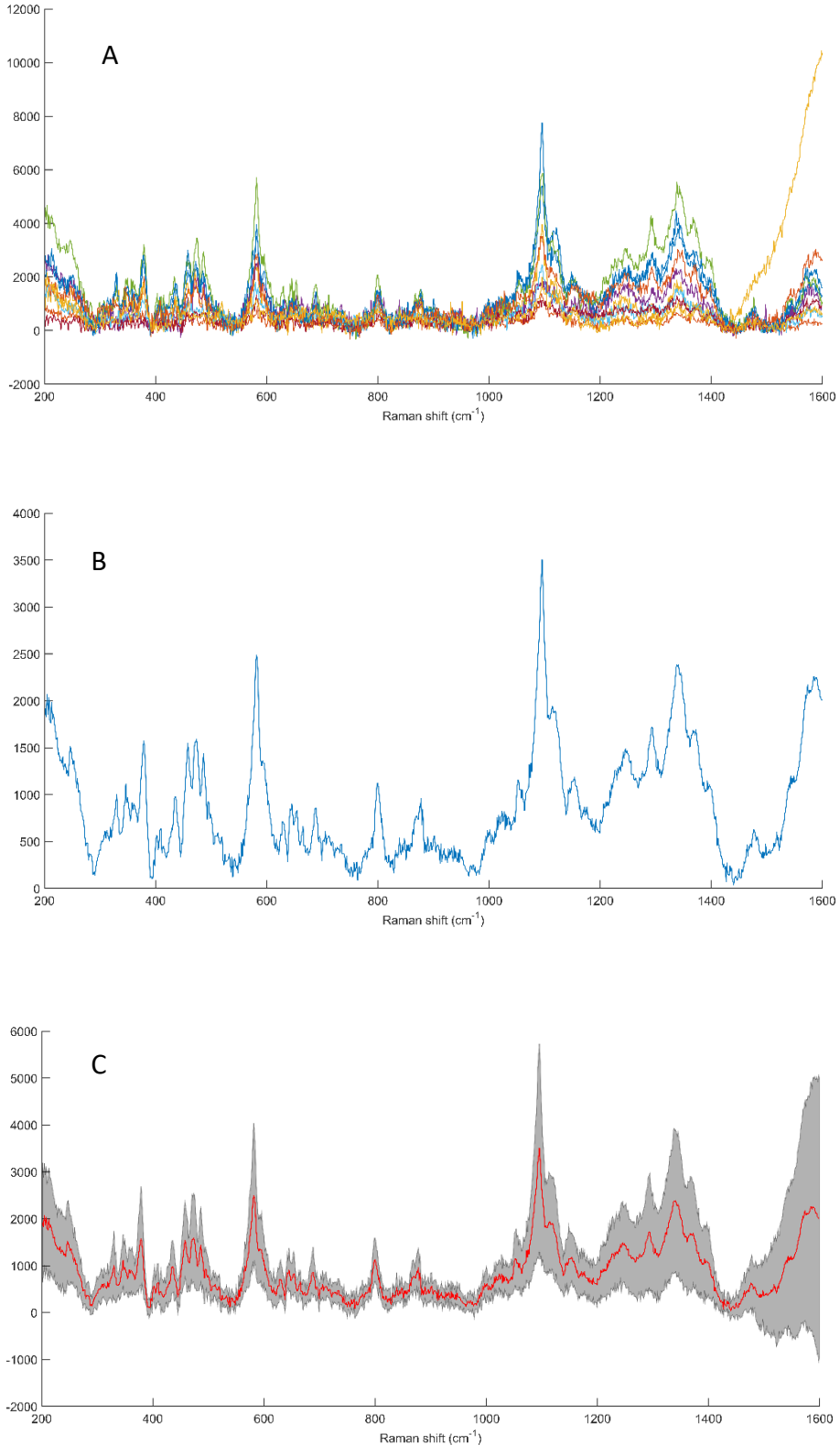

Figure 21: Collection of spectra collected from Jacket acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

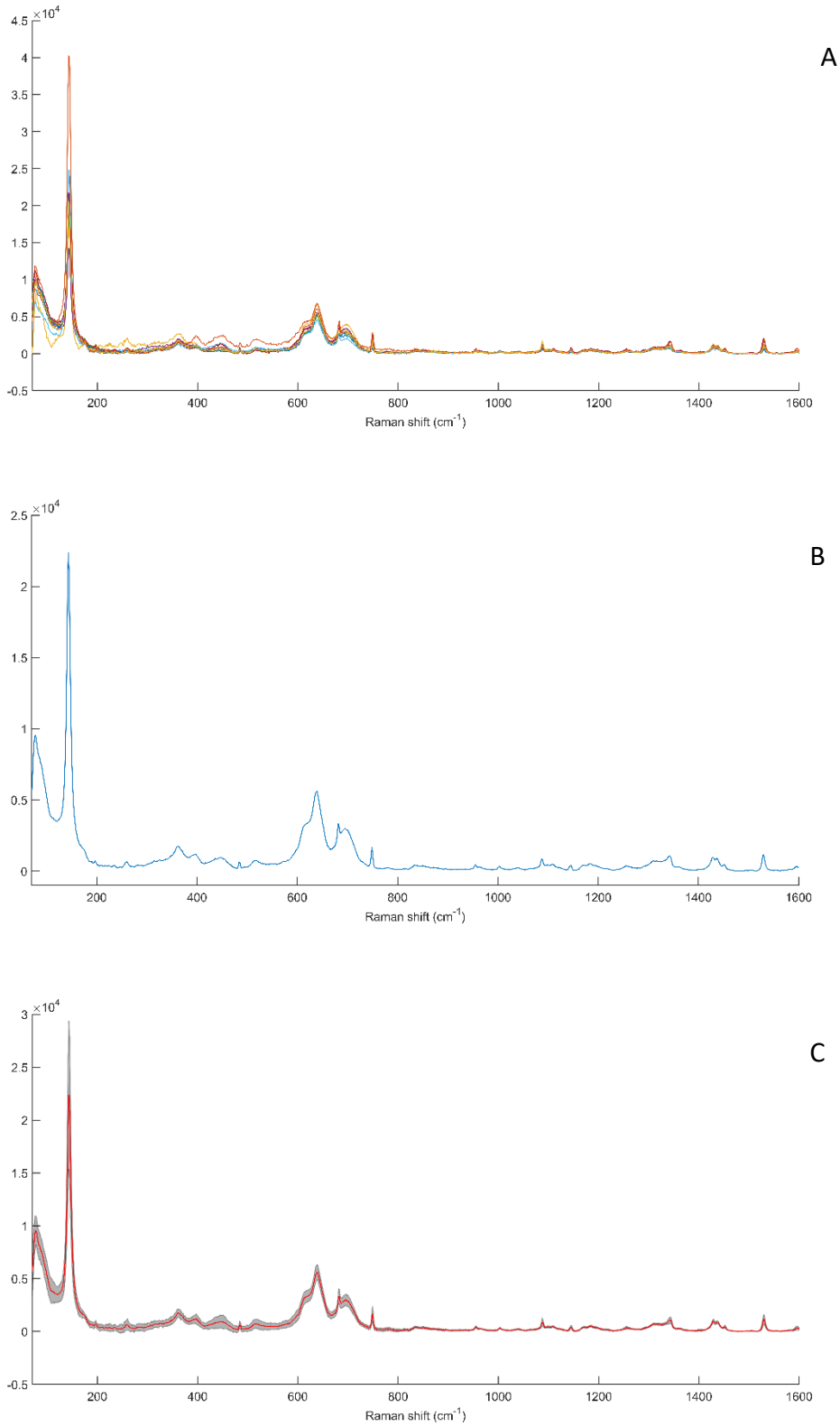

Figure 22: Collection of spectra collected from Lino acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

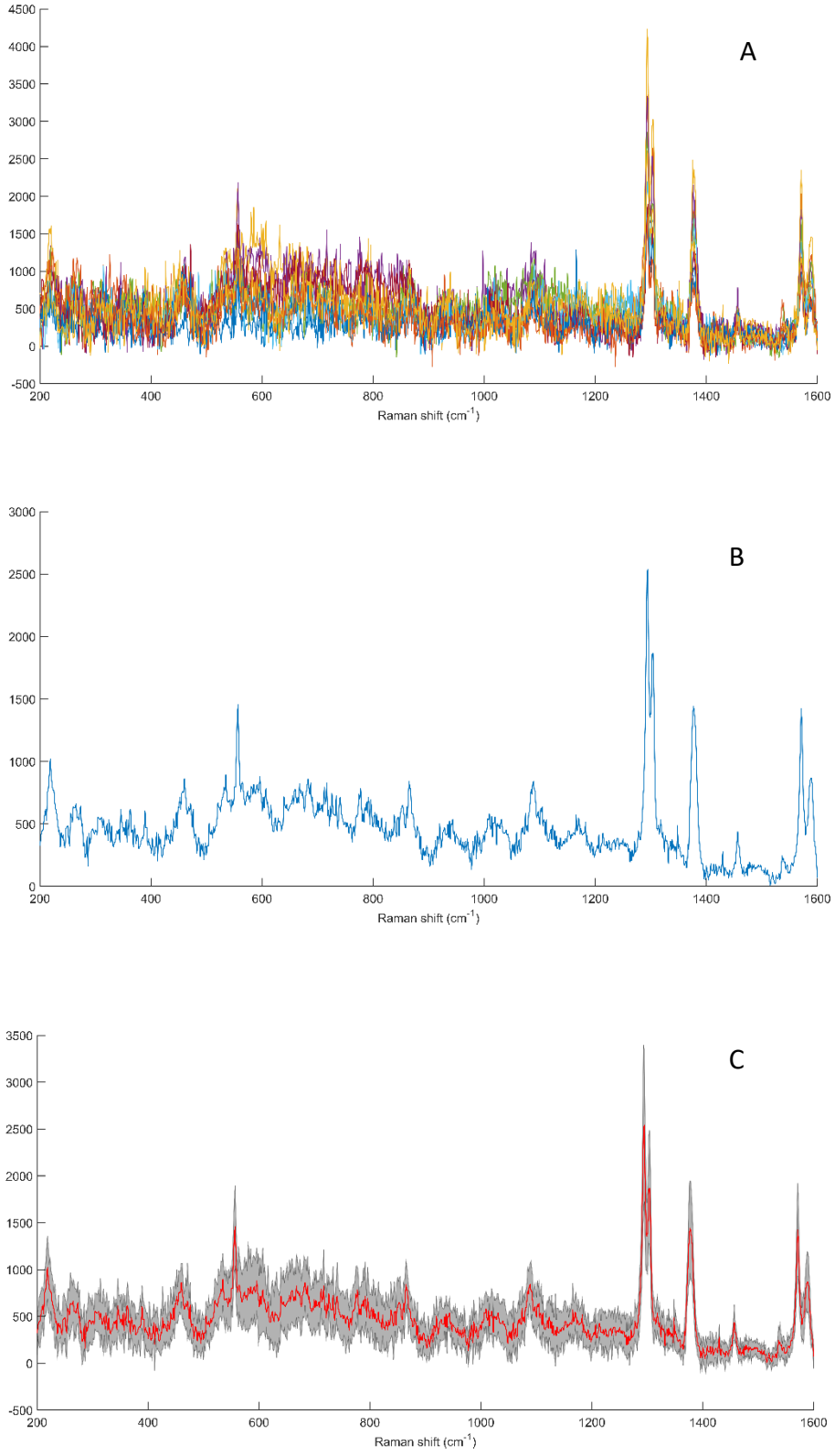

Figure 23: Collection of spectra collected from Metal tile acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 1% filter.

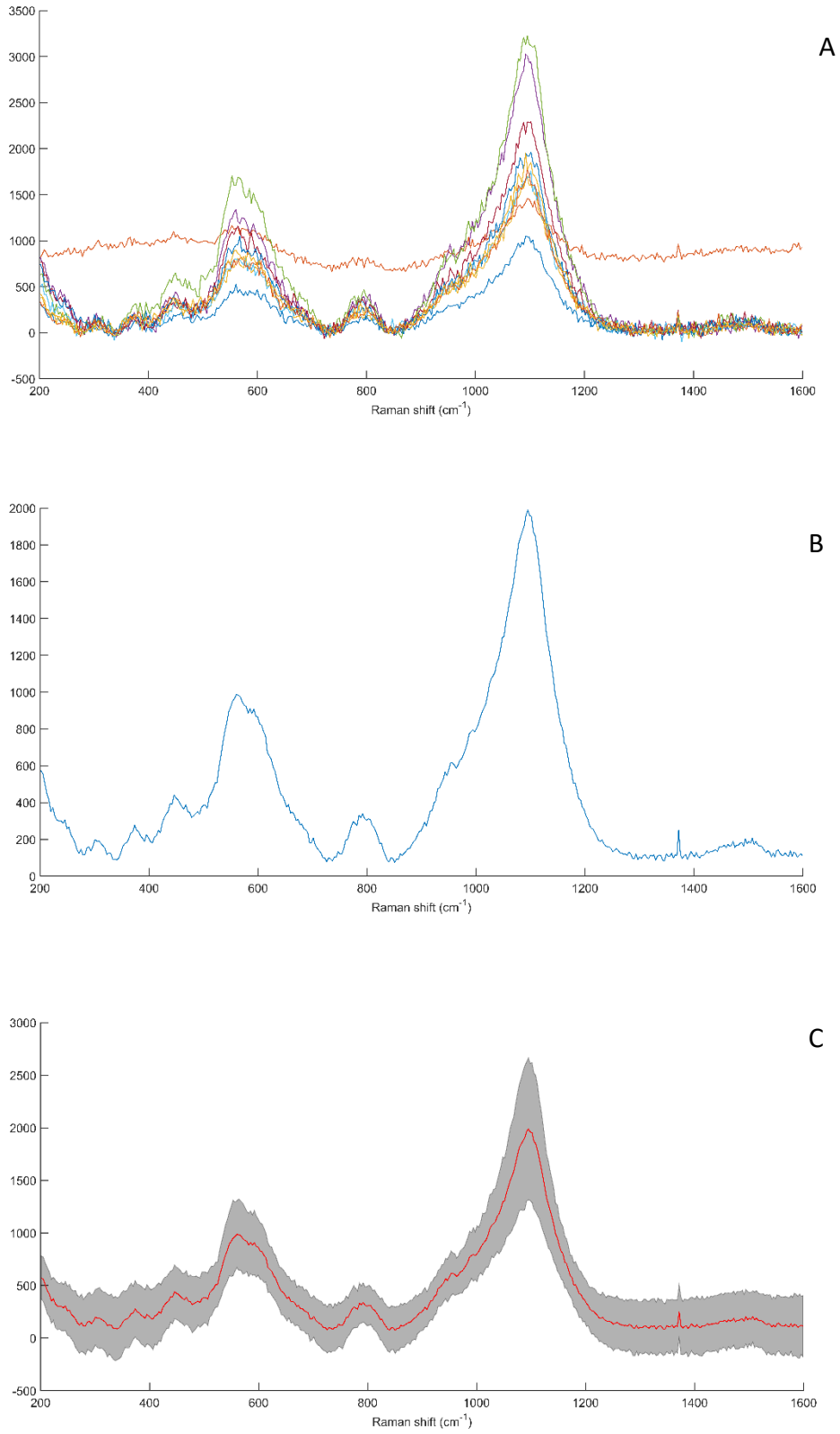

Figure 24: Collection of spectra collected from MirrIR slide acquired using a 532nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

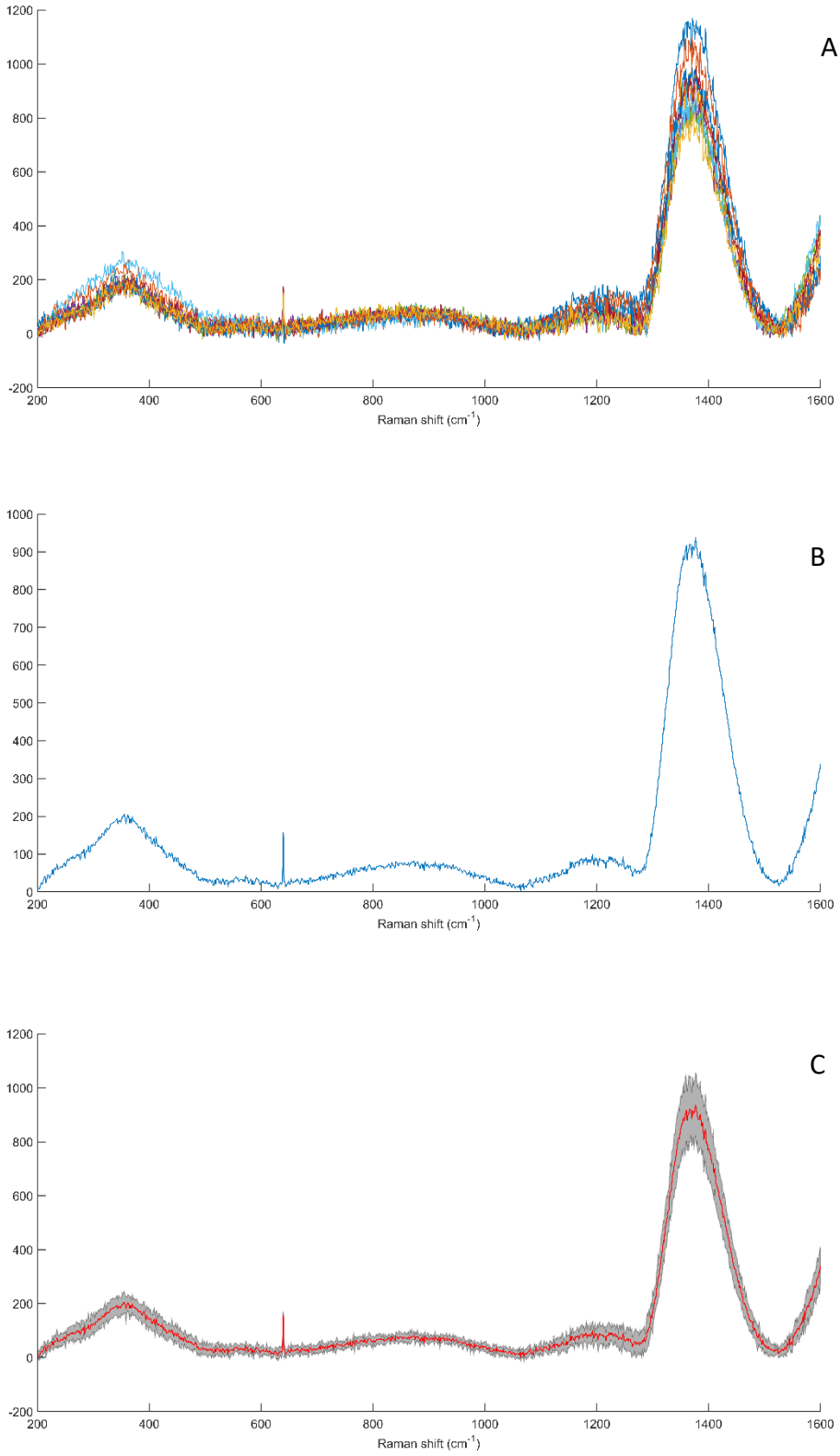

Figure 25: Collection of spectra collected from MirrIR acquired using a 785nm laser starting with an overlay of 10 spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 20 s acquisition time, 2 accumulations, 10% filter.

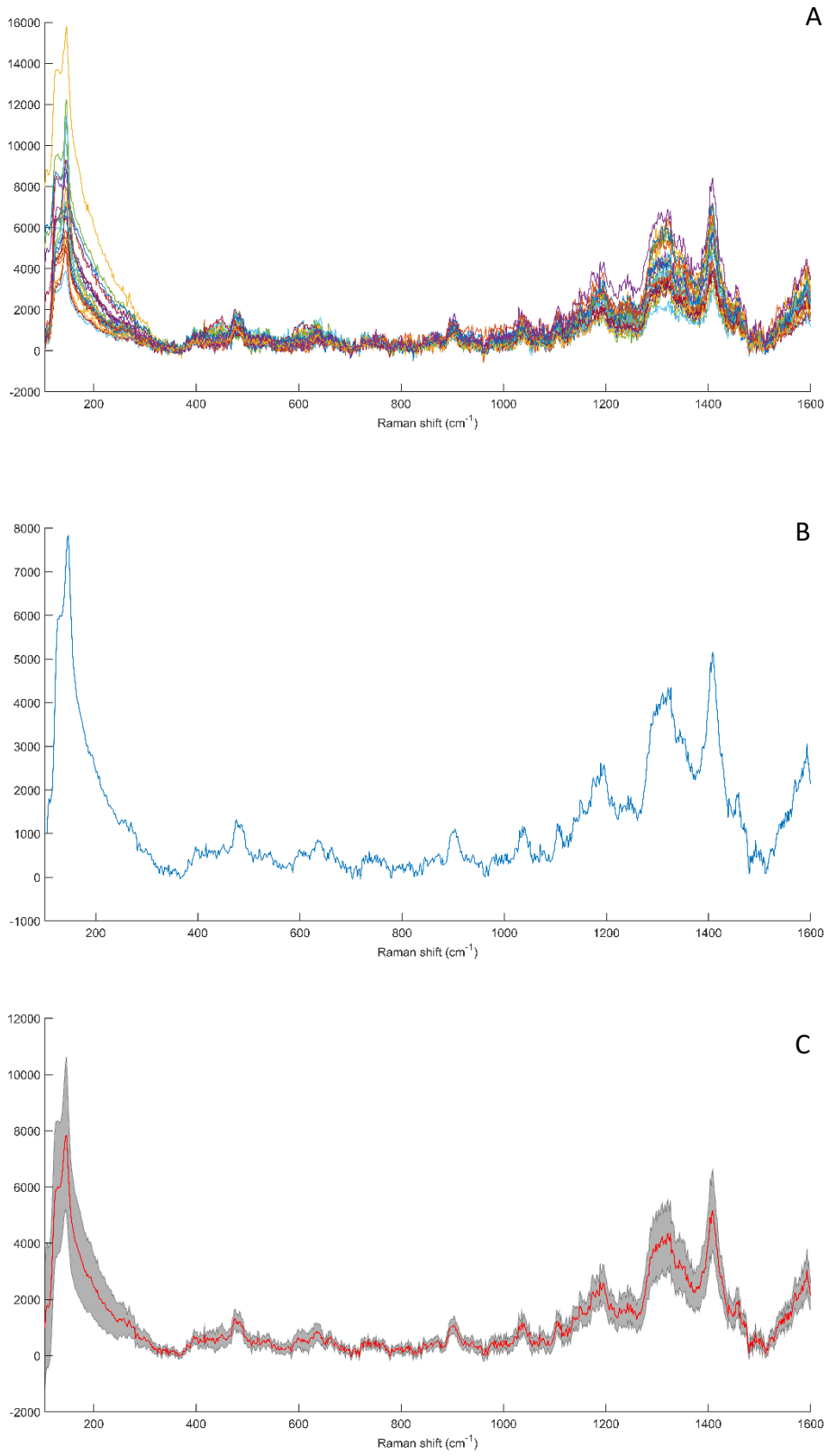

Figure 26: Collection of spectra collected from Boot acquired using a 633nm laser starting with an overlay of spectra (A)a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 15 s acquisition time, 5 accumulations 10% filter.

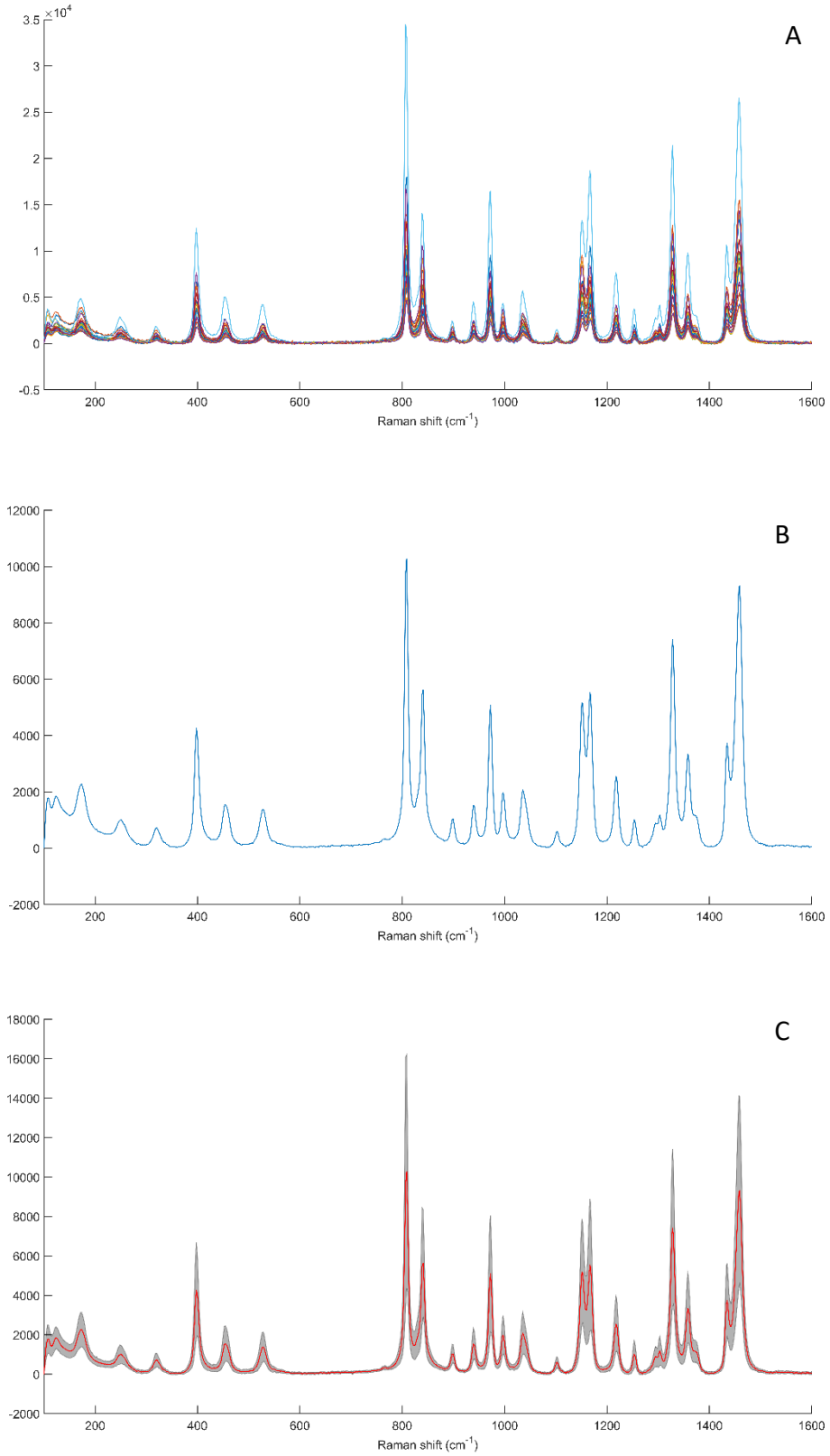

Figure 27: Collection of spectra collected from Clean filter acquired using a 633nm laser starting with an overlay of spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 5 s acquisition time, 5 accumulations 100% filter.

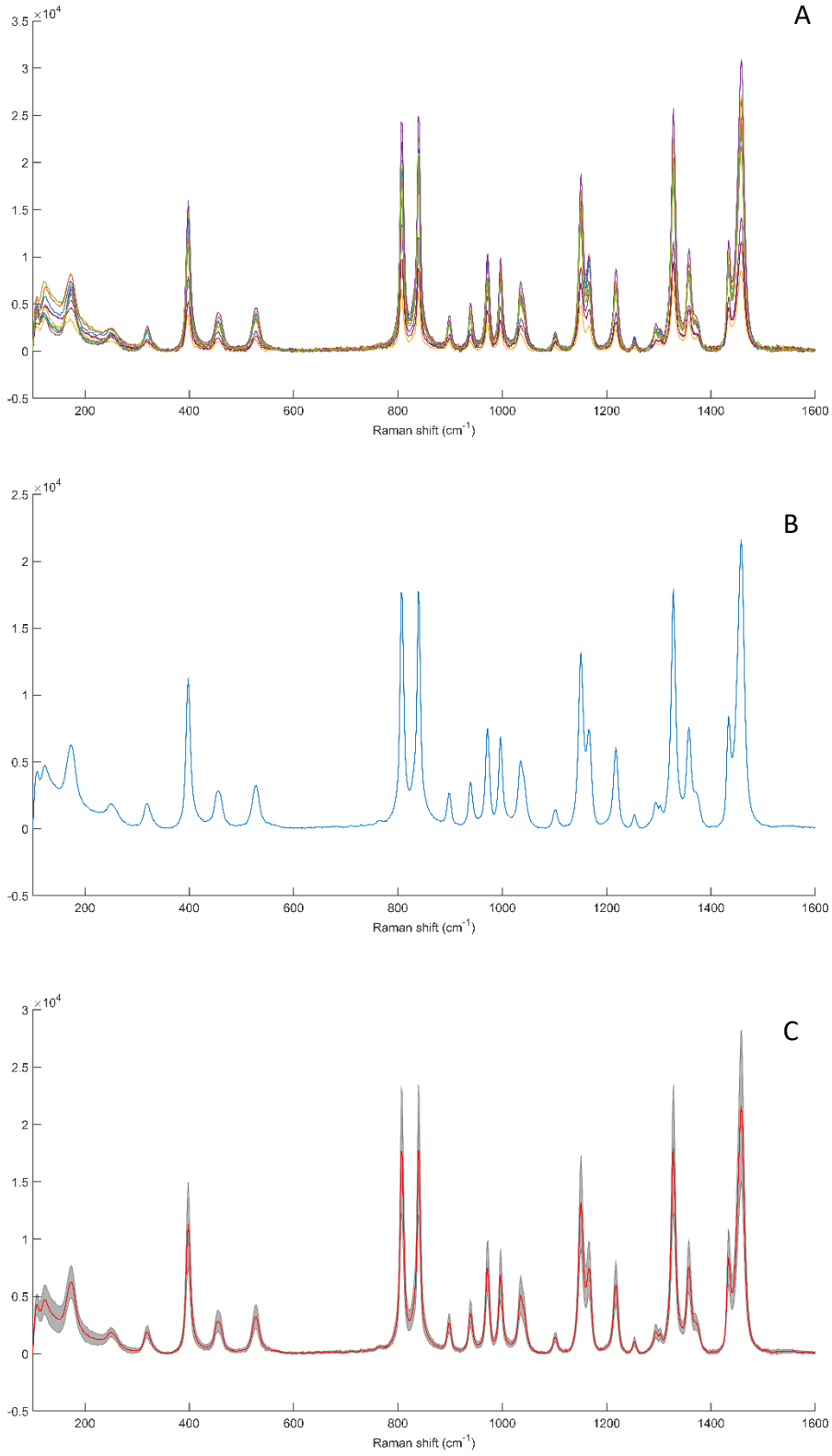

Figure 28: Collection of spectra collected from Dirty filter acquired using a 633nm laser starting with an overlay of spectra (A) a mean spectrum (B) and a mean spectrum shown in red surrounded by a grey standard deviation cloud (C). Parameters: 5 s acquisition time, 5 accumulations 100% filter.

Table 2 Wavenumbers and corresponding peak assignment for the Clean filter spectra collected using a 633nm laser [1,2,3,4]

Wavenumber cm^{-1}	Tentative peak assignment
633nm	
109, 123, 174	Lattice vibrations
249, 319, 398	δCC aliphatic chains
454	vS-S, vSi-O-Si
528	vS-S, vSi-O-Si, vC-I, vC-Br, vC-Cl
809, 898, 939, 973	vC-O-C, vCC alicyclic, aliphatic chains
839	vC-O-C, vO-O, vCC alicyclic, aliphatic chains
997, 1036	vCC alicyclic, aliphatic chains
1102, 1151	vC-O-C asymmetric
1167, 1303	vCC alicyclic, aliphatic chains
1218, 1253, 1294	Amide II
1328, 1358, 1372	$vC\text{-}NO_2$
1435, 1458	$\delta CH_2/\delta CH_3$ asymmetric

Table 3 Wavenumbers and corresponding peak assignment for the average spectra collected from each substrate using Raman with the 532, 633 and/or 785nm laser (where applicable) [1,2,3,4]

Wavenumber (cm^{-1})							Tentative Peak Assignments
Boot	Concrete	Dirty filter	Jacket	Lino	MirrIR	Metal tile	
130, 143, 147, 197	153, 155	108, 109, 123, 174		77, 143			Lattice vibrations
						219	vXmetal-O
397	279, 282	249, 250, 319, 324, 398, 401	379	260, 362, 398	301, 355, 374		δCC aliphatic chains
446, 476, 484, 513		454, 455	435, 458, 474, 487	484	446	460, 538, 557	vS-S, vSi-O-Si
		528, 530					vS-S, vSi-O-Si, vC-I, vC-Br, vC-Cl
			581		560, 592		vC-I, vC-Br, vC-Cl
				638, 681, 695, 749	792		vC-S aliphatic
599, 609, 635, 640, 662	712, 714	640		618			vCC alicyclic, aliphatic chains
903, 1036		806, 807, 811, 839, 898, 901, 939, 942, 970, 973, 974, 997, 998, 1001, 1035, 1036, 1038, 1044		955, 1004		865	vC-O-C, vCC alicyclic, aliphatic chains
		840, 843					vC-O-C, vO-O, vCC alicyclic, aliphatic chains
912			800, 887				vC-O-C

1104, 1149	1085, 1087	1102, 1105, 1151, 1152, 1154	1052, 1096, 1114, 1154	1087, 1146	1095	1089	vC-O-C asymmetric
1196		1165, 1167, 1170, 1303, 1306		1310		1303	vCC alicyclic, aliphatic chains
1241, 1243		1218, 1221, 1253, 1256, 1294	1246, 1293	1256		1294	Amide III
1321, 1337, 1369		1328, 1330, 1332, 1356, 1358, 1361, 1369, 1372	1338, 1373	1343	1371	1377	vC-NO_2
1396, 1409, 1413, 1459, 1471		1433, 1437, 1458, 1459, 1461, 1472	1397, 1477	1428, 1437, 1451		1457	$\delta CH_2/\delta CH_3$ asymmetric
		1491					vCC aromatic ring vibrations
			1543	1530		1537	Amide II
			1584			1571, 1589	vN=N aliphatic
1593, 1594							vC=C

Substrate removal results

Glucose study results: Beer-Lambert Method

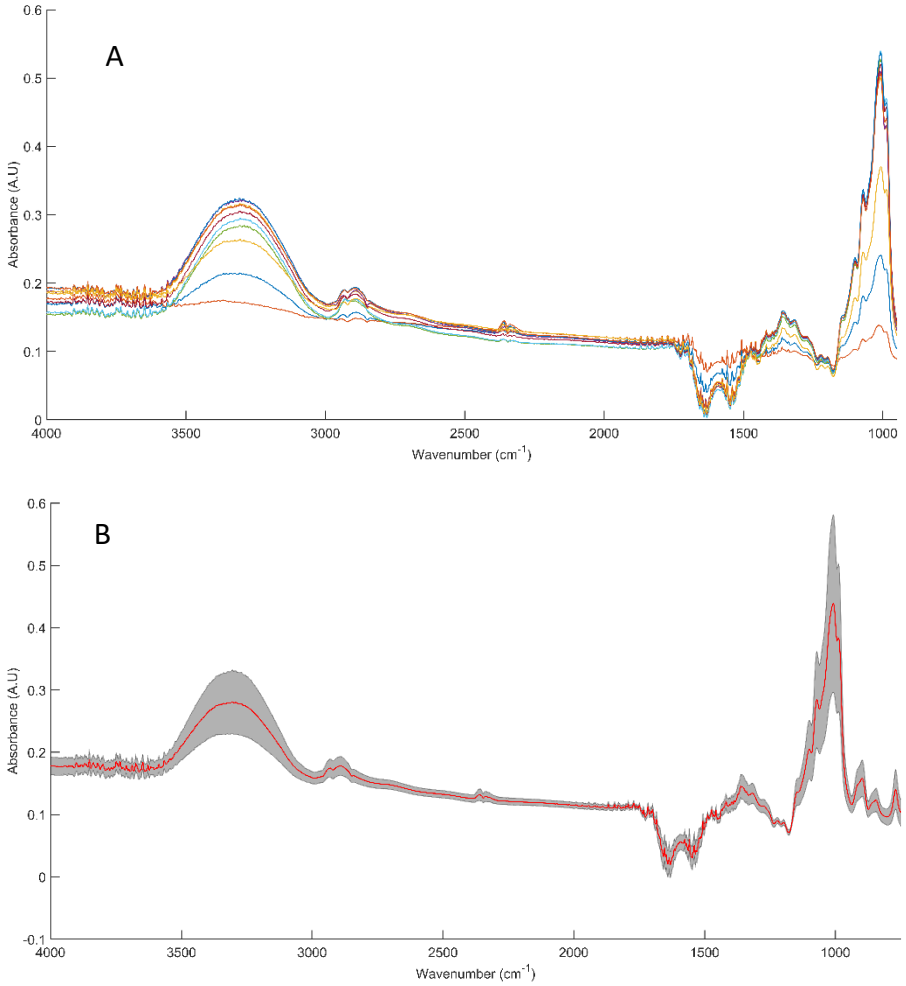

Figure 29: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Boot with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

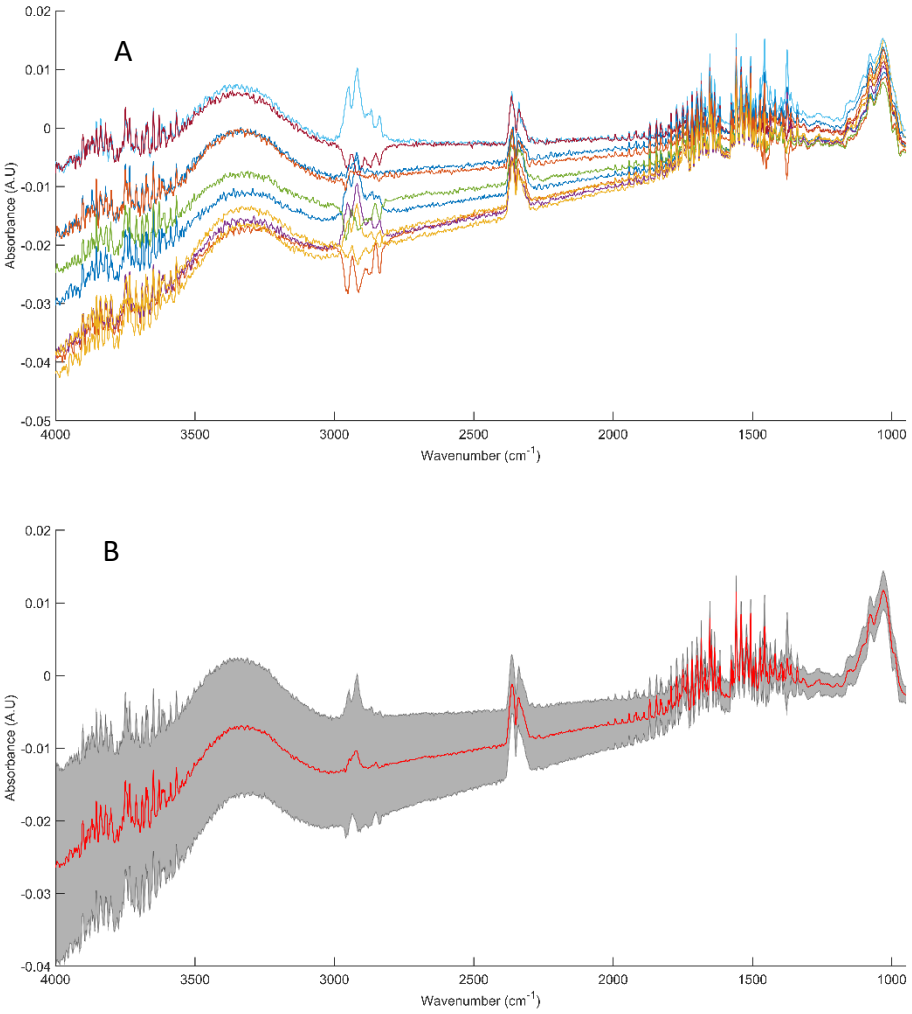

Figure 30: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Clean filter with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

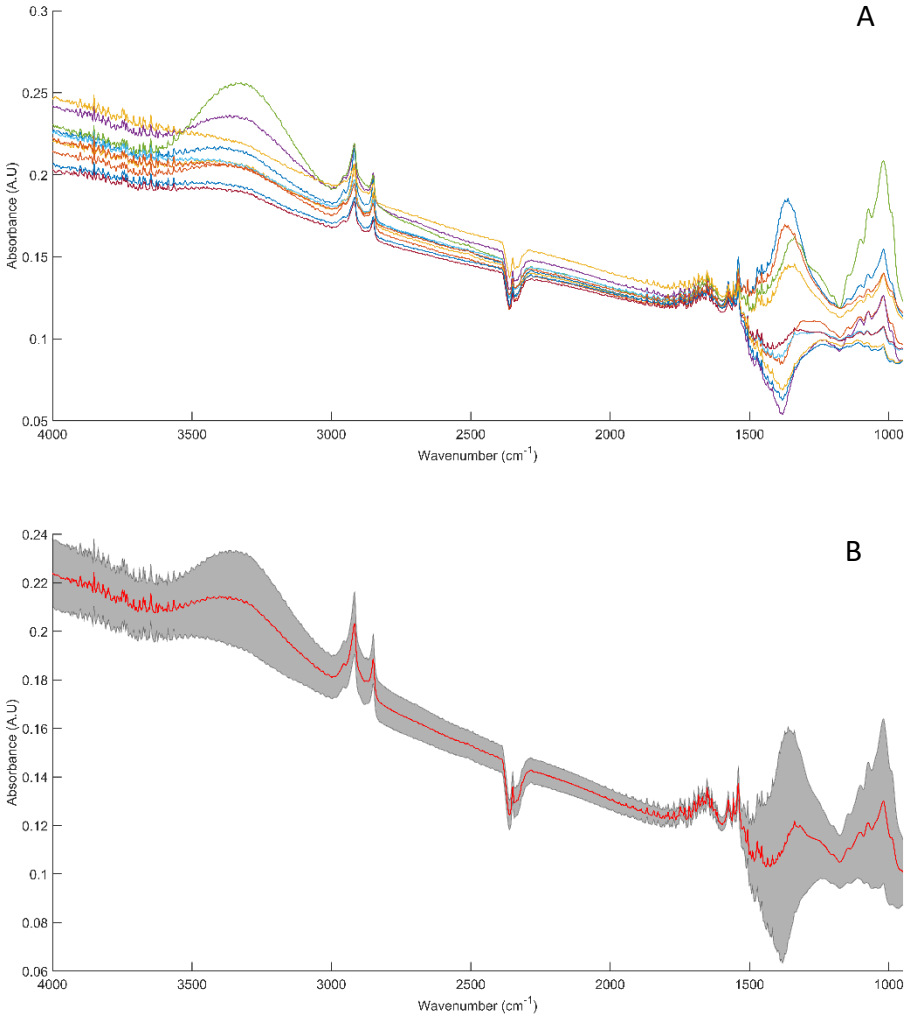

Figure 31: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Concrete with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

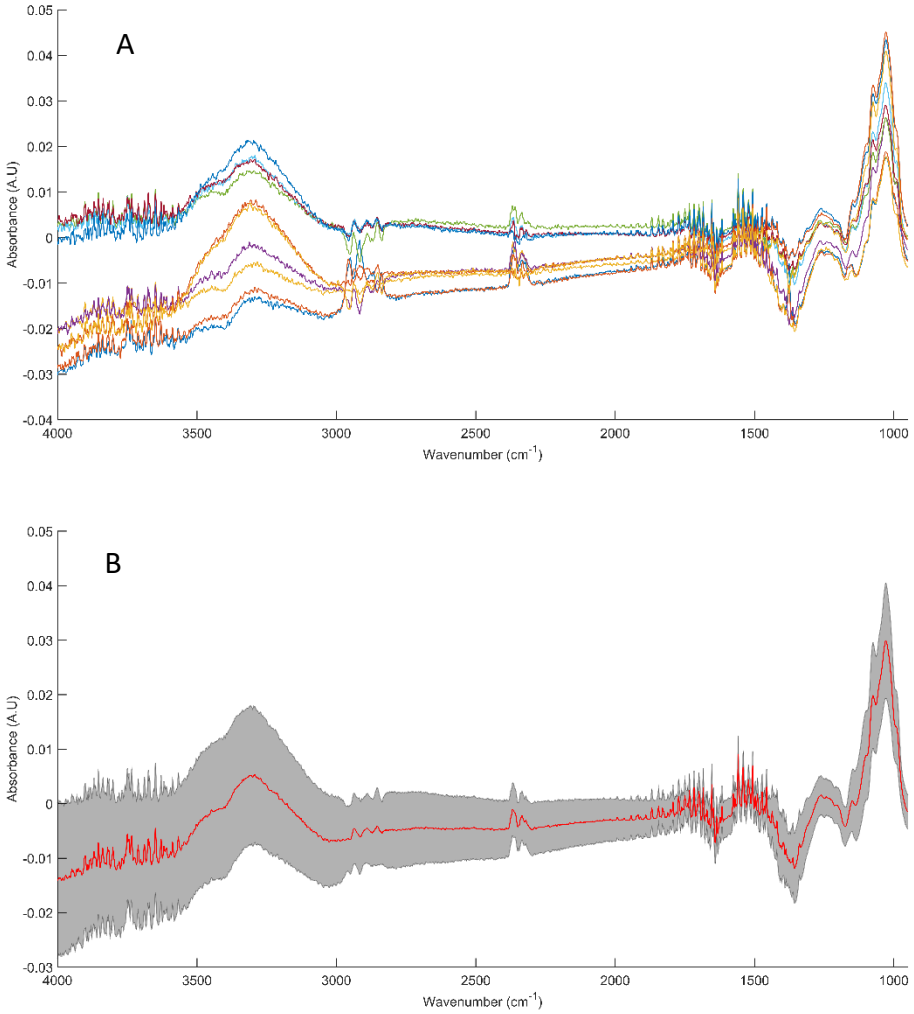

Figure 32: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Dirty filter with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

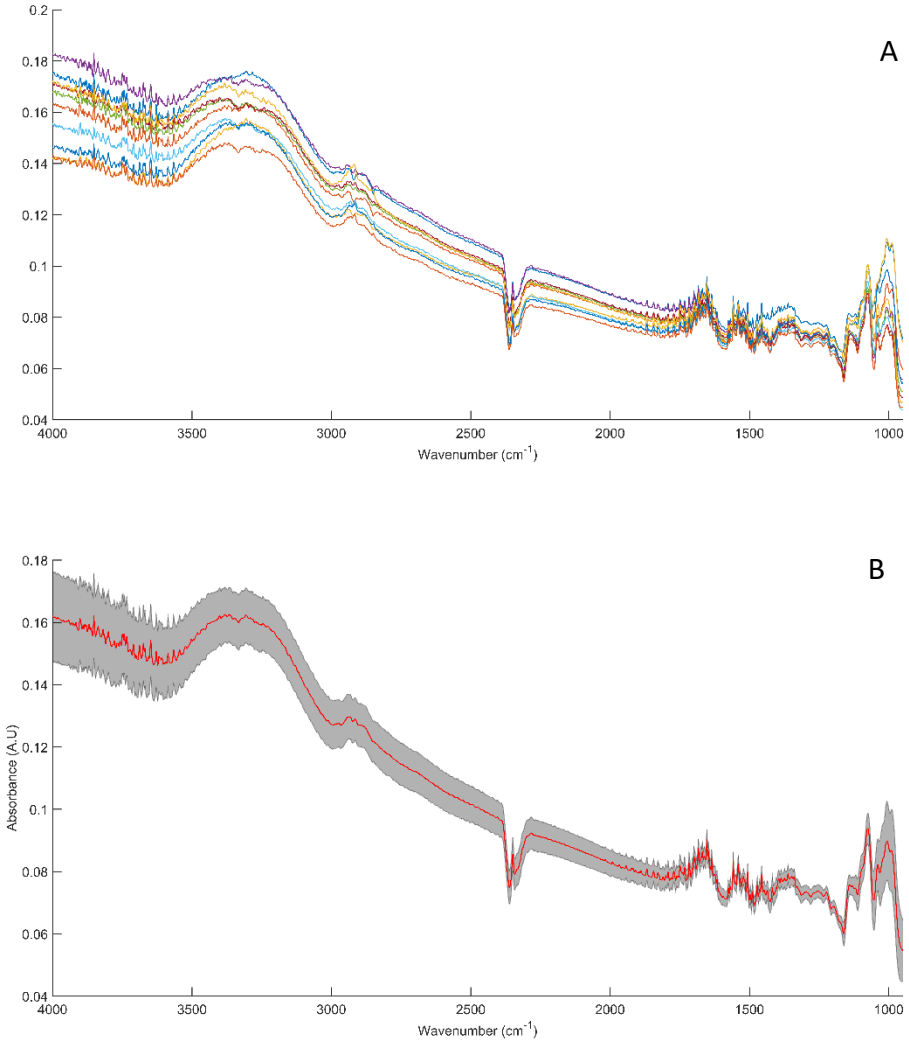

Figure 33: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Jacket with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

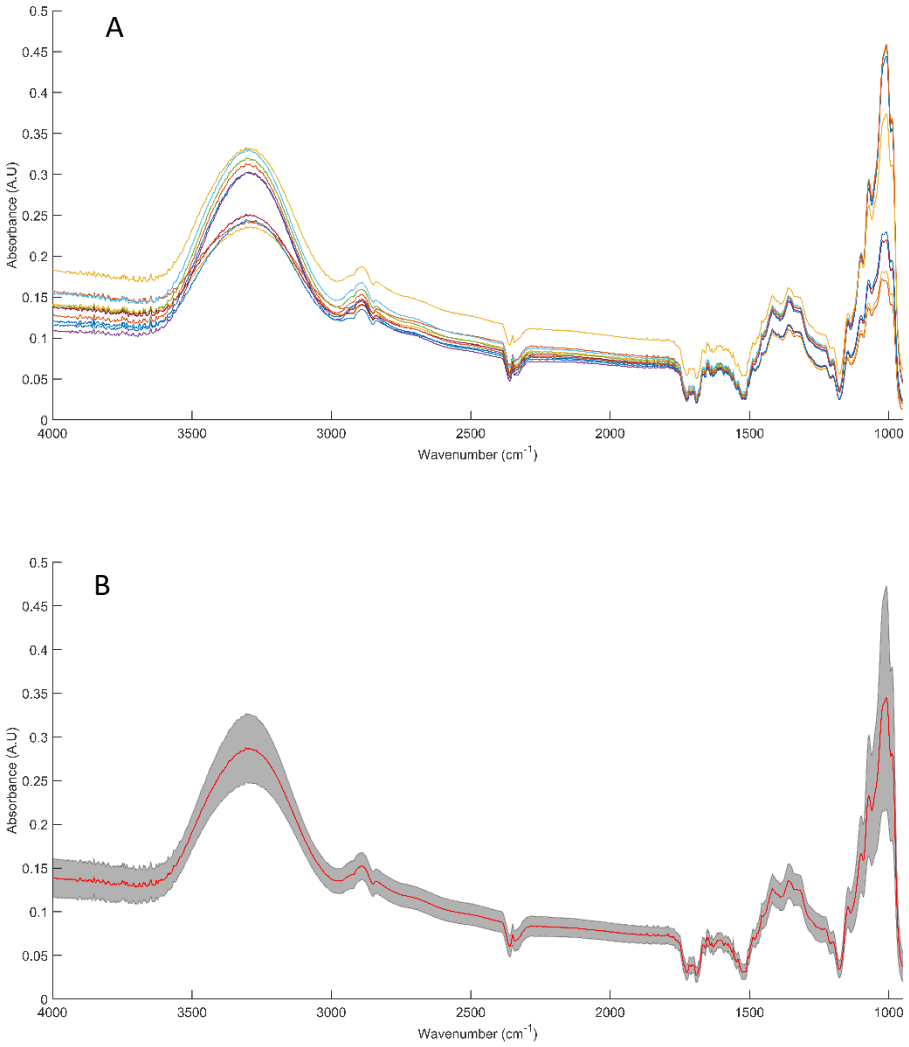

Figure 34: Spectra generated using the Beer-Lambert equation on single channel spectra collected from Metal tile with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

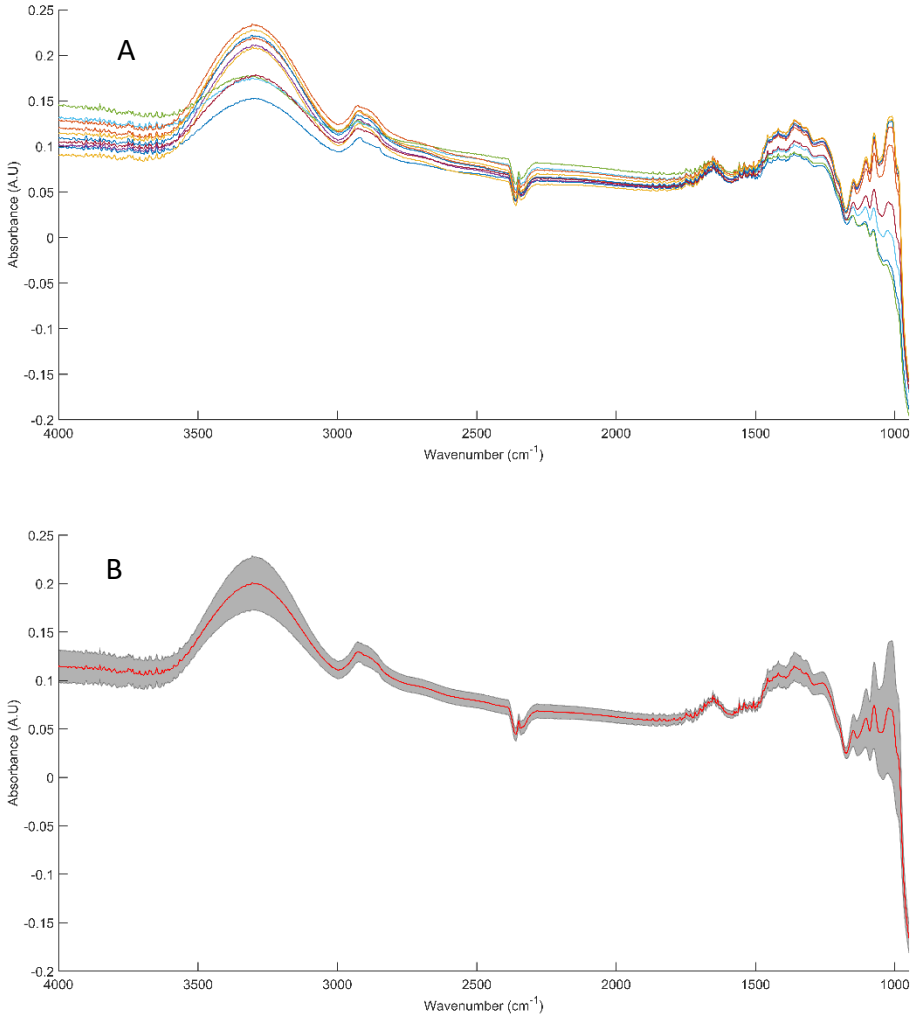

Figure 35: Spectra generated using the Beer-Lambert equation on single channel spectra collected from MirrIR slide with glucose applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

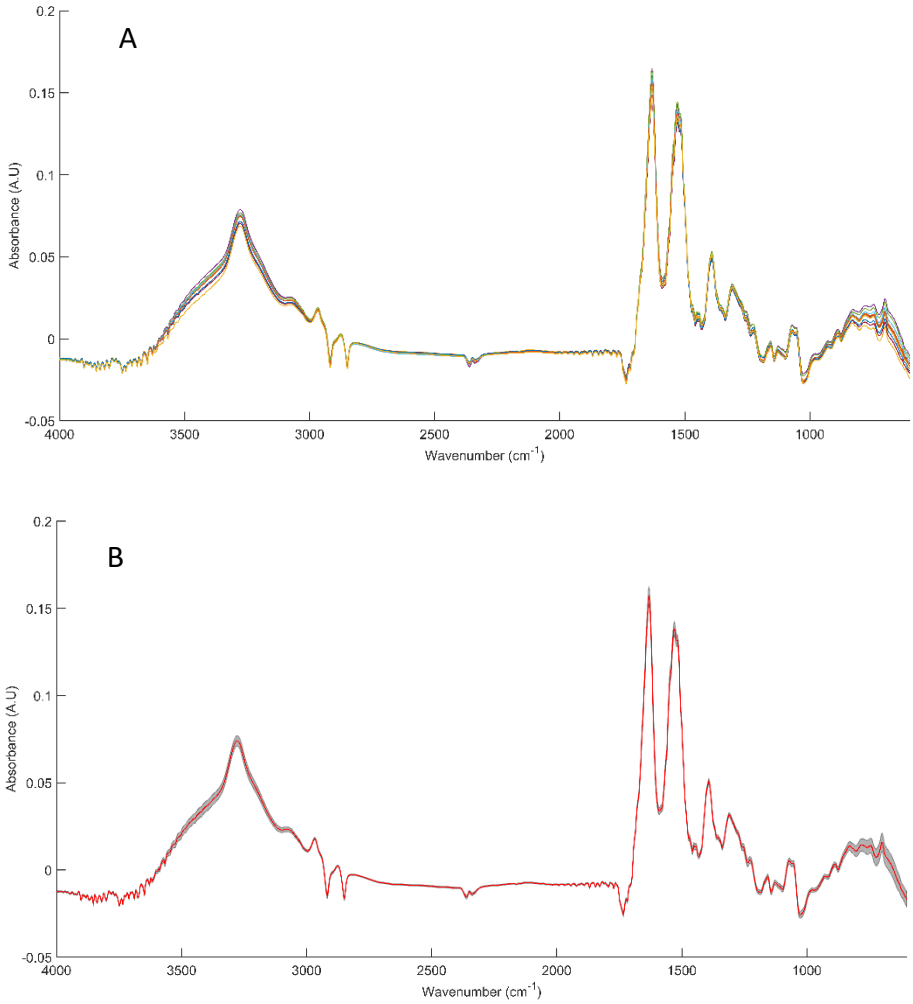

Figure 36: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Boot with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

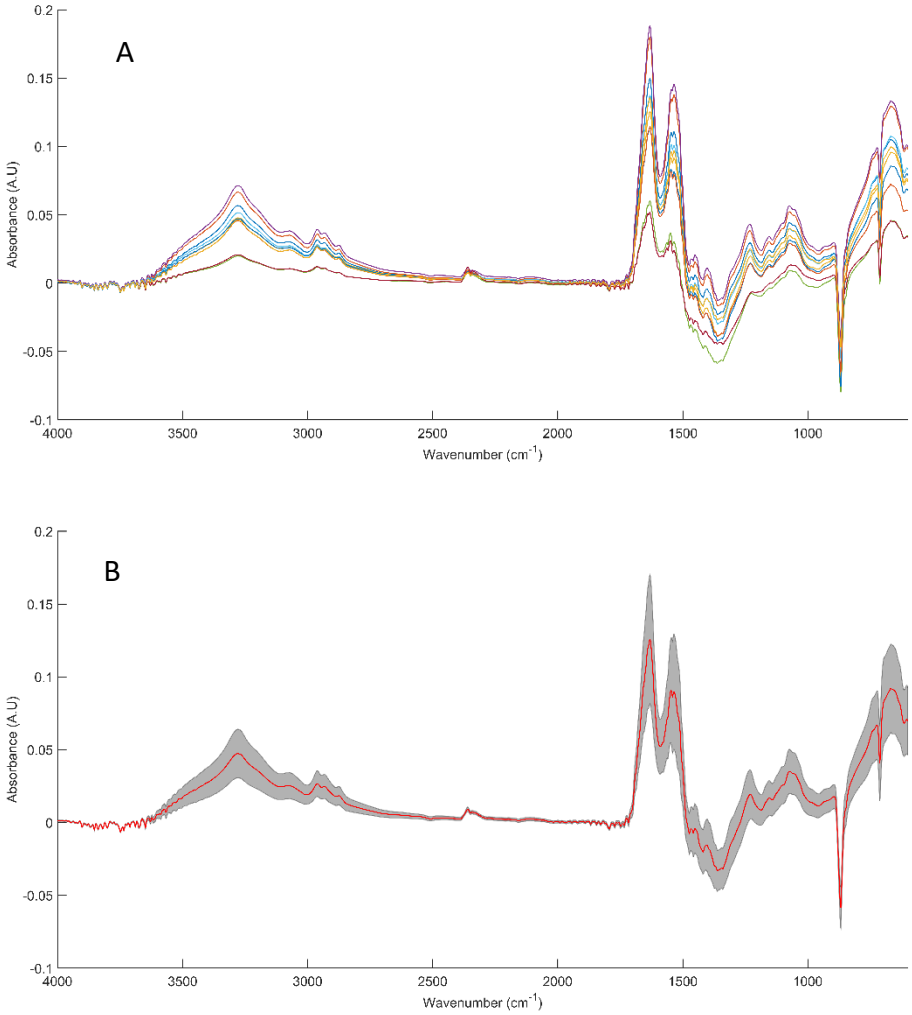

Figure 37: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Concrete with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

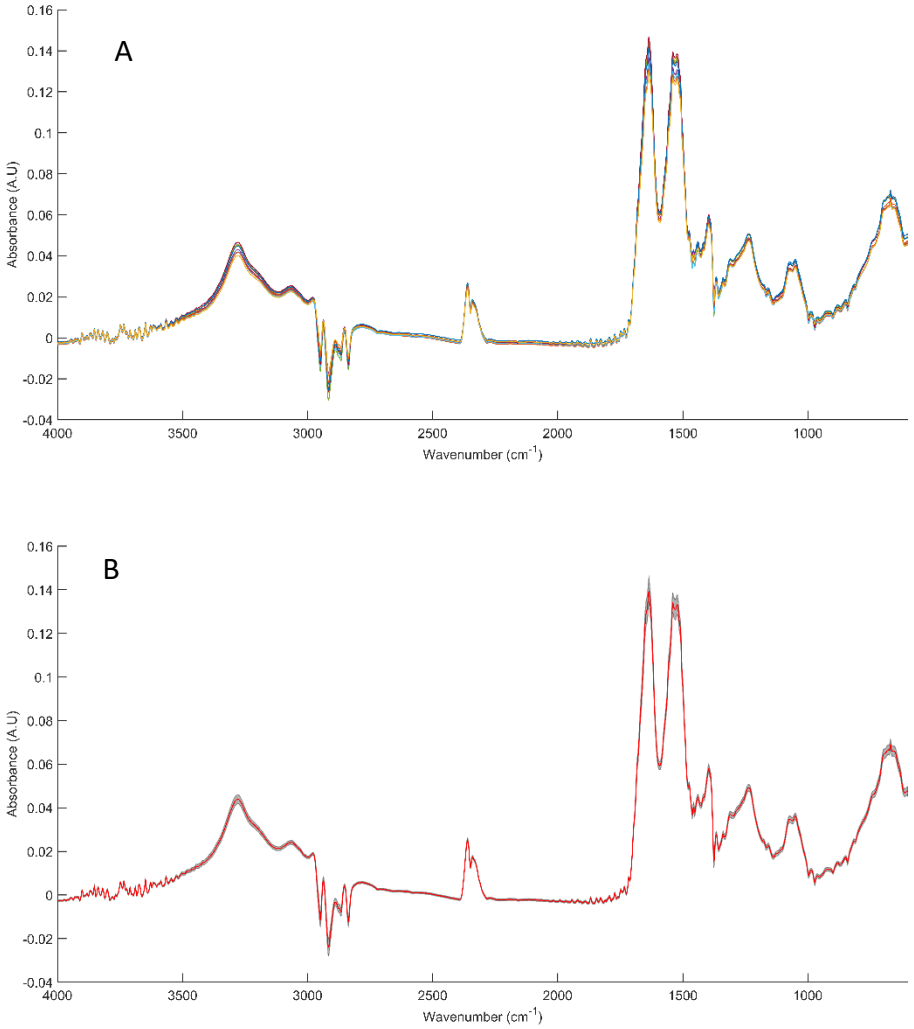

Figure 38: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Dirty filter with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

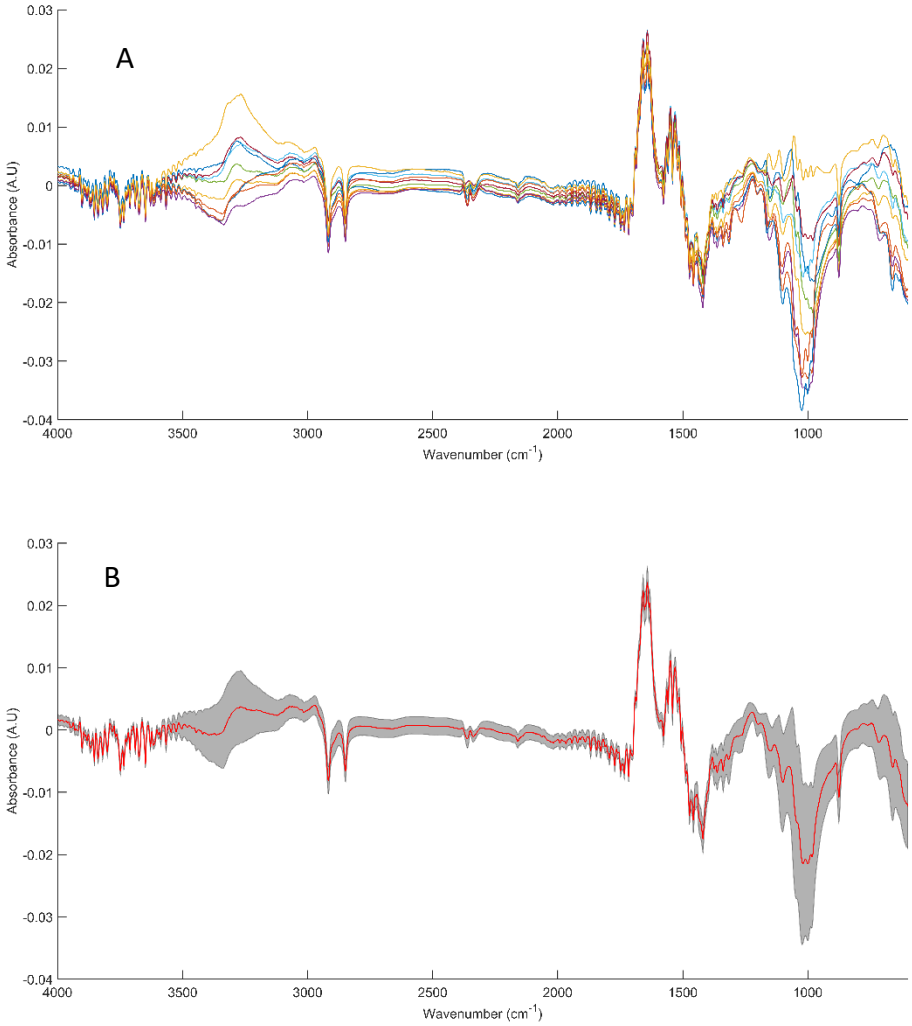

Figure 39: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Jacket with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

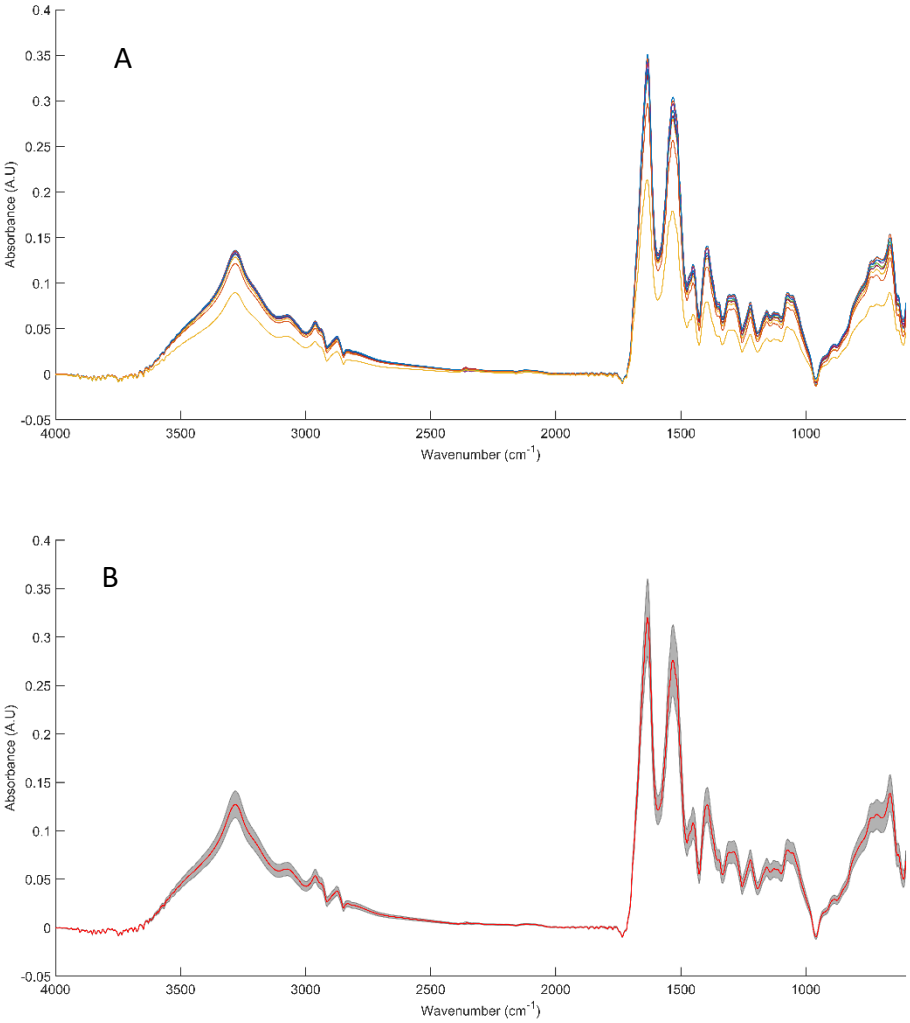

Figure 40: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Lino with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

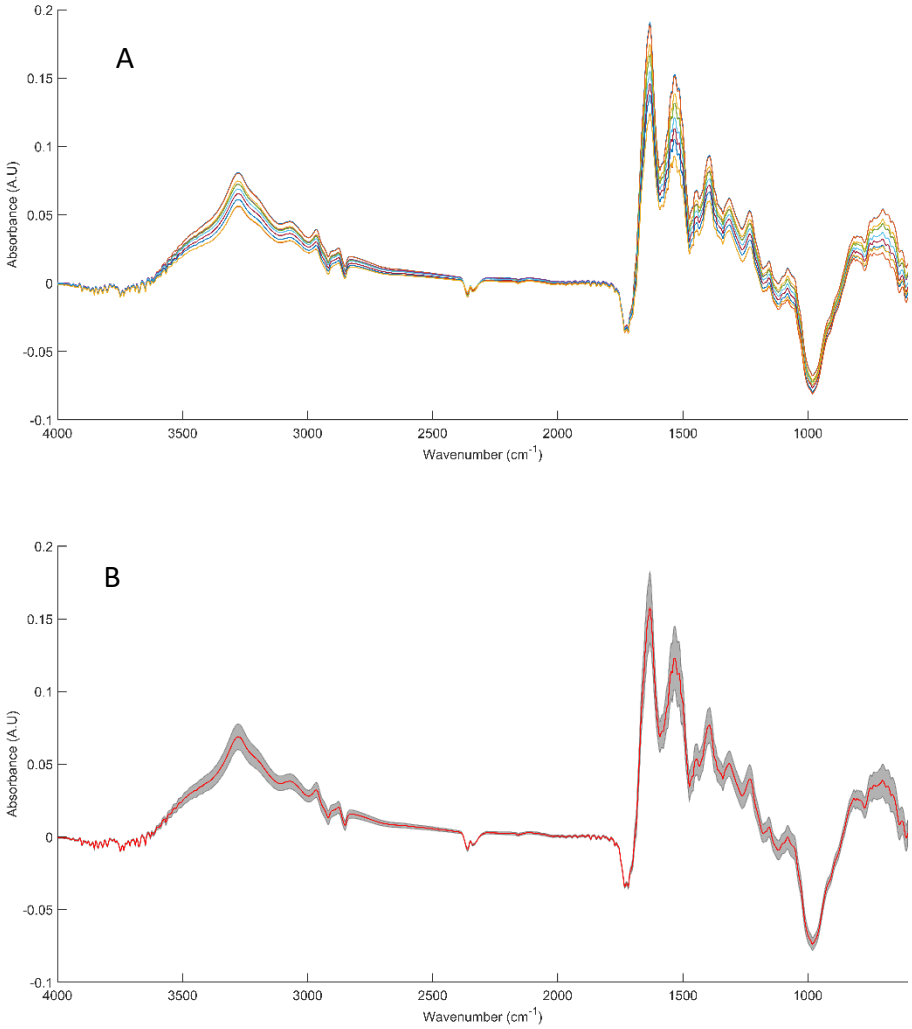

Figure 41: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from Metal tile with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

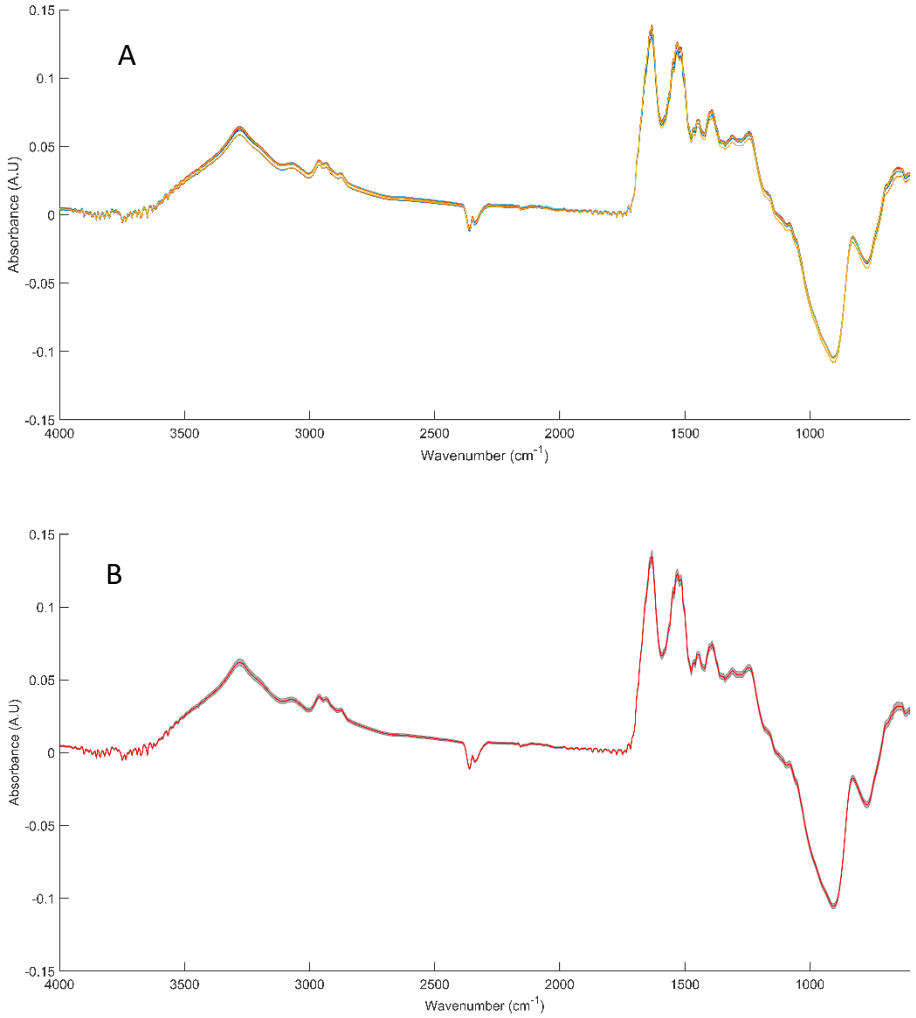

Figure 42: Spectra generated using the Beer-Lambert based method of substrate removal to generate absorbance spectra from MirrIR slide with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

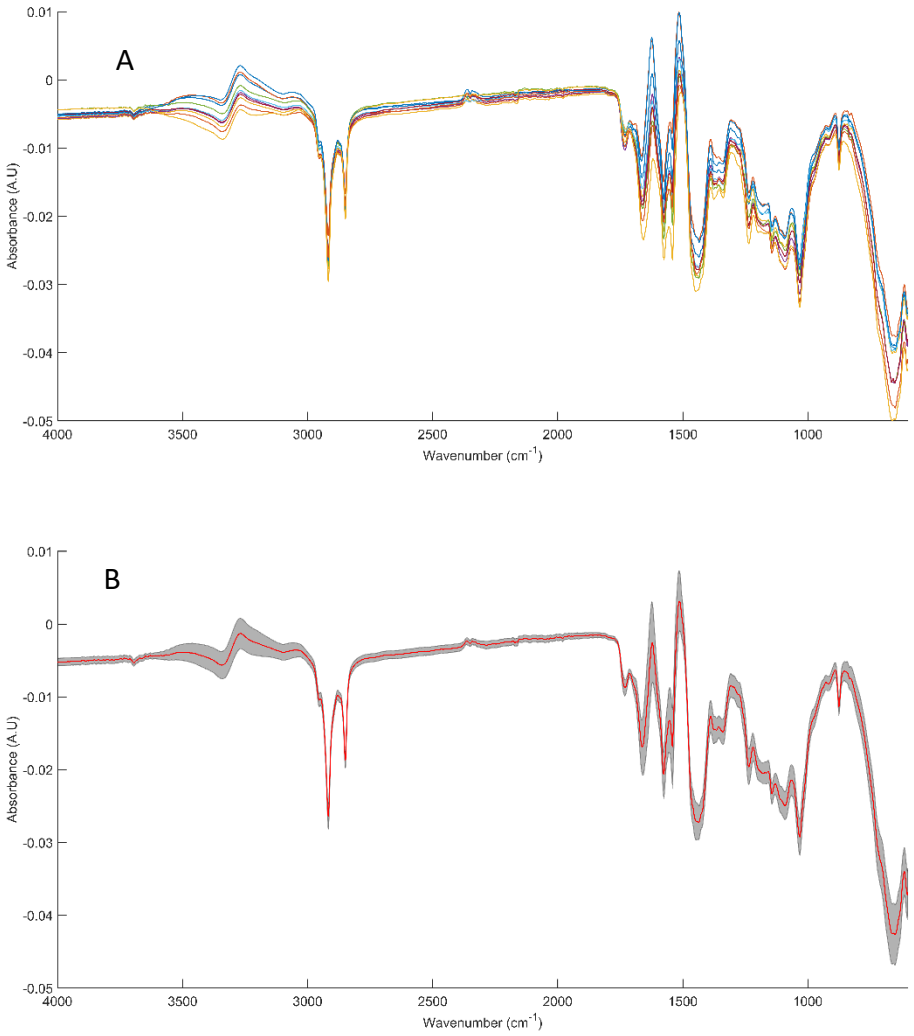

Figure 43: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Boot with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

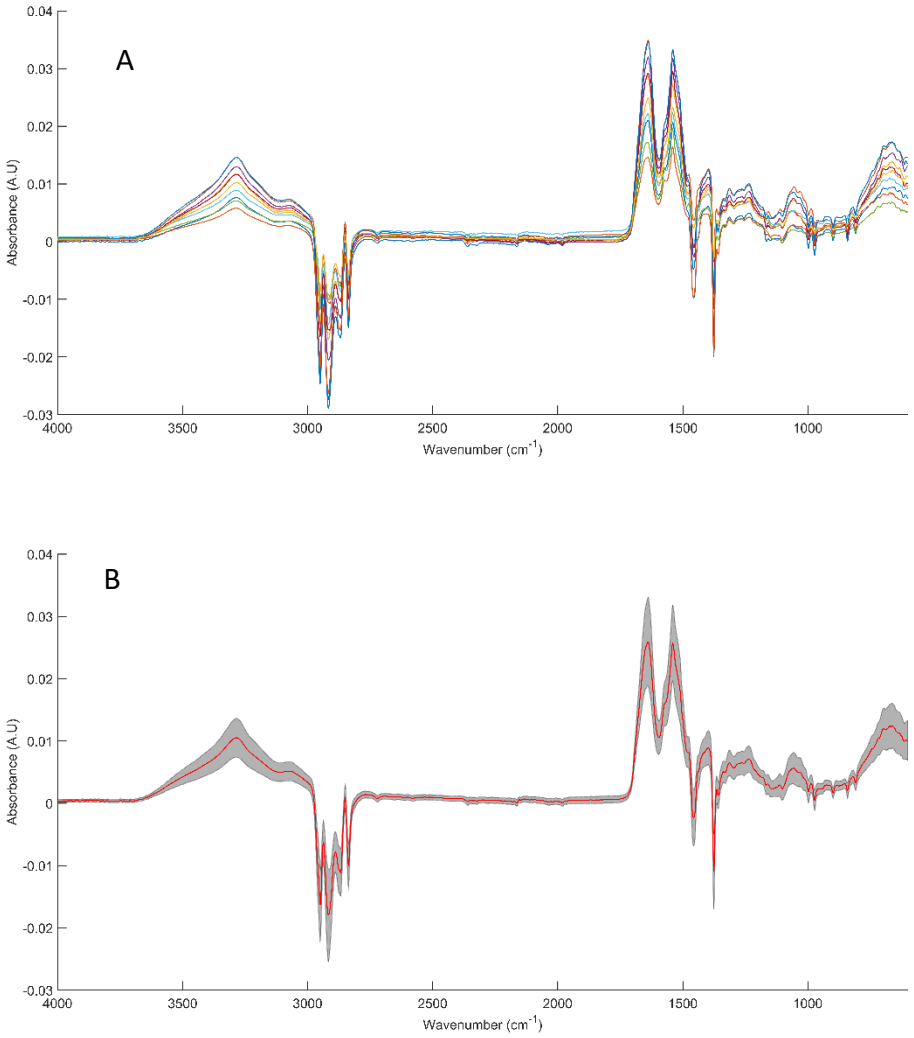

Figure 44: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Clean filter with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

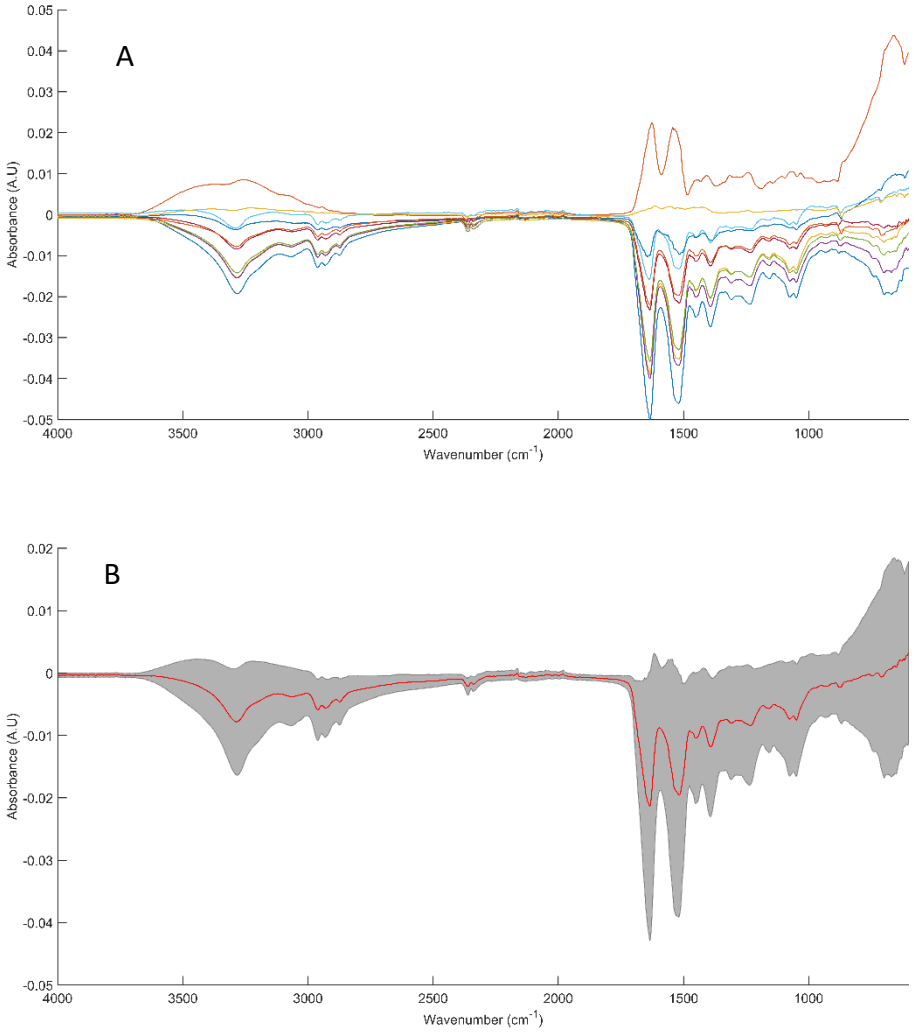

Figure 45: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Concrete with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

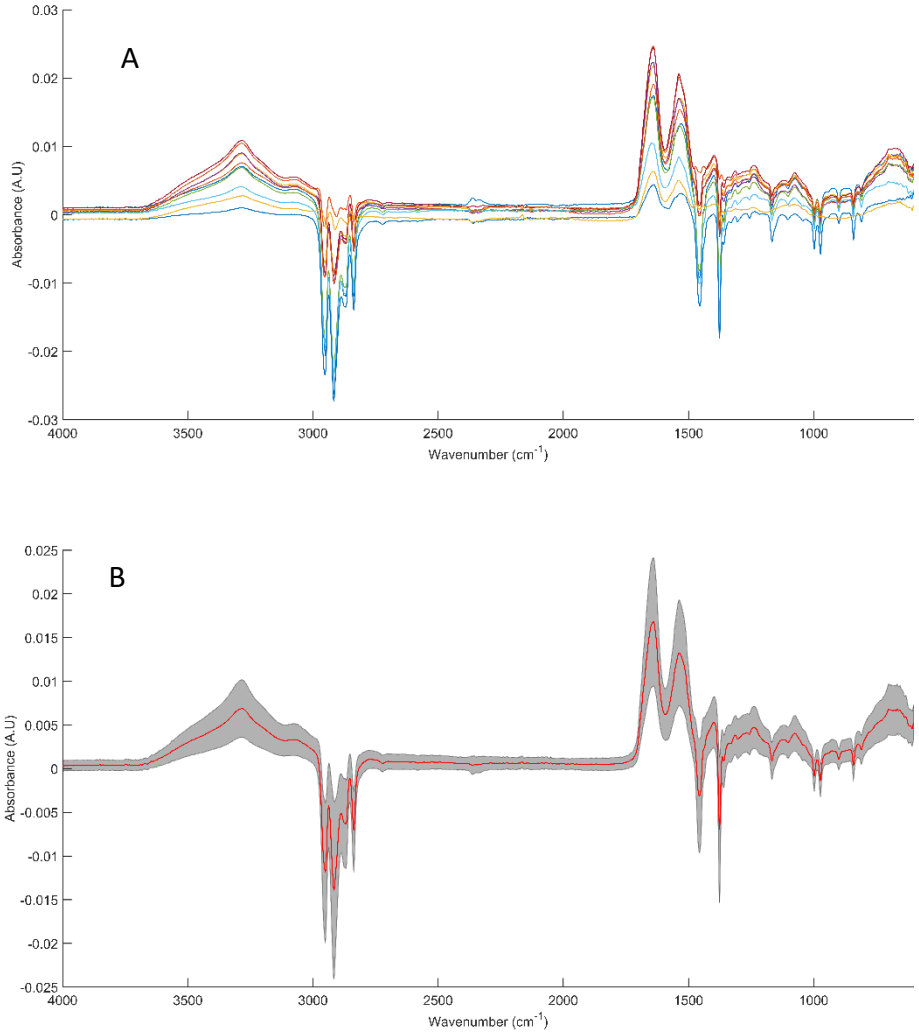

Figure 46: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Dirty filter with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

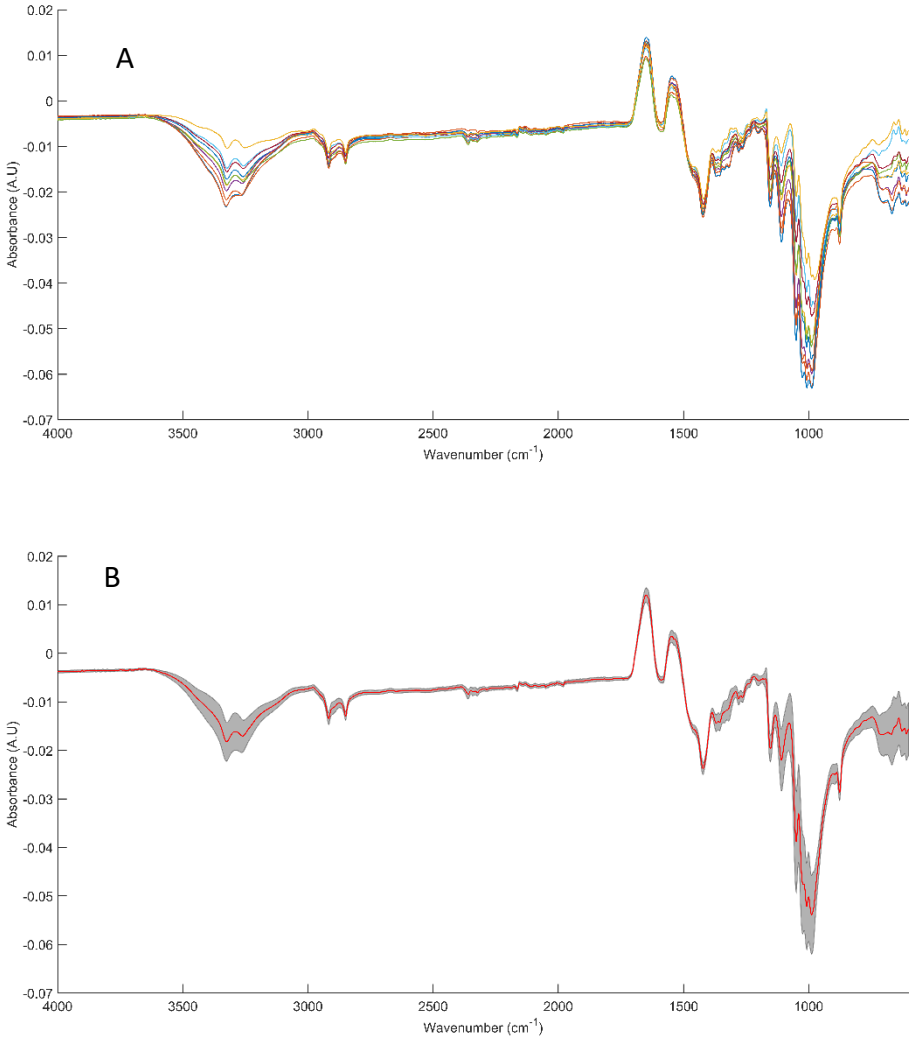

Figure 47: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Jacket with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

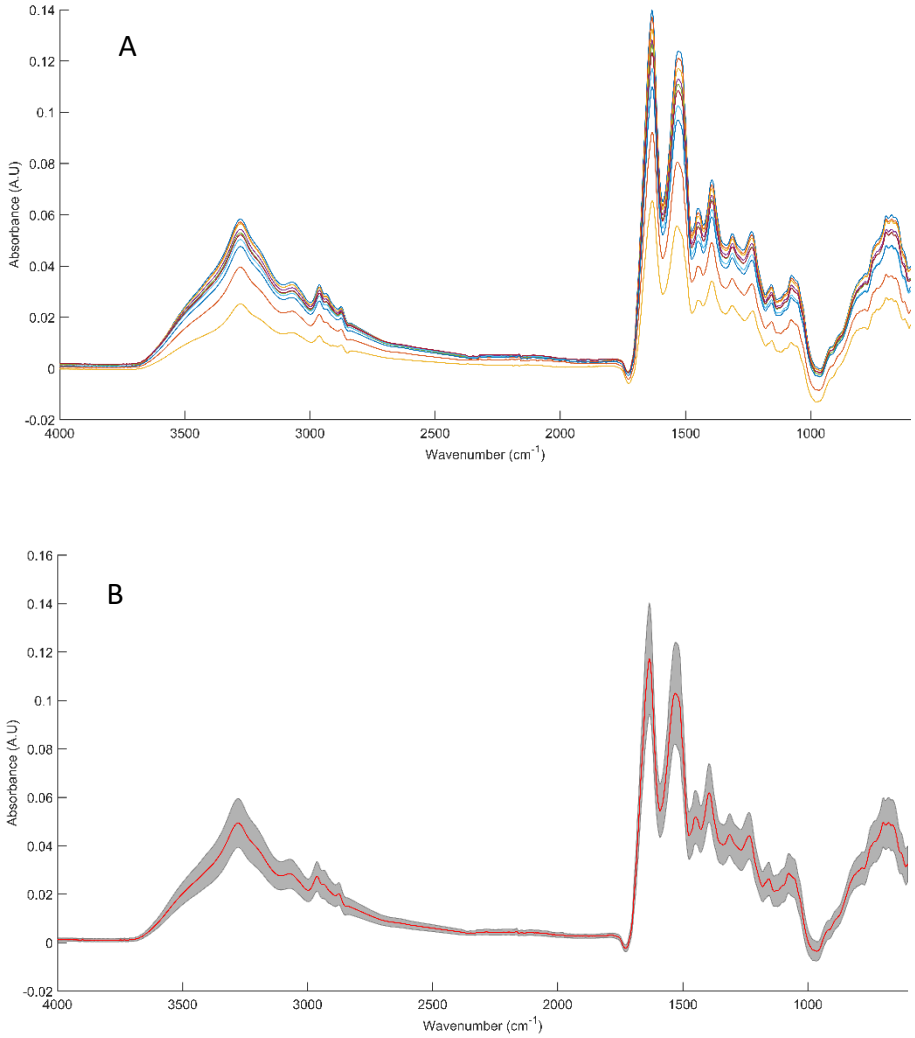

Figure 48: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from Metal tile with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

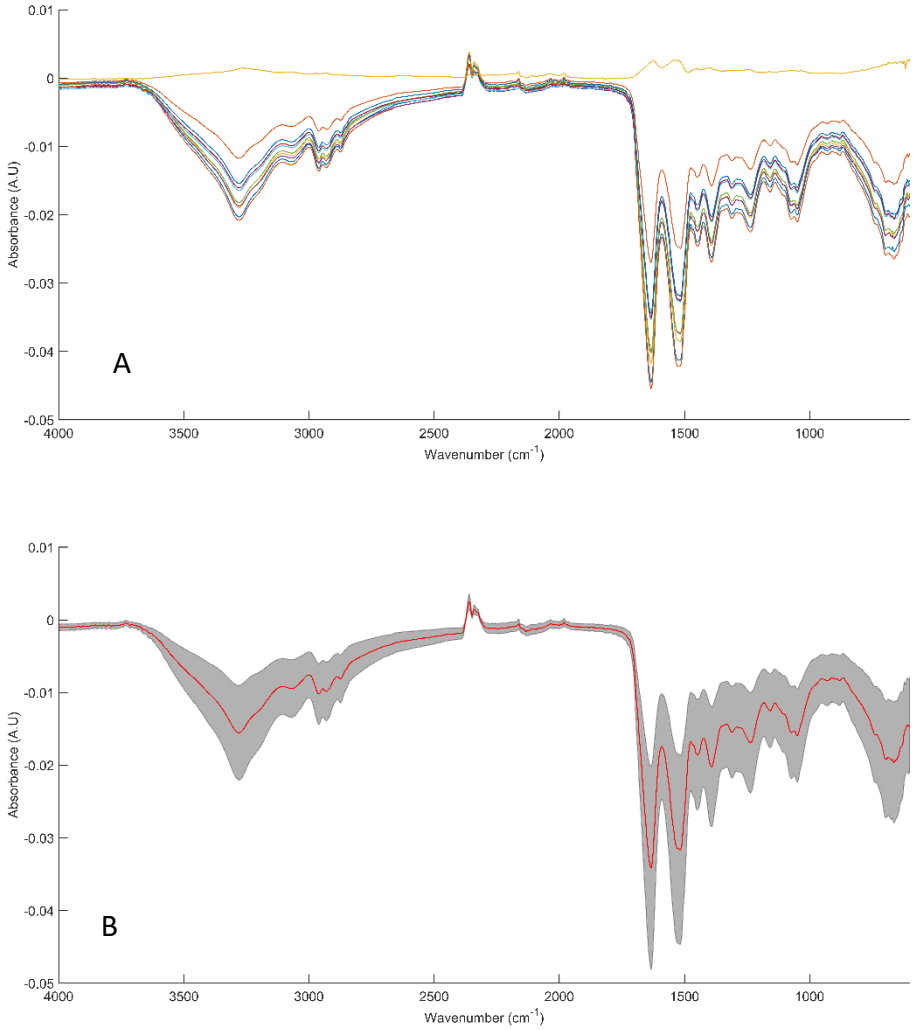

Figure 49: Spectra generated using the software-based method of substrate removal to generate absorbance spectra from MirrIR slide with ovalbumin applied, showing an overlay of 10 absorbance spectra (A) and the mean absorbance spectrum shown in red, surrounded by a grey standard deviation cloud (B).

References

1 Baker, M. J., Hughes, C. S. & Hollywood, K. A. *Biophotonics: Vibrational Spectroscopic Diagnostics*. (Morgan & Claypool, 2016).

2 Clemens, G. N., Hands, J. R., Dorling, K. & Baker, M. J. Vibrational Spectroscopic Methods for Cytology and Cellular Research. *Analyst* **139**, 4411-4444 (2014).

3 Beauchamp, P. *Spectroscopy Data Tables*, <http://www.cpp.edu/~psbeauchamp/pdf/424_spectra_tables.pdf> (2011).

4 Horiba. *Raman Data and Analysis*, <http://www.horiba.com/fileadmin/uploads/Scientific/Documents/Raman/bands.pdf>

Appendix B

Pre-processing Data

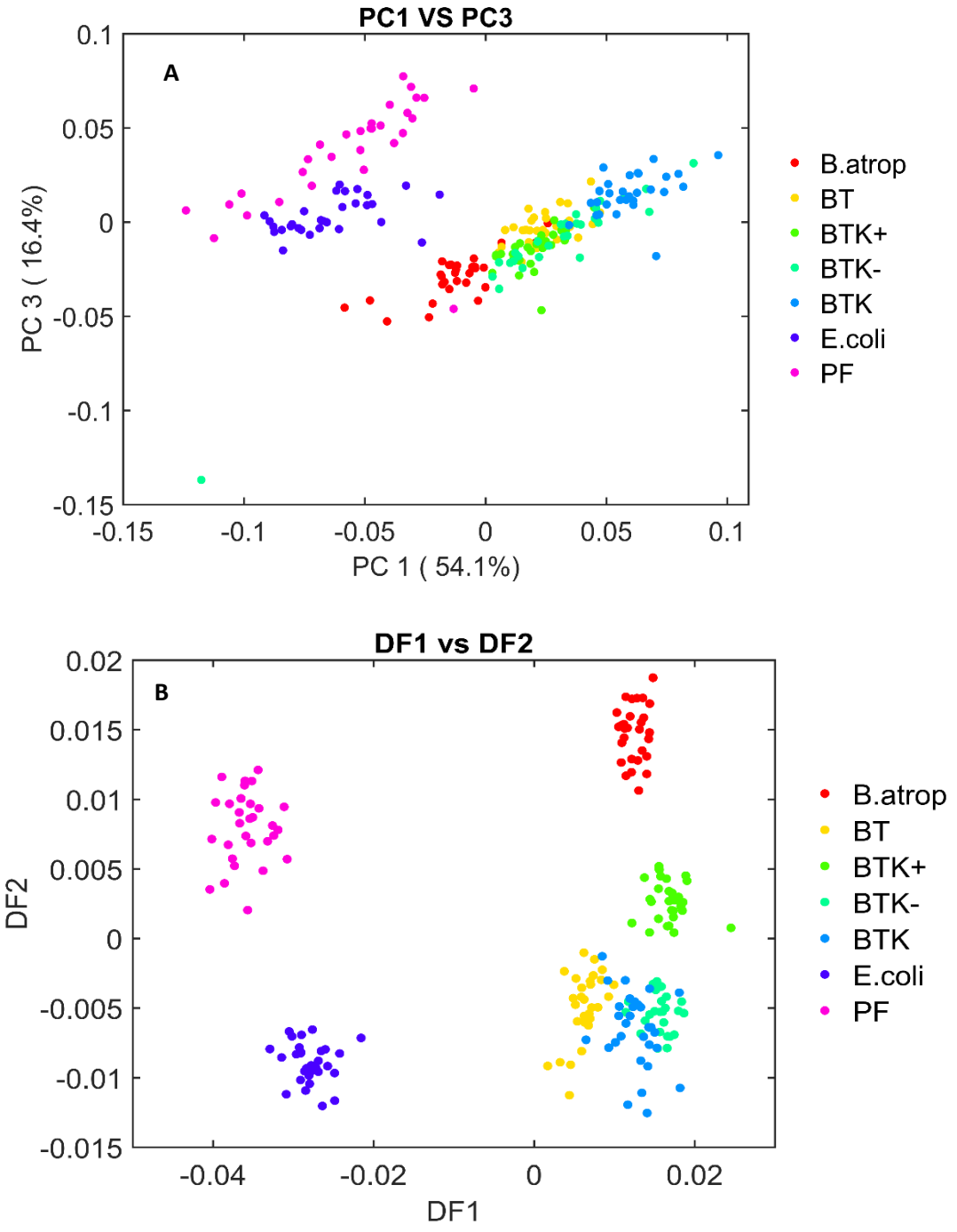

Figure 1: Figure showing a PCA plot using PC1 and PC3 (A) and a DFA plot using DF1 and DF2 (B) from data pre-processed using Method 1.

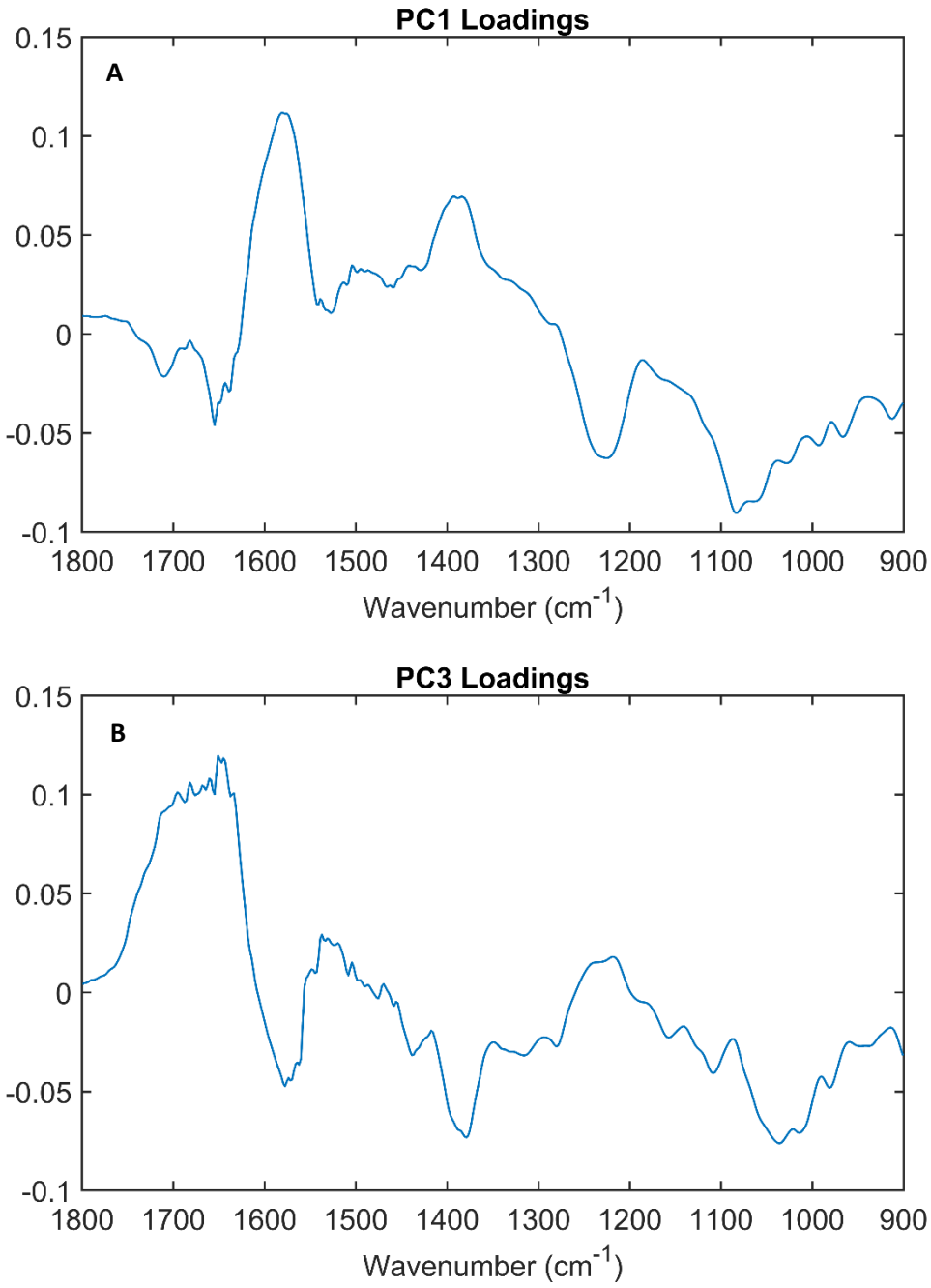

Figure 2: Figure showing a PC1 loadings plot (A) and a PC3 loadings plot (B) from data pre-processed using Method 1.

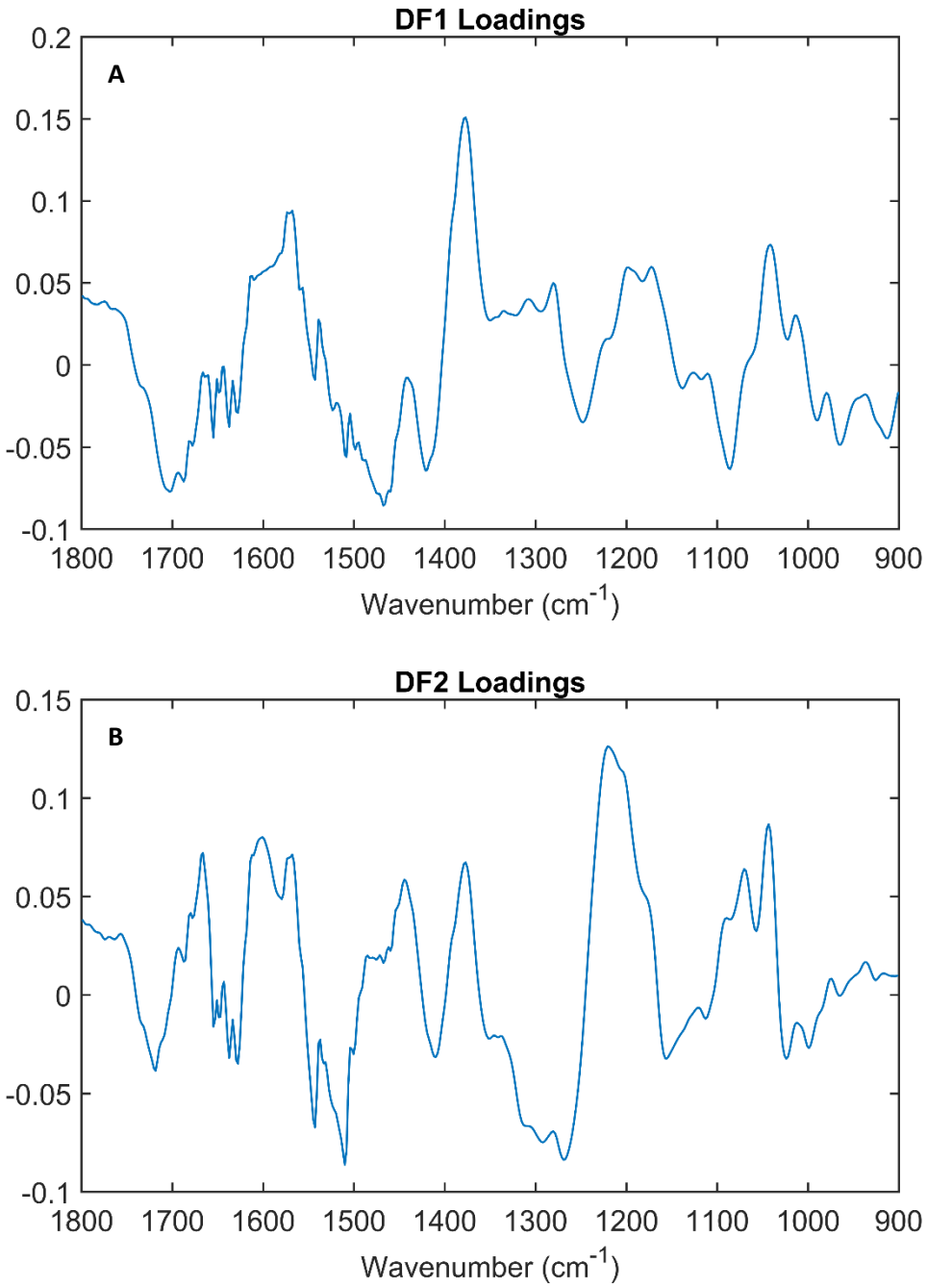

Figure 3: Figure showing a DF1 loadings plot (A) and a DF2 loadings plot (B) from data pre-processed using Method 1.

Table 1: Wavenumbers and corresponding peak assignments for PC1, PC3, DF1 and DF2 loadings from Method 1. [5,6,7,2,8,]

Wavenumber (cm^{-1})				Tentative peak assignment
PC1	PC3	DF 1	DF 2	
966, 993, 1028	1014	1173	1199, 1201	vC-O, vC-C, vC-O-H, vC-O-C vP=O symmetric of PO^-_2
	1036			Phenylalanine C-H in-plane
1065,1084		1041, 1086	1043, 1070	vP=O symmetric of PO^-_2
1225			1221, 1269, 1292, 1309	Amide III band of proteins
1385, 1392	1379, 1388	1379, 1468, 1419, 1460, 1475, 1491, 1498	1377, 1444	δCH_2 and δCH_3
		1508, 1568	1510, 1543, 1568	Amide II
1576, 1581		1574		Aspartate and glutamate carboxylate stretching
			1601, 1666	vC=O
		1613	1612	Tyrosine and Tryptophan
	1633			Amide I β-pleated sheet structures
	1645, 1651, 1660, 1668			Amide I
	1682, 1695	1687		Amide I band components resulting from antiparallel Pleated sheets from β-turns of proteins
		1703		vC=O

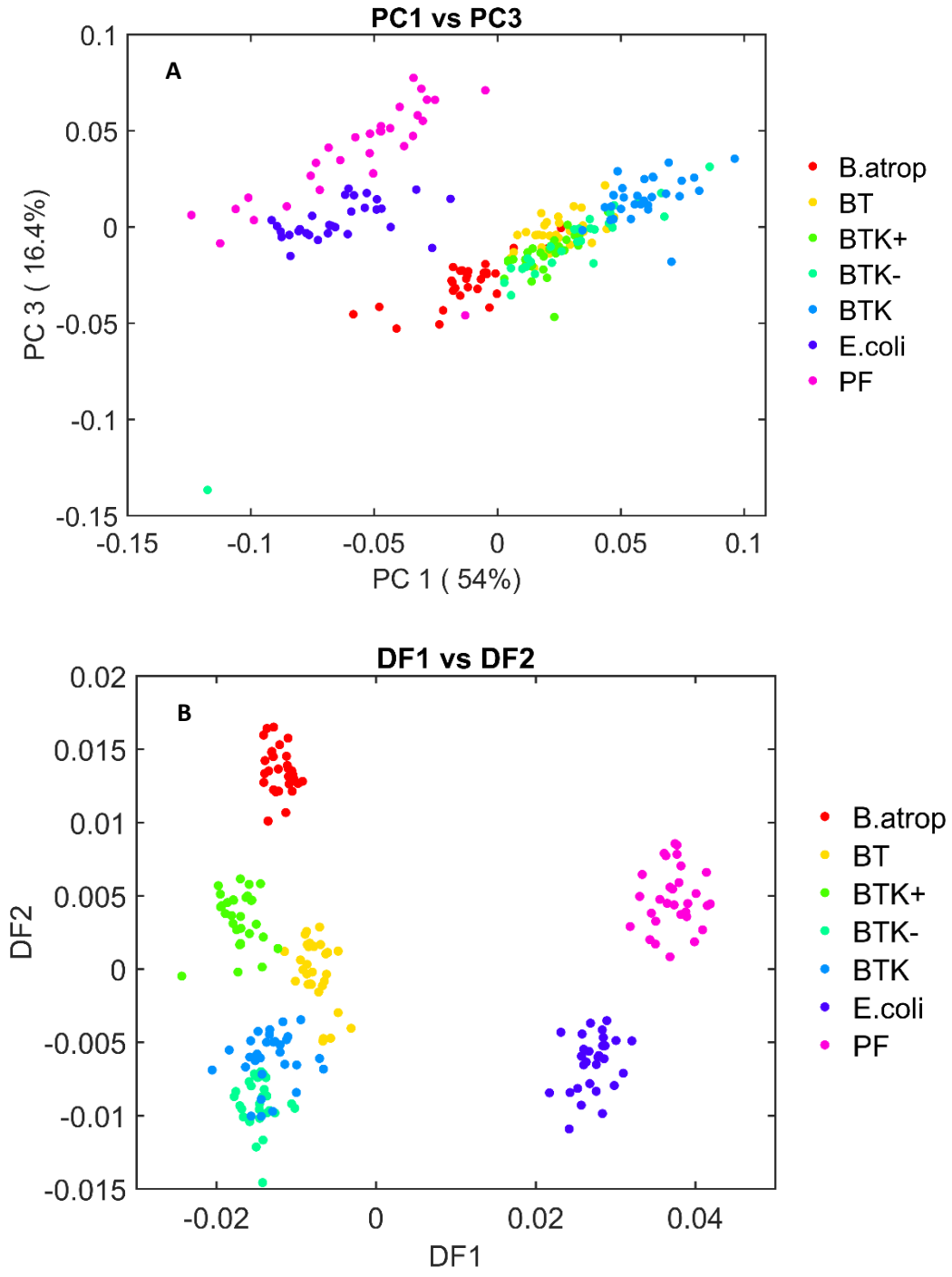

Figure 4: Figure showing a PCA plot using PC1 and PC3 (A) and a DFA plot using DF1 and DF2 (B) from data pre-processed using Method 2.

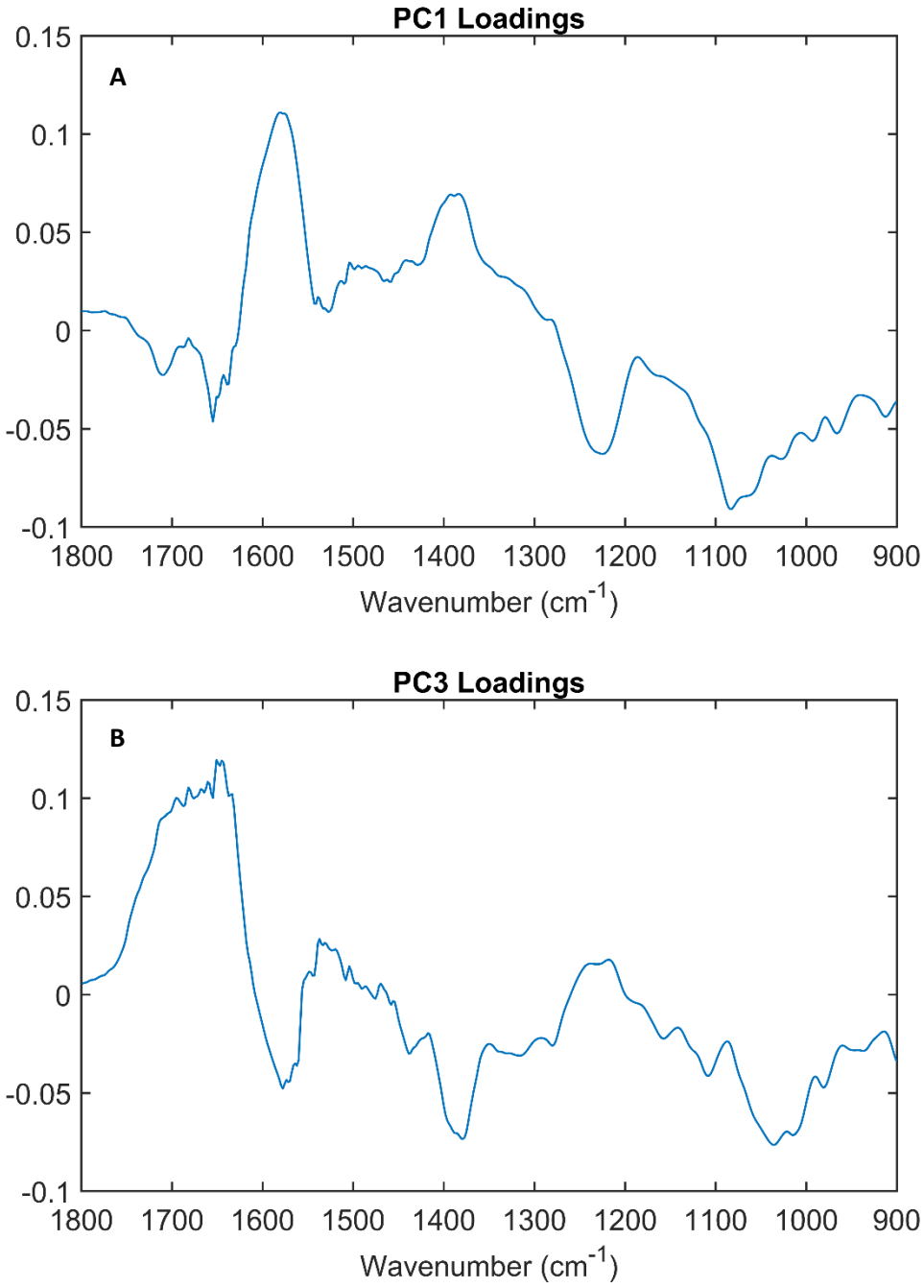

Figure 5: Figure showing a PC1 loadings plot (A) and a PC3 loadings plot (B) from data pre-processed using Method 2.

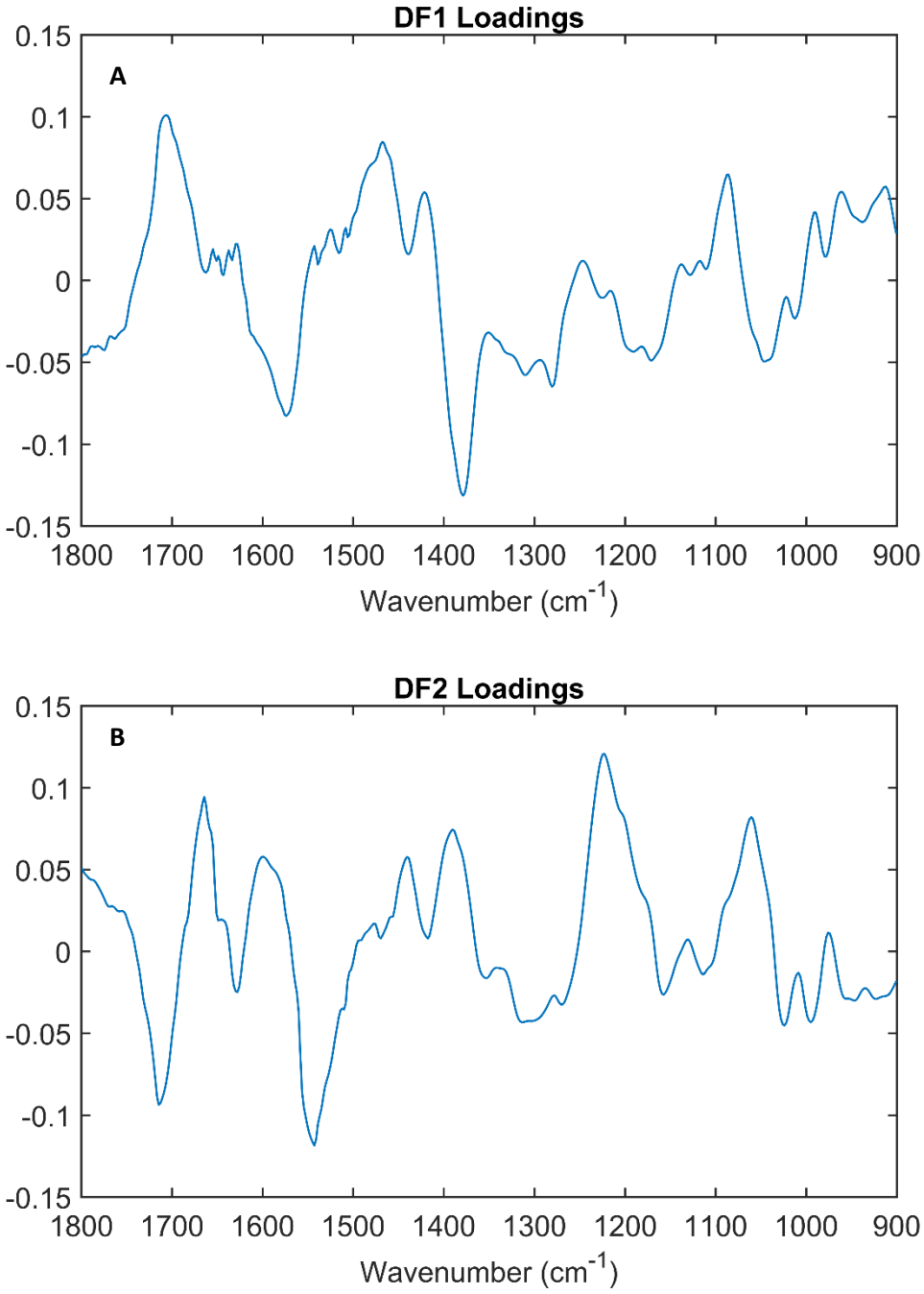

Figure 6: Figure showing a DF1 loadings plot (A) and a DF2 loadings plot (B) from data pre-processed using Method 2.

Table 2 Wavenumbers and corresponding peak assignments for PC1, PC3, DF1 and DF2 loadings from Method 2. [5,6,7,2,8,]

Wavenumber (cm^{-1})				Tentative peak assignment
PC1	PC3	DF 1	DF 2	
966, 993, 1028	1014	912, 960		νC-O, νC-C, νC-O-H, νC-O-C νP=O (symmetric) of PO^-_2
	1036			Phenylalanine C-H in-plane
			1061	C-N, C-C stretch
1059, 1084		1086		νP=O (symmetric) of PO^-_2
			1201	νPO^-_2 (asymmetric) of PO^-_2
1225		1280, 1309	1225	Amide III band of proteins
1383, 1392	1379, 1388	1379, 1468, 1421, 1483	1390, 1441	δCH_2 and δCH_3
			1543	Amide II
1579		1574	1599	Aspartate and glutamate carboxylate stretching
	1633, 1645			Amide I β-pleated sheet structures
	1645, 1651, 1660		1664	Amide I
			1657	Amide I α-helix
	1682, 1695			Amide I band components resulting from antiparallel Pleated sheets from β-turns of proteins
		1707	1714	νC=O

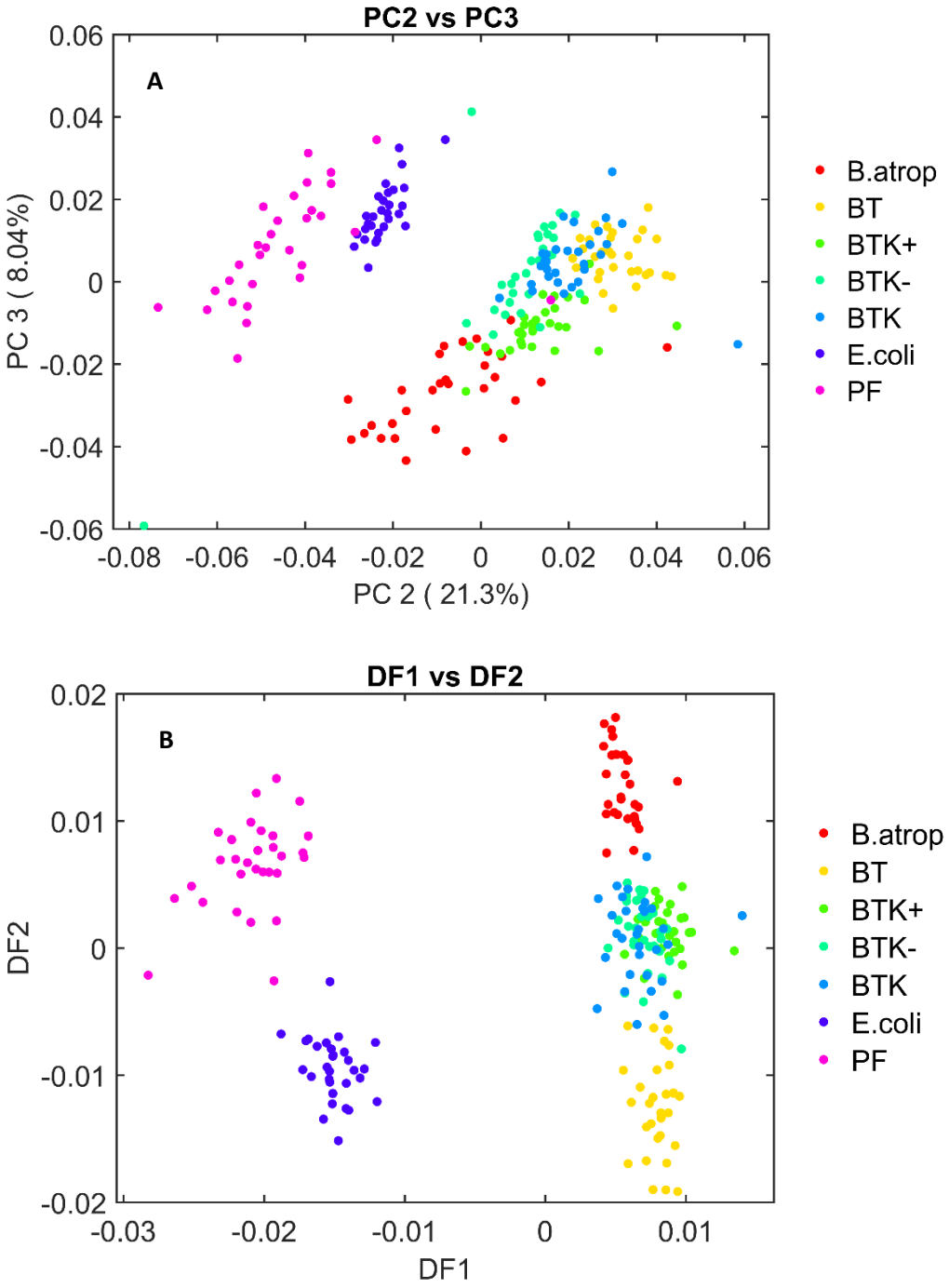

Figure 7: Figure showing a PCA plot using PC1 and PC3 (A) and a DFA plot using DF1 and DF2 (B) from data pre-processed using Method 4.

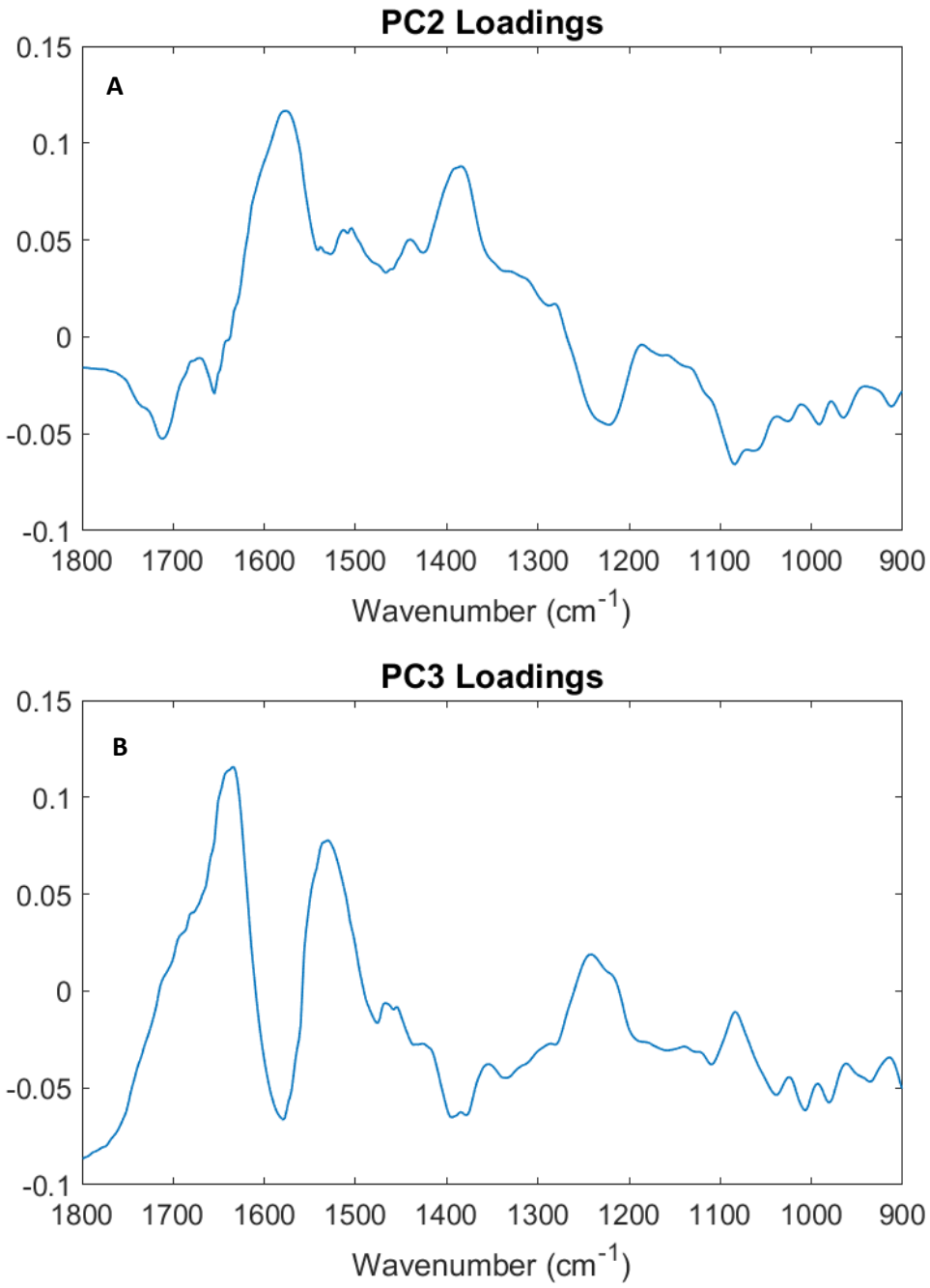

Figure 8: Figure showing a PC1 loadings plot (A) and a PC3 loadings plot (B) from data pre-processed using Method 4.

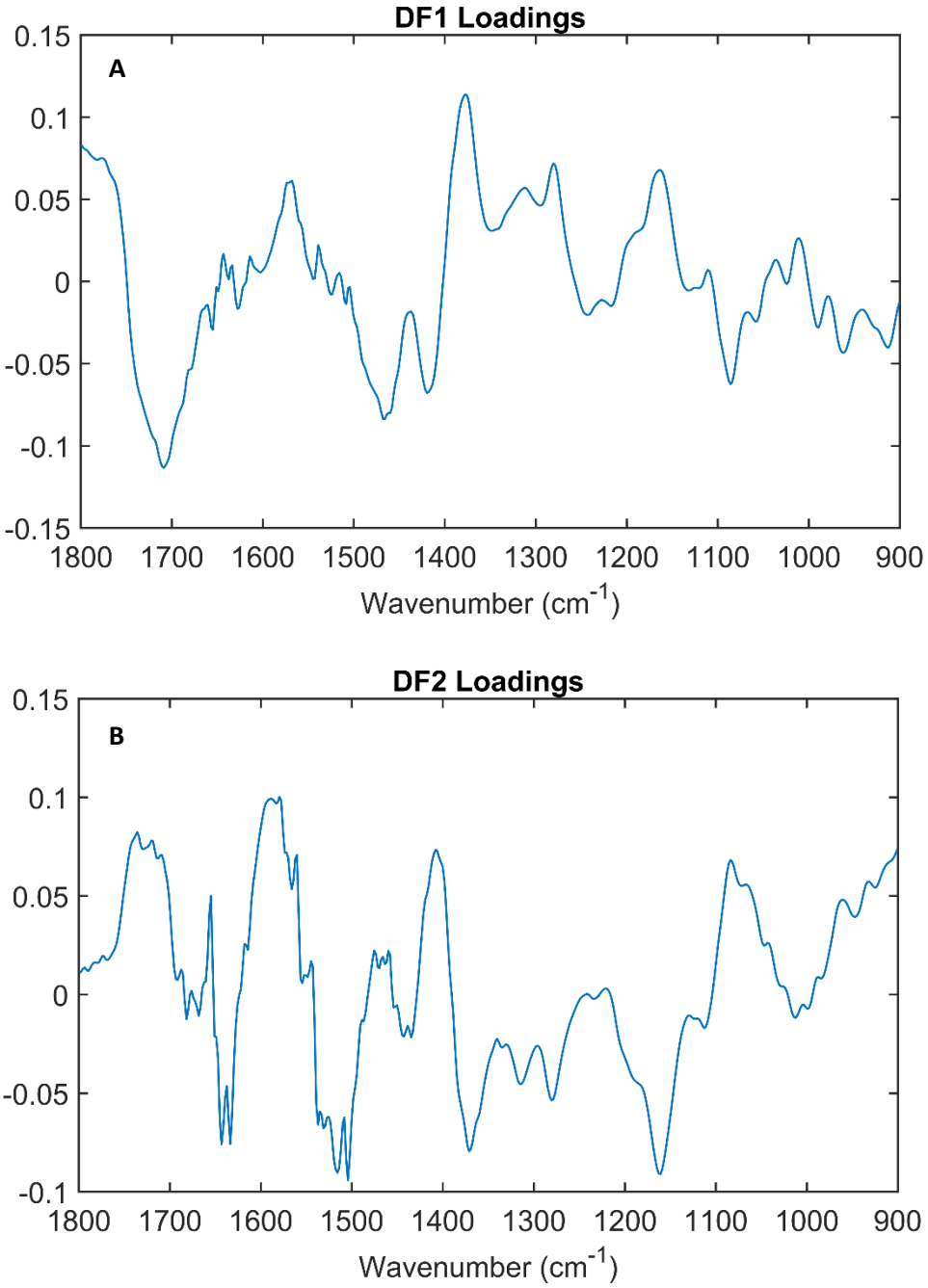

Figure 9: Figure showing a DF1 loadings plot (A) and a DF2 loadings plot (B) from data pre-processed using Method 4.

Table 3: Wavenumbers and corresponding peak assignments for PC1, PC3, DF1 and DF2 loadings from Method 4. [5,6,7,2,8,]

Wavenumber (cm^{-1})				Tentative peak assignment
PC2	PC3	DF 1	DF 2	
	980	1163	912, 931, 1161	vC-O, vC-C, vC-O-H, vC-O-C
	1007			Phenylalanine
	1039			Phenylalanine C-H in-plane
1064				vC-N, vC-C stretch
1084		1086	1065, 1084	vP=O (symmetric) of PO^-_2
		1280, 1311	1280	Amide III band of proteins
			1408	vC=O (symmetric) of COO^-
1385, 1392, 1441	1379, 1394	1377, 1419, 1460, 1468	1360, 1371, 1419	δCH_2 and δCH_3
1514			1516	Tyrosine
1504	1529, 1537		1504, 1531, 1537	Amide II
1577	1579	1568	1560, 1572, 1579, 1589	Aspartate and glutamate carboxylate stretching
	1633, 1641		1633, 1643	Amide I β-pleated sheet structures
	1651, 1658			Amide I α-helix
		1678		Amide I band components resulting from antiparallel Pleated sheets from β-turns of proteins
1712		1709	1709, 1720, 1736	vC=O

Derivative data

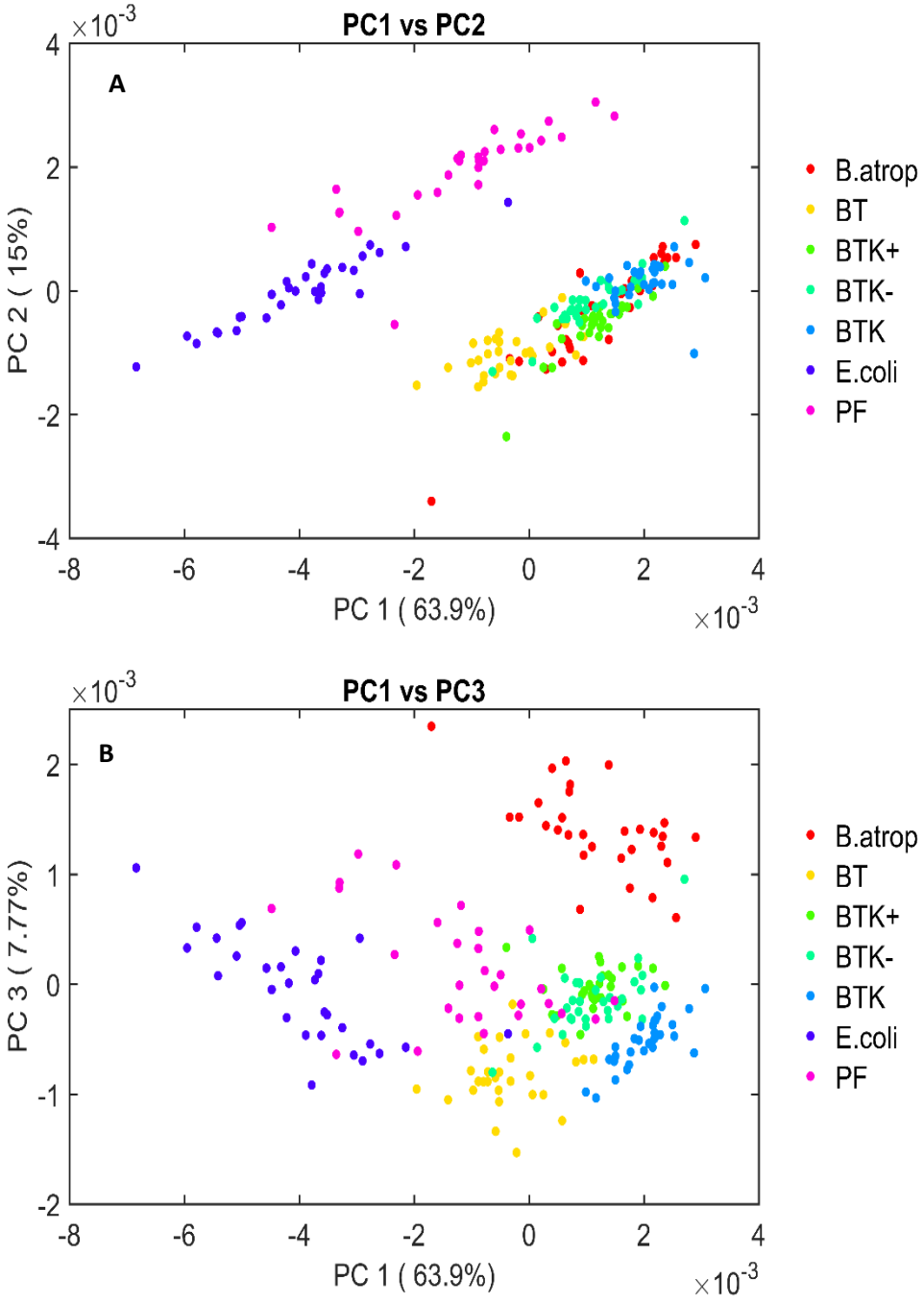

Figure 10: Figure showing a PCA plot using PC1 and PC2 (A) and a PCA plot using PC1 and PC3 (B) from 1st order derivative data.

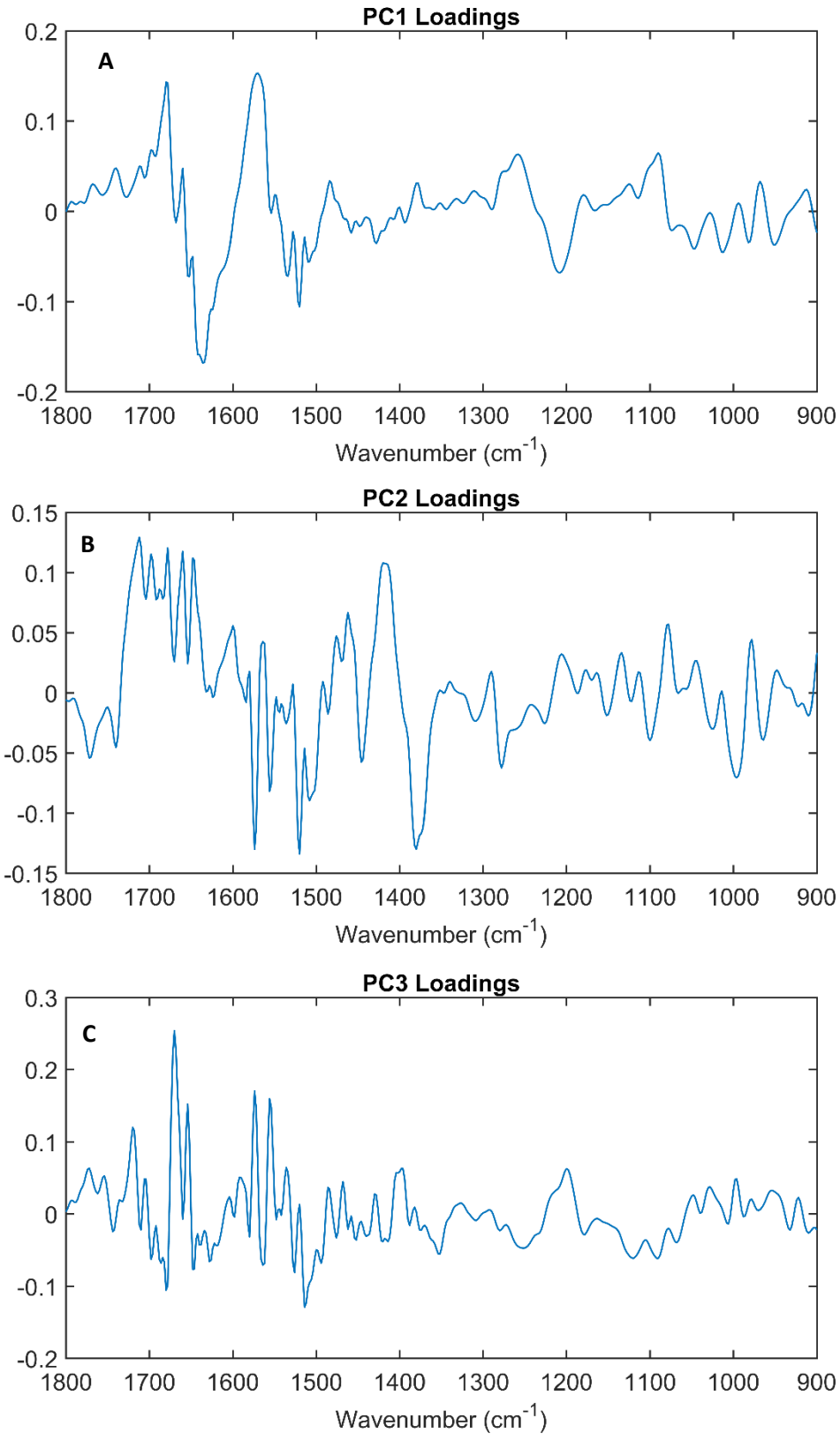

Figure 11: Figure showing a PC1 loadings plot (A) a PC2 loadings plot (B) and a PC3 loadings plot (C) from 1st order derivative data.

Table 4: Wavenumbers and corresponding peak assignments for PC1, PC2 and PC3 loadings from 1st order derivative data. [5,6,7,2,8,]

Wavenumber (cm^{-1})			Tentative peak assignment
PC1	PC2	PC3	
906, 930, 960, 975, 990, 998	904, 929, 954, 972, 984, 1036	917, 927, 938, 1012	vC-O, vC-C, vC-O-H, vC-O-C
		1003	Phenylalanine symmetric ring breathing
1080	1089, 1108, 1119	1058	vP=O (symmetric) of PO^-_2
	1126, 1144, 1158		vC-N, vC-C
		1187	Deoxyribose
1188, 1236	1217, 1286, 1296	1225, 1288, 1301, 1317	Amide III band of proteins
	1329	1338	CH_2 wagging
1388, 1470, 1492	1399, 1437, 1451, 1482, 1490, 1494	1380, 1385, 1391, 1409, 1425, 1433, 1464, 1472, 1480, 1489,	δCH_2 and δCH_3
1546		1519, 1522, 1531	Amide II
	1560, 1576	1560, 1569	Ring base
	1579, 1583	1579, 1583	vC-C of phenyl ring
1598, 1666, 1670		1610, 1650	Amide I
1657			Amide I α-helix
		1676	Amide I band components resulting from antiparallel Pleated sheets from β-turns of proteins
	1735	1701, 1709, 1713, 1739, 1747	vC=O

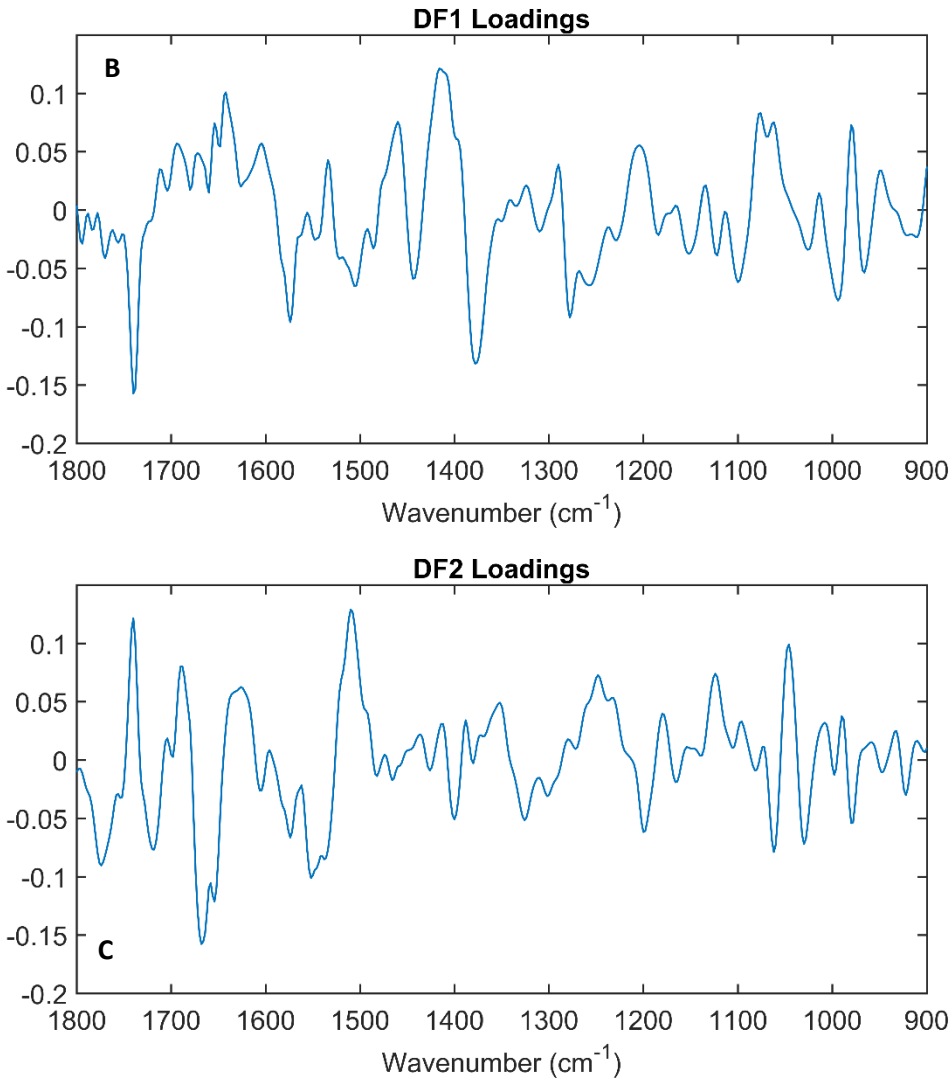

Figure 12: Figure showing a DF1 loadings plot (A) and DF2 loadings plot (B) from 1st order derivative data.

Table 5: Wavenumbers and corresponding peak assignments for DF1 and DF2 loadings from 1st order derivative data. [1,2,3,4,5]

Wavenumber (cm^{-1})		Tentative peak assignment
DF 1	DF 2	
904, 934, 956, 973, 985, 1010, 1017, 1163, 1169, 1190	916, 927, 942, 952, 971 985, 995, 1001, 1019, 1159, 1171	vC-O, vC-C, vC-O-H, vC-O-C
1042, 1087, 1220	1037, 1055, 1070, 1077, 1086, 1212	vP=O (symmetric) of PO^-_2
1130, 1140		vC-N, vC-C
1285, 1302, 1317, 1346	1342	Amide III band of proteins
1390, 1433, 1451, 1480	1392, 1407, 1422, 1430, 1454, 1471, 1476, 1486	δCH_2 and δCH_3
1529, 1539	1526	Amide II
1589		vC-C of phenyl ring
	1592	vC=N, vNH_2 adenine
	1601, 1611,1645	Amide I
	1678	Amide I antiparallel pleated sheets from β-turns of proteins
1718	1708, 1733, 1748	vC=O

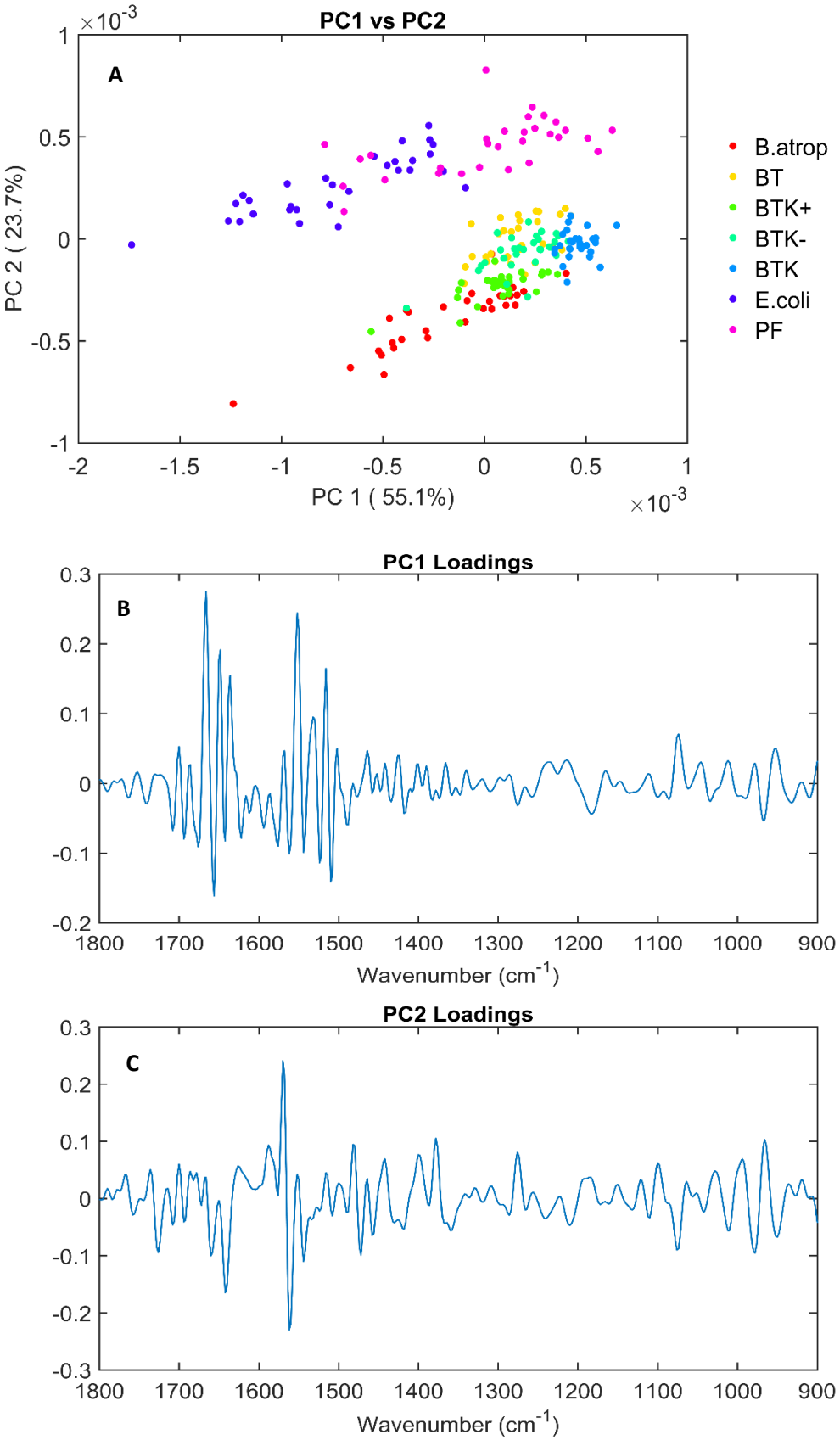

Figure 13: Figure showing a PCA plot using PC1 and PC2 (A) PC1 loadings plot (B) and a PC2 loadings plot (C) generated from the 2nd order derivative data.

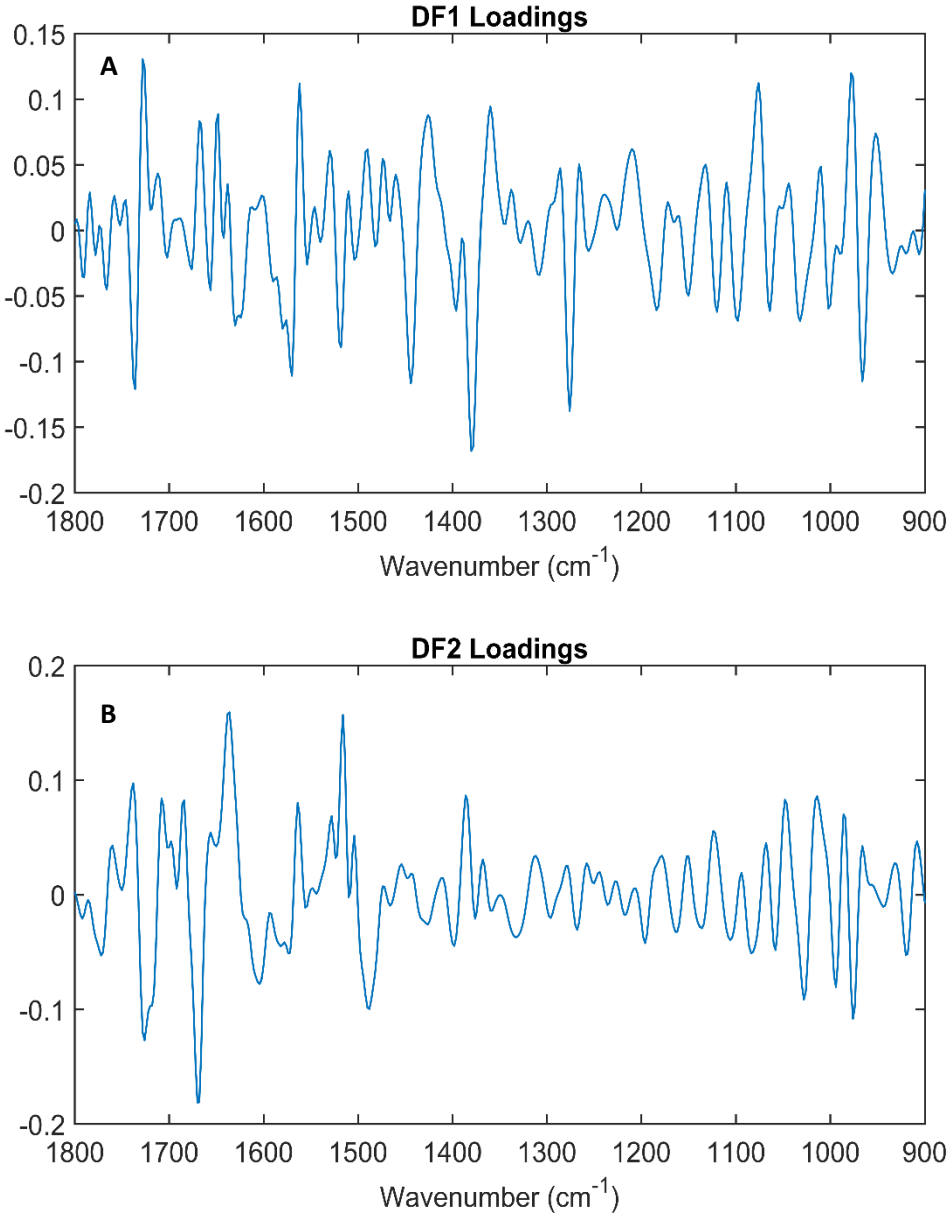

Figure 14: Figure showing a DF1 loadings plot (A) and a DF2 loadings plot (B) generated from the 2nd order derivative data.

Table 6: Wavenumbers and corresponding peak assignments for PC1, PC2, DF1 and DF2 loadings from 2nd order derivative data. [5,6,7,2,8,]

Wavenumber (cm^{-1})				Tentative peak assignment
PC1	PC2	DF 1	DF 2	
952, 968	950, 966, 978, 994, 1012	952, 966, 978, 1032	920, 976, 986, 994, 1014, 1028	vC-O, vC-C, vC-O-H, vC-O-C
		1002		Phenylalanine
		1120, 1132, 1150	1124	vC-N, vC-C
1074	1076, 1100	1064, 1076, 1098	1048, 1084	vP=O (symmetric) of PO^-_2
	1276	1184, 1210, 1266, 1276		Amide III band of proteins
1488	1360, 1378, 1400, 1418, 1442, 1458, 1472, 1482	1360, 1380, 1396, 1426, 1444, 1474, 1490	1386, 1488	δCH_2 and δCH_3
1516		1518	1516	Tyrosine
1502, 1510, 1524, 1532, 1544, 1552, 1562	1544, 1562	1504, 1530, 1562	1504, 1528, 1564	Amide II
1576	1570, 1588	1570, 1580	1574	Aspartate and glutamate carboxylate stretching
1594				vC-C of phenyl ring
1636			1636	Amide I β-pleated sheet structures
1622, 1642, 1648, 1656, 1666	1626, 1642, 1660	1624, 1630, 1648, 1668	1604, 1656, 1670	Amide I
1676, 1694			1684	Amide I antiparallel pleated sheets from β-turns of proteins
1700, 1708	1700, 1726, 1736	1728, 1736	1708, 1718, 1726, 1738, 1772	vC=O

Appendix C

DRIFTS Substrate Study

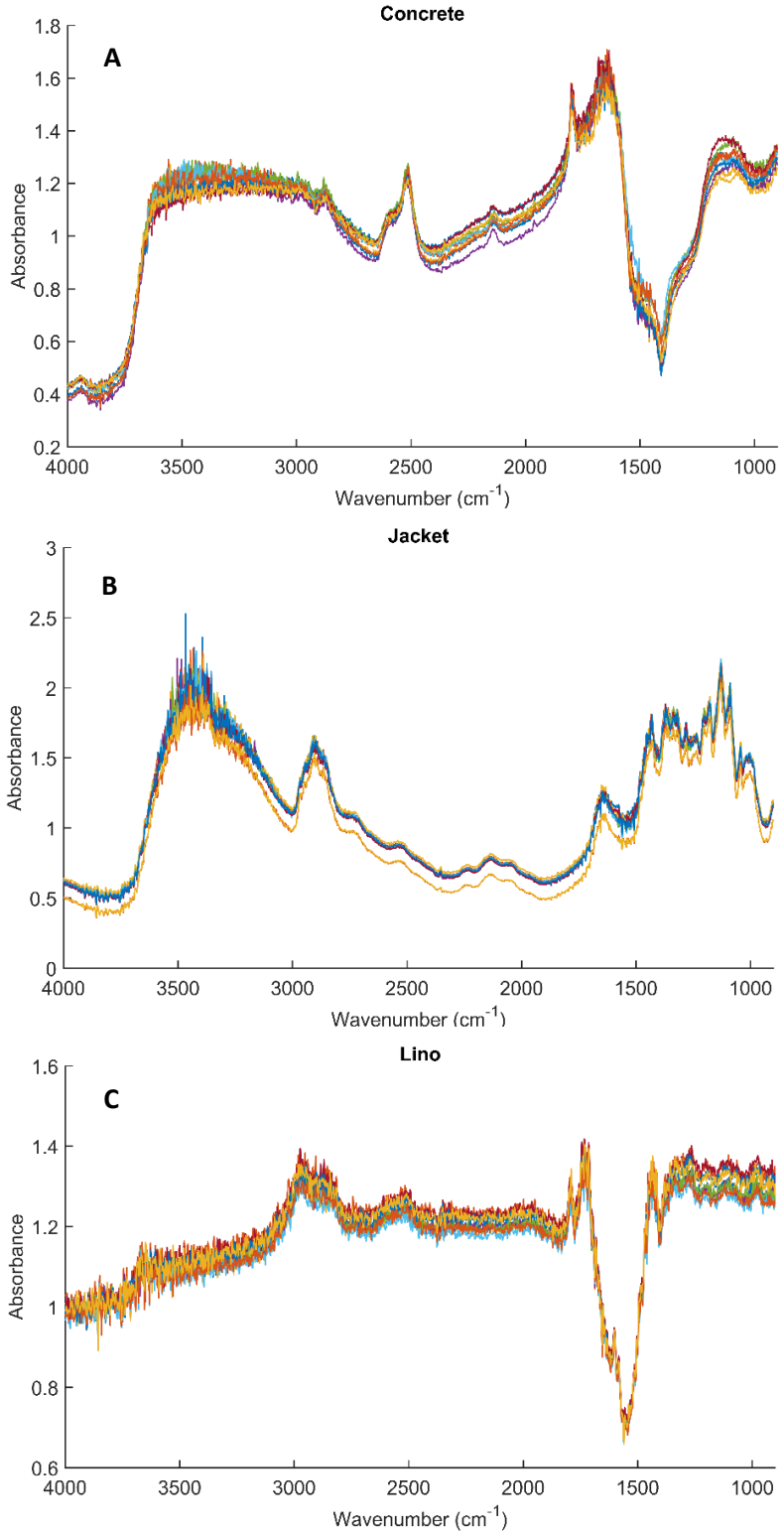

Figure 1: Spectra collected using DRIFTS showing an overlay of 10 spectra for Concrete (A), Jacket (B) and Lino (C).

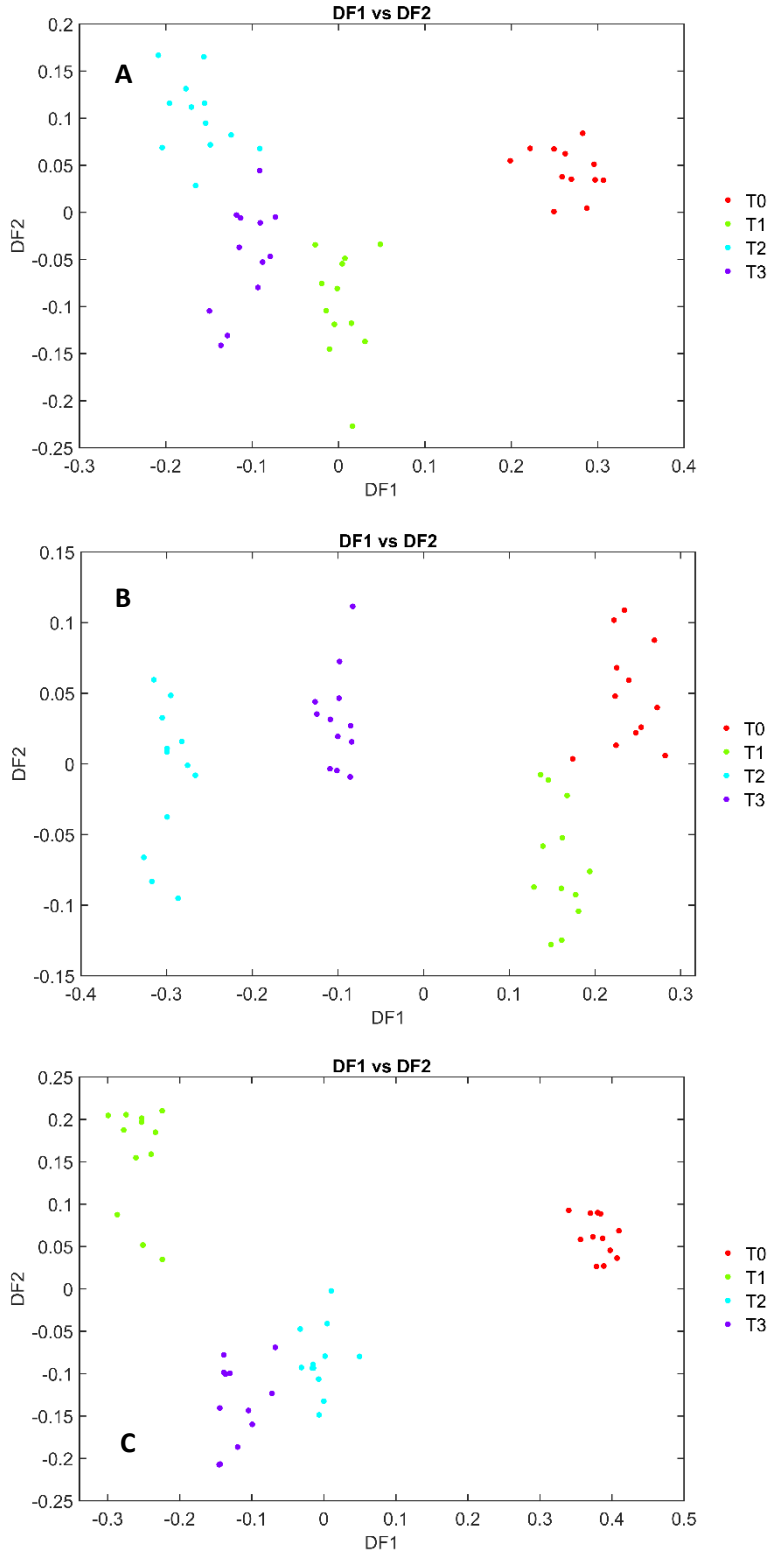

Figure 2: DFA plots showing separation of time point for spectra collected from BT deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

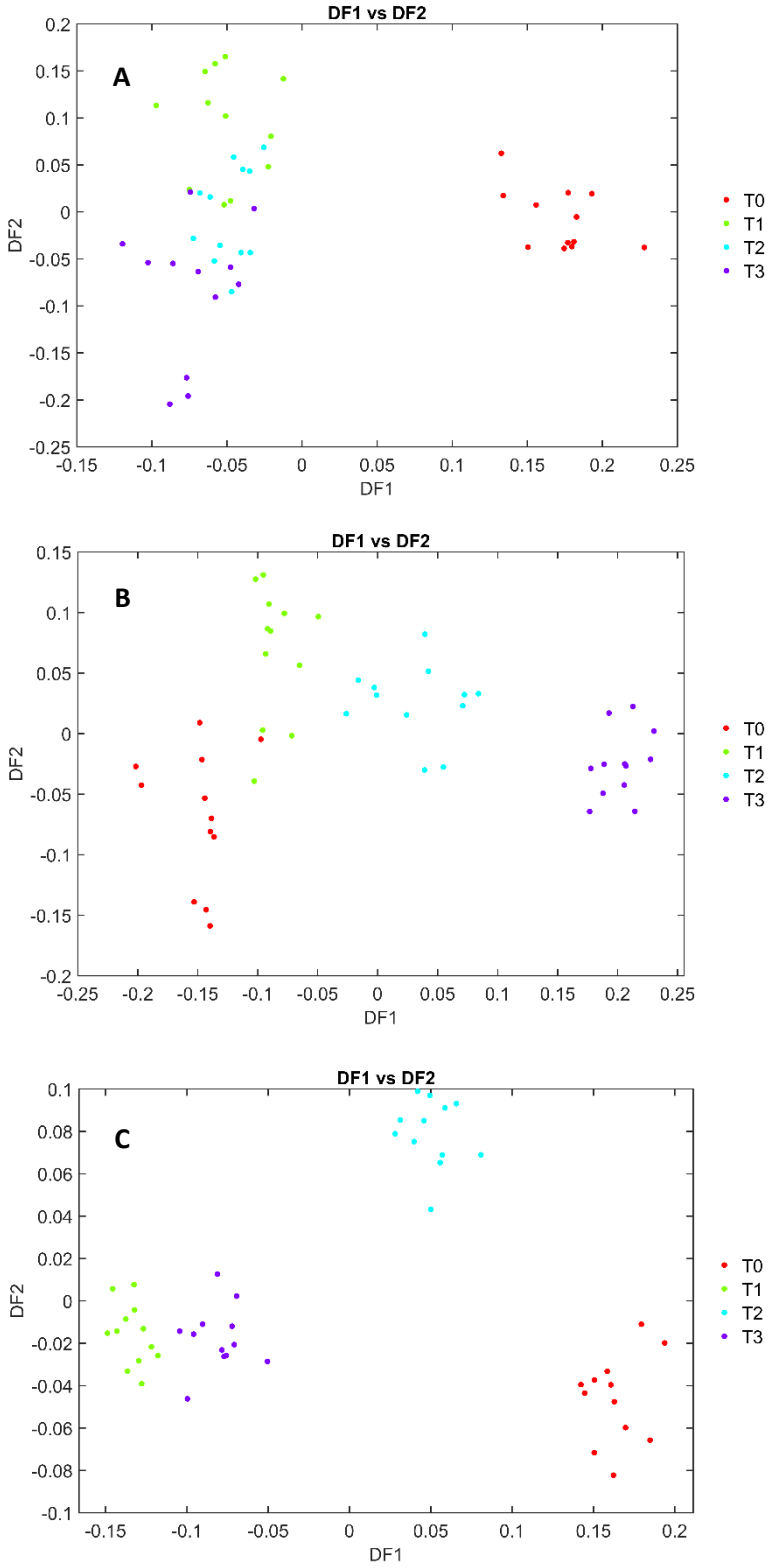

Figure 3: DFA plots showing separation of time point for spectra collected from BTK-deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

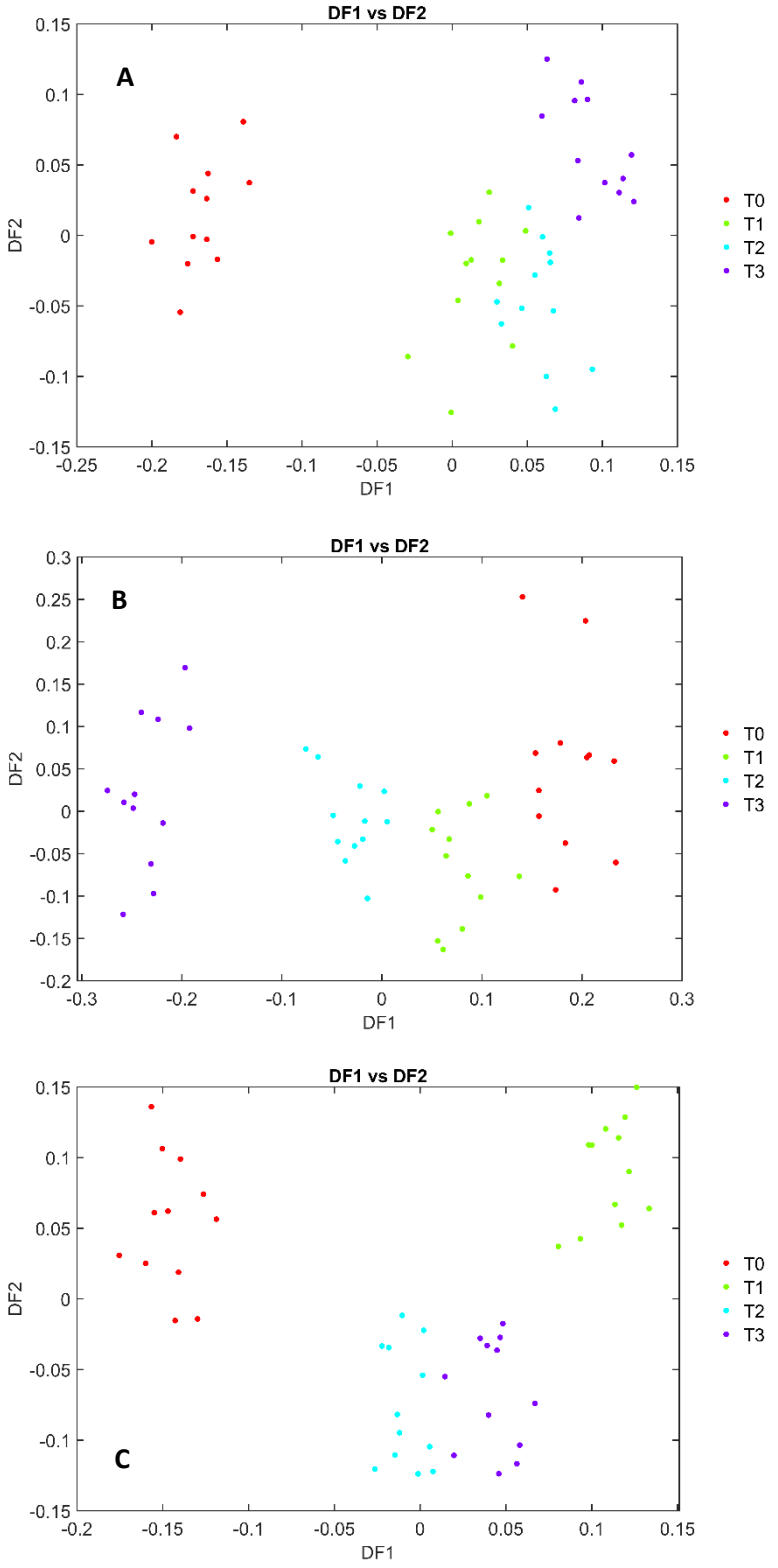

Figure 4: DFA plots showing separation of time point for spectra collected from BTK+ deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

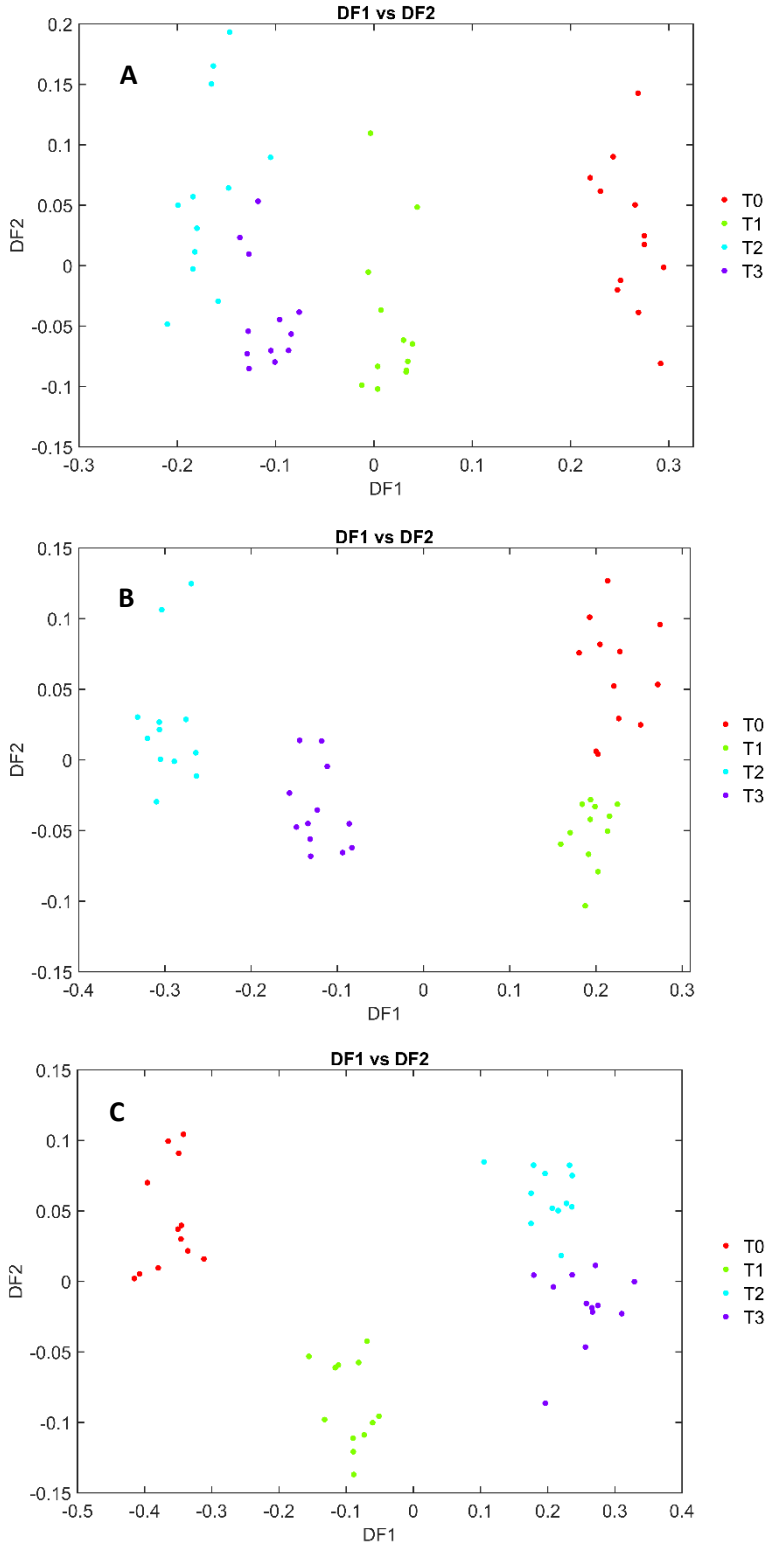

Figure 5: DFA plots showing separation of time point for spectra collected from E. coli deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

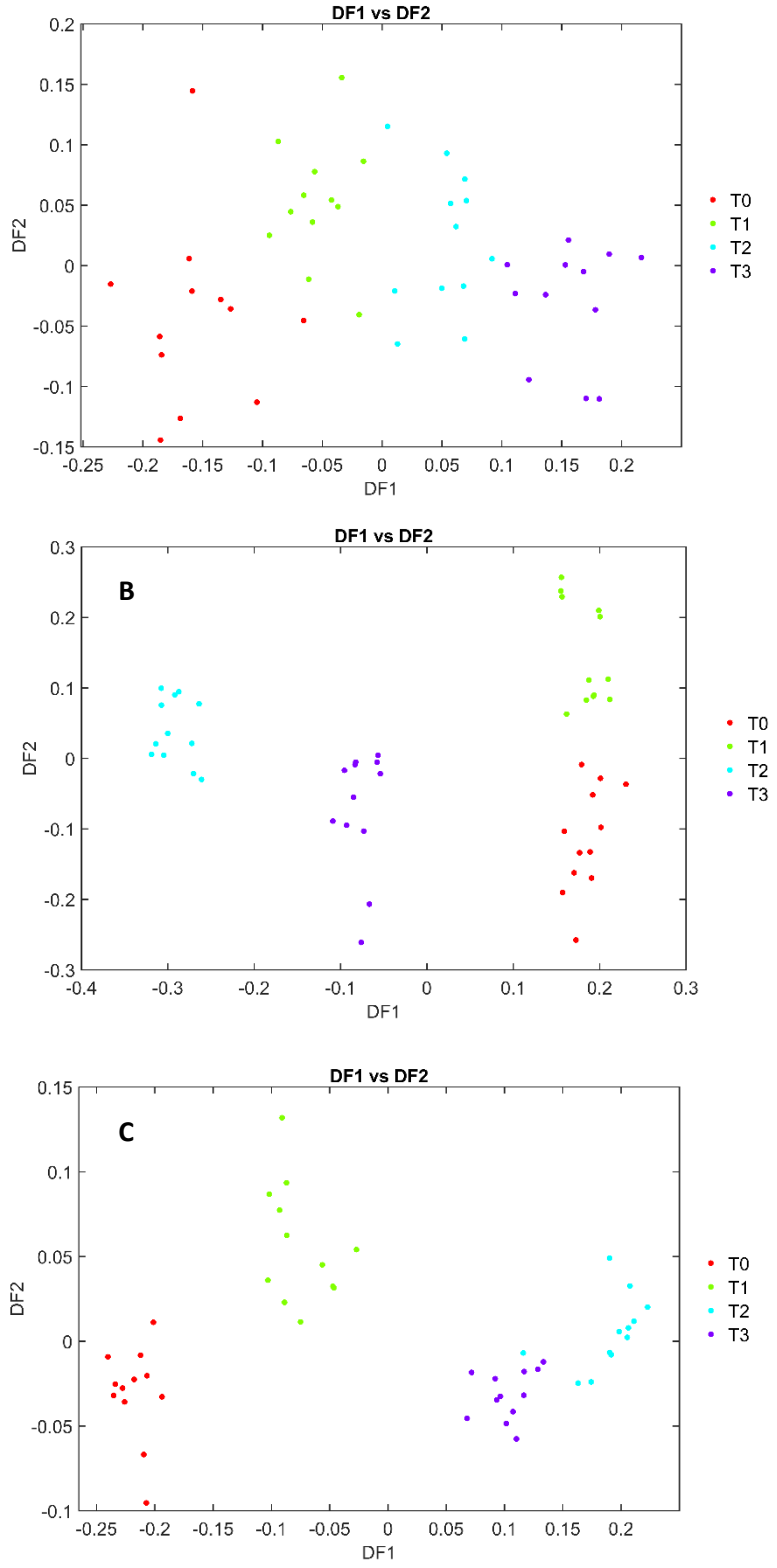

Figure 6: DFA plots showing separation of time point for spectra collected from PF deposited on boot during the 30°C study (A), the 37°C study (B) and the 49°C study (C).

Appendix D

High Throughput - Fourier Transform Infrared Environmental Conditioning Study

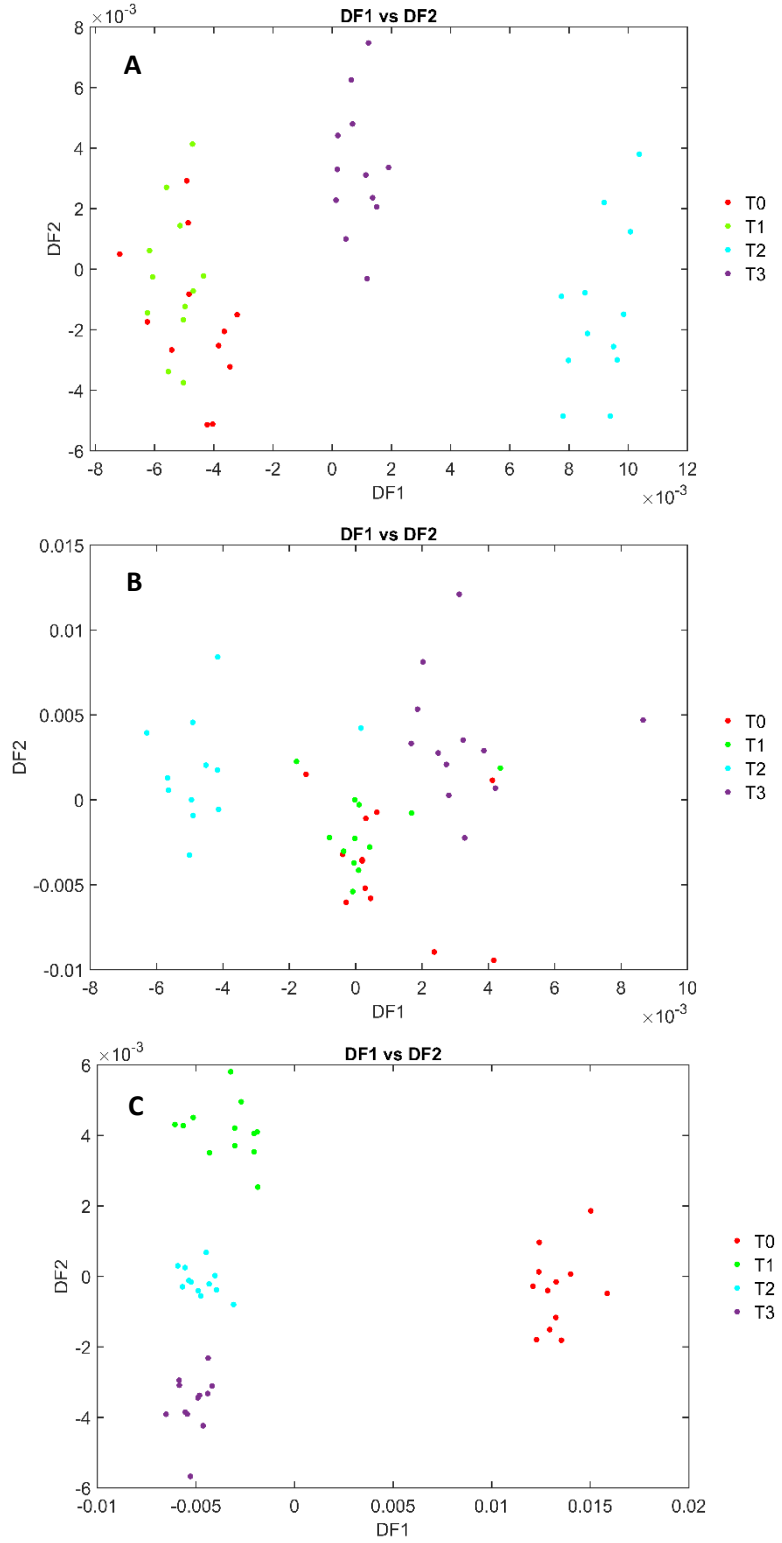

Figure 1: DFA plots showing separation based on the ageing of B. atrop across the three temperature studies showing T2 and T3 separating in the 30°C study (A), some separation of T2 in the 37°C study (B) and clustering and separation of all time points in the 49°C study (C).

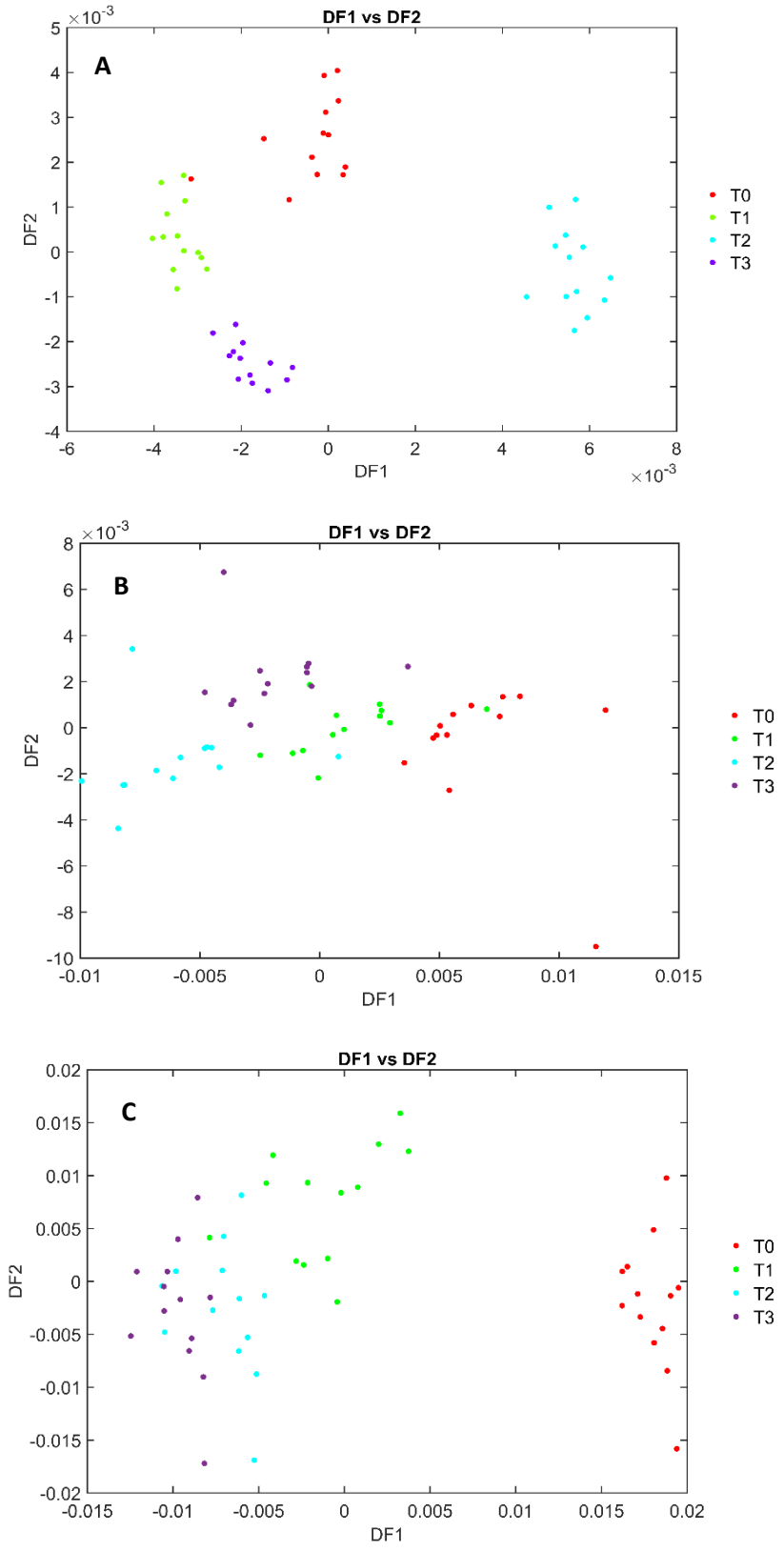

Figure 2: DFA plots showing separation based on the ageing of BTK across the three temperature studies showing clustering of all time points in the 30°C study (A), clustering and some separation of time points in the 37°C study (B) and separation of T0 and T1 in the 49°C study (C).

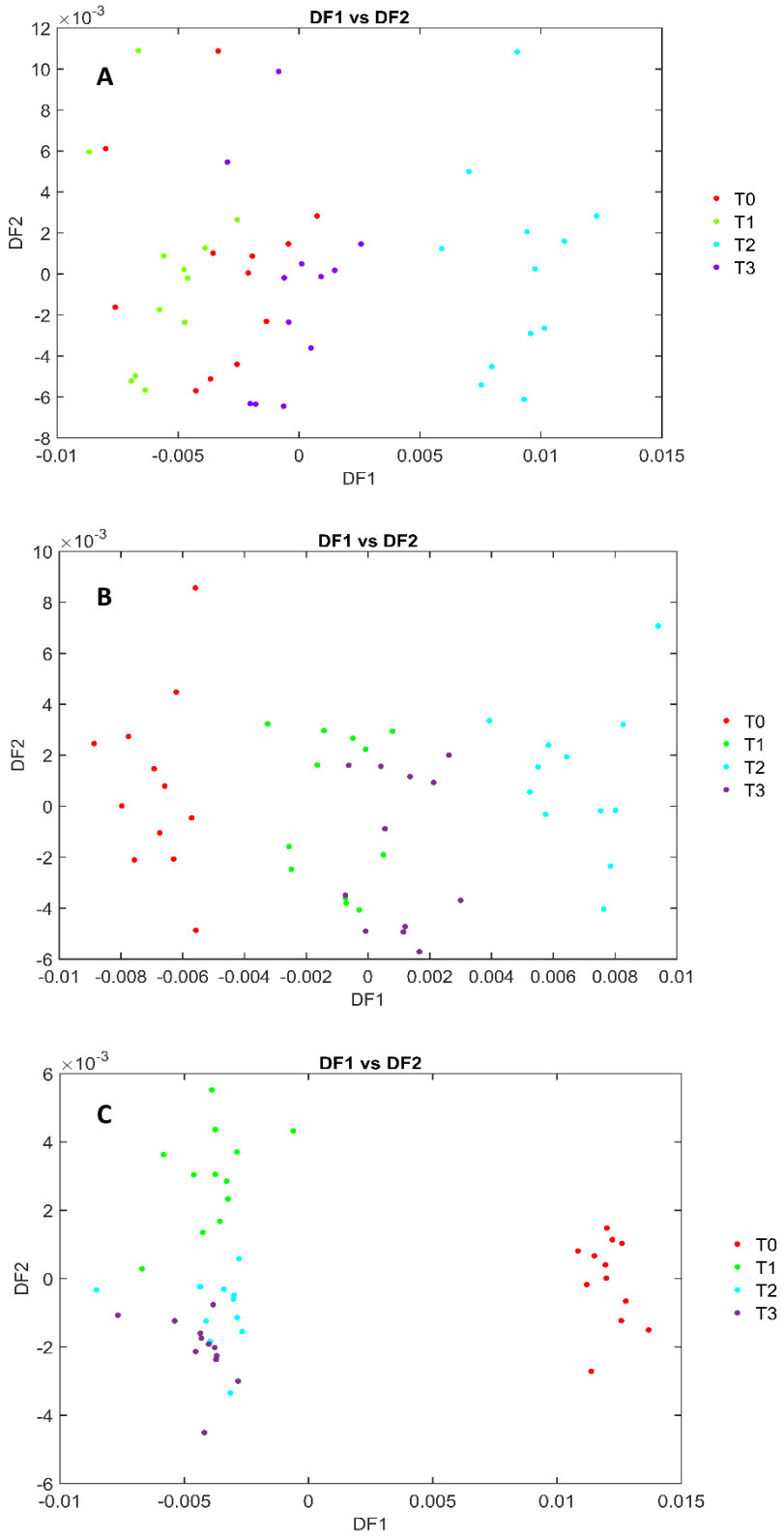

Figure 3: DFA plots showing separation based on the ageing of BTK+ across the three temperature studies showing separation of T2 in the 30°C study (A), some separation of time points in the 37°C study (B) and separation of T0 in the 49°C study (C).

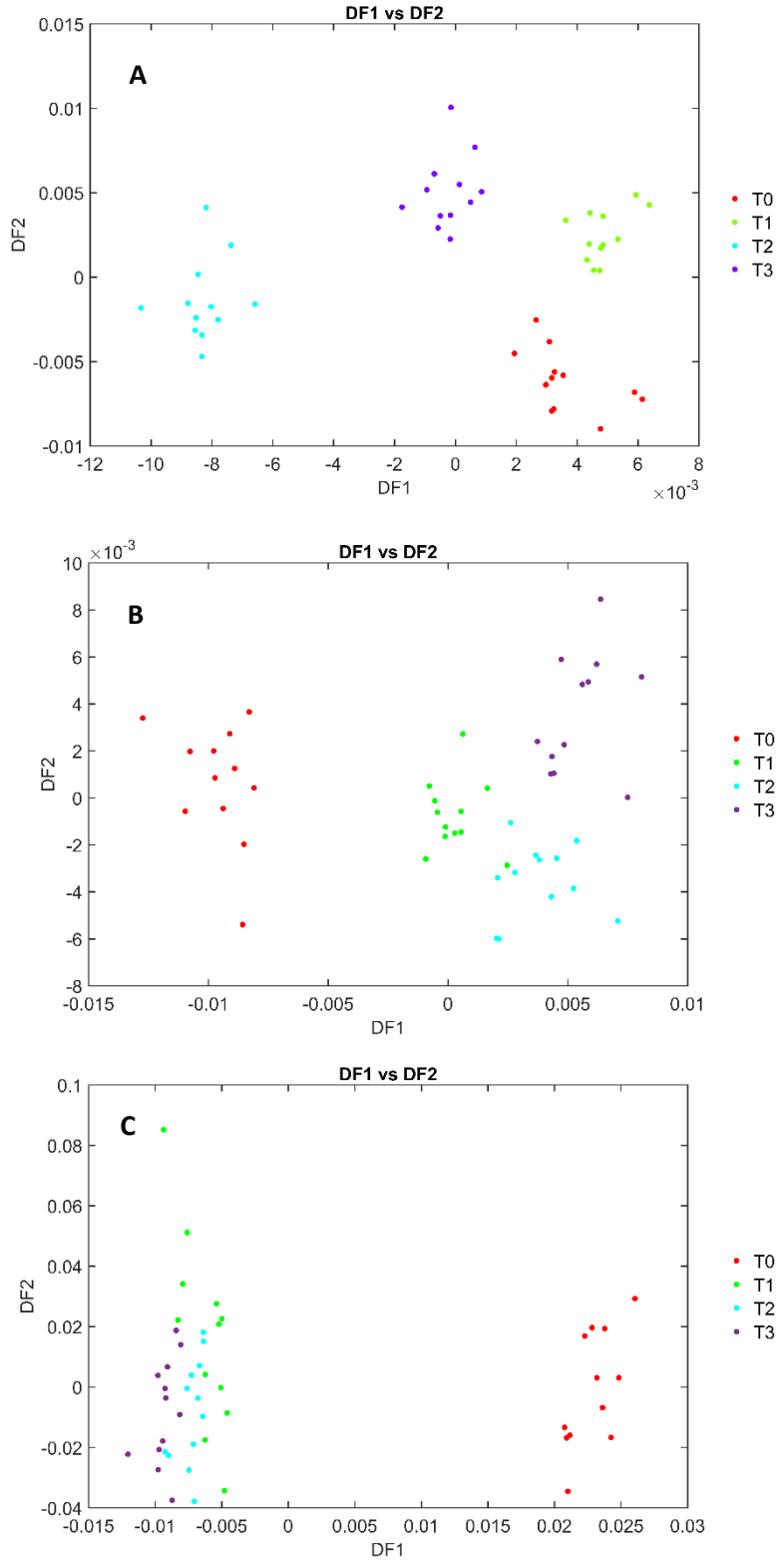

Figure 4: DFA plots showing separation based on the ageing of E. coli across the three temperature studies showing separation and clustering in the 30°C study (A), some separation and clustering of time points in the 37°C study (B) and separation of T0 in the 49°C study (C).

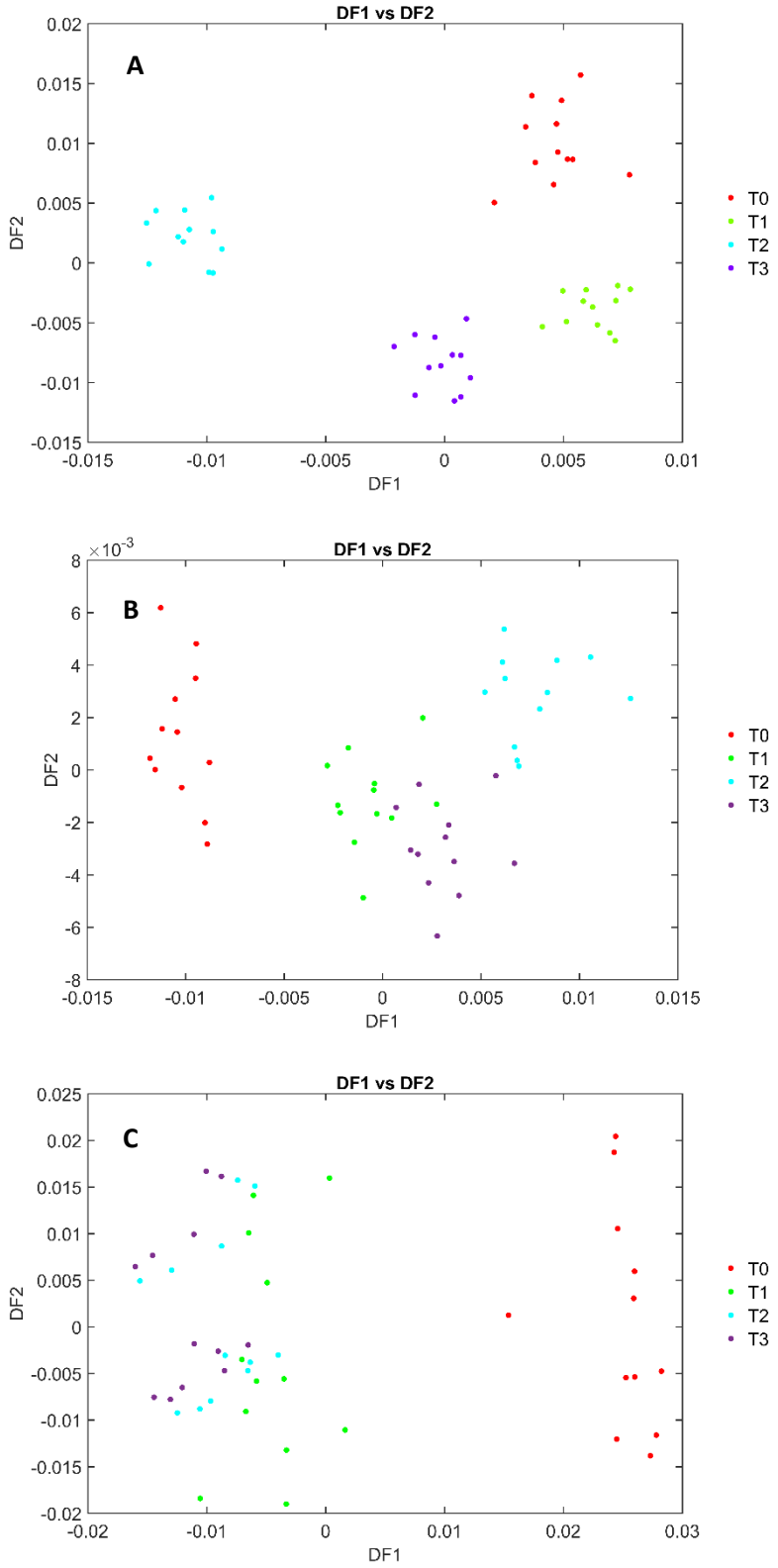

Figure 5: DFA plots showing separation based on the ageing of PF across the three temperature studies showing separation and clustering in the 30°C study (A), some separation and clustering of time points in the 37°C study (B) and separation of T0 in the 49°C study (C).

Multiblock PCA

30°C Study

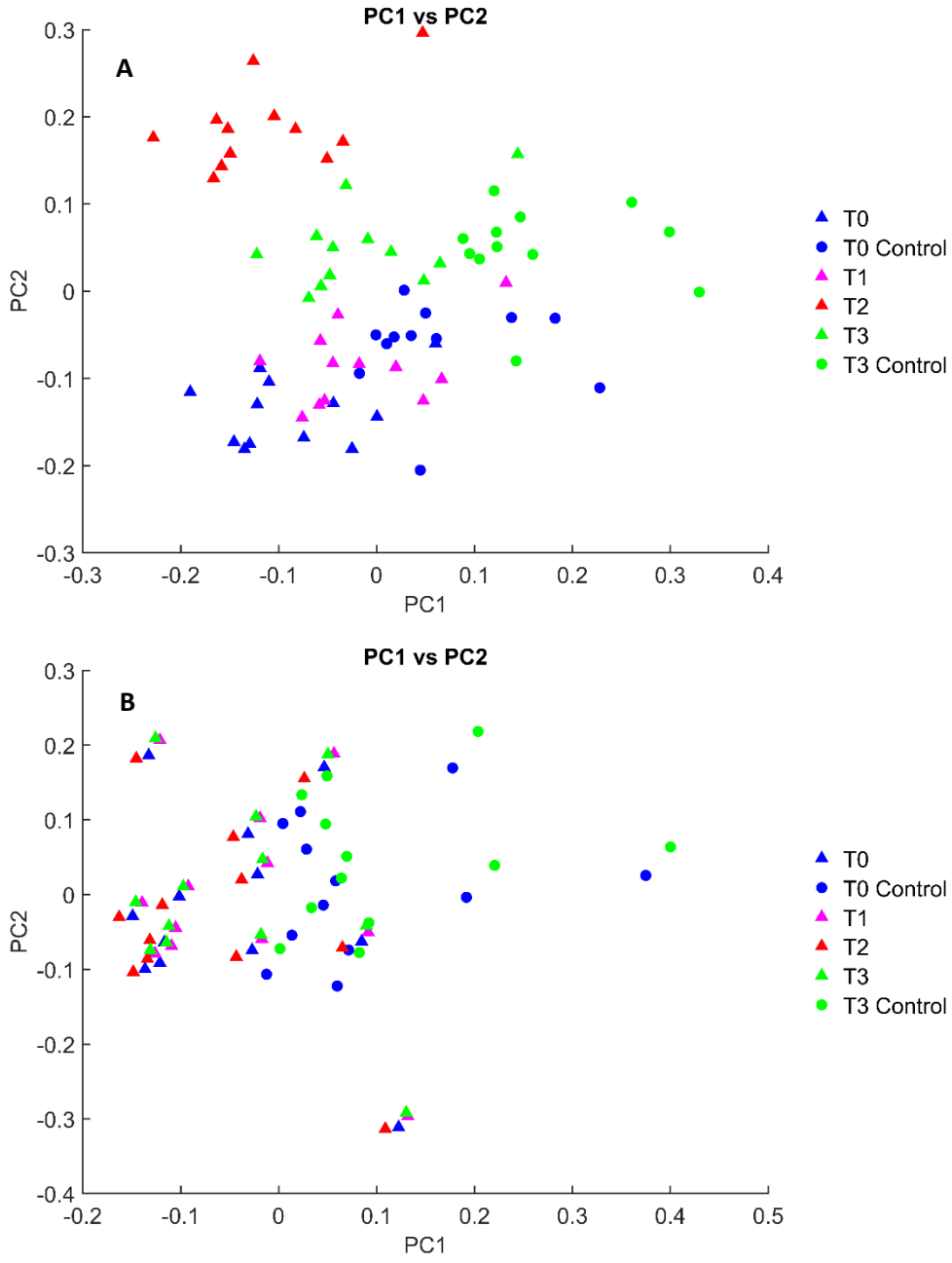

Figure 6: Multiblock PCA plots showing separation of the 30°C dataset that has been cut to the fingerprint region, based on time point not bacterial strain, showing the plot of PC1 vs PC2 for EMSC corrected data (A) and the PC1 vs PC2 plot for NR data (B).

37°C Study

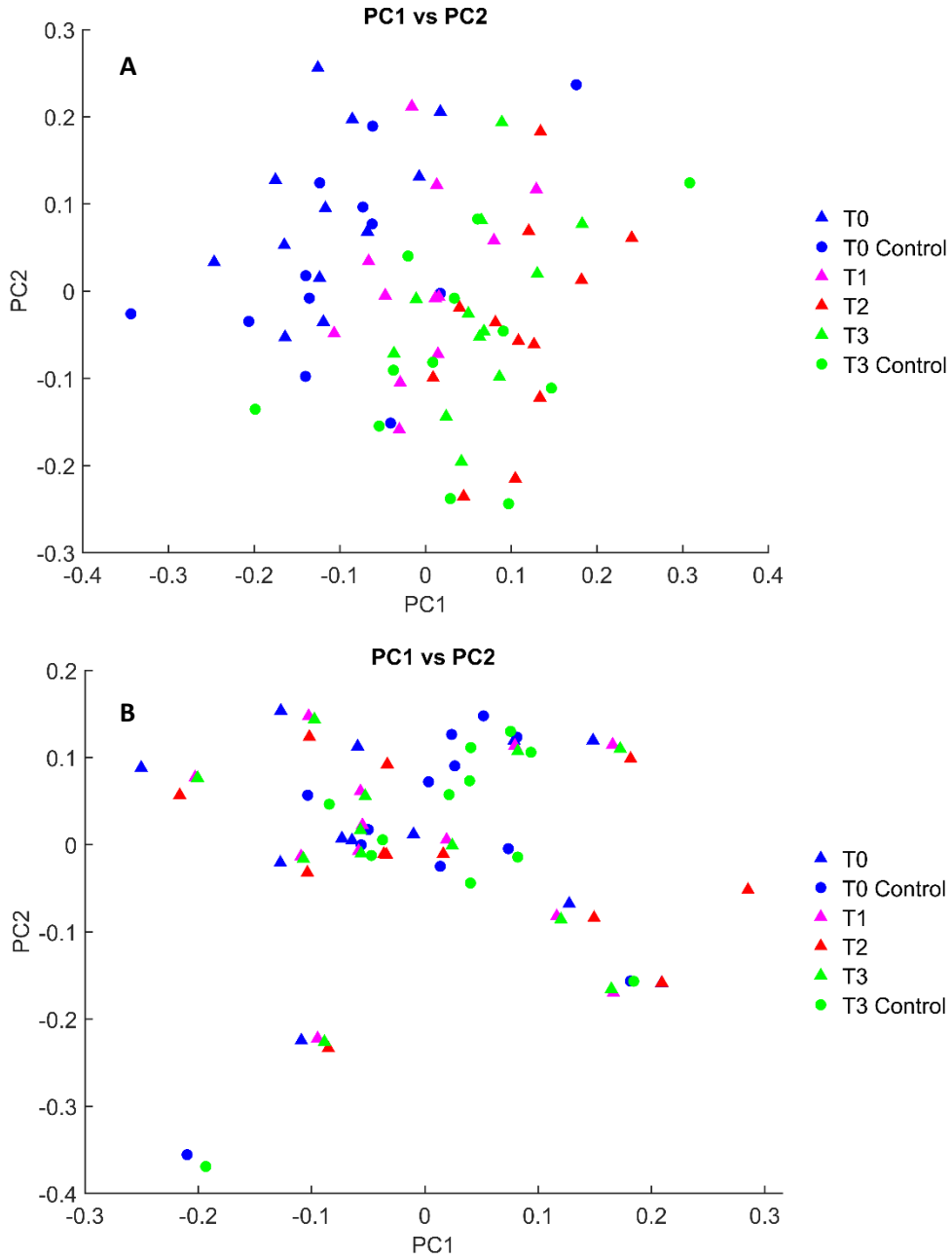

Figure 7: Multiblock PCA plots showing separation of the 30°C dataset that has been cut to the fingerprint region, based on time point not bacterial strain, showing the plot of PC1 vs PC2 for EMSC corrected data (A) and the PC1 vs PC2 plot for NR data (B).