Mycobacteriales: A Raman Characterisation Study and an Investigation into Diagnostic Avenues for Bovine and Human Tuberculosis

Michael Macgregor-Fairlie - [redacted]

Supervisor: Prof. Pola Goldberg-Oppenheimer, School of Chemical Engineering

Co-Supervisor: Dr. Timothy Overton, School of Chemical Engineering



UNIVERSITY OF BIRMINGHAM

A thesis submitted to fulfil the requirements for the degree of Doctor of Philosophy

College of Engineering and Physical Sciences,

School of Chemical engineering,

University of Birmingham,

June 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Acknowledgements:

First and foremost, I would like to extend my gratitude to my main supervisor, Prof. Pola Goldberg-Oppenheimer, whom, without, I would not be here right now, on the precipice of submitting the longest document I have ever written, and who has guided me throughout this entire process at the University of Birmingham, often in the face of great adversity in my personal life. Secondly, is my co-supervisor, Dr. Tim Overton, whose input has been critical to making sure that I organise my bacterial experiments in a methodical manner.

Next is Dr. Ed Bolt of the University of Nottingham, who enabled me to learn more about the potential of CRISPR. And, finally, for the academics, is Prof. Ilko Bald of the University of Potsdam, who enabled me to fulfil a decade long desire to return to Germany.

Of course, I would also like to thank the ANMSA Group and its members, but more specifically, Georgia, Martin, Emma Buchan and Emma McCarthy who have always been happy to provide a sense check and provide great conversation.

I'd also like to thank Dan Weston and Paulo De Gomes for enabling me to learn more about Python and programming in general. Without the help of both Paulo and Dan, none of statistical work in this thesis would have been possible.

On a personal level, I'd like to thank Anaïs and Max, who helped me find a home in Selly Oak during my time at the University of Birmingham. I'd like to thank Em, Sam, and Imogen for their support during trying times. And I'd like to thank all of my friends

whom I have been fortunate enough to meet during my time in Berlin and Birmingham.

I'd also like to thank the Birmingham Lions Ice Hockey Club, who have provided a welcoming environment for a mature PhD student, and the space to become a better ice hockey player (albeit, whilst injuring myself in the process).

I'd like to thank my Mum, Lillian, for her support throughout this entire process. I know she may not know what I'm doing, but I know she is still proud of me. And, lastly, I'd like to thank my late-father, Michael… who instilled a strong sense of ambition in me. He may no longer be with us, but I know he would be proud of me if he was still with us.

List of Publications:

Peer-Reviewed Journal Articles:

- M. Macgregor-Fairlie, D. Weston, P. De Gomes, J.J Rickard, P. Goldberg Oppenheimer; Hybrid use of Raman spectroscopy and artificial neural networks to discriminate *Mycobacterium bovis* BCG and members of the order Mycobacteriales. *PLOS one (awaiting peer review), 2023.* Pre-print DOI: https://doi.org/10.1101/2023.05.30.542797

- M. MacGregor-Fairlie, S. Wilkinson, G. S. Besra, P. Goldberg Oppenheimer; Tuberculosis diagnostics: overcoming ancient challenges with modern solutions. Emerg Top Life Sci 11 December 2020; 4 (4): 435–448. DOI: https://doi.org/10.1042/ETLS20200335

Posters:

- M. Macgregor-Fairlie, P. Goldberg Oppenheimer, T. Overton (September 2021) “Advancing the fight against an ancient foe – Raman profiling of Mycolic Acid” (Poster presentation). BioMedEng2021, Sheffield.

Abstract:

Despite the overshadowing threat of the COVID-19 pandemic, tuberculosis (TB) persists as a major global health challenge, responsible for an estimated 1.6 million deaths in 2021 and afflicting approximately one-third of the global population. The World Health Organization's urgent appeal for cost-effective, accessible, sensitive, and specific diagnostic tools underscores the pressing need for innovative solutions.

This thesis was initiated by exploring the diagnostic potential of two promising technological avenues: the gene-editing tool CRISPR-Cas12a and Raman spectroscopy. At the inception of the project, these technologies remained largely uncharacterised within the context of TB diagnosis.

The investigation conducted throughout this research subsequently revealed a significant gap in the current knowledge base pertaining to the application of Raman spectroscopy in TB diagnostics. Specifically, no prior research had considered the influence of the microbiome on the Raman spectra of test results.

This thesis delves deeper into this uncharted territory, focusing on the Raman characterisation of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG. Using *In silico* methods and examining the bacteria's cell wall lipids, the research illuminates the physical components contributing to the Raman spectra of these bacteria, unearthing new spectral elements in the process. Furthermore, deploying a collaboratively-developed Python package for multivariate statistical analysis, this study differentiates subtle variations between *M. bovis* BCG and other closely related bacteria using the bovine microbiome as an experimental model, an aspect not previously considered.

This research enhances our understanding of the Raman spectra of organisms similar to *M. bovis* BCG and, for the first time, takes into account the role of the microbiome. In the process, it broadens our knowledge of these complex organisms. By employing open-source statistical tools, the thesis also contributes to wider applications within the field.

Ultimately, this research hopes to facilitate the application of these diagnostic tools beyond tuberculosis, offering a potent weapon in humanity's ongoing battle against infectious diseases.

Contents:

Acknowledgements: i
List of Publications: iii
Peer-Reviewed Journal Articles: iii
Posters: iii
Abstract: iv
Contents: v
List of Figures: xiii
List of Tables: xviii
List of Abbreviations: xx
Chapter 1: 1
1. Introduction: 1
1.1. Aetiology of Tuberculosis: 1
1.2. Pathogenesis of Tuberculosis: 5
1.2.1. Initial Infection: 5
1.2.2. Latent Tuberculosis: 6
1.2.3. Active Tuberculosis: 7
1.2.4. Extrapulmonary Tuberculosis: 7
1.2.5. HIV-TB Co-infection: 9
1.3. Tuberculosis Treatment: 11
1.3.1. Standard Tuberculosis Treatment: 13
1.3.2. Drug-resistant Tuberculosis 14
1.3.3. Multi-drug Resistant Tuberculosis: 15
1.3.4. Extensively Drug Resistant Tuberculosis: 16
1.4. Tuberculosis Diagnostics: 17
1.4.1. Direct culture and observation: 18
1.4.2. Molecular and immunological detection: 20
1.4.3. Recent developments: 26
1.5. Epidemiology of Tuberculosis: 31
1.6. Bovine Tuberculosis: 35

1.7. TB Vaccination: 36
1.8. Mycobacterium Cell Wall Structure: 38
1.9. Scope of this thesis: 43
Chapter 2: 45
2.1. Bacteria, Molecules and Compounds: 45
2.1.1. *Mycobacterium bovis, Mycobacterium tuberculosis, & Mycobacterium smegmatis:* 46
2.1.2. *Mycoplasma bovis*: 47
2.1.3. *Mycoplasma hyosynoviae:* 48
2.1.4. *Pseudomona aeruginosa:* 49
2.1.5. *Pseudomonas fluorescens:* 49
2.1.6. *Bacillus cereus:* 50
2.1.7. *Escherichia coli:* 51
2.1.8. *Mannheimia haemolytica:* 52
2.1.9. *Moraxella catarrhalis:* 53
2.1.10. *Rhodococcus erythropolis:* 54
2.1.11. *Staphylococcus aureus*: 55
2.1.12. *Corynebacterium glutamicum:* 56
2.1.13. Chloramphenicol: 57
2.1.14. Ampicillin: 57
2.1.15. Kanamycin: 58
2.2. Sample Preparation Methodologies: 58
2.2.1. Bacterial Cultures & Purification Methodologies: 58
2.2.1.1. *Mycobacterium bovis BCG & Mycobacterium tuberculosis 7H9:* 58
2.2.1.2. *Mycobacterium smegmatis:* 58
2.2.1.3. *Mycoplasma bovis:* 59
2.2.1.4. *Mycoplasma hyosynoviae:* 59
2.2.1.4. *Pseudomonas fluorescens:* 60
2.2.1.5. *Pseudomona aeruginosa:* 60
2.2.1.6. *Bacillus cereus:* 60

2.2.1.7. *Escherichia coli:* 60
2.2.1.8. *Mannheimia haemolytica:* 61
2.2.1.9. *Moraxella catarrhalis:* 61
2.2.1.10. *Rhodococcus erythropolis:* 61
2.2.1.11. *Staphylococcus aureus:* 61
2.2.1.12. *Corynebacterium glutamicum:* 62
2.2.1.13. Bacterial preparation: 62
2.3. Molecular Methods: 62
2.3.1. Bacterial Transformation: 62
2.3.2. Plasmid Purification: 63
2.3.3. Digestion, Phosphorylation & Ligation: 63
2.3.4. Gel Electrophoresis & extraction: 63
2.4. *In silico* Methods: 64
2.4.1. Avogadro: 64
2.4.2. ORCA: 65
2.4.3. BlueBEAR: 65
2.5. Analytical Methods: 65
2.5.1. Optical Trapping: 65
2.5.2. Raman Spectroscopy: 67
2.5.3. WiRE: 69
2.6. Software for the data analyses: 70
2.7. Statistical methods used for the data analyses: 70
2.7.1. Principal Component Analysis (PCA): 70
2.7.2. PCA Quadrant Analysis: 71
2.7.3. Self-Organised Mapping (SOM): 71
2.7.4. Self-Organised Mapping Discrimination Index (SOMDI): 72
2.7.5. Self-Optimising Kohohen Index Network (SKiNET): 72
2.7.6. Principal Component Analysis – Linear Discriminant Analysis (PCA-LDA): 72
2.7.7. Confusion Matrix: 73

2.7.8. Two-Way Analysis of Variance (ANOVA): 73
Chapter 3: 75
3. *In silico* analysis of the Mycobacterial cell wall: 75
3.1. Introduction: 75
3.1.1. European Molecular Biology Laboratories: 76
3.1.2. Avogadro: 77
3.1.3. BlueBEAR: 78
3.1.4. ORCA: 79
3.2. Methods: 80
3.2.1. European Molecular Biology Laboratories: 80
3.2.2. Avogadro: 84
3.2.3. BlueBEAR/ANSMA group computer: 84
3.2.4. ORCA: 85
3.3. Results: 85
3.3.1. (2R)-2-[(1R)-1-Hydroxy-12-2-[14-(2-Icosylcyclopropyl)-Tetradecyl]-Cyclopropyl-Dodecyl]-Hexacosanoate: 85
3.3.2. (2R)-2-[(1R)-1-hydroxy-12-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyldodecyl]hexacosanoic acid: 86
3.3.3. (2R)-2-[(1R)-1-hydroxy-14-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyltetradecyl]hexacosanoic acid: 87
3.3.4. (2R)-2-[(1R)-1-hydroxy-16-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylhexadecyl]hexacosanoate: 88
3.3.5. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate: 89
3.3.6. (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoic acid: 90
3.3.7. (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoate: 91
3.3.8. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoic acid: 92
3.3.9. (2R)- 2- [(1R)- 1- hydroxy- 18- 2- [10- (2-octadecylcyclopropyl)decyl]cyclopropyloctadecyl]tetracosanoic acid: 93

3.3.10. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate: 94
3.3.11. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid: 95
3.3.12. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid: 96
3.3.13. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoate: 97
3.3.14. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoic acid: 98
3.3.15. (2R)-2-[(1R)-1-hydroxy-22-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyldocosyl]hexacosanoate: 99
3.3.16. (2R)-2-[(1R)-14-2-[(15Z)-hexatriacont-15-en-1-yl]cyclopropyl-1-hydroxytetradecyl]hexacosanoate: 100
3.3.17. α-mycolic acid: 101
3.3.18. α,α'-trehalose 6-mycolate: 102
3.3.19. β-D-Araf-(1→2)-α-D-Araf-(1→3)-α-D-Araf-(1→5)-α-D-Araf-OMe: 103
3.3.20. Keto Mycolic Acid: 103
3.3.21. Methoxy Mycolic Acid: 104
3.4. Discussion: 107
Chapter 4: 112
4. Optical trapping & Raman analysis of Mycobacterial cell wall lipids: 112
4.1. Introduction: 112
4.2. Methods: 114
4.2.1. Reagents: 114
4.2.2. Lipid extraction: 114
4.2.3. Raman Spectroscopy: 116
4.2.4. Optical trapping: 116
4.2.5. Principal Component Analysis: 116
4.2.6. Code Availability: 117
4.3. Results: 118
4.3.1. Average Spectra: 118

4.3.1.1. Polar lipid extract: 118
4.3.1.2. Non-Polar lipid extract: 119
4.3.2. PCA Quadrant analysis & loadings: 121
4.4. Discussion: 125
Chapter 5: 130
5. Raman analysis of the Respiratory Tract Microbiome of *Bos taurus*: 130
5.1. Introduction: 130
5.2. Methods: 135
5.2.1. Reagents: 135
5.2.2. Raman Spectroscopy: 136
5.2.3. Principal Component Analysis: 136
5.2.4. Principal Component Analysis & Linear Discriminant Analysis: 136
5.2.5. Self-Organising Maps: 137
5.2.6. Raman Peak Analysis: 138
5.2.7. Code Availability: 138
5.3. Results: 138
5.3.1. Average Raman spectra: 138
5.3.2. PCA analysis: 141
5.3.3. PCA-quadrant analysis: 143
5.4. Discussion: 153
Chapter 6: 158
6. Detection of *Mycobacterium bovis* BCG in an Artificial Sputum 158
6.1. Introduction: 158
6.2. Methods: 160
6.2.1. Reagents: 160
6.2.2. Raman Spectroscopy: 161
6.2.3. Principal Component Analysis: 161
6.2.4. Code Availability: 162
6.3. Results: 163
6.3.1. Average spectra: 163

6.3.2 PCA of spectra: 165
6.4. Discussion: 167
Chapter 7: 170
7. Discussions, Conclusions & Outlook 170
Appendix: 176
A.1. Use of CRISPR-Cas for the detection of Tuberculosis: 176
A.1.1 Introduction: 176
A.1.1.1 CRISPR-Cas: 178
A.1.1.1.1. gRNA: 179
A.1.1.1.2. Cas9: 179
A.1.1.1.4. Cas12a: 183
A.1.1.1.5. Other Cas enzymes: 184
A.1.1.1.6. Transfection methods: 184
A.1.1.2. Molecular detection with CRISPR 187
A.1.1.3. Genetic Elements for Molecular Detection: 189
A.1.1.3.1. Bacterial and Viral Bioinformatics Resource Center (BV-BRC): 192
A.1.1.3.2. CHOPCHOP: 192
A.1.1.3.3. Basic Local Alignment Search Tool (BLAST): 193
A.1.1.3.4. Addgene: 193
A.1.2. Methods: 193
A.1.2.1. gRNA design: 194
A.1.2.2. Bacterial culture & transformation: 195
A.1.2.3. Vector digest & ligation: 196
A.1.3. Results: 196
A.1.4. Discussion: 197
A.2. SEM preparation protocol for bacteria: 199
A.3. Genetic elements identified in *M. tuberculosis* H37Rv: 200
A.4. pCRISPR, pCAS9, and pMA-RQ plasmid maps: 201
A.5. Oligonucleotide sequences ordered from IDT: 202
A.6. Example of BlueBEAR input file: 202

A.7. Example of .BASH script submitted to BlueBEAR: 207
A.8. Raman spectra of individual bacteria and peak assignment: 207
A.8.1. *S. aureus*: 208
A.8.2. *R. erythropolis*: 209
A.8.3. *P. fluorescens*: 210
A.8.4. *P. aeruginosa*: 211
A.8.5. *M. hyosynoviae*: 212
A.8.6. *Mycoplasma bovis*: 213
A.8.7. *M. bovis* BCG: 215
A.8.8. *M. catarrhalis*: 216
A.8.9. *M. haemolytica*: 218
A.8.10. *E. coli*: 220
A.8.11. *C. glutamicum*: 221
A.8.12. *B. cereus*: 222
A.9 Optical Density of bacteria present in bacterial mix: 224
References: 225

List of Figures:

Figure 1 - The Mycobacterium tuberculosis bacterial complex. Created with BioRender.com 2
Figure 2 - Symptoms of pulmonary TB. Created with BioRender.com 4
Figure 3 - Estimated TB incidence in 2021, for countries with at least 100 000 incident cases (1) 32
Figure 4 - Global number of deaths by cause, 2019 (149) 34
Figure 5 -The cell envelope layers are populated by proteins that build and modulate the cell envelope 40
Figure 6 - Generalised formula for mycolic acids 42
Figure 7 - Generalised formulas for the Mycolic acids in the cell walls of *Mycobacterium sp*. (155) 42
Figure 8 - Scanning Electron Micrograph of *M. bovis* BCG taken at 7,015x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the entire picture. An individual cell is highlighted. 46
Figure 9 - Scanning Electron Micrograph image of *Mycoplasma bovis* taken at 20,000x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the centre of the picture. Individual cells can be readily visualised. 47
Figure 10 - Scanning Electron Micrograph image of *P. aeruginosa* isolates attaching to glass surfaces. Isolates can be observed in the top two boxes, whilst more clustered cells can be observed in C and D. Magnification was not stated in the paper. Taken from Deligianni *et al.* (181) 49
Figure 11 - Scanning Electron Micrograph image of *B. cereus* taken at 5,000x magnification. Individual cells can be readily visualised. Taken from Filho *et al.* (182) 50
Figure 12 - Scanning Electron Micrograph image of *E. coli* taken at 10,000x magnification. A cluster of bacteria dominates the picture. Individual cells can be readily visualised. Taken from Krishnakumar (183) 51
Figure 13 - Scanning Electron Micrograph image of *M. haemolytica* taken at 8,000x magnification. A cluster of bacteria dominates the picture. Note the extracellular matrix that the bacteria have generated. Individual cells can be readily visualised. Two cells undergoing fission are highlighted. Taken from Haig (184) 52
Figure 14 - Scanning Electron Micrograph of *M. catarrhalis*, taken at 2000x magnification. A vast biofilm forms in the centre of the micrograph. The arrow present points to an individual cell. Taken from Tan *et al.* (185) 53
Figure 15 - Scanning Electron Micrograph of *R. erythropolis* taken at 2,736x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the entire picture. An individual cell is highlighted in the picture, with an arrow emphasising the cell. 54

Figure 16 - Scanning Electron Micrograph of *S. aureus* taken at 10,000x magnification. A cluster of bacteria dominates the e picture. An individual cell is highlighted in the picture. Taken from Zeng *et al.* (186).....55
Figure 17 - Scanning Electron Micrograph of *C. glutamicum* taken at 1,350x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A biofilm of bacteria dominates the picture. A small cluster of bacteria undergoing fission is highlighted, with an arrow pointing to them. 56
Figure 18 - Schematic of Optical trapping within the context of this project. Created by Biorender.com 66
Figure 19 – An illustration of the theory or Raman spectroscopy. For reference, E_R = the energy of a Rayleigh scattered photon, E_S = the energy of a Stokes photon E_{AS} denotes the energy of an anti-Stokes photon and E_{vib} = the energy involved in a molecular vibrational transition. Created with Biorender.com 67
Figure 20 - *In silico* Raman spectra of (2R)-2-[(1R)-1-Hydroxy-12-2-[14-(2-lcosylcyclopropyl)-Tetradecyl]-Cyclopropyl-Dodecyl]-Hexacosanoate 85
Figure 21 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-12-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyldodecyl]hexacosanoic acid 86
Figure 22 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-14-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyltetradecyl]hexacosanoic acid 87
Figure 23 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-16-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylhexadecyl]hexacosanoate.....88
Figure 24 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate 89
Figure 25 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoic acid 90
Figure 26 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoate 91
Figure 27 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoic acid 92
Figure 28 - *In silico* Raman spectra of (2R)- 2- [(1R)- 1- hydroxy- 18- 2- [10- (2-octadecylcyclopropyl)decyl]cyclopropyloctadecyl]tetracosanoic acid 93
Figure 29 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate 94
Figure 30 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid 95
Figure 31 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid 96
Figure 32 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoate.....97

Figure 33 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoic acid 98
Figure 34 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-22-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyldocosyl]hexacosanoate 99
Figure 35 - *In silico* Raman spectra of (2R)-2-[(1R)-14-2-[(15Z)-hexatriacont-15-en-1-yl]cyclopropyl-1-hydroxytetradecyl]hexacosanoate 100
Figure 36 - *In silico* Raman spectra of α-mycolic acid 101
Figure 37 - *In silico* Raman spectra of α,α'-trehalose 6-mycolate 102
Figure 38 - *In silico* Raman spectra of β-D-Araf-(1→2)-α-D-Araf-(1→3)-α-D-Araf-(1→5)-α-D-Araf-OMe 103
Figure 39 - *In silico* Raman spectra of Keto Mycolic Acid 103
Figure 40 - *In silico* Raman spectra of Methoxy Mycolic Acid 104
Figure 41 - SEM micrographs of *E. coli*: (a) control and (b) 10, (c) 20 and (d) 30 μg ml^{-1} AgNPs. (Arrows indicate depression and extensive damage of bacterial cells) (Ansari *et al.* 2014) (253) 109
Figure 42 - Polar lipid Average Raman spectra, based on the average spectra of 100x technical repeats. The coloured line is representative of the average, whilst the greyed-out area represents the standard deviation in each sample. 118
Figure 43 - Non-polar lipid Average Raman spectra. Based on the average spectra of 100x technical repeats. The coloured line is representative of the average, whilst the greyed-out area represents the standard deviation in each sample. 119
Figure 44 - PCA Quadrant analysis of Polar lipid extract. **(A)**. Quadrant analysis of the PCA data, illustrating the clustering in the quadrants: Q2 (Orange), Q3 (Green), & Q4 (Red). **(B)**. Clustering plot illustrating the most influential spectra in determining the clustering effects observed within the respective quadrants Q1, Q2, Q3, Q4. **(C)**. Raman loadings plot demonstrating the most influential parts of the spectra in the PCA. 121
Figure 45 - PCA Quadrant analysis of non-polar lipid extract. **(A)**. Quadrant analysis of the PCA data, illustrating the clustering in the quadrants: Q1 (Blue), Q2 (Orange), Q3 (Green), & Q4 (Red). **(B)**. Clustering plot illustrating the most influential spectra in determining the clustering effects observed within the respective quadrants Q1, Q2, Q3, Q4. **(C)**. Raman loadings plot demonstrating the most influential parts of the spectra in the PCA. 122
Figure 46 - Average Raman spectra of the bacterial species analysed in this chapter. Average is based on 100 spectra per bacterial species, arising from 100 technical repeats. The slight blurring of the line is representative of the standard deviation in the sample. 139
Figure 47 - Average Raman spectra of the Mycobacteriales analysed in this chapter. Average is based on 300 spectra per bacterial species. The bold line represents the average spectra. The blurred lines represent the standard variation in each biological repeat. 140
Figure 48 - PCA of Raman spectra of all bacteria analysed in this project with two components, and 100x technical repeats, with one biological repeat each. Note the broad

overlap between the majority of the bacteria. *Mycobacterium bovis* BCG, *Moraxella catarrhalis* and *Rhodoccus erythropolis*.....142
Figure 49 - PCA Analysis of the Mycobacteriales utilising two principal components following three biological repeats of the Mycobacteriales utilised in this project. 100x technical repeats were used in each biological repeat leading to N=300 spectra per bacterial species. Note the widespread of each bacterial species and the crossover of the Raman spectra. ..143
Figure 50 - (a) Quadrant analysis of the PCA data, illustrating the clustering in Q3 (Green). (b) Clustering plot illustrating the most influential spectral elements in Q1, Q2, Q3 and Q4. (c) Raman loadings plot demonstrating the most influential parts of the spectra in the respective principal component (LPC1, LPC2). 145
Figure 51 - (a) SOM trained to analyse the respective bacteria illustrating the clustering of the spectra. (b) SOMDI showing the peaks which have the greatest influence in activating the respective neurons in SOM. 146
Figure 52 - Box plots comparing the dominant Raman spectral peaks of the studied bacterial species (NS = not significant, * = p<0.05, ** = p<0.01, ***= p<0.001, ****= p<0.0001). 148
Figure 53 - LDA Plot of the Mycobacteriales, displaying vast separation between the three bacterial species.....149
Figure 54 - Confusion matrix following the 60:40 training and test data split during the PCA-LDA. (c). Characteristic derived spectral barcodes of the C. glutamicum, M. bovis BCG and R. erythropolis after the application of a Savitzky-Golan filter and reduction of the noise from the double differentiation. 150
Figure 55 - Characteristic derived spectral barcodes of the *C. glutamicum*, *M. bovis* BCG and *R. erythropolis* after the application of a Savitzky-Golay filter and reduction of the noise from the double differentiation. 151
Figure 56 - Combined average Raman spectra of *Mycobacterium bovis* BCG at various concentrations of 10^X CFUs/ml. The coloured line represents the average spectra. The greyed lines represent the standard variation in each biological repeat.....163
Figure 57 - Combined average Raman spectra of Mycobacterium bovis BCG at various concentrations of 10^X CFUs/ml and the standardised mix of bacteria. The coloured line represents the average spectra. The greyed lines represent the standard variation in each biological repeat. 163
Figure 58 - PCA of the bacterial suspensions utilising two principal components following three technical repeats of the numerous suspensions utilised in this project. 3x technical repeats were used in each technical repeat leading to N=3 spectra per suspension. Note the clustering of each bacterial suspension and the crossover of the Raman spectra. 165
Figure 59 – Raman loadings for the PCA of the bacterial mixes demonstrating the most influential parts of the spectra in the respective principal component (LPC1, LPC2).....166
Figure 60 - **(A)** The most widely used engineered CRISPR-Cas system utilizes a fusion between a crRNA and part of the tracrRNA sequence. This single gRNA complexes with Cas9

to mediate cleavage of target DNA sites that are complementary to the 5' 20 nt of the gRNA and that lie next to a PAM sequence. **(B)** Example sequences of a gRNA. (Adapted from Sander & Young (306)) 181
Figure 61 - Plasmid maps for pCRISPR, pCAS9 and pMA-RQ, acquired from Addgene & Thermo Scientific (172) 201
Figure 62 - Average Raman Spectra of *S. aureus* 208
Figure 63 - Average Raman Spectra of *R. erythropolis* 209
Figure 64 - Average Raman Spectra of *P. fluorescens* 210
Figure 65 - Average Raman Spectra of *P. aeruginosa* 211
Figure 66 - Average Raman Spectra of *M. hyosynoviae* 212
Figure 67 - Average Raman Spectra of *Mycoplasma bovis* 213
Figure 68 - Average Raman Spectra of *M. bovis* BCG 215
Figure 69 - Average Raman Spectra of *M. catarrhalis* 216
Figure 70 - Average Raman Spectra of *M. haemolytica* 218
Figure 71 - Average Raman Spectra of *E. coli* 220
Figure 72 - Average Raman Spectra of *C. glutamicum* 221
Figure 73 - Average Raman Spectra of *B. cereus* 222

List of Tables:

Table 1 - Currently used diagnostic techniques, recognised by the WHO, with the associated strengths and weaknesses, and their associated sensitivity and specificity (alongside a 95% confidence interval), taken from Macgregor-Fairlie *et al.* 2020 (71) 25
Table 2 - Diagnostic techniques currently in development with the associated strengths and weaknesses, and their associated sensitivity and specificity (alongside a 95% confidence interval), taken from Macgregor-Fairlie *et al.* 2020 (71).....31
Table 3 - Variations in size and structure of mycolic acid within the mycolic acid containing genera of the order Actinomycetales (179).....43
Table 4 - Compound table listing the compounds identified *via* ChEBI alongside SMILES format (246).....83
Table 5 - Table displaying the spectral assignment for the most prominent peaks, and the compound they are most prominent in according to Talari *et al.* (247). Spectra associated with lipids are highlighted in orange. 107
Table 6 - Predicted Raman peaks for cellular lipids follow *In silico* analysis of *M. tuberculosis* related compounds.....114
Table 7 - Raman peak assignment for the most prominent peaks in the Raman spectra of non-polar and polar lipid extracts 125
Table 8 - Table illustrating which named species have been detected in three papers) (278, 279, 280) 133
Table 9 - Table illustrating bacteria acquired for this project; displayed with full name, culture reference and original source. 134
Table 10 - Summary of the prominent peaks with the greatest statistical significance determined in **Figure** 54. Peaks attributions are taken from Movasaghi *et al.* (285) & Talari *et al. (247)* 148
Table 11 - Mean confusion matrix generated from the PCA-LDA. The values are rounded to the nearest integer. 150
Table 12 - Summary of the prominent peaks with the greatest statistical significance determined in **Figure** 54. Peaks attributions are taken from Movasaghi *et al.* (285) & (247) 152
Table 13 - Peak assignment for the most prominent peaks within the Raman spectra generated in this chapter.....167
Table 14 - List of CRISPR/Cas transfection methods 186
Table 15 - Methodology for timings and fixing agents for SEM sample preparation 199
Table 16 - Table illustrating the genetic sequences of the genes identified in *via* BV-BRC, highlighted sequences were used as basis for gRNA design and “protected” when ordered from Thermo Fisher. 200

Table 17 - Sequence of Oligonucleotides ordered from IDT with gRNA sequences highlighted 202
Table 18 - Raman peak assignment for *S. aureus* 208
Table 19 - Raman peak assignment for *R. erythropolis* 209
Table 20 - Raman peak assignment for *P. fluorescens* 211
Table 21 - Raman peak assignment for *P. aeruginosa* 212
Table 22 - Raman peak assignment for *M. hyosynoviae* 213
Table 23 - Raman peak assignment for *Mycoplasma bovis* 215
Table 24 - Raman peak assignment for *M. bovis* BCG 216
Table 25 - Raman peak assignment for *M. catarrhalis* 217
Table 26 - Raman peak assignment for *M. haemolytica* 219
Table 27 - Raman peak assignment for *E. coli* 221
Table 28 - Raman peak assignment for *C. glutamicum* 222
Table 29 - Raman peak assignment for *B. cereus* 223
Table 30 - Illustration of the bacteria in the bacterial mix used in Chapter 7, their recorded Optical Densities and their proportions in the final mix 224

List of Abbreviations:

°C	degrees Celsius
µg	Micrograms
µL	Microlitres
µm	Micrometre
AIDS	Acquired Immune Deficiency Syndrome
ATCC	American Type Culture Collections
BCG	Bacillus Calmette–Guérin
BHI	Brain-Heart Infusion
BLAST	Basic Local Alignment Search Tool
BlueBEAR	Blue Birmingham Environment for Academic Research
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
CDC	Center for Disease Control
CFU	Colony forming units
cm	Centimetre
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
crRNA	CRISPR RNA
ddPCR	Digital Droplet Polymerase Chain Reaction
DFT	Density Function Theory
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA#
ECACC	European Collection of Authenticated Cell Cultures
e.g.	*Exempli gratia* (Latin) – For example
EMA	European Medicines Agency
EMBL	European Molecular Biology Laboratories
Etc.	*Et cetera* (Latin – and so forth)
EU	European Union
FDA	Food and Drugs Administration
GDP	Gross Domestic Product
gRNA	Guide Ribonucleic Acid
HDJ	Homology Directed Repair
HIV	Human Immunodeficiency Virus
HPC	High-performance PC
HPV	Human Papillomavirus
I.e.	*Id est* (Latin) – That is

IR	Infrared
IRD	The Influenza Research Database
L	Litre
LAMP	Loop-Mediated Isothermal Amplification
LDA	Linear Discriminant Analysis
LJ	Lowenstein-Jensen
LMIC	Low- and Middle-Income Countries
LPA	Line Probe Assay
LTBI	Latent Tuberculosis Infection
M	Moles
MALDI-TOF	Matrix-Assisted Laser-Desoprtion-Ionisation Time-Of-Flight
MDR-TB	Multi-drug resistant - Tuberculosis
mg	Miligrams
MGIT	Mycobacteria Growth Indicator Tube
MHRA	Medicines and Healthcare products Regulatory Agency
ml	Mililitre
mol	Moles
MS	Mass Spectroscopy
N.B	*Note bene* (Latin) – Please note
NCTC	National Collection of Tissue Cultures
NHEJ	Non-homologous end joining
NHS	National Health Service
NICE	National Institute For Health and Care Excellence
OADC	Oleic acid, Albumin, Dextrose, Catalase
PAM	Protospace adjacent motif
PATRIC	Pathosystems Resource Integration Center
PBE	Perdew-Burke-Ernzerhof exchange
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCR-LDA	Principal Component Analysis-Linear Discriminant Analysis
qPCR	Real-Time Polymerase Chain Reaction
RIF	Rifampicin
RNA	Ribonucleic acid
RPA	Recombinase Polymerase Amplification
RPM	Revolutions per minute
SERS	Surface-Enhanced Raman Spectroscopy
SKiNET	Self-Optimising Kohonen Index Network

SNPs	Single Nucleotide Polymorphisms
SOM	Self-Organising Map
SOMDI	Self-Organising Map Discriminant Index
ssDNA	Single-strand DNA
SVP	Split valence polarisation
TB	Tuberculosis
TNF-α	Tumour Necrosis Factor Alpha
tracrRNA	Trans-activating RNA
tRNA	Tranporter-RNA
TZVP	Triple Zeta Variance Potential
UK	United Kingdom
UV	Ultraviolet
ViPR	Virus pathogen database and analysis
WHO	World Health Organization
XDR-TB	Extensively drug-resistant TB

Chapter 1:

1. Introduction:

1.1. Aetiology of Tuberculosis:

Long considered a disease of the Victorian age, Tuberculosis (TB) has never disappeared and continues to be a very much lived reality for the vast majority of the global population and persists in infecting countless individuals on an annual basis across the world (1). A significant cause of mortality and morbidity in many countries, TB has plagued humanity since time immemorial.

Evidence for TB infection can be found as far back as Ancient Egypt and beyond, with the oldest infected remains having been identified in Hungary and believed to be approximately 9000 years old (2).

The disease itself is caused by ten bacteria which form the *Mycobacterium tuberculosis* bacterial complex (MTBC), all of which belong to the genus,

Mycobacterium (3). **Figure 1** illustrates these organisms, below:

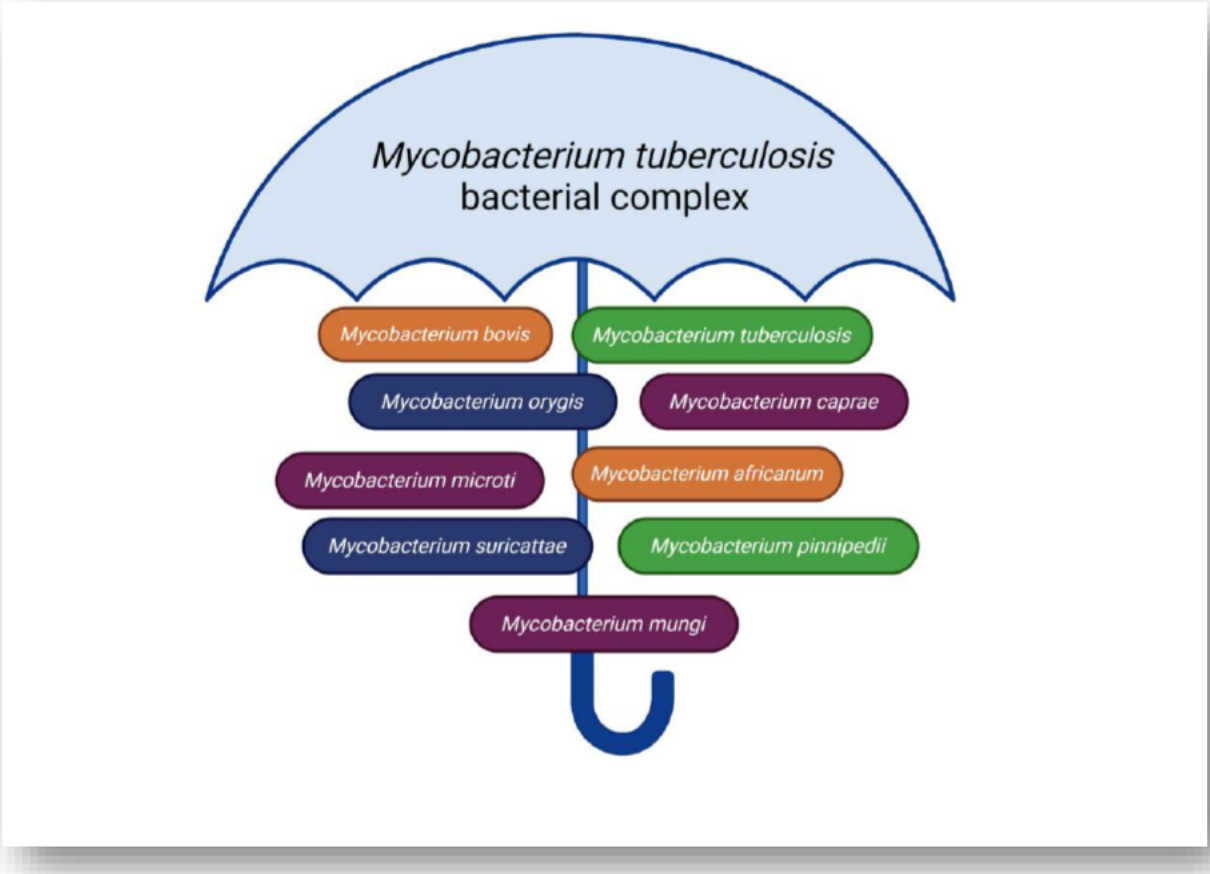


Figure 1 - The Mycobacterium tuberculosis bacterial complex. Created with BioRender.com

Primarily, TB is a respiratory disease caused by bacteria; often referred to as pulmonary TB. It can exfiltrate the lungs and establish infection elsewhere in the body, leading to extrapulmonary TB. Perhaps a more insidious aspect of TB infection is its ability to infect a person and then remain latent (Hereto referred as “Latent TB Infection” or LTBI). Though unable to infect other people in the latent phase of disease, the disease can become active and begin to cause symptoms (4). In the active stage of infection, the bacteria can then proliferate and spread between

individuals. This occurs in approximately 10% of individuals who are infected with the TB causing bacteria (5), though it is thought to be significantly higher in patients co-infected with Human Immunodeficiency Virus (HIV) (6). Latent TB infection is also a contraindication for certain biological therapies such as infliximab, which inhibits Tumour Necrosis Factor Alpha (TNF-α), a critical cytokine in the immunological control of TB infection (7).

In 2021, the World Health Organization (WHO) estimate that approximately 1 in 3 people worldwide are infected with TB causing bacteria and approximately 1.6 million people lost their lives due to the disease (1). These statistics must also be considered within the context of the ongoing COVID-19 pandemic and it is of a growing concern that COVID-19 is likely to be a major setback in the fight against tuberculosis (8). Specifically, there are concerns relating to treatment and diagnosis access, as well as individual behaviour where people are reticent to undertake treatment if they begin to present with symptoms due to a perceived risk of COVID-19. Additionally, the free proliferation of TB infection has significant healthcare and economic costs (9, 10, 11).

The symptoms of pulmonary tuberculosis are illustrated in **Figure 2** opposite (4):

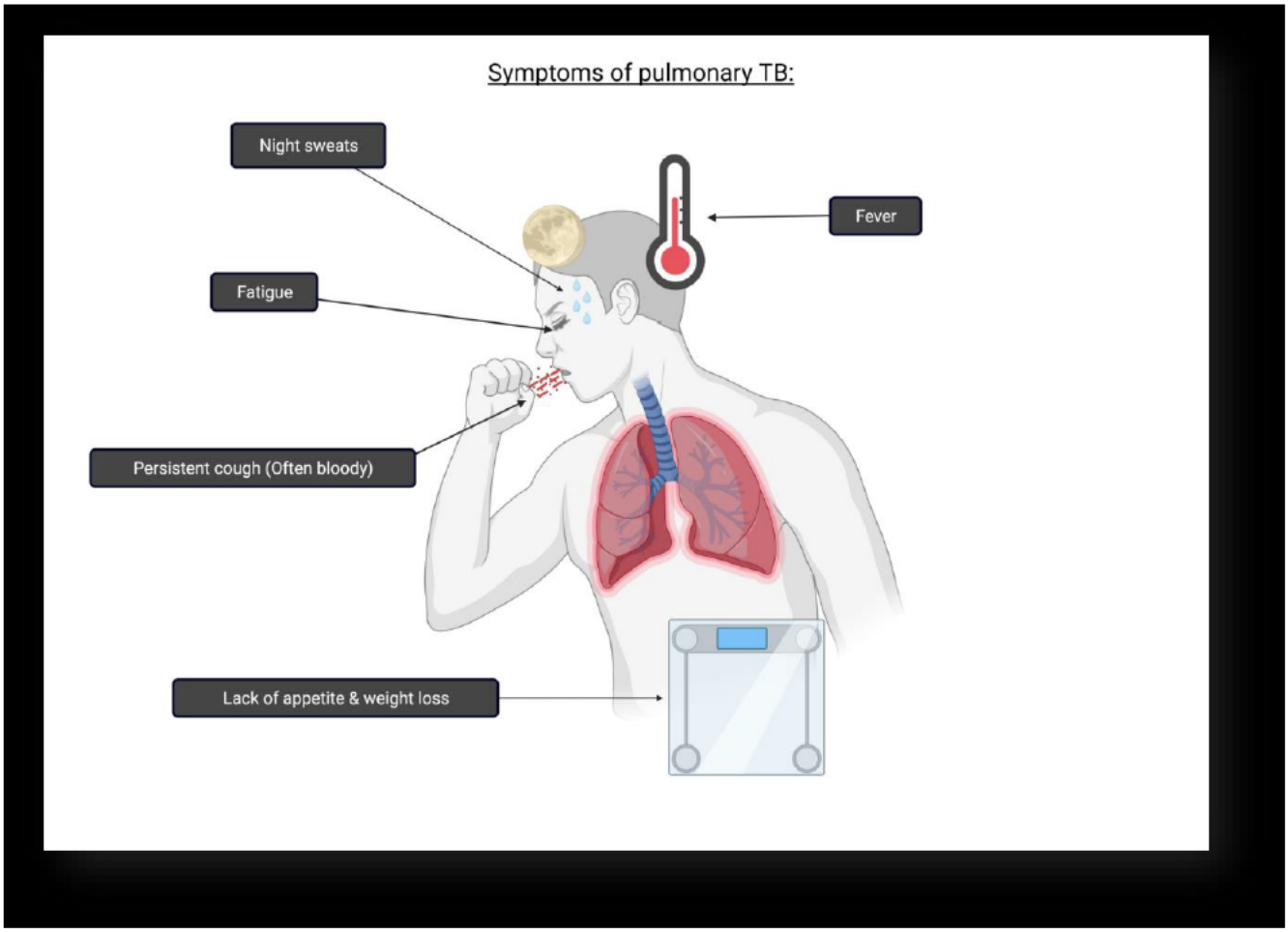


Figure 2 - Symptoms of pulmonary TB. Created with BioRender.com

Should the bacteria exfiltrate the lungs, symptoms often present based on the locale of the extrapulmonary infection, *i.e.* if the infection disseminates into the bones, it will present as osteomyelitis (12).

Common sites of extrapulmonary infection include the lymph nodes, the central nervous system, the genital tract, the skeletal system, and the gastrointestinal tract. Naturally, upon presentation, these extrapulmonary TB possesses many issues when attempting treatment based upon the ease of access of the location of the extrapulmonary infection.

1.2. Pathogenesis of Tuberculosis:

1.2.1. Initial Infection:

Upon inhalation, the bacteria travel to the lungs where they are immediately besieged by macrophages present within the alveoli, though this is to the bacteria's benefit. Once engulfed by macrophages, the bacteria can begin to multiply. It has been theorised that, in a hostile environment, the bacteria are able to change their metabolic state and use the macrophages as a nutrient source (13). However, it is also at this point that the body's adaptive immunity starts (14).

The bacteria travel to the lymph nodes where they are beset by CD4+ T-cells alongside dendritic cells (15). In this moment, the bacteria are fighting an uphill battel, where, the infection then proceeds in one of three directions:

1. Complete resolution
2. Latent infection
3. Active infection

All of which are dependent on various factors such as general health of the individual, whether they are malnourished and the effectiveness of their immune system as well as the number of bacteria inhaled and the duration of exposure.

Complete resolution of infection does indeed occur, but it is worth noting that it is an exception rather than a rule when considering TB infection (16, 17). As a result, the majority of people develop a latent infection.

1.2.2. Latent Tuberculosis:

Within the context of tuberculosis, when one discusses “Latent” infection or Latent Tuberculosis Infection (LTBI), it is in reference to the causative pathogen establishing an infection within an individual yet remaining in a relative state of homeostasis. The immune system actively attacks the infection but remains unable to fully eliminate it. Meanwhile, the immune defence is of a substantial enough level that it prevents the bacteria from fully propagating around the body.

For approximately 90% of people with TB, this is where the infection stays. As for the underlying molecular mechanisms for this process, it has been covered in extensive detail by various researchers in many papers (18, 19, 20). However, it tends to arise from a variety of attributes that *Mycobacteria* sp. possess which are unique to the genus.

For the sake of clarity, latent infection consists of a variety of factors continuously attacking the TB causing bacteria.

From the outset and upon engulfment by macrophages, the bacteria are transported to lymph nodes in the periphery. Once here, the adaptive immune response commences, and the bacteria are attacked by macrophages and CD4+ T cells in a bid to isolate the infection. Concurrently, an entire cascade of chemokines, cytokines, and complement proteins are released in a bid to stem the spread of the infection and draw other immune cells to the site of the bacteria.

Over time, granulomas form. At the centre of a granuloma is the causative agent of TB: A member of the MTBC. However, on the exterior, one would see a collection of macrophages, and other immune cells such as neutrophils and dendritic cells (21).

Histologically, if one were to extensively examine a TB granuloma, one would see that there is a mixture of intact and disrupted cells, living and dead cells and cellular debris, with a greater number of dead cells as one gets closer to the centre. Towards the centre, living cells from the MTBC are found in hypoxic conditions. The borders of these granulomas are often well-defined in cases of LTBI (22).

1.2.3. Active Tuberculosis:

Should the immune system fail to contain the bacteria within a granuloma (be this immediately following infection or due to factors within an individual's life which influence the ability to contain the bacteria), the bacteria spill out into the respiratory tract and it is here that they are able to contaminate sputum and disseminate to other individuals and, thus, perpetuate the cycle of infection (23).

Meanwhile, following the activation of the bacteria, an infected individual will begin to present with the symptoms listed in **Figure 2.**

1.2.4. Extrapulmonary Tuberculosis:

TB does not always affect the lungs. In approximately 15% of active cases, the bacteria are able to leave the lungs and go on to infect other organs and tissues (24), causing what is known as extrapulmonary TB. This primarily affects lymph nodes which are not directly associated with the lungs (specifically, the cervical lymph nodes in the neck where the infection has been previously known as scrofula).

It can also affect the central nervous system, the genital tract, the skeletal system (where it has been known as Pott's disease) and the gastrointestinal tract (25).

The underlying mechanism of extrapulmonary infection has not yet been fully established though there are several theories pertaining to how the bacteria are able to exfiltrate the lungs (24).

These primarily consist of the "trojan horse" theory, where the bacteria infect a macrophage and are then carried away to a distant site where they proliferate (24). Additionally, there are also theories relating to MTBC bacteria being able to infiltrate and spread *via* endothelial cells as well as dendritic cells (26).

Diagnosis of extrapulmonary TB can also present with its own issues, due to the differing location of the bacteria. A key example of this would be the use of smear microscopy (a technique where the healthcare worker takes a sputum sample; stains it and then examines it under a microscope) and its failure to detect extrapulmonary TB if the patient does not present with pulmonary symptoms (27).

Conversely techniques relying on an immunological response (such as the Tuberculin skin test or the Quanti-FERON Gold test) present with relatively accurate results (though these tests present with their own limitations which shall be discussed later on) (28). Additionally, fine needle biopsy of the site of a suspected extrapulmonary infection can yield successful diagnosis (29).

For most individuals extrapulmonary TB is of minimal concern, however, the prevalence of extrapulmonary TB vastly increases in populations co-infected with Human Immunodeficiency Virus (HIV) (30) .

1.2.5. HIV-TB Co-infection:

Human Immunodeficiency Virus (HIV) is a retrovirus which attacks the immune system (specifically CD4+ T-cells), ultimately leaving those who are infected with the virus unable to fend off infections that an individual with a fully functioning immune system can easily defeat (most commonly known as opportunistic infections) (31).

When a person infected with HIV is no longer able to fend off such infections, they proceed on to Acquired Immune Deficiency Syndrome (AIDS) (32). Historically, HIV/AIDS was a death sentence and the vast majority of people infected with HIV would develop AIDS and then pass away from complications arising from opportunistic infections (33).

Fortunately, HIV/AIDS treatment has improved drastically since the virus' advent in the 1980's and anti-retroviral therapy can reduce viral titre to such an extent that an individual cannot spread the virus to others, and they are exceedingly unlikely to suffer complications from the virus (34). As a result of the successful developments in treatment, a HIV patient in a high-income country is likely to enjoy a high standard of living and a lifespan comparable to someone with type 1 diabetes (35, 36).

The same cannot be said for those living in Sub-Saharan Africa, where HIV/AIDS is of the greatest burden (37) and whilst some countries such as Botswana have successfully sought to defeat the scourge of HIV/AIDS (Botswana is set to achieve

the WHO's goal of 90-90-90[1] by 2024, even when 20% of the population has HIV/AIDS) (38), others continue to have an exceptionally high case rates of HIV/AIDS without any imminent prospect of completely resolving the situation (A key example being South Africa where anti-retroviral therapy rates are plateauing and the country continues to struggle in the aftermath of the HIV/AIDS denialist policies of the Mbeki government) (39).

HIV/AIDS within the context of TB is a co-morbidity where HIV patients are significantly more likely to develop active TB (an approximately 26-fold increase)(40) and extra-pulmonary TB (41), whilst also increasing the chances of mortality (42).

TB alone is a horrific disease, however, co-infection with HIV is an important consideration given the spread of TB within countries where HIV is prevalent, and the detrimental impact HIV can have on a person's ability to ward off TB. From a public health perspective, HIV/TB co-infection means that co-infected patients are more likely to develop an active infection and spread TB to others; it is therefore critical to successfully identify and treat co-infected patients in a bid to stem TB infections (as well as HIV).

Furthermore, co-infected patients represent a unique subset of TB patients where special considerations must be given to treatment given the hepatotoxic nature of both ART and anti-mycobacterial drugs (43) as well as such considerations being

[1] 90-90-90 = 90% of the population knowing their HIV status, 90% of positive individuals on anti-retroviral therapy, 90% with suppressed viral loads.

made for diagnoses within HIV/TB patients due to the diminished ability for the body to mount an immune response (44).

Due to this diminished ability, the diagnostic criteria for some diagnostic tests must be adjusted accordingly (a prime example being the tuberculin skin test where the reaction is graded as positive if the size of the respective weal is >5mm, whereas a test is considered positive if the weal is >15mm in a “healthy” individual). However, adjusting the diagnostic criteria for testing is only possible if one knows their HIV status (45).

Lastly, another factor to consider when discussing HIV/TB co-infection is the lack of infrastructure within the countries where co-infection is prevalent. This includes rail/road infrastructure, but also electrical, telecommunications and medical/scientific infrastructure. Due to this lack of development, issues relating to HIV and TB are more pronounced because of inaccessibility to diagnostics and treatment (46).

This contrasts strongly with high-/very high-income countries such as the United Kingdom, Germany and Norway, where healthcare is readily accessible to the majority of the population (47). Therefore when a case of TB/HIV co-infection does present, it can be readily addressed, thus reducing the probability of the infection being spread or a patient going from a latent infection to an active infection (48).

1.3. Tuberculosis Treatment:

Unlike the majority of bacteria, *Mycobacterium sp.* occupy a rather specific phylogenetic *niche* where they are neither Gram positive or Gram negative. Instead, *Mycobacterium sp.* are considered acid-fast (49).

That is to say that they are unable to be colourised by components of the Gram stain; instead, they appear to acquire some of the Gram stain but not all of it, thus making it difficult to differentiate them from, either Gram positive or Gram negative bacteria (49).

This is largely due to *Mycobacterium sp.* having a vastly different cell wall structure from most other pathogenic bacteria, which makes developing drugs for the bacteria more challenging. As a result, the drugs utilised against *Mycobacterium sp.* often differ from those utilised in other infections (50, 51, 52).

For instance, beta-lactam antibiotics (such as penicillin) work by binding irreversibly to penicillin-binding proteins, said proteins are important for cell wall synthesis. With these proteins inhibited, the cell wall can no longer be maintained, and the bacteria is then uncontrollably flooded by water, causing cell lysis (53).

Beta-lactam antibiotics are routinely used on Gram positive bacteria, however, they are ineffective against Gram negative bacteria and *Mycobacterium sp.*. Instead, drugs targeting protein synthesis and other metabolic pathways are used (54).

Initially, these drugs were used as a monotherapy for TB, though these fell out of favour as it became clear that monotherapy was not suitable for TB treatment. This was noted at the time, as some trial participants relapsed following attempts to treat TB with streptomycin, with such a relapse presenting with streptomycin resistance (55).

Over time, as collective knowledge grew, the repertoire of antibiotics which could be used on TB expanded, and this included drugs such as rifampicin and isoniazid. Eventually, combination therapy began to be utilised, as it was noted that the level of resistance dropped significantly when drugs were used in combination and it is this combination therapy which is still used today (56).

1.3.1. Standard Tuberculosis Treatment:

Due to the increased success of combination therapy, TB is now treated in accordance with several different drug regimens which are constantly being studied, reviewed and modified.

As for the basis of this treatment, the first course of treatment has a tendency to be based on two or three of five drugs, pending upon the phase of treatment (57):

1. Rifampicin
2. Isoniazid
3. Pyrazinamide
4. Ethambutol
5. Streptomycin

Assuming an individual has no contra-indications for therapy and is presenting with active disease, duration of treatment can vary between 4 and 12 months. As for the progression of the course of drugs, this tends to vary by jurisdiction. For instance, the guidance provided by the Centre for Disease Control in the USA varies from the guidance provided by the National Institute for Health and Clinical Excellence (NICE)

in the UK (58). Additionally, low- and middle- income countries also need to consider accessibility and cost of certain drugs (59).

Overall, the WHO recommends several treatment regimens to successfully treat TB; the overall course will be influenced by the accessibility and cost of drugs as well as the prevailing attitude of the prescribing doctor. Additionally, treatment may vary should a patient have extrapulmonary TB or a HIV infection.

1.3.2. Drug-resistant Tuberculosis

Unfortunately, like a lot of infectious diseases that are caused by bacteria, increased resistance to antibiotics is a substantial concern when discussing tuberculosis.

Thankfully diagnostic methods are continuously improving at detecting drug resistance, though they are far from perfect. Drug resistant TB tends to result in poorer healthcare outcomes as well as increased disability-adjusted life years for the patient and a high economic burden in the countries where it has taken root (60).

As previously mentioned, drug resistance in *M. tuberculosis* was first noted during the trials of streptomycin in the 1940's. One of the issues with the use of macrolide antibiotics is that they are primarily bacteriostatic in nature (they inhibit growth), but they do not always deliver the mortal blow (61).

The mechanism of action of macrolides is very similar to other antibiotics utilised against the MTBC. That is to say that they target specific enzymes within certain metabolic pathways and, ultimately, this leads to inhibition of growth though not necessarily destruction of the cell (62).

However, said antibiotics can only target bacteria that are actively growing. Due to this, there is a correlation between bacterial growth rate and the efficiency of the respective antibiotic (63).

MTBC bacteria meanwhile are very fickle by nature. Due to their slow growth and their cell wall differing from most pathogenic bacteria, researchers are already fighting an uphill battle due to a lack of drug targets. When this is combined with their propensity to undertake DNA methylation as well as the development of Single Nucleotide Polymorphisms (SNPs), MTBC bacteria become even more formidable (64).

As a result of these factors, MTBC readily develop drug resistance. That is, MTBC bacteria can withstand a barrage of antibiotics and a notable proportion of TB cases present as somewhat drug resistant to one of the first-line antibiotics that patients are treated with (65).

1.3.3. Multi-drug Resistant Tuberculosis:

Multi-drug resistant TB (MDR-TB) presents as TB which displays resistance to two or more of the first-line antibiotics used against TB (66). This means that treatment is often protracted and drawn out, fostering the dissemination of the MDR-TB between individuals.

MDR-TB tends to result in poorer prognosis and has a higher rate of remission when compared to TB where multi-drug resistance is not present. Additionally, such cases present with further increased disability-adjusted life years for the patient and a high economic burden in the countries where it has taken root (67).

Whilst MDR-TB is a cause for concern in the countries where TB is the most prevalent, the burden of MDR-TB is particularly notable in countries of the former USSR (68).

1.3.4. Extensively Drug Resistant Tuberculosis:

MDR-TB is only the tip of the iceberg. As time has progressed, the emergence of extensively drug resistant TB (XDR-TB) has emerged (69).

Quite simply, the WHO defines this as (1):

"XDR-TB is TB that is resistant to rifampicin, plus any fluoroquinolone, plus at least one of the drugs bedaquiline and linezolid"

However, whilst XDR-TB exacerbate the issues related to economic cost, disability and quality of life as previously mentioned, it does not appear to be spreading as rapidly as MDR-TB (70). Collectively, approx. 5% of the bacteriologically confirmed diagnoses of TB express some form of drug resistance (1).

To summarise, drug resistance represents a significant public health concern when it comes to managing tuberculosis. With the advent of the COVID-19 pandemic, cases of drug resistance are becoming increasingly more difficult to identify and there is great trepidation that recent progress may be lost (1, 71, 72, 73).

For the vast majority of patients, TB can be treated, provided that treatment is adhered to and cases are rapidly identified. It is therefore of the utmost primacy that diagnostic methods are rapid, cheap, and accurate whilst also identifying any drug resistance.

1.4. Tuberculosis Diagnostics:

NB: The following sections will seek to incorporate and update the work covered by Macgregor-Fairlie *et al.* 2020 (74).

Treatment represents only one side of the coin when it comes to resolving the issues that arise from TB worldwide and the World Health Organization holds both treatment and diagnosis with the same parity of esteem (1).

Diagnosis is typically underpinned by a combined use of molecular detection methods, immunological based assays and direct culture and observation (75). These methods, however, do have known limitations such as, being less effective in individuals infected with HIV (due to their reduced immune response) (76), those with a latent infection (77), and in children unable to produce significant sputum containing MTBC (78). Furthermore, these methods often require a steady supply of electricity, well-trained personnel and lab modules, which are seldom available in the often under-resourced laboratories in low- and middle-income countries (LMIC) (79, 80). In terms of diagnosing extrapulmonary TB, diagnosis often remains the same as with pulmonary or latent TB, though it is often combined with the use of invasive biopsies taken from the suspected site of infection.

Timely and effective diagnosis for combatting and eradicating TB is at the heart of WHO guidelines with management of latent infections being considered a key component in eradication (81). However, this goal is poorly supported by current technologies which fall short of the portable diagnostic needs, exhibiting poor-sensitivity, special-handling requirements, and complicated, costly procedures.

There is, therefore, an urgent need to develop diagnostic methods that can accurately, rapidly, and cheaply diagnose TB particularly, in the rural areas of the Global South.

1.4.1. Direct culture and observation:

The ‘gold-standard’ for positive diagnosis of TB is the use of traditional plate culture (82), typically conducted in a high containment laboratory, due to the inherent risk associated with working with MTBC. The culturing of MTBC raises several issues, mainly relating to the growth rate of MTBC, which is substantially slower than other pathogenic organisms, most probably due to a selective pressure at which faster growing *Mycobacterium sp*. induces a greater immune response (83). It can often take more than four weeks for the MTBC to produce colonies using the traditional Lowenstein-Jensen medium (LJ) (84). An advantage of plate culture is that drug sensitivity testing can take place concurrently, enabling clinicians to guide antimicrobial therapy more effectively (85).

Mycobacteria growth indicator tubes (MGIT) seek to overcome this using a fluorescent probe, which is quenched by dissolved oxygen within a selective medium. Upon successful cultivation of a member of the MTBC, the oxygen is depleted, causing the probe to be activated. This is subsequently detected using a light sensor. The MGIT has been shown to reduce the mean detection time from 38.6 days to 21.4 days (84). Meanwhile, a patient with an active TB continues to (unknowingly) disseminate the infection. It is worth noting that to successfully culture

the organism requires high level of expertise and MGIT only offer a rather minor advantage over the traditional culture methods (86).

Since the MTBC is a group 3 Hazard according to the Advisory Committee on Dangerous Pathogens (87), defined as a biological agent, which *"…can cause severe human disease and may be a serious hazard to employees…may spread to the community…"* (87), it requires the corresponding containment facilities to be safely handled. However, this is not always possible within LMICs and, even less so within the more rural settings. Any such facility would also require the relevant expertise and logistics to be run, which is not possible in many of the circumstances (88). Furthermore, sampling methods of MTBC can be inconsistent, with the subsequent sensitivity varying between 16-60% (89). Sampling methods can involve for instance, a patient producing sputum and ejecting it into a sterile container, or more invasive methods such as, bronchoalveolar lavage and bronchial washing (90). Both exhibiting a significant difference in their sensitivity of 85.7% *versus* 50.0%, respectively (90). This can consequently have an impact on the detection rate in local settings, and consideration must be also given to the impact the lavage or washing methods may have on the patients themselves (91).

The use of smear microscopy is often used in tandem with culture in resource poor settings. The method requires the use of the Ziehl-Neelsen stain that dyes MTBC and can often result in low sensitivity (56%) in the early active cases (92). Whilst both culture and smear microscopy can be very useful when used together, their efficacy drops significantly in paediatric (93) and HIV or immunocompromised

patients due to the lower levels of MTBC in the sputum as well as a weakened ability to produce sputum (94, 95).

Further methods have been combined with culture to successfully diagnose TB including, the chest X-ray and the Tuberculin skin test. Nevertheless, their efficacy was found to be relatively poor when compared to other more modern methods, which utilise molecular or immunological techniques (96, 97). The diagnostic flaws of these methods are often attributed to the effect of the human error in interpretation of the results (98). To overcome this, several companies are developing computational algorithms to better interpret chest X-ray results and significantly boost the detection rate whilst this still varies considerably between patient populations due to the co-morbidities (99).

1.4.2. Molecular and immunological detection:

Historically, the above-described direct detection methods have formed the basis of TB diagnostics. However recently, a shift has occurred towards, either molecular or immunological methods. These rely on the detection of molecules or compounds associated with MTBC (as opposed to the entire organism) or the detection of components of the immune system reactive to MTBC (*e.g.,* antibodies). There is currently a growing number of detection products on the market utilising these approaches. The two most prominent widely used assays are the molecular/Polymerase Chain Reaction (PCR)-based, GeneXpert MTB/RIF® assay and the immunological-based, QuantiFERON-TB gold® assay. GeneXpert first pellets bacteria from sputum and then uses PCR to test for the presence of *rpoB,* a

gene in MTBC responsible for producing RNA polymerase B. It also tests for resistance to Rifampicin, a first-line antibiotic.

Gamma interferon release assays (*e.g.,* the QuantiFERON-TB gold®) have been developed to test for latent TB by analysing the levels of gamma interferon released by immune cells (taken from a blood sample) in the presence of TB antigens. If the level of release is found to be above a specified threshold, a patient is considered positive for latent TB.

Whilst the GeneXpert has been successfully tested to detect other members of the MTBC *in silico*, it has not been tested in the field and, as such, has not been approved by the WHO, FDA, Medicines and Healthcare Regulatory Agency or EMA for the diagnosis of other MTBC organisms (100). It should be noted that, in theory, GeneXpert will detect other MTBC organisms and the treatment of TB which is caused by other MTBC remains broadly the same as TB caused by *M. tuberculosis* (101), but it is not a definitive diagnostic tool against other members of the MTBC.

Gamma interferon release assays (*e.g.* the QuantiFERON-TB gold®) have been developed to test for TB infection by analysing the levels of gamma interferon released by immune cells (taken from a blood sample) in the presence of TB antigens. If the level of release is found to be above a specified threshold, a patient is considered positive for TB. These assays possess broad detection and often produce false positives for *M. tuberculosis* specifically (102, 103).

There are several additional diagnostic methods for diagnosing TB, though their uptake is not as prevalent (104). Notably, this includes loop-mediated isothermal

amplification (LAMP) (105), which uses isothermal amplification procedure, in contrast to the PCR's multiple and cycling reactions, to produce a result that can be visualised by the naked eye. Although the sensitivity of LAMP is questionable in smear negative samples (106) and the infrastructure required to carry out these tests can often be prohibitively expensive (107), this method is recognised by the WHO (108). Another example is the line probe assay (LPA), which uses PCR to amplify DNA from MTBC and subsequently, applies the output to immobilised oligonucleotides on a strip. These oligonucleotides in turn, emit a colourimetric signal, which signifies the presence of gene mutations conferring drug resistance within the sample. LPAs are useful in detecting drug resistance in addition to the MTBC in smear positive samples (109) but are not recommended for use in smear negative samples due to their poor detection rate (108, 110) as the result of low numbers of individual cells being below the limit of detection. Further the WHO approved diagnostic method is urine-based lipoarabinomannan assay, which seeks to detect the presence of the MTBC compound, the *Lipoarabinomannan*, in the urine of TB patients. Although these assays have been demonstrated as useful, they are contraindicated in HIV and immunocompromised patients due to their poor sensitivity in these patient groups (111).

Molecular or immunological approaches can produce results in considerably shorter times than conventional culture. These tests are known to be prone to false positives (112), and thus a combined use with culture methods enables a more accurate diagnosis. Though limitations become prominent when used in HIV,

immunocompromised, paediatric or latent patients (113, 114). Additionally, molecular or immunological assays are costly to deploy and require a good level of expertise and suitable infrastructure to be implemented as evident by their high uptake in high-income countries such as the United Kingdom (115) yet, almost negligible uptake in countries such as, India or Eswatini (116), which often lack such infrastructure.

Overall, the inherent weaknesses of the current diagnostic methods as well as the reasons for their failings are broad and various. To briefly summarise the limitations of the current repertoire of TB tests include being highly time-consuming (in either acquisition or generation of results), requiring specialist laboratory equipment and well-trained personnel, being open to wide interpretation, exhibiting low specificity, which is further exacerbated in HIV and immunocompromised patients as well as children, or infrastructure issues within LMICs. Therefore, recent developments have been sought to overcome many of these limitations by both reducing the equipment and infrastructural requirements and by utilising different sample types.

To further summarise, the repertoire of current diagnostic tools; their inherent strengths and weaknesses can be seen in **Error! Reference source not found.**, opposite:

Diagnostic method	Sensitivity (95% CI)[2]	Specificity (95% CI)	Strengths	Weaknesses
Bacterial culture	100%[117]	100%[117]	+ Gold standard of sensitivity and specificity + Drug sensitivity testing can take place in tandem (through use of drug diffusion disks) + Cheaper than molecular/immunological methods	– Requires high level biosafety containment laboratory – Generation of results is time-consuming (>28 days)
MGIT (Becton Dickinson and Company, USA)	86-93%[118]	99.99%[118]	+ Faster than conventional culture + High degree of specificity and sensitivity	– More expensive than conventional culture – Requires specialist training – Method is labour intensive – Requires high-containment lab
Smear microscopy	60-69%[119]	97-98%[119]	+ Rapid, on-site delivery + Cheap + Few reagents are required	– Requires training – Reagents are toxic (*e.g.* phenol)
Chest X-ray	73-79%[120]	60-63%[120]	+ Readily available in healthcare settings + Non-invasive	– Low specificity & sensitivity – High initial costs, although has multiple applications
Tuberculin skin test (Sanofi, France)	48-78%[121]	57-81%[121]	+ Cheap + Can be used for LTBI + Can be rapidly deployed + Requires no handling of *Mtb*	– Results take approx. 5 days to appear – Requires repeated visits to healthcare professional – Highly variable sensitivity and specificity
GeneXpert MTB/RIF (Cepheid, USA)	82-88%[122]	96-98%[122]	+ Can test for *Mtb* and rifampicin resistance + Rapid turnaround (<2 hours) + Cartridges can be stored at room temperature	– High initial start-up cost (>$10,000) – Variable sensitivity in HIV/Immunocompromised patients (73 - 87%) – Low sensitivity in smear-negative patients (64 – 75%)
QuantiFERON-TB Gold (Qiagen, USA)	61-86%[121]	57-81%[121]	+ Can be used for LTBI + Blood sample (which is easier to acquire than a sputum sample) + Rapid detection	– Less sensitive in HIV/Immunocompromised individuals (43-49%) – Less sensitive in children (70-76%) – Requires handling of blood samples – Requires specialist training – Relatively expensive
TB LAMP (Eiken, Japan)	85.6-92.6%[123]	91.0-96.1%[123]	+ Sensitivity and specificity comparable to PCR testing + Cheaper to run than PCR + Visual readout + Rapid detection	– Infrastructure required can be prohibitively expensive – Cannot be used for LTBI – Presents a significant contamination risk if run in a molecular biology laboratory

[2]95% Confidence interval

Line probe assays for *Mtb* (Hain Lifescience, Germany & Nipro, Japan)	95.6-97.5%[124]	98.7-99.5%[124]	+ Can detect resistance to isoniazid and/or rifampicin + Rapid detection (<6 hours)	– Less sensitive and specific in smear-negative samples (62.5-74%) and (84.9-92.8%) – Requires electricity – Reagents require refrigeration/freezing – Cannot be used for LTBI
Determine TB LAM Ag test (Abbott, USA)	13-93%[125]	87-99%[125]	– Non-invasive – Rapid detection – Useful in immunocompromised/HIV patients who are seriously ill[3]	– Large variability in sensitivity – Not recommended in immunocompetent individuals

Table 1 - Currently used diagnostic techniques, recognised by the WHO, with the associated strengths and weaknesses, and their associated sensitivity and specificity (alongside a 95% confidence interval), taken from Macgregor-Fairlie *et al.* 2020 (71)

[3] Defined as *"respiratory rate >30 breaths per min, body temperature >39°C, heart rate >120 beats per min, or unable to walk unaided"*

1.4.3. Recent developments:

The above represents the current repertoire of diagnostic techniques for TB. As always, researchers and commercial interests are always in the process of enhancing current techniques or attempting a new approach entirely.

For instance, digital droplet PCR (ddPCR) is an advancement of PCR which offers absolute quantification rather than the relative quantification which is offered by real-time PCR (qPCR). ddPCR is a more recent innovation which partitions the amplification reaction seen in PCR (in the form of dropwise reactions) and this way, it provides an absolute quantification of gene expression rather than a relative one. Due to this, its sensitivity is higher than qPCR, and this method is capable to detect single copies of DNA (126). This technique has been already used to monitor mixed populations of cells to better understand drug-resistance (127). Whilst ddPCR offers a greater amount of sensitivity than quantitative or qPCR methods and can be used to test for MTBC infection in both sputum and blood samples (128, 129), it often can be prohibitively expensive (130). However, ddPCR could greatly expand the number of tests which can be concurrently run, as has happened with SARS-CoV-2 (131).

CRISPR/Cas9 is a technology that is often thought of as a gene-editing technique. Essentially, it has acted as a programmable enzyme that can “cut” DNA in specific places, though recent discoveries have identified more enzymes which possess collateral activity as well as an ability to target single stranded DNA, as is the case with Cas12a (132).

Contemporary research has demonstrated that this technique could be used to detect MTBC DNA (133). This is particularly noteworthy due to its potential to be lyophilised onto a lateral flow stick with attomolar sensitivity, enabling it to be readily deployed in areas where electricity is seldom available (134). Ai *et al.* (133) have utilised a recombinase polymerase amplification (RPA) to increase the amount of genetic material before placing the amplification product onto the lateral flow stick. RPA is a DNA/RNA amplification technique that can be run at a relatively low temperature (*ca.* 37^{o}C). The speed of the amplification of RPA is proportional to the temperature (*e.g.,* the higher the temperature, the faster the reaction) and, in this instance, it was utilised to amplify *IS6110,* a highly conserved genetic element in the MTBC (135). Due to the adaptable nature of the assay, the template to be amplified and analysed using CRISPR/Cas12a can be modified, enabling its extension to identify resistance mechanisms. The overarching principles of CRISPR/Cas are summed up in **A.1. Use of CRISPR-Cas for the detection of Tuberculosis**:Error! Reference source not found.**.**

Complimenting this, is the ability CRISPR/Cas to be used in tandem with PCR to significantly bolster the limit of detection in more resource-rich settings. Furthermore, it has been demonstrated that the technique has a better overall diagnostic performance when compared with the GeneXpert MTB/RIF® assay as well as a faster turnaround (136).

Next-generation sequencing is also a potential tool in the fight against TB. Historically, the thought of large devices such as a Sanger or Pyro sequencer being

deployed to the field would not have been taken seriously. However, recent developments have meant that sequencing technology has been miniaturised further, with the advent of techniques such as Nanopore sequencing which have been deployed on the front lines of disease outbreaks such as the Ebola outbreak in Africa in the mid-2010s (137).

The development of artificial intelligence systems is another way that we are improving our repertoire of techniques, by enhancing the level of accuracy and sensitivity to TB when chest X-ray and smear microscopy is employed (138).

The last technique to be discussed is that of Raman spectroscopy. Raman is a spectroscopic and (potentially) portable analytical tool, which could be readily deployed into rural areas. Raman spectroscopy has been utilised in many other applications, such as the diagnosis of cancer (139) and to identify bacterial infections. The technique utilises the phenomenon of Raman scattering to detect the unique molecular fingerprints of various bacteria when subjected to an excitation with a certain wavelength of light. Preliminary research has demonstrated its ability to detect TB alone (both drug-sensitive and drug-resistant serotypes), however no research exists as to its efficacy outside of a monoculture sample (a critical objective if such a device were to make it to market). The successful development of an Raman-based method would enable clinicians to rapidly screen through a vast number of patients due to Raman's ability to scan a sample within a matter of minutes, as opposed to many hours molecular methods use. Given the benefits of Raman spectroscopy and the accessibility of it throughout the duration of this

project, this is the main technique that shall be utilised and discussed in this thesis (140).

The techniques mentioned here are by no means exhaustive of the current methods in development, but an indication of where the field of TB diagnostics could be going. In general, demands are being placed on scientists to develop methods which can be readily deployed to rural areas where general infrastructure may be lacking. Therefore, these techniques need to prove sensitive, accurate and, ideally, able to function without mains electricity or extensive amounts of training.

To further summarise, the repertoire of current diagnostic tools; their inherent strengths and weaknesses can be seen in **Table 2** - Diagnostic techniques currently in development with the associated strengths and weaknesses, and their associated sensitivity and specificity (alongside a 95% confidence interval), taken from Macgregor-Fairlie *et al.* 2020 (71)**Error! Reference source not found.**, opposite:

Diagnostic method	Reported Sensitivity	Reported Specificity	Strengths	Weaknesses
ddPCR [129]	NA[4]	NA	+ Reported sensitivity down to 1 copy of DNA per sample + Provides absolute quantification of sample + Flexibility to analyse sputum and blood	– Prohibitively expensive ($10,000+) – Requires uninterrupt ble power supply
CRISPR/Cas12a	90% [133]	98% [133]	+ Rapid turnaround (<1.5 hours) + High sensitivity and specificity + Reaction can take place at room temperature + Reaction can be readily visualised + Reaction can be readily adapted to cover other serovars of *Mtb* and drug-resistant strains + Culture-free + Can use multiple sample sites	– Non-specific targeting is feasible – Efficacy not established for paediatric or HIV/Immunocompromised patients
Deeplex-MycTB (Genoscreen, France)	NA	NA	+ Detection of Rifampicin resistance supersedes GeneXpert MTB/RIF[141] + Provides resistance profiles to 15 drugs [142] + Can be multiplexed to analyse up to 384 samples [142] + Identifies more than 100 different mycobacterial species [142] + Culture-free	– Requires storage at -20ºC – Longer turnaround compared to other methods (48 hours) – Requires PCR to be performed before sequencing – Requires access to web application – Requires molecular biology lab – Requires expensive sequencing equipment – May miss novel sequences conferring resistance – Requires sputum
Whole Genome Sequencing (WGS)	>95%[143]	>95%[143]	+ Detects drug resistance + Whilst price is high, it is continuously decreasing + Enables global surveillance of strains	– Longer turnaround compared to other methods (up to 72 hours) – Some techniques are prohibitively expensive (>$95 per sample) – Requires uninterrupt ble power supply – Requires sputum – May miss novel sequences conferring resistance
MinION Nanopore sequencing (Oxford Nanpore technologies, UK)	94.8%[144]	98%[144]	+ Portable + Can detect drug resistance + Cheaper than WGS (<$71 per sample) + Has been deployed in the field[145] + Quick turnaround (<6 hours)	– Requires uninterrupt ble power supply – High initial start-up costs – Full characterisation to take place – Requires sputum
miRNA	24.7 – 39.9% [146](95% CI[5])	>90%[146]	+ Culture-free + Quick turnaround (<24 hours) + Blood sample	– High variability in results – Effective miRNA yet to be uncovered – Must be utilised regularly
Aeonose (eNose BV, Netherlands)	75-92% [147] (95% CI)	44-65% [147] (95% CI)	+ Non-invasive + Highly portable + Rapid turnaround (<10 mins) + Battery powered	– Lower sensitivity and specificity in men, smokers and people with a high Body Mass Index – Requires internet access – Cannot determine drug sensitivity – Efficacy in LTBI and immunocompromised yet to be established
Raman spectroscopy	Active TB: 84.62% [148] LTBI: 86.84%[148]	Active TB: 89.47%[148] LTBI: 65%[148]	+ Portable + Battery powered + Low cost per sample (<$7) + Quick turnaround (<2 hours) + Blood sample	– No efficacy established for HIV/Immunocompromised and paediatric patients – Small sample size in cited paper (n = 118) – No analysis of drug-resistance

AI Processing	68-96%[96] (95% CI)	72-85%[96] (95% CI)	+ Can be analysed on remote servers + Non-invasive + Diminishes human error	– Requires mobile phone data access – Preliminary results vary between populations
Arginine film	NA	NA	+ Quick turnaround (<2 hours) + Culture-free + PCR free + Reported sensitivity down to 4.4nM of DNA[149]	– Requires sputum sample – Requires un interruptible electricity supply

Table 2 - Diagnostic techniques currently in development with the associated strengths and weaknesses, and their associated sensitivity and specificity (alongside a 95% confidence interval), taken from Macgregor-Fairlie *et al.* 2020 (71)

1.5. Epidemiology of Tuberculosis:

The spread of TB across the globe is currently focused within South-East Asia (in 2021, this region accounted for 45% of TB cases globally (1)). Of the countries most affected by TB, approximately two thirds of all global cases can be accounted for by a mere eight countries (150):

1. India
2. Indonesia
3. China
4. The Philippines
5. Pakistan
6. Nigeria
7. Bangladesh
8. The Democratic Republic of the Congo

[4] Not Available
[5] Confidence interval

As illustrated in **Figure 3,** below:

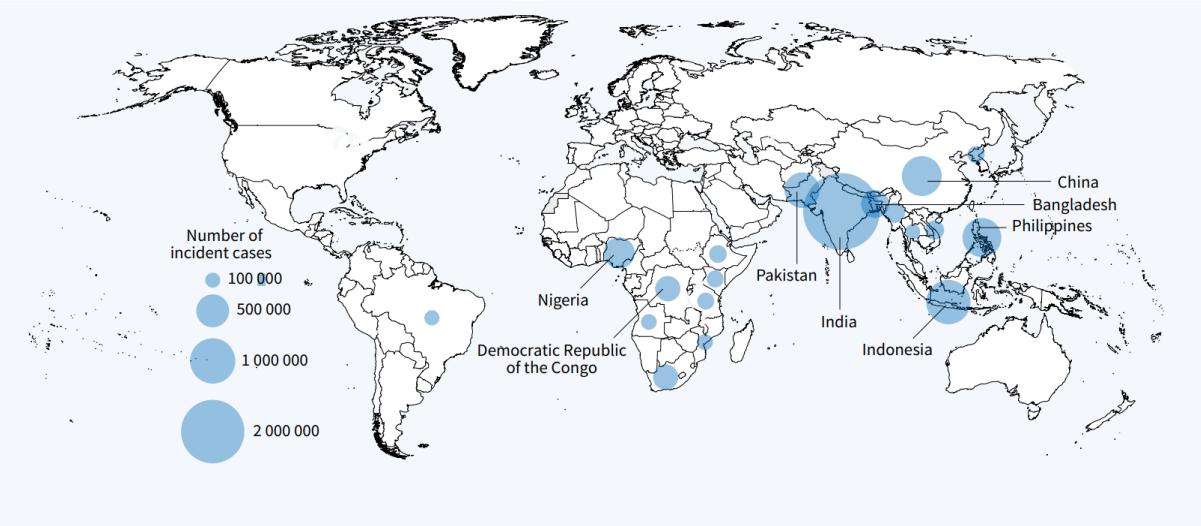


Figure 3 - Estimated TB incidence in 2021, for countries with at least 100 000 incident cases (1)

However, of all the cases in 2021, approximately 6.7% occurred in people living with HIV (1), though the prevalence of co-infection can vary greatly between countries and regions. For instance, in some parts of Southern Africa, this can exceed 50%.

When discussing drug resistance, three countries account for 42% of global cases of MDR-TB (1):

1. India
2. The Russian Federation
3. Pakistan

This often presents as rifampicin resistance. However, other forms of resistance are increasingly emerging, both from a genotype and phenotype perspective. Additionally, due to the ongoing COVID-19 pandemic, the level of resistant TB

notifications has marginally increased, likely due to a reluctance of patients to pursue treatment and diagnosis (151).

The WHO notes that there is a significant gap between TB notifications and estimates; this primarily arises due to a lack of surveys, and it has been proposed that such surveys should be held more frequently. Other objectives that countries should be striving to achieve is to regularly confirm the presence of TB using bacteriological methods, though this presents with a great number of issues due to infrastructural and logistical demands.

In terms of overall mortality, the majority of TB deaths arise from four countries (1):

1. India
2. Indonesia
3. Myanmar
4. The Philippines

Globally, TB is the 13th leading cause of death ((152) **Figure 4**, opposite), and, prior to the emergence of COVID-19, it was the single deadliest infectious agent on an annual basis. COVID-19, however, has also impacted the fight against TB. As a result, the WHO has reported an increased in global deaths from TB, despite downward trends prior to the pandemic. With the primary reason being that those suffering from TB being unlikely to attend healthcare facilities due to concerns arising from COVID-19 (71).

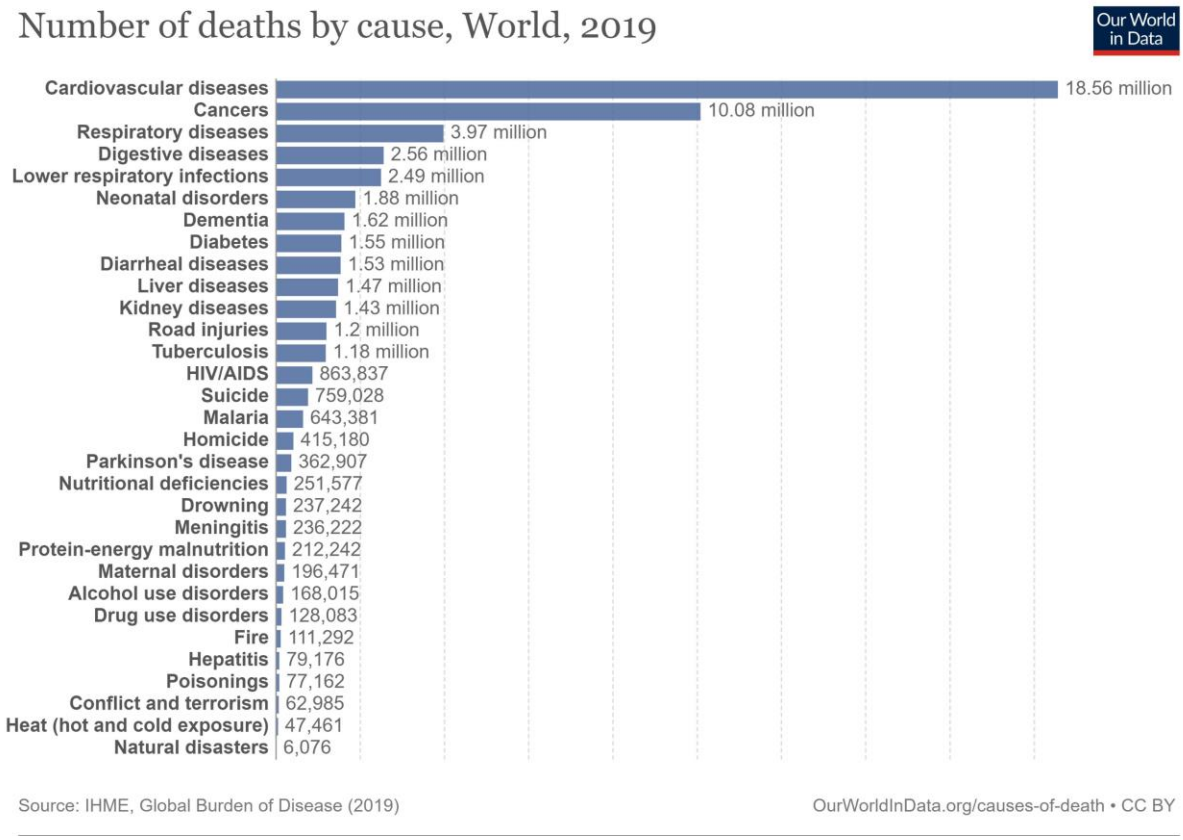


Figure 4 - Global number of deaths by cause, 2019 (152)

The interaction between COVID-19 and latent TB is still being elucidated, and it is not known if COVID-19 infection can increase the likelihood of latent TB becoming active, though there are case studies demonstrating that this is feasible, but it does not appear to be widespread (153).

Lastly, in terms of the spread of TB, two notable predictors of TB prevalence are Gross Domestic Product (GDP)(Per capita) and malnutrition. Wherever there is a low GDP per capita combined with high prevalence of malnutrition, the probability of developing TB disease is substantially higher, though this is multi-faceted. For instance, a decline in income could make an individual less likely to pursue health care (154, 155) .

1.6. Bovine Tuberculosis:

So far only human tuberculosis has been discussed at length. However, often neglected is the impact of bovine tuberculosis. The aetiology of bovine TB is similar to human TB with one exception: The primary causative agent is *M. bovis* (156). Though symptoms, even in animals, present in a similar fashion to humans: Bloody cough, and weight loss.

Bovine TB can be spread to humans *via* infected animals (either directly, or to abattoir workers during slaughter and the subsequent aerosolisation of the bacteria, as well as unpasteurised dairy products (157).

The primary diagnostic method for bovine TB is similar to humans: Gamma interferon release assays or a tuberculin skin test (157). On the other hand, treatment differs considerably.

Naturally, in humans, there is a focus on treatment until there is complete resolution of symptoms. However, the same cannot be said for bovine TB. As a result, when a cow is positively identified as having bovine TB, the movement of the cattle is severely restricted and, often, the infected animals are culled (158), in accordance with EU directive 2016/429 EEC (159), which applies to both the EU and UK.

Due to the above, bovine TB often has a significant impact upon farmers and the wider economy. In the UK alone, it was estimated that bovine TB costs the UK taxpayer approximately £150 million per year (160), and this is only set to increase as time progresses.

Additionally, within the context of the above and with some farmers still struggling to recover following the foot and mouth disease outbreak in 2001 (161), it has become the policy of the British government to undertake a highly controversial badger cull (162).

Whilst bovine TB predominantly affects cattle, *M. bovis* can infect other animals and has been documented in numerous species, including dogs, cats, deer, camels and, most importantly, badgers (163).

The data on the impact of mass badger culling is inconclusive, however, one of the more recent meta-analyses on the topic has concluded that the impact of a cull is exceptionally limited (164). Meanwhile, other researchers believe that it may have a contradictory effect due to the rapid dispersion of infected badgers in areas where culling has taken place (164).

Regardless, the economic impact of bovine TB cannot be denied. *M. bovis* is significant for a variety of other reasons. Primarily, vaccination for TB has relied upon the use of an attenuated strain of *M. bovis* called Bacille Calmette-Guérin (BCG), which shall be discussed below.

1.7. TB Vaccination:

Vaccination is the process of inducing immunity within an individual by introducing a component (or derivative) of a pathogen into a person in the hopes of reducing the severity of or preventing an infection from establishing itself. Relevant to TB is the use of *M. bovis* BCG as mentioned above.

M. bovis BCG is the result of approximately 230 passages of *M. bovis* taken from the udder of an infected cow and then grown on glycerin, potato and ox bile medium over the course of 11 years. Following the years of effort by Albert Calmette and Camille Guérin, a significantly less pathogenic yet somewhat immunogenic organism had been generated (165).

However, the BCG vaccine has several problems associated with it. First and foremost, the vaccine is relatively poor in terms of efficacy in preventing infection (*ca.* 30%)(166). As for its ability to prevent an active infection arising from a latent infection, its efficacy is approximately 71% (167). The vaccine itself is also contraindicated in patients who are immunosuppressed (*e.g.* HIV patients) (168), which is counterproductive given that HIV patients are significantly more likely to develop active disease and are therefore prime candidates for a TB vaccine.

Secondly, the vaccine tends to form an oozing ulcer at the site of the injection which often leaves a scar (169). This occurs at such a frequency that the presence of such a scar is often considered the hallmark of TB vaccination.

Of course, there are multiple vaccine candidates in development for TB, however, at the time of writing none have come close to being released into the market, and this does not seem likely to change within the near future (170).

Ultimately, the BCG vaccine is a tool that can be utilised in the fight against TB, but it is worth noting that it is not the ultimate preventative measure against TB as vaccination is with other diseases (such as measles). Additionally, it is only most beneficial when used in populations with a high infection rate of TB but cannot be

used in populations where it is contraindicated (for instance, populations with a high HIV burden) and it is this latter fact where it becomes more difficult to reconcile BCG vaccination with a high level of HIV infection, as is the case in some countries such as South Africa.

Regardless, the WHO recommends the use of the BCG vaccination in children, as a combative effort to wipe out TB.

In terms of physiology, *M. bovis* BCG possesses several physical and genetic elements which make it distinct from *M. bovis* and, ultimately, reduces its pathogenicity. As a result, unlike *M. bovis* and *M. tuberculosis*, *M. bovis* BCG can be handled safely within a laboratory with a biosafety level of 2. As a result, *M. bovis* BCG shall be used as a model organism for bovine and human tuberculosis within this project (171).

1.8. Mycobacterium Cell Wall Structure:

Mycobacterium sp., possess a rich cell wall which differs greatly from most other pathogenic bacteria. In terms of the actual cell wall structure, (49) covers this extensively; this is further added to by Dulberger *et al.* (172) as per **Figure 5**, opposite.

Research has often considered the use of Raman spectroscopy on MTBC but this has a tendency to focus primarily on mycolic acid as the main spectral element that allows for detection of the bacteria (173); this could present with false positives given that there are other bacteria within the respiratory tract that also contain mycolic acid within their cell walls, prior to the publication of this thesis, no research has been

performed that compares the spectra from *Mycobacterium sp.* with other mycolic acid containing species of bacteria from the order Mycobacteriales.

Given the conserved nature of mycolic acids within the cell walls across MTBC and other Mycobacteriales (174), it is plausible that the Raman spectra generated from the analysis of these bacteria could broadly overlap. It is therefore necessary to consider other molecules which may contribute to building a better picture of the Mycobacterial cell wall, as both Dulberger *et al.* (172) and Cook *et al.* (49) detail.

MTBC generally present with an especially thick cell wall that contains several peptidoglycan-arabinogalactan polymers that are unique to *Mycobacterium sp.* (175) and these compounds (as well as their interaction with the mycolic acid on the cell surface) are more likely to be a viable diagnostic target to identify within the cell wall of any *Mycobacterium sp*. By identifying these and incorporating their unique Raman spectra into a spectral barcode which enables healthcare staff to identify spectral

elements which are unique to the bacteria within the MTBC, the prospect of a rapid Raman (or spectral)-based diagnostic test gets closer to being a reality.

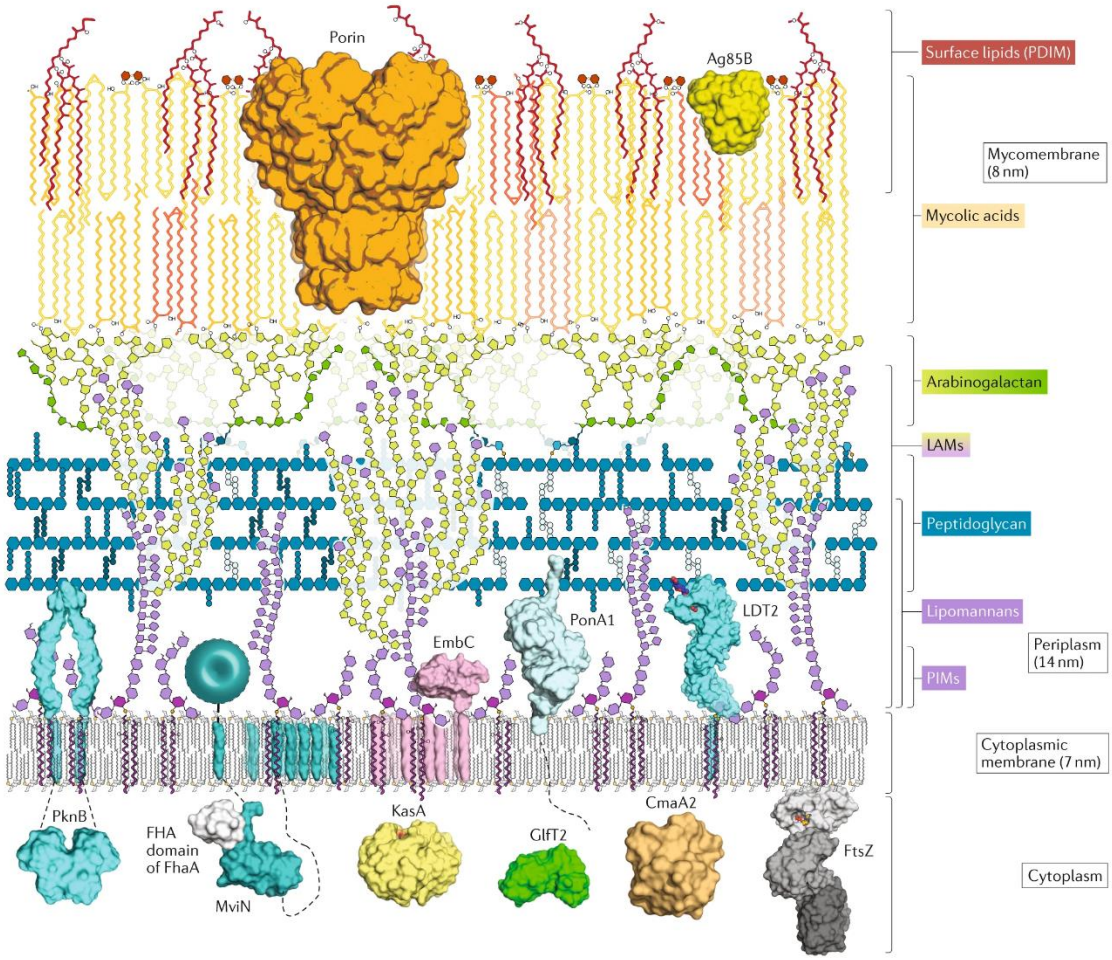


Figure 5 -The cell envelope layers are populated by proteins that build and modulate the cell envelope.

Proteins discussed in the Review for which structural information is available are PknB (Protein Data Bank (PDB) entries 2KUI and 1MRU), Ag85B (PDB entry 1F0P), GlfT2 (PDB entry 4FIY), KasA (PDB entry 4C6X), FtsZ (PDB entry 4KWE), PonA1 (PDB entries 5CXW and 5HLD), EmbC (PDB entry 3PTY); MviN–FhaA forkhead-associated (FHA) domain complex (PDB entry 3OUN), CmaA2 (PDB entry 1KPI), L,D-transpeptidase 2 (LDT2; PDB entries 3TUR, 5DUJ) and MspA porin (PDB entry 1UUN). Dashed lines represent protein regions that are not resolved in crystal structures. Phosphoinositol mannosides (PIMs) are dark purple (inositol) and light purple (mannose) and are short. Lipomannans (LMs) are derived from PIMs and are shown with more light purple mannose residues. Lipoarabinomannans (LAMs) are similar to LMs but have branched, light green arabinose chains. Peptidoglycan is blue; dark blue crosslinks are D–D, light blue crosslinks are L–D. Arabinogalactan is dark green (galactan) and light green (arabinan). Trehalose is represented by two dark orange hexagons and is linked to the mycolic acids in trehalose monomycolate (TMM) and trehalose dimycolate (TDM). For mycolic acids, light yellow indicates α-cis/cis-cyclopropane; dark yellow indicates keto, trans-cyclopropane; light orange indicates keto, cis-cyclopropane; orange indicates methoxy, trans-cyclopropane; and dark orange indicates methoxy, cis-cyclopropane. PDIM is red. Many classes of surface and cytoplasmic membrane lipids are not included. (172)

Other structures and compounds that are present within the cell wall of MTBC are also worth considering; these could include specific proteins such as MmpL3 (176), lipids (177), or trehalose esters or glycolipids (178). It therefore becomes necessary that any diagnostic technique which considers the physical characteristics of the MTBC (such as Matrix-Assisted Laser Desorption/Ionisation-Time-Of-Flight or MALDI-TOF) must also consider the entire physical structure as opposed to a specific element present within the bacteria and must be fully characterised, as well as noting particularly unique spectral elements of the cell wall membrane.

Mycolic acids form the largest constituent part of the cell wall of MTBC; they are also the definitive spectral elements that Mühlig *et al.* (173) uses to identify *Mycobacterium sp.* in the vast amount of current literature. The more up-to-date Stöckel *et al.* (179) discusses these spectral motifs within the context of some bacteria that reside within the microbiome, however the paper fails to consider the effects of other mycolic acids within other bacteria that can reside within the microbiome as well as failing to further analyse other.

Mycolic acids themselves are long-chain 2-alkyl branched, 3-hydroxy fatty acids and are ubiquitous within the outer coat of the Mycobacteriales order. On the cell wall of *Mycobacterium sp.* they can be found, either, as free acids, or esterified to arabinogalactan, trehalose, glucose or glycerol (180).

The structure of the acids themselves vary between species; Marrakchi *et al.* (181) discusses these at length, and the generalised structure of these compounds is illustrated in the **Figure 6**, opposite:

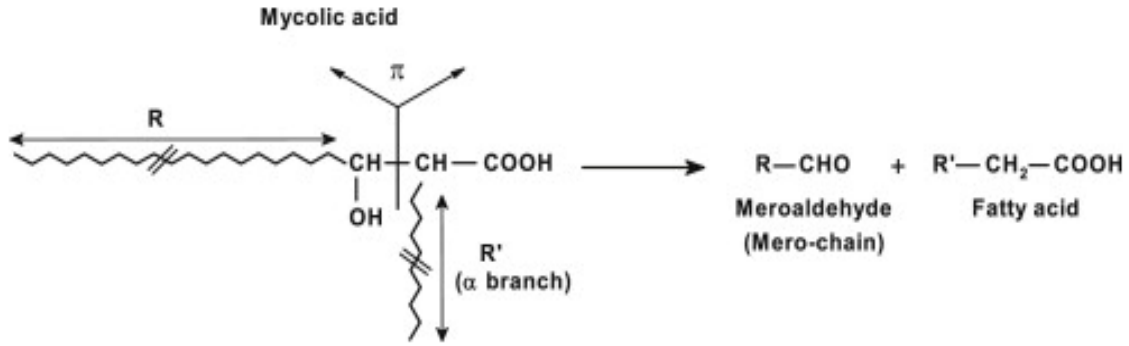


Figure 6 - Generalised formula for mycolic acids

In terms of structures specific to *Mycobacterium spp*., (182) addresses these directly, in **Figure 7**, opposite:

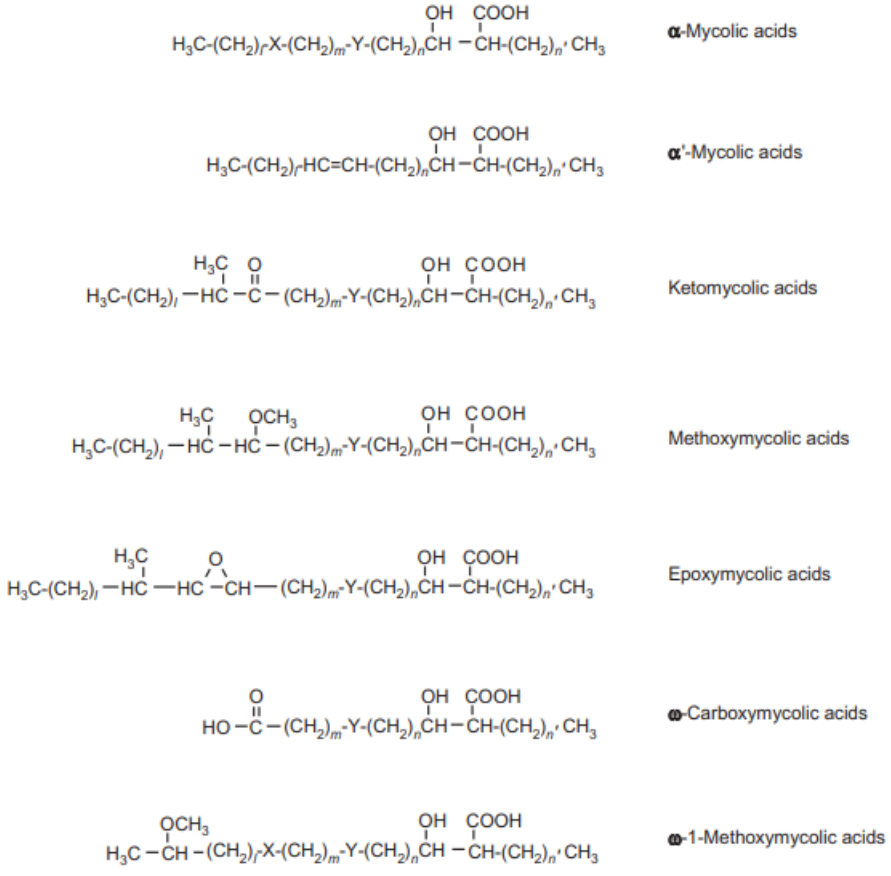


Figure 7 - Generalised formulas for the Mycolic acids in the cell walls of *Mycobacterium sp*. (155)

The proportion of the above mycolic acids on a bacterium can allow researchers to characterise the bacteria that is being analysed (183). However, the extent to which this allows scientists to differentiate between *Mycobacterium sp*. and other mycobacteriales has not been established. Yassin *et al.* (182) goes on to illustrate that there is an overlap within the size and structure of these compounds in the **Table 3** below:

Genus:	**Total number of carbons:**	**Number of double bonds:**	**Pyrolysis products:**
Corynebacterium	22-36	0-2	8-16
Mycobacterium	60-90	2	22-26
Rhodoccocus	30-56	0-4	12-18

Table 3 - Variations in size and structure of mycolic acid within the mycolic acid containing genera of the order Actinomycetales (182)

Mycobacteriales contain a variety of compounds and components which are likely to produce Raman spectra that are unique to those organisms, though it is plausible that the mycolic acid make up the primary constitutive part of the spectra.

Therefore, if any Raman-based assay were to attempt to incorporate mycolic acids into a test for TB, it would need to consider the configuration of mycolic acids and be able to differentiate these by applying chemometric methods.

1.9. Scope of this thesis:

M. tuberculosis and the MTBC continues to be a pathogen of significant concern across the globe and claims the lives of millions of people every year. Effective diagnosis remains an issue in identifying TB cases due to a wide variety of reasons

as identified above and in Macgregor-Fairlie *et al.* (74). Though the vast majority of these issues arise due to a lack of infrastructure and access to specialist facilities.

Vaccination in and of itself is not a completely effective tool in preventing the proliferation of TB between individuals, especially those co-infected with HIV. Meanwhile, treatment (for the most part) remains incredibly effective but for it to be truly effective, we must improve on and add to our diagnostic methods for TB.

The WHO recommends the combined approach of prevention, monitoring/diagnosis and treatment, with the bulk of the focus being on effective diagnosis. Raman spectroscopy is a potential workaround for this, and there is potential for this technique to be deployed in the field for the effective diagnosis of TB.

Additionally, whilst Raman spectroscopy has been utilised in characterising *M. tuberculosis* and other *Mycobacterium sp.,* no research has been undertaken in understanding the Raman spectra of the bacteria that occupy the same ecological niche nor other similarly related bacteria. In essence, previous research has failed to look at the bigger picture.

Therefore, the primary aims of this thesis are as such:

1. The construction of an *in silico* Raman library of key compounds present on the cell wall of *Mycobacterium sp.*
2. The Raman analysis of cell wall extracts of *M. tuberculosis, M. smegmatis* and *M. bovis* using optical trapping to consolidate the information provided by the *in silico* analysis above.

3. Identification and Raman analysis of bacterial genera inhabiting the respiratory tract of *B. taurus,* and a further comparitive analysis of the Raman spectra of *M. bovis* BCG and other Mycobacteriales.
4. An investigation into the limit-of-detection of Raman spectroscopy in a representative artificial sputum sample, containing the bacteria identified in point 3.
5. The design of a theoretical CRISPR/Cas-based assay which seeks to identify *M. tuberculosis*, in addition to genes which confer drug resistance. ***NB:*** This part of the thesis was initially covered at the outset of this project, however, due to the COVID-19 pandemic, it was later abandoned. Due to the work that was carried out, this part of the project is now contained in the appendix.

The discussion sections of each chapter will be limited to technical evaluations of the data generated, with further discussion on key aspects of the results within the context of the published literature and current diagnostics in chapter 7.

Chapter 2:

2.1. Bacteria, Molecules and Compounds:

N.B: Scanning Electron Micrographs (where uncited) were generated by the author under guidance from Regan Doherty and Catherine Griffiths at the Southampton Biomedical Imaging Unit, University Hospital Southampton NHS trust. For the methodology used to acquire the SEM images, see

A.2. SEM preparation protocol for bacteria:.

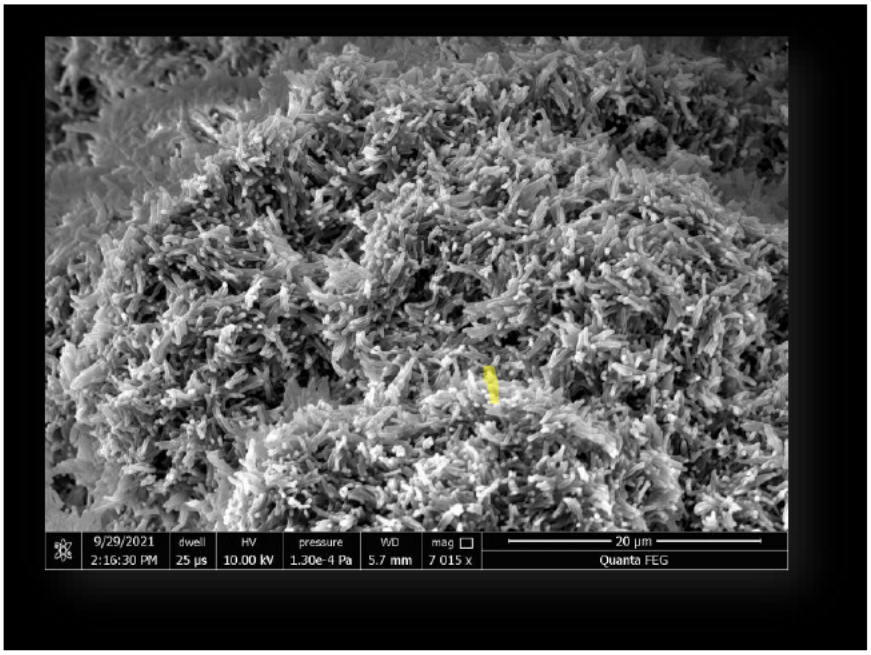


Figure 8 - Scanning Electron Micrograph of *M. bovis* BCG taken at 7,015x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the entire picture. An individual cell is highlighted.

2.1.1. *Mycobacterium bovis, Mycobacterium tuberculosis, & Mycobacterium smegmatis:*

Mycobacterium bovis BCG (NCTC 5692) makes up the most significant bacterial species that this project will consider. Whilst *Mycobacterium tuberculosis* is the most clinically significant species within the *Mycobacterium* genus, it is a pathogen with safety requirements beyond those that the School of Chemical Engineering at the University of Birmingham can provide and it therefore cannot be safely handled within the context of this project.

Genetic and structural elements from *M. tuberculosis* shall be considered in **Chapter 3, 4, 5, 6 & A.1**.

Meanwhile, *M. bovis* BCG will be considered in **Chapters 5, 6 & A.1**.

M. bovis BCG (NCTC 5692) was acquired from the National Collection of Tissue Cultures (NCTC), UK Health Security Agency (Formerly Public Health England), Porton Down, UK.

Cell wall lipids of *M. tuberculosis* H37Rv and *M. smegmatis* Mc2 155 were kindly provided by the Bhatt group from the University of Birmingham's Institute of Microbiology and Infection.

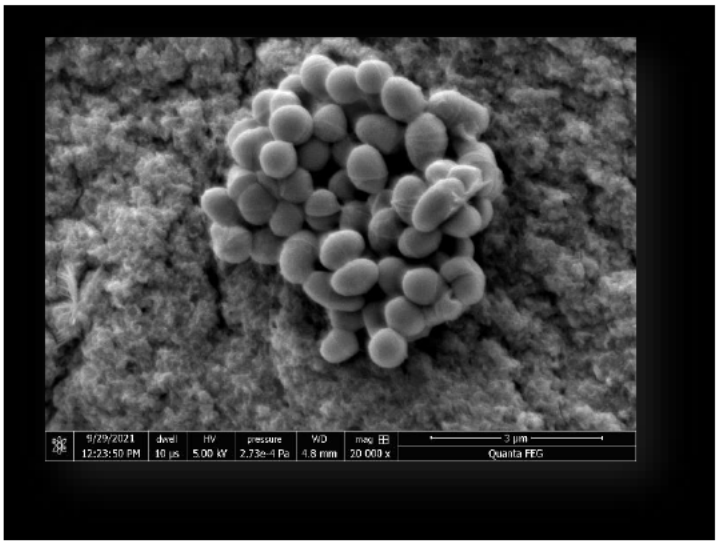


Figure 9 - Scanning Electron Micrograph image of *Mycoplasma bovis* taken at 20,000x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the centre of the picture. Individual cells can be readily visualised.

2.1.2. *Mycoplasma bovis*:

Mycoplasma sp. are a remarkable genus of bacteria, in that they do not possess a cell wall. Making up a significant proportion of the microbiome of *B. taurus*, they

are opportunistic pathogens and are resistant to a notable number of antibiotics. For the sake of clarity, *Mycoplasma bovis* shall be referred to using its full name to avoid confusion with *Mycobacterium bovis*.

Mycoplasma bovis was readily utilised in **Chapters 5 & 6**.

Mycoplasma bovis (NCTC 10131) was acquired from the National Collection of Tissue Cultures (NCTC), UK Health Security Agency (Formerly Public Health England), Porton Down, UK.

2.1.3. *Mycoplasma hyosynoviae:*

Mycoplasma hyosynoviae is another member of the *Mycoplasma* genus which forms part of the microbiome of *B. taurus.* Like other members of the *Mycoplasma* genus, they are opportunistic pathogens and resistant to antibiotics. They also lack a cell wall.

M. hyosynoviae was utilised in **Chapters 5 & 6**.

M. hyosynoviae were kindly gifted by Andrew Rycroft, the Royal Veterinary College, London, UK. Identify was confirmed *via* PCR of the conserved hypothetical protein mhp165 and through whole genome sequencing.

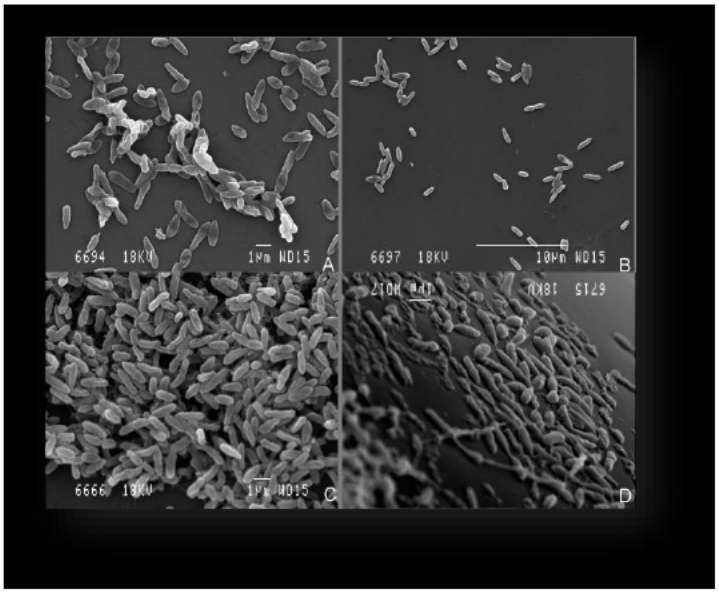


Figure 10 - Scanning Electron Micrograph image of *P. aeruginosa* isolates attaching to glass surfaces. Isolates can be observed in the top two boxes, whilst more clustered cells can be observed in C and D. Magnification was not stated in the paper. Taken from Deligianni *et al.* (184)

2.1.4. *Pseudomona aeruginosa:*

Pseudomonas aeuginosas is a common commensal of animals, which is often an opportunistic pathogen. *P. aerguinosa* was utilised in **Chapters 5 & 6**.

P. aeruginosa was acquired from the Overton group in the School of Chemical Engineering at the University of Birmingham, UK.

2.1.5. *Pseudomonas fluorescens:*

Pseudomonas fluorescens is a common commensal of animals, which is often an opportunistic pathogen. *P. fluorescens* was utilised in **Chapters 5 & 6**.

P. fluorescens was acquired from the Overton group in the School of Chemical Engineering at the University of Birmingham, UK.

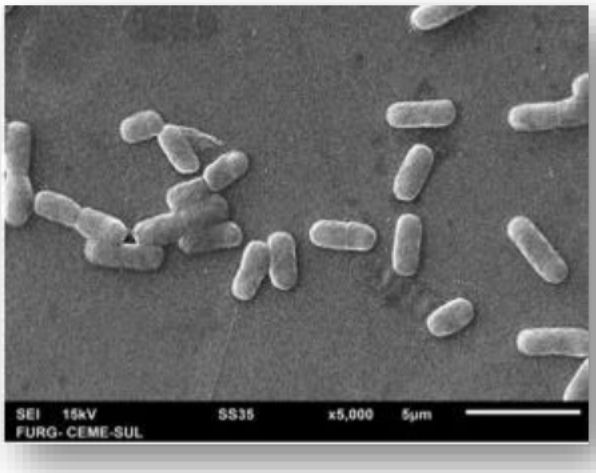


Figure 11 - Scanning Electron Micrograph image of *B. cereus* taken at 5,000x magnification. Individual cells can be readily visualised. Taken from Filho *et al.* (185)

2.1.6. *Bacillus cereus:*

Bacillus cereus (NCTC 11143) is a common commensal of animals and can cause food-related illness in humans. Additionally, the bacteria is well-known to be spore-forming. *B. ceresus* was utilised in **Chapters 5 & 6.**

B. cereus was acquired from the biology teaching labs in the School of Chemical Engineering at the University of Birmingham, UK.

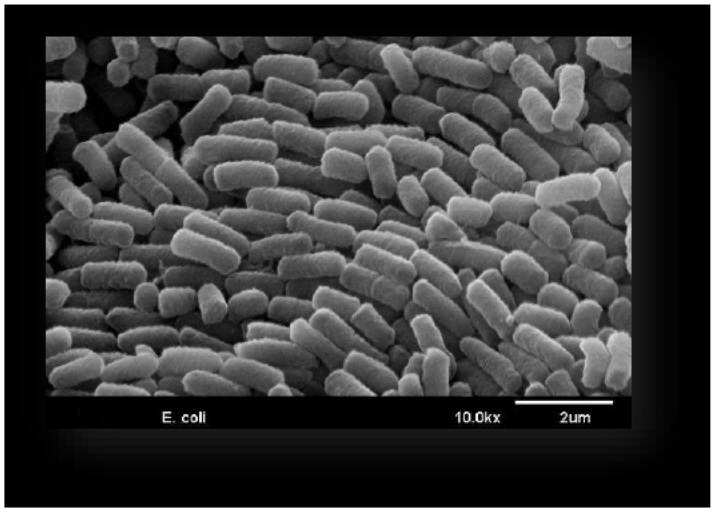


Figure 12 - Scanning Electron Micrograph image of *E. coli* taken at 10,000x magnification. A cluster of bacteria dominates the picture. Individual cells can be readily visualised. Taken from Krishnakumar (186)

2.1.7. *Escherichia coli:*

Escherichia coli is a common commensal bacteria which colonises the digestive tract of both animals and humans. Occasionally, it can cause disease, as well as opportunistic infections such as cystitis. *E. coli* was utilised in **Chapters 5 & 6.**

E. coli TOP10 were used in this project and acquired from Fisher Scientific, Loughborough, UK. *E. coli* TOP10 is a proprietary strain of *E. coli* that incorporates several genetic elements which foster the ability of the *E. coli* to transform and clone heterologous genes.

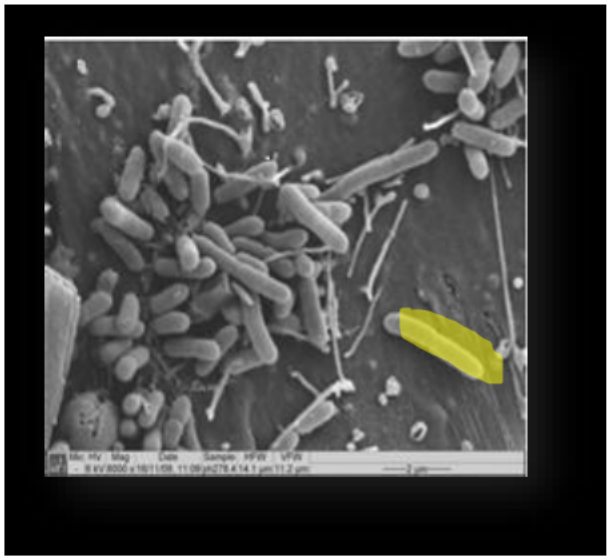


Figure 13 - Scanning Electron Micrograph image of *M. haemolytica* taken at 8,000x magnification. A cluster of bacteria dominates the picture. Note the extracellular matrix that the bacteria have generated. Individual cells can be readily visualised. Two cells undergoing fission are highlighted. Taken from Haig (187)

2.1.8. *Mannheimia haemolytica:*

Mannheimia haemolytica (NCTC 10638) is a causative agent of bovine respiratory disease and is capable of causing disease within cattle. It is often present within the microbiome. *M. haemolytica* was used in **Chapters 5 & 6.**

M. haemolytica was purchased from NCTC, UK Health Security Agency, Porton down, UK.

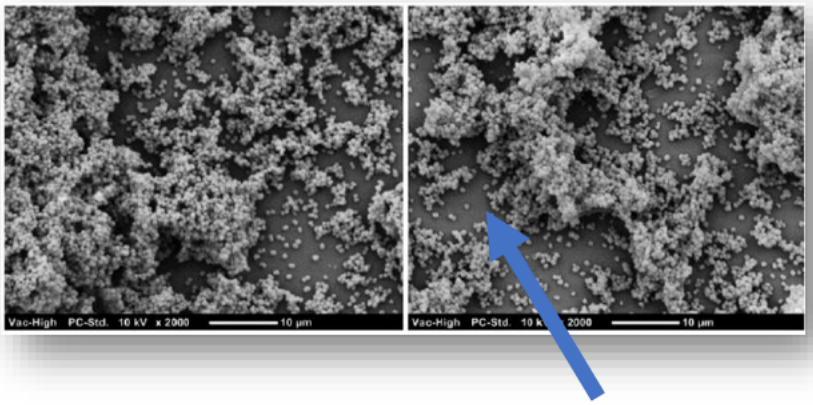


Figure 14 - Scanning Electron Micrograph of *M. catarrhalis*, taken at 2000x magnification. A vast biofilm forms in the centre of the micrograph. The arrow present points to an individual cell. Taken from Tan *et al.* (188)

2.1.9. *Moraxella catarrhalis:*

Moraxella catarrhalis (NCTC 11020) is a commensal bacteria of the bovine respiratory tract, and is capable of causing disease in both humans and cattle. *M. catarrhalis* was used in **Chapters 5 & 6**.

M. catarrhalis was purchased from the NCTC, UK Health Security Agency, Porton down, UK.

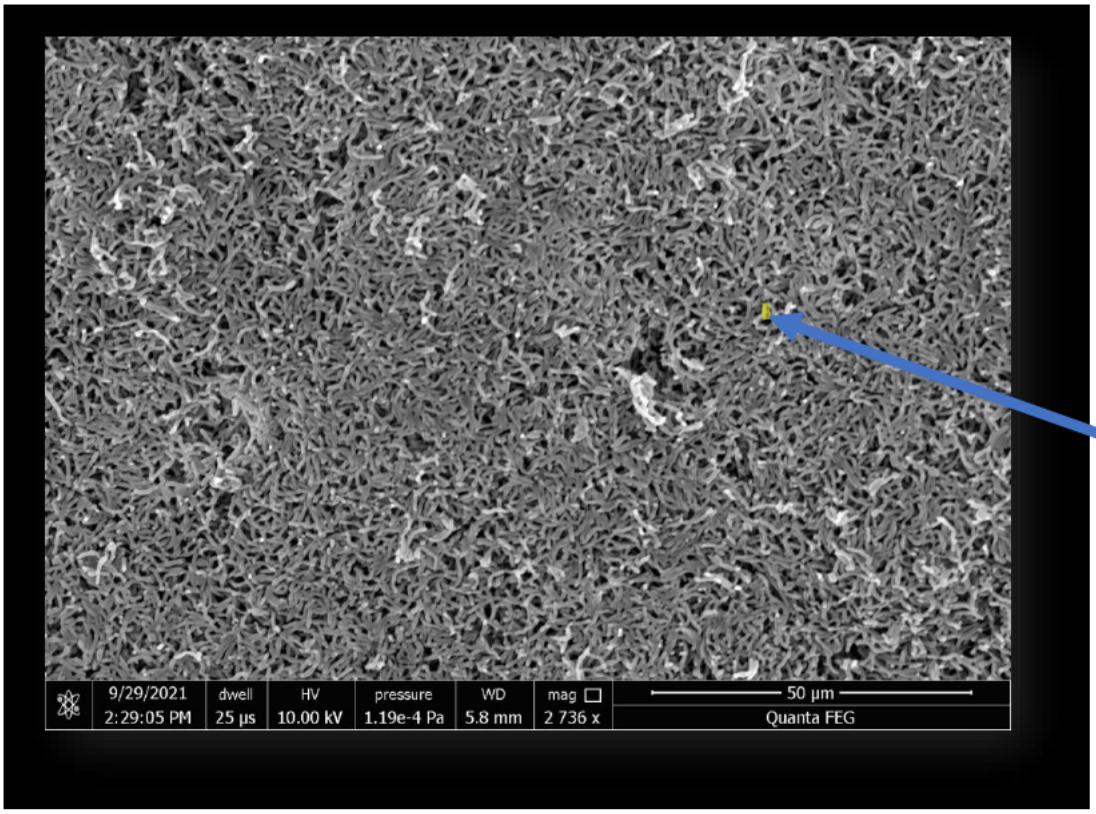


Figure 15 - Scanning Electron Micrograph of *R. erythropolis* taken at 2,736x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A cluster of bacteria dominates the entire picture. An individual cell is highlighted in the picture, with an arrow emphasising the cell.

2.1.10. *Rhodococcus erythropolis:*

Rhodococcus erythropolis (American Type Culture Collection (ATCC 4277)) is an often soil dwelling bacteria that has been known to cause opportunistic infections in humans. The genus often forms part of the microbiome of cattle. *R. erythropolis* was used in **Chapters 5 & 6**.

R. erythropolis was acquired from the biology teaching labs in the School of Chemical Engineering at the University of Birmingham, UK.

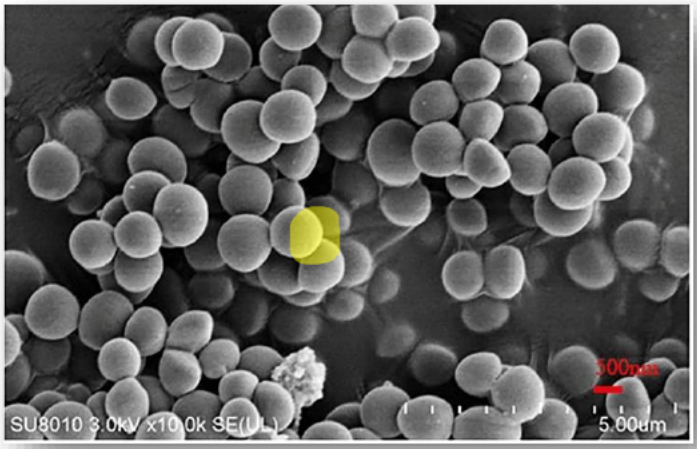


Figure 16 - Scanning Electron Micrograph of *S. aureus* taken at 10,000x magnification. A cluster of bacteria dominates the e picture. An individual cell is highlighted in the picture. Taken from Zeng *et al.* (189)

2.1.11. *Staphylococcus aureus*:

Staphylococcus aureus (NCTC 10788) is a prolific organism which is present in the microbiome of both humans and animals. It is capable of causing destructive disease and known to acquire antibiotic resistance. Generally speaking, the organism resides on the skin as well as in the upper and lower respiratory tracts. *S. aureus* was utilised in **Chapters 5 & 6.**

S. aureus was purchased from the NCTC, UK Health Security Agency, Porton down, UK.

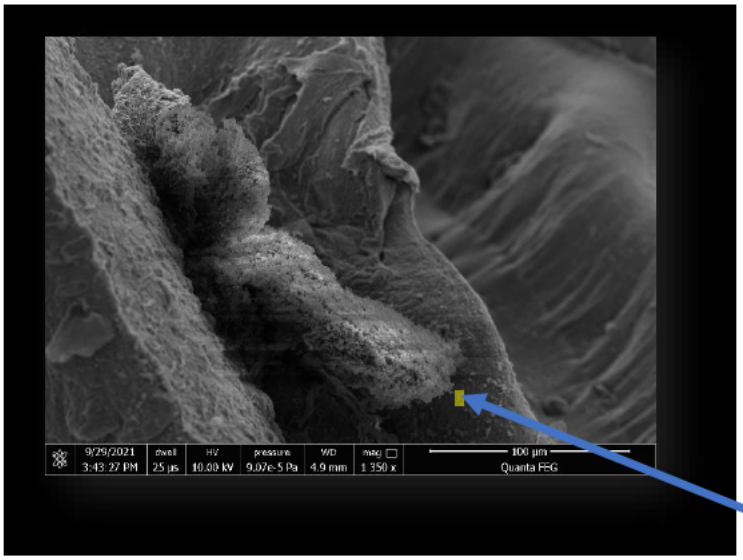
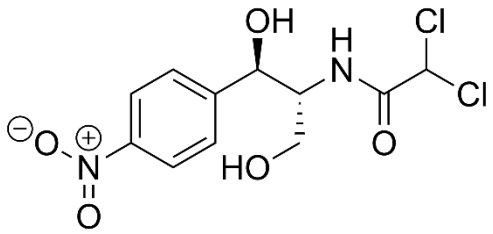


Figure 17 - Scanning Electron Micrograph of *C. glutamicum* taken at 1,350x magnification. Acquired from the University of Southampton Biomedical Imaging Unit at University Hospital, Southampton. A biofilm of bacteria dominates the picture. A small cluster of bacteria undergoing fission is highlighted, with an arrow pointing to them.

2.1.12. *Corynebacterium glutamicum:*

Corynebacterium glutamicum (ATCC 13032) is an important organism that is often utilised in industrial production of proteins and amino acids. Additionally, the bacteria is physiologically similar to bacteria within the MTBC and other *Mycobacterium sp.;* as a result, it is often used in comparative studies. The genus is also known to cause opportunistic disease and to be a member of the microbiome. *C. glutamicum* was used in **Chapters 5 & 6.**

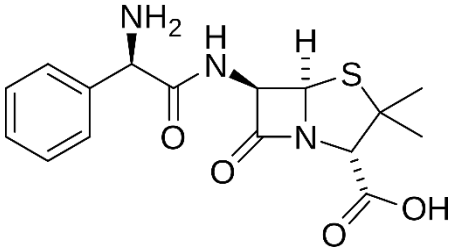
C. glutamicum was kindly donated by Professor Gurdyal Besra, School of Biosciences, University of Birmingham, UK.



2.1.13. Chloramphenicol:

Chloramphenicol is a broad spectrum and semi-synthetic antibiotic which has activity against a wide variety of Gram-negative and Gram-positive organisms. Chloramphenicol's mechanism of action is *via* inhibition of protein synthesis.

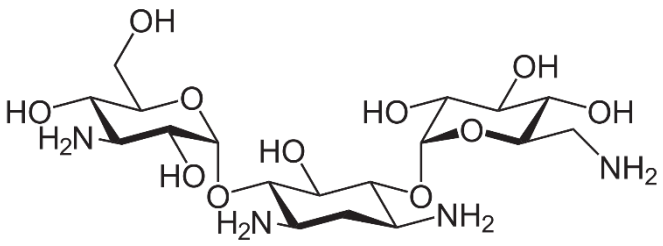
Chloramphenicol was acquired from Fisher Scientific, Loughborough, UK



2.1.14. Ampicillin:

Ampicillin is a β-lactam antibiotic which inhibits cell wall synthesis. Whilst most β-lactams target only Gram-positive bacteria, due to an extra amide group, ampicillin is able effective against Gram-negative bacteria.

Ampicillin was acquired from Fisher Scientific, Loughborough, UK.



2.1.15. Kanamycin:

Kanamycin is a protein synthase inhibitor. It primarily targets Gram-negative bacteria and inhibits relevant enzymes involved in protein synthesis.

Kanamycin was acquired from Fisher Scientific, Loughborough, UK.

2.2. Sample Preparation Methodologies:

2.2.1. Bacterial Cultures & Purification Methodologies:

2.2.1.1. *Mycobacterium bovis BCG & Mycobacterium tuberculosis 7H9:*

M. bovis BCG was initially cultured on Middlebrook 7H11 agar (Fisher scientific, Loughborough, UK) + 0.2% glycerol (VWR, Lutterworth, UK) + 10% OADC growth supplement (Merck, Gillingham, UK) and then incubated at 37˚C for 4 weeks.

Colonies were then transferred to Middlebrook 7H9 (Fisher scientific, Loughborough, UK) + 0.2% glycerol + 10% OADC growth supplement and then incubated at 37˚C and 150RPM for 1 week (190).

M. tuberculosis H37Rv was cultured as above by the Bhatt group.

2.2.1.2. *Mycobacterium smegmatis:*

M. smegmatis was cultured under recommended culture conditions by the Bhatt group.

2.2.1.3. *Mycoplasma bovis:*

Mycoplasma bovis was cultured as per Cook *et al. (191)*, though 5% CO_2 was not added to the incubator. Antibiotics were also omitted.

Yeast extract was made as per the recommendations of the ATCC under “Yeast Extract #3”. Yeast extract was prepared by dissolving 150g of store bought freeze-dried baker’s yeast (*Saccharomyces Cerevisiae*. Allinson’s, Peterborough, UK) in 1000ml de-ionised water and then autoclaved at 121˚C before being allowed to settle overnight. The supernatant was then transferred into centrifuge bottles and spun at 9000RPM for 50 mins. The supernatant was then aliquoted into 100ml bottles and the precipitate discarded. The supernatant was then further autoclaved at 121˚C and the aliquots were then stored at 2-8˚C.

2.2.1.4. *Mycoplasma hyosynoviae:*

M. hyosynoviae was cultured as per Cook *et al. (191)*, though 5% CO_2 was not added to the incubator. Antibiotics were also omitted.

Yeast extract was made as per the recommendations of the ATCC under “Yeast Extract #3”. Yeast extract was prepared by dissolving 150g of store bought freeze-dried baker’s yeast (*Saccharomyces Cerevisiae*. Allinson’s, Peterborough, UK) in 1000ml de-ionised water and then autoclaved at 121˚C before being allowed to settle overnight. The supernatant was then transferred into centrifuge bottles and spun at 9000RPM for 50 mins. The supernatant was then aliquoted into 100ml bottles and the precipitate discarded. The supernatant was then further autoclaved at 121˚C and the aliquots were then stored at 2-8˚C.

2.2.1.4. *Pseudomonas fluorescens:*

P. fluorescens was cultured on solid nutrient agar media (Oxoid, Basingstoke, UK) at 30˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 30˚C and 150RPM for 24 hours (192).

2.2.1.5. *Pseudomona aeruginosa:*

P. aeruginosa was cultured on solid nutrient agar media (Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours (193).

2.2.1.6. *Bacillus cereus:*

B. cereus was cultured on solid nutrient agar media (Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours (194).

2.2.1.7. *Escherichia coli:*

E. coli was cultured on solid nutrient agar media (Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours (195).

Selective nutrient agar was also used. Respective selective nutrient agars contained 100µg/mL ampicillin, 50µg/mL kanamycin, and 12.5µg/mL

chloramphenicol. Antibiotics were dissolved in de-ionised water and then filtered through a 0.22µm syringe filter.

2.2.1.8. *Mannheimia haemolytica:*

M. haemolytica was cultured on solid colombia blood agar (196)(Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours.

2.2.1.9. *Moraxella catarrhalis:*

M. catarrhalis was cultured on brain-heart infusion agar (Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to brain-heart broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours (197).

2.2.1.10. *Rhodococcus erythropolis:*

R. erythropolis was cultured on brain-heart infusion agar (Oxoid, Basingstoke, UK) at 37˚C for 48 hours before being transferred to brain-heart broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 48 hours (198).

2.2.1.11. *Staphylococcus aureus:*

S. aureus was cultured on solid nutrient agar media (Oxoid, Basingstoke, UK) at 37˚C for 24 hours before being transferred to liquid nutrient broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 24 hours (199).

2.2.1.12. *Corynebacterium glutamicum:*

C. glutamicum was cultured on Tryptic soy agar (Oxoid, Basingstoke, UK) at 37˚C for 48 hours before being transferred to Tryptic soy broth (Oxoid, Basingstoke, UK) and subsequently incubated at 37˚C and 150RPM for 48 hours.

2.2.1.13. Bacterial preparation:

For Raman analysis, cells were centrifuged at 3900 rpm (Eppendorf Centrifuge 5810/5810 R. Eppendorf, Stevenage, UK) for 10 mins and the supernatant was removed. The cells were then resuspended in 1ml of sterile, distilled water and spun again (173). Cells were aseptically removed from the respective pellet using a 10µL inoculating loop and spread on to an autoclaved aluminium coated glass slide. The cells were then allowed to air dry under laminar flow for two hours. For the analysis of *C. glutamicum, R. erythropolis* and *M. bovis* BCG, three biological repeats (defined as the use of different, independently grown or prepared samples to assess the biological variability of an experiment) were carried out (20), whilst only one biological repeat was carried out for the remaining bacteria.. For all samples, 100x technical repeats (defined as the repetition of a measurement using the same sample and conditions) were carried out per biological repeat.

2.3. Molecular Methods:

2.3.1. Bacterial Transformation:

E. coli TOP10 were transformed in accordance with the protocol available from New England Biolabs (NEB) “High Efficiency Transformation Protocol (C2987H/C2987I)” (200).

2.3.2. Plasmid Purification:

Plasmid purification was undertaken in accordance with the with the protocol available from Thermo Scientific “GeneJET Plasmid Miniprep Kit #K0502, #K0503” (201).

2.3.3. Digestion, Phosphorylation & Ligation:

Plasmid digestion was carried out in accordance with the protocol available from New England Biolabs titled “Optimizing Restriction Endonuclease Reactions” with the addition of 0.5µg 100x BSA to make a total reaction volume of 50µL (202).

Phosphorylation and ligation of oligonucleotides took place in accordance with Marrafini *et al.* (203). Oligonucleotides were ordered from Integrated DNA Technologies, Inc. (Leuven, Belgium). Plasmids containing the relevant genes of interest were ordered from Thermo Fisher, with ampicillin as a resistance marker.

2.3.4. Gel Electrophoresis & extraction:

Gel electrophoresis was carried out in 1% agarose gel containing GelRed (BioTium, Fremont, USA) and 1x TBE buffer (Tris base 10.8gL^{-1}, boric acid 5.5gL^{-1}, EDTA 1.16gL^{-1}). pCRISPR was pooled and then 10µL BlueJuice™ Gel-Ladepuffer (10x) (Invitrogen, Inchinnan, UK) added.

The gel was then run at 120V for 60 mins.

Gel extraction was carried out in accordance with the protocol from Thermo Scientific “GeneJET Gel extraction kit, #K0691, #K0692” (204).

2.4. *In silico* Methods:

2.4.1. Avogadro:

Avogadro is an advanced molecule editor and visualiser which can generate input for computational chemistry programs such as ORCA and acts as a Graphic User Interface (GUI) for said programmes (205). Avogadro can visualise structural files of a compound in .CML format. It can also generate *in silico* Raman data following the submission of ORCA output files containing the computational information of a compound after such a file has been generated by ORCA.

Structural files were acquired in .CML format from Chemical Entities of Biological Interest (ChEBI) directory, available from the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI). Input files for ORCA were generated and optimised for use on ORCA *via* Avogadro utilising BP86 and Split valence polarization with diffuse functions.

These files were then modified with the insertion of script designed to facilitate the generation of output files remotely *via* BlueBEAR (Blue Birmingham Environment for Academic Research. The University of Birmingham's high performance computer).

Output files were then visualised on Avogadro, as was Raman spectral data. Avogadro v1.2.0 was used throughout the duration of this project and the associated work was carried out in **Chapter 3**.

2.4.2. ORCA:

ORCA is an *ab initio* quantum chemistry package that is able to generate *in silico* measurements of compounds. It has been widely cited in literature and was developed by the Max Planck Institute for carbon research (206, 207, 208). Specifically, the package has been optimised for the generation of spectroscopic readouts (209).

ORCA 5.0.3 was utilised throughout the duration of this project and was operated *via* BlueBEAR. The work associated with ORCA was carried out in **Chapter 3**.

2.4.3. BlueBEAR:

Blue Birmingham Environment for Academic Research or BlueBEAR is the University of Birmingham's high-performance computer which is capable of undertaking the large calculations required for the generation of output files which visualise the *in silico* Raman spectra of the compounds discussed in **Chapter 3.**

The computations therefore described herein were performed using the University of Birmingham's BlueBEAR HPC service, which provides a High Performance Computing service to the University's research community.
See http://www.birmingham.ac.uk/bear for more details.

2.5. Analytical Methods:

2.5.1. Optical Trapping:

Optical trapping (also known as optical tweezers) relies on the principle of radiation pressure exerted by light (210). When a laser beam passes through a

microscopic object, it exerts a force on the object due to a transfer of momentum from the photons contained within that beam (211). By utilising this force and focusing the light on to a specific area, objects can be trapped and held in place (See **Figure 18**). This trapping effect can be used on all manner of materials, ranging from single atoms(212) to cellular organisms (213).

Optical trapping took place using a 1064nm Cobolt 04-01 Series diode laser (HÜBNER Photonics, Germany) fed directly into the Leica lens via a fibre optic cable and was utilised in **Chapter 4**.

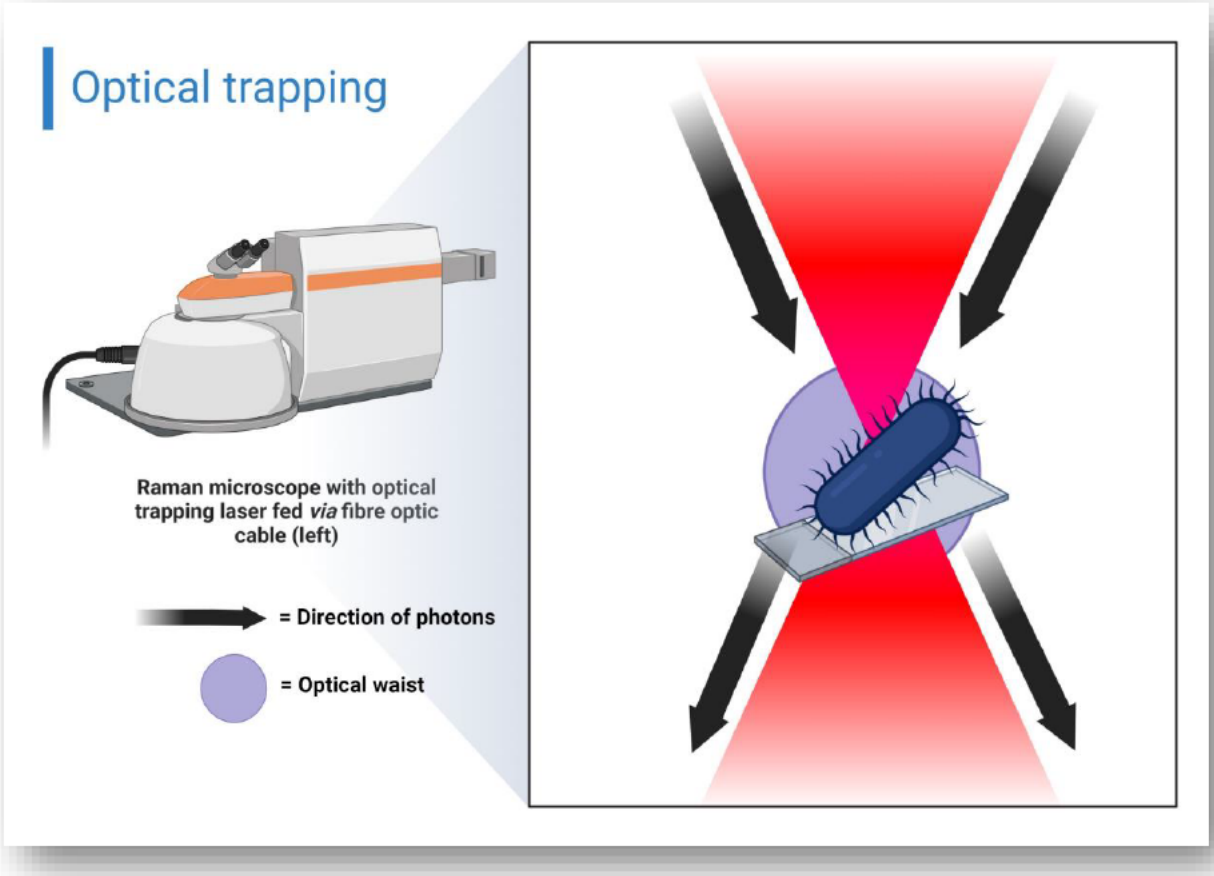


Figure 18 - Schematic of Optical trapping within the context of this project. Created by Biorender.com

2.5.2. Raman Spectroscopy:

Raman spectroscopy is a technique that measures the ability of a compound to reflect a known wavelength of light; it measures the way light is scattered when an object is hit by a laser at a specific wavelength. When this is done, it generates a fingerprint and allows the identification of an object to take place (214); this also extends to pathogens.

The underlying theory of Raman spectroscopy is based on several principles. An overview of how Raman works is illustrated in **Figure 19** below:

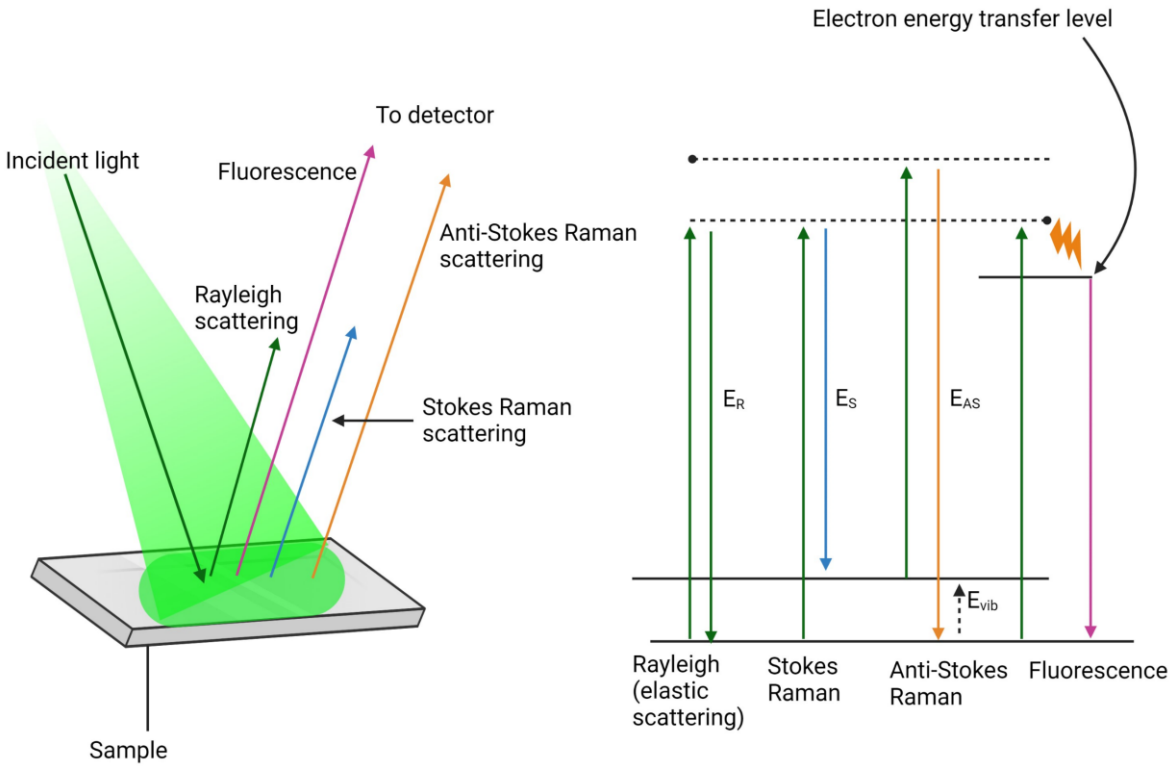


Figure 19 – An illustration of the theory or Raman spectroscopy. For reference, E_R = the energy of a Rayleigh scattered photon, E_S = the energy of a Stokes photon E_{AS} denotes the energy of an anti-Stokes photon and E_{vib} = the energy involved in a molecular vibrational transition. Created with Biorender.com

When photons strike a sample, it interacts with the molecules in various ways. Some of this light may bounce off the surface as a reflection or scatter in a way that does not change its energy (Rayleigh scattering), meaning it keeps the same wavelength as the incoming light. If the energy of the incoming photons aligns with an electronic energy level of the molecule, these photons can trigger the molecule to reach this state, and then the molecule releases a photon with a longer wavelength during its return to the normal state (a process known as fluorescence). Raman scattering, which involves energy exchange, can also occur when incoming photons cause specific molecular vibrations (unique to each chemical group). In the Stokes Raman scattering scenario, the molecule takes in energy and emits a photon with a wavelength longer than the incoming photon. Conversely, if the molecule starts from an already vibrationally excited state, it can lose vibrational energy and pass it onto the scattered photon (anti-Stokes Raman scattering), causing the emitted photon to have a shorter wavelength than the initial one.

Raman is able to rapidly identify pathogens, even when compared to other spectrometry techniques such as Matrix-Assisted Laser Desoprtion/Ionisation – Time-Of-Flight (MALDI-TOF) (<10 mins vs 60+ mins (214, 215)), this allows for a rapid turnaround time. A Raman spectroscopy setup also requires substantially less room when compared to MALDI-TOF. Furthermore, given the use of only one wavelength of light such a device can be miniaturised into a highly portable

version (216). Raman is also highly sensitive and able to detect 1-4 x10^2 cells/mL within a sample (217).

Due to its inherent ability to be miniaturised as well as the ability to rapidly identify bacteria, it is therefore feasible to develop a high-throughput screening technology for TB and bovine TB using Raman spectroscopy. This is particularly useful when there are large cohorts of patients, or a large herd of cattle that need to be screened within a small amount of time.

A Raman spectrometer is composed of a laser that is used to illuminate the object of interest, this then reflects, absorbs and scatters the electromagnetic radiation. Following this, the scattered radiation is fed into a spectrometer and filtered to eliminate the wavelength of the laser (Rayleigh scattering) then fed into a computer *via* a charge-coupled device (CCD) and interpreted.

An InVia Raman spectrometer system by Renishaw (Wooton-under-edge, UK) equipped with a Leica confocal microscope DM2500 and 100x Leica objective lens with a 785nm diode laser was used to acquire the Raman spectra in **Chapters 4, 5 & 6.** Further specifics are expanded upon within their respective chapter.

2.5.3. WiRE:

Windows-based Raman Environment (WiRE) is the proprietary software package from Renishaw, which acts as the GUI for the Raman spectrometer. WiRE allows users to control the Raman instrument, as well as view samples and configure measurements. Additionally, WiRE also enables users to perform basic

processing and analysis of data, such as background fluorescence removal, noise removal, cosmic ray removal, *etc.*

WiRE 5.2 was used in the analysis of the Raman data generated in **Chapters 4, 5 & 6.**

2.6. Software for the data analyses:

Figures, where uncited, were created using biorender.com and Inkscape 1.2.1 . References were organised with EndNote 20.4.1. Microsoft Office Excel 365 and Plotly were used to for data plotting. The Python script used for the generation of average spectra, PCA, PCA quadrant analysis, SOM and SOMDI was collaboratively developed with Paulo De Gomes and is readily accessible *via* Vettigli *et al.* (218) . The Python script used for the generation PCA-LDA, confusion matrix, and spectral barcodes was collaboratively developed with Daniel P Weston and is readily accessible *via* Weston (219). The customised Python scripts for statistical analyses were adapted for use in this project and locally deployed *via* Jupyter Notebook 6.4.8.

2.7. Statistical methods used for the data analyses:

2.7.1. Principal Component Analysis (PCA):

Principal Component Analysis is a dimensionality reduction technique that transforms data into a new coordinate system, capturing its variance in a few principal components. It simplifies complex datasets, revealing patterns and reducing computational complexity.

When being applied to Raman spectral data, PCA can unravel underlying patterns within the intensity variations across the spectra. It identifies dominant spectral features and reduces data dimensionality, aiding in visualising sample similarities, detecting outliers, and highlighting key molecular contributors (220).

2.7.2. PCA Quadrant Analysis:

PCA Quadrant Analysis extends PCA by categorising data points into quadrants based on their scores along principal components. In this instance, two principal components were used. It enhances interpretation by highlighting relationships between variables, aiding in the identification of distinct clusters or trends.

Quadrant analysis enhances PCA's utility in Raman spectra by categorising spectra based on their principal component scores. This can unveil distinct chemical or structural groups within a dataset, providing a more intuitive understanding of how different components contribute to spectral variations (221).

In the work presented in this thesis, the PCA quadrant analysis has afforded us the ability to see which spectral components have the greatest amount of influence in the distribution of our PCA datapoints.

2.7.3. Self-Organised Mapping (SOM):

Self-Organised Mapping is an unsupervised machine learning technique that creates a low-dimensional representation of high-dimensional data. It organises similar input data into clusters on a map, preserving the topological relationships present in the original data.

SOM is valuable for organising and clustering Raman spectra based on similarities in their vibrational patterns. It helps visualise spectral relationships, identifying regions of similar composition. SOM is particularly useful when exploring complex molecular mixtures or subtle variations in chemical structures, such as those present in and on bacteria (222).

2.7.4. Self-Organised Mapping Discrimination Index (SOMDI):

SOMDI assesses the discriminatory power of a Self-Organised Map. It measures the separation between clusters on the SOM, providing insights into how well the map distinguishes different classes or groups within the data.

SOMDI assesses how well a SOM distinguishes between different groups or classes in Raman data. It quantifies the separation between clusters on the SOM, indicating the map's effectiveness in capturing relevant chemical differences among samples (223, 224).

2.7.5. Self-Optimising Kohohen Index Network (SKiNET):

SKiNET is an additional form of multivariate analysis that simultaneously provides dimensionality reduction, feature extractions and multi-class classification, by combining both SOM and SOMDI into a single tool (224).

2.7.6. Principal Component Analysis – Linear Discriminant Analysis (PCA-LDA):

PCA-LDA combines PCA and Linear Discriminant Analysis to enhance class separation in multivariate data. PCA identifies relevant features, and LDA

optimally projects these features for effective classification, improving the performance of pattern recognition algorithms.

PCA-LDA optimises the classification of Raman spectra. PCA identifies critical spectral features, and LDA builds a decision boundary to maximize class separation. This combined approach is powerful for discriminating between different sample types based on their Raman signatures (225).

2.7.7. Confusion Matrix:

A Confusion Matrix is a table that evaluates the performance of a classification algorithm. It compares predicted and actual class labels, displaying true positives, true negatives, false positives, and false negatives. It is a crucial tool for assessing model accuracy and identifying misclassifications.

In Raman spectroscopy, a confusion matrix assesses the accuracy of a classification model. It quantifies the success of identifying specific chemical or biological components in a sample based on their Raman spectra, providing insights into the model's strengths and areas for improvement (226).

2.7.8. Two-Way Analysis of Variance (ANOVA):

A two-way ANOVA, applied to Raman spectral data, investigates how two categorical factors influence the spectral intensities of bacterial species. The factors could represent different species (Factor 1) and an additional experimental condition (Factor 2), like varying treatments or time points. The analysis assesses whether variations in spectral intensities are due to each factor individually, as well as their potential interaction. Significant results for Factor 1 indicate overall

differences in Raman spectra between bacterial species, while Factor 2 reveals effects from the additional condition. An interaction effect suggests non-additive influences. In Raman data interpretation, this method elucidates how experimental variables jointly impact spectral characteristics (227).

Chapter 3:

3. *In silico* analysis of the Mycobacterial cell wall:

3.1. Introduction:

N.B: Due to the COVID-19 pandemic, this work was undertaken during the 6-month lab closure at the University of Birmingham.

Over the last years it has become more and more feasible to undertake *In silico* analysis using various types of computers (228, 229, 230, 231). Large universities such as the University of Birmingham often possess a supercomputer, which utilise large processors with numerous cores and memory systems of considerable size (232).

Additionally, due to the COVID-19 pandemic, there has been a large interest in moving away from lab work to *in silico* work, often due to the social distancing measures as well as lockdowns around the world (233). For instance, a straightforward albeit rudimentary search for "*in silico"* on ScienceDirect, one can see a steady upward trend across the years, with accelerated increases during the years 2020 and 2021.

As discussed, the ability to work remotely with *in silico* work *via* a university's supercomputer is one of the benefits which have come into play during the COVID-19 pandemic. However, what has not been discussed yet is the ability to analyse compounds which are not readily available, or compounds which would take a substantial amount of time to fully analyse.

Additionally, *in silico* work can enable scientists to work with more complex systems, such as protein interactions (234), drug interactions (235) and enzyme kinetics (236). Specifically, for this project, *in silico* work was undertaken to generate Raman spectra for compounds present within the cell wall of *M. tuberculosis.* The aim of this was to generate evidence which would support the analysis of the *Mycobacterium sp.* in the following chapter as well as enable the author of this thesis to pick out relevant Mycobacterium peaks in the *in vitro* research in the chapters following this one.

3.1.1. European Molecular Biology Laboratories:

As such, the structural data was initially acquired from the European Molecular Biology Laboratories (EMBL). EMBL is a world-renowned research institution focused on molecular biology, genetics, and structural biology. It was established in 1974 and is currently headquartered in Heidelberg, Germany. With collaboration between several sites in 6 European countries, with sites in Germany, France, Italy, Spain, and the United Kingdom (237).

EMBL's research focuses on understanding the fundamental principles of life, with the ultimate goal of improving human health and addressing societal challenges. Their research includes studying the structure and function of biomolecules, investigating gene expression and regulation, and understanding cellular processes and interactions.

One of the key areas of research at EMBL is the characterisation of compounds of interest. EMBL researchers utilise a range of techniques to characterise molecules

& compounds, including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry, and electron microscopy. The EMBL maintains a structural database of compounds, in the form of the Chemical Entities of Biological Interest database (ChEBL), and it is this database from which structural information was taken for this chapter. ChEBL utilises user-submitted data from various institutions, which it then verifies utilising cross-referencing from other researchers, as well as verification from researchers at the ChEBL (238).

3.1.2. Avogadro:

Avogadro is an open-source, cross-platform computational chemistry software program designed for molecular modelling and visualization. The program was first released in 2006 and has since become a widely used tool in the computational chemistry community (205).

Avogadro is used for a variety of applications in computational chemistry, including molecular modelling, quantum chemistry, molecular dynamics simulations and analysing *in silico* Raman data with input from ORCA. The program allows users to build, edit, and manipulate molecular structures, and to calculate properties such as energies, vibrational frequencies, and molecular orbitals, as well as the display of artificial Raman and Infrared (IR) data.

One of the key features of Avogadro is its flexibility and extensibility. The program can be easily customised and expanded with plugins, allowing users to add new features and capabilities to the software.

Avogadro has been used in a wide range of research studies and applications in computational chemistry. For example, generating illustrative outputs for the physical properties of carbon nanotubes (239), and the construction of & optimisation of the *in silico* structure of potential antimalarial compounds (240). For this specific project, as detailed above, it was used to optimise the geometry from the compounds of interest and generate suitable input data for ORCA.

3.1.3. BlueBEAR:

The BlueBEAR supercomputer is a high-performance computing facility at the University of Birmingham. The facility is designed to support a range of research activities across many scientific fields, including physics, chemistry, engineering, and life sciences.

BlueBEAR is built around a hybrid architecture, which combines traditional central processing units (CPUs) with graphics processing units (GPUs) and field-programmable gate arrays (FPGAs). This architecture allows researchers to tackle computationally challenging problems that require massive amounts of parallel processing power, such as simulations of complex systems, data analysis, and machine learning.

The supercomputer is also equipped with a high-speed network, which enables researchers to transfer large amounts of data quickly and efficiently between the computer nodes and storage systems. The facility provides researchers with a range of storage options, including high-performance parallel file systems, object storage, and archival storage, to support data management needs.

In terms of capabilities, BlueBEAR has been used in published research across many different fields. For instance, the simulation of pandemics (241), deployment in understanding metabolomics in HepaRG cells (242) and an investigation into the properties of molecular crystals (243).

BlueBEAR's hybrid architecture, high-speed network, and advanced storage options provide researchers with the computational resources needed to tackle complex problems and generate new insights into a wide range of scientific phenomena.

3.1.4. ORCA:

ORCA (short for "Open-Source Computational Chemistry Software") is a powerful computational chemistry software package that allows researchers to model and simulate a wide range of chemical systems using quantum mechanics. Developed by a team of researchers at the Max Planck Institute for Chemical Energy Conversion, ORCA has become a popular tool for studying chemical reactions and properties, with a growing user base around the world (207).

ORCA employs a range of advanced algorithms and techniques to perform calculations on molecules and materials, including density functional theory (DFT), wavefunction-based methods, and semi-empirical methods. The software package can be readily modified, allowing researchers to tailor their calculations to their specific research questions and computational resources.

In recent years, ORCA has been used in a variety of scientific applications, including drug discovery (244), materials science (245, 246), and energy research.

Specifically, a recent study used ORCA to investigate the catalytic mechanism of an enzyme involved in the production of biofuels (247), as well as another study which sought to understand critical reactions used in the production of ammonia (248).

Overall, ORCA is a powerful tool in generating *in silico* data and it is through the combined use of BlueBEAR, Avogadro & ORCA that this chapter becomes feasible.

3.2. Methods:

3.2.1. European Molecular Biology Laboratories:

EMBL was utilised *via* ChEBL by searching for the term "*Mycobacterium tuberculosis*" with the option for three-star verification being taken. 157 entries were present. Of these, 21 compounds (See **Table 4** opposite) with a direct source from *M. tuberculosis* were identified, and their associated MolFile were downloaded for further use with Avogadro.

	Compound name:	SMILES:
#1	(2R)-2-[(1R)-1-Hydroxy-12-2-[14-(2-Icosylcyclopropyl)-Tetradecyl]-Cyclopropyl-Dodecyl]-Hexacosanoate	**C(CCCCCC1C(C1)CCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCCC)CCCCCCCCC2C(C2)CCCCCCCCCCCCCCCCCCCC**
#2	(2R)-2-[(1R)-1-hydroxy-12-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyldodecyl]hexacosanoic acid	**C(CCCCCC1C(C1)CCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCCCC)CCCCCCCCC2C(C2)CCCCCCCCCCCCCCCCCCCC**
#3	(2R)-2-[(1R)-1-hydroxy-14-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyltetradecyl]hexacosanoic acid	**C(CCCCCC1C(C1)CCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCCCC)CCCCCCCCC2C(C2)CCCCCCCCCCCCCCCCCCCC**
#4	(2R)-2-[(1R)-1-hydroxy-16-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylhexadecyl]hexacosanoate	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCC**
#5	(2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCCC**
#6	(2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoic acid	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCCC**
#7	(2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCCC**
#8	(2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoic acid	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O) =O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCCC**

#9	(2*R*)-2-[(1*R*)-1-hydroxy-18-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyloctadecyl]tetracosanoic acid	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCC**
#10	(2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCCC**
#11	(2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCC**
#12	(2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCC**
#13	(2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoate	**C1(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCC)C2)C(C1)CCCCCCCCCCCCCCCCC**
#14	(2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoic acid	**C1(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C(O)=O)CCCCCCCCCCCCCCCCCCCCCC)C2)C(C1)CCCCCCCCCCCCCCCCC**
#15	(2R)-2-[(1R)-1-hydroxy-22-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyldocosyl]hexacosanoate	**C(CCCCCCCCC1C(CCCCCCCCCCC2C(CCCCCCCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCCC)C2)C1)CCCCCCCCC**
#16	(2R)-2-[(1R)-14-2-[(15Z)-hexatriacont-15-en-1-yl]cyclopropyl-1-hydroxytetradecyl]hexacosanoate	**C(CCCCC1C(C1)CCCCCCCCCCCC[C@@H](O)[C@H](C([O-])=O)CCCCCCCCCCCCCCCCCCCCCC)CCCCCCCC/C=C\CCCCCCCCCCCCCCCCCC**
#17	α-mycolic acid	**CCCCCCCCCCCCCCCCCCCCCCCC(C(O)CCCCCCCCCCCCCCC1CC1CCCCCCCCCCCCCC1CC1CCCCCCCCCCCCCCCCC)C(O)=O**

#18	α,α'-trehalose 6-mycolate	**CCCCCCCCCCCCCCC(C(O)CCCCCCC\C=C\CCCCCC)C(=O)OC[C@H]1O[C@H](O[C@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)[C@H](O)[C@@H](O)[C@@H]1O**
#19	β-D-Araf-(1→2)-α-D-Araf-(1→3)-α-D-Araf-(1→5)-α-D-Araf-OMe	**CO[C@H]1O[C@H](CO[C@H]2O[C@H](CO)[C@@H](O[C@H]3O[C@H](CO)[C@@H](O)[C@@H]3O[C@@H]3O[C@H](CO)[C@@H](O)[C@@H]3O)[C@@H]2O)[C@@H](O)[C@@H]1O**
#20	Keto Mycolic Acid	**CCCCCCCCCCCCCCCCCCCCCCCC(C(O)CCCCCCCCCCCCCCCCCCC1CC1CCCCCCCCCCCCCCCC(=O)C(C)CCCCCCCCCCCCCCCCCC)C(O)=O**
#21	Methoxy Mycolic Acid	**CCCCCCCCCCCCCCCCCCCCCCCC(C(O)CCCCCCCCCCCCCCCCCCC1CC1CCCCCCCCCCCCCCCC C(OC)C(C)CCCCCCCCCCCCCCCCCC)C(O)=O**

Table 4 - Compound table listing the compounds identified *via* ChEBI alongside SMILES format (249)

3.2.2. Avogadro:

MolFiles were visualised with Avogadro. Geometry was then optimised and an ORCA input was generated. This input sought to undertake split valence polarisation (SVP), and BP86 as a Density functional Theory (DFT) functional, due to its greater agreement with the geometry of compounds over other methods such as B3LYP. These were incorporated into an XYZ file. Tight Self-consistent field (SCF) was maintained to better calculate the electron density of the orbitals. Triple Zeta Variance Potential (TZVP) was also used to better calculate the vibrational density. Perdew-Burke-Ernzerhof exchange (PBE) was also conducted to further understand agreement of atomic energies.

These files were then saved as .inp files and then submitted to BlueBEAR for analysis with ORCA. An example of the .inp files can be found in **A.6. Example of BlueBEAR input file**:.

Avogadro was used once more once ORCA had generated output files to visualise the Raman data. The visualised data was then exported as a .txt document to Notepad++ for processing, and then visualised on WiRE before being submitted to Plotly to generate plots.

3.2.3. BlueBEAR/ANSMA group computer:

The above files were instructed to be run in parallel using 16 cores from BlueBEAR, with a run time of one week. Output files were saved in the same partition relevant to the research group of the author and this could then be exported to a local computer.

BlueBear's IceLake processor was used. An example of the .bash script can be found in **A.7. Example of .BASH script submitted to BlueBEAR:.**

3.2.4. ORCA:

For the generation of *in silico* data, ORCA 5.0.3 was used throughout.

3.3. Results:

3.3.1. (2R)-2-[(1R)-1-Hydroxy-12-2-[14-(2-Icosylcyclopropyl)-Tetradecyl]-Cyclopropyl-Dodecyl]-Hexacosanoate:

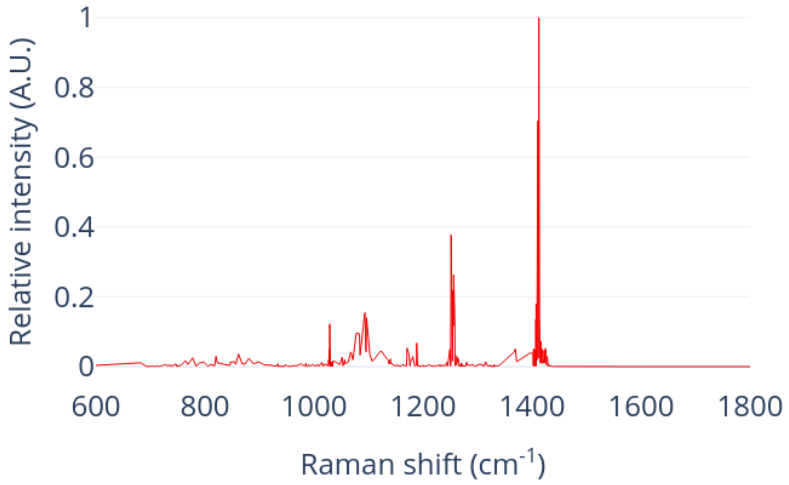


Figure 20 - *In silico* Raman spectra of (2R)-2-[(1R)-1-Hydroxy-12-2-[14-(2-Icosylcyclopropyl)-Tetradecyl]-Cyclopropyl-Dodecyl]-Hexacosanoate

3.3.1. Presents with notable peaks at 1029, 1078, 1093, 196, 1123, 1170, 1251 – 1256 and 1252cm^{-1}. Most of these peaks likely arise due to differing formation of carbon chains within the compound, however, *Talari et al.* (250) identifies the peaks at 1123 as glycogen and 1251-1256 as proteins, groups which simply do not exist on

this compound. The peaks at *ca.* 1400cm^{-1} arise due to the carboxyl group present on the molecule (250).

3.3.2. (2R)-2-[(1R)-1-hydroxy-12-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyldodecyl]hexacosanoic acid:

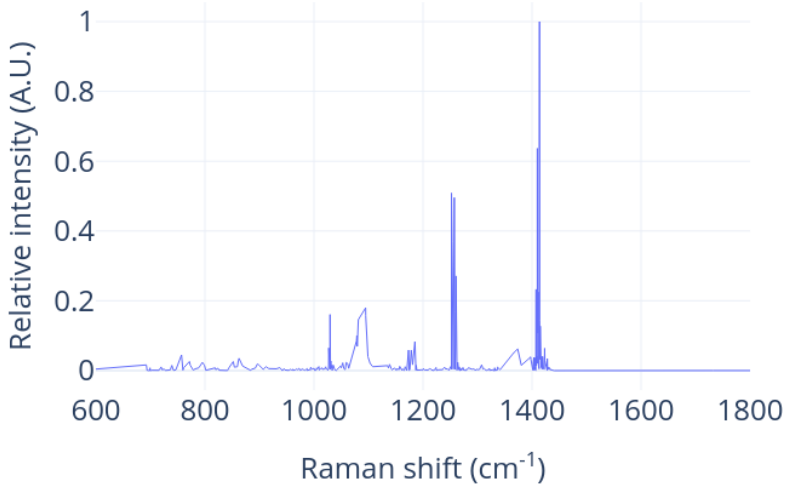


Figure 21 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-12-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyldodecyl]hexacosanoic acid

The Raman spectra of 3.3.2. presents in a similar fashion to 3.3.1, albeit the peak at 1094 is more smoothed out. This is likely due to the presence of the extra H^+ present within the COO^- functional group (250).

3.3.3. (2R)-2-[(1R)-1-hydroxy-14-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyltetradecyl]hexacosanoic acid:

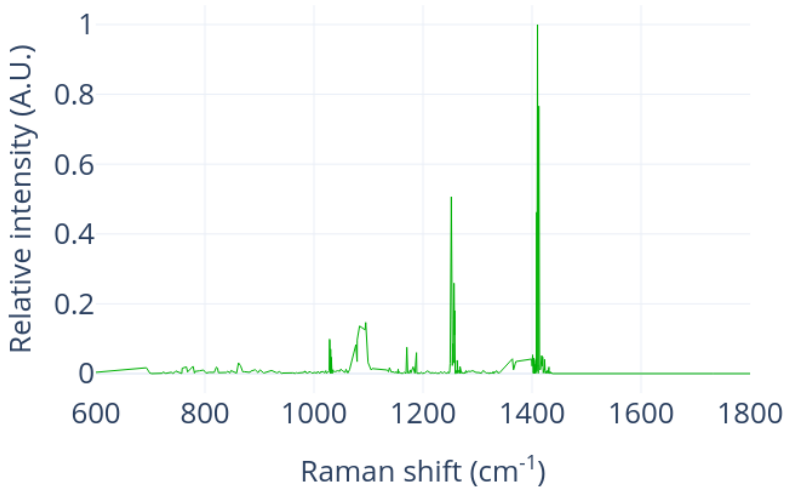


Figure 22 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-14-2-[14-(2-icosylcyclopropyl)tetradecyl]cyclopropyltetradecyl]hexacosanoic acid

The Raman spectra of 3.3.3. presents in a similar fashion to 3.3.1 & 3.3.2, albeit the peaks at 1170 & 1188 contrast slightly from their intensity. This could be as a result of the extra C in an intermediate carbon chain in the molecule (250).

3.3.4. (2R)-2-[(1R)-1-hydroxy-16-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylhexadecyl]hexacosanoate:

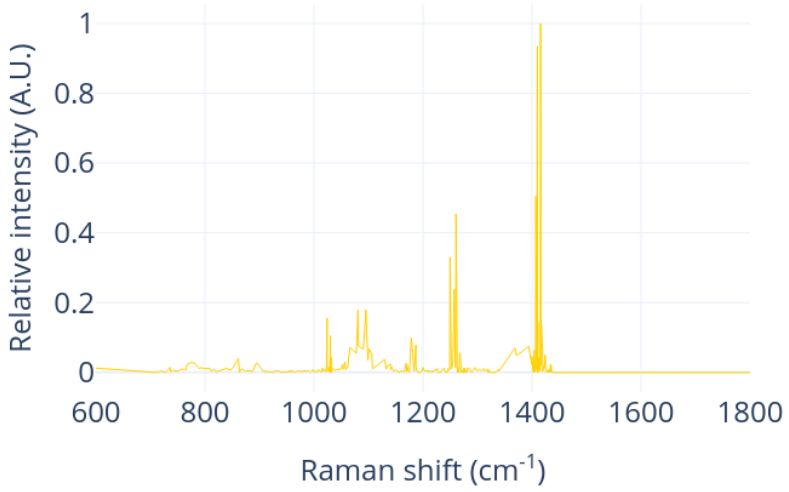


Figure 23 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-16-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylhexadecyl]hexacosanoate

The Raman spectra of 3.3.4. presents in a similar fashion to its predecessors, with notable peaks at 1024, 1029, 1080, 1095, 1178, 1260, 1369, 1393, and 1409cm^{-1}. The peaks at 1029, 1095, 1260 and 1393cm^{-1} arise due to differing formation of carbon chains and methyl groups, as well as the theoretical rocking of the chains. The peak at 1369cm^{-1} arises due to the compound being a fatty acid. The peak at 1409cm^{-1} arises due to the COO^- group present on the chain. However, the remaining peaks cannot be assigned based on the peak assignment data provided in Talari *et al* (250).

3.3.5. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate:

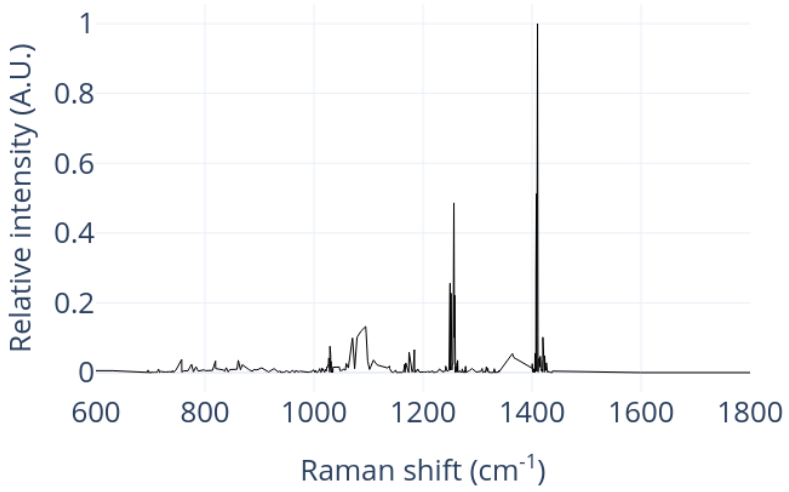


Figure 24 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate

The Raman spectra for 3.3.5. presents with notable peaks at 1029, 1070, 1086, 1095, & 1409 cm^{-1}. Which can be assigned to various C-C bonds as well as the presence of the lipid properties of the compound (250). However, whilst there are also notable peaks at 1174, 1184, 1249, 1256 & 1364 cm^{-1}, these peaks are often assigned to proteins, nucleotides or no functional group, which cannot be explained given that the compound in 3.3.5 possesses none of these properties.

3.3.6. (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoic acid:

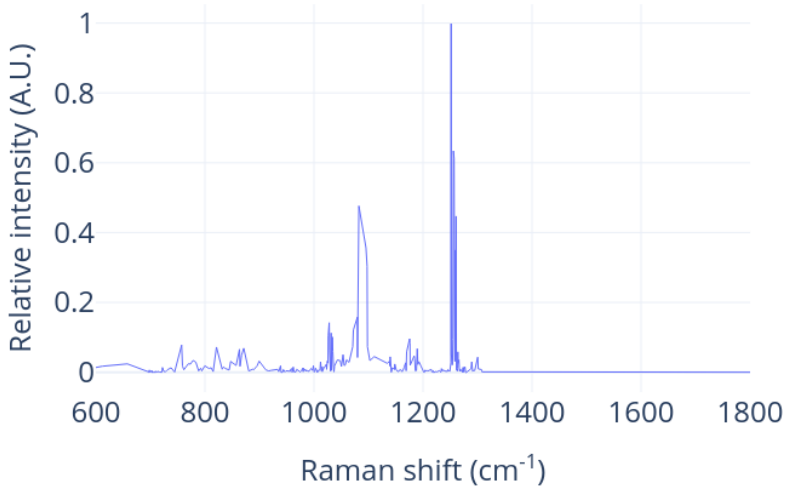


Figure 25 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoic acid

3.3.6. presents with notable peaks across the spectrum. Of note are the peaks at 870, 1030, 1031, 1251 and 1255cm^{-1}, which can be assigned to C-H related chains and the fact that the compound is a fatty acid (250). However, the peaks at 757, 821, 1080, & 1082cm^{-1} cannot be assigned. Additionally, the peak at 1257cm^{-1} is often associated with amides, functional groups which are not present on this compound.

3.3.7. (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoate:

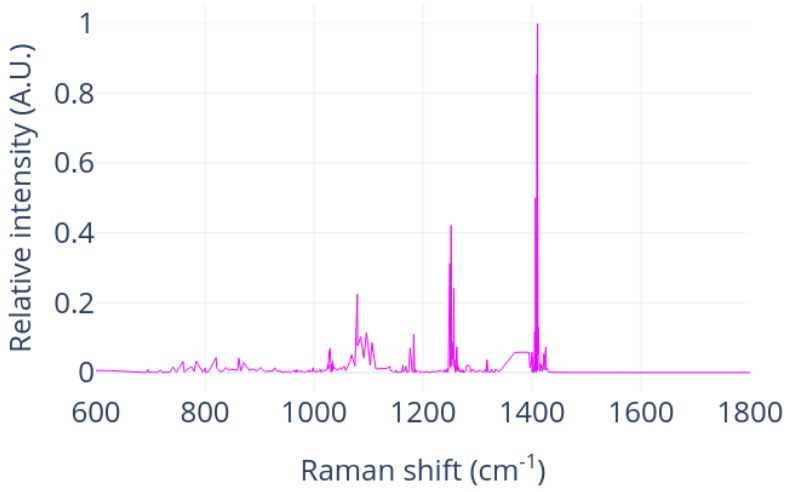


Figure 26 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-[2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyl]octadecyl]hexacosanoate

The peaks present for 3.3.7. represent a variety of different functional groups. The key peaks which can be readily assigned are 1029cm^{-1}, which can be assigned to O-H stretching. Meanwhile, the peaks at 1369, 1253 & 1256cm^{-1} are indicative of the presence of lipids, which is an accurate reflection of the compound. The peaks at 1406 & 1409cm^{-1} indicate the presence of a COO^- group (250). However, peaks at 1176 & 1183 cm^{-1} indicate the presence of amine groups which is impossible given the nature of this compound, as is the peak at 1096cm^{-1} which indicates the presence of a phosphate.

3.3.8. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoic acid:

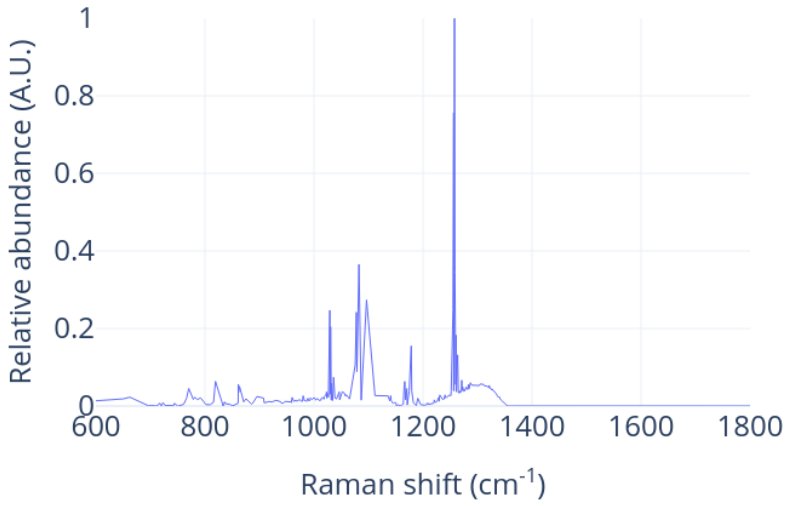


Figure 27 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoic acid

3.3.8. presents with a particularly intense peak between 1257 and 1264 cm^{-1} which are indicative of the presence of a lipid. The presence of a lipid is further confirmed by the presence of the peak at 1032 cm^{-1} . There is also a peak at 1029 cm^{-1} which indicates the presence of a COO^- group (250). However, there are peaks at 770, 820, 860, and 1178 cm^{-1} which suggest the presence of proteins and oligonucleotides, once more.

3.3.9. (2R)- 2- [(1R)- 1- hydroxy- 18- 2- [10- (2- octadecylcyclopropyl)decyl]cyclopropyloctadecyl]tetracosanoic acid:

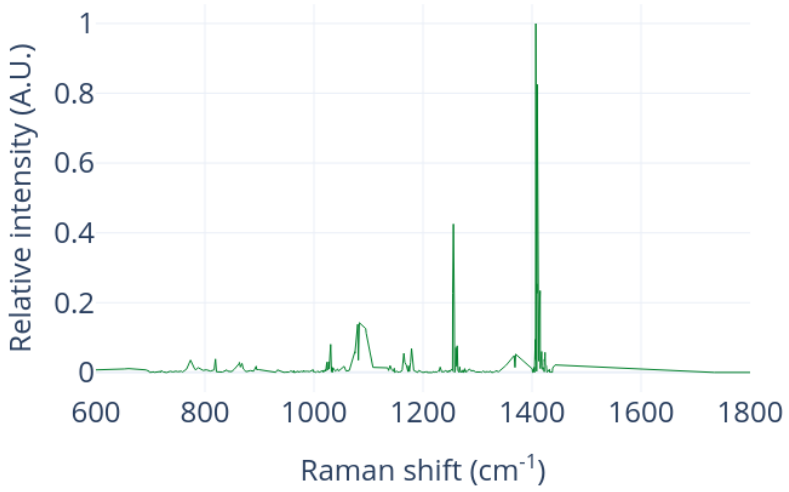


Figure 28 - *In silico* Raman spectra of (2R)- 2- [(1R)- 1- hydroxy- 18- 2- [10- (2- octadecylcyclopropyl)decyl]cyclopropyloctadecyl]tetracosanoic acid

3.3.9. possesses particularly prominent features at 1029, 1030, 1253, 1371, 1258 and 1409 cm^{-1} which are indicative of the presence of O-CH_3 stretching, lipids, CH_2 and COO-, respectively (250). Meanwhile, the spectra at 1080 cm^{-1} is indicative of phospholipids, of which there are none on this compound.

3.3.10. (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate:

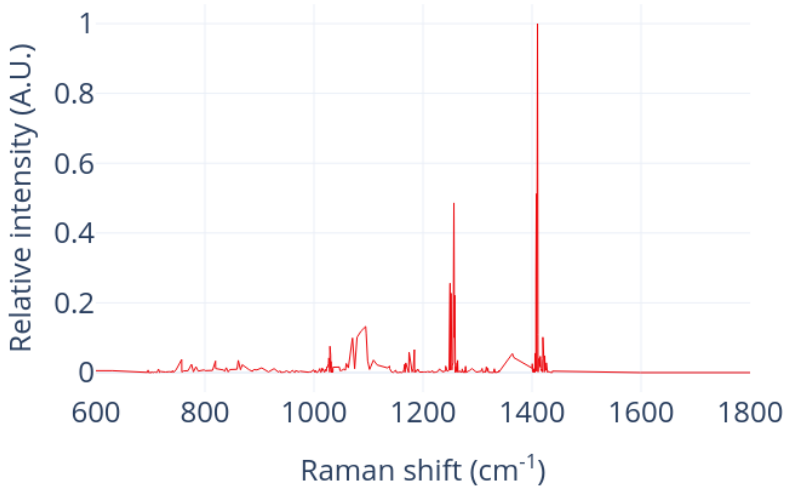


Figure 29 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-18-2-[10-(2-icosylcyclopropyl)decyl]cyclopropyloctadecyl]hexacosanoate

3.3.10. encompasses features at 1070, 1095, and 1068 cm^{-1} indicative of the presence of lipids. Meanwhile, the peak at 1029 cm^{-1} represents O-CH_3 stretching. The peaks at 1174 cm^{-1} is indicative of a C-H bend, whilst 1258, and 1410 cm^{-1} are indicative of the bending or stretching of $CH_{2\ (250)}$.

3.3.11. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid:

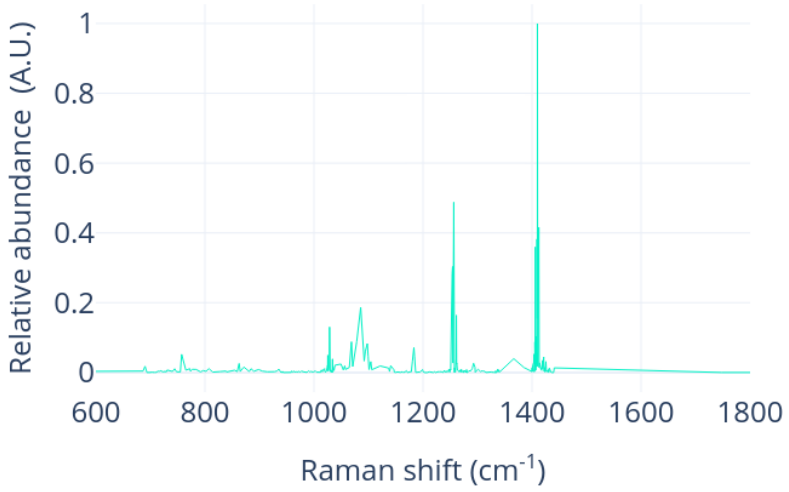


Figure 30 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-nonadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid

3.3.11. possesses a particularly prominent peak at 1400 cm^{-1} which is indicative of CH bending. Following this there is also a peak at 1367 cm^{-1} indicating the presence of CH_3 in a lipid. Meanwhile, peaks at 1260 and 1255 cm^{-1} confer the presence of lipids once more. A peak at 1029 cm^{-1} suggests the presence of O-CH_3 stretching and the presence of the peaks at 1290 cm^{-1} indicates the presence of CH bending. The peak at 1086 cm^{-1} indicates the presence of a $_{\nu}$C-C bond (250). Lastly, peaks at 1068 & 1098 cm^{-1} have no assignment and the peak at 1183 cm^{-1} indicates the presence of oligonucleotides though there is no nitrogen within this compound.

3.3.12. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid:

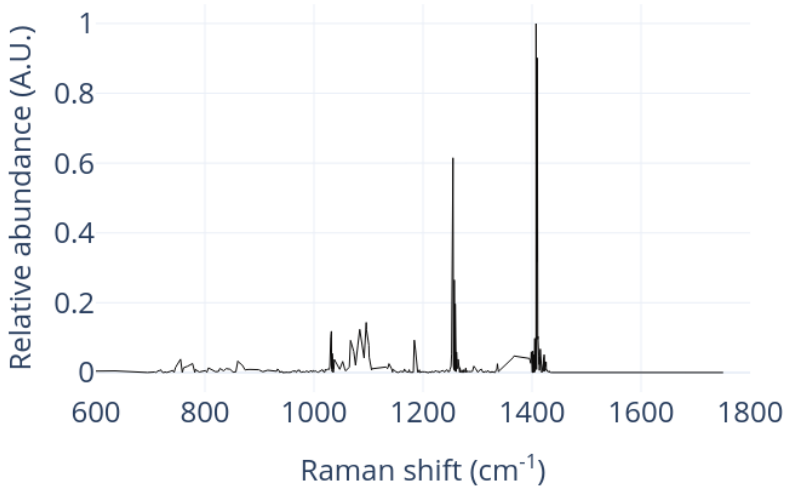


Figure 31 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]hexacosanoic acid

3.3.12. possesses prominent peaks at 1031 and a series of clustered peaks at approximately 1070 cm^{-1} which are indicative the presence of C-H bending and triglycerides. The peak at 1184 cm^{-1} indicates the presence of oligonucleotides, though this is impossible given the nature of this compound. The peak centred around 1254 cm^{-1} is indicative of the presence of lipids. Meanwhile, the peak at 1366 cm^{-1} is associated with phospholipids, though this, once again, is impossible given the nature of the compound. Lastly, the peaks centred around 1407 cm^{-1} is likely due to the presence of COO- (250).

3.3.13. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoate:

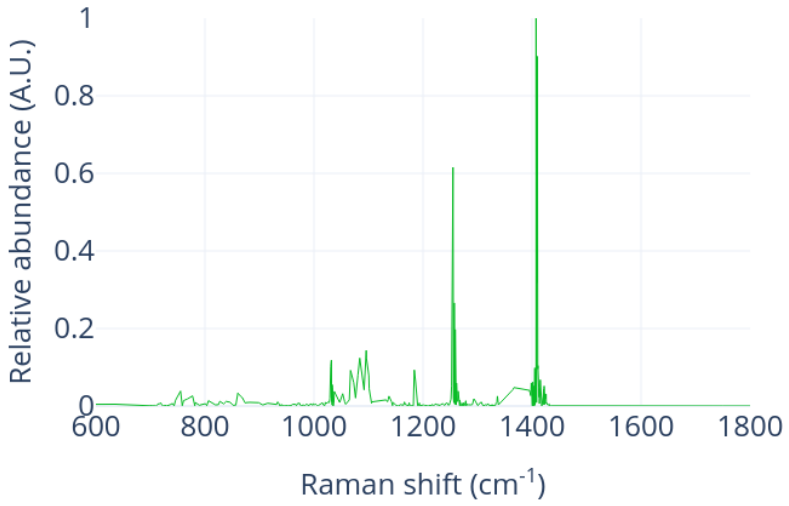


Figure 32 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoate

3.3.13. possesses a prominent peak at 1031 cm^{-1} which is indicative the presence of C-H bending. The peaks centred around 1370 may be due to triglycerides, once more. The peak at 1083 cm^{-1} is indicated to be present due to C-N stretching from, either, proteins or, potentially lipids, however, there is no nitrogen present within this compound. Peaks at 1095 and 1254 cm^{-1} arise due to the compound being lipid in nature. Whilst the peak at 1184 cm^{-1} is, once again, indicative of the presence of oligonucleotides, something which is impossible. Meanwhile, the peak at 1366 cm^{-1} is associated with phospholipids. Lastly, the peaks centred around 1407 cm^{-1} is likely due to the presence of COO- (250).

3.3.14. (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoic acid:

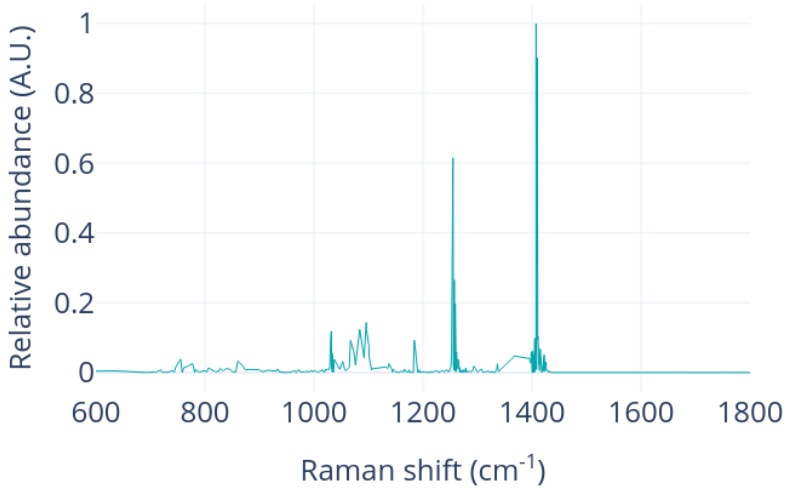


Figure 33 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-20-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropylicosyl]tetracosanoic acid

3.3.14., akin to 3.3.14, possesses a prominent peak at 1031 cm^{-1} which is indicative the presence of C-H bending. The peaks centred around 1370 may be due to triglycerides, once more. The peak at 1083 cm^{-1} is indicated to be present due to C-N stretching from, either, proteins or, potentially lipids, however, there is no nitrogen present within this compound. Peaks at 1095 and 1254 cm^{-1} arise due to the compound being lipid in nature. Whilst the peak at 1184 cm^{-1} is, once again, indicative of the presence of oligonucleotides, something which is impossible. Meanwhile, the peak at 1366 cm^{-1} is associated with phospholipids. Lastly, the peaks centred around 1407 cm^{-1} is likely due to the presence of COO- (250).

3.3.15. (2R)-2-[(1R)-1-hydroxy-22-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyldocosyl]hexacosanoate:

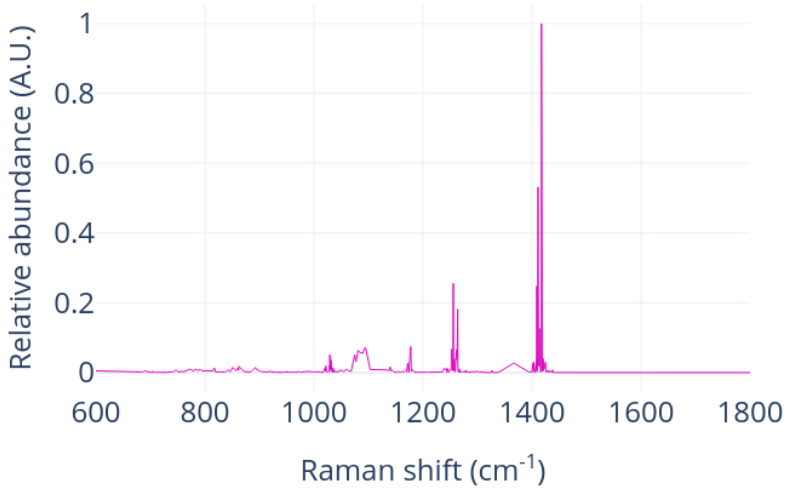


Figure 34 - *In silico* Raman spectra of (2R)-2-[(1R)-1-hydroxy-22-2-[10-(2-octadecylcyclopropyl)decyl]cyclopropyldocosyl]hexacosanoate

3.3.15. has noable peaks at 1029 cm^{-1} which likely arises as a result of O-CH_3 stretching. There is also a notable peak at 1093 cm^{-1} which theoretically arises due to the presence of C-N on a protein, though there are no proteins here. The peak at 1095 & 1255 cm^{-1} arises due to the presence of lipids. There is also a peak at 1177 cm^{-1} which X states is likely due to oligonucleotides. Peaks present at 1365 and 1411 cm^{-1} are of unknown origin but could be due to broadening around CH_3 and COO- respectively. The final peak at 1417 cm^{-1} is likely due to C=C stretching (250).

3.3.16. (2R)-2-[(1R)-14-2-[(15Z)-hexatriacont-15-en-1-yl]cyclopropyl-1-hydroxytetradecyl]hexacosanoate:

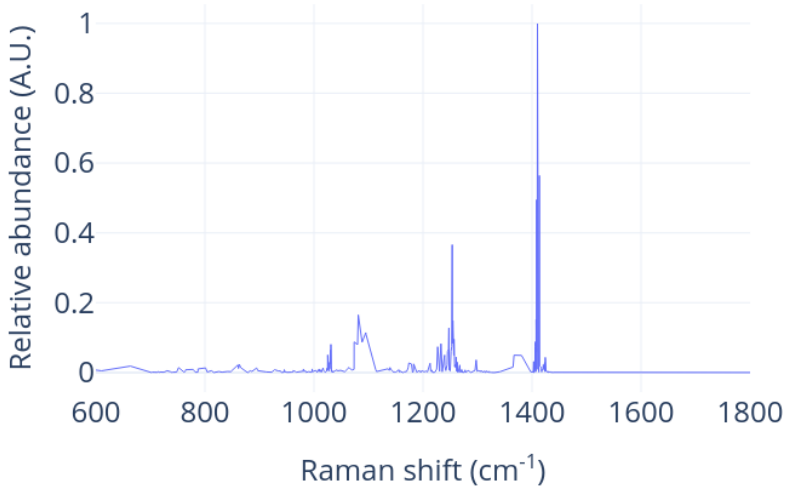


Figure 35 - *In silico* Raman spectra of (2R)-2-[(1R)-14-2-[(15Z)-hexatriacont-15-en-1-yl]cyclopropyl-1-hydroxytetradecyl]hexacosanoate

Peaks for 3.3.16. arise at 1025 cm^{-1} which is indicative of the presence of carbohydrates, though this compound is not a carbohydrate. This is also true for the peak at 1073 cm^{-1} . The peak at 1031 cm^{-1} is as a result of C-H bending, in plane. The presence of the peaks at 1076 & 1095 cm^{-1} arise due to the lipid nature of this compound. The peak at 1081 cm^{-1} is likely as a result of C-C in a lipid. The broad clustering of peaks at 1125 – 1255 cm^{-1} is centred around 1255 cm^{-1}, indicating the presence of lipids. The broad peak around 1380 is likely due to CH_3. The peak at 1407 cm^{-1} is likely due to the presence of COO- (250).

3.3.17. α-mycolic acid:

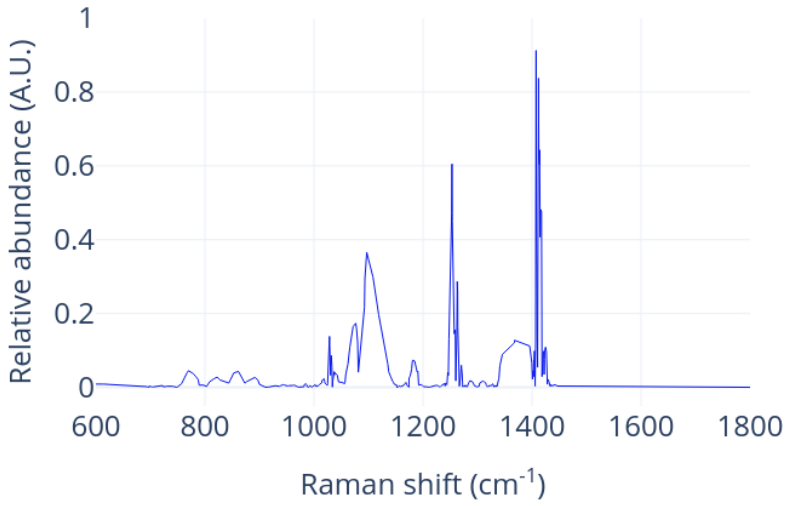


Figure 36 - *In silico* Raman spectra of α-mycolic acid

The peak at 1029 cm^{-1} is likely due to the stretching of O-CH_3 of the methoxy group. Peaks at 1076 and 1095 cm^{-1} are likely due to lipids. The peak at 1184 cm^{-1} is proposed to be due to oligonucleotides, but this is not possible due to the nature of the compound. Peaks at 1253 & 1368 cm^{-1} are associated with lipids. And the last peak at 1407 cm^{-1} is likely due to COO- (250).

3.3.18. α,α'-trehalose 6-mycolate:

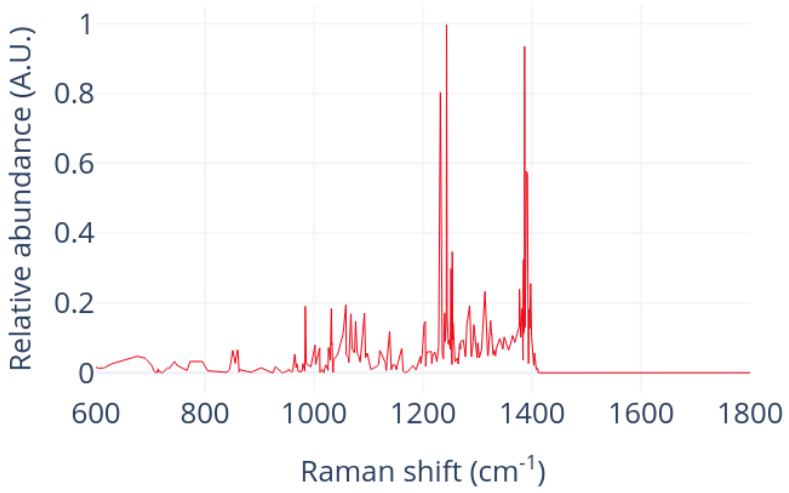


Figure 37 - *In silico* Raman spectra of α,α'-trehalose 6-mycolate

The small cluster of peaks at approximately 984 cm^{-1} has no assignment. Peaks around 1031 cm^{-1} are indicative of C-H bending. Peaks at 1057 and 1313 cm^{-1} indicate the presence of lipids and a CH_3CH_2 twisting mode which is present in lipids. The peaks at 1204, 1231, and 1243 is indicative of the presence of amides, though these functional groups are not present. The peak at 1253 cm^{-1} is indicative of a C-O aromatic stretch and the peak at 1386 cm^{-1} indicates the presence of a CH_3 band (250).

3.3.19. β-D-Araf-(1→2)-α-D-Araf-(1→3)-α-D-Araf-(1→5)-α-D-Araf-OMe:

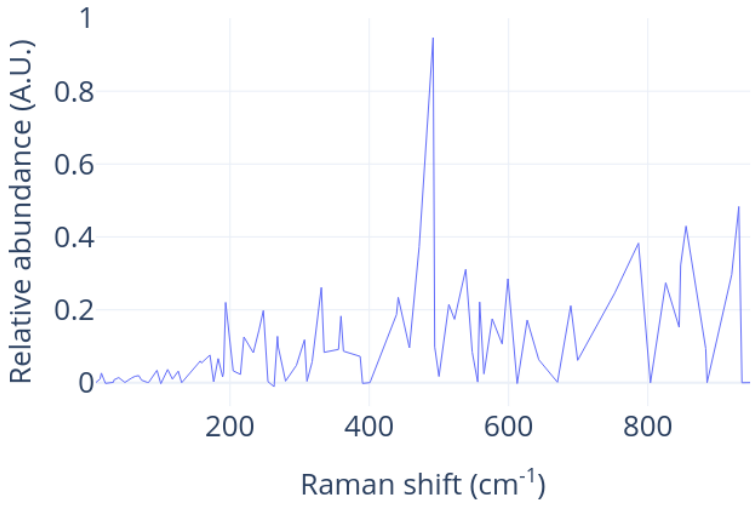


Figure 38 - *In silico* Raman spectra of β-D-Araf-(1→2)-α-D-Araf-(1→3)-α-D-Araf-(1→5)-α-D-Araf-OMe

The most prominent peak within this Raman spectra is the one around 477 cm^{-1} which is indicative of the presence of a multi-chained polysaccharide (250).

3.3.20. Keto Mycolic Acid:

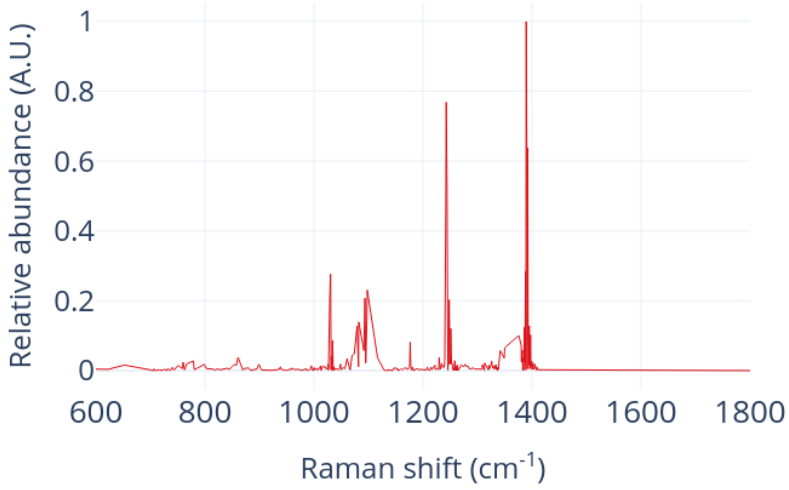


Figure 39 - *In silico* Raman spectra of Keto Mycolic Acid

3.3.20. features prominent beginning at 1030 cm^{-1} which is indicative of C-H bending, this is also true for the peak at 1176 cm^{-1}. Adding to this, there are also prominent peaks at 1082 cm^{-1}, indicating the presence of carbohydrates. The prominent peak at 1242 cm^{-1} indicates the presence of amides, though this is not possible given that this compound is a fatty acid. Peaks present at 1389 & 1097 cm^{-1} remain unassigned (250).

3.3.21. Methoxy Mycolic Acid:

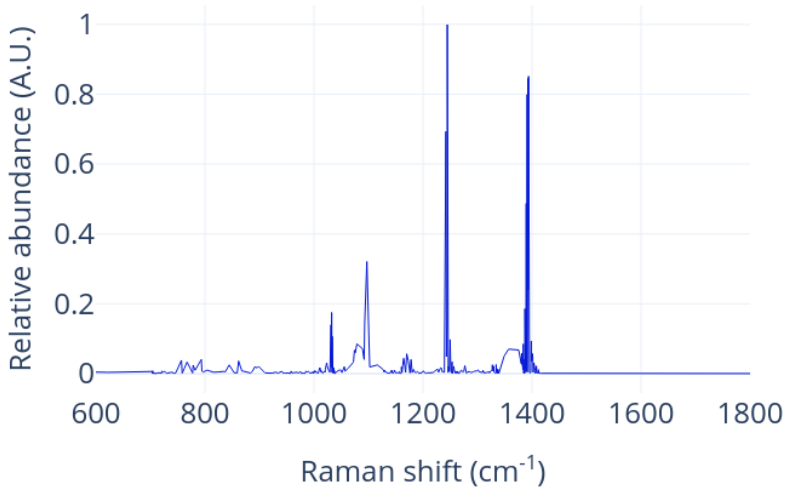


Figure 40 - *In silico* Raman spectra of Methoxy Mycolic Acid

Peaks for methoxy mycolic acid are particularly prominent at 1032 cm^{-1} indicating the presence of carbohydrates, which is not possible, due to the lower amount of oxygen present on this compound. The peak at 1078 cm^{-1} arises due to a C-C or C-H bond in formation, as part of a lipid. The peak 1097 cm^{-1} and 1244 cm^{-1} cannot be assigned. The peak at 1242 cm^{-1} has been theorised to be due to the presence of an

amide but, once again, this is impossible. The peaks at 1391 and 1392 cm^{-1} are due to C-H rocking and stretching, respectively (250).

To summarise, the following peaks were detected and are assigned in **Table 5** below:

Raman Peak (cm^{-1})	Functional Group / Band Assignment	Prominent in compound #
477	Polysaccharides	19
757	Unassigned	6
770	Unassigned	8
820	C-C Stretching	8
821	Unassigned	6
860	Phosphate group/Phophatidic acid	8
870	Lipid	6
984	Unassigned	18
1024	Unassigned	4
1025	Carbohydrates	16
1029	O-CH_3 stretching of methoxy groups	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 15, 17
1030	C-H bending	6, 9, 12, 20
1031	δC-H, phenylalanine C-H in-plane bending mode of phenylalanine	13, 14, 17
1032	C-C skeletal stretching	8, 21
1057	Lipids	18
1068	Unassigned	10, 11
1070	Triglycerides	5, 10, 12
1073	Triglycerides	16
1076	Triglycerides	16, 17
1078	ν(C-C) or ν(C-O) stretch (lipid)	1, 2, 3, 21
1080	Amide II	4, 6, 9
1081	$\nu_1CO_3^{2-}$,$\nu_3PO_4^{3-}$,ν(CC)skeletal of acyl backbone in lipid (*gauche* conformation)	16
1082	Carbohydrates residues of collagen Carbohydrates peak for solutions Nucleic acids	6, 20
1083	C-N Stretching mode of proteins (and lipid mode of lesser degree)	14
1086	ν(C-C) *gauche*	5, 11

1093	Symmetric PO_2- stretching vibration of the DNA backbone; phosphate backbone vibration as a marker mode for the DNA concentration; C-N of proteins	1, 15
1094	DNA C-N Stretching	2
1095	Lipid	4, 5, 10, 13, 14, 15, 16, 17
1096	Phosphodioxy (PO_2^-) groups	1, 2, 3, 7
1097	Unassigned	20, 21
1098	Unassigned	11
1123	C-C Stretching mod of lipids and protein	1, 2, 3
1125	Unassigned	16
1170	C-H in-plane bending mode of tyrosine	1, 2, 3
1174	Tyrosine, phenylalanine, C-H bend (protein)	5, 16
1176	C-H bending tyrosine (proteins)	7, 20
1177	Cytosine, guanine	15
1178	Unassigned	4, 8
1183	Unassigned	7, 11
1184	Unassigned	5, 13, 14, 16, 17
1188	Unassigned	3
1204	Amide III and CH_2 wagging vibrations from glycine backbone and proline side chains Collagen Tyrosine, phenylalanine (IgG)	18
1231	Unassigned	18
1242	Amide III	20, 21
1243	Amide III C-O_4 aromatic stretch	18
1244	Unassigned	21
1249	Unassigned	5
1251	Unassigned	1, 2, 3, 6, 16
1252	C-O_4 aromatic stretch	1, 2, 3
1253	Unassigned	7, 9, 17, 18
1254	Formalin contamination on fixed tissues C-N in plane stretching	12, 13, 14, 16
1255	Lipids	6, 11, 15
1256	Unassigned	1, 2, 3, 5, 7
1257	Adenine, thymine, Amide III	6, 8

1258	Adenine, cytosine, amide III	9
1260	Amide III, proteins, CH_2 in-plane deformation (Lipids)	4, 11
1264	Unassigned	8
1290	Cytosine	11
1313	CH_3CH_2 twisting mode of collagen/lipid	18
1364	Unassigned	5
1365	Tryptophan	15
1366	Unassigned	12, 13, 14
1367	ν_S (CH_3) (Phospholipids)	11
1368	Unassigned	17
1369	Guanine, TRP (protein), porphyrins, lipids	4, 7, 14
1370	The most pronounced saccharide band	13, 14
1371	CH_3 Stretching	9, 14
1380	Unassigned	16
1386	CH_3 band	18
1389	Unassigned	20
1391	CH rocking	21
1392	C-N stretching, quinoid ring-benzoid ring-quinoid ring	21
1393	CH rocking	4
1400	NH in-plan deformation (to 1430) ν(C=O)O^-(amino acids, aspartic and glutamic acid)	11
1406	ν_S COO^-	7
1407	Unassigned	12, 13, 14, 16, 17
1409	ν_S COO^-	1, 2, 3, 4, 5, 7, 9
1411	Unassigned	15
1417	C=C stretching in quinoid ring	15

Table 5 - Table displaying the spectral assignment for the most prominent peaks, and the compound they are most prominent in according to Talari *et al.* (250). Spectra associated with lipids are highlighted in orange.

3.4. Discussion:

The compounds analysed *in silico* within this chapter are primarily derivatives of meromycolate chains, because of this, the compounds listed present with potential Raman spectra which are similar in nature to each other.

Of note are the peaks which regularly appear at *ca.* 1400 cm^{-1} & 1255 cm^{-1}, which are closely associated with the presence of COO^- and lipids, respectively. We also see a great number of peaks at *ca.* 1070 cm^{-1}, indicating lipids. This makes sense given that, once again, these are derivatives of meromycolates and fatty acids. One can therefore assume that these are omnipresent within the cell wall of Mycobacteria in general, and the literature would support this assertion. Specifically, the work of Perumal *et al. (251)* & Rivera-Betancourt *et al.* (252) demonstrates the presence of some of these peaks within their Surface Enhanced Raman Spectroscopy (SERS) analysis of *M. tuberculosis* and other *Myocbacterium* species. SERS is a technique that is Raman-based but seeks to utilise a plasmonic interface in a bid to increase the scattering of the electrons when an object is hit by a specified wavelength of light (253). In doing so, the resolution of the Raman spectra is greatly amplified.

The use of SERS has been widely documented by researchers for the detection of pathogenic bacteria (254). However, the number of SERS surfaces is limited and a notable proportion of the current literature have focused on the use of silver (173) and gold (255), both materials of which are highly bactericidal as Ansari *et al. (256)* and Cui *et al.* (257) demonstrate. The use of SERS could be used on the compounds analysed within this chapter as well as those in Chapter 5. However, its use may become limited in circumstances reflective of those in Chapters 6 & 7 due to these compounds damaging the cell wall of bacteria, as Ansari *et al.* illustrates in **Figure 41** opposite.

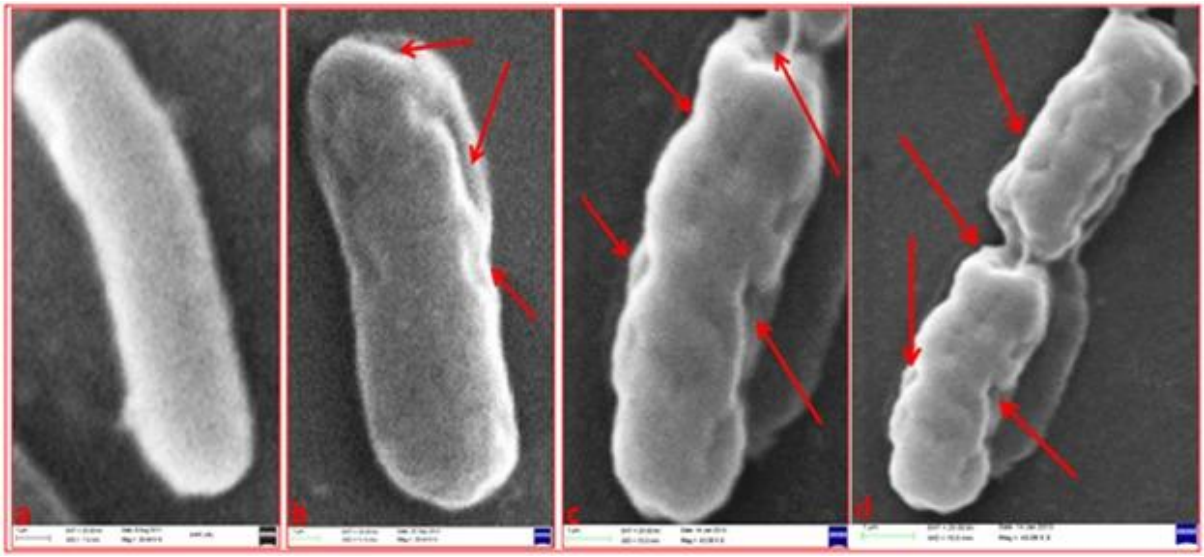


Figure 41 - SEM micrographs of *E. coli*: (a) control and (b) 10, (c) 20 and (d) 30 μg ml^{-1} AgNPs. (Arrows indicate depression and extensive damage of bacterial cells) (Ansari *et al.* 2014) (256)

This phenomenon has yet to be observed within SERS related literature, but it is feasible that damage to the bacteria may result if bacteria are exposed to a SERS surface for an extended period of time, as is likely within the field and this would, therefore, change the spectra generated. Though this could be mitigated through the use of other methods, such as the use of different SERS materials (258) or the use of single-cell channels where cells momentarily pass through a SERS hotspot (259).

The analysis of the arabinogalactan derivative in 3.3.19 highlights the key peak at 477 cm^{-1} which is present due to the polysaccharide nature of the compound. This is likely to be present when analysing the cell wall as well as the whole cell, due to the widespread proliferation of arabinogalactan on the cell wall.

This then raises the question about the presence of other peaks present within the compounds analysed in this chapter. For instance, what about the peaks inferring

the presence of amides, oligonucleotides, carbohydrates or phospholipids? The reasons for this are numerous.

First and foremost, whilst *in silico* analysis can afford researchers a great deal of information without having to expend reagents or put researchers at risk of harm from potentially harmful compounds/biological entities, it is far from perfect.

ORCA is well-known to have slight discrepancies between the proposed Raman spectra of a compound and the observed spectra in a compound and this arises due to several reasons. One such reason could be due to an absence of full optimisation of the angles between different parts of the compound (260).

The impact of this would mean that when the compound is submitted to BlueBEAR, ORCA is unable to comprehensively calculate the Raman spectra. Though it is worth noting that this is likely to be a limitation of the hardware used rather than a user error.

This is not to disparage the data which has been acquired for this chapter. The data acquired can be utilised in chapter 5 and foster an even more in-depth understanding of the cell wall of these bacteria by comparing *in silico* data with *in vitro* data.

Additionally, one must not neglect the original aims of this chapter, which were to simply acquire the *in silico* Raman spectra of compounds associated with the cell wall of *M. tuberculosis* and other *Mycobacterium* species, which we have succeeded in doing.

To further expand upon these aims, one could consider the analysis of other compounds which are present on the cell wall but would be required to ascertain and validate the structure of said compounds as well as ensuring they are suitably optimised for use with ORCA and Avogadro, something which was not within the remit of this project.

An *in silico* approach provides a valuable computational framework for understanding spectral signatures, a crucial aspect necessitating further exploration is the correlation between the predictions generated through *in silico* simulations contained herein and the experimental analyses of extracted material.

Moving forward, the overall properties of these compounds as well as their Raman spectra will be noted upon further in Chapters 4, 5 & 6 and they shall be referenced within these respective chapters, given that the *in vitro* analysis of the cell wall and whole-cell bacteria is likely to present with spectral elements that confer the data acquired within this chapter.

Chapter 4:

4. Optical trapping & Raman analysis of Mycobacterial cell wall lipids:

4.1. Introduction:

The use of optical trapping (also known as optical tweezers) has an established history dating as far back as the 1970s, however, it is only within the last decade or so that the combined use of optical trapping with Raman spectroscopy has gained momentum (261). Through this combined use, it has become feasible to elucidate the properties of bacteria (213), viruses (262), *etc.* that would not always be possible using standard Raman spectroscopy alone. Additionally, the use of optical trapping has potential implications for the observation of pathogens in the field.

Optical trapping itself is carried out by exerting small amounts of force *via* a laser beam, where the narrowest point forms a beam waist (263). Within the beam waist is an extremely strong and concentrated electrical field, which interacts with the surface of an object and enables researchers to lock an item into place, providing the electrical field is applied equally.

By utilising optical trapping, it is possible to explore bacteria in greater detail by locking them into place, enabling extended analysis (264). It also enables easier analysis for smaller compounds/objects. For example, the combined use of Raman and optical trapping in Ghanashyam *et al.* enabled researchers to discern the subtle differences between oxygenated erythrocytes and deoxygenated erythrocytes (265).

Within the context of this project, the use of optical trapping will be applied to understand the cell wall surface of polar lipid and non-polar lipid extracts from the cell walls of *M. tuberculosis*, *M. bovis* BCG & *Mycobacterium smegmatis*.

M. smegmatis belongs to a group of fast-growing Mycobacteria and is therefore distinct from members of the MTBC, though it is readily used as a model organism in research due to possessing closer physiology than other potential model organisms such as *E. coli*. *M. smegmatis* is also of a substantially lower pathogenicity when compared to *M. tuberculosis* and *M. bovis*. Due to its faster growing nature, the mycolic acid profile and the profile of the overall cell wall will also be different to that of the *M. tuberculosis* & *M. bovis* extracts, whilst the cell wall of both *M. tuberculosis* & *M. bovis* BCG are likely to be more similar in nature, due to their closer phylogeny (266).

In undertaking this work, it was hoped that the data acquired will enable the corroboration between the *in silico* data generated in Chapter 3 and consolidate the ability to identifiy the compounds found on the cell wall in Chapter 4.

Specifically, whilst carrying out the Raman analysis of cell wall lipids, one would expect the following peaks to appear, if the *in silico* analysis were to be accurate (**Table 6**, opposite):

Raman Peak (cm^{-1})	Functional Group / Band Assignment
870	Lipid
1057	Lipids
1070	Triglycerides
1073	Triglycerides
1076	Triglycerides
1078	$_{\nu}$(C-C) or $_{\nu}$(C-O) stretch (lipid)
1081	$_{\nu 1}CO_3^{2-}$,$_{\nu 3}PO_4^{3-}$,$_{\nu}$(CC)skeletal of acyl backbone in lipid (*gauche* conformation)
1095	Lipid
1123	C-C Stretching mod of lipids and proteins
1255	Lipids
1260	Amide III, proteins, CH_2 in-plane deformation (Lipids)
1313	CH_3CH_2 twisting mode of collagen/lipid
1369	Guanine, TRP (protein), porphyrins, lipids

Table 6 - Predicted Raman peaks for cellular lipids follow *In silico* analysis of *M. tuberculosis* related compounds

Lastly, given their prevalence within the *in silico* data, one would also expect peaks at *ca.* 1407 cm^{-1} & 1409 cm^{-1} to also appear.

4.2. Methods:

4.2.1. Reagents:

Bacteria were cultured by the Bhatt group within the institute of microbiology and infection at the University of Birmingham, in accordance with the recommendations from the European Collection of Authenticated Cell Cultures (ECACC, Porton Down, UK).

4.2.2. Lipid extraction:

Lipid extraction was undertaken by the Bhatt lab under the following conditions: Cells were centrifuged for 10 mins at 4500 rpm and the supernatant discarded. The pellet was then resuspended in PBS and dried in a speed vacuum for approximately 20 mins. The tube was then weighed. 2mL of methanol within 0.3% NaCl was added, followed

by 1mL of petroleum ether (60-80). This was mixed on a rotator for 15 mins and then centrifuged at 3000 rpm for 3 mins. The supernatant was then placed in a pre-weighed tube (Tube B), the tube with the precipitate in is Tube A. 1ml of petroleum ether was added to tube A and this was mixed for 15 mins on a rotator and centrifuged once at 3000rpm for 3 mins.

The supernatant was placed in tube B and the tubes were then dried and weighed. 2.3 mL of chloroform:methanol: 0.3% NaCl (90:100:30) was added to the pellet in tube A and mixed for 60 mins before being centrifuged for 3 mins at 3500rpm.

The top layer was then transferred into tube C. 750ul of chloroform: methanol: 0.3% NaCl was added to the pellet and this was mixed for 30 mins on a rotator, before being spun for 3 mins at 3500rpm. The top layer was then transferred into tube C. This was repeated once more.

Tube A then consisted of non-extractable lipids and was discarded. To tube C, 1.3mL of chloroform and 1.3mL of 0.3% NaCl was added. This was then mixed for 5 mins on a rotator and then centrifuged for 3 mins at 3500rpm. Following centrifugation, a multi-layered liquid remained and the top layer was discarded. The bottom layer consisted of polar lipids, and this was transferred into a new tube then dried.
Both sets of polar and non-polar lipids were then stored at $4^{o}C$ until needed.

4.2.3. Raman Spectroscopy:

Spectral acquisitions were performed using an InVia confocal Raman (Renishaw). The spectrometer was calibrated prior to each use with silicon (520.7cm^{-1}). A 100x Leica objective and a 1200l/mm grating were used for all measurements in the range of 750-1750cm^{-1}. Raman map scans, consisting of 3 accumulations and 10 second exposure time, were acquired over a 10µm x 10µm square grid using a 785nm excitation laser with laser power of 10-14mW. 100 spectra were collected for each lipid extract of each bacterial species. Generating a total of *n*=600 spectra (3x bacterial species (*M. bovis* BCG, *M. tuberculosis* H37v & *M. smegmatis* MC22) and 2x lipid extracts (polar and non-polar) from each species). Wire 5.1 (Renishaw Plc) software was used for baseline subtraction and cosmic ray removal (267).

4.2.4. Optical trapping:

Optical trapping took place using a 1064nm Cobolt 04-01 Series diode laser (HÜBNER Photonics, Germany) fed directly into the Leica lens *via* a fibre optic cable.

4.2.5. Principal Component Analysis:

Principal component analysis (PCA) was used to interpret the *minute* differences in spectra between the different non-polar lipid and polar lipid extracts from the three bacterial species analysed. As a multivariate data analysis technique PCA was able to reduce the dimensionality of the dataset into the most relevant components to maximize the variance among different samples by projecting the data into a space with N orthogonal basis vectors, arranged in descending order with respect to

variance. Each axis, or principal component, represented a fraction of the total data variance. The dimensionality reduction was achieved by selecting the number of principal components that explains at least the required fraction of the dataset variance. PCA analysis was used to determine how the separated clusters separated between different species and to identify what they had in common. PCA loadings were calculated using the eigenvectors and eigenvalues obtained from matrix operations to find the PC. The loadings contained information specific to the initial Raman data and differed depending on the PC space. In a 2D PCA system, the loadings of the different PC quadrants were expressed as PC1 > 0 and PC2 > 0, PC1 0 and PC2, and PC1 > 0 and PC2 0, resulting in four different loading fingerprints related to the four quadrants (268). PCA analysis was deployed locally using a customised Python script (267)

4.2.6. Code Availability:

The customised written Python algorithm can be downloaded from Vettigli *et al.* (218).

4.3. Results:

4.3.1. Average Spectra:

4.3.1.1. Polar lipid extract:

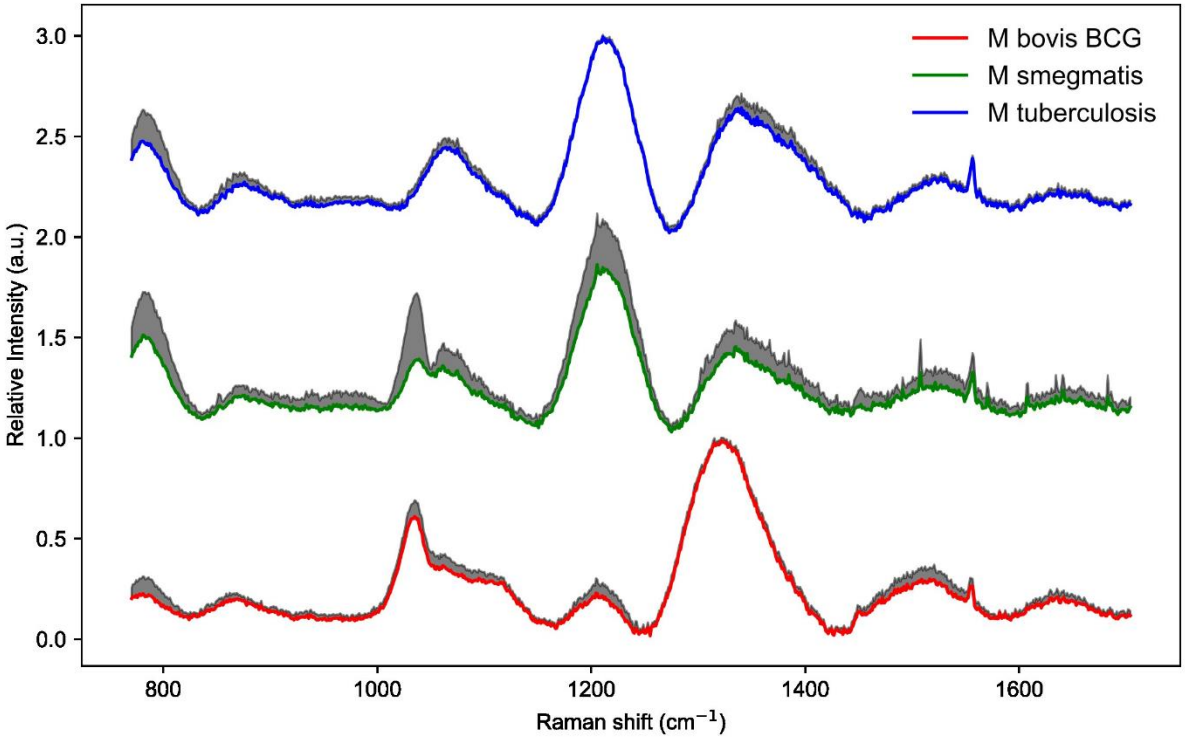


Figure 42 - Polar lipid Average Raman spectra, based on the average spectra of 100x technical repeats. The coloured line is representative of the average, whilst the greyed-out area represents the standard deviation in each sample.

The Raman spectra of the polar lipid extracts from the bacterial species vary between the different species. However, there are particularly noteworthy peaks at *ca.* 750 cm^{-1} as well as *ca*. 870 cm^{-1}. After this, there is a large, intense peak present on all three spectra at approximately 1057 cm^{-1}, this then broadens and decreases in intensity until *ca.* 1180 cm^{-1}.

It is at this point where the polar lipid extracts begin to vary considerably more. *M. tuberculosis* and *M. smegmatis* both possess more intense peaks around 1210 cm^{-1}, in comparison to *M. bovis* BCG. Conversely, *M. bovis* BCG possesses a more intense peak at *ca.* 1350 cm^{-1} , whilst *M. smegmatis* and *M. tuberculosis* possess a somewhat diminished peak at this point. There is then a small yet broad peak at *ca.* 1514 cm^{-1} with a particularly notable peak at *ca*. 1580 cm^{-1}.

4.3.1.2. Non-Polar lipid extract:

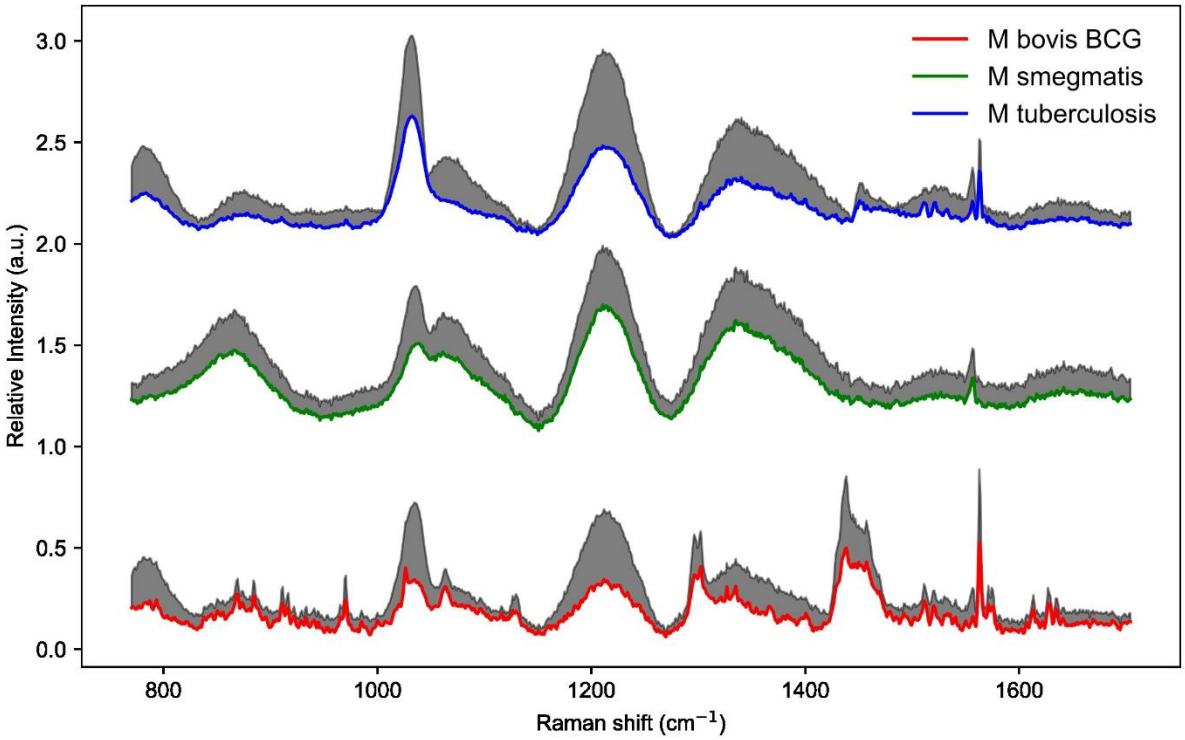


Figure 43 - Non-polar lipid Average Raman spectra. Based on the average spectra of 100x technical repeats. The coloured line is representative of the average, whilst the greyed-out area represents the standard deviation in each sample.

Meanwhile, the non-polar lipid extracts present in a somewhat different manner with a large amount of agreement between the three bacterial species. Though, upon first

glance, it seems that there is a greater level of agreement between *M. tuberculosis* and *M. smegmatis*, with *M. bovis* BCG appearing to be more of an anomaly.

Akin to the polar lipids, there are some particularly notable peaks at *ca.* 750, 870, 1057, 1210, and 1350 cm^{-1}. However, *M. tuberculosis* & *M. bovis* BCG appears to possess a particular unique peak at *ca.* 1420 cm^{-1}, which is substantially more prominent in *M. bovis* BCG. Shortly after this, there is a much milder peak towards *ca.* 1465 cm^{-1} .

After this, there is a particularly prominent peak towards 1606 cm^{-1} and this is of a greater intensity in *M. bovis* BCG.

4.3.2. PCA Quadrant analysis & loadings:

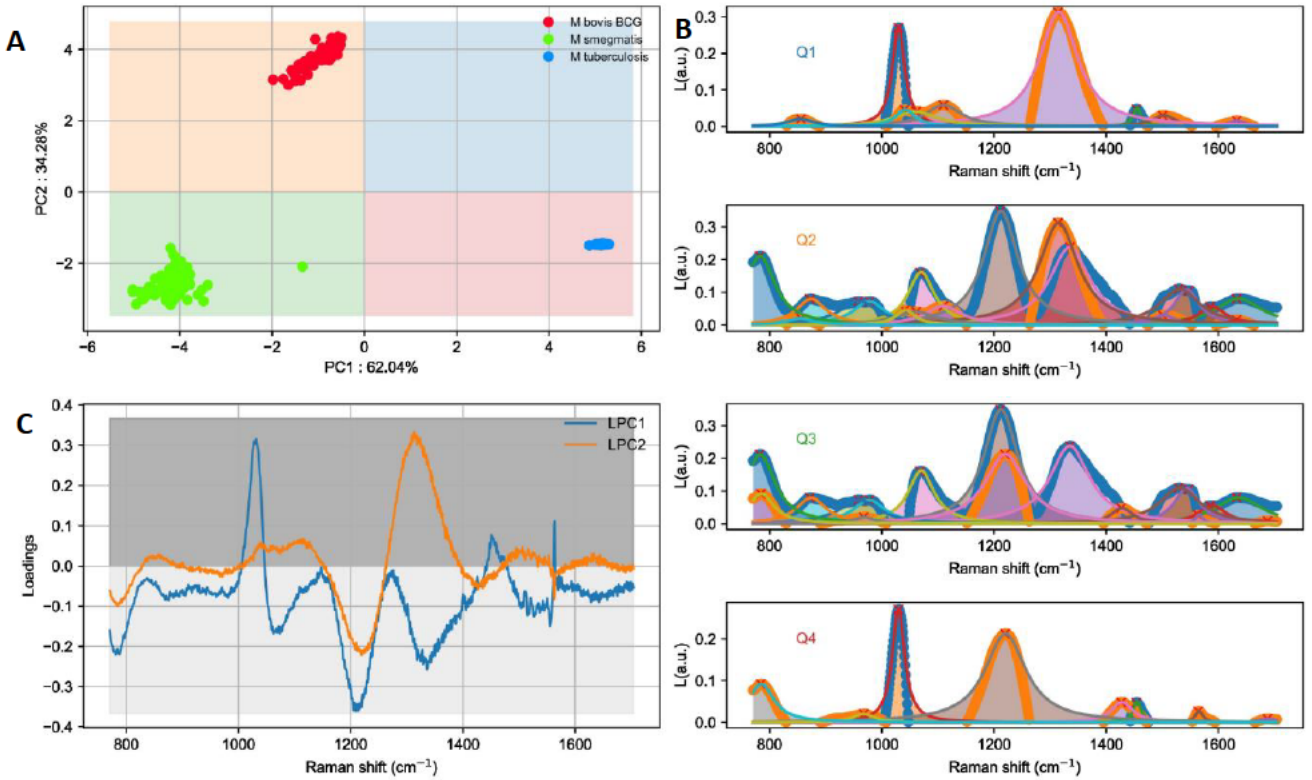


Figure 44 - PCA Quadrant analysis of Polar lipid extract. **(A)**. Quadrant analysis of the PCA data, illustrating the clustering in the quadrants: Q2 (Orange), Q3 (Green), & Q4 (Red). **(B)**. Clustering plot illustrating the most influential spectra in determining the clustering effects observed within the respective quadrants Q1, Q2, Q3, Q4. **(C)**. Raman loadings plot demonstrating the most influential parts of the spectra in the PCA.

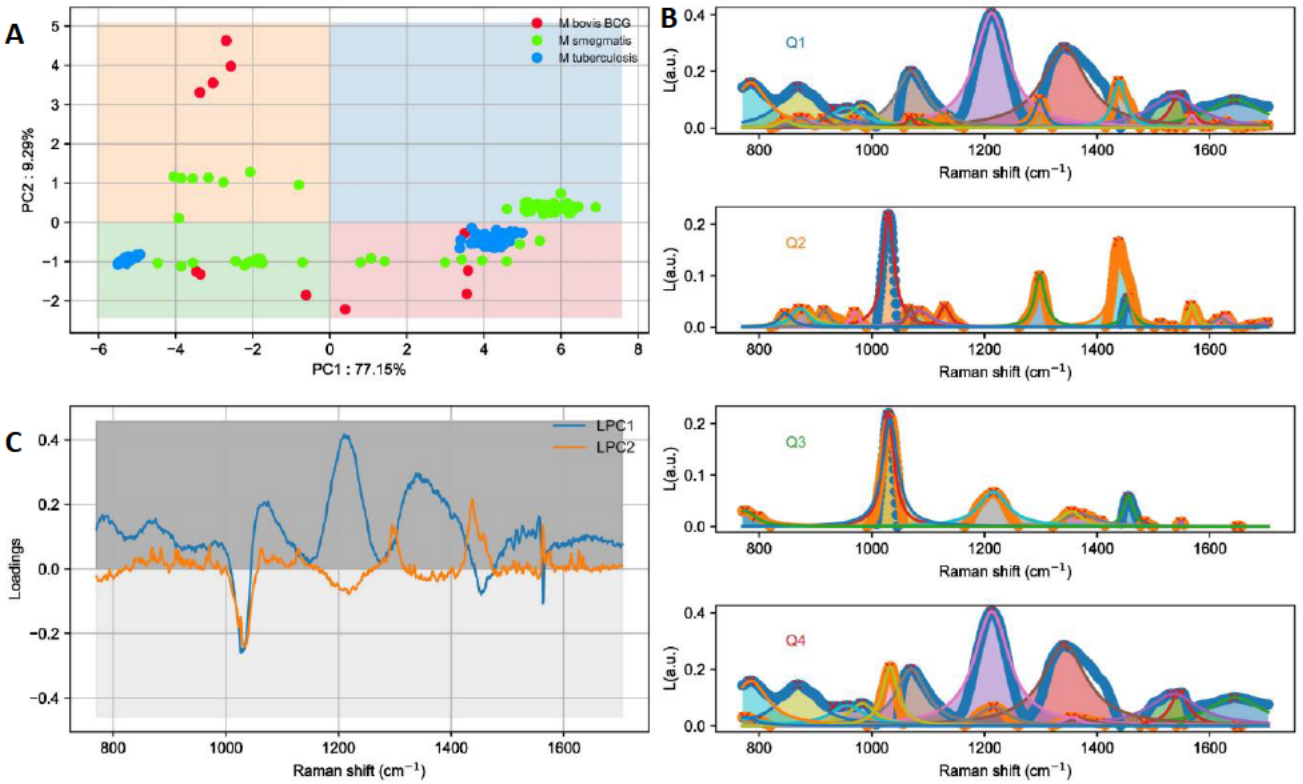


Figure 45 - PCA Quadrant analysis of non-polar lipid extract. **(A)**. Quadrant analysis of the PCA data, illustrating the clustering in the quadrants: Q1 (Blue), Q2 (Orange), Q3 (Green), & Q4 (Red). **(B)**. Clustering plot illustrating the most influential spectra in determining the clustering effects observed within the respective quadrants Q1, Q2, Q3, Q4. **(C)**. Raman loadings plot demonstrating the most influential parts of the spectra in the PCA.

Upon considering the PCA quadrant analysis of both the polar and non-polar lipid extracts of the bacterial species, one can observe that there is far greater amount of consistency between the polar extracts than the non-polar extracts, demonstratable by the tight clusters for each bacterial species which are present on the PCA, suggesting that the spectral data is readily identifiable for each species. This cannot be said for the non-polar lipid extracts where the data points are spread out over the course of the entire PCA. Regardless, by undertaking PCA, one can account for 96.32% of the variance in the polar lipids, and 86.43% for the non-polar lipids.

Moving on to the loadings, it is particularly noteworthy that the most influential peaks for the polar lipids are the peaks present at *ca.* 1057, 1210, and 1350 cm^{-1}. However, due to the great separation of these datapoints, we can begin to dissect the data to a greater extent.

Specifically, if one begins to examine the polar lipid extract quadrant containing *M. tuberculosis* and *M. bovis* BCG, one can observe to observe that the peaks at *ca.* 750, 1168, 1250, and 1350 cm^{-1} are the most influential in determining the presence of polar lipids from *M. bovis* BCG. Meanwhile, we can observe that the peaks centred at *ca.* 750, 1057, and 1250 cm^{-1} are the most significant for *M. tuberculosis*.

Moving on to the non-polar lipid PCA quadrant analysis, it becomes substantially more difficult to determine which peaks have the greatest influence. However, based on the loadings, one can conclude that it is largely the peaks at *ca.* 1057, 1250, 1350, & 1450 cm^{-1} which hold the greatest amount of influence on the PCA.

When considering the quadrants, we can see that the datapoints for *M. smegmatis* are widely spread across all quadrants. *M. tuberculosis* is present in two clusters whilst *M. bovis* BCG is present in all but one quadrant.

However, it appears that there is a large cluster of principal components from *M. smegmatis* in Q1. As such, the peaks at *ca.* 1210, 1095, and 1350 cm^{-1} could therefore be indicative of a a non-polar lipid composition which is more similar to *M. smegmatis.*

Moving into Q2, a particularly intense peak at 1057 cm^{-1} is observed and, given the absence of this peak in Q1, it is likely that this is present due to non-polar lipids from *M. bovis* BCG though the influence of other datapoints cannot be totally excluded (Whilst the division of the PCA into quadrants is useful for the purposes of processing data, it does not completely sever the influence of other datapoints in the generation of spectral loading schemes and nearby clusters of data will influence the quadrant loading graphs). There is also another peak at *ca.* 1450 cm^{-1} .

Q3 presents with a more intense peak at 1057 cm^{-1}. Additionally, there is a small peak at *ca.* 1250 cm^{-1} which was absent in Q2. This is likely because of the presence of *M. tuberculosis* and *M. bovis* BCG.

Q4 presents with multiple peaks, and these peaks are likely to be influenced by the presence of all three bacterial species. Of particular note are the peaks at *ca.* 750, 850, 1057, and 1095 cm^{-1}. This is before the more intense peaks at 1250 and 1350 cm^{-1} is observed.

To consolidate the spectral elements detected in this chapter, the peak assignments are presented in the **Table 7**, below:

Raman Peak (cm^{-1})	Functional Group / Band Assignment
750	$CH_{2,6}$ out-of-plane bending, observed in the spectra of a single human erythrocyte
780	Uracil-based ring breathing mode
870	Lipid
1057	Lipid
1095	Lipid
1168	Lipids, ν(C=C) δ(COH) (lipid assignment), ν(C-C), carotenoids
1210	Unassigned
1255	Lipids
1420-1450	CH_2 scissoring vibration (lipid band)
1450	CH_2 Bending
1465	Lipids
1514	ν(C=C) Carotenoids
1580	C-C Stretching
1606	C=C Bending

Table 7 - Raman peak assignment for the most prominent peaks in the Raman spectra of non-polar and polar lipid extracts

4.4. Discussion:

As one would expect, most of the peaks which confer the highest loading values (based on their peaks for quadrant analysis) within the PCA for this chapter are closely associated with lipids. Of the loading values, some of those most influential as principal components were previously detected in the *in silico* analysis in Chapter 4 and the data generated within this chapter corroborates the theoretical data obtained in Chapter 4, thus contributing to its validation. Also observed is an expansion of the Raman spectra beyond *ca.* 1450 cm^{-1}, which the *in silico* analysis was unable to do.

The Raman spectra of these three bacterial species does indeed vary (after all, they are three distinct species of bacteria), and the profile of the spectra of the polar and non-polar lipids also varies. This can be due to a variety of reasons. Primarily, as has been previously discussed, the defining feature of the *Mycobacterium sp*. cell wall is the presence of mycolic acids, and the isomers of these mycolic acids vary in ratio based upon a variety of factors (*e.g.* growth phase, bacterial species, nutrition, *etc.*), therefore the Raman spectra will vary based upon these ratios (251)

It is also likely that the data will vary somewhat based upon when the cells are analysed within their life cycle (269). Therefore, future research could consider time-based experiments where the lipid extraction takes place at pre-defined times within the cell life cycle. This is something which has previously been undertaken by Verma *et al.* (183) albeit to the exclusion of the species analysed in this chapter.

Additionally, if one were to repeat these experiments, it would prove worthwhile in analysing the lipids from several repeat cultures, something which was not possible during these experiments due to the limited amount of lipid extracts that were available. And to incorporate other bacteria, to elucidate which Raman peaks are unique to *Mycobacterium sp*.

For further consideration, there is also the presence of carotenoids. Carotenoids are pigmented compounds that are fat-soluble and would be a by-product of the extraction process contained within this chapter given their lipid solubility (270).

Carotenoids also provide valuable insight into the life cycle of a *Mycobacterium sp.* cell. As the cell ages, the presence of carotenoids on the cell wall decreases and this

could provide a valuable standard for the understanding of the cell wall when it is being analysed. Suffice to say, if one were able to correlate the relative intensity of the carotenoid-associated peaks within the Raman spectra, it would enable researchers to standardise the analysis of cell wall lipids and *Mycobacterium sp.* in general (270).

The data acquired from this analysis is deeply insightful and enables a greater understanding of the profile of the bacteria these lipids have been extracted from. Of note, is the omnipresent Raman peaks which can be easily assigned to lipids as detailed in **Table 7**. More specifically is the peak at approximately 1210 cm^{-1}, which remains unassigned within this context. Talari *et al.* (250) does assign this peak, however, the consensus is that this peak is assigned to C-C_6H_5 stretching modes of tyrosine and phenylalanine, or $_{v18}(\delta:C_mH)$, observed in the spectra of a single erythrocyte.

Naturally, one could formulate an argument that such a peak could arise due to $_{v18}(\delta:C_mH)$ and that the observation of this peak in erythrocytes is a pure coincidence. Instead, it could be that this peak is ubiquitous across any Raman spectra where large hydrocarbon chains are present. However, the updated Ghanashyam *et al.* (265) fails to illustrate the presence of such a peak in oxygenated and deoxygenated erythrocytes. The work undertaken by Perumal *et al.* (251) does indeed highlight a peak at *ca.* 1210 cm^{-1} at this point in the SERS spectra which are associated with Keto- and methoxy-mycolic acid but not alpha-mycolic acid.

The presence of said peaks has not been noted in other research which has considered the whole organism such as that conducted in Stöckel *et al.* (179) and Mühlig *et al.* (173). It is therefore possible that, when considering the full organism, the peak itself is drastically reduced as other spectral elements are more prominent.

Additionally, the work of Perumal *et al.* (251), Stöckel *et al.* (179) and Mühlig *et al.* (173) predominantly takes place under Surface-Enhanced Raman Spectroscopy (SERS) conditions rather than Raman, it is therefore possible that the peak observed in this chapter may be less prominent under SERS or it may be that it shifts somewhat, as is common when comparing SERS and Raman spectra in *Mycobacterium sp.* (173). When considering Raman, of the research considered, Hanson *et al.* (269) and Baron *et al.* (270) give contradictory results, where the peak at *ca.* 1210 cm^{-1} is not present in the former, but is present in the latter.

The deployment of SERS combined with optical trapping is something that future experiments may wish to consider as well, given that SERS can often enable a greater amount of resolution compared to Raman spectroscopy (271).

As for the cause of this peak, given that it is present in the lipid extract of *Mycobacteria sp.*, it is likely a rich lipid, though one could not draw a conclusion of the precise structure of the compound without a greater amount of analytical data.

Secondly, whilst the peak at *ca.* 750 cm^{-1} initially appears to be associated with $CH_{2,6}$ out-of-plane bending, it is my belief that this is a relatively broad peak, which could incorporate peaks corresponding to the presence of lipids. This namely consists of peaks present at *ca.* 727/728 cm^{-1} which have been assigned to lipids.

One way to investigate this would be to expand upon the range one acquires the Raman spectra, if one were to repeat these experiments, as well as consider the use of SERS.

To further characterise this peak, one could also utilise other analytical techniques such as Liquid Chromatography – Mass Spectrometry (LC-MS) and Infrared (IR) Spectrometry. With the former offering insight into the composition of the lipids in the cell wall in terms of isomers present, and the latter affording more structural information on the lipids themselves.

However, as we see a shift from purified lipid extracts and on to whole cells as per the next chapter, one may observe that the peak at *ca.* 1210 cm^{-1} may disappear or diminish due to the amount of other compounds present on the cell wall and in the cell (*i.e.* nucleic acids, proteins, carbohydrates, *etc.*). Regardless, this peak at 1210 cm^{-1} is of great interest and needs to be considered moving forward (as well as with the other notable Raman peaks in **Table 7**).

Chapter 5:

NB: At the time of writing, the following chapter forms the bulk of Macgregor-Fairlie *et al.* (267) which is currently undergoing peer review for PLOS One.

5. Raman analysis of the Respiratory Tract Microbiome of *Bos taurus*:

5.1. Introduction:

Like humans, there are a large number of microorganisms residing within cattle; this are approximately $10^{10} - 10^{11}$ cells/mL in the gut (272). However, just like humans, the airway differs slightly from the gastrointestinal tract in terms of microbes present (273, 274). The bacteria present within the respiratory system also varies pending upon where a sample is acquired from (*i.e.* the number of and variety of bacteria in the pharynx of a cow will vary from those acquired from the nose) (275). There is also a strong consensus that there is often little difference between the number of species of bacteria within an area, but rather the relative abundance of certain bacterial species that cause disease (276).

Whilst this can cause issues when it comes to identifying some respiratory diseases such as Bovine Respiratory Disease (BRD) or pneumonia, where it is likely a result of dysbiosis rather than an infection *per se* (277), *M. bovis* is an obligate pathogen (278). Therefore its presence is confirmation of bovine TB.

The issue with direct detection of the pathogen stems from other bacteria that are present within the microbiota of the animal. When culturing, this is overcome by using selective media that seeks to isolate *M. bovis* using various selective agents.

However, this selectivity cannot be provided when attempting to directly detect bacteria from patient samples which will contain a multitude of other bacterial species. As a result, when attempting to detect *M. bovis* using any spectroscopic method, other bacteria that may be present must be taken into consideration as they are likely to be detected alongside *M. bovis*, this is something that other researchers have failed to undertake in the past.

Zeineldin, Lowe and Aldridge (276) go on to review several studies which have sequenced the genome of the various bacteria that inhabit the microbiome of the respiratory system of cows. The studies covered provide a great deal of insight into the microbiome and allow us to understand the relative abundance of organisms as well as specific organisms.

Consequently, when developing a Raman-based assay, it is necessary to construct a library of Raman spectra generated from the named organisms as well as *M. bovis*. In doing so, differences can be detected and a part of the spectra that is unique to *M. bovis* can be identified. This library must consist of the organisms which occur most frequently in the respiratory tract of cows.

The studies covered by Zeineldin, Lowe and Aldridge (276) only discriminate to the genus level, though phylum discrimination remains consistent throughout the studies highlighted, as well as other studies brought to attention through a literature search. Of the phylum's present within the microbiome of cattle, Ternericutes, Proteobacteria, Firmicutes, Bacteriodetes, and Actinobacteria dominate, though the relative abundance of these varies from study to study.

Through breaking these down to the genus level a better picture can be acquired and we begin to see the domination by specific genera of bacteria: *Moraxella sp., Sphingomonas sp., Mycoplasma sp., Corynebacterium sp., Aggregatibacter sp., Psychrobacter sp., Ureaplasma sp., Pastuerella sp., Bacteroides sp., Mannheimia sp.,* and *Ureaplasma sp.*. Some studies have also indicated the presence of *Streptococcus spp.*, *Staphylococcus sp.*, *Pseudomonas sp.* and *Bacillus sp. (276, 279, 280)*. Whilst these studies do not identify species, it is possible to build up a better picture of what the proposed library should consist of.

Out of the studies covered in the above reviews, as well as other sources that were found during a literature search, only three papers could be found that illustrate which bacteria were detected down to the species level (281, 282, 283), though it is worth noting that there is a large degree of variability between cattle samples and, thus, it is likely that species present within cows will vary between herds, especially when discussing relative abundance. It is possible that several different species of bacteria may be present within the same animal albeit from the same genus, *e.g. Staphylococcus equorum* and *Staphylococcus sciuri* as documented by (282).

Of the literature that then goes on to discuss the species of bacteria that are present within the microbiome of cows, we can establish that the bacteria in the **Table 8** opposite are present across multiple herds:

Genus	**Species:**
Aggregatibacter	No Species stated
Bacillus	*B. cereus, B. flexus*
Bacteroides	*B. eggerthii, B. coprophilus, B. fragilis, B. plebius, B. uniformis*
Corynebacterium	*C. variabile, C. pilosum*
Mannheimia	*M. haemolytica*
Moraxella	*M. bovoculi*
Mycoplasma	*M. bovis, M. bovirhinis, M, dispar, M. hyopneumoniae*
Pastuerella	*P. multocida,*
Pseudomonas	*P. fluorescens, P. putida, P. syringae*
Psychrobacter	*P. Sanguinis, P. pacificensis*
Sphingomonas	*S. yabuuchiae, S. wittichii*
Staphylococcus	*S. equorum, S. sciuri*
Streptococcus	*S. salivarius*
Ureaplasma	*U. diversum*

Table 8 - Table illustrating which named species have been detected in three papers) (281, 282, 283)

When considering the above species, the ability to acquire the bacteria as well as containment restrictions were also considered within the limitations of this project at the University of Birmingham's School of Chemical Engineering. Therefore, when acquiring the bacteria, a compromise was sought between the containment facilities available, the bacteria which could be safely contained and the demands of this project. A heavy focus was placed on acquiring *Mycoplasma sp.* due to it forming the overwhelming majority of microbiome in the respiratory tract (*ca.* 80% of bacteria) (283) Having previously been detected in cattle and due to the presence of mycolic acid within its cell wall, *Rhodococcus equi* or a similar species from the genus was also sought. Ultimately, the following bacteria were acquired as per **Table 9** opposite.

Genus	Species	Culture collection ref	Source
Bacillus	*cereus*	NCTC 11143	Elaine Mitchell, University of Birmingham
Corynebacterium	*glutamicum*	ATCC 13032	Gurdyal Besra, University of Birmingham
Escherichia	*coili*	Top 10	Thermofisher
Mannheimia	*haemolytica*	NCTC 10638	NCTC, UK Health Security Agency
Moraxella	*catarrhalis*	NCTC 11020	NCTC, UK Health Security Agency
Mycobacterium	*bovis* BCG	NCTC 5692	NCTC, UK Health Security Agency
Mycoplasma	*bovis* *hyosynoviae*	NCTC 10131 ATCC27095	NCTC, UK Health Security Agency Andrew Rycroft, Royal Veterinary college
Pseudomonas	*fluorescens* *aeruginosa*	NCIMB 9046 NCTC 10662	Tim Overton, University of Birmingham
Rhodococcus	*erythropolis*	NCIMB 10422	Elaine Mitchell, University of Birmingham
Staphylococcus	*aureus*	NCTC 10788	NCTC, UK Health Security Agency

Table 9 - Table illustrating bacteria acquired for this project; displayed with full name, culture reference and original source.

Of greater relevance to this project is the analysis of *M. bovis* BCG, *R. erythropolis* & *C. glutamicum* which belong to the Mycobacteriales genus and, as such, possess similar physical properties. It was the Raman analysis of these bacteria that were presented in Macgregor-Fairlie *et al.* (267) and it is this work and techniques which shall be discussed in this chapter, alongside the bacteria acquired in **Table 9**. Initial

investigation sought to investigate the Raman spectra of all bacteria identified above and then scrutinise the Raman spectra of Mycobacteriales to a greater extent, due to their similar physical properties and the hypothetical potential for spectral crossover.

This chapter therefore aims to establish the Raman spectra of bacterial genera which have a confirmed presence within the microbiome of cattle. It then seeks to use a multivariate analysis to, firstly, investigate the potential for spectral crossover of *M. bovis* BCG with other bacteria considered in this study. It will then further analyse the spectra (*via* the use of more advanced tools) of the relevant Mycobacteriales and identify specific spectral elements which are unique to *M. bovis* BCG and discern it apart from *C. glutamicum* & *R. erythropolis*, due to them all being Mycobacteriales.

The ultimate goal of this particular chapter is that the methods and materials can be readily utilised and applied to more clinically relevant models which would incorporate *M. tuberculosis*. In the meantime, the work conducted herein is a relevant disease model for bovine TB, a disease which causes significant economic harm to farmers in the UK, as discussed in chapter 1.

5.2. Methods:

5.2.1. Reagents:

Bacteria were prepared as per Chapter 2.2.1. The cells were then allowed to air dry under laminar flow for two hours (173).

5.2.2. Raman Spectroscopy:

Spectral acquisitions were performed using an InVia confocal Raman (Renishaw). The spectrometer was calibrated prior to each use with silicon (520.7cm^{-1}). A 100x Leica objective with a 0.8 aperture and a 1200l/mm grating were used for all measurements in the range of 750-1750cm^{-1}. Raman map scans, consisting of 3 accumulations and 10 second exposure time, were acquired over a 10µm x 10µm square grid using a 785nm excitation laser with laser power of 10-14mW. 100 spectra were collected for each bacterium with 3 repeats for *M. bovis* BCG, *C. glutamicum*, *R. erythrpolis*, generating a total of *n*=300 spectra for these three bacterial species. For the remaining bacterial species, 1 repeat was undertaken which presented with 100 spectra for each species. Wire 5.1 (Renishaw Plc) software was used for baseline subtraction and cosmic ray removal.

5.2.3. Principal Component Analysis:

Principal component analysis (PCA) was used to interpret the *minute* differences in spectra between the different bacterial strains, in a similar fashion to Chapter 5. This was initially performed on all bacterial species, and then followed up by further comparative analysis on the repeats of *C. glutamicu*m, *R. erythropolis, & M. bovis* BCG. A further quadrant analysis was applied as in **Chapter 5.**

5.2.4. Principal Component Analysis & Linear Discriminant Analysis:

Further, PCA was used as a pre-processing tool to reduce the dimensionality of the Mycobacteriale dataset and then used as input data to the linear discriminant analysis (LDA). The PCA pre-processing step yielded an inherent benefit of all the

output data being orthogonal, which circumvented a weakness of LDA *i.e.,* the collinearity. LDA, like the PCA, sought to maximise variance between components however, instead of within the dataset, LDA maximised variance between groups. This meant that it has projected the data into a space such that the variance between each Mycobacteriales was maximised. The comparison of the pure PCA and PCA-LDA approaches were compared for their relative effectiveness in classifying spectra and to the self-organising maps classification approach (267).

5.2.5. Self-Organising Maps:

Self-organising maps (SOMs) were used for multivariate analysis (284)-(268) of the Mycobacteriale dataset. SOMs are single-layer artificial neural networks that are represented as a two-dimensional (2D) hexagonal array of neurons. Inspired by the visual cortex in the brain, the SOM is trained to activate neighbouring neurons based on similar inputs, in this case Raman spectra from multiple repeats. Each neuron has a weight vector with a length equal to the number of variables in a spectrum. The weights are gradually adjusted to be similar to the input data by exposing the network to training samples over several iterations, so that each neuron only activates on a specific spectral signature. The result is a 2D projection of hyperspectral data that can be seen as visible clustering based on type, group, and state. SOM employs the self-organizing map discriminant index (SOMDI), which appends a set of label vectors to each neuron and allowed us to study the most prominent features that cause the activation of a specific neuron to a class label. Following that, a supervised learning step was introduced to optimize the network

and the class label associated with each neuron was used to quickly identify new data presented to the SOM. SOM was utilised solely for the analysis of the Mycobacteriales due to the wider availability of training and test data as well as an increased amount of variability between datasets, thus being more reflective of real-world circumstances (224, 267).

5.2.6. Raman Peak Analysis:

The most important peaks from the Mycobacteriales were extracted by applying a multi-Lorentzian peak fitting method (268) followed by a peak comparison among the different samples and selecting the common peaks between samples to perform statistical analysis. A box plot of the Raman intensity of these peaks was performed and compared among samples using multiple pairwise comparisons, Tukey's honestly significant difference test (Tukey's HSD) to investigate the differences. Each spectrum was baseline subtracted and normalized between 0 and 1 using the asymmetric least square method (267).

5.2.7. Code Availability:

The customised written Python algorithm can be downloaded from Vettigli *et al.* (218) & Weston (219).

5.3. Results:

5.3.1. Average Raman spectra:

A broad average representative Raman spectra in **Figure 46** shows the spectral differences between the bacteria analysed in this study, whilst **Figure 47** shows the broad average spectrum of the Mycobacteriales.

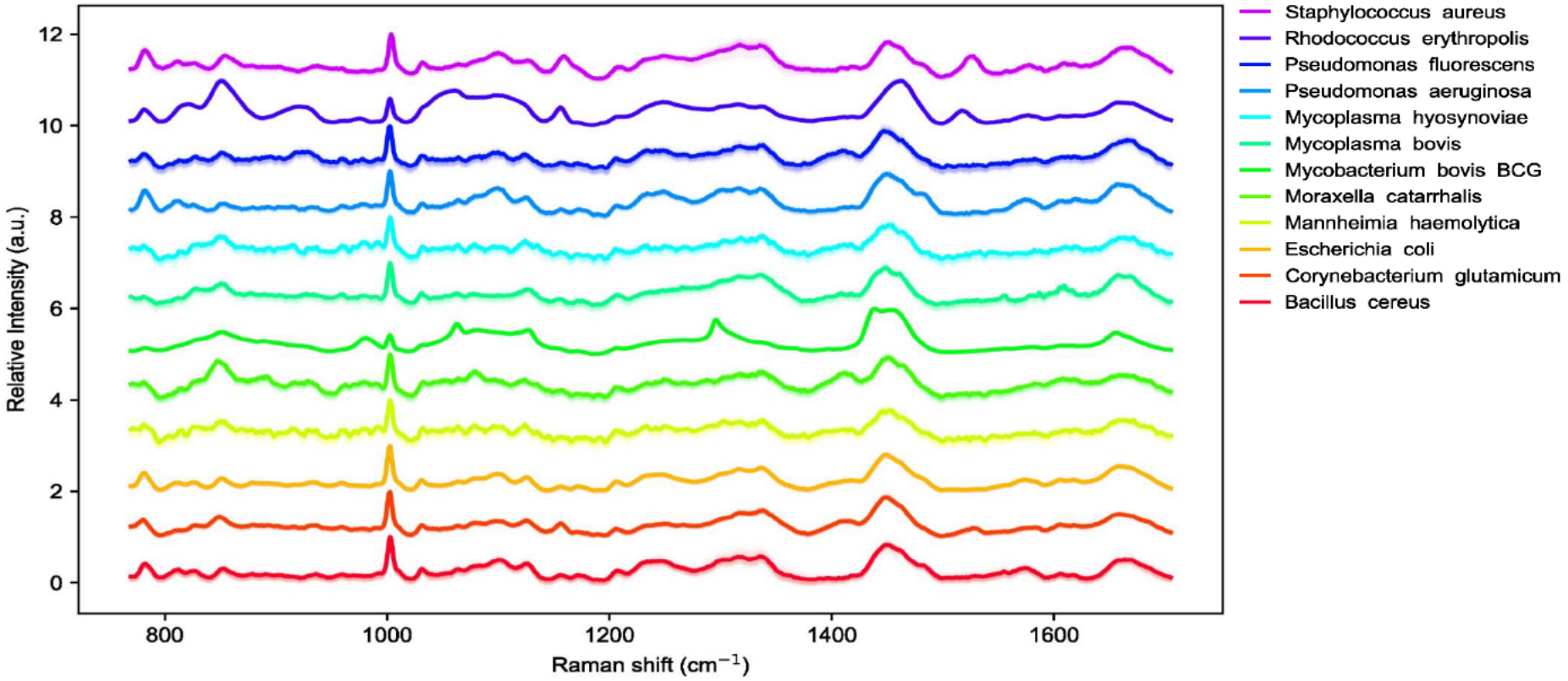


Figure 46 - Average Raman spectra of the bacterial species analysed in this chapter. Average is based on 100 spectra per bacterial species, arising from 100 technical repeats. The slight blurring of the line is representative of the standard deviation in the sample.

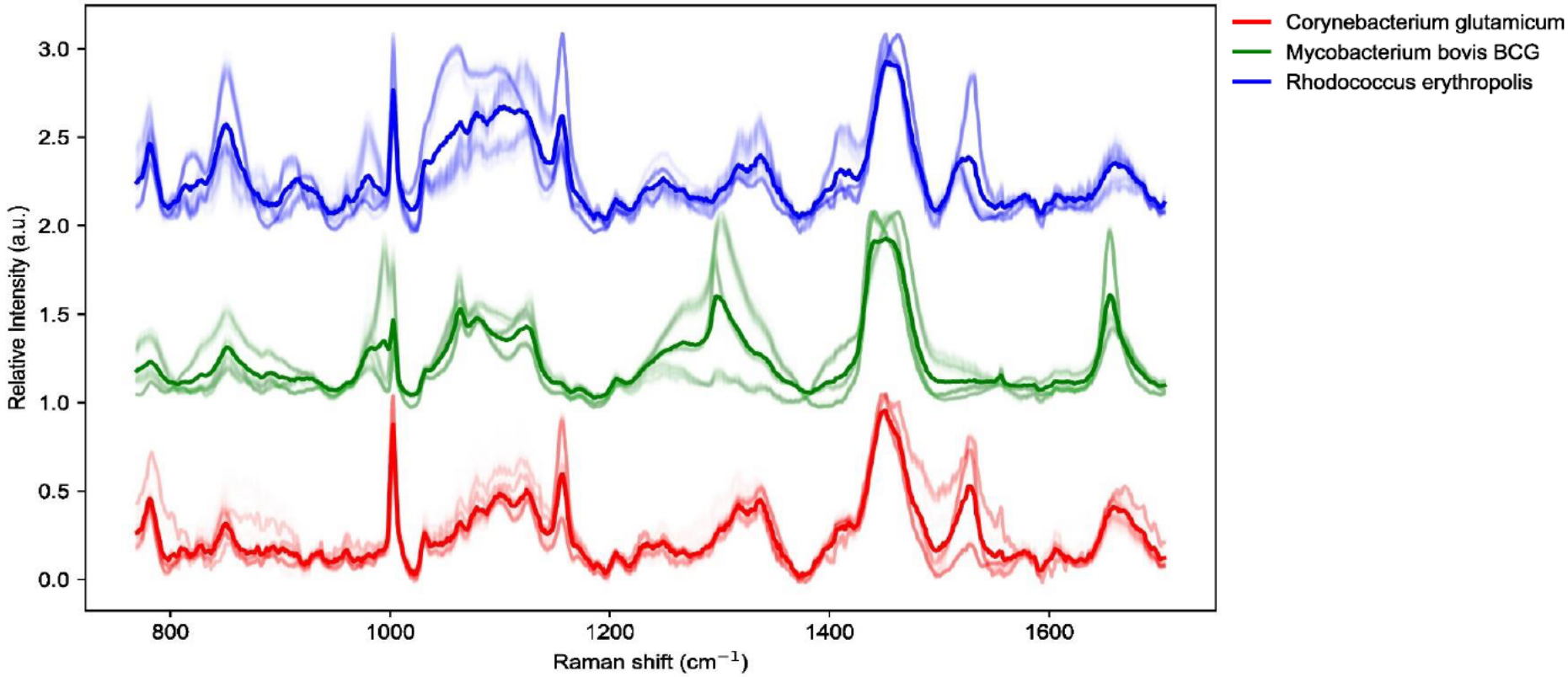


Figure 47 - Average Raman spectra of the Mycobacteriales analysed in this chapter. Average is based on 300 spectra per bacterial species. The bold line represents the average spectra. The blurred lines represent the standard variation in each biological repeat.

Within these spectra, several prominent peaks can be observed. Broadly speaking, notable peaks occur at 1002 cm^{-1} & 1450 cm^{-1} for all bacteria. The peak at 1002 cm^{-1} is likely due to, either, C-C stretching (250), all of which one would expect to observe in bacteria, due to the structures present (285, 286, 287). The peak occurring at *ca.* 1450 is likely due to CH_2 bending, or $\delta(CH_2)$.

For the remaining peaks, the following intense peaks can be assigned, based on the Average Raman spectra (For greater resolution of the average Raman spectra of the non-Mycobacteriale bacteria, the spectra and peak assignment are viewable in the appendix (

A.8. Raman spectra of individual bacteria and peak **assignment**:**).** (250)).

5.3.2. PCA analysis:

Clustering of the different bacteria obtained *via* PCA (**Figure 48**, opposite) can be observed. The initial PCA analysis of the data also shows a great amount of spectral crossover, with a large degree of agreement between all bacteria studied, except for *M. bovis* BCG, which acts as an outlier. Though only 54.77% of the variance can be explained. Further PCA of the Mycobacteriales and their respective repeats illustrates a greater amount of variability with widespread clusters (**Figure 49**, page 143) with approximately 62.91% of the variance explained by using two principal components. The greatest cluster domain spread is observed in *R. erythropolis* and *M. bovis* BCG, which is also evident from the Raman spectra.

Whilst this method is widely used across literature, limitations exist when using it. Specifically, the data can become unwieldy when using multiple samples (such as the number of bacteria utilised herein). As the dataset size increases, the computation of eigenvectors and eigenvalues becomes more resource-intensive. Additionally, PCA assumes linear relationships between variables, which might not hold in complex datasets. Non-linearity, outliers, and skewed distributions can distort results. Another concern is interpretability; while PCA identifies patterns, interpreting the biological or practical significance of these patterns may be challenging. Moreover, PCA may not capture subtle non-linear relationships, limiting its ability to unveil the full complexity of large and intricate datasets. Therefore, while PCA is valuable, an awareness of these limitations must be maintained.

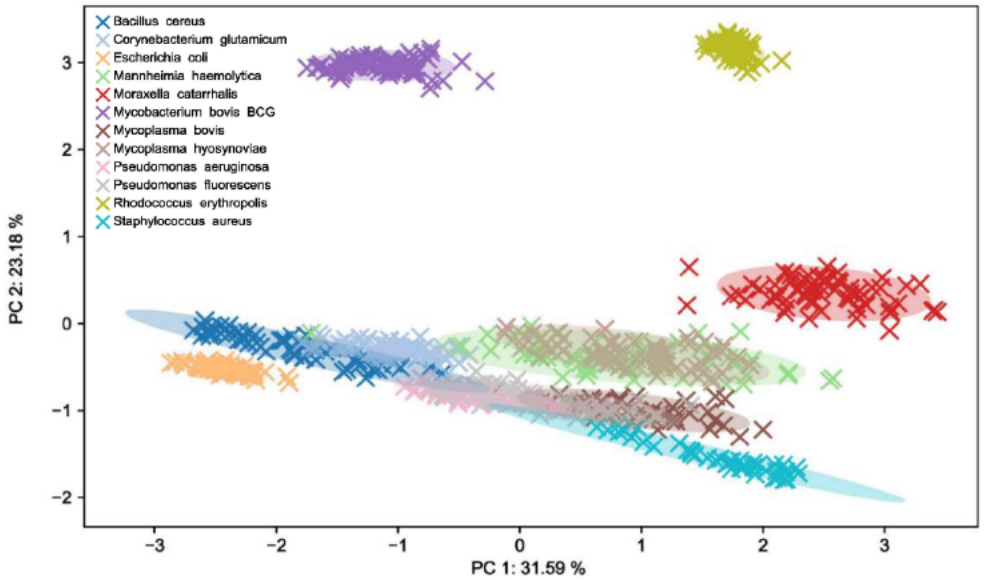


Figure 48 - PCA of Raman spectra of all bacteria analysed in this project with two components, and 100x technical repeats, with one biological repeat each. Note the broad overlap between the majority of the bacteria. *Mycobacterium bovis* BCG, *Moraxella catarrhalis* and *Rhodoccus erythropolis*.

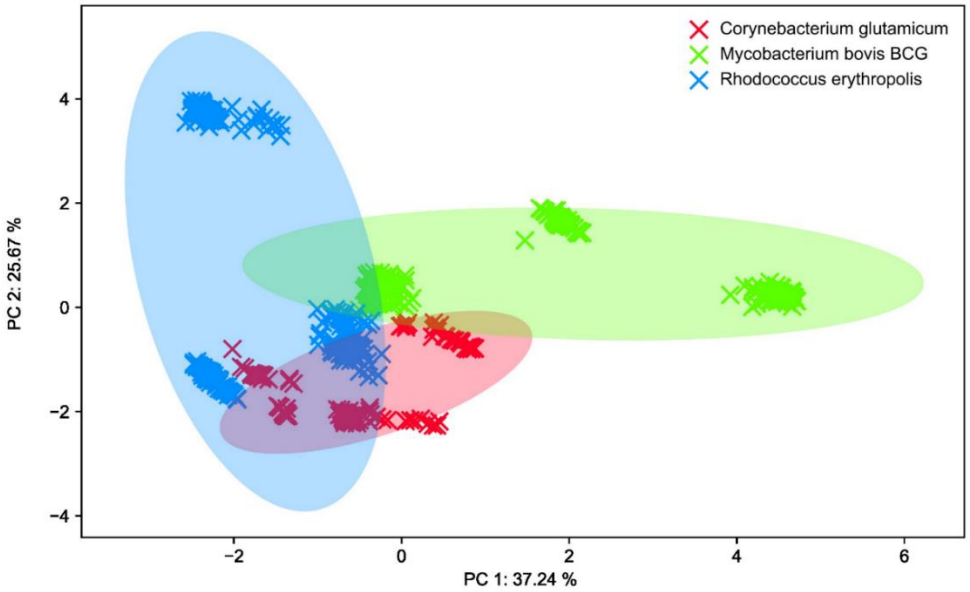


Figure 49 - PCA Analysis of the Mycobacteriales utilising two principal components following three biological repeats of the Mycobacteriales utilised in this project. 100x technical repeats were used in each biological repeat leading to N=300 spectra per bacterial species. Note the widespread of each bacterial species and the crossover of the Raman spectra.

5.3.3. PCA-quadrant analysis:

Figure 50a (page 145) illustrates a population of *R. erythropolis* and *C. glutamicum* samples in the 3rd, green quadrant (Q3) was observed with the corresponding loading spectra (**Figure 50b**) found to be related to the Q3 loading identified by the presence of multiple peaks, highlighting the relative similarities between both species of bacteria. *M. bovis* BCG located in Q1, shows a high signal intensity for the 1050cm^{-1} and 1300 cm^{-1} bands resulting in greater separation of the *M. bovis* BCG relative to the other bacteria. *C. glutamicum also* exhibited a cluster in Q4 due to the presence of a high-intensity peaks at 1150cm^{-1} and the 1300cm^{-1} region and the *R. erythropolis* shows a separated cluster in Q2, which is related to the presence of a high-intensity

peak at 1050cm^{-1} combined with other peaks present in Q3 yet, with the absence of a strong peak at 1300cm^{-1}.

A region where most samples converge was identified between [-3,1] in PC1 and [-3,1] in PC2 thus, revealing a commonality among different samples although both *M. bovis* BCG and *R. erythropolis* have a larger variation in spectral signal which enables them to populate other regions of the PC space. This has also been identified from the Raman loadings plot (**Figure 50c**, opposite), where a few spectral lines above and below the sample average were indicative of the Raman signal variance.

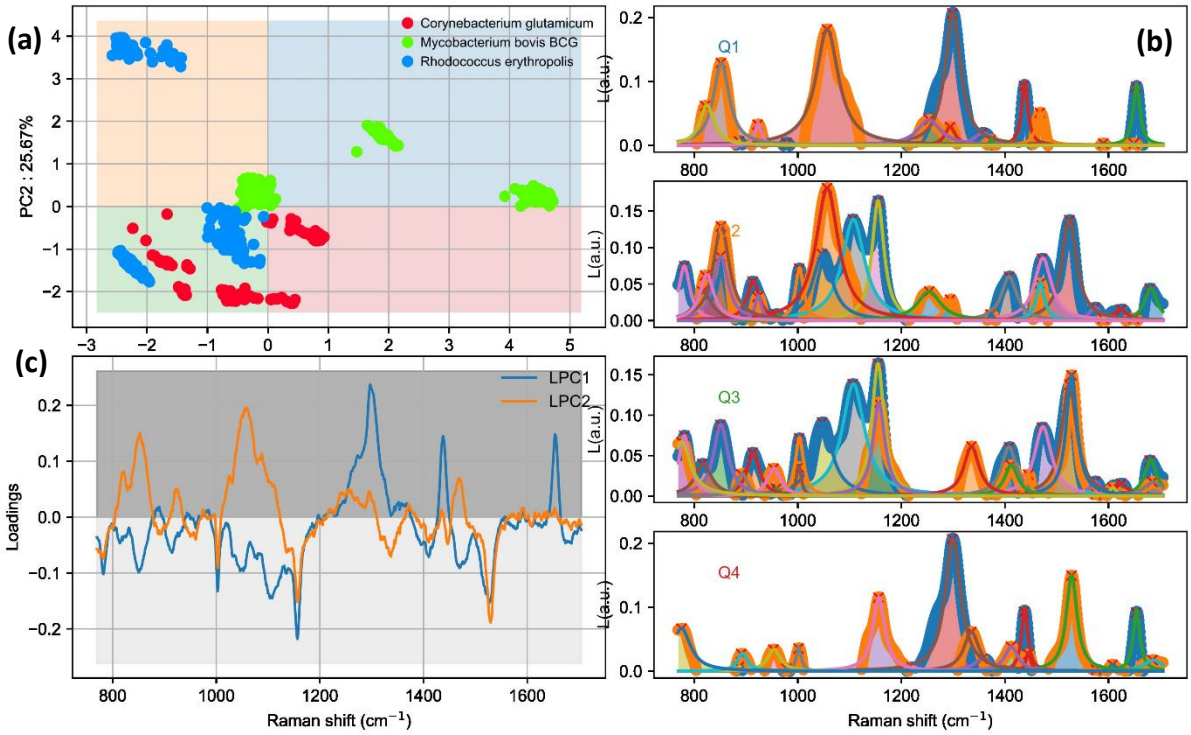


Figure 50 - (a) Quadrant analysis of the PCA data, illustrating the clustering in Q3 (Green). (b) Clustering plot illustrating the most influential spectral elements in Q1, Q2, Q3 and Q4. (c) Raman loadings plot demonstrating the most influential parts of the spectra in the respective principal component (LPC1, LPC2).

The variance highlighted in **Figure 50** is highlighted *via* SOM (**Figure 51a**, opposite), where the sample separations clustering is identified as a colour intensity. The darker green hexagon is indicative of a more distinct spectral region for *M. bovis* BCG with the green hexagon sandwiched by the blue hexagons representing a degree of crossover between *M. bovis* BCG, and *R. erythropolis,* respectively, further consolidating the PCA observations. *R. erythropolis* (blue) and *C. glutamicum* (red) appear mostly similar except the different peak intensities in the 1000 cm^{-1}, 1200 cm^{-1} region as well as the 850 cm^{-1} and 1550cm^{-1} peaks and *M. bovis* BCG (green) appears to be more distinctive especially, in the 1200cm^{-1} & 1400 cm^{-1} region along with an absence of a peak at 1550 cm^{-1} and a more intense peak at 1700 cm^{-1}.

The activation peaks (**Figure 51,** below) correspond to the peak statistical analysis with *R. erythropolis* (blue) and *C.* glutamicum (red) being predominantly similar whilst *M. bovis* BCG (green) exhibiting is more distinct spectral fingerprint particularly, in the 1300 cm^{-1} & 1650 cm^{-1} region.

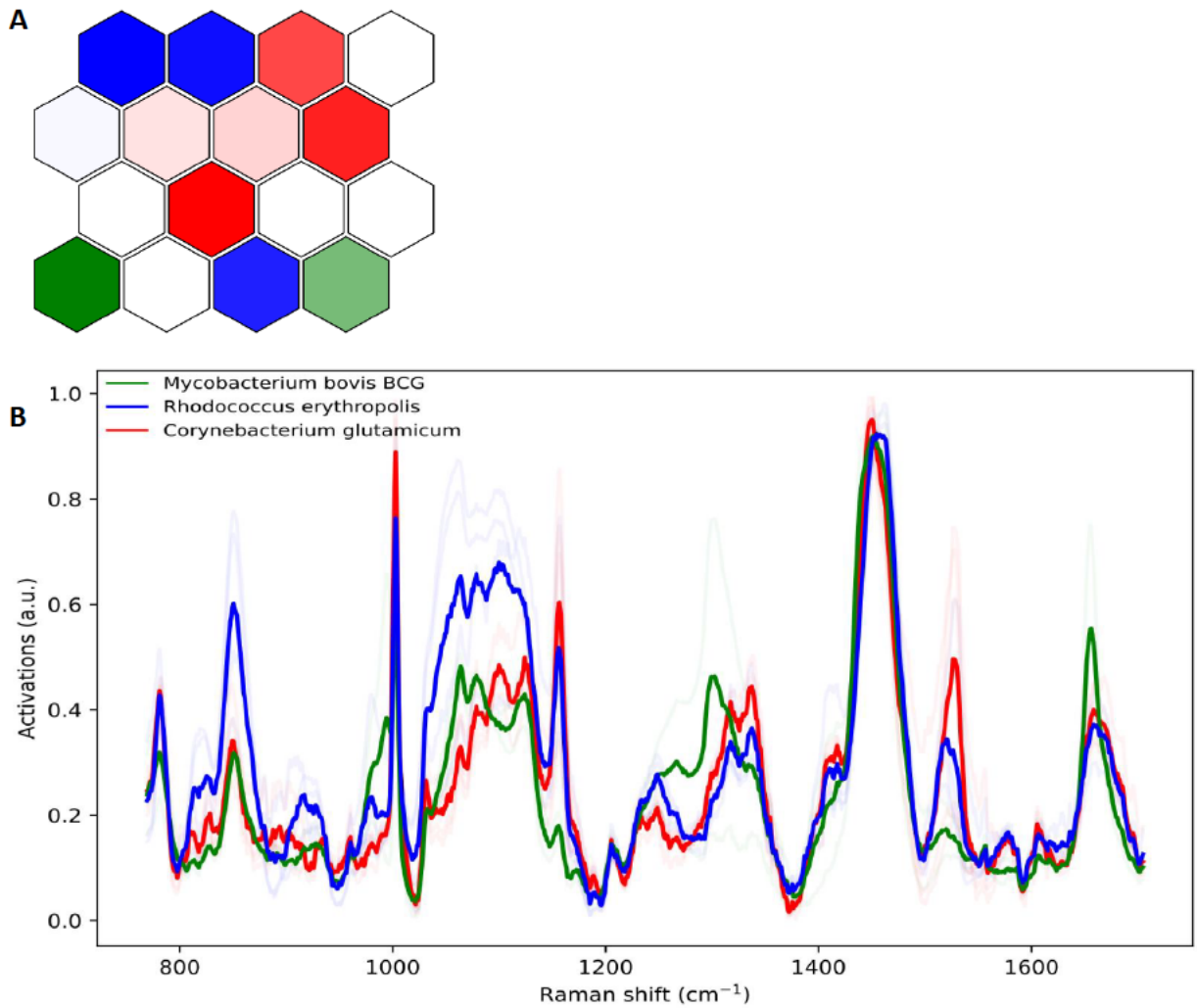


Figure 51 - (a) SOM trained to analyse the respective bacteria illustrating the clustering of the spectra. (b) SOMDI showing the peaks which have the greatest influence in activating the respective neurons in SOM.

Subsequently the most dominant spectral peaks for the Mycobacteriales were identified and statistically analysed. Boxplots comparing the peaks shown in **Figure 52** (opposite) show the main differences between the peaks for each bacteria studied. The 780cm^{-1} peak is found to be similar for the *C. glutamicum* and *R. erythropolis* and significantly different for *M. bovis* BCG. This peak corresponds to an uracil containing ring, with the difference caused by the varying levels of uracil utilisation between the cells. Similar differences are also observed for 1155, 1450, 1525, and 1610cm^{-1} peaks. Highly significant peaks identified among all samples including the 1000, 1575 and 1660cm^{-1} are characteristic of Mycobacteriales. On the other hand, the peak at 1210 was found to be not statically significant. Peak differentiation can be achieved, mainly *via* the 850cm^{-1} band, which has been found to be statistically significant for the *R. erythropolis* (blue) but not for the *M. bovis* BCG (green) and *C. glutamicum* (red). We can, therefore, accurately identify the various bacterial species within the mixture. Specifically, *R. erythropolis* (blue) *via* the prominent intensity of the 850 cm^{-1} peak, associated with tyrosine, *C. glutamicum* (red) from the *M. bovis* BCG (green) *via* the

780, 1000, 1155, 1525 and 1610 cm^{-1} peaks, attributed to the uracil-based breathing ring, phenylalanine, carotenoids, in-plane vibrations of conjugated -C=C- and cytosine, respectively (**Table 10**, opposite). Additionally, by generating the box plot and holding the spectra to closer scrutiny, one can observe that the aforementioned peaks at *ca.* 1210 cm^{-1} are not statistically significant in helping to identify the Raman spectra of the bacteria.

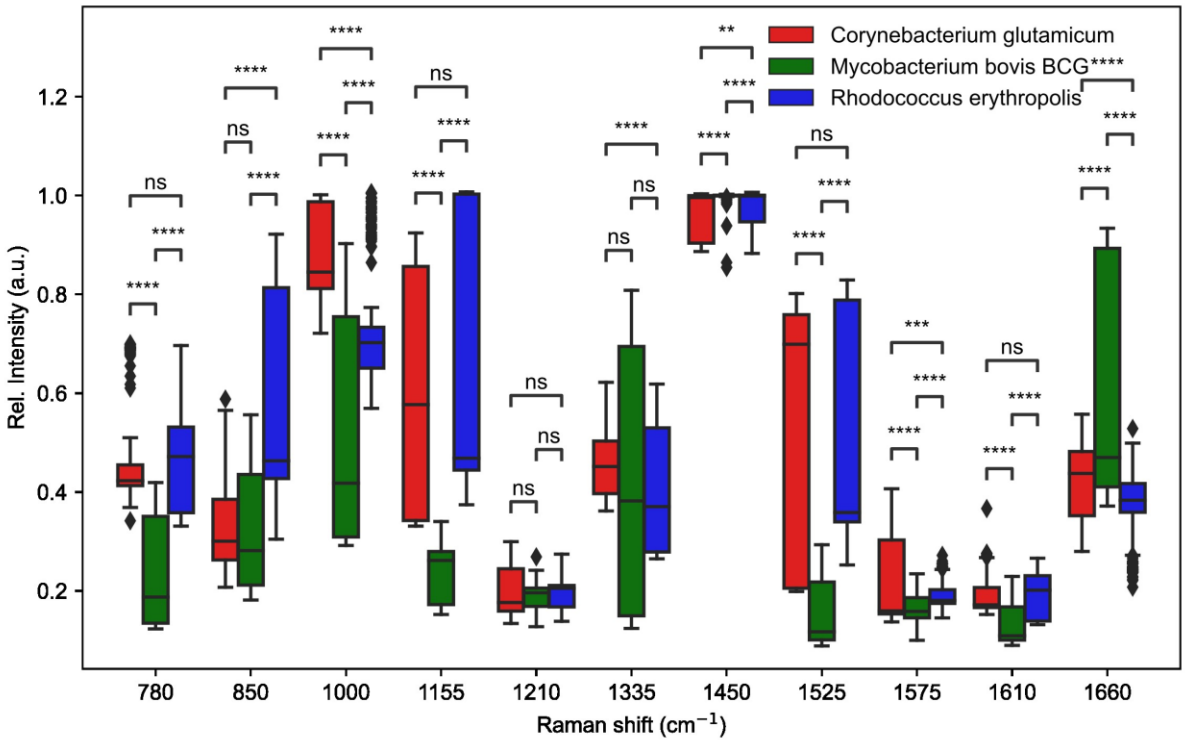


Figure 52 - Box plots comparing the dominant Raman spectral peaks of the studied bacterial species (NS = not significant, * = p<0.05, ** = p<0.01, ***= p<0.001, ****= p<0.0001).

Raman Peak (cm^{-1})	Functional Group / Band Assignment	Prominent In:		
		C. Glutamicum	*R. Erythropolis*	*M. Bovis BCG*
780	Uracil-based breathing ring	X	X	
850	Tyrosine		X	
1000	Phenylalanine	X	X	X
1155	Carotenoids	X	X	
1210	Stretching in tyrosine and Phenylalanine	X	X	X
1450	CH_2 bending	X	X	
1525	In-plane vibrations of conjugated -C=C-	X	X	
1575	DNA bases	X	X	X
1610	Cytosine	X	X	
1660	Lipids	X	X	X

Table 10 - Summary of the prominent peaks with the greatest statistical significance determined in **Figure** 54. Peaks attributions are taken from Movasaghi *et al.* (288) & Talari *et al. (250)*

With the dataset pre-processed by PCA, the LDA data was subject to cross-validation, with 4 folds used in analysis. The mean accuracy ($\pm 1\sigma$) was $99.11 \pm 0.63\%$, showing

excellent discrimination between the different species (**Figure 543**, below), where each species is clustered significantly apart from the rest with the overall detailed analysis partitioning the PCA-processed dataset into 40% test data and 60% training.

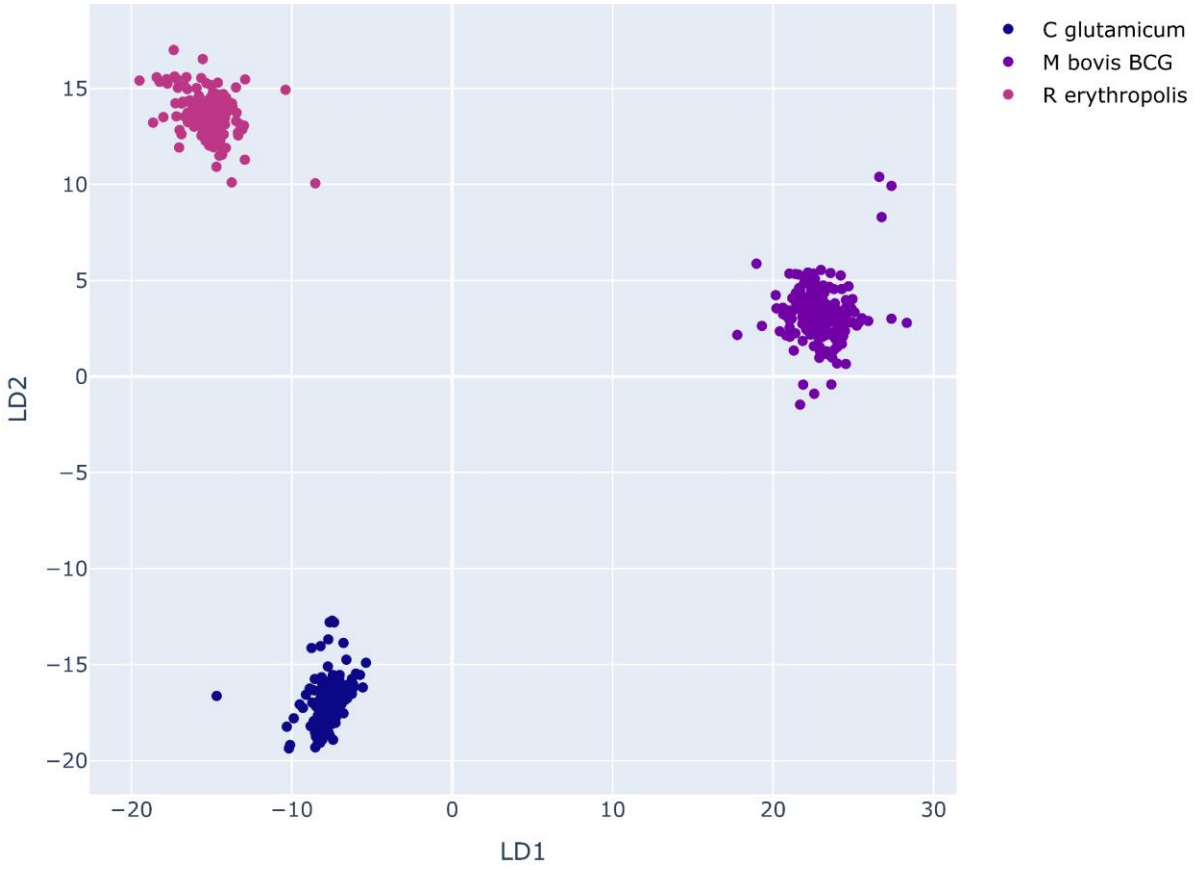


Figure 53 - LDA Plot of the Mycobacteriales, displaying vast separation between the three bacterial species.

The derived confusion matrix (**Figure 54** and **Table 11**, below) shows that on average the PCA-LDA approach is highly effective at discerning between different bacterium species. The off-diagonal values represent the incorrect identification occurrence, where on average each of these values was below 1. This means that in the worst performance case scenario, the PCA-LDA approach will misidentify two samples from a pool of three bacteria.

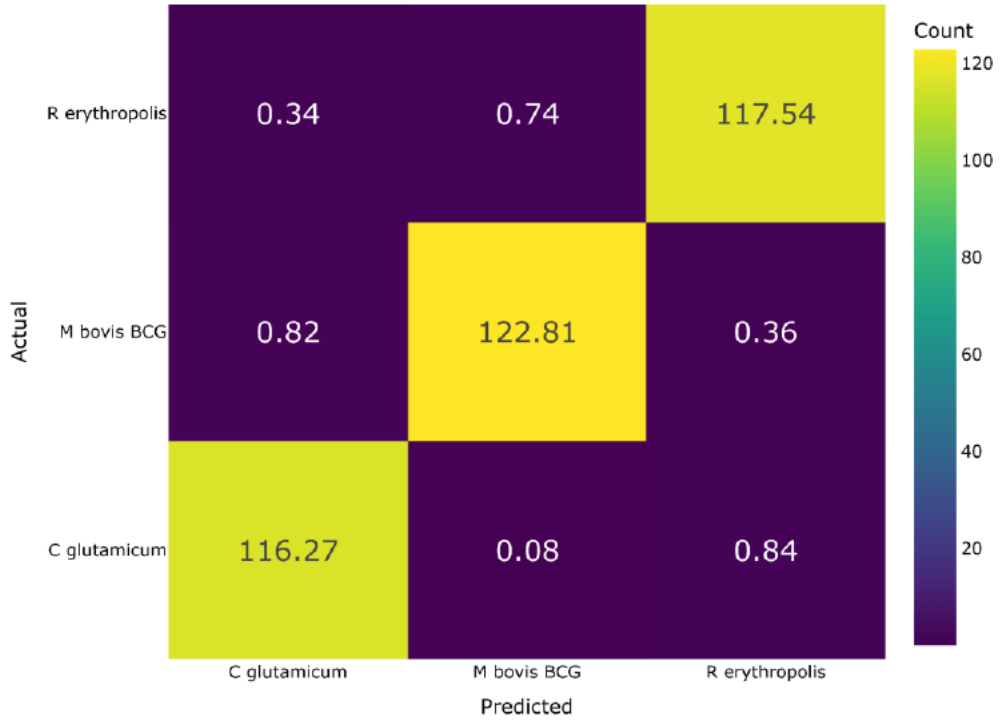


Figure 54 - Confusion matrix following the 60:40 training and test data split during the PCA-LDA. (c). Characteristic derived spectral barcodes of the C. glutamicum, M. bovis BCG and R. erythropolis after the application of a Savitzky-Golan filter and reduction of the noise from the double differentiation.

	Precision	**Recall**	**F1-score**	**Support**
C. Glutamicum	0.987	0.993	0.990	121
M. Bovis (BCG)	0.994	0.992	0.993	118
R. Erythropolis	0.993	0.988	0.990	121
Accuracy			0.991	360
Macro average	0.991	0.991	0.991	360
Weighted average	0.991	0.991	0.991	360

Table 11 - Mean confusion matrix generated from the PCA-LDA. The values are rounded to the nearest integer.

The obtained spectra were further classified to derive the characteristic fingerprint spectral barcode for each bacterium derived from the second derivatives, after application of a suitable threshold to discriminate between signal and noise (**Figure 55,** opposite). Initially, a Savitzky-Golay filter was used to smooth the signal and reduce the noise from double differentiation and subsequently, a grid approach was used to select the window size and polynomial order with optimal values for these were identified as 5 and 2, respectively. Following filtering, the second derivatives were

assigned binary values, +1 if their absolute value was greater than or equal to 5% of the maximum value for each species, and 0 otherwise. In our case, this threshold, building upon the work in Velioglu *et al.* (289), was increased to 5% since the 1% value applied in Shivakoti *et al.* (35) unnecessarily included noise. If the value was below the threshold, it was assigned 0. Bars were plotted for each wavenumber and their position on the Plotly "Greys" colourmap determined the colour of each bar, where the greater the count, the darker was the bar.

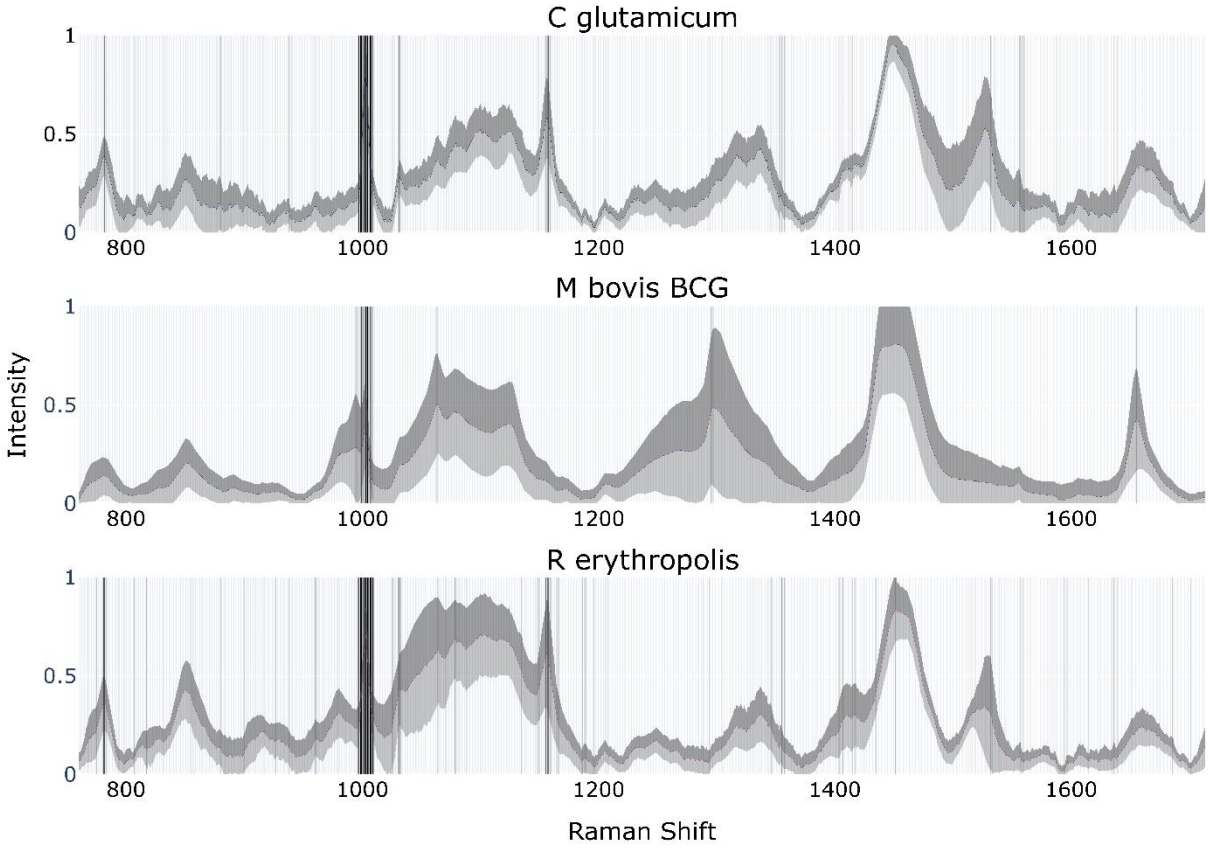


Figure 55 - Characteristic derived spectral barcodes of the *C. glutamicum*, *M. bovis* BCG and *R. erythropolis* after the application of a Savitzky-Golay filter and reduction of the noise from the double differentiation.

These were subsequently, overlaid over the averaged spectra for each species, with a standard deviation value being shaded. A common feature for all three species was

a strong second derivative significant value at *ca*. 1000 cm^{-1}, also identified by the PCA and SOM, attributed to the phenylalanine peak (**Table 11**, below).

Raman Peak (cm^{-1})	Functional Group / Band Assignment	Prominent In:		
		C. Glutamicum	***R. Erythropolis***	***M. Bovis BCG***
780	Uracil-based breathing ring	X	X	
850	Tyrosine		X	
1000	Phenylalanine	X	X	X
1155	Carotenoids	X	X	
1210	C-C_6H_5 Stretching in tyrosine and Phenylalanine	X	X	X
1450	CH_2 bending	X	X	
1525	In-plane vibrations of conjugated -C=C-	X	X	
1575	DNA bases	X	X	X
1610	Cytosine	X	X	
1660	Lipids	X	X	X

Table 12 - Summary of the prominent peaks with the greatest statistical significance determined in **Figure** 54. Peaks attributions are taken from Movasaghi *et al.* (288) & (250)

Interestingly, *M. bovis* BCG lacks many features shared by the other two species based on the derived barcodes. This was observed previously *via* PCA where *M. Bovis* (BCG) had a different sign for PC1 and *via* the PCA-LDA, where its sign for LD1 contrasted strongly to the others. This indicates that *M. bovis* BCG has a substantially different spectral fingerprint to the other two species analysed, regardless of the method utilised. Discerning between *C. glutamicum* and *R. erythropolis* samples *via* the barcode approach would rely on the lighter shaded barcodes, which correspond to the characteristic traits of each bacterium and their respective spectra.

Overall, standard spectral classification *via* the artificial neural network SKiNET enables high levels of discrimination when analysing multivariate spectral data

generated by analysis of bacteria. In contrast, standard PCA is found to exhibit low levels of accuracy. However, this accuracy can be further enhanced using a PCA quadrant analysis as well as LDA, which corroborates the deductions generated from SOM analysis and SKiNET. Further enhancing the possibility to identify and characterise bacteria is the ability to discern distinct elements of the Raman spectra generated as a result of statistical peak analysis, which affords insight into which components of the spectra yield the greatest significance in terms of determining which peaks have the greatest influence in determining and differentiating the exact organism.

5.4. Discussion:

The combined use of the above techniques allows one to investigate and dissect the often differences in Raman spectra which arise as a result of investigating biological organisms. Within this chapter several bacteria have been analysed with Raman spectroscopy for the very first time. However, for the sake of brevity, a focus will be placed on the Mycobacteriales.

Considering the initial PCAs, one can observe that there is a great deal of spectral crossover between the majority of the bacteria being analysed, with *M. bovis* BCG, *M. catarrhalis* and *R. erythropolis* acting as outliers. However, following two additional repeats and focusing further on the Mycobacteriales, a greater amount of crossover can be observed and this could be due to the inherent variability present within biological organisms, as has previously been discussed (290).

By using a PCA quadrant analysis, once more, the bacteria being analysed in this research can be held to more scrutiny and we can begin to examine and understand the unique characteristics of each bacterial species.

Meanwhile, deployment of SKiNET allows for an easily recognisable output that is easier to interpret with a cursory glance. The use of the box plots helps researchers understand the important components of a Raman analysis and to understand which spectral elements are the most important when trying to identify bacteria.

As for the generation of the PCA-LDA, by undertaking this, a highly discriminatory algorithm can be deployed and discern spectral elements apart to a much higher degree than the human eye and that of PCA, as evidenced by the confusion matrix.

Perhaps most critically though, the generation of a spectral barcode can afford information that could contribute towards a spectral library that, one day, could very well be used to enable Raman spectroscopy to be used for the diagnosis of the pathogenic infections. More specifically, this could involve the use of a cloud-based storage system where the Raman spectra of bacteria are readily uploaded in a fashion similar to the sequencing data available *via* BLAST and BV-BRC. Once acquired, a Raman analysis could take place and check if the spectra matches with other bacterial species available, thus allowing a rapid diagnosis of infectious diseases.

The statistical techniques in this chapter are highly valuable to the field of chemometrics and, through the deployment of the right Python script, highly advanced statistical tests can be applied to highly complex multivariate data. In

doing this, we have established the ability to discriminate against various bacterial species, specifically those which are closely related and this is a consideration which has not been previously reported in literature, outside of a genus level. Ultimately, these are prized tools which can enable scientists to undertake a vast amount of analysis without having to be a specialised data scientist.

Of course, the data acquired within this chapter does come with limitations. The non-Mycobacteriale data acquired in this chapter did not utilise any repeats due to time constraints and this is something which should be repeated in the future to provide more comprehensive data. As demonstrated by the PCA in **Figure 49,** there is a notable amount of variation between samples of even the same species and, therefore, it may be the case that differences are only observed following further repetitions. More specifically, special consideration should be given to the use of SERS. However, of a particular note is the persistent Raman peak at *ca.* 1210 cm^{-1} .

Through statistical analysis, it can be established that this peak in particular is not significant for the determination and discrimination of the Raman spectra unique to *M. bovis* BCG within the presence of *R. erythropolis* and *C. glutamicum*. Based on the data acquired from Talari *et al.* (250), the peak assignment is believed to be due to the presence of C-C_6H_5 stretching in tyrosine and phenylalanine, however these compounds are unlikely to be present when considering the lipid analysis and most certainly would not be present when considering the *in silico* data, therefore the presence of these peaks may not be solely due to these compounds.

Whilst the statistical significance of this peak has been called into question due to the box plot analysis, its presence suggests that it is a legitimate peak, but the reason for this presence is not due to C-C_6H_5 stretching in tyrosine and phenylalanine and must be due to some other compound which has not yet been characterised by Raman spectroscopy.

Future research should consider repeating the Raman analysis of the other bacterial species within this chapter. For the sake of brevity, a focus was placed on characterising the Mycobacteriales, however, most of the organisms considered in this chapter have not been fully characterised using Raman spectroscopy. Additionally, whilst a large number of tools have been presented within this chapter, it would be superfluous to deploy complex data processing tools on data that is rather limited in nature, especially if one were to expect significant results.

Perhaps of the greatest consideration is the *Mycoplasma sp.* which have received a greater amount of interest as of late due to their proclivity to infect cell cultures, as well as their propensity to cause disease which is often difficult to fully treat due to the physiology of *Mycoplasma sp*.

Additionally, one should consider expanding the species of organisms analysed and could deploy the techniques within this chapter on bacteria which are more clinically significant. One could also attempt to use these techniques to discriminate against different serovars, with special consideration being given to the use of SERS, though this also presents with its own limitations, as previously covered. All of which would

acquire a great amount of useful information and could help to consolidate the use of Raman as a diagnostic tool, not only for tuberculosis but also for other diseases.

In the next chapter, the information and some of the techniques acquired from this chapter and the ones preceding it shall be used to establish if it is possible to identify *M. bovis* BCG within an artificial “sample” of sputum which contains other bacteria.

Chapter 6:

6. Detection of *Mycobacterium bovis* BCG in an Artificial Sputum

6.1. Introduction:

As per Chapter 5, one can see that the microbiome is made up of multitudes of bacteria and this is often in flux. Often the populations of respective bacteria within the microbiome can change. Sometimes this can lead to disease, but other times this is simply a result of a change in diet, stress, or various other factors (277), and this is also true for humans as much as it is for animals (291)

The information acquired in Chapter 6 is without use unless one begins to investigate the potential of other bacteria to interfere with the Raman spectra of that generated by *M. bovis* BCG, given that this study has been using this as a model organism.

To further investigate and characterise the bacteria within the microbiome and to establish whether Raman spectroscopy could detect TB or other members of the MTBC, it becomes necessary to create an artificial matrix of bacteria which is representative of a theoretical sample of saliva as well as an artificial version of the saliva itself. In essence, an artificial sample.

Naturally, one could attempt to acquire field samples, however, the acquisition of which was not proposed at this project's outset and, therefore, ethical approval would have been unlikely to have been acquired in time.

Additionally, whilst the acquisition of field samples is an obstacle that future researchers will have to overcome at some point, field samples cannot be controlled for (*e.g.* the ratio of bacteria in each sample will vary considerably). Therefore, the use of an artificial sample is more optimal for these initial experiments.

To this end, it therefore stands to reason to utilise an artificial sputum and to test the limit-of-detection of *M. bovis* BCG within an artificial sample that contains known concentrations of different bacteria, as well as varying amounts of *M. bovis* BCG.

Unlike some bacteria, such as *S. aureus* or *E. coli*, *M. bovis* (and other MTBC bacteria) are obligate pathogens, as previously mentioned. That is, if *M. bovis* is present within a sample, then it is causing disease, unlike the aforementioned organisms which form part of the microbiome of both humans and animals (292, 293, 294, 295).

For this section, aliquots of an artificial sputum were prepared with the bacteria aforementioned in chapter 6. These aliquots were then “spiked” with varying amounts of *M. bovis* BCG in a bid to establish what the limit-of-detection would be for this methodology and an artificial sample.

Critically, this chapter shall attempt to establish the feasibility of Raman spectroscopy to identify *M. bovis* BCG within a scenario more reflective of the real-world and influence future work which will influence future work with the more relevant (albeit more dangerous) real-world counterpart, *M. bovis*, as well as other MTBC bacteria.

6.2. Methods:

6.2.1. Reagents:

An artificial sputum was prepared as per Kloß *et al.* 2015 (296). The artificial sputum was prepared as follows: 400 mg DNA from fish sperm (Sigma Life Science) and 3 g mucin from porcine stomach (type II) (Sigma Life Science) were each suspended in 25 mL deionized water and dissolved in a rotation mixer over night at 4 °C. One gram bovine serum albumin (BSA) (VWR Prolabo) was dissolved in 10 mL deionized water. Five hundred milligram sodium chloride (NaCl), 225 mg potassium chloride (KCl), and 0.6 mg diethylenetriaminepentaactetic acid (DTPA) (Merck) were also dissolved in 10 mL deionized water. All solutions were combined, and 5 mL egg yolk emulsion (VWR Prolabo) was added. The solution was filled up to 85 mL with deionized water, then the pH was adjusted to 7.0 with 1 M tris(hydroxymethyl)aminomethane hydrochloride (TRIS × HCl pH 8), and finally, the solution was filled up to 100 mL with deionized water and was vortexed. The artificial sputum was syringe filtered *via* a 0.2 µm syringe filter into 1ml aliquots then stored at −20 °C until use.

Bacteria were prepared as per Chapter 2.2.1. Liquid cultures were centrifuged for 5 mins at 3900 RPM then resuspended in 1ml of sterile artificial sputum, with the exception of *M. bovis* BCG which was centrifuged at 3900 RPM for 10 mins. Optical Density was then measured at 600nm (see **A.9 Optical Density of bacteria present in bacterial mix**:) and bacterial density was calculated per ml (297). Aliquots were then made up to contain a total of 1×10^6 CFU/ml of bacteria, with the proportions dictated by the relevant literature (281, 282, 283). *Mycoplasma sp.* in this

section were diluted in accordance with Ramos *et al.* (298), due to a lack of literature on *Mycoplasma bovis* and *Mycoplasma hyosynoviae.*

Serial dilutions of *M. bovis* BCG were made up from 10^6 to 10^1 CFU/ml and inoculated into the artificial sputum. Samples containing only serial dilutions of *M. bovis* BCG and of the other bacteria with *M. bovis* BCG were investigated.

A sample was removed from the aliquot with a 10 µl inoculating loop, and this was then spread on sterile aluminium foil and allowed to airdry for 2 hours under laminar flow.

6.2.2. Raman Spectroscopy:

Spectral acquisitions were performed using an InVia confocal Raman (Renishaw). The spectrometer was calibrated prior to each use with silicon (520.7cm^{-1}). A 100x Leica objective and a 1200l/mm grating were used for all measurements in the range of 750-1750cm^{-1}. Raman data, consisting of 3 accumulations and 10 second exposure time, was acquired using a 785nm excitation laser with laser power of 10-14mW. Wire 5.1 (Renishaw Plc) software was used for baseline subtraction and cosmic ray removal, with the results visualised as below (267). 3 technical repeats were carried out per sample.

6.2.3. Principal Component Analysis:

Principal component analysis (PCA) was used to interpret the *minute* differences in spectra between the bacterial mixes containing varying amounts of *M. bovis* BCG and bacteria as well as the samples containing varying amounts of *M. bovis* BCG only. As a multivariate data analysis technique PCA was able to reduce the dimensionality of

the dataset into the most relevant components to maximize the variance among different samples by projecting the data into a space with N orthogonal basis vectors, arranged in descending order with respect to variance. Each axis, or principal component, represented a fraction of the total data variance. The dimensionality reduction was achieved by selecting the number of principal components that explains at least the required fraction of the dataset variance. PCA analysis was used to determine how the separated clusters separated between different species and to identify what they had in common. PCA loadings were calculated using the eigenvectors and eigenvalues obtained from matrix operations to find the PC. The loadings contained information specific to the initial Raman data and differed depending on the PC space. In a 2D PCA system, the loadings of the different PC quadrants were expressed as PC1 > 0 and PC2 > 0, PC1 0 and PC2, and PC1 > 0 and PC2 0. PCA analysis was deployed locally using a customised Python script (267)

6.2.4. Code Availability:

The customised written Python algorithm can be downloaded from Vettigli *et al.* (218).

6.3. Results:

6.3.1. Average spectra:

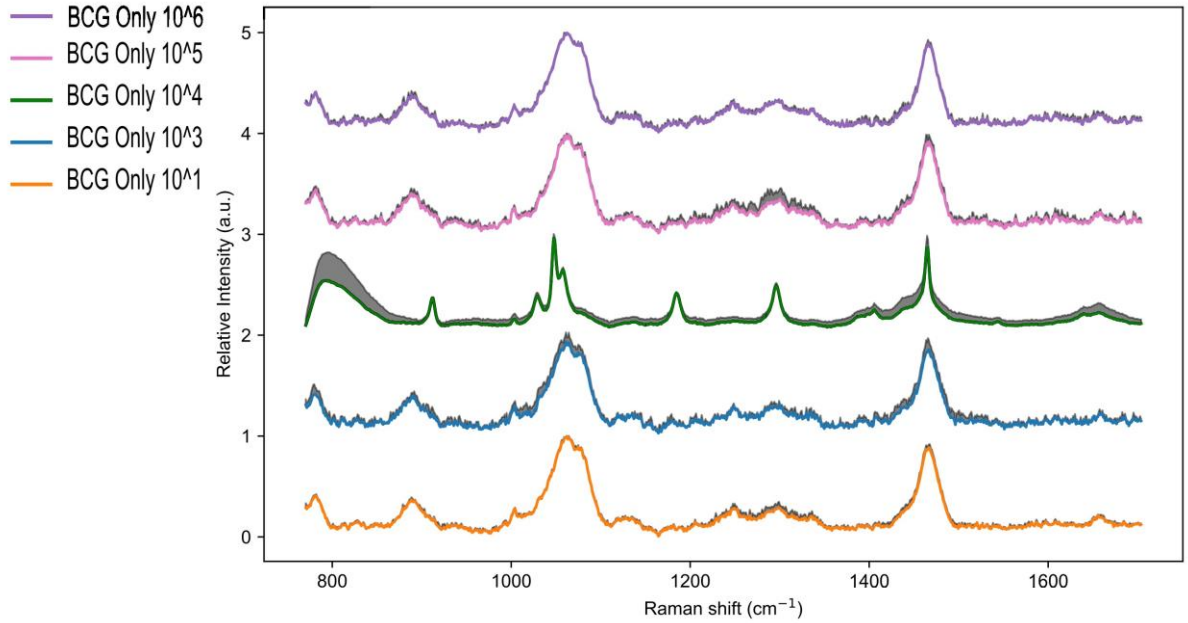


Figure 56 - Combined average Raman spectra of *Mycobacterium bovis* BCG at various concentrations of 10^X CFUs/ml. The coloured line represents the average spectra. The greyed lines represent the standard variation in each biological repeat.

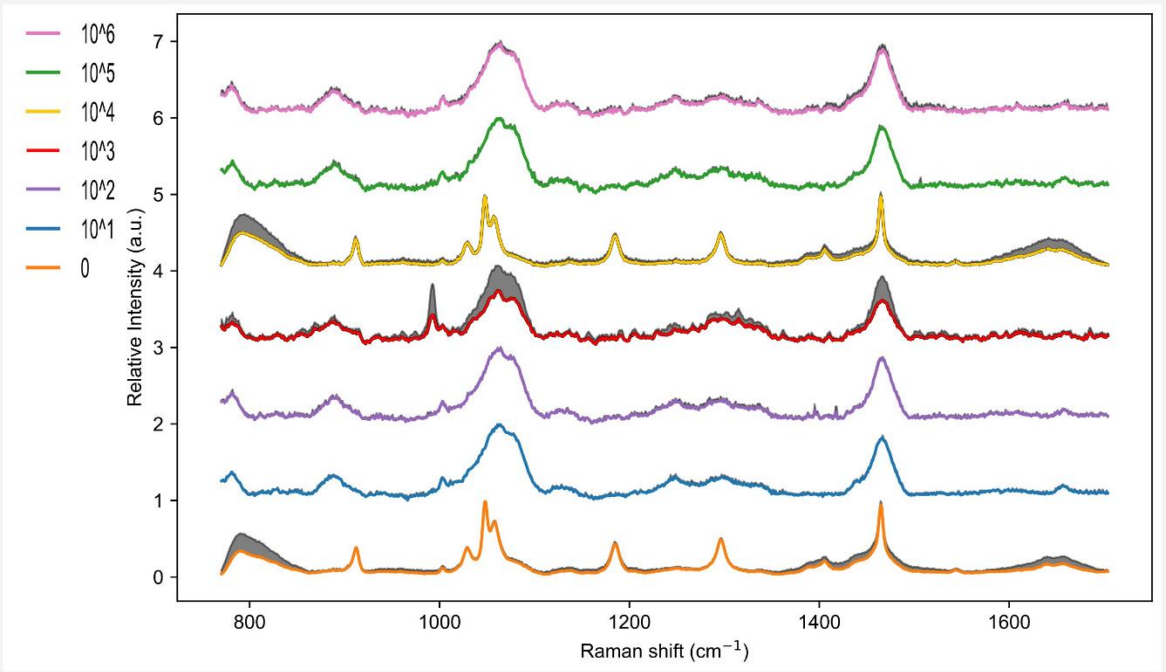


Figure 57 - Combined average Raman spectra of Mycobacterium bovis BCG at various concentrations of 10^X CFUs/ml and the standardised mix of bacteria. The coloured line represents the average spectra. The greyed lines represent the standard variation in each biological repeat.

Upon a superficial glance the average spectra (**Figure 56** & **Figure 57**, above) of the bacterial mix broadly overlaps, though there are exceptions to this. Most notably, the readings from 0 CFU/ml, 10^4 CFU/mL (BCG Only), and 10^3 CFU/ml. No data was acquired for 10^2 CFU/ml (BCG Only) due to disruption of the sample and time constraining attempts to repeat the sample.

Of note is a particularly intense peak which is omnipresent across all samples at *ca*. 1450 cm^{-1} and 1155 cm^{-1} conferring the presence of CH_2 bending and carotenoids, respectively. For the more anomalous samples, peaks are observed at 780, 900, 1057, 1095, 1195, and 1300 cm^{-1}. Most notably, there is also a somewhat unique peak present on the spectra for 10^3 CFU/mL at 995 cm^{-1}.

6.3.2 PCA of spectra:

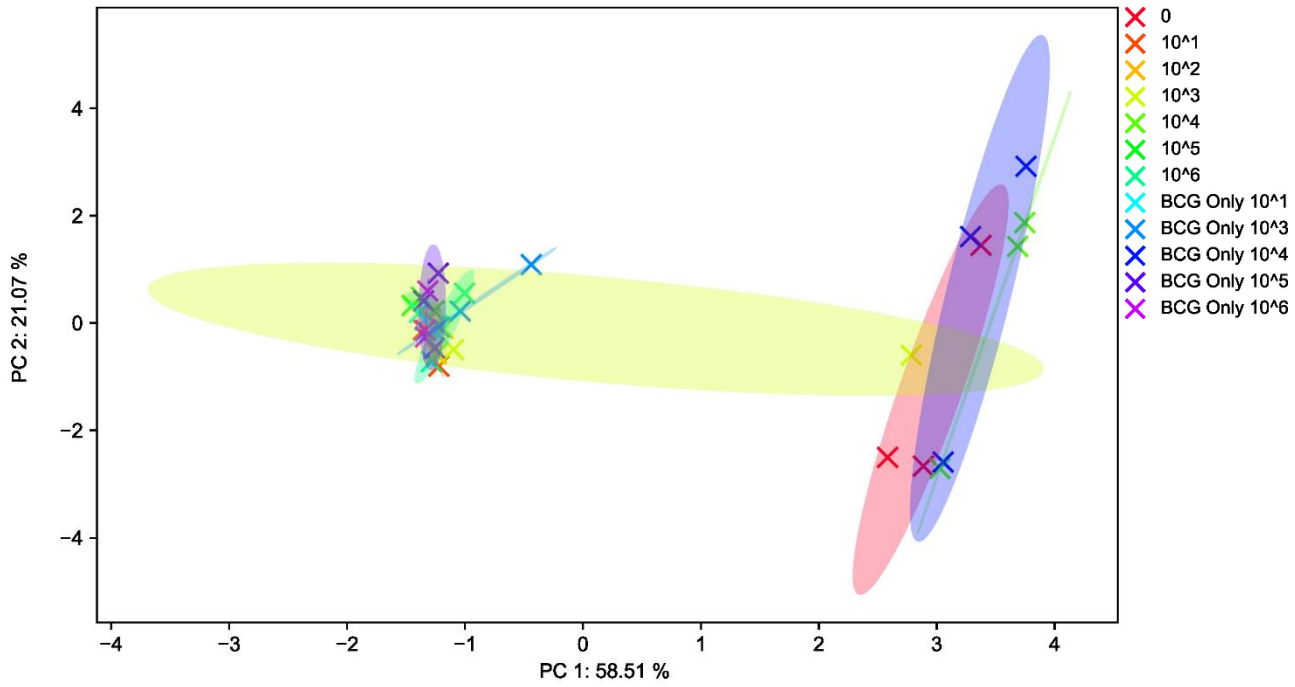


Figure 58 - PCA of the bacterial suspensions utilising two principal components following three technical repeats of the numerous suspensions utilised in this project. 3x technical repeats were used in each technical repeat leading to N=3 spectra per suspension. Note the clustering of each bacterial suspension and the crossover of the Raman spectra.

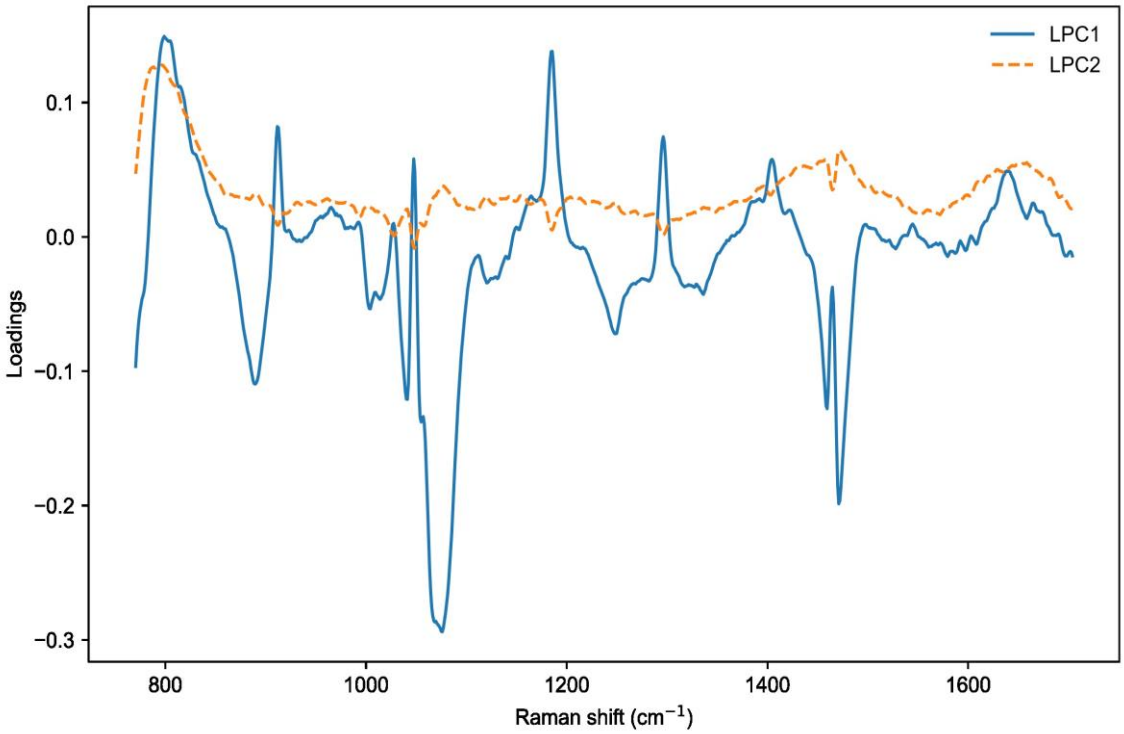


Figure 59 – Raman loadings for the PCA of the bacterial mixes demonstrating the most influential parts of the spectra in the respective principal component (LPC1, LPC2).

Upon examination of the PCA (**Figure 58**, page 113) one can observe that there is indeed a great amount of crossover of the spectra of the various bacteria in this study. The PCA is able to differentiate some of the bacteria, however, this is in line with the more anomalous spectra observed when considering the average spectra. The PCA is able to account for 79.58% of the differences between the spectra by using two principal components and this contrasts strongly with the monoculture work in the previous chapter where only 55% of the differences in spectra can be accounted for using two components for PCA.

When dissecting the loadings (**Figure 59**, above), the spectra at 800, 900, 1057, 1095, 1168, 1255, 1300, 1400, 1465, and 1606 are the most influential in

determining PC1 and are also influential in determining PC2. The aforementioned Raman peaks can therefore be summarised as such:

Raman Peak (cm^{-1})	Functional Group / Band Assignment
800	$CH_{2,6}$ out-of-plane bending, observed in the spectra of a single human erythrocyte
	Uracil-based ring breathing mode
900	Lipid
1057	Lipid
1095	Lipid
1168	Lipids, ν(C=C) δ(COH) (lipid assignment), ν(C-C), carotenoids
1255	Lipids
1300	Fatty acids, $\delta(CH_2)$: Lipids, CH_2 twisting (lipids)
1400	(-1430) $\nu(C=O)O^-$
1465	Lipids
1606	C=C Bending

Table 13 - Peak assignment for the most prominent peaks within the Raman spectra generated in this chapter

6.4. Discussion:

Bacteria are complex organisms and, as has been previously discussed, their Raman spectra can and does indeed vary based on numerous factors (299). Unfortunately, the research contained in this chapter was unable to determine and readily differentiate between samples containing strictly *M. bovis* BCG and samples containing *M. bovis* with other samples.

This chapter is based solely off of one repeat of one experiment and uses tools which are not completely adequate for the use of the discrimination of multivariate data (267). As discussed in Chapter 6, there are more efficient tools that can be used to pick apart spectral data. However, without the use of multiple repeats of the same experiment, their use is somewhat redundant as it is unlikely that they would

provide an accurate insight into the data acquired in this chapter and would certainly not be suitable to draw any viable conclusions which may influence future research. As Macgregor-Fairlie *et al.* (267) has illustrated, there is often a large amount of variation of Raman spectra even between the same samples, and this would be further complicated by Raman spectra where the sole defining characteristic is the level of enumeration of the respective bacteria. This complication is further exacerbated by the limited amount of literature that is currently available which seeks to expand the detection of bacteria *via* Raman spectroscopy beyond monoculture or single-cell detection.

Therefore, the data acquired here should be taken with a large pinch of salt, as further repeats could prove fruitful. Though it does offer some insight into the potential ramifications of using Raman spectroscopy without consideration being given to the other bacteria that may be present within a sample.

As mentioned, there is also anomalous data here. This data could arise due to a variety of reasons, however, one thing to note is that *M. bovis* BCG have a tendency to cluster in liquid media (300) and, therefore, their disruption and resuspension for use in a liquid suspension can be prone to great difficulty. Naturally, steps were taken to prevent this following centrifugation (Thorough mixing and resuspension), but it is possible that these steps would have completely prevented the clustering of the bacteria on a microscopic scale. A potential way to avoid this in the future would be to use flow cytometry (specifically, cell impedance flow cytometry) in a bid to create a pre-determined cell suspension (301).

Overall, it becomes difficult to draw any meaningful conclusions from this set of limited experiments, besides highlighting the crossover over the bacterial species used in this set of experiments, as well as key points within the Raman spectra.

Though what can be observed is consistencies between the Raman spectra of bacteria and how these are relatively ubiquitous, but, as displayed in chapter 6, it is this ubiquity which can enable researchers and scientists to successfully identify bacteria in a sample.

Moving forward, one should consider further repeats of the experiments conducted in this chapter and then hope to deploy the statistical analyses in chapter 6.

Chapter 7:

7. Discussions, Conclusions & Outlook

This thesis considered the use of CRISPR/Cas as a diagnostic tool for tuberculosis before it began to explore the current gap in knowledge that is present within the field of Raman diagnostics and the potential of interference from the microbiome, something which previous researchers have often neglected, in the hopes of this information being fed into the field and enabling future researchers to utilise it in a bid to develop a Raman based diagnostic tool.

Chapter 3 focused on the use of *in silico* Raman analysis of compounds associated with the cell wall surface of *M. tuberculosis* and other *Mycobacterium sp*. through the combined use of the quantum chemistry software package, ORCA, and the computational chemistry Graphic User Interface package, Avogadro *via* the University of Birmingham's supercomputer BlueBEAR. In doing so a novel and unassigned Raman peak was detected. This is the first time such work has been conducted.

Chapter 4 sought to utilise both optical trapping on both polar and non-polar cell wall lipids extracted from *M. smegmatis, M. tuberculosis* & *M. bovis* BCG to identify novel Raman spectra which arise strictly due to the lipid rich cell wall. This is the first time such work has been conducted. During the conduct of this work, several novel motifs were identified that have not been discussed in literature. Specifically, this includes a lipid-rich area of the spectra at approximately 1210 cm^{-1}, the presence of which was also conferred by the work undertaken in Chapter 5.

Chapter 5 then used Raman spectroscopy to analyse several bacterial genera which are associated with the bovine microbiome in a bid to establish an understanding of the Raman spectra of said bacteria and foster research which could be used to deploy Raman spectroscopy for the detection of bovine TB and, ultimately, human TB.

The research contained in this chapter also sought to fill the current gap in knowledge where bacteria within the same family as *Mycobacterium sp*. have not been considered in the published literature.

Through the repeat analysis of two other Mycobacteriales, *R. erythropolis & C. glutamicum,* a multivariate approach which employed the use of a Python script that incorporated tools for statistical work was used and the ability to actively discriminate the three organisms from one another was established. This work also sought to establish a spectral barcode for each organism, with the hope that it be repeated on other organisms in the future and uploaded on to a cloud so that it can be used globally.

Chapter 6 then sought to compile an artificial matrix of bacteria which was proposed to resemble a “real-world” use to a greater extent than analysing the bacteria in isolation. The preliminary results presented in this chapter were inconclusive in nature. However, valuable information can be gleamed from an examination of the Raman spectra presented. This work could be improved by undertaking the suggestions mentioned at the end of the chapter. Specifically, this could be

investigated more thoroughly simply by culturing the bacteria once more and repeating the experiments to increase the repeats.

Chapter A.1 considered the use of CRISPR/Cas as a diagnostic tool, however, shortly after this work commenced, similar research was published in Ai *et al.* (133) in which *IS6110* was targeted by CRISPR/Cas gRNA. The work produced in this chapter sought to consider other genetic elements for detection, however, this was interrupted by the lab shutdown prompted by the UK government's COVID-19 response and subsequent work was completely abandoned due to the University of Birmingham's COVID-19 mitigation strategy which prevented lab access across the multiple laboratories required to undertake this particular work.

The work undertaken in this thesis seeks to fill a gap that had previously been identified in research: A lack of consideration afforded to the bacteria inhabiting the microbiome. This is and remains a critical consideration that scientists must make if there is hope for the use of Raman spectroscopy as a diagnostic tool for tuberculosis and other infectious diseases. By undertaking the work in this thesis, it is hoped that it will help contribute towards changing the direction of the conversation surrounding Raman diagnostics. Additionally, it is hoped that the published work (as well as the work soon-to-be published) will help proliferate a greater variety of tools that scientists can use to help process multivariate data such as that produced during Raman analysis.

Moving forward, future research should consider investigation into other organisms, as well as undertaking repeats of the work presented in this thesis. To summarise,

the work undertaken in **Chapter A.1.** could be expanded upon to consolidate the gRNAs utilised to detect *M. tuberculosis* and the MTBC as well as SNPs conferring drug resistance and other drug-resistance mechanisms.

The work undertaken in **Chapter 3** could expand to include cellular components which are associated with other *Mycobacterium sp*. as well as mycolic acids associated with different bacteria. However, this work would rest solely on the isolation of said compounds by other researchers. Unless one were to build the structure of these compounds manually in Avogadro and then optimise their geometry. Other methods of *in silico* analysis could be undertaken, such as IR and LC-MS.

To expand upon the work covered in **Chapter 4**, one could seek to isolate the cell wall lipids of other bacterial species and consider other analytical techniques which could be used within this context such as LC-MS and IR spectroscopy. This would add to our knowledge of the compounds present on the cell wall further and allows researchers to consolidate our knowledge of the isomers of the lipids present. SERS is another potential consideration, however, efforts must be made to understand the damaging effects of the materials over time, or use of different materials/methods must be considered.

With regards for **Chapter 5**, one should seek to investigate the other bacteria utilised in this chapter to a larger degree and undertake repeats of those where no repeat was undertaken. Additionally, if a higher containment laboratory were to be available, the bacteria used in this section could be discarded and other pathogenic bacteria

could be considered, such as *M. bovis* or drug-resistant variants of *S. aureus.* Additionally, the use of a higher containment laboratory would enable an investigation into bacteria more relevant to human health.

In **Chapter 6**, work was undertaken to see if a limit-of-detection could be established within a mixture of bacteria and the results generated within this chapter are inconclusive. Future work should undertake repeats of this work in a bid to effectively deploy the methods used in **Chapter 5**.

Overall, the work in this thesis attempts to illustrate and construct a cohesive body of work which seeks to enable a better understanding of the Raman spectra of *M. tuberculosis* and *M. bovis* in a bid to enable future work which would seek to utilise a portable Raman spectrometer as a tool which can be deployed to more rural areas of low- and middle-income countries. The work in this thesis takes a logical workflow by analysing *in silico* compounds, isolated cell wall compounds, then whole-cell bacteria in isolation before attempting to understand the Raman spectra of a mixture of bacteria, which is more reflective of a real-world scenario.

In doing this, it is hoped that such a tool would enable quicker and more effective diagnosis of tuberculosis from oral swabs, especially if it were to be combined with the work undertaken by Kawaseeko *et al.* (148), where a combined blood test (as in the aforementioned paper) and confirmatory sputum test (which this thesis hopes to contribute towards) would enable such a diagnostic tool to be used.

As previously stated, past research has never considered other bacteria in the microbiome, and the body of work presented in this thesis attempts to rectify this; by

analysing genera of bacteria that inhabit the respiratory tract of cattle, a model can be established which would enable a similar body of knowledge to be built on human tuberculosis. Additionally, this body of work could be utilised to combact bovine TB, a disease which causes untold economic hardship to a great number of farmers.

By adding in a spectral barcode library of various bacteria, as proposed in this thesis, it also becomes possible to accumulate a vast amount of knowledge which would enable healthcare facilities to form a network, which would foster the freeflow of spectral information related to bacteria. Specifically, such a network could be updated in real-time, allowing for drug-resistant variants to be recognised quickly.

Raman spectroscopy itself is not without its limitations though. At the time of writing, no diagnostic device that is Raman-based has been approved for deployment into a healthcare setting. Additionally, any device would have to be thoroughly investigated and characterised before it were to be used for diagnostic purposes, specifically, impurities within sputum, blood, and other bodily fluids would have to be analysed, as organic tissue is associated with autofluorescence (302). The presence of which would interfere with the effective diagnosis of disease.

However, Raman spectroscopy represents a very promising diagnostic approach that could overstep a great number of the obstacles in the way of effectively deploying current technologies (namely, its speed and portability) and, with the advent of AI technologies as well as a shift to programming methods which can be readily disseminated to budding scientists, it may prove to be the case that this well-established technique is just about to enter its golden era.

Appendix:

A.1. Use of CRISPR-Cas for the detection of Tuberculosis:

N.B: Work undertaken for this chapter was later abandoned due to the COVID-19 pandemic which resulted in lab closures and continuation of the work was not feasible due to the University of Birmingham's social distancing measures. This chapter was originally Chapter 3 of this thesis and has been placed in the appendix at the request of the markers.

A.1.1 Introduction:

Clustered Regularly Interspaced Palindromic Repeats (or CRISPR for short) is a rapidly evolving tool that enables researchers to modify the genome of cells.

Initially discovered as part of a primitive immune system of *Streptococcus thermophilus* against invading bacteriophages, scientists have been able to harness the power of CRISPR and the associated Cas9 endonuclease for a variety of uses (303).

In effect, the CRISPR/Cas9 system works by storing a small piece of an invading bacteriophage's genome within the genome of *S. thermophilus*; the Cas9 enzyme then utilises this small segment of RNA as a guide and acts as a proof-reading enzyme; should it find a segment of complimentary DNA that has been inserted into the genome of *S. thermophilus* as a result of bacteriophage infection, it then cuts it out; this leads to the bacteria then being able to fend off phage infection and living another day.

However, one of the greatest strengths of the CRISPR/Cas9 system is that it is completely programmable. As a result, the price of genetically modifying a cell has crashed. Meaning that modifying the genome of cells has become readily accessible to anyone with a basic molecular laboratory.

More recently, researchers have continued to discover more Cas nucleases which are able to modify the genome of organisms in different ways; this includes targeting RNA as opposed to DNA and enabling researchers to modify the genome of cells at a nucleotide level.

The CRISPR system is an adaptive immune mechanism was first discovered in 2007 in *S. thermophilus (304) ,* though CRISPR motifs were unknowingly first discovered in *Escherichia coli* in 1987 (305). However, the development of the CRISPR system clearly wasn't something that started overnight.

The oceans are often considered the starting point of all life on earth and it's here where our battle begin. Each mL of water contains between $6.56x10^7$ and $1.59x10^6$ bacteriophages(306); these are viruses that infect bacteria.

Contrasting with this it is estimated that each mL of ocean water contains between $6.88x10^6$ and $1.59x10^5$ of prokaryotes; these act as the main hosts of bacteriophages and enable them (phages) to replicate. Of these bacteria, up to 50% are killed daily (307).

Of course, Darwinian selection favours bacteria which can successfully fend off phages which has led to the ubiquitous spread of the CRISPR system within approximately 85% of archaea and 45% of bacteria (308).

A.1.1.1 CRISPR-Cas:

Often when discussed in literature, “CRISPR” refers to the use of the CRISPR motif as well as the Cas9 endonuclease. However, the successful use of CRISPR is a little more complex than adding an artificial sequence of nucleotides, an enzyme and hoping for the best.

The CRISPR/Cas9 complex requires several different components before it can be effectively used in cells; this consists of:

- The Cas9 endonuclease
- The guide RNA (gRNA), though this requires several features in order to be effective in the Cas9 system:
    - A target RNA sequence (crRNA)
    - A Cas9 nuclease-recruiting sequence (tracrRNA)
    - Proto-Spacer Motif (PAM)
- A transfection method

For the sake of brevity, it is worth noting that there are many variables surrounding these core components, but they can be tweaked to increase the level of expression within cells. Additionally, the method used for transfection needs to be adapted

based upon the type of cells that are being worked with (*e.g.* Fungi, bacterial, mammalian, *etc.*).

A.1.1.1.1. gRNA:

Guide RNA (gRNA) is composed of several components. When being utilised for a Cas9 endonuclease reaction the gRNA consists a crRNA of approximately 20 nucleotides and a tracrRNA sequence. There is also the requirement of a PAM sequence that matches the 5'-N[6]GG motif within *Streptococcus pyogenes* (The bacteria that Cas9 comes from)(309) ; this PAM acts as a spacer sequence which allows the Cas9 enzyme to differentiate between phage DNA and the DNA of the cell (310), thus preventing the cleavage of endogenous DNA. Overall, the size of a Cas9 gRNA is in the region of 100 nucleotides.

The structure of the gRNA varies considerably and tends to change when different Cas enzymes are being utilised. For instance, in the case of Cas12a the gRNA can be as little as 20 nucleotides.

A.1.1.1.2. Cas9:

The Cas9 enzyme which has been isolated from *S. pyogenes*, though it should be noted that there are several Cas9 enzymes from different bacteria, such as those isolated from *Neisseria Meningitidis* and *Staphylococcus aureus* which carry out similar functions to the Cas9 enzyme from *S. pyogenes* though they have different physical properties, such as size(311).

[6] N = any nucleotide

Garcia-Doval and Jinek (312) go on to discuss the physical characteristics of the Cas9 enzyme; in essence:

"Cas9 is a bilobed molecule consisting of nuclease (NUC) and alpha-helical recognition (REC) lobes. The NUC lobe contains the HNH and RuvC nuclease domains as well as the C-terminal Wedge (WED) and PAM-interacting (PI) domains"(312).

Of course, it is this enzyme which forms a critical part of the CRISPR/Cas9 complex; without it, DNA cannot not be modified.

However, it's worth noting that this is all well and good should one possess a Cas9 enzyme as well as a well-designed gRNA, but nothing can be done unless the CRISPR/Cas9 complex (or any CRISPR/Cas complex for that matter) is inserted into a cell, and this shall be discussed later.

When the Cas9 enzyme has a suitable gRNA inserted into it, and the complimentary DNA is detected by the enzyme, it cleaves the target DNA, causing a double-strand break within the target DNA. **Figure 60**, opposite, gives a broad overview of this:

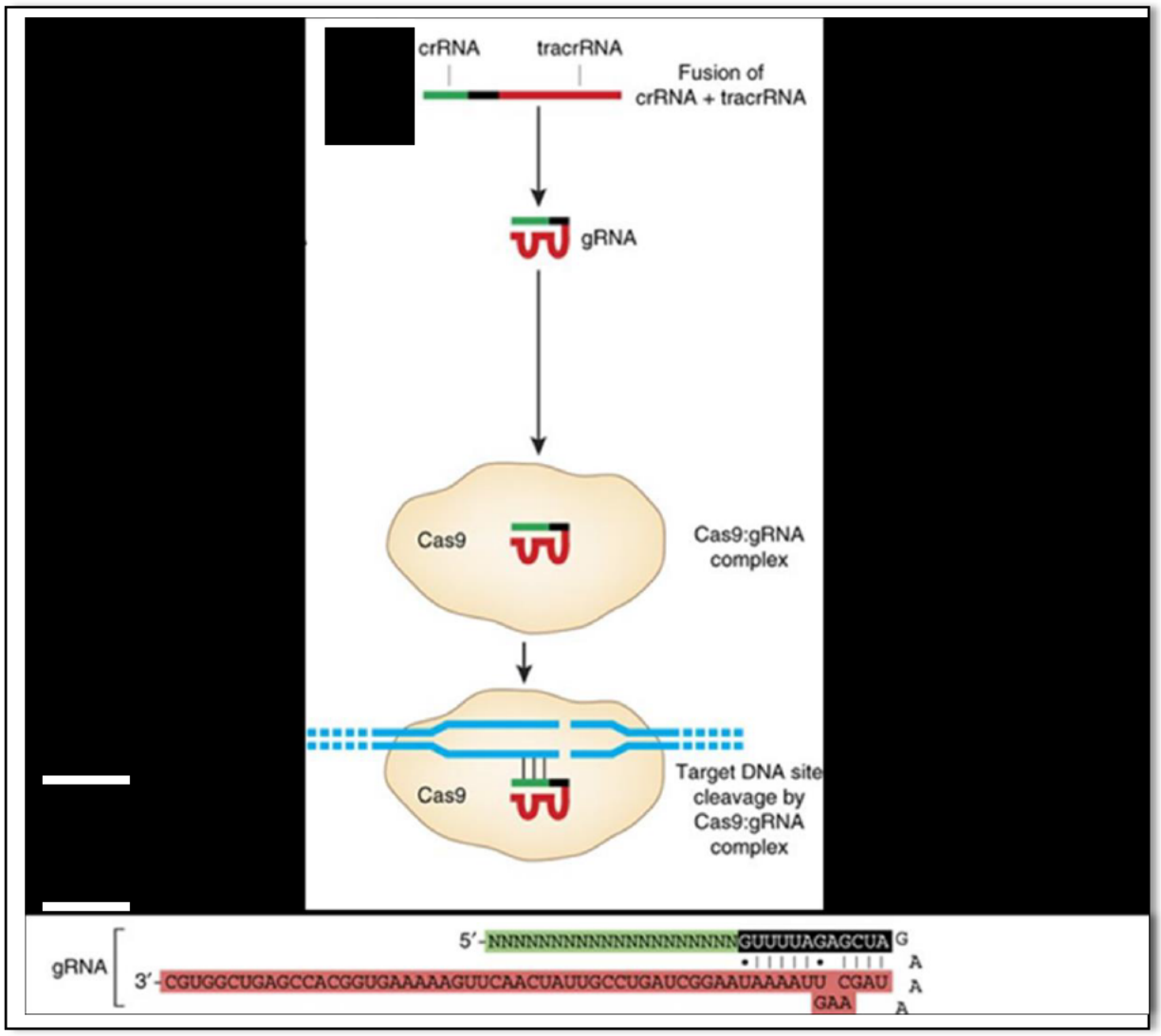


Figure 60 - (A) The most widely used engineered CRISPR-Cas system utilizes a fusion between a crRNA and part of the tracrRNA sequence. This single gRNA complexes with Cas9 to mediate cleavage of target DNA sites that are complementary to the 5′ 20 nt of the gRNA and that lie next to a PAM sequence. **(B)** Example sequences of a gRNA. (Adapted from Sander & Young (309))

Following the CRISPR/Cas9 reaction, the cell will begin to try and repair the damage to the DNA. It is possible to manipulate this process. The repair process takes two forms:

1. Non-homologous end joining (NHEJ) (313)
2. Homology directed repair (HDR) (314)

NHEJ is a quick fix for the cell that is being modified with CRISPR/Cas9. As a result, it is often error prone, which leads to insertions or deletions (indels) of random nucleotides into the genome as the cell seeks to repair its DNA. These indels can sometimes lead to a frameshift, causing a change in the expression of a gene.

Meanwhile, HDR is a guided process which enables cells to alter gene expression. This requires the insertion of a repair template which has homology surrounding the sequence being targeted by the CRISPR/Cas complex.

NHEJ is a substantially more efficient process than HDR, however, NHEJ is often random, though it is incredibly useful when attempting to generate knockout mutants. HDR, on the other hand, has lower efficiency but is useful for generating mutants where one is attempting to change the genome of a cell completely (*e.g.* replacing a certain sequence, or introducing a fluorophore, *etc.*).

Other factors one needs to consider when utilising CRISPR/Cas is the prospect of off-target effects. These are effects achieved when CRISPR/Cas is used, and it accidentally targets other parts of the genome. The best way to avoid this is effective gRNA design and ensuring that the sequence being used within the gRNA is not present in other places within the genome of the targeted cells.

Cas9 is the most widely characterised Cas enzyme simply due to being discovered first (315), however, for the sake of this project, other enzymes need to be considered, primarily in the form of Cas12a.

A.1.1.1.4. Cas12a:

Cas12a is similar to Cas9 in that it cleaves DNA, however one of the major differences is that it is able to create a staggered break, which allows for the insertion of products that have been designed to have sticky ends and enables the novel product to be inserted into the DNA of a living organism thus changing the way that genes are expressed.

Furthermore, unlike the Cas9 enzyme, Cas12a does not require the use of a synthetic crRNA. It is able to use pre-crRNA available within the genome of the targeted cell to use as its own crRNA (316). As a result, the Cas12a enzyme can be applied in circumstances where a smaller product is required during transfection, *i.e.* DNA viruses, certain cell types, *etc.*

One of the other benefits of this staggered cleaving is that the Cas12a enzymes also possess collateral activity. That is, once they have been activated, they are able to cleave other genetic material in the periphery. This can enable synthetic DNA products with fluorescent markers to be inserted into the reaction and therefore generate an output following the successful cleaving of ssDNA, as per (317) where detection of Human Papilloma virus (HPV, a double-stranded DNA virus) DNA was studied and compared with then found to rival the traditional PCR, though it should be noted that this is a mere preliminary study and requires further investigation.

The detection of specific sequences within nucleic acids is, of course, something that brings about a lot of interest and is something that (134) used in the aforementioned paper; where the above technology was also used to detect dsDNA,

A.1.1.1.5. Other Cas enzymes:

Given their nomenclature, one would be correct in assuming that Cas9 and Cas12a are not the only Cas enzymes utilised for CRISPR/Cas reactions. However, they are the most significant to discuss within the confines of this project. Moving forward, the most critical takeaway is that Cas9 and Cas12a do not exist in isolation.

A.1.1.1.6. Transfection methods:

The use of CRISPR/Cas is completely redundant unless one possesses a delivery system for the cells used in an experiment. As may be expected, delivery methods vary just as much as our cells; the choice of delivery method varies considerably on whether one is attempting to use CRISPR/Cas9 on eukaryotes, prokaryotes, fungi, and viruses.

Lino *et al.* (318) goes on to cover this topic extensively; breaking this down into two main categories: cargo and delivery vehicles. In essence, “cargo” refers to ways to bundle the system by carrying a product which encodes the Cas9 protein, and the gRNA; this could be in the form of a plasmid which encodes both (319), an mRNA with a separate guide RNA (320), or a Cas9 protein with gRNA as a ribonucleoprotein complex (321).

In terms of delivery vehicles, this can vary considerably, and depends on the size of the product as well as other facts. Lino *et al.* (318) goes on to cover this extensively in **Table 14** opposite:

Delivery vehicle	Composition	Most common cargo	Capacity	Advantages	Limitations	Ease of use
Microinjection[322, 323]	Needle	DNA plasmid;mRNA (Cas9+sgRNA);Protein (RNP)	nM levels of Cas9 and sgRNA	Guaranteed delivery into cell of interest	Time-consuming; difficult; generally, in vitro only	****
Electroporation; nucleofection[324]	Electric current	DNA plasmid;mRNA (Cas9+ sgRNA)	nM levels of Cas9 and sgRNA	Delivery to cell population; well-known technique	Generally, in vitro only; some cells not amenable	*
Hydrodynamic delivery[325, 326]	High-pressure injection	DNA plasmid; Protein (RNP)	nM levels of Cas9 and sgRNA	Virus-free; low cost; ease	Non-specific; traumatic to tissues	**
Adeno-associated virus (AAV) [327, 328]	Non-enveloped, ssDNA	DNA plasmid	<5kb nucleic acid	Minimal immunogenicity	Low capacity	***
Adenovirus[329]	Non-enveloped, dsDNA	DNA plasmid	8kb nucleic acid	High efficiency delivery	Inflammatory response; difficult scaled production	***
Lentivirus[330, 331]	Enveloped, RNA	DNA plasmid	10kb, up to 18kb nucleic acid	Persistent gene transfer	Prone to gene rearrangement; transgene silencing	***
Lipid nanoparticles/ liposomes/lipoplexes[332, 333]	Natural or synthetic lipids or polymers	mRNA (Cas9+sgRNA); Protein (RNP)	nM levels of Cas9 and sgRNA	Virus-free; simple manipulation; low cost	Endosomal degradation of cargo; specific cell tropism	**

Delivery vehicle	Composition	Most common cargo	Capacity	Advantages	Limitations	Ease of use
Cell-penetrating peptides (CPPs) [334]	Short amino acid sequences	Protein (RNP)	nM levels of Cas9 and sgRNA	Virus-free; can deliver intact RNP	Variable penetrating efficiency	**
DNA nanoclew[335]	DNA spheroid	Protein (RNP)	nM levels of Cas9 and sgRNA	Virus-free	Modifications for template DNA needed	****
iTOP[336]	Hyperosmolality + transduction compound	Protein (RNP)	nM levels of Cas9 and sgRNA	Virus-free; high-efficiency	Non-specific; no in vivo use yet reported	***
Gold nanoparticles (AuNPs)[337]	Cationic arginine-coated AuNP	Protein (RNP)	nM levels of Cas9 and sgRNA	Inert; membrane-fusion-like delivery	Nonspecific inflammatory response	**
SLO[338]	Bacterial pore-forming toxin	100kDa proteins and complexes	Unknown for CRISPR	Reversible pore formation; no impact on cell viability	Not yet proven with CRISPR	***
MENDs[339]	Poly-lysine core, lipid coating, CPP decoration	Nucleic acids	Unknown for CRISPR	Customizable; readily modified for precise delivery	Not yet proven with CRISPR	****
Lipid-coated mesoporous silica NPs[340]	Mesoporous Si coated with lipid	Small molecules and short RNA sequences	Unknown for CRISPR	Inert; easy modification with targeting moieties	Not yet proven with CRISPR	***
Inorganic NPs[341]	NPs of various compositions (carbon, silica)	Large proteins, nucleic acids	Unknown for CRISPR	Inert; used for similar applications	Not yet proven with CRISPR	**

Table 14 - List of CRISPR/Cas transfection methods

A.1.1.2. Molecular detection with CRISPR

With the ability to selectively target specific genetic elements, it becomes feasible to design highly discriminatory diagnostic tools which can recognise the genetic material from pathogens.

This has taken multiple approaches, including the combined use of two deactivated cas9 enzymes bound to a luciferase sub-unit and two gRNAs which are set closely apart (342). Upon successful binding of the targeted gene, the luciferase can form a whole unit and emit fluorescence, a measurable output.

The above luciferase-Cas hybrid is a lone example, but it is worth noting that there are multitudes of other ways to undertake molecular detection with CRISPR, as Yuan *et al.* (343) and Shin *et al.* (344) cover in extensive detail.

Whilst there is plenty of relevance to discuss for this project, the ability of Cas12a to collaterally and indiscriminately cleave other oligonucleotides is of particular interest, specifically those bound to a fluorophore and quencher (345). Through cleaving these oligonucleotides, the fluorophore is released from its quencher and able to emit a fluorescent output.

This was demonstrated by Chen *et al.* in 2018 and, at the outset of this project, had not yet been undertaken with relevance to *M. tuberculosis* or other members of the MTBC.

One of the drawbacks of the research published in Chen *et al.* is that they required the use of Recombinase Polymerase Amplification (RPA) and an extraction process, a process that takes time.

RPA is like PCR in that it results in the same output: an amplification of nucleic acid. However, in contrast to PCR, it is an isothermal reaction (requiring a single temperature as opposed to the cycling required in PCR) and it is a single tube reaction (346).

Thankfully, the temperatures required in RPA are lower than those in PCR (insofar that they can be readily obtained in almost all settings) and the rate of the reaction is proportional to the temperatures obtained. The optimal temperature for RPA is approximately 37-42^{o}C. The reaction can take place at temperatures barely exceeding room temperature, albeit at a lower rate.

Therefore, it can be readily used in rural settings where more austere equipment may be the norm. Additionally, RPA can withstand ambient temperatures for days at a time, though long-term storage at 20^{o}C is recommended.

With these advantageous properties, RPA can present as a strong alternative to PCR. Taking note of this, the use of RPA and CRISPR/Cas detection is still a multi-step process, and the removal of RPA is likely to enable a greater proliferation of molecular detection *via* CRISPR/Cas.

Regardless, once the genetic material has been amplified, it can then be tested using the CRISPR/Cas array. In the case of Cas12a, the enzyme makes a staggered

cut and then carries out the collateral activity mentioned above ; providing that there are oligonucleotides in the periphery where there is a fluorophore and quencher, these are separated and will result in a fluorescent output (134).

This is simply one example of the Cas12a enzyme being utilised in this manner, however, other researchers have sought to use the enzyme in order to present with a colorimetric output (albeit with a higher Limit-of-Detection) (347)rather than a fluorescent output and this is likely to be of great significance in settings where a plate reader is not available.

A.1.1.3. Genetic Elements for Molecular Detection:

When identifying genetic elements for the identification of bacteria, it is critical to establish that these elements are not present within other bacterial species or other bacteria in general. Paradoxically, it is also critical to establish that the genetic element is not so niche that it is only able to identify a specific serovar and no other.

So far, the most prominent genetic element used for the identification of MTBC in any molecular test is *IS6110* and it is this genetic element which is used in the GeneXpert MTB test (and the vast majority of other molecular tests (348). Additionally, the GeneXpert test has been somewhat modified to include the genes conferring resistance to rifampicin.

Furthermore, an updated test, the GeneXpert TB/Rif Ultra, includes primers and probes which expand the genes amplified from a test sample and confers a greater degree of accuracy and sensitivity when compared to the standard GeneXpert (349).

The GeneXpert TB/Rif series, as is clear by its very namesake, has been designed to detect some bacteria from the MTBC and rifampicin resistance despite other types of resistance being present within MTBC. The former limitation is not as significant as its inability to detect other genes which confer drug resistance (*i.e.* Isoniazid, ethambutol, *etc.*). More so, whilst rifampicin resistance is notable, it is not the only type of drug resistance and, in some settings, it is not the most prevalent type of drug resistance (350). Therefore the inclusivity of other types of drug resistance in TB is a critical step in strengthening the fight against this awful disease.

Additionally, whilst the *IS6110* is highly conserved across multiple variants of *M. tuberculosis* this conservation is not applicable to all strains of the bacteria. Sankar *et al.* (351) discusses this in length. The consensus is that *IS6110* gene is repeated throughout a notable number of *M. tuberculosis* though this repeat varies greatly and that new genetic elements must be considered.

Therefore, the presence of the gene can be well below the limit-of-detection in some cases and, in others, it may be highly fragmented or not even present at all.

There is justification to start to examine other genetic targets within the *M. tuberculosis* genome and consider methods which can be combined in order to make a diagnostic test more comprehensive.

This chapter will therefore focus on the design of CRISPR diagnostics utilising novel genetic elements which will identify the presence or absence of *M. tuberculosis*. With this having been done, it is hoped that such a test could be multiplexed in order to confirm the presence or absence of *M. tuberculosis.*

As a proof of concept, it was hoped that *E. coli* could be heterologously transformed with non-pathogenic genes from *M. tuberculosis*. The same *E. coli* would then be transformed using plasmids conferring the production of Cas9 and appropriate gRNA. All three plasmids would confer different types of antibiotic resistance as a marker.

Therefore, following the successful use of Cas9, the plasmid containing the genes from *M. tuberculosis* would undergo a frameshift mutation, thus impacting ampicillin resistance and making the cells sensitive to ampicillin once more.

Following the confirmed successful design of gRNA for Cas9, the gRNA would have then been optimised for use with Cas12a and multiplexed with poly-A probes, these probes would undergo cleavage by Cas12a and release a fluorescent signal when in the presence of *M. tuberculosis* DNA.

Future work could have then focused on the development of an assay which recognises SNPs as well as other genetic elements which confer drug resistance.

This work commenced in autumn of 2018 but was later superseded by the work of Ai *et al.* 2019. The combination of this and the ongoing COVID-19 pandemic then led to this work being abandoned.

In essence, Ai *et al.* 2019 is able to demonstrate the feasibility of DNA extraction from both *in vitro* and patient samples, and then utilise CRISPR-Cas12a for the detection of *M. tuberculosis* DNA. The paper discusses how this was able to be performed in 1.5 hours, however there is a need for further multi-centre prospective

research to deepen understanding and validate clinical applications. Additionally, consideration must also be given to drug resistance mechanisms as well and how these can be incorporated into future tests.

A.1.1.3.1. Bacterial and Viral Bioinformatics Resource Center (BV-BRC):

The Bacterial and Viral Bioinformatics Resource Center (BV-BRC) was formed in 2019 through a merger of three BRC resources: the PAThossystems Resource Integration Center (PATRIC), the Influenza Research Database (IRD) and the Virus Pathogen Database and Analysis Resource (ViPR)(352).

Having been cited 1000s of times, the joint venture between the University of Chicago, the University of Virginia, as well as the J Craig Venter Institute is a priceless resource that contains a great deal of genomic information with over 600000 annotated bacterial genomes.

This information is critical to understand the genetic phylogeny of bacteria. Naturally, such a resource is valuable when attempting to identify genetic elements for a molecular-based assay to identify a pathogen.

A.1.1.3.2. CHOPCHOP:

CHOPCHOP is a gRNA design tool developed by researchers at Harvard University. Optimised for the use of Cas9, Cas12a and Cas13b, the tool enables researchers to optimise their CRISPR/Cas reactions by identifying potential targets within an organism’s genome and ranking the results based on efficiency (353).

A.1.1.3.3. Basic Local Alignment Search Tool (BLAST):

The Basic Local Alignment Search Tool is a well-established aid to researchers that wish to validate proposed genetic sequences. BLAST enables researchers to search for a nucleotide sequence and to check its origin, as well as its alignment to a wide-array of organisms.

BLAST has a history going as far back as the early 90s and has been cited countless times since then. As time has worn on, BLAST has expanded its uses to incorporate utilities such as protein alignment, primer design for PC and to check for multiple alignment across various domains. BLAST is one of the resources owned and maintained by the US government's National Center for Biotechnology Information (354).

A.1.1.3.4. Addgene:

Addgene is a well-reputed non-profit plasmid repository that maintains thousands of various plasmid and viral vectors with multiple uses. Addgene itself has received endorsements from many highly respected journals such as Nature, PLOS One, the Proceedings of the National Academy of Sciences (PNAS) and receives its funding from various government grants as well as New England Biolabs, the Chan Zuckerberg Initiative and other commercial partners (355).

A.1.2. Methods:

A.1.2.1. gRNA design:

Genomes were narrowed down by isolating all strains of *M. tuberculosis* from their phylogenetic *niche*. 143 fully sequenced genomes were identified; individual genes were then identified and selected for based on size and prevalence amongst the 143 sequenced strains. Genes which did not appear in all 143 strains were omitted, as were genes with polymorphisms. Following this, 5 genes were identified:

1. 4FE-4S ferredoxin: An iron-sulphur protein that facilitates electron transfer
2. Translation elongation factor TU protein: Functional in the formation of tRNA
3. Succinyl CoA ligase (ADP forming) beta chain: An enzyme that catalyses the formation of Guanosine diphosphate from Guanosine triphosphate
4. LSU riboprotein L7p/L12p: Large Sub-Unit Riboprotein; critical for the formation of ribosomes and protein transport
5. Glycosyltransferase: Enzyme responsible for the catalyse the formation of glycosidic bonds between glycosyl groups

Once sequences had been obtained from the reference strain *M. tuberculosis* H37Rv (See **A.3. Genetic elements identified in *M. tuberculosis* H37Rv**:) and sequence alignment had been confirmed, artificial plasmids containing the above nucleic acid sequences were then ordered from Thermo Fisher (**A.4. pCRISPR, pCAS9, and pMA-RQ plasmid maps**:). Sequences were codon optimised for cloning in *E. coli*. CRISPR motifs identified through CHOPCHOP (as per **A.5. Oligonucleotide sequences ordered from IDT**:) were protected and therefore reflect theoretical

motifs in *M. tuberculosis*. Synthetic, complimentary oligonucleotides were ordered from Integrated DNA technologies (Leuven, Belgium) in accordance with Jiang *et al.* (311) to facilitate upstream ligation.

gRNA and Cas9 were to be generated using the plasmids pCRISPR and pCAS9 (203), respectively; acquired from Addgene (**Figure 61 - Plasmid maps for pCRISPR, pCAS9 and pMA-RQ, acquired from Addgene & Thermo Scientific** (172)). pCRISPR was to have each CRISPR motif ligated into it and the whole plasmid was to be used to transform separate cultures of *E. coli,* giving rise to five different cultures of *E. coli* which contain the five gRNAs previously mentioned.

A.1.2.2. Bacterial culture & transformation:

Five aliquots of *E. coli* TOP10 were transformed with five separate plasmids containing the genes identified above, according to “Gibson Assembly Chemical Transformation Protocol (E5510)” available *via* New England Biolabs (356).

Once transformed, bacteria were cultured on 20ml of selective Lysogeny broth (LB) 1% agar with 50µl/ml of ampicillin, alongside a negative control of 20ml selective LB 1% agar with 100µl/ml kanamycin.

Plates were then incubated overnight at 37^{o}C.

Following incubation, colonies were picked and placed in 50ml of selective Lysogeny broth with 50µl/ml of ampicillin and incubated overnight in a shaker at 37^{o}C and 200RPM.

After liquid culture, the bacteria were centrifuged at 3000RPM for 5 mins and the supernatant was then removed. The pelleted bacteria were then resuspended in 5ml of 50:50 Glycerol:LB and then stored at -80^{o}C until required.

E. coli containing pCRISPR and pCAS9 were shipped in bacterial stabs from Addgene; bacteria were then prepared as above, albeit pCRISPR plates/broths contained 100µl/ml kanamycin, and pCAS9 plates/broths contained 50µl/ml chloramphenicol.

A.1.2.3. Vector digest & ligation:

Vector digest and ligation was attempted as per Jiang *et al.* (203) . Oligonucleotides from IDT were resuspend according to package instructions in RNAse free water. No resistant colonies were present following attempt at ligation.

Due to initial failure, SnapGene (GSL Biotech LLC, San Diego) was consulted and used to simulate the ligation with the proposed oligonucleotides; SnapGene concluded that no such ligation could take place.

It was then proposed that the gRNA insert could be synthesised *via* PCR with BsaI restriction enzyme sites straddling the PCR product and the respective motif from pMA-RQ being used as a template; this would facilitate the creation of sticky ends suitable for ligation into the pCRISPR plasmid.

A.1.3. Results:

Unfortunately, due to the COVID-19 pandemic and the emergence of Ai *et. al* 2019, the above work was abandoned and therefore no results can be presented.

A.1.4. Discussion:

The theoretical work presented in this chapter, as well as the subsequent work presented by Ai *et. al* 2019 illustrates the potential of CRISPR as a diagnostic tool for TB, however, beyond this initial paper, there has not been a substantial effort to establish such a method for use against TB. This could be as a result of the COVID-19 pandemic which began soon after the publication of this particular paper.

The adoption and investigation into a similar methodology for the diagnosis of COVID-19 has shown that it is indeed feasible to use CRISPR/Cas based diagnostics in the field (357); with comparable sensitivity and specificity to other techniques (358). However, at the time of writing, no CRISPR/Cas based test has made it to market.

Perhaps more so, the rapid development of CRISPR/Cas based diagnostics for COVID-19 has shown that even rapidly emerging pathogens can be identified using the technique. Therefore the technique could be adopted to identify emerging drug resistance mechanisms shortly after they begin to emerge and have been identified.

With this in mind, Ai *et al.* have illustrated the feasibility of the use of CRISPR/Cas as a diagnostic method on clinical patient samples. The proposed work within this chapter suggests the use of other genetic elements within the *M. tuberculosis* genome in a bid to ensure accuracy, as there can be genetic drift within *M. tuberculosis (359)*, which PCR can accommodate but CRISRP cannot, due to its specificity.

Ai *et al.* themselves go on to state that the sole reason *IS6110* was used because it is already recognised as a genetic element within TB testing and one which could be rapidly utilised within their proposed test.

Moving forward, future research should therefore focus on investigating the use of other genetic elements within *M. tuberculosis* as well as other members of the MTBC. Additionally, it would prove worthwhile to consider a multiplexing test which can be readily deployed to rural areas as well as a more comprehensive test which can readily detect SNPs conferring drug resistance and other genetic elements which confer drug resistance.

Ultimately, the use of a CRISPR based test is a step in the right direction, since it can already be placed onto a paper or paper-like material which can be used at room temperature. Presently, the primary challenges are to ensure that it is comprehensive enough to identify various MTBC bacteria and to ensure that it can test for drug resistance, though there is a degree of certainty that, following the wake of COVID-19, there is likely to be renewed interest in tuberculosis and the future is starting to look positive.

Lastly, the overall goal of any researchers developing such a test would be to overcome the reticence of some regulatory bodies such at the FDA and MHRA when it comes to establishing novel methods of diagnostics such as CRISPR and this may prove to be the most difficult hurdle for researchers to conquer.

A.2. SEM preparation protocol for bacteria:

Main Fixative: Consists of 3% by volume glutaraldehyde & 0.15% Alcian blue in 0.1 M cacodylate buffer pH 7.2.

Buffer: 0.1 M cacodylate buffer at pH 7.2

Post fixative: 1% Osmium tetroxide in 0.1 M cacodylate buffer at pH 7.2

Method:

Agent	Time (Mins)
Main fixative	60
Buffer rinse	10
Buffer rinse	10
Post fixative	60
Buffer rinse	10
Buffer rinse	10
30% ethanol	10
50% ethanol	10
70% ethanol	10
95% ethanol	10
Absolute ethanol	20
Absolute ethanol	20

Table 15 - Methodology for timings and fixing agents for SEM sample preparation

Samples are then dried at their critical point in CO_2 before they are mounted on carbon films and sputter coated in 1nm of platinum using an argon ion beam coater (360).

Samples were then imaged on an FEI Quanta 250 Scanning Electron Microscope (FEI Company, USA).

A.3. Genetic elements identified in *M. tuberculosis* H37Rv:

Gene	Sequence
Succinyl-CoA Ligase	ATGGATCTTTTCGAGTATCAAGCCAAGGAGTTATTCGCCAAGCACAACGTGCCCAGCACGCCGGGTCGGGTGACCGACACAGCCGAGGGTGCCAAGGCTATCGCCACGGAGATCGGGCGT CCGGTGATGGTCAAAGCGCAGGTCAAGATCGGCGGCCGGGGCAAGGCCGGTGGCGTCAAATACGCCGCGACCCCACAAGACGCGTACGAGCACGCCAAGAAACATCCTCGGCCTGGACATC AAAGGACACATCGTCAAGAAACTGCTGGTCGCTGAGGCTAGCGATATCGCCGAGGAGTACTACCTATCCTTCCTGCTCGACCGGGCCAACCGCACCTACCTGGCGATGTGCTCGGTGGAG GGCGGCATGGAGATCGAAGAGGTAGCGGCCACCAAACCCGAGCGGCTCGCCAAAGTCCCGGTGAATGCCGTCAAGGGCGTTGACCTAGATTTCGCGCGGTCCATCGCCGAACAGGGTCAT CTTCCGGCCGAGGTGCTCGACACCGCAGCGGTCACCATCGCCAAGCTGTGGGAGCTCTTCGTCGCCGAGGACGCGACGCTGGTTGAGGTCAACCCGTTGGTGCGGACGCCTGACCACAAG ATCCTCGCGCTGGATGCCAAGATCACCCTCGACGGCAACGCCGATTTCCGTCAGCCTGGCCATGCCGAGTTCGAGGATCGAGCTGCCACCGATCCACTGGAGTTGAAGGCCAAGGAGCAC GACCTCAACTACGTCAAGCTGGACGGTCAGGTGGGGATCATCGGCAATGGCGCGGGCTTGGTGATGTCGACTCTCGACGTCGTCGCGTATGCCGGTGAGAAGCACGGCGGAGTCAAGCCG GCCAACTTCCTGGATATCGGCGGCGGCGCTTCGGCCGAGGTGATGGCCGCGGGTCTGGACGTGGTGCTGGGCGACCAGCAGGTCAAGAGCGTGTTCGTCAACGTCTTCGGTGGCATCACC TCGTGCGATGCGGTGGCGACCGGGATCGTCAAGGCGCTGGGCATGCTGGGTGACGAAGCCAACAAGCCGCTGGTGGTTCGGCTCGACGGCAACAACGTCGAGGAAGGCCGTCGCATCCTG ACCGAGGCCAACCACCCCCTGGTGACACTGGTGGCGACGATGGACGAAGCCGCCGACAAGGCCGCTGAGCTGGCGAGCGCCTGA
4Fe-4S Ferredoxin	GTGACGTACACGATCGCCGAACCCTGTGTCGACATCAAGGACAAGGCATGCATTGAGGAGTGCCCGGTCGATTGCATCTACGAGGGCGCCCGGATGCTGTATATCCACCCCGACGAATGC GTCGACTGTGGGGCTTGCGAGCCGGTCTGCCCCGTTGAAGCTATCTTCTACGAAGACGATGTGCCCGAACAGTGGAGCCATTACACCCAGATCAACGCCGATTTCTTCGCCGAGCTGGGA TCGCCGGGCGGTGCGGCCAAGGTTGGCATGACCGAGAACGACCCGCAAGCGGTCAAGGATCTGGCGCCGCAGAGCGAGGACGCCTGA
Translation Elongation Factor Tu	GTGGCGAAGGCGAAGTTCCAGCGGACCAAGCCCCACGTCAACATCGGGACCATCGGTCACGTTGACCACGGCAAGACCACCCTGACCGCGGCTATCACCAAGGTCCTGCACGACAAATTC CCCGATCTGAACGAGACGAAGGCATTCGACCAGATCGACAACGCCCCCGAGGAGCGTCAGCGCGGTATCACCATCAACATCGCGCACGTGGAGTACCAGACCGACAAGCGGCACTACGCA CACGTCGACGCCCCTGGCCACGCCGACTACATCAAGAACATGATCACCGGCGCCGCGCAGATGGACGGTGCGATCCTGGTGGTCGCCGCCACCGACGGCCCGATGCCCCAGACCCGCGAG CACGTTCTGCTGGCGCGTCAAGTGGGTGTGCCCTACATCCTGGTAGCGCTGAACAAGGCCGACGCAGTGGACGACGAGGAGCTGCTCGAACTCGTCGAGATGGAGGTCCGCGAGCTGCTG GCTGCCCAGGAATTCGACGAGGACGCCCCGGTTGTGCGGGTCTCGGCGCTCAAGGCGCTCGAGGGTGACGCGAAGTGGGTTGCCTCTGTCGAGGAACTGATGAACGCGGTCGACGAGTCG ATTCCGGACCCGGTCCGCGAGACCGACAAGCCGTTCCTGATGCCGGTCGAGGACGTCTTCACCATTACCGGCCGCGGAACCGTGGTCACCGGACGTGTGGAGCGCGGCGTGATCAACGTG AACGAGGAAGTTGAGATCGTCGGCATTCGCCCATCGACCACCAAGACCACCGTCACCGGTGTGGAGATGTTCCGCAAGCTGCTCGACCAGGGCCAGGCGGGCGACAACGTTGGTTTGCTG CTGCGGGGCGTCAAGCGCGAGGACGTCGAGCGTGGCCAGGTTGTCACCAAGCCCGGCACCACCACGCCGCACACCGAGTTCGAAGGCCAGGTCTACATCCTGTCCAAGGACGAGGGCGGC CGGCACACGCCGTTCTTCAACAACTACCGTCCGCAGTTCTACTTCCGCACCACCGACGTGACCGGTGTGGTGACACTGCCGGAGGGCACCGAGATGGTGATGCCCGGTGACAACACCAAC ATCTCGGTGAAGTTGATCCAGCCCGTCGCCATGGACGAAGGTCTGCGTTTCGCGATCCGCGAGGGTGGCCGCACCGTGGGCGCCGGCCGGGTCACCAAGATCATCAAGTAG
Glycosyltransferase	ATGTCTGCTCTGCGCTCGGTGTTGCTGCTGTGCTGGCGCGACATCGGGCACCCCCAGGGGGGCGGGAGCGAAGCCTATCTGCAACGCATCGGGGCTCAGTTGGCCGCATCGGGCATTGCA GTCACGTTGCGCACCGCTCGCTATCCCGGTGCGCCACGGCATGAACTGGTCGACGGGGTGCGGATCAGTCGTGCCGGCGGGCGCTACTCGGTGTATCTATGGGCGTTGCTGGCGATGGCC GCAGCCCGATGTGGGCTTGGGCCGCTGCGCCGAGTGCGCCCGGATGTGGTCGTCGATACCCAAAACGGCTGGCCGTTTGTGGCCCGGCTGTTGTATGGCCGGCGGTCGCTGGTACTGGTA CACCATTGCCACCGTGAGCAGTGGCCGGTGGCCGGGCGGATGATGGGTCGGCTCGGCTGGTATGTCGAGTCGATGTTGTCGCCACGGCTACACCGGCGCAACCAGTACGTGACGGTGTCG CTGCCGTCGGCGCGGGATCTGATCGCCCTCGGTGTGGACAGCGAGCGGATCGCTGTGGTGCGCAACGGCCTCGACGAGGCGCCGTCGCCAACGTTGTCCGGCCCACGTGCGCCCACGCCG CGTGTGGTGGTGCTCTCCCGGCTGGTGCCGCACAAGCAGATCGAGGACGCGTTGGCAGCGGTCGCGGAGCTACAGCCTCGGATACCGGGCCTGCACCTAGACATCGTCGGCGGTGGCTGG TGGCGGCAGCGCCTCGTTGACCATGTGCACCGGCTCGACATTGCTGACGCCGTTACCTTTCACGGGCATGTCGACGATGTGACCAAACACCATGTGCTGCAAAGCTCCTGGGTGCACTTG TTGCCCTCACGTAAAGAGGGATGGGGGCTCGCGGTCATCGAGGCGGCCCAGCACGGCGTGCCCACCATCGGGTACAGATCCTCCGGTGGTTTGGCGGACTCGATCGTCGACGGGGTGACC GGCATATTGGTCGACGACCGGGCCGAATTGGTGGCTTGGCTCGAACAACTGCTGTCCGATTCGGTGCTGCGTGACCAACTCGGCGCCAAGGCACAGGCGCGTAGCGGTGAGTTCTCCTGG CGGCAAAGCGCCGAAGCGCTGCGCAGCGTGTTGGAGGCAGTGCAGGCCAGCCGTTTTGTCAGCGGCGTGGTTTGA
LSU Riboprotein L7p/L12p	ATGGCAAAGCTCTCCACCGACGAACTGCTGGACGCGTTCAAGGAAATGACCCTGTTGGAGCTCTCCGACTTCGTCAAGAAGTTCGAGGAGACCTTCGAGGTCACCGCCGCCGCTCCAGTC GCCGTCGCCGCCGCCGGTGCCGCCCCGGCCGGTGCCGCCGTCGAGGCTGCCGAGGAGCAGTCCGAGTTCGACGTGATCCTTGAGGCCGCCGGCGACAAGAAGATCGGCGTCATCAAGGTG GTCCGGGAGATCGTTTCCGGCCTGGGCCTCAAGGAGGCCAAGGACCTGGTCGACGGCGCGCCCAAGCCGCTGCTGGAGAAGGTCGCCAAGGAGGCCGCCGACGAGGCCAAGGCCAAGCTG GAGGCCGCCGGCGCCACCGTCACCGTCAA AAAACGCTTTAACAAAGGCCGGACCCGGAGTTCCT

Table 16 - Table illustrating the genetic sequences of the genes identified in *via* BV-BRC, highlighted sequences were used as basis for gRNA design and “protected” when ordered from Thermo Fisher.

A.4. pCRISPR, pCAS9, and pMA-RQ plasmid maps:

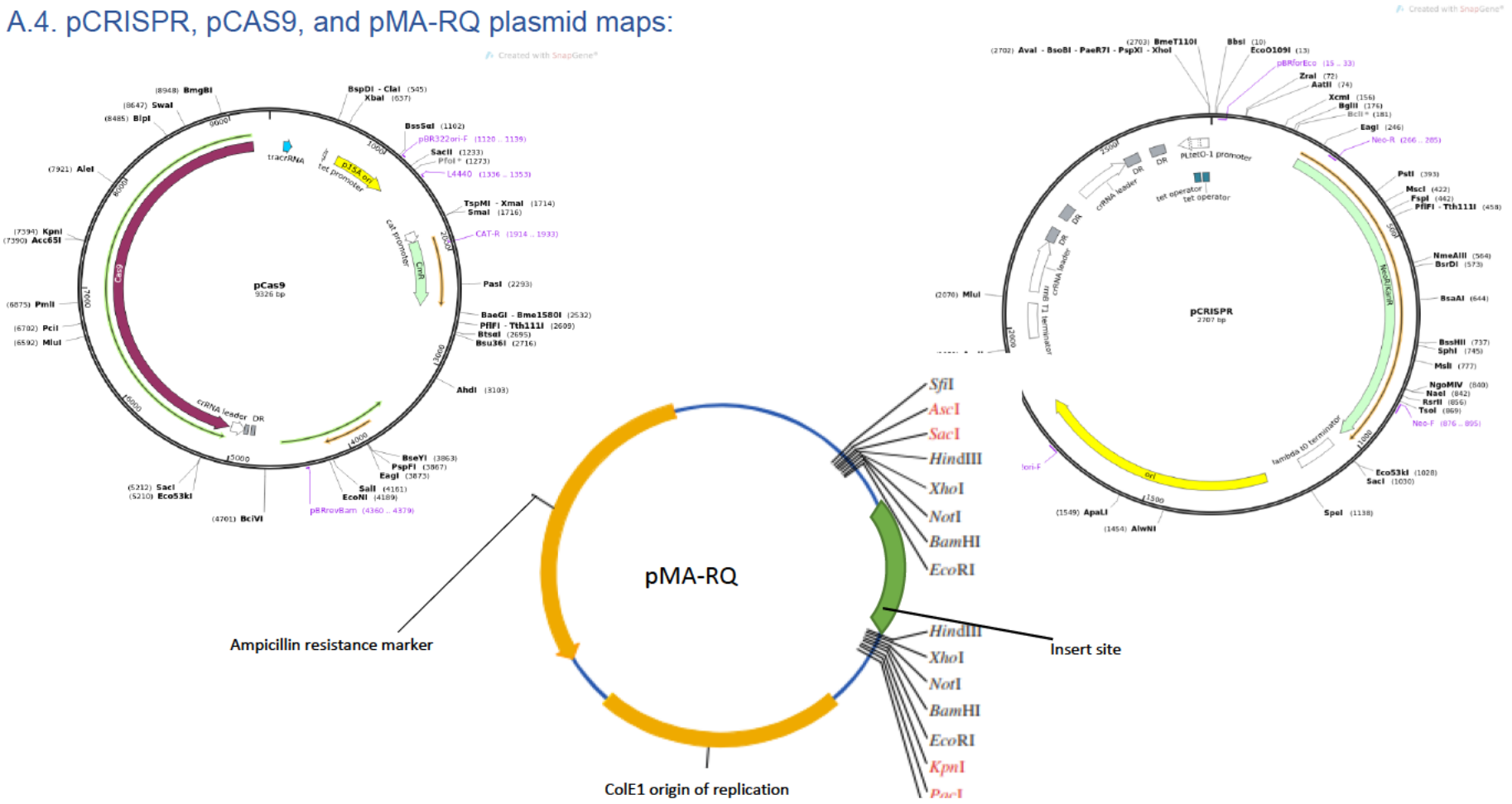


Figure 61 - Plasmid maps for pCRISPR, pCAS9 and pMA-RQ, acquired from Addgene & Thermo Scientific (172)

A.5. Oligonucleotide sequences ordered from IDT:

Oligo	Sequence
LSU Forward	AAACGAAGATCGGCGTCATCAAGGTGGTCCGGGAG
LSU Reverse	AAAACTCCCGGACCACCTTGATGACGCCGATCTTC
Glycosyltransferase Forward	AAACGTAGCGGTGAGTTCTCCTGGCGGCAAAGCGG
Glycosyltransferase Reverse	AAAACCGCTTTGCCGCCAGGAGAACTCACCGCTAC
Elongation Factor Forward	AAACTGTCGAGGAACTGATGAACGCGGTCGACGAG
Elongation Factor Reverse	AAAACTCGTCGACCGCGTTCATCAGTTCCTCGACA
Ferredoxin Forward	AAACAAGACGATGTGCCCGAACAGTGGAGCCATTG
Ferredoxin Reverse	AAAACAATGGCTCCACTGTTCGGGCACATCGTCTT
CoA Ligase Forward	AAACTATGCCGGTGAGAAGCACGGCGGAGTCAAGG
CoA Ligase Reverse	AAAACCTTGACTCCGCCGTGCTTCTCACCGGCATA

Table 17 - Sequence of Oligonucleotides ordered from IDT with gRNA sequences highlighted

A.6. Example of BlueBEAR input file:

```
# avogadro generated ORCA input file
# Basic Mode
# Alpha-Mycolic-Acid
! BP86 def2-SVP
%PAL NPROCS 16 END
! UKS PBE TZVP
! Direct
! PModel
! XYZFile
! TightSCF
! PAL16

! Opt
! NumFreq

%elprop Polar 1
       Tol 1e-7
end

%method
 Grid        2
 FinalGrid   4
end

%output
 Print [ P  Hirshfeld ] 1
 Print [ P_Mayer ] 1
end


* xyz 0 1
   C       1.11580      -0.04009       0.04288
   C       0.59297      -1.26595      -0.70107
   C      -0.93276      -1.28383      -0.72013
   C      -1.45384      -2.50512      -1.47232
   C      -2.97923      -2.52290      -1.49056
   C      -3.50211      -3.73948      -2.24903
```

C	-5.02688	-3.77703	-2.22615
C	-5.55832	-4.99362	-2.97862
C	-7.07005	-5.10943	-2.81390
C	2.64200	-0.02088	0.05667
C	3.16815	1.20636	0.79683
C	4.69492	1.22472	0.80765
C	5.22509	2.45150	1.54684
C	6.75253	2.46792	1.55622
C	7.28707	3.69380	2.29513
C	8.81526	3.70830	2.30348
C	9.35348	4.93250	3.04168
C	-8.63870	-7.16030	-2.92319
C	-9.85697	-7.58462	-3.67982
C	-11.00963	-6.62282	-3.39627
C	-12.26938	-7.01722	-4.15796
C	-13.43645	-6.11192	-3.77209
C	-14.70485	-6.50710	-4.52071
C	-15.88082	-5.62983	-4.10093
C	-17.14759	-6.02996	-4.84991
C	-18.32944	-5.15904	-4.43371
C	-19.58764	-5.56418	-5.19418
C	-20.77846	-4.69536	-4.79854
C	-22.02106	-5.11364	-5.57732
C	-23.22825	-4.25139	-5.21687
C	-24.44694	-4.69731	-6.02032
C	-25.68136	-3.86245	-5.69602
C	-26.92821	-4.27984	-6.42296
C	-26.97497	-5.43226	-7.40511
C	-27.08272	-4.01695	-7.90292
C	-30.01631	-10.66578	-5.64438
C	-28.80312	-9.74388	-5.66917
C	-29.10242	-8.48895	-6.48364
C	-27.89624	-7.55855	-6.51640
C	-28.16039	-6.34459	-7.40625
C	10.87323	4.94928	3.05148
C	-29.72218	-11.93811	-4.85887
C	-30.94692	-12.84405	-4.84184
C	-30.66853	-14.11986	-4.05861
C	-31.90414	-15.00988	-4.04382
C	-31.63946	-16.28554	-3.25729
C	-32.88139	-17.16505	-3.24764
C	-32.62600	-18.44306	-2.46320
C	-33.86976	-19.31879	-2.46736
C	-33.61754	-20.60551	-1.69607
C	-34.87018	-21.47446	-1.70201
C	-34.62254	-22.81920	-1.01676
O	-34.14265	-22.59258	0.31108
C	-35.88505	-23.68734	-0.96699
C	-36.34707	-24.00538	-2.37913
O	-35.63111	-24.94883	-3.02642
O	-37.27379	-23.42118	-2.91875
C	-35.66699	-24.93695	-0.10413
C	-36.91660	-25.80281	-0.02144
C	-36.62150	-27.06566	0.77280

C	-37.84611	-27.96375	0.79358
C	-37.54361	-29.25322	1.53772
C	-38.76806	-30.15280	1.52917
C	-38.48671	-31.43830	2.28746
C	-39.72376	-32.32194	2.28714
C	-39.45990	-33.59975	3.06539
C	-40.70328	-34.47518	3.07264
C	-40.44406	-35.75131	3.85673
C	-41.68655	-36.62867	3.86526
C	-46.58511	-47.46267	7.83363
C	-45.34225	-46.57571	7.82234
C	-45.59717	-45.29012	7.04015
C	-44.35640	-44.40296	7.02821
C	-44.61635	-43.12001	6.24580
C	-43.37716	-42.23327	6.23371
C	-43.64029	-40.95253	5.45141
C	-42.40134	-40.06733	5.44023
C	-42.66508	-38.78822	4.65792
C	-41.42496	-37.90633	4.64839
C	-46.33597	-48.75036	8.61485
C	-47.57143	-49.63464	8.62751
C	-7.62732	-6.26113	-3.59015
C	-7.27609	-7.69474	-3.26872
H	0.73682	-0.04354	1.07216
H	0.73941	0.87121	-0.43769
H	0.96649	-2.17769	-0.21904
H	0.97442	-1.26404	-1.72945
H	-1.30712	-0.36978	-1.19700
H	-1.31480	-1.29226	0.30805
H	-1.08009	-3.41939	-0.99518
H	-1.07214	-2.49711	-2.50048
H	-3.35527	-1.60619	-1.96094
H	-3.35869	-2.53735	-0.46141
H	-3.10554	-4.65522	-1.79347
H	-3.14599	-3.71260	-3.28586
H	-5.42946	-2.86172	-2.67703
H	-5.37414	-3.80408	-1.18591
H	-5.07675	-5.90257	-2.59815
H	-5.29845	-4.91922	-4.04166
H	-7.55402	-4.18904	-3.16332
H	-7.31372	-5.20664	-1.74894
H	3.01962	-0.02052	-0.97315
H	3.01757	-0.93182	0.53859
H	2.79136	1.20525	1.82690
H	2.79137	2.11691	0.31518
H	5.07023	1.22452	-0.22301
H	5.07049	0.31335	1.28875
H	4.84959	2.45125	2.57738
H	4.84912	3.36250	1.06545
H	7.12648	2.46712	0.52507
H	7.12699	1.55576	2.03667
H	6.91261	3.69411	3.32605
H	6.91222	4.60555	1.81429
H	9.18830	3.70679	1.27199

H	9.18870	2.79533	2.78324
H	8.98615	4.93602	4.07452
H	8.98581	5.84761	2.56306
H	-8.84534	-7.03108	-1.86522
H	-10.13391	-8.59810	-3.36747
H	-9.65631	-7.62142	-4.75746
H	-10.71722	-5.60060	-3.66838
H	-11.21736	-6.61174	-2.31859
H	-12.52495	-8.06052	-3.93711
H	-12.08523	-6.94866	-5.23690
H	-13.18580	-5.06736	-3.99393
H	-13.61190	-6.17929	-2.69124
H	-14.94045	-7.55916	-4.31876
H	-14.53956	-6.41378	-5.60095
H	-15.64821	-4.57712	-4.30241
H	-16.04562	-5.72431	-3.02060
H	-17.37752	-7.08366	-4.64936
H	-16.98183	-5.93559	-5.93013
H	-18.10035	-4.10494	-4.63165
H	-18.50092	-5.25739	-3.35493
H	-19.81674	-6.61782	-4.99203
H	-19.40951	-5.47469	-6.27299
H	-20.55169	-3.64147	-4.99970
H	-20.96554	-4.78890	-3.72206
H	-22.24644	-6.16705	-5.36899
H	-21.82278	-5.03465	-6.65344
H	-23.00948	-3.19764	-5.42656
H	-23.43766	-4.33503	-4.14374
H	-24.65066	-5.75374	-5.80431
H	-24.21845	-4.62378	-7.09046
H	-25.48490	-2.80423	-5.90573
H	-25.87897	-3.93415	-4.61898
H	-27.80698	-4.08279	-5.81498
H	-26.07187	-5.96244	-7.68480
H	-26.23809	-3.61258	-8.45010
H	-28.04448	-3.66846	-8.26153
H	-30.30398	-10.92623	-6.67049
H	-30.86819	-10.14205	-5.19327
H	-27.94638	-10.27272	-6.10482
H	-28.52621	-9.46725	-4.64453
H	-29.96164	-7.96254	-6.05043
H	-29.37849	-8.77345	-7.50625
H	-27.01817	-8.10197	-6.88814
H	-27.65318	-7.22995	-5.49796
H	-28.35384	-6.67472	-8.43348
H	-29.05881	-5.81828	-7.06168
H	11.27161	4.06079	3.55171
H	11.23901	5.83247	3.58410
H	11.27128	4.97735	2.03210
H	-28.87548	-12.46812	-5.31195
H	-29.43415	-11.68375	-3.83162
H	-31.23451	-13.09753	-5.86971
H	-31.79399	-12.31203	-4.39149
H	-30.37877	-13.87021	-3.03084

H	-29.82791	-14.65931	-4.51159
H	-32.19207	-15.26182	-5.07176
H	-32.74617	-14.46774	-3.59599
H	-31.35265	-16.03653	-2.22841
H	-30.80060	-16.83206	-3.70529
H	-33.16809	-17.41324	-4.27700
H	-33.72087	-16.61804	-2.80110
H	-32.34709	-18.19775	-1.43111
H	-31.78452	-18.99006	-2.90571
H	-34.15314	-19.55577	-3.50027
H	-34.70956	-18.77481	-2.01778
H	-33.32795	-20.36435	-0.66637
H	-32.78271	-21.15340	-2.14973
H	-35.19291	-21.63229	-2.73722
H	-35.67498	-20.93691	-1.18490
H	-33.82690	-23.35302	-1.55057
H	-34.76002	-21.98576	0.75767
H	-36.69137	-23.11383	-0.48804
H	-34.90532	-25.35433	-2.51820
H	-34.82563	-25.52813	-0.48431
H	-35.39261	-24.63821	0.91556
H	-37.26047	-26.08130	-1.02370
H	-37.72828	-25.24010	0.45562
H	-36.32681	-26.80623	1.79684
H	-35.77890	-27.60129	0.31809
H	-38.15518	-28.19430	-0.23359
H	-38.68237	-27.44111	1.27375
H	-37.24954	-29.02811	2.57008
H	-36.69948	-29.76990	1.06456
H	-39.05116	-30.38542	0.49496
H	-39.61635	-29.62872	1.98659
H	-38.19438	-31.20594	3.31884
H	-37.64706	-31.97035	1.82372
H	-40.00688	-32.56469	1.25552
H	-40.56630	-31.78110	2.73540
H	-39.17093	-33.35530	4.09494
H	-38.62201	-34.14477	2.61376
H	-40.99034	-34.72145	2.04298
H	-41.54129	-33.92680	3.52033
H	-40.15689	-35.50295	4.88594
H	-39.60503	-36.29813	3.40912
H	-41.97426	-36.87489	2.83579
H	-42.52459	-36.08142	4.31406
H	-46.87242	-47.70558	6.80327
H	-47.42314	-46.91504	8.28197
H	-44.50308	-47.12011	7.37239
H	-45.05395	-46.32939	8.85153
H	-46.43579	-44.74401	7.48922
H	-45.88478	-45.53458	6.01025
H	-43.51661	-44.94591	6.57769
H	-44.06760	-44.15541	8.05695
H	-45.45553	-42.57526	6.69548
H	-44.90439	-43.36566	5.21629
H	-42.53687	-42.77549	5.78295

```
  H    -43.08801    -41.98510      7.26233
  H    -44.47964    -40.40826      5.90132
  H    -43.92831    -41.19864      4.42202
  H    -41.56105    -40.60990      4.98992
  H    -42.11260    -39.81911      6.46894
  H    -43.50412    -38.24319      5.10752
  H    -42.95281    -39.03434      3.62844
  H    -40.58502    -38.45065      4.19953
  H    -41.13721    -37.65802      5.67739
  H    -46.05182    -48.51126      9.64624
  H    -45.50103    -49.30195      8.16729
  H    -48.41600    -49.12013      9.09692
  H    -47.37501    -50.55174      9.19096
  H    -47.86229    -49.91502      7.61009
  H     -7.75167     -6.01416     -4.63963
  H     -7.18769     -8.39184     -4.09430
  H     -6.59279     -7.89350     -2.45073
*
```

A.7. Example of .BASH script submitted to BlueBEAR:

```
#!/bin/bash

#

#SBATCH --nodes=1
#SBATCH --ntasks=16


#SBATCH --job-name=orca-Alpha-Mycolic-Acid

#SBATCH --output=output.%j.orca-Alpha-Mycolic-Acid
#SBATCH --time=7-23:59

set -e

module purge; module load bluebear

module load ORCA/4.2.1-gompi-2019b

/rds/bear-apps/2019b/EL8-has/software/ORCA/4.2.1-gompi-2019b/orca Alpha-
Mycolic-Acid.inp>& Alpha-Mycolic-Acid.out
```

A.8. Raman spectra of individual bacteria and peak assignment:

N.B: All peak assignments are based on Talari *et al. (250)*

A.8.1. *S. aureus*:

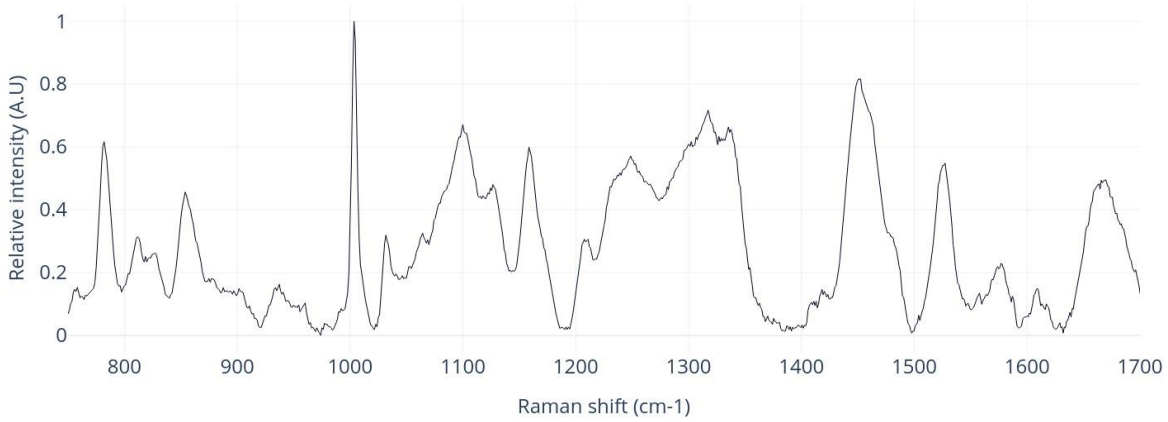


Figure 62 - Average Raman Spectra of *S. aureus*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
780	Uracil-based ring breathing mod
820	Protein band
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
950	Most probably due to single bond stretching vibrations for the amino acids proline and valine and polysaccharides
1002	Phenylalanine, C-C aromatic ring stretching
1018	Stretching C-O ribose
1065	Fatty acid
1100	C-C Vibration mode of the *gauche*-bonded chain
1123	C-C stretching mode of lipids and protein, C-N stretch glucose
1160	C-C/C-N Stretching (proteins)
1210	$C\text{-}C_6H_5$ stretching mode in tyrosine and phenylalanine
1250	Amide III
1320	Guanine (DNA/RNA), CH deformation (proteins)
1340	Nucleic acid modes
1450	CH_2 bending, or $\delta(CH_2)$
1525	Carotenoids
1580	C-C Stretching
1610	Cytosine (NH_2)
1660	Amide I, ν(C=C) *cis,* lipids, fatty acids

Table 18 - Raman peak assignment for *S. aureus*

A.8.2. *R. erythropolis*:



Figure 63 - Average Raman Spectra of *R. erythropolis*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
780	Uracil-based ring breathing mode
820	Protein band
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
938	C-C stretch backbone
1002	Phenylalanine, C-C aromatic ring stretching
1070	Triglycerides (Fatty acids)
1150	Carotenoid
1250	Amide III
1355	Guanine
1450	CH_2 bending, or $\delta(CH_2)$
1525	In-plane vibrations of the conjugated -C=C-
1575	Nucleic acid mode
1660	Amide I, ν(C=C) *cis,* lipids, fatty acids

Table 19 - Raman peak assignment for *R. erythropolis*

A.8.3. *P. fluorescens*:



Figure 64 - Average Raman Spectra of *P. fluorescens*

Raman Peak (cm^{-1})	**Functional Group / Band Assignment**
760	Trypophan *δ*(ring)
780	Uracil-based breathing mode
820	Protein band
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
920	C-C stretch of proline ring/glucose/lactic acid
1002	Phenylalanine, C-C aromatic ring stretching
1065	Palmitic acid, fatty acid, proline
1082	Nucleic acid, carbohydrates residues of collagen, carbohydrates peaks for solutions
1100	C-C Vibration mode of the *gauche*-bonded chain
1120	The strong C-O band of ribose
1140	Unassigned
1210	C-C_6H_5 stretching mode in tyrosine and phenylalanine
1240	RNA, differences in collagen content, asymmetric phosphate stretching modes
1320	Guanine, CH deformation
1337	Amide III, CH_2 wagging vibrations from glycine backbone
1409	$_{v,\ s}$COO-
1445	*δ*(CH_2), *δ*(CH_3), collagen (protein and lipid assignment)
1465	Lipids
1528	Carotenoid

1542	Amide II
1660	Amide I, ν(C=C) cis, lipids and fatty acids, C=C in unsaturated fatty acids

Table 20 - Raman peak assignment for *P. fluorescens*

A.8.4. *P. aeruginosa*:

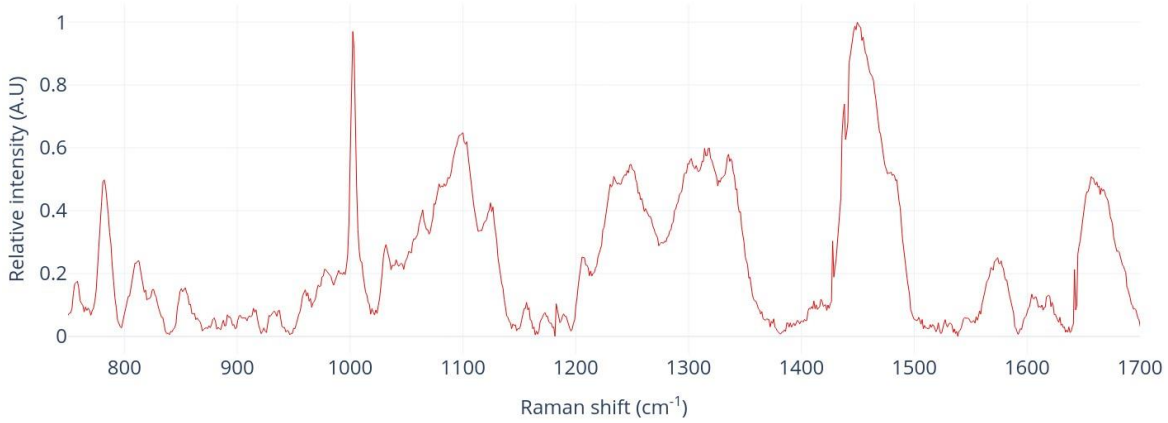


Figure 65 - Average Raman Spectra of *P. aeruginosa*

Raman Peak (cm^{-1})	**Functional Group / Band Assignment**
780	Uracil-based breathing mode
815	Proline, stretching of nucleic acid
824	Out-of-plan ring breathing, tyrosine
860	Phosphate group
960	Quinoid ring in-plane deformation
980	C-C stretching *β*-sheet (proteins)
1002	Phenylalanine, C-C aromatic ring stretching
1030	C-H bending
1060	C-C skeletal stretching
1080	Amide II, typical phospholipids
1100	C-C Vibration mode of the *gauche*-bonded chain
1120	Carotene
1210	C-C_6H_5 stretching mode in tyrosine and phenylalanine
1238	Amide III, CH_2 wagging vibrations from glycine backbone
1245	Amide III
1300	--(CH_2)n; in-plane twist vibration (lipid band), fatty acids, CH_2 twisting and wagging of lipids
1320	Guanine, CH deformation
1340	Nucleic acids
1440	Lipids, CH_2 bending

1450	CH_2 bending, or $\delta(CH_2)$
1580	C-C stretching
1600	Amide I
1620	ν(C=C)
1660	Amide I, ν(C=C) cis, lipids and fatty acids, C=C in unsaturated fatty acids

Table 21 - Raman peak assignment for *P. aeruginosa*

A.8.5. *M. hyosynoviae*:

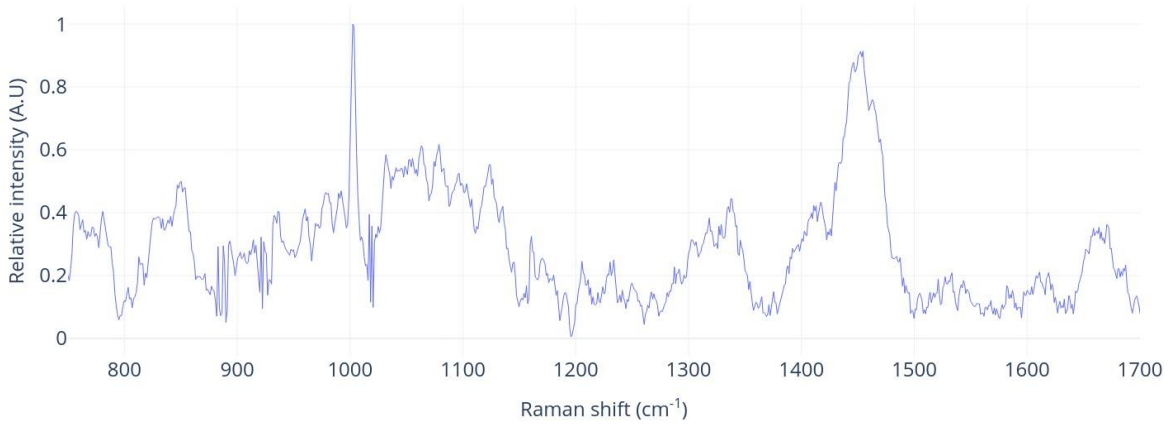


Figure 66 - Average Raman Spectra of *M. hyosynoviae*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
760	Trypophan δ(ring)
780	Uracil-based breathing mode
815	Proline, stretching of nucleic acid
840	α-anomers, glucose-saccharide band
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
898	Monosaccharides (β-glucose), C-O-C skeletal mode, adenine
938	C-C stretch backbone
960	Quinoid ring in-plane deformation
980	C-C stretching β-sheet (proteins)
996	C-O ribose, C-C
1002	Phenylalanine, C-C aromatic ring stretching
1035	Collagen
1060	C-C skeletal stretching
1080	Amide II, phospholipids
1100	C-C Vibration mode of the *gauche*-bonded chain

1120	Carotene
1160	C-C/C-N stretching (proteins)
1170	C-H in-plane bending mode of tyrosine
1210	C-C_6H_5 stretching mode in tyrosine and phenylalanine
1235	Amide III
1245	Amide III
1288	Phosphodiester groups in nucleic acids
1300	Lipids
1320	Guanine, CH deformation
1340	Nucleic acids
1420	CH_2 scissoring vibration (lipid band)
1460	CH_2/CH_3 deformation of lipids and collagen, CH_2 wagging
1520	Carotene
1532	Unassigned
1543	$_{\nu}11$ ($C_{\beta}C_{\beta}$)
1582	Δ(C=C), phenylalanine
1586	Phenylalanine
1610	Cytosine
1618	$_{\nu}$(C=C)
1650	(C=C) amide I
1658	Amide I
1669	Carbonyl stretch (C=O)

Table 22 - Raman peak assignment for *M. hyosynoviae*

A.8.6. *Mycoplasma bovis*:

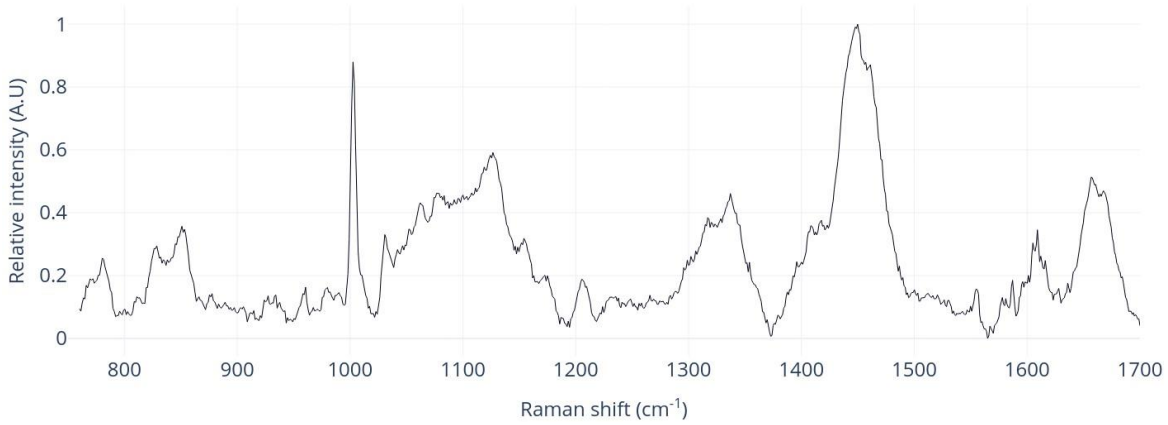


Figure 67 - Average Raman Spectra of *Mycoplasma bovis*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
780	Uracil-based ring breathing mode
830	C-H out-of-plane bending in benzoid ring
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
877	C-O-C ring (Carbohydrates), C-C-N+ (Symmetric stretching (lipids))
922	C-C stretch
934	C-C backbone (collagen assignment)
957	Carotenoid, cholesterol
980	C-C stretching β-sheet
1002	Phenylalanine, C-C aromatic ring stretching
1030	C-H bending
1053	C-O stretching, C-N stretching
1061	C-C in-plane bending
1078	C-C or C-O for lipids
1117	Glucose
1126	C-N stretching
1153	Carbohydrate peak for solutions
1171	Tyrosine
1205	Difference in collagen content
1235	Amide III
1248	Amide III
1263	Lipids
1317	Guanine
1336	Polynucleotide chain (DNA purine bases), $_{\delta}(CH_3)_{\delta}(CH_2)$ twisting, collagen (protein assignment)
1349	Unassigned
1396	B-carotene
1409	$_{v,\ s}COO^-$
1417	C=C stretching in quinoid ring
1449	C-H vibration
1460	CH_2/CH_3 deformation of lipids and collagen
1499	C-C stretching in benzenoid ring
1554	Amide II
1573	Guanine, adenine
1576	Nucleic acids
1586	Phenylalanine
1605	Cytosine
1609	Cytosine
1616	C=C stretching mode of tyrosine and tryptophan

1628	$C_{\alpha}=C_{\alpha}$ stretch
1637	Amide I
1656	C=C (lipids)
1669	Carbonyl stretch (C=O)

Table 23 - Raman peak assignment for *Mycoplasma bovis*

A.8.7. *M. bovis* BCG:

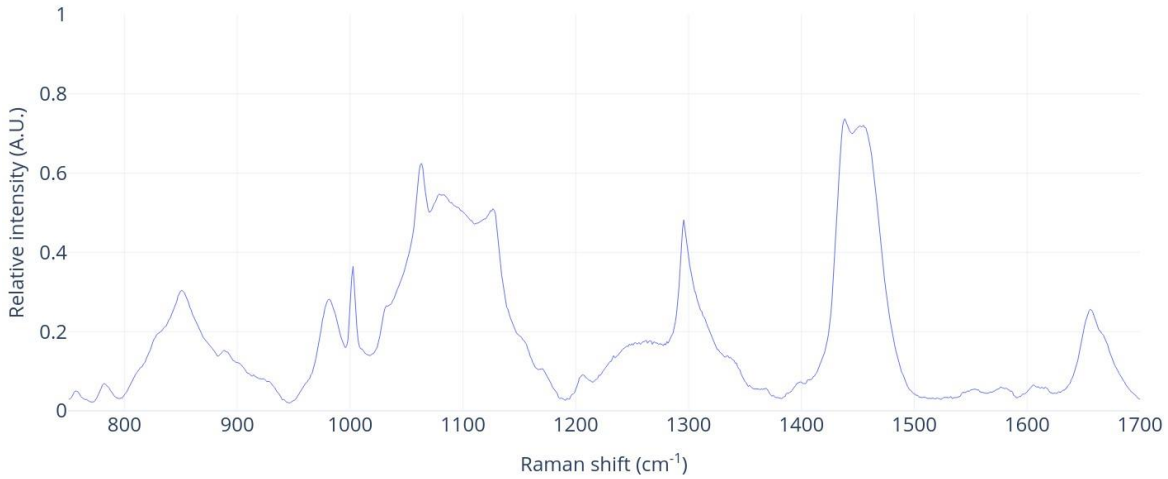


Figure 68 - Average Raman Spectra of *M. bovis* BCG

Raman Peak (cm^{-1})	Functional Group / Band Assignment
780	Uracil-based ring breathing mode
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
885	C-O-C skeletal mode
980	C-C stretching β-sheet
1002	Phenylalanine, C-C aromatic ring stretching
1061	C-C in-plane bending
1080	Amide II
1126	ν (C-C) skeletal of acyl backbone in lipid (trans conformation)
1205	Difference in collagen content
1260	Amide III, proteins, CH_2 in-plane deformation (Lipids)
1296	CH_2 deformation
1398	CH_2 deformation
1437	CH_2 deformation in lipids
1455	$\delta(CH_2)$
1552	ν (C=C), tryptophan
1575	Ring breathing modes in DNA bases

1607	C=C phenylalanine, tyrosine
1658	Amide I

Table 24 - Raman peak assignment for *M. bovis* BCG

A.8.8. *M. catarrhalis*:



Figure 69 - Average Raman Spectra of *M. catarrhalis*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
757	Unassigned
766	Pyrimidine ring breathing mode
776	Phosphatidylinositol
780	Uracil-based ring breathing mode
815	Proline, stretching of nucleic acid
847	Monosaccharides (α-glucose), (C-O-C) skeletal mode Disaccharide (maltose), (C-O-C) skeletal mode
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
890	Protein bands
915	Ribose vibration, one of the distinct RNA modes (with 915 and 974 cm^{-1})
928	$_{\nu}$(C-C), stretching; probably in amino acids proline and valine (protein band)
960	Quinoid ring in-plane deformation
980	C-C stretching β-sheet (proteins)
996	C O ribose, C C
1002	Phenylalanine, C-C aromatic ring stretching
1032	CH_2CH_3 bending modes of collagen and phospholipids, C-C skeletal stretch (one of C-C ring

	vibration to be expected in aromatic structure of xylene)
1048	Glycogen
1063	C-C skeletal stretch random conformation
1078	Lipids, ν(C-C) skeletal, C-C or C-O (lipids)
1099	C-C vibration mode of the *gauche*-bonded chain
1105	Carbohydrates peak for solutions
1124	ν(C-C) skeletal of acyl backbone in lipid (trans conformation)
1131	Fatty acid
1157	β-Carotene accumulation (C-C stretch mode)
1205	Differences in collagen content
1234	Concerted ring mode
1247	Amide III
1288	Phosphodiester groups in nucleic acids
1300	Lipid
1318	Guanine
1330	Typical phospholipids, regions associated with DNA, collagen
1335	CH_3CH_2 wagging, collagen, nucleic acids
1344	δ(CH)
1412	*vs* COO^-
1451	CH_2CH_3 deformation
1463	Fermi interaction $\delta(CH_2)$ and $\gamma(CH_2)$
1499	C-C stretching in benzenoid ring
1544	Amide II
1580	C-C stretching
1607	Tyrosine, phenylalanine ring vibration
1618	ν(C=C), tryptophan (protein assignment)
1659	Amide I
1685	Amide I

Table 25 - Raman peak assignment for *M. catarrhalis*

A.8.9. *M. haemolytica*:

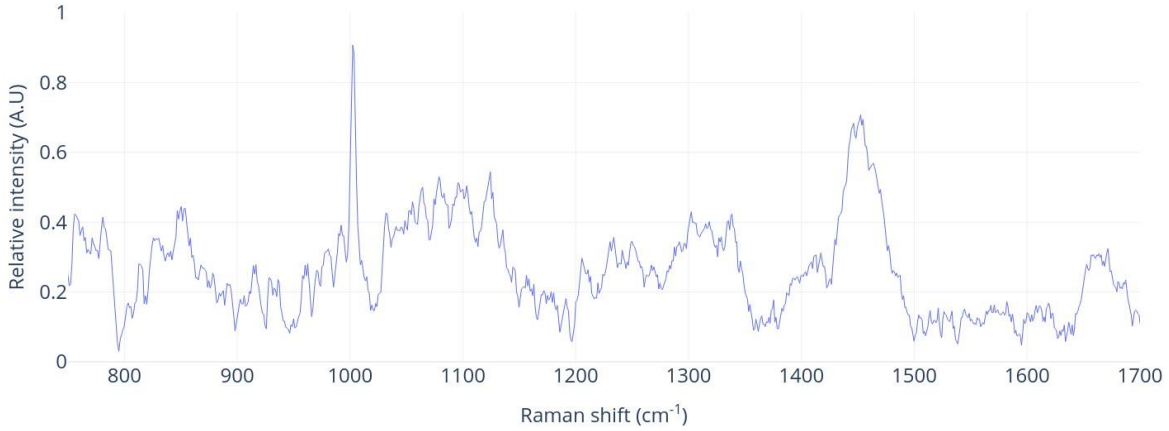


Figure 70 - Average Raman Spectra of *M. haemolytica*

Raman Peak (cm^{-1})	**Functional Group / Band Assignment**
759	Tryptophan
780	Uracil-based ring breathing mode
802	Uracil-based ring breathing mode
815	Proline, hydroxyproline, tyrosine, $_{v2}$ PO_2^- stretch of nucleic acids
826	O-P-O stretch DNA
852	Proline, hydroxyproline, tyrosine
867	Ribose vibration, one of the distinct RNA modes (with 915 and 974 cm^{-1})
877	C-C-N^+ symmetric stretching (lipids), C-O-C ring (carbohydrate)
893	Backbone, C-C skeletal
915	Ribose vibration, one of the distinct RNA modes (with 915 and 974 cm^{-1})
922	C-C stretch
937	C-C backbone (collagen assignment)
956	Carotenoids
980	C-C stretching β-sheet (proteins)
1002	Phenylalanine, C-C aromatic ring stretching
1032	CH_2CH_3 bending modes of collagen and phospholipids, C-C skeletal stretch
1057	Lipids
1061	C-C skeletal stretch random conformation
1080	Amide II
1095	Lipid, $_{v}$(C-N)

1101	O-P-O backbone stretch of DNA
1124	ν(C-C) skeletal of acyl backbone in lipid
1131	Palmitic acid, fatty acid, C-C skeletal stretching
1149	Carbohydrates peak for solids
1157	In-plane vibrations of the conjugated =C-C=, β-Carotene accumulation (C-C stretch mode)
1172	δ(CH), tyrosine (protein assignment)
1199	Tryptophan ring breathing
1205	Differences in collagen content
1230	Antisymmetric phosphate stretching vibration
1240	One of the two most distinct peaks for RNA (with 813 cm^{-1}), asymmetric phosphate [PO_2^-(asym.)] stretching modes, Collagen
1248	Amide III
1268	Δ(=C-H) (phospholipids)
1290	Cytosine
1300	Lipids
1315	Guanine
1330	Typical phospholipids, region associated with DNA and phospholipids, collagen, nucleic acids and phosphates
1339	C-C stretch of phenyl (1) and C_3-C_3stretch and C_5-O_5 stretch CH_α in-plane bend
1347	Unassigned
1360	Tryptophan
1373	Thymine, Adenine, Guanine
1392	C-N stretching, quinoid ring benzoid ring quinoid ring
1408	*vs* COO^-
1417	C=C stretching in quinoid ring
1444	$\delta(CH_2)$, lipids, fatty acids
1454	$\delta(CH_2)$, deoxyribose
1465	Lipids
1506	N-H bending
1525	In-plane vibrations of the conjugated -C=C-
1544	Amide II
1573	Guanine, adenine
1580	C-C stretching, C=C bending mode of phenylalanine
1614	Tyrosine
1618	ν(C=C), tryptophan
1662	Nucleic acids

Table 26 - Raman peak assignment for *M. haemolytica*

A.8.10. *E. coli*:

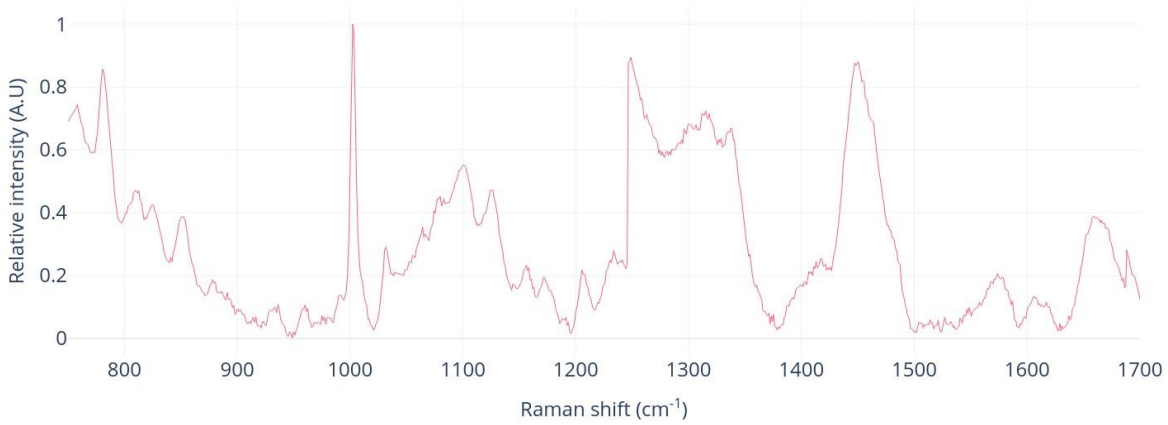


Figure 71 - Average Raman Spectra of *E. coli*

Raman Peak (cm^{-1})	**Functional Group / Band Assignment**
759	Tryptophan
780	Uracil-based ring breathing mode
811	O-P-O stretching RNA
826	O-P-O stretch DNA
852	Proline, hydroxyproline, tyrosine
877	C-C-N^+ symmetric stretching (lipids), C-O-C ring (carbohydrate)
934	C-C backbone (collagen assignment)
960	Symmetric stretching vibration of $\nu_1PO_4^{3-}$(phosphate of HA)
996	C-O ribose, C-C
1002	Phenylalanine, C-C aromatic ring stretching
1030	C-H Bending
1062	C-C Skeletal stretch random confirmation
1080	Amide II
1095	Lipid
1124	ν(C-C) skeletal of acyl backbone in lipid
1155	C-C (and C-N) stretching of proteins (also carotenoids)
1172	δ(C-H), tyrosine (protein assignment)
1204	Amide III and CH_2 wagging vibrations from glycine back bone and proline side chains
1234	A concerted ring mode
1247	Amide III
1300	Lipids

1315	Guanine
1337	CH_2/CH_3wagging, twisting and/or bending mode of collagens, lipids, nucleic acids and tryptophan
1417	C=C stretching in quinoid ring
1450	CH_2 bending
1524	Carotenoids
1548	Tryptophan
1573	Guanine, adenine
1606	C=C bending
1617	C=C phenylalanine, tyrosine
1659	Amide I
1685	Amide I

Table 27 - Raman peak assignment for *E. coli*

A.8.11. *C. glutamicum*:



Figure 72 - Average Raman Spectra of *C. glutamicum*

Raman Peak (cm^{-1})	**Functional Group / Band Assignment**
759	Tryptophan
780	Uracil-based ring breathing mode
813	C- C stretching (collagen assignment)
827	Proline, hydroxyproline, tyrosine, $_{v2}PO_2^-$ stretch of nucleic acids
850	Most probably due to single-bond stretching vibrations for the amino acids and valine and polysaccharides
918	Proline, hydroxyproline & Glycogen and lactic acid

933	Proline, hydroxyproline, ν(C-C) skeletal of collagen backbone
1002	Phenylalanine, C-C aromatic ring stretching
1031	δ(C-H), C-H in-plane bending mode of phenylalanine
1063	C-C skeletal stretch random conformation
1078	Lipids
1100	C-C vibration mode of the *gauche*-bonded chain
1124	ν(C-C) skeletal of acyl backbone in lipid (*trans*-conformation)
1154	*β*-Carotenes
1205	Differences in collagen content
1235	Amide III
1246	Amide III
1317	Guanine
1336	Purine bases, $\delta(CH_3)\delta(CH_2)$ twisting ,collagen (protein assignment
1409	*vs* COO^-
1448	CH_2CH_3 deformation, CH_2 deformation
1528	Carotenoids
1576	Nucleic acids
1602	Phenylalanine, δ(C=C),phenylalanine (protein assignment)
1618	Tryptophan
1659	Amide I

Table 28 - Raman peak assignment for *C. glutamicum*

A.8.12. *B. cereus*:

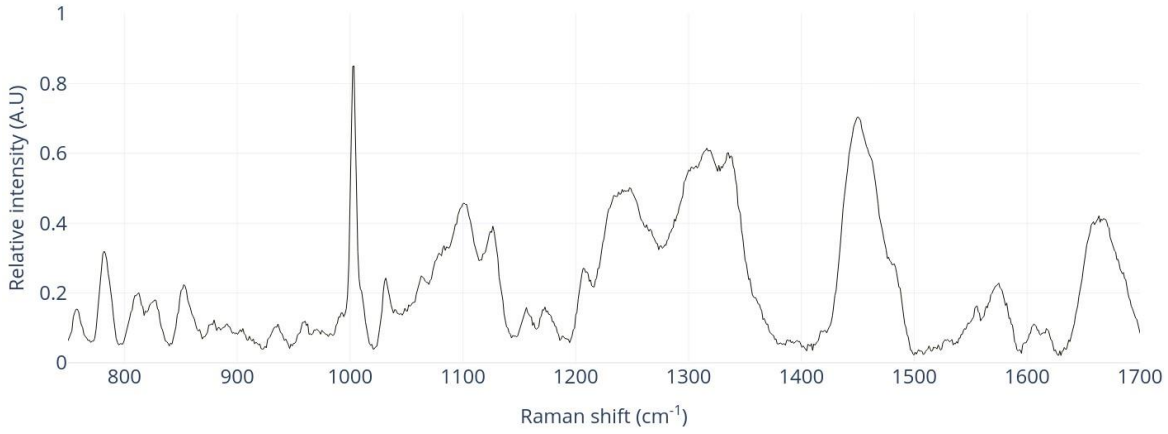


Figure 73 - Average Raman Spectra of *B. cereus*

Raman Peak (cm^{-1})	Functional Group / Band Assignment
759	Tryptophan
780	Uracil-based ring breathing mode
811	O-P-O stretching RNA
827	Proline, hydroxyproline, tyrosine, $_{v2}PO_2^-$ stretch of nucleic acids
852	Proline, hydroxyproline, tyrosine
996	C-O ribose, C-C
1002	Phenylalanine, C-C aromatic ring stretching
1031	$_{\delta}$(C-H), C-H in-plane bending mode of phenylalanine
1063	C-C skeletal stretch, random conformation
1083	C-N Stretching mode of proteins (and lipid mode of lesser degree)
1100	C-C vibration mode of the *gauche*-bonded chain
1130	C-C skeletal stretch transconformation
1156	C-C, C-N stretching (protein)
1173	Cytosine, guanine, tyrosine
1205	Difference in collagen content
1242	Amide III
1317	Guanine
1336	Purine bases, $_{\delta}(CH_3)_{\delta}(CH_2)$ twisting ,collagen (protein assignment
1447	CH_2 bending mode of proteins and lipids
1554	Amide II
1573	Guanine, adenine
1606	C=C bending
1620	$_{\nu}$(C=C), porphyrin
1663	DNA

Table 29 - Raman peak assignment for *B. cereus*

A.9 Optical Density of bacteria present in bacterial mix:

Bacteria	**OD_{600nm} 600nm (CFU/ml)**	**Measured OD_{600nm}:**	**Proportion of final mixture:**	**Reference for OD**
Mycobacterium bovis BCG	0.39 (10^6)	0.38	N/A	(361)
Mycoplasma bovis	1.00 (10^7)	0.98	20%	(298)
Mycoplasma hyosynoviae	1.00 (10^7)	0.99	20%	(298)
Pseudomonas fluorescens	1.00 (10^9)	0.99	1%	(362)
Pseudomonas aeruginosa	1.00 (10^8)	0.97	1%	(363)
Bacillus cereus	0.50 (10^9)	0.46	2%	(364)
Escherichia coli	0.62 (10^8)	0.60	2%	(365)
Mannheimia haemolytica	0.6 (10^{10})	0.61	14%	(366)
Moraxella catarrhalis	0.20 (10^7)	0.20	35%	(367)
Rhodococcus erythropolis	0.80 (10^7)	0.79	2%	(368)
Staphylococcus aureus	1.00 (10^9)	0.99	1%	(369)
Corynebacterium glutamicum	1.00 (10^8)	1.00	2%	(370)

Table 30 - Illustration of the bacteria in the bacterial mix used in Chapter 7, their recorded Optical Densities and their proportions in the final mix

References:

1. World Health Organization. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization; 2022.
2. Zink AR, Molnár E, Motamedi N, Pálfy G, Marcsik A, Nerlich AG. Molecular history of tuberculosis from ancient mummies and skeletons. International Journal of Osteoarchaeology. 2007;17(4):380-91.
3. Kanabalan RD, Lee LJ, Lee TY, Chong PP, Hassan L, Ismail R, et al. Human tuberculosis and Mycobacterium tuberculosis complex: A review on genetic diversity, pathogenesis and omics approaches in host biomarkers discovery. Microbiological Research. 2021;246:126674.
4. Glickman MS, Jacobs WR, Jr. Microbial Pathogenesis of Mycobacterium tuberculosis: Dawn of a Discipline. Cell. 2001;104(4):477-85.
5. Ahmad S. Pathogenesis, Immunology, and Diagnosis of Latent Mycobacterium tuberculosis Infection. Clinical and Developmental Immunology. 2011;2011:814943.
6. Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV Co-Infection. PLOS Pathogens. 2012;8(2):e1002464.
7. Agarwal A, Kedia S, Jain S, Gupta V, Bopanna S, Yadav DP, et al. High risk of tuberculosis during infliximab therapy despite tuberculosis screening in inflammatory bowel disease patients in India. Intest Res. 2018;16(4):588-98.
8. Dheda K, Perumal T, Moultrie H, Perumal R, Esmail A, Scott AJ, et al. The intersecting pandemics of tuberculosis and COVID-19: population-level and patient-level impact, clinical presentation, and corrective interventions. The Lancet Respiratory Medicine. 2022;10(6):603-22.
9. Collins D, Hafidz F, Mustikawati D. The economic burden of tuberculosis in Indonesia. The International Journal of Tuberculosis and Lung Disease. 2017;21(9):1041-8.
10. Ukwaja KN, Modebe O, Igwenyi C, Alobu I. The economic burden of tuberculosis care for patients and households in Africa: a systematic review [Review article]. The International Journal of Tuberculosis and Lung Disease. 2012;16(6):733-9.
11. Chandra A, Kumar R, Kant S, Parthasarathy R, Krishnan A. Direct and indirect patient costs of tuberculosis care in India. Tropical Medicine & International Health. 2020;25(7):803-12.
12. Peirse M, Houston A. Extrapulmonary tuberculosis. Medicine. 2017;45(12):747-52.
13. Vrieling F, Kostidis S, Spaink HP, Haks MC, Mayboroda OA, Ottenhoff THM, et al. Analyzing the impact of Mycobacterium tuberculosis infection on primary human macrophages by combined exploratory and targeted metabolomics. Scientific Reports. 2020;10(1):7085.
14. Gupta A, Kaul A, Tsolaki AG, Kishore U, Bhakta S. Mycobacterium tuberculosis: Immune evasion, latency and reactivation. Immunobiology. 2012;217(3):363-74.
15. Sholeye AR, Williams AA, Loots DT, Tutu van Furth AM, van der Kuip M, Mason S. Tuberculous Granuloma: Emerging Insights From Proteomics and Metabolomics. Frontiers in Neurology. 2022;13.

16. Millington KA, Gooding S, Hinks TSC, Reynolds DJM, Lalvani A. Mycobacterium tuberculosis-Specific Cellular Immune Profiles Suggest Bacillary Persistence Decades after Spontaneous Cure in Untreated Tuberculosis. The Journal of Infectious Diseases. 2010;202(11):1685-16849.
17. Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One. 2011;6(4):e17601.
18. Ferraris DM, Miggiano R, Rossi F, Rizzi M. Mycobacterium tuberculosis Molecular Determinants of Infection, Survival Strategies, and Vulnerable Targets. Pathogens. 2018;7(1):17.
19. Elkington PT, D'Armiento JM, Friedland JS. Tuberculosis immunopathology: the neglected role of extracellular matrix destruction. Sci Transl Med. 2011;3(71):71ps6.
20. Elkington P, Polak ME, Reichmann MT, Leslie A. Understanding the tuberculosis granuloma: the matrix revolutions. Trends in Molecular Medicine. 2022;28(2):143-54.
21. Lin PL, Flynn JL. Understanding latent tuberculosis: a moving target. J Immunol. 2010;185(1):15-22.
22. Khabibullina NF, Kutuzova DM, Burmistrova IA, Lyadova IV. The Biological and Clinical Aspects of a Latent Tuberculosis Infection. Tropical Medicine and Infectious Disease. 2022;7(3).
23. Sia JK, Rengarajan J. Immunology of Mycobacterium tuberculosis Infections. Microbiology Spectrum. 2019;7(4):7.4.6.
24. Moule MG, Cirillo JD. Mycobacterium tuberculosis Dissemination Plays a Critical Role in Pathogenesis. Frontiers in Cellular and Infection Microbiology. 2020;10.
25. Gopalaswamy R, Dusthackeer VNA, Kannayan S, Subbian S. Extrapulmonary Tuberculosis—An Update on the Diagnosis, Treatment and Drug Resistance. Journal of Respiration. 2021;1(2):141-64.
26. Ryndak MB, Laal S. Mycobacterium tuberculosis Primary Infection and Dissemination: A Critical Role for Alveolar Epithelial Cells. Frontiers in Cellular and Infection Microbiology. 2019;9.
27. Simieneh A, Tadesse M, Kebede W, Gashaw M, Abebe G. Combination of Xpert® MTB/RIF and DetermineTM TB-LAM Ag improves the diagnosis of extrapulmonary tuberculosis at Jimma University Medical Center, Oromia, Ethiopia. PLOS ONE. 2022;17(2):e0263172.
28. Kim YJ, Kang JY, Kim SI, Chang MS, Kim YR, Park YJ. Predictors for false-negative QuantiFERON-TB Gold assay results in patients with extrapulmonary tuberculosis. BMC Infectious Diseases. 2018;18(1):457.
29. Ulain N, Ali A, Khan M, Ullah Z, Shaheen L, Shareef N, et al. Improving diagnosis of tuberculous lymphadenitis by combination of cytomorphology and MPT64 immunostaining on cell blocks from the fine needle aspirates. PLOS ONE. 2022;17(10):e0276064.
30. Arpagaus A, Franzeck FC, Sikalengo G, Ndege R, Mnzava D, Rohacek M, et al. Extrapulmonary tuberculosis in HIV-infected patients in rural Tanzania: The prospective Kilombero and Ulanga antiretroviral cohort. PLOS ONE. 2020;15(3):e0229875.

31. Naif HM. Pathogenesis of HIV Infection. Infect Dis Rep. 2013;5(Suppl 1):e6.
32. Tsukamoto T. Hematopoietic Stem/Progenitor Cells and the Pathogenesis of HIV/AIDS. Frontiers in Cellular and Infection Microbiology. 2020;10.
33. Jelliman P, Porcellato L. HIV is Now a Manageable Long-Term Condition, But What Makes it Unique? A Qualitative Study Exploring Views About Distinguishing Features from Multi-Professional HIV Specialists in North West England. Journal of the Association of Nurses in AIDS Care. 2017;28(1):165-78.
34. Eisinger RW, Dieffenbach CW, Fauci AS. HIV Viral Load and Transmissibility of HIV Infection: Undetectable Equals Untransmittable. JAMA. 2019;321(5):451-2.
35. Livingstone SJ, Levin D, Looker HC, Lindsay RS, Wild SH, Joss N, et al. Estimated Life Expectancy in a Scottish Cohort With Type 1 Diabetes, 2008-2010. JAMA. 2015;313(1):37-44.
36. Gueler A, Moser A, Calmy A, Günthard HF, Bernasconi E, Furrer H, et al. Life expectancy in HIV-positive persons in Switzerland: matched comparison with general population. AIDS. 2017;31(3):427-36.
37. Sartorius B, VanderHeide JD, Yang M, Goosmann EA, Hon J, Haeuser E, et al. Subnational mapping of HIV incidence and mortality among individuals aged 15–49 years in sub-Saharan Africa, 2000–18: a modelling study. The Lancet HIV. 2021;8(6):e363-e75.
38. Marukutira T, Scott N, Kelly SL, Birungi C, Makhema JM, Crowe S, et al. Modelling the impact of migrants on the success of the HIV care and treatment program in Botswana. PLOS ONE. 2020;15(1):e0226422.
39. Modisenyane SM, Hendricks SJH, Fineberg H. Understanding how domestic health policy is integrated into foreign policy in South Africa: a case for accelerating access to antiretroviral medicines. Global Health Action. 2017;10(1):1339533.
40. Wong NS, Leung CC, Chan KCW, Chan WK, Lin AWC, Lee SS. A longitudinal study on latent TB infection screening and its association with TB incidence in HIV patients. Scientific Reports. 2019;9(1):10093.
41. Shivakoti R, Sharma D, Mamoon G, Pham K. Association of HIV infection with extrapulmonary tuberculosis: a systematic review. Infection. 2017;45(1):11-21.
42. da Silva Escada RO, Velasque L, Ribeiro SR, Cardoso SW, Marins LMS, Grinsztejn E, et al. Mortality in patients with HIV-1 and tuberculosis co-infection in Rio de Janeiro, Brazil - associated factors and causes of death. BMC Infectious Diseases. 2017;17(1):373.
43. Araújo-Mariz C, Lopes EP, Acioli-Santos B, Maruza M, Montarroyos UR, Ximenes RAdA, et al. Hepatotoxicity during Treatment for Tuberculosis in People Living with HIV/AIDS. PLOS ONE. 2016;11(6):e0157725.
44. Salindri AD, Auld SC, Schechter MC, Gandhi NR, Magee MJ. Negative tuberculin skin test result predicts all-cause mortality among tuberculosis patients with HIV and diabetes comorbidity. Annals of Epidemiology. 2019;33:72-8.e4.
45. Carranza C, Pedraza-Sanchez S, de Oyarzabal-Mendez E, Torres M. Diagnosis for Latent Tuberculosis Infection: New Alternatives. Frontiers in Immunology. 2020;11.
46. Anku PJ, Amo-Adjei J, Doku D, Kumi-Kyereme A. Challenges of scaling-up of TB-HIV integrated service delivery in Ghana. PLOS ONE. 2020;15(7):e0235843.

47. Jun M, Squires DA, Osborn R, Thomson S, editors. International profiles of health care systems, 20122012.
48. Bruchfeld J, Correia-Neves M, Källenius G. Tuberculosis and HIV Coinfection. Cold Spring Harb Perspect Med. 2015;5(7):a017871.
49. Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, et al. Physiology of Mycobacteria. In: Poole RK, editor. Advances in Microbial Physiology. 55: Academic Press; 2009. p. 81-319.
50. Jankute M, Cox JAG, Harrison J, Besra GS. Assembly of the Mycobacterial Cell Wall. Annual Review of Microbiology. 2015;69(1):405-23.
51. Favrot L, Ronning DR. Targeting the mycobacterial envelope for tuberculosis drug development. Expert Review of Anti-infective Therapy. 2012;10(9):1023-36.
52. Chiaradia L, Lefebvre C, Parra J, Marcoux J, Burlet-Schiltz O, Etienne G, et al. Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. Scientific Reports. 2017;7(1):12807.
53. Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiology Reviews. 2008;32(2):234-58.
54. Pfyffer GE, Palicova F. Mycobacterium : General Characteristics, Laboratory Detection, and Staining Procedures. Manual of Clinical Microbiology2011. p. 472-502.
55. Furtos NC, Doane EA. Transmission of Streptomycin-Resistant Tubercle Bacilli in Man. Journal of the American Medical Association. 1949;140(16):1274-5.
56. Riva MA. From milk to rifampicin and back again: history of failures and successes in the treatment for tuberculosis. The Journal of Antibiotics. 2014;67(9):661-5.
57. Johnston JC, Cooper R, Menzies D. Chapter 5: Treatment of tuberculosis disease. Canadian Journal of Respiratory, Critical Care, and Sleep Medicine. 2022;6(sup1):66-76.
58. Gill CM, Dolan L, Piggott LM, McLaughlin AM. New developments in tuberculosis diagnosis and treatment. Breathe. 2022;18(1):210149.
59. Siapka M, Vassall A, Cunnama L, Pineda C, Cerecero D, Sweeney S, et al. Cost of tuberculosis treatment in low- and middle-income countries: systematic review and meta-regression. The International Journal of Tuberculosis and Lung Disease. 2020;24(8):802-10.
60. Auld SC, Gandhi NR, Brust JCM. Drug-Resistant Tuberculosis and HIV. In: Sereti I, Bisson GP, Meintjes G, editors. HIV and Tuberculosis: A Formidable Alliance. Cham: Springer International Publishing; 2019. p. 203-37.
61. Du Preez I, Loots DT. Novel insights into the pharmacometabonomics of first-line tuberculosis drugs relating to metabolism, mechanism of action and drug-resistance. Drug Metabolism Reviews. 2018;50(4):466-81.
62. Jindani A, Aber VR, Edwards EA, Mitchison DA. The Early Bactericidal Activity of Drugs in Patients with Pulmonary Tuberculosis. American Review of Respiratory Disease. 1980;121(6):939-49.
63. Gold B, Nathan C. Targeting Phenotypically Tolerant Mycobacterium tuberculosis. Microbiology Spectrum. 2017;5(1):10.1128/microbiolspec.tbtb2-0031-2016.

64. Sherman DR, Gagneux S. Estimating the mutation rate of Mycobacterium tuberculosis during infection. Nature Genetics. 2011;43(5):400-1.
65. Pourakbari B, Mamishi S, Mohammadzadeh M, Mahmoudi S. First-Line Anti-Tubercular Drug Resistance of Mycobacterium tuberculosis in IRAN: A Systematic Review. Frontiers in Microbiology. 2016;7.
66. Günther G. Multidrug-resistant and extensively drug-resistant tuberculosis: a review of current concepts and future challenges. Clinical Medicine. 2014;14(3):279-85.
67. van den Hof S, Collins D, Hafidz F, Beyene D, Tursynbayeva A, Tiemersma E. The socioeconomic impact of multidrug resistant tuberculosis on patients: results from Ethiopia, Indonesia and Kazakhstan. BMC Infectious Diseases. 2016;16(1):470.
68. Manshina AV, Toskin I, Khalfin RA, Madianova VV, Getahun H, Blondeel K. Epidemiology of HIV-associated tuberculosis in Eastern European countries and the Russian Federation in 2004-2014: A review of literature. Profilakticheskaya Meditsina. 2017;20(1):50-6.
69. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, et al. Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal. PLOS Medicine. 2015;12(9):e1001880.
70. Cohen KA, Manson AL, Abeel T, Desjardins CA, Chapman SB, Hoffner S, et al. Extensive global movement of multidrug-resistant M. tuberculosis strains revealed by whole-genome analysis. Thorax. 2019;74(9):882-9.
71. Geric C, Saroufim M, Landsman D, Richard J, Benedetti A, Batt J, et al. Impact of COVID-19 on Tuberculosis Prevention and Treatment in Canada: A Multicenter Analysis of 10 833 Patients. The Journal of Infectious Diseases. 2021;225(8):1317-20.
72. Lakoh S, Jiba DF, Baldeh M, Adekanmbi O, Barrie U, Seisay AL, et al. Impact of COVID-19 on Tuberculosis Case Detection and Treatment Outcomes in Sierra Leone. Tropical Medicine and Infectious Disease. 2021;6(3):154.
73. Chilot D, Woldeamanuel Y, Manyazewal T. Real-Time Impact of COVID-19 on Clinical Care and Treatment of Patients with Tuberculosis: A Multicenter Cross-Sectional Study in Addis Ababa, Ethiopia. Annals of Global Health. 2021.
74. MacGregor-Fairlie M, Wilkinson S, Besra GS, Goldberg Oppenheimer P. Tuberculosis diagnostics: overcoming ancient challenges with modern solutions. Emerging Topics in Life Sciences. 2020;4(4):435-48.
75. Mukherjee S, Perveen S, Negi A, Sharma R. Evolution of tuberculosis diagnostics: From molecular strategies to nanodiagnostics. Tuberculosis. 2023;140:102340.
76. Ganesan K, Mwesigwa R, Dear N, Esber AL, Reed D, Kibuuka H, et al. Epidemiology of Tuberculosis Among People Living With HIV in the African Cohort Study From 2013 to 2021. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2023;92(5):359-69.
77. Gong W, Wu X. Differential Diagnosis of Latent Tuberculosis Infection and Active Tuberculosis: A Key to a Successful Tuberculosis Control Strategy. Frontiers in Microbiology. 2021;12.

78. Sartoris G, Seddon JA, Rabie H, Nel ED, Schaaf HS. Abdominal Tuberculosis in Children: Challenges, Uncertainty, and Confusion. Journal of the Pediatric Infectious Diseases Society. 2020;9(2):218-27.
79. Maehira Y, Spencer RC. Harmonization of Biosafety and Biosecurity Standards for High-Containment Facilities in Low- and Middle-Income Countries: An Approach From the Perspective of Occupational Safety and Health. Frontiers in Public Health. 2019;7.
80. Kouriba B, Ouwe Missi Oukem-Boyer O, Traoré B, Touré A, Raskine L, Babin FX. Installing biosafety level 3 containment laboratories in low- and middle-income countries: challenges and prospects from Mali's experience. New Microbes and New Infections. 2018;26:S74-S7.
81. Abria SA, Kasaevab T, Miglioric GB, Golettid D, Zennerf D, Denholmg J, et al., editors. Tools to implement the World Health Organization End TB Strategy: Addressing common challenges in high and low endemic countries International Journal of Infectious Diseases2020.
82. Du J, Shu W, Liu Y, Wang Y, Zhan Y, Yu K, et al. Development and validation of external quality assessment panels for mycobacterial culture testing to diagnose tuberculosis in China. European Journal of Clinical Microbiology & Infectious Diseases. 2019;38(10):1961-8.
83. Feng Z, Bai X, Wang T, Garcia C, Bai A, Li L, et al. Differential Responses by Human Macrophages to Infection With Mycobacterium tuberculosis and Non-tuberculous Mycobacteria. Frontiers in microbiology. 2020;11:116-.
84. Ma Y, Fan J, Li S, Dong L, Li Y, Wang F, et al. Comparison of Lowenstein-Jensen medium and MGIT culture system for recovery of Mycobacterium tuberculosis from abscess samples. Diagnostic Microbiology and Infectious Disease. 2020;96(4):114969.
85. Kenaope L, Ferreira H, Seedat F, Otwombe K, Martinson NA, Variava E. Sputum culture and drug sensitivity testing outcome among X-pert Mycobacterium tuberculosis/rifampicin-positive, rifampicin-resistant sputum: A retrospective study — Not all rifampicin resistance is multi-drug resistant. Journal of Global Antimicrobial Resistance. 2020;21:434-8.
86. Nambiar R, Chatellier S, Bereksi N, van Belkum A, Singh N, Barua B, et al. Evaluation of Mycotube, a modified version of Lowenstein-Jensen (LJ) medium, for efficient recovery of Mycobacterium tuberculosis (MTB). Eur J Clin Microbiol Infect Dis. 2017;36(10):1981-8.
87. Advisory Committee on Dangerous Pathogens. The Approved List of biological agents. In: Department for Environment Food & Rural Affairs AaPHA, Department of Health and Social Care, editor. 3rd ed. London, UK: HM Government; 2013.
88. Maehira Y, Spencer RC. Harmonization of Biosafety and Biosecurity Standards for High-Containment Facilities in Low- and Middle-Income Countries: An Approach From the Perspective of Occupational Safety and Health. Front Public Health. 2019;7:249-.
89. Ahmad M, Ibrahim WH, Sarafandi SA, Shahzada KS, Ahmed S, Haq IU, et al. Diagnostic value of bronchoalveolar lavage in the subset of patients with negative sputum/smear and mycobacterial culture and a suspicion of pulmonary tuberculosis. International Journal of Infectious Diseases. 2019;82:96-101.

90. Kim YW, Kwon BS, Lim SY, Lee YJ, Cho YJ, Yoon HI, et al. Diagnostic value of bronchoalveolar lavage and bronchial washing in sputum-scarce or smear-negative cases with suspected pulmonary tuberculosis: a randomized study. Clinical Microbiology and Infection. 2020;26(7):911-6.
91. Neves CP, Costa AG, Safe IP, de Souza Brito A, Jesus JS, Kritski AL, et al. The role of mini-bronchoalveolar lavage fluid in the diagnosis of pulmonary tuberculosis in critically ill patients. BMC Infectious Diseases. 2020;20(1):229.
92. Chadha VK, Anjinappa SM, Rade K, Baskaran D, Narang P, Kolappan C, et al. Sensitivity and specificity of screening tools and smear microscopy in active tuberculosis case finding. Indian Journal of Tuberculosis. 2019;66(1):99-104.
93. Kunkel A, Abel Zur Wiesch P, Nathavitharana RR, Marx FM, Jenkins HE, Cohen T. Smear positivity in paediatric and adult tuberculosis: systematic review and meta-analysis. BMC infectious diseases. 2016;16:282-.
94. Getahun H, Gunneberg C, Granich R, Nunn P. HIV Infection—Associated Tuberculosis: The Epidemiology and the Response. Clinical Infectious Diseases. 2010;50(Supplement_3):S201-S7.
95. Park JH, Choe J, Bae M, Choi S, Jung KH, Kim MJ, et al. Clinical Characteristics and Radiologic Features of Immunocompromised Patients With Pauci-Bacillary Pulmonary Tuberculosis Receiving Delayed Diagnosis and Treatment. Open Forum Infect Dis. 2019;6(2):ofz002-ofz.
96. Qin ZZ, Sander MS, Rai B, Titahong CN, Sudrungrot S, Laah SN, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep. 2019;9(1):15000-.
97. Ruhwald M, Aggerbeck H, Gallardo RV, Hoff ST, Villate JI, Borregaard B, et al. Safety and efficacy of the C-Tb skin test to diagnose Mycobacterium tuberculosis infection, compared with an interferon γ release assay and the tuberculin skin test: a phase 3, double-blind, randomised, controlled trial. The Lancet Respiratory Medicine. 2017;5(4):259-68.
98. Saktiawati AMI, Subronto YW, Stienstra Y, Sumardi, Supit F, van der Werf TS. Sensitivity and specificity of routine diagnostic work-up for tuberculosis in lung clinics in Yogyakarta, Indonesia: a cohort study. BMC Public Health. 2019;19(1):363.
99. Sathitratanacheewin S, Sunanta P, Pongpirul K. Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability. Heliyon. 2020;6(8):e04614.
100. Pepple KL, Van Gelder R, Forooghian F. Caveats about QuantiFERON-TB gold in-tube testing for uveitis. Am J Ophthalmol. 2014;157(4):752-3.
101. Silva ML, Cá B, Osório NS, Rodrigues PNS, Maceiras AR, Saraiva M. Tuberculosis caused by Mycobacterium africanum: Knowns and unknowns. PLoS Pathog. 2022;18(5):e1010490.
102. Yamasue M, Komiya K, Usagawa Y, Umeki K, Nureki S-i, Ando M, et al. Factors associated with false negative interferon-γ release assay results in patients with tuberculosis: A systematic review with meta-analysis. Scientific Reports. 2020;10(1):1607.

103. Quintana-Ortega C, Mendez-Echevarria A, del Rosal T, Gonzalez-Muñoz M, Baquero-Artigao F. False-positive Results of Quantiferon-Tb-Gold Assay in Children. The Pediatric Infectious Disease Journal. 2020;39(7):620-3.
104. Paton NI, Borand L, Benedicto J, Kyi MM, Mahmud AM, Norazmi MN, et al. Diagnosis and management of latent tuberculosis infection in Asia: Review of current status and challenges. International Journal of Infectious Diseases. 2019;87:21-9.
105. Thapa J, Maharjan B, Malla M, Fukushima Y, Poudel A, Pandey BD, et al. Direct detection of Mycobacterium tuberculosis in clinical samples by a dry methyl green loop-mediated isothermal amplification (LAMP) method. Tuberculosis. 2019;117:1-6.
106. Toonkomdang S, Phinyo P, Phetsuksiri B, Patumanond J, Rudeeaneksin J, Klayut W. Pragmatic accuracy of an in-house loop-mediated isothermal amplification (LAMP) for diagnosis of pulmonary tuberculosis in a Thai community hospital. PLOS ONE. 2020;15(7):e0236496.
107. Gelaw B, Shiferaw Y, Alemayehu M, Bashaw AA. Comparison of loop-mediated isothermal amplification assay and smear microscopy with culture for the diagnostic accuracy of tuberculosis. BMC infectious diseases. 2017;17(1):79-.
108. World Health Organization. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization; 2019.
109. Yasmin M, Refregier G, Siddiqui RT, Iqbal R, Abbasi SA, Tahseen S. Reverse line probe assay for cheap detection of Single Nucleotide Polymorphisms in Mycobacterium tuberculosis. Tuberculosis. 2018;110:52-5.
110. Singh BK, Sharma SK, Sharma R, Sreenivas V, Myneedu VP, Kohli M, et al. Diagnostic utility of a line probe assay for multidrug resistant-TB in smear-negative pulmonary tuberculosis. PloS one. 2017;12(8):e0182988-e.
111. Bjerrum S, Schiller I, Dendukuri N, Kohli M, Nathavitharana RR, Zwerling AA, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV. Cochrane Database of Systematic Reviews. 2019(10).
112. Banaei N, Gaur RL, Pai M. Interferon Gamma Release Assays for Latent Tuberculosis: What Are the Sources of Variability? J Clin Microbiol. 2016;54(4):845-50.
113. Ayubi E, Doosti-Irani A, Sanjari Moghaddam A, Sani M, Nazarzadeh M, Mostafavi E. The Clinical Usefulness of Tuberculin Skin Test versus Interferon-Gamma Release Assays for Diagnosis of Latent Tuberculosis in HIV Patients: A Meta-Analysis. PLOS ONE. 2016;11(9):e0161983.
114. Benachinmardi KK, Sangeetha S, Rao M, Prema R. Validation and Clinical Application of Interferon-Gamma Release Assay for Diagnosis of Latent Tuberculosis Infection in Children. Int J Appl Basic Med Res. 2019;9(4):241-5.
115. National Institute for Health and Care Excellence. Tuberculosis (NG33). London, UK: National Institute for Health and Care Excellence; 2016.
116. Subbaraman R, Nathavitharana RR, Satyanarayana S, Pai M, Thomas BE, Chadha VK, et al. The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. PLOS Medicine. 2016;13(10):e1002149.

117. Nambiar R, Bereksi N, Gonzalez R, de Cozar A, Loubet M, Shetty A, et al. Performance of bioMérieux Lowenstein–Jensen slopes in plastic tube packaging, compared to existing phenotypic methods, for efficient recovery of the Mycobacterium tuberculosis complex. Journal of Medical Microbiology. 2019;68(3):398-401.
118. Cruciani M, Scarparo C, Malena M, Bosco O, Serpelloni G, Mengoli C. Meta-analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without solid media, for detection of mycobacteria. J Clin Microbiol. 2004;42(5):2321-5.
119. Davis JL, Cattamanchi A, Cuevas LE, Hopewell PC, Steingart KR. Diagnostic accuracy of same-day microscopy versus standard microscopy for pulmonary tuberculosis: a systematic review and meta-analysis. The Lancet Infectious Diseases. 2013;13(2):147-54.
120. Piccazzo R, Paparo F, Garlaschi G. Diagnostic Accuracy of Chest Radiography for the Diagnosis of Tuberculosis (TB) and Its Role in the Detection of Latent TB Infection: a Systematic Review. The Journal of Rheumatology. 2014;91:32-40.
121. De Keyser E, De Keyser F, De Baets F. Tuberculin skin test versus interferon-gamma release assays for the diagnosis of tuberculosis infection. Acta Clinica Belgica. 2014;69(5):358-66.
122. Li S, Liu B, Peng M, Chen M, Yin W, Tang H, et al. Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. PLOS ONE. 2017;12(7):e0180725.
123. Nagai K, Horita N, Yamamoto M, Tsukahara T, Nagakura H, Tashiro K, et al. Diagnostic test accuracy of loop-mediated isothermal amplification assay for Mycobacterium tuberculosis: systematic review and meta-analysis. Sci Rep. 2016;6(1):39090.
124. Nathavitharana RR, Cudahy PGT, Schumacher SG, Steingart KR, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. European Respiratory Journal. 2017;49(1):1601075.
125. Minion J, Leung E, Talbot E, Dheda K, Pai M, Menzies D. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. European Respiratory Journal. 2011;38(6):1398-405.
126. Nyaruaba R, Mwaliko C, Kering KK, Wei H. Droplet digital PCR applications in the tuberculosis world. Tuberculosis. 2019;117:85-92.
127. Rigouts L, Miotto P, Schats M, Lempens P, Cabibbe AM, Galbiati S, et al. Fluoroquinolone heteroresistance in Mycobacterium tuberculosis: detection by genotypic and phenotypic assays in experimentally mixed populations. Scientific Reports. 2019;9(1):11760.
128. Luo J, Luo M, Li J, Yu J, Yang H, Yi X, et al. Rapid direct drug susceptibility testing of Mycobacterium tuberculosis based on culture droplet digital polymerase chain reaction. The International Journal of Tuberculosis and Lung Disease. 2019;23(2):219-25.
129. Yang J, Han X, Liu A, Bai X, Xu C, Bao F, et al. Use of Digital Droplet PCR to Detect Mycobacterium tuberculosis DNA in Whole Blood-Derived DNA Samples from Patients with

Pulmonary and Extrapulmonary Tuberculosis. Frontiers in cellular and infection microbiology. 2017;7:369-.
130. Kuypers J, Jerome KR. Applications of Digital PCR for Clinical Microbiology. J Clin Microbiol. 2017;55(6):1621-8.
131. Wagner K, Fox P, Gordon E, Hahn W, Olsen K, Markham A, et al. A multiplexed, paired-pooled droplet digital PCR assay for detection of SARS-CoV-2 in saliva. Scientific Reports. 2023;13(1):3075.
132. Li S-Y, Cheng Q-X, Liu J-K, Nie X-Q, Zhao G-P, Wang J. CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. Cell Research. 2018;28(4):491-3.
133. Ai J-W, Zhou X, Xu T, Yang M, Chen Y, He G-Q, et al. CRISPR-based rapid and ultra-sensitive diagnostic test for Mycobacterium tuberculosis. Emerging microbes & infections. 2019;8(1):1361-9.
134. Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science. 2018;360(6387):439-44.
135. Coros A, DeConno E, Derbyshire KM. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. J Bacteriol. 2008;190(9):3408-10.
136. Sam IK, Chen Y-y, Ma J, Li S-y, Ying R-y, Li L-x, et al. TB-QUICK: CRISPR-Cas12b-assisted rapid and sensitive detection of Mycobacterium tuberculosis. Journal of Infection. 2021;83(1):54-60.
137. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
138. Orjuela-Cañón AD, Jutinico AL, Awad C, Vergara E, Palencia A. Machine learning in the loop for tuberculosis diagnosis support. Frontiers in Public Health. 2022;10.
139. Traynor D, Behl I, O'Dea D, Bonnier F, Nicholson S, O'Connell F, et al. Raman spectral cytopathology for cancer diagnostic applications. Nature Protocols. 2021;16(7):3716-35.
140. Wang L, Liu W, Tang J-W, Wang J-J, Liu Q-H, Wen P-B, et al. Applications of Raman Spectroscopy in Bacterial Infections: Principles, Advantages, and Shortcomings. Frontiers in Microbiology. 2021;12.
141. Ngabonziza JCS, Decroo T, Migambi P, Habimana YM, Van Deun A, Meehan CJ, et al. Prevalence and drivers of false-positive rifampicin-resistant Xpert MTB/RIF results: a prospective observational study in Rwanda. The Lancet Microbe. 2020;1(2):e74-e83.
142. Genoscreen Deeplex Myc-TB. Deeplex Myc-TB: From clinical samples to drug resistance profiles Lille, France: Genoscreen; 2020 [cited 2020 14/08]. Available from: https://www.genoscreen.fr/images/genoscreen-services/deeplex/technical_note_20200706_CE.pdf.
143. Makhado NA, Matabane E, Faccin M, Pinçon C, Jouet A, Boutachkourt F, et al. Outbreak of multidrug-resistant tuberculosis in South Africa undetected by WHO-endorsed commercial tests: an observational study. The Lancet Infectious Diseases. 2018;18(12):1350-9.

144. Tafess K, Ng TTL, Lao HY, Leung KSS, Tam KKG, Rajwani R, et al. Targeted-Sequencing Workflows for Comprehensive Drug Resistance Profiling of Mycobacterium tuberculosis Cultures Using Two Commercial Sequencing Platforms: Comparison of Analytical and Diagnostic Performance, Turnaround Time, and Cost. Clinical Chemistry. 2020;66(6):809-20.
145. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. Emerg Infect Dis. 2016;22(2):331-4.
146. Gupta RK, Turner CT, Venturini C, Esmail H, Rangaka MX, Copas A, et al. Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. The Lancet Respiratory Medicine. 2020;8(4):395-406.
147. Saktiawati AMI, Stienstra Y, Subronto YW, Rintiswati N, Sumardi, Gerritsen J-W, et al. Sensitivity and specificity of an electronic nose in diagnosing pulmonary tuberculosis among patients with suspected tuberculosis. PLOS ONE. 2019;14(6):e0217963.
148. Kaewseekhao B, Nuntawong N, Eiamchai P, Roytrakul S, Reechaipichitkul W, Faksri K. Diagnosis of active tuberculosis and latent tuberculosis infection based on Raman spectroscopy and surface-enhanced Raman spectroscopy. Tuberculosis. 2020;121:101916.
149. Eloi P, Nascimento GA, Córdula C, Visani V, Castelletti H, Bezerra G, et al. Toward a point-of-care diagnostic for specific detection of Mycobacterium tuberculosis from sputum samples. Tuberculosis. 2020;121:101919.
150. Chakaya J, Khan M, Ntoumi F, Aklillu E, Fatima R, Mwaba P, et al. Global Tuberculosis Report 2020 – Reflections on the Global TB burden, treatment and prevention efforts. International Journal of Infectious Diseases. 2021;113:S7-S12.
151. Nachega JB, Kapata N, Sam-Agudu NA, Decloedt EH, Katoto PDMC, Nagu T, et al. Minimizing the impact of the triple burden of COVID-19, tuberculosis and HIV on health services in sub-Saharan Africa. International Journal of Infectious Diseases. 2021;113:S16-S21.
152. Ritchie HS, Fiona; Roser, Max. Causes of Death 2018 [updated 12.2019. Available from: https://ourworldindata.org/causes-of-death.
153. Leonso A-A, Brown K, Prol R, Rawat S, Khunger A, Bromberg R. A Rare Case of Latent Tuberculosis Reactivation Secondary to a COVID-19 Infection. Infectious Disease Reports. 2022;14(3):446-52.
154. Darnton-Hill I, Mandal PP, de Silva A, Bhatia V, Sharma M. Opportunities to prevent and manage undernutrition to amplify efforts to end TB. Int J Tuberc Lung Dis. 2022;26(1):6-11.
155. Janssens J-P, Rieder HL. An ecological analysis of incidence of tuberculosis and per capita gross domestic product. European Respiratory Journal. 2008;32(5):1415-6.
156. Domingo M, Vidal E, Marco A. Pathology of bovine tuberculosis. Research in Veterinary Science. 2014;97:S20-S9.
157. de la Rua-Domenech R, Goodchild AT, Vordermeier HM, Hewinson RG, Christiansen KH, Clifton-Hadley RS. Ante mortem diagnosis of tuberculosis in cattle: A review of the

tuberculin tests, γ-interferon assay and other ancillary diagnostic techniques. Research in Veterinary Science. 2006;81(2):190-210.
158. Brooks-Pollock E, Roberts GO, Keeling MJ. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature. 2014;511(7508):228-31.
159. Loria GR, Ruocco L, Ciaccio G, Iovino F, Nicholas RAJ, Borrello S. The Implications of EU Regulation 2016/429 on Neglected Diseases of Small Ruminants including Contagious Agalactia with Particular Reference to Italy. Animals (Basel). 2020;10(5).
160. Hogarth PS, Jason. New APHA tools for the fight against bovine TB: Animal and Plant Health Agency; 2023 [Available from: https://aphascience.blog.gov.uk/2023/03/24/tb-day-2023/.
161. Vial F, Miguel E, T. Johnston W, Mitchell A, Donnelly CA. Bovine Tuberculosis Risk Factors for British Herds Before and After the 2001 Foot-and-Mouth Epidemic: What have we Learned from the TB99 and CCS2005 Studies? Transboundary and Emerging Diseases. 2015;62(5):505-15.
162. Cressey D. UK official defends badger cull. Nature. 2013.
163. Palmer MV. Mycobacterium bovis: Characteristics of Wildlife Reservoir Hosts. Transboundary and Emerging Diseases. 2013;60(s1):1-13.
164. Langton TES, Jones MW, McGill I. Analysis of the impact of badger culling on bovine tuberculosis in cattle in the high-risk area of England, 2009–2020. Veterinary Record. 2022;190(6):e1384.
165. Luca S, Mihăescu T. History of BCG Vaccine. Maedica. 2013;8 1:53-8.
166. Katelaris AL, Jackson C, Southern J, Gupta RK, Drobniewski F, Lalvani A, et al. Effectiveness of BCG Vaccination Against Mycobacterium tuberculosis Infection in Adults: A Cross-sectional Analysis of a UK-Based Cohort. The Journal of Infectious Diseases. 2019;221(1):146-55.
167. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. Jama. 1994;271(9):698-702.
168. Barreto ML, Pereira SM, Ferreira AA. BCG vaccine: efficacy and indications for vaccination and revaccination. Jornal de pediatria. 2006;82 3 Suppl:S45-54.
169. Vítková E, Galliová J, Křepela K, Kubín M. Adverse reactions to BCG. Central European journal of public health. 1995;3 3:138-41.
170. Schrager LK, Vekemens J, Drager N, Lewinsohn DM, Olesen OF. The status of tuberculosis vaccine development. The Lancet Infectious Diseases. 2020;20(3):e28-e37.
171. Yang T, Zhong J, Zhang J, Li C, Yu X, Xiao J, et al. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. Frontiers in Microbiology. 2018;9.
172. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope — a moving target. Nature Reviews Microbiology. 2020;18(1):47-59.

173. Mühlig A, Bocklitz T, Labugger I, Dees S, Henk S, Richter E, et al. LOC-SERS: A Promising Closed System for the Identification of Mycobacteria. Analytical Chemistry. 2016;88(16):7998-8004.
174. Collins MD, Goodfellow M, Minnikin DE. A Survey of the Structures of Mycolic Acids in Corynebacterium and Related Taxa. Microbiology. 1982;128(1):129-49.
175. Alderwick LJ, Harrison J, Lloyd GS, Birch HL. The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan. Cold Spring Harbor Perspectives in Medicine. 2015;5(8).
176. Su CC, Klenotic PA, Bolla JR, Purdy GE, Robinson CV, Yu EW. MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. Proceedings of the National Academy of Sciences of the United States of America. 2019;166(23):11241-6.
177. Minnikin DE, Brennan PJ. Lipids of Clinically Significant Mycobacteria. In: Goldfine H, editor. Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids. Cham: Springer International Publishing; 2020. p. 33-108.
178. Welsh KJ, Hunter RL, Actor JK. Trehalose 6,6′ -dimycolate – A coat to regulate tuberculosis immunopathogenesis. Tuberculosis. 2013;93:S3-S9.
179. Stöckel S, Meisel S, Lorenz B, Kloß S, Henk S, Dees S, et al. Raman spectroscopic identification of Mycobacterium tuberculosis. Journal of Biophotonics. 2017;10(5):727-34.
180. Tahiri N, Fodran P, Jayaraman D, Buter J, Witte MD, Ocampo TA, et al. Total Synthesis of a Mycolic Acid from Mycobacterium tuberculosis. Angewandte Chemie International Edition. 2020;59(19):7555-60.
181. Marrakchi H, Lanéelle M-A, Daffé M. Mycolic Acids: Structures, Biosynthesis, and Beyond. Chemistry & Biology. 2014;21(1):67-85.
182. Yassin A-AF. 10 - Detection and Characterization of Mycolic Acids and Their Use in Taxonomy and Classification. In: Rainey F, Oren A, editors. Methods in Microbiology. 38: Academic Press; 2011. p. 207-37.
183. Verma T, Podder S, Mehta M, Singh S, Singh A, Umapathy S, et al. Raman spectroscopy reveals distinct differences between two closely related bacterial strains, Mycobacterium indicus pranii and Mycobacterium intracellulare. Analytical and Bioanalytical Chemistry. 2019;411(30):7997-8009.
184. Deligianni E, Pattison S, Berrar D, Ternan NG, Haylock RW, Moore JE, et al. Pseudomonas aeruginosa Cystic Fibrosis isolates of similar RAPD genotype exhibit diversity in biofilm forming ability in vitro. BMC Microbiology. 2010;10(1):38.
185. Filho JC, Jorge S, Kremer FS, de Oliveira NR, Campos VF, da Silva Pinto L, et al. Complete genome sequence of native Bacillus cereus strains isolated from intestinal tract of the crab Ucides sp. Data in Brief. 2018;16:381-5.
186. Krishnakumar K. Comparative Study On Durability Properties of Bacterial Concrete. International Research Journal of Engineering and Technology. 2016;03:129-32.
187. Haig S-J. Adherence of Mannheimia haemolytica to ovine bronchial epithelial cells. Bioscience Horizons: The International Journal of Student Research. 2011;4(1):50-60.

188. Tan A, Li W-S, Verderosa AD, Blakeway LV, D. Mubaiwa T, Totsika M, et al. Moraxella catarrhalis NucM is an entry nuclease involved in extracellular DNA and RNA degradation, cell competence and biofilm scaffolding. Scientific Reports. 2019;9(1):2579.
189. Zeng J, Chen D, Lv C, Qin K, Zhou Q, Pu N, et al. Antimicrobial and anti-biofilm activity of Polygonum chinense L.aqueous extract against Staphylococcus aureus. Scientific Reports. 2022;12(1):21988.
190. Painter H, Prabowo SA, Cia F, Stockdale L, Tanner R, Willcocks S, et al. Adaption of the ex vivo mycobacterial growth inhibition assay for use with murine lung cells. Sci Rep. 2020;10(1):3311.
191. Cook BS, Beddow JG, Manso-Silván L, Maglennon GA, Rycroft AN. Selective medium for culture of Mycoplasma hyopneumoniae. Vet Microbiol. 2016;195:158-64.
192. UK Health Security Agency. Pseudomonas fluorescens 2023 [Available from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+10038&collection=nctc.
193. UK Health Security Agency. Pseudomonas aeruginosa Department of Health; 2023 [Available from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+10332&collection=nctc.
194. UK Health Security Agency. Bacillus cereus 2023 [Available from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC%2010320&collection=nctc.
195. Elbing KL, Brent R. Recipes and Tools for Culture of Escherichia coli. Curr Protoc Mol Biol. 2019;125(1):e83.
196. Catry B, Dewulf J, de Kruif A, Vanrobaeys M, Haesebrouck F, Decostere A. Accuracy of Susceptibility Testing of Pasteurella multocida and Mannheimia haemolytica. Microbial Drug Resistance. 2007;13(3):204-11.
197. Chan C, Ng D, Schryvers AB. The Role of the Moraxella catarrhalis CopB Protein in Facilitating Iron Acquisition From Human Transferrin and Lactoferrin. Front Microbiol. 2021;12:714815.
198. Singhi D, Goyal A, Gupta G, Yadav A, Srivastava P. Rhodoccoccus erythropolis Is Different from Other Members of Actinobacteria: Monoploidy, Overlapping Replication Cycle, and Unique Segregation Pattern. J Bacteriol. 2019;201(24).
199. Missiakas DM, Schneewind O. Growth and laboratory maintenance of Staphylococcus aureus. Curr Protoc Microbiol. 2013;Chapter 9:Unit 9C.1.
200. New England Biolabs Inc. High Efficiency Transformation Protocol (C2987H/C2987I) 2022 [Available from: https://international.neb.com/protocols/0001/01/01/high-efficiency-transformation-protocol-c2987.
201. Thermo Scientific. GeneJET Plasmid Miniprep Kit #K0502, #K0503 2014 [Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0012655_GeneJET_Plasmid_Miniprep_UG.pdf.

202. New England Biolabs Inc. Optimizing Restriction Endonuclease Reactions 2014 [Available from: https://international.neb.com/tools-and-resources/usage-guidelines/optimizing-restriction-endonuclease-reactions.
203. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology. 2013;31(3):233-9.
204. Thermo Scientific. GeneJET Gel Extraction Kit #K0691, #K0692 2015 [Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0012661_GeneJET_Gel_Extraction_UG.pdf.
205. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics. 2012;4(1):17.
206. Neese F. The ORCA program system. WIREs Computational Molecular Science. 2012;2(1):73-8.
207. Neese F. Software update: the ORCA program system, version 4.0. WIREs Computational Molecular Science. 2018;8(1):e1327.
208. Neese F, Wennmohs F, Becker U, Riplinger C. The ORCA quantum chemistry program package. The Journal of Chemical Physics. 2020;152(22).
209. Zavatski S, Bandarenka H, Hetmańczyk Ł, Hetmańczyk J, Vorobyeva M, Arynbek Y, et al. Model phospholipid interaction with cholesterol and melatonin: Raman spectroscopy and density functional theory study. Journal of Raman Spectroscopy. 2022;53(9):1540-50.
210. Ashkin A. Acceleration and Trapping of Particles by Radiation Pressure. Physical Review Letters. 1970;24(4):156-9.
211. Ashkin A, Dziedzic JM, Bjorkholm JE, Chu S. Observation of a single-beam gradient force optical trap for dielectric particles. Opt Lett. 1986;11(5):288-90.
212. Liu A, Liu J, Peng W, Xu X-B, Chen G-J, Ren X, et al. Multigrating design for integrated single-atom trapping, manipulation, and readout. Physical Review A. 2022;105(5):053520.
213. Zhang Z, Kimkes TEP, Heinemann M. Manipulating rod-shaped bacteria with optical tweezers. Scientific Reports. 2019;9(1):19086.
214. Pahlow S, Meisel S, Cialla-May D, Weber K, Rösch P, Popp J. Isolation and identification of bacteria by means of Raman spectroscopy. Advanced Drug Delivery Reviews. 2015;89:105-20.
215. Huang T-S, Lee C-C, Tu H-Z, Lee SS-J. Rapid identification of mycobacteria from positive MGIT broths of primary cultures by MALDI-TOF mass spectrometry. PLOS ONE. 2018;13(2):e0192291.
216. Khandasammy SR, Fikiet MA, Mistek E, Ahmed Y, Halámková L, Bueno J, et al. Bloodstains, paintings, and drugs: Raman spectroscopy applications in forensic science. Forensic Chemistry. 2018;8:111-33.
217. Liu Y, Zhou H, Hu Z, Yu G, Yang D, Zhao J. Label and label-free based surface-enhanced Raman scattering for pathogen bacteria detection: A review. Biosensors and Bioelectronics. 2017;94:131-40.

218. Vettigli G. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map 2018 [Available from: https://github.com/JustGlowing/minisom/.
219. Weston D. Mycolic Acid: Github; 2022 [cited 2023 18/06]. Available from: https://github.com/DanW97/mycolic-acid.
220. Greenacre M, Groenen PJF, Hastie T, D'Enza AI, Markos A, Tuzhilina E. Principal component analysis. Nature Reviews Methods Primers. 2022;2(1):100.
221. Zhang S, Zhang X, Yu W, Lin Z, Chen D. Infection biomarkers in assisting the judgement of blood stream infection and patient prognosis: a retrospective study incorporating principal components analysis. Ann Transl Med. 2020;8(23):1581.
222. Galvan D, Effting L, Cremasco H, Conte-Junior CA. The Spread of the COVID-19 Outbreak in Brazil: An Overview by Kohonen Self-Organizing Map Networks. Medicina. 2021;57(3):235.
223. Makmuang S, Terdwongworakul A, Vilaivan T, Maher S, Ekgasit S, Wongravee K. Mapping hyperspectral NIR images using supervised self-organizing maps: Discrimination of weedy rice seeds. Microchemical Journal. 2023;190:108599.
224. Banbury C, Mason R, Styles I, Eisenstein N, Clancy M, Belli A, et al. Development of the Self Optimising Kohonen Index Network (SKiNET) for Raman Spectroscopy Based Detection of Anatomical Eye Tissue. Scientific Reports. 2019;9(1):10812.
225. Ali S, Hassan M, Kim JY, Farid MI, Sanaullah M, Mufti H. FF-PCA-LDA: Intelligent Feature Fusion Based PCA-LDA Classification System for Plant Leaf Diseases. Applied Sciences. 2022;12(7):3514.
226. Zhao Y, Tian S, Yu L, Zhang Z, Zhang W. Analysis and Classification of Hepatitis Infections Using Raman Spectroscopy and Multiscale Convolutional Neural Networks. Journal of Applied Spectroscopy. 2021;88(2):441-51.
227. García-Timermans C, Props R, Zacchetti B, Sakarika M, Delvigne F, Boon N. Raman Spectroscopy-Based Measurements of Single-Cell Phenotypic Diversity in Microbial Populations. mSphere. 2020;5(5):10.1128/msphere.00806-20.
228. Bhat GR, Sethi I, Rah B, Kumar R, Afroze D. Innovative in Silico Approaches for Characterization of Genes and Proteins. Frontiers in Genetics. 2022;13.
229. Madden JC, Enoch SJ, Paini A, Cronin MTD. A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications. Alternatives to Laboratory Animals. 2020;48(4):146-72.
230. Hemmerich J, Ecker GF. In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. WIREs Computational Molecular Science. 2020;10(4):e1475.
231. Macalino SJY, Billones JB, Organo VG, Carrillo MCO. In Silico Strategies in Tuberculosis Drug Discovery. Molecules. 2020;25(3):665.
232. Fernández Á, Fernández C, Miguel-Dávila JÁ, Conde MÁ, Matellán V. Supercomputers to improve the performance in higher education: A review of the literature. Computers & Education. 2019;128:353-64.
233. Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. Computers in Biology and Medicine. 2021;137:104851.

234. Velihina YS, Obernikhina NV, Pilyo SG, Kachaeva MV, Kachkovsky OD. In silico study the interaction of heterocyclic bases with peptide moieties of proteins in “fragment-to-fragment” approach. Ukrainica Bioorganica Acta. 2021;16(1):34-43.
235. Chen H, Zhang Z, Zhang J. In silico drug repositioning based on the integration of chemical, genomic and pharmacological spaces. BMC Bioinformatics. 2021;22(1):52.
236. Urbizo-Reyes U, Liceaga AM, Reddivari L, Kim K-H, Anderson JM. Enzyme kinetics, molecular docking, and in silico characterization of canary seed (Phalaris canariensis L.) peptides with ACE and pancreatic lipase inhibitory activity. Journal of Functional Foods. 2022;88:104892.
237. Cantelli G, Bateman A, Brooksbank C, Petrov AI, Malik-Sheriff Rahuman S, Ide-Smith M, et al. The European Bioinformatics Institute (EMBL-EBI) in 2021. Nucleic Acids Research. 2021;50(D1):D11-D9.
238. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Research. 2016;45(D1):D945-D54.
239. Simpson S, Lonie DC, Chen J, Zurek E. A Computational Experiment on Single-Walled Carbon Nanotubes. Journal of Chemical Education. 2013;90(5):651-5.
240. Salifu EY, Abugri J, Rashid IA, Osei F, Ayariga JA. In silico identification of potential inhibitors of acyl carrier protein reductase and acetyl CoA carboxylase of Plasmodium falciparum in antimalarial therapy. Frontiers in Drug Discovery. 2023;3.
241. Alexiadis A, Albano A, Rahmat A, Yildiz M, Kefal A, Ozbulut M, et al. Simulation of pandemics in real cities: enhanced and accurate digital laboratories. Proc Math Phys Eng Sci. 2021;477(2245):20200653.
242. Malinowska JM, Palosaari T, Sund J, Carpi D, Weber RJM, Lloyd GR, et al. Derivation of metabolic point of departure using high-throughput in vitro metabolomics: investigating the importance of sampling time points on benchmark concentration values in the HepaRG cell line. Archives of Toxicology. 2023;97(3):721-35.
243. Bhattacharya B, Michalchuk AAL, Silbernagl D, Yasuda N, Feiler T, Sturm H, et al. An atomistic mechanism for elasto-plastic bending in molecular crystals. Chemical Science. 2023;14(13):3441-50.
244. Fatima A, Khanum G, Savita S, Pooja K, Verma I, Siddiqui N, et al. Quantum computational, spectroscopic, Hirshfeld surface, electronic state and molecular docking studies on sulfanilic acid: An anti-bacterial drug. Journal of Molecular Liquids. 2022;346:117150.
245. Kusumawati Y, Ivansyah AL, Ali BTI, Kurnia KA, Hutama AS, Fansuri H. Photophysical properties of ammonium, pyrrolidinium, piperidinium, imidazolium, and pyridinium as a guide to prepare ionic-organic hybrid materials. Heliyon. 2022;8(3):e09121.
246. Lu Z. Computational discovery of energy materials in the era of big data and machine learning: A critical review. Materials Reports: Energy. 2021;1(3):100047.
247. López Barreiro D, Martin-Martinez FJ, Torri C, Prins W, Buehler MJ. Molecular characterization and atomistic model of biocrude oils from hydrothermal liquefaction of microalgae. Algal Research. 2018;35:262-73.

248. Malloum A, Conradie J. Accurate binding energies of ammonia clusters and benchmarking of hybrid DFT functionals. Computational and Theoretical Chemistry. 2021;1200:113236.
249. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008;36(Database issue):D344-50.
250. Talari ACS, Movasaghi Z, Rehman S, Rehman Iu. Raman Spectroscopy of Biological Tissues. Applied Spectroscopy Reviews. 2015;50(1):46-111.
251. Perumal J, Dinish US, Bendt AK, Kazakeviciute A, Fu CY, Ong ILH, et al. Identification of mycolic acid forms using surface-enhanced Raman scattering as a fast detection method for tuberculosis. Int J Nanomedicine. 2018;13:6029-38.
252. Rivera-Betancourt OE, Karls R, Grosse-Siestrup B, Helms S, Quinn F, Dluhy RA. Identification of mycobacteria based on spectroscopic analyses of mycolic acid profiles. Analyst. 2013;138(22):6774-85.
253. Akanny E, Bonhommé A, Bessueille F, Bourgeois S, Bordes C. Surface enhanced Raman spectroscopy for bacteria analysis: a review. Applied Spectroscopy Reviews. 2021;56(5):380-422.
254. Zhou X, Hu Z, Yang D, Xie S, Jiang Z, Niessner R, et al. Bacteria Detection: From Powerful SERS to Its Advanced Compatible Techniques. Advanced Science. 2020;7(23):2001739.
255. Deb M, Hunter R, Taha M, Abdelbary H, Anis H. Rapid detection of bacteria using gold nanoparticles in SERS with three different capping agents: Thioglucose, polyvinylpyrrolidone, and citrate. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2022;280:121533.
256. Ansari MA, Khan HM, Khan AA, Ahmad MK, Mahdi AA, Pal R, et al. Interaction of silver nanoparticles with Escherichia coli and their cell envelope biomolecules. Journal of Basic Microbiology. 2014;54(9):905-15.
257. Cui Y, Zhao Y, Tian Y, Zhang W, Lü X, Jiang X. The molecular mechanism of action of bactericidal gold nanoparticles on Escherichia coli. Biomaterials. 2012;33(7):2327-33.
258. Krajczewski J, Ambroziak R, Kudelski A. Substrates for Surface-Enhanced Raman Scattering Formed on Nanostructured Non-Metallic Materials: Preparation and Characterization. Nanomaterials (Basel). 2020;11(1).
259. Heck C, Kanehira Y, Kneipp J, Bald I. Placement of Single Proteins within the SERS Hot Spots of Self-Assembled Silver Nanolenses. Angewandte Chemie International Edition. 2018;57(25):7444-7.
260. Bursch M, Mewes J-M, Hansen A, Grimme S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry**. Angewandte Chemie International Edition. 2022;61(42):e202205735.
261. Bradac C. Nanoscale Optical Trapping: A Review. Advanced Optical Materials. 2018;6(12):1800005.
262. Lupovitz I, Durham ND, Munro JB, Sorkin R. Revealing virus-host membrane fusion mechanisms using optical tweezers. Biophysical Journal. 2022;121(3, Supplement 1):75a.

263. Pesce G. Optical tweezers: Theory and practice. Reference Module in Materials Science and Materials Engineering: Elsevier; 2022.
264. Min TL, Mears PJ, Chubiz LM, Rao CV, Golding I, Chemla YR. High-resolution, long-term characterization of bacterial motility using optical tweezers. Nature Methods. 2009;6(11):831-5.
265. Ghanashyam C, Shetty S, Bharati S, Chidangil S, Bankapur A. Optical Trapping and Micro-Raman Spectroscopy of Functional Red Blood Cells Using Vortex Beam for Cell Membrane Studies. Analytical Chemistry. 2021;93(13):5484-93.
266. Sundarsingh T JA, J R, Rajan A, Shankar V. Features of the biochemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis. Journal of Infection and Public Health. 2020;13(9):1255-64.
267. MacGregor-Fairlie M, Gomes PD, Weston D, Rickard JJS, Oppenheimer PG. Hybrid Raman spectroscopy and artificial neural network algorithm discriminating mycobacterium bovis BCG and members of the order mycobacteriales. bioRxiv. 2023:2023.05.30.542797.
268. De Carvalho Gomes P, Hardy M, Tagger Y, Rickard JJS, Mendes P, Oppenheimer PG. Optimization of Nanosubstrates toward Molecularly Surface-Functionalized Raman Spectroscopy. The Journal of Physical Chemistry C. 2022;126(32):13774-84.
269. Hanson C, Bishop MM, Barney JT, Vargis E. Effect of growth media and phase on Raman spectra and discrimination of mycobacteria. Journal of Biophotonics. 2019;12(11):e201900150.
270. Baron VO, Chen M, Hammarstrom B, Hammond RJH, Glynne-Jones P, Gillespie SH, et al. Real-time monitoring of live mycobacteria with a microfluidic acoustic-Raman platform. Communications Biology. 2020;3(1):236.
271. Deng R, Zhang Y, Wang X, Xie X, Song Y, Bu J, et al. In situ intracellular Raman spectroscopic detection with graphene-based thermoelectric optical tweezers. Sensors and Actuators B: Chemical. 2022;361:131722.
272. Matthews C, Crispie F, Lewis E, Reid M, O'Toole PW, Cotter PD. The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency. Gut Microbes. 2019;10(2):115-32.
273. Durso LM, Harhay GP, Smith TPL, Bono JL, DeSantis TZ, Harhay DM, et al. Animal-to-Animal Variation in Fecal Microbial Diversity among Beef Cattle. Applied and Environmental Microbiology. 2010;76(14):4858.
274. Lima SF, Teixeira AGV, Higgins CH, Lima FS, Bicalho RC. The upper respiratory tract microbiome and its potential role in bovine respiratory disease and otitis media. Scientific Reports. 2016;6:29050.
275. McDaneld TG, Kuehn LA, Keele JW. Evaluating the microbiome of two sampling locations in the nasal cavity of cattle with bovine respiratory disease complex (BRDC). J Anim Sci. 2018;96(4):1281-7.
276. Zeineldin M, Lowe J, de Godoy M, Maradiaga N, Ramirez C, Ghanem M, et al. Disparity in the nasopharyngeal microbiota between healthy cattle on feed, at entry processing and with respiratory disease. Veterinary Microbiology. 2017;208:30-7.

277. Chai J, Capik SF, Kegley B, Richeson JT, Powell JG, Zhao J. Bovine respiratory microbiota of feedlot cattle and its association with disease. Veterinary Research. 2022;53(1):4.
278. Madacki J, Mas Fiol G, Brosch R. Update on the virulence factors of the obligate pathogen Mycobacterium tuberculosis and related tuberculosis-causing mycobacteria. Infection, Genetics and Evolution. 2019;72:67-77.
279. Hall JA, Isaiah A, Estill CT, Pirelli GJ, Suchodolski JS. Weaned beef calves fed selenium-biofortified alfalfa hay have an enriched nasal microbiota compared with healthy controls. PLOS ONE. 2017;12(6):e0179215.
280. Timsit E, Workentine M, Crepieux T, Miller C, Regev-Shoshani G, Schaefer A, et al. Effects of nasal instillation of a nitric oxide-releasing solution or parenteral administration of tilmicosin on the nasopharyngeal microbiota of beef feedlot cattle at high-risk of developing respiratory tract disease. Research in Veterinary Science. 2017;115:117-24.
281. Gaeta NC, Lima SF, Teixeira AG, Ganda EK, Oikonomou G, Gregory L, et al. Deciphering upper respiratory tract microbiota complexity in healthy calves and calves that develop respiratory disease using shotgun metagenomics. Journal of Dairy Science. 2017;100(2):1445-58.
282. Nicola I, Cerutti F, Grego E, Bertone I, Gianella P, D'Angelo A, et al. Characterization of the upper and lower respiratory tract microbiota in Piedmontese calves. Microbiome. 2017;5(1):152.
283. McMullen C, Orsel K, Alexander TW, van der Meer F, Plastow G, Timsit E. Comparison of the nasopharyngeal bacterial microbiota of beef calves raised without the use of antimicrobials between healthy calves and those diagnosed with bovine respiratory disease. Veterinary Microbiology. 2019;231:56-62.
284. Saha C, Baruah N, Nayak SK, editors. Implementation of Self-Organizing Map and Logistic Regression in Dissolved Gas Analysis of Transformer oils. 2021 IEEE International Conference on the Properties and Applications of Dielectric Materials (ICPADM); 2021 12-14 July 2021.
285. Teufel R, Mascaraque V, Ismail W, Voss M, Perera J, Eisenreich W, et al. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. Proc Natl Acad Sci U S A. 2010;107(32):14390-5.
286. Vila E, Hornero-Méndez D, Azziz G, Lareo C, Saravia V. Carotenoids from heterotrophic bacteria isolated from Fildes Peninsula, King George Island, Antarctica. Biotechnol Rep (Amst). 2019;21:e00306.
287. López G-D, Álvarez-Rivera G, Carazzone C, Ibáñez E, Leidy C, Cifuentes A. Bacterial Carotenoids: Extraction, Characterization, and Applications. Critical Reviews in Analytical Chemistry. 2021:1-24.
288. Movasaghi Z, Rehman S, Rehman IU. Raman Spectroscopy of Biological Tissues. Applied Spectroscopy Reviews. 2007;42(5):493-541.
289. Velioglu SD, Ercioglu E, Temiz HT, Velioglu HM, Topcu A, Boyaci IH. Raman Spectroscopic Barcode Use for Differentiation of Vegetable Oils and Determination of Their

Major Fatty Acid Composition. Journal of the American Oil Chemists' Society. 2016;93(5):627-35.
290. Allen DM, Einarsson GG, Tunney MM, Bell SEJ. Characterization of Bacteria Using Surface-Enhanced Raman Spectroscopy (SERS): Influence of Microbiological Factors on the SERS Spectra. Anal Chem. 2022;94(26):9327-35.
291. VanEvery H, Franzosa EA, Nguyen LH, Huttenhower C. Microbiome epidemiology and association studies in human health. Nature Reviews Genetics. 2023;24(2):109-24.
292. Laux C, Peschel A, Krismer B. Staphylococcus aureus Colonization of the Human Nose and Interaction with Other Microbiome Members. Microbiology Spectrum. 2019;7(2):10.1128/microbiolspec.gpp3-0029-2018.
293. Haag AF, Fitzgerald JR, Penadés JR. Staphylococcus aureus in Animals. Microbiology Spectrum. 2019;7(3):10.1128/microbiolspec.gpp3-0060-2019.
294. Martinson JNV, Walk ST. Escherichia coli Residency in the Gut of Healthy Human Adults. EcoSal Plus. 2020;9(1).
295. Vasco K, Nohomovich B, Singh P, Venegas-Vargas C, Mosci RE, Rust S, et al. Characterizing the Cattle Gut Microbiome in Farms with a High and Low Prevalence of Shiga Toxin Producing Escherichia coli. Microorganisms. 2021;9(8).
296. Kloß S, Lorenz B, Dees S, Labugger I, Rösch P, Popp J. Destruction-free procedure for the isolation of bacteria from sputum samples for Raman spectroscopic analysis. Analytical and Bioanalytical Chemistry. 2015;407(27):8333-41.
297. Beal J, Farny NG, Haddock-Angelli T, Selvarajah V, Baldwin GS, Buckley-Taylor R, et al. Robust estimation of bacterial cell count from optical density. Communications Biology. 2020;3(1):512.
298. Ramos EI, Das K, Harrison AL, Garcia A, Gadad SS, Dhandayuthapani S. Mycoplasma genitalium and M. pneumoniae Regulate a Distinct Set of Protein-Coding Genes in Epithelial Cells. Front Immunol. 2021;12:738431.
299. Witkowska E, Niciński K, Korsak D, Szymborski T, Kamińska A. Sources of variability in SERS spectra of bacteria: comprehensive analysis of interactions between selected bacteria and plasmonic nanostructures. Analytical and Bioanalytical Chemistry. 2019;411(10):2001-17.
300. Cantillon D, Wroblewska J, Cooper I, Newport MJ, Waddell SJ. Three-dimensional low shear culture of Mycobacterium bovis BCG induces biofilm formation and antimicrobial drug tolerance. NPJ Biofilms Microbiomes. 2021;7(1):12.
301. Tracey H, Coates N, Hulme E, John D, Michael DR, Plummer SF. Insights into the enumeration of mixtures of probiotic bacteria by flow cytometry. BMC Microbiology. 2023;23(1):48.
302. Huang N, Short M, Zhao J, Wang H, Lui H, Korbelik M, et al. Full range characterization of the Raman spectra of organs in a murine model. Opt Express. 2011;19(23):22892-909.
303. Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V. crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. RNA Biology. 2013;10(5):841-51.

304. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007;315(5819):1709-12.
305. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide Sequence of the iap Gene, Responsible for Alkaline. Journal of bacteriology. 1987;169(12):5429-33.
306. Parikka KJ, Le Romancer M, Wauters N, Jacquet S. Deciphering the virus‐to‐prokaryote ratio (VPR): insights into virus–host relationships in a variety of ecosystems. Biological reviews. 2017;92(2):1081-100.
307. Kariginov F, Hannon G. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mollecular Cell. 2010;37(1):1-23.
308. Rath D, Amlinger L, Rather A, Lundgren M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie. 2015;117(1):119-28.
309. Sander J, Joung J. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology. 2014;32(1):347-55.
310. Mali P, Esvelt K, Church G. Cas9 as a versatile tool for engineering biology. Nature Methods. 2013;10(1):957-63.
311. Jiang F, Doudna J. CRISPR–Cas9 Structures and Mechanisms. Annual review of Biophysics. 2017;46(1):509-29.
312. Garcia-Doval C, Jinek M. Molecular architectures and mechanisms of Class 2 CRISPR-associated nucleases. Current Opinion in Structural Biology. 2017;47(1):157-66.
313. Yang K, Guo R, Xu D. Non-homologous end joining: advances and frontiers. Acta Biochimica et Biophysica Sinica. 2016;48(7):632-40.
314. Vines AJ, Cox K, Leland BA, King MC. Homology-directed repair involves multiple strand invasion cycles in fission yeast. Molecular Biology of the Cell. 2022;33(4):ar30.
315. Ishino Y, Krupovic M, Forterre P. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. J Bacteriol. 2018;200(7).
316. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature. 2016;532:517.
317. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science. 2018.
318. Lino C, Harper J, Carney J, Timlin J. Delivering CRISPR: a review of the challenges and approaches. Drug delivery. 2018;25(1):1234-57.
319. Kornete M, Marone R, Jeker L. Highly Efficient and Versatile Plasmid-Based Gene Editing in Primary T Cells. Journal of Immunology. 2018;200(7):2489-501.
320. Sato M, Koriyama M, Watanabe S, Ohtsuka M, Sakurai T, Inada E, et al. Direct Injection of CRISPR/Cas9-Related mRNA into Cytoplasm of Parthenogenetically Activated Porcine Oocytes Causes Frequent Mosaicism for Indel Mutations. International Journal of Molecular Sciences. 2015;16(8):17838-56.
321. Liang X, Potter J, Kumae S, Zou Y, Quintanilla R, Sridharan M, et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. Journal of Biotechnology. 2015;208(1):44-53.

322. Chuang C, Chen C, Huang C, Su Y, Peng S, Lin T, Yang TS, et al. Generation of GGTA1 Mutant Pigs by Direct Pronuclear Microinjection of CRISPR/Cas9 Plasmid Vectors. Animal Biotechnology. 2017;3(28):174-81.
323. Raveux A, Vandormael-Pournin S, Cohen-Tannoudji M. Optimization of the production of knock-in alleles by CRISPR/Cas9 microinjection into the mouse zygote. Scientific Reports. 2017;7:42661.
324. Rathbone T, Ates I, Fernando L, Addlestone E, Lee CM, Richards VP, et al. Electroporation-Mediated Delivery of Cas9 Ribonucleoproteins Results in High Levels of Gene Editing in Primary Hepatocytes. Crispr j. 2022;5(3):397-409.
325. Chen S, Lee L, AYF, Modzelweski A, He L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. Journal of Biological Chemistry 2016.
326. Paquet D, Kwart D, Chen, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature Research. 2016;533:125-9.
327. Dong C, Qu L, Wang H, Wei L, Dong Y, Xiong S. Targeting hepatitis B virus cccDNA by CRISPR/Cas9 nuclease efficiently inhibits viral replication. Antiviral Research. 2015;118:110-7.
328. Guan, Y, Ma Y, Li Q, Sun Z, Ma L, et al. CRISPR/Cas9‐mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. EMBO Molecular Medicine. 2016;8:477-88.
329. Long C, Amoasii L, Mireault A, McAnally J, Li H, Sanchez-Ortiz E, et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science. 2016;351(6271):400-3.
330. Voets O, Tielen F, Elstak E, Benschop J, Grimbergen M, Stallen J, et al. Highly efficient gene inactivation by adenoviral CRISPR/Cas9 in human primary cells. PLOS ONE. 2017;12(8):e0182974.
331. Maggio I, Stefanucci L, Janssen JM, Liu J, Chen X, Mouly V, et al. Selection-free gene repair after adenoviral vector transduction of designer nucleases: rescue of dystrophin synthesis in DMD muscle cell populations. Nucleic Acids Research. 2016;44(3):1449-70.
332. Miller JB, Zhang S, Kos P, Xiong H, Zhou K, Perelman SS, et al. Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA. Angewandte Chemie. 2017;129(4):1079-83.
333. Yin H, Song C-Q, Dorkin JR, Zhu LJ, Li Y, Wu Q, et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. Nature Biotechnology. 2016;34:328.
334. Axford DS, Morris DP, McMurry JL. Cell penetrating peptide-mediated nuclear delivery of Cas9 to enhance the utility of CRISPR/Cas genome editing. The FASEB Journal. 2017;31(1_supplement):909.4-.4.
335. Sun W, Ji W, Hall JM, Hu Q, Wang C, Beisel CL, et al. Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR–Cas9 for Genome Editing. Angewandte Chemie. 2015;127(41):12197-201.

336. D'Astolfo Diego S, Pagliero Romina J, Pras A, Karthaus Wouter R, Clevers H, Prasad V, et al. Efficient Intracellular Delivery of Native Proteins. Cell. 2015;161(3):674-90.
337. Mout R, Ray M, Yesilbag Tonga G, Lee Y-W, Tay T, Sasaki K, et al. Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. ACS Nano. 2017;11(3):2452-8.
338. Teng KW, Ishitsuka Y, Ren P, Youn Y, Deng X, Ge P, et al. Labeling Proteins Inside Living Cells Using External Fluorophores for Fluorescence Microscopy. eLife. 2017;6:e25460.
339. Nakamura T, Akita H, Yamada Y, Hatakeyama H, Harashima H. A Multifunctional Envelope-type Nanodevice for Use in Nanomedicine: Concept and Applications. Accounts of Chemical Research. 2012;45(7):1113-21.
340. Su J, Sun H, Meng Q, Zhang P, Yin Q, Li Y. Enhanced Blood Suspensibility and Laser-Activated Tumor-specific Drug Release of Theranostic Mesoporous Silica Nanoparticles by Functionalizing with Erythrocyte Membranes. Theranostics. 2017;7(3):523-37.
341. Luo G-F, Chen W-H, Liu Y, Lei Q, Zhuo R-X, Zhang X-Z. Multifunctional Enveloped Mesoporous Silica Nanoparticles for Subcellular Co-delivery of Drug and Therapeutic Peptide. Scientific Reports. 2014;4:6064.
342. Zhang Y, Wang Y, Xu L, Lou C, Ouyang Q, Qian L. Paired dCas9 design as a nucleic acid detection platform for pathogenic strains. Methods. 2022;203:70-7.
343. Yuan B, Yuan C, Li L, Long M, Chen Z. Application of the CRISPR/Cas System in Pathogen Detection: A Review. Molecules. 2022;27(20):6999.
344. Shin J, Miller M, Wang Y-C. Recent advances in CRISPR-based systems for the detection of foodborne pathogens. Comprehensive Reviews in Food Science and Food Safety. 2022;21(3):3010-29.
345. Kim H, Lee S, Yoon J, Song J, Park HG. CRISPR/Cas12a collateral cleavage activity for simple and rapid detection of protein/small molecule interaction. Biosensors and Bioelectronics. 2021;194:113587.
346. Li J, Macdonald J, von Stetten F. Review: a comprehensive summary of a decade development of the recombinase polymerase amplification. Analyst. 2019;144(1):31-67.
347. Moon J, Kwon H-J, Yong D, Lee I-C, Kim H, Kang H, et al. Colorimetric Detection of SARS-CoV-2 and Drug-Resistant pH1N1 Using CRISPR/dCas9. ACS Sensors. 2020;5(12):4017-26.
348. Goig GA, Torres-Puente M, Mariner-Llicer C, Villamayor LM, Chiner-Oms Á, Gil-Brusola A, et al. Towards next-generation diagnostics for tuberculosis: identification of novel molecular targets by large-scale comparative genomics. Bioinformatics. 2020;36(4):985-9.
349. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, et al. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. Cochrane Database of Systematic Reviews. 2021(2).
350. Dookie N, Rambaran S, Padayatchi N, Mahomed S, Naidoo K. Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care. Journal of Antimicrobial Chemotherapy. 2018;73(5):1138-51.

351. Sankar S, Kuppanan S, Balakrishnan B, Nandagopal B. Analysis of sequence diversity among IS6110 sequence of Mycobacterium tuberculosis: possible implications for PCR based detection. Bioinformation. 2011;6(7):283-5.
352. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res. 2023;51(D1):D678-d89.
353. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Research. 2019;47(W1):W171-W4.
354. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7:12.
355. Vandenberghe LH. Addgene: Molecular Therapy Interview with Melina Fan and Karen Guerin. Molecular Therapy. 2019;27(1):6-9.
356. New England Biolabs Inc. Gibson Assembly® Chemical Transformation Protocol (E5510) Ipswich, MA, USA: New England Biolabs Inc,; 2020 [cited 2020 14.05]. Available from: https://international.neb.com/protocols/2012/12/11/gibson-assembly-transformation-protocol-e5510.
357. Ooi KH, Liu MM, Tay JWD, Teo SY, Kaewsapsak P, Jin S, et al. An engineered CRISPR-Cas12a variant and DNA-RNA hybrid guides enable robust and rapid COVID-19 testing. Nature Communications. 2021;12(1):1739.
358. Nimsamer P, Mayuramart O, Rattanaburi S, Chantaravisoot N, Saengchoowong S, Puenpa J, et al. Comparative performance of CRISPR-Cas12a assays for SARS-CoV-2 detection tested with RNA extracted from clinical specimens. J Virol Methods. 2021;290:114092.
359. Brites D, Gagneux S. The Nature and Evolution of Genomic Diversity in the Mycobacterium tuberculosis Complex. In: Gagneux S, editor. Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control. Cham: Springer International Publishing; 2017. p. 1-26.
360. Erlandsen SL, Kristich CJ, Dunny GM, Wells CL. High-resolution visualization of the microbial glycocalyx with low-voltage scanning electron microscopy: dependence on cationic dyes. J Histochem Cytochem. 2004;52(11):1427-35.
361. Peñuelas-Urquides K, Villarreal-Treviño L, Silva-Ramírez B, Rivadeneyra-Espinoza L, Said-Fernández S, León MBd. Measuring of Mycobacterium tuberculosis growth: a correlation of the optical measurements with colony forming units. Brazilian Journal of Microbiology. 2013;44.
362. Kahli H, Béven L, Grauby-Heywang C, Debez N, Gammoudi I, Moroté F, et al. Impact of Growth Conditions on Pseudomonas fluorescens Morphology Characterized by Atomic Force Microscopy. Int J Mol Sci. 2022;23(17).
363. Dong-ju K. Relation of microbial biomass to counting units for Pseudomonas aeruginosa. African Journal of Microbiology Research. 2012;6(21).
364. Biesta-Peters EG, Reij MW, Joosten H, Gorris LGM, Zwietering MH. Comparison of Two Optical-Density-Based Methods and a Plate Count Method for Estimation of Growth

Parameters of Bacillus cereus. Applied and Environmental Microbiology. 2010;76(5):1399-405.
365. Mira P, Yeh P, Hall BG. Estimating microbial population data from optical density. PLoS One. 2022;17(10):e0276040.
366. Dassanayake RP, Shanthalingam S, Herndon CN, Lawrence PK, Frances Cassirer E, Potter KA, et al. Mannheimia haemolytica serotype A1 exhibits differential pathogenicity in two related species, Ovis canadensis and Ovis aries. Veterinary Microbiology. 2009;133(4):366-71.
367. Bair KL, Campagnari AA. Moraxella catarrhalis Promotes Stable Polymicrobial Biofilms With the Major Otopathogens. Frontiers in Microbiology. 2020;10.
368. Deriase S, El-Gendy N. Mathematical Correlation between Microbial Biomass and Total Viable Count for Different Bacterial Strains Used in Biotreatment of Oil Pollution. Biosciences Biotechnology Research Asia. 2014;11:61-5.
369. Chang Y-C, Yang C-Y, Sun R-L, Cheng Y-F, Kao W-C, Yang P-C. Rapid single cell detection of Staphylococcus aureus by aptamer-conjugated gold nanoparticles. Scientific Reports. 2013;3(1):1863.
370. Costa-Riu N, Burkovski A, Krämer R, Benz R. PorA Represents the Major Cell Wall Channel of the Gram-Positive Bacterium Corynebacterium glutamicum. Journal of Bacteriology. 2003;185(16):4779-86.