

**Development of novel mass spectrometric methods for the
characterisation and identification of microorganisms**

A thesis submitted by

Nicole Strittmatter

For the degree of Doctor of Philosophy (PhD)

Imperial College London
Department of Surgery and Cancer

Abstract

This study describes the development of a novel, REIMS-based characterisation and identification tool for unicellular organisms. A bipolar handheld sampling probe was optimised and characterised on several instrumental platforms before being applied to the acquisition of a large scale spectral database of bacteria and yeasts. Using this database, the specificity of the method was characterised using multivariate statistics and found to yield comparable identification results as MALDI-TOF-MS based techniques for a set of 28 clinically relevant bacterial species. Further tests were performed on sub-species level such as ribotype in case of *C. difficile*, serotype in case of *S. pneumoniae*, antibiotic resistance in *K. pneumoniae* and strain-level differentiation for *E. coli*. The method could be further applied to yeasts and yielded excellent results for a set of five *Candida* species.

To extend the range of the methodology to cell line cultures, the method was further tested for reproducibility and robustness initially using three cell lines. Subsequently, REIMS profiles were collected for the whole NCI60 cancer cell line panel and investigated for their spectral reproducibility, clustering behaviour with regards to tissue type of origin and comparison with spectra of corresponding bulk cancer tissue specimens. REIMS profiles were additionally correlated with publicly available gene and protein expression data in order to elucidate the sensitivity of this REIMS-based approach.

Finally, taxon-specific bacterial biomarkers were derived from a dataset containing 228 bacterial species by finding spectral features that show specificity for a certain phylogenetic group of bacteria. Using these markers, bacteria were detected in tissue sections of both cancerous and healthy colorectal tissue previously acquired using DESI-MSI. Findings were in good agreement with data obtained using 16S rRNA gene sequencing-based analysis and relevant literature.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

To everyone who supported me on the way.

Statement of originality

I certify that this thesis and the research to which it refers are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Nicole Strittmatter

Table of Contents

Abstract.....	2
Statement of originality.....	5
List of Figures.....	11
List of Tables.....	19
List of Appendices.....	21
Background and Objectives.....	22
Background.....	22
Objectives.....	23
1. Introduction.....	25
1.1. Microorganisms.....	25
1.2. Identification and characterisation of microorganisms using mass spectrometry.....	29
1.2.1. 1970-1995.....	29
1.2.2. 1996 onwards: soft ionisation techniques.....	31
1.2.3. 2004 onwards: Ambient mass spectrometry techniques.....	39
1.3. Statistical methods for analysis of mass spectrometric profiling data.....	46
1.3.1. Biomarker discovery using univariate statistical methods.....	46
1.3.2. Multivariate statistical methods.....	47
2. Materials and Methods.....	52
2.1. REIMS analysis of microorganisms.....	52
2.1.1. Culturing of microorganisms.....	52
2.1.2. Rapid evaporative ionisation mass spectrometry.....	52
2.1.3. Determination of the amount of microbial wet weight needed for REIMS analysis.....	55
2.1.4. Data analysis.....	55

2.1.5.	Bacterial metabolite identification.....	58
2.2.	Analysis of cell cultures using REIMS.....	59
2.2.1.	Culturing of cell lines.....	59
2.2.2.	REIMS analysis of cell lines.....	59
2.2.3.	Data Analysis of REIMS cell line data.....	60
2.3.	Desorption Electrospray Ionisation Mass Spectrometry.....	61
2.3.1.	Desorption Electrospray Ionisation Setup.....	61
2.3.2.	Desorption Electrospray Ionisation Mass Spectrometry Imaging.....	63
2.3.3.	Analysis of bacteria using DESI.....	66
2.4.	Determination of taxon-specific markers.....	66
2.4.1.	Sample set.....	67
2.4.2.	Data analysis.....	67
2.5.	Detection of taxon-specific markers in human colorectal tissue specimens.....	70
3.	Method Development and Characterisation.....	71
3.1.	Experimental setup – Ion source.....	71
3.1.1.	Comparison of spectral content of REIMS and DESI datasets.....	71
3.2.	REIMS Sampling Probe.....	75
3.3.	Influence of the electrosurgical generator.....	79
3.4.	Experimental setup - Instrumental parameters.....	82
3.4.1.	Comparison of positive and negative ion mode.....	82
3.4.2.	Optimisation of ion transfer and ion optical settings.....	83
3.4.3.	Suitability of different mass analysers.....	87
3.4.4.	Instrument contamination.....	88
3.5.	Experimental setup – Adaptation to Xevo G2-XS instrumental platform.....	91
3.5.1.	Different atmospheric pressure interfaces.....	91
3.5.2.	Aerosol pick up using isopropyl alcohol.....	95

3.6.	Characterisation of spectral reproducibility	100
3.7.	Analysis of liquid cultures.....	106
4.	Analysis of microorganisms	108
4.1.	Building of REIMS spectral database	108
4.2.	Characterisation of REIMS spectral content.....	109
4.2.1.	Bacterial glycerophospholipids detected using REIMS	111
4.2.2.	Mycolic acids in Corynebacterinaceae suborder	116
4.2.3.	Sphingolipids in the phylum Bacteroidetes	123
4.2.4.	Lipid A in <i>H. pylori</i>	127
4.2.5.	Quorum-sensing molecules in <i>P. aeruginosa</i>	129
4.2.6.	Rhamnolipids in <i>P. aeruginosa</i>	132
4.2.7.	Lipopolypeptides in <i>Bacillus</i> species.....	134
4.2.8.	Polyhydroxyalkanoate polymers detected using REIMS	137
4.3.	Specificity assessment of REIMS-based microbial ID method	138
4.3.1.	Specificity on different taxonomic levels	138
4.3.2.	Dependence of the correct identification performance on spectral resolution	142
4.3.3.	Inter-platform comparison	147
4.3.4.	Comparison of spectral profiles for bacteria cultured on liquid and solid culturing medium.....	149
4.3.5.	Testing of sensitivity and specificity at sub-species level	153
4.3.6.	Analysis of yeasts	160
4.3.7.	Analysis of bacterial mixtures	163
5.	Analysis of cell lines using REIMS	168
5.1.	Introduction	168
5.2.	Assessment of robustness of REIMS spectral profiles.....	170
5.3.	REIMS profiles for the NCI60 cell line set.....	174

5.4.	Comparison of REIMS spectral profiles of bulk cancer tissue samples and cell lines	178
5.5.	NCI60 cell line panel - Correlation with gene expression data.....	181
5.6.	Analysis of mycoplasma infected cell lines	188
6.	Taxon-specific markers.....	192
6.1.	Introduction: Identification of bacteria in human tissues.....	192
6.2.	Assessment of feasibility of a taxon-specific marker approach.....	195
6.3.	Biomarkers specific for different bacterial taxonomic levels	196
6.4.	Detection of bacteria in human colorectal tissue specimens.....	212
6.4.1.	Detection of bacteria in necrotic tissue	212
6.4.2.	Detection of bacteria in healthy mucosa.....	216
6.4.3.	Detection of bacteria in entire colorectal tissue sample set.....	216
	Summary	224
	Impact	226
	Future work.....	228
	Microorganisms	228
	Cell lines	229
	Acknowledgements.....	231
	List of publications	233
	Published.....	233
	Submitted.....	235
	List of conference presentations	236
	Podium presentations	236
	Poster presentations	237
	References.....	238
	Appendix.....	254

Permissions to reprint277

List of Figures

Figure 1. Artistic impression of bacterial cell wall for Gram-positive and Gram-negative bacteria.....	28
Figure 2. Most common glycerolphospholipid structures encountered in prokaryotic and eukaryotic cells	29
Figure 3. Schematic representation of ionisation processes occurring during Matrix-assisted laser desorption ionisation (MALDI) and Electrospray ionisation (ESI).	32
Figure 4. Routine MALDI Biotyper workflow used in clinical microbiology laboratories for the identification of bacteria.	35
Figure 5. Schematic representation of Desorption Electrospray Ionisation.	42
Figure 6. Schematic representation of REIMS instrumentation and data collection.	44
Figure 7. Ionisation mechanism for REIMS for the example of Gram-negative bacterial cells	45
Figure 8. Schematic representation of experimental REIMS setup using bipolar tool as applied for all instrumental platforms used in this study.....	53
Figure 9. Schematic representation depicting processes during bipolar REIMS ionisation....	54
Figure 10. Schematic representation for selection of spectra for downstream data analysis using a total ion count (TIC) threshold.	56
Figure 11. Data processing workflow for specificity assessment on different phylogenetic levels	57
Figure 12. Setup for REIMS analysis of cell line pellets.....	60
Figure 13. Home-built DESI source mounted on Exactive instrument	62
Figure 14. Schematic representation and photograph of a DESI sprayer assembly as used in this study.....	62
Figure 15. Photograph of the DESI spray head used in this study showing relative position of gas and solvent capillaries.	63

Figure 16. Schematic representation displaying processes underlying DESI mass spectrometry imaging.....	65
Figure 17. Data analysis workflow used for the determination of taxon-specific markers.	67
Figure 18. Workflow used for visualising taxon-specific markers in human colorectal tissue sections.....	69
Figure 19. Spectral profiles of same clinical isolate of <i>Bacteroides fragilis</i> obtained using REIMS and DESI using negative ion mode.	73
Figure 20. Comparison of spectral profiles of the same clinical isolate of <i>Pseudomonas aeruginosa</i> obtained using REIMS and DESI	74
Figure 21. Spectra obtained from <i>Escherichia coli</i> K12 derivatives using DESI-MS and REIMS.	75
Figure 22. Monopolar and bipolar setups tested for the analysis of microorganisms using REIMS.	76
Figure 23. Schematic representation of experimental setup and electrical current density using a monopolar and a bipolar tool	77
Figure 24. Spectra of <i>Staphylococcus aureus</i> obtained using monopolar and bipolar setup...	77
Figure 25. REIMS spectral profiles obtained for Luria-Bertani agar and blood agar.	79
Figure 26. Dependence of spectral profile of <i>Staphylococcus aureus</i> ATCC 25923 on RF power setting of electrosurgical generator.....	80
Figure 27. REIMS spectral profiles for a clinical isolate of <i>Proteus mirabilis</i> acquired using two different electrosurgical generators as power supply.....	81
Figure 28. Principal components analysis plot of dataset comprising three different bacterial species recorded using two different electrosurgical power supplies.....	82
Figure 29. Comparison of REIMS profiles obtained in positive and negative ion mode for <i>Escherichia coli</i> , <i>Pseudomonas aeruginosa</i> , <i>Staphylococcus aureus</i>	83
Figure 30. Total ion current for repeated measurements of <i>Staphylococcus similans</i> as a function of capillary voltage, capillary temperature, skimmer voltage, and tube lens voltage.	85

Figure 31. Spectral appearance in dependence on atmospheric pressure interface skimmer voltage and capillary temperature.....	85
Figure 32. Spectral appearance for three consecutive REIMS measurements of <i>Staphylococcus similans</i> recorded at 34°C capillary temperature.....	86
Figure 33. Spectrum of <i>Escherichia coli</i> strains analysed on three different instruments with different mass analysers.....	88
Figure 34. REIMS spectra of <i>Pseudomonas aeruginosa</i> recorded on a Thermo Scientific Orbitrap Discovery XL hybrid instrument using the LTQ XL analyser without prior cleaning.....	89
Figure 35. Different atmospheric pressure interfaces used in this study.....	92
Figure 36. Schematic representation of ion flow in Waters Stepwave ion guide.....	93
Figure 37. Spectral profiles of <i>Bacteroides fragilis</i> obtained using different atmospheric pressure interfaces.....	94
Figure 38. Photograph of apparatus used for introduction of isopropanol using bipolar sampling probe and direct introduction into the mass spectrometer.....	96
Figure 39. Spectral profiles for <i>Candida albicans</i> using different experimental setups.....	97
Figure 40. Spectral profiles of <i>Proteus mirabilis</i> (Gram-negative) using different experimental setups.....	97
Figure 41. REIMS spectra obtained for <i>Bacteroides fragilis</i> using IPA mediated aerosol pick-up setup operated with different IPA flow rates.....	98
Figure 42. REIMS spectra obtained for <i>Bacteroides fragilis</i> using aerosol pick-up setup operated with different isopropanol flow rates.....	99
Figure 43. REIMS spectral profiles for each five different clinical isolates of <i>Staphylococcus aureus</i> , <i>Pseudomonas aeruginosa</i> and <i>Escherichia coli</i> cultured on Columbia blood agar..	101
Figure 44. REIMS spectral profiles obtained for <i>Staphylococcus aureus</i> ATCC 25923, <i>Pseudomonas aeruginosa</i> ATCC 27853 and <i>Escherichia coli</i> ATCC 25922, each grown on five different solid growth media.....	102

Figure 45. REIMS spectral profiles obtained for <i>Staphylococcus aureus</i> ATCC 25923, <i>Pseudomonas aeruginosa</i> ATCC 27853 and <i>Escherichia coli</i> ATCC 25922, each grown on Columbia blood agar for 48hrs at 37°C under different atmospheres	103
Figure 46. Regions of mass spectra of <i>S. aureus</i> grown under aerobic and anaerobic conditions.....	105
Figure 47. REIMS spectral profiles obtained for clinical isolate of <i>Proteus mirabilis</i> with and without solid support.....	107
Figure 48. Overlaid REIMS spectral profiles as obtained for nine different bacterial species for m/z 500-1900 and m/z 600-900 in zoom	111
Figure 49. PCA plot and RMMC plot of the processed dataset, mass range m/z 500-1,900.	113
Figure 50. Full REIMS spectrum of <i>Staphylococcus epidermidis</i> ATCC 1228 cultured for 48hrs on BHI agar.....	114
Figure 51. Fragmentation spectra for cardiolipins at m/z 1352, m/z 1394 and m/z 1423 from <i>Staphylococcus epidermidis</i> ATCC12228	115
Figure 52. Structure of Corynomycolic acid, a short-chain mycolic acid found in <i>Corynebacterium</i> and <i>Rhodococcus</i> spp., and collision-induced dissociation fragmentation pathway observed in negative ion mode.....	116
Figure 53. REIMS spectral profiles for <i>Corynebacterium afermentans</i> (clinical isolate), <i>Rhodococcus equi</i> (ATCC 6939), <i>Nocardia asteroides</i> (ATCC 19247) and <i>Mycobacterium avium</i> (clinical isolate).....	117
Figure 54. Zoom into REIMS profile for five different <i>Corynebacterium</i> species	118
Figure 55. Tandem mass spectra obtained for mycolic acids of an isolate of <i>Corynebacterium striatum</i>	119
Figure 56. Tandem mass spectra of mycolic acids obtained for <i>Nocardia asteroides</i> ATCC19247.....	122
Figure 57. Fragmentation spectrum and tentative scheme of fragmentation for oxidised ceramide signal at m/z 590 obtained from <i>B. fragilis</i>	125
Figure 58. Fragmentation spectra obtained for peaks at m/z 705, m/z 691 and m/z 677 from <i>Bacteroides fragilis</i> assigned as ceramide phosphorylethanolamines.	126

Figure 59. Fragmentation spectra obtained for peaks at m/z 932, 946 and 960 from <i>Parabacteroides distonasis</i> assigned as C15:0 substituted phosphoglycerol dihydroceramides (subPG-DHC).	126
Figure 60. Structures of the endotoxin Lipid A for <i>Escherichia coli</i> and <i>Helicobacter pylori</i>	127
Figure 61. REIMS spectrum obtained from a clinical isolate of <i>Helicobacter pylori</i>	128
Figure 62. A clinical isolate of <i>Pseudomonas aeruginosa</i> grown as a lawn and as single colonies showing markedly different REIMS profiles	130
Figure 63. Tandem mass spectra of quorum-sensing molecules produced by <i>P. aeruginosa</i>	131
Figure 64. Rhamnolipid species commonly produced by <i>P. aeruginosa</i> strains.....	133
Figure 65. Fragmentation spectra for rhamnolipids species detected in <i>Pseudomonas aeruginosa</i>	133
Figure 66. Mass spectra obtained from lipopolypeptide producing strains of the <i>Bacillus</i> genus recorded using REIMS.	135
Figure 67. REIMS fragmentation spectrum of m/z 1034 corresponding to Surfactin(C15)..	136
Figure 68. REIMS spectral profile of a clinical isolate of <i>Bacillus cereus</i> displaying polyhydroxybutyrate production.....	137
Figure 69. Pseudo-2-dimensional PCA and RMMC plots of model comprising 28 different clinically relevant bacterial species.	140
Figure 70. Hierarchical cluster analysis (HCA) of the dataset shown in Table 18 and Figure 69.....	141
Figure 71. Cross-validation results for the dataset shown in Figure 69 as a function of the bin size on species-, genus-, and Gram stain-level	142
Figure 72. Results of unsupervised and supervised analysis for the dataset shown in Figure 69 using 0.01Da, 0.1Da, and 1Da bin size.....	143
Figure 73. Principal components analysis plots from a dataset comprising 217 bacterial species different phylogenetic levels	145

Figure 74. RMMC plot showing the combined data acquired on three different instruments (Exactive, Orbitrap, and Xevo).....	148
Figure 75. REIMS spectral profiles obtained from a clinical isolate of <i>P. aeruginosa</i> grown in solid and liquid culturing medium.	151
Figure 76. REIMS spectral profiles obtained from <i>Staphylococcus aureus</i> ATCC 6538 grown in solid and liquid culturing medium.	151
Figure 77. PCA plots of datasets comprising only solid cultures, only liquid cultures and both.	152
Figure 78. PCA-LDA model of seven different <i>E. coli</i> strains.....	154
Figure 79. PCA and PLS plot for three different <i>S. pneumoniae</i> serotypes.	156
Figure 80. PCA and RMMC multivariate statistical plots for three different ribotypes of <i>C. difficile</i>	157
Figure 81. RMMC plot of carbapenem-resistant and sensitive <i>Klebsiella pneumoniae</i>	159
Figure 82. PCA plot generated from a dataset comprising a variety of bacteria and yeast...	161
Figure 83. PCA and RMMC plot of <i>C. albicans</i> , <i>C. glabrata</i> , <i>C. lusitaniae</i> , <i>C. parapsilosis</i> and <i>C. tropicalis</i>	162
Figure 84. ANOVA plot for the model shown in Figure 83 comparing <i>C. glabrata</i> against the remaining 4 <i>Candida</i> species	163
Figure 85. Comparison of spectral profiles obtained for binary bacterial mixtures of <i>Staphylococcus aureus</i> and <i>Pseudomonas aeruginosa</i>	164
Figure 86. Full scan REIMS spectra obtained from different binary mixtures of <i>Escherichia coli</i> and <i>Bacteroides fragilis</i>	167
Figure 87. Experimental scheme to assess spectral reproducibility and self-identity of three different cell lines.	171
Figure 88. Mass spectral profiles between <i>m/z</i> 600-900 as obtained for HeLa, MES-SA, and SNB-19 cell line pellets, and 3-dimensional PCA plot for dataset comprising HeLa, MES-SA and SNB-19 cell lines over the spectral mass range of <i>m/z</i> 600-900.....	172
Figure 89. Hierarchical cluster dendrogram of the NCI60 cell lines including biological replicates	175

Figure 90. PCA plot of REIMS spectral profiles obtained for NCI60 dataset..	177
Figure 91. 2-dimensional PCA plot of averaged REIMS data collected from the NCI60 cells and cancer tissue samples	179
Figure 92. Box plot visualisation of <i>m/z</i> values found to be significantly increased in either bulk cancerous tissue or cancer cell lines using ANOVA.	180
Figure 93. Correlation analysis between <i>fads2</i> gene expression and mass spectral data	182
Figure 94. Binned peak intensity of <i>m/z</i> 786.54 as a function of scaled <i>fads2</i> gene expression, ratio of intensity values for peaks at <i>m/z</i> 768.55 and 770.57 as a function of <i>fads2</i> gene expression, Relative abundances of peaks <i>m/z</i> 768.55 and 770.57 in raw data for cell lines HL-60 and HT-29	183
Figure 95. PE(38:3)/PE(38:2) peak intensity ratio and PE(38:3) peak intensity as a function of FADS2 protein expression	183
Figure 96. Correlation analysis between <i>ugcg</i> gene expression and mass spectral data	184
Figure 97. MS/MS spectra for signals at <i>m/z</i> 842, 844, and 846 identified as glycosylated ceramides GlyCer(d18:1/24:2), GlyCer(d18:1/24:1) and GlyCer(d18:1/24:0)	186
Figure 98. Mass spectrometric signal intensity as a function of <i>ugcg</i> gene expression for glycosylates ceramide species listed in Table 26.	187
Figure 99. Number of binned <i>m/z</i> signals significantly higher in <i>Mycoplasma</i> -infected versus <i>Mycoplasma</i> -free samples in HEK and HeLa cell lines. Untreated <i>Mycoplasma</i> -free, <i>Mycoplasma</i> -infected, <i>Mycoplasma</i> -free and Plasmocin TM -treated, <i>Mycoplasma</i> -infected and Plasmocin TM -treated; HEK and HeLa samples as a function of PC1 and PC2 of PCA transformed untreated samples in the space of the 18 overlapping <i>m/z</i> signals.	188
Figure 100. Intensities of TIC nomalised and log transformed signal at <i>m/z</i> 819.52 in <i>Mycoplasma</i> -free or infected and Plasmocin TM -treated samples in HEK and HeLa cell lines.	190
Figure 101. DESI-MS ion image displaying tissue type distribution in colorectal tissue specimen A35 and H&E stained and histopathologically annotated section post-DESI.	212
Figure 102. Full scan mass spectra for colorectal adenocarcinoma, tumour surrounding stroma and necrotic tissue of same tissue section shown in Figure 101.	213

Figure 103. Single ion images for known taxon-specific markers (TSM) of sample A35 center of tumour.....215

Figure 104. Single ion images of taxon-specific markers (TSM) detected in healthy colorectal tissue specimen A56 10cm and H&E image of same tissue section.....217

List of Tables

Table 1. Instrumental parameters of Orbitrap Discovery and Xevo G2-S instruments used in this study.	55
Table 2. Thermo Exactive instrumental parameters used for DESI-MS imaging experiments.	65
Table 3. Results of experiment to determine amount of biomass needed for analysis.....	78
Table 4. Technical details of tested generators Covidien Force Triad and Erbe ICC300.	81
Table 5. Size of spectral database acquired as part of the presented study	109
Table 6. Number of spectral features detected for 28 of the most common clinical pathogens.	110
Table 7. Identified phospholipids detected in the mass range m/z 600-900 for all analysed species.	112
Table 8. Cardiolipin species that were identified for <i>Staphylococcus epidermidis</i> ATCC 12228.....	116
Table 9. Identified mycolic acids as detected in different <i>Corynebacterium</i> species.....	119
Table 10. Identified mycolic acids as detected in <i>Rhodococcus</i> species.	120
Table 11. Identified mycolic acids as detected in <i>Nocardia</i> species.	121
Table 12. Identified mycolic acids as detected in different <i>Mycobacterium</i> species.....	123
Table 13. Identified sphingolipid species in members of the Bacteroidetes phylum.	124
Table 14. Identified quorum-sensing molecules in <i>Pseudomonas aeruginosa</i>	130
Table 15. Rhamnolipid species commonly produced by <i>P. aeruginosa</i> strains.	133
Table 16. Surfactin species detected in positive and negative ion mode for <i>Bacillus subtilis</i> and <i>Bacillus pumilus</i>	135
Table 17. Lichenysin compounds detected in <i>Bacillus licheniformis</i>	137
Table 18. Dataset used to assess the specificity of a REIMS-based identification system ...	139
Table 19. Details of sample set analysed in the cross-platform comparison (Figure 74).....	147

Table 20. Confusion matrices for blind identification tests using different experimental platforms	148
Table 21. Sample set compiled to investigate the effect of liquid and solid culturing medium.	150
Table 22. Cell lines and respective number of replicates recorded for reproducibility assessment sorted by measurement day.	171
Table 23. Cross-validation results for SNB-19, HeLa and MES-SA cell lines based on PCA model comprising the first 4 principal components and using 3 nearest neighbour as classifier	174
Table 25. Cross-validation results for replicated NCI60 cell lines based on PCA model shown in Figure 90	178
Table 26. Mass spectrometric signals that show strong positive correlation with the <i>ugcg</i> gene expression for the NCI60 dataset.	186
Table 27. Values and tentative annotations of <i>m/z</i> signals that were found to be significantly higher in <i>Mycoplasma</i> -infected samples as compared to <i>Mycoplasma</i> -free samples in both HEK and HeLa cell line.	189
Table 28. Taxon-specific markers (given as <i>m/z</i> values) obtained for sample set detailed in Appendix 3.	199
Table 29. Molecular assignment of taxon-specific markers labelled with an asterisk in Table 28.	211
Table 30. Comparison of number of taxon-specific markers detected by DESI-MSI and copy numbers of bacteria detected by 16S rRNA sequencing-based community analysis for the same colorectal tissue specimens on phylum-level.	219
Table 31. Comparison of number of taxon-specific markers detected by DESI-MSI and copy numbers of bacteria detected by 16S rRNA sequencing-based community analysis for the same colorectal tissue specimens on class-level.	221

List of Appendices

Appendix 1. Determination of biomass needed for REIMS measurements for <i>S. warneri</i> , <i>P. aeruginosa</i> and <i>H. influenzae</i> . Measurement pre- and post-biomass uptake each performed in triplicate.	254
Appendix 2. List of bacterial species and number of respective spectral entries in REIMS spectral database as determined on 14.01.2015.	255
Appendix 3. List of fungal species and number of respective spectral entries in REIMS spectral database as determined on 14.01.2015.	261
Appendix 4. Sample set compiled for deriving of bacterial taxonomic markers.	262
Appendix 5. Detailed list of taxon-specific markers detected in colorectal tissue specimens using DESI-MSI.	271

Background and Objectives

Background

Mass spectrometry-based identification of microorganisms has been shown to be a universal tool and is in the process of revolutionising clinical microbiology. This is especially true for matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) which can be found in an increasing number of clinical microbiology laboratories around the planet. Although MALDI-TOF-MS significantly improves routine workflows by providing microbial identifications more quickly and at a lower cost per sample, it becomes increasingly apparent that the commercially available systems only provide a partial solution to the needs of clinical microbiologists. Knowledge of the identity of a pathogen causing disease facilitates adequate medical treatment, however, in routine clinical settings, more information exceeding the sole taxonomical identity of an organism is required. This is highlighted by the increasing prevalence of antibiotic resistance among bacterial pathogens which make alternative antibiotic treatments necessary than those that would be administered as standard choice of care. A quick decision on the most suitable antibiotic treatment can significantly shorten the duration of the infection and in some cases such as meningitis or sepsis directly impact the chances of survival of a patient.

Extensive research was performed using MALDI-TOF-MS to address this and other shortcomings in subspecies typing and led to solutions being developed for some of these problems such as the detection of β -lactamase activity to determine susceptibility to β -lactam antibiotics. However, these solutions are coming at the price of additional sample-preparation and culturing steps which do not currently fit into routine microbiological practice. Therefore many of the proposed solutions are unlikely to be implemented into routine clinical laboratories in the near future. In addition to limitations in subspecies typing, it is currently unlikely that MALDI-TOF-MS would find application directly to clinical samples without any sample preparation or culturing due to the strong background generated by the human proteome.

Thus, there is still an unmet need for a microbial identification method that would lead to species-level information while ideally simultaneously providing the crucial information on microbial phenotypes required by physicians to choose the adequate antimicrobial treatment. It is also becoming increasingly apparent that ribosomal protein-profiling might not be able to

deliver the desired sensitivity and specificity. Likely alternatives are sequencing of specific gene targets supplementing 16S rRNA ribosomal gene sequencing, or the use of small-molecular microbial cell constituents, the metabolome, which is more closely related to the phenotype of a given microorganism. Among those, cell membrane lipids have long been known to contain information that can be directly linked to microbial taxonomy.(1) This close connection is further highlighted by the specificity and longevity of fatty acid profiling. Fatty acid composition of bacterial cells however is strongly influenced by culturing conditions while the composition of intact phospholipids was shown to be more reproducible. Although it was often proven that species-specific lipid fingerprints were obtained for different types of bacteria, general applicability of this approach has so far not been demonstrated on a sufficiently large dataset. This is likely to be associated with intrinsic shortcomings of the lipidomic techniques applied such as excessive sample preparation, limited information content of mass spectra, limited reproducibility or limited sensitivity of the mass spectrometric technique applied.

The recently developed ambient mass spectrometry technique rapid evaporative ionisation mass spectrometry (REIMS) was demonstrated to generate highly specific phospholipid profiles of different biological tissue types. Spectral profiles showed the required reproducibility and specificity to distinguish between histological and histopathological tissue types using a training set of 302 patients and test sets of 81 patients covering large biological variance.(2) Therefore, it is a promising methodology for the development of a lipid-based, sample-preparation-free microbial identification system.

Objectives

- Development of a REIMS-based method for the lipidomic characterisation of microorganisms including bacteria and fungi. This comprises development of a sampling probe for analysing microbial cultures directly from the agar plate and characterising this method for instrumental parameter dependence.
- Build a large scale spectral database of mass spectra acquired from pure microbial cultures that contains the most common human pathogens. Approximately 5000 database entries will be acquired from the 100 most common clinical isolates. Microbial cultures will include bacteria (aerobic, anaerobic, microaerophilic) and

yeasts. For this purpose, a collaboration has to be established allowing regular access to microbial isolates from clinical origin.

- Assess the specificity of the method by characterising and identifying microorganisms from different phylogenetic levels using multivariate statistical methods. Specificity should be comparable to specificity obtained using commercially available MALDI-TOF-MS systems (>95% on species-level). Smaller sample sets will be analysed to assess specificity on subspecies-level such as antibiotic resistance, serotype or ribotype.
- Develop a workflow for the detection of bacteria directly in clinical specimens and in complex microbial mixtures. A taxon-specific biomarker approach is proposed, that consists of finding bacterial biomarkers that show specificity for a certain bacterial taxon such as genus, class, or phylum. Consequently these univariate biomarkers can be used to detect bacteria in complex matrices using REIMS or other compatible mass spectrometric techniques.
- Adapt REIMS-based method to the characterisation of cell lines to generate shotgun lipidomic profiles of the NCI60 cancer cell line panel. Characterise spectral information by correlating REIMS spectral profiles to protein and gene expression data.

1. Introduction

1.1. Microorganisms

Although largely invisible to the naked eye, prokaryotes (single-celled organisms) are an essential component of the world as we know it. Their total number in our ecosystem was estimated to be $4\text{-}6\cdot 10^{30}$ cells of which $1.2\cdot 10^{29}$, $2.6\cdot 10^{29}$, $3.5\cdot 10^{30}$ and $0.2\text{-}2.5\cdot 10^{30}$ are located in the open ocean, in soil, and in oceanic and terrestrial subsurfaces, respectively.(3) Prokaryotes comprise the two kingdoms Archaea and Bacteria, however, microorganisms is a wider category which is used for microscopic living organisms and also includes algae, fungi, protozoa, and viruses.

Microorganisms are ubiquitous and live in every part of the biosphere. They fulfil a range of essential tasks such as the decomposition of biomass for recycling nutrients within the ecosystem or the generation of atmospheric oxygen by algae as well as being a vital part of the nitrogen cycle. Bacteria and yeasts also play an important role in modern biotechnological applications such as food and beverage production (alcohol production, dairy products, etc.), biogas reactors, degradation of environmental pollutants, and the production of antibiotics, bioplastics, specific enzymes and commodity chemicals. A comparably small proportion of microorganisms are defined as pathogens and can cause acute and chronic illnesses to otherwise healthy higher ordered organisms. However, there is also increasing awareness of the vital role of microorganisms as symbionts and the resulting beneficial effects on the health of the same organisms including plants, insects, animals and humans. A paradigm change was the treatment of the microbiome as an entity in itself. This change was initiated by the discovery that a bacterium, *Helicobacter pylori*, was the causative agent of peptic ulcer disease, a disease which was previously thought to be stress-related. This led to significant improvements for correct diagnosis, prevention and treatment of the disease and was rewarded with a Nobel prize in physiology or medicine in 2005.(4-6) More and more research is conducted trying to elucidate host-microbial interactions and determine the contribution of the microbiome to health and disease development.

In humans, microbial genes outnumber human genes by about 100-fold and the majority of symbiotic bacteria are located on mucosal surfaces, especially within the gut (termed gut microbiota).(7) The gut microbiome co-develops with the human host from birth. The initial set of microorganisms is transferred from mother to child via a so-called “seed ecology”

which the child takes up by swallowing the vaginal microbiota during natural birth. In case of caesarean sections this initial set of microorganisms was found to be similar to the maternal skin microbiota.(8,9) The subsequent composition of the gut microbiota is influenced by the host genome, diet and life-style, disease state and use of antibiotics.(10) The composition of the gut microbiome was associated with several chronic diseases such as obesity(11), type 2 diabetes(12), inflammatory bowel disease(13), or cancer, especially in such organs that are continuously exposed to microorganisms such as the gastrointestinal tract.(14)

Due to the increasing importance associated with both microbial pathogens and commensals, the development of fast and reliable identification systems for microorganisms is a rapidly developing field.(15) Although a number of concepts have been proposed, none have proven to be specific, universal, fast and at the same time cheap enough to find widespread application. Until today, in routine clinical microbiology settings, identification of an isolate is still mostly accomplished by observing phenotypic characteristics such as colonial morphology, Gram-stain behaviour, and different enzymatic properties or carbon source utilisation patterns. However, these techniques are time-consuming, need experienced personnel and often lack specificity.

Bacterial species are defined by their 16S rRNA gene sequence, thus sequencing of the 16S rRNA encoding gene serves as the gold standard for bacterial identification and classification. The 16S rRNA is a part of the 30S small subunit in prokaryotic ribosomes and is approximately 1.5kbp in size. It is used for bacterial phylogeny due to its slow evolution rate which makes it a highly conserved region. It contains nine hypervariable regions which differ between bacterial species and are used to distinguish different taxa.(16) Although disputed, an often applied definition of a bacterial species is taken as organisms with at least 97% homology in their 16S rRNA gene sequences.(17) Partial or full 16S rRNA gene sequencing has the advantage of being culture-independent and thus is especially valuable for fastidious microorganisms. However, the sensitivity and specificity for direct sample applications varies considerably.(18) Despite of its role as gold standard, in some cases bacterial species cannot be confidently identified by their 16S rRNA gene sequence which can still make the application of additional techniques such as sequencing of further gene targets necessary.(19,20) One such example are the *Edwardsiella* type strains which share 99.35-99.81% of their 16S rRNA gene sequence while only showing 28-50% relatedness when the entire DNA is considered.(19) In addition, genotypic methods generally need extensive and careful sample preparation, they are comparably expensive and still need at

least several hours for identification. Due to these reasons, sequencing methods are rarely applied in routine clinical settings but mostly find application in reference laboratories.

Although not a taxonomical classification, the Gram-stain of bacteria is a widely used classification system, especially in clinical microbiological practice. Using Gram-staining, bacteria can generally be subdivided into two groups, Gram-positive and Gram-negative staining bacteria. The Gram-staining behaviour is determined by the capabilities of the peptidoglycan layer of the bacterial cell wall to retain the crystal violet dye. Peptidoglycan is a polymer consisting of glycan strands cross-linked by short peptides that forms a mesh-like layer outside the cytoplasmic membrane of most bacteria which constitutes the cell wall. There is a high diversity in the composition and sequence of the peptides in the peptidoglycan observed for different bacterial species and its structure can significantly vary with growth conditions. The function of peptidoglycan is providing structural strength, as well as counteracting the osmotic pressure of the cytoplasm. Gram-positive bacteria have very thick cell walls that consist of up to 90% dry weight of peptidoglycan which effectively retains the purple crystal violet molecules. In comparison, the cell envelope of Gram-negative bacteria usually contains only about 10% of peptidoglycan and does not retain the Gram-stain molecules.(21) As Figure 1 shows, Gram-positive bacteria have only one membrane formed of a phospholipid bilayer which is then surrounded by a thick layer of peptidoglycan (30-100 nanometers). Threading through the peptidoglycan structure are long anionic polymers called teichoic acids, which are composed largely of glycerol phosphate, glucosyl phosphate, or ribitol phosphate repeats. Wall teichoic acids are covalently attached to the peptidoglycan itself while lipoteichoic acids are anchored to the lipid head groups of the cytoplasmic membrane bilayer.(22)

In contrast, Gram-negative bacteria are diderm containing an inner and outer cell membrane separated by a thin peptidoglycan layer (7-8 nanometers). The outer membrane has an asymmetric structure where lipopolysaccharides (LPS) make up the outer monolayer of most Gram-negative bacteria. LPS consist of a glucosamine disaccharide with six or seven acyl chains (lipid A), a polysaccharide core and an extended polysaccharide chain that is called the O-antigen. LPS triggers innate immune response in humans which can lead to life-threatening endotoxic shock in bacterial septicemia.(23) Although not discussed here, both inner and outer bacterial membranes are hosting a variety of proteins that are embedded into the lipid matrix. Unlike eukaryotic cells, bacteria lack intracellular organelles, and

consequently, all of the membrane-associated functions found in eukaryotic organelles are thus performed on the inner membrane.

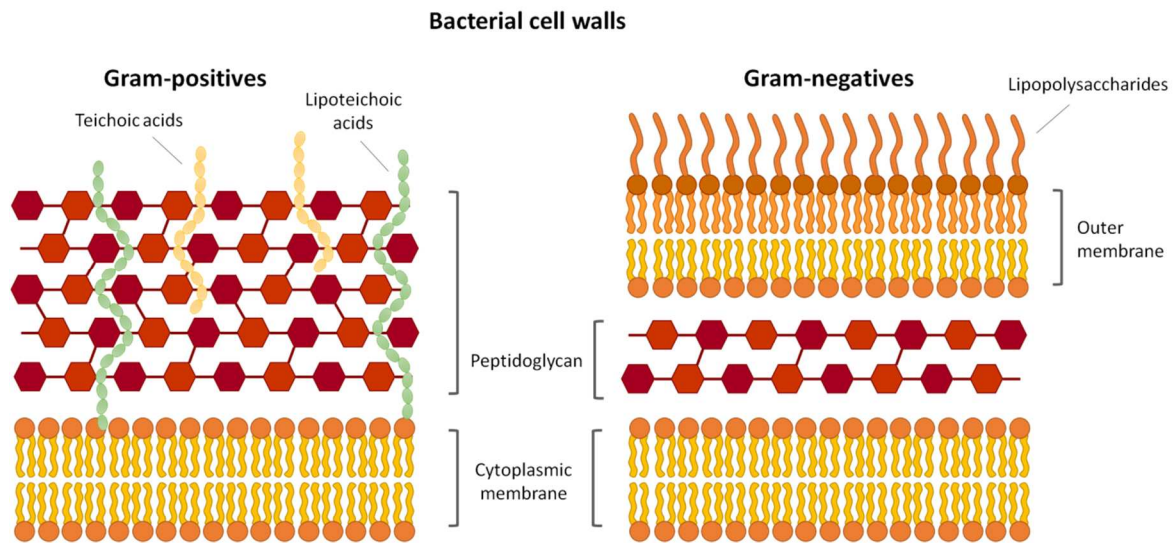


Figure 1. Artistic impression of bacterial cell wall for Gram-positive (left) and Gram-negative (right) bacteria (omitting proteins).

While the Gram-stain behaviour largely correlates with the main bacterial phyla (Gram-positive Actinobacteria and Firmicutes, Gram-negative Proteobacteria, Cyanobacteria, Fusobacteria, Bacteroidetes), this is not universally true. In addition, some Gram-variable and Gram-indeterminate groups have to be differentiated. Gram-stains are of significant importance in clinical microbiology settings as Gram-positives generally are more susceptible to antibiotics due to the lack of the outer cell membrane.

Bacteria are reliant on their cell envelope to protect them from unexpected and often hostile environmental conditions. The survival of bacteria depends especially on membrane lipid homeostasis and the ability to adjust lipid composition to acclimatise the bacterial cell to different environments. Although there is a considerable diversity of phospholipid structures in the bacterial world, most membrane phospholipids are glycerolipids that contain two acyl chains. These phospholipid acyl chains determine the fluidity of the membrane, which in turn influences many crucial membrane-associated functions, such as the passive permeability of hydrophobic molecules, active solute transport and protein–protein interactions.(24,25)

The most commonly encountered glycerophospholipids are phosphatidic acids (PAs), phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), phosphatidylcholines (PCs), phosphatidylinositols (PIs) and phosphatidylserines (PSs), as displayed in Figure 2. They share a common phosphatidylglycerolphosphate backbone but differ in the chemical nature of

their respective headgroups. Two fatty acids are attached to the glycerol backbone in sn1- and sn2-position and usually are present in various chain lengths of between 14 and 20 carbons. The formation of cyclopropane rings by methylation of *cis* double bonds and branched methyl groups are commonly occurring modifications among bacterial lipids. Bacteria have evolved a number of different mechanisms to control the *de-novo* formation of fatty acids and modify the structure of existing fatty acids in order to adjust membrane fluidity to the environmental requirements. These membrane lipid alterations are of importance not only for environmental bacteria, but also for pathogens that must adapt to the host environment.(25)

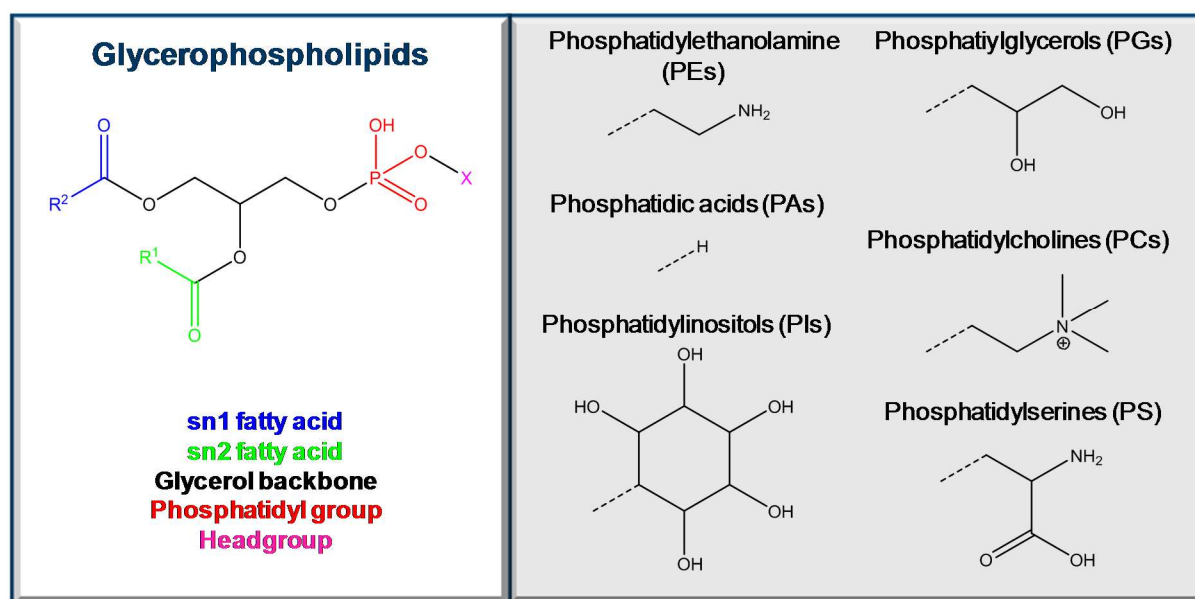


Figure 2. Most common glycerophospholipid structures encountered in prokaryotic and eukaryotic cells. X symbolises headgroup, R¹ and R² fatty acyl chain in sn1- and sn2-position, respectively.

1.2. Identification and characterisation of microorganisms using mass spectrometry

1.2.1. 1970-1995

Mass spectrometry (MS) gained attention for microbial identification more than four decades ago due to its intrinsic advantages of fast data acquisition, high sensitivity and high specificity.(26-28) In mass spectrometry, molecular compounds are transformed into gaseous ions which are subsequently separated according to their mass over charge ratio using electromagnetic fields. Different methodological approaches can differ in both the method used for ion generation (ionisation) and the method used for mass analysis (mass analyser).

Earlier studies reporting mass spectrometric signals generated from bacterial biomass for the purpose of bacterial identification were published in the first half of the 1970s and were using pyrolysis (thermal degradation of molecules in an inert atmosphere) followed by electron impact (EI) or chemical ionisation.(29,30) Meuzelaar and Kistemaker demonstrated in 1973 striking differences in the mass spectra of three bacterial species by coupling Curie-Point pyrolysis at 510°C to EI ionisation and a quadrupole mass analyser. However, due to the high pyrolysis temperatures, only small molecular weight ions were observed that were tentatively assigned to fragments originating from proteins, carbohydrates, lipids, nucleic acids, and porphyrins. All analysed bacterial strains could be readily distinguished using these mass spectra, however, generally the same mass spectrometric signals were observed even for taxonomically distant species.

Although the concept of bacterial fingerprinting was suggested in these early studies already, the widely accepted first milestone was set by Anhalt and Fenselau in 1975.(26) Using lower pyrolysis temperatures of 300-350°C, they reported the production of characteristic mass spectra from lyophilised pathogenic bacteria. Significantly larger molecules up to m/z 800 were observed especially from Gram-negative bacteria. These spectra showed a significantly higher amount of complexity both in molecular weight of the ions observed as in variance of mass spectral appearance. The observed signals were assigned to fragments of phospholipids and ubiquinones. This was a key development for bacterial identification using mass spectrometry as it was demonstrated that biomolecules from different pathogenic bacteria allow taxonomic distinctions to be made based on the characteristic mass spectral “fingerprint” signatures for individual organisms.

Another major methodological breakthrough that persists until this day was the introduction of fatty acid profiling of bacteria using gas-chromatography coupled to a flame ionisation detector (GC-FID).(31) In this methodology, volatile fatty acid methyl esters (FAMES) are prepared from whole bacterial cells using hydrolysis by sodium hydroxide (saponification) and subsequent acid-catalysed methylation before chromatographic analysis is carried out on the organic extracts.(32) Although there is a limited variability among the fatty acid species a certain type of bacterium will produce, the relative fatty acid distribution can vary considerably in relation to culturing medium, temperature, atmosphere, culture age and growth phase. Hence, standard growth procedures have to be followed for successful and reliable identification results. This system was commercialised by Microbial Identification Inc. (MIDI, Newark, Delaware, USA) under the name of Sherlock Microbial ID System

(SherlockMIS) and records both retention time and quantitative information on fatty acids with 9-20 carbons in length. However, the overall process is rather time-consuming with the chromatographic separation alone taking 20-30 minutes. More rapid methods were reported that use pyrolysis to perform an *in-situ* thermal hydrolysis methylation (THM) during pyrolysis followed by EI mass spectrometry for detection of FAMES to reduce the overall analysis time to less than 10 minutes. This system has been implemented into a field-portable mass spectrometer and was proposed for the detection of biological warfare agents.(33,34)

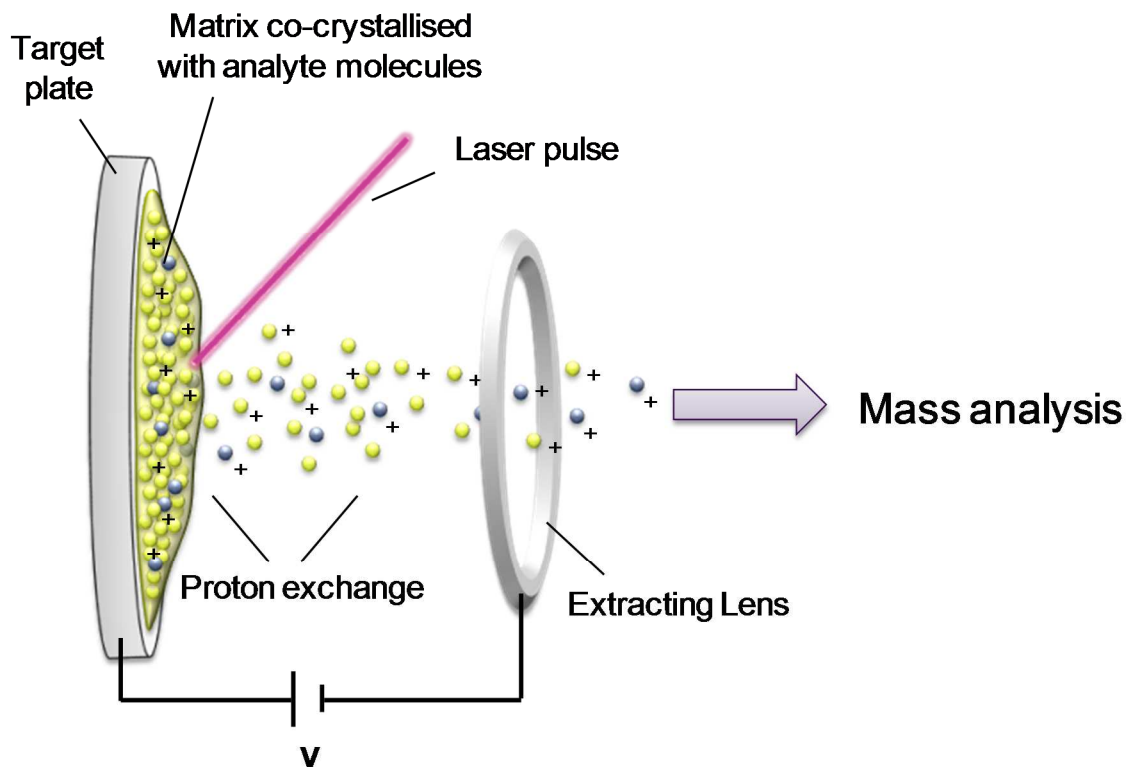
Throughout the 1980s, emerging desorption ionisation techniques (plasma desorption, laser desorption, fast atom bombardment) were evaluated for the generation of molecular biomarker ions from microorganisms. Considerable success was achieved using fast atom bombardment mass spectrometry (FAB-MS) which allowed monitoring of larger biomolecules such as intact complex glycerophospholipid species desorbed directly from lysed bacterial cells.(35,36) Each three Gram-positive and Gram-negative bacteria were analysed and could be distinguished based on their phospholipid content. Mass spectral signals up to m/z 1400 were observed, however, signal intensity for these higher molecular weight species was comparably low. An advantage of intact phospholipid analysis as compared to fatty acid analysis is that phospholipid patterns are expected to be more conserved than the fatty acid composition alone. While bacteria are able to change their membrane lipid composition by altering their fatty acid components, the phospholipid headgroup and thus the actual type of lipid will stay the same.(25)

1.2.2. 1996 onwards: soft ionisation techniques

1.2.2.1. Matrix-assisted laser desorption ionisation mass spectrometry

The end of the 1980s saw the introduction of two new major ionisation methods that soon became the most widely used ionisation techniques for mass spectrometric applications, matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI).(37,38) Both techniques soon enabled ionisation of very large molecules such as peptides, proteins and oligonucleotides.(39,40) The basic ionisation principles of the two techniques are depicted in Figure 3.

A) Matrix-assisted laser desorption ionisation



B) Electrospray ionisation

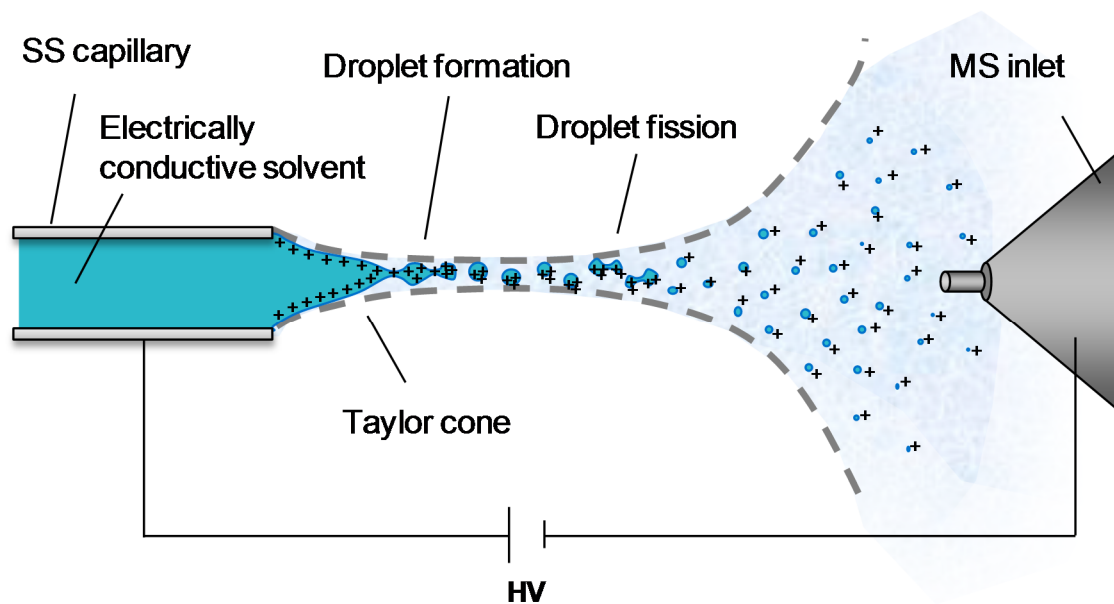


Figure 3. Schematic representation of ionisation processes occurring during A) Matrix-assisted laser desorption ionisation (MALDI) and B) Electrospray ionisation (ESI).

In MALDI MS, the sample is mixed with a thin layer of a matrix compound, most commonly small aromatic organic acids such as α -cyano-4-hydroxycinnamic acid (CHCA) or 2,5-dihydroxybenzoic acid (DHB). During application of the matrix layer, molecular constituents from the sample surface are solvated and co-crystallise with the matrix molecules. Subsequently, the sample is irradiated with a pulsed laser (most commonly a UV laser such as nitrogen (N_2) laser operating at 337nm). The matrix compound absorbs the laser energy which leads to ablation of matrix and analyte compounds from the sample surface. Protons are transferred from the matrix compound to the analyte which results in the formation of predominantly singly charged ions(41) which are subsequently separated according to their m/z ratio in the mass analyser. An alternative mode of ionisation for molecules which are residing as pre-existing ions within the matrix lattice (as is the case for many highly polarised compounds such as glycerophospholipids) is described by the lucky survivor model.(42) Upon laser irradiation and matrix evaporation, these pre-existing ions turn into gaseous ions which are subsequently pulled through a plume of gaseous matrix decomposition products by the extracting voltage. During this process, the majority of these ions are neutralised apart from a small proportion which are called the ‘lucky survivors’ and subsequently undergo mass analysis.

In electrospray ionisation (ESI), a solution of the sample in an electrically conductive solvent system is delivered through a capillary. When high electric potential is applied to the liquid sample, it leads to the formation of a Taylor-cone at the end of the capillary, which emits multiply charged fine droplets. The aerosol formation is often facilitated by an annular gas jet. Ionisation occurs either via the direct emission of ions from the surface of nanodroplets (‘ion evaporation model’) or the by the complete evaporation of the solvent leaving the charged residue behind (‘charged residue model’).(43) Based on chemical nature and size of the analyte molecules, singly and/or multiply charged ions can be generated. The electrospray plume is subsequently sampled by the mass spectrometer for mass analysis. While ionisation using MALDI is usually performed under vacuum conditions, electrospray ionisation is performed at atmospheric pressure while only mass analysis is performed under vacuum conditions.

In the mid 1990s, a decade after their invention, both techniques found their first application on the field of bacterial identification.(44-50) MALDI analysis of bacterial samples soon gained momentum and became the most widely used technique for bacterial characterisation and identification. This can be explained by the rich spectral content, high sensitivity, small

required sample amounts, and short sample preparation and acquisition times.(51) The inherent disadvantage of the electrospray method is that it requires clear liquid samples, which makes extraction and filtration/centrifugation steps necessary for bacterial biomass to avoid clogging of the electrospray tip. In contrast, MALDI allows the direct analysis of bacterial smears that are distributed on a target plate and coated by MALDI matrix and thus does not require any sample extraction or lysis. In most of the cases, the organic solvents and acidic conditions of the MALDI matrix will lead to lysis of the bacterial cells *in-situ*.(51)

In the context of bacterial characterisation, the vast majority of MALDI applications are based on the analysis of cellular proteins, predominantly protein profiling in the mass range of 2-20kDa. Proteins are the most abundant class of molecules in microorganisms, they are estimated to make up 60-70% of the dry weight in *E. coli* and bacilli, they ionise easily and therefore are readily observed using MALDI mass spectrometry. In addition, protein primary structure is directly determined by the DNA sequence which makes them more characteristic biomarkers on a genotypic level than lipids or metabolites which are strongly influenced by the phenotype of the organism. While in the next years there was continuing interest in using MALDI-MS for the characterisation and identification of microorganisms, little was known about the nature of the protein biomarkers in MALDI spectra of whole bacterial cells. Thousands of proteins are encoded by the genome, but only a small subset of these proteins actually are detected using MALDI-MS measurements. To understand why some proteins are detected and so many others are not, the molecular masses of proteins detected from *E. coli* K-12 were matched against masses predicted for proteins based on its genome. It was shown, that the biomarkers detected from whole cells by MALDI share the properties of high abundance, strong basicity, and medium hydrophobicity. In the case of *E. coli*, all the detected proteins originated from the cell interior, with about half of those coming from the ribosome.(52) In case of bacterial spores, comparably little mass spectrometric signals are detected, which are predominantly originating from the small acid-soluble protein family.(53,54)

As is the case for lipids, the growth medium had a significant influence on the proteins/peptides detected.(47) However, a common subset of bacterial protein signals remained detected and led to correct identification results when an adapted identification algorithm developed at the Pacific Northwest National Laboratory (Richland, WA, USA) was used.(55,56)

The two main commercial MALDI-TOF platforms were developed for routine clinical identification of microbes, the Bruker BioTyper instruments (Bruker Daltonics, Bremen, Germany) and the VitekMS series originating from a strategic partnership between bioMérieux Inc. (Marcy-l'Étoile, France) and Shimadzu (Kyoto, Japan). The inherent speed of analysis, high sensitivity and specificity combined with the good agreement with 16S rRNA sequencing led to the widespread use of MALDI-MS for the identification of microorganisms, both in research and in clinical microbiology laboratories.(15) These commercial MALDI-TOF microbial identification systems are approved for clinical microbiology routine use in the European Union since 2009, as well as in many other countries around the world and most recently gained approval by the US Food and Drug Administration (FDA).

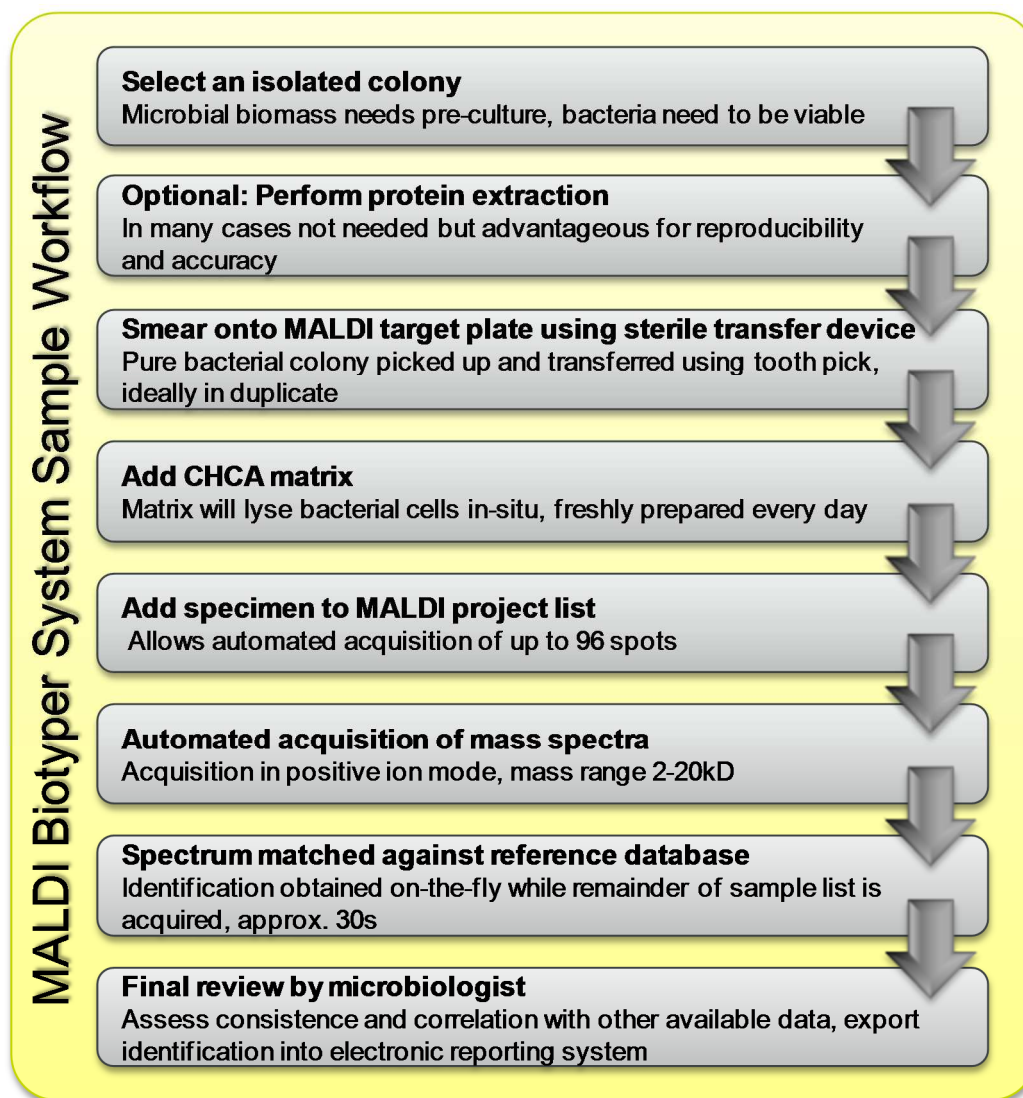


Figure 4. Routine MALDI Biotyper workflow used in clinical microbiology laboratories for the identification of bacteria.

The Bruker Biotyper is arguably the most widely used mass spectrometric microbial ID platform in clinical microbiology laboratories worldwide(15) and does give a simple identification result to the microbiologist based on a score between 0-3. Four values are required for the calculation of the identification score: the ratio of matched peaks to the number of peaks in the unknown spectrum, the ratio of matched peaks to the number of peaks in the reference spectrum, a peak weight ranging from 0 to 100 indicating species specificity and the correlation factor of the matching peak intensities. A maximum of seventy of the most intensive peaks are taken into account in the standard identification method. The obtained score is then converted into a log score, where the highest score 1000 is represented by a log score of 3. According to the manufacturer's recommendations, a log score <1.7 indicates non-reliable identification, a log score between 1.7 to 2.0 indicates a reliable ID on genus-level, while a log score above 2.0 indicates species-level identification.(57)

The general routine workflow proposed by Bruker Daltonics for the identification of microorganisms is displayed in Figure 4. The turn-around time for the displayed workflow can be as low as six minutes for a single isolate. However, due to the multiplexing of up to 96 organisms on a single MALDI target plate, actual reporting times from receipt of the pre-cultured plates are between 1-2hrs.

The success of this technology is due to the fact that these commercialised systems were demonstrated to give comparable or even superior results to conventional biochemical identification systems such as Vitek or API® (bioMerieux) in a number of validation studies.(58-61) Combined with short reporting times, these systems are able to reduce the average turnover time in clinical microbiology laboratories by about one day (24hrs). In addition, analysis costs are a fraction of those of conventional techniques once the instrument is installed.(61,62) For the majority of bacterial cells, MALDI analysis can be directly performed on intact cells and is providing excellent identification results. However, short additional extraction steps significantly increase the identification accuracy, especially in case of yeasts and Gram-positive bacteria.(58,63) Misclassifications usually occur due to missing database entries, insufficient data quality, mixed cultures, or too close taxonomical relationships.(58-61) Spectral databases are continuously improved and the Bruker database now includes 4,613 reference entries (version 3.3.1.0). Custom database entries can be included into a local database and were often shown to further improve identification results for under-represented species.(64-66)

In the past, relatively little work was performed on viruses, due to their high molecular weight (>20kDa) and limited number of expected proteins for a given virus. In addition, there is no sample preparation protocol of similar simplicity as for bacteria since viruses require a cell substrate to be cultivated *in vitro* which requires separation of viral proteins from substrate proteins once the cells are lysed.(67)

One major drawback of the routine MALDI-TOF-MS protocols involving bacterial protein profiling is that the method is not directly applicable to human samples as the extensive human protein background and low bacterial counts complicate bacterial detection. However, extensive research is performed to develop direct “on sample” applications and successful attempts were reported for samples such as positive blood culture bottles, urine, and cerebrospinal fluids. In healthy individuals, these are considered sterile samples and a large proportion of infections in such samples are caused by a single organism, which facilitates detection by MALDI-TOF-MS as the components of bacterial mixtures currently cannot be reliably identified. All these protocols involve a range of lysis and centrifugation steps to lyse the interfering human cells and separate and concentrate the bacterial cells for further analysis.(68-71) The largest impact on clinical practice can probably be obtained in case of sepsis, where earlier detection of the pathogen was shown to significantly improve survival rates.(72) Ferroni *et al.* developed a protocol that leads to possible identification of the pathogen in 30 minutes from the point when the blood culture flags positive if the workflow is considered in case of a single sample only (longer times required when analysis is multiplexed).(73)

Although the introduction of MALDI-TOF-MS for bacterial identification made a huge impact on clinical microbiology routine, there are still areas where the approach of profiling bacterial protein patterns does not give the desired results as in case of bacterial mixtures, or sub-species typing such as serotyping or ribotyping. In some cases, tandem mass spectrometry (MS/MS) can be utilised to complement the direct MS analysis of intact proteins for the detection and identification of bacteria.(74-76) In case of the top-down proteomics approach the intact biomarker protein is fragmented to provide sequence-specific fragment ions. In a following homology search, the MS/MS spectrum is compared to spectra generated *in silico* from all the proteins sequences stored in a database whose masses match the observed *m/z* value. This way the unknown microorganism is identified based on the identification of an individual biomarker. Using this approach, the individual components of a mixture of two *Bacillus* spores could be successfully identified.(77) This example also

exemplifies that tandem MS may be particularly advantageous when only a few protein biomarkers are present in a regular mass spectrum by providing further specificity. However, as MALDI ionisation leads to predominantly singly charged ions, obtained fragmentation spectra generally lead to comparatively low protein sequence coverage due to lower fragmentation efficiency.

In case of a bottom-up proteomics approach, proteins fractionated from lysed cells are cleaved to peptides which are fragmented and analysed in a MS/MS experiment. Identified peptides provide identification of proteins and, in turn, proteins provide identification of the microorganism. This bottom-up approach requires residue-specific cleavage of biomarker proteins as an extra step carried out enzymatically or chemically *in-situ*. This strategy provides another practical solution to the challenges of mixtures of microorganisms, but works best in such samples exhibiting relatively low spectral complexity. Since this approach requires lengthy sample preparation, top-down proteomics is favoured in most of the cases. In addition, the probability of false matching of a potential microorganism is generally lower in case of a top-down approach.(75) In both cases, identification relies on the detection and identification of unique, strain-specific biomarker ions versus peaks that can be observed in all bacteria of a species or genus.(74) Using a bottom-up methodology, differentiation of *Lactobacillus* subspecies was successfully demonstrated with an approach termed shotgun mass mapping (SMM) by performing a tryptic digest of isolates and subsequently comparing the spectra to a home-built reference library.(78)

1.2.2.2. Electrospray Ionisation Mass Spectrometry

The most important application in the field of clinical microbiology using electrospray ionisation mass spectrometry is arguably rapid ESI-MS analysis of polymerase chain reaction (PCR) products. This is in sharp contrast with the typical several hour time-scale of a gel electrophoresis experiment. DNA sequence constitutes the most unique biomarker for each kind of organism, but DNA copy numbers per cell are low (typically one copy per cell) without amplification which complicates its detection. In 1996, Wunschel *et al.* were the first to report the ionisation and detection of a double stranded PCR product by ESI-MS.(48) The mass measurements using ESI-FTICR-MS matched the theoretical masses and were consistent with results obtained by gel electrophoresis. Based on this methodology, the TIGER approach was developed. TIGER stands for “triangulation identification for the genetic evaluation of risks” and combines nucleic acid amplification with high-resolution

ESI-MS detection.(79,80) The essential aspect of this approach is the use of PCR primers targeting broadly conserved regions of microbial genomes that flank variable regions, namely ribosomal and housekeeping protein genes. The PCR products are in the size range of 80–140 base pairs and can be measured using high resolution FTICR- or TOF-MS in a high-throughput mode. The mass accuracy must be better than 20 ppm to confirm the correct base composition. The microorganisms are identified by comparison of the derived base composition with available genome sequences in databases. To enhance confidence, several genomic regions are analysed to identify the organism. The general workflow comprises extraction of the microbial DNA, broad-range PCR amplification, ESI-MS of the amplicons and computerised triangulation.(81) Besides bacteria, the approach is equally well suited to the identification and quantification of viruses, fungi, and protozoa directly from clinical samples.(15,79) This system was developed as Ibis T5000 by Ibis Biosciences Inc., and was succeeded by the PlexID system (Abbott Molecular Inc., IL, USA). However, due to high costs, large space requirements and the limited acceptance by the microbiology community, this system eventually got discontinued in September 2012.(15) Based on the same principle, Abbott recently launched the IRIDICA platform for the detection of >750 bacterial and *Candida* species, more than 200 species of fungi, and >130 viruses in less than 6hrs from receipt of the clinical sample according to the manufacturer. The process includes automated DNA extraction, PCR set-up, amplification and amplicon purification followed by ESI-TOF-MS for amplicon identification. Its universality, especially for fastidious organisms, is a huge advantage compared to other diagnostic approaches, however, the instrumentation and cost per sample are comparably expensive and the instrument has very high space requirements.

1.2.3. 2004 onwards: Ambient mass spectrometry techniques

Ambient mass spectrometry is defined as mass spectrometric analysis performed without active sampling or sample preparation, while the object of interest is not removed from its ambient environment. The first ambient ionisation method was desorption electrospray ionisation mass spectrometry (DESI-MS) introduced in 2004.(82,83) Definition of ambient mass spectrometry implicates that samples of arbitrary shape can be analysed in their native state without introduction into the mass spectrometer vacuum system. In case of microbiological applications this means that bacteria can be analysed directly from the Petri-dish. Many ambient ionisation methods furthermore are minimally invasive and hence particularly well suited for *in-vivo* studies. Due to its versatility and electrospray-like

ionisation characteristics, DESI still today is the most commonly used of these techniques, however, other techniques that are commonly used include Direct Analysis in Real-Time (DART)(84), Low-Temperature Plasma Ionisation (LTP)(85), Easy-Sonic Spray Ionisation (EASI)(86) and Laser Ablation Electrospray Ionisation (LAESI)(87).

Until 2009 alone, more than 30 individual ambient ionisation methods have been published(88) and many of those were used for bacterial characterisation,(89) resulting in an ever increasing number of studies involving a wide range of emerging ionisation techniques that indicate the breadth of possible applications of these methods in microbiology. These applications include the analysis of bacterial metabolites and phospholipids *in-vivo* and directly from the Petri-dish using DESI(90) and liquid microjunction-electrospray setups.(91). Nanospray desorption electrospray ionisation (nano-DESI) is a surface liquid microjunction-electrospray approach, which was successfully applied to monitor intact molecular species at single colony-level and was further used in numerous studies demonstrating the power of networking algorithms for high-throughput identification and characterisation of molecules secreted into the surrounding medium by bacterial colonies. In addition, it is suited for the two-dimensional mapping of bacterial metabolites directly from the Petri dish as demonstrated in case of the spatially resolved detection of glycolipids and bacterial metabolites directly from living colonies and the study of interactions between neighbouring colonies and in mixed-species biofilms.(92-94) Methodologically similar to other liquid junction technologies, however, strictly speaking not an ambient ionisation method, 'contact' liquid extraction surface analysis (LESA) followed by nanospray ionisation (NSI) was used to identify bacterial proteins from *Escherichia coli* colonies directly off the Petri-dish using top-down proteomics.(95) This technique might allow monitoring of protein changes in response to environmental stimulants in real-time. Electrospray techniques are generally better suited for top-down proteomics compared to MALDI due to the multiple charge states of proteins leading to more information rich protein fragmentation spectra. Parsiegl *et al.* were using LAESI-MS to detect phycobilisomal antenna proteins in small cyanobacterial populations and were able to determine the ratio of phycocyanin and allophycocyanin in the antenna complex, the subunit composition of the phycobiliproteins, and the tentative identity of over 30 metabolites and lipids.(96)

Direct analysis in real-time (DART) is the most commonly used plasma-based ambient ionisation technique and was used for direct profiling of fatty acid methyl esters (FAMES) from whole bacterial cells, thus enabling fast FAMES profiling without the need for

chromatographic separation.(97) The methodology is expected to give lower specificity than a chromatography-based approach; however, this might be circumvented using ion mobility to separate isobaric fatty acids. DART-MS was also applied to monitor production of prenylated indole alkaloids in the hyphae of *Malbranchea graminicola* fungus grown on agar plates which led to the detection of two novel chlorinated metabolites.(98)

Most ambient mass spectrometric methods strongly favour the detection of low molecular weight species with m/z ratios of less than 2000 and thus are not well suited for the detection of proteins. However, a growing interest in lipidomics, and the advent of ambient ionisation techniques as an easy means to generate lipid profiles gave new momentum to lipid profile-based identification of microorganisms. Many different ionisation techniques, including fast atom bombardment (FAB)(35), electrospray ionisation (ESI)(44,49,99), MALDI(100,101), desorption electrospray ionisation (DESI)(102,103) and paper spray(104) have been used among many others to demonstrate that different bacteria have species-specific phospholipid profiles. However, none of these methods have been shown to possess the specificity and robustness required to serve as the basis for a general lipid-based identification system.(89)

1.2.3.1. Desorption Electrospray Ionisation Mass Spectrometry

Desorption Electrospray Ionisation (DESI) was first described in 2004 and proved a landmark in mass spectrometry as the first ambient ionisation technique.(82,83) In DESI, a pneumatically-assisted electrospray is directed onto the sample from a short distance. Due to the impact of the primary electrospray droplets, a liquid film forms on the surface of the sample and dissolves the analyte molecules from the sample surface. Upon the impact of further primary droplets into this liquid film, secondary droplets get ejected and aspirated by an extended inlet capillary. This process is schematically depicted in Figure 5.

Since 2004, DESI MS is commercialised by ProSolia Inc. (Indianapolis, IN, US) and found various applications as for the detection of explosives,(105,106) characterisation of pharmaceuticals,(107) screening of biological fluids for drugs and illicit drugs,(108,109) direct analysis of microorganisms(90,103,110) or hyphenation to solid phase extraction(111,112) and solid-phase microextraction techniques.(109,113) However, the most prominent application of DESI-MS is arguably metabolic tissue imaging where (after MALDI-MSI and SIMS-MSI) it constitutes the third most often used imaging technique.(114-118) Unlike other tissue imaging methods such as staining procedures (as haematoxylin and eosin staining, immunohistochemical stains or fluorescent microscopy),

mass spectrometry imaging (MSI) has the major advantage of being able to detect thousands of different molecular constituents in a single experiment in a spatially resolved fashion. It is an especially valuable tool when combined with high-resolution mass analysers which significantly increase the specificity of the methodology by resolving nominally isobaric compounds. Thousands of compounds present in a sample can hence be analysed and visualised without *a priori* knowledge.

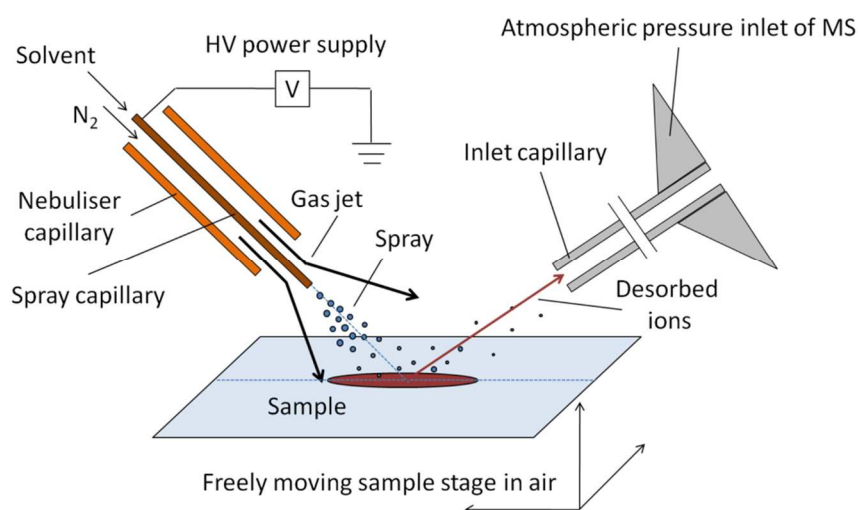


Figure 5. Schematic representation of Desorption Electrospray Ionisation.

Since none of the MSI techniques include chromatographic separation, detection of certain analytes of interest is often impeded by ion suppression. Due to the histological variance of suppression effects, the MSI signal intensity distribution does not necessarily follow the actual concentration-distribution of the analyte. Due to this fact, mass spectrometric imaging experiments are at best semi-quantitative. For this reason, many mass spectrometry imaging studies rely on multivariate statistical pattern recognition algorithms to distinguish between tissue types rather than relying on the different distributions of univariate markers.(119-122) DESI imaging studies were performed for various tissue types as normal breast tissue and breast cancer,(120,123) lymph node metastasis,(124) colorectal cancer,(119) or drug and drug metabolite distribution studies.(117,125,126)

Unlike other commonly applied techniques for tissue imaging such as MALDI or SIMS-MSI, DESI can be used without matrix-application and transfer of the samples into vacuum. However, ambient analysis in turn requires the samples to be stable at room temperature. Spectral resolution, although initially restricted to 200 μm and still much lower than those reported for MALDI and SIMS-MSI, by now is comparable to commercially available standard MALDI sources (50-100 μm). The smallest feature detected on rat brain using DESI

was determined to be 35 μ m(127) and further improvements in spatial resolution are to be expected by further decreasing solvent flow rates and nebulising gas pressures and optimising sprayer geometry, however, ultimately spatial resolution is limited by Coulombic repulsion of the charged primary droplets.

1.2.3.2. Rapid Evaporative Ionisation Mass Spectrometry

Rapid evaporative ionisation mass spectrometry (REIMS) was first reported in 2009 with the main rationale of intra-surgical tissue identification.(128) It allows near real-time characterisation of human tissue *in-vivo* by analysis of the aerosol ('surgical smoke') released during an electrosurgical dissection. Although ambient mass spectrometric techniques in general allow the analysis of samples in their native state and environment, none has been found applicable for *in-vivo* applications directly on patients yet. REIMS is efficiently closing this gap using electrosurgical instruments as ion sources. This coupling and its application for intra-surgical tissue identification is known as the intelligent knife (iKnife).

Two major groups of electrosurgical tools are used in clinical practice which follow a monopolar or bipolar principle as depicted in Figure 6. In case of monopolar electrosurgical tools, dissection is made using a sharp handheld electrode while the sample (or patient) is lying on a large surface counter electrode (see Figure 6A). This setup is employed in most open surgeries where the electrosurgical knife is used to dissect tissues and to cauterise blood vessels. In case of bipolar electrosurgery, both electrodes are handheld following a forceps-like configuration (Figure 6B and C). This setup allows more accurate and less invasive operation and is predominantly used for surgical interventions on the brain.

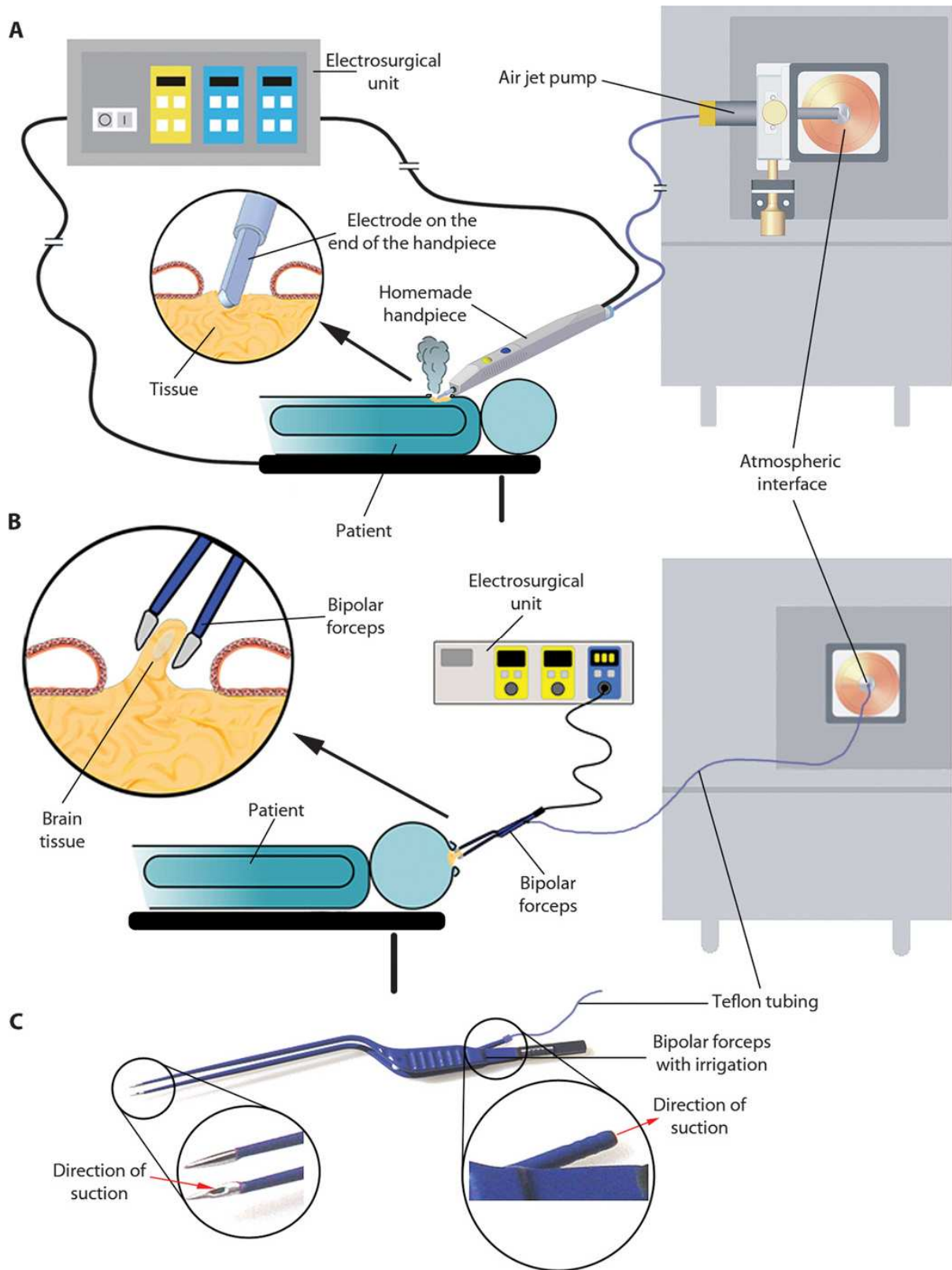


Figure 6. Schematic representation of REIMS instrumentation and data collection. A) monopolar and b) bipolar surgical ion source and ion transfer setups for REIMS experiments. C) Aerosol aspiration using commercially available bipolar forceps. From Ref. (2). Reprinted with permission from AAAS.

Both setups are using radiofrequency alternating current to evaporate tissue material by Joule heating, which produces an aerosol containing tissue decomposition products. In clinical practice until today, this aerosol is largely regarded as potentially hazardous by-product of electrosurgical interventions. However, REIMS might induce a paradigm shift by using this formerly disregarded aerosol to derive histological tissue identification and assess tumour margins during surgical interventions.(129) This is accomplished by introducing the surgical aerosol into a mass spectrometer for analysis to obtain a tissue fingerprint that can then be compared to a database of REIMS spectra of different tissue types for identification.(130) This process can be completed in less than 0.5-2.5 seconds and thus offers an attractive alternative to intra-operative tissue histology commonly used during cancer resections. Intra-surgical frozen section histology requires a minimum of 20-45 minutes in which the operation has to be paused while in parallel a large strain is posed on histopathology departments to process the specimens and determine whether the tumour was completely resected.(2,129) REIMS-based tissue identification has recently been demonstrated to give at least equivalent tissue identification accuracy as compared to frozen section histology on a set of 81 intra-operative interventions. In addition to intra-operative tissue ID, REIMS spectral profiles were shown to vary between distinct histological tumour types and also between primary and metastatic tumours.(2)

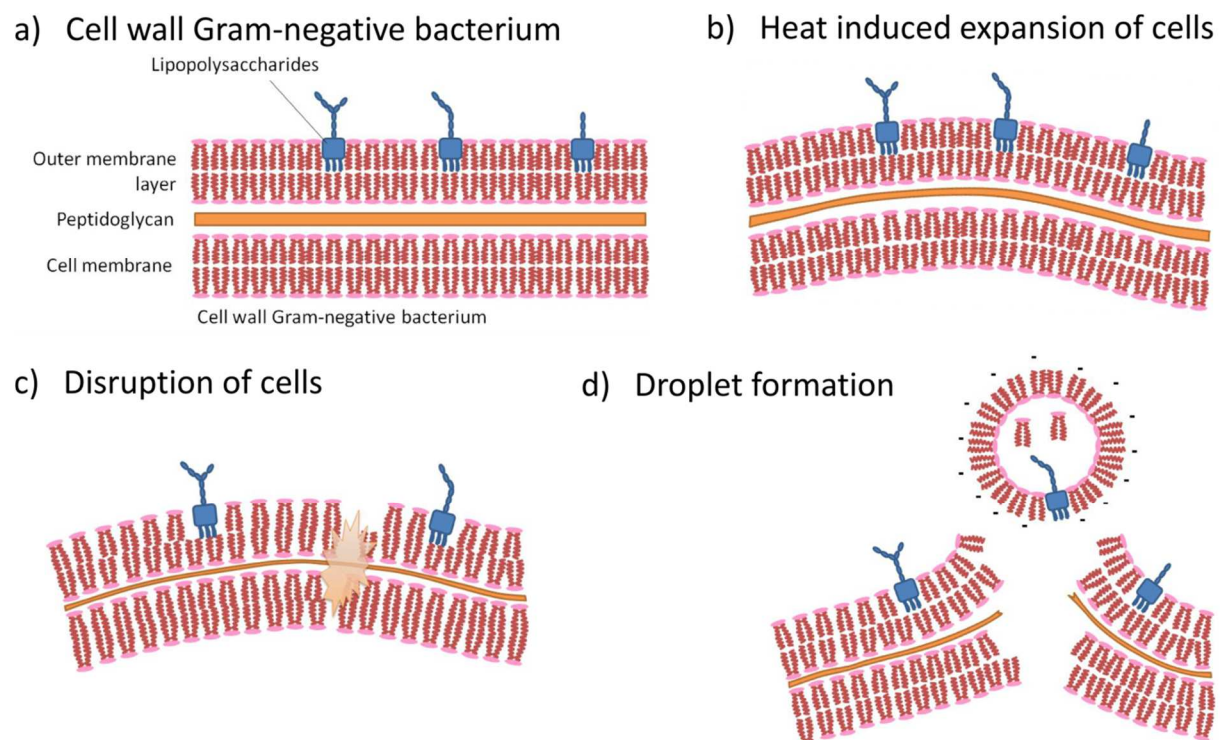


Figure 7. Schematic representation of desorption mechanism in REIMS for the example of Gram-negative bacterial cells.

No mechanistic studies on the REIMS mechanism have been published so far, however, the tentative rapid evaporative ionisation mechanism can be considered as heat-induced disruption of cells as depicted in Figure 7 for the cell wall of a Gram-negative bacterium. Upon exposure to the radiofrequency (RF) alternating current, the aqueous part inside any biological cellular material expands due to the high temperature. Water will evaporate and the cells will expand until they burst and lipid covered aqueous droplets are formed. The lipids originating from the cell membranes will form droplets due to their amphiphilic character. These droplets (molecular clusters) are aspirated by the mass spectrometer and subsequently undergo surface collisions inside the ion optics of the instrument where they form molecular ions. Due to the nature of the ionisation mechanism, ionisation of molecules is preferred if they are either charged prior to ionisation or highly polarised.

The REIMS method shows high methodological resemblance to pyrolysis approaches, however, unlike pyrolysis systems, REIMS is working under ambient pressure and provides soft ionisation yielding predominantly molecular ions. The observed soft ionisation is tentatively associated with collisional cooling of the produced ions occurring under atmospheric conditions.

1.3. Statistical methods for analysis of mass spectrometric profiling data

1.3.1. Biomarker discovery using univariate statistical methods

Biomarker discovery is usually performed using univariate tests in which m/z values in each sample are compared against each other in a one-by-one fashion. For a two class system that can be performed using traditional t -test, however, for multi-class systems, analysis of variance (ANOVA) is better suited. The result of a one-way ANOVA analysis is a p -value for each m/z feature and gives a measure of the probability of the observed difference being created by chance. A p -value of 0.05 refers to a false positive rate of 5% which is generally considered as significant.(131) However, even if there is a significant difference detected between the group means, there might be still be a considerable overlap between the individual data points which determines the diagnostic value of a biomarker. For this purpose, box-plots are supplied together with ANOVA results throughout this work to assess group overlap.

While ANOVA determines whether there is a significant difference between at least two of the different group means, it does not provide information about which group is significantly different. Thus, a post-hoc test such as the Tukey's honest significant difference test (Tukey's HSD test) is necessary in order to assign the observed difference to a certain group.

1.3.2. Multivariate statistical methods

Multivariate statistical analysis methods on the other hand are more suitable if groups are to be distinguished using correlated changes in relative abundance of several m/z features simultaneously. They are tools to identify correlations in high dimensionality datasets and reduce their complexity by identifying those features of discriminatory power. Multivariate statistical methods taking the relative distributions of several features simultaneously into account allow for finding similarity and differences in a dataset which cannot be captured by collection of individual biomarkers, but only by a combination of the distribution of a multitude of markers. This is particularly of use when analysing closely related biological specimens such as different bacterial subspecies, or different human cancer cell lines. In such cases, most metabolic (gene expression, protein expression, metabolite expression) constituents are expected to be present in most samples of the dataset in question, thus finding a single univariate biomarker to unambiguously identify every sample group is in many cases unlikely. Balog *et al.* found that for REIMS spectra of mammalian tissues, each m/z feature investigated was shared by >80% of individual samples analysed.(130)

1.3.2.1. Principal component analysis

Unsupervised statistical methods are used in order to identify the largest variance in a given dataset without introducing bias. Principal component analysis (PCA) is a dimensionality reduction technique aiming to reduce the number of variables (in case of mass spectrometry profiling data these equate to m/z features) while keeping the loss of information to a minimum by grouping variables with high covariance. A linear orthogonal transformation is performed in order to maximise variance. This results in orthogonal principal components (PCs) describing the variance observed in the dataset. The first principal component (PC1) describes the largest observed variance, while PC2 describes the next largest, PC3 the next largest and so forth.(131) The number of principal components to be visualised is either chosen by the user or determined by the amount of variance to be retained (for instance >95% of variance). Data after PCA is usually presented in the form of score plots (commonly

displayed as scatter plots), spectral loading plots and percentage of explained variance for a given component. The latter is of particular importance as it forms a measure for how much a given PC contributes to explaining the overall variance in the data. The loading plot contains the contribution to the overall explained variance as a function of m/z value for a given principal component. Contribution can be both positive and negative. However, the largest variance in the data is not necessarily related to differences between the individual sample cohorts in the data, but can also originate from other, biologically non-relevant factors such as systematic spectral variation introduced by sample preparation. In case of analysis of microorganisms, such factors can originate from different culturing conditions such as liquid versus solid culturing medium, culture age, culturing atmospheres, different culturing media or even different batches of culturing medium. Thus, using PCA is an excellent way to inspect the data for the presence of such factors that might introduce a bias into the dataset.

Due to the applicability of PCA as a dimensionality reduction strategy, it is often applied as a data pre-processing method before the use of further downstream data analysis algorithms, such as supervised multivariate methods. By retaining only the components explaining the largest variance in the data, the assumption is that contribution to the variance by spectral noise is removed due to its uncorrelated nature.

PCA has been applied to a variety of mass spectral profiling techniques and is commonly used for metabolomics approaches.⁽¹³²⁾ Most profiling studies cited in this work will contain data analysis using PCA, both for profiling of bacteria and other tissues using REIMS, DESI and MALDI^(102,110,128,130,133,134) as well as for imaging mass spectrometry data sets⁽¹³¹⁾ as it implies to the reader how suitable a certain methodology is for a given purpose. The larger the proportion of variance in the produced data is generated by the biological phenomenon in question, the better the presented methodology is suited for assessing this biological phenomenon.

Data visualisation of multivariate datasets can be challenging. Principal component analysis is particularly well suited to analyse sample sets of high complexity containing several hundreds of data points using two- and three-dimensional scatter plots allowing simultaneous visualisation of two and three principal components, respectively. However, although principal components do not have to be consecutive components, no more than three components can be displayed at one time thus limiting the extend of the variability that can be visualised at a time.

1.3.2.2. Hierarchical cluster analysis

Clustering methods such as hierarchical cluster analysis (HCA) are commonly used to assess the similarity of profiles produced. It is a common tool for displaying phylogenetic relationships in unicellular organisms based on their DNA(135) or 16S rRNA gene sequences(136), protein(137) or gene expression profiles(138), or metabolic profiles.(134,139) Similarity is displayed in a multi-level tree-like structure, a so-called dendrogram, where more similar profiles (samples/strains) are located on the same branch of the tree. Larger and smaller variation is explained by the different branching levels where clusters on one level are joined as clusters on the next level. To produce these dendrograms it requires the choice of a similarity measure and a 'linkage' calculation method. While the former determines the similarity between every pair of objects (here mass spectral profiles) in the data set, the latter determines how lower-level clusters are located on higher-order clusters by determining the proximity of objects to each other and grouping the closest pairs into binary clusters until all the data points are linked together in a hierarchical tree. The most common similarity measure is the determination of the Euclidean distance between the data objects. In a final step the data is partitioned by determining the number of clustering levels either by detecting natural groupings in the hierarchical tree or by arbitrary user-defined selection.

Hierarchical cluster analysis is often performed after data normalisation and data reduction techniques in order to reduce the biologically non-relevant information such as spectral noise.(131) Several branching levels in a single diagram allow for better visualisation of multi-dimensionality in the data than three components might allow in PCA which makes HCA dendrograms an intuitive way to visualise similarity between data points. However, it is best suited for comparatively small numbers of individual data points as visualisation of dozens to hundreds of data points quickly gets challenging and removes the intuitive nature by overcrowding the dendrogram.

1.3.2.3. Discriminant analysis

Supervised statistical methods such as discriminant analysis (DA) take the user-defined group membership of the individual data points into account. Unlike unsupervised PCA, DA will not search for the largest variance in the data in general, but it will instead search for the largest variance in the data explaining the group membership of the data points (thus it is called supervised). Linear discriminant analysis (LDA) attempts to find those linear

combinations of m/z features that discriminate between the different given groups (this can be two or more) by maximising the inter-class difference while simultaneously minimising intra-class variability.(140) The loadings function will give information on those m/z features that contribute the most (both positively and negatively) to discriminating between the different groups. It is often used in combination with prior performed PCA thus only subjecting a certain number of principal components to LDA. However, this combination is based on the assumption that the variance between the different groups is also one of the main contributors to the overall variance in the data. The number of linear discriminant functions calculated is usually determined as the number of groups minus one. Thus, one linear discriminant function is used to distinguish between two groups, two discriminant functions are used to distinguish between three groups and so on. A discriminant model determined in this manner can then be used to classify new samples. For this purpose, often a training set and a test set are composed. The usually smaller training set is comprised of different samples from a given group system (for instance different species of bacteria or different human tissue types) from which a discriminant model is built and then used to classify unrelated samples into those groups it is built to distinguish. The quality of the model is then determined by the number of correct classifications of unknown samples. In such cases where there are not enough data points available per group to built independent training and test sets, often cross-validation methods are used instead. In leave-one-out cross-validation, each data point is left out of the model building process at a time and subsequently projected into the resulting model and classified. This can be performed sequentially for all data points or a subset of randomly selected data points only (for instance 10% of the overall data points). Model quality and robustness again is quantified based on the number of accurate classifications.

A PCA-LDA combination has been previously used for most REIMS studies on mammalian tissue specimen. The 300-dimensional mass spectral data (m/z 600-900 at 1Da bin size) was subjected to PCA in order to reduce the data to the first 60 principal components. This data was then subjected to LDA for tissue classification experiments.(2,130) In one such study by Balog *et al.*, 97% tissue identification accuracy was achieved using separate training and test sets. Although the method was proven sensitive to systematic data variation such as caused by age or different diets in rats, tissue identification capabilities were not impaired.(130)

Maximum margin criterion (MMC) is another supervised multivariate algorithm that was initially developed for face analysis.(141) Unlike in LDA, its feature-extracting capacity does not depend on the nonsingularity of the within-class scatter matrix which is especially

advantageous in case of small sample sets. This methodology was adapted to a Recursive Maximum margin criterion (RMMC) by Veselkov *et al.* and shown to be applicable to lipidomics-based mass spectrometry imaging data.(122) Class-specific m/z signatures are derived based on weighted combinations of m/z feature patterns. Simultaneously, as in LDA, the ratio of inter-class to intra-class variability is maximised. However, while LDA maximises the distance between the individual group means, in RMMC the margin between individual groups is increased by maximising the distance between their most outer data points thus improving group separation.(122) The RMMC approach avoids selection of an optimal number of PCA-derived components before applying LDA and is therefore avoiding model over- or under-fitting. RMMC is used in this thesis alongside PCA-LDA as main supervised data analysis strategies.

2. Materials and Methods

2.1. REIMS analysis of microorganisms

2.1.1. Culturing of microorganisms

All culturing of microorganisms was performed at the Department of Microbiology, Imperial College Healthcare NHS Trust, Charing Cross Hospital (London, UK). Clinical isolates were routinely collected from clinical microbiology work by trained NHS staff. Most of the microorganisms analysed during this study were previously isolated from blood cultures, identified using a Bruker Biotyper instrument, and stored on beads and in glycerol broth in a -80°C freezer. For REIMS analysis, microorganisms were grown either from fresh cultures or from beads on a range of solid agar-based media commonly used in clinical microbiology settings. Media were purchased from Oxoid (Basingstoke, UK) or E&O Laboratories Ltd. (Bonnybridge, UK). The bacteria were incubated under appropriate atmospheric conditions at 37°C before analysis. Atmospheric conditions included aerobic (plates stored in a hot room), anaerobic (plates stored in incubator), microaerophilic (plates stored in a jar containing microaerophilic conditions inside hot room), and aerobic conditions containing 5% CO₂ (plates stored in a humidified incubator). Microaerophilic conditions (5-10% oxygen, 8-10% CO₂) were generated using a Whitley Jar Gassing System (Don Whitley Scientific Ltd., Shipley, UK).

2.1.2. Rapid evaporative ionisation mass spectrometry

The day before REIMS analysis was scheduled to take place, the instrument intended to be used was vented to perform a standard maintenance procedure aimed at cleaning possible sources of contamination inside the atmospheric pressure interface. In case of Thermo Fisher Scientific instruments (Orbitrap Discovery, LTQ XL and Exactive instrument) the atmospheric pressure interface housing was removed and inlet capillary, tube lens and skimmer were subjected to cleaning by immersion in a series of organic solvents starting with methanol/water (1:1 v/v), acetonitrile, acetone, dichloromethane and finally methanol in an ultrasonic bath for 5 minutes. In case of Waters instruments (Synapt G2, Xevo G2-S, Xevo G2-XS), the custom-made atmospheric pressure interfaces were removed and the collision

surface was wiped off using a cotton swab wetted with methanol. The inlet capillary was flushed using methanol and acetone. In addition, the Stepwave ion guide was removed and cleaned by immersion in methanol/water (3:1 v/v) for 5 minutes in an ultrasonic bath. Subsequently, the instruments were reassembled and pumped down. Orbitrap-based instruments were additionally baked out for 12hrs overnight. Each instrument was calibrated according to the manufacturer's instructions before use utilising a home-built nanoelectrospray ionisation source.

For REIMS analysis, two handheld electrodes in form of a forceps (termed bipolar forceps) were used as the sampling probe (Erbe Elektromedizin GmbH, Tübingen, Germany). A Valleylab Force EZc electrosurgical unit (Covidien, Dublin, Ireland) was used at 60W power setting in bipolar mode as RF alternating current power supply (470kHz, sinusoid alternating current). An approximately 1.5m long 1/8in. outer diameter, 1/16in. inner diameter PTFE tubing (Fluidflon PTFE tubing; LIQUID-scan GmbH Co. KG, Überlingen, Germany) was applied to connect the embedded fluid line of the bipolar forceps and the inlet capillary of the mass spectrometer. In each case the inherent vacuum system of the mass spectrometer was used for aspiration of the aerosol. This setup is shown in Figure 8 while instrumental settings are given in Table 1.

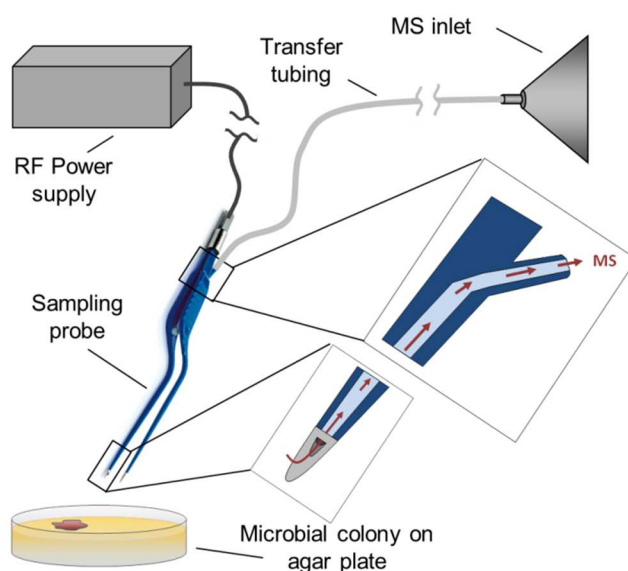


Figure 8. Schematic representation of experimental REIMS setup using bipolar tool as applied for all instrumental platforms used in this study. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Mass spectrometric analysis of the microorganisms was performed directly from the solid culture medium (Figure 8). 0.1-1.5mg of microbial biomass was scraped off the agar surface

using one of the electrodes of the bipolar forceps. The two electrodes were subsequently brought into close proximity (i.e. by pinching the biomass between the tips of the forceps as indicated in Figure 9) and the RF power supply was triggered using a foot switch. The microbial biomass is rapidly heated up due to its non-zero impedance and an aerosol containing the analytes is produced and transferred directly into the mass spectrometer. Where possible, five individual measurements were performed for each strain and averaged as a database entry.

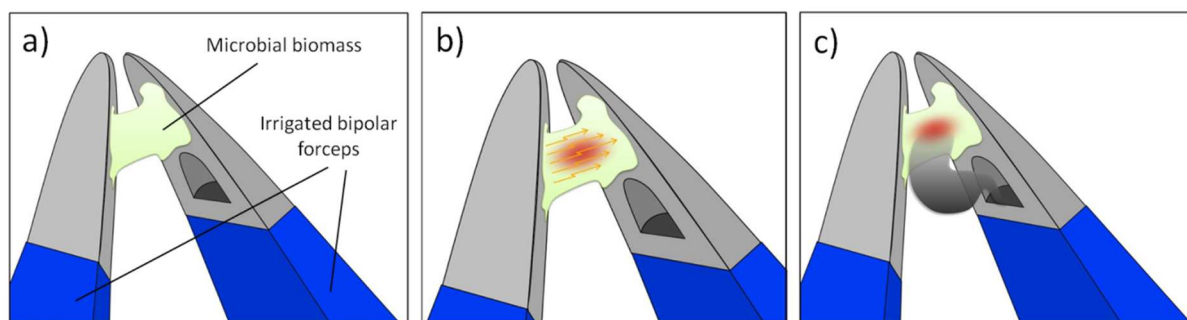


Figure 9. Schematic representation depicting processes during bipolar REIMS ionisation. a) Microbial biomass is held between the two electrodes of the forceps, b) RF alternating current is applied, rapidly heating up the microbial biomass, c) aerosol is produced and aspirated into embedded fluid line. Adapted from Ref. (133) with permission from The Royal Society of Chemistry.

To avoid any infection caused by aerosolised pathogenic bacteria, the analysis site was enclosed into a Class II safety level glove box compartment equipped with UV light source and HEPA filters.

The device consists of three separate parts that are subject to potential blocking: a) the aspiration port inside the forceps, b) the polymer transfer line, c) the inlet capillary of the mass spectrometer. In practice, the only part that is observed to block regularly is the aspiration port within the forceps. Preventional cleaning is applied after on average every 50th individual measurement to avoid blocking. This was done by disconnecting the forceps from the transfer line and connecting it to a 20mL plastic syringe filled with 70% *v/v* ethanol to wash out eventual deposits. The aspiration line is then dried using nitrogen. The process of maintenance cleaning takes about 1-2 minutes.

Table 1. Instrumental parameters of Orbitrap Discovery and Xevo G2-S instruments used in this study.

Thermo Orbitrap Discovery		Exactive	Waters Xevo G2-S	
Parameter	Setting	Setting	Parameter	Setting
Injection time	1000ms	1000ms	Scan time	1000ms
Microscans	1	1	Scan Mode	Sensitivity
Mass analyser	FTMS ^a	FTMS ^b	Mass analyser	TOF
Ion mode	negative	Negative	Ion mode	negative
Mass range	150-2000	150-2000	Mass range	150-2000
Tube Lens Voltage	-120V	-160V	Sampling Cone	30V
Capillary Voltage	-40V	-50V	Source Offset	80V
Skimmer Voltage	n/a	-24V	Source Temperature	150°C
Capillary Temperature	250°C	250°C		
Automatic Gain Control	Off	On (5x10 ⁶)		

a: Orbitrap Discovery instrument is working at a resolution of 30,000 at m/z 400, b: Mass analyser was used at a resolution of 50,000 (m/z 200)

2.1.3. Determination of the amount of microbial wet weight needed for REIMS analysis

For determination of the necessary amount of bacterial biomass, bacteria were picked up as for mass spectrometric analysis and scraped off the sampling probe electrodes using a sterile cotton swab tip. The cotton swabs were transferred into a microcentrifuge tube and the tube weighed in triplicate using a Sartorius AG 180D (Göttingen, Germany) analytical balance. Microcentrifuge tube and cotton swab tip were additionally weighed in triplicate before taking up the bacterial biomass. Each five replicates were analysed per strain. Mean values for microcentrifuge tubes with and without bacterial biomass were calculated and subtracted from each other to obtain the net amount of bacterial wet weight that was picked up using the bipolar sampling probe.

2.1.4. Data analysis

2.1.4.1. Determination of the number of peaks

The amount of m/z signals observed using REIMS was assessed by randomly selecting a database entry for 28 of the most common bacterial pathogens (files sourced from dataset used for specificity assessment, Table 18). Each one file per species was peak picked in MATLAB environment (MathWorks, Natick, MA, USA) using the ‘mspeaks’ function. This function is part of the MATLAB suite and first applies smoothing to the mass spectral data followed by identification of putative peak location and intensity values and subsequent

application of post-filtering criteria. Height filter was chosen as a post-filtering criterion with a threshold of 200. Mass spectral features above this threshold were found to have a signal to noise ratio >3.

2.1.4.2. Specificity assessment using multivariate statistical tools

Raw mass spectrometric files were transcoded to mzML format by the ProteoWizard msconvert tool (version 3.0.4043)(143) and imzML format using imzML Converter (version 1.0.5)(144) and imported into MATLAB for data pre-processing, pattern recognition analysis and visualisation. A typical total ion chromatogram as obtained during data acquisition for a database entry is shown in Figure 10. As REIMS is a pulsed ion source, individual measurements are represented by a sharp rise and subsequent fall in total ion current (TIC). The database file acquisition process takes on average 0.8-1.2 minutes with an individual measurement lasting for 3-6 seconds (see spikes in Figure 10). Constructed bacterial database contains mass spectrometric raw files grouped by bacterial species. For specificity assessment, an appropriate sub-set of files was chosen and converted to mzML format, followed by conversion to imzML format. All spectra acquired for an individual sample were extracted from the imzML format and subsequently summed up to lead to one sum spectrum per isolate. Spectra were chosen based on an average total ion current threshold of 120% as indicated by the blue line in Figure 10.

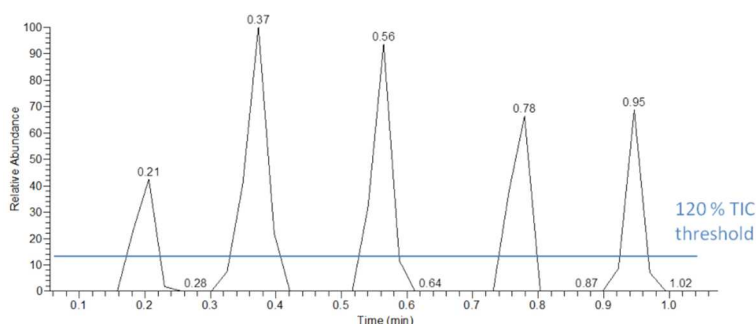


Figure 10. Schematic representation for selection of spectra for downstream data analysis using a total ion count (TIC) threshold. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

All REIMS spectra were subsequently linearly interpolated to a common sampling interval of 0.01Da. Since a binning strategy was used in this study, recursive segment-wise peak alignment was used to remove small mass shifts in peak positions across spectral profiles.(145) The aligned data were subjected to data normalisation (median-fold change or TIC normalisation as stated in text) to account for sample to sample variation in overall

signal intensity unrelated to molecular patterns.(146) Log-based transformation was subsequently applied to stabilise variance as a function of increased signal intensity (variance stabilisation normalisation) to ensure that the noise structure was consistent with the downstream application of multivariate statistical techniques.(146) A schematic representation of the applied data analysis strategy is shown in Figure 11 and was used unless stated otherwise.

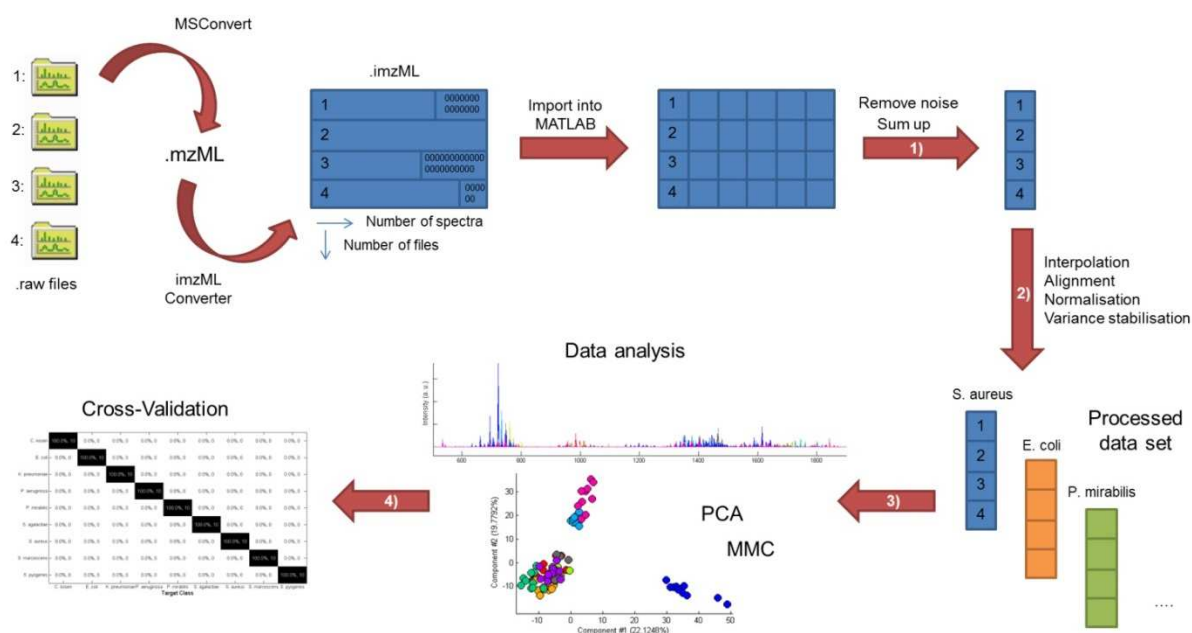


Figure 11. Data processing workflow for specificity assessment on different phylogenetic levels.

Reproduced from Ref. (133) with permission from The Royal Society of Chemistry.

Principal component analysis (PCA) was initially applied to map high-dimensional REIMS data into an uncorrelated set of components capturing the majority of variation in the dataset. Graphical representations of the first few generally most informative components were used to explore the overall similarities/differences in molecular ion composition between bacterial species. Unless stated otherwise, a supervised recursive maximum margin criterion algorithm(122,141) (RMMC) was subsequently applied to derive components with enhanced capacity for discriminating between bacterial types by taking the microbiological assignment of specimens into account. Alternatively, linear discriminant analysis (LDA) on the components derived from PCA was performed for supervised analyses. The final reduced set of discriminating components was equal to the number of bacterial species (classes) minus one. The discriminating models were validated using appropriate cross-validations (CV) as stated in text.

Hierarchical Cluster Analysis (HCA) shown in Figure 70 was performed using Euclidean pairwise distance calculation with a complete linkage metric. 3x3 strains of the original median-fold change normalised dataset shown in Figure 69A and B were averaged for each bacterial species to form the dataset which was then subjected to HCA. This step was undertaken in order to facilitate visualisation while still incorporating a maximum of the biological variance among strains of a certain species.

2.1.4.3. Inter-platform data processing

Thermo .raw files and Waters .raw folders were converted to .mzXML file format using MSConvert (ProteoWizard 3.0.4601). Intensity values of Exactive, Orbitrap and Xevo mass spectra were interpolated to a common mass range of m/z 200-2000 with 0.01Da mass resolution. Mass spectra originating from the same replicates were averaged and subjected to peak picking. Peak lists of averaged mass spectra were normalised to their unit vector and binned to a common mass range of m/z 200-2000 with 1Da mass resolution. Waters Xevo G2-S data was denoised by threshold detection according to Donoho and Johnstone.(147) Binned peak lists were then subjected to multivariate statistical analysis.

2.1.5. Bacterial metabolite identification

Bacterial metabolites were primarily identified based on exact mass measurements and literature references on compounds with the same exact mass that were found in the same or closely related bacterial species. Mass deviations were calculated using the following formula

$$\Delta m[ppm] = \left| \frac{m_{exp} - m_{th}}{m_{th}} 10^6 \right|$$

with Δm = mass deviation (in ppm)

m_{exp} = experimental exact mass, 4 decimal places accuracy

m_{th} = theoretical exact mass, 4 decimal places accuracy

Mass accuracies of <3ppm were regarded as supporting the proposed the sum formula. Further structural identifications were only made by supporting literature references and confirmed by additional tandem mass spectrometry measurements if signal intensity was found sufficiently high and reference spectra were available. Fragmentation experiments were performed on either on a Thermo LTQ XL or Xevo G2-S/XS instrument and using collision induced dissociation as fragmentation mechanism.

2.2. Analysis of cell cultures using REIMS

2.2.1. Culturing of cell lines

All cells were supplied by the Research group of Dr Gergely Szakacs (Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary) as part of a collaboration with the research group of Prof Zoltan Takats at Imperial College London. The NCI60 cell panel was obtained from the DCTD Tumor Repository (National Cancer Institute, Frederick, MD, USA); HEK, HeLa and MES-SA cell lines were from the ATCC collection (Manassas, VA, USA). All cells were cultured in RPMI 1640 medium with the exception of HEK and HeLa cells in the Mycoplasma study which were cultured in Gibco DMEM medium (Invitrogen, Carlsbad, CA, USA). In all cases, media were supplemented with 10% (*v/v*) fetal bovine serum, 2mM glutamine, 100 units/mL penicillin, and 100mg/mL streptomycin (Invitrogen). Cells were incubated in 75cm² tissue culture flasks under conditions of humidified 37°C, 5% carbon dioxide atmosphere. Cell lines were regularly screened for Mycoplasma contamination using the MycoAlert™ Mycoplasma Detection Kit (Lonza Group Ltd, Basel Switzerland).

At 80-90% confluence, cells were rinsed with Phosphate Buffered Saline (PBS, pH: 7.2) and detached using 0.1% trypsin/EDTA for 10 minutes. The trypsin was subsequently neutralised with excess culture medium. The cell suspensions were centrifuged at 250×g for five minutes. After centrifugation the cells were re-suspended and washed two times in 10mL PBS. A third wash was performed in a microcentrifuge tube with only 1mL PBS. The cell pellets were frozen and stored at -80°C until further analysis.

2.2.2. REIMS analysis of cell lines

The REIMS experimental setup is identical to the setup used for the analysis of microbial cultures with the exception that analysis is performed from a microcentrifuge tube rather than a Petri dish (see Figure 12). Mass spectrometric analysis of the cell line biomass was performed directly on the thawed cell pellet without further sample pre-processing steps and cell biomass was picked up using one of the electrodes of the bipolar forceps. The two electrodes were subsequently brought into close proximity (i.e. by pinching the biomass between the tips of the forceps) and the RF power supply was triggered using a foot switch.

The produced aerosol was directly transferred into the mass spectrometer. Multiple technical replicates were recorded per cell line.

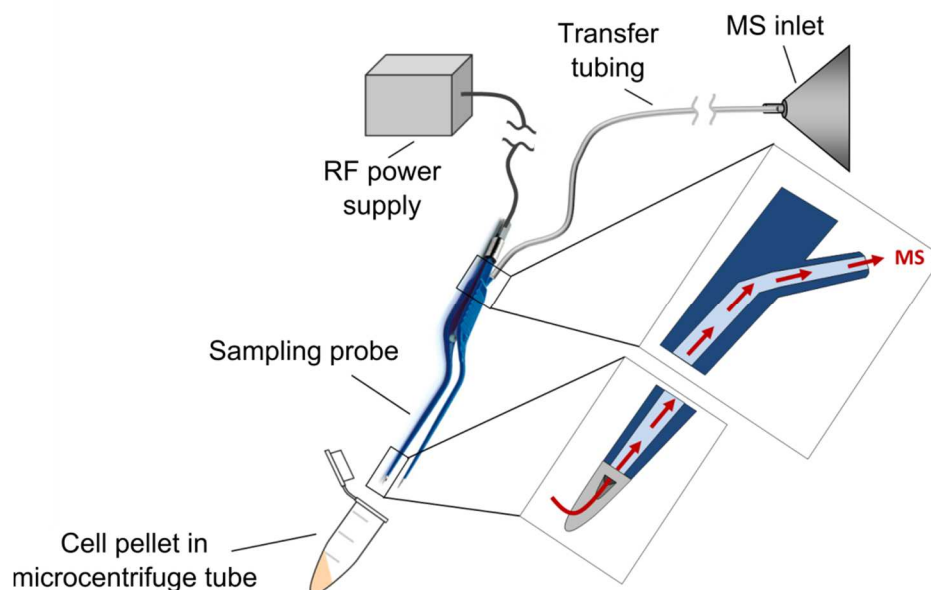


Figure 12. Setup for REIMS analysis of cell line pellets. All analyses were performed on a Thermo Exactive instrument. Adapted with permission from Ref. (257). Copyright (2016) American Chemical Society.

2.2.3. Data Analysis of REIMS cell line data

2.2.3.1. Unsupervised multivariate statistics and cross-validations

Raw mass spectrometric files were converted into mzML format using the MSConvert tool (part of the ProteoWizard suite, version 3.0.4043) and subsequently imported as imzML format into MATLAB (MathWorks) for data pre-processing.(142-144) All REIMS spectra were linearly interpolated to a common sampling interval of 0.01Da. Recursive segment-wise peak alignment was then used to remove small mass shifts in peak positions across spectral profiles.(148) The aligned data were subjected to total ion count (TIC) data normalisation and log-based transformation. Pattern recognition analysis and visualisation were performed either in MATLAB (MathWorks) or in RStudio (Boston, MA, USA, see also www.r-project.com).(149) The mass range of m/z 150-1000 was used for data analysis in all studies. For self-identity experiments, the data set was filtered to keep a reduced set of m/z values: a m/z value was kept, if the difference between the available samples were significantly different at $\alpha=0.01$ threshold level based on the Kruskal-Wallis test. For each cross-validation run, a principal component analysis (PCA) transformation of the training data set

with pre-determined number of principal components (PCs) was calculated in R and prediction score was calculated for each test sample using the 3 nearest neighbor (3-NN) method. The training data in the 'spectral reproducibility set' was selected as follows: for each measurement day, a cell line with defined passage (p) and flask number (A/B) was kept as part of the training data (i. e. HeLa p4 A) if samples were available from at least two different biological replicates (i.e. A1-3). Such sets from each of the three cell lines were combined randomly to produce balanced training data where each cell line is represented by similar number of samples. All of the remaining samples constituted the test set. In the 'NCI60 data set', where applicable, one of the biological replicates was omitted while the remaining data constituted the training set.

2.2.3.2. Correlation with gene expression data

Gene expression data for the NCI60 cell line panel was sourced from the CellMiner website(150,151). For each available gene and filtered m/z value the gene expression and the binned signal intensity across the overlapping 58 cell lines was correlated using Pearson's correlation coefficient with 1000 iterations. The bootstrapped correlation value was defined as the lower 95% confidence interval level of the 1000 iterations, resulting in a total of 26065 (genes) x 17878 (filtered m/z) = 465990070 values.

2.2.3.3. Bootstrapped correlation analysis

For each available gene and sufficiently varied binned m/z value a bootstrapped correlation coefficient was calculated, resulting in a total of 26065 x 5452 = 142106380 correlation values. In the case of each pair of gene expression and m/z value, the gene expression and the binned signal intensity across the overlapping 58 cell lines were correlated using Pearson's correlation coefficient with 1000 iterations. The bootstrapped correlation value was defined as the lower 95% confidence interval level of the 1000 iterations.

2.3. Desorption Electrospray Ionisation Mass Spectrometry

2.3.1. Desorption Electrospray Ionisation Setup

A home-built automated DESI ion source as shown in Figure 13 was used for all DESI experiments in this study. This DESI stage involved a home-built sprayer assembly mounted

on a 3D rotating stage which in turn was mounted to a 3D moving stage allowing movement in *x*-, *y*- and *z*-direction.(119) Samples were mounted on an aluminium sample table using double-sided tape and subjected to DESI analysis. Collection of ions was performed by a home-built extended capillary inlet.

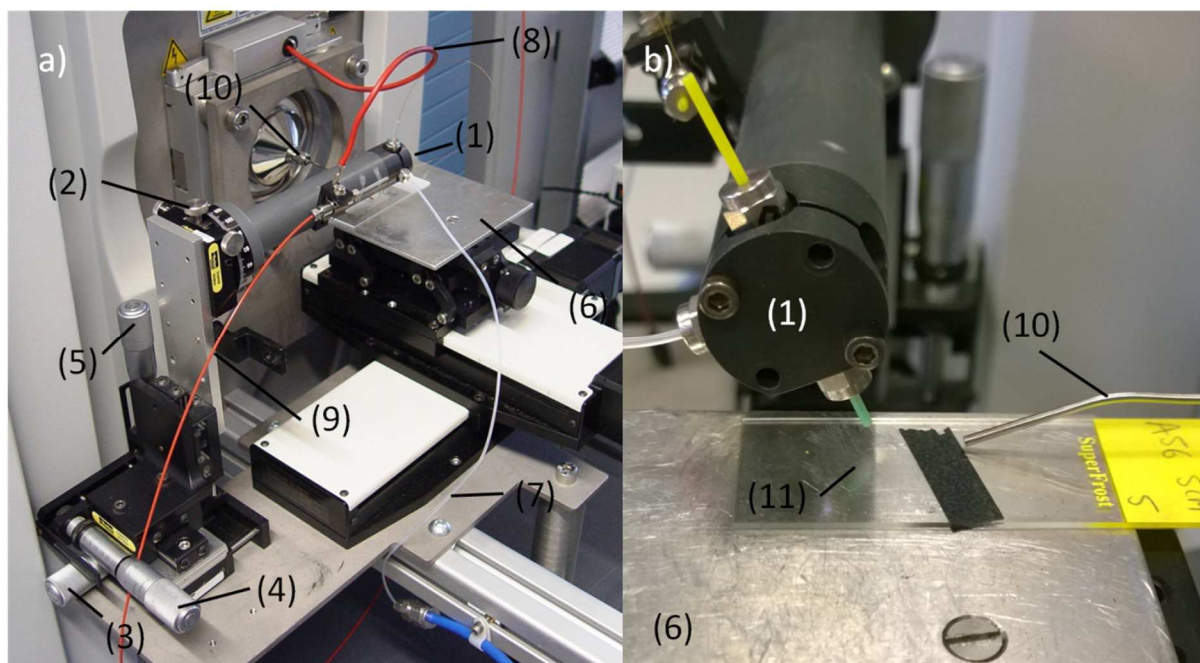


Figure 13. Home-built DESI source mounted on Exactive instrument. Electrosonic sprayer (1) secured onto rotating stage (2) which in turn is mounted on 3D moving stage; (3)-(5) screws for adjustment in *x*-, *y*-, and *z*-direction; (6) sample table, movable in *x*-, *y*-, and *z*-direction; (7) gas supply; (8) high voltage supply; (9) solvent supply; (10) atmospheric inlet of mass spectrometer with extended bent inlet capillary (sniffer); (11) tissue section on glass slide.

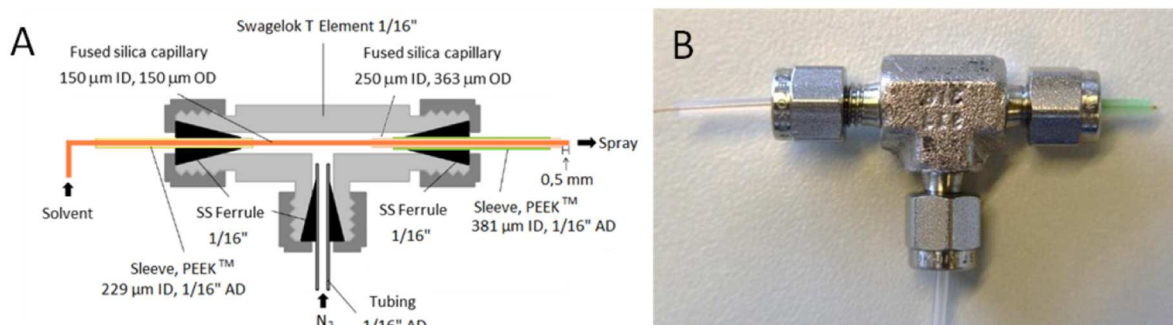


Figure 14. A) Schematic representation and B) photograph of a DESI sprayer assembly as used in this study. Reproduced from Ref. (152) with permission from The Royal Society of Chemistry.

A schematic representation of the home-built DESI sprayer setup used in this study is shown in Figure 14 and employs a stainless steel (SS) Swagelok T element as sprayer body and

fused silica capillaries held in place by sleeves and ferrules to deliver gas and solvent, respectively.

A 50 μm ID, 150 μm OD fused silica capillary was used to deliver the electrospray solvent and it was inserted into an approximately 1.5cm long 250 μm ID, 363 μm OD fused silica capillary delivering the nebulising gas. Since perfect coaxial orientation of the two fused silica capillaries proved difficult to achieve, the two capillaries were oriented with the inner capillary touching the outer capillary at its top edge when the sprayer is directed towards the sample (see Figure 15) at an angle up to 90°. This setup resulted in an elliptical spray point facing the inlet capillary of the mass spectrometer and an increased ion yield. The polyimide coating of both fused silica sprayer capillaries was removed to enable accurate shaping of the capillary tips using a ceramic wafer. This was found to be crucially important for obtaining a stable and symmetrical electrospray plume which in turn was found to be essential for image quality and ion yield. Best image quality was found with a spray head featuring the solvent capillary extruding approximately 0.5mm from the gas capillary.(152)

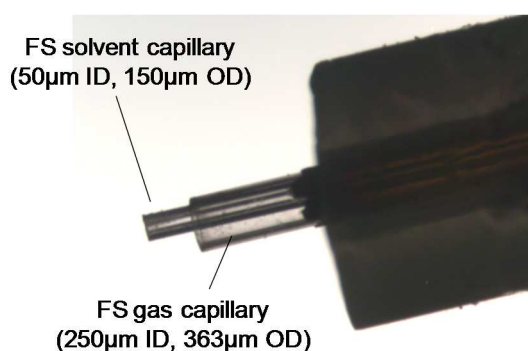


Figure 15. Photograph of the DESI spray head used in this study showing relative position of gas and solvent capillaries; 5x magnification. Amended from Ref. (152) with permission from The Royal Society of Chemistry.

2.3.2. Desorption Electrospray Ionisation Mass Spectrometry Imaging

2.3.2.1. Sample collection

Tissue specimens and related clinico-pathological data were collected between November 2011 and September 2012 with informed and written consent from patients undergoing planned surgical resection for confirmed colorctal cancer (CRC) at a single cancer referral centre (St Mary's Hospital (London, UK). Fresh tissue samples were harvested from centre of tumour, junction of tumour and healthy tissues, 5cm distance to tumour and 10cm distance to

tumour. Patients undergoing left sided or anterior resections were prepared using laxatives Senna and Citramag to clean the bowel before the intervention, while right-sided resection were performed without further pre-treatment. Tissue harvesting was performed in the pathology department by a single senior gastrointestinal pathologist (Prof Robert D. Goldin) and acquired samples were immediately transferred to a freezer at -80°C prior to processing. This study was granted full ethical approval by the institutional review board at Imperial College Healthcare NHS Trust (REC reference number 07/H0712/112).

2.3.2.2. Sample preparation

Tissue samples were cryosectioned to 15µm thickness with either a Bright 5030 cryotome (Bright Instruments, Cambridgeshire, UK) or a Thermo Fisher Scientific HM550 cryotome. If samples were embedded to facilitate cryosectioning, 3% carboxy-methyl cellulose was used as embedding medium. However, the majority of samples were sectioned without any embedding medium. For this process, bulk tissue samples were fixed to a sample holder using a droplet of distilled water. Samples were cryosectioned at thicknesses ranging from 12-15µm and subsequently thaw-mounted onto either plain glass slides or Superfrost glass slides and stored at -80°C until DESI-MS imaging analysis.

2.3.2.3. DESI-MS imaging

Tissue sections were subjected to DESI-MS imaging analysis using an Exactive mass spectrometer (Thermo Fisher Scientific Inc., Bremen, Germany). Exactive instrument parameters are listed in Table 2. A photo of the home-build automated DESI source is shown in Figure 13.

Methanol/water (95:5 v/v) was used as the electrospray solvent at a flow-rate of 1.5µL/min. Nitrogen N4.8 was used as nebulising gas at a pressure of 7bars. All solvents used were of LC-MS grade (Chromasolv, Sigma Aldrich, St Louis, MO, USA). The height distance between the DESI sprayer and the sample surface was set to 2mm with the distance between the sprayer and inlet capillary set to 14mm. The distance between the sample surface and the inlet capillary of the mass spectrometer was <<1mm. The angle between the sprayer tip and the sample surface was set at 80°. The collection angle between inlet capillary and sample was set to 10°.

Table 2. Thermo Exactive instrumental parameters used for DESI-MS imaging experiments.

Parameter	Setting.
Polarity	negative
Resolution	100,000
Mass range	200-1050
Spray voltage	- 4.5kV
Capillary temperature	250°C
Capillary voltage	- 50V
Tube lens voltage	- 150V
Skimmer Voltage	- 24V
Max. injection time	1000ms
Microscans	1
AGC target	5e6

The general principle underlying imaging processes using DESI-MS is displayed in Figure 16. Rather than point-by-point sampling, horizontal line scans are performed over the tissue surface by moving the automated sampling platform at a speed that covers the area determined as a pixel (spatial resolution) in the time the mass spectrometer requires to complete one scan (acquire one mass spectrum). This results in one file per row of the resulting image (number of rows determined by sample height divided by spatial resolution).

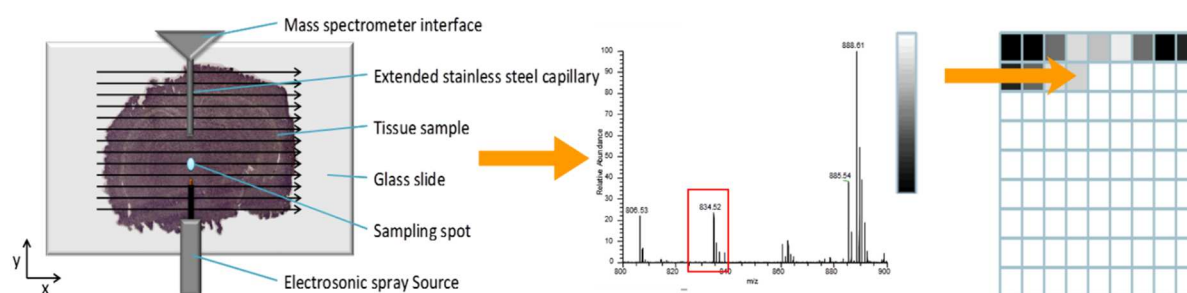


Figure 16. Schematic representation displaying processes underlying DESI mass spectrometry imaging. Tissue sections are continuously sampled using horizontal line scans and followed by single ion image generation.

2.3.2.4. Imaging data pre-processing

For image analysis, individual horizontal line scans were converted into centroided .imzML files using the imzML Converter Version 1.1.4.5 (www.maldi-msi.org) and subsequently imported into MATLAB (MathWorks) for pre-processing and multivariate analysis.(144) Molecular ion peaks within a m/z range smaller than the native accuracy of the mass spectrometer (<4ppm in this case) were assigned to the same molecular ion species uniformly for all pixels on a tissue section. The resulting sample-specific lists of m/z values were aligned between samples by means of a dynamic programming based peak-matching algorithm.(153) Median-fold change pixel-wise normalisation was used to remove overall

intensity variation between tissue spectra caused by technical factors. A log-based variance stabilising transformation was subsequently applied to account for intensity variation between small and large peaks.(154)

2.3.2.5. Histological assignment and spectral pre-processing

The same tissue sections that were used for DESI-MSI analysis were subsequently stained using hematoxylin and eosin (H&E) and a cover-slide was applied. Following staining, tissue sections were digitally scanned at high resolution using a NanoZoomer 2.0-HT digital slide scanner (Hamamatsu Corp., Japan).

Histological and MSI features were optically aligned in a user-friendly MATLAB imaging toolbox using a rigid image-alignment algorithm.(154) Histological assignments were made by a single histopathologist (Abigail V. M. Speller) by choosing morphologically uniform regions-of-interest. Corresponding spectra were subsequently extracted and stored in a database for further down-stream analyses.

2.3.3. Analysis of bacteria using DESI

For analysis of microorganisms, bacterial biomass was scraped off the agar surface using a sterile cotton-swab and smeared onto the surface of a standard microscope glass slide. Subsequent spectral acquisition was performed in negative ion mode on a Thermo Exactive instrument using the mass range m/z 150-2000 and other instrumental parameters as listed for DESI imaging experiments (see Table 2). Same DESI parameters were used as for imaging experiments with the exceptions of a solvent flow rate of 2.5 μ L/min, a shorter distance between sprayer and inlet capillary of 7mm and point-to-point rather than continuous movement.

2.4. Determination of taxon-specific markers

The general workflow applied for the identification of taxon-specific markers is displayed in Figure 17. All data was compiled and a representative subset of the overall >3,000 strains recorded on the Exactive instrument (see entire database entries in Appendix 2) was generated. This was followed by a peak-picking and peak-matching routine and an ANOVA-based biomarker discovery to derive the taxon-specific markers. The overall workflow for detection of bacteria in DESI imaging sample sets is displayed in Figure 18. Once taxon-

specific bacterial markers are derived, bacteria are subsequently detected in imaging datasets and displayed with the corresponding histological boundaries.

2.4.1. Sample set

A dataset was created comprising 228 different bacterial species belonging to 80 genera, 47 families, 23 orders, 11 classes and 5 phyla (596 strains in total). A maximum of five to seven individual strains was used for calculations for badly represented genera, while a maximum of 3 strains per species was used for well represented genera (10 species per genus or more). A detailed sample set can be found in Appendix 4. Different culturing media, culturing ages and culturing atmospheres were included into the sample set where possible.

2.4.2. Data analysis

To derive taxon-specific markers, the data were processed in centroid rather than profile mode. For this, raw mass spectrometric files were transcoded to mzXML format by the ProteoWizard msconvert tool (version 3.0.4043)(143) and imported into MATLAB environment (MathWorks).

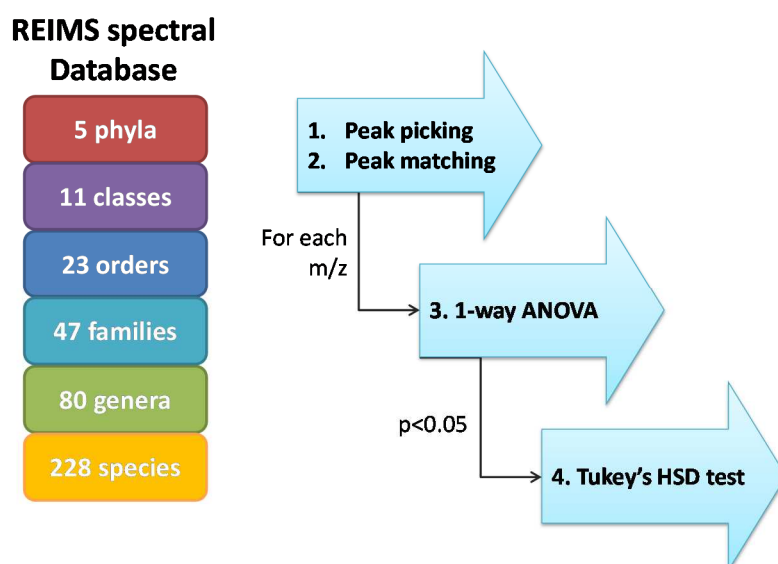


Figure 17. Data analysis workflow used for the determination of taxon-specific markers.

2.4.2.1. Peak-picking and peak-matching

Peak picking was performed on individual spectra within each separate mzXML file, with processing being performed on each set of m/z values and corresponding intensity values.

Local maxima and peak boundaries were determined according to local minima or zero intensity values to either side of the maximum. Peak intensity is calculated as its sum between the two boundaries, with boundary intensities being halved (important only for unresolved peaks; resolved peak boundaries are 0). The peak's 'centre of mass' is determined by means of an intensity-weighted average. After processing all spectra of a single mzXML file, the individual m/z and intensity vectors are aligned using a dynamic programming algorithm. The user specifies maximum peak shift (in ppm) that acts as the maximum distance over which peaks can be matched. If multiple files are to be compared, the peak-matching algorithm is re-run using the same parameters.

2.4.2.2. Biomarker discovery

Each variable is analysed independently, initially by 1-way analysis of variance (ANOVA), to determine if there is a significant difference between the means of the defined groups (here, bacterial species). Significant differences (e.g. $p < 0.05$) are further analysed by Tukey's honestly significant difference (HSD) post-hoc method. HSD calculates the confidence intervals for differences between the mean of two groups and if this interval spans across zero then the difference between two means is regarded as insignificant. A difference for each variable between all groups is determined. Thus, a variable that differs significantly between one and all other groups is considered a likely biomarker candidate. Results for each variable are returned as the ANOVA p-value. The emphasis has been placed on variables that are significantly higher in one group than in others, rather than variables that are absent in all but one taxonomical group. This approach enables finding markers that would have otherwise been excluded in case of groups containing unrelated misclassified bacterial species which do not show the same spectral feature.

For each taxonomical level, both a list of the top 20 most significant peaks per taxon and 150 most significant peaks in the dataset were created and manually compiled into markers that are specific on different taxonomical levels. To further aid analysis, box-and-whisker plots were generated showing the 25th, 50th (median) and 75th percentiles of the marker in the taxon of question and other taxons at the same phylogenetic level. Whiskers indicate 1.5 times the interquartile range. In a subsequent manual step, such markers were filtered out which were significantly increased for a certain bacterial species but were not specific enough to be used as univariate markers. As these imaging data sets were recorded using a mass range of m/z 200-1000, determination of taxon-specific markers was restricted to the same mass range.

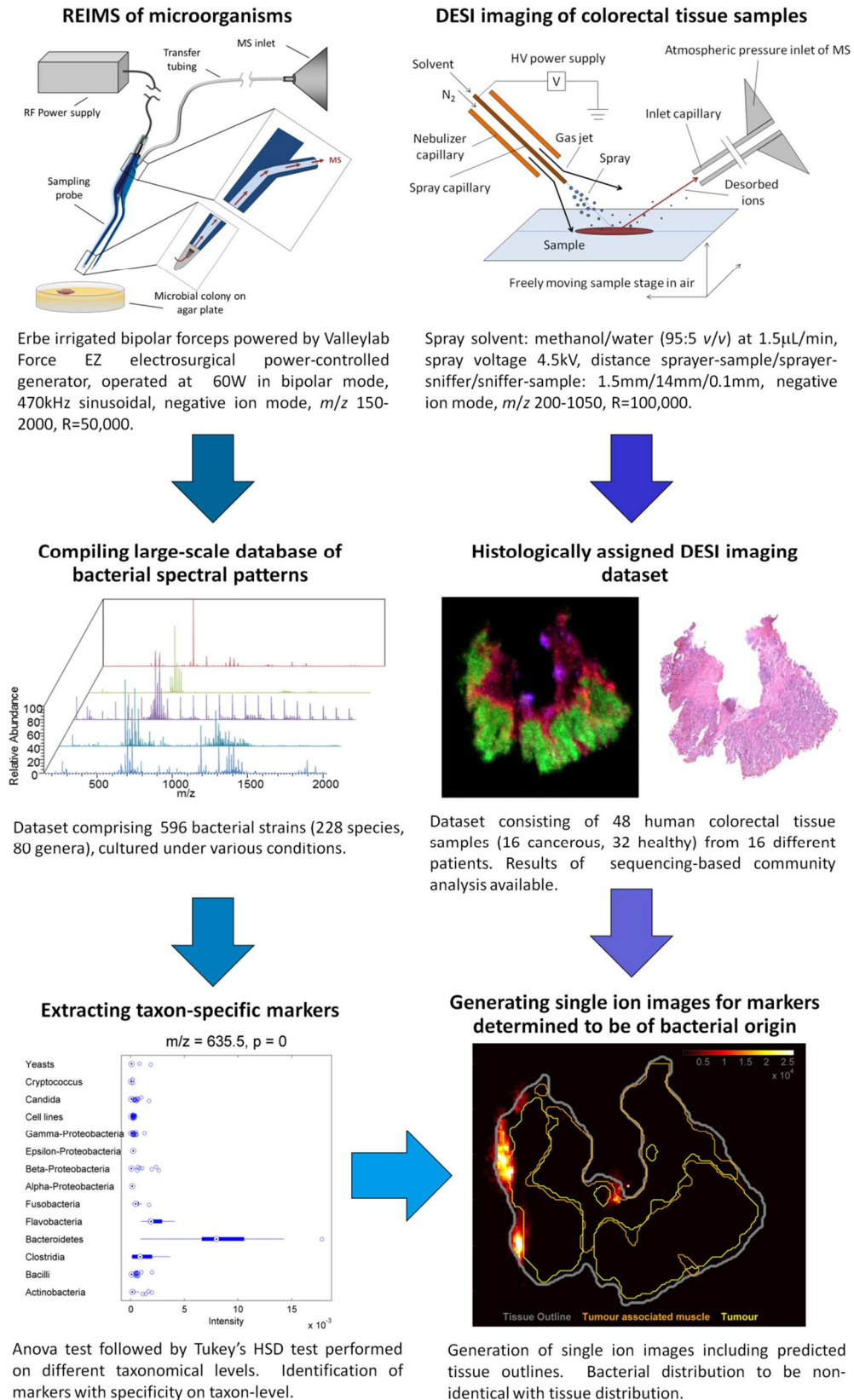


Figure 18. Workflow used for visualising taxon-specific markers in human colorectal tissue sections. A REIMS spectral database is generated and used to derive taxon-specific markers for bacteria. DESI imaging of tissue sections is performed and bacterial markers are visualised in DESI imaging datasets.

2.5. Detection of taxon-specific markers in human colorectal tissue specimens

Each set of three tissue samples (center of tumour, 5cm distant and 10cm distant from tumour) from 16 colorectal cancer patients that underwent 16S rRNA sequencing-based community analysis were used to assess feasibility of detection of taxon-specific markers in human tissue specimen. This was performed by generating single ion images of the taxon-specific markers given in Table 28 (page 199) and filtering said images for such structures that are not present in background and show no correlation of distribution with histological tissue types.

RGB images were generated using MSiReader Version 0.05(155) with linear interpolation (order 1) and 0.01Da bin size. In-house processing of DESI imaging data files was performed as described in Ref. (122) and under Materials and Methods - Desorption Electrospray Ionisation Mass Spectrometry. To produce ion images of the previously derived taxon-specific markers (TSM) containing outlines of histological tissues, first a binary segmentation method(156) was used to delineate the tissue object from the background pixels, which is shown by the thick grey line in Figure 18. In order to define the tissue type boundaries, a one-against-all multivariate analysis was performed on a training set comprising a small subset (20-30 pixels) of pixels of the same tissue section which were previously annotated by a histopathologist as being morphologically homogenous. A probability for each pixel belonging to each class was determined. The tissue type boundaries, as depicted by the coloured lines, delineate those pixels with a probability of greater than 0.90 of belonging to that class. The one-against-all procedure makes it possible for a pixel to have more than one probability greater than 0.90, and pixels that are encompassed by multiple regions can be considered to be at the junction between two (or more) tissue types. Ion images were subsequently generated using a m/z window of 10ppm.

3. Method Development and Characterisation

3.1. Experimental setup – Ion source

3.1.1. Comparison of spectral content of REIMS and DESI datasets

Desorption electrospray ionisation mass spectrometry (DESI-MS) has been reported in literature as a suitable tool to characterise and differentiate different types of bacteria.(90,103,110,157) For this reason, as a part of this study, the proposed REIMS methodology was compared to the more established DESI methodology in case of both bacterial and fungal analysis. Results were evaluated for spectral information content, ease of workflow and general applicability.

Studies involving DESI published in literature were mostly performed by the research group of Prof R. Graham Cooks (Purdue University, IN, USA) and were focused on the analysis of *E. coli*, *Salmonella* spp., *S. aureus*, and *Bacillus* spp.. Bacteria were analysed both in negative and positive ion mode and from m/z 50-1600. Especially in the study presented by Zhang *et al*, good spectral diversity was observed for both Gram-positive and Gram-negative bacteria, however, analysis was not performed directly on the bacterial cells but on lysed cells after addition of 70% ethanol. In studies involving direct analysis of bacteria, significantly better signal intensity was observed for Gram-negative bacteria due to their thinner cell wall as compared to Gram-positive organisms. Comparatively poor phospholipid intensity was observed for *S. aureus* and *B. subtilis* and strong intensity signals could only be observed for extracellular metabolites such as the lipopolypeptide surfactin.(90,102) No studies were published on the analysis of yeasts.

A range of bacterial reference strains for both Gram-positive and Gram-negative bacteria were analysed using both DESI and REIMS methodologies in-house. While analysis directly from the Petri-dish was possible in case of REIMS, in case of DESI very poor spectral quality was observed when the electrospray was directed onto bacteria growing on the agar surface. This was associated with significant deformations of the soft agar surface under the gas jet from the DESI spray caused by the pressure of the nitrogen stream. The deformations inflicted on the agar during analysis cause the sampling surface to lower relative to the height

of the extended MS capillary inlet hindering sampling of the secondary droplets produced during the DESI process as the droplets are unable to leave the deformation crater. As collection angles in DESI should ideally be close to zero degrees,(158) this effectively prevents generation of good quality spectral data. Thus, for DESI analyses of bacteria it was decided to instead deposit bacterial biomass on a solid glass surface using a cotton swab. Analysis was performed without further sample preparation and good spectral quality was obtained for several species such as *Bacteroides fragilis* or *Fusobacterium necrophorum*, *Pseudomonas aeruginosa*, *Bacillus cereus*, *Micrococcus luteus*, *Streptococcus agalactiae*, *Stenotrophomonas maltophilia* or *Clostridium difficile*, among others. This shows that good quality spectral data can be obtained both from Gram-positive and Gram-negative bacteria using both methodologies. However, while REIMS gives good quality data for all types of bacteria and yeasts, only comparably poor spectral quality were observed in case of the DESI analysis of certain Gram-positive species such as *Staphylococcus aureus* or all yeast species (*C. krusei*, *C. albicans*) which is tentatively associated with their cell wall properties. Interestingly, for some of those bacterial species that proved problematic using direct DESI analysis, particularly strong spectral intensity was observed using REIMS, such as *Staphylococcus* species.

A comparison of spectral content reveals similar phospholipid profiles in the mass range m/z 600-900 using both DESI and REIMS. Although relative signal intensities are varying between the two techniques, in general the same phospholipid species are visible. However, while DESI offers more mass spectral signals in the lower mass range of m/z 150-500, REIMS gives significantly more spectral constituents above m/z 1000, with clearly detectable mass spectral signals up to m/z 2000. Comparison of spectral profiles highlighting this difference between DESI and REIMS spectra of *Bacteroides fragilis* and *Pseudomonas aeruginosa* are shown in Figure 19 and Figure 20, respectively. A further complication of spray techniques as compared to REIMS is an increased spectral complexity as introduced by different adducts for the same molecular species as observed in case of ceramide species (m/z 590, 604 and 618) in *B. fragilis* which are detected almost exclusively as $[M+Cl]^-$ adduct in case of REIMS while DESI spectra feature the same ion both as $[M+Cl]^-$ and $[M-H]^-$ (m/z 554, 568 and 582).

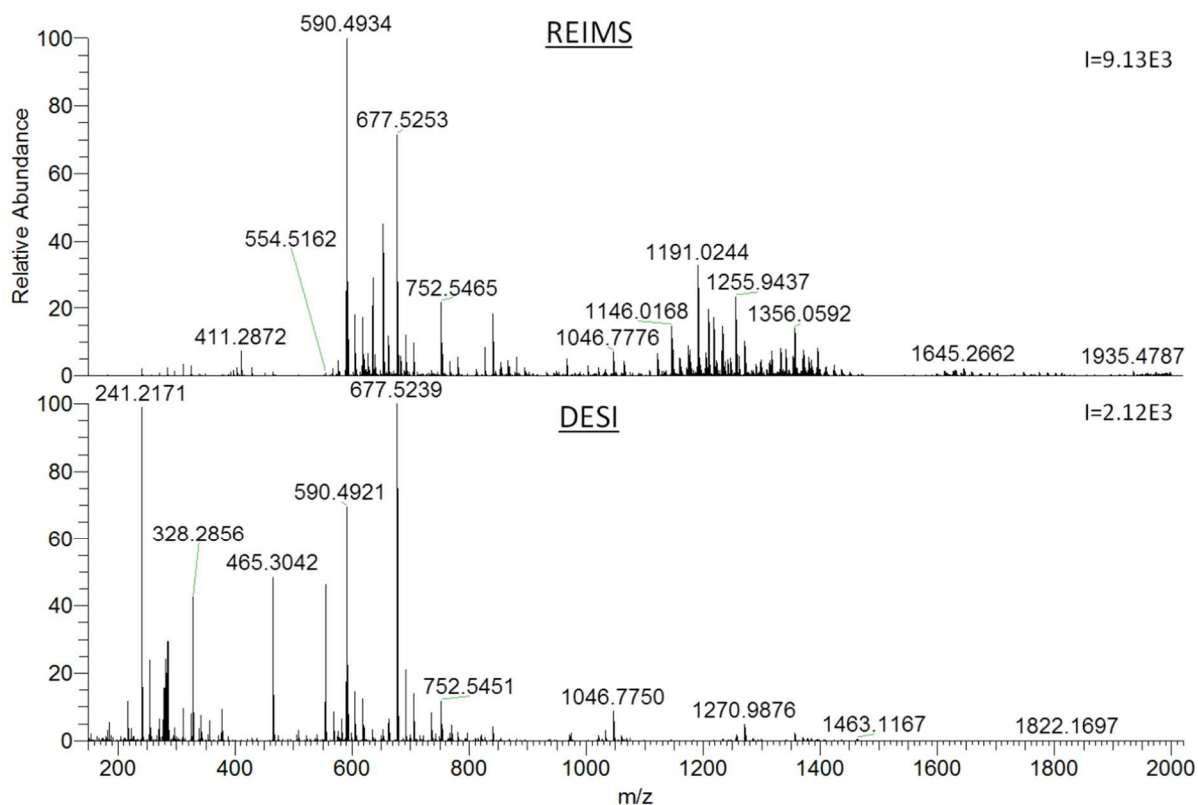


Figure 19. Spectral profiles of same clinical isolate of *Bacteroides fragilis* obtained using REIMS (top) and DESI (bottom) using negative ion mode.

A further advantage of REIMS compared to DESI is the lack of spray solvent peaks which greatly helps data analysis and partially explains the enhanced spectral content below m/z 500 observed by DESI. The use of solvents during the ionisation process also enhances the chance of introducing spectral contaminants. Largely identical spectral features are observed for m/z 150-1000 in both DESI and REIMS, however, DESI largely favours ionisation of smaller more polar molecular constituents as visible in both Figure 19 and Figure 20 (bottom panel each). This in turn can lead to significant signal suppression of other analytes as shown in the example of *Pseudomonas aeruginosa* where small molecular weight quorum-sensing molecules suppress the phospholipid signal.

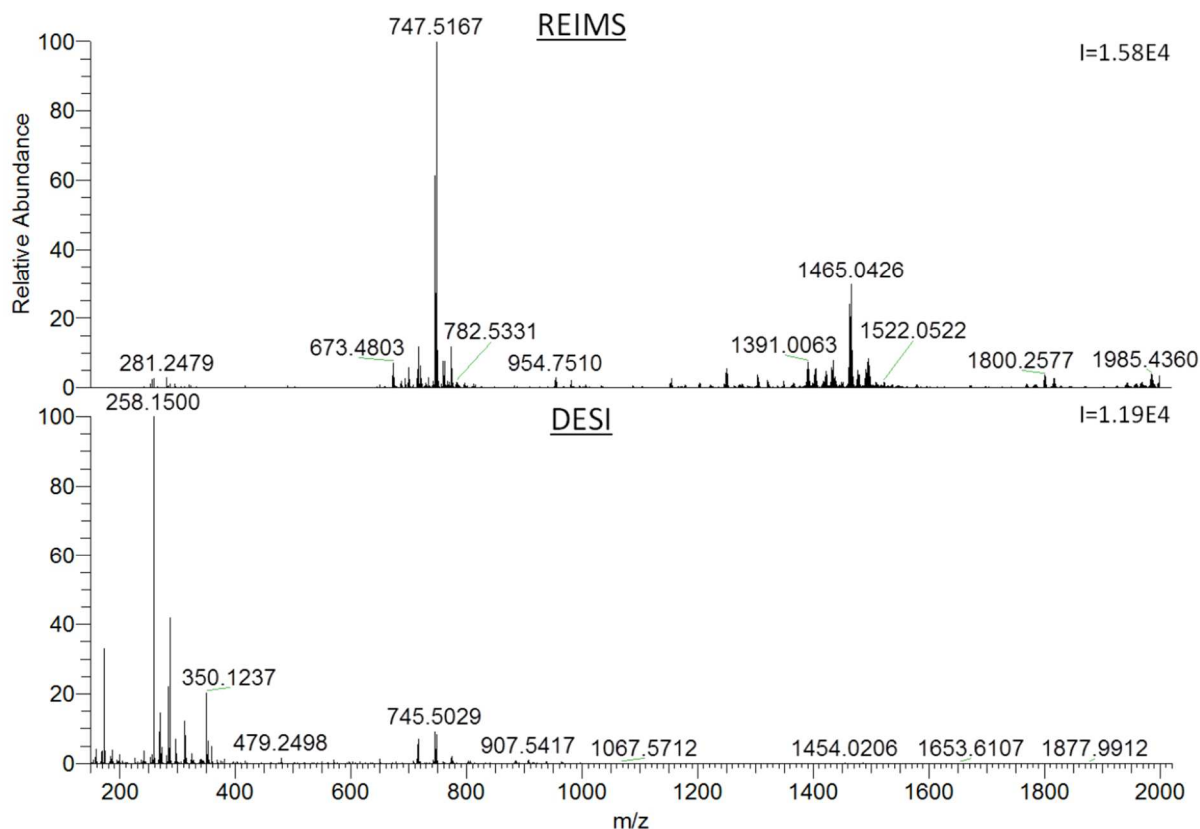


Figure 20. Comparison of spectral profiles of the same clinical isolate of *Pseudomonas aeruginosa* obtained using REIMS (top) and DESI (bottom).

In earlier lipid characterisation studies of bacteria using DESI, the detection of high abundance lyso-phospholipids (both lyso-PEs and lyso-PGs) was reported for *Escherichia coli* K12 which were found to be the spectral base peaks. The reported spectral profile was found by the authors to be identical with spectra obtained using ESI-MS from cell extracts of the same strain (see Figure 21A and B).(103) However, these were found to be significantly lower for in-house performed DESI experiments as shown in Figure 21C, which suggests that the study reported by Zhang *et al.* involved instrument parameters that caused a high degree of fragmentation of intact phospholipid species. Neither lyso-PEs nor lyso-PGs were observed in REIMS spectra of *Escherichia coli* strains (see Figure 21D) suggesting comparably soft ionisation characteristics of the REIMS method.

REIMS was found to give equally good spectral results for a variety of different bacterial and yeast species while being able to analyse bacteria directly from the Petri-dish without any further sample pre-treatment steps. In addition, significantly increased spectral variability and stability was observed for REIMS. Therefore, REIMS was chosen as preferred technique for large-scale database building.

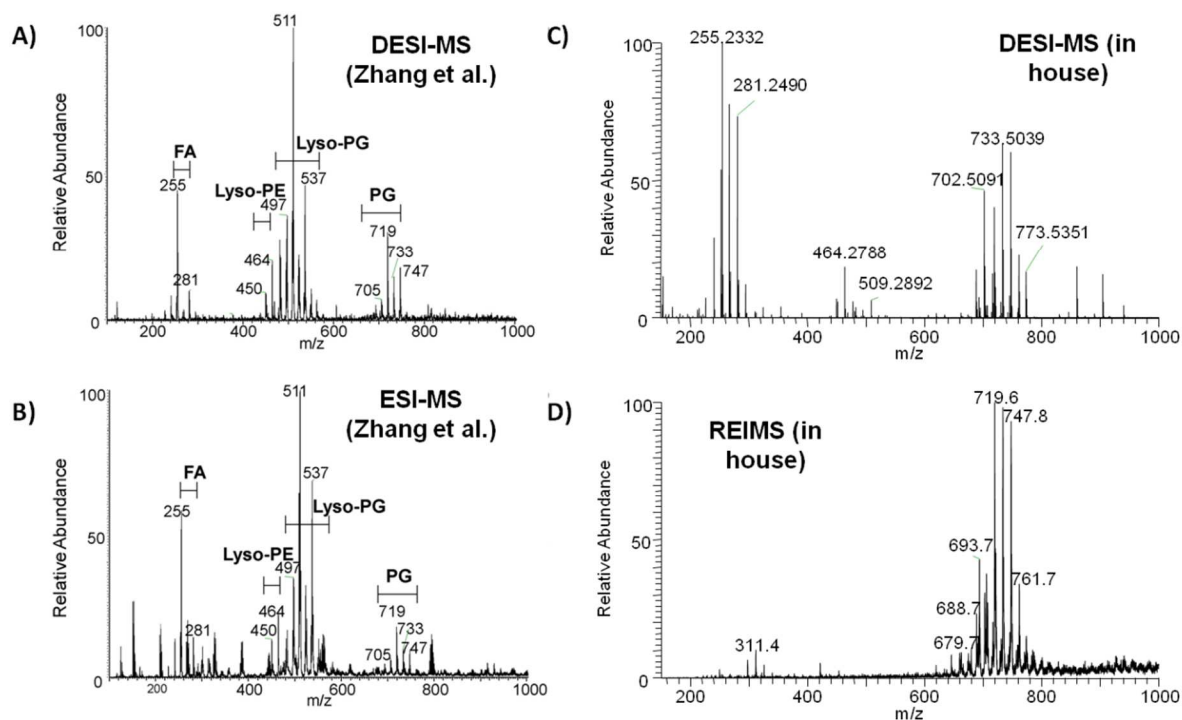


Figure 21. Spectra obtained from *Escherichia coli* K12 derivatives using DESI-MS and REIMS. Data recorded using A) DESI-MS by Zhang *et al.*, Purdue University, Ref. (103), B) ESI-MS by Zhang *et al.*, C) DESI-MS performed at Imperial College London, D) REIMS performed at Imperial College London. A and B are reprinted from Ref. (103), with permission from Elsevier.

3.2. REIMS Sampling Probe

Different experimental setups following the scheme of monopolar and bipolar diathermy were tested for their suitability as ion sources. Both setups (shown in Figure 22) could generate mass spectral fingerprints of intact bacterial cells. For the monopolar setup, bacterial biomass is placed on a large surface area electrode and the electrical circuit is closed by touching the sample with the sharp, handheld counter electrode. Alternatively, the agar can be removed from the Petri-dish and placed on the counter electrode. The monopolar setup produces a very high electric current density in the proximity of the handheld electrode touching point, which results in high effective evaporation temperatures and excessive solid aerosol (soot) formation. In agreement with these observations, the spectral quality is poor in this case and spectra do not feature intensive signals above m/z 1000 (see Figure 24) This observation was tentatively associated with a high degree of charring and effective combustion of the sample which cannot be circumvented by decreasing the power output. In addition, the large amount of smoke produced (and aspirated by the instrument) results in significant contamination of the ion optics. In contrast, the bipolar setup uses a pair of

handheld electrodes in the form of forceps as the sampling probe and offers a more evenly distributed electrical current between the two electrodes (see Figure 23). This setup proved to be superior in comparison to the monopolar setup for the analysis of minor amounts of biomass.

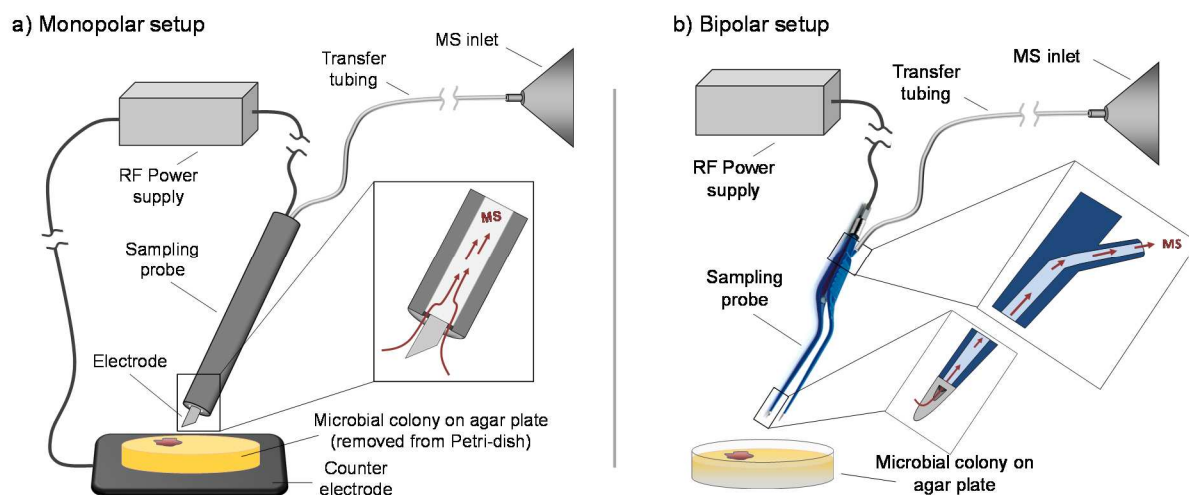


Figure 22. a) Monopolar and b) bipolar setups tested for the analysis of microorganisms using REIMS. Both setups were also tested with external transfer line. Reproduced from Ref. (133) with permission from The Royal Society of Chemistry.

The bipolar setup provided easier handling, almost complete elimination of memory effects and less frequent blocking of transfer devices. In addition, less fragmentation and better sensitivity for higher masses was observed using the bipolar tool. The most effective and reproducible aspiration of the produced aerosols was achieved using an electrode setup with an embedded aspiration line (as displayed in Figure 23B). The comparatively lower amount of aerosol produced by the bipolar tool is due to smaller amounts of biomass needed for analysis (higher sensitivity) and lower current density due to the parallel geometry of the electrodes. Secondary electrospray ionisation was tested using the experimental setup which was described earlier for the post-ionisation of surgical aerosols. In agreement with earlier studies, further increase in sensitivity was obtained only in positive ion mode.(159)

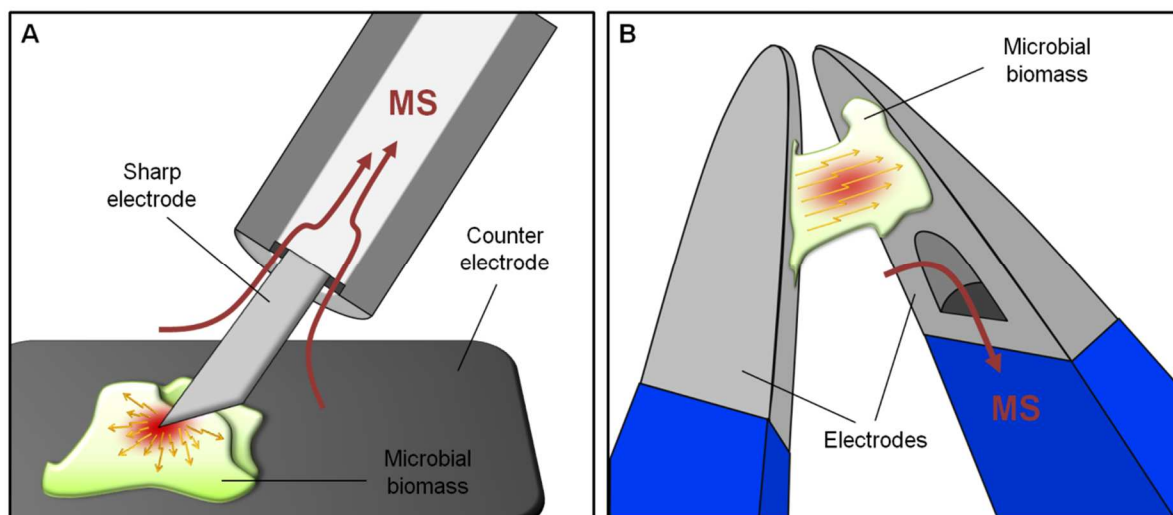
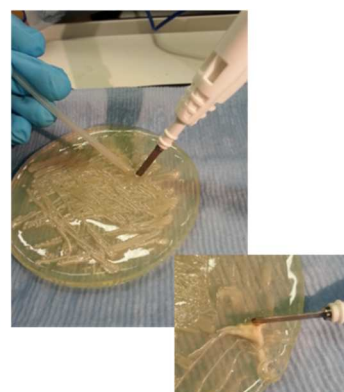
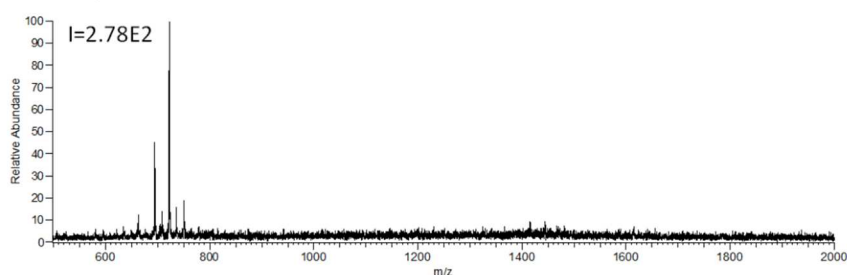


Figure 23. Schematic representation of experimental setup and electrical current density using A) a monopolar and B) a bipolar tool. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

monopolar



bipolar (irrigated)

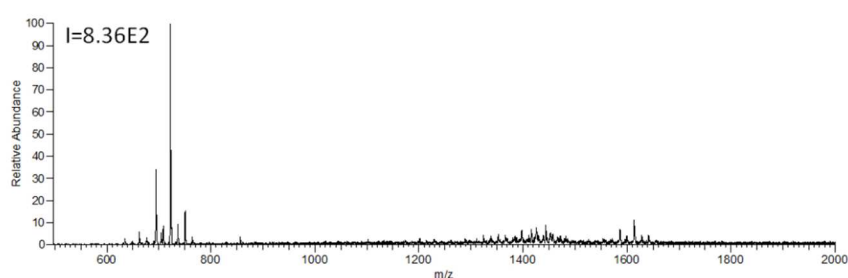


Figure 24. Spectra of *Staphylococcus aureus* obtained using monopolar (top) and bipolar setup (bottom).

The use of bipolar setups provides better handling and sensitivity and increased detection of high molecular weight lipid species. I=Intensity value at 100% relative abundance. Amended with permission from Ref. (142). Copyright 2014 American Chemical Society.

The amount of biomass needed for sensitive REIMS analysis using the bipolar sampling tool with embedded aspiration line was determined using three strains of bacteria, *Staphylococcus warneri*, *Pseudomonas aeruginosa* and *Haemophilus influenzae*. These cover Gram-positive and Gram-negative bacteria, as well as bacteria that can be detached easily

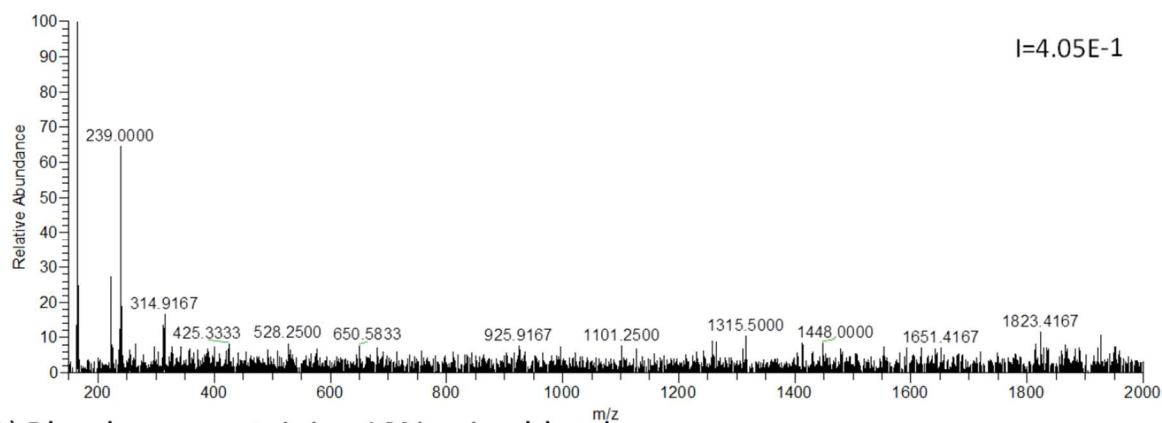
from the agar surface and those which are more challenging to remove from the agar growth plate (*P. aeruginosa*). Gram positive *S. warneri* was found to be rather dense as compared to Gram-negatives *P. aeruginosa* and *H. influenzae*, with especially colonies of the latter looking very moist. Results of all five replicates are shown in Table 3 (for more details see Appendix 1). The amounts needed were found to be comparable in weight for all three bacterial strains analysed despite their different physical properties. For *S. warneri*, between 0.2-0.9mg were needed, while for *P. aeruginosa* and *H. influenzae* 0.6-1.0mg and 0.2-0.6mg were required, respectively.

Table 3. Results of experiment to determine amount of biomass needed for analysis. Each value was measured in triplicate and the mean calculated. Five replicates were performed for each strain.

	<i>Staphylococcus warneri</i>	<i>Pseudomonas aeruginosa</i>	<i>Haemophilus influenzae</i>
Replicate 1	0.5mg	1.0mg	0.3mg
Replicate 2	0.9mg	0.6mg	0.4mg
Replicate 3	0.2mg	0.9mg	0.6mg
Replicate 4	0.7mg	0.7mg	0.3mg
Replicate 5	0.6mg	0.7mg	0.2mg

A major concern in case of direct bacterial REIMS analysis was the potential influence of the growth medium on the spectral content. This problem is obvious in case of bacterial growth media such as blood agar containing a significant amount of lipids itself. Growth medium alone was analysed using the bipolar sampling tool and the analysis has not produced any signals that could be attributed to the growth medium (see Figure 25). This observation was tentatively associated with the nature of the carbohydrate (agarose) matrix, which undergoes condensation reactions via water losses, resulting in extensive cross linking and eventually charring on heating which in turn will hinder the REIMS mechanism. Below, blank spectra of blood agar and Luria-Bertani agar are shown and show only polysiloxane instrument contaminants as indicated by the 78.0183 Da mass differences (measured using the Orbitrap mass analyser). Since the mass spectra otherwise do not contain any growth medium related mass spectral features and the agar surface was not disrupted while picking up the microbial biomass, clustering analysis in this study was generally performed excluding growth medium as a blank or quality control.

A) Luria-Bertani medium



B) Blood agar containing 10% ovine blood

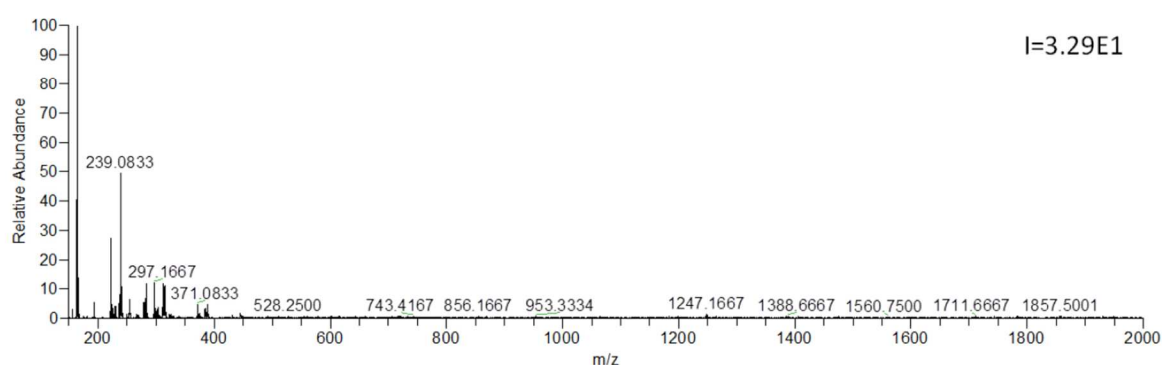


Figure 25. REIMS spectral profiles obtained for A) Luria-Bertani agar (Paul trap mass analyser) and B) blood agar (Orbitrap mass analyser). I=Intensity value at 100% relative abundance.

3.3. Influence of the electrosurgical generator

Throughout the entire study, different commercially available electrosurgical generators were used as radiofrequency power supplies. For most of the data presented, a Valleylab Force EZc generator from Covidien was used. This generator was working at 470kHz frequency (sinusoid) and – similarly to all other power supplies used – does only allow the adjustment of the output power. Based on the information from the manufacturer, the current densities are in the order of 10^4 A/m² at the given power settings. Since the power supply is power controlled, the current does change with the amount of biomass between the electrode tips. This will most likely have an influence on the effective evaporation temperatures which in turn might affect spectral appearance. However, the evaporation event itself takes a few seconds only, which makes the measurement of the temperature challenging, taking into account the quick change of temperature, the boiling of the material and the few mm³ sample volume.

At the practical level, the appearance of the data did not show any significant dependence on RF frequency or evaporating power when the volumes were kept constant. Furthermore, the highly aqueous bacterial biomass cannot be heated freely above the boiling temperature of water, because as soon as there is no aqueous phase between the electrodes, the electric circuit opens. Instead of having higher temperature, the rate of evaporation will increase as the amount of sample is decreased. There is a given electrosurgical generator-dependent threshold power which is necessary to produce spectra, however, the spectral patterns do not show significant dependence on the power as shown in Figure 26.

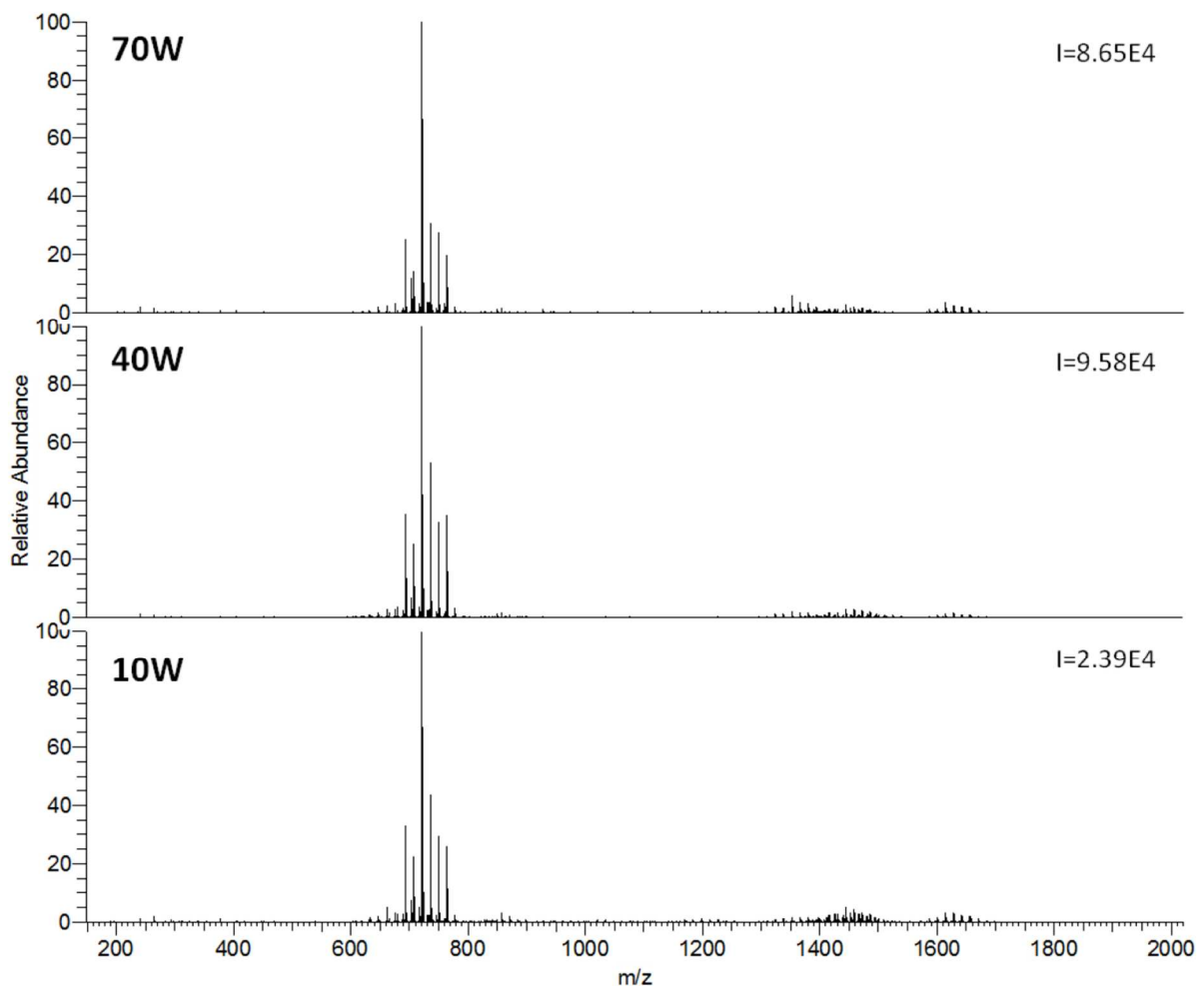


Figure 26. Dependence of spectral profile of *Staphylococcus aureus* ATCC 25923 on RF power setting of electrosurgical generator. I=Intensity value at 100% relative abundance. Data acquired using Thermo Exactive instrument.

Two electrosurgical generators from two different manufacturers were additionally tested for their influence on spectral appearance: a Force Triad generator (Covidien, Dublin, Ireland) with very similar RF frequency to the Force EZc generator and an Erbe ICC300 (Erbe

Elektromedizin, Tübingen, Germany). A comparison of the technical details of both generators in bipolar mode is shown in Table 4.

Table 4. Technical details of tested generators Covidien Force Triad and Erbe ICC300.

Setting	Force Triad	Erbe ICC300
RF frequency	472 kHz	330 kHz
Waveform	sinusoidal	sinusoidal
Peak-to-Peak Voltage	500 V	380 V
Power setting	95	80
Mode	Bipolar, low	Bipolar

Five strains of three closely related bacterial species were analysed using either generator as RF power supply. *Klebsiella pneumoniae* (5 strains), *Proteus mirabilis* (4 strains) and *Serratia marcescens* (5 strains) are all part of the closely related Enterobacteriaceae family and frequently show misclassifications in cross-validation results. Each strain was sequentially analysed using either generator on a Thermo LTQ XL instrument and the resulting dataset (comprising 28 data points) subjected to principal components analysis (PCA) to assess the influence of the electrosurgical generator on the spectral profiles. A comparison of REIMS profiles for an identical strain of *Proteus mirabilis* using both power supplies is shown in Figure 27 and displays an overall very similar appearance.

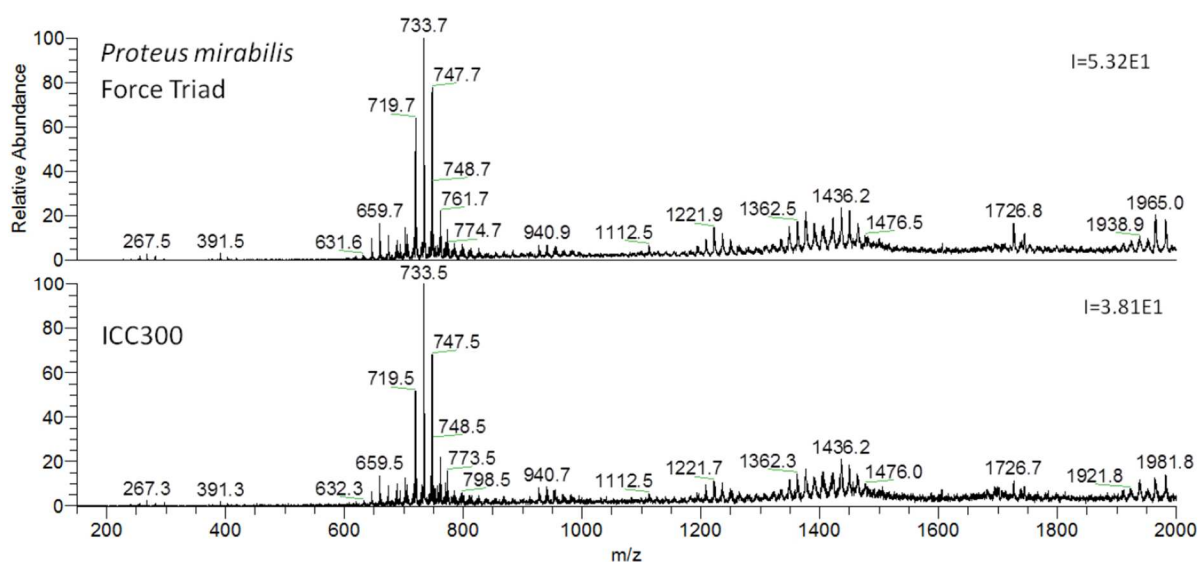


Figure 27. REIMS spectral profiles for a clinical isolate of *Proteus mirabilis* acquired using two different electrosurgical generators as power supply (top: Force Triad, bottom: Erbe ICC300). I=Intensity value at 100% relative abundance. Data acquired on Paul trap instrument.

The resulting PCA plot of the entire dataset is shown in Figure 28 and shows a clear separation of the three different bacterial species. No systematic variation can be observed

originating from the two different power supplies. These results indicate that an arbitrary electrosurgical generator with a bipolar application mode can be used as a power supply and will deliver spectral profiles comparable to those obtained by other models or brands. However, each generator requires optimisation with regard to power output to achieve optimal ionisation characteristics (minimal charring and an even and moderate power output for a large variety of bacterial species). The Force Triad generator generally led to higher signal intensities compared to the Erbe ICC300, however, the Erbe generator had better ionisation characteristics for a wider variety of bacterial species.

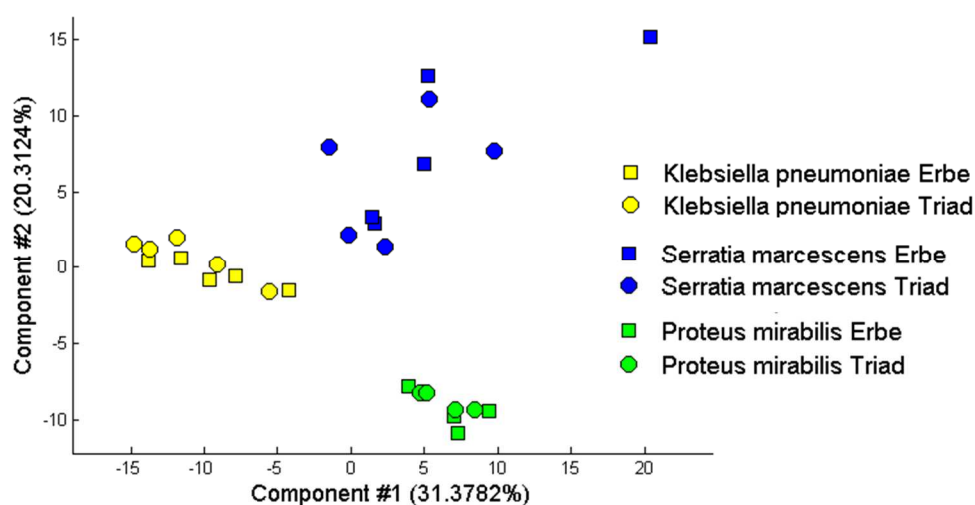


Figure 28. Principal components analysis plot of dataset comprising three different bacterial species recorded using two different electro-surgical power supplies.

3.4. Experimental setup - Instrumental parameters

3.4.1. Comparison of positive and negative ion mode

During the rapid evaporative ionisation processes, positively and negatively charged species are generated in equal amounts. However, as Figure 29 shows comparably poor signal intensity is observed for bacterial species when analysed in positive ion mode. REIMS predominantly ionises hydrophobic molecular species that are either present as ions in the sample material already or such species with very high tendency to form ionic adducts. Therefore the detection of complex lipid species is highly favoured in REIMS mechanism. Bacterial membranes of most bacterial species mostly consist of phosphatidylglycerols, phosphatidylethanolamines and cardiolipins which all undergo ionisation in negative ion mode. The poor signal intensity in positive ion mode is tentatively attributed to the low

abundance of lipid species in bacteria that would readily ionise in positive ion mode due to the absence of phosphatidylcholins and sphingomyelins in most bacteria.(1) Whereas negative ion mode gives strong intensity spectra for all bacterial and fungal species analysed, positive ion mode yields comparable spectral quality for only a limited number of species including *P. aeruginosa*, which features strong quorum-sensing molecule signals or *B. fragilis*, which contains high amounts of sphingolipids. Thus, for the generation of a robust identification method, negative ion mode was chosen due to its constantly high spectral quality for a wide variety of bacterial species.

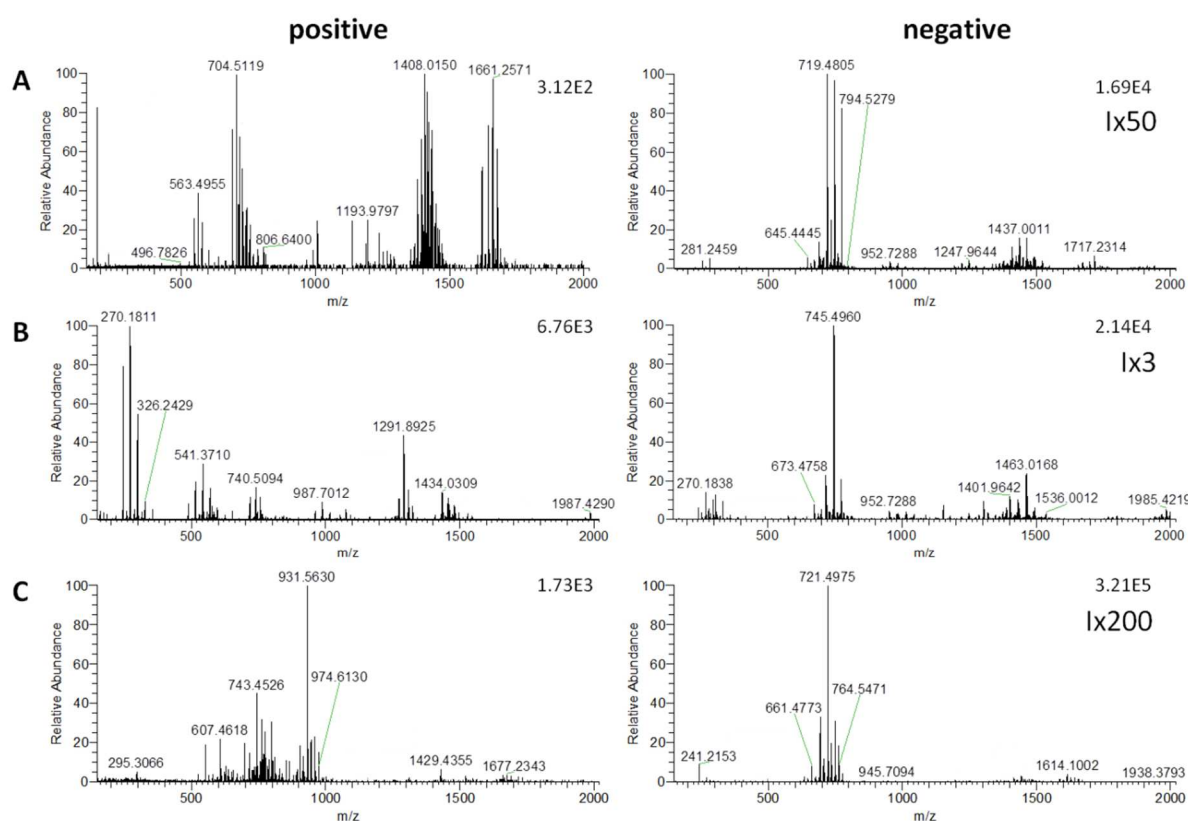


Figure 29. Comparison of REIMS profiles obtained in positive and negative ion mode for A) *Escherichia coli*, B) *Pseudomonas aeruginosa*, C) *Staphylococcus aureus*. I=Intensity. Amended with permission from Ref. (142). Copyright 2014 American Chemical Society. I=Intensity at 100% relative abundance.

3.4.2. Optimisation of ion transfer and ion optical settings

In a common capillary-skimmer-type atmospheric pressure interface (used by Thermo LTQ XL, Orbitrap Discovery and Exactive instruments), there are a small number of parameters that can be adjusted and have influence on spectral appearance. These are capillary voltage and temperature, tube lens (a lens surrounding the inlet capillary close to its internal opening)

voltage, and skimmer voltage. All these parameters were studied over their whole range in detail using *Staphylococcus similans*.

No influence on spectral appearance and total ion current was observed in case of tube lens voltage and capillary voltage settings (see Figure 30A and D). Fluctuations in TIC were observed but were in line with analytical variations observed due to differences in amount of biomass sampled. However, significant effects on signal intensity and spectral appearance were observed in case of capillary temperature and skimmer voltage (see Figure 30B and C). The lack of effect in case of tube lens and capillary voltage is in accordance with earlier studies where no dependence on tube lens voltage was found in case of REIMS while strong dependence on tube lens voltage was observed for electrospray post-ionisation of REIMS aerosols.⁽¹⁵⁹⁾ The ionisation mechanism was investigated using rhodamine 6G pick-up experiments performed in-house on a Thermo Orbitrap Discovery instrument by depositing rhodamine 6G onto several potential collision surfaces inside the mass spectrometer (all ion guiding parts of the atmospheric pressure interface and the first square quadrupole ion guide) and determining its ionisation response during REIMS analysis. Highest rhodamine 6G abundance was found when the inner surface of the skimmer was studied, which suggests that collision of the clusters in the REIMS aerosol occurs predominantly at this position (E. A. Jones and O. Golf *et al.*, data not published). This observation confirms that no desolvated molecular ions pass through the inlet capillary that could be affected by a change in either capillary voltage or tube lense voltage and the actual ion formation occurs in the proximity of the skimmer electrode.

In agreement with this hypothesis, the skimmer potential setting was found to have a significant influence on both spectral appearance and ion yield. When decreasing skimmer voltages from 0 to -50V, mass spectra can be observed at all values, however, a clear onset value for efficient ion generation was observed at -15V. From then onwards, no influence on total ion current was observed by further increasing the skimmer voltage. However, skimmer voltage was further found to influence the spectral appearance over the entire range tested. Fragmentation was observed at very and high skimmer voltages low (<-5V, >-45V, see Figure 31, left) as indicated by changes in the abundance of a peak at m/z 241 which corresponds to fatty acid C15:0 which is part of all major phospholipids observed for *Staphylococcus* species (see Table 7). Furthermore, large molecular weight ions ($m/z > 1000$) were found to decrease in relative intensity for very high skimmer voltages.

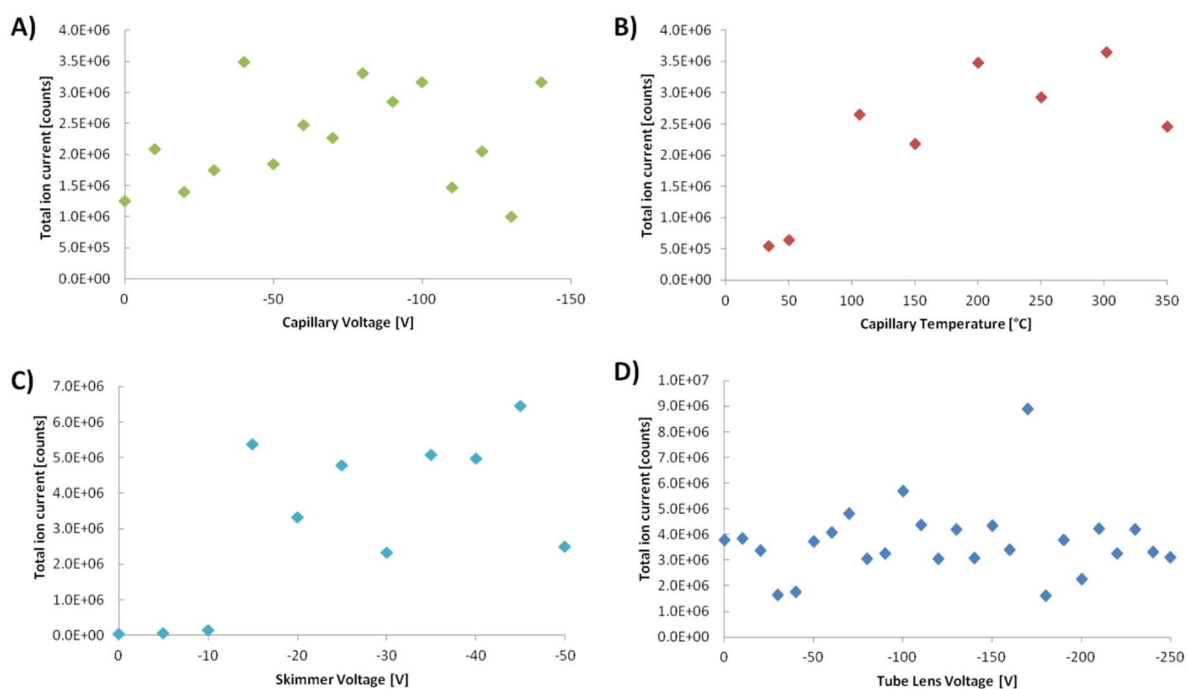


Figure 30. Total ion current for repeated measurements of *Staphylococcus similans* as a function of A) capillary voltage, B) capillary temperature, C) skimmer voltage, and D) tube lens voltage.

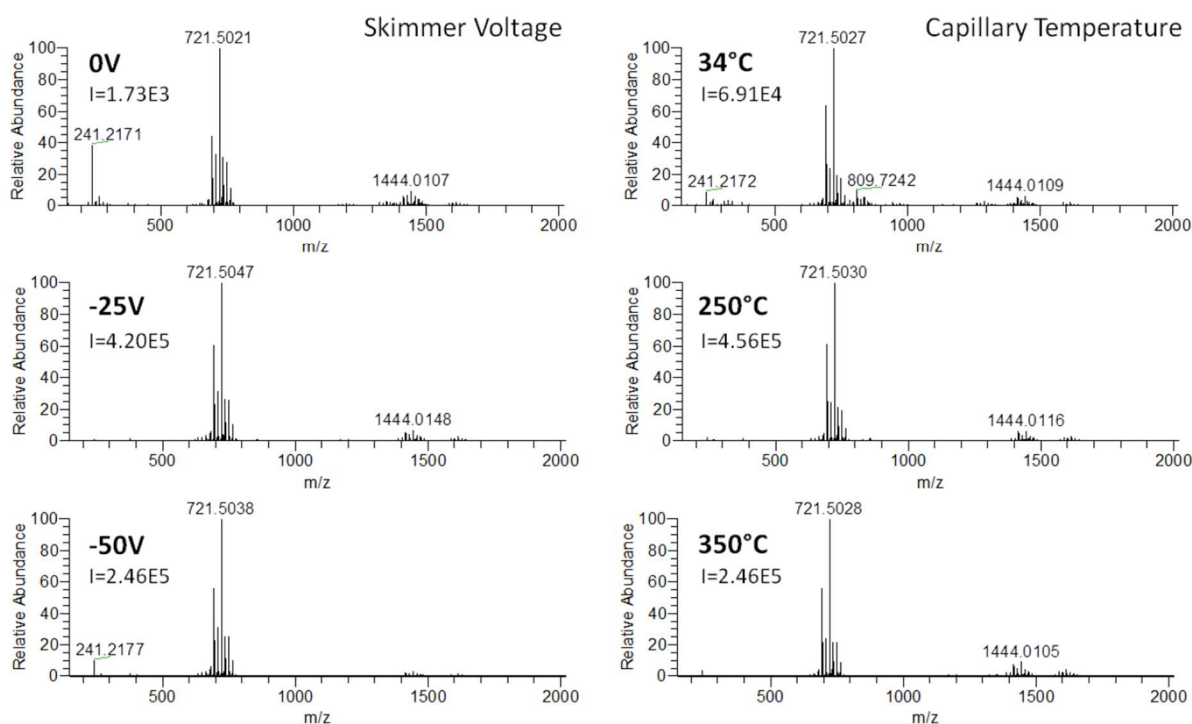


Figure 31. Spectral appearance in dependence on atmospheric pressure interface skimmer voltage (left) and capillary temperature (right). I=Intensity value at 100% relative abundance.

Although the viscosity of aspirated air is increased at higher capillary temperatures resulting in a net drop of sample introduced, the observed signal intensities were found to be higher.

The temperature likely has an effect on the internal energy of molecular clusters, influencing the efficiency of desolvation on the surface collision event discussed above (see Figure 30B and Figure 32, right). Significantly higher signal intensities were observed for capillary temperatures above 100°C. The total ion current shows minimal further improvement and a plateau effect towards even higher capillary temperatures.

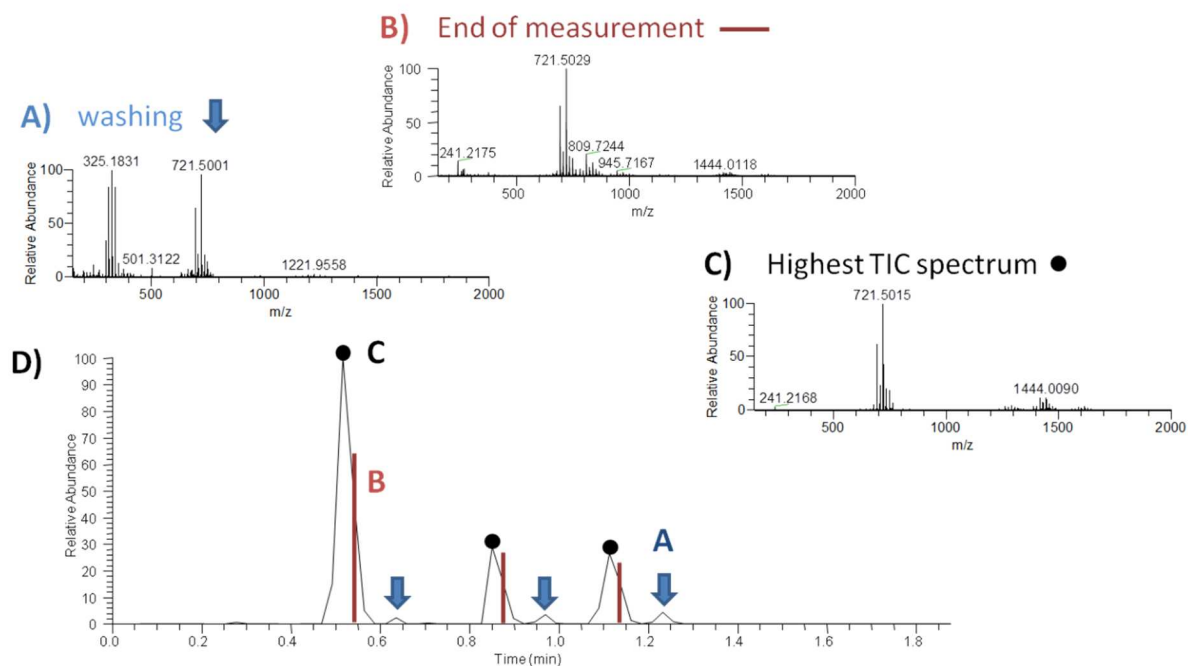


Figure 32. Spectral appearance for three consecutive REIMS measurements of *Staphylococcus similans* recorded at 34°C capillary temperature. A) Spectrum obtained during washing of sampling probe using 70% ethanol, B) spectrum obtained towards the end of an individual sampling event, C) highest TIC spectrum of a sampling event. D) Total ion count (TIC) trace of three consecutive sampling events.

However, the capillary temperature was found to influence also the spectral appearance. For temperatures below 100°C, high variance of spectral profiles was observed during a single REIMS measurement as demonstrated in Figure 32. While for each individual measurement event the spectrum with highest TIC looks identical to those recorded at higher capillary temperatures, an additional set of peaks can be observed between m/z 800-900 towards the end of each measurement. The influence of these additional peaks disappears towards higher temperatures, and it is virtually absent from 250°C onwards. Very low capillary temperatures (< 100°C) cause the appearance of signals obtained during the cleaning procedure of the electrode tips (tips are wiped off using 70% ethanol after each analysis). This unfavourable effect is also disappearing towards higher capillary temperatures.

For these reasons, 250°C capillary temperature was found to constitute an optimum for stable spectral appearance, sensitivity and suitability for possibly thermally unstable analytes.

3.4.3. Suitability of different mass analysers

In order to assess reproducibility of REIMS profiles of microorganisms, the developed methodology was tested on different instrumental platforms. Overall six different mass spectrometers were tested comprising three different types of mass analysers (1x Paul trap, 2x Orbitrap, 2xToF, 1x IMS-ToF). Exemplary REIMS spectral profiles obtained for *Escherichia coli* strains using a Paul trap (B), Orbitrap (A) and time-of-flight (C) mass analyser are shown in Figure 33.

All different instrument platforms and mass analysers were found applicable for the analysis of microorganisms using REIMS and were found to yield similar mass spectral profiles dominated by the same mass spectral features in the mass range m/z 600-800, and smaller peak groups at m/z 1200-1300, m/z 1350-1550, and m/z 1650-1750. Peaks at m/z 1200-1300 are not visible in the mass spectrum obtained on the Paul trap, however, these are likely covered by the baseline noise. Abundant signals in the low mass region below m/z 500 are observed in the time-of-flight mass analyser, indicating a larger degree of fragmentation. However, the observed lipid fragmentation products are more likely to be generated by the different atmospheric pressure interfaces used (skimmer-type interface in ion trapping instruments, custom-built interface with inserted collision sphere in time-of-flight instruments). The effect of different interfaces between REIMS setup and mass analysers is further discussed under the point Experimental setup – Adaptation to Xevo G2-XS instrumental platform - Different atmospheric pressure interfaces.

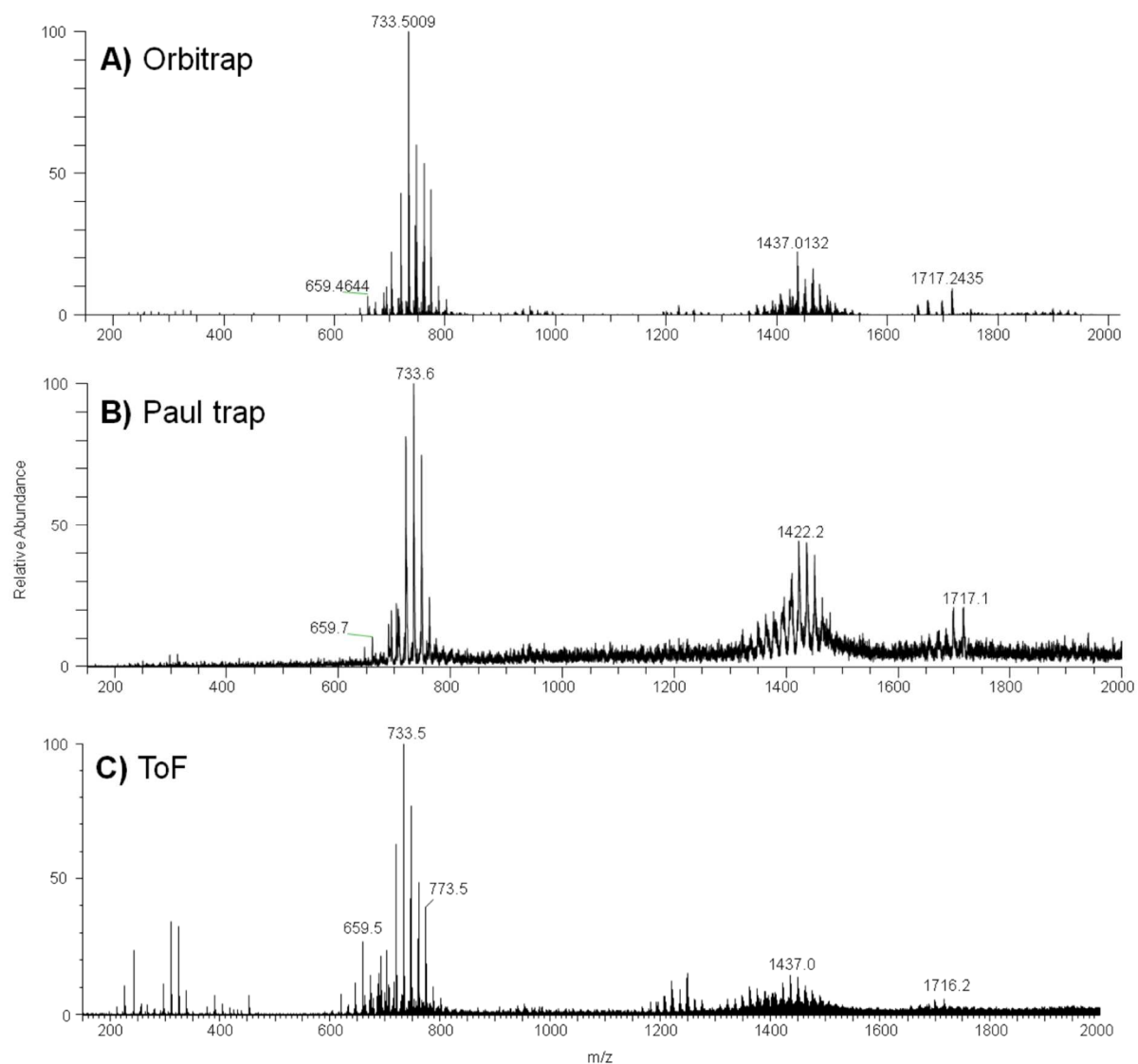


Figure 33. Spectrum of *Escherichia coli* strains analysed on three different instruments with different mass analysers. A) Orbitrap-based mass analyser (Thermo Exactive), B) Paul trap mass analyser (Thermo LTQ XL), C) Time-of-Flight (TOF) mass analyser (Waters Xevo G2S).

3.4.4. Instrument contamination

In theory, every spectral feature detected using the presented direct infusion REIMS setup is originating from the sampled material as no other material (solvent or matrix) is introduced into the mass spectrometer than the produced aerosol. This greatly helps data analysis as compared to spray methods and matrix-assisted desorption methods that often require background-subtraction before further downstream data analysis methods could be applied.

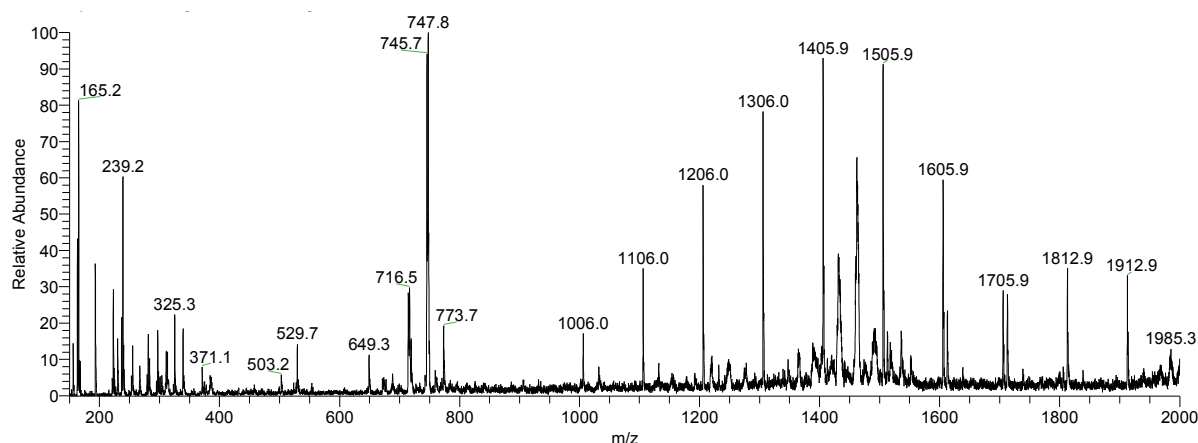


Figure 34. REIMS spectra of *Pseudomonas aeruginosa* recorded on a Thermo Scientific Orbitrap Discovery XL hybrid instrument using the LTQ XL analyser without prior cleaning. Ultramark 1621 contamination can be seen as polymeric structure ranging from m/z 1000-2000. Further contaminants include polysiloxanes at $m/z < 400$.

However in practice, due to the ionisation mechanism involving surface collisions inside the mass spectrometer atmospheric pressure interface, contaminants are likely to be observed using REIMS. Commonly observed contaminants comprised C16:0 and C18:0 fatty acids, polysiloxane contaminants ($\Delta m = 74$) and such contaminants that were actively introduced into the instrument during REIMS analysis (bacterial metabolites) and during instrument calibration. Thermo instruments were calibrated using the commercially available Thermo Fisher Scientific Pierce™ calibration mix which in negative ion mode consists of sodium dodecyl sulphate (SDS), taurocholic acid and polymeric Ultramark 1621. While no spectral contaminants were observed that originated from SDS and taurocholic acid, extensive contamination were observed due to Ultramark as displayed in Figure 34 showing the polymeric distribution with a distinct mass spacing of $\Delta m = 100\text{Da}$.

To avoid detecting features originating from a contaminated atmospheric pressure interface, any instrumentation used in the presented work with a cold collision surface (this excludes the heated coil custom-made API used in conjunction with the Waters Xevo G2-XS instrument) was cleaned prior to REIMS analyses to ensure that the spectral influence of contaminants was negligible. Instrument contaminants are less critical in such methodologies where ions are generated outside of the instrument (such as ESI and DESI) and thus the instruments in those cases need less frequent cleaning.

In case of the Thermo Scientific Exactive, LTQ XL and Orbitrap Discovery XL instruments, the instrument is vented and inlet capillary, tube lens and skimmer were cleaned to ensure all contaminants from previous experiments are removed. The skimmer was found to be the

main origin of contamination and was thus cleaned (together with the tube lens) by initially wiping off contaminants using a cotton swab and methanol before subjecting it to ultrasonication in five different solvents of increasing hydrophobicity. These were methanol/water (1:1 v/v), acetonitrile, acetone, dichloromethane and finally methanol. Sonicating the skimmer in a halogenated solvent was found to be essential to remove the perfluorinated Ultramark polymer. The inlet capillary was ultrasonicated in a mixture of methanol/water/acetone (1:1:1 v/v/v). Other accessible surfaces of the atmospheric pressure interface were wiped off using a cotton swab and methanol before the API was reassembled and reinserted into the instrument, followed by a minimum of 12hrs bakeout to achieve sufficiently good vacuum for mass analysis. All FTMS instruments used during this study do require calibration after maintenance procedures to obtain sufficiently high mass accuracy. As calibration using the commercial HESI ion source led to returning Ultramark contaminations in the REIMS spectra, calibration was performed using nanospray ionisation devices. For this purpose, gold-coated nanospray needles with 5µm inner diameter tips were used (obtained from DNU-MS, Berlin, Germany). Spray durations were kept as short as possible (below 30s) and only the required ion mode was calibrated which was found to leave no significant contamination to the instrument.

To reduce the effect of memory effects caused by compounds introduced during REIMS analysis, only a limited number of spectrally similar samples were analysed on a single day (max 45, depending on species) and were randomised with at least an equal number of a selection of taxonomically unrelated species. Such bacterial species that were empirically found to be producing metabolites that strongly contaminate the instrument were analysed in the end of the day in the order of increasing severity of induced contaminations. Only a limited number of database entries could be recorded per day in case of *Bacteroides* spp. (n<5, corresponding to 25 individual REIMS measurements) before carry-over of ceramide signals was observed, while only a single database entry of lipopeptides-producing *Bacillus* spp. (*B. subtilis*, *B. licheniformis*) could be recorded with as little as 3 individual REIMS measurements leading to persistent instrument contamination. While some cleaning success can be obtained without venting the instrument in case of time-of-flight analysers by introduction of hydrophobic solvents into the cooled atmospheric pressure interface, this is not an option for FTMS instruments as the evaporating solvent will destroy the ultrahigh vacuum.

Although contaminations introduced during calibration could be avoided by use of a customised calibration solution where Ultramark 1621 was replaced by compounds inducing less memory effects, the contamination to the skimmer surfaces during REIMS analysis by microbial species is a persisting problem that in case of the available skimmer-type Thermo Scientific instrumentation is very difficult to solve without substantial customisation of the atmospheric pressure interface and following ion transfer optics. Waters Corporation (Milford, MA, USA) instrumentation equipped with a Stepwave™ ion guide was found ideally suited for development of an atmospheric pressure interface as by default they do not contain any possible deposition and collision surfaces, they are equipped with a time-of-flight mass analyser requiring less stringent vacuum conditions, and any modifications of the atmospheric pressure interface can be made outside of the instrument without the need for major reconstruction.

3.5. Experimental setup – Adaptation to Xevo G2-XS instrumental platform

3.5.1. Different atmospheric pressure interfaces

While no influence on spectral appearance and signal intensity was observed based on using different instruments and mass analysers *per se*, significant differences were observed related to the atmospheric pressure interfaces (API). A schematic of the four different types of APIs used in this study are shown in Figure 35 and include a) traditional commercial skimmer-based API as used in Thermo Scientific instruments, b) commercial Stepwave™-based interface as used in Waters Corporation instruments, c) customised API for Waters instruments comprising a cold sphere as collision surface and d) customised API for Waters instruments comprising a heated coil-shaped collision surface.

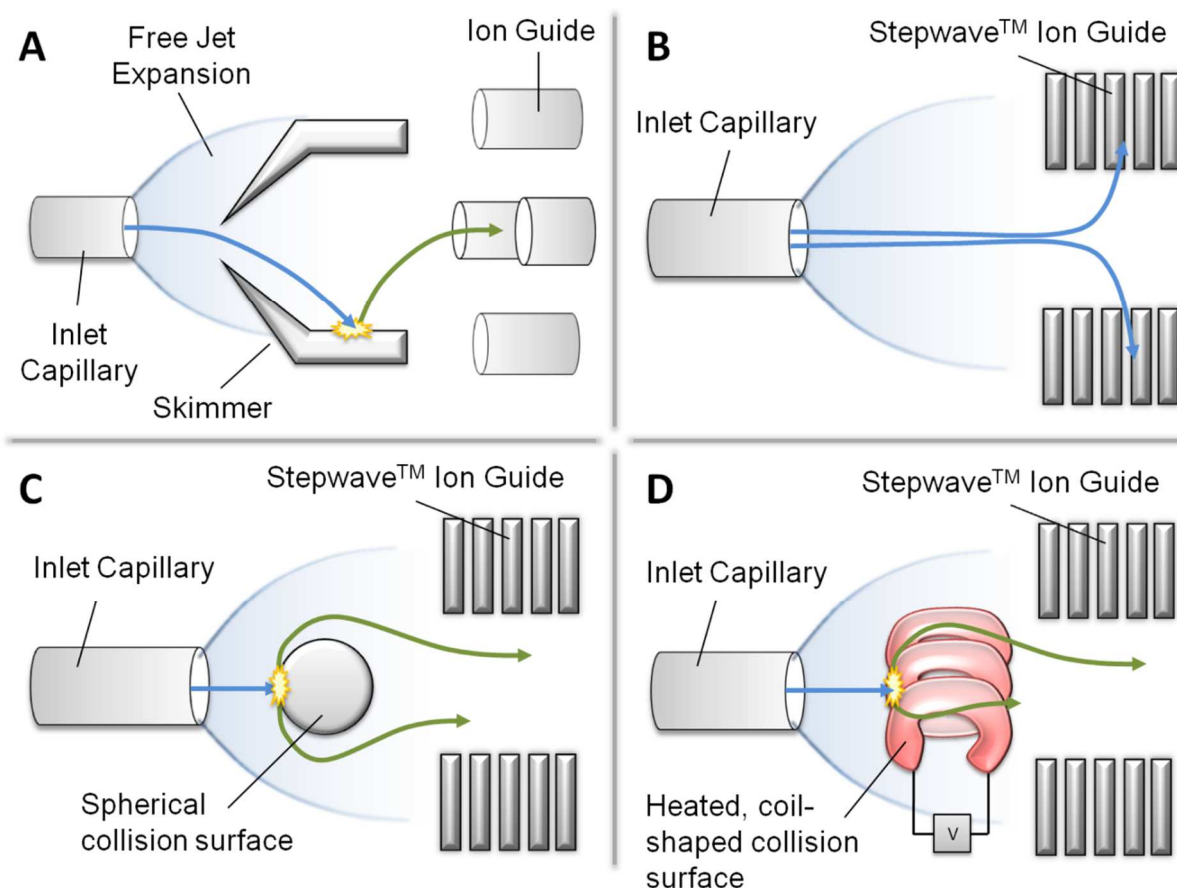


Figure 35. Different atmospheric pressure interfaces used in this study. A) Commercial skimmer interface of Thermo Exactive instrument, B) commercial Stepwave interface of Waters instruments, C) and D) custom-built atmospheric pressure interfaces for Waters instruments. Blue arrows indicate flight paths of neutral species while green arrows indicate the flight path of generated ionic species.

The surface collision mechanism in the case of the default skimmer-type interface is displayed in Figure 35A. The actual surface collision involved in the REIMS ionisation mechanism were found not to occur on the outer conical surface of the skimmer but rather at the inner surface directly on front of further ion optical elements. This is due to the free-jet expansion that is occurring at the opening of the inlet capillary inside the first differential pumping stage (gas pressure approx. 1mbar). Due to this expansion, the trajectory of a small subset of molecular clusters generated during REIMS ionisation leads through the skimmer opening and collides with its inner surfaces from where the ions continue through the ion optics towards the mass analyser. This mechanism also implies that large parts of the actual molecular clusters are not undergoing this secondary ionisation step by not entering through the skimmer opening. Thus, a significant increase in sensitivity is expected if a larger part of the sampled aerosol could be induced to undergo surface collisions.

While the unmodified capillary-skimmer interface (Thermo Scientific instruments) allows the REIMS analysis of biological samples, the standard Stepwave-based API of the Waters instruments was found not to be suited for REIMS analysis because of a lack of mass spectral response. Both capillary-skimmer and Stepwave setups are meant to separate neutral from ionic species, however, while the skimmer eliminates neutrals by having them collide with its outer conical surface directly in between the opening of the inlet capillary and further ion optics, in the Stepwave ion guide charged particles are diverted from the straight flight path at a remote location from both further ion optics and inlet capillary. In this setup (shown in Figure 36), ions generated from surface collisions of large neutral clusters have no chance of following the flight path of pre-existing ionic species as possible surface collisions will occur at a large distance from any ion optical elements. The lack of REIMS signal observed in instruments deploying these commercially available interfaces again shows that the rapid evaporative ionisation process in itself does not create ions *per se* but that the ions are rather generated inside the mass spectrometer.

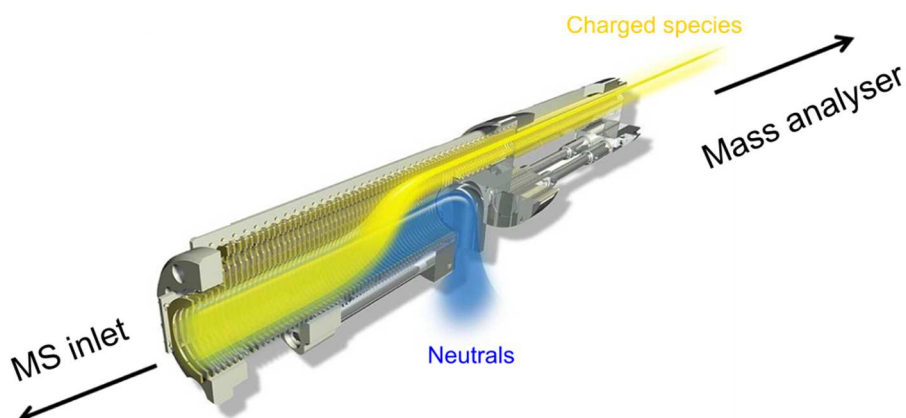


Figure 36. Schematic representation of ion flow in Waters Stepwave ion guide (yellow: ions, blue: neutrals). Reprinted and modified with permission of Waters Corporation (Milford, MA, USA).

This ionisation mechanism is further proven by the generation of ions upon introduction of a collision surface between the end of the inlet capillary and the front end of the Stepwave ion guide as depicted for two different setups in Figure 35C and D. Figure 35C depicts an atmospheric pressure interface with a spherical stainless steel collision surface mounted at 6mm distance from the end of the inlet capillary. This setup was found to give identical spectral profiles to those obtained by the commercial skimmer-type API for a wide variety of bacterial species and thus enabling REIMS analysis on instruments using the Stepwave technology. However, this interface as well as the skimmer-type interface was found to suffer from severe carryover effects in case of the repeated measurements of certain analytes such

as sphingolipids or lipopolypeptides. As outlined in Method Development and Characterisation - Experimental setup - Instrumental parameters - Instrument contamination this could largely be avoided by cleaning the atmospheric pressure interface before analysis and by randomising different bacterial species during analysis of larger sample sets. However, these steps are not applicable for an unsupervised and ideally automated clinical microbiological setting.

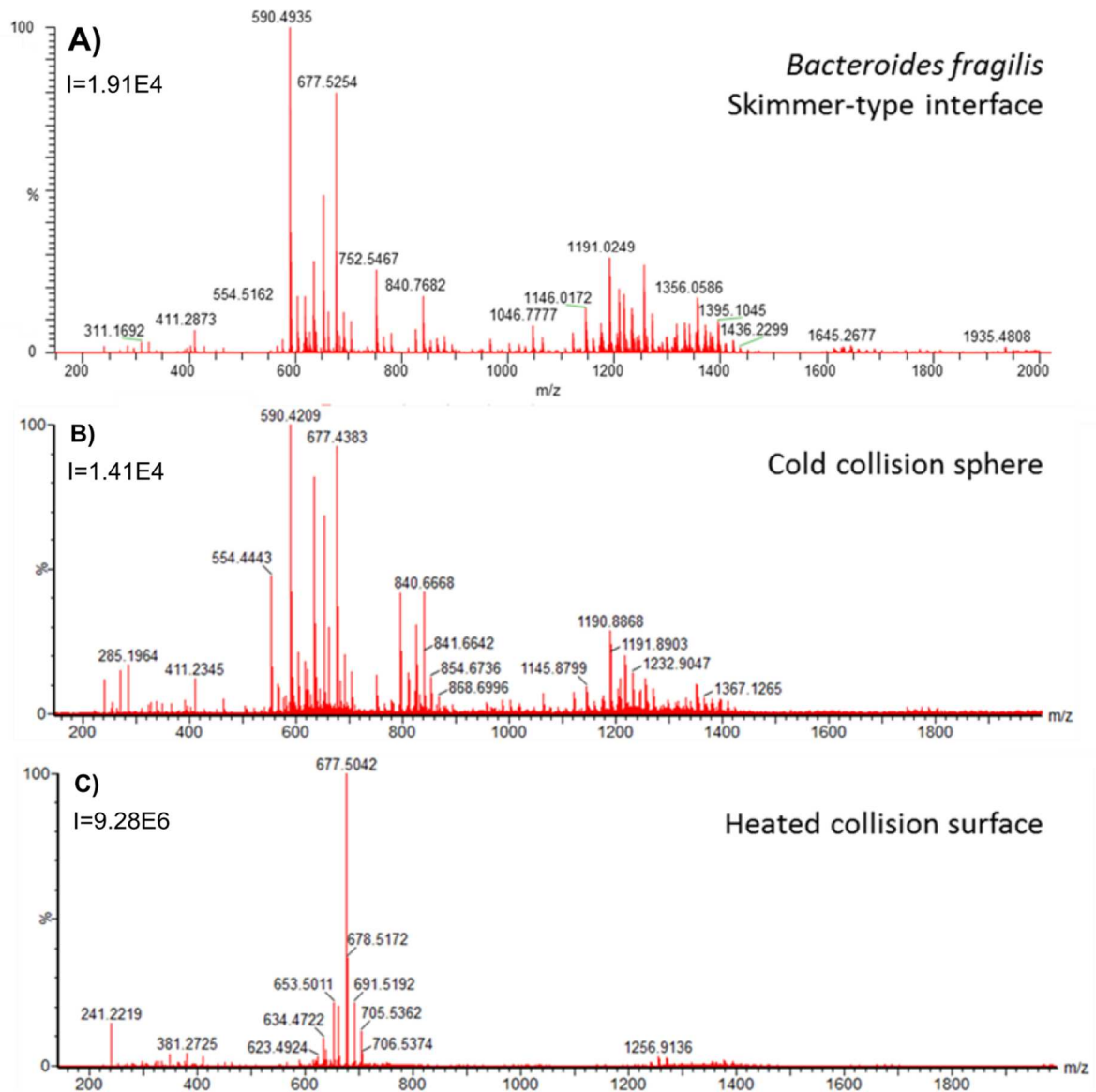


Figure 37. Spectral profiles of *Bacteroides fragilis* obtained using different atmospheric pressure interfaces depicted in Figure 35. A) Skimmer-type interface was recorded on Exactive instrument, while B) cold collision sphere and C) heated collision surface were recorded on Waters Xevo G2 instrument. I=Intensity at 100% relative abundance.

To avoid cross-contamination, a collision surface was tested that could be heated up to several hundreds °C. By heating the collision surface, any organic deposits will react with the oxygen introduced through the inlet capillary and be converted to CO₂ and thus avoid contamination build-up. A coil-shaped collision surface was used instead of a stainless steel spherical collision surface as it was found to be the most beneficial configuration to ensure equal heat distribution when developed for intra-surgical applications. The coil-shaped collision element was manufactured out of Kanthal S, a heat-resistant iron-chromium-aluminium (FeCrAl) alloy, allowing operating temperatures in the range of 700-800°C and supplied by the Waters Research Center (Budapest, Hungary). Using this heated collision surface, memory effects and thus the frequency of instrument cleaning procedure could be greatly reduced. Hundreds of database entries could be recorded without memory effects and even prolonged exposure to lipopolypeptides did not result in any observed carry-over.

Nevertheless, spectral profiles recorded using this heated coil interface in some cases showed markedly different appearance compared to those recorded using cold collision surfaces as illustrated in Figure 37. This indicates that not all spectral constituents are thermally stable enough for this type of heated surface induced ionisation. The effect of the heated coil seems to be especially strong on phosphatidic acid (common in fungi such as *C. albicans*, Figure 39) and some sphingolipid species (common in phylum Bacteroidetes, Figure 37) while generally little effect on spectral appearance can be observed for phosphatidylglycerol and phosphatidylethanolamines (main phospholipid species in *Proteus mirabilis*, Figure 40). The heated coil was also observed to lead to a loss of spectral information for *Pseudomonas aeruginosa* where the loss of low mass quorum-sensing molecules (important virulence factors) was observed.

3.5.2. Aerosol pick up using isopropyl alcohol

Further setups were tested that rely on introduction of isopropyl alcohol to aid ionisation by partially or fully solvating the analyte molecules and thus reducing intermolecular forces which would otherwise negatively affect ionisation hence reducing sensitivity. The introduction of isopropyl alcohol was first tested for the heated interface dedicated for intra-surgical tissue identification which is using a Venturi pump for aerosol transfer from the monopolar sampling device to the inlet capillary of the mass spectrometer. Aerosol sampling by the mass spectrometer in this case occurs orthogonally to the outlet of the Venturi pump. For the bipolar forceps sampling probe, the application of a Venturi pump is disadvantageous

compared to the aspiration by the vacuum system of the instrument, since it leads to considerable dilution of the sampled aerosols and thus loss in sensitivity.

In comparison to an API applying cold collision surface (skimmer-type API, custom-built interface as shown in Figure 35A and C), the heated collision surface was found to eliminate certain spectral features such as ceramides in *Bacteroides fragilis* (see Figure 37). Introduction of isopropanol into the sampled aerosol before entering the mass spectrometer was found to restore those spectral features and generate mass spectral fingerprints similar to those obtained by an atmospheric pressure interface with cold collision surface. In addition, the introduction of isopropanol to the aerosol led to similar or higher signal intensity as compared to direct aerosol introduction and thus theoretically enabled the use of a Venturi pump for aerosol transport. However, increases in sensitivity will be neutralised by the introduction of a Venturi setup. Thus, a further increase in sensitivity was expected for direct introduction of isopropanol into the aerosol line without applying additional nebulising setups. For this purpose, a T-piece was constructed as depicted in Figure 38. This T-piece was directly attached onto an extended mass spectrometer inlet capillary, thus still enabling sampling using the inherent vacuum system of the instrument. Increasing isopropanol flow rates were tested between 0-0.25mL/min and the optimum flow rate was determined to be 0.1mL/min.

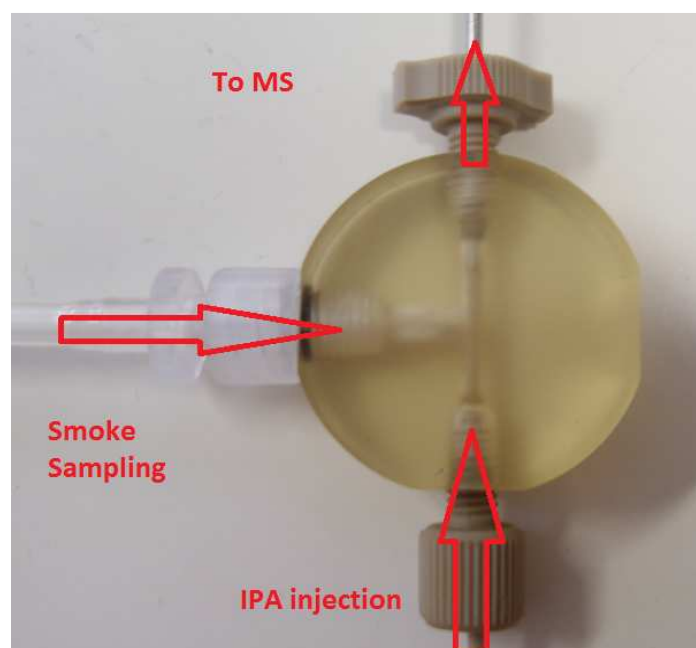


Figure 38. Photograph of apparatus used for introduction of isopropanol using bipolar sampling probe and direct introduction into the mass spectrometer.

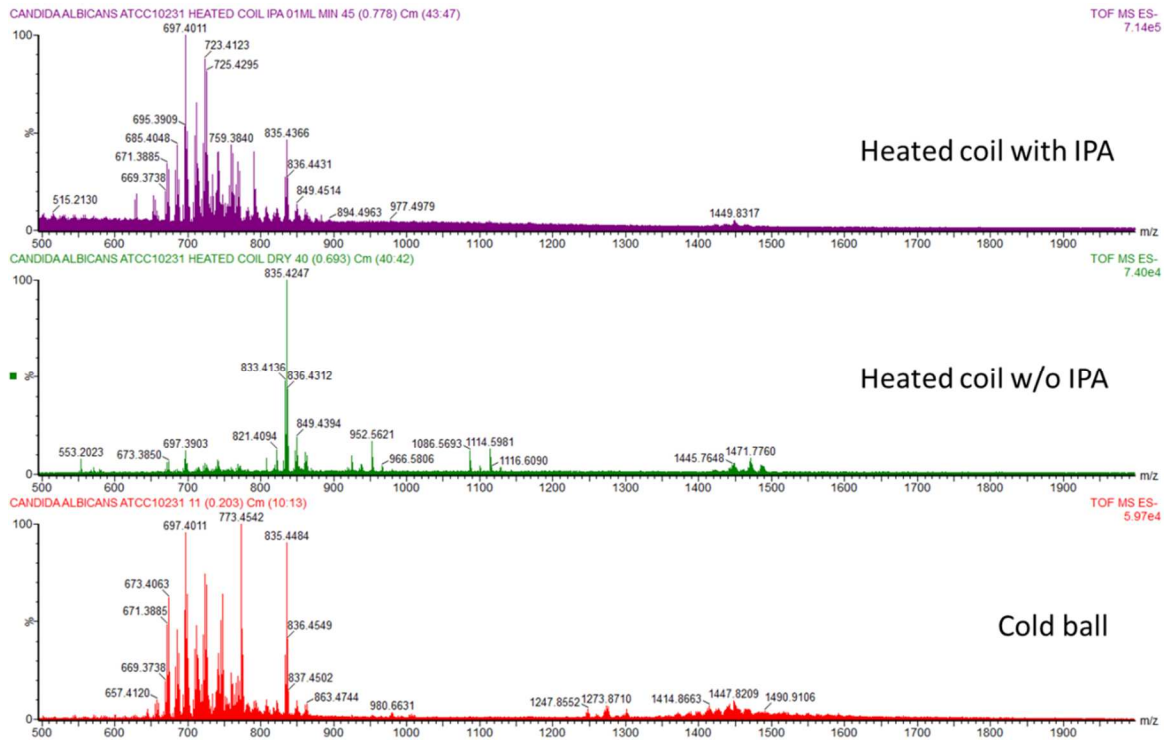


Figure 39. Spectral profiles for *Candida albicans* (yeast) using different experimental setups: heated coil with introduction of IPA, heated coil without IPA introduction, and cold collision sphere. All data recorded on Xevo G2XS instrument.

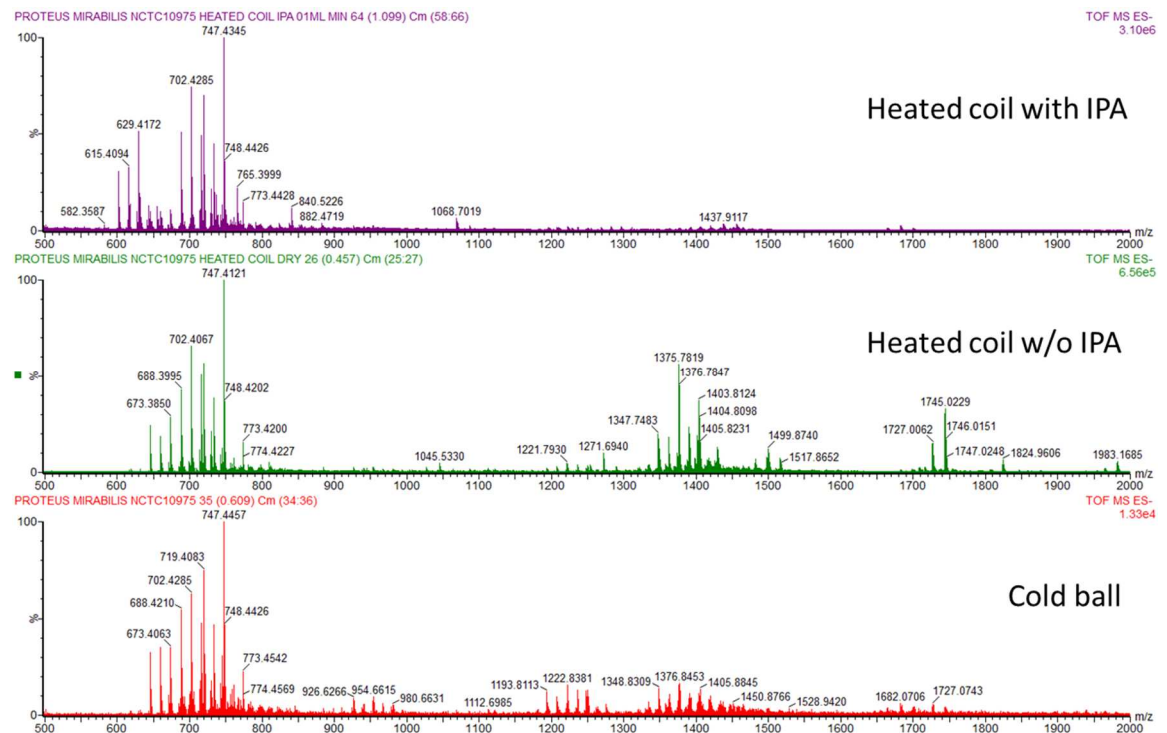


Figure 40. Spectral profiles of *Proteus mirabilis* (Gram-negative) using different experimental setups: heated coil with introduction of IPA, heated coil without IPA introduction, and cold collision sphere. All data recorded on Xevo G2XS instrument.

The effect of the different setups on spectral appearance is shown for *Candida albicans* and *Proteus mirabilis* in Figure 39 and Figure 40, respectively. Both examples clearly demonstrate that using a heated instead of a cold interface results in significant changes of the spectral appearance with many spectral features in *Candida albicans* being significantly reduced in relative intensity or disappearing altogether. Introduction of isopropanol into the system helps to circumvent this problem and creates a spectrum more similar to cold collision surface interfaces. However, one observed disadvantage is the observation of a rising baseline towards lower masses effectively reducing signal-to-noise ratio. Additionally, the use of isopropanol results in significant loss of mass spectral information above m/z 1000 as clearly visible in case of *Proteus mirabilis* (Figure 40).

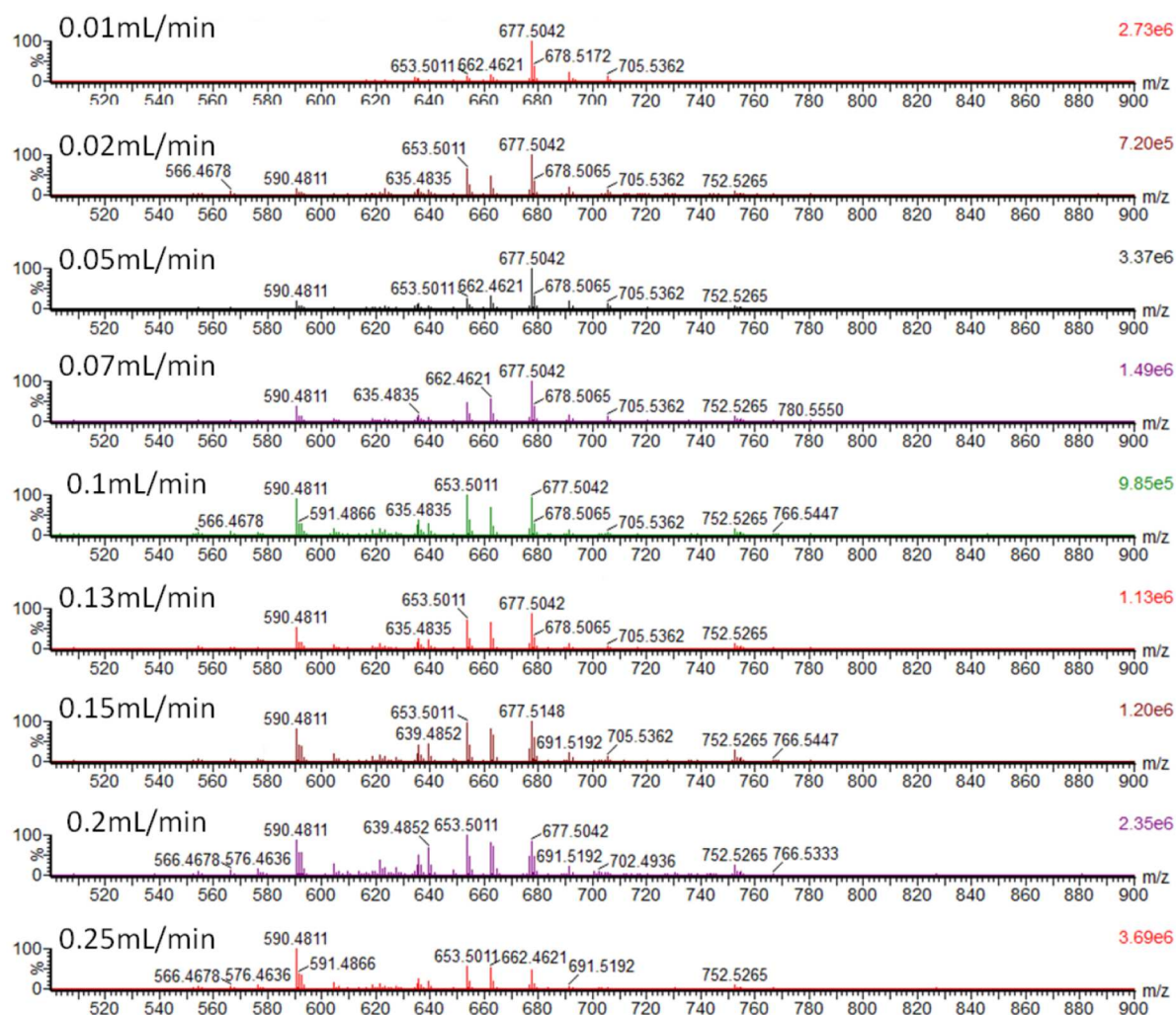


Figure 41. REIMS spectra obtained for *Bacteroides fragilis* using IPA mediated aerosol pick-up setup operated with different IPA flow rates. Zoom into mass range of m/z 500-900 covering phospholipid region.

The effect on spectral appearance for different isopropanol flow rates was determined for *Bacteroides fragilis* and is displayed in Figure 41 and Figure 42. A direct effect of the isopropanol being present is detectable from 0.02mL/min upwards as clearly visible based in case of the appearance of m/z 590 (ceramide species) and m/z 752 (α -Galactosylceramide). These species were found to increase in their relative abundance with increasing isopropanol flow rates. As soon as the appearance of m/z 590 and 752 sets in, peaks in the very high mass region $m/z >2000$ were found to disappear (see Figure 42) thus indicating a negative influence on heavier spectral features.

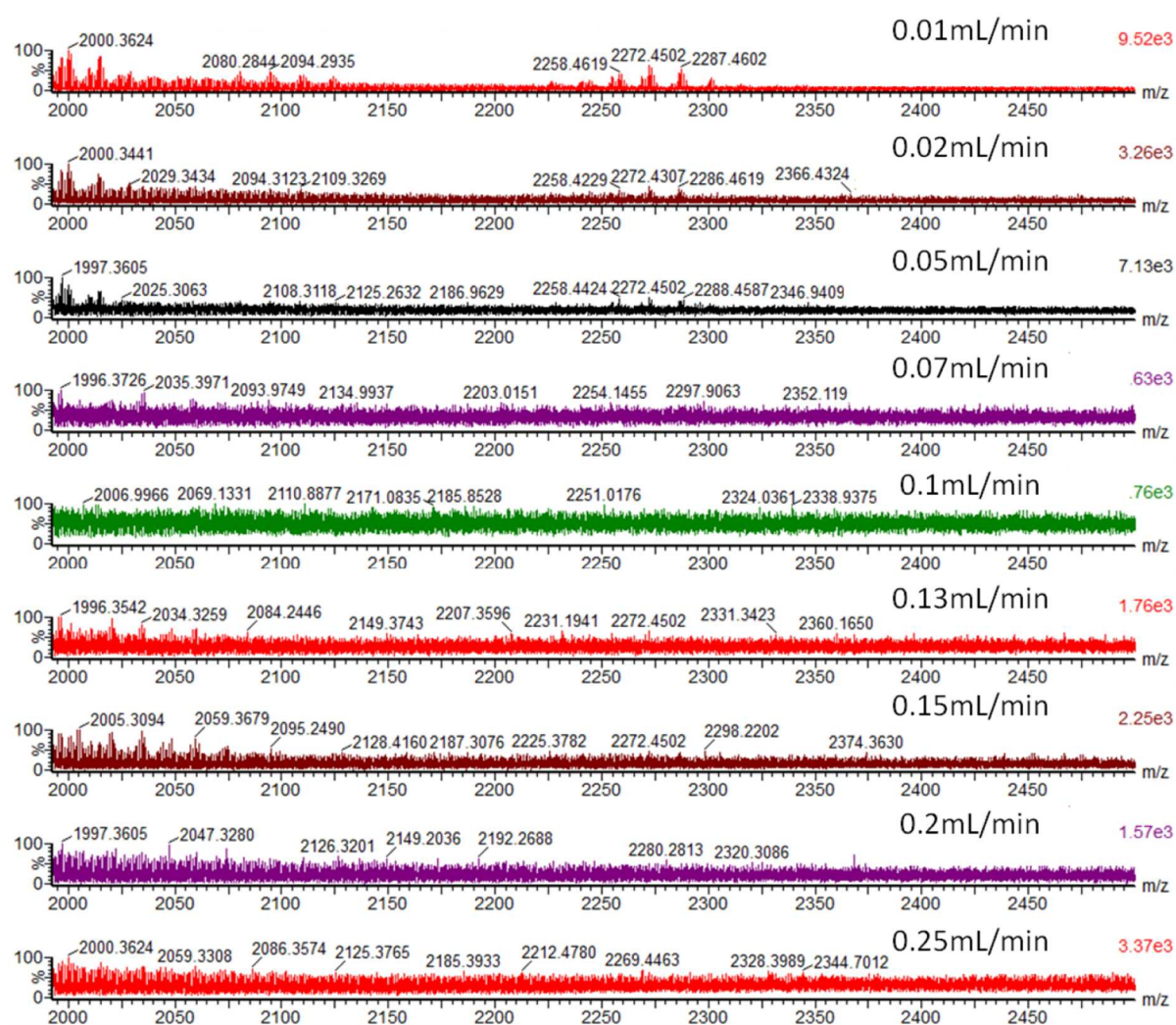


Figure 42. REIMS spectra obtained for *Bacteroides fragilis* using aerosol pick-up setup operated with different isopropanol flow rates. Zoom into mass range of m/z 2000-2500 showing effect on very high mass ions.

While a heated atmospheric pressure interface utilising isopropanol and a Venturi air jet pump is the method of choice in case of intra-surgical applications, this setup is not suited for the identification of bacteria. This is due to strong limitations in spectral content originating

from the use of a heated collision surface, which leads to the exclusion of thermally labile compounds, loss in sensitivity (due to dilution occurring because of the Venturi pump), and loss of high mass features and reproducibility due to the introduction of isopropanol. Thus, the optimum setup for a commercial platform for the identification of bacteria would be a setup utilising direct introduction of the aerosol and a cold collision surface. In order to avoid cross-contamination, this collision surface should be exposed to short periods of heating between mass spectral analyses rather than being constantly heated up in order to obtain optimum spectral quality in an unsupervised clinical microbiology setting.

3.6. Characterisation of spectral reproducibility

The inter-species variance of data has to be larger than intra-species variance for any microbial identification tool. The biological variance introduced to the overall spectral appearance by different culturing media, different culturing atmospheres and different strains of the same species was assessed and the results are shown in Figure 43, Figure 44 and Figure 45, respectively for the three most common and most extensively studied pathogenic species encountered in clinical microbiology, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Escherichia coli*.

Figure 43 shows REIMS spectral profiles obtained for five different clinical isolates of the same bacterial species cultured under the same conditions. Excellent pattern stability can be observed for *S. aureus*, while more variability was observed for *P. aeruginosa* and *E. coli*. In case of the latter, some changes in relative signal intensity were observed between major phosphatidylglycerol species in the mass range m/z 600-800 while a very high degree of similarity was observed for the high mass range. In case of *P. aeruginosa*, three of five strains exhibit production of extracellular metabolites such as quorum-sensing molecules and rhamnolipids in the mass range of m/z 200-350 and m/z 500-680, respectively. Furthermore, the same strains exhibit a group of signals in the range of m/z 900-1100. However, the phospholipid region between m/z 680-800 and masses above m/z 1100 show good agreement between all five clinical isolates. These three bacterial species represent three very different behaviours, with *S. aureus* displaying very stable profiles, *E. coli* showing slight changes in phospholipid signal distributions and *P. aeruginosa* showing presence or absence of entire compound classes.

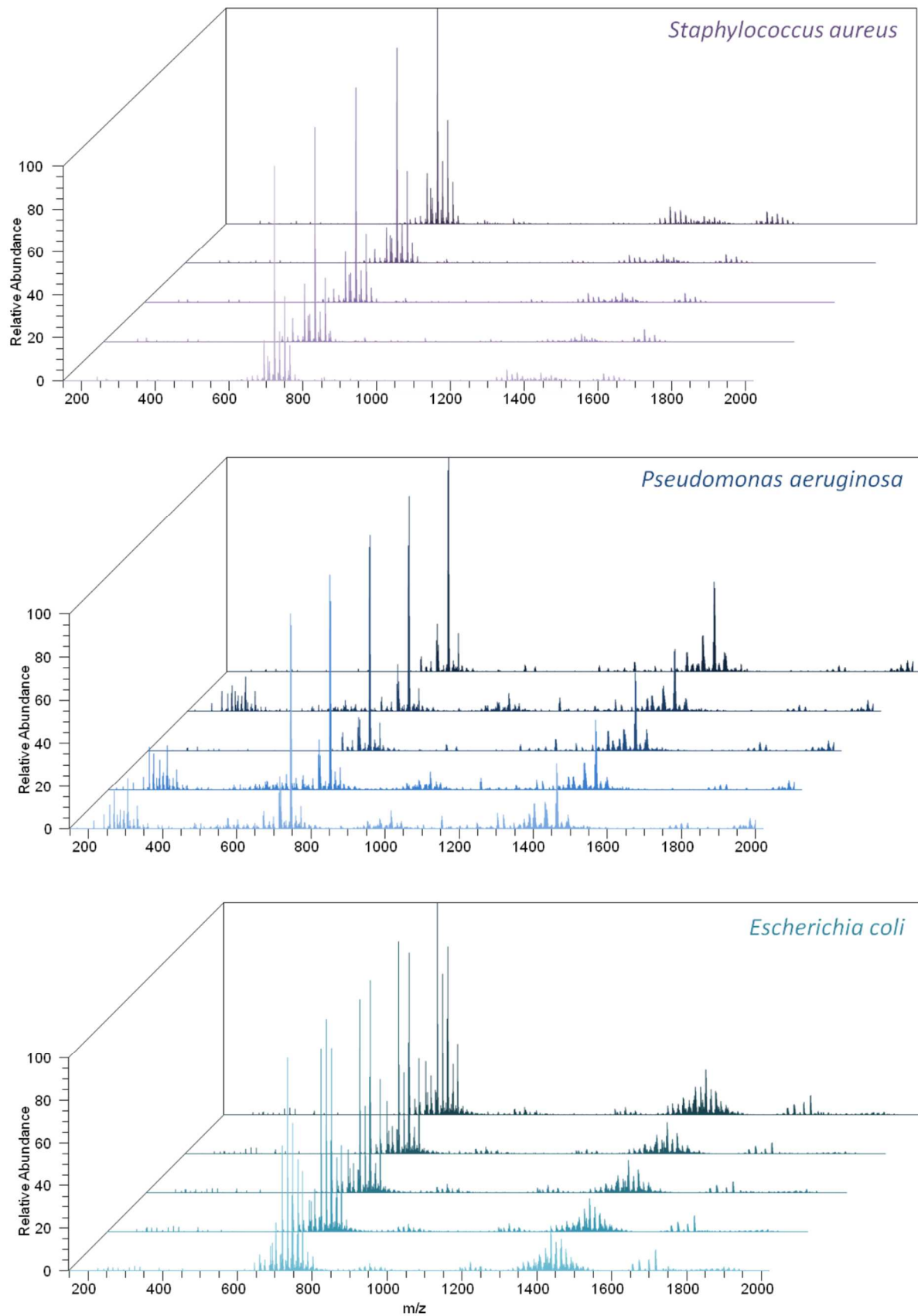


Figure 43. REIMS spectral profiles for each five different clinical isolates of *Staphylococcus aureus* (top), *Pseudomonas aeruginosa* (middle) and *Escherichia coli* (bottom) cultured on Columbia blood agar.

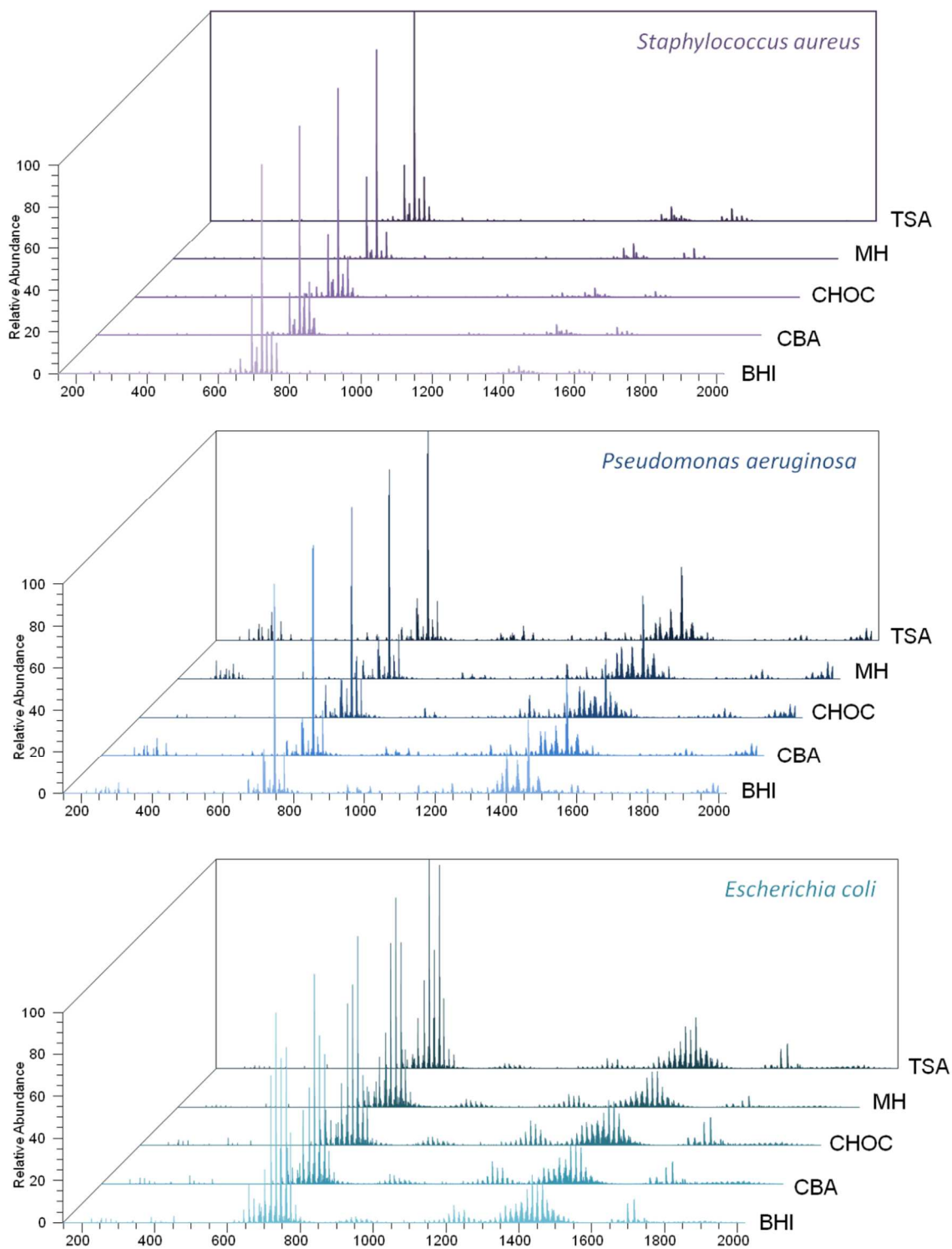


Figure 44. REIMS spectral profiles obtained for *Staphylococcus aureus* ATCC 25923, *Pseudomonas aeruginosa* ATCC 27853 and *Escherichia coli* ATCC 25922, each grown on five different solid growth media. From front to back: Brain-heart infusion agar (BHI), Columbia horse blood agar (CBA), chocolate agar (CHOC), Mueller-Hinton agar (MH), Trypticase soy agar (TSA). Amended with permission from Ref. (142). Copyright 2014 American Chemical Society.

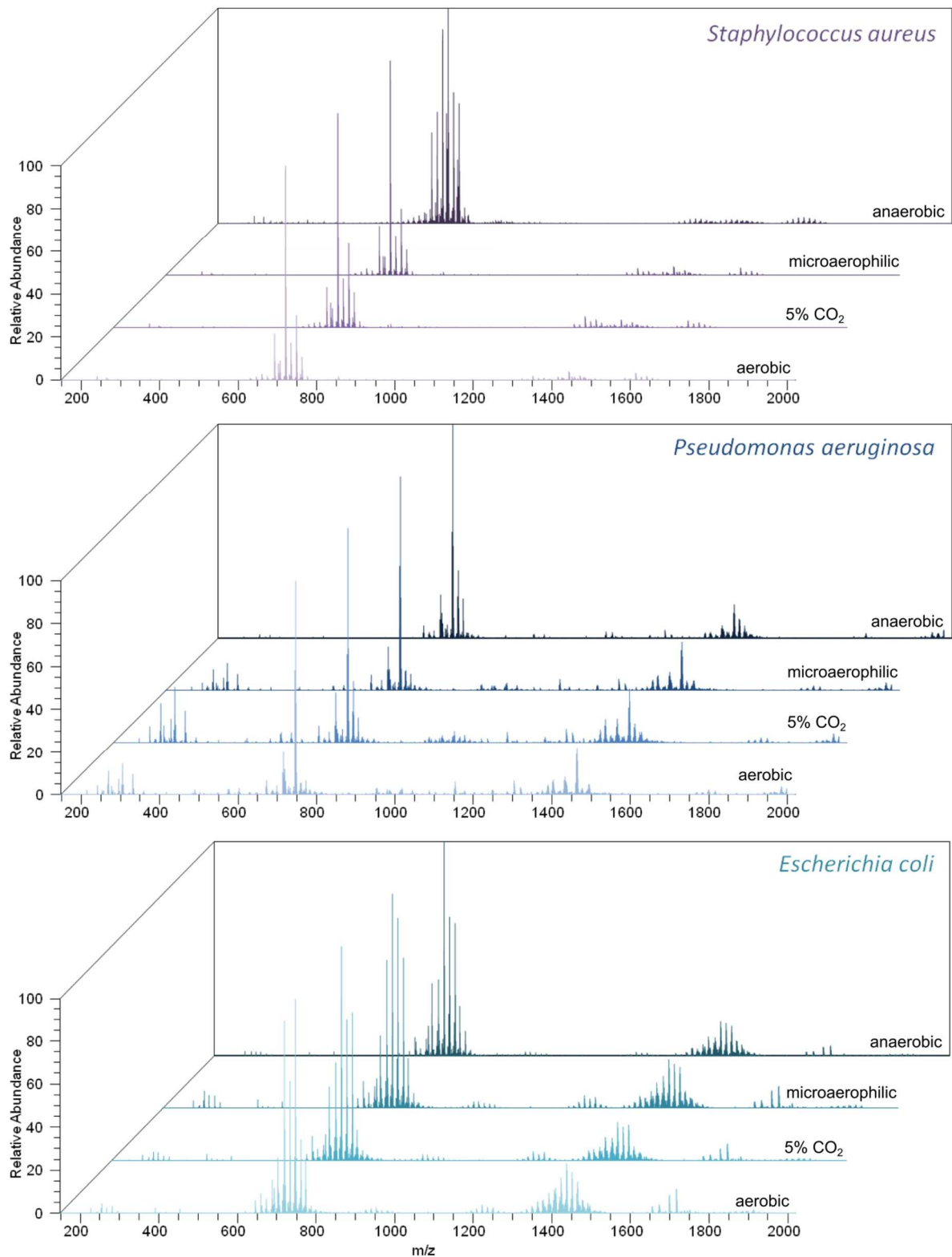


Figure 45. REIMS spectral profiles obtained for *Staphylococcus aureus* ATCC 25923, *Pseudomonas aeruginosa* ATCC 27853 and *Escherichia coli* ATCC 25922, each grown on Columbia blood agar for 48hrs at 37°C under different atmospheres. From front to back: aerobic, aerobic containing 5% CO₂, microaerophilic, anaerobic.

However, in all cases it is apparent that large parts of the spectral information remains conserved, especially for masses above m/z 1100. This highly conserved nature of signals at $m/z >1000$ has been largely ignored by other lipid based studies, which predominantly focused on the mass range below m/z 1000. This might be due to difficulties in detection of those higher mass compounds for technologies such as DESI and a limited knowledge on the chemical nature of these heavy bacterial lipid species.

Similar spectral behaviour was observed if a single bacterial strain was cultured on five different commonly used nutrient-rich solid growth media (see Figure 44). In this case, very similar mass spectral profiles were obtained within all three species, with very limited variance for *S. aureus*, different quorum-sensing expression profiles for *P. aeruginosa* as well as slight changes for the lipid species at m/z 900-1100, while in case of *E. coli* slight variations of the major phosphatidylglycerols species were observed. However, none of the organisms is fastidious and all growth media are rich in nutrients providing the bacterium with all necessary nutrients to achieve optimum membrane lipid composition. Changes in production of extracellular metabolites in case of *P. aeruginosa* was additionally observed for different colony morphologies. Single colonies of *P. aeruginosa* generally did not display large amounts of quorum-sensing molecules and rhamnolipids while the same strain grown as a lawn would greatly exhibit signals originating from these molecules. This is tentatively associated with production of both types of molecules being cell density-regulated.(160) This highlights the need for standardised culturing and sampling conditions if studies are aimed at the differences in expression of extracellular metabolites in correlation with proteomic or genomic information.

All three bacterial species are facultative anaerobes and can thus survive and grow under a variety of different atmospheric compositions. Figure 45 shows the REIMS profiles obtained after culturing the bacteria under the four most commonly used atmospheric compositions. *P. aeruginosa* and *E. coli* show similar mass spectral features when cultured under all four different conditions. However, significant changes were observed in the spectral profile of *S. aureus* grown under anaerobic conditions with a strong shift of spectral features towards higher molecular weight species, both for the low mass range of m/z 650-800 as for higher mass species in the range of m/z 1300-1700. This effect is more clearly visualised in the zoomed mass spectral regions in Figure 46. This can be explained by results obtained in a study analysing the fatty acid composition of *S. aureus* grown under aerobic and anaerobic conditions using gas chromatography.

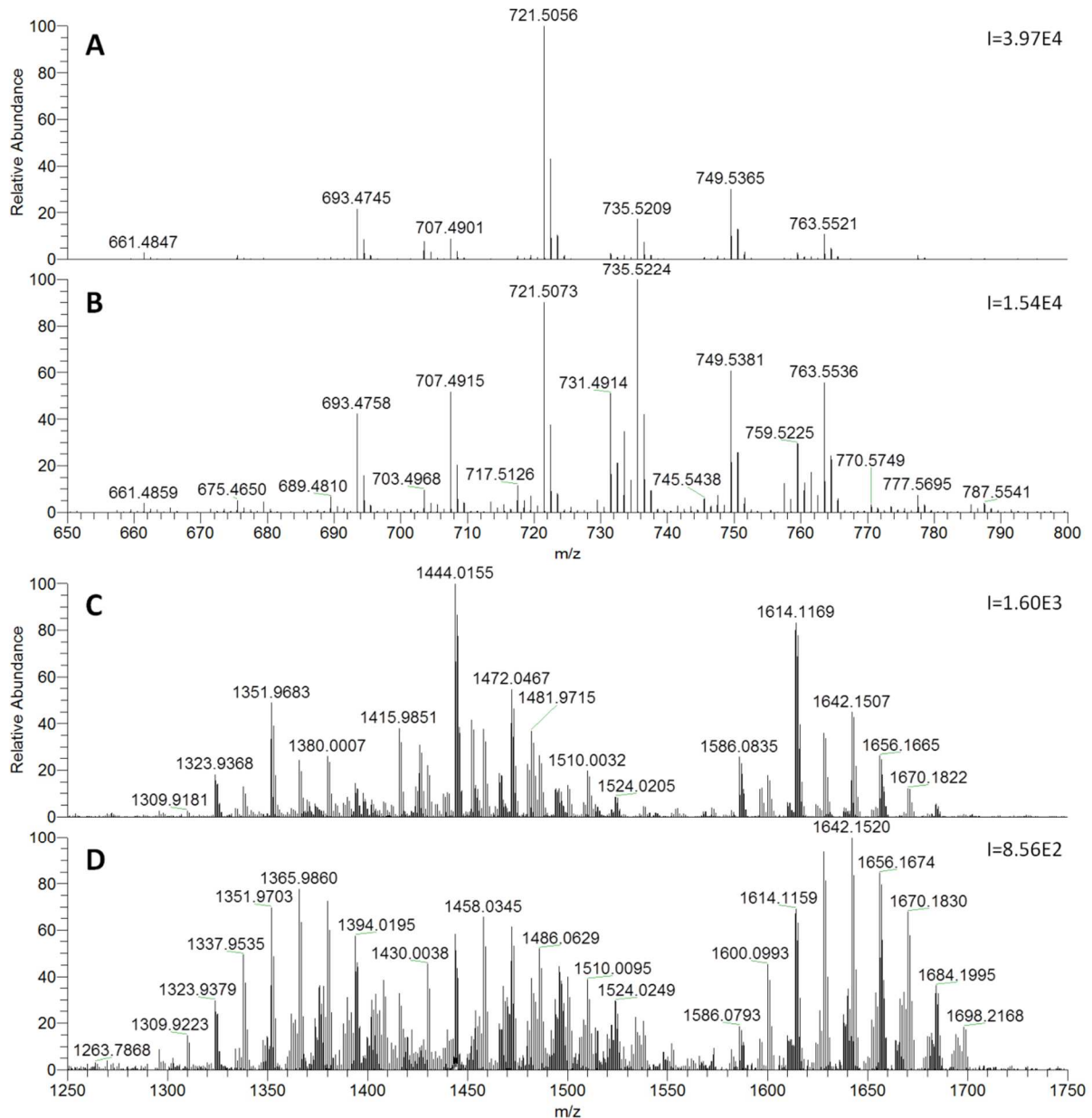


Figure 46. Regions of mass spectra of *S. aureus* grown under aerobic and anaerobic conditions. A) m/z 650-800, aerobic, B) m/z 650-800, anaerobic, C) m/z 1250-1750, aerobic, and D) m/z 1250-1750, anaerobic. I=Intensity value at 100% relative abundance.

While under aerobic conditions, fatty acids with 15, 18, and 20 carbon atoms account for 40.86%, 3.7%, and 21.84% of the overall fatty acid content, under anaerobic conditions these numbers change to 16.26%, 22.38% and 37.65%, respectively.(161) This was associated with the membrane-bound electron transport system being inactive under anaerobic conditions. As *S. aureus* was found to contain less than 1% free fatty acids, these fatty acids are predominantly built into membrane lipid species and thus lead to a significant increase in heavier phospholipid species which can also be observed using REIMS, both for

phosphatidylglycerol species in the mass range of m/z 650-800, and for cardiolipins and other heavy molecular weight species in the mass range of m/z 1250-1750 (see Figure 46).

3.7. Analysis of liquid cultures

For many applications in research, diagnostics or biotechnology, liquid cultures are preferred over solid cultures due to their higher degree of homogeneity regarding concentration of nutrients and cell density. In addition, liquid culturing conditions resemble more the conditions bacteria experience in the human body. The bipolar setup with incorporated aspiration port as used for solid cultures is little suited for the analysis of liquid cultures. Simple immersion of the electrodes into the liquid medium leads to aspiration of liquid culturing medium into the mass spectrometer which in turn leads to impaired vacuum conditions and instrument contamination as the aspiration port is too closely situated to the electrode tips. An external aspiration port could also be used to circumvent the aspiration of liquid into the instrument; however, limitations to the efficiency of the aerosolisation process prevail as the aerosolisation temperature is limited due to the excess of water in the sample. This, together with dilution of cells by the culturing medium considerably limits sensitivity. However, the technique could be adapted to the analysis of liquid cultures by spinning the microbial liquid cultures in a centrifuge and discarding the supernatant.

For most bacterial species, a centrifugation step of 10mins at 3500g is sufficient to form a cell pellet suitable for REIMS analysis. However, certain bacterial species such as *Klebsiella pneumoniae* or *Pseudomonas aeruginosa* need to be centrifuged under more rigorous conditions such as 10mins at 12500g to allow effective removal of the supernatant. This was tentatively associated with the mucoid nature of these cultures. The remaining pellet after centrifugation can directly be analysed by taking up microbial biomass from the pellet between the electrodes of the sampling probe as shown in Figure 9. If the obtained microbial pellet contains only small amounts of biomass or excess liquid, better sensitivity can be obtained if the microbial biomass is instead analysed from a solid support material such as the tip of a cotton swab. To achieve this, some of the microbial biomass is picked up from the cell pellet using the swab and parts of the wetted swab were squeezed between the electrodes of the sampling probe for REIMS analysis. As Figure 47 shows, largely identical spectral profiles were obtained using either direct analysis or REIMS analysis aided by a solid support. No spectral features originating from the cellulose of the cotton were observed although the cotton underwent significant carbonisation. This effect is similar to the one

observed for the agar matrix, which was reasoned to undergo extensive charring without producing any mass spectral signals in the studied mass range. Since both cellulose and agarose are linear polysaccharides, this effect is likely to be of similar origin.

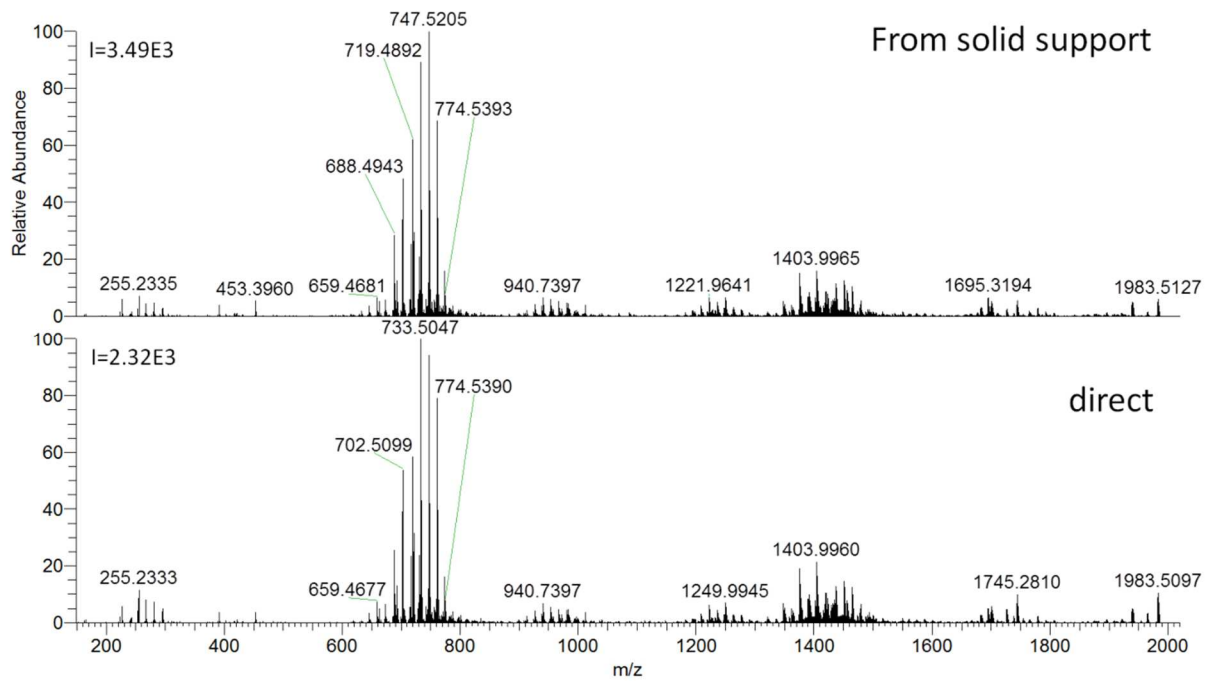


Figure 47. REIMS spectral profiles obtained for clinical isolate of *Proteus mirabilis* with and without solid support. Spectra obtained from ionisation from the bipolar forceps electrodes (bottom, direct) and analysing samples from the surface of a cotton swab (top, from solid support). I=Intensity value at 100% relative abundance.

4. Analysis of microorganisms

4.1. Building of REIMS spectral database

A large scale spectral database was built with the intention to produce a spectral library comprising entries for the 100 most common pathogenic species and with an overall size of approx. 5,000 entries. Spectral acquisition was distributed over several instruments in order to be able to test robustness and reproducibility of the proposed methodology more comprehensively. All strains were analysed using settings as described in Materials and Methods - REIMS analysis of microorganisms. An overview of the numbers of strains acquired on each instrumental platform is shown in Table 5. The largest database for testing specificity on different phylogenetic levels was based on data recorded on the Thermo Exactive instrument and eventually comprised 3,308 entries of 263 bacterial and fungal species belonging to 95 genera (as determined on 14.01.2015) at the end of the project. A comprehensive list of individual species of bacteria and yeast and respective number of entries that were part of this database are given in Appendix 2 and Appendix 3. The database comprised 43 reference and type strains that were either sourced from the quality control organisms at the clinical microbiology laboratory (Charing Cross Hospital, London, UK), through a collaboration with Dr Anne McCartney (Food and Nutritional Sciences, University of Reading, UK) and Dr Lesley Hoyles (University of Westminster and Imperial College London), or cultured from Culti-loops (Oxoid, Basingstoke, UK) or Vitroids (Sigma Aldrich, St Louis, MO, USA). The vast majority of other strains featured in the database were obtained from the routine work of the clinical microbiology laboratory and were identified using the MALDI Biotyper instrument. The majority of yeasts (210 strains) were sourced from an internal study of the Clinical Microbiology Laboratory that assessed the robustness of the MALDI Biotyper for yeast speciation and were confirmed using pyrosequencing if discrepancies were found between MALDI and conventional identification results.

Spectral database entries for an overall number of 221 bacterial species were recorded on the Exactive instrumental platform, however, only a small proportion of these database entries can be used for specificity assessment using multivariate statistical models such as PCA and LDA as these methods require multiple datapoints per data group or class. 77 bacterial species were represented by a single database entry only while 154 species were represented by <5 database entries. Overall, 55 (24.9%) bacterial species were represented by >10

database entries. Thus, only the most common pathogenic species that were represented with sufficient amount of database entries were used for specificity assessment by multivariate statistical methods.

Table 5. Size of spectral database acquired as part of the presented study as determined on 14.01.2015

	Instrument used for analysis			
	Thermo Exactive	Thermo Orbitrap Discovery XL	Waters Xevo G2-S	Waters Synapt G2
Bacteria				
no of strains	3054	704	662	252
no of genera	85	23	42	11
no of species	221	39	90	13
Yeasts				
no of strains	254	5	76	0
no of genera	10	2	5	0
no of species	42	3	17	0
Total no of strains (database entries):		5,007		

4.2. Characterisation of REIMS spectral content

The number of mass spectrometric features observed using REIMS was determined for 28 of the most commonly detected bacterial pathogens. For this purpose, REIMS spectra were peak picked as described in Materials and Methods - Data analysis - Determination of the number of peaks. No spectral deconvolution was performed as all signals detected were singly charged. Isotope signals were not removed to keep the nature of a profiling mass spectrometric method. Results are shown in Table 6. The smallest number of m/z features was detected in case of *E. coli* (434 features) while the highest number of features detected in *K. oxytoca* (1590 features). This leads to a range of approximately 400-1600 m/z signals detected in REIMS measurements with the average amount of m/z signals being around 900. Within the *Staphylococcus* genus, spectral appearance is comparatively similar, however, the amount of m/z features detected varies between 783 for *S. capitis* and 1481 for *S. haemolyticus*. This suggests differences in number of features detected to be correlated with general quality of the REIMS measurement and signal to noise ratios. While comparatively small number of high abundance mass spectrometric signals (400-500 features) determine the overall appearance, approximately 1000 low abundance signals can be detected if favourable signal to noise levels are obtained. This large variance in the number of detected m/z signals

might pose a difficulty for data analysis using multivariate statistical methods as it might lead to increased intra-species variance.

Table 6. Number of spectral features detected for 28 of the most common clinical pathogens.

Bacterial Species	Number of m/z signals
<i>Bacteroides fragilis</i>	1076
<i>Burkholderia cepacia complex</i>	871
<i>Citrobacter koseri</i>	897
<i>Clostridium difficile</i>	545
<i>Enterobacter cloacae</i>	1156
<i>Enterococcus faecalis</i>	994
<i>Enterococcus faecium</i>	1196
<i>Escherichia coli</i>	434
<i>Haemophilus influenzae</i>	535
<i>Klebsiella oxytoca</i>	1590
<i>Klebsiella pneumoniae</i>	966
<i>Micrococcus luteus</i>	587
<i>Moraxella catarrhalis</i>	677
<i>Morganella morganii</i>	408
<i>Neisseria gonorrhoeae</i>	1141
<i>Proteus mirabilis</i>	817
<i>Pseudomonas aeruginosa</i>	451
<i>Serratia marcescens</i>	1148
<i>Staphylococcus aureus</i>	1047
<i>Staphylococcus capitis</i>	783
<i>Staphylococcus epidermidis</i>	1337
<i>Staphylococcus haemolyticus</i>	1481
<i>Staphylococcus hominis</i>	1059
<i>Stenotrophomonas maltophilia</i>	1108
<i>Streptococcus agalactiae</i>	709
<i>Streptococcus pneumoniae</i>	772
<i>Streptococcus pyogenes</i>	851

The REIMS ionisation mechanism strongly favours the detection of compounds undergoing ionic dissociation in aqueous media. Most signals that have previously been reported using REIMS were limited to the mass range of $m/$ 600-900 and were originating from phosphatidic acids (PAs), phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), phosphatidylinositols (PIs), phosphatidylserines (PSs), sphingomyelins (SMs), phosphatidylcholines (PCs), diacylglycerols (DAG-H₂O) and triacylglycerols (TAGs).(128,159) These compounds were found to originate mostly from the lipid bilayer of the cell membranes. The detection of polar lipids is highly favoured compared to other compounds present in the sample, similarly to other ionisation methods. This is associated with the combination of hydrophobic acyl-chains and ionic head groups making desorption ionisation energetically favourable.

The chemical nature of the vast majority of REIMS signals was found to be deprotonated, quasi-molecular $[M-H]^-$ species. However, $[M+Cl]^-$ adducts and in some cases $[M-H_2O]^-$ for phosphatidylglycerol species were also observed. $[M+Cl]^-$ adducts were frequently observed for sphingolipid species where the $[M-H]^-$ signal was negligible. Water loss was observed in case of high abundance phosphatidylglycerol species, as observed in *Staphylococcus* species. However, these features were detected with low spectral intensity as compared to the intact $[M-H]^-$ species (<5% relative abundance as compared to parent ion).

4.2.1. Bacterial glycerophospholipids detected using REIMS

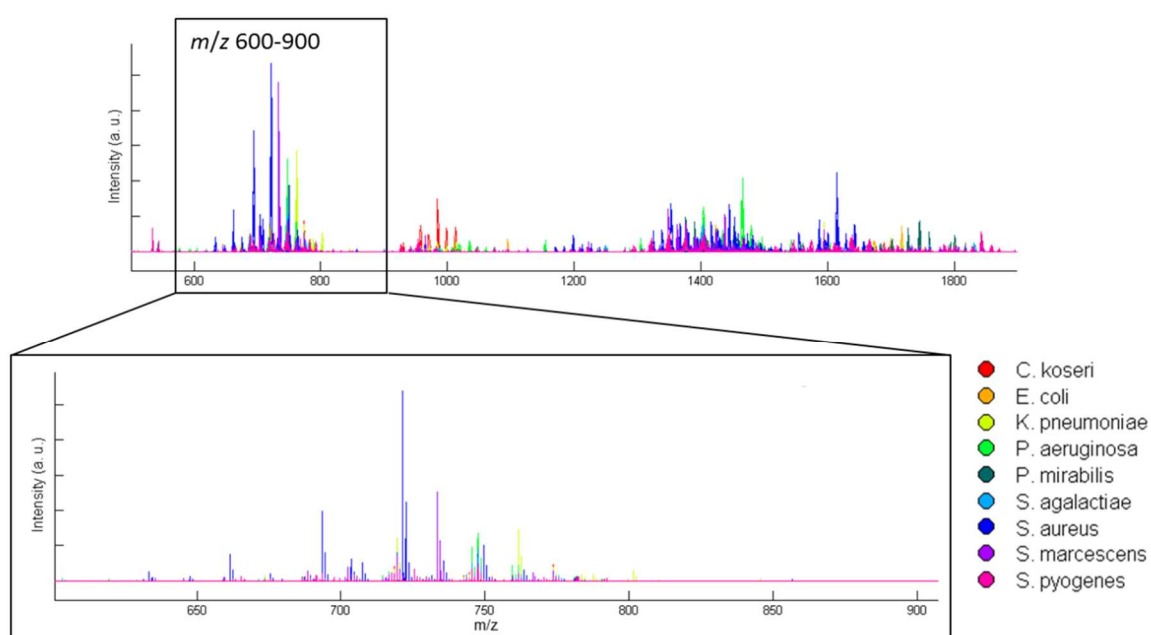


Figure 48. Overlaid REIMS spectral profiles as obtained for nine different bacterial species for m/z 500-1900 and m/z 600-900 in zoom. Adapted from Ref. (133) with permission from The Royal Society of Chemistry.

The majority of spectral profiles obtained from biological samples using REIMS are dominated by intact phospholipid species as shown in Figure 48 that can be ascribed to PAs, PEs and PGs. This can clearly be seen in Table 7, which shows the qualitative phospholipid distribution as obtained for nine different bacterial pathogenic species using exact mass measurements and tandem mass spectrometric data acquired during REIMS experiments.(133) Only high abundance signals (>5% relative abundance) were included. Distinct peak patterns can be obtained for all bacterial species, even for those that are closely related such as different *Streptococcus* spp. or members of the Enterobacteriaceae family (*E. coli*, *C. koseri*, *K. pneumoniae*, *S. marcescens*, *P. mirabilis*). Most spectral patterns for both

Gram-negative and Gram-positive species are dominated by high abundance deprotonated phosphatidylglycerol (PG) signals.

Table 7. Identified phospholipids detected in the mass range m/z 600-900 for all analysed species. Only phospholipids with relative abundances >5% and only the most abundant acyl chain combination were included. Solid growth media on which bacteria were grown is given in parentheses. ID based solely on exact mass when lipid composition given as sum carbon number rather than individual acyl chains. Reproduced from Ref. (133) with permission from The Royal Society of Chemistry.

Nominal mass m/z	<i>C. koseri</i> (CBA)	<i>E. coli</i> (CBA)	<i>K. pneumoniae</i> (LB)	<i>P. mirabilis</i> (MCC)	<i>P. aeruginosa</i> (LB)	<i>S. marascens</i> (MCC)	<i>S. aureus</i> (CBA)	<i>S. agalactiae</i> (CBA)	<i>S. pyogenes</i> (CBA)
645									PA(32:1)*
659			PA(16:0/17:1)	PA(16:0/17:1)		PA(16:0/17:1)			
661							PA(33:0)*		
665									PG(12:0/16:0)
671									PA(34:2)*
673				PA(16:0/18:1)	PA(16:0/18:1)				PA(16:0/18:1)*
675							PG(15:0/15:0-H ₂ O)		PG(30:0-H ₂ O)*
688	PE(16:1/16:0)			PE(16:1/16:0)					
691									PG(14:0/16:1)
693	PG(16:0/14:0)		PG(16:0/14:0)				PG(15:0/15:0)	PG(15:0/15:0)	PG(14:0/16:0)
697									PA(36:3)*
699									PA(18:1/18:1)*
701								PG(32:1)-H ₂ O*	PG(32:1)-H ₂ O*
702	PE(16:0/17:1)	PE(16:0/17:1)	PE(16:0/17:1)	PE(16:0/17:1)		PE(16:0/17:1)			
707							PG(15:0/16:0)		
716	PE(18:1/16:0)			PE(18:1/16:0)	PE(18:1/16:0)	PE(17:0/17:1)			
717								PG(32:2)*	PG(16:1/16:1)
719	PG(16:1/16:0)	PG(16:1/16:0)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)	PG(16:0/16:1)
721							PG(15:0/17:0)	PG(15:0/17:0)	PG(16:0/16:0)
725									PA(16:1/18:2)
727									PG(16:1/18:1)-H ₂ O
729								PG(16:0/18:1)-H ₂ O*	PG(16:0/18:1)-H ₂ O
730				PE(16:0/19:1)					
733	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)	PG(16:0/17:1)			
735							PG(15:0/18:0)		
743								PG(16:0/18:3)	PG(16:1/18:2)
745	PG(16:1/18:1)	PG(16:1/18:1)	PG(16:1/18:1)		PG(16:1/18:1)	PG(16:1/18:1)	PG(16:0/18:2)*	PG(16:0/18:2)*	PG(16:1/18:1)
747	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)	PG(16:0/18:1)
749							PG(15:0/19:0)	PG(15:0/19:0)	PG(16:0/18:1)*
752									
759		PG(17:1/18:1)	PG(17:1/18:1)		PG(17:1/18:1)	PG(17:1/18:1)			
761		PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)	PG(16:0/19:1)			
763							PG(15:0/20:0)		
770									PE(38:2)*
771								PG(36:3)*	PG(18:1/18:1)*
773	PG(18:1/18:1)	PG(18:1/18:1)	PG(17:1/19:1)		PG(17:1/19:1)	PG(18:1/18:1)	PG(36:2)*	PG(36:2)*	PG(18:1/18:1)
775							PG(36:1)*	PG(36:1)*	PG(18:0/18:1)
787			PG(18:1/19:1)						
801			PG(19:1/19:1)						

* Signal intensity not sufficient to obtain meaningful MS/MS data; Abbreviations: PG = phosphatidylglycerol, PE = phosphatidylethanolamine, CBA = Columbia blood agar, LB = Luria-Bertani agar, MCC = McConkey agar.

Generally, Gram-negative species display a higher percentage of unsaturated phospholipid species and a higher relative amount of PEs. This is in good agreement with literature on the bacterial phospholipid composition.(1) *Staphylococcus aureus* (and other *Staphylococcus*

spp.) are clearly distinguished from other bacterial species by the fact that they exclusively show signals arising from saturated phospholipid species.

These differences in phospholipid composition are mirrored in the multivariate statistical plots as shown in Figure 49.(133) Gram-positive and Gram-negative bacterial species are separated from each other along the first principal component. *S. aureus* is further separated from other Gram-positive species by their markedly different phospholipid profile.

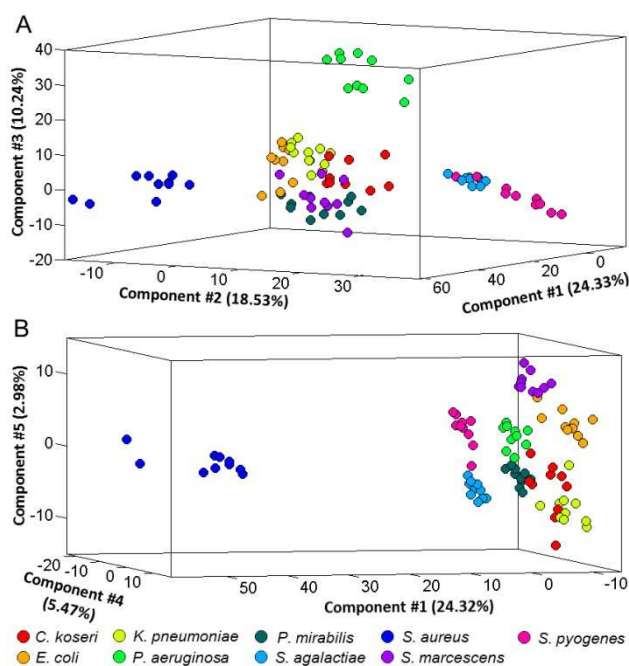


Figure 49. A) PCA plot, and B) RMMC plot of the processed dataset, mass range m/z 500-1,900.

Reproduced from Ref. (133) with permission from The Royal Society of Chemistry.

Phospholipid composition of *E. coli* cells in their exponential growth phase were reported to consist of 75% phosphatidylethanolamines (PEs), 20% phosphatidylglycerols (PGs), and 5% cardiolipins (CLs).(162) Although identical phospholipid species regarding class and carbon chain length of lipids were detected using REIMS, the relative abundances of the corresponding signals do not reflect the amounts detected using chromatographic methods. While based on Ref. (162), PEs would be expected to be the dominant lipid class in REIMS spectra of *E. coli*, in practice phosphatidylglycerol and cardiolipins prove to be the most abundant spectral features for the vast majority of bacteria as it is shown in Figure 48 and Table 7 for nine common bacterial pathogens including *E. coli*. This data suggests a favourable ionisation behaviour for phosphatidylglycerols as compared to phosphatidylethanolamines.

In addition to regular glycerophospholipid species, high spectral abundance PE and PG plasmalogen species have also been detected in REIMS spectra of anaerobic bacteria such as *Clostridium* spp., *Fusobacterium* spp., and *Veillonella* spp.. While plasmalogens are major constituents of the cell wall of many anaerobic bacteria, they are absent in aerobic or facultatively anaerobic bacteria. They are also rarely found in fungi although they are widely distributed among mammalian species.(163)

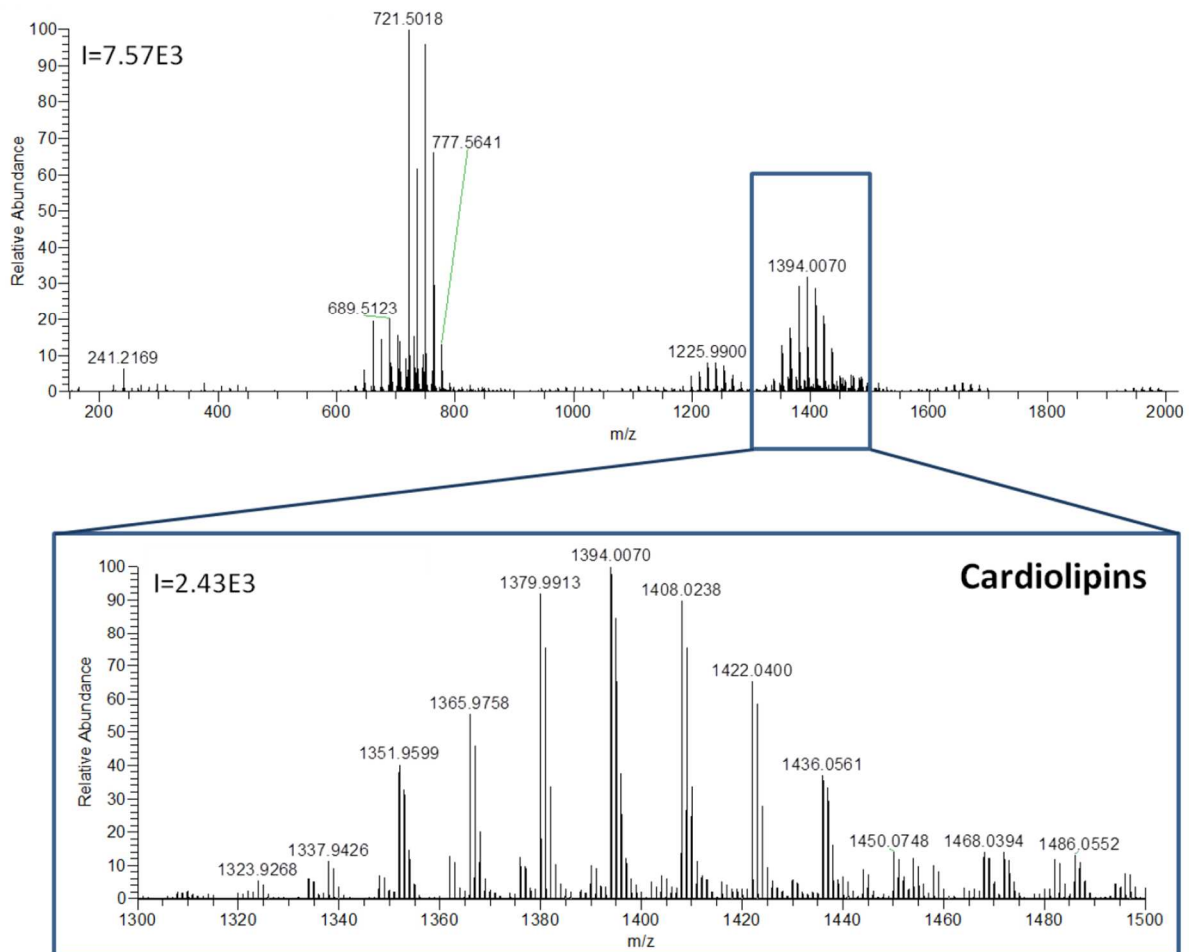


Figure 50. Full REIMS spectrum of *Staphylococcus epidermidis* ATCC 1228 cultured for 48hrs on BHI agar. Zoomed region shows cardiolipin region. Asterixes indicate those species that were confirmed by MS/MS measurements. I=Intensity at 100% relative abundance.

Cardiolipins (1,3-bis(sn-3'-phosphatidyl)-sn-glycerols) are complex diphosphatidylglycerol lipids containing four fatty acid chains that can differ in length and degree of unsaturation. In bacteria, they are predominantly found in the plasma membranes and although there is significant potential for complexity in cardiolipin structure considering the presence of four fatty acyl chains, the cardiolipin profiles are generally found to be simple and reproducible.(164) Bacterial cardiolipins predominantly exhibit shorter carbon chain lengths

with mostly saturated or mono-unsaturated fatty acids. In contrast, eukaryotic cells feature predominantly longer chain polyunsaturated fatty acids as building blocks of their cardiolipins. Using REIMS, cardiolipins were detected and identified in a wide range of both Gram-positive and Gram-negative bacteria, based on exact mass measurements. They were typically detected in the mass range between m/z 1300-1450 which corresponds to overall chain lengths of CL(60:0)-CL(72:0).

Figure 50 shows the overall spectral profile and the cardiolipins region in particular for *Staphylococcus epidermidis* ATCC 12228. For this strain, the presence of cardiolipins was further confirmed using tandem mass spectrometric fragmentation patterns. For this purpose, the strain was cultured on BHI medium which was found to increase the relative spectral intensity for cardiolipins as compared to horse blood agar. Representative fragmentation spectra are shown in Figure 51. Detected fragmentation patterns are in good agreement with those reported for mammalian cardiolipins of higher molecular weight.(165)

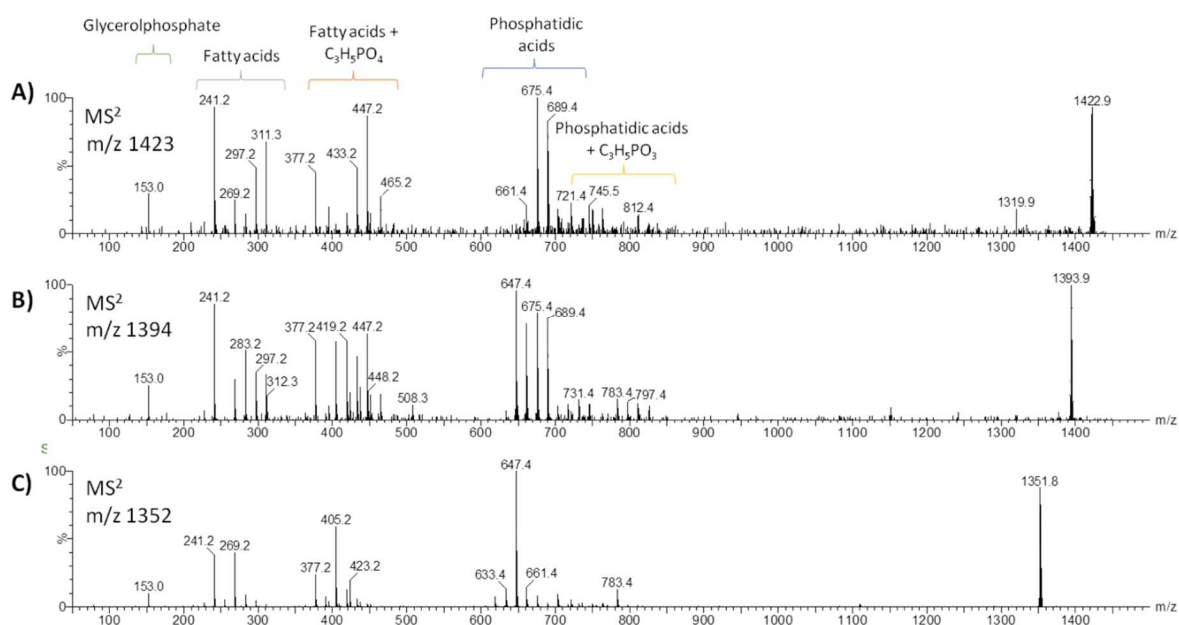


Figure 51. Fragmentation spectra for cardiolipins at A) m/z 1352, B) m/z 1394 and C) m/z 1423 from *Staphylococcus epidermidis* ATCC12228. Peak identifications as listed in Table 8.

In good agreement with the acylation pattern of phosphatidylglycerol species, (see Table 7), only cardiolipins with saturated fatty acyl groups were detected in case of the *Staphylococcus* genus (see Table 8). This is expected as phosphatidylglycerols moieties form the building blocks of cardiolipins.

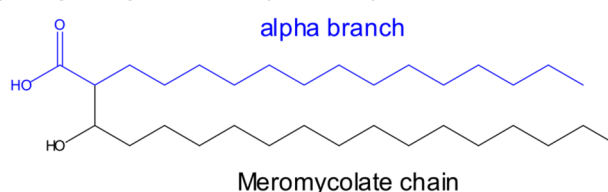
Table 8. Cardiolipin species that were identified for *Staphylococcus epidermidis* ATCC 12228.

Compound	Sum formula	Exact mass [M-H] ⁻	Exp. mass	Mass Deviation
CL(62:0)	C ₇₁ H ₁₃₈ O ₁₇ P ₂	1323.9335	1323.9268	5.0 ppm
CL(63:0)	C ₇₂ H ₁₄₀ O ₁₇ P ₂	1337.9492	1337.9426	4.9 ppm
CL(64:0)	C ₇₃ H ₁₄₂ O ₁₇ P ₂	1351.9649	1351.9601	3.6 ppm
CL(65:0)	C ₇₄ H ₁₄₄ O ₁₇ P ₂	1365.9806	1365.9758	3.5 ppm
CL(66:0)	C ₇₅ H ₁₄₆ O ₁₇ P ₂	1379.9962	1379.9913	3.5 ppm
CL(67:0)	C ₇₆ H ₁₄₈ O ₁₇ P ₂	1394.0119	1394.0070	3.5 ppm
CL(68:0)	C ₇₇ H ₁₅₀ O ₁₇ P ₂	1408.0275	1408.0238	2.6 ppm
CL(69:0)	C ₇₈ H ₁₅₂ O ₁₇ P ₂	1422.0432	1422.0400	2.3 ppm
CL(70:0)	C ₇₉ H ₁₅₄ O ₁₇ P ₂	1436.0588	1436.0561	1.9 ppm
CL(71:0)	C ₈₀ H ₁₅₆ O ₁₇ P ₂	1450.0745	1450.0748	0.2 ppm
CL(72:0)	C ₈₁ H ₁₅₈ O ₁₇ P ₂	1464.0900	1464.0970	4.8 ppm

4.2.2. Mycolic acids in *Corynebacterinaceae* suborder

The suborder *Corynebacterineae* forms a large group of actinomycete species characterised by the presence of mycolic acids. Mycolic acids are complex lipids composed of a longer beta-hydroxy chain (meromycolate chain) and shorter alpha-alkyl side chain (α -branch, see Figure 52). The variability of their chain lengths and the complexity of their structures contribute to the definition of the genera, from the simplest corynomycolic acids of *Corynebacterium* to the most complex mycolic acids of the *Mycobacterium* genus.(166) The number of carbon atoms and the degree of desaturation (refers to the number of double bonds and/or cyclopropane rings) of main and side chains vary according to the taxonomical classification of bacteria.

A) Corynomycolic acid (m/z 495)



B) Fragmentation pathway

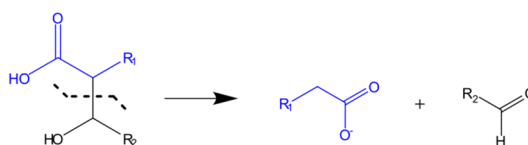


Figure 52. A) Structure of Corynomycolic acid, a short-chain mycolic acid found in *Corynebacterium* and *Rhodococcus* spp., B) Collision-induced dissociation fragmentation pathway observed in negative ion mode. Fragmentation by neutral loss of meromycolate chain as aldehyde.

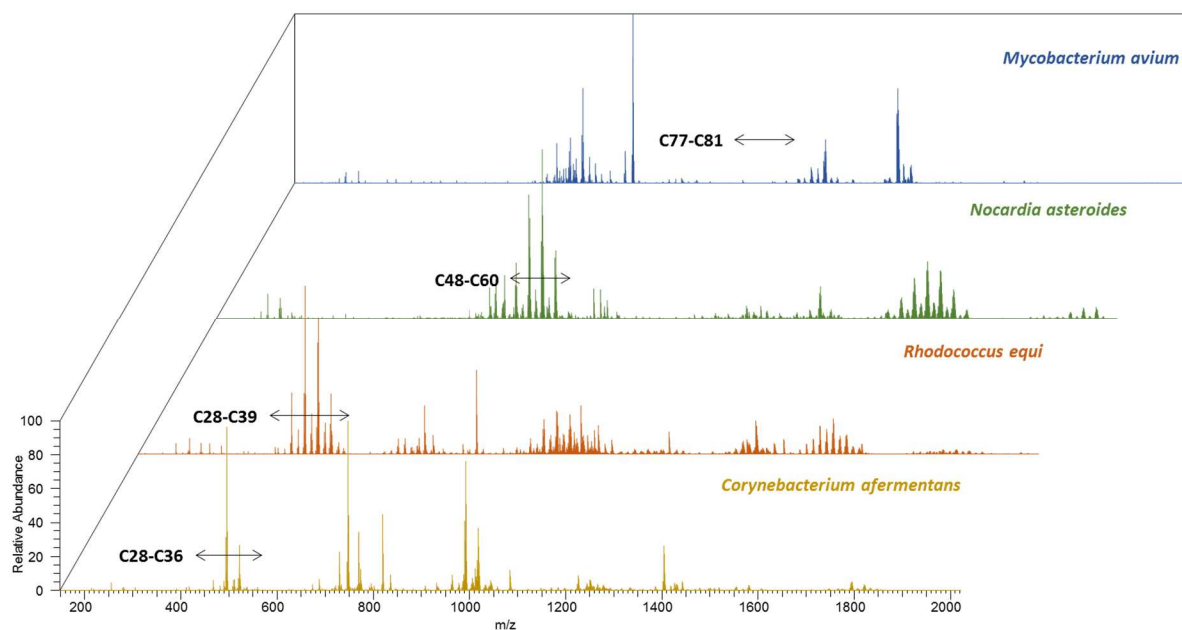


Figure 53. REIMS spectral profiles for *Corynebacterium afermentans* (clinical isolate), *Rhodococcus equi* (ATCC 6939), *Nocardia asteroides* (ATCC 19247) and *Mycobacterium avium* (clinical isolate).

Figure 53 displays REIMS spectral profiles that were obtained for members of the four most common genera of the Corynebacterineae suborder as encountered in clinical microbiology. In their REIMS spectra, *Corynebacterium* spp. display mycolic acids of chain lengths ranging from C28 to C36, whereas C28 to C39, C48 to C60 and C77 to C81 were found in *Rhodococcus* spp., *Nocardia* spp., and *Mycobacterium* spp., respectively.

4.2.2.1. *Corynebacterium* spp.

Species identification within the *Corynebacterium* genus can be challenging using existing methods. Partial 16S rRNA gene sequencing does not lead to sufficient specificity (ideally the less commonly used full sequence is needed), the API Coryne system (bioMerieux, Durham, NC) takes 24hrs and may yield unreliable results and they further have very similar protein patterns in the Biotyper database.(167,168)

Corynebacteria contain mycolic acids with the shortest overall chain lengths of 22-38 carbon atoms. Mycolic acids in *Corynebacterium* spp. are composed of an even number of carbon atoms with the degree of desaturations reported to be between 0 and 2.(169) Results obtained using REIMS are in good agreement with these findings with only even carbon numbered mycolic acids being present in the REIMS spectra showing the expected range of unsaturations (see Table 9).

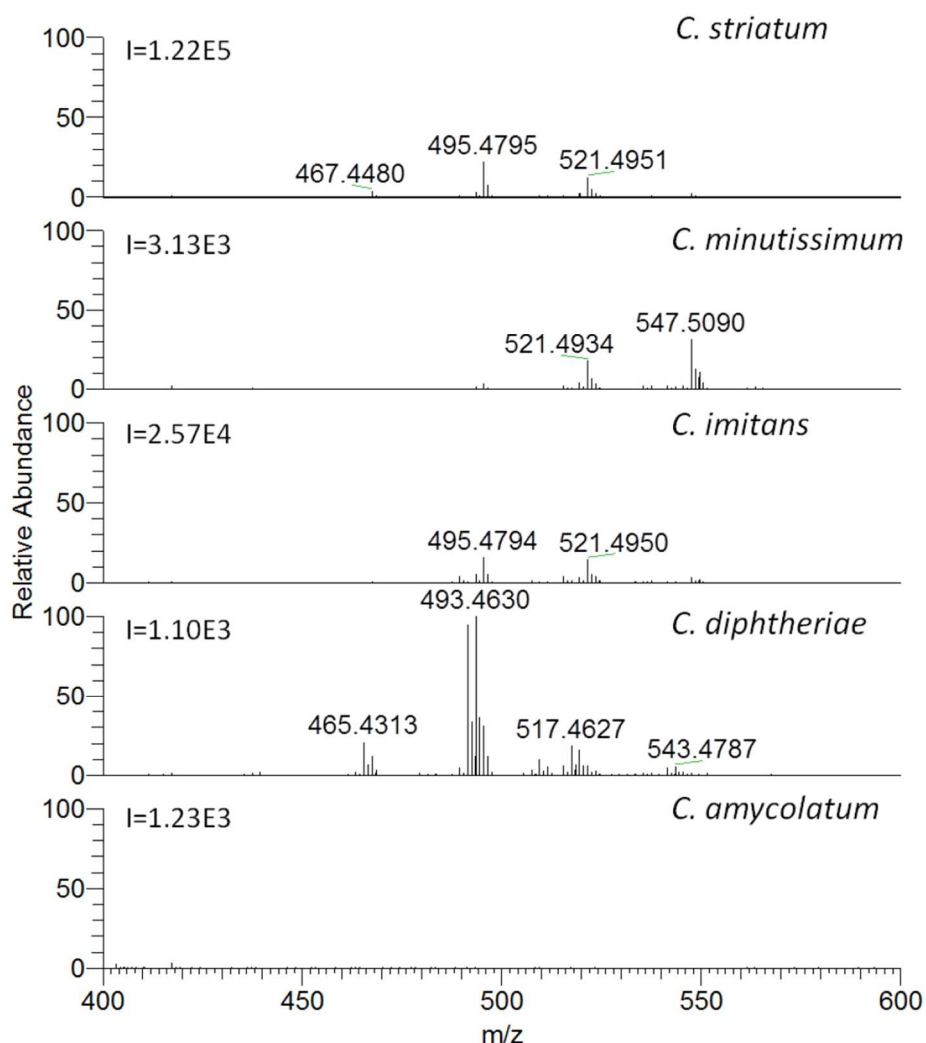


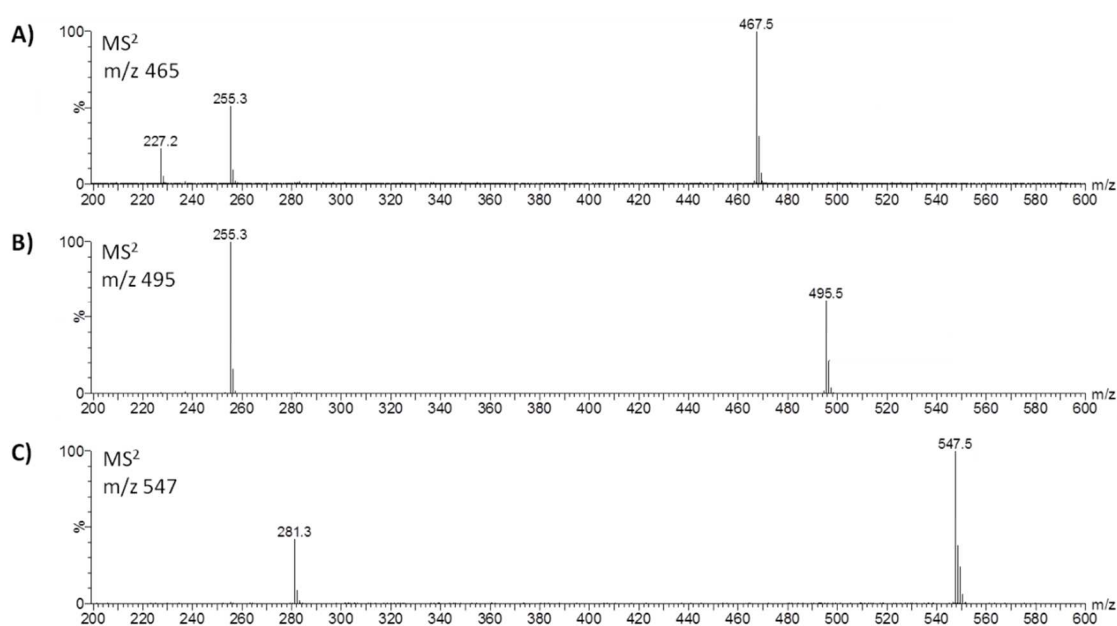
Figure 54. Zoom into REIMS profile for five different *Corynebacterium* species reveals different mycolic acid patterns; spectra normalised to base peak. I=Intensity at 100% relative abundance.

De Briel *et al.* proposed a HPLC-based mycolic acid profiling method for speciation within the *Corynebacterium* genus.(170) Distinctively different mycolic acid patterns showing good correlation with taxonomic classification were obtained also using REIMS (see Figure 54). *Corynebacterium amycolatum*, an emerging opportunistic pathogen, can be easily distinguished by its absence of mycolic acid signals. Mycolic acid patterns obtained using REIMS might therefore offer a promising approach to identify *Corynebacterium* species.

Table 9. Identified mycolic acids as detected in different *Corynebacterium* species.

Compound	Sum formula	Exact mass [M-H] ⁻	Exp. mass	Mass Deviation
Mycolic acid C28:0	C ₂₈ H ₅₅ O ₃	439.415669	439.4159	0.5 ppm
Mycolic acid C30:0	C ₃₀ H ₅₉ O ₃	467.446969	467.4473	0.7 ppm
Mycolic acid C32:1	C ₃₂ H ₆₁ O ₃	493.462619	493.4634	1.6 ppm
Mycolic acid C32:0	C ₃₂ H ₆₃ O ₃	495.478269	495.4786	0.7 ppm
Mycolic acid C34:2	C ₃₄ H ₆₃ O ₃	519.478269	519.4788	1.0 ppm
Mycolic acid C34:1	C ₃₄ H ₆₅ O ₃	521.493919	521.4942	0.5 ppm
Mycolic acid C36:2	C ₃₆ H ₆₇ O ₃	547.509569	547.5102	1.2 ppm

Using collision induced dissociation, mycolic acids fragment at the bond between the meromycolate chain and the α -branch (see Figure 52B). While the meromycolate chain forms an aldehyde and leaves the parent ion as a neutral entity, the α -branch of the mycolic acid forms a negatively charged carboxylate ion, allowing the structural assignment of the parent ion.(171) Fragment ions observed (see Table 9) were in good agreement with those reported in literature for *C. glutamicum* and confirmed chemical assignment.(171) All detected mycolic acids could be assigned to alpha- (or alpha'-) mycolic acids based on exact mass measurements. However, no further assignments were made than overall number of carbons as mycolic acids were frequently observed to be mixtures of several isobaric compounds and fatty acid composition derived from a single strain is likely insufficient to reflect composition within the whole genus.

**Figure 55. Tandem mass spectra obtained for an isolate of *Corynebacterium striatum*, A) m/z 467, B) m/z 495 and C) m/z 547. Signals were attributed to mycolic acids as listed in Table 9.**

4.2.2.2. *Rhodococcus* spp.

Mycolic acids constituents were determined for *Rhodococcus equi* ATCC 6939. According to Nishiuchi *et al.*,(172) *Rhodococcus* spp. contain mycolic acids between 30 and 54 carbon chain lengths and 0-2 unsaturations. For the analysed *Rhodococcus* strain, mycolic acids with overall chain length of 28-39 were found with 0-3 unsaturations. The overall chain length for this particular strain is in good agreement for *R. equi* species as reported in literature.(172) However, a higher amount of unsaturations were found in the presented study (see Table 10).

Table 10. Identified mycolic acids as detected in *Rhodococcus* species.

Compound	Sum formula	Exact mass [M-H] ⁻	Exp. mass	Mass Deviation
Mycolic acid C28:0	C ₂₈ H ₅₆ O ₃	439.4157	439.4159	0.5 ppm
Mycolic acid C30:1	C ₃₀ H ₅₈ O ₃	465.4313	465.4315	0.4 ppm
Mycolic acid C30:0	C ₃₀ H ₆₀ O ₃	467.4470	467.4472	0.4 ppm
Mycolic acid C31:1	C ₃₁ H ₆₀ O ₃	479.4470	479.4473	0.6 ppm
Mycolic acid C31:0	C ₃₁ H ₆₂ O ₃	481.4626	481.4630	0.8 ppm
Mycolic acid C32:2	C ₃₂ H ₆₀ O ₃	491.4470	491.4475	1.0 ppm
Mycolic acid C32:1	C ₃₂ H ₆₂ O ₃	493.4626	493.4634	1.6 ppm
Mycolic acid C32:0	C ₃₂ H ₆₄ O ₃	495.4783	495.4786	0.6 ppm
Mycolic acid C33:2	C ₃₃ H ₆₂ O ₃	505.4626	505.4630	0.8 ppm
Mycolic acid C33:1	C ₃₃ H ₆₄ O ₃	507.4783	507.4785	0.4 ppm
Mycolic acid C33:0	C ₃₃ H ₆₆ O ₃	509.4939	509.4943	0.8 ppm
Mycolic acid C34:3	C ₃₄ H ₆₂ O ₃	517.4626	517.4632	1.2 ppm
Mycolic acid C34:2	C ₃₄ H ₆₄ O ₃	519.4783	519.4788	1.0 ppm
Mycolic acid C34:1	C ₃₄ H ₆₆ O ₃	521.4939	521.4944	1.0 ppm
Mycolic acid C34:0	C ₃₄ H ₆₈ O ₃	523.5096	523.5100	0.8 ppm
Mycolic acid C35:3	C ₃₅ H ₆₄ O ₃	531.4783	531.4784	0.2 ppm
Mycolic acid C35:2	C ₃₅ H ₆₆ O ₃	533.4939	533.4946	1.3 ppm
Mycolic acid C35:1	C ₃₅ H ₆₈ O ₃	535.5096	535.5100	0.7 ppm
Mycolic acid C35:0	C ₃₅ H ₇₀ O ₃	537.5252	537.5259	1.3 ppm
Mycolic acid C36:3	C ₃₆ H ₆₆ O ₃	545.4939	545.4944	0.9 ppm
Mycolic acid C36:2	C ₃₆ H ₆₈ O ₃	547.5096	547.5102	1.1 ppm
Mycolic acid C36:1	C ₃₆ H ₇₀ O ₃	549.5252	549.5260	1.5 ppm
Mycolic acid C36:0	C ₃₆ H ₇₂ O ₃	551.5409	551.5424	2.7 ppm
Mycolic acid C37:3	C ₃₇ H ₆₈ O ₃	559.5096	559.5102	1.1 ppm
Mycolic acid C37:2	C ₃₇ H ₇₀ O ₃	561.5252	561.5257	0.9 ppm
Mycolic acid C37:1	C ₃₇ H ₇₂ O ₃	563.5409	563.5418	1.6 ppm
Mycolic acid C37:0	C ₃₇ H ₇₄ O ₃	565.5565	565.5573	1.4 ppm
Mycolic acid C38:4	C ₃₈ H ₇₄ O ₃	571.5096	571.5098	0.3 ppm
Mycolic acid C38:3	C ₃₈ H ₇₄ O ₃	573.5252	573.5261	1.6 ppm
Mycolic acid C38:2	C ₃₈ H ₇₄ O ₃	575.5409	575.5415	1.0 ppm
Mycolic acid C38:1	C ₃₈ H ₇₄ O ₃	577.5565	577.5579	2.4 ppm
Mycolic acid C39:2	C ₃₈ H ₇₆ O ₃	589.5565	589.5578	2.2 ppm

This might be due to a lack of sensitivity in the cited literature source as these highly unsaturated species detected in the REIMS spectra are of low spectral intensity. All ionic species in Table 10 could be assigned to alpha- (or alpha'-) mycolic acids based on exact mass measurements.

4.2.2.3. *Nocardia* spp.

In literature sources, *Nocardia* spp. were reported to contain mycolic acids with chain lengths between 48 and 60 carbon atoms with a degree of unsaturation between 0 and 3.(169,173) These findings could be confirmed for a *Nocardia* sp. isolated from a sample of respiratory origin and type-strain *Nocardia asteroides* ATCC 19247. Detected mycolic acid species are listed in Table 11 and could be assigned to alpha- (or alpha'-) mycolic acids based on exact mass measurements.

Table 11. Identified mycolic acids as detected in *Nocardia* species.

Compound	Sum formula	Exact mass [M-H]	Exp. mass	Mass Deviation
Mycolic acid C48:3	C ₄₈ H ₉₀ O ₃	713.6817	713.6797	2.8 ppm
Mycolic acid C48:2	C ₄₈ H ₉₂ O ₃	715.6974	715.6959	2.1 ppm
Mycolic acid C50:3	C ₅₀ H ₉₄ O ₃	741.7130	741.7114	2.2 ppm
Mycolic acid C50:2	C ₅₀ H ₉₆ O ₃	743.7287	743.7285	0.3 ppm
Mycolic acid C52:3	C ₅₂ H ₉₄ O ₃	769.7443	769.7430	1.7 ppm
Mycolic acid C52:2	C ₅₂ H ₉₆ O ₃	771.7600	771.7588	1.6 ppm
Mycolic acid C53:3	C ₅₃ H ₉₆ O ₃	783.7600	783.7596	0.5 ppm
Mycolic acid C53:2	C ₅₃ H ₉₄ O ₃	785.7756	785.7754	0.3 ppm
Mycolic acid C54:4	C ₅₄ H ₉₆ O ₃	795.7600	795.7594	0.8 ppm
Mycolic acid C54:3	C ₅₄ H ₉₈ O ₃	797.7756	797.7739	2.1 ppm
Mycolic acid C54:2	C ₅₄ H ₁₀₀ O ₃	799.7913	799.7902	1.4 ppm
Mycolic acid C55:4	C ₅₄ H ₁₀₂ O ₃	809.7756	809.7748	1.0 ppm
Mycolic acid C55:3	C ₅₄ H ₁₀₄ O ₃	811.7913	811.7907	0.7 ppm
Mycolic acid C55:2	C ₅₄ H ₁₀₆ O ₃	813.8069	813.8061	1.0 ppm
Mycolic acid C56:5	C ₅₆ H ₁₀₂ O ₃	821.7756	821.7748	1.0 ppm
Mycolic acid C56:4	C ₅₆ H ₁₀₄ O ₃	823.7913	823.7907	0.7 ppm
Mycolic acid C56:3	C ₅₆ H ₁₀₆ O ₃	825.8069	825.8053	1.9 ppm
Mycolic acid C56:2	C ₅₆ H ₁₀₈ O ₃	827.8226	827.8213	1.6 ppm
Mycolic acid C57:4	C ₅₇ H ₁₀₆ O ₃	837.8069	837.8050	2.3 ppm
Mycolic acid C57:3	C ₅₇ H ₁₀₈ O ₃	839.8226	839.8215	1.3 ppm
Mycolic acid C58:5	C ₅₈ H ₁₀₆ O ₃	849.8069	849.8068	0.1 ppm
Mycolic acid C58:4	C ₅₈ H ₁₀₈ O ₃	851.8226	851.8218	0.9 ppm
Mycolic acid C58:3	C ₅₈ H ₁₁₀ O ₃	853.8382	853.8375	0.8 ppm
Mycolic acid C59:3	C ₅₉ H ₁₁₂ O ₃	867.8539	867.8537	0.2 ppm
Mycolic acid C60:4	C ₆₀ H ₁₁₂ O ₃	879.8539	879.8537	0.2 ppm
Mycolic acid C60:3	C ₆₀ H ₁₁₄ O ₃	881.8695	881.8683	1.4 ppm

Representative fragmentation spectra for different chain length mycolic acids produced by *Nocardia asteroides* ATCC19247 are shown in Figure 56. Observed mycolic acid α -branch fragments are of same length as observed in case of the shorter mycolic acids of the *Corynebacterium* and *Rhodococcus* genera, indicating that major changes in length and number of unsaturations of aliphatic chains predominantly takes place in the meromycolate chain.

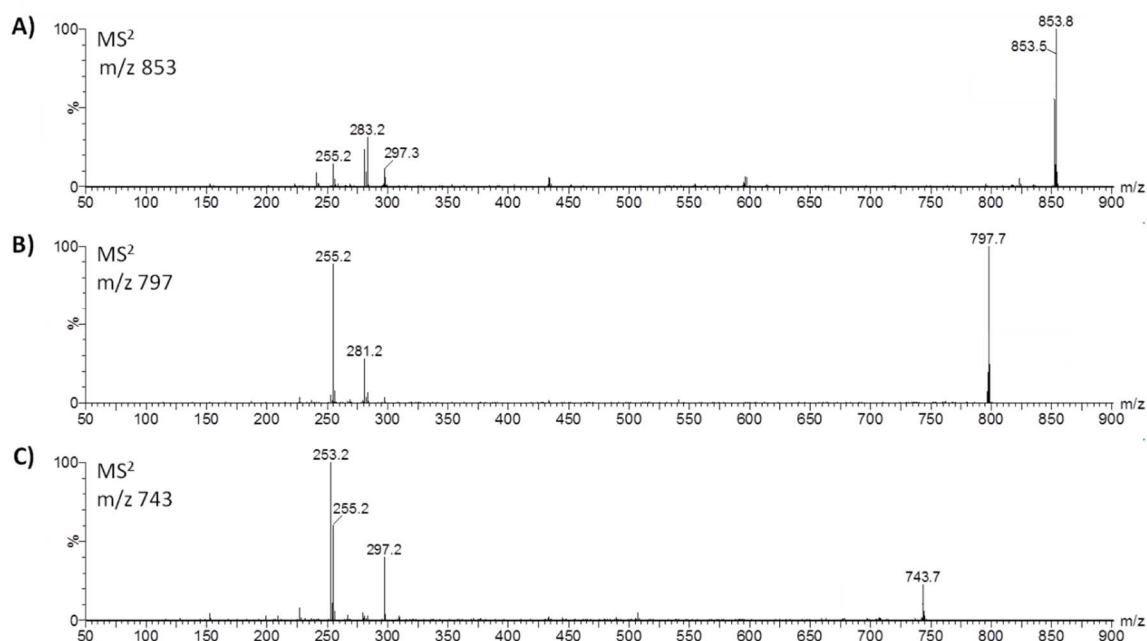


Figure 56. Tandem mass spectra obtained for *Nocardia asteroides* ATCC19247, A) m/z 853, B) m/z 797 and C) m/z 743. Signals were attributed to mycolic acids as listed in Table 11.

4.2.2.4. *Mycobacterium* spp.

One isolate of *Mycobacterium avium*, *M. fortuitum* and *M. peregrinum* were obtained from clinical respiratory specimens and analysed using REIMS. The detected mycolic acid species are listed in Table 12. Mycobacteria are reported to contain mycolic acids with chain lengths between 60 and 90 carbon atoms and 1-2 unsaturations. While other members of the Corynebacterineae suborder mainly contain unfunctionalised mycolic acids, there is a larger structural variability in Mycobacteria ranging from unfunctionalised alpha-mycolic acids to methoxy-, keto- and epoxy-functionalities as part of the beta-hydroxy side chain. Apart from *Segniliparus* spp., no other member of the Corynebacterineae suborder contains similarly long mycolic acids. However, while lower chain length mycolic acids were found with high intensity, mycolic acids in case of *Mycobacterium* spp. were found with comparably low spectral intensity (see Figure 53). This might either be due to lower general abundance in the membrane of Mycobacteria as compared to other genera or more likely decreasing ionisation efficiency with increasing mycolic acid chain lengths. Mycolic acids of at least two different classes were detected (see Table 12), however, further discrimination could not be made based on exact mass measurements only. Tandem mass spectra could not be recorded as signal intensity was not sufficient for fragmentation analysis.

Table 12. Identified mycolic acids as detected in different *Mycobacterium* species.

Compound	Sum formula	Exact mass [M-H]	Exp. mass	Mass Deviation
α/α' -Mycolic acid C77:2	C ₇₇ H ₁₅₀ O ₃	1122.1512	1122.1525	1.2 ppm
α/α' -Mycolic acid C78:2	C ₇₈ H ₁₅₂ O ₃	1136.1669	1136.1684	1.3 ppm
α/α' -Mycolic acid C79:2	C ₇₉ H ₁₅₄ O ₃	1150.1825	1150.1833	0.7 ppm
Epoxy/keto-Mycolic acid C79:1 or Methoxy-Mycolic acid C79:2	C ₇₉ H ₁₅₄ O ₄	1166.1774	1166.1769	0.4 ppm
Epoxy/keto-Mycolic acid C80:1 or Methoxy-Mycolic acid C80:2	C ₈₀ H ₁₅₆ O ₄	1180.1931	1180.1897	2.9 ppm
Epoxy/keto-Mycolic acid C81:1 or Methoxy-Mycolic acid C81:2	C ₈₁ H ₁₅₈ O ₃	1194.2087	1194.2102	1.3 ppm

Based on exact mass measurements the exact type of mycolic acid cannot be determined as α - and α' -mycolic acids as well as epoxy-, keto- and methoxy-mycolic acids respectively do have the same exact mass. All mycolic acids in case of *Corynebacterium*, *Rhodococcus* and *Nocardia* spp. were found to be of the α - or α' -type respectively, while *Mycobacterium* species showed a wider variety of mycolic acids as shown in Table 12.

4.2.3. Sphingolipids in the phylum Bacteroidetes

Sphingolipids are a class of lipids built around a sphingosine backbone with a long-chain amino alcohol and an amide-linked fatty acid attached to it. Sphingolipids are ubiquitous among eukaryotes and serve as structural and signalling components in both mammalian and yeast cells.(174,175) However, sphingolipid production is present in only a few bacterial genera.(176) Species of the Gram-negative bacterial phylum Bacteroidetes are unusual in that they produce sphingolipids with up to 40-70% of their total membrane phospholipid content.(177) Genera within the phylum Bacteroidetes include *Bacteroides*, *Parabacteroides*, *Prevotella*, *Tannerella* and *Porphyromonas*. In eukaryotes, sphingolipids serve also as signaling molecules that play a key role in modulating cell differentiation, gene expression regulation and cell apoptosis.(174,175) Although structurally similar, bacterial sphingolipids are differing from eukaryotic sphingolipid species regarding acyl chain lengths (mammalians are typically 18-20 carbon vs. maximum 19 carbons in bacterial sphingolipids), acyl chain nature (bacterial chains often branched), degree of saturation (mammalian lipids typically contain an unsaturated amino alcohol moiety, while in bacteria often saturated), and sphingolipid headgroup (mammalian cells typically contain phosphocholine while this headgroup is absent in bacteria).(24,178) Sphingolipids that were described in literature and identified in REIMS spectra of members of the phylum Bacteroidetes include free ceramides,

phosphoethanolamine dihydroceramides, as well as substituted and unsubstituted phosphoglycerol dihydroceramides.(179-181) An overview of these detected compounds is given in Table 13. It was recently shown that *Bacteroides fragilis* NCTC 9343 additionally produces an isoform of α -galactosylceramides, a sponge-derived sphingolipid that serves as ligand for the host immune receptor CD1d.(181) These compounds were found in REIMS spectra of *B. fragilis* strains at m/z 752, 766 and 780 and were not found in any other analysed species within the Bacteroidetes class.

Table 13. Identified sphingolipid species in members of the Bacteroidetes phylum.

Formula	Experimental mass	Exact mass	Mass Deviation	Observed in
Ceramide Phosphorylethanolamine/Phosphoethanolamine Dihydroceramides (PE-DHC)				
C ₃₆ H ₇₄ N ₂ O ₇ P ⁻	677.5253	677.5239	2.0	<i>B. fragilis</i> , <i>B. ovatus</i> , <i>B. thetaiotaomicron</i> , <i>B. uniformis</i> , <i>B. vulgatus</i> , <i>P. bivia</i> , <i>P. distonasis</i>
C ₃₇ H ₇₆ N ₂ O ₇ P ⁻	691.5411	691.5396	2.2	
C ₃₈ H ₇₈ N ₂ O ₇ P ⁻	705.5569	705.5552	2.4	
Free ceramides				
C ₃₄ H ₆₉ NO ₄ Cl ⁻	590.4934 ^a	590.4921	2.2	<i>B. fragilis</i> , <i>B. ovatus</i> , <i>B. thetaiotaomicron</i> , <i>B. uniformis</i> , <i>B. vulgatus</i> , <i>P. bivia</i> , <i>P. distonasis</i>
C ₃₅ H ₇₁ NO ₄ Cl ⁻	604.5090	604.5077	2.1	
C ₃₆ H ₇₃ NO ₄ Cl ⁻	618.5246	618.5234	1.9	
<i>Bacteroides fragilis</i> α -Galactosylceramides				
C ₄₀ H ₇₉ NO ₉ Cl ⁻	752.5465	752.5449	2.1	<i>B. fragilis</i>
C ₄₁ H ₈₁ NO ₉ Cl ⁻	766.5623	766.5605	2.3	
C ₄₂ H ₈₃ NO ₉ Cl ⁻	780.5781	780.5762	2.4	
C15:0 substituted Phosphoglycerol Dihydroceramides (subPG-DHC)				
C ₅₀ H ₁₀₀ O ₁₀ NP	904.7007	904.7028	2.3	<i>B. fragilis</i> , <i>B. ovatus</i> , <i>B. thetaiotaomicron</i> , <i>B. uniformis</i> , <i>B. vulgatus</i> , <i>P. distonasis</i>
C ₅₁ H ₁₀₂ O ₁₀ NP	918.7163	918.7185	2.4	
C ₅₂ H ₁₀₄ O ₁₀ NP	932.7324 ^b	932.7337	1.4	
C ₅₃ H ₁₀₆ O ₁₀ NP	946.7481 ^b	946.7484	0.3	
C ₅₄ H ₁₀₈ O ₁₀ NP	960.7637 ^b	960.7624	1.3	
Unsubstituted Phosphoglycerol Dihydroceramides (unPG-DHC)				
C ₃₇ H ₇₆ O ₉ NP	708.5184	708.5199	2.1	<i>P. distonasis</i>
C ₃₉ H ₈₀ O ₉ NP	736.5497	736.5484	1.8	

a: see Fragmentation spectra in Figure 57. b: see fragmentation spectra in Figure 59

Free ceramides were not observed to be identical with those observed for samples of mammalian origin but were found to contain an additional oxygen molecule; a tentative fragmentation mechanism of the [M+Cl]⁻ ion at m/z 590 is shown in Figure 57 explaining the two major fragments observed from collision induced dissociation of the parent ion.

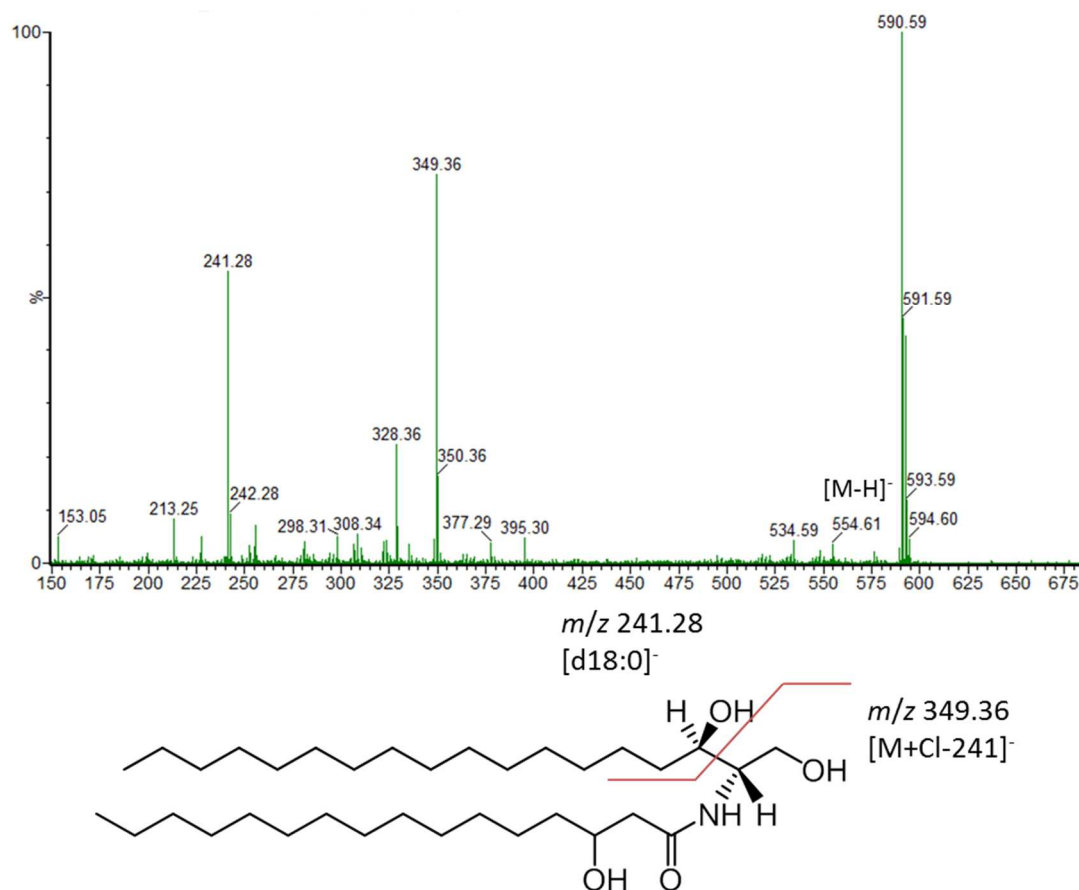


Figure 57. Fragmentation spectrum and tentative scheme of fragmentation for oxidised ceramide signal at m/z 590 obtained from *B. fragilis*. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Figure 58 shows tandem mass spectra obtained for ceramide phosphorylethanolamine $[M-H]^-$ ions that are observed at m/z 677, 691 and 705 within members of the Bacteroidetes class. In agreement with literature spectra, these compounds show a characteristic fragment at m/z 140 which corresponds to the phosphoethanolamine headgroup,(182) as well as a ion corresponding to PO_3^- at m/z 79. All three parent ions furthermore show a fragment at m/z 241 corresponding to a common fatty acid residue with 15 carbons and a neutral loss of $\Delta m = 268$.

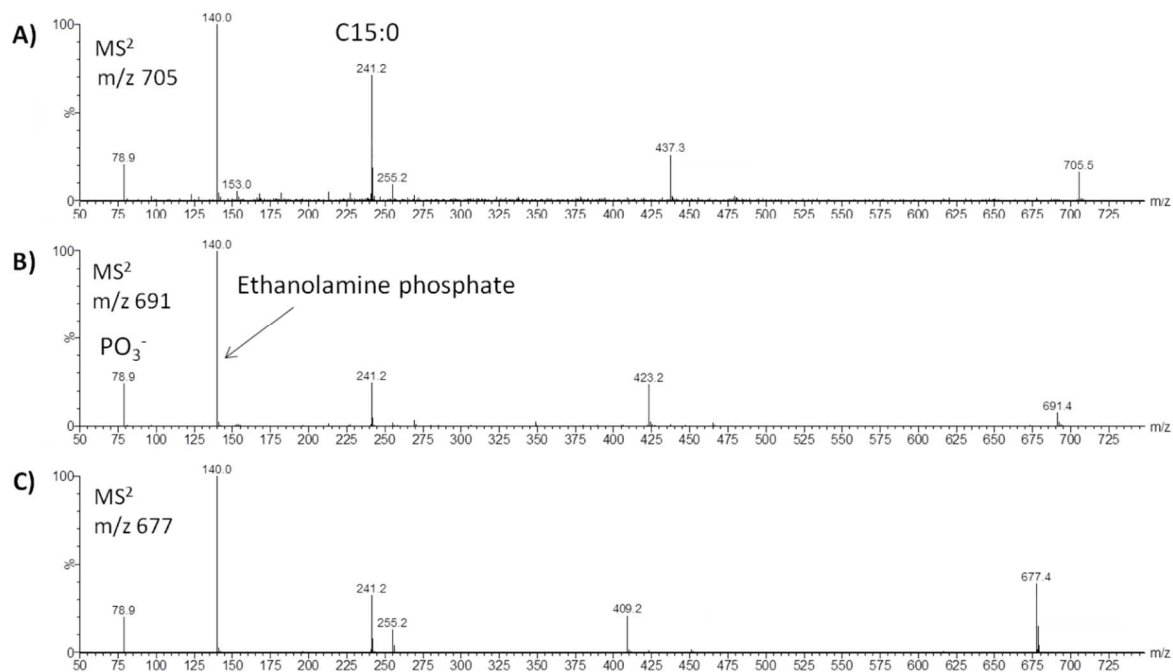


Figure 58. Fragmentation spectra obtained for peaks at A) m/z 705, B) m/z 691 and C) m/z 677 from *Bacteroides fragilis* assigned as ceramide phosphorylethanolamines.

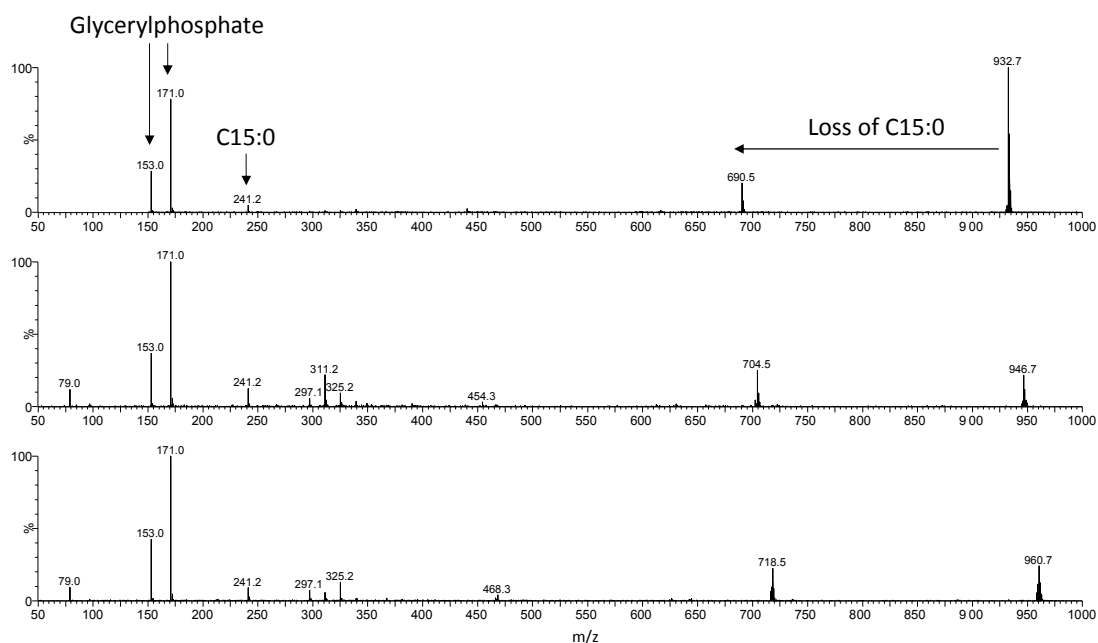


Figure 59. Fragmentation spectra obtained for peaks at m/z 932, 946 and 960 from *Parabacteroides distonasis* assigned as C15:0 substituted phosphoglycerol dihydroceramides (subPG-DHC).

Figure 59 shows fragmentation spectra obtained for $[M-H]^-$ ions of C15:0 substituted phosphoglycerol dihydroceramides that were detected in members of the *Parabacteroides* genus. Main fragments at m/z 153 and 171 can be ascribed to the glycerolphosphate headgroup while m/z 241 can be attributed to C15:0 acyl chains. A neutral loss of 242Da

form the parent ion can additionally be ascribed to the C15:0 acyl chain. These compounds have been described in literature for *Porphyromonas gingivalis* (not part of the present database) and thus seem specific for the *Porphyromonadaceae* family.(180) These compounds were reported to penetrate into human tissues and were found in blood, vascular tissues and brain.(180)

4.2.4. Lipid A in *H. pylori*

Lipopolysaccharides (LPS) are a major component of the outer membrane of Gram-negative bacteria and generally consist of three structural elements, the O-antigen, a core (both consisting of polysaccharide chains) and lipophilic Lipid A (also called endotoxin). Lipid A anchors the lipopolysaccharide in the outer bacterial cell membrane and consists of a usually phosphorylated glucosamine disaccharide unit which acts as a backbone and is substituted with fatty acid residues via either ester or amide linkers. Some of these acyl chains are β -hydroxylated and can carry a further acyl residue attached to the β -hydroxyl group.

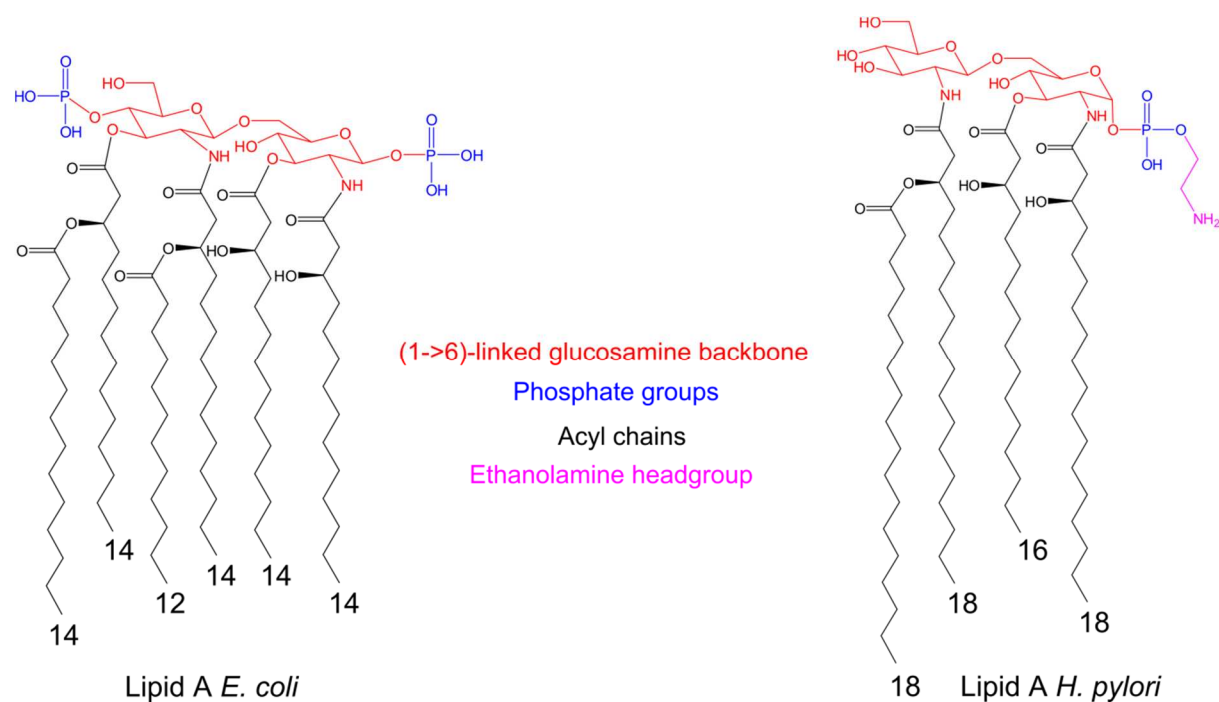


Figure 60. Structures of the endotoxin Lipid A for *Escherichia coli* (left) and *Helicobacter pylori* (right).

Considerable structural variations can be found for different bacterial species, especially in the nature and number of the fatty acid residues attached to the disaccharide backbone. While Lipid A of *E. coli* consists of two carbohydrates, each bearing a phosphate group and six fatty acid residues with chain lengths of 12-14 carbons, *Helicobacter pylori* Lipid A generally

consists of a phosphorylethanolamine group linked to the disaccharide backbone which contains four further acyl substituents with chain lengths of 16-18 carbons (see Figure 60). In the human body, Lipid A binds to the Toll-like receptor 4 (TLR4) complex where it can trigger a potent reaction by the innate immune system. Lipid A is thus a potent virulence factor and absence or modulation of any of its structural components can lead to significant loss in virulence.(183)

H. pylori is the most predominant cause of gastritis and peptic ulcers in humans. The degree of the pro-inflammatory cytokine response to the presence of *H. pylori* is reported to be dependent on the Lipid A acylation patterns. The most abundant form of *H. pylori* Lipid A is shown in Figure 60 and comprises one 3-hydroxyhexadecanoic, two octadecanoic and one 3-hydroxyoctadecanoic fatty acids substituents.(184) This Lipid A variant was detected using REIMS as shown in Figure 61 and confirmed by exact mass measurements with a mass deviation from the theoretical mass of 1.8ppm. Lipid A is the major component of the bacterial outer membrane and would thus be expected as a highly abundant signal in REIMS spectra. Instead, it is detected at intensities below 10% relative abundance. This discrepancy is tentatively associated with its large molecular weight (MW = 1547 g/mol) and associated low ionisation efficiency.

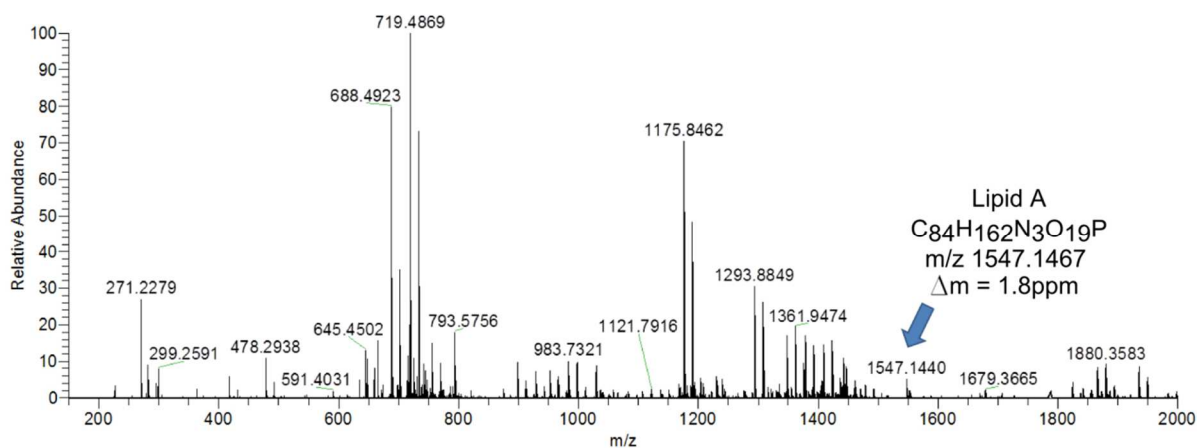


Figure 61. REIMS spectrum obtained from a clinical isolate of *Helicobacter pylori*. Corresponding *m/z* signal highlighted in spectrum. Mass deviation of theoretical and experimental mass determined as 1.8ppm.

Some bacteria were found to modify their Lipid A structures in response to environmental conditions such as magnesium deficiency or such conditions as present during infection of a mammalian host.(185) These structural modifications are necessary in order to maintain outer cell membrane integrity and can lead to significant change in LPS mediated recognition by the host. It was shown that *Pseudomonas aeruginosa* is synthesising structurally different

Lipid A variants depending on the stage of cystic fibrosis (CF) infection. While strains isolated from non-cystic fibrosis show a predominantly a penta-acylated β -(1,6)-glucosamine disaccharide structure, a hexa-acylated Lipid A that contains palmitate and aminoarabinose is observed in CF patients, while a hepta-acylated Lipid A was isolated from CF patients with acute pulmonary disease.(185-187) Since REIMS is principally able to detect Lipid A as shown in case of *H. pylori*, it offers a unique opportunity to identify a bacterial isolate while simultaneously allowing the structural determination of a main virulence factor in a single analysis.

4.2.5. Quorum-sensing molecules in *P. aeruginosa*

Cell-to-cell communication via so-called quorum-sensing (QS) mechanisms is ubiquitous in the bacterial world. These QS systems rely on synthesis of small molecules that diffuse in and out of cells where they promote collective behaviour by modulating gene expression.(188) These molecules are also called autoinducers as among others they also promote expression of genes that lead to their own synthesis. QS molecules are also known to influence the behaviour of eukaryotic cells in order to facilitate bacterial survival.(189) Several chemically distinct molecular families of QS molecules have been described of which the N-acylhomoserine lactone (AHL) family in Gram-negative bacteria have arguably been best studied. *P. aeruginosa* is known to employ both AHL and a more unique system which is biochemically linked to the first and which utilises 2-heptyl-3-hydroxy-4(1H)-quinolone as main functional entity (*Pseudomonas* quinolone signal, PQS). PQS and structurally similar derivatives are known to regulate the production of virulence determinants such as elastase, rhamnolipids, the galactophilic lectin, and the pigment pyocyanin. It was further reported to influence biofilm development and maturation.(190) Fifty-six 4-hydroxy-a-alkylquinolines and related compounds were detected in the culture supernatant of *P. aeruginosa* and structurally assigned by Lepine *et al.* using HPLC-MS/MS.(191)

Abundant signals were observed in REIMS spectra of *P. aeruginosa* that correlate with the PQS-based quorum-sensing system. These were found in different abundances and compositions based on the type of culturing medium, culture age, growth mode and strain. An example for different growth modes observed in a number of *P. aeruginosa* strains is given in Figure 62 and shows spectra of cells of the same strain grown on the same plate in small single colonies (<1mm diameter) and as a lawn. While no signals corresponding to QS

molecules were observed in case of spectra obtained from single colonies, abundant QS molecules were observed for those cells grown in a lawn pattern. Production of QS molecules and resulting gene expression is known to be depending on cell density and was found increased in biofilms of *P. aeruginosa*(192); the observed findings were thus correlated with this cell density-correlated effect. The same effect was observed in case of rhamnolipid production (see next point).

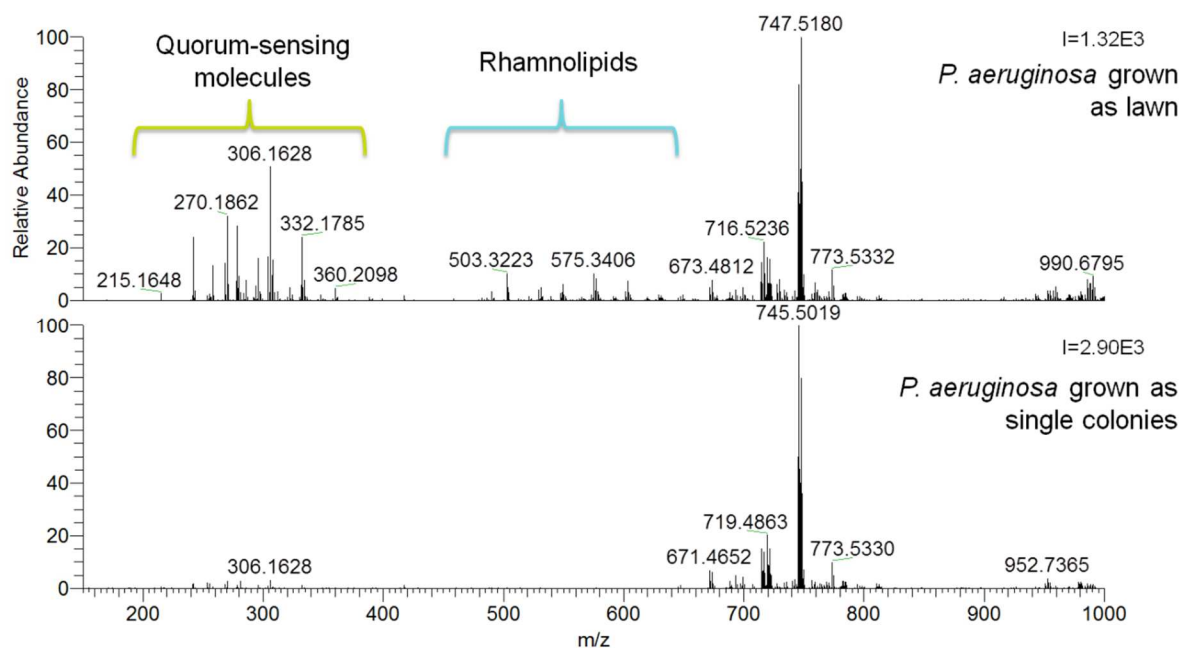


Figure 62. A clinical isolate of *Pseudomonas aeruginosa* grown as a lawn (top) and as single colonies (bottom) showing markedly different REIMS profiles. I=Intensity value at 100% relative abundance.

Table 14. Identified quorum-sensing molecules in *Pseudomonas aeruginosa*.

Compound	Sum formula	Exact mass	Exp. mass	Mass Deviation
2-Heptylquinoline-4(1H)-one	C ₁₆ H ₂₁ NO	[M-H] ⁻ = 242.1550	242.1552	-0.8 ppm
2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS)	C ₁₆ H ₂₁ NO ₂	[M-H] ⁻ = 258.1499	258.1502	-1.2 ppm
Hydroxynonenylquinoline	C ₁₈ H ₂₃ NO	[M-H] ⁻ = 268.1707	268.1711	-1.5 ppm
Hydroxynonylquinoline	C ₁₈ H ₂₅ NO	[M-H] ⁻ = 270.1863	270.1868	-1.9 ppm
Hydroxyundecenylquinoline	C ₂₀ H ₂₆ NO	[M-H] ⁻ = 296.2020	296.2023	-1.0 ppm

Compounds 2-heptylquinoline-4(1H)-one and 2-heptyl-3-hydroxy-4(1H)-quinolone (PQS) have been confirmed by comparison with tandem mass spectra of standard compounds (see Figure 63). Hydroxynonenylquinoline (*m/z* 268), hydroxynonylquinoline (*m/z* 270) and hydroxyundecenylquinoline (*m/z* 296) show similar fragmentation patterns and can thus be ascribed to structurally similar compounds. Common fragments include *m/z* 143, 157 and

170. Tandem mass spectra of these compounds featured in the literature are only reported for the $[M+H]^+$ ion. However, the fragments observed in negative ion mode (m/z 157 and 170) seem to correlate with the fragments observed in positive ion mode (m/z 159 and 172) and are indicative of 4-hydroxy-2-alkylquinolines.(191)

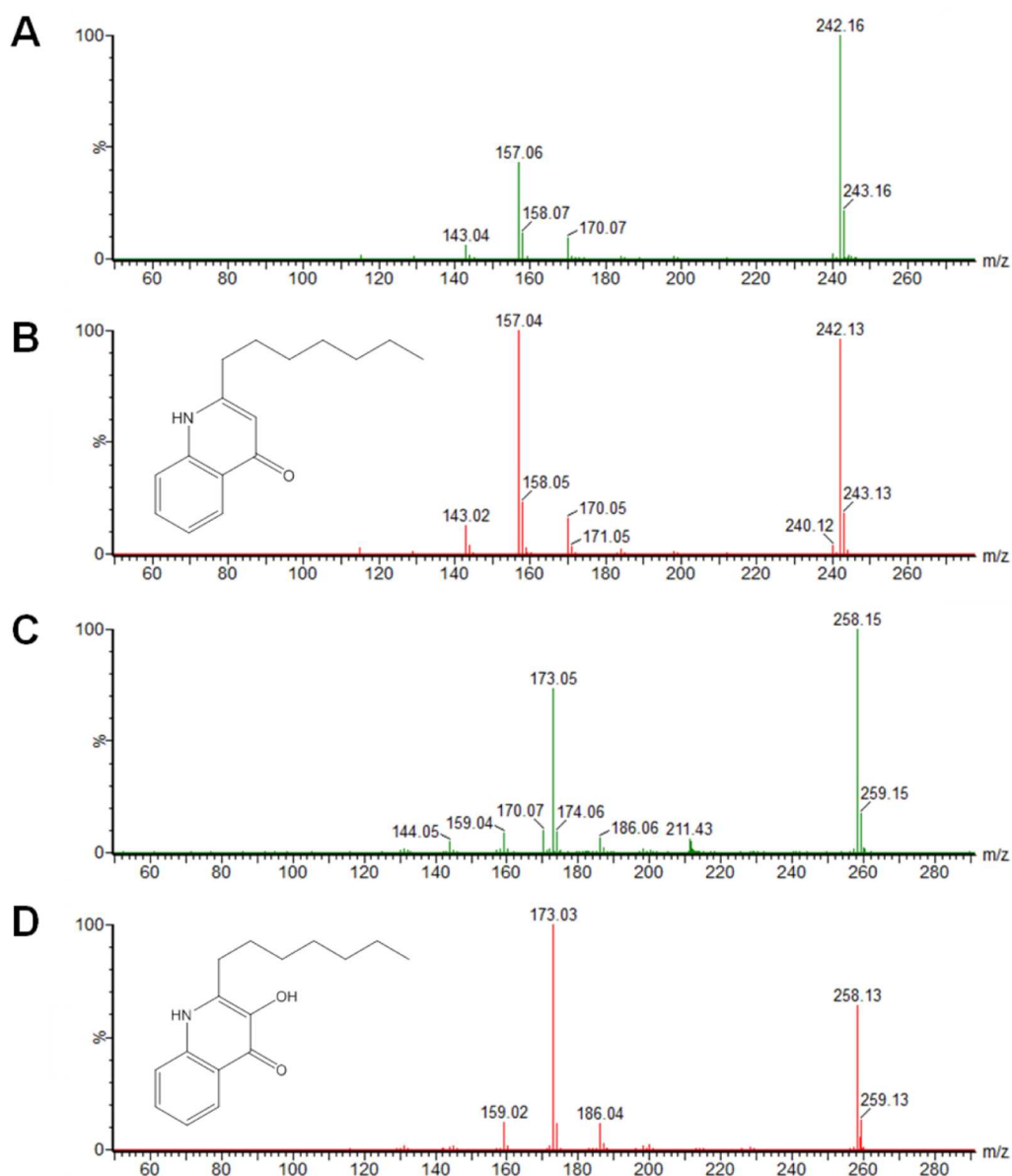


Figure 63. Tandem mass spectra of quorum-sensing molecules produced by *P. aeruginosa*. A) m/z 242, B) 2-Heptylquinoline-4(1H)-one standard, C) m/z 258, D) 2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS).

Fragments for PQS (m/z 258) are all showing a mass shift of 16Da as compared to this common set of fragments which can be explained by the additional hydroxyl group attached to the quinoline moiety. Structurally confirmed quorum-sensing molecules are listed in Table 14. Furthermore, m/z signals at 306 and 332 show fragments at m/z 270 and 296 (and a

common fragment at m/z 157), respectively (loss of 36Da) and a isotopic pattern indicative of $[M+Cl]^-$ adducts of hydroxynonylquinoline and hydroxyundecenylquinoline.

4.2.6. Rhamnolipids in *P. aeruginosa*

Microorganisms like bacteria, yeasts, and fungi are known to produce various types of biosurfactants. Rhamnolipids are a class of surface-active glycolipids containing one or two 3-hydroxy fatty acids of various lengths, linked to a mono- or dirhamnose (Rha) moiety. They can be found in different concentrations and composition patterns in a variety of *Pseudomonas* species. Some *Burkholderia* spp. have also been shown to produce rhamnolipids with longer alkyl chains than those produced by *P. aeruginosa*.⁽¹⁹³⁾ Rhamnolipid production could be linked to physiological functions such as biofilm formation, uptake and biodegradation of poorly soluble substrates, surface motility, displaying antimicrobial activity against both Gram-negative and Gram-positive species as well as a range of fungal species.⁽¹⁹⁴⁾ Rhamnolipids probably also contribute to the inflammation-related tissue damage observed in the lungs of CF patients. In fact, rhamnolipid concentrations in CF patients are high: up to 8 μ g/mL rhamnolipid concentration was found in sputum samples obtained from *P. aeruginosa* colonised CF patients⁽¹⁹⁵⁾ and as much as 65 μ g/mL were found in secretions of a lung removed from a CF patient.⁽¹⁹⁶⁾

Although initially described as a mixture of four congeners, the development of more sensitive analytical techniques has led to the further discovery of about 60 rhamnolipid homologues.⁽¹⁹³⁾ A typical rhamnolipid profile as obtained from *Pseudomonas aeruginosa* isolates is shown in Figure 64. Identified rhamnolipid species are listed in Table 15. Overall spectral intensity of rhamnolipids depends on a number of factors, e.g. single colonies were observed to display less rhamnolipids than when the same bacterial strain was grown as a lawn (this effect was correlated with the detection and production of QS molecules, see previous point). Rhamnolipid production is linked to the QS apparatus, a fact that can be observed in REIMS spectra of *P. aeruginosa* as high abundance of rhamnolipids are usually accompanied by abundant QS signals as shown in Figure 63. Rhamnolipid intensity was also seen to increase significantly with increasing age of the culture. However, not all of the analysed *P. aeruginosa* strains were observed to produce measurable amounts of rhamnolipids even under favourable conditions.

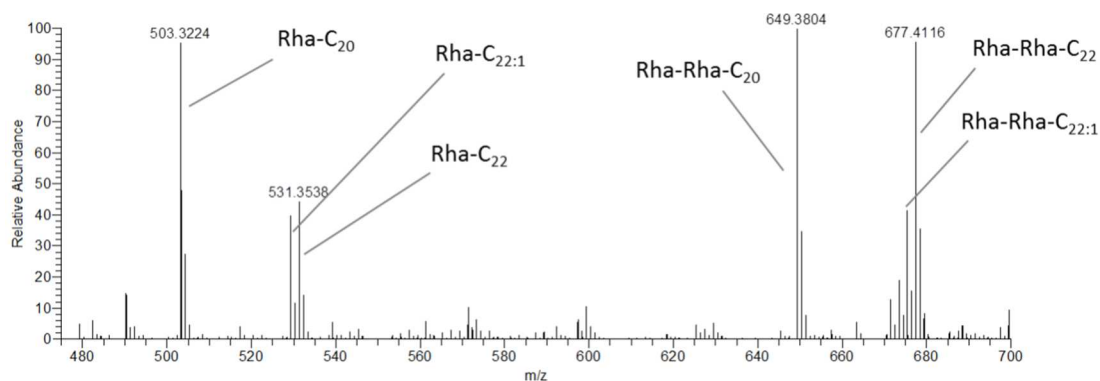


Figure 64. Rhamnolipid species commonly produced by *P. aeruginosa* strains. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Table 15. Rhamnolipid species commonly produced by *P. aeruginosa* strains.

Compound	Sum formula	Exact mass [M-H] ⁻	Exp. mass	Mass Deviation
Rha-C ₂₀	C ₂₆ H ₄₈ O ₉	503.3225	503.3224	0.2 ppm
Rha-C _{22:1}	C ₂₈ H ₅₀ O ₉	529.3382	529.3384	-0.4 ppm
Rha-C ₂₂	C ₂₈ H ₅₂ O ₉	531.3539	531.3538	0.2 ppm
Rha-Rha-C ₂₀	C ₃₂ H ₅₈ O ₁₃	649.3805	649.3804	0.2 ppm
Rha-Rha-C ₂₂	C ₃₄ H ₆₂ O ₁₃	677.4118	677.4116	-0.3 ppm
Rha-Rha-C _{22:1}	C ₃₄ H ₆₀ O ₁₃	675.3961	675.3965	-0.6 ppm

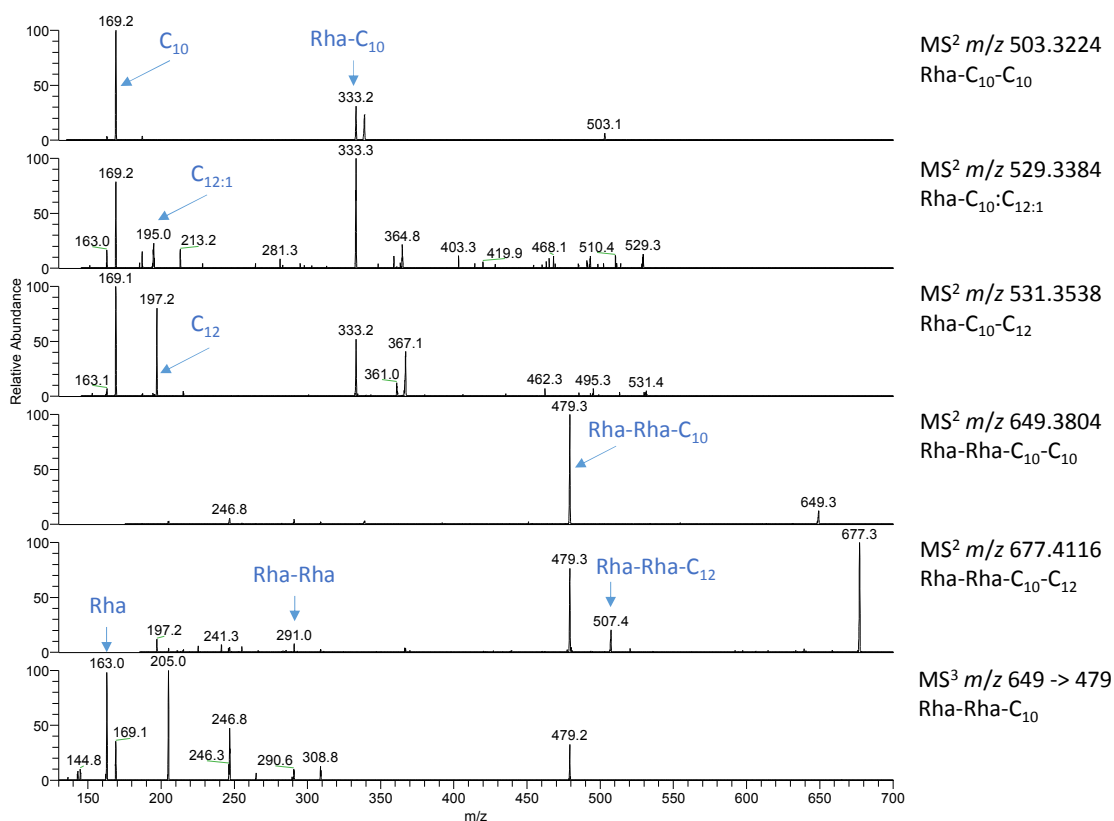


Figure 65. Fragmentation spectra for rhamnolipids species detected in *Pseudomonas aeruginosa* recorded using a Thermo LTQ XL ion trap instrument using collision induced dissociation at collision energy setting of 25.

Fragmentation spectra of rhamnolipids species listed in Table 15 is shown in Figure 65. Rhamnolipids at m/z 503 and m/z 649 were confirmed by comparison to tandem mass spectra published in literature.(197,198) Other rhamnolipids were assigned based on similar fragmentation patterns such as the loss of rhamnose moieties and the loss of one of the acyl chains and exact mass measurements.

Rhamnolipids have several potential biotechnological applications, especially as biodegradable surfactants for use in industry or medicine or as a possible alternative for production of rhamnose.(199,200) Therefore, a REIMS-based rhamnolipid screening method could be applied for the identification of optimal rhamnolipid producers.

4.2.7. Lipopolypeptides in *Bacillus* species

Bacillus spp. are important Gram-positive model organisms that produce a variety of antibiotics of which the most important are lipopolypeptides such as the lipopeptide surfactin ($C_{13}H_{27}HCO|CH_2CO-Glu-Leu-D-Leu-Val-Asp-D-Leu-Leu|$, where the symbols ‘|’ represent the cyclic bonding sites) and its homologues. Surfactins are produced by strains of *Bacillus subtilis* and *Bacillus pumilus*, while structurally closely related lichenysins are produced by *Bacillus licheniformis*. Due to their hydrophobic side chains, these compounds can be incorporated into the phospholipid bilayer where they induce permeabilisation and perturbation of the cell membranes by reorienting the lipid headgroups towards the cell membrane interior.(201) They showed a variety of properties such as anticoagulating and immunosuppressing activity.(202) The various polar functional groups of the surfactin molecule allow straightforward ionisation in both positive and negative ion modes. REIMS spectra obtained in negative ion mode for *B. subtilis*, *B. pumilus* and *B. licheniformis* are shown in Figure 66. Surfactin production was observed both in *B. subtilis* and *B. pumilus*, however, the acyl-chain lengths were markedly different. While *B. subtilis* produced shorter chain length surfactins with side chains containing 13-16 carbon atoms (highest abundance for surfactin(C15)), *B. pumilus* produced longer chain length surfactins with 14-18 carbon atoms (highest abundance surfactin(C16)). Other minor related antibiotic compounds produced by *B. subtilis* include fengycin and iturin, however, these compounds were not observed in the REIMS spectra, presumably due to concentrations below the limit of detection in the tested strains.

Bacillus subtilis mass spectra are dominated by species clustering around m/z 1034 in negative ion mode and m/z 1059 in positive ion mode. This cluster can be ascribed to [Surfactin(C15)-H]⁻ and [Surfactin(C15)+Na]⁺ and C13 and C14 homologues, respectively, as shown in Table 16. The protonated quasi-molecular ions corresponding to [M+H]⁺ adducts were not observed in positive ion mode. This is due to the high affinity of surfactin for sodium cations and high abundance of sodium in living organisms and the culturing medium. The same surfactin species although with different relative abundances are found in strains of *B. pumilus*.

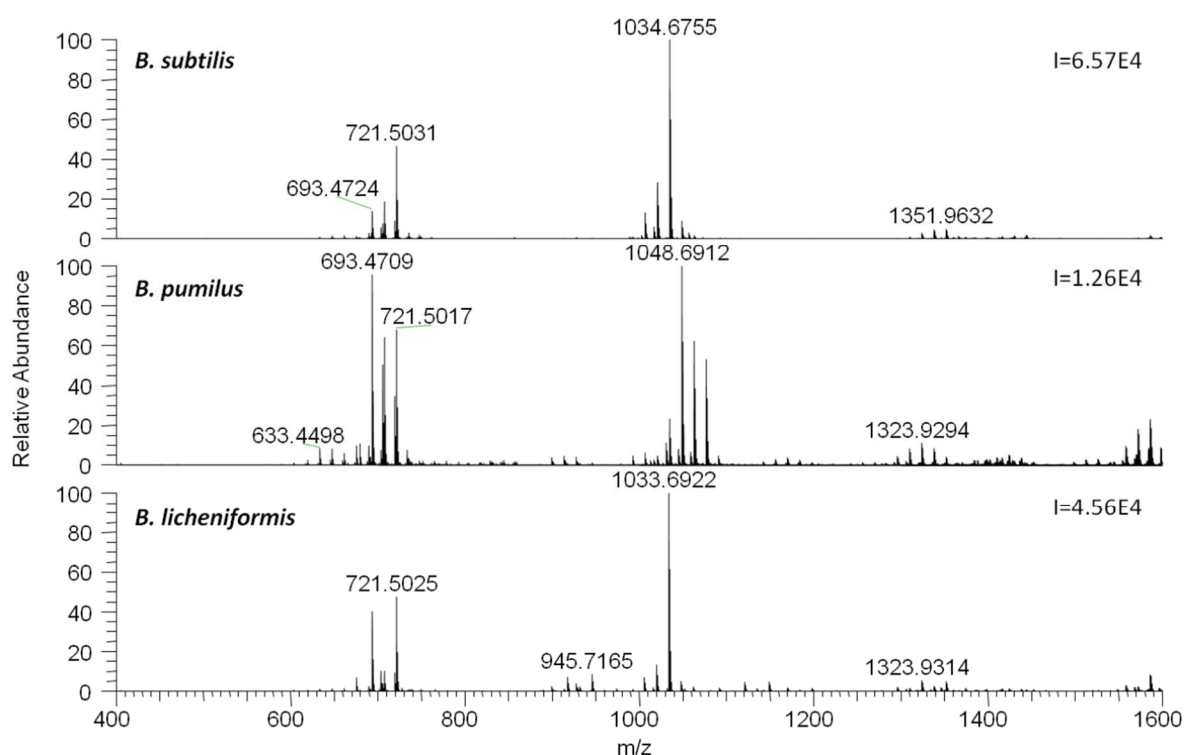


Figure 66. Mass spectra obtained from lipopolypeptide producing strains of the *Bacillus* genus recorded using REIMS.

Table 16. Surfactin species detected in positive and negative ion mode for *Bacillus subtilis* and *Bacillus pumilus*.

Compound	Negative ion mode			Positive ion mode		
	Exp. mass	Exact mass [M-H] ⁻	Δppm	Exp. mass	Exact mass [M+Na] ⁺	Δppm
Surfactin(C13)	1006.6453	1006.6440	1.3	1030.6389	1030.6416	2.6
Surfactin(C14)	1020.6604	1020.6597	0.7	1044.6545	1044.6573	2.7
Surfactin(C15)	1034.6754	1034.6753	0.1	1058.6702	1058.6729	2.6
Surfactin(C16)	1048.6912	1048.6910	0.2			
Surfactin(C17)	1062.7061	1062.7066	0.4			
Surfactin(C18)	1076.7216	1076.7222	0.6			

Doubly charged negative ions of surfactin species (as reported in literature in case of DESI-MS) were not observed using REIMS.(102) This underlines the strong tendency of REIMS to produce singly charged ionic species. Furthermore, significantly higher amounts of phospholipids were detected using REIMS than in a comparable DESI study,(102) indicating better cell disruption and less influence of ion suppression observed in REIMS. The presence of surfactin and its homologues was further confirmed by tandem mass spectrometry measurements. Fragmentation patterns observed in negative ion mode (see Figure 67) correspond well with fragmentation patterns reported in literature.(102) Fragments can be assigned to water loss from the $[M-H]^-$ parent ion, neutral loss of the cyclic polypeptide moiety (fragment at m/z 241), and partial peptide losses.

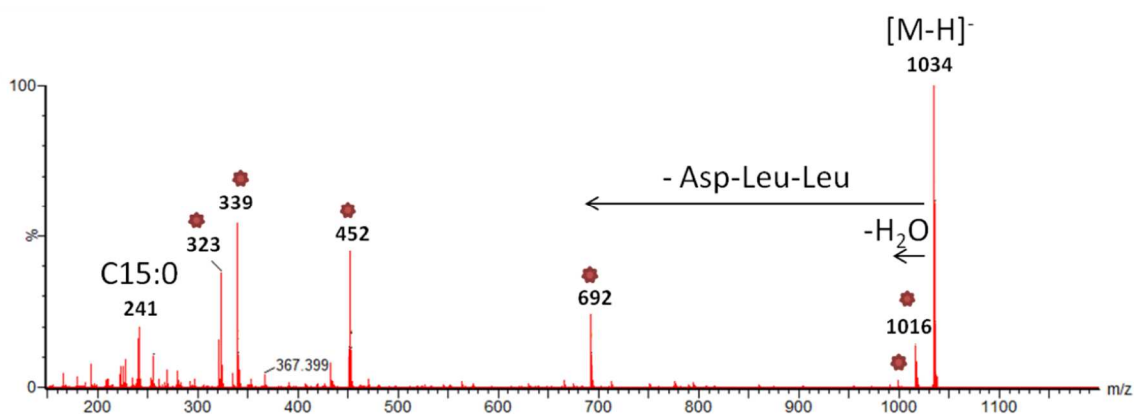


Figure 67. REIMS fragmentation spectrum of m/z 1034 corresponding to Surfactin(C15). Red asterisk indicates fragments reported in Ref. (102) for same compound.

Lichenysin was observed with alkyl chain lengths of 13-16 carbon atoms with the most abundant species being Lichenysin(C15). Detected compounds are listed in Table 17 for negative ion mode. Detection of lipopolypeptides surfactin and lichenysin suggests that REIMS ionisation mechanism extends to peptide-like structures if compounds possess sufficient amphiphilic properties. In conclusion, REIMS offers a rapid and specific identification for the *Bacillus subtilis* complex of organisms which are otherwise challenging to differentiate based on traditional phenotypic tests and phylogenetic analysis of the 16S rRNA sequences.(203)

Table 17. Lichenysin compounds detected in *Bacillus licheniformis*.

Compound	Exp. mass	Exact mass [M-H] ⁻	Δppm
Lichenysin (C13)	1005.6594	1005.6600	0.6
Lichenysin (C14)	1019.6748	1019.6756	0.8
Lichenysin (C15)	1033.6906	1033.6913	0.7
Lichenysin (C16)	1047.7055	1047.7070	1.4

4.2.8. Polyhydroxyalkanoate polymers detected using REIMS

Polyhydroxyalkanoates (PHAs) are linear polyesters produced naturally by bacterial fermentation of sugar or lipids, usually under a shortage of a non-carbonous nutrient, e.g. nitrogen. They are produced by the bacteria to store carbon and energy. These polyesters are biodegradable and are used in the production of bioplastics. The simplest and most commonly occurring form of PHA is poly-β-hydroxybutyrate (poly-3-hydroxybutyrate, P3HB). However, more than 100 different monomers have been reported as PHA constituents. Generally, PHAs are classified into three different classes according to monomer carbon chain length: short (C3-C5), medium (C6-C14) and long (C>14) chain PHAs.(204) More than 90 genera of archaea and eubacteria (both Gram-positive and Gram-negative) have been reported to produce PHAs.(205)

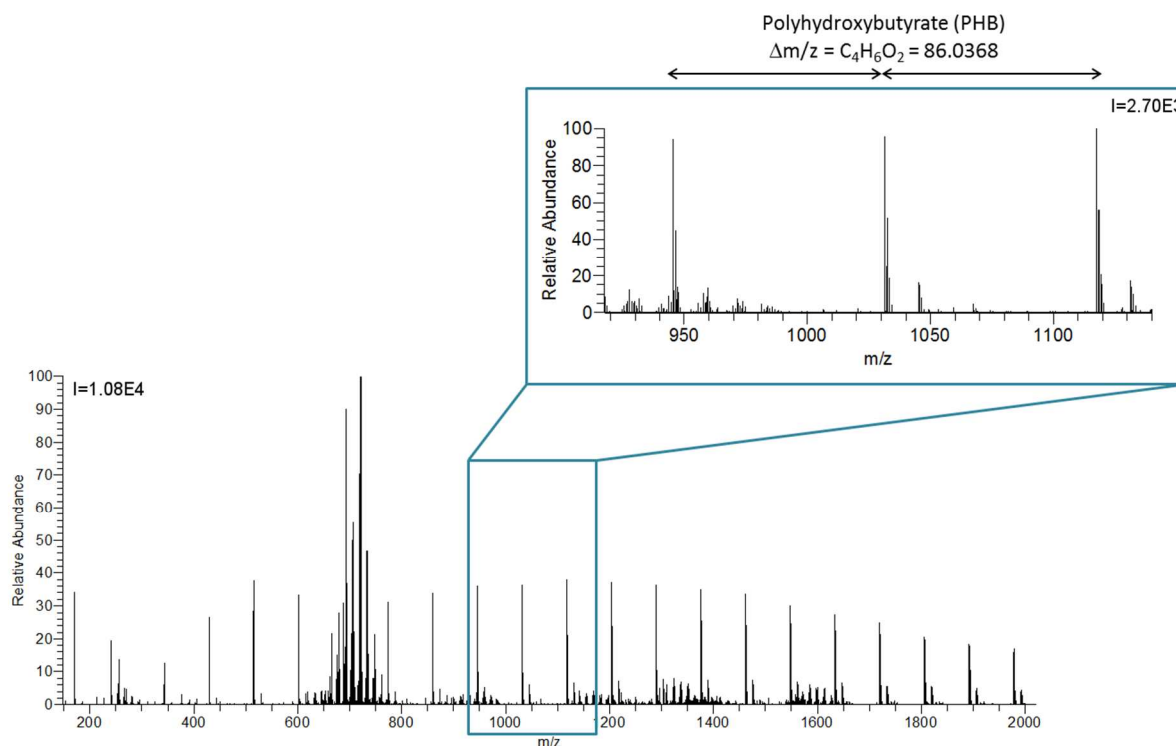


Figure 68. REIMS spectral profile of a clinical isolate of *Bacillus cereus* displaying polyhydroxybutyrate production. I=Intensity value at 100% relative abundance.

All polymers detected using REIMS were consisting of C₄H₆O₂ (polyhydroxybutyrate) monomers, however, the exact nature of polymer (P3HB or P4HB) could not be determined based on exact mass measurements only. These polymers were found in different amounts in strains of *Bacillus cereus* (see Figure 68), *Delftia acidovorans*, *Burkholderia cepacia* complex and *Achromobacter xylosoxidans*.

The ability of REIMS to detect biopolymer production in several bacterial species offers an opportunity for biotechnological applications to rapidly screen organisms for successful production of these polymers, average polymer chain lengths and modifications and/or best polymer producers. This is equally true for production of biosurfactants as shown in case of lipopolypeptides and rhamnolipids.

4.3. Specificity assessment of REIMS-based microbial ID method

4.3.1. Specificity on different taxonomic levels

Suitability of the proposed REIMS method for bacterial identification requires that inter-species spectral variance needs to be larger than the intra-species variance for different strains and phenotypes of the same species. In order to investigate whether this basic condition is fulfilled, a dataset was created comprising 15 different clinical isolates for each of 28 different bacterial species (see Table 18).

Results of multivariate statistical analyses of the generated dataset are given in Figure 69A and B, respectively. Generally the plots resulting from the supervised and unsupervised analysis of REIMS data show high similarity. This is due to the fact that REIMS features exclusively signals originating from the sample (i.e. not from any chemical background). This is an advantage compared to DESI- or MALDI-MS where solvent- and matrix-related signals significantly contribute to the overall spectral content.

Gram-positive and Gram-negative species are separated along the first multivariate component in both PCA and RMMC analysis (Figure 69A and B). Compared to Gram-negative species, Gram-positive bacteria generally show a higher amount of saturated phospholipid species and lower relative abundance of phosphatidylethanolamines. These observations are in agreement with the bacterial cell membrane composition reported in literature.(1)

Table 18. Dataset used to assess the specificity of a REIMS-based identification system. 15 clinical isolates were used per species. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Gram-stain	Family	Genus	Species	Growth conditions
negative	Pseudomonadaceae	<i>Pseudomonas</i>	<i>aeruginosa</i>	CBA, aerobic
		Enterobacteriaceae	<i>Citrobacter</i>	<i>koseri</i>
	<i>Enterobacter</i>		<i>aerogenes</i>	CBA, aerobic
			<i>cloacae</i>	CBA, aerobic
	<i>Klebsiella</i>		<i>oxytoca</i>	CBA, aerobic
			<i>pneumoniae</i>	CBA, aerobic
	<i>Escherichia</i>		<i>coli</i>	CBA, aerobic
	<i>Proteus</i>		<i>mirabilis</i>	CBA, aerobic
	<i>Morganella</i>		<i>morganii</i>	CBA, aerobic
	<i>Serratia</i>		<i>marcescens</i>	CBA, aerobic
	Pasteurellaceae		<i>Haemophilus</i>	<i>influenzae</i>
	Burkholderiaceae	<i>Burkholderia</i>	<i>cepacia complex</i>	CBA, aerobic (5 % CO ₂)
	Xanthomonadaceae	<i>Stenotrophomonas</i>	<i>maltophilia</i>	CBA, aerobic
	Bacteroidaceae	<i>Bacteroides</i>	<i>fragilis</i>	CBA, anaerobic
Moraxellaceae	<i>Moraxella</i>	<i>catarrhalis</i>	CBA, aerobic	
Neisseriaceae	<i>Neisseria</i>	<i>gonorrhoeae</i>	CBA, aerobic (5 % CO ₂)	
positive	Staphylococcaceae	<i>Staphylococcus</i>	<i>aureus</i>	CBA, aerobic
			<i>epidermidis</i>	CBA, aerobic
			<i>capitis</i>	CBA, aerobic
			<i>haemolyticus</i>	CBA, aerobic
			<i>hominis</i>	CBA, aerobic
			<i>faecalis</i>	CBA, aerobic (5 % CO ₂)
	Enterococcaceae	<i>Enterococcus</i>	<i>faecium</i>	CBA, aerobic (5 % CO ₂)
			<i>difficile</i>	CBA, anaerobic
	Clostridiaceae	<i>Clostridium</i>	<i>luteus</i>	CBA, aerobic
	Streptococcaceae	<i>Streptococcus</i>	<i>agalactiae</i>	CBA, aerobic(5 % CO ₂)
			<i>pyogenes</i>	CBA, aerobic (5 % CO ₂)
<i>pneumoniae</i>			CBA, aerobic (5 % CO ₂)	

Hierarchical cluster analysis was performed (HCA) in order to investigate how well the REIMS spectral profiles follow the bacterial taxonomy as determined by 16S rRNA gene sequences. HCA was implemented using Euclidean pairwise distance calculation with a complete linkage metric. Three sets of each three strains of the original dataset shown in Figure 69A and B were averaged for each bacterial species to form the dataset which was subjected to HCA. This step was undertaken in order to facilitate visualisation while still incorporating a maximum of the biological variance among strains of a certain species. Figure 70 shows that spectral profiles of closely related bacterial species are grouped closely together while rather unrelated bacterial species group separately. For the Gram-positive species this is visible for each the *Staphylococcus* spp. (*S. aureus*, *S. capitis*, *S. epidermidis*, *S. hominis* and *S. haemolyticus*), *Streptococcus* spp. (*S. agalactiae*, *S. pneumoniae*, *S.*

pyogenes) and two *Enterococcus* spp. (*E. faecalis* and *E. faecium*). *Streptococcus* and *Enterococcus* spp. which both belong to the *Lactobacillales* order are further situated on the same cluster in the HCA.

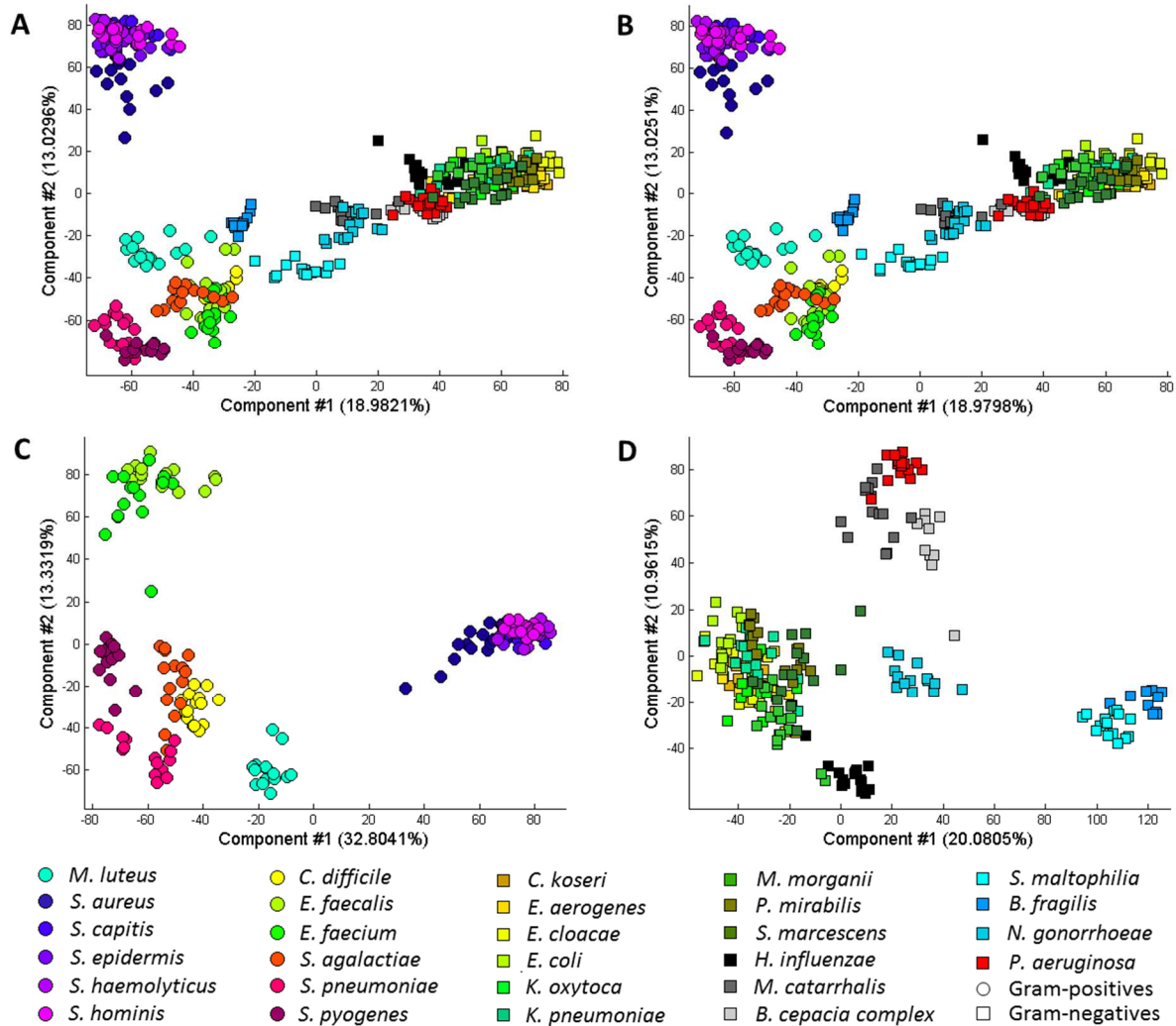


Figure 69. Pseudo-2-dimensional A) PCA and B) RMMC plots of model comprising 28 different clinically relevant bacterial species. Gram-positives are indicated by circles (○), Gram-negatives are represented by squares (□). C) and D) are showing PCA plots for only the Gram-positive and Gram-negative species, respectively. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Regarding Gram-negative species, all members of the *Enterobacteriaceae* family (members of the genera *Escherichia*, *Citrobacter*, *Enterobacter*, *Proteus*, *Morganella*, *Klebsiella* and *Serratia*) grouped closely together. Furthermore, *Pseudomonas aeruginosa*, *Moraxella catarrhalis* and *Burkholderia cepacia complex* strains are all located together in a separate cluster when compared to the other Gram-negative species. *Pseudomonas* spp. and *Moraxella* spp. are both part of the *Pseudomonadales* order. Although *Burkholderia cepacia complex* strains belong to β -Proteobacteria today, they were previously classified into the

Pseudomonas genus(206), thus indicating a high phenotypic similarity between *Pseudomonas* spp. and *Burkholderia* spp. which explains their proximity on the HCA dendrogram (see Figure 70). The same trends were observed in the PCA plots for Gram-positive and Gram-negative-species only (see Figure 69C and D). Similarly to HCA, these results demonstrate that REIMS spectral profiles largely follow taxonomical trends. The overall agreement with the bacterial taxonomy is expected to increase with larger coverage of bacterial diversity among different phyla, classes and orders.

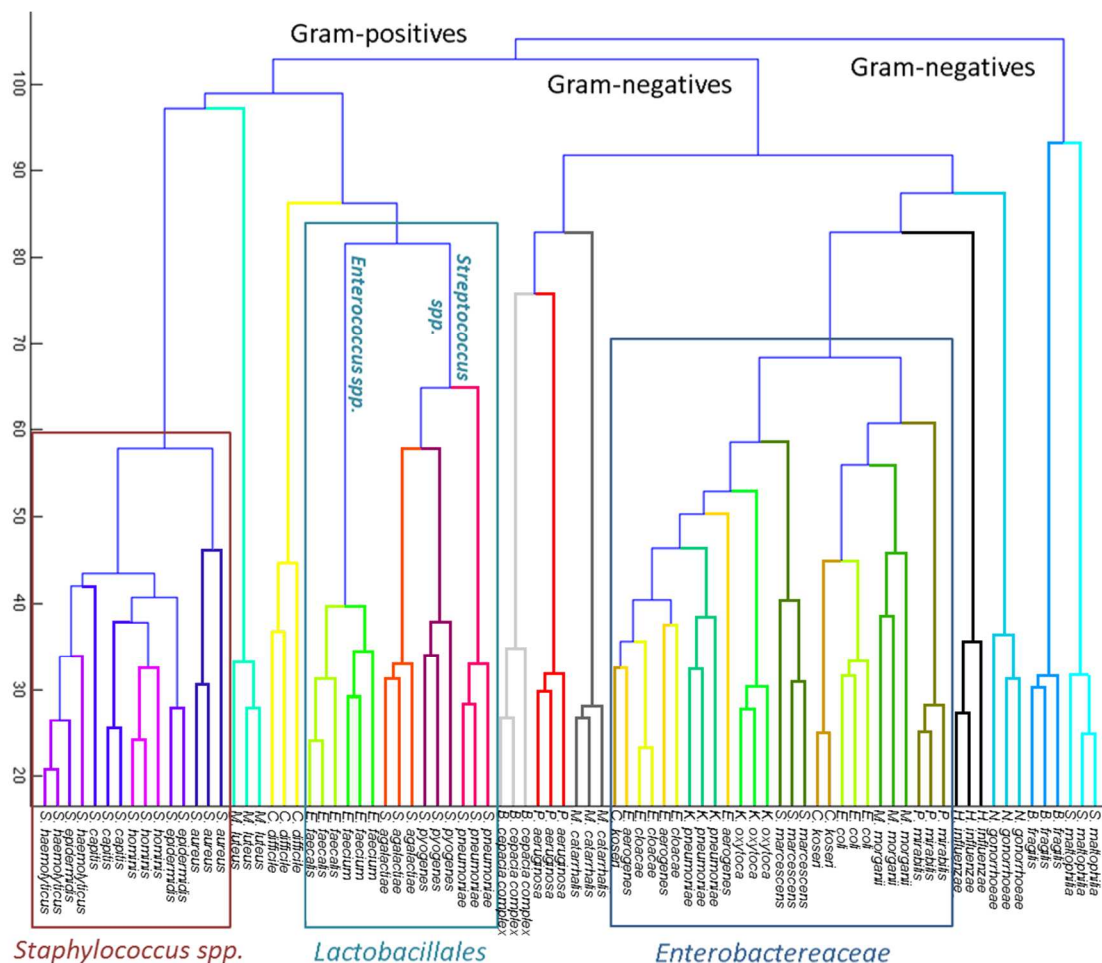


Figure 70. Hierarchical cluster analysis (HCA) of the dataset shown in Table 18 and Figure 69. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

The presented RMMC model (Figure 69B) was cross-validated in order to assess the specificity of the REIMS method on Gram-, genus- and species-level. Cross-validation was performed using leave-one-out cross-validation and three nearest neighbours as classifier. In leave-one-out cross-validation, each data point is left out from the supervised model once and then projected into the generated data space and classified according to a given criterion (here according to its 3 nearest neighbours). Therefore, this type of cross-validation shows

similarity with blind identification tests and is applied here to assess the specificity of the method. However, for Gram-level cross-validation an entire bacterial species was left out at a time. Cross-validation results yield 95.9%, 97.8% and 100% correct classification at species-, genus- and Gram-level, respectively. Misclassifications were only observed for closely related bacterial species and comprised three misclassifications within the *Staphylococcus* genus, five misclassifications within the *Enterococcus* genus and nine misclassifications within the *Enterobacteriaceae* family. Misclassifications are defined in comparison to identifications provided by a commercial MALDI-TOF-MS instrument as obtained during routine clinical work. The correct identification statistics is therefore not expected to be better than those reported in literature for commercial MALDI MS systems.

4.3.2. Dependence of the correct identification performance on spectral resolution

In order to assess the mass resolution required to reliably distinguish between the different bacterial species as shown in Figure 69, the original data were binned using varying bin sizes. Aligned raw data were normalised to the total ion count (TIC) and log-transformed for each bin size individually before unsupervised and supervised multivariate analyses were carried out (for respective PCA plots obtained for bin sizes 0.01, 0.1 and 1Da see Figure 72A, B and C, respectively). The minimum resolution required for accurate classification using REIMS was assessed by comparing the results of cross-validations for each bin size. As for the model shown in Figure 69, leave-one-out cross-validation was used at species- and genus-level, whereas leave-species-out cross-validation was performed at Gram-stain-level.

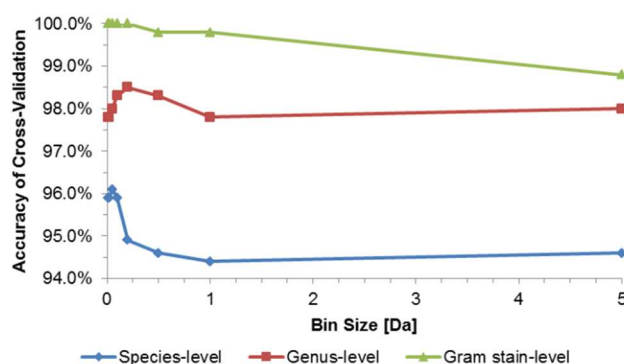


Figure 71. Cross-validation results for the dataset shown in Figure 69 as a function of the bin size on species-, genus-, and Gram stain-level. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

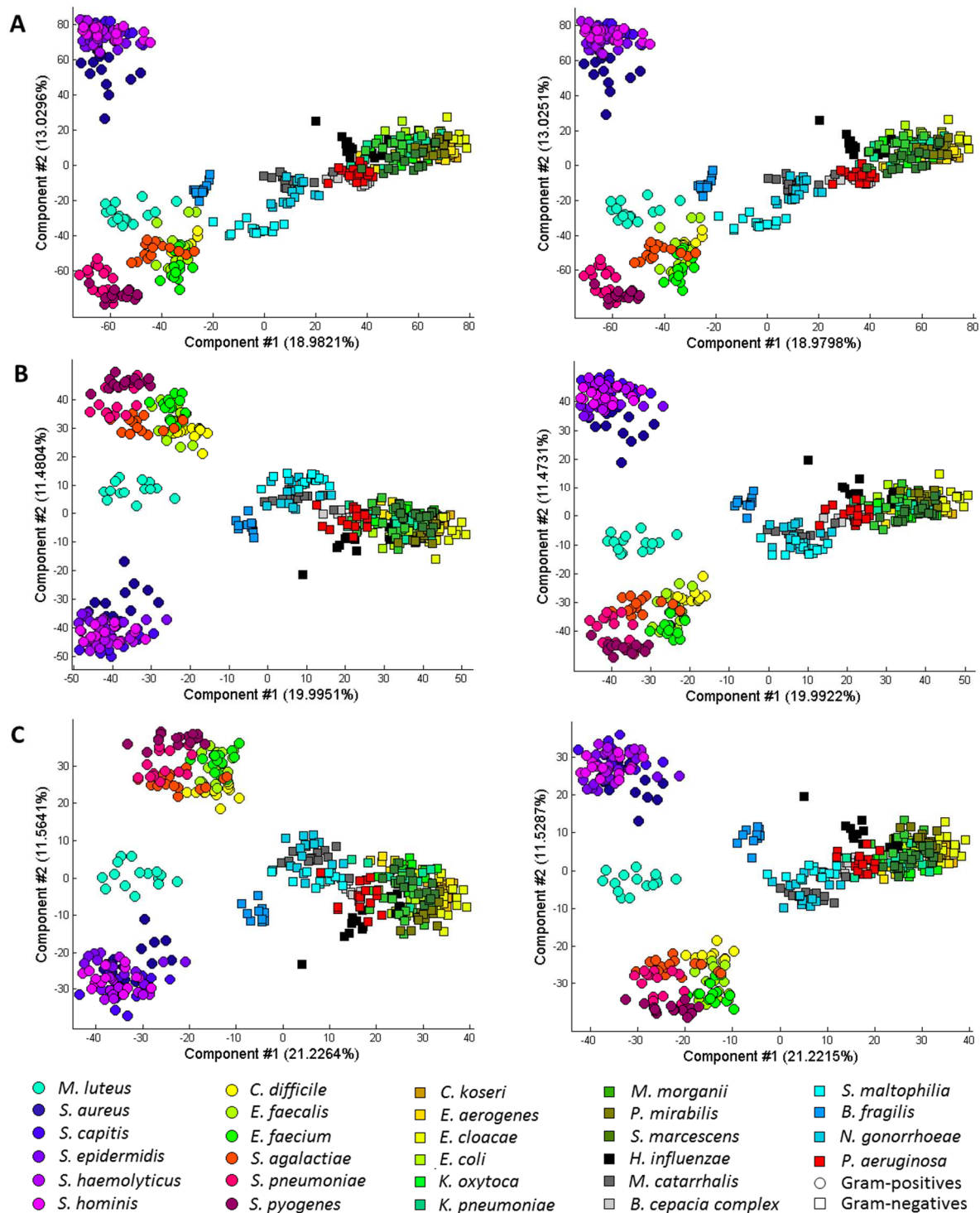


Figure 72. Results of unsupervised (left) and supervised (right) analysis for the dataset shown in Figure 69 using A) 0.01Da, B) 0.1Da, and C) 1Da bin size. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

The results given in Figure 71 show that no considerable loss in identification accuracy was observed neither at species-, genus- nor Gram-stain-level for different bin sizes between 0.01 and 1Da. These data strongly suggests that a mass analyser working at unit resolution would

be sufficient for the routine identification of unknown bacteria. Even a bin size of 5Da does not result in a significant loss of identification accuracy, clearly indicating the presence of major spectral differences even between closely related species.

Nevertheless, performing analysis on a high resolution mass spectrometer facilitates identification of unknown spectral features by exact mass measurements. In addition, using high-resolution data leads to more compact data groups in PCA and improved information recovery as can be seen in case of Lactobacillales in Figure 72, where separation of the *S. agalactiae* group from other *Streptococcus* spp. gets less pronounced towards higher bin sizes. Similarly, *S. maltophilia* and *B. fragilis* are observed to move towards the group of Gram-negative bacteria suggesting a loss of significant features towards larger bin sizes.

To test how well the REIMS spectra are following the phylogenetic tree of bacteria on a more comprehensive dataset, a reduced dataset was created from the entire database listed in Table 5 containing only one strain for each of the 217 bacterial species listed in Appendix 4 to enable visualisation using multivariate statistical methods.

PCA was applied to initially assess the similarity between REIMS spectral profiles of bacterial species on different higher taxonomical levels such as Gram-, phylum-, class- and order-level. HCA analysis is traditionally more commonly used for this purpose; however, PCA is better suited for a more intuitive visualisation of large datasets as the one presented here. In Figure 73A, all 217 species are coloured based on their Gram-stain behaviour. Figure 73B and C each show the same PCA plot, however, the labelling is according to phylum (B) and class (C) of the phylogenetic tree and are intended to facilitate the visualisation of the clustering on different taxonomic levels. In Figure 73A, an overall good separation can be observed for Gram-positive and Gram-negative species along the first principal component. However, Gram-negative species are clearly separated into two distinct clusters while Gram-positive species show a tentative separation into three different clusters. For Gram-negative species this separation can be attributed to the different phyla Proteobacteria and Bacteroidetes (see Figure 73B). Bacteroidetes are clearly separated from all other bacterial species along the second principal component indicating their very special membrane lipid composition containing large amounts of otherwise rare sphingolipids. Furthermore, Flavobacteria and Bacteroidetes show a good separation on class-level along the second principal component in the plot displayed in Figure 73D for Gram-negative species only. Although further trends are visible for separation of Fusobacteria and Alpha-Proteobacteria, no further distinct separation is visible along the first two principal components in Figure

73D indicating a large amount of heterogeneity among the different bacterial species and making it unlikely to find specific biomarkers at high taxonomic levels such as phylum, class and order for non-Bacteroidetes species.

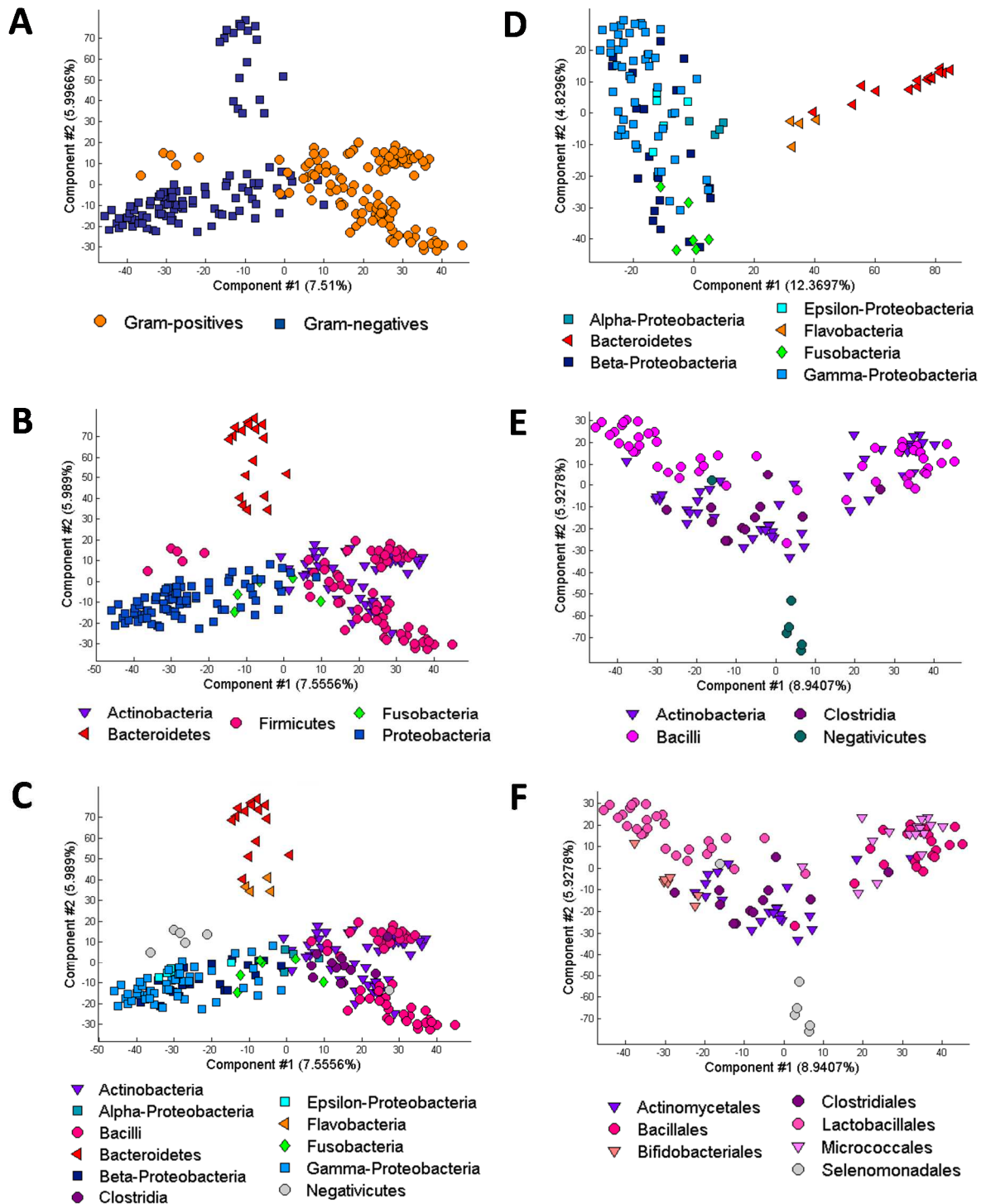


Figure 73. Principal components analysis plots from a dataset comprising 217 bacterial species different phylogenetic levels. A) Gram-level, B) phylum-level, C) Class-level, D) Class-level, Gram-negative species only, E) Class-level, Gram-positive species only, F) Order-level, Gram-positive species only.

Unlike in case of Gram-negative bacteria, the separation among Gram-positive bacteria cannot be correlated with systematic spectral differences on phylum-level. The members of the Negativicutes class are outliers from the centre of the Gram-positive cluster and are located next to members of the Gram-negative phylum Proteobacteria, suggesting a markedly different lipid composition from their Gram-positive counterparts. In fact, although Negativicutes are a class of firmicute bacteria, due to their peculiar cell wall composition their Gram-staining behaviour is negative. Negativicutes are the only diderm members of the Firmicutes phylum and as such possess inner and outer membranes in the same way as Gram-negative bacteria. Nevertheless, although phylogenetic analysis based on DNA sequences reveals significant heterogeneity within the Firmicutes phylum, no close relationship was found between Negativicutes and Proteobacteria.(207) No further separation is visible when investigating Gram-positive classes only. However, if strains are labelled according to bacterial order (Figure 73F), more clear clustering is visible within the different phyla. The class of Bacilli separates into the two orders of Bacillales and Lactobacillales. Bacillales are forming a cluster with Micrococcales, all of which were observed to contain large amounts of saturated phospholipid species in their REIMS spectral profiles.

The clustering results based on REIMS spectral profiles do not exactly match those expected based on 16S rRNA sequences, but rather as shown earlier in this study follow phenotypic characteristics which are rather determined by the environmental niche a bacterium lives in. Significant heterogeneity is observed especially among the very large phyla Proteobacteria, Actinobacteria and Firmicutes. However, better clustering results are obtained on lower taxonomic levels (order, family, genus). These results suggest that a general identification algorithm based on a decision-tree of phylogenetic relationships between bacteria is not likely to yield reliable identification results.

The appearance of occasional outliers is expected as the strains used to build this dataset were clinical isolates identified only based on their MALDI-TOF-MS results on the Bruker BioTyper instrument and were not further validated using molecular methods. MALDI-TOF-MS was performed on bacterial samples that were prepared using the direct smear technique which further reduces the correct identification rates.(208)

4.3.3. Inter-platform comparison

An independent set of clinical isolates for blind identification tests was cultured on various different growth media (see Table 19 for details) and subsequently analysed using an Orbitrap Discovery or a Xevo G2-S Q-TOF instrument (the latter being equipped with a modified atmospheric interface). In all cases, a minimum of two different culturing media were used per species. Only enriched (blood containing) media were investigated in case of *N. gonorrhoeae* and *S. agalactiae* due to limitations in growth on other media. However, chocolate/non-chocolate and selective agars were included in this study.

Acquired spectra (full mass range) were classified based on a subset of nine different species of the previously generated model shown in Figure 69 which were all cultured on Columbia horse blood agar. Since a 1Da sampling interval proved sufficiently specific, these blind identification tests were performed at a common bin size of 1Da. This further compensates for the differences in mass resolution of the different instruments.

Table 19. Details of sample set analysed in the cross-platform comparison (Figure 74). Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Gram-stain	Family	Genus	Species	Growth media ^a
negative	Pseudomonadaceae	<i>Pseudomonas</i>	<i>aeruginosa</i>	CBA, BAB, CHOC, LB, MCC, ISO, MH, MH+B, TS
	Enterobacteriaceae	<i>Escherichia</i>	<i>coli</i>	BHI, BAB, CBA, CHOC, ISO, MH+B, MH, TS
	Moraxellaceae	<i>Moraxella</i>	<i>catarrhalis</i>	BHI, CBA, CHOC, MH, MH+B, TS
	Neisseriaceae	<i>Neisseria</i>	<i>gonorrhoeae</i>	CBA, CHOC, VCAT, ISON
	Xanthomonadaceae	<i>Stenotrophomonas</i>	<i>maltophilia</i>	BHI, CBA, CHOC, MH, MH+B, TS
positive	Staphylococcaceae	<i>Staphylococcus</i>	<i>aureus</i>	AZT, BAB, BHI, CBA, ISO, LB, MH, TS
	Streptococcaceae	<i>Streptococcus</i>	<i>agalactiae</i>	CBA, ISON
	Clostridiaceae	<i>Clostridium</i>	<i>difficile</i>	CBA, Braziers
	Micrococcaceae	<i>Micrococcus</i>	<i>luteus</i>	CHOC, BHI, TS, CBA, MH+B, BAB, MH

Abbreviations: LB = Luria-Bertani agar, BAB = Blood agar base, MCC = McConkey agar, CBA = Columbia horse blood agar, BHI = Brain-heart infusion agar, TS = Trypticase soy agar, CHOC = Chocolate agar, ISO = Iso-sensitest agar (for antimicrobial susceptibility testing), MH = Mueller-Hinton agar, MH+B = Mueller-Hinton agar containing horse blood, ISON = Iso-sensitest agar with blood and NAD, Braziers = Braziers *Clostridium difficile* selective agar, AZT = Aztreonam blood agar (selective agar for Gram-positives), VCAT = Chocolate agar including Vancomycin, Colistin, Amphotericin and Trimethoprim (selective for *N. gonorrhoeae*)

The described experiment yielded 100% correct blind identification results for strains acquired on the Orbitrap instrument (same mass analyser as training set) and 97.8% correct identification for the files acquired using the Xevo Q-TOF instrument (see confusion matrix in Table 20). The two misclassifications are tentatively attributed to the differences in noise structure between the different mass analysers used for training and test set. In addition, as the RMMC plot shown in Figure 74 demonstrates, very little separation originating from the

different instruments and or mass analysers is shown in case of supervised analysis and the inter-species variance clearly dominates the plot. This demonstrates that the spectral profiles obtained using REIMS are highly reproducible, even considering the conceptually different atmospheric pressure interfaces of the two different types of mass analysers.

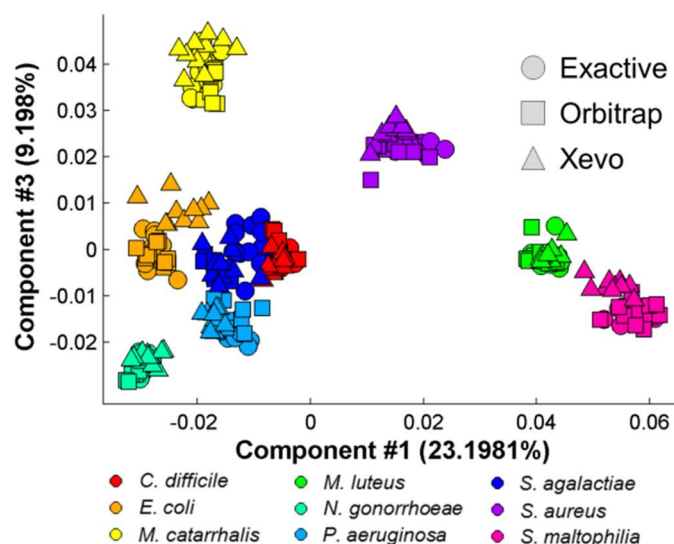


Figure 74. RMMC plot showing the combined data acquired on three different instruments (Exactive, Orbitrap, and Xevo). Bin size 1Da, m/z 200-2000. Confusion matrices for cross-platform blind identification results are given in Table 20. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Table 20. Confusion matrices for blind identification tests using different experimental platforms. Both Orbitrap and Xevo data was classified based on the model recorded on the Exactive instrument. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

Target Class, Exactive	<i>C. difficile</i>	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>E. coli</i>	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. catarrhalis</i>	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>M. luteus</i>	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>N. gonorrhoeae</i>	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>P. aeruginosa</i>	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
	<i>S. agalactiae</i>	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
	<i>S. aureus</i>	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
	<i>S. maltophilia</i>	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	8
		<i>C. difficile</i>																	
	<i>E. coli</i>																		
	<i>M. catarrhalis</i>																		
	<i>M. luteus</i>																		
	<i>N. gonorrhoeae</i>																		
	<i>P. aeruginosa</i>																		
	<i>S. agalactiae</i>																		
	<i>S. aureus</i>																		
	<i>S. maltophilia</i>																		10
		<i>C. difficile</i>	<i>E. coli</i>	<i>M. catarrhalis</i>	<i>M. luteus</i>	<i>N. gonorrhoeae</i>	<i>P. aeruginosa</i>	<i>S. agalactiae</i>	<i>S. aureus</i>	<i>S. maltophilia</i>	<i>C. difficile</i>	<i>E. coli</i>	<i>M. catarrhalis</i>	<i>M. luteus</i>	<i>N. gonorrhoeae</i>	<i>P. aeruginosa</i>	<i>S. agalactiae</i>	<i>S. aureus</i>	<i>S. maltophilia</i>
		Predicted Class, Xevo									Predicted Class, Orbitrap								

4.3.4. Comparison of spectral profiles for bacteria cultured on liquid and solid culturing medium

In routine clinical microbiology settings the vast majority of strains are isolated and cultured on solid agar-based culturing medium. Solid culturing medium has several advantages compared to liquid culturing media, of which the most important is the visual information that can be obtained by observing colony size and morphology, swarming behaviour, colony colour or discolouration in the agar surface by lysis of agar constituents or by excretion of coloured secondary metabolites. It also facilitates the detection of possible contaminants and enables standardised antibiotic sensitivity testing. Advantages of liquid culturing medium on the other hand include more homogenous nutrient distribution and better reproducibility of culturing conditions. The long-term goal of a REIMS-based microbial identification tool however is the development of a method that is applicable to the direct analysis of clinical samples such as blood, urine, sputum or pus – all of which more closely resemble liquid culturing conditions.

To assess the influence of the two different culturing modes, 15 bacterial species were cultured both in liquid (5mL volume in 25mL flask) and solid brain-heart-infusion medium (standard 9cm Petri-dishes; both commercially available from E&O Laboratories Ltd.). The detailed sample set is listed in Table 21. All strains were incubated for 48 hours at 37°C. Liquid cultures were grown on a static shelf due to restrictions imposed by working within a clinical microbiology laboratory handling patient samples. The lids of the liquid culturing flasks were only loosely attached to avoid contaminations, but enable gas exchange by diffusion. *B. fragilis* and *C. difficile* were grown under anaerobic conditions, *S. pneumoniae* and *S. agalactiae* as well as *E. faecium* and *E. faecalis* were grown aerobically in an atmosphere containing 5% CO₂. All other isolates were cultured under standard aerobic conditions. Although 10 individual strains were prepared per bacterial species in each liquid and solid medium, the actual numbers of isolates per species deviate from this number. This is due to exclusion of contaminated specimens or a lack in growth or availability. In case of *S. pneumoniae* no growth was observed on all ten solid cultures while good growth was observed in liquid culture medium. Numbers of isolates of >10 include back-up strains which were cultured and analysed to replace other strains that fail to achieve necessary quality.

Table 21. Sample set compiled to investigate the effect of liquid and solid culturing medium.

Gram-stain	Family	Genus	Species	number of strains		
				liquid culture	solid culture	
negative	Pseudomonadaceae	<i>Pseudomonas</i>	<i>aeruginosa</i>	9	10	
		<i>Moraxella</i>	<i>catarrhalis</i>	9	9	
	Bacteroidaceae	<i>Bacteroides</i>	<i>fragilis</i>	5	9	
	Enterobacteriaceae	<i>Escherichia</i>	<i>coli</i>	9	11	
		<i>Enterobacter</i>	<i>cloacae</i>	9	10	
		<i>Klebsiella</i>	<i>pneumoniae</i>	10	9	
		<i>Proteus</i>	<i>mirabilis</i>	11	10	
positive	Staphylococcaceae	<i>Staphylococcus</i>	<i>aureus</i>	10	11	
			<i>hominis</i>	6	10	
			<i>epidermidis</i>	9	11	
	Streptococcaceae	<i>Streptococcus</i>	<i>agalactiae</i>	7	10	
			<i>pneumoniae</i>	6	-	
			<i>Enterococcus</i>	<i>faecalis</i>	10	10
			<i>Enterococcus</i>	<i>faecium</i>	10	13
	Clostridiaceae	<i>Clostridium</i>	<i>difficile</i>	10	9	

Direct analysis of liquid cultures resulted in poor spectral quality due to the restrictions imposed on evaporation temperature and the dilution of bacterial cells by the culturing medium. Thus, bacterial cultures were centrifuged at 12500g for 10 minutes prior to analysis and the supernatant discarded. No further sample preparation was performed and the cell pellets were directly subjected to REIMS analysis. The observed spectral profiles were visually assessed for major spectral differences and subsequently subjected to multivariate statistical analysis.

For about half of the bacterial species analysed (*P. aeruginosa*, all members of the Enterobacteriaceae family, *Moraxella catarrhalis*, and *Streptococcus agalactiae*), no systematic variation in the spectral profiles associated with the culturing mode was observed. However, in the rest of the cases clear differences could be observed originating from differential relative signal intensity of certain phospholipid species. Figure 75 and Figure 76 display the two different cases of largely identical spectral profiles in case of *P. aeruginosa* and major changes to spectral appearance in case of *Staphylococcus aureus*.

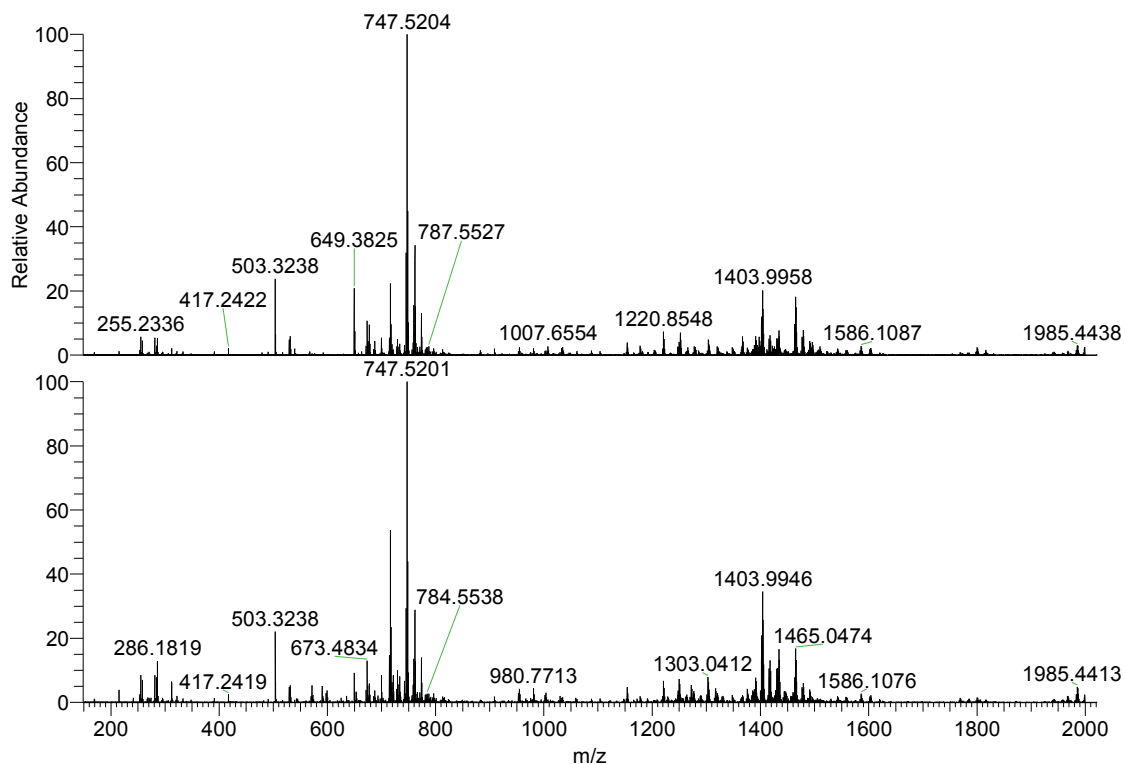


Figure 75. REIMS spectral profiles obtained from a clinical isolate of *P. aeruginosa* grown in solid (top) and liquid (bottom) culturing medium.

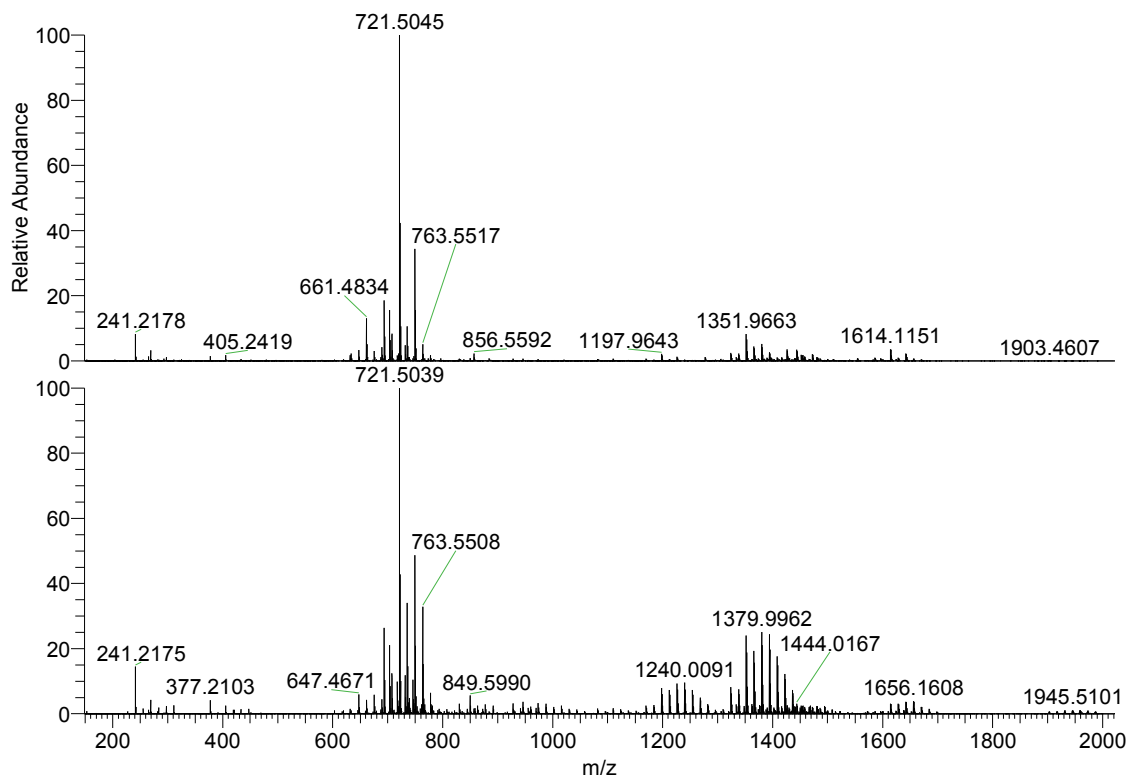


Figure 76. REIMS spectral profiles obtained from *Staphylococcus aureus* ATCC 6538 grown in solid (top) and liquid (bottom) culturing medium.

In case of all *Staphylococcus* species, liquid culturing conditions were often observed to result in a relative increase of saturated phosphatidylglycerol species of longer chain lengths and an increase in relative intensity of cardiolipins. A very similar behaviour was observed when *S. aureus* strains were grown on solid medium under anaerobic conditions (see Characterisation of spectral reproducibility, Figure 45). This suggests the possibility of reduced oxygen availability in liquid culture (especially considering that culture was performed on a static shelf) as origin of the spectral changes rather than the actual fact that culturing was performed in liquid medium.

However, no general assumptions can be deduced regarding the bacterial phenotypes expressed under different culturing conditions since each bacterial species showed characteristic spectral differences. In addition, biological variations were also observed between different strains of the same bacterial species. Thus, principal component analysis was used to further investigate whether the extent of changes in spectral appearance is exceeding the spectral variance between different bacterial species.

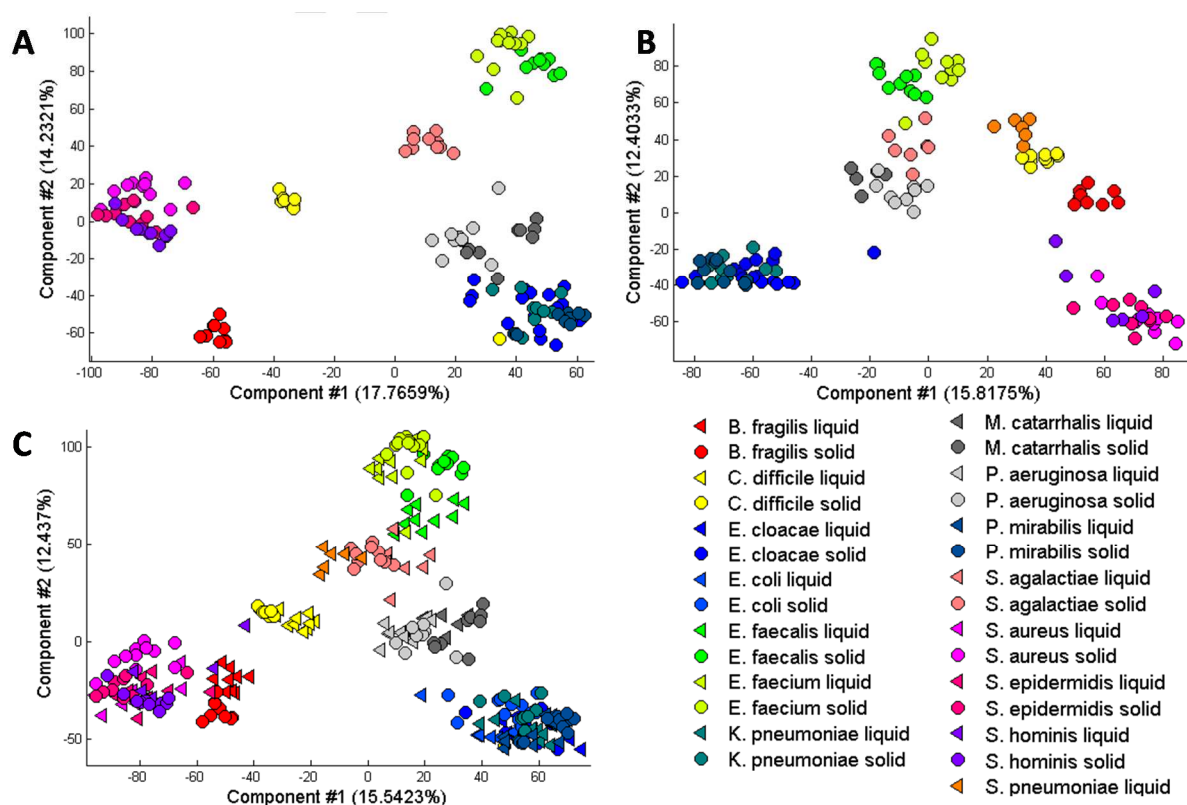


Figure 77. PCA plots of datasets comprising A) only solid cultures, B) only liquid cultures and C) both.

The resulting PCA score plots are shown in Figure 77 for bacteria cultured on solid medium (A), liquid medium (B) and both datasets combined (C). Differences in the score plots can be observed for liquid and solid cultures regarding the compactness of groups and their distance

from each other. In addition, slight differences can also be observed in relative positioning of the groups. This is particularly clear in case of data points corresponding to *C. difficile*, *M. catarrhalis* and *P. aeruginosa*. *C. difficile* datapoints observed from solid culture form a very distinctive cluster located equally distant from all other datapoints (apart from single isolate which clusters with Enterobacteriaceae and is therefore most likely a contaminant) while in liquid culture they are closely located to the *Streptococcus* spp. data points. Similarly, *M. catarrhalis* and *P. aeruginosa* (both Pseudomonadaceae) are located close to Gram-negative Enterobacteriaceae when grown on solid culturing medium while they are closer located to Streptococcaceae when grown on liquid medium. However, overall the similarity between both PCA plots is high and suggests no significant changes of spectral profiles but conserved spectral features that differ in relative intensities. This is further demonstrated by the PCA plot of the resulting combined data set which shows an overall very similar appearance to the previous two PCA plots and clearly shows that the same closely related bacterial species continue to cluster together while unrelated strains show greater distance to each other. Figure 77C suggests that spectral differences are negligible in case of *P. aeruginosa*, *M. catarrhalis* and members of the Enterobacteriaceae family while there are systematic intra-species changes in case of *C. difficile*, *B. fragilis* and *Enterococcus* and *Staphylococcus* spp. which while still clustering close to each other show two distinct data clusters for those strains grown on solid and on liquid culturing medium.

These results indicate that there are effects on the spectral profiles if culturing is performed in liquid culturing medium rather than on solid medium. As most biological fluids and tissues have to be regarded as liquid rather than solid medium, it is thus important to include bacteria grown on liquid medium into the database for future direct on-sample applications.

4.3.5. Testing of sensitivity and specificity at sub-species level

4.3.5.1. Strain-level specificity

A set of seven standard *E. coli* laboratory strains (NCM3722, MG1655, MC1000, MC4100, DH5 α , C600, and OP50) was analysed in order to test whether the specificity of the presented method goes beyond genus- and species-level and allows the identification of bacteria at strain-level. Strains NCM3722, MG1655, MC1000, MC4100, DH5 α and C600 are all derived from the K-12 parent strain, whereas OP50 is a B-strain derivative. The effect of culture conditions was further investigated by forming a dataset including the strains grown to

different culture ages (1-4 days, on Luria-Bertani agar) and on five different solid culture media (Luria-Bertani agar, blood agar base, brain-heart infusion agar, trypticase soy agar, MacConkey agar). Data acquisition was performed using the linear ion trap of the LTQ Orbitrap Discovery instrument working at unit resolution. Data was TIC normalised and reduced to the mass range of m/z 600-900 which was found to improve identification performance. While no significant changes in spectra could be observed for different culture ages between 1-4 days, changes in relative phospholipid signal ratios were observed for different culturing media.

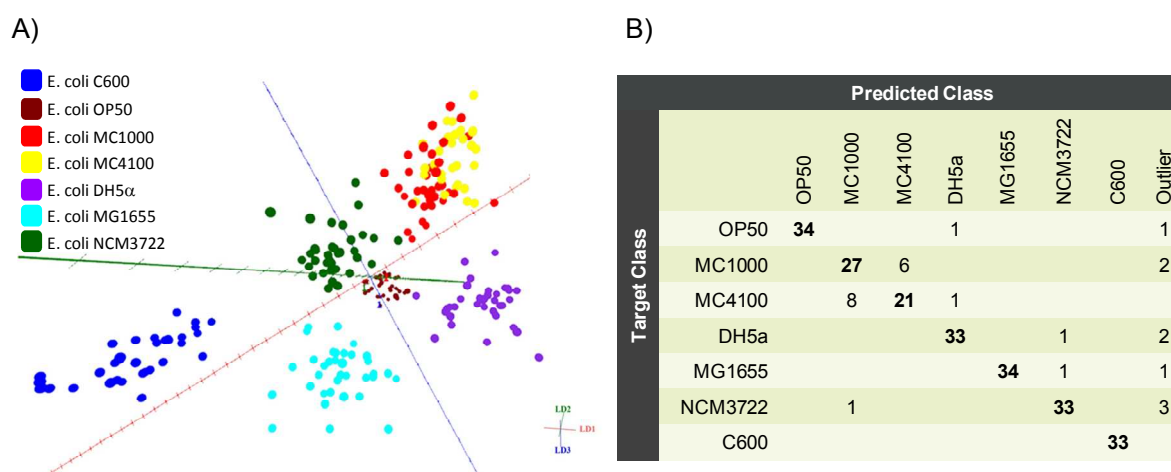


Figure 78. A) PCA-LDA model of seven different *E. coli* strains. Data was TIC normalised and reduced to m/z 600-900. B) Cross-validation results CV=87.3%. Data points classified as outliers if located outside 3x the standard deviation from the group mean.

Application of a supervised PCA-LDA classification algorithm leads to 87.3% accurate identification results using leave-one-out cross-validation (LDA plot shown in Figure 78A). As Figure 78 shows, OP50 as only B-strain is clearly separated from the K-12 derivatives along the first principal component. The RMMC algorithm in this case was found to be inferior to the PCA-LDA approach, only yielding 85.3% correct cross-validation results. This observation (together with the reduction of mass range) suggests that different identification algorithms used at different levels of identification (subspecies-level vs. higher taxonomical levels) might further improve the identification performance. Figure 78B shows the confusion matrix of the cross-validation results obtained from the model shown in Figure 78A. The majority of misclassifications occurred between two strains (MC1000 and MC4100). Therefore, the 5 of 7 analysed strains self-identify largely correctly despite being cultured on different media and for different durations. This suggests that REIMS generally can deliver

strain-level specificity, with likely better identification outcomes if culturing was performed under standardised conditions.

4.3.5.2. Serotyping of *Streptococcus pneumoniae*

Streptococcus pneumoniae is a Gram-positive bacterium that causes diseases such as bacteraemia, meningitis and infections of the respiratory tract, especially in young children and the elderly with most cases occurring in the developing world. *Streptococcus pneumoniae* cells are covered with layers of polysaccharides forming a capsule which acts as an essential factor in virulence. 93 distinct pneumococcal serotypes have been identified, however, only a comparably small number of these serotypes are accounting for most cases of infection. Vaccines are available for this pathogen, but for guiding vaccine composition development, epidemiological information on serotypes causing infectious outbreaks is critically important.(209) Identification of *S. pneumoniae* serotypes is most commonly performed using the Quellung reaction which involves adding an antibody solution to a broth of *S. pneumoniae* and observing a positive reaction indicated by “swelling” of the bacterial cells which is observed using a light microscope. This test is laborious and time-consuming and consists of a range of subsequent individual tests until a serotype is unambiguously identified. Usually antibody solutions are added in mixtures of several antibodies at a time to reduce amount of tests necessary.(210) Molecular serotyping methods involving PCR tests were found to be more cost-effective, however, are still comparably laborious and time-consuming.(211) Therefore, a straightforward way to distinguish between different pneumococcal serotypes without the need to introduce further sample processing steps besides those needed for species-level identification would have a huge impact on daily microbiological practice.

A collaboration was established with Dr. Irene Burkhardt, a consultant microbiologist from the Universitätsklinikum Heidelberg, Germany who kindly agreed to send a set of each 10 strains of three different serotypes for analysis using REIMS. Strains were grown on horse blood agar for 48hrs under an atmosphere containing 5% CO₂ and subsequently analysed using a Thermo Exactive instrument. Results of multivariate statistical methods applied to the dataset are shown in Figure 79. Trends toward separation can be seen in the PCA plot and better separation is visible for the supervised PLS analysis. Cross-validation results for the obtained dataset give 77.8% correct classification into the three different serotypes. The confusion matrix shown in Figure 79 shows that while there are 3 misclassifications between

serotype 19A and each serotype 14A and 3, there are no misclassifications observed between serotypes 14 and 3. These results suggest that there is some information concerning serotypes in the lipid-based REIMS spectral data which in turn suggests that the bacterium makes slight modifications in membrane lipid composition to accommodate different polysaccharide moieties. Targeted improvements in sampling and data analysis procedures as well as increasing the size of the dataset might lead to improvements on classification. Further experiments using a larger number of strains and serotypes will be required to fully assess specificity and investigate the origin of the larger difference between serotypes 14 and 3 in comparison to serotype 19A. Alternative chromatography-based profiling of the saccharide moieties and correlation with gene/protein analyses can additionally be used to further assess this phenomenon.

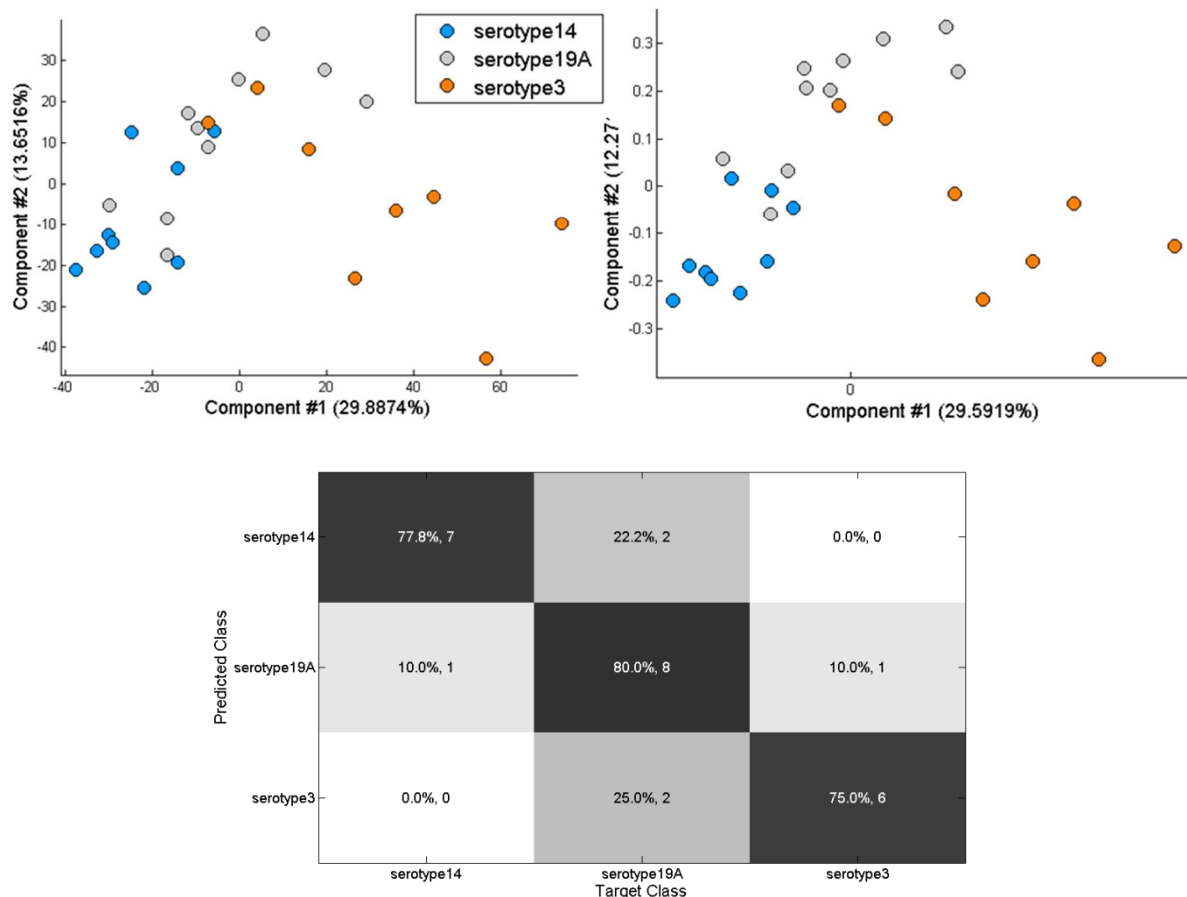


Figure 79. PCA (top left) and PLS (top right) plot for three different *S. pneumoniae* serotypes. Cross-validation results are 77.8% (bottom).

4.3.5.3. Ribotyping of *C. difficile*

Clostridium difficile is a Gram-positive anaerobic bacterium and infections are often nosocomially acquired after broad-band antibiotic treatment, which allows excessive growth

of this hardy spore-forming species. More than 300 PCR ribotypes are recognised, however, in routine clinical microbiology, severe *C. difficile* outbreaks are often associated with certain ribotypes such as ribotype 027 or 078 which are thought to be particularly pathogenic. Therefore it is especially interesting for clinical microbiology laboratories to establish whether an infection was acquired during the hospital stay (nosocomial – all patients would be expected to have been infected by a strain of the same ribotype) or whether the infection was caused by a strain acquired before entering the hospital.

Ribotyping of *C. difficile* is routinely performed by isolating *C. difficile* on species-specific media (such as Braziers medium), DNA extraction and subsequently performing PCR amplification of the 16S-23S intergenic spacer region to determine the ribotype pattern. This process is very time consuming and labour intensive, thus, the specificity of the REIMS technique was investigated for this particular problem.

Each 10 strains of three different ribotypes of *C. difficile* were cultured on Columbia horse blood agar for 24hrs under anaerobic conditions. The ribotypes included 002 and 014 which are regarded to be less pathogenic and the more pathogenic ribotype 078. Strains were analysed on the Thermo Exactive instrument under standard conditions in a randomised fashion. *C. difficile* strains were further randomised with unrelated bacterial species. Data was subsequently subjected to multivariate statistical analysis.

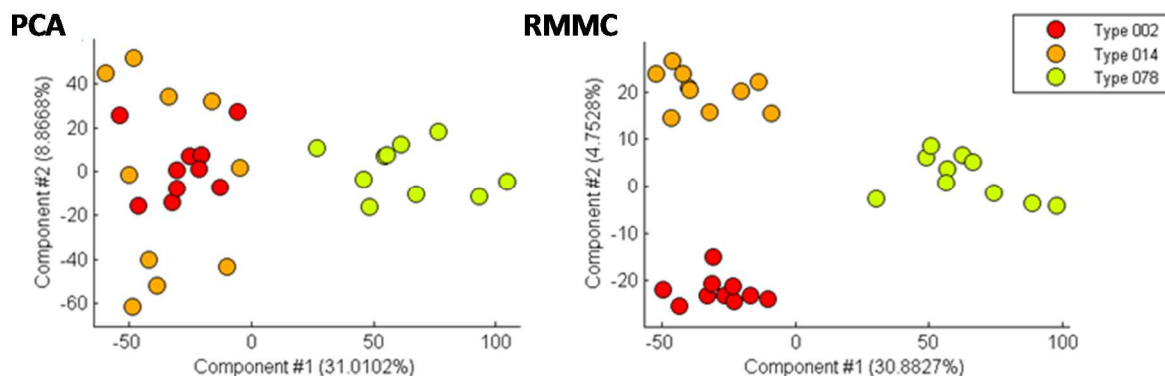


Figure 80. PCA (left) and RMMC (right) multivariate statistical plots for three different ribotypes of *C. difficile*. Overall classification accuracy is 90%.

Both in supervised and unsupervised analysis, clear separation trends were observed between the less (002 and 014) and the more pathogenic ribotype (078) along the first principal component (see Figure 80). When performing cross-validations of the supervised RMMC model, no misclassifications are obtained between more and less pathogenic strains, but the only misclassifications observed are between the less pathogenic 002 and 014 ribotypes.

Overall cross-validation accuracy is 90%. These results suggest that *C. difficile* does show different lipid profiles, especially between those ribotypes that show differences with regard to pathogenicity. However, further experiments including larger numbers of ribotypes are required to validate this separation.

4.3.5.4. Antibiotic resistance in *K. pneumoniae*

Antibiotic resistance of microorganisms is a global problem of increasing significance that often complicates treatment of infections. The protein profiles that are acquired during routine MALDI-TOF-MS analysis do not deliver sufficient information on the antibiotic resistance pattern.(212) This means that although MALDI-TOF-MS allows the rapid identification of an organism in clinical settings, the determination of antibiotic sensitivity still has to be performed separately using conventional culturing-based methods. For this purpose, bacteria are streaked in a film onto one to three agar plates and an array of 5-15 antibiotic discs will be placed on top of the agar surface. Antibiotic susceptibility is determined based on the size of the inhibition zone around the antibiotic disk. This process is expensive in terms of personnel and consumables and usually requires incubation for another 24-48hrs based on the organism. Numerous MALDI-based techniques were developed over the last 15 years,(212) and the detection of β -lactamase activity is arguably the most developed method. This method is applicable to β -lactam antibiotics such as ampicillin, piperacillin, cefotaxime, ceftazidime, ertapenem, imipenem, and meropenem. The bacterium is placed into liquid growth medium containing low concentrations of the antibiotic. Subsequently, not the protein profile of the bacterium is analysed, but the bacterial hydrolysis product in the culture supernatant that is produced from a resistant organism.(213) Usually 2-6 hours of incubation are sufficient to detect the drug metabolite due to the high turn-over rate of β -lactamases which results in resistance information being available the same day as the MALDI organism identity. This methodology did not gain approval for clinical use yet and requires the MALDI instrument to be used under non-routine settings, does require a non-routinely performed liquid culturing step and only works for a single antibiotic (or group of related antibiotics) at a time.

The possible application of a REIMS-based test to determine antibiotic sensitivity patterns was assessed in case of carbapenemase-producing *Klebsiella pneumoniae*, an emerging group of pathogens that show resistance to carbapenem antibiotics. Carbapenems are members of the β -lactam class and frequently used to treat infections known or suspected to be caused by

multidrug-resistant bacteria. Although it was originally detected in *K. pneumoniae* and thought to be genus-specific, production of carbapenemase is an important resistance mechanism for an increasing number of Gram-negative bacteria and poses significant challenges to clinicians due to limited number of antibiotics available for treatment of affected patients. Although carbapenemase-production is not the only mechanism inducing carbapenem-resistance, it is of particular interest, since it is commonly missed by routine antibiotic sensitivity testing and it may induce resistance to virtually all β -lactam antibiotics and thus it is associated with high mortality rates.(214)

A test to determine carbapenemase production using MALDI-TOF-MS was demonstrated to identify producers of NDM-1, VIM-1, KPC, OXA-48 and OXA-162 carbapenemases in a quick and cost-effective test. Nevertheless, this test requires additional sample preparation steps and 2h incubation with the antibiotic reagent.(215) To assess the feasibility of determining antibiotic sensitivity using REIMS, sampling was performed on 22 strains of *K. pneumoniae* (sensitive n=10, resistant n=12) grown for 18hrs on ISO sensitest agar (Oxoid, Basingstoke, UK) with a single disk of 10mcg ertapenem (ETP 10) close to the rim of the Petri-dish (approx. 2cm from the rim towards center of the dish). Two sampling points were taken from each plate, one at the border of bacterial growth towards the antibiotic disk, and a second one at the opposite end of the Petri-dish. Carbapenem-resistant strains were molecularly analysed to identify the resistance mechanism (2x*Klebsiella pneumoniae* carbapenemase (KPC), 5xNDM metallo beta lactamase, 3xOXA-48 carbapenemase, 2xVIM metallo beta lactamase) as part of a project performed at the Department of Microbiology (Charing Cross Hospital, London, UK).

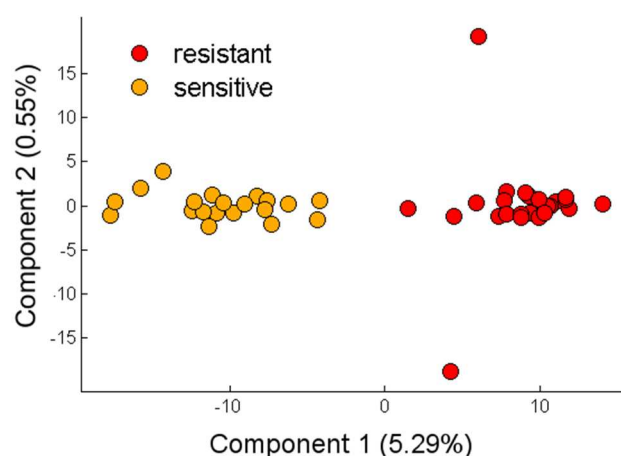


Figure 81. RMMC plot of carbapenem-resistant (red) and sensitive (orange) *Klebsiella pneumoniae*. Cross-validation results 75% accuracy.

No signal corresponding to the antibiotic ertapenem or its metabolites could be detected in the REIMS spectra, thus REIMS spectral profiles were subjected to multivariate statistical tools for pattern recognition analysis. While no separation could be observed in PCA analysis, using supervised RMMC analysis leads to separated datagroups for datapoints corresponding to resistant and sensitive *K. pneumoniae* (see Figure 81). When performing leave-strain-out cross-validation, an overall classification accuracy of 75.0% was obtained. Misclassifications were randomly distributed over sensitive and resistant strains. For resistant strains, misclassifications could not be associated with a particular resistance mechanism. Although overall accuracy was not very good, this might be improved using larger sample sets as is commonly observed for pattern recognition algorithms.(216) However, the current results suggest that REIMS spectral profiles contain some degree of information on antibiotic sensitivity and might be a useful tool to help clinicians decide on suitable antibiotic treatment more quickly than using standard clinical microbiology workflows by providing indicative treatment suggestions without any additional hands-on work required. This would allow for the right treatment to be given while more time-consuming follow-up tests could be performed.

4.3.6. Analysis of yeasts

Clinical microbiologists are increasingly being confronted with infections caused by fungal pathogens, especially in immuno-compromised patients such as those suffering from HIV or cystic fibrosis (CF).

Identification of yeasts using MALDI-TOF-MS requires the pre-treatment of the yeast sample prior to mass spectrometric analysis in order to give reliable identification performances (score >2.0). While the recommended sample pre-treatment for MALDI-TOF-MS comprises the complete extraction of the fungal material using formic acid and acetonitrile,(63) intact yeast species can directly be analysed by REIMS without any modifications in experimental setup or analysis workflow and without compromising spectral intensity. 209 strains of 16 different genera of pathogenic yeasts have been recorded until today, however, statistically significant numbers of strains per species (>10 strains) have only been recorded for the following five *Candida* species: *C. albicans*, *C. glabrata*, *C. parapsilosis*, *C. lusitaniae* and *C. tropicalis*.

Figure 82 shows the PCA plot of a dataset comprising 298 strains of bacteria (87 species, both Gram-negative and Gram-positive) and 209 strains of yeasts (16 species). No overlap between bacteria and fungi was observed in the PCA data space and cross-validation shows 100% correct classification using a leave-genus-out approach for the supervised RMMC model. This was attributed to the markedly different phospholipid composition between bacteria and fungi. Whereas REIMS spectra obtained from bacteria mostly feature PGs, PEs and low abundance PAs, the spectra obtained from yeasts mainly feature highly abundant PAs, PEs and phosphatidylinositols (PIs), with the latter being very rare among bacteria. Of the >228 bacterial species recorded until the end of this project, PIs have only been found in some members of the Actinobacteria phylum such as *Corynebacterium* spp., *Nocardia* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Actinomyces* spp., *Rothia* spp., *Propionibacterium* spp..

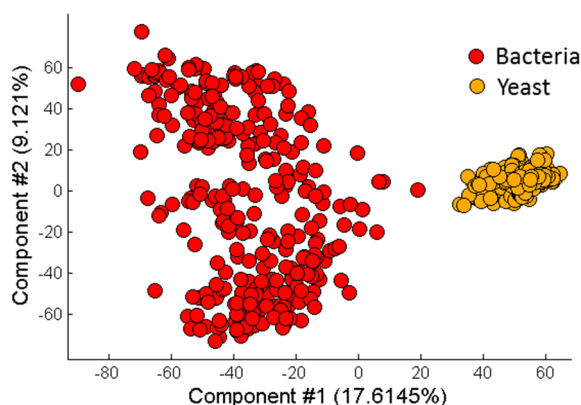


Figure 82. PCA plot generated from a dataset comprising a variety of bacteria and yeast. Reprinted with permission from Ref. (142). Copyright 2014 American Chemical Society.

As it is shown by the PCA plot in Figure 82, the data cloud representing bacteria exhibits a much larger extension than the one representing yeasts. This in turn means that the phospholipid constituents of bacteria display a much wider variety and variability than for yeasts. Indeed, REIMS spectral profiles of bacteria are differing extensively, whereas the REIMS spectra of yeasts, similarly to mammalian tissue spectra, have an overall very conserved appearance with observed differences largely arising from different phospholipid ratios rather than lipid species being present or completely absent.

To determine whether REIMS is able to distinguish yeast species despite the apparently smaller spectral variability, five different clinically relevant *Candida* species were analysed. Figure 83 shows the separation of the resulting data groups both in supervised and unsupervised analysis. *C. glabrata* is clearly separated from the four other species along the

first principal component while *C. lusitaniae*, *C. parapsilosis*, *C. tropicalis* and *C. albicans* are largely separated along the second principal component. This separation into two groups is tentatively attributed to the fact that *C. glabrata* was found to belong to a separate phylogenetic clade than the other *Candida* species based on DNA sequences.(135) Two misclassifications were observed when performing leave-one-out cross-validation resulting in 98.8 % correct classification.

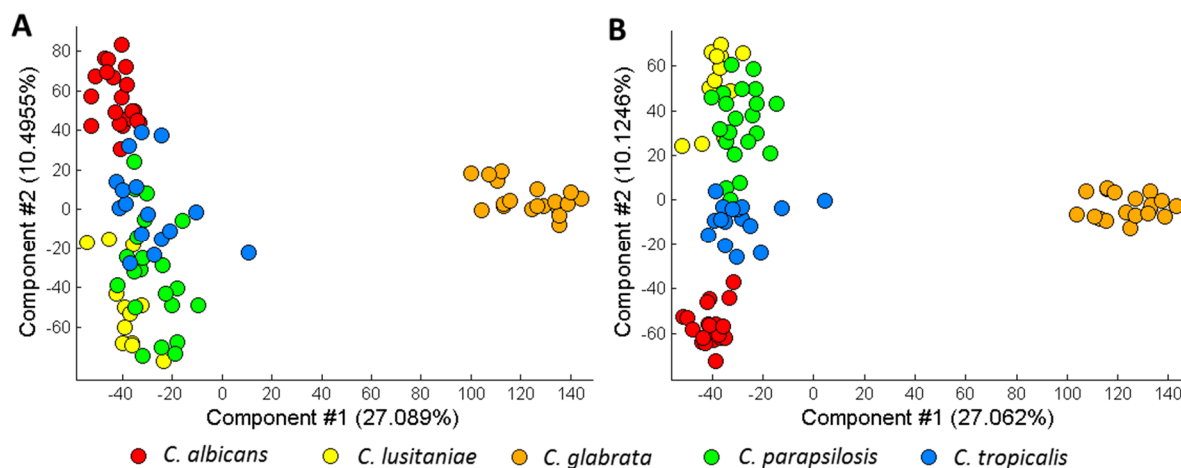


Figure 83. A) PCA and B) RMMC plot of *C. albicans* (n=20), *C. glabrata* (n=19), *C. lusitaniae* (n=12), *C. parapsilosis* (n=19) and *C. tropicalis* (n=16). CV = 98.8 %. Reprinted with permission from Ref. (142).

Copyright 2014 American Chemical Society.

The clear separation of *Candida glabrata* from all the other *Candida* spp. cannot be explained by variation of a single molecular species but is a result of an overall different spectral profile. This can clearly be seen in the ANOVA plot displayed below (Figure 84). However, certain molecular species show a significantly different distribution, such as the PI(36:1) species at m/z 863.5655 or the PA(36:2) species at m/z 699.4991 that are both highly abundant within *Candida glabrata* while being a minor spectral component in the other *Candida* species. However, it will need the addition of further *Candida* species to the model to determine whether these are generally specific features for *C. glabrata* and whether the separation observed does indeed follow the clustering observed by Butler *et al.* using DNA sequencing.(135)

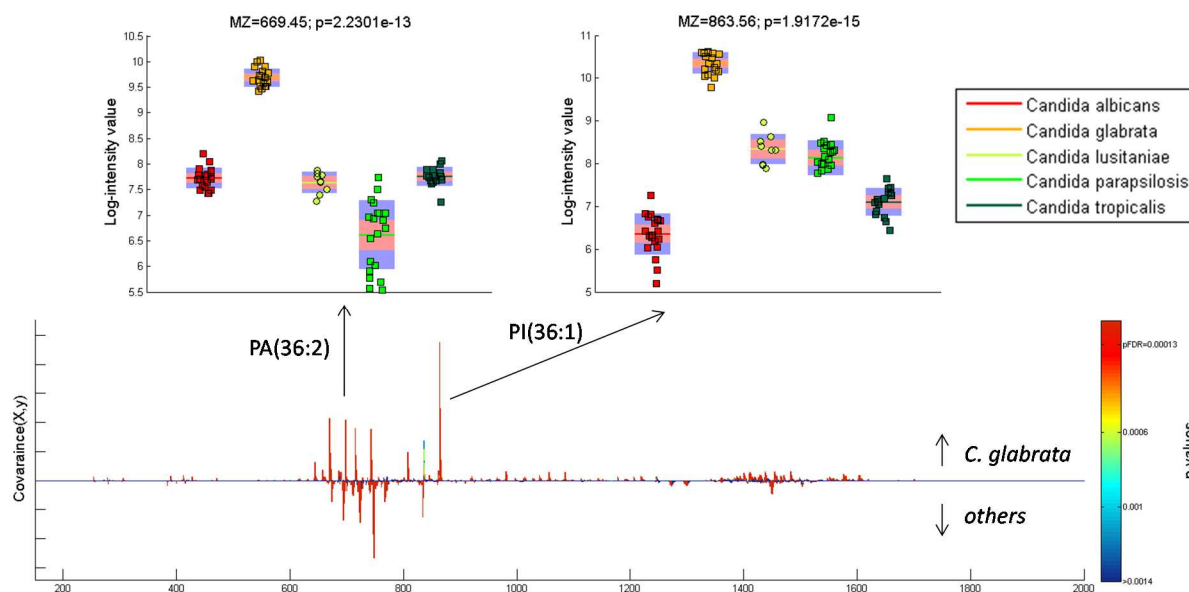


Figure 84. ANOVA plot (threshold $p=0.05$) for the model shown in Figure 83 comparing *C. glabrata* against the remaining 4 *Candida* species. Peak statistics plots are shown for two molecular species specific for *C. glabrata*.

4.3.7. Analysis of bacterial mixtures

While some clinical sample types are expected to contain a single causative pathogen at high concentrations, other clinical specimens are expected to contain a mixture of different bacteria either belonging to commensal microflora or pathogenic species. Specimens of the latter type include faeces and mucosal swabs. Thus, for direct on sample applications it is important to analyse bacterial communities by REIMS. As an initial experiment, binary mixtures of bacteria were created at different ratios and analysed by REIMS. Bacterial biomass was cultured for 48hrs on Columbia horse blood agar and subsequently scraped off using a 10 μ L loop and collected in a microcentrifuge tube for *Pseudomonas aeruginosa*, *Bacteroides fragilis*, *Staphylococcus aureus* and *Escherichia coli*.

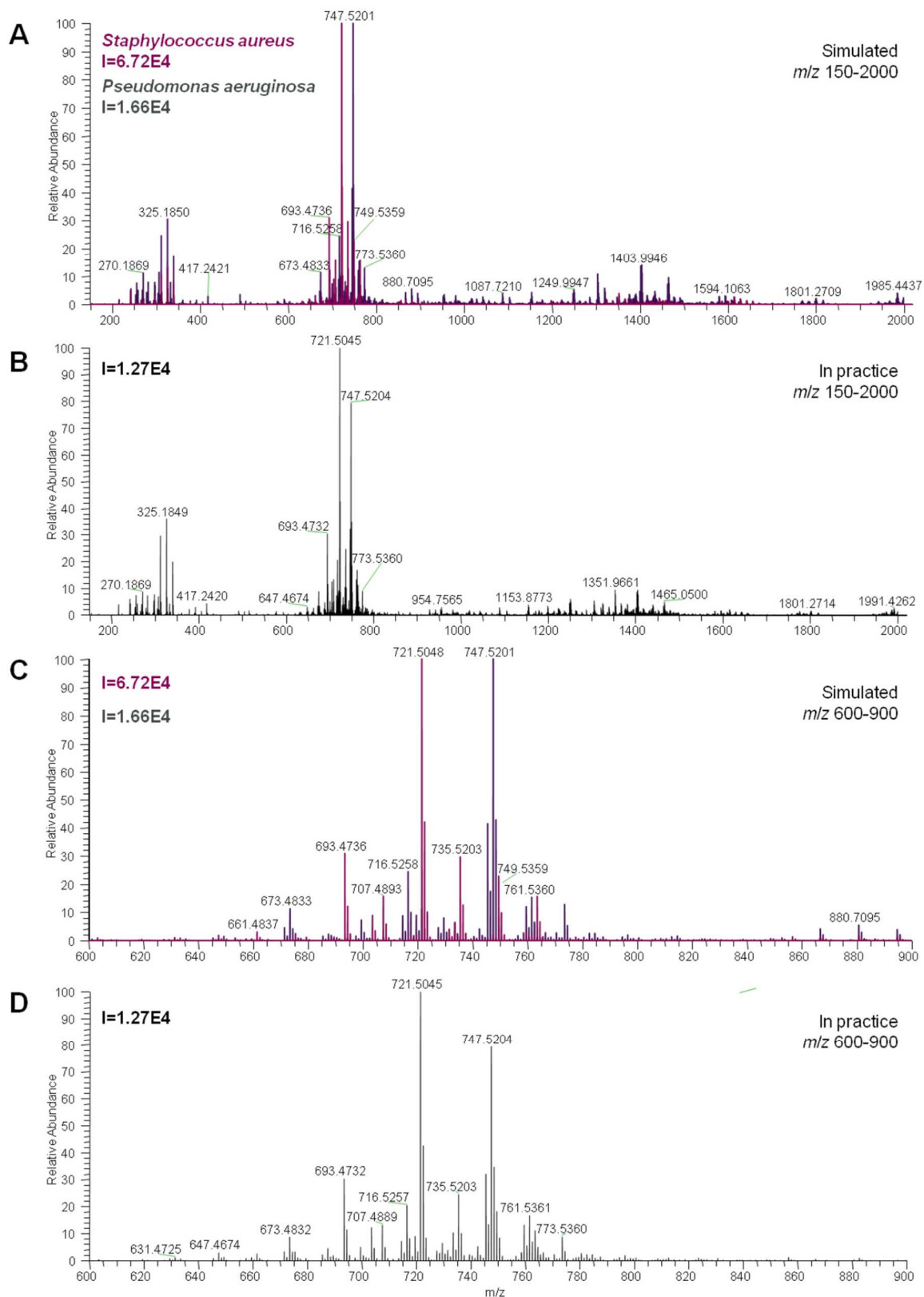


Figure 85. Comparison of spectral profiles obtained for binary bacterial mixtures of *Staphylococcus aureus* and *Pseudomonas aeruginosa*. **A** and **C**) Simulated REIMS spectrum of a binary mixture (1:1 w/w) assuming identical ionisation, **B** and **D**) REIMS spectrum of a binary mixture of both bacteria. **A** and **B**) show spectra over the full mass range m/z 150-2000, while **C** and **D**) show mass range m/z 600-900 only.

Using an analytical balance, bacteria were weighed into a fresh microcentrifuge tube in ratios of 1:5, 1:3, 1:1, 3:1, and 5:1 (weight:weight) and homogenised by stirring with a loop. Artificially generated mixtures from bacteria grown on separate plates do not resemble properties of microbial communities where presence of other species of bacteria might alter the production of metabolites (such as would be expected for rhamnolipids and quorum-sensing molecules in case of *P. aeruginosa*). However, this method of growing pure bacterial species on separate plates, collecting biomass and subsequent mixing was chosen as it allows most controlled generation of mixture ratios.

Results obtained by the analysis of the 1:1 mixture of *S. aureus* and *P. aeruginosa* can be seen in Figure 85. Figure 85A shows a simulated mass spectrum of a 1:1 mixture of *S. aureus* (purple) and *P. aeruginosa* (dark grey) in which it was assumed that the ionisation response factor for both bacterial species is identical. Figure 85B shows the REIMS spectrum of a binary mixture of both bacterial species. The spectrum of the mixture shows peaks of both bacterial species and within each species retains the peak ratios as observed in REIMS spectra of a pure culture (within the tolerated biological variance observed in routine REIMS measurements). Figure 85C and D show the same data for the reduced mass range of m/z 600-900. Besides a group of peaks observed in pure *P. aeruginosa* between m/z 850-900 which is reduced in intensity in the mixture with *S. aureus*, the overall spectral appearance shows an additive behaviour. Peak ratios between *P. aeruginosa*-originating and *S. aureus*-originating spectral features are not 1:1 but rather show a ratio of 0.8:1. This was partly anticipated as *S. aureus* in general shows significantly better response to ionisation than *P. aeruginosa* (approx. factor of 2 to 4 difference in overall response factors). It is interesting to note that this difference in ionisation response seems partly eliminated by having both bacteria in a mixture. This raises the possibility that differences in ionisation response in case of *S. aureus* are not solely due to the fact that it has a thicker cell wall with less variety in its lipid composition, but instead might produce extracellular compounds that increase the response to ionisation. When producing a mixture of *S. aureus* with other bacteria this compound will get distributed among both species and thus increase ionisation of both.

A slightly more complex behaviour was observed in a mixture of *Bacteroides fragilis* and *Escherichia coli* (see Figure 86). While in case of pure *B. fragilis* the signal at m/z 590 only has a relative intensity of 80% compared to the base peak of the heavier sphingolipid at m/z 653, in case of mixtures with *E. coli* this changes to m/z 590 becoming the dominant spectral feature. This is tentatively associated with the nature of the m/z 590 ion; as a $[M+Cl]^-$ ion its

response might be affected by changes in chloride ion content which is a likely side effect when preparing mixtures of different bacteria. Besides the effect of changes in response of chlorinated to quasi-molecular ions, response to ionisation correlates well with the ratio of the mixture. In case of Figure 86C and D, spectral profile of *B. fragilis* are visible at about 35 and 20% for mixture ratios of 1:3 and 1:5, respectively. In Figure 86B (*B. fragilis* : *E. coli* (3:1), very little *E. coli* is detected (at about 10% relative intensity), however, the effect of changes associated with chloride adduct formation was still clearly visible. Low detection of *E. coli* in this sample was tentatively associated with poor homogeneity of the mixture.

REIMS spectral profiles acquired of binary mixtures suggest an additive nature of m/z features. This is an important prerequisite for identification of microorganisms in microbial communities and is paramount for the success of a taxon-specific marker approach as ion suppression does not seem to significantly alter REIMS profiles. Although overall signal abundance cannot be compared to that of pure cultures, relative abundances between spectral features seem conserved.

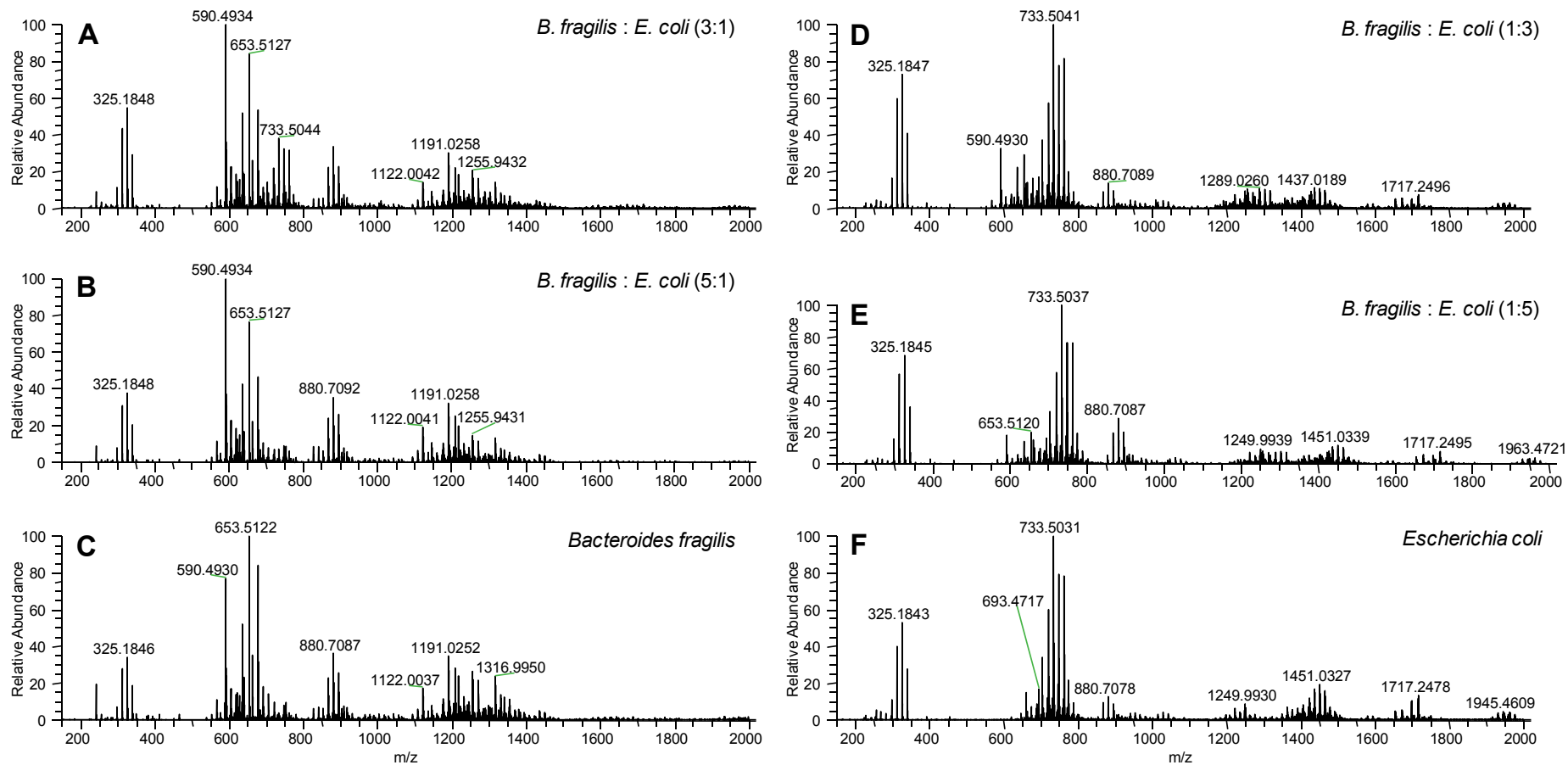


Figure 86. Full scan REIMS spectra obtained from different binary mixtures of *Escherichia coli* and *Bacteroides fragilis*. A) *B. fragilis* : *E. coli* (3:1), B) *B. fragilis* : *E. coli* (5:1), C) *Bacteroides fragilis* (100%), D) *B. fragilis* : *E. coli* (1:3), E) *B. fragilis* : *E. coli* (1:5), and F) *Escherichia coli* (100%).

5. Analysis of cell lines using REIMS

5.1. Introduction

The proposed REIMS-based microbial characterisation and identification tool and workflow allows for adaptation to the analysis of other unicellular forms of life such as cell lines. Cell lines are popular models to study various biochemical and disease processes *in vitro*. In case of cancer studies, cell lines provide a means to study cancer development and progression as well as the investigation of pathobiochemical processes as close to the human body as possible while still allowing free manipulation of experimental parameters. The most extensively characterised cell line collection up to date is the NCI60 cell line panel compiled by the National Cancer Institute as part of the In Vitro Cell Line Screening Project (IVCLSP). The panel comprises 60 human cancerous cell lines from nine different organs of origin, namely leukaemia, melanoma, cancers of the lung, colon, brain, ovary, breast, prostate, and kidney. A large variety of data is available for these cell lines including drug sensitivity patterns for more than 100,000 compounds and natural products, global protein and gene expression data and common mutations associated with cancer.(137,217-219) Large parts of the data are publicly available and accessible via the CellMiner database (<http://discover.nci.nih.gov/cellminer/home.do>).(150)

However, the associated metabolomics and lipidomics data is comparatively sparse. Although many publications are available investigating the lipid composition of certain types of cell lines using both chromatography-based(220,221) and shotgun mass spectrometry methods,(89,222-225) to our knowledge there is no work available characterising the lipidome of the NCI60 cell line panel as a whole using a single set of conditions. This represents a striking gap in the cancer-related biochemical data, as complex lipids are the main constituents of cell membranes and play important functional, structural, and metabolic roles by acting as signaling molecules (e.g., PI phosphates, ceramides, lysophosphatidic acids) or as precursors for secondary messengers (e.g., inositol triphosphate/diacylglycerol). Changes in the membrane lipid composition can regulate function of membrane proteins and affect cell signaling and transport mechanisms.

Mammalian tissue studies using REIMS revealed that the information content of the spectral profiles is not limited to the histological classification of tissues but contains additional information on the phenotype of the organism examined.(2,142) REIMS is therefore expected

to serve as a lipidomics-based shotgun phenotyping technique, however, the general sensitivity and specificity of the method to certain changes (as changes in protein expression patterns) in a biological system still needs to be assessed. For this purpose, the analysis of cell lines is significantly better suited than *ex-vivo* tissue specimens.

The traditional technique of choice for lipid profiling is liquid chromatography-mass spectrometry (LC-MS). However, even using state-of-the-art ultra-high performance liquid chromatography (UPLC-MS), run-time per sample is still in the range of 10-20 minutes and the analysis requires extensive sample preparation (homogenisation, extraction, etc.). Several mass spectrometric profiling methodologies have been developed in the recent past and the field has gained additional momentum with the introduction of ambient mass spectrometry in 2004.(89) Ambient mass spectrometric methods such as desorption electrospray ionisation mass spectrometry (DESI-MS), offer the capabilities to analyse samples in their native state, without any significant sample preparation steps.(82) These ambient lipid profiling technologies have recently been deployed in cancer studies to characterise the lipid composition of breast cancer compared to normal breast tissue(226), the identification of cancer metastasis within lymph nodes (124), colorectal cancer compared to normal mucosa(119), and brain cancer(227) among others. While these studies provided excellent basis for the development of novel diagnostic approaches, the molecular background of the differentiation between different histological classes remains unclear, due to the limited coverage of tissue metabolome/lipidome/proteome and lack of functional tests (e.g. gene silencing) in case of human samples. A complementary approach for studying the molecular background of histologically specific lipid profiles involves the use of cell lines, which solves the problems regarding sample availability and standardisation of sampling, lifts most of the ethical constraints and also allows functional testing including gene silencing or metabolic flux analysis.

Very promising results for high throughput metabolite profiling of cell lines was obtained using laser-assisted electrospray ionisation mass spectrometry (LAESI-MS). Combined with multivariate statistical tools this methodology was applied for the characterisation of human T-lymphotrophic virus (HTLV)-induced alterations of the biochemical processes of the host cells. Spectral differences were observed that could be directly correlated to one of either virus type studied (HTLV1 vs. HTLV3) and Tax protein expression (Tax1 vs. Tax3).(225) Furthermore, this approach was applied to the analysis of Kaposi's sarcoma-associated herpes virus (KSHV) in latently infected B-lymphocytes (BCBL-1).(224) In both cases, differences

in mass spectral profiles were found to be largely due to relative changes in phospholipid composition, highlighting the power of an ambient mass spectrometry-based shotgun-lipidomic profiling approach.

5.2. Assessment of robustness of REIMS spectral profiles

REIMS analysis has been applied to different types of biological samples and showed good spectral pattern reproducibility in case of human tissues and microorganisms including both bacteria and fungi.^(2,142) To combine these two application areas and to test the general hypothesis that REIMS spectral patterns are reproducible and sufficiently specific to differentiate between different human cancer cell lines, three different cell lines (HeLa – cervical adenocarcinoma, MES-SA – uterine sarcoma, and SNB-19 – glioblastoma) were analysed in an experiment designed to test spectral reproducibility. The experimental scheme is depicted in Figure 87 and accounts for variance introduced by different culture batches or passage numbers and for analytical variance introduced by multiple measurements. Replicates were randomly analysed over three analysis days in order to assess the analytical variance and robustness of a REIMS-based lipid profiling method. In addition, the influence of one freeze-thaw cycle on spectral variance was investigated and was found to be insignificant. A detailed sampling schedule can be found in Table 22.

Raw REIMS profiles of all three cell lines show a high degree of similarity (Figure 88). As analysis is performed in negative ion mode, the spectral content is similar to that described in earlier studies reporting REIMS profiles of various biological materials, featuring predominantly glycerophospholipid-type membrane lipid components such as phosphatidylethanolamines (PEs), phosphatidylinositols (PIs), phosphatidylglycerols (PGs), phosphatidic acids (PAs) and phosphatidylserines (PSs) as well as other complex lipids including ceramides and glycosylated ceramide species.^(2,128,133) Similarly to earlier observations, all detected ions carried a single negative charge, the vast majority by forming the quasi-molecular $[M-H]^-$ ion. In addition, thermal degradation product $[M-NH_3-H]^-$ was observed in case of PEs. Sphingolipid species were commonly detected as $[M+Cl]^-$ ions.

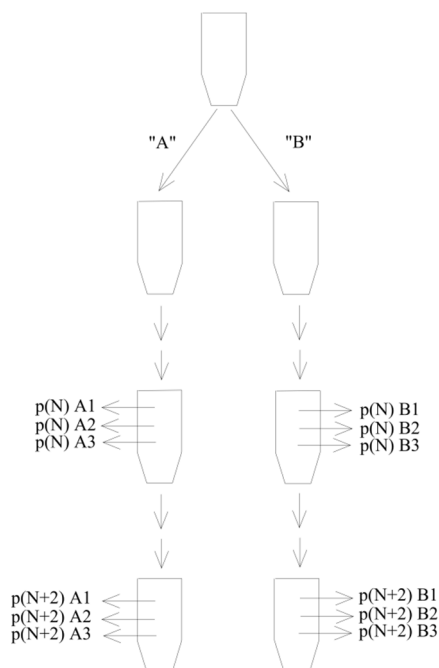


Figure 87. Experimental scheme to assess spectral reproducibility and self-identity of three different cell lines. p(N) indicates passage number, A and B different culture batches, while A/B1-3 indicates replicates of same culture batch. Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

Table 22. Cell lines and respective number of replicates recorded for reproducibility assessment sorted by measurement day. Number of technical replicates recorded are given in parantheses.

Measurement day 1	Measurement day 2	Measurement day 3
HeLa p4 A1 (6)	HeLa p4 A3 (3)*	HeLa p4 A2 (6)
HeLa p4 A3 (3)	HeLa p4 B1 (3)*	HeLa p4 B3 (6)
HeLa p4 B1 (3)	HeLa p4 B2 (6)	HeLa p6 A2 (6)
HeLa p6 A1 (3)	HeLa p6 A1 (3) *	HeLa p6 B2 (6)
HeLa p6 B1 (3)	HeLa p6 A3 (6)	
HeLa p6 B3 (6)	HeLa p6 B1 (3)*	
MES-SA p8 A1 (3)	MES-SA p8 A1 (3)*	MES-SA p8 A2 (6)
MES-SA p8 A3 (6)	MES-SA p8 B2 (6)	MES-SA p8 B3 (6)
MES-SA p8 B1 (3)	MES-SA p10 A2 (3)*	MES-SA p10 A1 (6)
MES-SA p10 A2 (3)	MES-SA p10 B1 (3)*	MES-SA p10 A3 (6)
MES-SA p10 B1 (3)		MES-SA p10 B2(5)
MES-SA p10 B3 (6)		
SNB p4 A1 (6)	SNB p4 A2 (3)*	SNB p4 A3 (6)
SNB p4 A2 (3)	SNB p4 B1 (6)	SNB p4 B3 (6)
SNB p4 B2 (3)	SNB p4 B2 (2)*	SNB p6 A2 (5)
SNB p6 A3 (3)	SNB p6 A1 (5)	SNB p6 B3 (4)
SNB p6 B1 (2)	SNB p6 A3 (2)*	
	SNB p6 B1 (2)*	

* underwent a planned freeze-thaw cycle

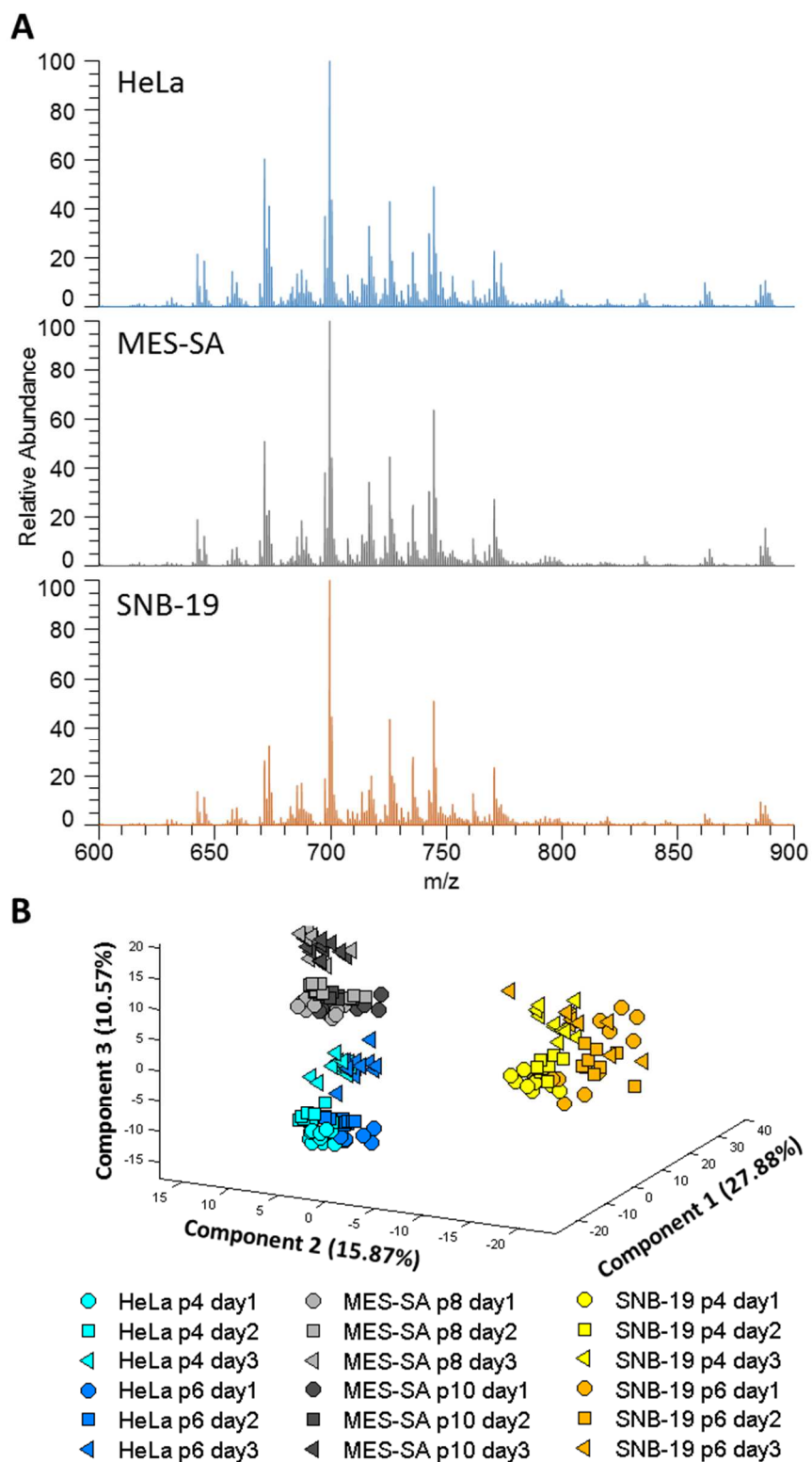


Figure 88. A) Mass spectral profiles between m/z 600-900 as obtained for HeLa, MES-SA, and SNB-19 cell line pellets, B) 3-dimensional PCA plot for dataset comprising HeLa, MES-SA and SNB-19 cell lines over the spectral mass range of m/z 600-900. Data were binned to 0.01Da, median normalised and log-transformed. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

For clinically oriented applications, REIMS spectral profiles were commonly analysed using supervised multivariate statistical methods such as linear discriminant analysis (LDA) to explore the feasibility of differentiation of various tissue types including healthy and diseased tissues.(2,130) In this study, data analysis was intentionally restricted to exploratory unsupervised analysis methods to assess whether REIMS profiles would reproducibly cluster into different groups corresponding to cell line identities. Cross-validation was performed to further validate clustering behaviour of the cell line REIMS patterns.

Spectral profiles as obtained for each cell line using the proposed methodology (Figure 12, page 60) are shown in Figure 88A and demonstrate high inter-strain similarity. However, differences in relative signal abundances were observed for each cell line leading to three clearly differentiated data groups corresponding to the cell types when performing principal component analysis on the respective dataset (see Figure 88B). SNB-19 cells are clearly differentiated from HeLa and MES-SA cells along the first principal component. Second and third principal component are required to fully separate HeLa and MES-SA cell lines from each other. A slight separation along the second principal component due to passage numbers can be seen for HeLa and SNB-19 cells, but is less pronounced for the MES-SA cell lines. Further separation correlating with the analysis day was observed for all three cell lines along the third principal component and it is especially pronounced for the third analysis day for MES-SA and HeLa cell lines while being less pronounced for SNB-19. However, these analytical and biological variances are small compared to the inherent spectral differences between the different cell lines. These results suggest that although there is an expected biological and analytical variance in the data, the REIMS spectral profiles show sufficient reproducibility and specificity to characterise and distinguish human cancer cell lines.

Cross-validations were performed based on PCA models of a training set created as described in Materials and Methods - Analysis of cell cultures using REIMS - Data Analysis of REIMS cell line data. Following this procedure, three different training sets were generated and subjected to PCA analysis. These training sets comprised between 25-28 sampling points, which represent <15% of the overall amount of samples (n=203). The composition of the training sets and corresponding cross-validation results are shown in Table 23. Consistently, the cross-validation results show $\geq 98\%$ correct classification for SNB-19 samples, with only a single misclassification into MES-SA observed for the third training set which was associated with the unequal sample size for MES-SA (n=12) and SNB-19 (n=7). This might lead to a bias of the PCA calculations towards the larger sample subset and thus more likely

misclassifications into the MES-SA group. While the first test set proved to be very robust for the classification of all three cell lines, the second and third test set do lead to occasional misclassifications between MES-SA and HeLa cells, respectively. This suggests a more close spectral relationship between HeLa (cervical origin) and MES-SA (uterine origin) compared to the SNB-19 cells (CNS origin) which is in agreement with the PCA analysis results of the whole dataset as described above.

Table 23. Cross-validation results for SNB-19, HeLa and MES-SA cell lines based on PCA model comprising the first 4 principal components and using 3 nearest neighbour as classifier. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

kept in test set	predicted			correct
	HeLa	MES-SA	SNB-19	
HeLa p6 A day2 (9 samples)	63	-	-	100%
MES-SA p8 A day1 (9 samples)	2	57	-	97%
SNB-19 p4 A day1 (9 samples)	-	-	54	100%
HeLa p4 B day2 (9 samples)	62	1	-	98%
MES-SA p10 B day1 (9 samples)	17	42	-	71%
SNB-19 p4 B day2 (7 samples)	-	-	56	100%
HeLa p4 A day1 (9 samples)	49	14	-	78%
MES-SA p10 A day3 (12 samples)	-	56	-	100%
SNB-19 p6 A day2 (7 samples)	-	1	55	98%

5.3. REIMS profiles for the NCI60 cell line set

Following the confirmation that REIMS spectral patterns are able to distinguish between three cancer cell lines, we set out to profile the entire NCI60 panel, consisting of 60 human different cancer cell lines. Based on amount of available biomass after culture, we recorded 4 to 15 individual measurement points for each cell line, and included eight biological replicates (see Table 24 for detailed sample set). Hierarchical cluster analysis (HCA; Figure 89, all technical replicates per cell line averaged) and principal component analysis (PCA, Figure 90) of the resulting dataset indicated that the 60 cells are characterised by unique REIMS profiles. Biological replicates showed the expected level of similarity as indicated by the cluster analysis (see Figure 89) showing close clustering in the HCA dendrogram. This is true for all cell lines except IGROV-1 and SF-268. In case of IGROV-1, a comparably small number of 3 and 5 REIMS spectra were available for each respective replicate, which was tentatively associated with the incorrect classification. The spectral difference observed in the

case of SF-268 raises the possibility of strong biological variance or a potentially compromised biological replicate.

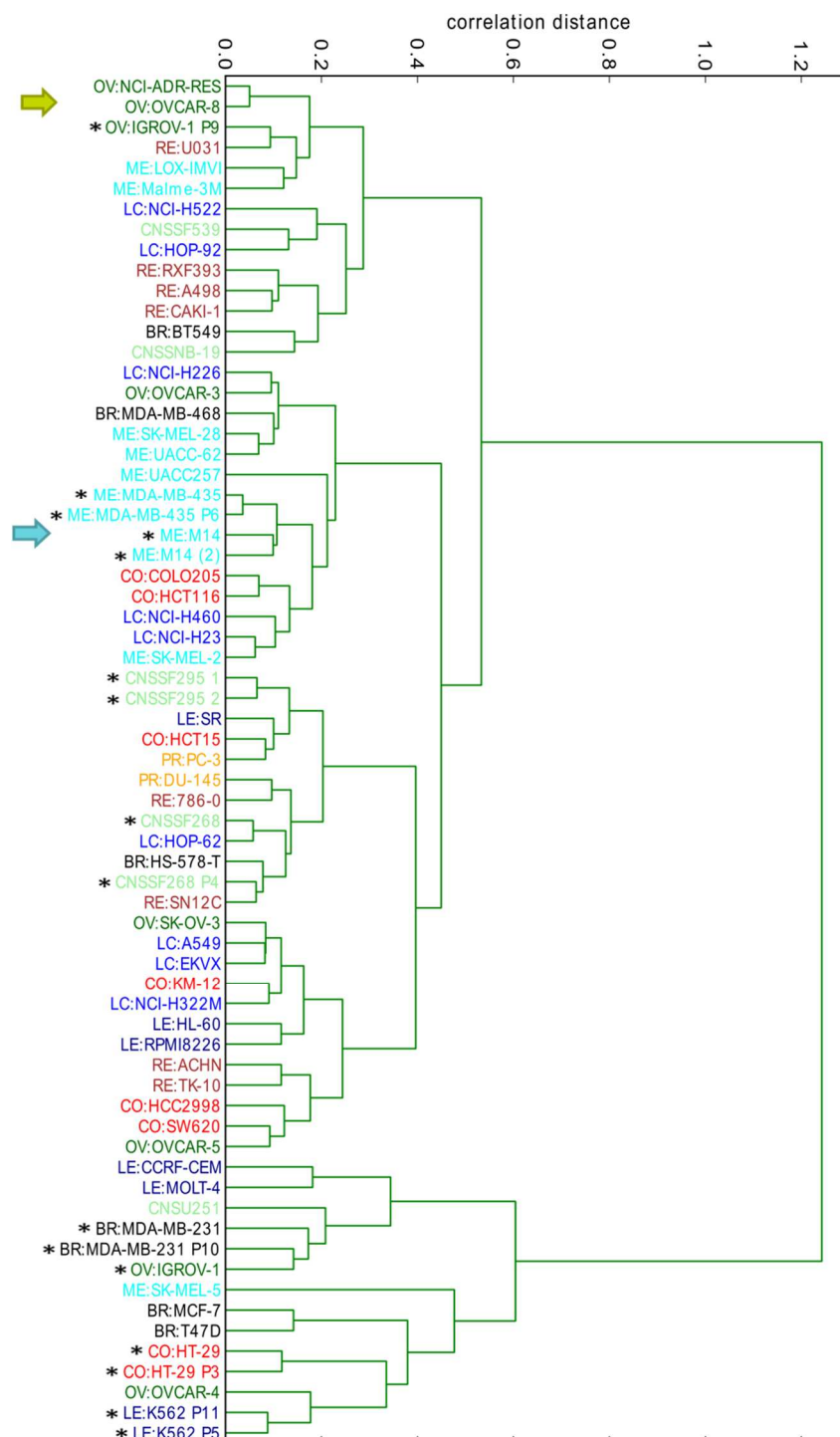


Figure 89. Hierarchical cluster dendrogram of the NCI60 cell lines including biological replicates (highlighted by asterisks). Arrows indicate closely related cell lines, blue MDA-MB-435 and M14, green NCI-ADR-RES and OVCAR-8. Distance was calculated using Pearson correlation and agglomeration via the Ward metric. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

Table 24. Cell lines and respective number of replicates used in this study. Numbers of respective biological replicates are given in parantheses. Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

Tissue of origin	Cell line	No of biological replicates	No of technical replicates
Breast	BT549	1	4
	HS-578-T	1	6
	MCF-7	1	12
	MDA-MB-231	2	10 (6+4)
	MDA-MB-468	1	
	TD-47-D	1	11
CNS	SF268	2	18 (8+10)
	SF295	2	18 (8+10)
	SF539	1	8
	SNB-19	1 (3)	10 (10+?+?)
	SNB-75	not measured	
	U251	1	9
Colon	COLO-205	1	10
	HCC2998	1	7
	HCT-15	1	7
	HCT-116	1	11
	HT-29	2	9 (5+4)
	KM-12	1	7
	SW-620	1	12
	Leukaemia	CCRF-CEM	1
HL-60		1	7
K562		1 (2)	9 (9+?)
MOLT-4		1	6
RPMI-8226		1	5
SR		1	8
Melanoma	LOX-IMVI	1	9
	M14	2	15 (7+8)
	Malme-3M	1	6
	MDA-MB-435	2	14 (4+10)
	SK-MEL-2	1	10
	SK-MEL-5	1	8
	SK-MEL-28	1	9
	UACC62	1	6
	UACC257	1	9
Non small cell lung cancer	A549	1	10
	EKVX	1	9
	HOP-62	1	6
	HOP-92	1	13
	NCI-H23	1	11
	NCI-H226	1	9
	NCI-H322M	1	7
	NCI-H460	1	9
	NCI-H522	1	9
	Ovarian	IGROV-1	2
NCI-ADR-RES		1	9
OVCAR-3		1	6
OVCAR-4		1	7
OVCAR-5		1	10
OVCAR-8		1	8

	SK-OV-3	1	9
Prostate	DU-145	1	9
	PC-3	1	11
Renal	786-0	1	9
	A498	1	11
	ACHN	1	10
	CAKI-1	1	6
	RXF393	1	6
	SN-12-C	1	8
	TK-10	1	9
	UO31	1	7

Similar results were obtained for cross-validation results performed on the PCA model (see Table 25 for identification accuracies based on PCA model shown in Figure 90) which reveal partial misclassification for IGROV-1 and no correct results for SF-268.

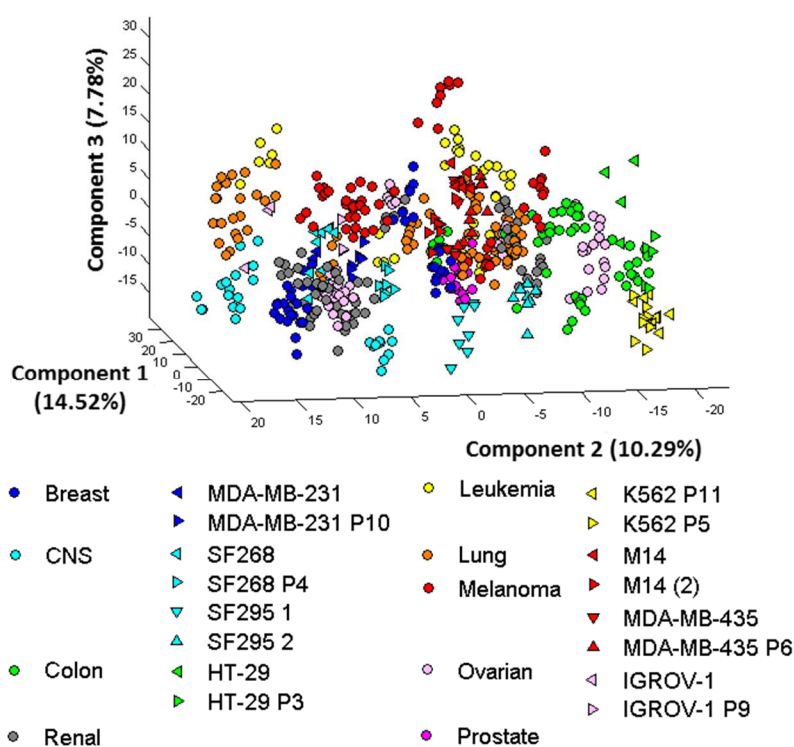


Figure 90. PCA plot of REIMS data obtained for the NCI60 dataset. Biological replicates are highlighted by triangular markers. Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

Gene expression profiling revealed that the MDA-MB-435 cells more closely resembled melanoma cell lines than the other breast tumour lines.(228) Consistent with further gene expression,(229) high-density single-nucleotide polymorphism(230) and karyotype analyses,(231) the REIMS profiles also indicate that cell lines MDA-MB-435 and M14 are closely related (Figure 89, blue arrow, right). Karyotyping has also found that NCI-ADR-RES is in fact a drug-resistant derivative of OVCAR-8.(232) As shown in Figure 89, these

cell lines (indicated by blue arrow, left) also show close similarity based on their REIMS profiles. Taken together, these results confirm that REIMS profiles are strongly associated with the biological identity of cancer cell lines.

Gene and protein expression patterns of the NCI-60 panel were found to largely correlate with tissue type of origin to some extent,(137,229) whereas metabolic signatures did not differentiate well between tissue origins.(219) Clustering of the cell lines based on their REIMS lipid profile showed extensive heterogeneity within most tissue types, except for melanoma samples (Figure 89, light blue coloured cell lines).

Table 25. Cross-validation results for replicated NCI60 cell lines based on PCA model shown in Figure 90. Three nearest neighbour as classifier. Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

kept in test set	predicted	correct
K562 P5 (9 samples)	LE:K562 (9)	100%
K562 P11 (9 samples)	LE:K562 (9)	100%
MDA-MB-231 (6 samples)	BR:MDAMB231 (4)	100%
MDA-MB-231 P10 (4 samples)	BR:MDAMB231 (6)	100%
SF295 1 (10 samples)	CNS:SF295 (4), CNS:SNB19 (4)	50%
SF295 2 (8 samples)	CNS:SF295 (10)	100%
M14 (7 samples)	ME:M14 (5), ME:MDAMB435 (1), ME:SKMEL28 (2)	63%
M14 (2) (8 samples)	ME:M14 (1), ME:MDAMB435 (6)	14%
MDA-MB-435 (10 samples)	ME:MDAMB435 (4)	100%
MDA-MB-435 P6 (4 samples)	ME:MDAMB435 (10)	100%
SF268 (8 samples)	LC:HOP62 (7), RE:UO31 (2), BR:HS578T (1)	0%
SF268 p4 (10 samples)	RE:SN12C (8)	0%
HT-29 (5 samples)	CO:HT29 (3), OV:OVCAR5 (1)	75%
HT-29 P3 (4 samples)	CO:HT29 (4) CO:COLO205 (1)	80%
IGROV-1 (5 samples)	OV:IGROV1 (3)	100%
IGROV-1 P9 (3 samples)	OV:OVCAR8 (4) RE:UO31 (1)	0%

5.4. Comparison of REIMS spectral profiles of bulk cancer tissue samples and cell lines

In previous studies, the REIMS spectral profiles of *ex-vivo* human tissue specimens were shown to be highly specific for the histological/anatomical tissue type.(2)

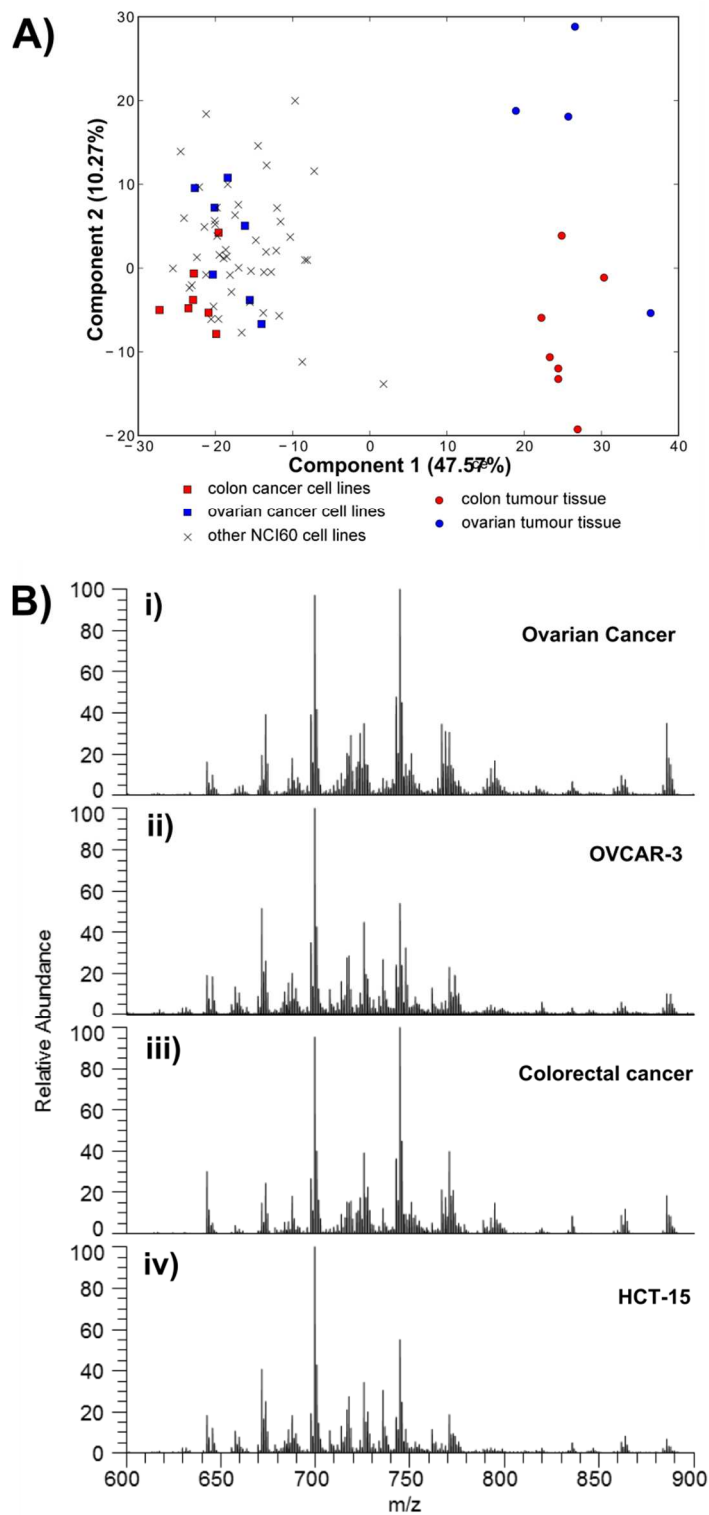


Figure 91. A) 2-dimensional PCA plot of averaged REIMS data collected from the NCI60 cells (squares) and cancer tissue samples (circles). The tissue of origin is indicated by red (colon) and blue (ovarian) colors. **B)** Comparison of spectral profiles for bulk tissue samples and cell lines of the corresponding tissue type of origin for i) ovarian cancer and ii) ovarian cancer cell line, and iii) colorectal cancer and iv) colon cancer cell line. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

To test how well cell line lipid profiles correspond to those of bulk tissue specimens, REIMS profiles of the NCI60 panel were compared to *ex-vivo* tissue samples of ovarian and colorectal adenocarcinomas analysed using the same bipolar experimental setup. A PCA plot of the resulting dataset is shown in Figure 91A and reveals clear differences between cell lines and bulk tissue specimens along the first principal component suggesting strong differences among their membrane lipid composition. A tentative separation according to tissue type of origin can be observed for both cell lines and tissue specimens, although more pronounced in the latter. Only a small number of tissue specimens (n=4) were available in case of ovarian tumours, but based on previous studies a significant increase in separation power can be expected for larger sample sets.(2) Nevertheless, the direction of separation is similar in both cases, suggesting similar lipidomic differences.

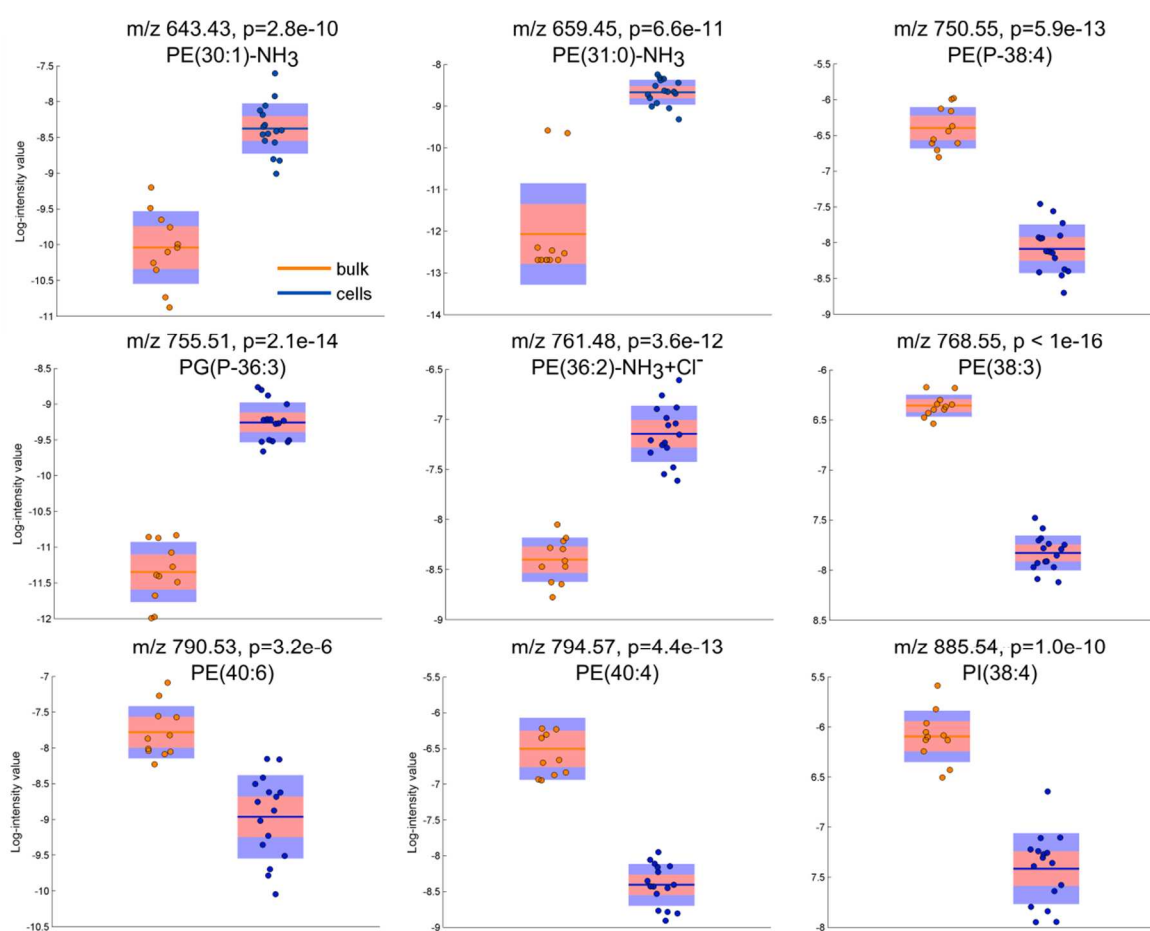


Figure 92. Box plot visualisation of *m/z* values (incl. p-value and annotation) found to be significantly increased in either bulk cancerous tissue (orange, comprising colorectal and ovarian cancer) or cancer cell lines (blue, comprising colorectal and ovarian cancer cell lines) using ANOVA. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

Representative spectral profiles of both ovarian and colon cancer tissue and cell line samples are shown in Figure 91B. Bulk tissues display larger amounts of long-chain phosphatidylinositols such as PI(38:4) at m/z 885.55. Similar trends were observed in case of certain phosphatidylethanolamines, for instance the peaks detected in the mass range of m/z 790-794 corresponding to PE(40:6)-PE(40:4) species or those occurring at m/z 766.54 and 768.55 corresponding to PE(38:4) and PE(38:3), respectively. On the other hand, m/z 645.45, corresponding to PA(32:1), was found in significantly higher proportions in cell lines. We speculate that the characteristic differences in lipid composition are due to the uniform lipid content of the culturing medium, which does not represent the complex lipid source of real tumours that rely on dietary and liver-synthesised lipids as well as *de-novo* lipid synthesis (see Figure 92 for a statistical analysis of distinct spectral features associated with cell lines and bulk tumours). This data adds to the controversy around the applicability of cell lines to model the complex behaviour of bulk cancer specimens. Thus, careful consideration around the translatability of results of cell line experiments is necessary in order to avoid wrong or simplified conclusions about bulk cancer. Experiments using bulk human cancer specimen, cell lineages derived from the same bulk tumour and xenografts of these same cell lines (patient-derived xenografts) can be conducted in order to further investigate this phenomenon.

5.5. NCI60 cell line panel - Correlation with gene expression data

The NCI60 cell panel has been extensively characterised and global transcriptomics and protein expression datasets were made publicly available. In order to correlate the REIMS dataset to previously published gene expression data, the respective dataset was obtained from the CellMiner online query tool (150,151).

Using an exhaustive search strategy, each m/z intensity value was correlated with each gene expression across all the overlapping cell lines as described in Materials and Methods - Analysis of cell cultures using REIMS - Data Analysis of REIMS cell line data.

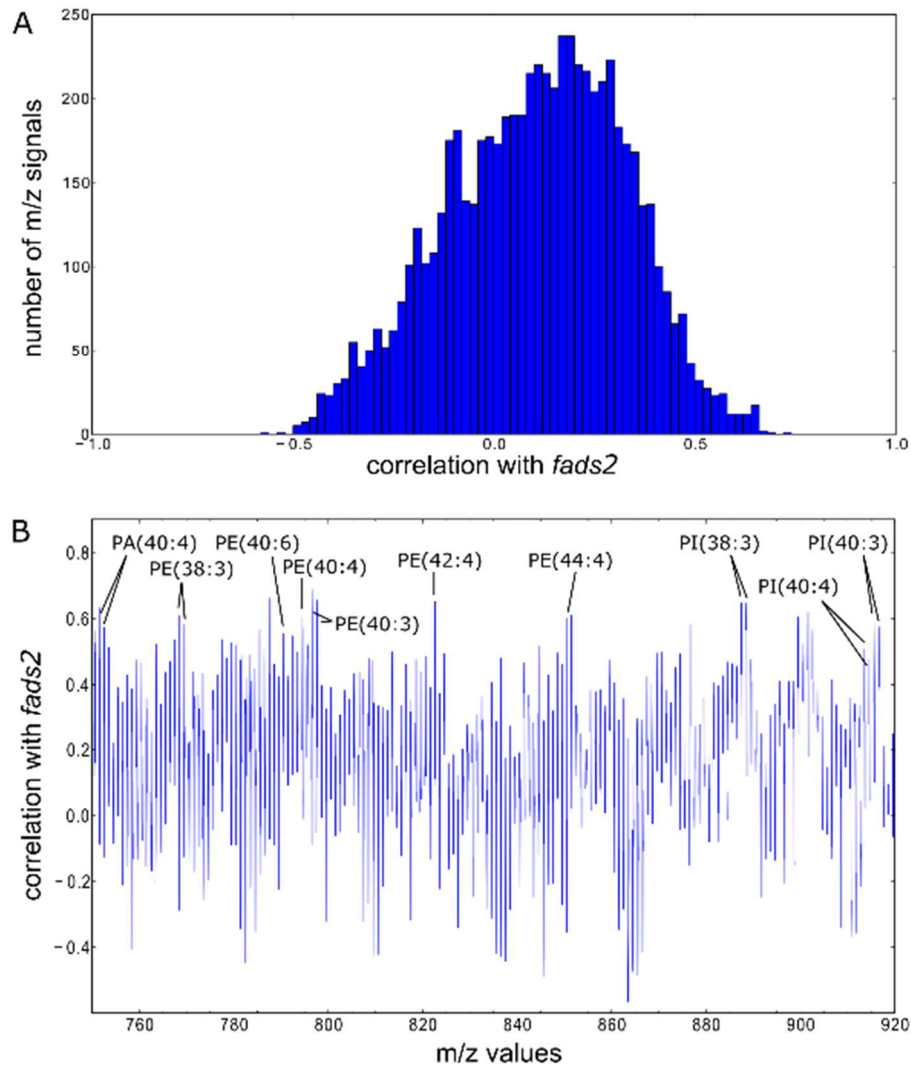


Figure 93. Correlation analysis between *fads2* gene expression and mass spectral data. A) Histogram of the correlation values between *fads2* gene expression and *m/z* peak intensities. B) Correlation values with *fads2* gene expression as a function of *m/z* values, and the putative lipids corresponding to the highly positively correlated *m/z* signals.

Strong positive correlation was obtained for the expression of genes playing an important role in lipid metabolism including the *fads2* gene, which is encoding the fatty acid desaturase enzyme (FADS2). The mammalian $\Delta 6$ -desaturase encoded by the *fads2* gene catalyses the biosynthesis of polyunsaturated fatty acids from precursor essential fatty acids such as linoleic acid C18:2n-6 and linolenic acid C18:3n-3. Other reported substrates include C16:0, C20:2n-6, C20:3n-3, C24:4n-6 and C24:5n-3.(233-235).

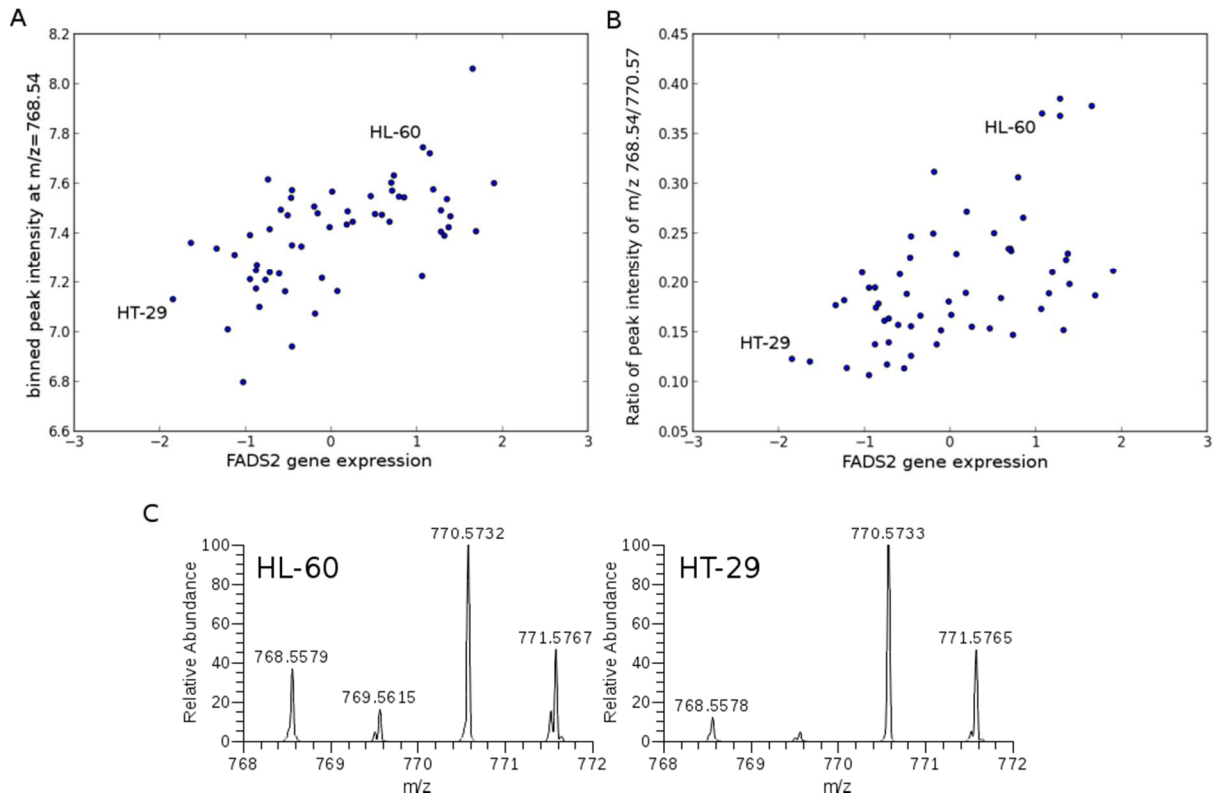


Figure 94. A) Binned peak intensity of m/z 786.54 as a function of scaled *fads2* gene expression, B) ratio of intensity values for peaks at m/z 768.55 and 770.57 as a function of *fads2* gene expression, C) Relative abundances of peaks m/z 768.55 and 770.57 in raw data for cell lines HL-60 (leukemia, left) and HT-29 (colon, right).

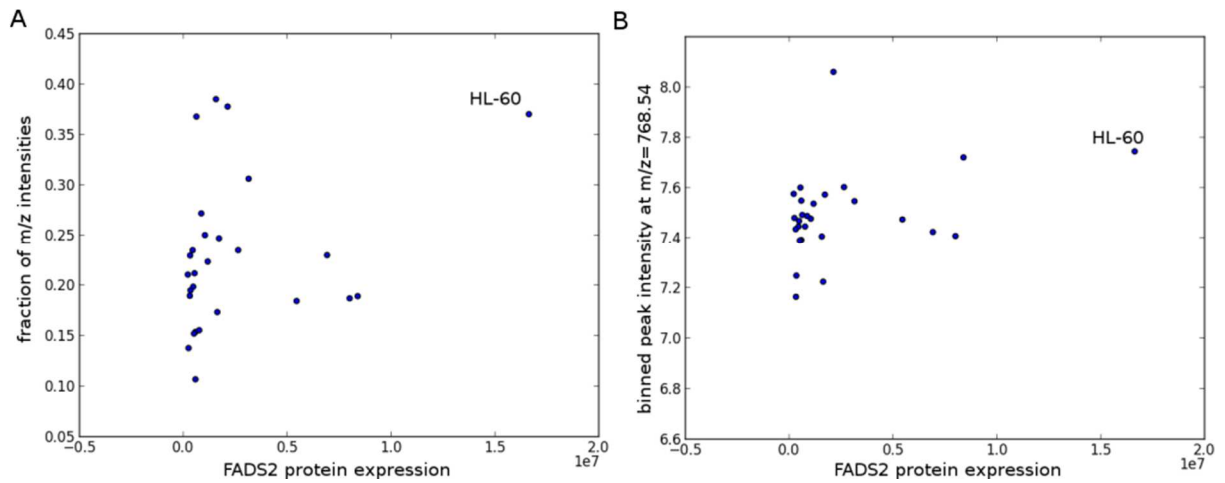


Figure 95. A) PE(38:3)/PE(38:2) peak intensity ratio and B) PE(38:3) peak intensity as a function of FADS2 protein expression as obtained from Gholami *et al.*(137)

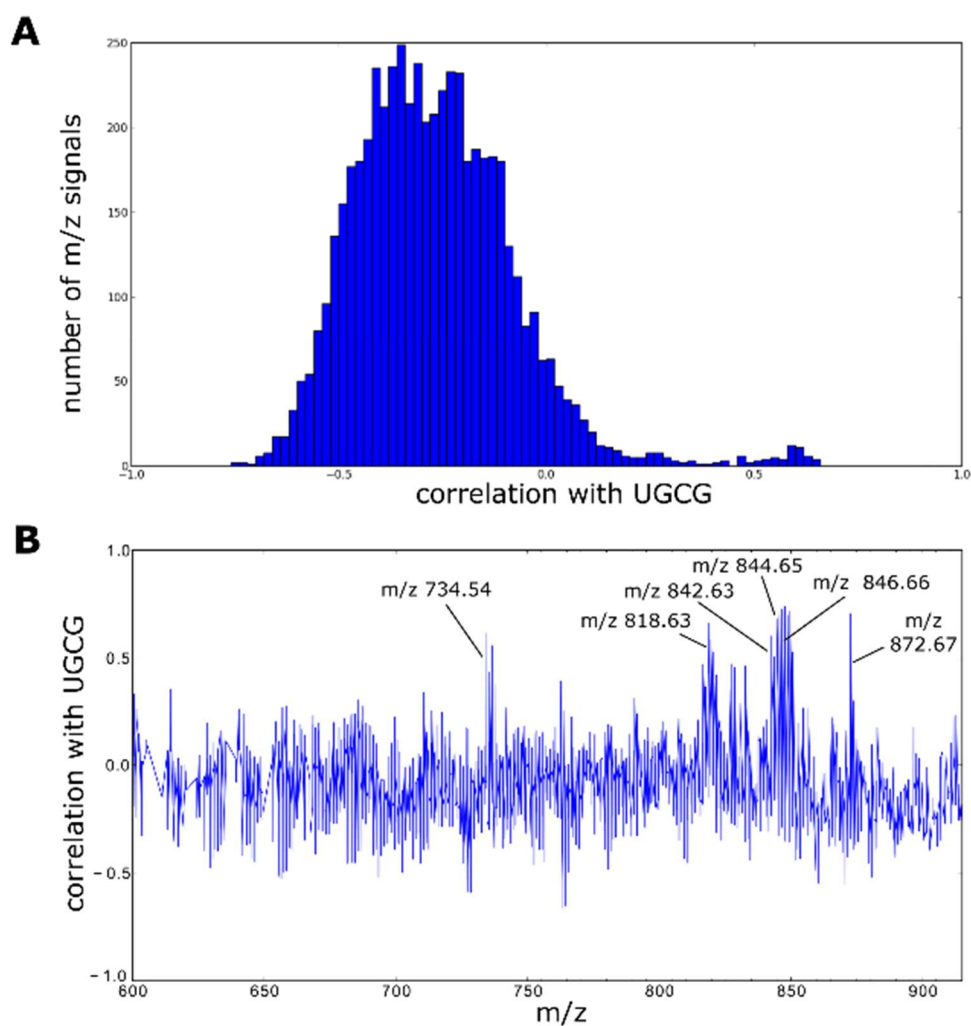


Figure 96. Correlation analysis between *ugcg* gene expression and mass spectral data. A) Histogram of the correlation values between *ugcg* gene expression and *m/z* peak intensities. B) Correlation values with *ugcg* gene expression as a function of *m/z* values including putative lipid identity for highly positively correlating *m/z* values.

The *fads2* gene expression data shows good positive correlation with the abundance of several signals in the REIMS data set (see Figure 93), among those *m/z* 768.54 (Figure 94A for correlation plot). This spectral feature was attributed to a peak at *m/z* 768.5578 in the corresponding raw data which in turn was assigned to PE(38:3) based on exact mass measurements, associated with PE(18:0/20:3) based on MS/MS data. Assuming that both enzyme substrate and product would be incorporated into the same phospholipid species, the putative substrate can indirectly be detected incorporated into PE(38:2) at *m/z* 770.5733. To investigate any changes in the relative abundances of PE(38:2) and PE(38:3) the ratio or raw signal intensities for both molecular species was determined for each cell line and plotted against the *fads2* gene expression (see Figure 94B). Similar correlation behaviour was

observed for both the signal at m/z 768.54 and the ratio of 768.54/770.57, with similar cell lines being situated at either end of the correlation curves. Figure 94C depicts the corresponding signal intensities in case of the HL-60 leukemia cell line and the colon HT-29 cell line which are showing high and low *fads2* expression, respectively. In case of the latter, PE(38:3) accounts for approximately 15% of relative abundance compared to PE(38:2), for HL-60 the relative abundance increases to approximately 40% of the PE(38:2).

Correlation between lipid profiles and protein expression data (sourced from Ref. (137)) was demonstrated by plotting both PE(38:3)/PE(38:2) intensity ratio and PE(38:3) abundance against the expression level values for the FADS2 protein (see Figure 95). However, in the database from Reference (137), FADS2 protein expression has only been determined in case of 29 cell lines, hence a comparably clear correlation pattern cannot be obtained. Nevertheless, the cell lines showing enzyme activity correspond to those that show high gene expression for the *fads2* gene. Unfortunately, no protein expression values were available for those cell lines that displayed low *fads2* gene expression.(137)

Strong correlations have also been found in case of *ugcg* gene expression (see Figure 96). This gene encodes the UDP-glucose ceramide glucosyltransferase (UGCG) enzyme, which catalyses the first glycosylation step in glycosphingolipid biosynthesis, having ceramides and UDP-glucose as substrates. Glycosylated lipids are enriched in lipid rafts/lipid microdomains and play fundamental roles in a variety of cellular processes.(236) Positive correlations were observed for several signals between m/z 700-900 and their respective isotopes as shown in Figure 96 and Table 26. Based on exact mass measurements and MS/MS experiments, these species were tentatively identified as homologous monohexosylated ceramides in form of $[M+Cl]^-$ ions. Isotopic patterns recorded for these species and MS/MS data support the presence of a chloride in the adduct ion. All ions show analogous fragmentation behaviour, displaying a loss of the chloride ion ($\Delta m = 35\text{Da}$) and additional loss of the hexose moiety ($\Delta m = 198\text{Da}$, loss of HCl and $\text{C}_6\text{H}_{10}\text{O}_5$). Using tandem mass spectrometry only it is not possible to differentiate between glucosylated and galactosylated ceramide species as they exhibit identical fragment ions. Exemplary tandem mass spectra and mass spectrometric signal intensities as a function of gene expression numbers can be found in Figure 97 and Figure 98, respectively. No negative correlation was observed for the *ugcg* gene expression with the precursor ceramide species nor the ratio of glucosylceramides to precursor ceramides. This is tentatively associated with the fact that

ceramides are ubiquitous precursors in sphingolipid biosynthesis and there are numerous biosynthetic pathways where ceramides act as substrates, intermediates or products.

Table 26. Mass spectrometric signals that show strong positive correlation with the *ugcg* gene expression for the NCI60 dataset.

Exp. mass	Exact mass	Δ ppm	Tentative ID	Formula	Adduct	Correlation coefficient
734.5355	734.5343	0.2	GlyCer(d18:1/16:0)	C ₄₀ H ₇₇ NO ₈	[M+Cl] ⁻	0.552
818.6295	818.6282	0.2	GlyCer(d18:1/22:0)	C ₄₆ H ₈₉ NO ₈	[M+Cl] ⁻	0.662
842.6312	842.6332	-0.2	GlyCer(d18:1/24:2)	C ₄₈ H ₈₉ NO ₈	[M+Cl] ⁻	0.602
844.6451	844.6439	0.1	GlyCer(d18:1/24:1)	C ₄₈ H ₉₁ NO ₈	[M+Cl] ⁻	0.668
846.6627	846.6595	0.4	GlyCer(d18:1/24:0)	C ₄₈ H ₉₃ NO ₈	[M+Cl] ⁻	0.688
872.6733	872.6752	-0.2	GlyCer(d18:1/26:1)	C ₅₀ H ₉₅ NO ₈	[M+Cl] ⁻	0.707

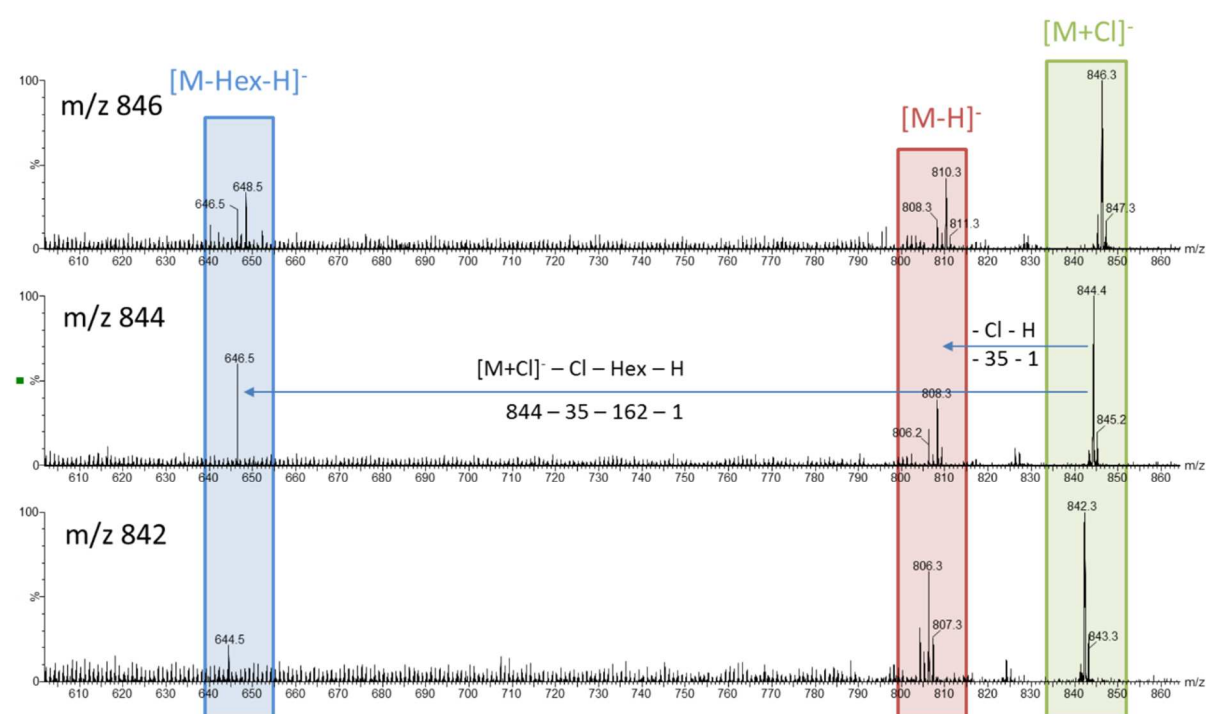


Figure 97. MS/MS spectra for signals at *m/z* 842, 844, and 846 identified as glycosylated ceramides GlyCer(d18:1/24:2), GlyCer(d18:1/24:1) and GlyCer(d18:1/24:0). Loss of chloride to form the [M-H]⁻ and loss of the hexose moiety to form [M-Hex-H]⁻ are observed.

Other putative sphingolipid species detected at *m/z* 860.6411, 862.6552 and 864.6775 showed positive correlations with the expression profiles of a number of genes encoding proteins involved in sphingolipid synthesis, e.g. *fa2h*, a gene encoding a protein that catalyses the synthesis of 2-hydroxysphingolipids. Further genes include the *degs2* gene, which encodes the Delta(4)-desaturase, sphingolipid 2 protein, and the *nr1h3* gene, which encodes the Liver X receptor alpha protein. Liver X receptors are part of the control system for transcriptional

programs involved in lipid homeostasis and inflammation and might thus indirectly influence the intracellular levels of certain sphingolipid species.

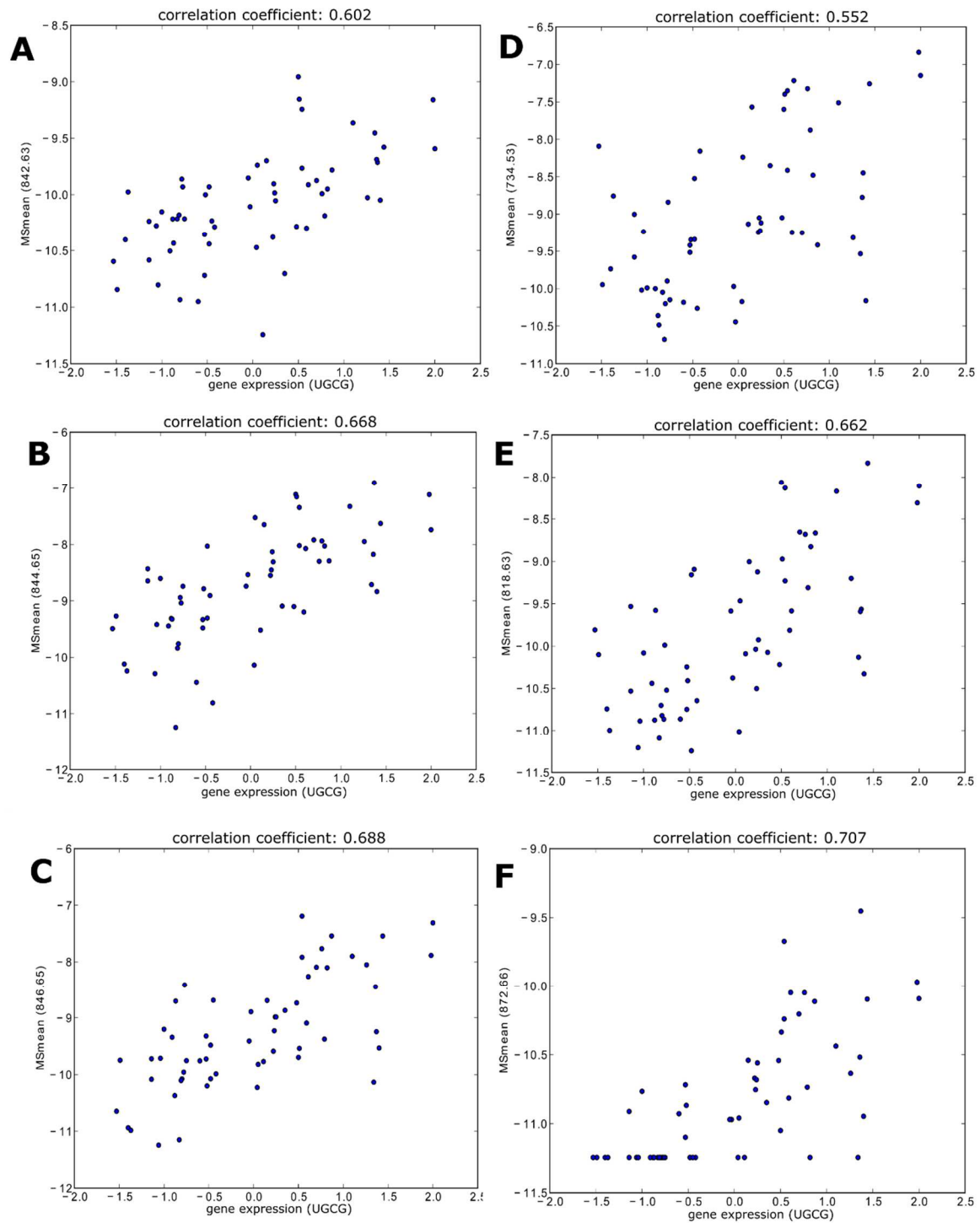


Figure 98. Mass spectrometric signal intensity (TIC normalised and log-transformed) as a function of *ugcg* gene expression for glycosylates ceramide species listed in Table 26. A) m/z 842.63, B) m/z 844.65, C) m/z 846.65, D) m/z 734.53, E) m/z 818.63, and F) m/z 872.66.

5.6. Analysis of mycoplasma infected cell lines

Cell cultures frequently get infected by *Mycoplasma*, a genus of bacteria that lack a cell wall encapsulating their cell membrane. *Mycoplasma* infection can alter many physiological processes and thus lead to misleading experimental results if a study is performed using infected cells. PlasmocinTM (InvivoGen, San Diego, CA, USA) is a commercially available antibiotic treatment that is frequently used to eradicate *Mycoplasma* infection in cell cultures.(237)

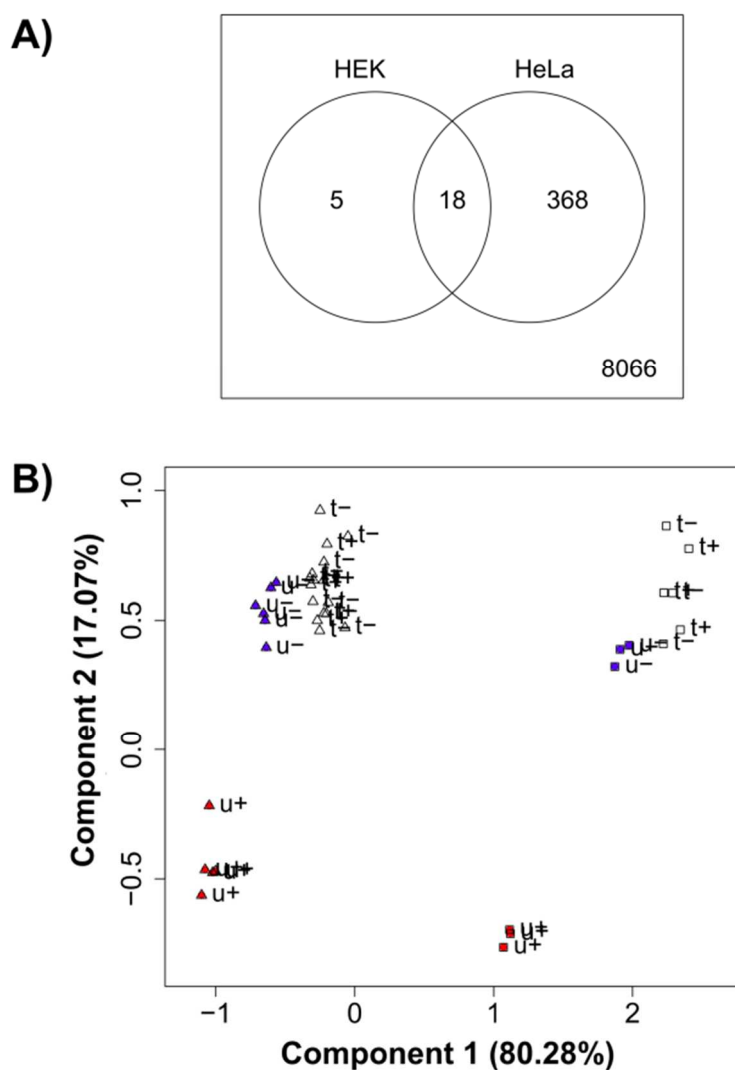


Figure 99. A) Number of binned m/z signals significantly higher in *Mycoplasma*-infected versus *Mycoplasma*-free samples in HEK and HeLa cell lines. B) Untreated *Mycoplasma*-free (blue, u-), *Mycoplasma*-infected (red, u+), *Mycoplasma*-free and PlasmocinTM-treated (white, t+), *Mycoplasma*-infected and PlasmocinTM-treated (white, t-); HEK (rectangle) and HeLa (triangle) samples as a function of PC1 and PC2 of PCA transformed untreated samples in the space of the 18 overlapping m/z signals.

Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

REIMS profiles of *Mycoplasma*-free, *Mycoplasma*-infected and PlasmocinTM-cured HEK and HeLa cell lines were recorded. The data was pre-processed as described for the NCI60 dataset. For both HEK and HeLa cell lines, the aim was to obtain *m/z* values that are significantly different between the untreated *Mycoplasma*-positive and -negative samples. Adjusted p-values were obtained using the adaptive Benjamini–Hochberg (BH) procedure to correct for multiple testing.(238)

Table 27. Values and tentative annotations of *m/z* signals that were found to be significantly higher in *Mycoplasma*-infected samples as compared to *Mycoplasma*-free samples in both HEK and HeLa cell lines. Adapted with permission from Ref. (134). Copyright (2016) American Chemical Society.

significantly different binned <i>m/z</i>	corresponding <i>m/z</i> signal	Annotation
687.54	687.5468	unknown
722.51	722.5156	PE(P-36:4)
733.53	733.5231	PE(P-38:4)
747.52	747.5193	PG(34:1)
748.53	748.5243	Isotope of <i>m/z</i> 747.52
753.51	753.5090	PG(P-36:4)
764.52	764.5264	PE(38:5)
764.53	764.5262	PE(38:5)
766.53	766.5412	PE(38:4)
773.54	773.5359	PG(36:2)
774.54	774.5391	PG(36:2), Isotope of <i>m/z</i> 773.54
774.55	774.5391	PG(36:2), Isotope of <i>m/z</i> 773.54
775.56	775.5520	PG(36:1)
776.56	776.5564	PG(36:1), Isotope of <i>m/z</i> 775.56
776.57	776.5564	PG(36:1), Isotope of <i>m/z</i> 775.56
819.52	819.5189	PG(40:7)
820.53	820.5268	PG(40:7), Isotope of <i>m/z</i> 819.52
820.54	820.5268	PG(40:7), Isotope of <i>m/z</i> 819.52

As shown in Figure 99A, 23 and 386 binned *m/z* signals were significantly higher in *Mycoplasma*-infected cells for HEK and HeLa cell lines, respectively. The higher number of significantly increased peaks found for HeLa cells may be explained with a higher number of sampling points (contributing to higher power in significance testing) or may reflect the increased reactivity of HeLa cells to *Mycoplasma* infection. Interestingly, we found only one signal showing reduced intensity in *Mycoplasma*-infected cell lines. Table 27 lists the 18 *m/z*

signals and their tentative annotations that were significantly higher in infected samples in both cell lines. An example for one of these values is shown in Figure 100 which visualises changes in intensity of *Mycoplasma*-free and -infected and PlasmocinTM-treated HEK and HeLa samples for the binned values around m/z 819.52 (identified as PG(40:7) based on exact mass measurements). This signal, as well as the signal corresponding to its isotope, was found significantly increased among *Mycoplasma*-infected cells for both HeLa and HEK cell lines but was found to return to the pre-infection level in the treated samples after successful PlasmocinTM treatment. Similar results were obtained for the other significant m/z signals. This change is tentatively associated with response of the host cell lipid metabolism to the *Mycoplasma* infection and highlights the possible negative influence of *Mycoplasma* infection on obtaining accurate and reliable data on cell line biochemistry.(239)

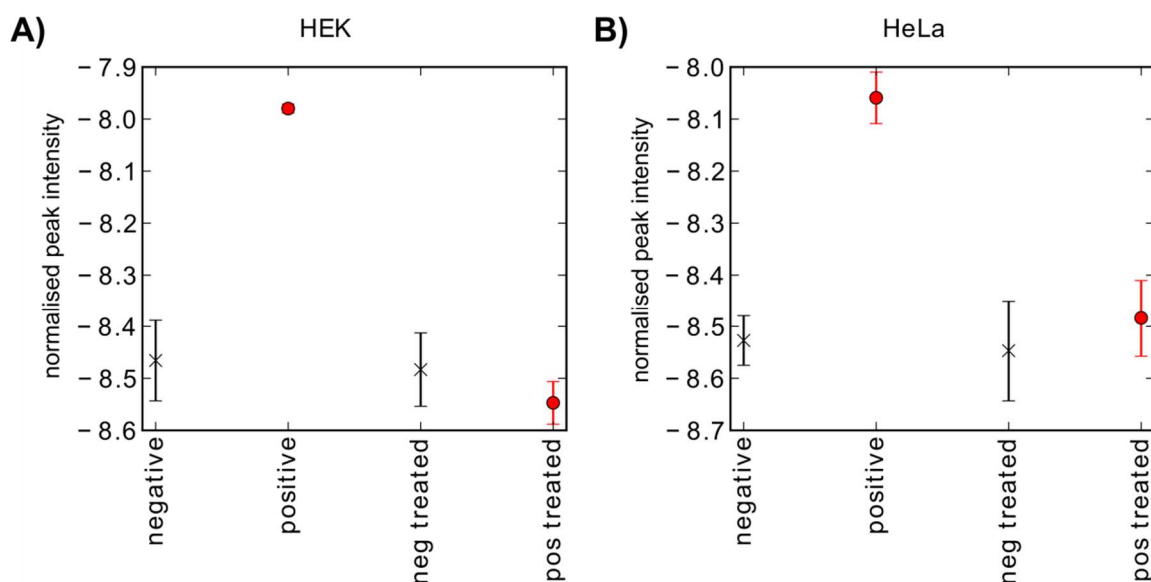


Figure 100. Intensities of TIC nomalised and log transformed signal at m/z 819.52 in *Mycoplasma*-free or infected and PlasmocinTM-treated samples in A) HEK and B) HeLa cell lines. Reprinted with permission from Ref. (134). Copyright (2016) American Chemical Society.

PCA was performed on the 18 distinguishing m/z features listed in Table 27 for *Mycoplasma*-infected and *Mycoplasma*-free HEK and HeLa samples (see Figure 99B). The first principal component (PC) corresponds to differences between cell lines while the second PC corresponds to differences between *Mycoplasma*-free and -infected samples. When plotting the PlasmocinTM-treated HEK and HeLa (both originally infected and healthy) samples, they are found to cluster together with the untreated and *Mycoplasma*-free samples of the corresponding cell line. Most of the significantly different signals between healthy and infected cell lines were found to be of phosphatidylglycerol origin and were found present

both infected and healthy cell lines. This underpins that *Mycoplasma* infection introduces a reversible change in cell line lipidome rather than direct detection of *Mycoplasma* lipids which would be expected absent in uninfected samples. In order to use this information in a systematic way to detect *Mycoplasma* infection, more cell lines and samples need to be collected and tested. Nevertheless, the approach demonstrates the applicability of the REIMS methodology to study changes during *Mycoplasma* infection and as possible method for *Mycoplasma* screening.

6. Taxon-specific markers

6.1. Introduction: Identification of bacteria in human tissues

The main application area for the detection of microorganisms in human subjects is in clinical microbiology for the identification of pathogenic species in clinical specimens such as sputum, faeces, pus, urine, cerebral spinal fluid or blood in order to establish diagnosis and commence adequate therapy. However, there is steadily growing awareness about the influence of microbial communities colonising the human body and their influence on human health. An increasing number of studies are focused on the effect and composition of the microbiome and its interaction with the human host system in both health and disease development. Through these efforts, several seemingly unrelated clinical conditions could be associated with infection by either a single pathogen, such as *Helicobacter pylori* infection causing the development of gastric ulcers and eventually gastric cancer,(5) or changes in overall microbiome composition, such as the change of ratio of Firmuctes to Bacteroidetes in the gut of obese and lean mice(240) or a loss of biodiversity in gut microbiome in case of chronic inflammatory conditions such as Crohn's disease.(241) There are several studies suggesting that the development of cancer might be influenced by the microbiome as well, such as the development of colorectal cancer (CRC). Evidence was presented that could correlate the occurrence of CRC with the presence of *Streptococcus gallolyticus*,(242) *Enterococcus faecalis*,(243) *Fusobacterium nucleatum*(244) and *Bacteroides fragilis*(245). In the light of these findings, it might be of considerable value to be able to characterise bacterial communities in a spatially-resolved manner and to associate local changes of the microbiome with the presence of a cancerous lesion in the gastrointestinal tract. Mass spectrometric methods such as imaging mass spectrometry of *ex-vivo* specimens(119,121,246) or REIMS incorporated into an endoscopic setup (so-called iEndoscope),(247) could be applied for this purpose.

Traditionally, presence of microbial species in tissues are determined using culture-based methods for instance in the case of isolating a single pathogen causing infection. Identification results are obtained by observing primarily phenotypic characteristics including colony morphology, carbon source utilisation pattern and expression of certain enzymes to define genera and species. Swabs are taken from the area to be analysed and microorganisms are cultured on a range of different culturing media. Although this methodology is untargeted

in nature, it comes with the inherent disadvantage that only those microorganisms can be identified, which grow on conventional microbiological media. It is estimated though that only about 1% of all microorganisms can be cultured using these methods.(248) Instead of identifying bacteria using biochemical tests and colony morphology, more recently bacteria are more and more often identified using their protein profiles obtained by MALDI-TOF-MS. This methodology shortened the time-demand needed for identification by at least 24hrs but it still requires a pure culture. Direct-on sample applications of MALDI were reported for urine and positive blood cultures, however, these protocols are considerably more complex compared to the identification of pure cultures as they require the separation of cells of bacterial origin from those of human origin. Using these culturing-based techniques, any information about spatial distribution of the organisms is lost as well as information about organisms that cannot be conventionally cultured.

Identification of bacteria using 16S rRNA gene sequencing can be performed directly from clinical and/or environmental samples and can thus resolve limitations posed by incorporation of culturing steps into the workflow. Two dominant technologies are generally used for culturing-independent characterisation of microbial communities, denaturing gradient gel electrophoresis (DGGE) and next generation sequencing (NGS).

In case of DGGE, similar length fragments of PCR-amplified fragments of DNA are separated in a gel containing a gradient of DNA denaturant or along a linear temperature gradient (TGGE) according to their physicochemical properties which in turn are dependent on the sequence of the fragments. The 16S rRNA gene is of similar size across bacterial phyla and thus DGGE is a valuable tool to assess diversity of 16S rRNA gene sequences. The number of bands at differing horizontal positions can be used to assess the biodiversity within a microbial community. Several microbial communities can be analysed on a single gel which allows for multiplexing and thus more time- and cost-efficient analysis. As analysis is qualitative only, bands can be cut from the gel and sequenced for more detailed analysis on the identity of organisms present.(249)

Using NGS methods, generic primer sets are directly added to a homogenate of the biological sample and the presence of bacterial taxa is determined by reading out the detected DNA fragments. This significantly increases the untargeted nature of sequencing-based techniques and this methodology is thus commonly applied to characterise microbiomes. This approach is currently the most commonly used tool in microbial community analysis. Analysis can be performed to either identify the bacterial taxa present in a sample or alternatively which

functional genes are present. Common NGS platforms allow sequencing of only comparably small regions of the 16S rRNA gene (Roche-454 pyrosequencing platform around 400 bases, Illumina MiSeq around 600 bases), allowing for cost-effective deep sequencing of large study cohorts, however, organism identification is less sensitive compared to sequencing the whole 16S rRNA gene. For this reason, studies applying NGS generally focus on characterising the microbial community as whole rather than individual taxa.(250) Similarly to conventional culturing techniques, these culturing-independent molecular methods lose spatial information due to sample homogenisation and thus do not reflect spatial compositional changes.

Fluorescent *in-situ* hybridisation (FISH) is a histological technique able to visualise the spatial distribution of given DNA sequences in tissue sections. Fluorescent DNA probes are applied which bind to complementary sequences of the DNA present in the sample. Fluorescence microscopy is subsequently used to detect and localise the fluorescent probes. rRNA-targeted oligonucleotide probes are available for a range of bacteria and can be used to selectively visualise bacteria in a complex microbial community or to detect bacterial and viral pathogens, including intracellular organisms such as *Chlamydia*.(251) The technique provides high spatial resolution and sensitivity, however, it is also highly targeted in nature and requires different probes for different species of bacteria and is thus not well suited for characterising microbial communities or untargeted detection of pathogens for clinical diagnostics.

While most outlined technologies rely on either the detection of unique bacterial proteins or unique DNA sequences, in this study we assessed the feasibility of detecting bacteria based on unique bacterial metabolites, particularly lipids. Although microbial REIMS phospholipid profiles were shown to be highly specific and reproducible, detection and identification of microorganisms based on their full lipidomic profile in a complex human matrix such as sputum or tissues is most likely not feasible as the deconvolution of signals of human and bacterial origin will be very challenging, especially in case of lipid species produced by both eukaryotes and prokaryotes. This task increases in difficulty in case of bacterial communities. A large part of potential specimen types (e.g. saliva, mucosal lining fluids) is usually not sterile and thus instead of just a single pathogenic species, the presence of an array of commensal pathogens is expected besides the pathogen. In this case the microbial profile would not only have to be separated from the human background profile but also the bacterial constituents of the bacterial community.

6.2. Assessment of feasibility of a taxon-specific marker approach

Due to limitations in detection of bacteria directly in human matrices, this study was planned to assess the applicability of a univariate biomarker approach to determine different types of bacteria present in a given sample. These samples may contain a mixture of different bacteria as in microbial communities or bacteria in a complex human matrix such as tissues or biofluids. In principle, a bacterium in a complex matrix could be detected in two ways using mass spectrometry, either by finding its multivariate spectral fingerprint within the complex matrix or by detecting the bacterium using a specific biomarker. The former option is associated with non-trivial problems with regard to data analysis, especially when considering the possible presence of multiple bacterial species with partially overlapping mass spectrometric signals. For detecting bacteria in human matrices it would also require a separate equally exhaustive database comprising possible human matrices such as blood, sputum, faeces etc. In cases such as faeces it will not be possible to get specimens containing no bacteria at all which further complicates the problem. On the other hand, detecting a chemically well-defined bacterial biomarker in a biological sample is a routine analytical exercise.

Therefore, a biomarker-based strategy is proposed that includes the identification of conservative lipid markers being specific for certain taxa of bacteria (phylum- down to species-level). These biomarkers were termed taxon-specific markers in the context of this work. A taxonomical marker-based approach would significantly reduce the bioinformatic requirements by reducing the multivariate profile-based methodology to the observation of presence or absence (and possible quantity) of biomarkers. These taxon-specific markers were identified using the REIMS-spectral database and subsequently used to detect bacteria in human matrices. These matrices can either be analysed using REIMS or by other mass spectrometric methods allowing the analysis of lipid compositions. In case of tissue-analysis, chromatography-based technologies offer high sensitivity however at the cost of loss of spatial information. For mapping of bacteria in tissue sections to assess their spatial distribution, mass spectrometry imaging (MSI) can be used. Commonly used MSI technologies are MALDI, SIMS and DESI-MS which all allow for the detection of complex lipid species and would thus be suited to detect taxonomic markers as obtained using REIMS. To assess feasibility of the proposed methodology, the visualisation of bacteria in colorectal tissue specimens using DESI-MSI was demonstrated. The tissue specimens were

cryosectioned and subsequently analysed using desorption electrospray ionisation mass spectrometry imaging as described in the Materials and Methods - Desorption Electrospray Ionisation Mass Spectrometry Imaging section. This mass spectrometry imaging-based approach is particularly interesting as using other techniques in this case would either lead to loss of spatial information (as with molecular or culturing-based methods) or loss of the untargeted nature (as in case of immunohistochemistry).

6.3. Biomarkers specific for different bacterial taxonomic levels

As observed for 28 bacterial species (shown in Figure 69 and Figure 70), closely related bacterial species are found to cluster together both in PCA and HCA analysis while unrelated bacterial species cluster separately. This indicates common spectral features among related bacteria which are not present in unrelated bacteria offering the possibility of finding biomarkers that are specific for a certain group. However, unsupervised multivariate analysis of a larger dataset shown in Figure 73 suggests that it might not be universally applicable to find taxonomic markers on all phylogenetic levels as the clustering results based on REIMS spectral profiles do not strictly match those of 16S rRNA sequences. Significant heterogeneity was observed especially among the very large phyla; however, better clustering results are obtained on lower taxonomic levels which would result in taxonomic markers providing more information on the exact nature of the bacteria present. Very distinct clustering behaviour of members of the phylum Bacteroidetes suggests that the biomarker approach will be well suited for this group of sphingolipid containing bacteria.

Taxon-specific markers (TSM) were derived as described in Materials and Methods - Determination of taxon-specific markers (see Figure 17) using the sample set detailed in Appendix 4. As predicted based on multivariate statistical analyses, specific bacterial biomarkers were found for a range of bacteria and on different phylogenetic levels. A list of taxon-specific markers found for different bacterial taxa is shown in Table 28. There is not necessarily a specific marker available at every taxonomical level for a certain group of bacteria. This is particularly true for very large and heterogeneous groups of bacteria, such as members of the phyla Actinobacteria, Firmicutes or Proteobacteria. However, taxonomical markers can often be found on low taxonomical levels such as species, genus or family when the group of bacteria is narrowed down to only such members having similar biochemical behaviour and environmental habitat. Markers in Table 28 which are prone to show interference by signals originating from the human tissue matrix are marked and require a

decision to be made on a case to case basis by taking distributions of other TSM and histological tissue types on the other hand into account when deciding whether markers are of bacterial or human origin. No markers were calculated for those bacterial groups listed in Table 28 which are highlighted in grey due to the insufficient size of the sample set which might lead to a bias in TSM generation due to spectral noise in such cases where only a single file represents the bacterial group. In cases where only a single taxon is included within a higher taxonomic level (e.g. only one class of bacteria within a phylum, as for Fusobacteria, or only a single genus within a family, as in Bacteroidaceae), found taxon-specific markers were attributed to the higher taxonomic level. The reasoning being that although the possibility exists that the detected TSM are specific for the lower taxon only (but dataset insufficient to prove), the lower taxon by definition will always be part of the higher taxon and thus all TSM found for lower levels will always be TSM for the presence of a higher (and less specific) taxonomic level as well.

There are two different types of taxonomical markers, those that are produced under arbitrary culturing conditions (conserved TSM) and others that are produced by certain strains only or under certain conditions only and hence often occur in highly variable amounts. Conserved taxon-specific markers are of particular interest as they might possibly be suitable to use for obtaining semi-quantitative information on certain bacterial species. Conserved (culture-condition independent) taxonomical markers include

- Mycolic acids for bacteria belonging to the Corynebacterineae suborder such as *Mycobacterium* spp., *Corynebacterium* spp. and *Rhodococcus* spp..

From REIMS experiments and from literature it is known that the mycolic acid chain length is very specific to the genus of bacteria within the Corynebacterineae suborder. It was further shown that mycolic acid patterns can be specific for certain species within a single genus. However, the REIMS dataset has to be further enlarged to validate species-specific mycolic acid patterns. The following mycolic acids have been detected from the corresponding genera:

- *Mycobacterium* spp.: C77-C81 (even and odd numbered, 0-2 unsaturations)
- *Corynebacterium* spp.: C28-C36 (even numbered, 0-2 unsaturations), excluding *C. amycolatum*
- *Nocardia* spp.: C48-C56 (even numbered, 0-3 unsaturations)
- *Rhodococcus* spp.: C28-C38 (even and odd numbered, 0-4 unsaturations)

- A variety of sphingolipid species were found to be specific for members of the phylum Bacteroidetes. These sphingolipids include oxidised ceramides species, phosphoethanolamine dihydroceramides and C15:0-substituted phosphoglycerol dihydroceramides and dihydroceramide.
Among those sphingolipid species, a series of galactosylated sphingolipids was found to be specific for *Bacteroides fragilis* species (*Bacteroides fragilis* alpha-Galactosylceramides).
- Among bacteria, plasmalogens are highly specific for anaerobic bacteria such as *Clostridium* spp. and *Fusobacterium* spp.. This is due to the fact that aerobic bacteria lost the biochemical pathway required for plasmalogen synthesis. Humans however are able to synthesise plasmalogens (although via a different biochemical pathway from anaerobes) although these were generally found to have longer chain lengths than bacterial plasmalogens.

A detailed assessment of the conservative nature of these markers lies outside of the scope of this project and hence conservative only refers to the detection of these signals in bacteria grown under arbitrary culturing conditions rather than stable detection levels. A thorough assessment of these properties will require a more comprehensive dataset with a larger number of strains per bacterial species as was used in the proof-of-principle study presented here.

Other taxonomical markers are less conserved and their detection is highly dependent on culturing conditions. These TSM are still indicative of a certain group of bacteria and can be used to detect the presence bacteria in complex matrices (not the absence), however, they might show very different production amounts and will thus not be suitable for quantification purposes. Such compounds include

- Lipopolypeptides that are produced specifically by certain *Bacillus* species such as surfactin for *B. subtilis* and *B. pumilus* and lichenysin for *B. licheniformis*. Production of these two molecules also enables straightforward differentiation of these otherwise very closely related bacteria.
- PQS-derived quorum-sensing molecules and mono- and di-rhamnolipid species found in *Pseudomonas aeruginosa*.

Table 28. Taxon-specific markers (given as m/z values) obtained for sample set detailed in Appendix 4. No markers were calculated for those levels in grey due to insufficient size of sample set. † marks an isotope peak, * marks a m/z value with molecular identity known (see Table 29), ‡ marks m/z value that is likely to experience interference from mammalian cells, m/z values in bold font indicate relative spectral abundance >50% in spectrum of pure bacterial culture.

	Phylum	Class	Order	Family	Genus	Species					
Gram-negative	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	<i>Bacteroides</i>	<i>Bacteroides acidifaciens</i>					
						<i>Bacteroides caccae</i>					
						<i>Bacteroides eggerthii</i>					
						<i>Bacteroides fragilis</i>					
						752.5419*					
						766.5574*‡					
						780.5732*‡					
						<i>Bacteroides helcogenes</i>					
						<i>Bacteroides ovatus</i>					
						<i>Bacteroides pyogenes</i>					
						<i>Bacteroides thetaiotaomicron</i>					
						<i>Bacteroides uniformis</i>					
						<i>Bacteroides vulgatus</i>					
								Porphyromonadaceae	<i>Parabacteroides</i>	<i>Parabacteroides distasonis</i>	
										<i>Parabacteroides johnsonii</i>	
				328.2857*	286.2466					814.7063	
				381.2765	627.4883					828.7232	
				552.4643‡	628.4913					840.6842	
				553.4674†,‡	465.3024*,‡					854.7022	
				590.4923*,‡	602.4783					858.6972	
				623.5024	604.5083*					872.7072	
				624.5054†	605.5113†					904.7051*	
				639.4954*	607.5074					918.7191*	
651.4953				932.7332*							
653.5113*				933.7362†							
654.5143†				944.7342*							
677.5238*				945.7372†							
691.5395*				946.7472*							
705.5562*				947.7502†							
				958.7461*							

Phylum	Class	Order	Family	Genus	Species
			959.7501 [†] 961.7661* 962.7691 [†]		
			Prevotellaceae 661.5283 675.5453 676.5503 [†] 870.8002 [†] 908.7401 922.7552	<i>Prevotella</i>	<i>Prevotella bivia</i>
			Rikenellaceae	<i>Alistipes</i>	<i>Alistipes onderdonkii</i>
	Flavobacteria 324.2545* 325.2576 [†] 333.2084[‡] 374.2365 392.2484 393.2504 [†] 564.4624 566.4794* 567.4834 ^{†,*} 568.4864 [†] 600.4664 604.4633 618.4773 619.4813 [†] 620.4883 [†] 639.5334 640.5363 [†]	Flavobacteriales	Flavobacteriaceae	<i>Chryseobacterium</i>	<i>Chryseobacterium indologenes</i> <i>Chryseobacterium sp</i>
				<i>Elizabethkingia</i>	<i>Elizabethkingia meningoseptica</i>
				<i>Myroides</i> 590.4474	<i>Myroides odoratimimus</i>
	Fusobacteria 227.2015 [‡] 644.4652 645.4633 [†]	Fusobacteria	Fusobacteriaceae	<i>Fusobacterium</i>	<i>Fusobacterium gonidiaformans</i> <i>Fusobacterium necrophorum</i> <i>Fusobacterium peridontiam</i> <i>Fusobacterium sp</i>

Phylum	Class	Order	Family	Genus	Species
646.4833* 647.4812† 882.6932 883.6982†					
Proteobacteria 768.5182‡ 782.5342‡ 783.5293†	Alpha-Proteobacteria 773.5332*‡	Caulobacterales 769.5502 770.5562† 771.5582† 795.5572‡ 797.5723†	Caulobacteraceae	<i>Brevundimonas</i>	<i>Brevundimonas diminuta</i>
		Rhizobiales 439.4155 440.4195† 739.5313‡ 784.5902‡ 785.5932‡,‡	Rhizobiaceae	<i>Rhizobium</i>	<i>Rhizobium radiobacter</i>
		Rhodospirillales 733.5752 857.7252	Acetobacteraceae	<i>Roseomonas</i>	<i>Roseomonas mucosa</i> <i>Roseomonas sp</i>
	Beta-Proteobacteria	Burkholderiales	Alcaligenaceae	<i>Achromobacter</i>	<i>Achromobacter sp</i> <i>Achromobacter xylosoxidans</i>
				<i>Alcaligenes</i>	<i>Alcaligenes faecalis</i>
			Burkholderiaceae 589.4013 590.4083† 591.4184 592.4214† 801.5662‡	<i>Burkholderia</i>	<i>Burkholderia cepacia complex</i>
			Comamonadaceae	<i>Acidovorax</i>	<i>Acidovorax temperans</i>
				<i>Comamonas</i>	<i>Comamonas kerstersii</i> <i>Comamonas sp</i>

Phylum	Class	Order	Family	Genus	Species
				<i>Delftia</i>	<i>Delftia acidovorans</i> <i>Delftia dentocariosa</i> <i>Delftia sp</i>
			Sutterellaceae 481.4274	<i>Sutterella</i>	<i>Sutterella wadsworthensis</i>
		Neisseriales 494.3855 502.3674 526.3673 527.3704 [†] 528.3653 544.3774	Neisseriaceae	<i>Eikenella</i>	<i>Eikenella corrodens</i>
				<i>Kingella</i>	<i>Kingella kingae</i> <i>Kingella sp</i>
				<i>Neisseria</i>	<i>Neisseria cineria</i> <i>Neisseria elongata</i> <i>Neisseria flavescens</i> <i>Neisseria gonorrhoea</i> <i>Neisseria lactamica</i> <i>Neisseria meningitidis</i> <i>Neisseria mucosa</i>
	Epsilon-Proteobacteria 730.5422 [‡] 731.5452 ^{‡,‡}	Campylobacterales	Campylobacteraceae 867.6582 993.8381 [†]	<i>Campylobacter</i>	<i>Campylobacter coli</i> <i>Campylobacter fetus</i> <i>Campylobacter jejuni</i> <i>Campylobacter sp</i>
			Helicobacteraceae 271.2284* [‡] 272.2305 [†] 299.2595*[‡] 300.2625 [†] 400.2644 492.3104 [‡] 794.5783 ^{†,‡}	<i>Helicobacter</i>	<i>Helicobacter pylori</i>
	Gamma-Proteobacteria	Aeromonadales	Aeromonadaceae	<i>Aeromonas</i>	<i>Aeromonas hydrophila</i>
		Cardiobacteriales 648.4603 649.4623 [†] 650.4653 [†] 793.4792	Cardiobacteriaceae	<i>Cardiobacterium</i>	<i>Cardiobacterium hominis</i>

Phylum	Class	Order	Family	Genus	Species			
		794.4802 [†]						
		Enterobacteriales	Enterobacteriaceae	<i>Citrobacter</i>	<i>Citrobacter amalonaticus</i> <i>Citrobacter braakii</i> <i>Citrobacter freundii</i> <i>Citrobacter koseri</i>			
		702.5083		<i>Enterobacter</i>	<i>Enterobacter</i>	<i>Enterobacter absuriae</i> <i>Enterobacter aerogenes</i> <i>Enterobacter amnigenus</i> <i>Enterobacter cloacae</i> <i>Enterobacter gergoviae</i>		
		703.5092 [†]						
		993.7282						
		994.7272 [†]						
							<i>Escherichia</i>	<i>Escherichia coli</i>
							<i>Hafnia</i>	<i>Hafnia alvei</i> <i>Hafnia paralvei</i> <i>Hafnia sp</i>
							<i>Klebsiella</i>	<i>Klebsiella oxytoca</i> <i>Klebsiella pneumoniae</i>
							<i>Morganella</i>	<i>Morganella morganii</i>
							<i>Panthoea</i>	<i>Panthoea sp</i>
							<i>Proteus</i>	<i>Proteus mirabilis</i> <i>Proteus vulgaris</i>
							<i>Providencia</i>	<i>Providencia rettgeri</i> <i>Providencia stuartii</i>
							<i>Raoultella</i>	<i>Raoultella ornitholytica</i> <i>Raoultella planticola</i>
			<i>Salmonella</i>				<i>Salmonella poona</i>	
			<i>Serratia</i>	<i>Serratia liquifaciens</i> <i>Serratia marcescens</i>				
			<i>Shigella</i>	<i>Shigella sonnei</i>				
		Pasteurellales	Pasteurellaceae	<i>Aggregatibacter</i>	<i>Aggregatibacter aphrophilus</i>			

Phylum	Class	Order	Family	Genus	Species
		688.4913 746.4503 823.5453 898.6921 915.6902 [†]		<i>Haemophilus</i>	<i>Haemophilus influenzae</i> <i>Haemophilus parahaemolyticus</i> <i>Haemophilus parainfluenzae</i>
				<i>Pasteurella</i>	<i>Pasteurella multocida</i>
		Pseudomonadales 745.5013 ^{*,‡} 746.5053 ^{*,‡}	Moraxellaceae	<i>Acinetobacter</i> 981.7171 982.7191 [†]	<i>Acinetobacter baumannii</i> <i>Acinetobacter iwoffii</i> 953.6811 <i>Acinetobacter johnsonii</i> <i>Acinetobacter junii</i>
				<i>Moraxella</i> 171.1395	<i>Moraxella catarrhalis</i> 771.5172 ^{*,‡} <i>Moraxella osloensis</i>
			Pseudomonadaceae 490.3304 514.3294	<i>Pseudomonas</i>	<i>Pseudomonas aearuginosa</i> 242.1554* 258.1508* 531.3654* <i>Pseudomonas luteola</i> <i>Pseudomonas monteilii</i> <i>Pseudomonas oryzihabitans</i> <i>Pseudomonas putida</i> <i>Pseudomonas stutzeri</i>
		Vibrionales 605.3823 607.3983 608.4013 [†] 633.4134 843.6042 [†]	Vibrionaceae	<i>Vibrio</i>	<i>Vibrio alginolyticus</i> <i>Vibrio cholerae</i> <i>Vibrio furnissii</i>
		Xanthomonadales 929.6852 930.6892 [†]	Xanthomonadaceae	<i>Stenotrophomonas</i>	<i>Stenotrophomonas maltophilia</i>
Gram-positive	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae 757.5403[‡]	<i>Actinobaculum</i> 287.2225

Phylum	Class	Order	Family	Genus	Species
				<i>Actinomyces</i>	<i>Actinomyces graevenitzii</i> <i>Actinomyces israelii</i> <i>Actinomyces odontolyticus</i> <i>Actinomyces oris</i> <i>Actinomyces sp</i> <i>Actinomyces turicensis</i> <i>Actinomyces viscosus</i>
			Corynebacteriaceae	<i>Corynebacterium</i>	<i>Corynebacterium afermentans</i> <i>Corynebacterium amycolatum</i> 169.0616 187.0735 999.7541 [†] <i>Corynebacterium diphtheriae</i> 491.4474* 492.4505 [†] 984.9051* <i>Corynebacterium imitans</i> <i>Corynebacterium minutissimum</i> <i>Corynebacterium sp</i> <i>Corynebacterium striatum</i>
			Microbacteriaceae	<i>Microbacterium</i>	<i>Microbacterium sp</i>
			Mycobacteriaceae	<i>Mycobacterium</i>	<i>Mycobacterium avium</i> <i>Mycobacterium fortuitum</i> <i>Mycobacterium peregrinum</i>
			Nocardiaceae	<i>Nocardia</i> 797.7762* 799.7842* 825.8083* 827.8162 [†] 828.8222 [†]	<i>Nocardia sp</i>

Phylum	Class	Order	Family	Genus	Species
				<i>Rhodococcus</i> 481.4630* 509.4943* 537.5259*	<i>Rhodococcus equi</i> <i>Rhodococcus sp</i>
			Propionibacteriaceae 361.2155 713.4752 779.5072 [‡] 877.5592 905.5842 990.7091 [†]	<i>Propionibacterium</i>	<i>Propionibacterium acnes</i>
		Bifidobacteriales 789.5293 [‡] 792.5502 ^{†,‡} 819.5783 [‡] 830.5622 [‡] 855.5272 [‡] 884.6092 885.6142 [†]	Bifidobacteriaceae	<i>Bifidobacterium</i> 827.4882 867.5942 [‡] 909.5752	<i>Bifidobacterium adolescentis</i> <i>Bifidobacterium bifidum</i> <i>Bifidobacterium breve</i> 805.5612 [‡] 923.5902 <i>Bifidobacterium infantis</i> <i>Bifidobacterium longum</i> <i>Bifidobacterium pseudocatenulatum</i>
				<i>Gardnerella</i> 766.5392 [‡]	<i>Gardnerella vaginalis</i>
		Micrococcales 913.5682 [‡] 914.5711 ^{†,‡} 915.5671 [†]	Micrococcaceae	<i>Arthrobacter</i>	<i>Arthrobacter creatinolyticus</i> <i>Arthrobacter sp</i>
				<i>Kokuria</i>	<i>Kokuria kristina</i> <i>Kokuria rhizophila</i> <i>Kokuria varians</i>
				<i>Micrococcus</i>	<i>Micrococcus luteus</i> <i>Micrococcus lylae</i>
				<i>Rothia</i>	<i>Rothia aeria</i> <i>Rothia amarne</i> <i>Rothia dentocariosa</i> <i>Rothia mucilaginoso</i> 885.5352

Phylum	Class	Order	Family	Genus	Species
					<i>Rothia sp</i>
			Micrococcineae	<i>Brevibacterium</i>	<i>Brevibacterium paucivorans</i> <i>Brevibacterium sp</i>
				<i>Dermabacter</i> 679.4953 707.5272 708.5282 [†] 722.5453 735.5563	<i>Dermabacter hominis</i> <i>Dermobacter sp</i>
Firmicutes	Bacilli	Bacillales	Bacillaceae	<i>Bacillus</i>	<i>Bacillus cereus</i> 677.4413 <i>Bacillus clausii</i> <i>Bacillus lichenformis</i> 1033.6927* <i>Bacillus pumilus</i> <i>Bacillus sonorensis</i> <i>Bacillus sp</i> <i>Bacillus subtilis</i> 1034.6766*
			Listeriaceae 675.9793 832.5352 [†]	<i>Listeria</i>	<i>Listeria monocytogenes</i>
			Paenibacillaceae 903.7221 914.7282	<i>Paenibacillus</i>	<i>Paenibacillus sp</i> <i>Paenibacillus unalis</i>

Phylum	Class	Order	Family	Genus	Species
			Staphylococcaceae 202.0525 228.0335 245.0596 763.5512[‡] 765.5482 [†]	<i>Staphylococcus</i>	<i>Staphylococcus aureus</i> <i>Staphylococcus capitis</i> <i>Staphylococcus caprae</i> <i>Staphylococcus cohnii</i> <i>Staphylococcus epidermis</i> <i>Staphylococcus haemolyticus</i> <i>Staphylococcus hominis</i> <i>Staphylococcus lugdunensis</i> <i>Staphylococcus pasteurii</i> <i>Staphylococcus pettenkoferi</i> <i>Staphylococcus saprophyticus</i> <i>Staphylococcus warneri</i>
		Lactobacillales 898.5391 923.5512 925.5671 926.5701 [†] 949.5672 951.5832 952.5861 [†] 953.5981 954.6011 [†] 955.5971 [†] 956.5971 [†] 979.6111	Aerococcaceae 163.0506	<i>Abiotrophia</i>	<i>Abiotrophia defectiva</i>
				<i>Aerococcus</i> 213.1245	<i>Aerococcus</i> sp <i>Aerococcus viridians</i> 160.0615 197.0829
			Carnobacteriaceae	<i>Granulicatella</i>	<i>Granulicatella adiacens</i>
			Enterococcaceae	<i>Enterococcus</i>	<i>Enterococcus avium</i> <i>Enterococcus casseliflavus</i> <i>Enterococcus cecorum</i> <i>Enterococcus faecalis</i> <i>Enterococcus faecium</i> <i>Enterococcus gallinarum</i> <i>Enterococcus raffinosus</i>
			Lactobacillaceae	<i>Lactococcus</i>	<i>Lactococcus lactis</i> <i>Lactococcus</i> spp

Phylum	Class	Order	Family	Genus	Species
			Leuconostocaceae	<i>Leuconostoc</i>	<i>Leuconostoc</i> sp
			Streptococcaceae 897.5351	<i>Lactobacillus</i>	<i>Lactobacillus gasseri</i> <i>Lactobacillus rhamnosus</i>
				<i>Streptococcus</i>	<i>Streptococcus agalactiae</i> <i>Streptococcus anginosus</i> <i>Streptococcus bovis</i> <i>Streptococcus canis</i> <i>Streptococcus constellatus</i> <i>Streptococcus cristatus</i> <i>Streptococcus dysgalactiae</i> <i>Streptococcus gallolyticus</i> <i>Streptococcus gordonii</i> <i>Streptococcus intermedius</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus milleri</i> <i>Streptococcus mitis</i> <i>Streptococcus mutans</i> 348.2184 372.2175 <i>Streptococcus oralis</i> <i>Streptococcus parasanguinus</i> <i>Streptococcus pneumoniae</i> <i>Streptococcus povas</i> <i>Streptococcus pseudoporcinus</i> 450.1635 <i>Streptococcus pyogenes</i> <i>Streptococcus salivarius</i> <i>Streptococcus sanguinis</i> <i>Streptococcus vestibularis</i> <i>Streptococcus viridans</i>

Phylum	Class	Order	Family	Genus	Species
	Clostridia 449.2685 703.4923 † 704.4953† 731.5253 † 732.5283 925.7262	Clostridiales	Clostridiaceae 649.4453 759.5583† 897.6951 969.7481 970.7541†	<i>Clostridium</i>	<i>Clostridium celerecrescens</i> <i>Clostridium difficile</i> <i>Clostridium histolyticum</i> <i>Clostridium innocuum</i> <i>Clostridium paraputrificum</i> <i>Clostridium perfringens</i> 427.2104 829.6192 <i>Clostridium ramosum</i> <i>Clostridium septicum</i> <i>Clostridium sporogenes</i> 896.7072† <i>Clostridium tertium</i>
			Peptostreptococcaceae 496.4124 497.4214 498.4244 635.3944† 645.4133 646.4173 681.3923	<i>Parvinomas</i> <i>Peptoniphilus</i>	<i>Parvinomas micra</i> <i>Peptoniphilus harei</i>
	Negativicutes 461.3394 628.4443 851.7352	Selenomonadales	Acidaminococcaceae 199.1696 643.4343 † 644.4383† 730.4652	<i>Acidaminococcus</i>	<i>Acidaminococcus fermentans</i>
			Veillonellaceae 213.1855 439.3795 641.4563 655.4713 † 669.4883†	<i>Dialister</i> <i>Veillonella</i>	<i>Dialister sp</i> <i>Veillonella atypica</i> <i>Veillonella dispar</i> <i>Veillonella parvula</i> <i>Veillonella ratti</i>

† Isotope peak, * molecular identity known (see Table 29), ‡ likely interference from mammalian cells, bold font indicates relative spectral abundance >50%

Table 29. Molecular assignment of taxon-specific markers labelled with an asterisk in Table 28.

<i>m/z</i>	TSM molecular identity	<i>m/z</i>	TSM molecular identity
242.1554	2-Heptylquinoline-4(1H)-one	766.5574	<i>B. fragilis</i> α -galactosylceramide
258.1508	2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS)	771.5172	PG(36:3)
271.2284	FA16:0 plus 16Da (O)	773.5332	PG(36:2)
299.2595	FA18:0 plus 16Da (O)	780.5732	<i>B. fragilis</i> α -galactosylceramide
324.2545	Dihydroceramide -H ₂	797.7762	Mycolic acid
328.2857	Dihydroceramide	799.7842	Mycolic acid
465.3024	Cholesterol sulfate	825.8083	Mycolic acid
481.4630	Mycolic acid	904.7051	C15:0 substituted PG Dihydroceramides
491.4474	Mycolic acid	918.7191	C15:0 substituted PG Dihydroceramides
509.4943	Mycolic acid	932.7332	C15:0 substituted PG Dihydroceramides
531.3654	Rhamnolipid	933.7362	C15:0 substituted PG Dihydroceramides, Isotope
537.5259	Mycolic acid	944.7342	C15:0 substituted PG Dihydroceramides
590.4923	Ceramide plus 16Da (O)	945.7372	C15:0 substituted PG Dihydroceramides, Isotope
604.5083	Ceramide	946.7472	C15:0 substituted PG Dihydroceramides
639.4954	Homologue <i>m/z</i> 653.5113 (-CH ₂)	947.7502	C15:0 substituted PG Dihydroceramides, Isotope
646.4833	PE Plasmalogen	958.7461	C15:0 substituted PG Dihydroceramides
653.5113	Sphingolipid	959.7501	C15:0 substituted PG Dihydroceramides, Isotope
677.5238	Ceramide Phosphorylethanolamine	961.7661	C15:0 substituted PG Dihydroceramides, Isotope
691.5395	Ceramide Phosphorylethanolamine	962.7631	C15:0 substituted PG Dihydroceramides
703.4923	PG Plasmalogen	984.9051	Mycolic acid dimer (<i>m/z</i> 491)
705.5562	Ceramide Phosphorylethanolamine	1033.6927	Lichenysin
731.5253	PG Plasmalogen	1034.6766	Surfactin
745.5013	PG(34:2)	1122.1512	Mycolic acid
752.5419	<i>B. fragilis</i> α -galactosylceramide	1166.1774	Mycolic acid

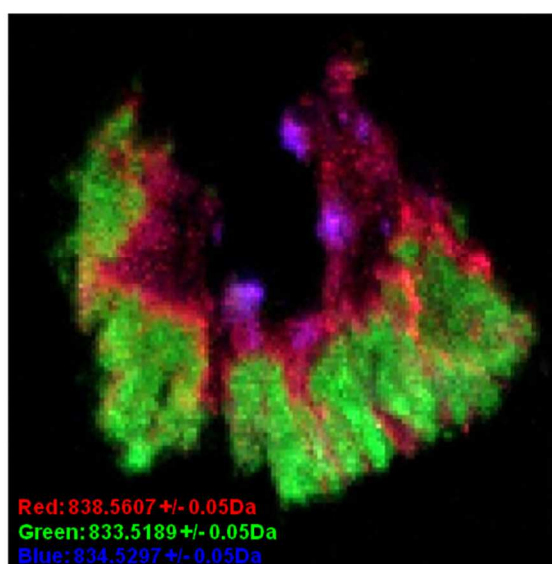
6.4. Detection of bacteria in human colorectal tissue specimens

To demonstrate the applicability of these taxonomical markers for the detection of bacteria in biological samples, we attempted to visualise the presence and distribution of bacteria in human colorectal tissue specimens. This was performed by generating single ion images for the taxonomical markers that are listed in Table 28. Bacteria are known to cover the mucosal membranes in the gut and the gut microbial community is arguably most extensively studied and characterised. Among cancerous specimens, bacteria were largely found localised in such areas that were identified as necrotic by histopathological examination of the H&E stained tissue sections. However, bacteria were also frequently detected along healthy mucosa. An example of both will be further discussed below.

6.4.1. Detection of bacteria in necrotic tissue

Figure 101A and B shows the tissue type-distribution of a cancerous tissue specimen A35 that originated from the centre of tumour dissected during a right hemicolectomy. Histopathological examination revealed the presence of cancerous and stromal tissue. Smaller amounts of necrotic tissue and lymphoid aggregates were observed.

A) DESI-MSI Single ion image



B) H&E of same tissue section

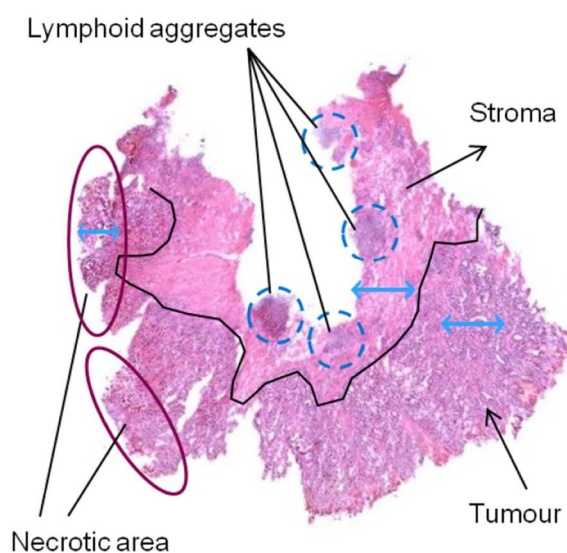


Figure 101. A) DESI-MS ion image displaying tissue type distribution in colorectal tissue specimen A35; green = tumour, red = stroma, blue = lymphoid aggregates. B) H&E stained and histopathologically annotated section post-DESI. Areas for average spectrum generation (see Figure 102) indicated using blue arrows.

Mass spectra of the necrotic tissue area as well as surrounding cancerous and stromal tissue are shown in Figure 102 and display a markedly different phospholipid composition for the necrotic area compared to viable human tissues, namely a significantly reduced glycerophospholipid content and a variety of lower molecular weight sphingolipid-derived taxonomic marker species in the mass range of m/z 500-700.

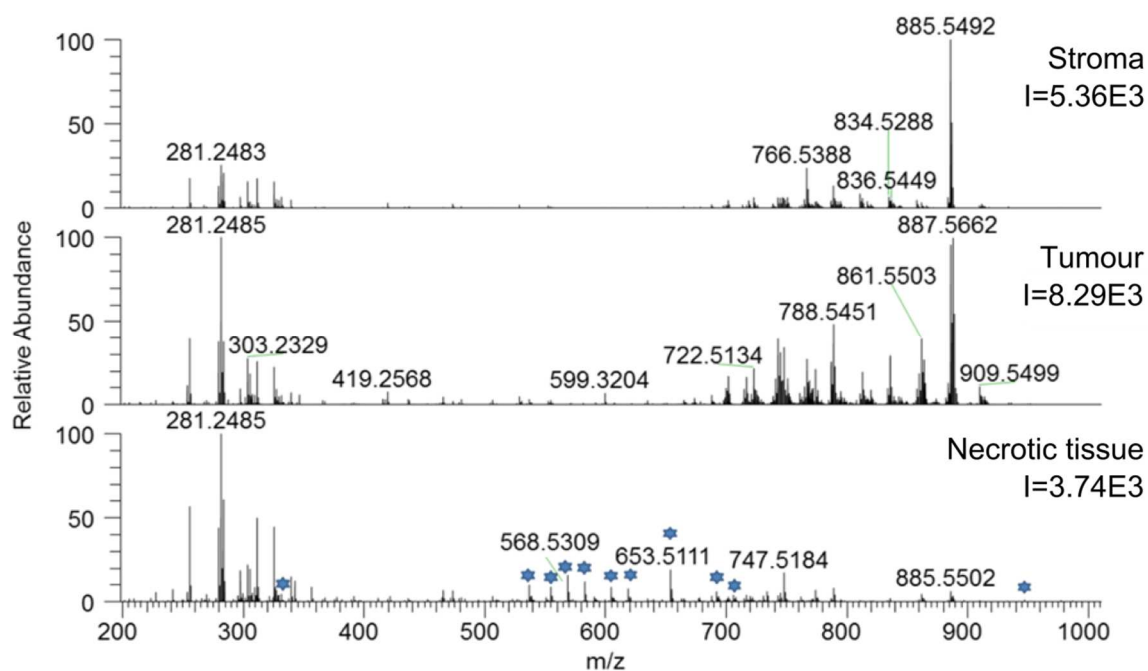


Figure 102. Full scan mass spectra for colorectal adenocarcinoma, tumour surrounding stroma and necrotic tissue of same tissue section shown in Figure 101. Blue stars indicate major taxonomic markers.

Areas chosen for generating average mass spectra indicated in Figure 101 B) by blue arrows.

An array of single ion images of taxon-specific markers is displayed in Figure 103 for markers of which the molecular identity is known. The majority of markers found could be attributed to sphingolipid species originating from members of the Bacteroidetes phylum. Dihydroceramide (m/z 381.2765) is found to be predominantly located into two areas on the left of the tissue specimen while sphingolipids at m/z 639.4954 and 653.5113 are additionally found in a third hotspot closer to the center of the tissue section. Ceramide phosphorylethanolamine species at m/z 677.5238, 691.5395 and 705.5560 are showing co-localised distribution to sphingolipids at m/z 639 and 653. Iso-C15:0-substituted phosphoglycerol dihydroceramides were found to be specific for the Porphyromonadaceae family (part of Bacteroidetes phylum), which in this study were only represented by *Parabacteroides* spp., however, these compounds are reportedly present in high abundance in *Porphyromonas gingivalis*,⁽¹⁸⁰⁾ suggesting general applicability of this marker for this

family. The distribution of the homologues series of signals at m/z 932.7332, 946.7472 and 962.7631 slightly differs from that of earlier mentioned sphingolipid compounds as it shows a closely situated but non-identical hotspot pattern. Notably, the signal at m/z 962.7631 shows a different localisation from its shorter chain homologues suggesting the influence of at least two different types of bacteria with different amounts of these compounds as compared to the composition reported for members of the *Parabacteroides* genus. Members of the Bacteroidetes phylum were reported in metagenomic studies to be accountable for up to 50% of the gut microbial community.(252,253) However, taxon-specific markers for *Bacteroidetes fragilis* were not detected suggesting that the Bacteroidetes bacteria present do not contain a high amount of the opportunistic pathogen *B. fragilis*.

Further groups of bacteria detected include members of the Clostridia class of Firmicutes bacteria (taxon-specific markers at m/z 703.4923 and 731.5253 corresponding to plasmalogens species). The marker at m/z 731.5253 showed the most extensive distribution pattern suggesting colonisation with strong growth of Clostridia over wider tissue areas than for Bacteroidetes. Further two hotspots of this marker can be seen at the bottom right periphery of the tissue section which did not seem to be colonised by detectable amounts of Bacteroidetes. By detecting a plasmalogens species at m/z 646.4833, Fusobacteria could be further detected in similar locations to the Bacteroidetes bacteria.

All of these detected bacterial classes are capable of living under anaerobic conditions and were reported to be major components of the human gut microbiome.(252) The large bacterial presence observed in the necrotic tissue areas is tentatively associated with the lack of immunoresponse of the human body, which enables bacteria to grow and multiply nearly uncontrolled. However, distributions of bacteria over the tissue section were seen to be non-identical for different types of bacteria which highlights the benefit of the proposed taxon-specific marker approach over 16S rRNA sequencing-based methods by retaining spatial information and allowing for more detailed study of compositional changes of the microbiome such as close to cancerous lesions. Although localisation of bacteria using FISH tests is expected to be more sensitive and allow for visualisation of even low densities of bacterial cells, the proposed methodology offers untargeted and multiplexed detection of bacteria.

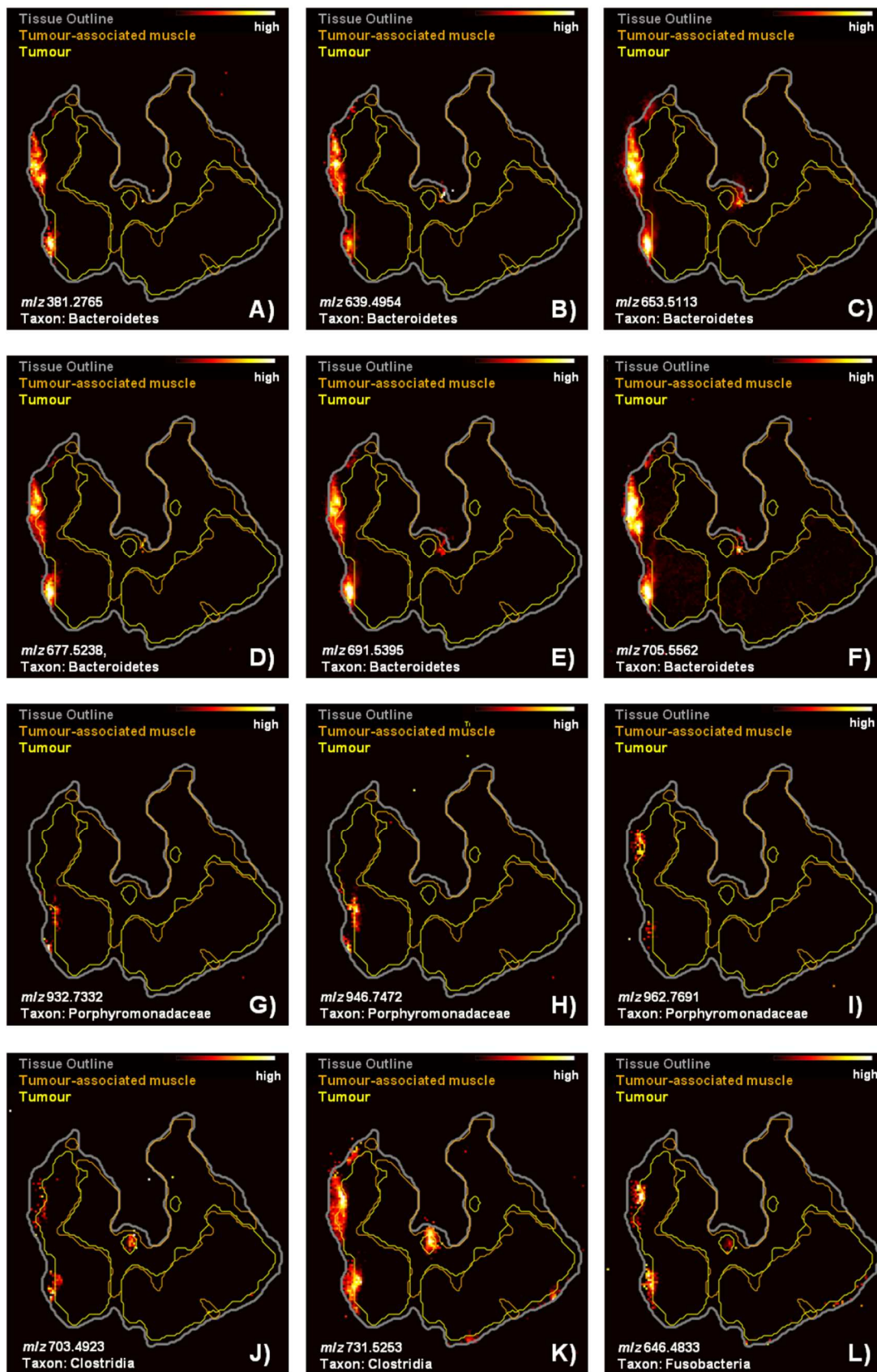


Figure 103. Single ion images for known taxon-specific markers (TSM) of sample A35 center of tumour. A to F) TSM for taxon Bacteroidetes, G to I) TSM for taxon Porphyromonadaceae, J and K) TSM for taxon Clostridia, L) TSM for taxon Fusobacteria. Spatial resolution: 65 μ m.

6.4.2. Detection of bacteria in healthy mucosa

Figure 104 shows the tissue type-distribution of a healthy tissue specimen A56 10cm that originated from a right hemicolectomy. It originated from healthy colon tissue 10cm distant from the centre of tumour. Histopathological examination revealed healthy mucosa and submucosa, divided by the muscularis mucosae layer. Figure 104 shows single ion images for a subset of taxon-specific markers that were detected in this sample. Taxon-specific markers were predominantly detected in several hotspots in two separated pockets of submucosa. Generally, far less and less intense TSM signals were observed in those submucosa regions than in case of necrotic tissue shown in the previous example. This is tentatively attributed with the healthy immunoresponse that restricts unlimited bacterial growth as is more likely to happen in the necrotic tissue specimen. The highest number of taxon-specific markers could again be detected for the Bacteroidetes phylum, however, markers for Proteobacteria and Clostridia were additionally detected.

6.4.3. Detection of bacteria in entire colorectal tissue sample set

To assess the correlation between the proposed taxon-specific marker approach and 16S rRNA sequencing-based community analysis, results of both methodologies were compared for samples from 16 patients and each three tissue locations (centre of tumour, 5cm and 10cm distant from tumour). Sequencing data was recorded as part of the Imperial College BRC Stratified Medicine project and was kindly provided by Mr Reza Mirnezami and Mr James Kinross. Results for samples with information available from both methodologies are shown in Table 30 and Table 31 for results on phylum- and class-level, respectively. Detailed lists of those taxon-specific markers detected in each sample can be found in Appendix 5.

Bacteria of at least one phylum could be visualised in 87.23% of analysed colorectal specimens both including healthy and cancerous tissues. No bacterial markers were detected in 6 out of a total of 47 samples (A17 5cm, A22 5cm, A31 5cm, A49 5cm, A51 CT, A55 5cm). In agreement with gut community compositions reported in literature,^(252,253) main bacterial species detected using next generation sequencing are Bacteroidetes, Proteobacteria and Firmicutes. A small number of samples feature larger copy numbers of Fusobacteria while generally small copy numbers of Actinobacteria are detected. Due to lack of taxon-specific markers of sufficient universal applicability in human matrices, no Fusobacteria were detected in this sample set using DESI-MSI.

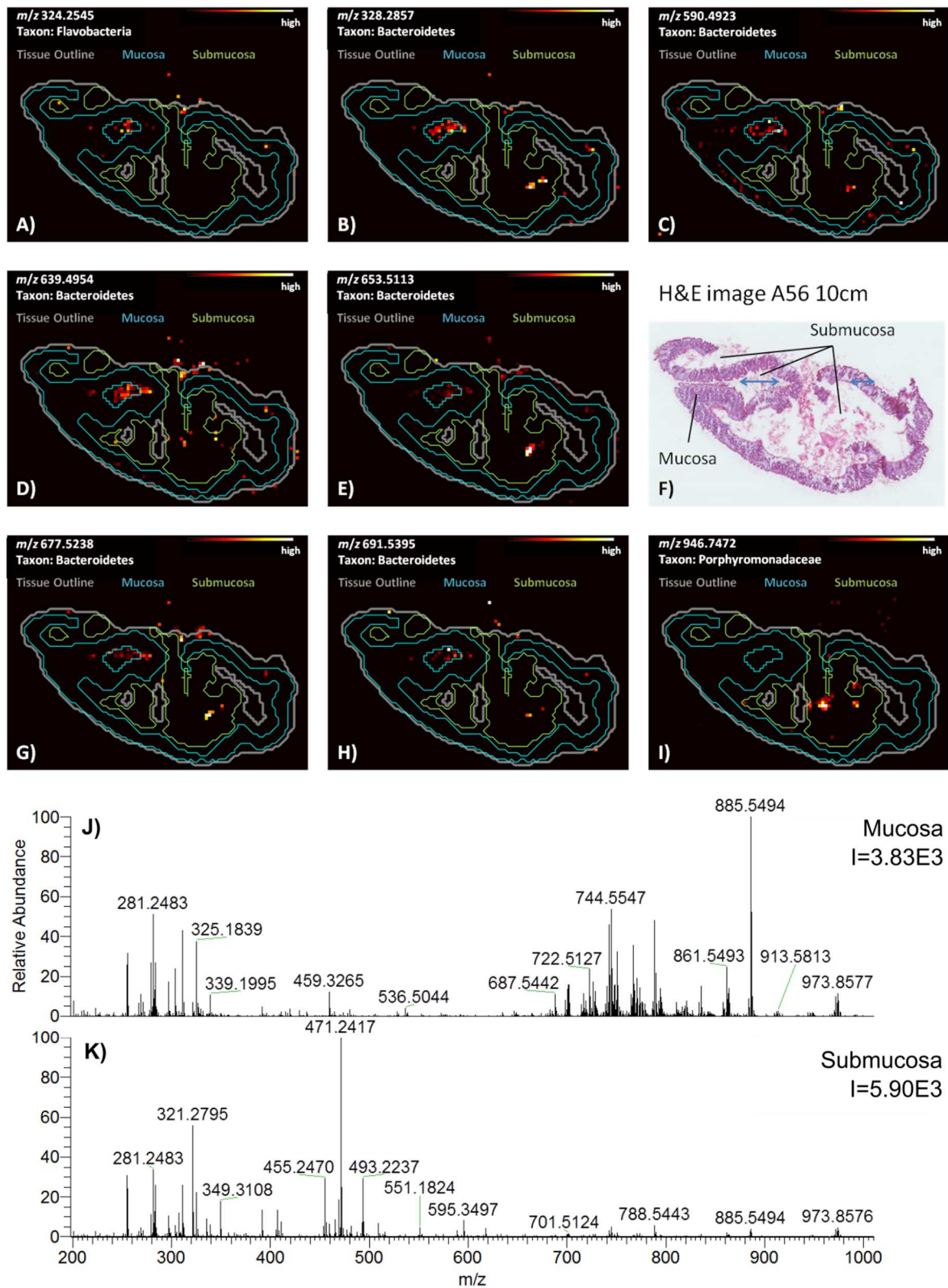


Figure 104. Single ion images of taxon-specific markers (TSM) detected in healthy colorectal tissue specimen A56 10cm and F) H&E image of same tissue section. A) TSM detected for taxon *Flavobacteria*, B to H) TSM detected for taxon *Bacteroidetes*, I) TSM detected for taxon *Porphyromonadaceae*. Spatial resolution: 110 μ m. J) and K) average spectra for mucosa and submucosa showing bacterial TSM, areas for spectrum generation indicated in F) by blue arrows.

In general, all bacterial phyla detected using DESI-MSI and the proposed taxon-specific marker approach were confirmed as present in the same patient sample using sequencing-based analysis (see Table 30). Although a certain degree of correlation between copy numbers and number of detected taxon-specific markers can be found in some patient samples, this is not true in all of the cases. However, due to the inherently different methodologies which were used on spatially different parts of the same sample, numbers given in Table 30 are not strictly comparable and sequencing-derived results are only displayed for validation of the presence of a certain taxon in question. The number of taxon-specific markers found cannot be directly compared to the copy-numbers found per bacterial taxon using molecular methods as the number of taxon-specific markers is not related to number of bacterial cells. Instead, the number of TSM found per taxon is an indication of the homogeneity of bacterial groups within each taxon and heterogeneity in comparison with other taxa. However, TSMs for the same taxon showing similar localisation patterns are increasing the confidence in detection of the bacterial group in question. The absolute intensity of each TSM in turn is potentially related to absolute bacterial cell numbers, however, additional experiments are required that would allow assessment of the quantitative nature of TSMs.

Bacteria that could be visualised using TSMs were in most cases found to be highly localised into histological features of less than a millimetre in size. As sequencing-based analysis involves tissue homogenisation, such features might get missed and/or diluted. DESI-MSI analysis of TSM clearly shows large spatial heterogeneity in bacterial composition and densities which complicates comparison of data obtained from two non-identical areas of tissues even within the same macroscopic tissue specimen. As bacterial markers detected by DESI were associated with tissue necrosis in many cases, detected bacteria most likely originate from a local overpopulation of areas with larger cell densities rather than the commensal bacterial community. To further investigate this, future work using FISH (fluorescent *in-situ* hybridisation) tests on serial tissue sections are necessary to assess whether bacteria are found over the entire tissue section but at significantly increased density in those areas where bacteria were detected using DESI-MSI or whether bacteria can indeed be only found in the latter areas. This will allow an assessment of sensitivity of the taxon-specific marker approach in direct infusion mass spectrometry techniques such as DESI-MSI.

Table 30. Comparison of number of taxon-specific markers detected by DESI-MSI and copy numbers of bacteria detected by 16S rRNA sequencing-based community analysis for the same colorectal tissue specimens on phylum-level.

	Number of taxon-specific markers detected by DESI-MSI				Bacteria detected by 16S rRNA sequencing-based community analysis				
	Bacteroidetes	Proteobacteria	Actinobacteria	Firmicutes	Bacteroidetes	Proteobacteria	Actinobacteria	Firmicutes	Fusobacteria
A14 10CM	1	3	0	3	608	836	6	2425	5
A14 5CM	2	3	0	0	297	197	2	686	2
A14 CT	3	6	2	4	557	1386	15	3453	171
A15 10CM	0	3	0	1	650	61	2	1129	62
A15 5CM	2	1	0	1	1105	114	18	1922	160
A15 CT	3	2	0	0	758	64	17	1745	986
A17 10CM	5	5	1	0	1965	283	2	2190	102
A17 5CM	0	0	0	0	661	146	7	917	60
A17 CT	13	9	2	3	2300	98	5	599	1782
A22 10CM	5	2	0	0	310	56	28	1297	18
A22 5CM	0	0	0	0	1185	216	25	2324	275
A22 CT	7	1	2	2	1057	212	32	2176	466
A31 10CM	0	0	0	0	248	1061	3	710	6
A31 5CM	1	1	0	0	805	1032	6	532	20
A31 CT	8	5	0	1	1597	733	19	1951	61
A36 10CM	8	9	0	3	1156	101	6	1258	6
A36 5CM	4	0	0	2	906	19	1	1379	0
A36 CT	2	9	4	1	876	23	2	1700	6
A37 10CM	7	9	2	1	236	83	14	1881	2
A37 5CM	7	4	0	4	255	277	46	1369	150
A37 CT	2	5	0	1	226	112	6	1469	4
A38 10CM	6	2	0	0	1478	1151	10	1421	2
A38 5CM	9	2	0	3	585	441	2	386	0
A38 CT	13	6	1	5	3572	102	11	621	749
A42 10CM	4	2	0	1	1207	210	23	5027	50
A42 5CM	5	3	0	1	584	222	3	2468	230
A42 CT	6	3	1	2	1432	394	8	2850	233
A49 10CM	8	6	2	2	551	414	33	1812	98
A49 5CM	0	0	0	0	829	195	30	2446	240
A49 CT	10	9	3	6	380	97	16	1188	392
A51 10CM	4	0	0	2	360	103	10	2368	39
A51 5CM	10	0	1	4	140	43	4	1007	27
A51 CT	0	0	0	0	163	125	7	2531	132
A55 10CM	4	4	0	2	378	1622	39	103	0
A55 5CM	0	0	0	0	858	2240	88	504	42

	Number of taxon-specific markers detected by DESI-MSI				Bacteria detected by 16S rRNA sequencing-based community analysis				
	Bacteroidetes	Proteobacteria	Actinobacteria	Firmicutes	Bacteroidetes	Proteobacteria	Actinobacteria	Firmicutes	Fusobacteria
A55 CT	0	2	0	0	307	919	10	362	1
A56 10CM	20	7	0	4	1035	340	2	1099	17
A56 5CM	8	2	0	2	2049	706	35	2002	31
A56 CT	17	3	0	1	1195	847	6	1488	40
A57 10CM	3	0	0	1	623	33	14	3167	0
A57 5CM	19	2	1	7	809	34	18	2435	4
A57 CT	6	2	1	2	607	16	9	1602	4
A62 5CM	6	0	0	1	1337	501	8	1146	0
A62 CT	6	2	0	3	1032	46	2	1330	2
A64 10CM	5	2	0	0	245	454	5	638	4
A64 5CM	24	11	4	6	994	1248	10	1969	6
A64 CT	23	7	3	7	569	1053	9	1108	14

Table 31. Comparison of number of taxon-specific markers detected by DESI-MSI and copy numbers of bacteria detected by 16S rRNA sequencing-based community analysis for the same colorectal tissue specimens on class-level.

	Number of taxon-specific markers detected by DESI-MSI										Bacteria detected by 16S rRNA sequencing-based community analysis										
	Bacteroidia	Flavobacteria	Alphaproteobacteria	Betaproteobacteria	Epsilonproteobacteria	Gammaproteobacteria	Actinobacteria (class)	Bacilli	Clostridia	Negativicutes	Bacteroidia	Flavobacteria	Alphaproteobacteria	Betaproteobacteria	Epsilonproteobacteria	Gammaproteobacteria	Actinobacteria (class)	Bacilli	Clostridia	Negativicutes	Fusobacteria (class)
A14 10CM	0	1	0	0	0	3	0	1	2	0	596	11	7	256	0	573	6	1	2421	0	5
A14 5CM	2	0	1	0	2	0	0	0	0	0	294	3	1	29	0	167	2	2	682	0	2
A14 CT	1	2	0	1	4	1	2	1	2	2	525	32	33	516	0	837	15	159	3286	0	171
A15 10CM	0	0	0	0	2	1	0	0	1	0	648	2	1	27	0	33	2	10	1119	0	62
A15 5CM	2	0	0	0	0	1	0	0	1	0	1105	0	2	51	3	58	18	23	1896	0	160
A15 CT	3	0	1	0	0	1	0	0	0	0	758	0	0	36	18	10	17	3	1742	0	986
A17 10CM	4	1	0	0	2	3	1	0	0	0	1940	25	40	130	5	108	2	280	1843	3	102
A17 5CM	0	0	0	0	0	0	0	0	0	0	635	23	33	39	17	57	7	48	849	0	60
A17 CT	10	3	0	0	4	5	2	1	1	1	2287	13	6	46	31	15	5	89	509	0	1782
A22 10CM	2	3	1	0	1	0	0	0	0	0	310	0	5	29	0	16	28	6	1288	0	18
A22 5CM	0	0	0	0	0	0	0	0	0	0	1143	41	55	107	0	54	25	14	2303	0	275
A22 CT	5	2	0	0	0	1	2	1	1	0	1056	0	47	125	0	40	32	36	2118	0	466
A31 10CM	0	0	0	0	0	0	0	0	0	0	237	10	6	447	0	608	3	142	321	0	6
A31 5CM	0	1	0	0	1	0	0	0	0	0	758	46	8	475	0	549	6	59	442	0	20
A31 CT	6	2	0	0	3	2	0	0	0	1	1594	3	1	291	1	440	19	42	1779	0	61
A36 10CM	7	1	3	0	0	6	0	2	1	0	1139	15	0	46	2	53	6	3	1229	0	6
A36 5CM	4	0	0	0	0	0	0	0	2	0	887	19	4	12	0	3	1	1	1337	0	0
A36 CT	2	0	2	1	4	2	4	0	1	0	834	40	0	4	4	14	2	14	1671	0	6
A37 10CM	2	5	4	1	3	1	2	0	1	2	233	2	38	3	0	42	14	4	1877	0	2
A37 5CM	5	2	1	0	0	3	0	1	2	1	160	16	83	59	9	94	46	89	1280	0	150
A37 CT	2	0	1	0	3	1	0	0	0	1	226	0	57	11	1	43	6	9	1460	0	4
A38 10CM	3	3	0	0	1	1	0	0	0	0	1463	4	2	260	0	886	10	268	1082	45	2
A38 5CM	6	3	0	0	1	1	0	0	2	1	582	0	0	58	0	383	2	76	295	12	0
A38 CT	8	5	0	0	1	5	1	0	4	1	3572	0	0	35	35	31	11	125	487	3	749

	Number of taxon-specific markers detected by DESI-MSI										Bacteria detected by 16S rRNA sequencing-based community analysis										
	Bacteroidia	Flavobacteria	Alphaproteobacteria	Betaproteobacteria	Epsilonproteobacteria	Gammaproteobacteria	Actinobacteria (class)	Bacilli	Clostridia	Negativicutes	Bacteroidia	Flavobacteria	Alphaproteobacteria	Betaproteobacteria	Epsilonproteobacteria	Gammaproteobacteria	Actinobacteria (class)	Bacilli	Clostridia	Negativicutes	Fusobacteria (class)
A42 10CM	3	1	0	0	1	1	0	0	0	1	1202	5	11	130	2	67	23	601	4392	0	50
A42 5CM	4	1	0	0	2	1	0	0	1	0	584	0	0	124	7	85	3	302	2150	0	230
A42 CT	4	2	0	0	2	1	1	0	1	1	1432	0	0	319	8	67	8	297	2523	0	233
A49 10CM	6	2	0	0	4	2	2	0	2	0	446	96	55	200	1	158	33	215	1537	5	98
A49 5CM	0	0	0	0	0	0	0	0	0	0	826	3	19	113	4	59	30	978	1299	15	240
A49 CT	5	5	0	2	4	3	3	0	4	2	379	1	0	30	51	16	16	717	430	0	392
A51 10CM	2	2	0	0	0	0	0	1	1	0	345	15	7	21	0	75	10	16	2349	0	39
A51 5CM	8	2	0	0	0	0	1	0	4	0	140	0	4	1	0	38	4	3	1002	0	27
A51 CT	0	0	0	0	0	0	0	0	0	0	160	3	14	50	1	60	7	20	2508	0	132
A55 10CM	3	1	1	1	0	2	0	0	2	0	28	347	216	638	0	768	39	6	96	0	0
A55 5CM	0	0	0	0	0	0	0	0	0	0	181	126	63	262	0	594	10	2	356	0	1
A55 CT	0	0	0	0	1	1	0	0	0	0	462	392	310	926	2	1002	88	82	412	0	42
A56 10CM	16	4	0	1	0	6	0	1	3	0	1029	6	0	153	1	186	2	12	1076	0	17
A56 5CM	7	1	0	0	1	1	0	0	2	0	2009	29	25	281	1	399	35	34	1937	0	31
A56 CT	14	3	0	0	1	2	0	0	1	0	1194	1	0	165	0	681	6	14	1467	0	40
A57 10CM	2	1	0	0	0	0	0	0	0	1	623	0	2	24	0	4	14	9	3158	0	0
A57 5CM	14	5	2	0	0	0	1	1	4	2	798	11	6	14	0	9	18	10	2420	0	4
A57 CT	4	2	0	1	0	1	1	0	1	1	607	0	0	14	0	2	9	1	1601	0	4
A62 5CM	4	2	0	0	0	0	0	0	1	0	1328	2	3	48	2	448	8	13	1127	0	0
A62 CT	4	2	0	0	1	1	0	1	2	0	1027	0	0	28	0	18	2	28	1297	0	2
A64 10CM	3	2	0	0	1	1	0	0	0	0	241	1	27	12	0	415	5	12	608	0	4
A64 5CM	20	4	1	1	4	5	4	0	4	2	972	6	73	25	0	1149	10	22	1885	0	6
A64 CT	17	6	1	0	2	4	3	1	5	1	556	3	75	17	0	961	9	12	1086	0	14

This study demonstrates that molecular species differ significantly between microbial lipidomes and the human tissue lipidome. Based on a database of REIMS lipid profiles of both microorganisms and human cell lines, taxon-specific markers were found for a variety of bacterial species at different taxonomic levels. These markers were shown to be absent in human tissues and can thus be used to visualise the presence of bacteria in human samples such as presented for human colorectal tissues. It was further demonstrated that taxonomic markers derived by the REIMS technique can be used in conjunction with other mass spectrometric ionisation techniques capable of detecting the same molecular species such as DESI, which is more suitable to the direct analysis of mucosal sampling devices or mass spectrometry imaging. The presented approach can be extended to a wide variety of applications as the characterisation of microbial communities during endoscopic interventions using REIMS, and for the detection of microbes in human sterile and non-sterile biofluid samples to facilitate appropriate antibiotic treatment. Although currently only a small number of taxon-specific markers was detected, this number is likely to increase with increasing number of database entries and by using molecular validation of microorganism identity during REIMS database building. Further markers are additionally expected in the higher mass range which was not analysed as part of this study (m/z 1000-2000). Eventually, larger number of entries per bacterial species and increased variability of culturing conditions will enable the establishment of conservative taxon-specific markers and allow the assessment of using these markers to quantify the amount of certain bacterial species.

Summary

A novel REIMS-based mass spectrometric method was developed for the purpose of identifying and characterising unicellular organisms. A pair of bipolar forceps with an irrigation port directly connected to the mass spectrometer inlet was found to be most suited for reproducible and sensitive generation of mass spectral data. Spectral profiles obtained using REIMS were found to be highly independent from most instrumental parameters such as the electrosurgical power supply model and its power settings, instrument tube lens and capillary voltages and mass analyser. However, adaptations were necessary for implementation on a Waters platform which was chosen as the instrument for commercialisation of the proposed technology. These included insertion of a heated collision surface into the atmospheric pressure interface to remove instrument contaminants.

The technology was found to be suitable to identify bacteria and yeasts directly from the agar culturing plate without any further sample pre-treatment and acquisition times of just 3-5 seconds. The technology was found to perform equally well for Gram-negative, Gram-positive bacteria and yeasts although the latter two contain far thicker cell walls than Gram-negative bacteria. Specificity for 28 bacterial species was found to be comparable to specificity obtained using the Bruker BioTyper at 95.8% on species-level. Species ID was obtained using the BioTyper which is why performance of REIMS was not expected to exceed performance by MALDI-TOF-MS.

Deviations in the sampling workflow were necessary in case of liquid cultures and cell lines in which case the organism biomass was centrifuged to a cell pellet before analysis. Mass spectrometric profiles were found to differ from those of the same species grown on solid culturing medium and thus it is suggested to include spectra obtained from liquid culture into a future spectral database. However, changes were largely found to be due to relative changes in intensity of spectral features rather than presence or absence of signals. The same was found to be true for different culturing media, although changes were found to be less significant in this case.

Different degrees of subspecies-specificity was found in several cases such as ribotypes in *C. difficile*, serotypes in *S. pneumoniae*, and antibiotic resistance for *K. pneumoniae*.

Direct on-sample microbial detection was found to be feasible using a taxon-specific marker approach in which bacteria are identified based on biomarkers that were determined to be uniquely specific at a certain taxonomic level. For this purpose, a bacterial spectral database was built comprising REIMS spectral profiles of 217 bacterial species. A strategy involving ANOVA tests followed by Tukeys Honestly Significant Difference test was used on all taxonomic levels to determine biomarker compounds. Using these taxon-specific markers, bacteria were detected in DESI-MS imaging datasets in both healthy and cancerous tissue specimens. Particularly high microbial density was found in necrotic tissue areas. Detected bacterial species were found to correlate with bacterial species known to be present in the gut and results of 16S rRNA-based community analysis. This shows that lipid-based microbial markers detected using REIMS can be equally applied to other mass spectrometric techniques that are a suitable means for obtaining information on the lipid content of a sample.

The methodology was further applied to the characterisation of the NCI60 human cancer cell line panel to obtain shotgun lipidomic information on all members of the panel. REIMS spectral profiles were shown to be sufficiently specific to identify different cancer cell lines as shown using PCA models. However, cell line lipid profiles were found to significantly differ from those obtained from *ex-vivo* tissue specimen. REIMS lipid profiles were correlated with gene expression data and good positive correlation was found with genes involved in lipid metabolism as shown for *fads2* and *ugcg* genes. Discussions are undergoing to make REIMS data publicly available by incorporating it into the CellMiner database.

Impact

The proposed REIMS-based microbial identification and characterisation tool was proven to be capable of performing similar tasks as the market-leading commercial Bruker BioTyper platform, namely specific identification of bacteria with >95% accuracy on species-level. Furthermore, the technique seems to be a promising platform to perform subspecies identifications such as strain-level identification, or detection of ribotype or antibiotic resistance. In many cases, this information can be found in the same spectral profiles without having to introduce further culturing or sample preprocessing steps. An inherent advantage of REIMS is its robustness and simple sampling probe making it amenable to automation. During the course of this PhD work and based on results obtained in this study, an automated sampling robot was developed in collaboration with the Waters Research Center (Budapest, Hungary) based on a modified Tecan Evo colony picking unit. This unit was installed at Imperial College for characterisation and testing in June 2015 and enables acquisition of REIMS spectral profiles and sampling for 16S rRNA sequencing of up to 1000 strains per day. This corresponds to the numbers of bacterial strains that are processed in a large clinical microbiology unit such as in Charing Cross Hospital, London per day. These developments make REIMS a promising alternative for bacterial identification when compared to the MALDI-based systems. However, initial culturing steps are still necessary, thus no further time-savings can be obtained as compared to MALDI-TOF-MS for species-level identification. However, potentially the time needed and costs associated with follow-up experiments for subspecies-typing could be significantly reduced using a REIMS-based workflow. One significant advantage of the automated REIMS method is that it is equally well suited to the analysis of bacteria and yeasts which currently require different sample pre-treatment steps in order to provide accurate species identification.

However, the largest impact on changing clinical practice could be obtained in case of direct detection of bacteria in clinical samples without the need for culturing steps which could reduce the time-demand needed for identification by a further 24hrs and would allow identification of causative agents on the same day as samples are received in the microbiology laboratory. Mass spectrometric technologies can be envisioned working in GP offices and would enable better diagnostic performance in minutes. The taxon-specific marker approach for detection of bacteria in complex matrices was found to be able to detect

bacteria in healthy and cancerous human tissue specimens analysed using DESI-MSI. This indicates that bacterial biomarkers derived from REIMS-spectral profiles can be detected using alternative and more sensitive means for lipid detection. While DESI and MALDI-MS are particularly interesting for tissue imaging applications, screening of taxonomic markers using LC-MS would be a promising alternative to identify causative agents of septicemia in blood samples or meningitidis in cerebrospinal fluids. The proposed taxon-specific marker approach can possibly also be applied for *in-vivo* REIMS applications such as in endoscopic examinations to identify overgrowth of certain species in a bacterial community. Technologies are currently developed at Imperial College that aim to enable the detection of mucosal abnormalities using DESI-MS of functionalised clinical swabs. This method combined with the taxonomical-marker approach could be used to identify for instance respiratory, genital or urinary tract infections.

An automated REIMS-based approach can be further used to screen for successful gene insertion into genetically modified bacteria and yeasts in which case thousands of colonies would have to be screened for successful production of the desired compounds. An example would be identification of best producers for biodegradable polymers such as polyhydroxyalkanoates or best producers of rhamnolipid biosurfactants for biosynthetic applications. Such applications are becoming more and more important in a society that becomes increasingly environmentally conscious.

Future work

Microorganisms

The presented work is proof-of-principle data which triggered the development of a commercially available REIMS-based microbial identification and characterisation tool. Waters Corporation expressed intentions in developing instrumentation and funded appropriate instrumentation and personnel to develop an extensive validated spectral database covering the most common human pathogens and commensals. This database will comprise >50,000 entries, each complemented by gene sequencing results, and relevant phenotypic test results such as antibiotic resistance, ribotypes or serotypes. Once comprehensive datasets will have been acquired, the database can be queried in order to test sensitivity and specificity for different applications. This will help establishing where this REIMS-based approach will fit into routine clinical microbiology.

An enlarged and detailed database will be the central piece of future studies as well as a commercial identification system. With increasing database size, more comprehensible sensitivity testing can be performed which will allow for a direct comparison with MALDI-TOF-MS and sequencing analyses. This will be accompanied by the development of an appropriate identification algorithm for accurate species identification. A comprehensive database is crucial for this development as the amount of species analysed in this study does not sufficiently reflect the complexity of thousands of different bacterial species. A detailed sampling protocol should be followed that includes initial identification of pure cultures of clinical isolates using MALDI-TOF-MS as part of clinical microbiology routine. Based on initial ID, samples will be sent for sequencing analysis to assign their species unambiguously. The same sample will be frozen in glycerol. Microorganisms will then be grown under several relevant culturing conditions and subjected to automated REIMS analysis. A large scale database will also facilitate the generation of robust taxon-specific markers by having a more populated phylogenetic tree structure. This will likely lead to some taxon-specific markers currently thought to be specific being lost, while other markers for currently underpopulated groups will be found.

As these taxonomical markers are not REIMS-specific they will be suitable for detection using any mass spectrometric technique that is suitable for the detection of lipids as presented here for DESI-MSI datasets. More such applications will be developed and an especially

promising application is the detection of a subset of most relevant taxonomic markers in biofluids using UPLC-MS. One example is detecting the causative agent of bacterial septicemia; it can be expected that both intracellular and extracellular bacterial metabolites can be detected using such a methodology in positive blood cultures by using a short gradient screening run in multiple reaction monitoring (MRM) mode searching only for a small number of taxonomic markers for the most common causative agents of sepsis. Although this would not be a shotgun approach but would require approximately 2-5 minutes per sample, it is still considerably quicker than culturing-based techniques or identification using MALDI-TOF-MS due to the lengthy extraction process required by the latter. This will require the selection of appropriate taxonomic markers and the development of a short UPLC-based screening run for these markers. Limits of detection will then have to be determined for a dilution series of the respective bacteria in blood cultures and in blood samples to determine suitability of this approach and bacterial cell numbers that can be detected.

As presented experiments showed that for the development of a commercial platform it would be beneficial for identification accuracy to include several different culturing conditions such as culturing atmospheres, different culturing media and solid as well as liquid cultures into the database. Especially the latter is expected to greatly enhance identification accuracy for bacteria grown in liquid cultures and directly from clinical specimens. The present sampling setup is not ideally suited for automated analysis of liquid cultures as it requires centrifugation of the liquid cultures to provide a solid cell pellet which can be subjected to REIMS analysis. Thus, a dedicated tool for analysis of such cultures will be developed which will feature separate aspiration ports for the analyte containing aerosol produced during REIMS analysis and the liquid culturing medium (and solution for subsequent liquid discharge). Alternatively, this could be achieved using a three-way valve. Such setup will also be suitable for automation which is an important feature for large scale database collection of several tens of thousands strains per year.

Cell lines

The present proof-of-principle study demonstrated the applicability of a REIMS-based shotgun lipidomic characterisation approach for human cancerous cell lines. Individual cancer cell lines were found to exhibit reproducible and cell line-specific spectral profiles while spectra could be acquired in less than 5s. This does not only allow rapid identification of cell

lines based on their spectral fingerprint, but also detailed characterisation of membrane lipid composition in order to study the changes in the cell membrane composition in different cancer phenotypes. The lipid analyses of the NCI60 cell line panel obtained using REIMS will be published as part of the CellMiner database where it will be available alongside drug sensitivity patterns, DNA mutation analysis results and gene and protein expression data.

By continued analysis of the correlation between REIMS spectral data and gene and protein expression data, the sensitivity of the REIMS spectral method for tumour phenotypic characterisation can be assessed in detail. In addition, this technique offers a unique possibility to investigate gene knock-out models for changes in lipid metabolism or the effects of feeding experiments with stable isotope-labelled nutrient sources. Better bioinformatic tools will be developed that link gene and protein expression profiles to the respective m/z signals in order to identify both positive and negative correlation behaviour for the NCI60 panel as well as for future studies involving cell lines. This will lead to robust links being found between lipid profiles as determined by REIMS and gene/protein expression which will shed light on the sensitivity and specificity that it observed using intra-operative REIMS. Such information will help to gain approval by regulatory bodies and increase confidence into lipid-profiling based diagnostic methods. Due to speed of acquisition and observed specificity, REIMS will have the potential to be a vital tool to better understand the relationship of lipid profiles between cell lines and solid tumours.

Further cell line systems will be explored such as epithelial-mesenchymal transitions in different cancer cell lines (in collaboration with Dr Emre Sayan, University of Southampton, UK), different stages of breast cancer (George Poulagiannis, ICR) and the effect of different cell cycle stages on the lipid composition (currently studied using mouse embryonic stem cells, in collaboration with Prof Amanda Fisher, Imperial College London, UK). All these studies will benefit from being able to correlate REIMS spectral profiles to gene and protein expression data to understand the biochemical basis of observed differences in lipid profiles.

Acknowledgements

Supervision

Prof Zoltan Takats

Dr Jake Bundy

Microbiology

Monica Rebec – for enabling me to work in the microbiology lab and for lots of advice

Ali Abdolrasouli and Tony Rickards – for microbiological training and support with culturing microorganisms

Michael Petrou – for supplying fungal specimens

Adam Burke – for help culturing and analysing bacterial samples

Irene Burkhard – for supplying 30 strains of *S. pneumoniae* of known serotype

Anne L McCartney, Lesley Hoyles – for supplying commensal bacterial species

All staff at Clinical Microbiology lab, CXH – for their kindness and help

Data analysis

Emrys A Jones – for help learning to use Matlab and data extraction scripts

Kirill A Veselkov – for supplying Matlab toolbox to perform multivariate analysis methods

Ottmar Golf – for help solving problems regarding data processing, especially inter-platform comparison

James S McKenzie – for help in biomarker recovery study

Cell lines

Anna Lovrics, Gergely Szakacs, Nora Kucsma, Judith Kessler – for supplying cell lines, help with data analysis and fruitful collaboration

DESI imaging

Reza Mirnezami – for collaboration in study using DESI MSI to characterise colorectal cancer (data recorded by Nicole Strittmatter and Anna Mroz as part of PhD thesis of Mr Reza Mirnezami)

Abigail Speller, Robert D Goldin – for histological assignments of colorectal sample set

Anna Mroz – for help crysectioning, running, staining and scanning DESI imaging samples

Technical support

Waters Research Centre in Budapest, Hungary (formerly Medimass Ltd.) for the technical support, supply of customised atmospheric interfaces, and supply of setup for introduction of IPA.

Steven Pringle for support trouble-shooting Waters instruments.

Funding sources

The work was funded by the European Research Council under the Starting Grant scheme (Contract No. 210356), the European commission FP7 intelligent surgical device project (Contract No. 3054940), the European Research Council Consolidator Grant (Contract No. 617896) and the BBSRC Industrial Partnership Award (Contract No. BB/L020858/1) and Waters Corporation.

List of publications

Published

- [1] **Strittmatter N**, Jones EA, Veselkov KA, Rebec M, Bundy JG, Takats Z. Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chemical Communications* **2013**, 49(55):6188-90. DOI: 10.1039/c3cc42015a.
- [2] **Strittmatter N**, Rebec M, Jones EA, Golf O, Abdolrasouli A, Balog J, Behrends V, Veselkov KA, Takats Z. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. *Analytical Chemistry* **2014**, 86(13): 6555-62. DOI: 10.1021/ac501075f.
- [3] Veselkov KA, Mirnezami R, **Strittmatter N**, Goldin RD, Kinross J, Speller AVM, Abramov T, Jones EA, Darzi A, Holmes E, Nicholson JK, Takats Z. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, 111(3): 1216-21. DOI: 10.1073/pnas.1310524111.
- [4] Abbassi-Ghadi N, Veselkov KA, Kumar S, Huang J, Jones EA, **Strittmatter N**, Kudo H, Goldin RD, Takats Z, Hanna GB. Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. *Chemical Communications* **2014**, 50(28): 3661-4. DOI: 10.1039/c3cc48927b.
- [5] Abbassi-Ghadi N, Jones EA, Veselkov KA, Huang J, Kumar S, **Strittmatter N**, Golf O, Kudo H, Goldin RD, Hanna GB, Takats Z. Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: a gateway for clinical applications. *Analytical Methods* **2015**, 7: 71-80. DOI: 10.1039/C4AY01770F.
- [6] Golf O, **Strittmatter N**, Karancsi T, Pringle SD, Speller AV, Mroz A, Kinross JM, Abbassi-Ghadi N, Jones EA, Takats Z. Rapid evaporative ionization mass spectrometry imaging platform for direct mapping from bulk tissue and bacterial growth media. *Analytical chemistry* **2015**, 87(5): 2527-34. DOI: 10.1021/ac5046752.
- [7] Liebeke M, **Strittmatter N**, Fearn S, Morgan AJ, Kille P, Fuchser J, Wallis D, Palchykov V, Robertson J, Lahive E, Spurgeon DJ, McPhail D, Takats Z, Bundy JG. Unique

metabolites protect earthworms against plant polyphenols. *Nature Communications* **2015**, DOI: 10.1038/ncomms8869.

[8] Guenther S, Muirhead LJ, Speller AV, Golf O, **Strittmatter N**, Ramakrishnan R, Goldin RD, Jones EA, Veselkov K, Darzi A, Takats Z. Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Research* **2015**. DOI: 0008-5472.CAN-14-2258.

[9] Swales JG, Tucker JW, **Strittmatter N**, Nilsson A, Cobice D, Clench MR, Mackay CL, Andren PE, Takáts Z, Webborn PJ. Mass spectrometry imaging of cassette-dosed drugs for higher throughput pharmacokinetic and biodistribution analysis. *Analytical Chemistry* **2014**, 86(16): 8473-80. DOI: 10.1021/ac502217r.

[10] Oetjen J, Veselkov K, Watrous J, McKenzie JS, Becker M, Hauberg-Lotte L, Kobarg JH, **Strittmatter N**, Mróz AK, Hoffmann F, Trede D, Palmer A, Schiffler S, Steinhorst K, Aichler M, Goldin R, Guntinas-Lichius O, von Eggeling F, Thiele H, Maedler K, Walch A, Maass P, Dorrestein PC, Takats Z, Alexandrov T. Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. *GigaScience* **05/2015**. DOI: 10.1186/s13742-015-0059-4.

[11] Ashton S, Song YH, Nolan J, Cadogan E, Murray J, Odedra R, Foster J, Hall P, Low S, Taylor P, Ellston R, Polanska U, Wilson J, Howes C, Smith A, Goodwin RJA, Swales JG, **Strittmatter N**, Takáts Z, Nilsson A, Andren P, Trueman D, Walker M, Reimer CL, Troiano G, Parsons D, De Witt D, Ashford M, Hrkach J, Zale S, Jewsbury S, Barry ST. Aurora kinase inhibitor nanoparticles target tumors with favorable therapeutic index in vivo. *Science Translational Medicine* **2016**, 325(8), p325ra17. DOI: 10.1126/scitranslmed.aad2355.

[12] **Strittmatter N**, Lovrics A, Sessler J, McKenzie JS, Bodai Z, Doria ML, Kucsma N, Szakacs G, Takats Z. Shotgun Lipidomic Profiling of the NCI60 Cell Line Panel Using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry* **2016**, 88(6), 7507-14. DOI: 10.1021/acs.analchem.6b00187.

[13] Abbassi-Ghadi N, Golf O, Kumar S, Antonowicz SSA, McKenzie JS, Huang J, **Strittmatter N**, Kudo H, Jones EA, Veselkov KA, Goldin RD, Takats Z, Hanna GB. Desorption Electrospray Ionization Mass Spectrometry Imaging of Esophageal Lymph Node Metastases. *Cancer Research* **2016**. DOI: 0.1158/0008-5472.CAN-16-0699.

Submitted

[14] Doria ML, McKenzie JS, Mroz A, Phelps D, Speller AVM, Rosini F, Strittmatter N, Golf O, Veselkov KA, Brown R, Ghaem-Maghami S, Takats Z. Epithelial ovarian carcinoma diagnosis by Desorption Electrospray Ionization Mass Spectrometry Imaging. *Scientific Reports* [submitted August 2016]

List of conference presentations

Podium presentations

1: **Strittmatter N**, McKenzie J, Burke A, Rickards T, Rebec M, Takats Z. Taxon-specific markers for the qualitative and quantitative detection of bacteria in human samples. Mass Spectrometry: Applications to the Clinical Lab Meeting 2015 (MSACL 2015 US), San Diego, CA, USA. (March 28 - April 1, 2015)

Young Investigator Travel Grant Awarded.

2: **Strittmatter N**, Rebec M, Jones EA, Golf O, Abdolrasouli A, Balog J, Behrends V, Veselkov KA, Takats Z. *Rapid Characterization and Identification of Clinically Relevant Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry*. Applications to the Clinical Lab Meeting 2014 (MSACL 2014 EU), Salzburg, Austria. (September 2-5, 2014)

Young Investigator Travel Grant Awarded.

3: **Strittmatter N**, Lovrics A, Jones EA, Golf O, Veselkov KA, Szakacs G, Takats Z. *Characterisation of Human Cell Lines Using Rapid Evaporative Ionization Mass Spectrometry*. International Mass Spectrometry Conference 2014, Geneva, Switzerland. (August 24-29, 2014)

4: **Strittmatter N**, Jones EA, Rebec M, Takats Z. *Identification of bacteria using rapid evaporative ionisation mass spectrometry*. 61st Conference of the American Society of Mass Spectrometry 2013 (ASMS), Minneapolis, MN, USA. (June 9-13, 2013)

5: **Strittmatter N**, Takats Z. *From TFME-DESI MS to in-vivo analysis using REIMS: Applications of ambient mass spectrometry to biological samples*. 11th EMSG Ardgour Symposium 2014, Ardgour, UK. (September 8-12, 2014)

6: **Strittmatter N**, McKenzie J, Speller AVM, Burke A, Pruski P, Mirnezami R, Marchesi J, Rebec M, Takats Z. *Qualitative and quantitative detection of bacteria in complex human matrices using taxon-specific markers*. The International Chemical Congress of Pacific Basin Societies 2015 (PacifiChem 2015), Honolulu, Hawaii, USA. (December 15-20, 2015)

Poster presentations

1: **Strittmatter N**, Rebec M, Jones EA, Golf O, McKenzie JM, Veselkov KA, Speller AVM, Takats Z. *Characterization and Identification of Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry and Visualisation of Bacteria in DESI-MS imaging datasets*. Imaging Mass Spectrometry Conference (OurCon2), Antalya, Turkey. (September 18-21, 2014)

2: **Strittmatter N**, Jones EA, Veselkov KA, Rebec M, Curtis S, Marchesi J, Bundy JG, Takats Z. *Identification of intact microorganisms using rapid evaporative ionisation mass spectrometry*. Annual Meeting of the British Mass Spectrometry Society 2013 (BMSS), Eastbourne, UK. (September 9-11, 2013)

3: **Strittmatter N**, Lovrics A, Jones EA, Veselkov KA, Szakacs G, Takats Z, *Characterisation of Human Cell Lines Using Rapid Evaporative Ionization Mass Spectrometry*. Annual British Mass Spectrometry Society Meeting 2014, Alderley Park, Cheshire, UK. (April 1-2, 2014)

4: **Strittmatter N**, Szekeres A, Körtvelyessy Győri I, Balog J, Szaniszló T, Kristof K, Takats Z. *Identification of Microorganisms by Rapid Evaporative Ionisation Mass Spectrometry*. 19th International Mass Spectrometry Conference 2012 (IMSC), Kyoto, Japan. (September 15-21, 2012)

References

- (1) Lechevalier, M. P.; Moss, C. W. Lipids in bacterial taxonomy - a taxonomist's view. *Critical Reviews in Microbiology* **1977**, *5*, 109-210.
- (2) Balog, J.; Sasi-Szabó, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezső, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takáts, Z. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Science Translational Medicine* **2013**, *5*, 194ra193.
- (3) Whitman, W. B.; Coleman, D. C.; Wiebe, W. J. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* **1998**, *95*, 6578-6583.
- (4) Warren, J. R.; Marshall, B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* **1983**, *1*, 1273-1275.
- (5) Cover, T. L.; Blaser, M. J. Helicobacter pylori infection, a paradigm for chronic mucosal inflammation: Pathogenesis and implications for eradication and prevention. *Advances in Internal Medicine* **1996**, *41*, 85-117.
- (6) McNamara, D.; El-Omar, E. Helicobacter pylori infection and the pathogenesis of gastric cancer: A paradigm for host-bacterial interactions. *Digestive and Liver Disease* **2008**, *40*, 504-509.
- (7) Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486*, 207-214.
- (8) Dominguez-Bello, M. G.; Costello, E. K.; Contreras, M.; Magris, M.; Hidalgo, G.; Fierer, N.; Knight, R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences* **2010**, *107*, 11971-11975.
- (9) Gronlund, M. M.; Lehtonen, O. P.; Eerola, E.; Kero, P. Fecal microflora in healthy infants born by different methods of delivery: Permanent changes in intestinal flora after cesarean delivery. *Journal of Pediatric Gastroenterology and Nutrition* **1999**, *28*, 19-25.
- (10) Nicholson, J. K.; Holmes, E.; Kinross, J.; Burcelin, R.; Gibson, G.; Jia, W.; Pettersson, S. Host-gut microbiota metabolic interactions. *Science* **2012**, *336*, 1262-1267.
- (11) Turnbaugh, P. J.; Ley, R. E.; Mahowald, M. A.; Magrini, V.; Mardis, E. R.; Gordon, J. I. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **2006**, *444*, 1027-1131.
- (12) Qin, J.; Li, Y.; Cai, Z.; Li, S.; Zhu, J.; Zhang, F.; Liang, S.; Zhang, W.; Guan, Y.; Shen, D.; Peng, Y.; Zhang, D.; Jie, Z.; Wu, W.; Qin, Y.; Xue, W.; Li, J.; Han, L.; Lu, D.; Wu, P.; Dai, Y.; Sun, X.; Li, Z.; Tang, A.; Zhong, S.; Li, X.; Chen, W.; Xu, R.; Wang, M.; Feng, Q.; Gong, M.; Yu, J.; Zhang, Y.; Zhang, M.; Hansen, T.; Sanchez, G.; Raes, J.; Falony, G.; Okuda, S.; Almeida, M.; LeChatelier, E.; Renault, P.; Pons, N.; Batto, J.-M.; Zhang, Z.; Chen, H.; Yang, R.; Zheng, W.; Li, S.; Yang, H.; Wang, J.; Ehrlich, S. D.; Nielsen, R.; Pedersen, O.; Kristiansen, K.; Wang, J. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **2012**, *490*, 55-60.
- (13) Frank, D. N.; St. Amand, A. L.; Feldman, R. A.; Boedeker, E. C.; Harpaz, N.; Pace, N. R. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* **2007**, *104*, 13780-13785.
- (14) Tjalsma, H.; Boleij, A.; Marchesi, J. R.; Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: Beyond the usual suspects. *Nature Reviews Microbiology* **2012**, *10*, 575-582.
- (15) Havlicek, V.; Lemr, K.; Schug, K. A. Current trends in microbial diagnostics based on mass spectrometry. *Analytical Chemistry* **2013**, *85*, 790-797.

- (16) Yarza, P.; Yilmaz, P.; Pruesse, E.; Glockner, F. O.; Ludwig, W.; Schleifer, K.-H.; Whitman, W. B.; Euzéby, J.; Amann, R.; Rossello-Mora, R. Uniting the classification of cultured and uncultured bacteria and archaea using 16s rRNA gene sequences. *Nature Reviews Microbiology* **2014**, *12*, 635-645.
- (17) Konstantinidis, K. T.; Ramette, A.; Tiedje, J. M. The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **2006**, *361*, 1929-1940.
- (18) Klouche, M.; Schroder, U. Rapid methods for diagnosis of bloodstream infections. *Clinical Chemistry and Laboratory Medicine* **2008**, *46*, 888-908.
- (19) Janda, J. M.; Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology* **2007**, *45*, 2761-2764.
- (20) Woo, P. C. Y.; Lau, S. K. P.; Teng, J. L. L.; Tse, H.; Yuen, K. Y. Then and now: Use of 16s rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection* **2008**, *14*, 908-934.
- (21) Silhavy, T. J.; Kahne, D.; Walker, S. The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology* **2010**, *2*, a000414.
- (22) Neuhaus, F. C.; Baddiley, J. A continuum of anionic charge: Structures and functions of d-alanyl-teichoic acids in gram-positive bacteria. *Microbiology and Molecular Biology Reviews* **2003**, *67*, 686-723.
- (23) Raetz, C. R.; Reynolds, C. M.; Trent, M. S.; Bishop, R. E. Lipid modification systems in gram-negative bacteria. *Annual Review of Biochemistry* **2007**, *76*, 295-329.
- (24) Parsons, J. B.; Rock, C. O. Bacterial lipids: Metabolism and membrane homeostasis. *Progress in Lipid Research* **2013**, *52*, 249-276.
- (25) Zhang, Y.-M.; Rock, C. O. Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology* **2008**, *6*, 222-233.
- (26) Anhalt, J. P.; Fenselau, C. Identification of bacteria using mass-spectrometry. *Analytical Chemistry* **1975**, *47*, 219-225.
- (27) Meuzelaar, H. L. C.; Kistemaker, P. G. Technique for fast and reproducible fingerprinting of bacteria by pyrolysis mass-spectrometry. *Analytical Chemistry* **1973**, *45*, 587-590.
- (28) Schulten, H. R.; Beckey, H. D.; Meuzelaar, H. L. C.; Boerboom, A. J. High-resolution field ionization mass-spectrometry of bacterial pyrolysis products. *Analytical Chemistry* **1973**, *45*, 191-195.
- (29) Meuzelaar, H. L.; Kistemaker, P. G. A technique for fast and reproducible fingerprinting of bacteria by pyrolysis mass spectrometry. *Analytical Chemistry* **1973**, *45*, 587-590.
- (30) Schulten, H. R.; Beckey, H. D.; Meuzelaar, H. L.; Boerboom, A. J. High resolution field ionization mass spectrometry of bacterial pyrolysis products. *Analytical Chemistry* **1973**, *45*, 191-195.
- (31) Moss, C. W. Gas-liquid chromatography as an analytical tool in microbiology. *Journal of Chromatography* **1981**, *203*, 337-347.
- (32) Miller, L. T. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *Journal of Clinical Microbiology* **1982**, *16*, 584-586.
- (33) Basile, F.; Beverly, M. B.; Voorhees, K. J.; Hadfield, T. L. Pathogenic bacteria: Their detection and differentiation by rapid lipid profiling with pyrolysis mass spectrometry. *Trends in Analytical Chemistry* **1998**, *17*, 95-109.
- (34) Hendricker, A. D.; Abbas-Hawks, C.; Basile, F.; Voorhees, K. J.; Hadfield, T. L. Rapid chemotaxonomy of pathogenic bacteria using in situ thermal hydrolysis and methylation as a sample preparation step coupled with a field-portable membrane-inlet quadrupole ion trap mass spectrometer. *International Journal of Mass Spectrometry* **1999**, *190-191*, 331-342.

- (35) Heller, D. N.; Cotter, R. J.; Fenselau, C. Profiling of bacteria by fast-atom-bombardment mass-spectrometry. *Analytical Chemistry* **1987**, *59*, 2806-2809.
- (36) Heller, D. N.; Murphy, C. M.; Cotter, R. J.; Fenselau, C.; Uy, O. M. Constant neutral loss scanning for the characterization of bacterial phospholipids desorbed by fast atom bombardment. *Analytical Chemistry* **1988**, *60*, 2787-2791.
- (37) Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984**, *88*, 4451-4459.
- (38) Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry* **1985**, *57*, 2935-2939.
- (39) Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988**, *60*, 2299-2301.
- (40) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.
- (41) Knochenmuss, R.; Dubois, F.; Dale, M. J.; Zenobi, R. The matrix suppression effect and ionization mechanisms in matrix-assisted laser desorption/ionization. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 871-877.
- (42) Karas, M.; Glückmann, M.; Schäfer, J. Ionization in matrix-assisted laser desorption/ionization: Singly charged molecular ions are the lucky survivors. *Journal of Mass Spectrometry* **2000**, *35*, 1-12.
- (43) Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry* **2000**, *35*, 804-817.
- (44) Smith, P. B. W.; Snyder, A. P.; Harden, C. S. Characterization of bacterial phospholipids by electrospray ionization tandem mass spectrometry. *Analytical Chemistry* **1995**, *67*, 1824-1830.
- (45) Claydon, M. A.; Davey, S. N.; Edwards-Jones, V.; Gordon, D. B. The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotechnology* **1996**, *14*, 1584-1586.
- (46) Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O., Jr. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 1227-1232.
- (47) Krishnamurthy, T.; Ross, P. L. Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 1992-1996.
- (48) Wunschel, D. S.; Fox, K. F.; Fox, A.; Bruce, J. E.; Muddiman, D. C.; Smith, R. D. Analysis of double-stranded polymerase chain reaction products from the bacillus cereus group by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 29-35.
- (49) Goodacre, R.; Heald, J. K.; Kell, D. B. Characterisation of intact microorganisms using electrospray ionisation mass spectrometry. *Fems Microbiology Letters* **1999**, *176*, 17-24.
- (50) Xiang, F.; Anderson, G. A.; Veenstra, T. D.; Lipton, M. S.; Smith, R. D. Characterization of microorganisms and biomarker development from global esi-ms/ms analyses of cell lysates. *Analytical Chemistry* **2000**, *72*, 2475-2481.
- (51) Fenselau, C.; Demirev, P. A. Characterization of intact microorganisms by maldi mass spectrometry. *Mass Spectrometry Reviews* **2001**, *20*, 157-171.
- (52) Ryzhov, V.; Fenselau, C. Characterization of the protein subset desorbed by maldi from whole bacterial cells. *Analytical Chemistry* **2001**, *73*, 746-750.

- (53) Hathout, Y.; Demirev, P. A.; Ho, Y. P.; Bundy, J. L.; Ryzhov, V.; Sapp, L.; Stutler, J.; Jackman, J.; Fenselau, C. Identification of bacillus spores by matrix-assisted laser desorption ionization-mass spectrometry. *Applied and Environmental Microbiology* **1999**, *65*, 4313-4319.
- (54) Hathout, Y.; Setlow, B.; Cabrera-Martinez, R. M.; Fenselau, C.; Setlow, P. Small, acid-soluble proteins as biomarkers in mass spectrometry analysis of bacillus spores. *Applied and Environmental Microbiology* **2003**, *69*, 1100-1107.
- (55) Jarman, K. H.; Cebula, S. T.; Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Kingsley, M. T.; Wahl, K. L. An algorithm for automated bacterial identification using matrix-assisted laser desorption/ionization mass spectrometry. *Analytical Chemistry* **2000**, *72*, 1217-1223.
- (56) Valentine, N.; Wunschel, S.; Wunschel, D.; Petersen, C.; Wahl, K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Applied and Environmental Microbiology* **2005**, *71*, 58-64.
- (57) Krásný, L.; Hynek, R.; Hochel, I. Identification of bacteria using mass spectrometry techniques. *International Journal of Mass Spectrometry* **2013**, *353*, 67-79.
- (58) Bizzini, A.; Durussel, C.; Bille, J.; Greub, G.; Prod'homme, G. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *Journal of Clinical Microbiology* **2010**, *48*, 1549-1554.
- (59) Cherkaoui, A.; Hibbs, J.; Emonet, S.; Tangomo, M.; Girard, M.; Francois, P.; Schrenzel, J. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *Journal of Clinical Microbiology* **2010**, *48*, 1169-1175.
- (60) van Veen, S. Q.; Claas, E. C. J.; Kuijper, E. J. High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *Journal of Clinical Microbiology* **2010**, *48*, 900-907.
- (61) Tan, K. E.; Ellis, B. C.; Lee, R.; Stamper, P. D.; Zhang, S. X.; Carroll, K. C. Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: A bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. *Journal of Clinical Microbiology* **2012**, *50*, 3301-3308.
- (62) Seng, P.; Drancourt, M.; Gouriet, F.; La Scola, B.; Fournier, P.-E.; Rolain, J. M.; Raoult, D. Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Infectious Diseases* **2009**, *49*, 543-551.
- (63) Cassagne, C.; Cella, A. L.; Suchon, P.; Normand, A. C.; Ranque, S.; Piarroux, R. Evaluation of four pretreatment procedures for maldi-tof ms yeast identification in the routine clinical laboratory. *Medical Mycology* **2013**, *51*, 371-377.
- (64) Lista, F.; Reubsæet, F.; De Santis, R.; Parchen, R.; de Jong, A.; Kieboom, J.; van der Laaken, A.; Voskamp-Visser, I.; Fillo, S.; Jansen, H.-J.; Van der Plas, J.; Paauw, A. Reliable identification at the species level of brucella isolates with maldi-tof-ms. *BMC Microbiology* **2011**, *11*, 267.
- (65) Mather, C. A.; Rivera, S. F.; Butler-Wu, S. M. Comparison of the bruker biotyper and vitek ms matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *Journal of Clinical Microbiology* **2014**, *52*, 130-138.
- (66) Schulthess, B.; Bloemberg, G. V.; Zbinden, R.; Böttger, E. C.; Hombach, M. Evaluation of the bruker maldi biotyper for identification of gram-positive rods: Development of a diagnostic algorithm for the clinical laboratory. *Journal of Clinical Microbiology* **2014**, *52*, 1089-1097.

- (67) Calderaro, A.; Arcangeletti, M.-C.; Rodighiero, I.; Buttrini, M.; Gorrini, C.; Motta, F.; Germini, D.; Medici, M.-C.; Chezzi, C.; De Conto, F. Matrix-assisted laser desorption/ionization time-of-flight (maldi-tof) mass spectrometry applied to virus identification. *Scientific Reports* **2014**, *4*.
- (68) Carbonnelle, E.; Mesquita, C.; Bille, E.; Day, N.; Dauphin, B.; Beretti, J. L.; Ferroni, A.; Gutmann, L.; Nassif, X. Maldi-tof mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clinical Biochemistry* **2011**, *44*, 104-109.
- (69) Ferreira, L.; Sánchez-Juanes, F.; González-Ávila, M.; Cembrero-Fuciños, D.; Herrero-Hernández, A.; González-Buitrago, J. M.; Muñoz-Bellido, J. L. Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* **2010**, *48*, 2110-2115.
- (70) Lagacé-Wiens, P. In *Sepsis*, Mancini, N., Ed.; Springer New York, 2015, pp 47-55.
- (71) Segawa, S.; Sawai, S.; Murata, S.; Nishimura, M.; Beppu, M.; Sogawa, K.; Watanabe, M.; Satoh, M.; Matsutani, T.; Kobayashi, M.; Iwadate, Y.; Kuwabara, S.; Saeki, N.; Nomura, F. Direct application of maldi-tof mass spectrometry to cerebrospinal fluid for rapid pathogen identification in a patient with bacterial meningitis. *Clinica Chimica Acta* **2014**, *435*, 59-61.
- (72) Kumar, A.; Roberts, D.; Wood, K. E.; Light, B.; Parrillo, J. E.; Sharma, S.; Suppes, R.; Feinstein, D.; Zanotti, S.; Taiberg, L.; Gurka, D.; Kumar, A.; Cheang, M. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine* **2006**, *34*, 1589-1596.
- (73) Ferroni, A.; Suarez, S.; Beretti, J.-L.; Dauphin, B.; Bille, E.; Meyer, J.; Bougnoux, M.-E.; Alanio, A.; Berche, P.; Nassif, X. Real-time identification of bacteria and candida species in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* **2010**, *48*, 1542-1548.
- (74) Demirev, P. A.; Fenselau, C. Mass spectrometry in biodefense. *Journal of Mass Spectrometry* **2008**, *43*, 1441-1457.
- (75) Demirev, P. A.; Fenselau, C. Mass spectrometry for rapid characterization of microorganisms. *Annual Reviews of Analytical Chemistry* **2008**, *1*, 71-93.
- (76) Wynne, C.; Fenselau, C.; Demirev, P. A.; Edwards, N. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. *Analytical Chemistry* **2009**, *81*, 9633-9642.
- (77) Demirev, P. A.; Feldman, A. B.; Kowalski, P.; Lin, J. S. Top-down proteomics for rapid identification of intact microorganisms. *Analytical Chemistry* **2005**, *77*, 7455-7461.
- (78) Schmidt, F.; Fiege, T.; Hustoft, H. K.; Kneist, S.; Thiede, B. Shotgun mass mapping of lactobacillus species and subspecies from caries related isolates by MALDI-MS. *Proteomics* **2009**, *9*, 1994-2003.
- (79) Ecker, D. J.; Sampath, R.; Massire, C.; Blyn, L. B.; Hall, T. A.; Eshoo, M. W.; Hofstadler, S. A. Ibis T5000: A universal biosensor approach for microbiology. *Nature Reviews Microbiology* **2008**, *6*, 553-558.
- (80) Hannis, J. C.; Manalili, S. M.; Hall, T. A.; Ranken, R.; White, N.; Sampath, R.; Blyn, L. B.; Ecker, D. J.; Mandrell, R. E.; Fagerquist, C. K.; Bates, A. H.; Miller, W. G.; Hofstadler, S. A. High-resolution genotyping of campylobacter species by use of PCR and high-throughput mass spectrometry. *Journal of Clinical Microbiology* **2008**, *46*, 1220-1225.
- (81) Emonet, S.; Shah, H. N.; Cherkaoui, A.; Schrenzel, J. Application and use of various mass spectrometry methods in clinical microbiology. *Clinical Microbiology and Infection* **2010**, *16*, 1604-1613.
- (82) Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **2004**, *306*, 471-473.

- (83) Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. Ambient mass spectrometry. *Science* **2006**, *311*, 1566-1570.
- (84) Cody, R. B.; Laramée, J. A.; Durst, H. D. Versatile new ion source for the analysis of materials in open air under ambient conditions. *Analytical Chemistry* **2005**, *77*, 2297-2302.
- (85) Harper, J. D.; Charipar, N. A.; Mulligan, C. C.; Zhang, X.; Cooks, R. G.; Ouyang, Z. Low-temperature plasma probe for ambient desorption ionization. *Analytical Chemistry* **2008**, *80*, 9097-9104.
- (86) Haddad, R.; Sparrapan, R.; Eberlin, M. N. Desorption sonic spray ionization for (high) voltage-free ambient mass spectrometry. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 2901-2905.
- (87) Nemes, P.; Vertes, A. Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry. *Analytical Chemistry* **2007**, *79*, 8098-8106.
- (88) Weston, D. J. Ambient ionization mass spectrometry: Current understanding of mechanistic theory; analytical performance and application areas. *The Analyst* **2010**, *135*, 661-668.
- (89) Chingin, K.; Liang, J.; Chen, H. Direct analysis of in vitro grown microorganisms and mammalian cells by ambient mass spectrometry. *RSC Advances* **2014**, *4*, 5768-5781.
- (90) Song, Y.; Talaty, N.; Datsenko, K.; Wanner, B. L.; Cooks, R. G. In vivo recognition of *Bacillus subtilis* by desorption electrospray ionization mass spectrometry (DESI-MS). *The Analyst* **2009**, *134*, 838-841.
- (91) Hsu, C.-C.; ElNaggar, M. S.; Peng, Y.; Fang, J.; Sanchez, L. M.; Mascuch, S. J.; Møller, K. A.; Alazeh, E. K.; Pikula, J.; Quinn, R. A.; Zeng, Y.; Wolfe, B. E.; Dutton, R. J.; Gerwick, L.; Zhang, L.; Liu, X.; Månsson, M.; Dorrestein, P. C. Real-time metabolomics on living microorganisms using ambient electrospray ionization flow-probe. *Analytical Chemistry* **2013**, *85*, 7014-7018.
- (92) Traxler, M. F.; Kolter, R. A massively spectacular view of the chemical lives of microbes. *Proceedings of the National Academy of Sciences* **2012**, *109*, 10128-10129.
- (93) Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein, P. C.; Kolter, R. Interspecies interactions stimulate diversification of the streptomyces coelicolor secreted metabolome. *MBio* **2013**, *4*.
- (94) Watrous, J.; Roach, P.; Heath, B.; Alexandrov, T.; Laskin, J.; Dorrestein, P. C. Metabolic profiling directly from the petri dish using nanospray desorption electrospray ionization imaging mass spectrometry. *Analytical Chemistry* **2013**, *85*, 10385-10391.
- (95) Randall, E. C.; Bunch, J.; Cooper, H. J. Direct analysis of intact proteins from *Escherichia coli* colonies by liquid extraction surface analysis mass spectrometry. *Analytical Chemistry* **2014**, *86*, 10504-10510.
- (96) Parsieglá, G.; Shrestha, B.; Carriere, F.; Vertes, A. Direct analysis of phycobilisomal antenna proteins and metabolites in small cyanobacterial populations by laser ablation electrospray ionization mass spectrometry. *Analytical Chemistry* **2012**, *84*, 34-38.
- (97) Pierce, C. Y.; Barr, J. R.; Cody, R. B.; Massung, R. F.; Woolfitt, A. R.; Moura, H.; Thompson, H. A.; Fernandez, F. M. Ambient generation of fatty acid methyl ester ions from bacterial whole cells by direct analysis in real time (dart) mass spectrometry. *Chemical Communications* **2007**, 807-809.
- (98) Watts, K. R.; Loveridge, S. T.; Tenney, K.; Media, J.; Valeriote, F. A.; Crews, P. Utilizing dart mass spectrometry to pinpoint halogenated metabolites from a marine invertebrate-derived fungus. *Journal of Organic Chemistry* **2011**, *76*, 6201-6208.
- (99) Vaidyanathan, S.; Kell, D. B.; Goodacre, R. Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 118-128.
- (100) Ishida, Y.; Madonna, A. J.; Rees, J. C.; Meetani, M. A.; Voorhees, K. J. Rapid analysis of intact phospholipids from whole bacterial cells by matrix-assisted laser desorption/ionization mass

spectrometry combined with on-probe sample pretreatment. *Rapid Communications in Mass Spectrometry* **2002**, *16*, 1877-1882.

(101) Shu, X.; Li, Y.; Liang, M.; Yang, B.; Liu, C.; Wang, Y.; Shu, J. Rapid lipid profiling of bacteria by online maldi-tof mass spectrometry. *International Journal of Mass Spectrometry* **2012**, *321-322*, 71-76.

(102) Song, Y.; Talaty, N.; Tao, W. A.; Pan, Z.; Cooks, R. G. Rapid ambient mass spectrometric profiling of intact, untreated bacteria using desorption electrospray ionization. *Chemical Communications* **2007**, 61-63.

(103) Zhang, J. I.; Talaty, N.; Costa, A. B.; Xia, Y.; Tao, W. A.; Bell, R.; Callahan, J. H.; Cooks, R. G. Rapid direct lipid profiling of bacteria using desorption electrospray ionization mass spectrometry. *International Journal of Mass Spectrometry* **2011**, *301*, 37-44.

(104) Hamid, A. M.; Jarmusch, A. K.; Pirro, V.; Pincus, D. H.; Clay, B. G.; Gervasi, G.; Cooks, R. G. Rapid discrimination of bacteria by paper spray mass spectrometry. *Analytical Chemistry* **2014**, *86*, 7500-7507.

(105) Cotte-Rodriguez, I.; Hernandez-Soto, H.; Chen, H.; Cooks, R. G. In situ trace detection of peroxide explosives by desorption electrospray ionization and desorption atmospheric pressure chemical ionization. *Analytical Chemistry* **2008**, *80*, 1512-1519.

(106) Takats, Z.; Cotte-Rodriguez, I.; Talaty, N.; Chen, H.; Cooks, R. G. Direct, trace level detection of explosives on ambient surfaces by desorption electrospray ionization mass spectrometry. *Chemical Communications* **2005**, 1950-1952.

(107) Kauppila, T. J.; Wiseman, J. M.; Ketola, R. A.; Kotiaho, T.; Cooks, R. G.; Kostianen, R. Desorption electrospray ionization mass spectrometry for the analysis of pharmaceuticals and metabolites. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 387-392.

(108) Kauppila, T. J.; Talaty, N.; Kuuranne, T.; Kotiaho, T.; Kostianen, R.; Cooks, R. G. Rapid analysis of metabolites and drugs of abuse from urine samples by desorption electrospray ionization-mass spectrometry. *The Analyst* **2007**, *132*, 868-875.

(109) Kennedy, J. H.; Aurand, C.; Shirey, R.; Laughlin, B. C.; Wiseman, J. M. Coupling desorption electrospray ionization with solid-phase microextraction for screening and quantitative analysis of drugs in urine. *Analytical Chemistry* **2010**, *82*, 7502-7508.

(110) Meetani, M. A.; Shin, Y.-S.; Zhang, S.; Mayer, R.; Basile, F. Desorption electrospray ionization mass spectrometry of intact bacteria. *Journal of Mass Spectrometry* **2007**, *42*, 1186-1193.

(111) Bianchi, F.; Gregori, A.; Braun, G.; Crescenzi, C.; Careri, M. Micro-solid-phase extraction coupled to desorption electrospray ionization-high-resolution mass spectrometry for the analysis of explosives in soil. *Analytical and Bioanalytical Chemistry* **2015**, *407*, 931-938.

(112) Denes, J.; Katona, M.; Hosszu, A.; Czuczy, N.; Takats, Z. Analysis of biological fluids by direct combination of solid phase extraction and desorption electrospray ionization mass spectrometry. *Analytical Chemistry* **2009**, *81*, 1669-1675.

(113) Strittmatter, N.; During, R. A.; Takats, Z. Analysis of wastewater samples by direct combination of thin-film microextraction and desorption electrospray ionization mass spectrometry. *The Analyst* **2012**, *137*, 4037-4044.

(114) Bennet, R. V.; Gamage, C. M.; Fernandez, F. M. Imaging of biological tissues by desorption electrospray ionization mass spectrometry. *Journal of Visualized Experimints* **2013**, e50575.

(115) Bennett, R. V.; Cleaves, H. J., 2nd; Davis, J. M.; Sokolov, D. A.; Orlando, T. M.; Bada, J. L.; Fernandez, F. M. Desorption electrospray ionization imaging mass spectrometry as a tool for investigating model prebiotic reactions on mineral surfaces. *Analytical Chemistry* **2013**, *85*, 1276-1279.

- (116) Janfelt, C. Imaging of plant materials using indirect desorption electrospray ionization mass spectrometry. *Methods in Molecular Biology* **2015**, *1203*, 91-97.
- (117) Vismeh, R.; Waldon, D. J.; Teffera, Y.; Zhao, Z. Localization and quantification of drugs in animal tissues by use of desorption electrospray ionization mass spectrometry imaging. *Analytical Chemistry* **2012**, *84*, 5439-5445.
- (118) Wiseman, J. M.; Ifa, D. R.; Venter, A.; Cooks, R. G. Ambient molecular imaging by desorption electrospray ionization mass spectrometry. *Nature Protocols* **2008**, *3*, 517-524.
- (119) Gerbig, S.; Golf, O.; Balog, J.; Denes, J.; Baranyai, Z.; Zarand, A.; Raso, E.; Timar, J.; Takats, Z. Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization mass spectrometric imaging. *Analytical and Bioanalytical Chemistry* **2012**, *403*, 2315-2325.
- (120) Guenther, S.; Muirhead, L. J.; Speller, A. V.; Golf, O.; Strittmatter, N.; Ramakrishnan, R.; Goldin, R. D.; Jones, E.; Veselkov, K.; Nicholson, J.; Darzi, A.; Takats, Z. Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Research* **2015**.
- (121) Mirnezami, R.; Spagou, K.; Vorkas, P. A.; Lewis, M. R.; Kinross, J.; Want, E.; Shion, H.; Goldin, R. D.; Darzi, A.; Takats, Z.; Holmes, E.; Cloarec, O.; Nicholson, J. K. Chemical mapping of the colorectal cancer microenvironment via maldi imaging mass spectrometry (maldi-msi) reveals novel cancer-associated field effects. *Molecular Oncology* **2014**, *8*, 39-49.
- (122) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V. M.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proceedings of the National Academy of Sciences* **2014**.
- (123) Calligaris, D.; Caragacianu, D.; Liu, X.; Norton, I.; Thompson, C. J.; Richardson, A. L.; Golshan, M.; Easterling, M. L.; Santagata, S.; Dillon, D. A.; Jolesz, F. A.; Agar, N. Y. Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *Proceedings of the National Academy of Sciences* **2014**, *111*, 15184-15189.
- (124) Abbassi-Ghadi, N.; Veselkov, K.; Kumar, S.; Huang, J.; Jones, E.; Strittmatter, N.; Kudo, H.; Goldin, R.; Takats, Z.; Hanna, G. B. Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. *Chemical Communications* **2014**, *50*, 3661-3664.
- (125) Kertesz, V.; Van Berkel, G. J.; Vavrek, M.; Koeplinger, K. A.; Schneider, B. B.; Covey, T. R. Comparison of drug distribution images from whole-body thin tissue sections obtained using desorption electrospray ionization tandem mass spectrometry and autoradiography. *Analytical Chemistry* **2008**, *80*, 5168-5177.
- (126) Swales, J. G.; Tucker, J. W.; Strittmatter, N.; Nilsson, A.; Cobice, D.; Clench, M. R.; Mackay, C. L.; Andren, P. E.; Takats, Z.; Webborn, P. J.; Goodwin, R. J. Mass spectrometry imaging of cassette-dosed drugs for higher throughput pharmacokinetic and biodistribution analysis. *Analytical Chemistry* **2014**, *86*, 8473-8480.
- (127) Campbell, D. I.; Ferreira, C. R.; Eberlin, L. S.; Cooks, R. G. Improved spatial resolution in the imaging of biological tissue using desorption electrospray ionization. *Analytical and Bioanalytical Chemistry* **2012**, *404*, 389-398.
- (128) Schaefer, K.-C.; Denes, J.; Albrecht, K.; Szaniszló, T.; Balog, J.; Skoumal, R.; Katona, M.; Toth, M.; Balogh, L.; Takats, Z. In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. *Angewandte Chemie-International Edition* **2009**, *48*, 8240-8242.
- (129) Takats, Z.; Denes, J.; Kinross, J. Identifying the margin: A new method to distinguish between cancerous and noncancerous tissue during surgery. *Future Oncology* **2012**, *8*, 113-116.
- (130) Balog, J.; Szaniszló, T.; Schaefer, K. C.; Denes, J.; Lopata, A.; Godorhazy, L.; Szalay, D.; Balogh, L.; Sasi-Szabo, L.; Toth, M.; Takats, Z. Identification of biological tissues by rapid evaporative ionization mass spectrometry. *Analytical Chemistry* **2010**, *82*, 7343-7350.

- (131) Jones, E. A.; Deininger, S.-O.; Hogendoorn, P. C. W.; Deelder, A. M.; McDonnell, L. A. Imaging mass spectrometry statistical analysis. *Journal of Proteomics* **2012**, *75*, 4962-4989.
- (132) Lindon, J. C.; Holmes, E.; Nicholson, J. K. Metabonomics techniques and applications to pharmaceutical research & development. *Pharmaceutical Research* **2006**, *23*, 1075-1088.
- (133) Strittmatter, N.; Jones, E. A.; Veselkov, K. A.; Rebec, M.; Bundy, J. G.; Takats, Z. Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chemical Communications* **2013**, *49*, 6188-6190.
- (134) Strittmatter, N.; Lovrics, A.; Sessler, J.; McKenzie, J. S.; Bodai, Z.; Doria, M. L.; Kucsma, N.; Szakacs, G.; Takats, Z. Shotgun lipidomic profiling of the nci60 cell line panel using rapid evaporative ionization mass spectrometry. *Analytical Chemistry* **2016**, *88*, 7507-7514.
- (135) Butler, G.; Rasmussen, M. D.; Lin, M. F.; Santos, M. A. S.; Sakthikumar, S.; Munro, C. A.; Rheinbay, E.; Grabherr, M.; Forche, A.; Reedy, J. L.; Agrafioti, I.; Arnaud, M. B.; Bates, S.; Brown, A. J. P.; Brunke, S.; Costanzo, M. C.; Fitzpatrick, D. A.; de Groot, P. W. J.; Harris, D.; Hoyer, L. L.; Hube, B.; Klis, F. M.; Kodira, C.; Lennard, N.; Logue, M. E.; Martin, R.; Neiman, A. M.; Nikolaou, E.; Quail, M. A.; Quinn, J.; Santos, M. C.; Schmitzberger, F. F.; Sherlock, G.; Shah, P.; Silverstein, K. A. T.; Skrzypek, M. S.; Soll, D.; Staggs, R.; Stansfield, I.; Stumpf, M. P. H.; Sudbery, P. E.; Srikantha, T.; Zeng, Q.; Berman, J.; Berriman, M.; Heitman, J.; Gow, N. A. R.; Lorenz, M. C.; Birren, B. W.; Kellis, M.; Cuomo, C. A. Evolution of pathogenicity and sexual reproduction in eight candida genomes. *Nature* **2009**, *459*, 657-662.
- (136) Özen, A. I.; Ussery, D. W. Defining the pseudomonas genus: Where do we draw the line with azotobacter? *Microbial Ecology* **2012**, *63*, 239-248.
- (137) Gholami, Amin M.; Hahne, H.; Wu, Z.; Auer, Florian J.; Meng, C.; Wilhelm, M.; Kuster, B. Global proteome analysis of the nci-60 cell line panel. *Cell Reports* **2013**, *4*, 609-620.
- (138) Scherf, U.; Ross, D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O.; Weinstein, J. N. A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* **2000**, *24*, 236-244.
- (139) Crocker, F. H.; Fredrickson, J. K.; White, D. C.; Ringelberg, D. B.; Balkwill, D. L. Phylogenetic and physiological diversity of arthrobacter strains isolated from unconsolidated subsurface sediments. *Microbiology* **2000**, *146*, 1295-1310.
- (140) Poston, W. L.; Marchette, D. J. Recursive dimensionality reduction using fisher's linear discriminant. *Pattern Recognition* **1998**, *31*, 881-888.
- (141) Li, H.; Jiang, T.; Zhang, K. MMC efficient and robust feature extraction by maximum margin criterion. *Proceedings of Advances in Neural Information Processing Systems* **2003**, 97-104.
- (142) Strittmatter, N.; Rebec, M.; Jones, E. A.; Golf, O.; Abdolrasouli, A.; Balog, J.; Behrends, V.; Veselkov, K. A.; Takats, Z. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. *Analytical Chemistry* **2014**, *86*, 6555-6562.
- (143) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012**, *30*, 918-920.
- (144) Race, A. M.; Styles, I. B.; Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *Journal of Proteomics* **2012**, *75*, 5111-5112.

- (145) Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. Recursive segment-wise peak alignment of biological 1h nmr spectra for improved metabolic biomarker recovery. *Analytical Chemistry* **2008**, *81*, 56-66.
- (146) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Analytical Chemistry* **2011**, *83*, 5864-5872.
- (147) Donoho, D. L.; Johnstone, I. M. Ideal denoising in an orthonormal basis chosen from a library of bases. *Comptes rendus de l'Académie des sciences. Série I, Mathématique* **1994**, *319*, 1317-1322.
- (148) Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. Recursive segment-wise peak alignment of biological 1h nmr spectra for improved metabolic biomarker recovery. *Analytical Chemistry* **2009**, *81*, 56-66.
- (149) Ihaka, R.; Gentleman, R. R. A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **1996**, *5*, 299-314.
- (150) Reinhold, W. C.; Sunshine, M.; Liu, H.; Varma, S.; Kohn, K. W.; Morris, J.; Doroshow, J.; Pommier, Y. Cellminer: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the nci-60 cell line set. *Cancer Research* **2012**, *72*, 3499-3511.
- (151) Shankavaram, U.; Varma, S.; Kane, D.; Sunshine, M.; Chary, K.; Reinhold, W.; Pommier, Y.; Weinstein, J. Cellminer: A relational database and query tool for the nci-60 cancer cell lines. *BMC Genomics* **2009**, *10*, 277.
- (152) Abbassi-Ghadi, N.; Jones, E. A.; Veselkov, K. A.; Huang, J.; Kumar, S.; Strittmatter, N.; Golf, O.; Kudo, H.; Goldin, R. D.; Hanna, G. B.; Takats, Z. Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (desi-ms) for the imaging analysis of human cancer tissue: A gateway for clinical applications. *Analytical Methods* **2015**, *7*, 71-80.
- (153) Jeffries, N. Algorithms for alignment of mass spectrometry proteomic data. *Bioinformatics* **2005**, *21*, 3066-3073.
- (154) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proceedings of the National Academy of Sciences* **2014**, *111*, 1216-1221.
- (155) Robichaud, G.; Garrard, K.; Barry, J.; Muddiman, D. Msireader: An open-source interface to view and analyze high resolving power ms imaging files on matlab platform. *Journal of The American Society for Mass Spectrometry* **2013**, *24*, 718-721.
- (156) Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9*, 62-66.
- (157) Song, S. H.; Park, K. U.; Lee, J. H.; Kim, E. C.; Kim, J. Q.; Song, J. Electrospray ionization-tandem mass spectrometry analysis of the mycolic acid profiles for the identification of common clinical isolates of mycobacterial species. *Journal of Microbiological Methods* **2009**, *77*, 165-177.
- (158) Costa, A. B.; Graham Cooks, R. Simulated splashes: Elucidating the mechanism of desorption electrospray ionization mass spectrometry. *Chemical Physics Letters* **2008**, *464*, 1-8.
- (159) Guenther, S.; Schäfer, K.-C.; Balog, J.; Dénes, J.; Majoros, T.; Albrecht, K.; Tóth, M.; Spengler, B.; Takáts, Z. Electrospray post-ionization mass spectrometry of electrosurgical aerosols. *Journal of the American Society for Mass Spectrometry* **2011**, *22*, 2082-2089.
- (160) Reis, R. S.; Pereira, A. G.; Neves, B. C.; Freire, D. M. G. Gene regulation of rhamnolipid production in pseudomonas aeruginosa – a review. *Bioresource Technology* **2011**, *102*, 6377-6384.

- (161) White, D. C.; Ferman, F. E. Fatty acid composition of the complex lipids of staphylococcus aureus during the formation of the membrane-bound electron transport system. *Journal of Bacteriology* **1968**, *95*, 2198-2209.
- (162) Oursel, D.; Loutelier-Bourhis, C.; Orange, N.; Chevalier, S.; Norris, V.; Lange, C. M. Lipid composition of membranes of escherichia coli by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1721-1728.
- (163) Goldfine, H. The appearance, disappearance and reappearance of plasmalogens in evolution. *Progress in Lipid Research* **2010**, *49*, 493-498.
- (164) Tyurina, Y. Y.; Domingues, R. M.; Tyurin, V. A.; Maciel, E.; Domingues, P.; Amoscato, A. A.; Bayir, H.; Kagan, V. E. Characterization of cardiolipins and their oxidation products by lc-ms analysis. *Chemistry and Physics of Lipids* **2014**, *179*, 3-10.
- (165) Maciel, E.; Domingues, P.; Domingues, M. R. M. Liquid chromatography/tandem mass spectrometry analysis of long-chain oxidation products of cardiolipin induced by the hydroxyl radical. *Rapid Communications in Mass Spectrometry* **2011**, *25*, 316-326.
- (166) Barry Iii, C. E.; Lee, R. E.; Mdluli, K.; Sampson, A. E.; Schroeder, B. G.; Slayden, R. A.; Yuan, Y. Mycolic acids: Structure, biosynthesis and physiological functions. *Progress in Lipid Research* **1998**, *37*, 143-179.
- (167) Alatoon, A. A.; Cazanave, C. J.; Cunningham, S. A.; Ihde, S. M.; Patel, R. Identification of non-diphtheriae corynebacterium by use of matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of Clinical Microbiology* **2012**, *50*, 160-163.
- (168) Konrad, R.; Berger, A.; Huber, I.; Boschert, V.; Hormansdorfer, S.; Busch, U.; Hogardt, M.; Schubert, S.; Sing, A. Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry as a tool for rapid diagnosis of potentially toxigenic corynebacterium species in the laboratory management of diphtheria-associated bacteria. *Euro Surveillance* **2010**, *15*.
- (169) Lanéelle, M.-A.; Launay, A.; Spina, L.; Marrakchi, H.; Laval, F.; Eynard, N.; Lemassu, A.; Tropis, M.; Daffé, M.; Etienne, G. A novel mycolic acid species defines two novel genera of the actinobacteria, hoyosella and amycolicococcus. *Microbiology* **2012**, *158*, 843-855.
- (170) De Briel, D.; Couderc, F.; Riegel, P.; Jehl, F.; Minck, R. High-performance liquid chromatography of corynomycolic acids as a tool in identification of corynebacterium species and related organisms. *Journal of Clinical Microbiology* **1992**, *30*, 1407-1417.
- (171) Yang, Y.; Shi, F.; Tao, G.; Wang, X. Purification and structure analysis of mycolic acids in corynebacterium glutamicum. *The Journal of Microbiology* **2012**, *50*, 235-240.
- (172) Nishiuchi, Y.; Baba, T.; Yano, I. Mycolic acids from rhodococcus, gordonia, and dietzia. *Journal of Microbiological Methods* **2000**, *40*, 1-9.
- (173) Nishiuchi, Y.; Baba, T.; Hotta, H. H.; Yano, I. Mycolic acid analysis in nocardia species: The mycolic acid compositions of Nocardia asteroides, N. farcinica, and N. nova. *Journal of Microbiological Methods* **1999**, *37*, 111-122.
- (174) Giussani, P.; Tringali, C.; Riboni, L.; Viani, P.; Venerando, B. Sphingolipids: Key regulators of apoptosis and pivotal players in cancer drug resistance. *International Journal of Molecular Sciences* **2014**, *15*, 4356-4392.
- (175) Montefusco, D. J.; Matmati, N.; Hannun, Y. A. The yeast sphingolipid signaling landscape. *Chemistry and Physics of Lipids* **2014**, *177*, 26-40.
- (176) Olsen, I.; Jantzen, E. Sphingolipids in bacteria and fungi. *Anaerobe* **2001**, *7*, 103-112.
- (177) Kunsman, J. E.; Caldwell, D. R. Comparison of the sphingolipid content of rumen bacteroides species. *Applied Microbiology* **1974**, *28*, 1088-1089.

- (178) An, D.; Na, C.; Bielawski, J.; Hannun, Y. A.; Kasper, D. L. Membrane sphingolipids as essential molecular signals for bacteroides survival in the intestine. *Proceedings of the National Academy of Sciences* **2011**, *108*, 4666-4671.
- (179) Miyagawa, E.; Azuma, R.; Suto, T.; Yano, I. Occurrence of free ceramides in bacteroides fragilis nctc 9343. *Journal of Biochemistry* **1979**, *86*, 311-320.
- (180) Nichols, F. C.; Yao, X.; Bajrami, B.; Downes, J.; Finegold, S. M.; Knee, E.; Gallagher, J. J.; Housley, W. J.; Clark, R. B. Phosphorylated dihydroceramides from common human bacteria are recovered in human tissues. *PLoS ONE* **2011**, *6*, e16771.
- (181) Wieland Brown, L. C.; Penaranda, C.; Kashyap, P. C.; Williams, B. B.; Clardy, J.; Kronenberg, M.; Sonnenburg, J. L.; Comstock, L. E.; Bluestone, J. A.; Fischbach, M. A. Production of α -galactosylceramide by a prominent member of the human gut microbiota. *PLoS Biology* **2013**, *11*, e1001610.
- (182) Nichols, F. C.; Riep, B.; Mun, J.; Morton, M. D.; Bojarski, M. T.; Dewhirst, F. E.; Smith, M. B. Structures and biological activity of phosphorylated dihydroceramides of porphyromonas gingivalis. *Journal of Lipid Research* **2004**, *45*, 2317-2330.
- (183) Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O. The structural basis of lipopolysaccharide recognition by the tlr4-md-2 complex. *Nature* **2009**, *458*, 1191-1195.
- (184) Zhou, P.; Hu, R.; Chandan, V.; Kuolee, R.; Liu, X.; Chen, W.; Liu, B.; Altman, E.; Li, J. Simultaneous analysis of cardiolipin and lipid a from helicobacter pylori by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Molecular Biosystems* **2012**, *8*, 720-725.
- (185) Ernst, R. K.; Hajjar, A. M.; Tsai, J. H.; Moskowitz, S. M.; Wilson, C. B.; Miller, S. I. Pseudomonas aeruginosa lipid a diversity and its recognition by toll-like receptor 4. *Journal of Endotoxin Research* **2003**, *9*, 395-400.
- (186) Ernst, R. K.; Yi, E. C.; Guo, L.; Lim, K. B.; Burns, J. L.; Hackett, M.; Miller, S. I. Specific lipopolysaccharide found in cystic fibrosis airway pseudomonas aeruginosa. *Science* **1999**, *286*, 1561-1565.
- (187) Karunaratne, D. N.; Richards, J. C.; Hancock, R. E. Characterization of lipid a from pseudomonas aeruginosa o-antigenic b band lipopolysaccharide by 1d and 2d nmr and mass spectral analysis. *Archives of Biochemistry and Biophysics* **1992**, *299*, 368-376.
- (188) Williams, P.; Winzer, K.; Chan, W. C.; Cámara, M. Look who's talking: Communication and quorum sensing in the bacterial world. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **2007**, *362*, 1119-1134.
- (189) Williams, P. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **2007**, *153*, 3923-3938.
- (190) Diggle, S. P.; Matthijs, S.; Wright, V. J.; Fletcher, M. P.; Chhabra, S. R.; Lamont, I. L.; Kong, X.; Hider, R. C.; Cornelis, P.; Cámara, M.; Williams, P. The pseudomonas aeruginosa 4-quinolone signal molecules hhq and pqs play multifunctional roles in quorum sensing and iron entrapment. *Chemistry & Biology* **2007**, *14*, 87-96.
- (191) Lépine, F.; Milot, S.; Déziel, E.; He, J.; Rahme, L. G. Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (haqs) produced by pseudomonas aeruginosa. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 862-869.
- (192) de Kievit, T. R. Quorum sensing in pseudomonas aeruginosa biofilms. *Environ Microbiol* **2009**, *11*, 279-288.
- (193) Abdel-Mawgoud, A.; Lépine, F.; Déziel, E. Rhamnolipids: Diversity of structures, microbial origins and roles. *Applied Microbiology and Biotechnology* **2010**, *86*, 1323-1336.

- (194) Benincasa, M.; Abalos, A.; Oliveira, I.; Manresa, A. Chemical structure, surface properties and biological activities of the biosurfactant produced by *Pseudomonas aeruginosa* IBI from soapstock. *Antonie van Leeuwenhoek* **2004**, *85*, 1-8.
- (195) Kownatzki, R.; Tümmler, B.; Döring, G. Rhamnolipid of *Pseudomonas aeruginosa* in sputum of cystic fibrosis patients. *The Lancet* **1987**, *329*, 1026-1027.
- (196) Read, R. C.; Roberts, P.; Munro, N.; Rutman, A.; Hastie, A.; Shryock, T.; Hall, R.; McDonald-Gibson, W.; Lund, V.; Taylor, G.; et al. *Effect of Pseudomonas aeruginosa rhamnolipids on mucociliary transport and ciliary beating*, 1992; Vol. 72, p 2271-2277.
- (197) G.V.A.O. Costa, S.; Nitschke, M.; Haddad, R.; Eberlin, M. N.; Contiero, J. Production of *Pseudomonas aeruginosa* IBI rhamnolipids following growth on Brazilian native oils. *Process Biochemistry* **2006**, *41*, 483-488.
- (198) Zgoła-Grzeškowiak, A.; Kaczorek, E. Isolation, preconcentration and determination of rhamnolipids in aqueous samples by dispersive liquid-liquid microextraction and liquid chromatography with tandem mass spectrometry. *Talanta* **2011**, *83*, 744-750.
- (199) Irfan-Maqsood, M.; Seddiq-Shams, M. Rhamnolipids: Well-characterized glycolipids with potential broad applicability as biosurfactants. *Industrial Biotechnology* **2014**, *10*, 285-291.
- (200) Linhardt, R. J.; Bakhit, R.; Daniels, L.; Mayerl, F.; Pickenhagen, W. Microbially produced rhamnolipid as a source of rhamnose. *Biotechnology and Bioengineering* **1989**, *33*, 365-368.
- (201) Heerklotz, H.; Wieprecht, T.; Seelig, J. Membrane perturbation by the lipopeptide surfactin and detergents as studied by deuterium NMR. *The Journal of Physical Chemistry B* **2004**, *108*, 4909-4915.
- (202) Tang, J. S.; Zhao, F.; Gao, H.; Dai, Y.; Yao, Z. H.; Hong, K.; Li, J.; Ye, W. C.; Yao, X. S. Characterization and online detection of surfactin isomers based on HPLC-MS(n) analyses and their inhibitory effects on the overproduction of nitric oxide and the release of TNF- α and IL-6 in LPS-induced macrophages. *Marine Drugs* **2010**, *8*, 2605-2618.
- (203) Rooney, A. P.; Price, N. P.; Ehrhardt, C.; Swezey, J. L.; Bannan, J. D. Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *Inaquosorum* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology* **2009**, *59*, 2429-2436.
- (204) Lee, S. Y. Plastic bacteria? Progress and prospects for polyhydroxyalkanoate production in bacteria. *Trends in Biotechnology* **1996**, *14*, 431-438.
- (205) Zinn, M.; Witholt, B.; Egli, T. Occurrence, synthesis and medical application of bacterial polyhydroxyalkanoate. *Advanced Drug Delivery Reviews* **2001**, *53*, 5-21.
- (206) Yabuuchi, E.; Kosako, Y.; Oyaizu, H.; Yano, I.; Hotta, H.; Hashimoto, Y.; Ezaki, T.; Arakawa, M. Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiology and Immunology* **1992**, *36*, 1251-1275.
- (207) Vesth, T.; Ozen, A.; Andersen, S. C.; Kaas, R. S.; Lukjancenko, O.; Bohlin, J.; Nookaew, I.; Wassenaar, T. M.; Ussery, D. W. *Veillonella*, *Firmicutes*: Microbes disguised as gram negatives. *Standards in Genomic Sciences* **2013**, *9*, 431-448.
- (208) Alatoon, A. A.; Cunningham, S. A.; Ihde, S. M.; Mandrekar, J.; Patel, R. Comparison of direct colony method versus extraction method for identification of gram-positive cocci by use of Bruker Biotyper matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* **2011**, *49*, 2868-2873.
- (209) Rodgers, G. L.; Klugman, K. P. The future of pneumococcal disease prevention. *Vaccine* **2011**, *29*, Supplement 3, C43-C48.
- (210) Habib, M.; Porter, B. D.; Satzke, C. Capsular serotyping of *Streptococcus pneumoniae* using the quellung reaction. *Journal of Visualized Experiments* **2014**, e51208.

- (211) Richter, S. S.; Heilmann, K. P.; Dohrn, C. L.; Riahi, F.; Diekema, D. J.; Doern, G. V. Evaluation of pneumococcal serotyping by multiplex pcr and quellung reactions. *Journal of Clinical Microbiology* **2013**, *51*, 4193-4195.
- (212) Hrabak, J.; Chudackova, E.; Walkova, R. Matrix-assisted laser desorption ionization-time of flight (maldi-tof) mass spectrometry for detection of antibiotic resistance mechanisms: From research to routine diagnosis. *Clinical Microbiology Reviews* **2013**, *26*, 103-114.
- (213) Sparbier, K.; Schubert, S.; Weller, U.; Boogen, C.; Kostrzewa, M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection of resistance against β -lactam antibiotics. *Journal of Clinical Microbiology* **2012**, *50*, 927-937.
- (214) Arnold, R. S.; Thom, K. A.; Sharma, S.; Phillips, M.; Kristie Johnson, J.; Morgan, D. J. Emergence of klebsiella pneumoniae carbapenemase-producing bacteria. *Southern Medical Journal* **2011**, *104*, 40-45.
- (215) Hrabák, J.; Študentová, V.; Walková, R.; Žemličková, H.; Jakubů, V.; Chudáčková, E.; Gniadkowski, M.; Pfeifer, Y.; Perry, J. D.; Wilkinson, K.; Bergerová, T. Detection of ndm-1, vim-1, kpc, oxa-48, and oxa-162 carbapenemases by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* **2012**, *50*, 2441-2443.
- (216) Wold, S.; Sjöström, M.; Eriksson, L. Pls-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109-130.
- (217) Ikediobi, O. N.; Davies, H.; Bignell, G.; Edkins, S.; Stevens, C.; O'Meara, S.; Santarius, T.; Avis, T.; Barthorpe, S.; Brackenbury, L.; Buck, G.; Butler, A.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Hunter, C.; Jenkinson, A.; Jones, D.; Kosmidou, V.; Lugg, R.; Menzies, A.; Mironenko, T.; Parker, A.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Smith, R.; Solomon, H.; Stephens, P.; Teague, J.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Reinhold, W.; Weinstein, J. N.; Stratton, M. R.; Futreal, P. A.; Wooster, R. Mutation analysis of 24 known cancer genes in the nci-60 cell line set. *Molecular Cancer Therapeutics* **2006**, *5*, 2606-2612.
- (218) Shoemaker, R. H. The nci60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer* **2006**, *6*, 813-823.
- (219) Su, G.; Burant, C. F.; Beecher, C. W.; Athey, B. D.; Meng, F. Integrated metabolome and transcriptome analysis of the nci60 dataset. *BMC Bioinformatics* **2011**, *12 Suppl 1*, S36.
- (220) Doria, M. L.; Cotrim, C. Z.; Simoes, C.; Macedo, B.; Domingues, P.; Domingues, M. R.; Helguero, L. A. Lipidomic analysis of phospholipids from human mammary epithelial and breast cancer cell lines. *Journal of Cell Physiology* **2013**, *228*, 457-468.
- (221) Fillet, M.; Van Heugen, J. C.; Servais, A. C.; De Graeve, J.; Crommen, J. Separation, identification and quantitation of ceramides in human cancer cells by liquid chromatography-electrospray ionization tandem mass spectrometry. *Journal of Chromatography A* **2002**, *949*, 225-233.
- (222) Bodzon-Kulakowska, A.; Cichon, T.; Golec, A.; Drabik, A.; Ner, J.; Suder, P. Desi-ms as a tool for direct lipid analysis in cultured cells. *Cytotechnology* **2015**, *67*, 1085-1091.
- (223) Fhaner, C. J.; Liu, S.; Ji, H.; Simpson, R. J.; Reid, G. E. Comprehensive lipidome profiling of isogenic primary and metastatic colon adenocarcinoma cell lines. *Analytical Chemistry* **2012**, *84*, 8917-8926.
- (224) Shrestha, B.; Sripadi, P.; Walsh, C. M.; Razunguzwa, T. T.; Powell, M. J.; Kehn-Hall, K.; Kashanchi, F.; Vertes, A. Rapid, non-targeted discovery of biochemical transformation and biomarker candidates in oncovirus-infected cell lines using laesi mass spectrometry. *Chemical Communications* **2012**, *48*, 3700-3702.
- (225) Sripadi, P.; Shrestha, B.; Easley, R. L.; Carpio, L.; Kehn-Hall, K.; Chevalier, S.; Mahieux, R.; Kashanchi, F.; Vertes, A. Direct detection of diverse metabolic changes in virally transformed and tax-expressing cells by mass spectrometry. *PLoS ONE* **2010**, *5*, e12590.

- (226) Calligaris, D.; Caragacianu, D.; Liu, X.; Norton, I.; Thompson, C. J.; Richardson, A. L.; Golshan, M.; Easterling, M. L.; Santagata, S.; Dillon, D. A.; Jolesz, F. A.; Agar, N. Y. R. Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *Proceedings of the National Academy of Sciences* **2014**, *111*, 15184-15189.
- (227) Eberlin, L. S.; Norton, I.; Orringer, D.; Dunn, I. F.; Liu, X.; Ide, J. L.; Jarmusch, A. K.; Ligon, K. L.; Jolesz, F. A.; Golby, A. J.; Santagata, S.; Agar, N. Y. R.; Cooks, R. G. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proceedings of the National Academy of Sciences* **2013**, *110*, 1611-1616.
- (228) Ross, D. T.; Scherf, U.; Eisen, M. B.; Perou, C. M.; Rees, C.; Spellman, P.; Iyer, V.; Jeffrey, S. S.; Van de Rijn, M.; Waltham, M.; Pergamenschikov, A.; Lee, J. C. F.; Lashkari, D.; Shalon, D.; Myers, T. G.; Weinstein, J. N.; Botstein, D.; Brown, P. O. Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics* **2000**, *24*, 227-235.
- (229) Scherf, U.; Ross, D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O.; Weinstein, J. N. A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* **2000**, *24*, 236-244.
- (230) Garraway, L. A.; Widlund, H. R.; Rubin, M. A.; Getz, G.; Berger, A. J.; Ramaswamy, S.; Beroukhi, R.; Milner, D. A.; Granter, S. R.; Du, J.; Lee, C.; Wagner, S. N.; Li, C.; Golub, T. R.; Rimm, D. L.; Meyerson, M. L.; Fisher, D. E.; Sellers, W. R. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **2005**, *436*, 117-122.
- (231) Rae, J. M.; Creighton, C. J.; Meck, J. M.; Haddad, B. R.; Johnson, M. D. Mda-mb-435 cells are derived from m14 melanoma cells--a loss for breast cancer, but a boon for melanoma research. *Breast Cancer Research and Treatment* **2007**, *104*, 13-19.
- (232) Liscovitch, M.; Ravid, D. A case study in misidentification of cancer cell lines: MCF-7/ADR cells (re-designated NCI/ADR-RES) are derived from OVCAR-8 human ovarian carcinoma cells. *Cancer Letters* **2007**, *245*, 350-352.
- (233) D'Andrea, S.; Guillou, H.; Jan, S.; Catheline, D.; Thibault, J. N.; Bouriel, M.; Rioux, V.; Legrand, P. The same rat delta6-desaturase not only acts on 18- but also on 24-carbon fatty acids in very-long-chain polyunsaturated fatty acid biosynthesis. *Biochemistry Journal* **2002**, *364*, 49-55.
- (234) Guillou, H.; Rioux, V.; Catheline, D.; Thibault, J.-N.; Bouriel, M.; Jan, S.; D'Andrea, S.; Legrand, P. Conversion of hexadecanoic acid to hexadecenoic acid by rat delta6-desaturase. *Journal of Lipid Research* **2003**, *44*, 450-454.
- (235) Park, W. J.; Kothapalli, K. S. D.; Lawrence, P.; Tyburczy, C.; Brenna, J. T. An alternate pathway to long-chain polyunsaturates: The FADS2 gene product delta8-desaturates 20:2n-6 and 20:3n-3. *Journal of Lipid Research* **2009**, *50*, 1195-1202.
- (236) Ishibashi, Y.; Kohyama-Koganeya, A.; Hirabayashi, Y. New insights on glucosylated lipids: Metabolism and functions. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2013**, *1831*, 1475-1485.
- (237) Uphoff, C. C.; Denkmann, S.-A.; Drexler, H. G. Treatment of mycoplasma contamination in cell cultures with plasmocin. *Journal of Biomedicine and Biotechnology* **2012**, *2012*, 8.
- (238) Benjamini, Y.; Hochberg, Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **2000**, *25*, 60-83.
- (239) Drexler, H.; Uphoff, C. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention. *Cytotechnology* **2002**, *39*, 75-90.
- (240) Ley, R. E.; Bäckhed, F.; Turnbaugh, P.; Lozupone, C. A.; Knight, R. D.; Gordon, J. I. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences* **2005**, *102*, 11070-11075.

- (241) Manichanh, C.; Rigottier-Gois, L.; Bonnaud, E.; Gloux, K.; Pelletier, E.; Frangeul, L.; Nalin, R.; Jarrin, C.; Chardon, P.; Marteau, P.; Roca, J.; Dore, J. Reduced diversity of faecal microbiota in crohn's disease revealed by a metagenomic approach. *Gut* **2006**, *55*, 205-211.
- (242) Boleij, A.; Tjalsma, H. The itinerary of streptococcus gallolyticus infection in patients with colonic malignant disease. *The Lancet Infectious Diseases* **2013**, *13*, 719-724.
- (243) Wang, X.; Yang, Y.; Huycke, M. M. Commensal bacteria drive endogenous transformation and tumour stem cell marker expression through a bystander effect. *Gut* **2015**, *64*, 459-468.
- (244) Rubinstein, Mara R.; Wang, X.; Liu, W.; Hao, Y.; Cai, G.; Han, Yiping W. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/ β -catenin signaling via its fada adhesin. *Cell Host & Microbe* **2013**, *14*, 195-206.
- (245) Wu, S.; Rhee, K.-J.; Albesiano, E.; Rabizadeh, S.; Wu, X.; Yen, H.-R.; Huso, D. L.; Brancati, F. L.; Wick, E.; McAllister, F.; Housseau, F.; Pardoll, D. M.; Sears, C. L. A human colonic commensal promotes colon tumorigenesis via activation of t helper type 17 t cell responses. *Nature Medicine* **2009**, *15*, 1016-1022.
- (246) Pevsner, P. H.; Melamed, J.; Remsen, T.; Kogos, A.; Francois, F.; Kessler, P.; Stern, A.; Anand, S. Mass spectrometry maldi imaging of colon cancer biomarkers: A new diagnostic paradigm. *Biomarkers in Medicine* **2009**, *3*, 55-69.
- (247) Balog, J.; Kumar, S.; Alexander, J.; Golf, O.; Huang, J.; Wiggins, T.; Abbassi-Ghadi, N.; Enyedi, A.; Kacska, S.; Kinross, J.; Hanna, G. B.; Nicholson, J. K.; Takats, Z. In vivo endoscopic tissue identification by rapid evaporative ionization mass spectrometry (REIMS). *Angewandte Chemie-International Edition* **2015**, *54*, 11059-11062.
- (248) Hugenholtz, P.; Goebel, B. M.; Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* **1998**, *180*, 4765-4774.
- (249) Muyzer, G.; Smalla, K. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek* **1998**, *73*, 127-141.
- (250) Weinstock, G. M. Genomic approaches to studying the human microbiota. *Nature* **2012**, *489*, 250-256.
- (251) Poppert, S.; Essig, A.; Marre, R.; Wagner, M.; Horn, M. Detection and differentiation of chlamydiae by fluorescence in situ hybridization. *Applied and Environmental Microbiology* **2002**, *68*, 4081-4089.
- (252) Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A. Diversity of the human intestinal microbial flora. *Science* **2005**, *308*, 1635-1638.
- (253) Turnbaugh, P. J.; Hamady, M.; Yatsunenko, T.; Cantarel, B. L.; Duncan, A.; Ley, R. E.; Sogin, M. L.; Jones, W. J.; Roe, B. A.; Affourtit, J. P.; Egholm, M.; Henrissat, B.; Heath, A. C.; Knight, R.; Gordon, J. I. A core gut microbiome in obese and lean twins. *Nature* **2009**, *457*, 480-484.

Appendix

Appendix 1. Determination of biomass needed for REIMS measurements for *S. warneri*, *P. aeruginosa* and *H. influenzae*. Measurement pre- and post-biomass uptake each performed in triplicate.

tube empty [g]				tube full [g]				biomass
Rep1	Rep2	Rep3	mean	Rep1	Rep2	Rep3	mean	[mg]
<i>Staphylococcus warneri</i>								
1.1890	1.1887	1.1891	1.1889	1.1894	1.1895	1.1894	1.1894	0.5000
1.1870	1.1872	1.1873	1.1872	1.1881	1.1880	1.1880	1.1880	0.8667
1.1944	1.1945	1.1950	1.1946	1.1947	1.1949	1.1949	1.1948	0.2000
1.1975	1.1975	1.1978	1.1976	1.1982	1.1984	1.1983	1.1983	0.7000
1.1863	1.1865	1.1863	1.1864	1.1869	1.1870	1.1870	1.1870	0.6000
<i>Pseudomonas aeruginosa</i>								
1.1936	1.1936	1.1938	1.1937	1.1947	1.1948	1.1946	1.1947	1.0333
1.2038	1.2037	1.2038	1.2038	1.2044	1.2044	1.2042	1.2043	0.5667
1.2005	1.2005	1.2000	1.2003	1.2011	1.2015	1.2012	1.2013	0.9333
1.1924	1.1914	1.1925	1.1921	1.1927	1.1928	1.1930	1.1928	0.7333
1.2114	1.2115	1.2117	1.2115	1.2120	1.2123	1.2123	1.2122	0.6667
<i>Haemophilus influenzae</i>								
1.1826	1.1827	1.1824	1.1826	1.1830	1.1829	1.1828	1.1829	0.3333
1.1978	1.1976	1.1975	1.1976	1.1981	1.1979	1.1982	1.1981	0.4333
1.1987	1.1988	1.1987	1.1987	1.1994	1.1993	1.1992	1.1993	0.5667
1.1947	1.1948	1.1947	1.1947	1.1952	1.1950	1.1949	1.1950	0.3000
1.1946	1.1947	1.1947	1.1947	1.1948	1.1949	1.1950	1.1949	0.2333

Appendix 2. List of bacterial species and number of respective spectral entries in REIMS spectral database as determined on 14.01.2015.

Species	Database entries
<i>Abiotrophia defectiva</i>	1
<i>Acetomicrobium faecale</i>	1
<i>Achromobacter spp</i>	6
<i>Achromobacter xylosoxidans</i>	6
<i>Acidaminococcus fermentans</i>	2
<i>Acidovorax temperans</i>	2
<i>Acinetobacter baumannii</i>	22
<i>Acinetobacter iwoffii</i>	15
<i>Acinetobacter johnsonii</i>	2
<i>Acintebacter junii</i>	1
<i>Acinetobacter pittii</i>	1
<i>Actinobaculum schaalii</i>	2
<i>Actinomyces gravenitzii</i>	1
<i>Actinomyces israelii</i>	1
<i>Actinomyces odontolyticus</i>	5
<i>Actinomyces oris</i>	5
<i>Actinomyces turicensis</i>	1
<i>Actinomyces sp</i>	1
<i>Actinomyces viscosus</i>	2
<i>Aerococcus sp</i>	1
<i>Aerococcus viridans</i>	2
<i>Aeromonas hydrophila</i>	1
<i>Aggregatibacter aphrophilus</i>	7
<i>Alcaligenes faecalis</i>	3
<i>Alistipes onderdonkii</i>	1
<i>Arthrobacter creatinolyticus</i>	1
<i>Arthrobacter sp</i>	1
<i>Bacillus cereus</i>	21
<i>Bacillus clausii</i>	8
<i>Bacillus licheniformis</i>	8
<i>Bacillus pumilus</i>	1
<i>Bacillus sonorensis</i>	1
<i>Bacillus subtilis</i>	7
<i>Bacillus spp</i>	5
<i>Bacteroides acidifaciens</i>	2
<i>Bacteroides caccae</i>	2
<i>Bacteroides eggerthii</i>	2
<i>Bacteroides fragilis</i>	39
<i>Bacteroides helcogenes</i>	1
<i>Bacteroides ovatus</i>	15
<i>Bacteroides pyogenes</i>	1
<i>Bacteroides thetaiotaomicron</i>	10
<i>Bacteroides uniformis</i>	7

Species	Database entries
<i>Bacteroides vulgatus</i>	26
<i>Bifidobacterium adolescentis</i>	1
<i>Bifidobacterium animalis</i>	1
<i>Bifidobacterium bifidum</i>	2
<i>Bifidobacterium breve</i>	4
<i>Bifidobacterium infantis</i>	1
<i>Bifidobacterium longum</i>	11
<i>Bifidobacterium pseudocatenulatum</i>	2
<i>Brevibacterium sp</i>	3
<i>Brevibacterium paucivorans</i>	1
<i>Brevundimonas diminuta</i>	2
<i>Burkholderia cepacia complex</i>	16
<i>Campylobacter coli</i>	2
<i>Campylobacter fetus</i>	3
<i>Campylobacter jejuni</i>	6
<i>Campylobacter sp</i>	15
<i>Cardiobacterium hominis</i>	4
<i>Chryseobacterium sp</i>	1
<i>Chryseobacterium indologenes</i>	3
<i>Citrobacter amalonaticus</i>	1
<i>Citrobacter braakii</i>	3
<i>Citrobacter freundii</i>	16
<i>Citrobacter koseri</i>	31
<i>Clostridium difficile</i>	702
<i>Clostridium celerecrescens</i>	1
<i>Clostridium hystolyticum</i>	2
<i>Clostridium innocuum</i>	5
<i>Clostridium paraputrificum</i>	2
<i>Clostridium perfringens</i>	11
<i>Clostridium ramosum</i>	3
<i>Clostridium septicum</i>	2
<i>Clostridium sordellii</i>	1
<i>Clostridium sporogenes</i>	2
<i>Clostridium tertium</i>	3
<i>Comamonas kerstersii</i>	2
<i>Comamonas sp</i>	1
<i>Corynebacterium afermentans</i>	2
<i>Corynebacterium amycolatum</i>	4
<i>Corynebacterium diphtheriae</i>	2
<i>Corynebacterium imitans</i>	3
<i>Corynebacterium minutissimum</i>	1
<i>Corynebacterium striatum</i>	5
<i>Corynebacterium spp</i>	11
<i>Delftia acidovorans</i>	4
<i>Delftia dentocariosa</i>	1

Species	Database entries
<i>Delftia sp</i>	2
<i>Dermabacter sp</i>	3
<i>Dialister sp</i>	1
<i>Eggerthella hongkongensis</i>	1
<i>Eggerthella lenta</i>	1
<i>Eikenella corrodens</i>	1
<i>Elizabethkingia meningoseptica</i>	4
<i>Enterobacter absuriae</i>	2
<i>Enterobacter aerogenes</i>	34
<i>Enterobacter amnigenus</i>	1
<i>Enterobacter cloacae</i>	76
<i>Enterobacter gergoviae</i>	1
<i>Enterococcus avium</i>	3
<i>Enterococcus casseliflavus</i>	2
<i>Enterococcus cecorum</i>	1
<i>Enterococcus faecalis</i>	34
<i>Enterococcus faecium</i>	31
<i>Enterococcus gallinarum</i>	5
<i>Enterococcus gallolyticus</i>	1
<i>Enterococcus raffinosus</i>	3
<i>Escherichia coli</i>	142
<i>Gardnerella vaginalis</i>	3
<i>Fusobacterium gonidiaformans</i>	3
<i>Fusobacterium necrophorum</i>	2
<i>Fusobacterium nucleatum</i>	20
<i>Fusobacterium peridotiam</i>	4
<i>Fusobacterium sp</i>	1
<i>Granulicatella adiacens</i>	1
<i>Haemophilus influenzae</i>	34
<i>Haemophilus parahaemolyticus</i>	2
<i>Haemophilus parainfluenzae</i>	1
<i>Hafnia alvei</i>	3
<i>Hafnia paralvei</i>	2
<i>Hafnia sp</i>	1
<i>Helicobacter pylori</i>	3
<i>Kingella kingae</i>	2
<i>Kingella sp</i>	1
<i>Klebsiella oxytoca</i>	35
<i>Klebsiella pneumoniae</i>	119
<i>Kokuria kristina</i>	2
<i>Kokuria rhizophila</i>	2
<i>Kokuria varians</i>	1
<i>Lactobacillus rhamnosus</i>	3
<i>Lactococcus gasseri</i>	2

Species	Database entries
<i>Lactococcus lactis</i>	1
<i>Lactococcus sp</i>	2
<i>Leuconostoc sp</i>	1
<i>Listeria monocytogenes</i>	14
<i>Microbacterium sp</i>	1
<i>Micrococcus luteus</i>	33
<i>Micrococcus lylae</i>	1
<i>Moraxella catarrhalis</i>	18
<i>Moraxella osloensis</i>	3
<i>Morganella morganii</i>	30
<i>Mycobacterium avium</i>	2
<i>Mycobacterium fortuitum</i>	1
<i>Mycobacterium peregrium</i>	1
<i>Myroides odoratimimus</i>	2
<i>Neisseria cineria</i>	1
<i>Neissera gonorrhoeae</i>	20
<i>Neisseria elongata</i>	2
<i>Neisseria flavescens</i>	5
<i>Neisseria lactamica</i>	3
<i>Neisseria meningitidis</i>	7
<i>Neisseria mucosa</i>	2
<i>Nocardia asteroides</i>	1
<i>Nocardia sp</i>	1
<i>Paenibacillus sp</i>	5
<i>Paenibacillus unalis</i>	1
<i>Pandoraea spp</i>	1
<i>Panthoea sp.</i>	1
<i>Parabacteroides distonasis</i>	7
<i>Parabacteroides johnsonii</i>	2
<i>Parvimonas micra</i>	1
<i>Pasteurella multocida</i>	1
<i>Peptoniphilus harei</i>	5
<i>Propionibacterium acnes</i>	14
<i>Prevotella bivia</i>	9
<i>Proteus mirabilis</i>	34
<i>Proteus vulgaris</i>	8
<i>Providencia rettgeri</i>	2
<i>Providencia stuartii</i>	2
<i>Pseudomonas aeruginosa</i>	268
<i>Pseudomonas luteola</i>	1
<i>Pseudomonas monteilii</i>	2
<i>Pseudomonas oryzihabitans</i>	2
<i>Pseudomonas putida</i>	1
<i>Pseudomonas stutzeri</i>	11
<i>Raoultella ornitholytica</i>	1

Species	Database entries
<i>Raoultella planticola</i>	1
<i>Rhizobium radiobacter</i>	5
<i>Rhodococcus equi</i>	1
<i>Rhodococcus sp</i>	2
<i>Roseomonas mucosa</i>	6
<i>Roseomonas sp</i>	1
<i>Rothia aerea</i>	3
<i>Rothia amarne</i>	1
<i>Rothia dentocariosa</i>	6
<i>Rothia mucilaginoso</i>	12
<i>Rothia sp</i>	1
<i>Salmonella poona</i>	1
<i>Serratia liquifaciens</i>	3
<i>Serratia marcescens</i>	41
<i>Shigella sonnei</i>	1
<i>Staphylococcus aureus</i>	80
<i>Staphylococcus capitis</i>	30
<i>Staphylococcus caprae</i>	1
<i>Staphylococcus cohnii</i>	4
<i>Staphylococcus epidermis</i>	34
<i>Staphylococcus haemolyticus</i>	21
<i>Staphylococcus hominis</i>	30
<i>Staphylococcus lugdunensis</i>	10
<i>Staphylococcus pasteurii</i>	3
<i>Staphylococcus pettenkoferi</i>	3
<i>Staphylococcus saprophyticus</i>	10
<i>Staphylococcus simulans</i>	5
<i>Staphylococcus warneri</i>	21
<i>Stenotrophomonas maltophilia</i>	30
<i>Streptococcus agalactiae</i>	94
<i>Streptococcus anginosus</i>	14
<i>Streptococcus bovis</i>	3
<i>Streptococcus canis</i>	1
<i>Streptococcus constellatus</i>	2
<i>Streptococcus cristitus</i>	2
<i>Streptococcus dysagalactiae</i>	17
<i>Streptococcus gallolyticus</i>	10
<i>Streptococcus gordonii</i>	6
<i>Streptococcus group C or G</i>	7
<i>Streptococcus intermedius</i>	5
<i>Streptococcus lutetiensis</i>	6
<i>Streptococcus milleri group</i>	19
<i>Streptococcus mitis</i>	5
<i>Streptococcus mutans</i>	7
<i>Streptococcus oralis</i>	24

Species	Database entries
<i>Streptococcus parasanguinus</i>	11
<i>Streptococcus pneumoniae</i>	88
<i>Streptococcus povas</i>	1
<i>Streptococcus pseudoporcinus</i>	2
<i>Streptococcus pyogenes</i>	32
<i>Streptococcus salivarius</i>	23
<i>Streptococcus sanguinus</i>	14
<i>Streptococcus vestibularis</i>	1
<i>Streptococcus viridans</i>	6
<i>Sutterella wadsworthensis</i>	2
<i>Veillonella atypica</i>	1
<i>Veillonella dispar</i>	1
<i>Veillonella parvula</i>	1
<i>Veillonella ratti</i>	1
<i>Vibrio alginolyticus</i>	1
<i>Vibrio cholerae</i>	1
<i>Vibrio furnissii</i>	1
<i>Vibrio parahaemolyticus</i>	1
<i>Virgibacillus proomii</i>	3
<i>Yersinia enterocolitica</i>	1
TOTAL	3054

Appendix 3. List of fungal species and number of respective spectral entries in REIMS spectral database as determined on 14.01.2015.

Species	Number of database entries
<i>Candida albicans</i>	56
<i>Candida dubliensis</i>	2
<i>Candida famata</i>	4
<i>Candida glabrata</i>	32
<i>Candida guilliermondii</i>	7
<i>Candida inconspicua</i>	4
<i>Candida kefyr</i>	4
<i>Candida kruzei</i>	14
<i>Candida lambica</i>	2
<i>Candida lipolytica</i>	3
<i>Candida lusitanae</i>	15
<i>Candida maris</i>	1
<i>Candida nivariensis</i>	1
<i>Candida parapsilosis</i>	44
<i>Candida pelliculosa</i>	1
<i>Candida rugosa</i>	1
<i>Candida sake</i>	1
<i>Candida sphaerica</i>	1
<i>Candida tropicalis</i>	17
<i>Candida utilis</i>	6
<i>Cryptococcus albidus</i>	1
<i>Cryptococcus gattii</i>	1
<i>Cryptococcus humicolus</i>	1
<i>Cryptococcus laurentii</i>	1
<i>Cryptococcus neoformans</i>	11
<i>Cryptococcus saito</i>	1
<i>Cryptococcus terreus</i>	1
<i>Geotrichum candidum</i>	2
<i>Hansenula apis</i>	1
<i>Hansenula polymorpha</i>	1
<i>Kloeckera apis</i>	1
<i>Kloeckera japonica</i>	1
<i>Lodderomyces elongisporus</i>	2
<i>Pichia kluyveri</i>	1
<i>Pichia manshurica</i>	1
<i>Rhodotorula mucilaginosa</i>	3
<i>Rhodotorula rubra</i>	1
<i>Saccharomyces cerevisiae</i>	2
<i>Trichosporon asahii</i>	1
<i>Trichosporon beigelii</i>	1
<i>Trichosporon cutaneum</i>	1
<i>Trichosporon mucoides</i>	2
TOTAL	254

Appendix 4. Sample set compiled for deriving of bacterial taxonomic markers.

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.	
Gram-negatives	Bacteroidetes	Bacteroidetes	Bacteroidales	Bacteroidaceae	<i>Bacteroides</i>	<i>Bacteroides acidifaciens</i>	2	
						<i>Bacteroides caccae</i>	2	
						<i>Bacteroides eggerthii</i>	2	
						<i>Bacteroides fragilis</i>	5	
						<i>Bacteroides helcogenes</i>	1	
						<i>Bacteroides ovatus</i>	3	
						<i>Bacteroides pyogenes</i>	1	
						<i>Bacteroides thetaiotaomicron</i>	3	
						<i>Bacteroides uniformis</i>	3	
						<i>Bacteroides vulgatus</i>	3	
				Porphyromonadaceae	<i>Parabacteroides</i>	<i>Parabacteroides distasonis</i>	5	
						<i>Parabacteroides johnsonii</i>	2	
				Prevotellaceae	<i>Prevotella</i>	<i>Prevotella bivia</i>	7	
				Rikenellaceae	<i>Alistipes</i>	<i>Alistipes onderdonkii</i>	1	
	Flavobacteria	Flavobacteriales	Flavobacteriaceae		<i>Chryseobacterium</i>	<i>Chryseobacterium indologenes</i>	3	
						<i>Chryseobacterium sp</i>	1	
					<i>Elizabethkingia</i>	<i>Elizabethkingia meningoseptica</i>	4	
					<i>Myroides</i>	<i>Myroides odoratimimus</i>	2	
	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	<i>Fusobacterium</i>	<i>Fusobacterium gonidiaformans</i>	3	
							<i>Fusobacterium necrophorum</i>	7
							<i>Fusobacterium peridontiam</i>	4
							<i>Fusobacterium sp</i>	1
	Proteobacteria	Alpha-Proteobacteria	Caulobacterales	Caulobacteraceae	<i>Brevundimonas</i>	<i>Brevundimonas diminuta</i>	2	
				Rhizobiales	Rhizobiaceae	<i>Rhizobium</i>	<i>Rhizobium radiobacter</i>	5
				Rhodospirillales	Acetobacteraceae	<i>Roseomonas</i>	<i>Roseomonas mucosa</i>	6
						<i>Roseomonas sp</i>	1	

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.	
		Beta-Proteobacteria	Burkholderiales	Alcaligenaceae	<i>Achromobacter</i>	<i>Achromobacter</i> sp	3	
						<i>Achromobacter xylosoxidans</i>	3	
					<i>Alcaligenes</i>	<i>Alcaligenes faecalis</i>	3	
				Burkholderiaceae	<i>Burkholderia</i>	<i>Burkholderia cepacia</i> complex	7	
				Comamonadaceae	<i>Acidovorax</i>	<i>Acidovorax temperans</i>	2	
					<i>Comamonas</i>	<i>Comamonas kerstersii</i>	2	
						<i>Comamonas</i> sp	1	
						<i>Delftia</i>	<i>Delftia acidovorans</i>	4
						<i>Delftia dentocariosa</i>	1	
					<i>Delftia</i> sp	2		
				Sutterellaceae	<i>Sutterella</i>	<i>Sutterella wadsworthensis</i>	2	
				Neisseriales	Neisseriaceae	<i>Eikenella</i>	<i>Eikenella corrodens</i>	1
						<i>Kingella</i>	<i>Kingella kingae</i>	3
							<i>Kingella</i> sp	1
		<i>Neisseria</i>	<i>Neisseria cineria</i>			1		
			<i>Neisseria elongata</i>			2		
			<i>Neisseria flavescens</i>			3		
			<i>Neisseria gonorrhoea</i>			4		
			<i>Neisseria lactamica</i>			3		
		<i>Neisseria meningitidis</i>	4					
<i>Neisseria mucosa</i>	2							
Epsilon-Proteobacteria	Campylobacterales	Campylobacteraceae	<i>Campylobacter</i>	<i>Campylobacter coli</i>	1			
				<i>Campylobacter fetus</i>	3			
				<i>Campylobacter jejuni</i>	3			
		<i>Campylobacter</i> sp	6					
		Helicobacteraceae	<i>Helicobacter</i>	<i>Helicobacter pylori</i>	3			
Gamma-Proteobacteria	Aeromonadales	Aeromonadaceae	<i>Aeromonas</i>	<i>Aeromonas hydrophila</i>	1			
	Cardiobacteriales	Cardiobacteriaceae	<i>Cardiobacterium</i>	<i>Cardiobacterium hominis</i>	4			

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.
			Enterobacteriales	Enterobacteriaceae	<i>Citrobacter</i>	<i>Citrobacter amalonaticus</i>	1
						<i>Citrobacter braakii</i>	3
						<i>Citrobacter freundii</i>	4
						<i>Citrobacter koseri</i>	4
					<i>Enterobacter</i>	<i>Enterobacter absuriae</i>	2
						<i>Enterobacter aerogenes</i>	3
						<i>Enterobacter amnigenus</i>	1
						<i>Enterobacter cloacae</i>	3
						<i>Enterobacter gergoviae</i>	1
					<i>Escherichia</i>	<i>Escherichia coli</i>	7
					<i>Hafnia</i>	<i>Hafnia alvei</i>	3
						<i>Hafnia paralvei</i>	2
						<i>Hafnia sp</i>	1
					<i>Klebsiella</i>	<i>Klebsiella oxytoca</i>	5
						<i>Klebsiella pneumoniae</i>	5
					<i>Morganella</i>	<i>Morganella morganii</i>	7
			<i>Panthoea</i>	<i>Panthoea sp</i>	1		
			<i>Proteus</i>	<i>Proteus mirabilis</i>	5		
				<i>Proteus vulgaris</i>	5		
			<i>Providencia</i>	<i>Providencia rettgeri</i>	2		
				<i>Providencia stuartii</i>	2		
			<i>Raoultella</i>	<i>Raoultella ornitholytica</i>	1		
				<i>Raoultella planticola</i>	1		
			<i>Salmonella</i>	<i>Salmonella poona</i>	1		
			<i>Serratia</i>	<i>Serratia liquifaciens</i>	3		
				<i>Serratia marcescens</i>	5		
			<i>Shigella</i>	<i>Shigella sonnei</i>	1		

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.
			Pasteurellales	Pasteurellaceae	<i>Aggregatibacter</i>	<i>Aggregatibacter aphrophilus</i>	5
					<i>Haemophilus</i>	<i>Haemophilus influenzae</i>	5
						<i>Haemophilus parahaemolyticus</i>	2
			<i>Haemophilus parainfluenzae</i>	1			
			<i>Pasteurella</i>	<i>Pasteurella multocida</i>	2		
			Pseudomonadales	Moraxellaceae	<i>Acinetobacter</i>	<i>Acinetobacter baumannii</i>	5
						<i>Acinetobacter iwoffii</i>	5
						<i>Acinetobacter johnsonii</i>	2
						<i>Acinetobacter junii</i>	1
				<i>Moraxella</i>	<i>Moraxella catarrhalis</i>	5	
					<i>Moraxella osloensis</i>	2	
Pseudomonadaceae	<i>Pseudomonas</i>	<i>Pseudomonas aearuginosa</i>	7				
		<i>Pseudomonas luteola</i>	1				
		<i>Pseudomonas monteilii</i>	2				
		<i>Pseudomonas oryzihabitans</i>	2				
		<i>Pseudomonas putida</i>	1				
		<i>Pseudomonas stutzeri</i>	5				
Vibrionales	Vibrionaceae	<i>Vibrio</i>	<i>Vibrio alginolyticus</i>	1			
			<i>Vibrio cholerae</i>	1			
			<i>Vibrio furnissii</i>	1			
Xanthomonadales	Xanthomonadaceae	<i>Stenotrophomonas</i>	<i>Stenotrophomonas maltophilia</i>	7			
Gram-positives	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	<i>Actinobaculum</i>	<i>Actinobaculum schaalii</i>	2

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.	
					<i>Actinomyces</i>	<i>Actinomyces graevenitzii</i>	1	
						<i>Actinomyces israelii</i>	1	
						<i>Actinomyces odontolyticus</i>	2	
						<i>Actinomyces oris</i>	5	
						<i>Actinomyces sp</i>	1	
						<i>Actinomyces turicensis</i>	1	
						<i>Actinomyces viscosus</i>	2	
				Corynebacteriaceae	<i>Corynebacterium</i>	<i>Corynebacterium afermentans</i>	2	
						<i>Corynebacterium amycolatum</i>	3	
						<i>Corynebacterium diphtheriae</i>	2	
						<i>Corynebacterium imitans</i>	3	
						<i>Corynebacterium minutissimum</i>	1	
						<i>Corynebacterium sp</i>	5	
						<i>Corynebacterium striatum</i>	3	
				Microbacteriaceae	<i>Microbacterium</i>	<i>Microbacterium sp</i>	1	
				Mycobacteriaceae	<i>Mycobacterium</i>	<i>Mycobacterium avium</i>	2	
						<i>Mycobacterium fortuitum</i>	1	
						<i>Mycobacterium peregrinum</i>	1	
				Nocardiaceae	<i>Nocardia</i>	<i>Nocardia sp</i>	1	
					<i>Rhodococcus</i>	<i>Rhodococcus equi</i>	1	
						<i>Rhodococcus sp</i>	2	
				Propionibacteriaceae	<i>Propionibacterium</i>	<i>Propionibacterium acnes</i>	7	
			Bifidobacteriales	Bifidobacteriaceae	<i>Bifidobacterium</i>	<i>Bifidobacterium adolescentis</i>	1	
							<i>Bifidobacterium bifidum</i>	2
							<i>Bifidobacterium breve</i>	3
							<i>Bifidobacterium infantis</i>	1
							<i>Bifidobacterium longum</i>	3
							<i>Bifidobacterium pseudocatenulatum</i>	2
							<i>Gardnerella</i>	<i>Gardnerella vaginalis</i>

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.
			Micrococcales	Micrococcaceae	<i>Arthrobacter</i>	<i>Arthrobacter creatinolyticus</i>	1
						<i>Arthrobacter sp</i>	1
					<i>Kokuria</i>	<i>Kokuria kristina</i>	2
						<i>Kokuria rhizophila</i>	2
						<i>Kokuria varians</i>	1
					<i>Micrococcus</i>	<i>Micrococcus luteus</i>	5
						<i>Micrococcus lylae</i>	2
					<i>Rothia</i>	<i>Rothia aeria</i>	3
						<i>Rothia amarne</i>	1
						<i>Rothia dentocariosa</i>	5
				<i>Rothia mucilaginoso</i>		5	
				<i>Rothia sp</i>		1	
				Micrococcineae	<i>Brevibacterium</i>	<i>Brevibacterium paucivorans</i>	1
						<i>Brevibacterium sp</i>	3
				<i>Dermabacter</i>	<i>Dermabacter hominis</i>	2	
<i>Dermobacter sp</i>	1						
Firmicutes	Bacilli	Bacillales	Bacillaceae	<i>Bacillus</i>	<i>Bacillus cereus</i>	3	
					<i>Bacillus clausii</i>	3	
					<i>Bacillus lichenformis</i>	3	
					<i>Bacillus pumilus</i>	1	
					<i>Bacillus sonorensis</i>	1	
					<i>Bacillus sp</i>	3	
					<i>Bacillus subtilis</i>	3	
			Listeriaceae		<i>Listeria</i>	<i>Listeria monocytogenes</i>	7
Paenibacillaceae	<i>Paenibacillus</i>	<i>Paenibacillus sp</i>	5				
		<i>Paenibacillus unalis</i>	1				

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.	
				Staphylococcaceae	<i>Staphylococcus</i>	<i>Staphylococcus aureus</i>	3	
						<i>Staphylococcus capitis</i>	3	
						<i>Staphylococcus caprae</i>	1	
						<i>Staphylococcus cohnii</i>	4	
						<i>Staphylococcus epidermis</i>	3	
						<i>Staphylococcus haemolyticus</i>	3	
						<i>Staphylococcus hominis</i>	3	
						<i>Staphylococcus lugdunensis</i>	3	
						<i>Staphylococcus pasteurii</i>	3	
						<i>Staphylococcus pettenkoferi</i>	3	
						<i>Staphylococcus saprophyticus</i>	3	
					<i>Staphylococcus warneri</i>	3		
			Lactobacillales	Aerococcaceae	<i>Abiotrophia</i>	<i>Abiotrophia defectiva</i>	1	
					<i>Aerococcus</i>	<i>Aerococcus sp</i>	1	
						<i>Aerococcus viridans</i>	2	
					Carnobacteriaceae	<i>Granulicatella</i>	<i>Granulicatella adiacens</i>	1
					Enterococcaceae	<i>Enterococcus</i>	<i>Enterococcus avium</i>	3
							<i>Enterococcus casseliflavus</i>	2
							<i>Enterococcus cecorum</i>	1
							<i>Enterococcus faecalis</i>	3
							<i>Enterococcus faecium</i>	3
							<i>Enterococcus gallinarum</i>	3
						<i>Enterococcus raffinosus</i>	3	
				Lactobacillaceae	<i>Lactococcus</i>	<i>Lactococcus lactis</i>	1	
						<i>Lactococcus spp</i>	2	
				Leuconostocaceae	<i>Leuconostoc</i>	<i>Leuconostoc sp</i>	1	
				Streptococcaceae	<i>Lactobacillus</i>	<i>Lactobacillus gasseri</i>	2	
						<i>Lactobacillus rhamnosus</i>	3	

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.
						<i>Streptococcus agalactiae</i>	3
						<i>Streptococcus anginosus</i>	3
						<i>Streptococcus bovis</i>	3
						<i>Streptococcus canis</i>	1
						<i>Streptococcus constellatus</i>	2
						<i>Streptococcus cristatus</i>	2
						<i>Streptococcus dysgalactiae</i>	3
						<i>Streptococcus gallolyticus</i>	3
						<i>Streptococcus gordonii</i>	3
						<i>Streptococcus intermedius</i>	3
						<i>Streptococcus lutetiensis</i>	3
					<i>Streptococcus</i>	<i>Streptococcus milleri</i>	3
						<i>Streptococcus mitis</i>	3
						<i>Streptococcus mutans</i>	3
						<i>Streptococcus oralis</i>	3
						<i>Streptococcus parasanguinus</i>	3
						<i>Streptococcus pneumoniae</i>	3
						<i>Streptococcus pofas</i>	1
						<i>Streptococcus pseudoporcinus</i>	2
						<i>Streptococcus pyogenes</i>	3
						<i>Streptococcus salivarius</i>	3
						<i>Streptococcus sanguinis</i>	3
						<i>Streptococcus vestibularis</i>	1
						<i>Streptococcus viridans</i>	3

Gram-stain	Phylum	Class	Order	Family	Genus	Species	No.	
		Clostridia	Clostridiales	Clostridiaceae	<i>Clostridium</i>	<i>Clostridium celerecrescens</i>	1	
						<i>Clostridium difficile</i>	4	
						<i>Clostridium histolyticum</i>	2	
						<i>Clostridium innocuum</i>	3	
						<i>Clostridium paraputrificum</i>	2	
		<i>Clostridium perfringens</i>	3					
		<i>Clostridium ramosum</i>	3					
		<i>Clostridium septicum</i>	2					
		<i>Clostridium sporogenes</i>	2					
						<i>Clostridium tertium</i>	3	
		Negativicutes	Selenomonadales	Peptostreptococcaceae	<i>Parvinomas</i>	<i>Parvinomas micra</i>	1	
					<i>Peptoniphilus</i>	<i>Peptoniphilus harei</i>	5	
				Acidaminococcaceae	<i>Acidaminococcus</i>	<i>Acidaminococcus fermentans</i>	2	
					<i>Dialister</i>	<i>Dialister sp</i>	1	
					Veillonellaceae	<i>Veillonella</i>	<i>Veillonella atypica</i>	1
							<i>Veillonella dispar</i>	1
							<i>Veillonella parvula</i>	1
							<i>Veillonella ratti</i>	1

Appendix 5. Detailed list of taxon-specific markers detected in colorectal tissue specimens using DESI-MSI.

Taxon	Level	TSM
Bacteroidetes	P	328.2857
Bacteroidetes	P	381.2765
Bacteroidetes	P	590.4923
Bacteroidetes	P	639.4954
Bacteroidetes	P	651.4953
Bacteroidetes	P	653.5113
Bacteroidetes	P	654.5143
Bacteroidetes	P	677.5238
Bacteroidetes	P	691.5395
Bacteroidetes	P	705.5562
Bacteroidia	C	286.2466
Bacteroidia	C	602.4783
Bacteroidia	C	604.5083
Bacteroidaceae	F	576.4764
Bacteroidaceae	F	683.5583
Bacteroidaceae	F	684.5613
Bacteroidaceae	F	824.5673
<i>B. fragilis</i>	S	752.5419
<i>B. fragilis</i>	S	780.5732
Porphyromonadaceae	F	918.7191
Porphyromonadaceae	F	932.7332
Porphyromonadaceae	F	933.7362
Porphyromonadaceae	F	944.7342
Porphyromonadaceae	F	945.7372
Porphyromonadaceae	F	946.7472
Porphyromonadaceae	F	947.7502
Porphyromonadaceae	F	961.7661
Porphyromonadaceae	F	962.7691
Prevotellaceae	F	675.5453
Prevotellaceae	F	676.5503
Flavobacteria	C	324.2545
Flavobacteria	C	325.2576
Flavobacteria	C	333.2084
Flavobacteria	C	374.2365
Flavobacteria	C	392.2484
Flavobacteria	C	393.2504
Flavobacteria	C	564.4624
Flavobacteria	C	566.4794
Flavobacteria	C	567.4834
Flavobacteria	C	604.4633
Flavobacteria	C	618.4773
Flavobacteria	C	619.4813
Flavobacteria	C	639.5334
Flavobacteria	C	640.5363
14 10CM		
14 5CM		
14 CT		
15 10CM		
15 5CM		
15 CT		
17 10CM		
17 5CM		
17 CT		
22 10CM		
22 5CM		
22 CT		
31 10CM		
31 5CM		
31 CT		
36 10CM		
36 5CM		
36 CT		
37 10CM		
37 5CM		
37 CT		
38 10CM		

Taxon	Firmicutes																							
Taxon	F	F	F	S	S	O	S	C	C	C	C	F	F	F	F	S	S	F	C	C	F	F	F	F
TSM	202.0525	763.5512	903.7221	1034.677	677.4413	925.7262	348.2184	703.4923	704.4953	731.5253	732.5283	649.4453	759.5583	969.7481	970.7541	427.2104	829.6192	635.3944	461.3394	628.4443	730.4652	213.1855	439.3795	655.4713
14 10CM				x						x								x						
14 5CM																	x							
14 CT	x									x	x								x					x
15 10CM										x														
15 5CM										x														
15 CT																								
17 10CM																								
17 5CM																								
17 CT						x									x									x
22 10CM																								
22 5CM																								
22 CT						x				x														
31 10CM																								
31 5CM																								
31 CT																							x	
36 10CM		x		x														x						
36 5CM										x	x													
36 CT										x														
37 10CM										x										x			x	
37 5CM			x							x	x												x	
37 CT																								x
38 10CM																								
38 5CM										x	x												x	

Permissions to reprint

Permissions to reprint were obtained for the following publications and can be found attached to this document:

1. N. Strittmatter, M. Rebec, E. A. Jones, O. Golf, A. Abdolrasouli, J. Balog, V. Behrends, K. A. Veselkov, Z. Takats, Characterization and Identification of Clinically Relevant Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry* **2014**, *86*, 6555-6562. DOI: 10.1021/ac501075f
2. N. Strittmatter, E. A. Jones, K. A. Veselkov, M. Rebec, J. G. Bundy, Z. Takats, Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chemical Communications* **2013**, *49*, 6188-6190. DOI: 10.1039/c3cc42015a
3. J. Balog, L. Sasi-Szabó, J. Kinross, M. R. Lewis, L. J. Muirhead, K. Veselkov, R. Mirnezami, B. Dezső, L. Damjanovich, A. Darzi, J. K. Nicholson, Z. Takáts, Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Science Translational Medicine* **2013**, *5*, 194ra193, DOI: 10.1126/scitranslmed.3005623
4. J. I. Zhang, N. Talaty, A. B. Costa, Y. Xia, W. A. Tao, R. Bell, J. H. Callahan, R. G. Cooks, Rapid direct lipid profiling of bacteria using desorption electrospray ionization mass spectrometry. *International Journal of Mass Spectrometry* **2011**, *301*, 37-44. DOI: 10.1016/j.ijms.2010.06.014
5. A Figure from the www.waters.com website depicting the StepWave ion guide: http://www.waters.com/webassets/cms/category/media/overview_images/StepWave_Figure1_lg_700.jpg
6. N. Abbassi-Ghadi, E. A. Jones, K. A. Veselkov, J. Huang, S. Kumar, N. Strittmatter, O. Golf, H. Kudo, R. D. Goldin, G. B. Hanna, Z. Takats. Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: a gateway for clinical applications. *Analytical Methods* **2015**, *7*: 71-80. DOI: 10.1039/C4AY01770F.
7. N. Strittmatter, A. Lovrics, J. Sessler, J. S. McKenzie, Z. Bodai, M. L. Doria, N. Kucsma, G. Szakacs, Z. Takats. Shotgun Lipidomic Profiling of the NCI60 Cell Line Panel Using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry* **2016**, *88*, 7507-7514. DOI: 10.1021/acs.analchem.6b00187.



RightsLink®

[Home](#)[Create Account](#)[Help](#)

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Characterization and Identification of Clinically Relevant Microorganisms Using Rapid Evaporative Ionization Mass Spectrometry

Author: Nicole Strittmatter, Monica Rebec, Emrys A. Jones, et al

Publication: Analytical Chemistry

Publisher: American Chemical Society

Date: Jul 1, 2014

Copyright © 2014, American Chemical Society

[LOGIN](#)

If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry

N. Strittmatter, E. A. Jones, K. A. Veselkov, M. Rebec, J. G. Bundy and Z. Takats, *Chem. Commun.*, 2013, **49**, 6188
DOI: 10.1039/C3CC42015A

If you are not the author of this article and you wish to reproduce material from it in a third party non-RSC publication you must [formally request permission](#) using RightsLink. Go to our [Instructions for using RightsLink page](#) for details.

Authors contributing to RSC publications (journal articles, books or book chapters) do not need to formally request permission to reproduce material contained in this article provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
Reproduced from Ref. XX with permission from the Centre National de la Recherche Scientifique (CNRS) and The Royal Society of Chemistry.
- For reproduction of material from PCCP:
Reproduced from Ref. XX with permission from the PCCP Owner Societies.
- For reproduction of material from PPS:
Reproduced from Ref. XX with permission from the European Society for Photobiology, the European Photochemistry Association, and The Royal Society of Chemistry.
- For reproduction of material from all other RSC journals and books:
Reproduced from Ref. XX with permission from The Royal Society of Chemistry.

If the material has been adapted instead of reproduced from the original RSC publication "Reproduced from" can be substituted with "Adapted from".

In all cases the Ref. XX is the XXth reference in the list of references.

If you are the author of this article you do not need to formally request permission to reproduce figures, diagrams etc. contained in this article in third party publications or in a thesis or dissertation provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the Centre National de la Recherche Scientifique (CNRS) and the RSC
- For reproduction of material from PCCP:
[Original citation] - Reproduced by permission of the PCCP Owner Societies
- For reproduction of material from PPS:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the European Society for Photobiology, the European Photochemistry Association, and RSC

- For reproduction of material from all other RSC journals:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry

If you are the author of this article you still need to obtain permission to reproduce the whole article in a third party publication with the exception of reproduction of the whole article in a thesis or dissertation.

Information about reproducing material from RSC articles with different licences is available on our [Permission Requests page](#).



RightsLink®

[Home](#)
[Account Info](#)
[Help](#)


Title: Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry

Author: Júlia Balog,László Sasi-Szabó,James Kinross,Matthew R. Lewis,Laura J. Muirhead,Kirill Veselkov,Reza Mirnezami,Balázs Dezső,László Damjanovich,Ara Darzi,Jeremy K. Nicholson,Zoltán Takáts

Publication: Science Translational Medicine

Publisher: The American Association for the Advancement of Science

Date: Jul 17, 2013

Copyright © 2013, Copyright © 2013, American Association for the Advancement of Science

Logged in as:
Nicole Strittmatter
Account #:
3000998825

[LOGOUT](#)

Order Completed

Thank you for your order.

This Agreement between Nicole Strittmatter ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

[Printable details.](#)

License Number	3977101439467
License date	Oct 27, 2016
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science Translational Medicine
Licensed Content Title	Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry
Licensed Content Author	Júlia Balog,László Sasi-Szabó,James Kinross,Matthew R. Lewis,Laura J. Muirhead,Kirill Veselkov,Reza Mirnezami,Balázs Dezső,László Damjanovich,Ara Darzi,Jeremy K. Nicholson,Zoltán Takáts
Licensed Content Date	Jul 17, 2013
Licensed Content Volume	5
Licensed Content Issue	194
Volume number	5
Issue number	194
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	
Title of your thesis / dissertation	Development of novel mass spectrometric methods for the characterisation and identification of microorganisms
Expected completion date	Apr 2016
Estimated size(pages)	275

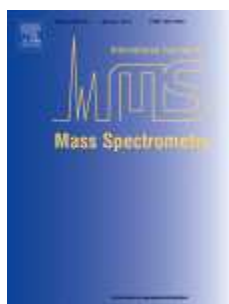
Requestor Location	Nicole Strittmatter 12 Church Street Haslingfield Cambridge, CB23 1JE United Kingdom Attn: Nicole Strittmatter
Billing Type	Invoice
Billing address	Nicole Strittmatter 12 Church Street Haslingfield Cambridge, United Kingdom CB23 1JE Attn: Nicole Strittmatter
Total	0.00 GBP

ORDER MORE**CLOSE WINDOW**

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



RightsLink®

[Home](#)
[Account Info](#)
[Help](#)


Title: Rapid direct lipid profiling of bacteria using desorption electrospray ionization mass spectrometry

Author: J. Isabella Zhang, Nari Talaty, Anthony B. Costa, Yu Xia, W. Andy Tao, Rebecca Bell, John H. Callahan, R. Graham Cooks

Publication: International Journal of Mass Spectrometry

Publisher: Elsevier

Date: 30 March 2011

Logged in as:
Nicole Strittmatter
Account #:
3000998825

[LOGOUT](#)

Copyright © 2010 Published by Elsevier B.V.

Order Completed

Thank you for your order.

This Agreement between Nicole Strittmatter ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

[Printable details.](#)

License Number	3977110173509
License date	Oct 27, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	International Journal of Mass Spectrometry
Licensed Content Title	Rapid direct lipid profiling of bacteria using desorption electrospray ionization mass spectrometry
Licensed Content Author	J. Isabella Zhang, Nari Talaty, Anthony B. Costa, Yu Xia, W. Andy Tao, Rebecca Bell, John H. Callahan, R. Graham Cooks
Licensed Content Date	30 March 2011
Licensed Content Volume	301
Licensed Content Issue	1-3
Licensed Content Pages	8
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Development of novel mass spectrometric methods for the characterisation and identification of microorganisms
Expected completion date	Apr 2016
Estimated size (number of pages)	275
Elsevier VAT number	GB 494 6272 12
Requestor Location	Nicole Strittmatter 12 Church Street Haslingfield Cambridge, CB23 1JE

United Kingdom
Attn: Nicole Strittmatter
0.00 GBP

Total

ORDER MORE

CLOSE WINDOW

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Request to reproduce an extract from a third party's published work

01.02.2016

Dear representative of Waters Corporation,

I am completing my PhD thesis at Imperial College London entitled 'Development of novel mass spectrometric methods for the characterisation and identification of microorganisms'.

I seek your permission to reprint, in my thesis a figure depicting the Stepwave ion guide which is published on the Waters Corporation homepage under the address http://www.waters.com/webassets/cms/category/media/overview_images/StepWave_Figure1_lg_700.jpg.

I would like to include this figure in my thesis which will be added to Spiral, Imperial's institutional repository <http://spiral.imperial.ac.uk/> and made available to the public under a Creative Commons Attribution-NonCommercial-NoDerivs licence.

If you are happy to grant me the permissions requested, please return a signed copy of this letter.

Yours sincerely,

Nicole Strittmatter

Permission granted for the use requested above:

I confirm that I am the copyright holder of the extract above and hereby give permission to include it in your thesis which will be made available, via the internet, for non-commercial purposes under the terms of the user licence.

Signed: 

Name: Dr Steve Pringle

Organisation: Waters Corp

Job title: Director of science

Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: a gateway for clinical applications

N. Abbassi-Ghadi, E. A. Jones, K. A. Veselkov, J. Huang, S. Kumar, N. Strittmatter, O. Golf, H. Kudo, R. D. Goldin, G. B. Hanna and Z. Takats, *Anal. Methods*, 2015, 7, 71
DOI: 10.1039/C4AY01770F

If you are not the author of this article and you wish to reproduce material from it in a third party non-RSC publication you must [formally request permission](#) using RightsLink. Go to our [Instructions for using RightsLink page](#) for details.

Authors contributing to RSC publications (journal articles, books or book chapters) do not need to formally request permission to reproduce material contained in this article provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
Reproduced from Ref. XX with permission from the Centre National de la Recherche Scientifique (CNRS) and The Royal Society of Chemistry.
- For reproduction of material from PCCP:
Reproduced from Ref. XX with permission from the PCCP Owner Societies.
- For reproduction of material from PPS:
Reproduced from Ref. XX with permission from the European Society for Photobiology, the European Photochemistry Association, and The Royal Society of Chemistry.
- For reproduction of material from all other RSC journals and books:
Reproduced from Ref. XX with permission from The Royal Society of Chemistry.

If the material has been adapted instead of reproduced from the original RSC publication "Reproduced from" can be substituted with "Adapted from".

In all cases the Ref. XX is the XXth reference in the list of references.

If you are the author of this article you do not need to formally request permission to reproduce figures, diagrams etc. contained in this article in third party publications or in a thesis or dissertation provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the Centre National de la Recherche Scientifique (CNRS) and the RSC

- For reproduction of material from PCCP:
[Original citation] - Reproduced by permission of the PCCP Owner Societies
- For reproduction of material from PPS:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the European Society for Photobiology, the European Photochemistry Association, and RSC
- For reproduction of material from all other RSC journals:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry

If you are the author of this article you still need to obtain permission to reproduce the whole article in a third party publication with the exception of reproduction of the whole article in a thesis or dissertation.

Information about reproducing material from RSC articles with different licences is available on our [Permission Requests page](#).



RightsLink®

[Home](#)[Create Account](#)[Help](#)

Title: Shotgun Lipidomic Profiling of the NCI60 Cell Line Panel Using Rapid Evaporative Ionization Mass Spectrometry

Author: Nicole Strittmatter, Anna Lovrics, Judit Sessler, et al

Publication: Analytical Chemistry

Publisher: American Chemical Society

Date: Aug 1, 2016

Copyright © 2016, American Chemical Society

[LOGIN](#)

If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#)
Comments? We would like to hear from you. E-mail us at customercare@copyright.com