

2017

Chemical, Physiological and Metabolic Interactions between Pseudomonas, Metals and Environmental Nutrients

Booth, Sean C.

Booth, S. C. (2017). Chemical, Physiological and Metabolic Interactions between Pseudomonas, Metals and Environmental Nutrients (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. doi:10.11575/PRISM/25315
<http://hdl.handle.net/11023/3658>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Chemical, Physiological and Metabolic Interactions between *Pseudomonas*, Metals and
Environmental Nutrients

by

Sean Cameron Booth

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

FEBRUARY, 2017

© Sean Cameron Booth 2017

Abstract

Environmental pollution is one of the major problems facing humanity. Bacteria are capable of removing pollutants from the environment through their metabolic activities. This works for organic pollutants, but metals inhibit the degradation process. *Pseudomonas pseudoalcaligenes* KF707 is a bacterium that is resistant to metals and is able to degrade pollutants such as polychlorinated biphenyls. In this thesis I present how interactions between the bacterium, its environment and metals affect the bacterium's physiology and metabolism of biphenyl. Chemical interactions with environmental components affect the toxicity of metals towards bacteria. By examining the tolerance of *Pseudomonas* species towards copper and aluminium in different media compositions I found that metal bioavailability and carbon source quality had a strong influence on the amount of metal they could withstand. Building on these data, I used metabolomics to understand how metals interfere with organic pollutant degradation. By quantifying the small molecules used and produced by the bacterial cell I was able to determine that metal toxicity is exacerbated by the oxidative stress of metabolizing an organic pollutant. *P. pseudoalcaligenes* KF707 can swim towards biphenyl but it was unknown how. By deleting genes that were expected to be involved in energy-taxis, a process that allows bacteria to swim to metabolizable carbon sources, I found that this was not how KF707 swims towards biphenyl. I did discover that some unexpected genes were involved in energy-taxis and also that the primary gene for this behavior, Aer, is actually a family of receptors with variable phylogenetic distribution in the genus *Pseudomonas*. These results provide new insight into the interactions between a bacterium and the nutrients and stressors in their environment.

Preface

Knowledge of how organisms interact with their environment is fundamental to understanding nature. The environment influences organisms in two key ways: by providing nutrients but also subjecting them to stress. While bacteria live in an incredibly diverse array of habitats, these tenets still hold true from the depths of the ocean to the guts of human beings. Due to their size, the habitats of bacteria can vary greatly across short distances but can also be described and defined in precise chemical detail. Bacteria can also influence their environment by using chemotaxis to swim to an optimum location or by secreting compounds that improve nutrient acquisition or limit stress. Growth of bacteria in a laboratory allows the exact concentrations of nutrients and stressors to be controlled, secretions to be measured and behavior monitored, as well as the use genetically manipulated organisms so that gene functions can be determined making it an ideal system for developing insights into how organisms interact with their environment.

Pseudomonas is one of the best studied genera of bacteria with over 247 different species isolated and described (including a great number of individual strains from many species), and 135 completely sequenced genomes. *Pseudomonas* was first described at the end of the 19th century by Walter Migula and named supposedly meaning ‘false unit’, though possibly actually meaning ‘false *Monas*’ a genus of nanoflagellates. Nowadays it acts as a model organism for studies of biofilm formation, chemotaxis and metabolism. The best studied and most infamous species is *P. aeruginosa*, well-known as an opportunistic pathogen, particularly in the lungs of Cystic Fibrosis patients. Many strains and pathovars of the plant pathogen *P. syringae* have been examined for their problematic infection of crops whereas *P. fluorescens* has many plant growth-promoting members. Other species, especially *P. putida* have been characterized for their bioremediation potential as organic pollutant degrading soil bacteria. This metabolic diversity, the ability to grow

rapidly using many different compounds as their sole source of carbon and energy is a defining feature of the genus.

Some species, such as *P. pseudoalcaligenes* KF707 are able to degrade polychlorinated biphenyls (PCBs), a group of toxic organic compounds that saw widespread use until they were banned in 1977. Despite this ban their high stability has allowed them to persist in contaminated environments. The ability to degrade PCBs is an interesting interaction between an organism and its environment as a stressor is also a nutrient. *P. pseudoalcaligenes* KF707 can also swim chemotactically towards biphenyl and has been found to be resistant to toxic metals, a potentially useful property as many polluted sites are co-contaminated with inorganic metals and organic carbon-based pollutants. These co-contaminated sites have been particularly problematic for bioremediation efforts as metals interfere with the degradation of organic pollutants, even in species that can withstand the metal or degrade the pollutant when challenged individually.

The interference of metal toxicity with organic pollutant degradation indicates that interactions between organisms and their environment are not independent; a stressor can influence the exploitation of nutrients. Due to their reactivity, it is possible for metals in an environment to interact directly, chemically, with nutrients and other components of the same environment. Their toxicity will also influence the physiology of organisms through the induction of stress, which could in turn affect the metabolism of the organism and therefore its interaction with nutrients. This thesis “Chemical, Physiological and Metabolic Interactions Between *Pseudomonas*, Metals and Environmental Nutrients” presents three studies aimed at investigating these topics. Each chapter is a self-contained manuscript as *Chapter Two* and *Chapter Three* have been published and *Chapter Four* is currently under review. *Chapter One* provides an introduction into the background literature. *Chapter Two* presents a foundational study of the effects of media

composition on metal toxicity (*Metallomics*, June 2013). This enabled the metabolomics work presented in *Chapter Three* where the biochemical effects of metal toxicity on biphenyl metabolism were determined (*Frontiers in Microbiology*, August 2015). *Chapter Four* presents a molecular biology study aimed at understanding energy-taxis in *Pseudomonas* as it was hypothesized that this was how *P. pseudoalcaligenes* KF707 was swimming towards biphenyl (deposited in the bioRxiv and submitted to *Molecular Microbiology*, October 2016). *Chapter Five* integrates the conclusions together and contextualizes the results in the grander scheme of environmental microbiology. Also included in this thesis are 3 appendices. *Appendix A* is a review of studies that used metabolomics to examine metal toxicity (*Metallomics*, November 2011). *Appendix B* is a review of computational tools for analyzing metabolomics data (*Computational and Structural Biology Journal*, January 2013). *Appendix C* is a book chapter detailing applications of NMR in environmental microbiology (*eMagRes*, December 2013). *Appendix D* contains supplementary methods for *Chapters Four* and *Five*. *Appendix E* contains copyright agreement information.

References to Papers Included in this Thesis

Booth SC, Workentine ML, Weljie AM, Turner RJ. 2011. Metabolomics and its application to studying metal toxicity. *Metallomics* **3**:1142.

Booth SC, George IFS, Zannoni D, Cappelletti M, Duggan GE, Ceri H, Turner RJ. 2013. Effect of aluminium and copper on biofilm development of *Pseudomonas pseudoalcaligenes* KF707 and *P. fluorescens* as a function of different media compositions. *Metallomics* **5**:723–35.

Booth SC, Weljie AM, Turner RJ. 2013. Computational Tools for the Secondary Analysis of Metabolomics Experiments. *Comput Struct Biotechnol J* **4**:e201301003.

Booth SC, Turner RJ, Weljie A. 2013. Metabolomics in Environmental Microbiology. *eMagRes* **2**.

Booth SC, Weljie AM, Turner RJ. 2015. Metabolomics reveals differences of metal toxicity in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on different carbon. *Front Microbiol* **6**:1–17.

Booth SC, Turner RJ. 2016. Aer is a family of Energy-Taxis Receptors in *Pseudomonas* which also involves Aer-2 and CttP. bioRxiv. Under review at *Molecular Microbiology*.

Acknowledgements

My journey to completing this PhD began a long time ago so I have many people I'd like to thank. First is my parents, thank you for fostering my interest in Science and supporting and encouraging me all these years. I want to thank my teachers in junior high, Mme. Littlejohn and in highschool Ms. Crichton who helped me develop my curiosity which led me here. I'd also like to thank my TAs during my undergraduate studies, especially Cynthia and Kim for getting me interested in graduate studies and introducing me to professors. Matt Workentine, who supervised me during my undergraduate project thanks for really getting me into science. For the past 6 years, Science has been a huge part of my life and I've really enjoyed becoming a scientist. If Ray hadn't given me the chance to do my undergraduate project in his lab I have no idea where I would be now. So thank you Ray for giving me an opportunity that others didn't.

Throughout my PhD I've had other mentors, especially the post-docs in the lab Joe Lemire and Denice Bay. I owe a huge debt of gratitude to Denice for teaching me so much about the reality of being a good scientist and putting up with all my questions all those years we were sitting back to back in the shared office. I also want to give a shout out to my co-grad students who made being in the lab so much fun and whom I'm proud to call my friends now. Thorin, Raymond, Marc, Matt, Lolita, hanging out with you guys at whisky Friday was one of the best parts of my PhD. Also Iain, who was first my student and then my partner in making Petri Dish Picasso a multi-year success, it's been great knowing and working with you. I should also thank my other students, Adriana, Emilie and Shelby, who were taxing at times but mostly just a joy to work with, thanks for keeping me excited and usually trying your best on your projects.

I'd also like to thank the professors I've gotten to know a bit over the years. Dr. Aalim Weljie thanks for helping make statistics and metabolomics understandable to me, Dr. Lisa Gieg thanks for being on my committee and helping me figure out how to become a post-doc. Dr. Michael Hynes and Dr. Joe Harrison, thank you for the useful scientific discussions. To my external examiners, Dr. Rebekah Devinney and Dr. Lindsay Eltis, thanks for making the time to read this thesis and attend my defence, I'm looking forward to meeting you.

I want to re-iterate my thanks to my family, and my girlfriend Jaspreet, for supporting me out of the lab, putting up with my complaints when things were down and tolerating my ego when things were up. There are many other people who helped me throughout the process, collaborators, fellow students who commiserated or made me realize things weren't that bad, senior students who showed me the way and professors who responded to cold calls for assistance that I want to recognize here for contributing to my success in intangible ways. Science is becoming an ever-more collaborative activity and I hope that I can assist others throughout my career as much as I've been helped.

Table of Contents

Abstract.....	ii
Preface.....	iii
Acknowledgements.....	vii
Table of Contents.....	ix
List of Tables.....	xv
List of Figures and Illustrations.....	xviii
List of Symbols, Abbreviations and Nomenclature.....	xxvi
Epigraph.....	xxix
CHAPTER ONE: INTRODUCTION.....	2
1.1 Remediation of Polluted Environments.....	2
1.2 Inhibition of Organic Pollutant Degradation by Metals.....	3
1.3 Metal Toxicity in Bacteria.....	6
1.3.1 Biochemical Mechanisms of Metal Toxicity.....	6
1.3.2 Direct, Specific Mechanisms.....	7
1.3.3 Direct, Non-Specific Mechanisms.....	9
1.3.4 Indirect, Specific Mechanisms.....	11
1.3.5 Indirect, Non-Specific Mechanisms.....	12
1.4 <i>P. pseudoalcaligenes</i> KF707: Studies on Biphenyl Metabolism and Metal toxicity.....	14
1.5 Chemotaxis.....	18
1.6 Research Goals and Specific Aims.....	23
CHAPTER TWO: EFFECT OF ALUMINIUM AND COPPER ON BIOFILM DEVELOPMENT OF <i>PSEUDOMONAS PSEUDOALCALIGENES</i> KF707 AND <i>P. FLUORESCENS</i> AS A FUNCTION OF DIFFERENT MEDIA COMPOSITIONS.....	25
2.1 Abstract.....	25
2.2 Introduction.....	26
2.3 Materials and Methods.....	29
2.3.1 Strains.....	29
2.3.2 Media and Solutions.....	29
2.3.3 Viability and Tolerance Testing.....	30
2.3.4 Solution State Modeling.....	31
2.3.5 Statistical Analysis.....	32
2.4 Results.....	34
2.4.1 Media Effects.....	34
2.4.2 Buffer Effects.....	36
2.4.3 Role of Carbon Source.....	38
2.4.4 Metal ion Speciation Analysis.....	40
2.5 Discussion.....	46
2.5.1 Inhibition of Biofilm Development.....	47
2.5.2 Comparison between Rich and Minimal Media.....	48
2.5.3 Effects of Carbon Source Complexity.....	50
2.5.4 Effect of Buffering Agent.....	50

2.5.5 Effect of Carbon Source	51
2.5.6 Computationally Modeled Chemical Speciation.....	53
2.5.6.1 Speciation with Different Buffering Reagents.....	54
2.5.6.2 Speciation with Different Carbon Sources	54
2.6 Conclusion	55
2.7 References.....	56

CHAPTER THREE: METABOLOMICS REVEALS DIFFERENCES OF METAL TOXICITY IN CULTURES OF *PSEUDOMONAS PSEUDOALCALIGENES* KF707 GROWN ON DIFFERENT CARBON SOURCES

3.1 Abstract.....	61
3.2 Introduction.....	62
3.3 Experimental Procedures	64
3.3.1 Culture Growth.....	64
3.3.2 Collection of Samples and Extraction of Metabolites	65
3.3.3 Derivatization and Analysis by GC-MS	65
3.3.4 Identification of Metabolites	66
3.3.5 Removal of derivatization artifacts	66
3.3.6 Quantification of Metabolites.....	67
3.3.7 Statistical Analysis: Pre-processing	68
3.3.8 Statistical Analysis	68
3.3.9 Identification of Unknown Analytes	69
3.3.10 Pathway Enrichment Analysis.....	69
3.4 Supplementary Material.....	70
3.4.1 Supplementary Methods.....	70
3.4.1.1 Removal of poor samples and outliers.....	70
3.4.1.2 Identification of Metabolites.....	70
3.4.1.3 Removal of Derivatization Artefacts	72
3.4.1.4 Quantification of Metabolites	73
3.4.1.5 Confirmation of Analyte Quantification Quality.....	73
3.4.1.6 Data Analysis: Pre-processing.....	73
3.4.1.7 Identification of Unknown Analytes.....	74
3.4.1.8 Use of p(corr) to determine correlation between metabolites and sample type.....	75
3.4.1.9 Generation and Interpretation of Secreted Metabolites SUS plots.....	75
3.4.1.10 Pathway Enrichment Analysis	76
3.4.1.11 Identification of Unknown Metabolites	79
3.5 Results and Discussion	87
3.5.1 Growth of <i>P. pseudoalcaligenes</i> KF707 in the presence of metals on succinate and biphenyl.....	87
3.5.2 Metabolomic characterization of cultures	90
3.5.3 Supervised Statistical Analysis.....	93
3.5.4 Pathway Enrichment Analysis.....	94
3.5.5 Toxicity effects of aluminum	99
3.5.6 Toxicity effects of copper.....	106
3.5.7 Effects of metal toxicity on biphenyl metabolism.....	114
3.5.8 Metabolic changes and implication of oxidative stress	116

3.6 Conclusions.....	117
3.7 References.....	119
CHAPTER FOUR: AER IS A FAMILY OF ENERGY-TAXIS RECEPTORS IN <i>PSEUDOMONAS</i> WHICH ALSO INVOLVES AER-2 AND CTP.....	132
4.1 Abstract.....	132
4.2 Introduction.....	133
4.3 Materials and Methods.....	135
4.3.1 Protein Sequences.....	135
4.3.2 Alignment.....	136
4.3.3 Sequence Harmony and Multi-Relief Analysis to Determine Groups.....	136
4.3.4 Comparison of Groups.....	137
4.3.5 Distribution of Groups.....	137
4.3.6 Detection of Evidence of Horizontal Gene Transfer.....	138
4.3.7 Culture Growth.....	138
4.3.8 Generation of Deletion Constructs and Mutants.....	138
4.3.9 Energy-taxis Swim Plates.....	140
4.3.10 Chemotaxis Swim Plates.....	140
4.4 Results.....	141
4.4.1 Bioinformatics Results.....	141
4.4.1.1 Grouping of Aer Homologs.....	145
4.4.1.2 Comparison of Groups.....	148
4.4.1.3 Distribution of Groups within <i>Pseudomonas</i>	149
4.4.1.4 Evidence for Horizontal Gene Transfer of Aer Homologs.....	152
4.4.2 Phenotypic Results.....	160
4.4.3 Discussion.....	169
4.4.3.1 Sequence and genetics classify Aer homologs into 5 groups.....	169
4.4.3.2 Aer.g1 is ancestral, prevalent and functional.....	170
4.4.3.3 Other Aer groups likely have unique functions.....	171
4.4.3.4 <i>aer</i> has been horizontally transferred.....	172
4.4.3.5 Aer2 and CttP contribute to energy-taxis in <i>P. pseudoaligenes</i> KF707.....	174
4.5 Conclusions.....	175
4.6 References.....	177
CHAPTER FIVE: OVERALL DISCUSSION AND CONCLUSIONS.....	184
5.1 Chemotaxis.....	184
5.1.1 Energy Taxis in <i>Pseudomonas pseudoalcaligenes</i> KF707.....	184
5.1.2 Chemotaxis towards Biphenyl.....	187
5.1.3 Chemotaxis in the Natural Environment.....	188
5.1.4 Function of Chemoreceptors in <i>Pseudomonas</i>	189
5.1.5 Environmental Relevance of Chemotaxis and Energy Taxis.....	191
5.1.6 Chemotaxis, a Key Interaction between Bacteria and their Environment.....	194
5.2 Growth Media: Bacteria's Environment in the Lab.....	195
5.2.1 LB: Lysogeny Broth.....	195
5.2.2 Growth Media: Mimicking an Organism's Environment in the Lab.....	196

5.3 Basic Research on how Metal Toxicity Affects Organic Pollutant Bioremediation	199
5.3.1 Further Investigations into the Effects of Xenobiotic Metabolism and Metal Toxicity	199
5.3.2 Connecting Fundamental and Applied Research	201
5.3.2.1 A Comment on Publishing and Reaching the Intended Audience.....	203
5.4 Final comments.....	204
APPENDIX A: METABOLOMICS AND ITS APPLICATION TO STUDYING METAL TOXICITY	207
A.1. Summary	207
A.2. Introduction.....	207
A.2.1. Metabolomics: A Systems Biology View of Metabolism	208
A.3. Experimental Overview of Metabolomics	211
A.3.1. Collection of Organisms.....	211
A.3.2. Metabolism Quenching and Sampling	212
A.3.3. Analytical Techniques	215
A.3.4. Data Processing	217
A.3.5. Data Interpretation.....	218
A.3.6. Bioinformatic Tools.....	219
A.4. Metabolomics of Metal Toxicity	219
A.4.1. Environmental Monitoring Organisms	221
A.4.2. Model Organisms	222
A.4.3. Plants	224
A.4.4. Bacteria.....	226
A.5. Concluding comments	228
A.6. References.....	232
APPENDIX B: COMPUTATIONAL TOOLS FOR THE SECONDARY ANALYSIS OF METABOLOMICS EXPERIMENTS	243
B.1. Abstract	243
B.2. Introduction.....	244
B.3. Background	245
B.4. Key Issues	248
B.5. Bioinformatic Basis.....	249
B.6. Metabolomics Secondary Analysis: Enrichment Analysis and Metabolite Mapping	251
B.7. Overviews of Metabolomics Secondary Analysis Tools	256
B.8. Comprehensive Platforms	260
B.8.1. MetaboAnalyst [57].	260
B.8.2. MeltDB [30]......	263
B.9. Enrichment Analysis	264
B.9.1. PAPI [24].	264
B.9.2. MBRole [27].	265
B.9.3. MPEA [29]......	266
B.9.4. TICL [68]......	266
B.9.5. IMPaLA [66]......	267

B.10. Metabolite Mapping	267
B.10.1. MetaMapp [25].	267
B.10.2. MassTriX [62].	269
B.10.3. PaintOmics [65].	270
B.10.4. VANTED [67].	270
B.10.5. Pathos [64].	271
B.10.6. ProMeTra [91].	272
B.11. Other	272
B.11.1. MetExplore [28].	272
B.11.2. Meta P-server [61].	273
B.12. Commercial Software	274
B.12.1. IPA.	274
B.12.2. MetaCore.	275
B.12.3. GeneSpring.	275
B.13. Summary and Outlook	276
B.14. References	279
APPENDIX C: METABOLOMICS IN ENVIRONMENTAL MICROBIOLOGY	291
C.1. Abstract	291
C.2. Introduction	292
C.2.1. Questions in environmental microbiology	292
C.2.2. Metabolomics	294
C.2.3. Methods	295
C.2.4. Acquisition parameters and spectral processing	297
C.2.5. Data pre-processing	299
C.2.6. Statistical Analysis	302
C.2.7. Biological Interpretation	304
C.3. Case Studies	304
C.3.1. Effect of tellurium oxyanion	304
C.3.2. Evaluation of phenotypic variants arising from a global regulator mutation	308
C.3.3. Evaluation of metal resistance in biofilms	311
C.3.4. Other Representative applications	316
C.4. Conclusions and Prospects	321
C.5. References	324
APPENDIX D: SUPPLEMENTARY METHODS FOR ENERGY TAXIS EXPERIMENTS	334
D.1. Detailed Molecular Biology Methods	334
D.1.1. PCR	334
D.1.2. Cloning	335
D.1.3. Conjugation	336
D.1.4. Counter Selection	336
D.1.5. Identification of Deletion Mutants	337
D.1.6. Complementation	337
D.2. References	341

APPENDIX E: COPYRIGHT PERMISSIONS	343
REFERENCES	347

List of Tables

Table 1-1: Summary of select studies that investigated the effects of metals on organic pollutant degradation.	5
Table 1-2: Characterized methyl-accepting chemotaxis proteins in <i>Pseudomonas</i> and BLAST hits in <i>P. pseudoalcaligenes</i> KF707. Reported values from BLAST are coverage/identity. Those in bold indicate likely homologous functions.	19
Table 2-1: OPLS statistics from models correlating biofilm and planktonic growth of <i>Pseudomonas</i> spp. in the presence of either aluminium or copper with computationally modeled chemical species concentrations in high phosphate, low phosphate and MOPS buffered media.	41
Table 2-2: OPLS statistics from models correlating biofilm and planktonic growth of <i>Pseudomonas</i> spp. in the presence of either aluminium or copper with computationally modeled chemical species concentrations in high phosphate buffered media with succinate, butyrate and benzoic acid as carbon sources.....	42
Table 3-1: Model statistics of PCA-Class models showing intraclass variation of GC-MS metabolomics quantification of samples from cells and spent media of cultures of <i>Pseudomonas pseudoalcaligenes</i> KF707 grown on either biphenyl or succinate as the sole carbon source and exposed to either control, Aluminum or copper.	71
Table 3-2: Analytes present in derivitization control samples. Retention time (RT), retention index (RI) and peak area are mean of all derivitization control samples run.	77
Table 3-3: Predicted fragmentation of 2-Phosphoglycolic acid (3TMS).....	79
Table 3-4: Functional group analysis to determine retention index of 2-phosphoglycolic acid (3TMS). Carboxyl group was interpreted as one >CO group and one –O– group.....	81
Table 3-5: Functional group analysis to determine retention index of 2-phosphoglycolic acid (3TMS). Carboxyl group was interpreted as one –CO-O-.....	82
Table 3-6: Predicted fragmentation of 2-hydroxymuconic semialdehyde(2TMS).....	83
Table 3-7: Functional group analysis to determine retention index of 2-hydroxymuconic semialdehyde (2TMS). Carboxyl group was interpreted as one –CO-O-.....	85
Table 3-8: Functional group analysis to determine retention index of 2-hydroxymuconic semialdehyde (2TMS). Carboxyl group was interpreted as one >CO group and one –O– group.	85
Table 3-9: Model statistics from principal component analysis and pairwise orthogonal partial least squares discriminant analysis (OPLS-DA) models comparing normalized, centered and scaled metabolite abundances in cells and spent media from <i>Pseudomonas pseudoalcaligenes</i> KF707 grown on either succinate or biphenyl with or without	

aluminum or copper. Components indicates number of predictive and orthogonal components present in the model, R2 indicates cumulative variance accounted for by the model, Q2 cumulative variance predicted by the model and CV-ANOVA p-value was obtained from 7-fold cross-validation analysis of variance. Models with a p-value <0.05 were considered significant..... 94

Table 3-10: Metabolic pathways affected by metal toxicity in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl (Bp) or succinate (Sc) as the sole carbon source in the presence of Al or Cu as determined by mBROLE. Presented here are modified p-values (i.e. multiple-testing corrected) indicating the probability that a pathway was affected under a particular condition. All reported pathways had an unmodified p-value <0.05 in at least one sample type. Metabolites were selected based on their VIP >0.8 as determined from OPLS-DA models comparing control to metal exposed cultures and were submitted to mBROLE. Only metabolites with a known KEGG ID were used. p-values are coloured according to significance: unmodified >0.05, not significant (black), unmodified <0.05 but modified >0.05 (grey), unmodified and modified <0.05 (white). Pathways are grouped into carbon metabolism (white), purine/pyrimidine metabolism (light grey), amino acid metabolism (grey) and sulfur containing metabolism (dark grey). 95

Table 4-1: Frequency of association of mobile elements within 5kb of *aer* genes. 151

Table 4-2: Frequency of occurrence of genes neighbouring *aer* homologs in *Pseudomonas*. ... 153

Table 4-3: Tukey Honest Significant Differences results comparing differences in normalized energy-taxis diameter in pyruvate or succinate plates. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant. 166

Table 4-4: P values from Tukey HSD results comparing differences in raw and normalized energy-taxis diameter growth rates in pyruvate or succinate plates. Growth rates were obtained by dividing the raw or normalized diameter at 48h by the value at 24h. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant..... 168

Table A-B-1: Selected Biochemical Databases 251

Table A-B-2: Select Platforms for Metabolomics Analysis and Interpretation..... 257

Table-A-C-1. Metabolic pathways enriched in response to copper exposure determined by enrichment analysis. Only pathways with a p-value <0.05 were included, and life-general metabolic pathways (such as aminoacyl-tRNA synthesis) and plant specific pathways were excluded. (reproduced from (21) with permission)..... 315

Table A-D-2: Plasmids used for energy-taxis experiments in *Pseudomonas pseudoalcaligenes* KF707..... 339

Table A-D-3: Primers used to amplify upstream and downstream regions to generate deletion constructs for deletion of Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP from *Pseudomonas pseudoalcaligenes* KF707. Restriction enzyme sites are in bold, annealing regions underlined. 340

Table A-D-4: Primers used to amplify Aer.g1 for complementation. PCR amplification product is 2228nt. After digestion with BamHI and HinDIII a 1623nt product is produced that contains only the Aer.g1 gene. Restriction enzyme sites are in bold, annealing regions underlined. 341

List of Figures and Illustrations

- Figure 2-1: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate in various Media. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown on either LB (circles, black) or minimal salts media (MSM) with either Succinate (squares, dark grey), Biphenyl (triangles, mid grey) or Benzoate (diamonds, light grey) as a sole carbon source, with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. Only *P. pseudoalcaligenes* was grown on biphenyl, and was grown for a total of 48 hours before quantification. 34
- Figure 2-2: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate under various Buffer Conditions. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown in Minimal Salts Media (MSM) containing 5mM succinate as a sole carbon source and buffered with either High Phosphate (black, circles), Low Phosphate (dark grey, triangles), with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. The MOPS buffered media and Low Phosphate media both contained the same concentration of phosphate necessary as a phosphorus source. The pH of all media was nominally adjusted to 7.0..... 38
- Figure 2-3: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate on various Carbon Sources. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown in Minimal Salts Media (MSM) containing 5mM of either Succinate (circles, black), Aspartate (squares, darkest grey), Butyrate (triangles, dark grey), Oxaloacetate (diamonds, light grey) or putrescine (stars, lightest grey) as a sole carbon source, buffered with high phosphate with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. The pH of all media was nominally adjusted to 7.0..... 40
- Figure 2-4: Correlation between bacterial growth and concentrations of chemical species predicted from computational modelling of high phosphate, low phosphate and MOPS buffered media. Data are from OPLS models correlating biofilm and planktonic growth of *P. spp.* in the presence of Al (A, B) or Cu (C) with either all predicted species (A, C)

or just the metal species (B). With Cu, a significant model could not be generated with just the metal species. The VIP (significance) indicates the importance of each chemical species to the model whereas pq represents the correlation with growth. Values presented were taken from the first predicted component. Only species with the highest VIP are shown. (D) The structures of the 5 carbon sources tested for growth: (i) Butyrate, (ii) putrescine, (iii) aspartate, (iv) succinate, (v) oxaloacetate. 43

Figure 2-5: Correlation between bacterial growth and concentrations of chemical species predicted from computational modelling of high phosphate buffered media with succinate, butyrate and benzoic acid as carbon sources. Data are from OPLS models correlating biofilm and planktonic growth of *Pseudomonas* spp. in the presence of Al (A,B) or Cu (C) with either all predicted species (A,C) or just the metal species (B,D). The VIP (significance) indicates the importance of each chemical species to the model whereas pq represents the correlation with growth. Values presented were taken from the first predicted component. Only species with the highest VIP are shown..... 45

Figure 3-1: PCA Scores plot from PCA-Class model of GC-MS metabolomic quantification of samples from cells of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate as the sole carbon source and exposed copper. 71

Figure 3-2: Structure of 2-phosphoglycolic acid 3TMS. 80

Figure 3-3: Gas chromatogram showing peak at RT:11.5787 (RI:1579). Peak is composed of ions 267 (yellow), 372 (red), 357 (blue), 255 (dark red), 239 (dark blue), 225 (cyan), 167 (dark green), 151 (light green). 80

Figure 3-4: Mass spectrum of unknown RT 11.58. 81

Figure 3-5: Structure of 2-hydroxymuconic semialdehyde. 83

Figure 3-6: Gas chromatogram showing peak at RT:11.37 (RI:1557). Peak is composed of ions 286 (blue), 133 (light green), 111 (cyan), 271 (red), 196 (yellow), 243 (dark red), 227 (dark blue). 84

Figure 3-7: Mass spectrum of unknown RT 11.37. 84

Figure 3-8: Growth (A) and culture pH (B) of *Pseudomonas pseudoalcaligenes* KF707 grown over 32 hours in minimal salts medium with either succinate (Sc) or biphenyl (Bp) as the sole carbon source, with either nothing, 3mM Al₂(SO₄)₃ (Al) or 60µM CuSO₄ (Cu). Points denote the mean of 3 biological replicates, error bars indicate SEM. 88

Figure 3-9: Principal component analysis (PCA) scores plot of GC-MS metabolite profiles of cells and spent media from cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl (Bp, purple) or succinate (Sc, teal) as the sole carbon source and exposed to either control (circles), 3mM Al (squares) or 60µM Cu (triangles). Results were normalized, scaled and centred before analysis. 89

Figure 3-10: Hierarchical clustering analysis and heatmap of GC-MS quantified metabolites from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate or biphenyl as the sole carbon source under control, Aluminum or Copper exposure. Data were only log transformed. Gold indicates high concentrations of metabolite while cyan indicates low, black being average (across ALL samples) for that metabolite. Distances between samples and variables were determined using euclidean distance and clusters (both of metabolites and samples) were determined by Ward's method. The color bars at the end of the sample dendrogram indicate the sample type. Color bars at the end of the metabolite dendrogram and beside the metabolite names indicate general metabolic class..... 91

Figure 3-11: Hierarchical clustering analysis and heatmap of GC-MS quantified metabolites from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate or biphenyl as the sole carbon source under control, Aluminum or Copper exposure. Data were normalized, mean-centered and scaled to unit variance to allow comparison of metabolites with varying dynamic ranges and variances. Gold indicates high concentrations of metabolite while cyan indicates low, black being average (across ALL samples) for that metabolite. Distances between samples and variables were determined using euclidean distance and clusters (both of metabolites and samples) were determined by Ward's method. The color bars at the end of the sample dendrogram indicate the sample type. Color bars at the end of the metabolite dendrogram and beside the metabolite names indicate general metabolic class. 92

Figure 3-13: OPLS-DA Scores Plots from pairwise models comparing GC-MS metabolomic quantification of samples from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as the sole carbon source and exposed to either control, aluminum or copper..... 96

Figure 3-14: OPLS-DA Scores Plots from pairwise models comparing GC-MS metabolomic quantification of samples from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate as the sole carbon source and exposed to either control, Aluminum or copper..... 97

Figure 3-15: Shared and unique structures plots showing comparison of changes to intracellular metabolites caused by Al (A,B) and Cu (C,D) in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either succinate or biphenyl as the sole carbon source. Coordinates were determined by the variable influence on projection (VIP) of each metabolite, as obtained from OPLS-DA models comparing control and metal exposed samples for either succinate (y-axes) or biphenyl (x-axes). Metabolites with a $VIP \geq 0.8$ (dashed lines) indicate a significant change occurred in the metal exposed samples, those that were below in both cases were omitted. The association of each metabolite with control or metal exposed samples was determined using $p(\text{corr})$, which indicates the degree of correlation of the metabolite with a sample type. Shapes were assigned that indicate how the metabolite was altered by metal exposure: increased in both biphenyl and succinate (squares), increased in biphenyl but decreased in succinate (diamonds), increased in succinate but decreased in biphenyl (triangle up) and decreased in both (triangle down)..... 98

Figure 3-16: Shared and unique structures plots comparing extracellular metabolic changes of identified metabolites caused by Aluminum (A) and Copper (B) in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either succinate or biphenyl as the sole carbon source. The data plotted here were derived from pairwise OPLS-DA models that compared control to metal exposed samples for each carbon source. Coordinates were determined by the p(corr) of each metabolite from each model, only those with a VIP>0.8 in at least one media type were plotted. Metabolites with a VIP>0.8 in biphenyl models are marked with a circle, in succinate an X. The larger the p(corr) the greater the correlation with metal exposure for that media type. To identify metabolites that were being secreted, data from the OPLS-DA models comparing the control and metal exposed cells samples for each media type were also used. Metabolites were determined to be secreted if the VIP was above 0.8 in the spent media models and either the VIP was below 0.8 in the cells models or the p(corr) was opposite what was observed in the spent media. I.E. if a metabolite was found to change significantly in both the cells and spent media and was correlated with the metal exposed samples in the media and the control in the cells, this metabolite was secreted in response to metal exposure. Conversely, if the opposite was true, i.e. the metabolite was correlated with the control samples in the spent media and the metal exposed cells, this metabolite's secretion decreased in response to metal exposure. The association of each metabolite with control or metal exposed cell samples was determined using p(corr), which indicates the degree of correlation of the metabolite with a sample type. Actual values were not used, only the directionality (i.e. above or below zero) as metabolites with high (>0.8) VIP scores tended to have high p(corr) absolute values. Using these values metabolites were assigned a shape that indicates how secretion of the metabolite was altered by metal exposure: squares were secreted more in both biphenyl and succinate, diamonds were secreted more in biphenyl but less in succinate, upwards pointed triangles were secreted more in succinate but less in biphenyl and downwards pointed triangles were secreted less in both in response to metal exposure. Metabolites were broadly classified according to chemical structure and/or biological function: those from the TCA cycle and energy producing pathways (light blue), other carboxylic acids (orange), small molecules (yellow), phosphate containing (green) (A and C), amino acids (dark blue), nucleotides and related pyridine/pyrimidine compounds (red) and polyamines (purple) (B and D). The top two panels show the effects of Aluminum (A and B) and the bottom two copper (C and D)..... 105

Figure 3-17: Variable Influence on Projection (VIP) of metabolites related to degradation of biphenyl from OPLS-DA models comparing control to metal exposed cells and spent media of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as a sole carbon source. Positive VIP values indicate a positive p(corr) value meaning an increase in metal exposed samples, negative a decrease. Transparent bars indicate a VIP<0.8, not significantly changed. Metabolites categorized as relating to biphenyl degradation are: biphenyl (orange), 2,3-dihydroxybiphenyl (light green), 2,3-dihydroxybenzoic acid (dark green), 3-hydroxybenzoic acid (yellow), salicylic acid (brown), benzene acetic acid (pink), o-toluic acid (red), catechol (cyan), 2-hydroxymuconic semialdehyde (purple), unknown RT: 11.41 (light purple). 110

Figure 3-18: Changes to metabolites derived from biphenyl degradation in cells (C) and spent media (M) of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as the sole carbon source and exposed to Al or Cu. Arrows indicate whether a metabolite was increased (↑) or decreased (↓) in response to each metal. Values were derived from VIP and p(corr) from OPLS-DA models comparing control to metal exposed samples. Empty boxes mean there was no significant change (VIP <0.8). For details, see Figure 3-16. Benzoic acid was detected but could not be accurately quantitated due to elevated concentrations resulting in detector saturation (*). Metabolite background shading denotes whether it is a canonical biphenyl degradation product and was detected (light grey) or not detected (white). Several metabolites that are not involved in biphenyl degradation but are structurally related were also detected (dark grey). 112

Figure 4-1: Maximum likelihood consensus tree showing phylogenetic relationship between Aer homolog protein sequences from select *Pseudomonas* species, rooted using the closest available non-*Pseudomonas* sequence. Tree was generated unrooted, then then rooted to the known outgroup. Numbers at nodes indicate bootstrap support values from 100 replicates. 142

Figure 4-2: Maximum likelihood consensus tree showing phylogenetic relationship between Aer homolog protein sequences from select *Pseudomonas* species. Colours indicate homolog group: Aer.g1 (red), Aer.g2 (yellow), Aer.g3 (cyan), Aer.g4 (purple) and Aer.g5 (blue). Groups were based on inspection of the alignment used to generate the phylogeny and other analyses, see the text for details. Tree was generated unrooted, then the root placed based on a similar tree that was rooted to the closes non-*Pseudomonas* sequence available. Branch lengths indicate number of AA substitutions per site. Numbers at nodes indicate bootstrap support values from 100 replicates. Dotted box indicates original separation of Aer.g1 into two groups. 143

Figure 4-3: Empirical cumulative distribution function of amino acid multi-relief scores for pair-wise comparisons of preliminary Aer homolog groups. Only scores above 0.5 are shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of amino acid positions for each comparison are above this cutoff. E.g. only 10% of amino acid positions in the 1vs2 comparison and even less in the 3Avs3B comparison are above this cutoff. Conversely, there is between 25% and 35% of amino acid positions above the cutoff for all other comparisons. 144

Figure 4-4: Empirical cumulative distribution function of amino acid multi-relief scores for pair-wise comparisons of adjusted Aer homolog groups. Only scores above 0.5 are shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of amino acid positions for each comparison are above this cutoff. All comparisons show that between 25% and 35% of amino acid positions are above the cutoff. 146

Figure 4-5: Amino acid conservation and multi-relief weight scores for all positions of Aer and domain architecture of Aer. Conservation was calculated across the entire alignment of all 144 Aer homolog sequences. Multi-relief scores were calculated for each pair-wise comparison of homolog groups. Both values have been smoothed by taking the

mean of the 3 preceding and following amino acids for each position. Domain architecture was obtained from the conserved domain database. Regions have been colour coded based on known features, and from features derived from the current study..... 147

Figure 4-6: Hierarchically clustered heatmap showing the presence and number of Aer homologs, Aer2 and Ctp in select *Pseudomonas* species. Species were clustered using the Bray-Curtis distance metric and average linkages. The number of homologs that each strain possess is indicated by the box colour: zero (black), one (blue), two (cyan). Pseudogenes are marked by asterisks (*). The homolog of *P. rhizosphearae* is more similar to non-*Pseudomonas* sequences and is marked by a pound (#)..... 149

Figure 4-7: Frequency of occurrence and orientation of genes upstream and downstream from *aer* homologs from select *Pseudomonas* species. Coloured and bolded genes are expected to be co-transcribed based on their orientation and high frequency of occurrence with *aer*. Numbers beside gene functions indicate the frequency that they were found within the 144 sequences examined. Gene lengths are approximate, *aer* is about 1.5kb long..... 151

Figure 4-8: Photographs of energy-taxis swim plates after 24h of growth at 30°C in 50mM pyruvate. Energy-taxis swim plates of *P. pseudoalcaligenes* strains with deletions in *aer-2*, *cttP* and *aer* homologs after 24h growth at 30°C in 50mM pyruvate. 159

Figure 4-9: Normalized energy-taxis growth diameters in 50mM pyruvate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer2*, *cttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey’s Honest Significant Differences test with a confidence value of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present..... 161

Figure 4-10: Normalized energy-taxis growth diameters in 50mM succinate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2*, *cttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey’s Honest Significant Differences test with a confidence value of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present..... 162

Figure 4-11: Mean raw and normalized growth rates of energy-taxis swim-growth-diameters of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2*, *cttP* and *aer* homologs. Growth rates were calculated by dividing the diameter at 48h by the diameter at 24h. Normalized rates were calculated from diameters that were normalized to the wild-type for each experiment at each time point. Stars indicate significant differences

from the corresponding wild-type according to Tukey's Honest Significant Differences test.	163
Figure 4-12: Chemotaxis swim assays of <i>P. pseudoalcaligenes</i> strains with deletions in <i>aer-2</i> , <i>cttP</i> and <i>aer</i> homologs. Strains were grown overnight, concentrated then spotted on minimal salts plates containing 0.3% agar. Either 50mM pyruvate in 1.5% agar or crystals of glucose were placed in the centre of the plates. Photographs were taken after 24h.....	164
Figure 5-1: Chemotaxis swim plates showing taxis of <i>P. pseudoalcaligenes</i> KF707 wild-type and deletion mutants towards biphenyl after 24h (A) and 48h (B).....	184
Figure 5-2: Normalized energy-taxis growth diameters in 50mM pyruvate or succinate strains of <i>P. pseudoalcaligenes</i> KF707 wild-type and quintuple <i>cttP/aer-2/aer.g1/aer.g2/aer.g4</i> mutant complemented with pSEVA342 or pSEVA342_Aer.g1. Bars indicate the average growth diameter, normalized to the wild-type, from 3 experimental replicates. Error bars indicate standard error.....	185
Figure 5-3: Energy-taxis swim plates showing exemplary results of <i>P. pseudoalcaligenes</i> KF707 wild-type and quintuple <i>cttP/aer-2/aer.g1/aer.g2/aer.g4</i> mutant complemented with pSEVA342 or pSEVA342_Aer.g1 in 50mM pyruvate with 1mM IPTG after 48h....	186
Figure A-A-1: Schematic Overview of Metabolomics Workflow Metal Toxicity Experiments. First organisms are grown and exposed to the metal(s). Next metabolism is quenched and samples are collected. Cells are lysed and metabolites are extracted prior to preparation for the analytical platform to be used. Samples are then analyzed by the platform(s) of choice, here only 1H-NMR (proton nuclear magnetic resonance spectroscopy) and GC-MS (gas chromatography mass spectrometry) are shown as they are most common. The data acquired in this fashion must then first be deconvoluted before being interpreted using various multivariate statistical and bioinformatic tools in order to generate conclusions of biological relevance.	214
Figure A-B-1: Flow chart showing possibilities of metabolomics secondary analysis. Beginning with a list of metabolites, and also in some cases associated with relative abundances or a comparable metric the data can be analyzed two different ways which may include intermediary steps. Also needed is a biochemical database to be used in annotating the biological (and/or chemical function) of the listed metabolites. The list of metabolites plus abundances can undergo statistical analysis in order to pre-screen for metabolites having significant differences between sample classes, or the abundances can be used to calculate how differently they're expressed between each sample class (difference metric) which is then used to rank the list from most different to most similar. Significances can be used for both metabolite mapping and enrichment analysis. Metabolite mapping is the visual attribution of specified metabolites within known, pre-defined metabolic pathways which can include further information like significance and abundances as node attributes such as size and colour. Enrichment analysis is a statistical calculation that uses biological annotation to attempt to discern out the input metabolites which higher level functional properties (pathways) are being affected. This	

can take the form of searching for particular annotations at the top/bottom of the ranked list or examining whether a particular set is over-represented in the significant list..... 253

Figure A-C-1: Targeted Profiling in Chenomx™. After phase and baseline correction, NMR spectra can be viewed in the Chenomx™ Profiler (28). Initially, just the original spectrum is visible (A), as a black line. Using the provided library of reference compounds, peaks at the appropriate chemical shift can be added (B), in blue. An additive spectrum of all profiled compound's peaks can then be viewed, as a red line. A subtractive spectrum, as a green line, is also provided which removes the profiled spectrum from the original to ensure that peaks have been properly identified and quantified. Profiled compounds are then matched up to peaks in the original spectrum (C) and manipulated until they match in amplitude to provide quantitative information (D). 301

Figure-A-C-2: SUS Plot of VIP values obtained from models comparing *P. pseudoalcaligenes* KF707 wild-type with and without tellurite and the T5 hyperresistant mutant with and without tellurite. The x-axis shows the metabolites that are important in the wild-type; the y-axis shows those that are important in the T5. Metabolites with a VIP > 1 were considered significant in distinguishing the control from tellurite exposed samples in each strain. Metabolites in the bottom right were important for the wild-type, in the top left for T5. Those in the top right were important in both and those the bottom left neither. (Figure adopted from (22), with permission) 306

Figure-A-C-3: Partial least squares discriminate analysis (PLS-DA) 3D scores plot from a model comparing *P. fluorescens* wild-type (CHA0), $\Delta gacS$ (CHA19) and the colony morphology phenotypic variants wrinkly spreader (WS) and small colony variant (SCV) obtained from the parental $\Delta gacS$ strain. Each data point represents metabolites extracted from a single sample. Each axis represents a component of variation that separates the four classes based on the concentrations of 32 metabolite concentrations. The clustering of all samples in each class indicates the similarity of the samples and the separation along each axis indicates the difference between classes. (Figure adapted from (23), with permission) 310

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
EPS	Extracellular Polymeric Substance
DNA	Deoxyribonucleic acid
LB	Lysogeny Broth
PCB	Polychlorinated Biphenyl
DMSO	Dimethyl Sulfoxide
MSM	Minimal Salts Medium
MOPS	3-(N-morpholino) propanesulfonic acid
MBEC™	Minimum Biofilm Eradication Constant
OPLS	Orthogonal Partial Least Squares
PLS	Partial Least Squares
DA	Discriminant Analysis
VIP	Variable Influence on Projection
MIC	Minimum Inhibitory Concentration
SEM	Standard Error of the Mean
NIST	National Institute of Standards and Technology
GMD	Golm Metabolome Database
TMS	Trimethylsilyl
GC-MS	Gas Chromatography Mass Spectrometry
PCA	Principal Component Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
CV-ANOVA	Cross Validation Analysis of Variance

ID	Inner Diameter
RPM	Rotations Per Minute
MSTFA	N-methyl-N-trimethylsilyltrifluoroacetamide
PQN	Probabilistic Quotient Normalization
RT	Retention Time
SUS	Shared and Unique Structures
ABC	ATP Binding Casette
HCL	Hierarchical Clustering Analysis
G6P	Glucose-6-phosphate
LPS	Lipopolysaccharide
R5P	Ribose-5-phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate
PPP	Pentose Phosphate Pathway
ATP	Adenosine Triphosphate
NADH	Nicotinamide adenine dinucleotide
ROS	Reactive oxygen species
2,4-D	2,4-dichlorophenoxyacetic acid
MCP	Methyl-accepting Chemotaxis Protein
FAD	Flavin Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
BLAST	Basic Local Alignment Search Tool
JTT	Jones-Taylor-Thornton
SHMR	Sequence Harmony Multi Relief

ECDF	Empirical Cumulative Distribution Function
AA	Amino Acid
CDD	Conserved Domain Database
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
Gm	Gentamycin
Cm	Chloramphenicol
PCR	Polymerase Chain Reaction
PAS	Pern Arnt Sin
GMP	Guanidine Monophosphate
MUSCLE	Multiple Alignment by Log-Expectation
COBALT	Constraint-Based Alignment Tool
ML	Maximum Likelihood
TBARS	Thiobarbutic acid reactive substance
PCB	Polychlorinated Biphenyl
PAH	Polyaromatic Hydrocarbon
1,2-DCA	1,2-Dichloroethane
NDR	NADH quinone oxidoreductase
EPR	Electron Paramagnetic Resonance
TM	Transmembrane
MA	Methyl-Accepting
LBR	Ligand Binding Region
TCA	Tricarboxylic Acid

Epigraph

“This is it: don’t get scared now” – *Kevin McCallister*

Chapter 1: Preface

This chapter introduces background information about various subjects relevant to the entire thesis. First is an introduction to the scope of anthropogenic pollution and the need for bioremediation. Next is an introduction into the problem of metals interfering with bioremediation and an in-depth discussion of known mechanisms of metal toxicity in bacteria. A brief review of studies on the main model organism of this thesis, *P. pseudoalcaligenes* KF707, is then provided before briefly explaining chemotaxis in bacteria and the many recent studies that have been performed using *Pseudomonas* species. Finally, there is an overview of the main goals and scientific questions addressed in this thesis.

Chapter One: **Introduction**

1.1 Remediation of Polluted Environments

The effects of human beings on the environment have become so dramatic that geologists are beginning to refer to the current epoch as the ‘anthropocene’ (1). While pre-industrial societies caused the extinction of many species, industrialization has resulted in a plethora of global-scale environmental problems including climate change, ocean acidification, eutrophication of huge areas such as the dead-zone in the Gulf of Mexico and contamination of the atmosphere, soils and hydrosphere with toxic xenobiotic compounds. Xenobiotics (compounds with solely anthropogenic sources) are not the only problem; though naturally occurring, heavy metals are introduced into the environment at high concentrations due to human activities (2). These compounds are produced for a variety of uses such as such as electronics, solvents, pharmaceuticals and multi-purpose industrial liquids like polychlorinated biphenyls (PCBs) (3) but through accidental spillage and improper disposal they have become pollutants and contaminate large quantities of soil and water (3). The problem with this contamination is that these compounds are often chronically or even acutely toxic to organisms in the environment and can accumulate up the food chain (4).

Pollutants can be divided into two broad classes: organic compounds made mostly of carbon and inorganic metal elements. Organic pollutants, such as petroleum products, can be chemically or biologically degraded into their constituent molecules. Their complete transformation into carbon dioxide and water is deemed mineralization. Conversely, metals such as aluminium and copper cannot be broken down, only their chemical form altered, say from solid salts to soluble ions. Organic and inorganic pollutants are potentially toxic. For example, PCBs have been associated with a variety of detrimental health effects in humans including disruption

of the nervous, endocrine and reproductive systems as well as being obesogenic and also carcinogenic in animals (5–8). Metals are also toxic to plants and animals but their toxicity towards bacteria is of particular interest as they can biochemically process hazardous metals into forms that are much less toxic to other organisms (9, 10).

Due to their toxicity, it is of great interest to public and environmental safety that steps be taken to clean up pollutants from the environment. Physical and chemical methods for cleaning up contaminated soils and marine sediments are generally expensive and potentially dangerous and ineffective (11). Bioremediation, the process of using living organisms to remediate pollution, is thus an important alternative as it is much cheaper, can be performed *in situ* and results in the complete mineralization of organic pollutants (12). Bacteria are particularly useful for bioremediation applications due to their broad metabolic capabilities (13) and their resistance to metal toxicity (14). Unfortunately, this is still not enough in many situations as sites co-contaminated with metals and organic pollutants are notoriously difficult to remediate (15). This is quite alarming as 40% of polluted sites are contaminated by metals and organic pollutants (U.S. EPA, www.epa.gov/superfund). In order to develop bioremediation techniques that can be applied to these co-contaminated sites an in-depth understanding of how metal toxicity affects bacteria is needed.

1.2 Inhibition of Organic Pollutant Degradation by Metals

Research in this area has focused on characterizing situations where a metal inhibited the microbial metabolism of an organic pollutant, but very few studies have investigated the underlying physiology and biochemistry of these interactions. This is likely due to most of these studies being performed by engineers and biotechnologists focused on developing functional applications for remediating the co-contaminated sites. Reviews on the subject focus on the

physicochemical parameters (pH, temperature, oxygen concentration) of the contaminated site and the metal's bioavailability (15–17). This concept of bioavailability of metals is quite important to metal toxicity studies as it affects laboratory and field experiments alike and can account for the wide ranges of reported inhibitory concentrations of metals (15). The other focus of these reviews is providing recommendations on what additives to use to improve bioremediation. This will not be discussed here as this thesis does not deal directly with practical applications.

A wide range of different metals have been found to inhibit the degradation of many pollutants by various bacteria (a selection of representative studies summarized in Table 1). Pollutants include phenol, dioxanes, chlorinated alkanes, petroleum products and polycyclic aromatic hydrocarbons (PAHs), commonly tested metals include aluminium, arsenic, cadmium, chromium, cobalt, copper, mercury, nickel, lead, and zinc, and bacteria include *Acintebacter*, *Bacillus*, *Burkholderia*, *Corynebacterium*, *Pseudonocardia*, *Pseudomonas*, and *Stenotrophomonas* as well as mixed communities. These results from these studies can generally be summarized in a few different fashions: (1) a certain concentration of some metal(s) reduced the efficacy of removal of an organic pollutant, (2) the metal increased the time needed to remove the organic pollutant, (3) manipulating the culture conditions (addition of extra nutrient, oxygen, pH adjustment etc.) reduced the harmful effects of the metal, or (4) the tested metal did not inhibit degradation of the tested organic pollutant. While potentially useful from an applications perspective, these studies provide no information with regards to why or how the metals inhibited the organic pollutant degradation. There is a dearth of fundamental investigations into the physiological effects of combining toxic metal stress with organic pollutant metabolism. However, there have been studies that have demonstrated the exact biochemical mechanisms of metal toxicity in bacteria under laboratory conditions.

Table 1-1: Summary of select representative studies wherein the effects of metals on organic pollutant degradation were investigated.

Pollutant	Metal(s)	Organism(s)	Comment	Ref
Chlorpyrifos	Cd (II)	<i>Stenotrophomonas</i> sp.	Adding gene for surface Cd-binding protein improved degradation	(18)
Phenol	Cr (VI)	<i>Acinetobacter</i> and <i>Bacillus</i>	Consortium can remove phenol and Cr from wastewater	(19)
Phenol	Cr (VI)	<i>Acinetobacter</i> sp.	Biostimulation with nutrients enabled degradation and Cr removal	(20)
TCE	Ni (II)	<i>Pseudomonas putida</i>	<i>P. putida</i> improves phytoremediation of co-contaminated soil	(21)
1,2-dichloroethane (1,2-DCA)	As (III), Cd (II), Hg (II), Pb (II)	Uncharacterized community	Higher concentrations of metals increase time to degrade 1,2-DCA	(22)
1,2-dichloroethane	As (III), Cd (II), Hg (II), Pb (II)	Characterized community + <i>Delftia</i> sp., <i>Pseudomonas</i> sp., <i>Cupriavidus</i> sp., <i>Stenotrophomonas</i> sp., <i>Xanthobacter autotrophicus</i>	Bioaugmentation of communities with highly metal resistant isolates and biostimulation with glucose decreased the inhibition caused by metals	(23)
Phenanthrene	Cd (II)	Uncharacterized consortia	Metal resistant consortium can degrade better, controlling media allows amount of bioavailable metal to be known	(24)
Phenanthrene	Al (III), Fe (III), Zn (II) +Cu (II)	Uncharacterized consortia	Al and Zn+Cu inhibited but Fe increased degradation	(25)
PAHs, chlorinated hydrocarbons	Co (II), Ni (II),	<i>Pseudomonas aeruginosa</i> <i>Burkholderia cepacia</i> <i>Corynebacterium kutscheri</i> <i>Rhodococcus</i> sp.	All strains could not degrade all substrates, but metals did not inhibit degradation likely due to media composition sequestering metals.	(26)
1,4-dioxane	Cd (II), Cu (II), Ni (II), Zn (II)	<i>Pseudonocardia dioxanivorans</i>	Cd, Cu and Ni inhibited degradation, ligands prevented inhibition	(27)

1.3 Metal Toxicity in Bacteria

Metals exert their toxicity on bacterial cells through a variety of biochemical mechanisms that target many different cellular entities. Individual metal elements can partake in several of these reactions, making the specific toxicity mechanisms for each metal difficult to define. Many metals can also react with oxygen and the by-products of aerobic respiration to cause oxidative stress. This makes it difficult to discern the exact mechanisms of toxicity, as well as determine which is the most important for inhibiting cell growth and causing death. Studies focusing on the mechanisms and targets of metal toxicity have used techniques ranging from physical to chemical to biological methods. This application of multiple techniques has enabled the detailed understanding of many specific mechanisms.

1.3.1 Biochemical Mechanisms of Metal Toxicity

Metals can be toxic to bacteria due to direct or indirect modes of action. Direct mechanisms involve the metal ion damaging a cellular target whereas indirect mechanisms cause toxicity by inducing the production of actively damaging chemical species or inhibiting the metabolism of an essential nutrient. As a key necessary metal, this mechanism of toxicity often involves iron in some way. Toxicity mechanisms can be further broken down into specific and non-specific processes. Specific mechanisms involve the disruption of some unique or small subset of targets while non-specific mechanisms damage a variety of cellular entities. Since individual metals can induce toxicity through multiple mechanisms, discerning which is most important for preventing growth and causing death is difficult and likely dependent on a number of extraneous factors. Iron containing enzymes are vulnerable to substitution (a direct mechanism), causing the release of iron which can lead to indirect toxicity. In aerobic systems, toxic reactive oxygen species (ROS) are

produced as an incidental by-product of normal respiration (28). Hydrogen peroxide and superoxide react with iron (II) and iron (III) to make the highly reactive hydroxyl radical (29). Many other metals can catalyze reactions that produce ROS which damage any available cellular target via non-specific oxidation. Metals that do not catalyze such reactions, but can displace iron can cause oxidative stress through a doubly indirect mechanism. Regardless of whether they are produced directly by the toxic metal or by endogenous iron liberated from cellular proteins, ROS production is the main mechanism of non-specific, indirect toxicity. Indirect specific toxicity also involves iron as these mechanisms generally occur through preventing proper iron usage within the cell.

1.3.2 Direct, Specific Mechanisms

Several important proteins are specifically damaged by metal stress. Specific mechanisms must be interpreted under the context of the model organism used for the experiment; it is quite likely that toxicity mechanisms in other organisms will differ from the model bacterium *Escherichia coli*. As the classically preferred substrate of *E. coli* (30), glucose is often used for growing cultures. 1,6-Fructose biphosphate aldolase (FbaA) catalyzes the key cleavage for producing two triose sugars and is inhibited by nickel, causing cell death when grown on glucose (31). Site directed mutants that changed the binding site of a non-catalytic zinc (II) rescued the phenotype, demonstrating that nickel specifically damages FbaA by replacing this metal. Sodium translocating NADH Quinone Oxidoreductase (Na⁺ NQR) is used by *Vibrio* species for generating a transmembrane sodium gradient which can be used for multiple purposes, similarly to a proton gradient (32). While this gradient is not necessary for growth, silver ions specifically disrupt the activity of NqrF in *V. harveyi* by interacting with a single cysteine residue (33). This contrasts with the non-specific mechanism of silver toxicity in *V. harveyi* where it causes membrane

depolarization through proton leakage (34). Metals can also act upon particular prosthetic groups, making their toxicity mechanism specific to proteins which contain the particular cofactor. Exposed iron sulphur clusters, particularly [4Fe-4S] clusters of dehydratases are vulnerable to several metals (35–37). Tellurite (TeO_3^{2-}), silver (I), cadmium (II), copper (I), mercury (II), and zinc (II) were all found to damage Fe-S containing proteins. Interestingly, the metals listed here show clear chemical differences, TeO_3^{2-} being an oxyanion, silver (I) and copper (I) being monovalent and the rest are divalent cations. Only TeO_3^{2-} damaged Fe-S clusters in an oxygen dependent manner, indicating that its toxicity was actually indirect as it was ROS produced by the TeO_3^{2-} that were disrupting the Fe-S clusters (36). The inactivating capacity of the other metals was confirmed under anaerobic conditions. For the divalent cations, the mechanism of toxicity is through the displacement of iron (II) ions in the cluster through ionic mimicry (35). It is less clear how silver (I) and copper (I) damage Fe-S clusters as none of the expected electron paramagnetic resonance (EPR) visible decomposition products were observed and only copper (I) mediated disruption could be repaired *in vitro*, which could also be achieved with the divalent cations. For all cases, only dehydratases such as isopropylmalate isomerase and fumarase which have exposed Fe-S clusters were inactivated whereas enzymes with buried clusters like glutamate synthetase were not (37). The specificity of this mechanism may also involve an interplay between the organism and the metal, as aluminium (III) toxicity in *Pseudomonas fluorescens* is caused by damaging the Fe-S cluster of aconitase (38). Non Fe-S containing dehydratases are also at risk; lead (II) has been demonstrated to specifically inhibit the delta-aminolevulinate dehydratase of *Pseudomonas putida* by displacing the active site zinc (39, 40). While these experiments have demonstrated a clear specific mechanism of metal toxicity, the effective specificity of these metals will depend upon the growth conditions as the necessity of the vulnerable enzymes can be obviated

by the addition of branched chain amino acids or anaerobic growth (37). The same is true for nickel toxicity as it was not active when cells were grown on substrates that don't require glycolysis (31).

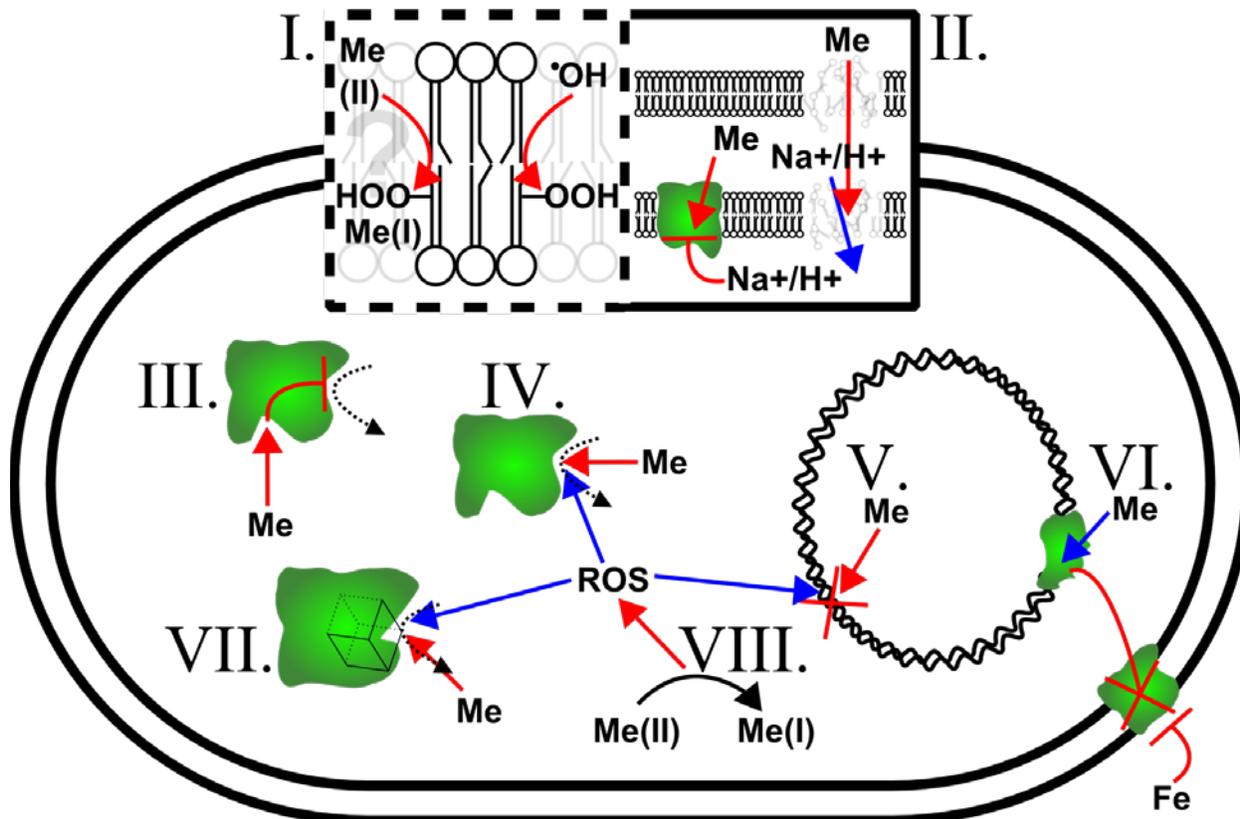


Figure 1-1: Summary of mechanisms and cellular targets of metal toxicity in bacteria. Me: a generic metal ion, see text for specific metals associated with each process. I. Lipid peroxidation either directly by the metal or by ROS produced by the metal. These are putative and disputed mechanisms, respectively, see the text for details. II. Membrane disruption resulting in loss of proton or sodium gradient by leakage. Inhibition of membrane protein stopping generation of a gradient. III. Protein dysfunction by allosteric inhibition or structural damage. IV. Protein dysfunction by replacement of catalytic metal or binding at catalytic site. V. DNA breakage by reaction with ROS. VI. Prevention of iron import by repressing necessary genes. VII. Disruption of exposed Fe-S clusters. VIII. Catalytic generation of ROS, which can also damage Fe-S clusters, proteins and DNA.

1.3.3 Direct, Non-Specific Mechanisms

Bacterial membranes and proteins are vulnerable to toxic effects from metals. As mentioned above, silver (I) can interfere with oxidative phosphorylation by causing proton leakage across the membrane (34). *E. coli* and *Staphylococcus aureus* exposed to silver nanoparticles and

copper surfaces showed signs of damaged cell membranes (41–43). These results were obtained from microscopy and the thiobarbituric acid reactive substance (TBARS) assay, which measures lipid peroxidation (44). Unsaturated lipids cannot propagate the radical chemistry necessary for lipid peroxidation to occur so it is only expected to be possible in the membranes of photosynthetic bacteria (28). The above chemical reaction information along with the inability to discern the cause and effect relationship between membrane disruption and cell death, and the possibility that the TBARS assay was measuring artifactual compounds has made other researchers question how metals damage membranes (14). The preponderance of reports on bacterial membranes being damaged by copper surfaces (41, 43, 45) and silver nanoparticles (42, 46, 47) suggests that some mechanism of toxicity is indeed occurring. Hong *et. al.* in particular addressed the possible shortcomings of the TBARS assay and concluded that despite the monounsaturated nature of bacterial membranes, ROS mediated lipid peroxidation is indeed possible and occurs upon metal exposure (41). The possibility that metals are directly initiating the peroxidation has yet to be addressed. As discussed above, many metals are semi-specific in their toxicity mechanisms, damaging exposed Fe-S clusters. While the aforementioned metals were demonstrated to damage a small number of proteins, nickel(II) has been demonstrated to directly replace key metals in a wide range of iron, zinc and copper containing proteins as well as inhibiting non-metalloproteins, possibly through active-site blocking or allosteric inhibition (48). Some of the enzymes damaged by nickel are particularly important to environmental pollution as nickel (II) was found to inhibit two catechol cleaving dioxygenases, which are necessary for aromatic degradation (49, 50). Other metals can inhibit such proteins, however not all are equally susceptible (51). Many proteins have been characterized as being inactivated by nickel, mostly due to its replacement for iron when crystallizing for structure determination as this prevents iron redox reactions from occurring while

maintaining protein structure (48). It is likely that other divalent metals will undergo similar activity inhibiting reactions *in vivo*, however their occurrence is obfuscated by more damaging mechanisms. Many proteins rely on sulphur atoms for correct functioning, e.g. correctly placed disulphide bridges. Cadmium (II) has a strong affinity for sulphur and has been demonstrated by transcriptomics to exert its toxicity by interfering with multiple sulphur mediated protein functions including Fe-S cluster assembly and disulfide bond repair (52). These mechanisms of toxicity were inferred based on changes of transcript profiles though so a conclusive mechanistic explanation remains to be had.

1.3.4 Indirect, Specific Mechanisms

Certain metals can interact with specific cellular targets, preventing their correct functioning and thereby causing indirect stress. This category of mechanisms could equally be called ‘disturbance of iron homeostasis’ as their toxicity is mediated by limiting uptake or use of iron. Gallium (III) and cobalt (II) interfere with iron assimilation through interaction with the uptake regulators PvdS in *P. aeruginosa* and Fur in *E. coli* (53, 54). Through an unknown mechanism gallium (III) represses *pvdS* expression which regulates the use of pyoverdine as a siderophore. This forces the cells to use other iron uptake mechanisms which do not selectively exclude gallium (III) like pyoverdine does. The end result is the replacement of redox inactive gallium (III) in enough iron-containing proteins to cause cell death. Normally when iron (II) is bound, the well-conserved regulator Fur represses the expression of many iron uptake genes (55). In low iron conditions the unbound Fur will allow gene expression. As an iron mimetic, cobalt (II) can bind Fur causing the repression of iron uptake gene expression even when iron levels are low (54). Copper (II) interferes with the biosynthesis of heme in *Neisseria gonorrhoeae* thereby amplifying oxidative stress as heme is a key cofactor in catalase and peroxidase (56). Without

functional versions of these enzymes hydrogen peroxide will cause overwhelming oxidation of DNA, proteins and lipids resulting in rapid cell death. Due to the widespread conservation of Fur, and also of heme biosynthetic pathways these mechanisms of metal toxicity could be expected to be found in most aerobic bacteria.

1.3.5 Indirect, Non-Specific Mechanisms

Metals can exert their toxicity by catalyzing the formation of Reactive Oxygen Species (ROS) which perpetrate the actual damaging activities. The best characterized reaction that occurs *in vivo* is the production of the hydroxyl radical OH^\cdot by the reaction of iron (II) with hydrogen peroxide (H_2O_2), which is called the Fenton reaction (29). Iron (II) is then regenerated by iron (III) reacting with superoxide $\text{O}_2^{\cdot-}$ allowing for further production of hydroxyl radicals. Due to the instability of the unpaired electron, OH^\cdot will react quickly with any available chemical entity including DNA, proteins and lipids, producing further radicals which results in a cascading chain reaction of oxidation (57). While iron is an essential element for normal growth, because of its ability to catalyze such damaging reactions it is almost all maintained in some kind of bound form and influx/efflux is carefully modulated to prevent excess amounts of free iron (58). Other metals can also catalyze ROS producing reactions or mediate the release of iron (II) thereby inducing Fenton chemistry, making it difficult to discern the specific mechanisms of toxicity of individual metals. Cobalt (II) for example was reported to cause toxicity by interfering with Fe-S clusters, however this was only in aerobic situations (54), under anaerobic conditions it did not (35). Instead it was likely substituting iron (II) in other proteins, causing free iron to be released eventually causing widespread oxidative stress that damaged the Fe-S clusters. Normally, superoxide and hydrogen peroxide are produced as unpreventable by-products of aerobic respiration through accidental reactions of O_2 with enzymes meant to be transferring electrons to other substrates (28).

Tellurite (TeO_3^{2-}) has been postulated to generate superoxide as cells exposed to it upregulate many oxidative stress response genes and superoxide can be produced *in vitro* by tellurite reduction (59). Other redox active metals (copper (II), cobalt (II), chromium (III)) can catalyze Fenton-like chemistry *in vitro* but determining whether the same reactions are occurring *in vivo* is difficult (60). These reactions require divalent metals to be reduced into the monovalent form by oxidizing some other compound within the cell (28). Once reduced they can react comparably to iron (II) with hydrogen peroxide to produce hydroxyl radicals. Such a mechanism of toxicity has been demonstrated for copper (II), but interestingly this reaction only occurs in the periplasm (61). This may explain why copper (II) exposed bacteria often show damaged membranes as the damaging metal is maintained in this region by efflux pumps protecting the interior of the cell (62).

Oxidative stress can damage many cellular targets. Superoxide can react with exposed Fe-S clusters, removing iron (II) and producing hydrogen peroxide (28). This makes proteins containing this cofactor particularly vulnerable as they are subject to both indirect and direct toxicity. Furthermore, this reaction causes the release of iron (II) into the cytoplasm freeing it to propagate oxidative stress. Non-specific protein inactivation caused by carbonylation and oxidation of sulphur containing amino acids (methionine, cysteine) which can cause incorrect or disrupted cross-linking have also been associated with oxidative stress (28). DNA can also be a target of ROS, which can either cause cell death or an increased mutation rate (28, 63). As it is negatively charged any free iron (II) in the cell can interact with DNA but still undergo Fenton chemistry (28). This results in the release of hydroxyl radicals at close proximity to DNA. When they interact the free radical can then either donate or accept a single electron to either the nucleotide or ribose moiety resulting in a variety of possible modifications (28). Non ROS-mediated DNA damage has also been suggested for other metals including nickel (II), lead (II) and

bismuth (II) though the mechanism in this case has been proposed as interference with normal DNA repair (48, 63). Oxidative stress can also be amplified by metals depleting the compounds that normally maintain the cytoplasm as a reducing environment. Glutathione (a tripeptide) is an important compound for this purpose. As mentioned above, cadmium (II) has a high affinity for sulphur allowing it to easily react with any available sulphur groups. Deleting the genes necessary for its synthesis made *E. coli* cells more susceptible to cadmium (II), copper (II) and zinc(II), indicating that metals can exert toxicity by depleting cellular antioxidants, indirectly affecting protein function (64). Another study found links between glutathione-based oxidative stress resistance mutants and silver (I), cobalt (II), and surprisingly, tellurite (65). This association of tellurite with multiple pathways of toxicity may explain its high toxicity to bacteria.

1.4 *P. pseudoalcaligenes* KF707: Studies on Biphenyl Metabolism and Metal toxicity

P. pseudoalcaligenes KF707 was originally isolated and studied for its ability to degrade biphenyl and PCBs (66). For the past 30 years it has been one of the two main model organisms for PCB degradation, the other being *Burkholderia xenovorans* LB400 (67). The complete biphenyl degradation pathway was thus characterized in these organisms as well as how the differential chlorination of PCB congeners affects the process (68, 69). *B. xenovorans* LB400 is able to degrade a wider array of congeners than *P. pseudoalcaligenes* KF707, but is inhibited by the chlorobenzoates it produces (67). *P. pseudoalcaligenes* thus became a better model for understanding stress tolerance during xenobiotic degradation as it is able to degrade fluorobiphenyl (70) and can also tolerate the toxic metalloid Tellurium at concentrations 100x higher than most bacteria (71, 72). This tolerance was first observed when it was used to track the population of KF707 cells in PCB-contaminated microcosms containing tellurite-sensitive communities (73).

These initial works led to the use of *P. pseudoalcaligenes* KF707 to investigate the mechanisms of tellurite toxicity and subsequently other metals.

Tellurite (TeO_3^{2-}) is an oxyanion of tellurium with high toxicity, though some bacteria such as *P. pseudoalcaligenes* KF707 are naturally more tolerant than others (71). It was expected that tellurite toxicity involved reactive oxygen species (ROS) but it was unclear how KF707 resisted these effects so experiments were undertaken to determine the cross-protectivity of tellurite and other ROS-inducing toxins (71). This study found that tellurite sensitized cells to ROS inducers, but that ROS-mediated induction of protective enzymes, such as super oxide dismutase, increased tellurite resistance. This indicated that tellurite toxicity involves ROS but does not induce protective pathways in KF707.

Surface-attached assemblages of bacterial cells living in a layer of self-secreted extracellular polymeric substance (EPS), called biofilms, are more resistant to stressors such as antibiotics and toxic metals than their free-swimming planktonic counterparts (74). *P. pseudoalcaligenes* KF707 was thus expected to have increased metal, metalloid and antibiotic tolerance when grown as a biofilm but surprisingly only had increased tolerance to antibiotics (75). Interestingly though, biofilms did reduce tellurite and selenite to their elemental forms as their planktonic counterparts did, indicating that biofilm though resistance was equivalent, in biofilms it involved a different mechanism. To better understand how the planktonic cells were coping with tellurite toxicity, a hyper resistant mutant was compared to the wild-type using metabolomics (76). Metabolomics is a technique whereby as many as possible of the small, low molecular weight compounds that are manipulated by an organism (metabolites) are quantified in order to obtain a biochemical snapshot of their physiology (for an in-depth introduction to metabolomics see Appendix A). By quantifying intracellular metabolites using nuclear magnetic resonance (NMR)

spectroscopy it was determined that the tellurite hyper-resistant mutant had altered its metabolism to survive the oxidative stress induced by tellurite.

These two findings: that biofilms and planktonic cultures differed in their resistance mechanisms to tellurite, and that metabolomics could provide insight into metal resistance mechanisms, led to my undergraduate project. This study combined NMR with gas-chromatography mass-spectrometry (GC-MS) metabolomics to determine that *P. fluorescens* biofilms responded to copper toxicity in a more concerted fashion to withstand the stress whereas planktonic cultures did not show an organized metabolic response (77). This work was influential towards my PhD research, though there were two more studies that also contributed.

P. pseudoalcaligenes KF707 degrades PCBs and is also highly resistant to tellurite, but it was unknown how metal toxicity affected its growth on PCBs. This was investigated by determining the tolerance of KF707 to many different metals, grown both as a biofilm and planktonically on either succinate or biphenyl (+/- PCBs) (78). This study provided many interesting findings: KF707 could form biofilms using biphenyl as the sole carbon source; PCBs did not inhibit growth or influence metal toxicity but growth on biphenyl as the carbon source could reduce tolerance; certain metals inhibited growth on biphenyl more than others; KF707 was chemotactic towards biphenyl but was inhibited by some metals. This work was followed up by a study that demonstrated that a *cheA* mutant of KF707 was not only deficient in chemotaxis as expected, but also had altered metabolism indicating a connection between metabolism and chemotaxis (79). Together these studies made a connection between biphenyl metabolism and chemotaxis with metal toxicity, driving the research questions I asked during my PhD and which are presented in this thesis. This will be elaborated upon below, after a brief introduction of chemotaxis.

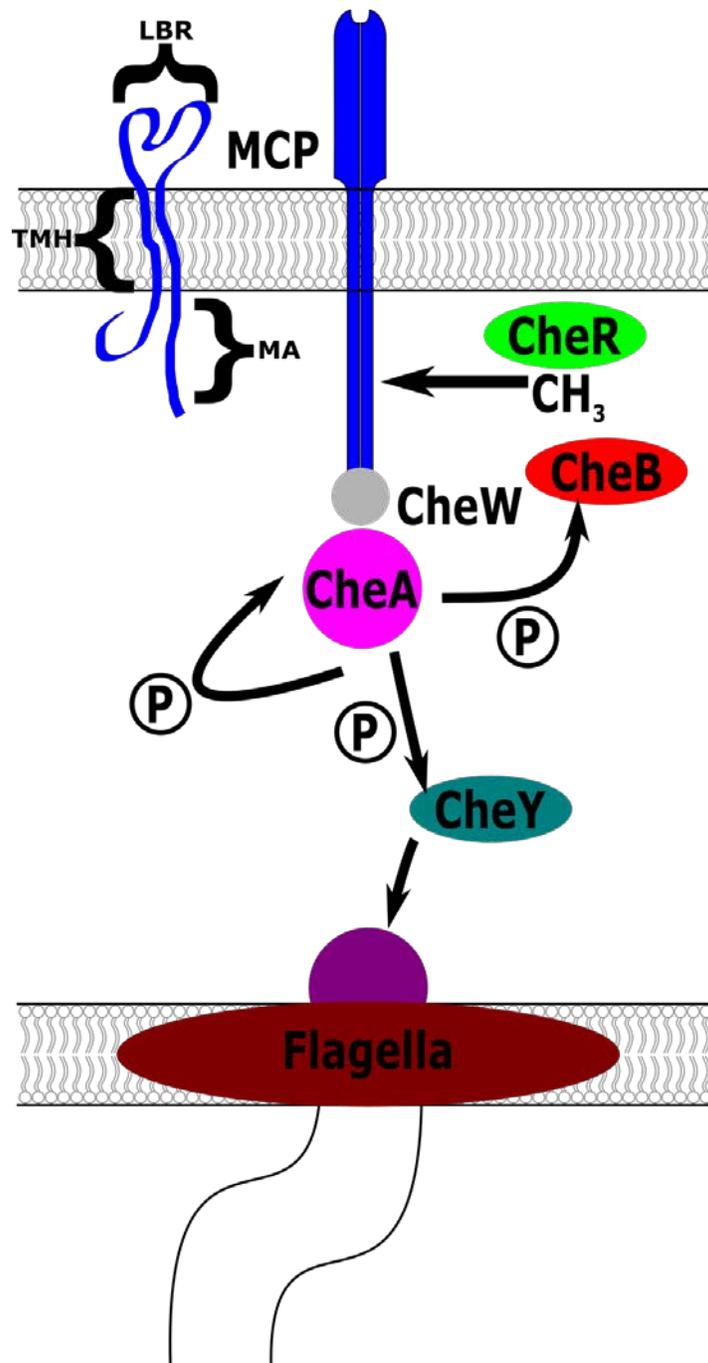


Figure 1-2: Schematic Representation of Chemotaxis Proteins and Methyl-Accepting Chemotaxis Protein Domain Architecture. Canonical MCPs consist of a ligand-binding region (LBR) flanked by two transmembrane helices (TMH) and finishing with a C-terminal methyl-accepting domain which also includes the CheW signalling interface region. When the cognate ligands binds to the LBR, the signal is transduced through CheW to CheA. CheA ceases autophosphorylating and instead phosphorylates CheY. CheY-P diffuses across the cell and interacts with the flagellar complex to change the rotational direction of the flagellum. CheA also phosphorylates CheB, which then removes the methyl groups that are added to the MCP by CheR.

1.5 Chemotaxis

Bacteria can swim along chemical gradients by sensing the concentrations of various compounds and using this information to control the rotation of their flagella, a process called chemotaxis (80). In *Pseudomonas*, like other bacteria, chemotaxis is controlled by Che proteins interacting with the flagellum (Figure 1-2) (81). When spun clockwise, the flagellum causes the bacterial cell to turn in random directions (tumble) but when spun counter-clockwise it moves the bacterium in the defined direction (run). By adjusting the frequency of tumbles and runs in response to attractant and repellent gradients the cell is able to move to more optimal environments. CheY, in its phosphorylated form, interacts with the flagellar machinery to decrease the frequency of tumbles. CheA is the kinase that phosphorylates CheY in response to the binding of an attractant to receptors called methyl-accepting chemotaxis proteins (MCPs). CheW is an adaptor protein that connects CheA to the MCP. The sensitivity of the response is controlled by CheB, CheR and CheZ. CheZ dephosphorylates CheY to ensure that continued action of CheA is needed to produce CheY-P. CheR is a methyltransferase that methylates MCPs. This methylation allows for 'memory' of the response to a particular attractant. As an MCP becomes more methylated, it increases the tendency of CheA to autophosphorylate rather than phosphorylate CheY, thus requiring ever higher concentrations of the attractant to produce the same levels of CheY-P. When phosphorylated by CheA, CheB-P demethylates the same residues of the MCP as CheR, increasing its sensitivity to its cognate attractant. Finally there is CheV, which in enteric bacteria has been characterized as an alternate adaptor protein, replacing CheW for certain MCPs (82). *Pseudomonas* species possess a homolog to CheV and it is likely to play a similar role as *Pseudomonas* species have far more MCPs than *E. coli* (>25 compared to 4) (Table 1-2).

Table 1-2: Characterized methyl-accepting chemotaxis proteins in *Pseudomonas* and BLAST hits in *P. pseudoalcaligenes* KF707. Reported values from BLAST are coverage/identity. Those in bold indicate likely homologous functions.

MCP	Accession	Species	Ligand	Ref	Best BLAST hit in KF707	Values
CtpH	NP_251251.1	<i>P. aeruginosa</i> PAO1	Phosphate [High]	(83, 84)	ELS27662.1	94/68
CtpL	BAA96489.1	<i>P. aeruginosa</i> PAO1	Phosphate (+PtsS periplasmic binding protein) [low], catechol (adventitious)	(83–85)	ELS26904.1	79/64
Aer-2 (McpB, TlpG)	NP_248866.1	<i>P. aeruginosa</i> PAO1	Oxygen	(86, 87)	ELS25788.1	100/69
Aer	NP_250252.1 NP_744260.1 ABQ79752.1	<i>P. aeruginosa</i> PAO1, <i>P. putida</i> PRS2000, KT2440, F1	FAD (Energy-taxis)	(86, 88–90)	ELS28791.1 ELS27852.1 ELS30001.1	100/80 100/53 97/43
McpP	NP_745005.1	<i>P. putida</i> KT2440	C2, C3 carboxylic acids	(91)	ELS27539.1	100/65
PA2652	NP_251342.1	<i>P. aeruginosa</i> PAO1	Malate	(92)	ELS26728.1	99/75
McpQ	NP_747121.1	<i>P. putida</i> KT2440	Citrate (Mg ²⁺)	(93)	ELS25378.1	100/41
McpR	NP_742484.1	<i>P. putida</i> KT2440	Malate, Fumarate	(94)	ELS29403.1	99/52
McpS	NP_746767.1	<i>P. putida</i> KT2440	Malate, Succinate, Fumarate, Citrate, Isocitrate,	(95)	ELS25378.1	99/39

			Oxaloacetate , Butyrate			
McfQ	ABQ81014.1	<i>P. putida</i> F1	Citrate, fumarate	(94)	ELS25378.1	100/41
McfR	ABQ76512.1	<i>P. putida</i> F1	Succinate, malate, fumarate	(94)	ELS29403.1	99/52
McfS	ABQ80640.1	<i>P. putida</i> F1	Succinate, malate, citrate, fumarate	(94)	ELS25378.1	99/40
Pfl01_0728	ABA72472.1	<i>P. fluorescens</i> Pf0-1	Malate, succinate, fumarate	(96)	None	N/A
Pfl01_3768	ABA75506.1	<i>P. fluorescens</i> Pf0-1	Malate, succinate	(96)	ELS26728.1	100/72
Pfl01_4431 (CtaA)	ABA76168.1	<i>P. fluorescens</i> Pf0-1	A, R, N, D, C, G, H, I, L, K, M, F, P, S, T, W, Y, V	(97)	ELS24941.1	100/74
Pfl01_0124 (CtaB)	ABA71868.1	<i>P. fluorescens</i> Pf0-1	A, R, N, D, C, Q, E, G, H, I, K, M, F, S, T, W, Y, V	(97)	ELS24941.1	99/65
Pfl01_0354 (CtaC)	ABA72098.1	<i>P. fluorescens</i> Pf0-1	R, C, G, M, T	(97)	ELS24941.1	100/67
PscA	WP_ 003382247.1	<i>P. syringae</i> pv. <i>Actinidiae</i> NZ-V13	D, E	(98)	ELS24941.1	99/63
PscB	WP_ 005738888.1	<i>P. syringae</i> pv. <i>Actinidiae</i> NZ-V13	A, N, Q, I, L, M, F, S, W, homoserine	(98)	ELS24941.1	100/63
PscC	WP_ 017684350.1	<i>P. syringae</i> pv. <i>Actinidiae</i> NZ-V13	I, P, γ - aminobutyric acid	(98)	ELS29386.1	99/62

PctA	NP_252999.1	<i>P. aeruginosa</i> PAO1	A, R, N, C, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V	(99)	ELS24941.1 ELS26892.1	100/76 98/32
PctB	NP_253000.1	<i>P. aeruginosa</i> PAO1	A, R, Q, M, K	(100)	ELS24941.1	100/75
PctC	NP_252997.1	<i>P. aeruginosa</i> PAO1	H, P, γ - aminobutyric acid	(100)	ELS24941.1	100/70
McpG	NP_743530.1	<i>P. putida</i> KT2440	γ - aminobutyric acid	(101)	ELS24941.1 ELS26892.1	100/72 98/32
McpC	ABQ76791.1	<i>P. putida</i> F1	Cytosine, nicotinic acid	(102, 103)	ELS29386.1	99/64
McpH	NP_742487.1	<i>P. putida</i> KT2440	Purine metabolism derivatives	(104)	ELS26892.1	100/65
PcaY	ABQ78288.1	<i>P. putida</i> F1	Aromatic acids	(105)	None	N/A
NbaY	BAF56675.1	<i>P. fluorescens</i> KU-7	Nitrobenzoic acid	(106)	ELS27539.1	100/63
NahY	YP_534842.1	<i>P. putida</i> G7	Naphthalene	(107)	ELS29403.1	75/48
McpT	YP_004750634.1 YP_004750609.1	<i>P. putida</i> DOT-T1E	Toluene	(108)	ELS29403.1	97/35
CttP (McpA)	NP_248870.1	<i>P. aeruginosa</i> PAO1	TCE	(109)	ELS25792.1	85/66

Chemotaxis to common aromatic hydrocarbon pollutants has been observed (108) and has been identified as a potential way to improve bioremediation making the search for MCPs that bind pollutants relevant to bioremediation (110). Bacterial MCPs come in a variety of shapes and sizes (110). The defining feature is a C-terminal cytoplasmic domain, known as the methyl accepting (MA) domain that interacts with CheA via CheW (111). The activation of this domain to stop CheA autophosphorylation depends on ligand binding in the N-terminal region of the MCP,

which is highly variable. MCPs have been categorized into 8 types, of which one type makes up 70% of bacterial MCPs (110). The most common, 'classical' MCP consists, from N to C-terminus of a transmembrane (TM) region, a periplasmic ligand binding region (LBR), a second TM region and the cytoplasmic methyl-accepting (MA) domain, which also includes the CheW interface. The least common types include variations with only one TM domain, the LBR on the cytoplasmic side, a variable number of TM regions but no clear LBR or no TM regions and no obvious LBR. So while the C-terminal region is extremely well conserved, the N-terminus is much more variable, even between MCPs of the same type. Additionally, MCPs that bind the same ligand have been demonstrated to show little sequence identity in the LBR in some cases (110).

Chemotaxis in *Pseudomonas* has been well studied and the functions of many MCPs have been characterized (81). Receptors have been described for intermediates of the tricarboxylic acid (TCA) cycle, amino acids, phosphate and aromatic compounds including the pollutants toluene and naphthalene (Table 2). Two other receptors, Aer and Aer-2 have been characterized for their ability to mediate aerotaxis (86). These receptors differ from all others as they do not have a periplasmic ligand binding domain, Aer-2 does not even have a transmembrane domain and is a cytosolic protein. While Aer-2 is thought to bind molecular oxygen, Aer does not sense any ligand directly, instead it detects energy production in the cell through a Flavin-adenine dinucleotide (FAD) prosthetic group (112). *Pseudomonas pseudoalcaligenes* KF707 appears to have homologs of both these receptors, as well as some MCPs that are likely organic and amino acid receptors (Table 2), though no MCP has been definitively characterized in this organism yet. It does not appear to have a homolog of McpT or NahY, the respective receptors for toluene and naphthalene. This implies that it may have a novel MCP that allows it to swim towards biphenyl, or it may be using its multiple homologs of Aer to swim towards biphenyl using energy-taxis.

1.6 Research Goals and Specific Aims

The scientific problems and studies summarized above led to the questions addressed in this thesis. The major problem I aimed to address was the inhibition of organic pollutant degradation by metals. Based on the studies that had recently been carried out in our lab, I developed my approach. The studies of *P. pseudoalcaligenes* KF707 showed that it was a good model for examining this issue as it could degrade and was chemotactic towards biphenyl, metals affected these interactions, and I knew I could understand the physiology of metal toxicity using metabolomics from studying tellurite toxicity and my work on copper toxicity in biofilms.

I thus had two goals in studying *P. pseudoalcaligenes* KF707:

(1) Determine how metal toxicity affects biphenyl degradation at a biochemical level.

(2) Determine how KF707 can swim towards this extremely hydrophobic compound.

These goals were divided into smaller questions, addressed in each of the three main chapters:

How does media composition influence metal toxicity? (*Chapter Two*).

What are the biochemical differences between cells grown on a simple carbon source compared to biphenyl and exposed to metals? (*Chapter Three*).

Does KF707 swim towards biphenyl using energy-taxis? (*Chapter Four*).

Chapter Two: Preface

Chapter 2 describes a high-throughput study for determining how media composition affects the metal tolerance of a bacterium. I built upon past works that had used microtitre plates and the Calgary Biofilm Device to determine how much metal an established biofilm could withstand, but made a key alteration by adding the metal at the same time as inoculation. To try and find a physicochemical explanation for the findings, I used geochemical modelling software to determine the speciation of the metals in the different media that were tested. This allowed me to correlate the concentration of various chemical species to the observed growth of the bacteria. This work demonstrated how media composition has a profound effect on the quantity of metal that bacteria can tolerate and provided the foundational data necessary for pursuing further experiments comparing cultures exposed to metals grown on different carbon sources. I designed the experiments and conducted them with technical assistance from Iain George. Gavin Duggan provided valuable advice on how to statistically analyze the growth data in combination with the speciation data. Davide Zannoni and Martina Capelletti provided editorial assistance in submitting the manuscript for publication in *Metallomics*.

Chapter Two: **Effect of aluminium and copper on biofilm development of *Pseudomonas pseudoalcaligenes* KF707 and *P. fluorescens* as a function of different media compositions**

Sean C. Booth, Iain George, Davide Zannoni, Martina Cappelletti, Gavin E. Duggan, Howard Ceri and Raymond J. Turner

Published in *Metallomics*, June 2013, 5:6 pp723-735

Copyright permissions for the reproduction of this manuscript can be found in Appendix E

2.1 Abstract

Bioremediation efforts worldwide are faced with the problem of metals interfering with the degradation of organic pollutants. There has been little systematic investigation into how the important environmental factors of media composition, buffering agent, and carbon source affect the exertion of metal toxicity on bacteria. This study aimed to systematically separate and investigate the influence of these factors by examining planktonic and biofilm establishment and growth. Two *Pseudomonads* were chosen, the PCB degrader *P. pseudoalcaligenes* KF707 and *P. fluorescens*. The two strains were grown in the presence of Al^{3+} and Cu^{2+} under different media conditions of carbon source (Lysogeny broth, biphenyl, succinate, aspartate, butyrate, oxaloacetate, putrescine and benzoate) and under different buffering conditions (high and low phosphate or MOPS). These experiments allowed for the elucidation of an effect of different metabolic conditions and metal speciation on planktonic bacteria growth and biofilm establishment and development under metal stress. Here we show that the nature of bacterial growth (planktonic and biofilm development) is dramatically affected by the interplay between toxic metals, carbon source and media composition. The capacity of a medium to bind toxic metals as well as quality

of carbon source greatly influences the amount of metal that bacteria can tolerate, depending on both the bacterium and metal. Future studies evaluating metal ion toxicity should consider these effects, as well as their interactions with specific environments into account in order to improve clean-up success.

2.2 Introduction

Environmental contamination of both organic and metal pollutants has become an important problem in North America and throughout the world (1). Progress has been made in applied bioremediation research of organic compounds (2) as well as in understanding microbial metal resistance (3). Basic research combining these two fields will provide a foundation for the development of practical applications for solving the issue of co-contamination.

In the environment, bacteria grow as biofilms (4). These single or multi-species assemblages of cells, which grow in a self-secreted layer of extracellular polymeric substance (EPS), are resistant or tolerant to antibacterials (5). The EPS, which consists of polysaccharides, protein, and DNA acts as a diffusional barrier to metals. It also includes functional groups which can bind ions, altering metal bioavailability (6). Thus, the interactions between metals and bacteria in their natural environment are affected by the properties of biofilms. Past studies observed differences between biofilm and planktonic cells in terms of susceptibility to toxic metals by exposing pre-grown cultures to gradients of metal and identifying the inhibitory concentrations (7-10). Despite their medical relevance, these experiments do not provide an accurate representation of bacterial growth in contaminated environments, where bacteria may need to grow in the presence of metal(s). Additionally, these past studies focused mainly on either genetic features of

the bacteria (7) or physicochemical properties of metals (9), but not on the growth environment and its influence on bacterial growth and biofilm development.

Media composition in laboratory experiments has been identified as one of the major factors influencing the metal speciation and hence the degree of toxicity observed (11). To study its contribution to metal toxicity, pure and well-characterized cultures growing in defined media should be preferable over microbial consortia growing in contaminated soil-based media to avoid misinterpretation due to the complexity of the system. Among laboratory media, many studies of bacterial metal resistance have used rich Lysogeny Broth (LB) media (7-9). Although this kind of medium ensures that cultures only experience metal toxicity without being subjected to nutritional stress, rich media can be problematic for metal toxicity experiments as undefined components such as yeast extract can chelate metals in an unknown manner (11). Moreover, LB is not well buffered and media pH is known to have a strong effect on metal speciation. These issues make defined, buffered minimal media more attractive for metal toxicity studies. The utilization of defined media also allows the application of software to computationally model the speciation of metals within a solution (12) which provides information on the amounts of each species in the media (13). Two additional factors influencing metal speciation are the buffering agent and the carbon source (11). The commonly used microbiological buffer, phosphate, is extremely efficient at chelating metals in solution and forming insoluble, non-bioavailable precipitates (14, 15). Various carbon sources can have metal binding functional groups. Thus all media components must be considered.

The current study aimed to methodically separate and investigate the influence of the above factors by examining biofilm development of two related Pseudomonads in the presence of two dissimilar metals (Al^{3+} and Cu^{2+}). *P. pseudoalcaligenes* KF707 is a poly-chlorinated biphenyl (PCB) degrader that has been well-studied over the past 20 years for its unique biodegradation

features, its interesting chemotactic abilities, and its capacity to grow on biphenyl as a biofilm (16-18). The effects of toxic metals in conjunction with PCBs on pre-grown biofilms have also been examined (10, 19). *P. fluorescens* has been studied for a long time in the context of biotechnological applications and several strains have been sequenced (20, 21). It is capable of metabolizing diverse compounds and has been used as a model organism for evaluating metal toxicity (9), signal response regulation (22), and metabolic stress responses (23). Together these two related soil organisms present a tool to compare the effects of media components on metal toxicity.

The various heavy metals contaminating the environment differ in their physicochemical properties, which dictates their interactions with media components and bacteria. For the purposes of this study, aluminium (Al^{3+}) and copper (Cu^{2+}) were chosen as model toxic metal ions due to their physicochemical differences from each other. The metabolic effects of Cu were recently studied in *P. fluorescens* (24) whose interactions with Al have also been very well characterized (25). Copper, in the form of Cu(II), while not quite a soft metal, prefers sulfur ligands such as cysteine and methionine but also nitrogen ligands like histidine (26). Aluminium on the other hand is far less redox active; it exists only in the Al^{3+} form and is a hard metal, preferring oxygen ligands like carboxylates (27). Cu is used as a cofactor in various proteins but Al is unused in any known life (28). Here the analysis of the effects of Cu and Al on bacterial growth in different media was performed providing insight into how media factors affect their toxicities and influences on microbial biofilm development. The biological data was also linked to computationally modeled speciation of these metals.

2.3 Materials and Methods

2.3.1 Strains

Pseudomonas pseudoalcaligenes KF707 (16) and *P. fluorescens* ATCC13525 stock cultures were stored at -80°C in 8% dimethyl sulfoxide (DMSO) in Lysogeny Broth (LB) media. Cultures were sub-cultured twice and were grown for 24 hrs at 30 °C before experimental inoculations. All streak plating for sub-cultures and spot-plating for recoveries was performed on 1.5% w/v agar plates with LB medium.

2.3.2 Media and Solutions

LB medium contained 5 g/L NaCl, 5 g/L yeast extract and 10 g/L tryptone. The pH was adjusted to 7.0 before autoclaving. Minimal Salts Medium (MSM) was made up in multiple parts; the base media containing phosphate (mono and dibasic), ammonium sulfate and 5 mM (nominal concentration, except biphenyl) of carbon source and the pH adjusted 7.0, then autoclaved. A concentrated mixture of trace metals was then added by sterile filtering through a 0.2 µm membrane. A 2X concentrated version of each media was also made for mixing with the tested metals. The final composition of all the MSM media (in g/L) was: (NH₄)₂SO₄, 2.6; MgSO₄·7H₂O, 0.4; CaSO₄·2H₂O, 0.0031; MnSO₄·H₂O, 0.05; FeSO₄·7H₂O, 0.1 (10). For the carbon source experiments, all MSM media contained 4.4 g/L K₂HPO₄ and 1.7 g/L KH₂PO₄. Carbon sources used were: sodium succinate hexahydrate, aspartic acid, butyric acid, oxaloacetic acid, putrescine and benzoic acid. For the buffer experiments the high phosphate media contained the same amount of phosphates whereas the low phosphate and MOPS [3-(N-morpholino) propanesulfonic acid] buffered media contained 0.26 g/L K₂HPO₄. For the MOPS-buffered media, a solution of MOPS was added by sterile filtering through a 0.2 µm membrane to give a final concentration of 4.18 g/L. For growth on biphenyl, the pegs of the Calgary Biofilm Device were twice immersed in 200

μL of 20 g/L biphenyl in hexane and allowed to dry. This led to the formation of crystals of biphenyl on the pegs and in the wells (10). MSM containing no additional carbon source was then used. For biofilm recoveries, 0.1% Tween-20 was added to LB media as a recovery solution. All serial dilutions were done in 0.9% w/v autoclave sterilized saline solution. Metal solutions were made up from $\text{Al}_2(\text{SO}_4)_3$ and $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ as stock solutions 4X more concentrated than the highest concentration tested (200 and 100 mM respectively). Stocks were filter sterilized through a 0.2 μm filter. A solution of ‘Universal Neutralizer’ was made up as a 50X concentrated stock consisting of (in g/L) L-histidine, 1.0; L-cysteine, 1.0; reduced glutathione 2.0. This solution was used to prevent residual soluble metals from carrying over into the recovery process thus limiting their toxicity (8).

2.3.3 Viability and Tolerance Testing

In order to examine the ability of bacteria to grow in the presence of metals, the established MBEC™ protocol (8, 29) was modified. Growth in 96-well microtitre plates was set up as follows. A decreasing gradient of metal concentration was prepared by adding 75 μL of media to all columns except column 2. To columns 2 and 3, 75 μL of a 1:1 mixture of 4X stock metal solution and 2X media was added. From column 3 onward $\frac{1}{2}$ serial dilutions were performed until the final column where 75 μL was discarded. All wells were then inoculated with a 75 μL 1/15 dilution of a 1.0 McFarland standard of culture in media, resulting in a microtitre plate containing 150 μL of inoculated media in each well with 11 decreasing metal concentrations as well as a control with no metal. The MBEC™ lid was then attached and cultures were incubated for 24hrs (biphenyl grown cultures were incubated an additional 24hrs to allow for slower growth on this carbon source (10)) at 30 °C, 95% humidity. *P. pseudoalcaligenes* KF707 was shaken at 100 rpm and *P. fluorescens* at 125 rpm. Inoculation densities were checked by 1/10 serially diluting the inoculants

7 times and spot-plated in order to be sure that the inoculants were in the correct range of a 1.0 McFarland standard (data not shown).

For quantifying viability, the established MBEC™ recovery protocol was used (8, 29). Briefly, the MBEC™ lid was twice rinsed in 200 μ L 0.9% saline for 1min, then placed in 200 μ L LB medium with 0.1% Tween-20 and 1/50 diluted universal neutralizer and sonicated for 10min at 60 Hz. The planktonic cultures were also recovered by adding 40 μ L of culture to 10 μ L of a 1/10 dilution of universal neutralizer in 0.9% saline. Both cultures were then serially diluted and spot plated onto LB agar plates, which were incubated for 36hrs at room temperature. Viability was then quantified by counting the highest dilution spot that had less than 50 colonies.

2.3.4 Solution State Modeling

Visual MINTEQ V3.0 (12) was used to model the solution state equilibria of possible chemical species at all nominal concentrations of metal for high, low phosphate, and MOPS buffered media and for the carbon sources succinate, butyric, and benzoic acid. Aspartic acid, putrescine and oxaloacetic acid were not included due to the lack of available complexation data. Taking as input the molar concentrations of all media components (as defined in the MSM), the equilibrium state concentrations of all chemical species in the solution were calculated using the included default MINTEQ databases. Each simulation was iterated 2000 times and was not terminated if the charge balance exceeded 30%. Activity correction was done by the Debye-Hückel method with a Davies b parameter of 0.3 and oversaturated solids were allowed to precipitate only after the final answer was reached. The pH was fixed at 7.0 since all media used here were adjusted before use and ionic strength was automatically calculated.

2.3.5 Statistical Analysis

Chemical species concentrations were scaled for unit variance using a custom R script (www.R-project.org). This process allowed species with very different dynamic ranges to be given equal weighting in further analysis. Species that did not exist in any particular media type were assigned a value of 1×10^{-32} , which approximates zero, but is amenable to log transformation. SIMCA P+ 12.0 (Umetrics, Umea, Sweden) was then used to perform Orthogonal Partial Least Squares (OPLS) analysis to investigate whether the observed culture viability (Y-variables) correlated with the predicted speciation concentrations (X-variables). The unifying variable, the type of media, and input concentration of metal, was used as sample names; the type of media determined the sample class. All chemical species were log transformed in order to approximate a normal distribution, an assumption of OPLS calculations. From the OPLS models, Variable Influence on Projection (VIP) was used to determine the importance of chemical species to the model phenomena at each concentration. Any variable with a VIP above 1 has an above average influence, however since VIP is a relative measure of influence, we avoided using an arbitrary threshold for VIP values. Instead, variables were ordered by VIP and any relative plateau in VIP was used to truncate the list of important variables. For the ensuing statistical analysis, separate analyses were done for each metal in understanding the effect of buffering agent and carbon source, respectively.

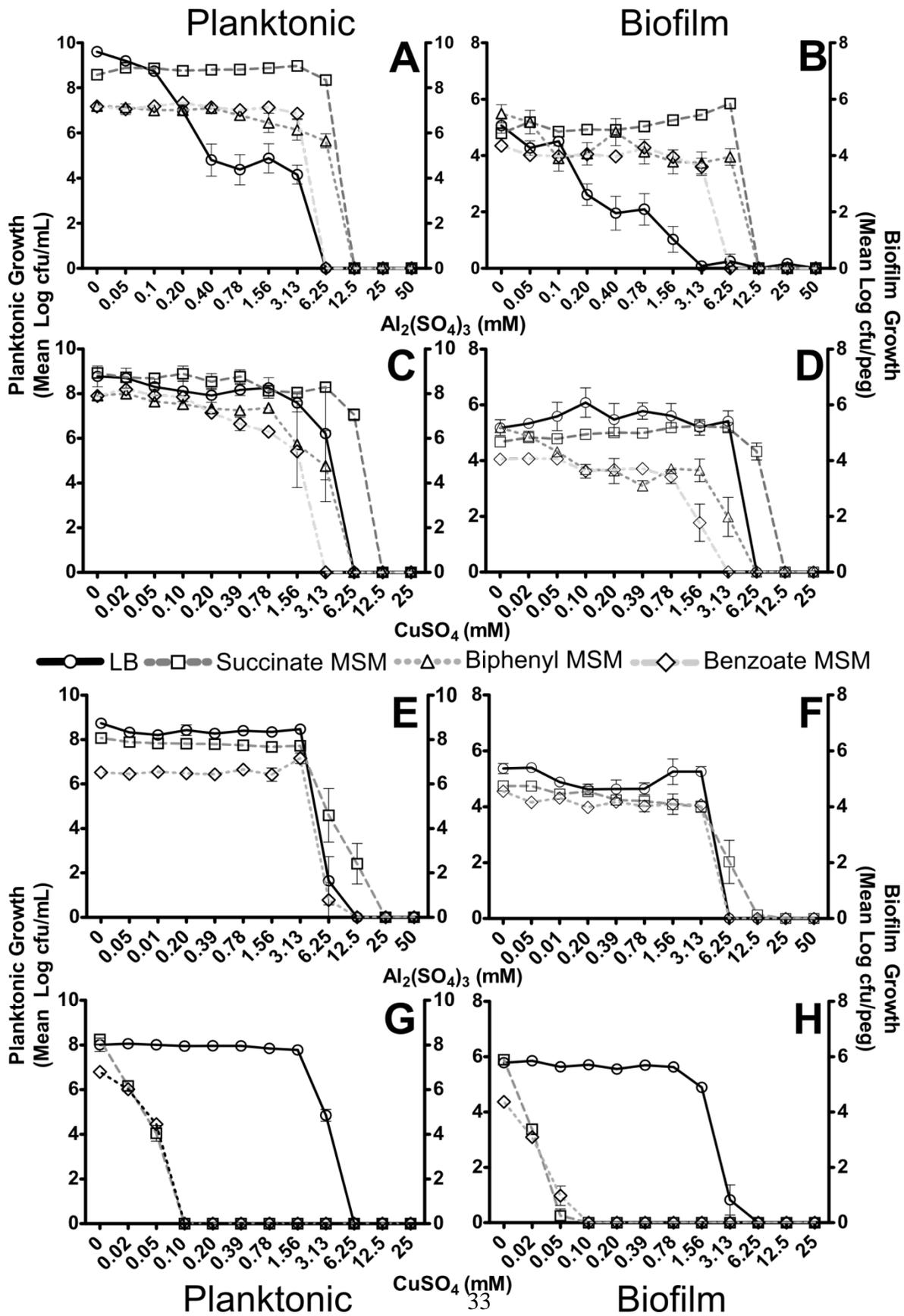


Figure 2-1: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate in various Media. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown on either LB (circles, black) or minimal salts media (MSM) with either Succinate (squares, dark grey), Biphenyl (triangles, mid grey) or Benzoate (diamonds, light grey) as a sole carbon source, with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. Only *P. pseudoalcaligenes* was grown on biphenyl, and was grown for a total of 48 hours before quantification.

2.4 Results

In the established MBEC™ protocol (8, 29), a planktonic inoculant is used to seed a biofilm, which is then allowed to grow to maturity, that is then transferred to a challenge plate with different concentrations of metals. This gives information on the response to an antimicrobial challenge. In the present study, the planktonic inoculant was added into media containing the metal. The analysis of this growth system provides information on planktonic survivability and growth at a given concentration of metal, as well as data about the planktonic bacteria's ability to adhere to the support peg and to grow as a biofilm in the presence of the metal challenge.

2.4.1 Media Effects

By growing the two *Pseudomonas spp.* in the presence of Al and Cu sulfate it was observed that the growth media affected the exertion of metal toxicity on planktonic viability and biofilm development. Under all conditions the same minimum concentration that prevented planktonic growth also prevented biofilm establishment and/or development (Figures 2-1, 2-2, 2-3). Generally, as metal concentration increased, biofilm growth tended to follow the same trend as planktonic growth. There were a few noteworthy cases where the establishment of biofilm growth diverged, as indicated below.

When comparing growth of the two strains in the presence of each metal with rich LB media and minimal media with succinate, biphenyl, or benzoate as a carbon source, a number of interesting differences are noticeable (Figure 2-1). To explore the effects of rich vs minimal media, LB was compared to succinate MSM. In most cases the MIC of each metal was always 2-fold higher in minimal media compared to LB. In the exception case, *P. fluorescens* with Cu, the MIC in LB was 64 times higher than in all MSM (Figure 2-1 G, H). Another stark difference between the two strains was observed when comparing Al tolerance between LB and MSM. When growing on LB, *P. pseudoalcaligenes* shows a steady decline in growth as metal concentration increased, whereas growth remained constant when growing on succinate MSM (Figure 2-1, A, B). There was an increase in biofilm growth on succinate MSM from about 5 to 6 log cfu/peg for Al (the highest tolerable concentration, Figure 2-1, B). This increase in biofilm growth as the metal concentration approaches inhibitory levels was also observed with some carbon sources similar to succinate, as noted below. The difference between LB and MSM with Al was not observed with *P. fluorescens* (Figure 2-1, E, F). Overall when comparing LB to MSM, the most drastic differences were the trend change in *P. pseudoalcaligenes* with Al (Figure 2-1, A) and the MIC difference in *P. fluorescens* with Cu (Figure 2-1, G).

The effect of carbon source complexity can also be observed from this set of experiments (Figure 2-1). While growth of both biofilm and planktonic cultures was generally lower when growing on biphenyl and benzoate, the overall MIC was comparable to the MIC in LB media. The growth trends differed as both planktonic and biofilm growth were up to 1 Log cfu/mL or peg lower at all concentrations. While growing on biphenyl or benzoate *P. pseudoalcaligenes* was more sensitive to concentrations approaching inhibitory levels. When comparing biphenyl to benzoate, growth trends were comparable as metal concentration increased, however with both

metals, the MIC of benzoate was 2-fold lower than that of biphenyl. When comparing benzoate to succinate, growth of *P. fluorescens* on benzoate was up to one Log cfu/mL or peg lower (Figure 2-1, E, H). *P. fluorescens* was not grown on biphenyl as it lacks the necessary degradation genes to use this carbon source.

2.4.2 Buffer Effects

In order to investigate whether the differences observed between LB media and succinate MSM were due to the phosphate buffer present in the MSM, alternative buffering conditions were examined. Generally, when the buffering agent was switched from high phosphate (42mM) to MOPS (20mM), the MIC decreased between 4-128 fold (Figure 2-2). For *P. pseudoalcaligenes*, this change was 16-fold with Al (Figure 2-2, A, B), whereas with Cu (Figure 2C,D) it was much larger, 128-fold. These contrast with *P. fluorescens*, which saw a lesser 8-fold decrease with Al (Figure 2-2, E, F). The 4-fold decrease observed with Cu (Figure 2-2, G, H), was less pertinent than the observation that without high phosphate no Cu was tolerable. For *P. pseudoalcaligenes* in the presence of Al, low phosphate allowed growth at a 2-fold higher concentration than MOPS (Figure 2-2, A, B). This increase was not observed in *P. fluorescens*, however at the concentration just before inhibitory, the MOPS buffered cultures showed a 4 log cfu/mL (Figure 2-2, E) and 2 log cfu/peg (Figure 2-2, F) decrease compared to the low phosphate buffered cultures. With Cu, the opposite effect was observed in *P. pseudoalcaligenes*; MOPS provided a 2-fold increase in inhibitory concentration over low phosphate (Figure 2-2, C,D). This was not observed in *P. fluorescens*, as without high phosphate, Cu was completely intolerable.

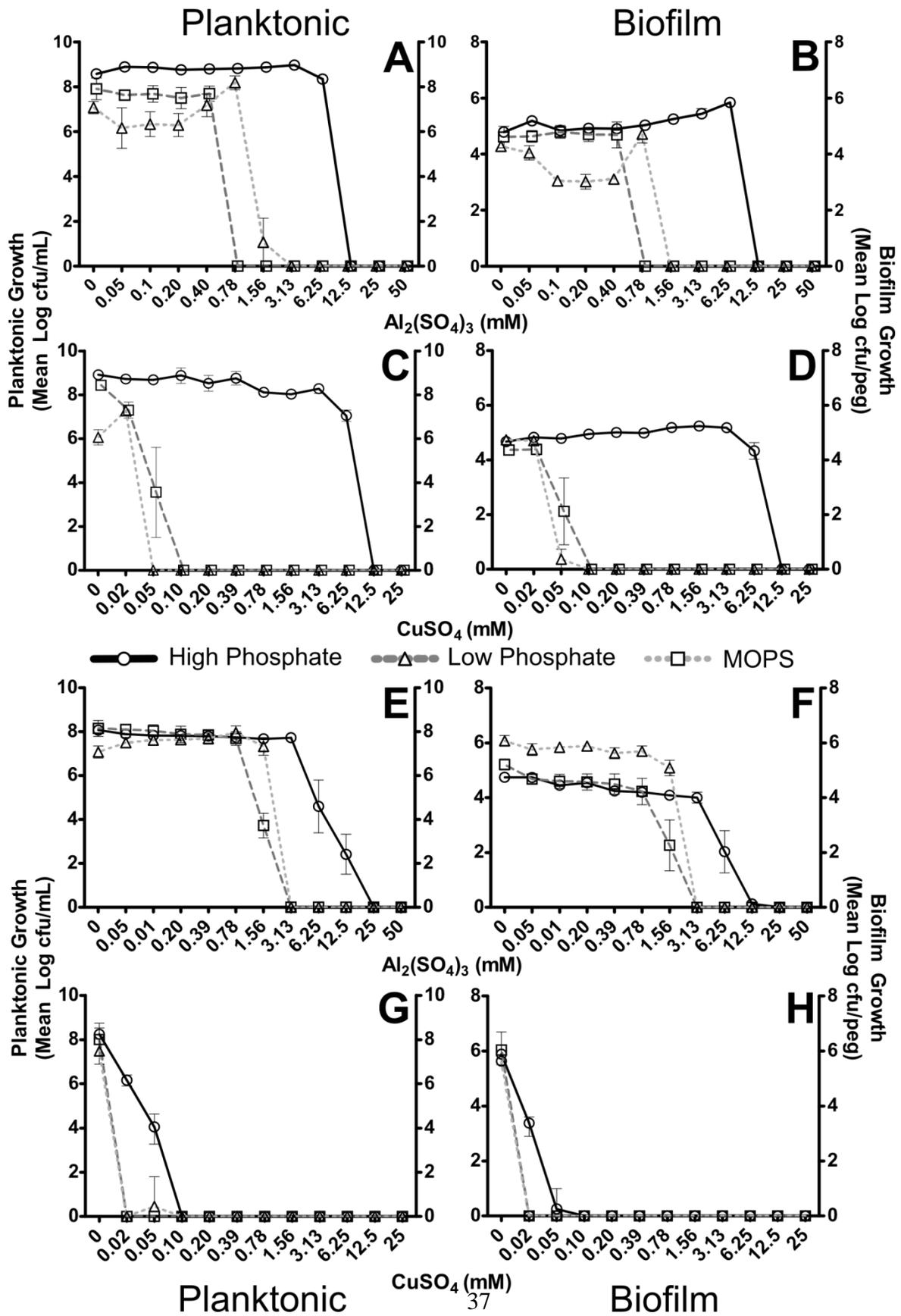


Figure 2-2: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate under various Buffer Conditions. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown in Minimal Salts Media (MSM) containing 5mM succinate as a sole carbon source and buffered with either High Phosphate (black, circles), Low Phosphate (dark grey, triangles), with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. The MOPS buffered media and Low Phosphate media both contained the same concentration of phosphate necessary as a phosphorus source. The pH of all media was nominally adjusted to 7.0.

2.4.3 Role of Carbon Source

After observing differential toxicity effects between biphenyl, benzoate, and succinate, a series of more chemically similar carbon sources were explored (for structures see Figure 2-4, D). In most cases the carbon source was not found to have a large effect on the MIC of the two metals tested. The growth trends did change as metal concentration increased. In the presence of Al, planktonic growth of *P. pseudoalcaligenes* on putrescine was about 1 log cfu/mL lower than other carbon sources (Figure 2-3, A), whereas growth of *P. fluorescens* on butyrate was similarly lower (Figure 2-3, E). The overall MIC of Al was only different for *P. pseudoalcaligenes* grown on oxaloacetate at 6.25mM, and for *P. fluorescens* growing on succinate at 25mM, otherwise it was 12.5mM. The trend of biofilm establishment and development was altered in the presence of Al (Figure 2-3, B, F). In *P. pseudoalcaligenes*, aspartate and oxaloacetate showed increases in growth comparable to succinate as the concentration of metal approached inhibitory levels (Figure 2-3, B). Growth on butyrate also showed a slight increase, however, along with putrescine, there was a ~1.5 log cfu/peg drop just before the MIC was reached. In *P. fluorescens* growth on aspartate and oxaloacetate was about 1 log cfu/peg higher than all other carbon sources at all concentrations. Al had a comparatively higher influence on growth on different carbon sources.

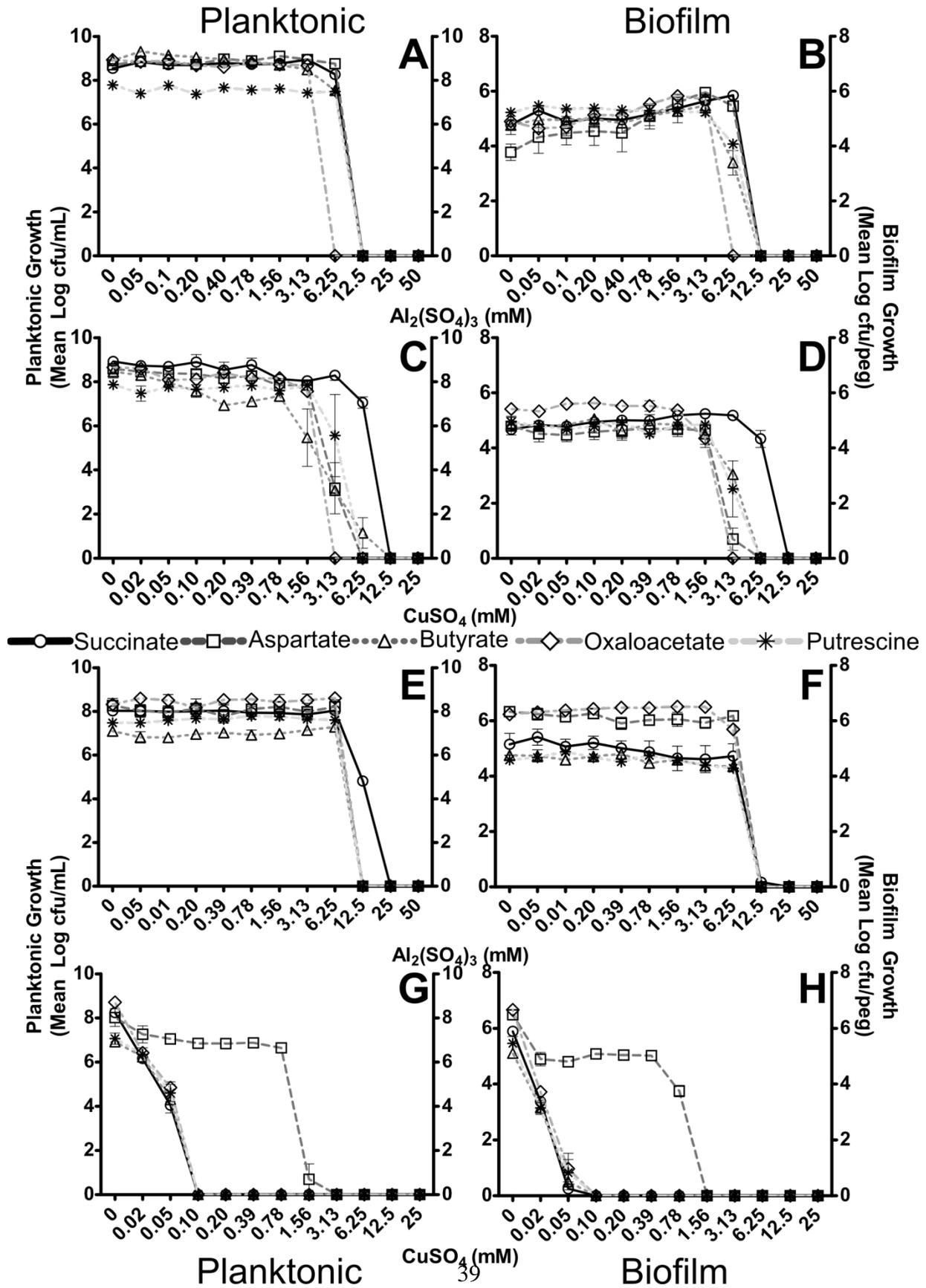


Figure 2-3: Planktonic and Biofilm Growth of *P. pseudoalcaligenes* KF707 and *P. fluorescens* in the presence of various Concentrations of Aluminum and Copper Sulfate on various Carbon Sources. Cultures were inoculated into media containing metal and incubated for 24 hours. Growth was quantified by spot-plating and is presented as log colony forming units (cfu) per mL (planktonic) and per peg (biofilm). Data points represent the mean of at least two experimental replicates which included at least two biological replicates ($n \geq 4$). Error bars represent standard error (SEM). *P. pseudoalcaligenes* (A, B, C, D) and *P. fluorescens* (E, F, G, H) were grown in Minimal Salts Media (MSM) containing 5mM of either Succinate (circles, black), Aspartate (squares, darkest grey), Butyrate (triangles, dark grey), Oxaloacetate (diamonds, light grey) or putrescine (stars, lightest grey) as a sole carbon source, buffered with high phosphate with either Al (A, B, E, F) or Cu (C, D, G, H) sulfate. The pH of all media was nominally adjusted to 7.0.

In *P. pseudoalcaligenes* with carbon sources other than succinate there was a decrease in planktonic and biofilm growth at Cu concentrations approaching inhibitory (Figure 2-3, C, D). *P. fluorescens* grown on aspartate had a MIC that was 16-fold higher than all other carbon sources (Figure 2-3, G, H). Even with this carbon source, *P. fluorescens* was still less tolerant to Cu than *P. pseudoalcaligenes* (Figure 2-3 C, D).

2.4.4 Metal ion Speciation Analysis

Orthogonal Partial Least Squares (OPLS) regression models were generated to elucidate any connections between chemical species in the growth media and the growth of the two organisms. The first OPLS model dealt with the effect of buffering conditions on Al. Using all the predicted chemical species, a model was generated which explained 79% of the variation in growth, as measured by cross-validation (Table 2-1). The first predictive component, which accounted for most of the variation, identified several variables for which a change in concentration correlated with growth. The chemical species with the highest variable influence on projection (VIP) from this initial analysis were examined closer.

Table 2-1: OPLS statistics from models correlating biofilm and planktonic growth of *Pseudomonas* spp. in the presence of either aluminium or copper with computationally modeled chemical species concentrations in high phosphate, low phosphate and MOPS buffered media.

Metal	Chemical Species	Component	R ² X	R ² Y	Q ²	CV-ANOVA (p)	
Aluminium	All predicted species	Model (Total)	0.976	0.825	0.788	4.49x10 ⁻⁹	
		Predicted (1)	0.363	0.793	0.745		
		Predicted (2)	0.390	0.0313	0.0431		
		Orthogonal	0.223	0.000962	N/A		
		Orthogonal (1)	0.101	0.0000579	N/A		
		Orthogonal (2)	0.123	0.000904	N/A		
	Only aluminium species	Model (Total)	0.999	0.771	0.746	1.11x10 ⁻¹⁰	
		Predicted (1)	0.364	0.745	0.746		
		Orthogonal	0.636	0.0259	N/A		
		Orthogonal (1)	0.62	0.000804	N/A		
		Orthogonal (2)	0.0152	0.0251	N/A		
	Copper	All predicted species	Model (Total)	0.976	0.791	0.738	7.51x10 ⁻¹³
			Predicted (1)	0.517	0.604	0.366	
			Predicted (2)	0.203	0.187	0.372	
Orthogonal			0.256	0.0000381	N/A		
Orthogonal (1)			0.256	0.0000381	N/A		
Only copper species		Model (Total)	1	0.228	0.18	0.944	
		Predicted (1)	0.631	0.218	0.179		
		Predicted (2)	0.059	0.00972	0.00133		
		Orthogonal	0.31	0	N/A		
		Orthogonal (1)	0.31	0	N/A		

The VIP of these species were plotted against their loading scores to identify those which were both significant and growth correlated/anti-correlated (Figure 2-4, A). Three types of species appeared on this plot, sulfate, succinate and hydroxylated Al. The succinate species were positively correlated with growth whereas the single Al species (Al₃(OH)₄⁵⁺) and sulfate species were

negatively correlated. Among many other predicted Al species, this was the only one that had a high VIP value. All the succinate species in the solution are accounted for. The remaining sulfate species not present were in low concentration.

Table 2-2: OPLS statistics from models correlating biofilm and planktonic growth of *Pseudomonas* spp. in the presence of either aluminium or copper with computationally modeled chemical species concentrations in high phosphate buffered media with succinate, butyrate and benzoate as carbon sources.

Metal	Chemical Species	Component	R ² X	R ² Y	Q ²	CV-ANOVA (p)	
Aluminium	All predicted species	Model (Total)	0.997	0.88	0.839	8.01x10 ⁻⁸	
		Predicted (1)	0.426	0.863	0.0819		
		Predicted (2)	0.233	0.0121	0.00635		
		Predicted (3)	0.0931	0.00442	0.0132		
		Orthogonal	0.246	0	N/A		
		Orthogonal (1)	0.157	0	N/A		
		Orthogonal (2)	0.0893	0	N/A		
	Only aluminium species	Model (Total)	0.996	0.691	0.646	1.73x10 ⁻⁷	
		Predicted (1)	0.26	0.691	0.646		
		Orthogonal	0.736	0.000697	N/A		
		Orthogonal (1)	0.736	0.000697	N/A		
	Copper	All predicted species	Model (Total)	0.998	0.774	0.71	1.01x10 ⁻⁴
			Predicted (1)	0.408	0.562	0.484	
Predicted (2)			0.119	0.209	0.226		
Orthogonal			0.472	0.00362	N/A		
Orthogonal (1)			0.318	0.000283	N/A		
Orthogonal (2)			0.0977	0.00175	N/A		
Orthogonal (3)			0.0564	0.00158	N/A		
Only copper species		Model (Total)	1	0.679	0.665	4.57x10 ⁻⁹	
		Predicted (1)	0.999	0.492	0.356		
		Predicted (2)	0.000181	0.187	0.309		

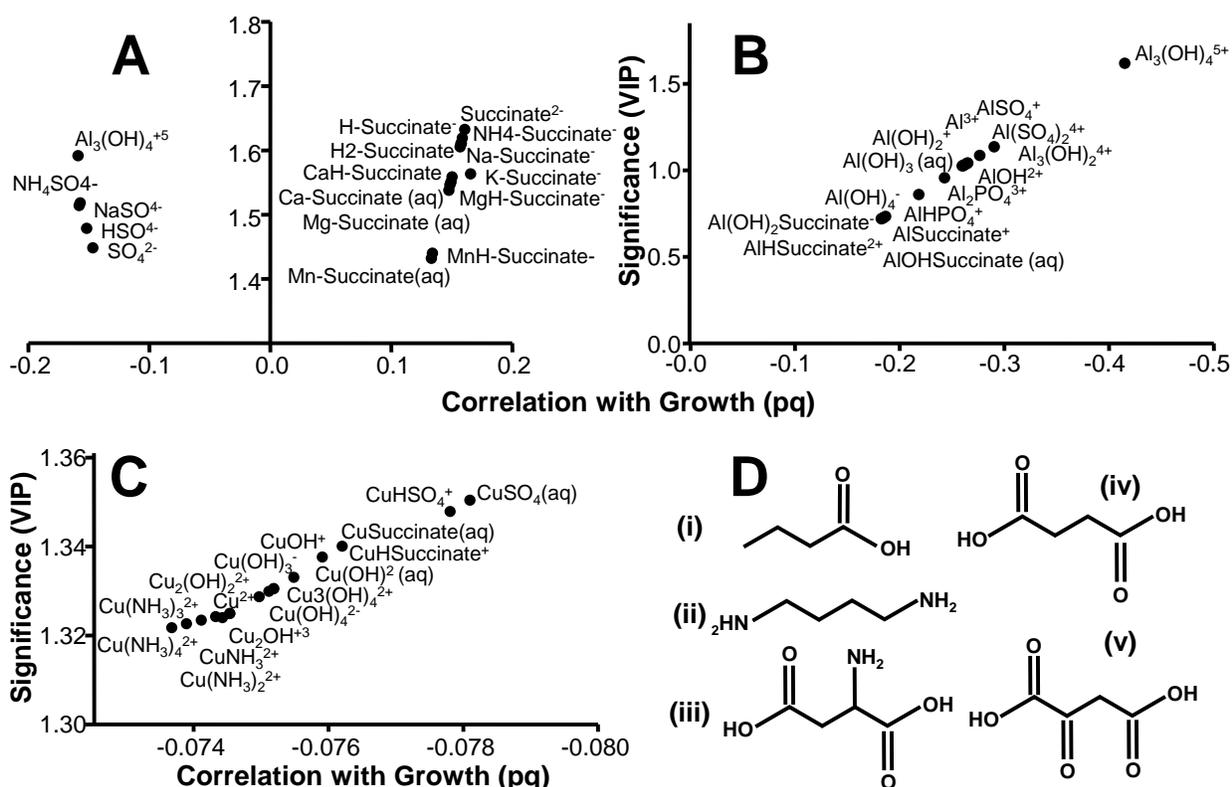


Figure 2-4: Correlation between bacterial growth and concentrations of chemical species predicted from computational modelling of high phosphate, low phosphate and MOPS buffered media. Data are from OPLS models correlating biofilm and planktonic growth of *Pseudomonas* spp. in the presence of Al (A, B) or Cu (C) with either all predicted species (A, C) or just the metal species (B). With Cu, a significant model could not be generated with just the metal species. The VIP (significance) indicates the importance of each chemical species to the model whereas pq represents the correlation with growth. Values presented were taken from the first predicted component. Only species with the highest VIP are shown. (D) The structures of the 5 carbon sources tested for growth: (i) butyrate, (ii) putrescine, (iii) aspartate, (iv) succinate, (v) oxaloacetate.

To further examine the relationship between the Al species and observed growth, a model using just the Al species was generated. Again a strong model was generated that accounted for 75% of the cross-validated variation in observed growth (Table 2-1). This model separated samples based on Al sulfate concentration, yet the differences between classes were minimized. Examining the loadings, all species showed similar values except for $\text{Al}_3(\text{OH})_4^{5+}$ (Figure 2-4, B). This species

was the most strongly anti-correlated with growth, differed mostly between Al containing samples rather than between the control and Al containing samples and had the highest VIP of all the Al species.

A similar set of models was generated for the Cu exposed cultures. With all predicted species a model was generated which accounted for 74% of the cross-validated variation in observed growth (Table 2-1). The first predictive component separated samples based on Cu concentration, and divided them into two groups: high phosphate was distinct from low phosphate and MOPS buffered samples. Along the first component the loadings again separated the chemical species into two groups, those correlated with growth, and those which were anti-correlated. To check which species were significant in their correlation or anti-correlation, VIP was plotted against loading value from the first component (Figure 2-4, C). Here we found that it was only the Cu species, except for CuHPO_4 (aq), that were most important for the model. All of these species were negatively correlated with growth. While the Cu sulfate species appear to be distinct from the rest of the group, all the Cu species have very similar VIP and loading values. To examine how strongly the Cu species were associated with growth, a model using just Cu species was generated but it had a very poor cross-validation predictive score (Table 2-1), thus it was not further analyzed.

The effect of different carbon sources on speciation was also examined by generating OPLS models. These correlated the growth of both *Pseudomonads* in the presence of Al and Cu on succinate, butyrate and benzoate in high phosphate buffered media with the computationally predicted chemical species in solution. The necessary thermodynamic and chemical equilibrium data to calculate speciation with all media components was unavailable for aspartate, oxaloacetate and putrescine. Therefore, we excluded these from our analysis. For each

(Figure 2-5, A). Using only the Al species common to all three media a model was produced (Table 2-2). The first predicted component accounted for 65% of the cross-validated variation in observed growth and separated samples based on Al concentration. The most important variable from this model was $\text{Al}(\text{SO}_4)^{2-}$ though AlSO_4^+ and $\text{Al}_3(\text{OH})_4^{5+}$ also had high VIP (Figure 2-5, B).

A link between Cu toxicity and the three computationally modeled carbon sources was also examined. When using all predicted species, it was the second predicted component that accounted for 20.9% of the variation of growth with a Q^2 value of 0.226 that considered attention (Table 2-2). This component mostly separated Cu containing samples from the control. The species which differentiated themselves were all Cu species that were anti-correlated with growth, but included none of the Cu carbon source complexes (Figure 2-5, C). When only the common Cu species were used to generate a model it was the first component that defined the trend (Table 2-2). This component accounted for 49.2% of the variation of growth and separated samples based on Cu concentration. When the VIP were examined, only CuHSO_4^+ and $\text{CuSO}_4(\text{aq})$ were significant (Figure 2-5, D).

2.5 Discussion

The co-contamination of organic and metal pollutants is an important global issue as metals interfere with the biodegradation of organic compounds (11, 30). Thus understanding how metals affect organic pollutant degradation would provide information for efficient bioremediation of these contaminated sites (2). The influence of metal toxicity on bioremediation processes has been shown to vary depending on the microorganism under analysis, the pollutant to be degraded, the type of soil or growth media, the concentration of the toxic metal and its bioavailability (11). In particular, differences in the bioavailable fraction of a toxic metal were shown to greatly influence

inhibitory concentrations (13). Bioavailability of a metal is determined by its distribution among free, ligand-bound, soluble, and precipitated forms, which is referred to as the chemical speciation (31).

Many studies have focused on identifying the concentrations of metals that interfere with microbial activities in contaminated soil microcosms (11). For a fuller understanding of the toxicity exerted by metals in the soil, environmental factors have to be taken into account. The factors of chemical speciation and local biogeochemistry can influence observed tolerance levels altering the perception of metal toxicity in bacteria. An understanding of the three-way interactions between the metal, the bacterium, and the environment would help clarify the biological response.

2.5.1 Inhibition of Biofilm Development

Under all experimental conditions, biofilm growth never occurred at a concentration of metal that did not accommodate planktonic growth. This observation was expected given that the planktonic cells would have to survive long enough to initiate attachment. Attachment is in part triggered by environmental stresses (32), that leads to physiological changes and a biofilm. Once initiated, the biofilms cells are more resistant to toxic metals (5). However, *P. pseudoalcaligenes* KF707 has been found to have equal susceptibility to metals regardless of growth mode (10). At high but tolerable concentrations of Al, biofilm cultures of *P. pseudoalcaligenes* growing in LB showed decreases from the control comparable to those observed in planktonic cultures. This indicated that the biofilm development was impaired under such stress conditions in rich media. In all cases where metal reduced planktonic growth, biofilm development was proportionally decreased. Similar effects were observed when *Staphylococcus aureus* was exposed to zero-valent bismuth nanoparticles as planktonic cell growth was reduced but biofilm formation was

completely inhibited (33). Additionally, using silver nanoparticles, formation of biofilms of *P. aeruginosa* could be inhibited with concentrations as low as 50 nM (34). While these applications of nanoparticles were beneficial in a medical context, the prevention of biofilm formation by metals observed in this study has a negative consequence for bioremediation.

Stressed cells are expected to form biofilms as a means of protection (32). It has been observed that exposure to sub-inhibitory levels of cadmium in *Rhizobium alarii* resulted in a decrease in planktonic cell density but an increase in biofilm formation (35). In the current study with *Pseudomonas spp.*, comparable growth was observed between the two modes. This suggests that biofilm cell density is more a function of cells attaching to a surface than a biofilm proliferating. Since the decreases of growth were observed at concentrations that approached inhibitory levels, some attachment may be affected hence the lower biofilm growth. If most of the quantified biofilm were from proliferation of attached cells, then the observed decrease in growth would only occur if the biofilm were comparably susceptible as the planktonic cultures. This may be the case for *P. pseudoalcaligenes* (10), however *P. fluorescens* biofilms are more resistant than their planktonic counterparts (9). Thus it can be concluded that attachment is a key factor in biofilm development and that the presence of toxic metals likely disrupts this process.

2.5.2 Comparison between Rich and Minimal Media

While metals will form various complexes in solution with components of media, these species can still remain bioavailable depending on the complex (36). Phosphate, which was in high concentration in the MSM, is an excellent chelator of Al, perhaps explaining the better growth in MSM compared to LB (14). Differences between complexes become biologically relevant when the two different *Pseudomonads* growing with Al in LB, and Cu in MSM are compared. In LB with Al, both bacteria were exposed to the same collection of metal complexes and displayed the

same MIC, but only *P. pseudoalcaligenes* showed a reduction in growth at higher concentrations. In MSM with Cu, the tolerance of *P. fluorescens* was almost negligible, especially compared to *P. pseudoalcaligenes*. These differences exemplify the importance of interactions between media components and metals in determining toxicity and also demonstrate the relevance of organism-specific resistance mechanisms to individual metals. Here the Cu data suggests that it is the free, soluble ion that is responsible for toxicity, even if this idea has been critiqued (31). Cu prefers sulfur and nitrogen ligands, which are abundant in LB (26, 37). Thus, the free ion concentration of Cu will be much higher in MSM, where *P. fluorescens* could not tolerate it. Heavy metal resistance of planktonic cells is largely reliant on efflux (3) with some examples of inhibition of uptake (38). The observed difference in tolerance between the two organisms in different media suggests that they are managing trans-membrane passage of free metal and/or one or more metal-ligand species differently. Cu will interact strongly with amino acids, while Al will to a lesser degree (39). Both strains tolerated Cu in LB comparably indicating parity in their Cu/Cu-ligand transport systems. *P. pseudoalcaligenes* was much more sensitive than *P. fluorescens* to Al in LB, indicating perhaps an inferior ability to (re)-exclude Al/Al-ligand species. Additional explanations arise in different biochemical resistance mechanisms. *P. fluorescens* has a very well-characterized mechanism for Al resistance where the glyoxylate shunt is used for energy, anti-oxidant NADPH production and oxalate is produced as an Al chelator (23, 40). This system was characterized in a minimal salts media, where both organisms tolerated high Al. Given the careful metabolic balancing required (25), *P. pseudoalcaligenes* may not be capable of this mechanism when growing on LB, hence the decreased growth. Finally, the low tolerance of *P. fluorescens* to Cu in MSM may be caused by nutritional stress. Metabolomic analysis of this organism growing in LB has shown that Cu stress causes the dysregulation of many amino acid metabolic pathways, likely

due to increased protein synthesis (24). Further to this, the lack of simple amino acids in the MSM could contribute to the lack of Cu tolerance observed in the present study.

2.5.3 Effects of Carbon Source Complexity

Increases in carbon source complexity resulted in a decrease in growth when metal concentrations approached inhibitory levels. While complexity, as well the investment required to enter central metabolism decreases in the order of biphenyl > benzoate > succinate; metal sensitivity decreased in the order of succinate < biphenyl < benzoate. The multi-step process of metabolizing biphenyl down to acetyl-CoA for entry into central metabolism requires the expression of a large number of genes/proteins (41). Thus, it is no surprise that metal toxicity shows more of an effect in biphenyl grown cultures, as there are more steps to interfere with and metal resistance strategies are typically energy demanding, leaving less opportunity for the cell to grow (42). The observation that growth on benzoate was more sensitive than biphenyl was unexpected, as it is an intermediate between biphenyl and succinate in complexity. The absolute amount of biphenyl available was much higher than that of benzoate, however, biphenyl is sparingly soluble in water. The lower amount of carbon per molecule and overall lower quantity may have led to a combined effect of metal toxicity and relative starvation, resulting in lower growth than on biphenyl.

2.5.4 Effect of Buffering Agent

The differences between rich and minimal media were ascribed to the formation of insoluble metal-phosphate precipitates. To confirm this, cultures were grown with MOPS buffered and unbuffered succinate MSM. Phosphate strongly interacts with Al (14), and also with Cu, but MOPS is not expected to interact with metals (15). Hence, with both metals, the removal of high phosphate resulted in lower inhibitory concentrations as a result of increased bioavailability of the

metal (14). For environmental purposes, this suggests that monitoring phosphate content of contaminated soils is very important. Indeed, the interactions between heavy metals and phosphates in soils have been found to be relevant to microbial growth (43). While the results for MOPS buffered and low-phosphate media were quite comparable, as was expected from a non-interacting buffering agent, the slight increase in tolerance observed in low-phosphate cultures was surprising. Since both media had the same amount of phosphate, differences cannot be from phosphate-metal precipitation. The low-phosphate media was essentially unbuffered, which would allow the bacterial culture to alter the media's pH. For both metals, the highly bioavailable free ionic form will decrease as pH increases (14, 44). The MOPS buffer could have prevented the cultures from increasing media pH, which would maintain the bioavailability of toxic Al.

2.5.5 Effect of Carbon Source

Following observations from carbon source complexity, a series of chemically similar carbon sources were investigated in order to investigate if other nutritional factors affect metal toxicity. Succinate is a simple dicarboxylic acid, which is a component in the highly conserved central metabolic TCA cycle, making it a very simple carbon and energy source. Aspartate is the amino acid analog of succinate, and being one of the simplest, is easily used for biosynthesis of more complex amino acids. Butyrate is the monocarboxylic acid version of succinate, making it the shortest fatty acid that can enter metabolism via β -oxidation. Oxaloacetate is also part of the TCA cycle, and is the same as succinate, except for the additional double bonded oxygen. Of the 4-carbon members of the TCA cycle, only succinyl-CoA is separated by more enzymatic steps than oxaloacetate and succinate. Putrescine is the only carbon source of those studied which does not have a carboxylic acid functional group. Instead it has two amine groups, leading it to enter metabolism quite differently than the other carbon sources. Since these compounds were the sole

source of carbon and energy, their different metabolic features were expected to result in differing levels of metal tolerance.

While only one case produced drastically different inhibitory levels, there were some alterations in growth at concentrations approaching inhibitory. Also, a number of non-metal related differences were observed. Growth of *P. pseudoalcaligenes* on putrescine, and *P. fluorescens* on butyrate, was reduced compared to other carbon sources, likely due to the added enzymatic steps necessary to integrate these compounds into central metabolism. Oddly, the decreases in growth were not under the same condition in both organisms, suggesting differences in metabolic efficiency between the two strains regarding these carbon sources. Biofilm growth of *P. fluorescens* was higher with the two TCA cycle intermediates (which are the simplest of those examined) compared to the other carbon sources, indicating it is more sensitive to quality of carbon source compared to *P. pseudoalcaligenes*.

There was another difference between the two organisms, as *P. pseudoalcaligenes* showed an increase in biofilm growth at concentrations of Al approaching inhibitory levels with succinate, aspartate, and oxaloacetate. Putrescine and butyrate instead showed decreased growth just before inhibitory levels were reached. Biofilm growth is expected to increase under environmental stress (4), though not necessarily in all organisms, with all metals or under all nutritional conditions, as evidenced here. The nutritional differences are logical as succinate, aspartate, and oxaloacetate enter metabolism far easier than putrescine and butyrate, thereby leaving more energy for metal resistance. The lack of increased growth in *P. fluorescens* suggests differences between the environmental stress response and biofilm production gene regulation in the two organisms. The trend of increasing growth was only observed with Al. The opposite effect, a decrease in both biofilm and planktonic growth, was observed as Cu concentrations approached inhibitory levels.

This could indicate a difference on the effect of attachment for biofilm establishment between the two metals.

Oxaloacetate was observed to allow an increase in biofilm growth; it was also the worst carbon source for metal tolerance in *P. pseudoalcaligenes*. It would be tempting to expect this difference to be due to it directly chelating metals resulting in a Trojan horse of metal and carbon source. Oxaloacetate is the most electronegative carbon source, which would make it an excellent Al chelator (log K data is unfortunately unavailable [NIST 8.0]). No difference in *P. fluorescens* tolerance was observed, but this may be related to superior Al resistance mechanisms, which have already been demonstrated (25).

The biggest difference in metal toxicity differences was observed with *P. fluorescens* growing on aspartate in the presence of Cu. The 16-fold increase in tolerance was unparalleled and unexpected.

2.5.6 Computationally Modeled Chemical Speciation

Computationally modeling of defined media with toxic metals allows for an estimation of the different chemical species that are present (13). Since a specific metal species may be responsible for the inhibition of bacterial growth, pairing speciation modeling with observed bacterial viability and biofilm development at various concentrations should show a connection. The large number of predicted metal species, among a much greater number of benign and beneficial chemical species produces a complex system to make correlations. OPLS was used here as this statistical technique can deconvolute large, multi-variate systems, thus aiding in understanding underlying differences and simplifying correlations (45). This approach allowed exploration of possible correlations between the observed culture viability and biofilm development in MSM and metal speciation, which was predicted by Visual MINTEQ v3.0.

2.5.6.1 Speciation with Different Buffering Reagents

The model of growth in the presence of Al with various buffering agents was expected to show which Al species were most important for toxicity since the amount of non-phosphate bound Al was dramatically different between the three media. With both models, $\text{Al}_3(\text{OH})_4^{5+}$ was the most significant and correlated Al species. The tridecameric $\epsilon\text{-Al}_{13}$ hydroxylated complex has been a cause of concern for fish and phytotoxicity (46). This complex is formed from the joining of smaller polynuclear hydrated complexes, including $\text{Al}_3(\text{OH})_4^{5+}$ (47). Given that the $\epsilon\text{-Al}_{13}$ complex was not included in the speciation model, as well as other semi-characterized polynuclear hydrated species (48), the $\text{Al}_3(\text{OH})_4^{5+}$ complex was likely best correlated with toxicity as it was one of two polynuclear complexes predicted by the computational speciation. Given the frequent production of such complexes, especially in waste water treatment, where Al is used as a coagulant (47). Our data demonstrate it is important to understand how these species interact and exert toxicity in microorganisms. For Cu, a correlation was not found.

2.5.6.2 Speciation with Different Carbon Sources

Both metals showed comparable results when growth on the three carbon sources was modeled. The variation, which was not due to differences in carbon source, was mostly modeled in the second component. For both metals this resulted in metal-sulfate species being the most correlated with toxicity. The most likely explanation for this phenomenon is that these species were in higher proportion to all others as sulfate was by far the most common ligand in solution. The fact that metal-carbon source species only contributed to making the models worse was noteworthy, as this indicates that metal entering the cell when bound to the only available carbon source was not important for toxicity.

2.6 Conclusion

Media factors have a very important, though undefined influence on metal toxicity. Here we have demonstrated that media richness, ease of assimilation of carbon source as well as the amount and type of buffering agent will influence metal toxicity. These experimentally manipulated factors were found to have different effects with regards to the two *Pseudomonas* species that were examined as well as between the two metals. The differences observed between rich and minimal media, with quantity and type of buffering agent and with carbon source were most pronounced in *P. fluorescens* and with copper, the carbon source providing the largest influence. Using matching computationally modeled speciation data, $Al_3(OH)_4^{5+}$ was correlated with the decrease in bacterial growth. Such a correlation was not made with any of the carbon sources, nor with copper. However, the most important observation was that the two metals and organisms were affected differently by the manipulated media factors. While some effects can be accounted for by physicochemical considerations of the media, others were related to the particular metal and microorganism involved. This indicates that these environmental factors, along with bacterial growth mode, must be taken into careful consideration when studying the effects of metal concentrations on organic pollutant degradation or when monitoring metal toxicity levels. In these situations, it will be key to understand the three-way interactions between the microorganisms, the metals, and the nutritional and physicochemical properties of the contaminated site. Further basic research into understanding how the environmental factors enabling bacterial growth under toxic metal stress such as nitrogen, phosphorus and carbon availability as well as physical parameters such as temperature affect metal toxicity will enable improved applied solutions for pollutant remediation.

2.7 References

1. **Maier RM, Sandrin TR.** 2003. Impact of Metals on the Biodegradation of Organic Pollutants: A Review. *Environmental Health Perspectives*.
2. **Shukla KP, Singh NK, Sharma S.** 2010. Bioremediation: Developments, current practices and perspectives. *Genetic Engineering and Biotechnology Journal* **2010**.
3. **Nies DH.** 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiology Reviews* **27**:313-339.
4. **López D, Vlamakis H, Kolter R.** 2010. Biofilms. *Cold Spring Harbor perspectives in biology* **2**:a000398.
5. **Harrison JJ, Ceri H, Turner RJ.** 2007. Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology* **5**:928-938.
6. **Langley S, Beveridge TJ.** 1999. Metal binding by *Pseudomonas aeruginosa* PAO1 is influenced by growth of the cells as a biofilm. *Canadian Journal of Microbiology* **45**:616-622.
7. **Harrison JJ, Tremaroli V, Stan MA, Chan CS, Vacchi-Suzzi C, Heyne BJ, Parsek MR, Ceri H, Turner RJ.** 2009. Chromosomal antioxidant genes have metal ion-specific roles as determinants of bacterial metal tolerance. *Environmental Microbiology* **11**:2491-2509.
8. **Harrison JJ, Turner RJ, Ceri H.** 2005. High-throughput metal susceptibility testing of microbial biofilms. *BMC Microbiology* **5**:1-11.
9. **Workentine ML, Harrison JJ, Stenroos PU, Ceri H, Turner RJ.** 2008. *Pseudomonas fluorescens*' view of the periodic table. *Environmental Microbiology* **10**:238-250.
10. **Tremaroli V, Vacchi Suzzi C, Fedi S, Ceri H, Zannoni D, Turner RJ.** 2010. Tolerance of *Pseudomonas pseudoalcaligenes* KF707 to metals, polychlorobiphenyls and chlorobenzoates: effects on chemotaxis-, biofilm- and planktonic-grown cells. *FEMS Microbiology Ecology* **74**:291-301.
11. **Sandrin TR, Hoffman DR.** 2007. Bioremediation of Organic and Metal Cocontaminated Environments: Effects of Metal Toxicity, Speciation, and Bioavailability on Biodegradation, p 1-34. *In* Singh SN, Tripathi RD (ed), *Bioremediation: A Novel Technology*. Springer-Verlag, Berlin.
12. **Gustafsson JP.** 2010. Visual MINTEQ, v3.0 Beta. Royal Institute of Technology, Land and Water Resource Engineering, Stockholm, Sweden.,
13. **Twiss MR, Errécalde O, Fortin C, Campbell PGC, Jumarie C, Denizeau F, Berkelaar E, Hale B, Van Rees K.** 2001. Coupling the use of computer chemical speciation models and culture techniques in laboratory investigations of trace metal toxicity. *Chemical Speciation and Bioavailability* **13**:9-24.
14. **Berthon G.** 2002. Aluminium speciation in relation to aluminium bioavailability, metabolism and toxicity. *Coordination Chemistry Reviews* **228**:319-341.
15. **Mash HE, Chin YP, Sigg L, Hari R, Xue H.** 2003. Complexation of copper by zwitterionic aminosulfonic (good) buffers. *Analytical Chemistry* **75**:671-677.
16. **Furukawa K, Miyazaki T.** 1986. Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. *Journal of Bacteriology* **166**:392-398.

17. **Tremaroli V, Fedi S, Tamburini S, Viti C, Tatti E, Ceri H, Turner RJ, Zannoni D.** 2011. A histidine-kinase *cheA* gene of *Pseudomonas pseudoalcaligenes* KF707 not only has a key role in chemotaxis but also affects biofilm formation and cell metabolism. *Biofouling* **27**:33-46.
18. **Triscari-Barberi T, Simone D, Calabrese FM, Attimonelli M, Hahn KR, Amoako KK, Turner RJ, Fedi S, Zannoni D.** 2012. Genome sequence of the polychlorinated-biphenyl degrader *pseudomonas pseudoalcaligenes* KF707. *Journal of Bacteriology* **194**:4426-4427.
19. **Tremaroli V, Fedi S, Turner RJ, Ceri H, Zannoni D.** 2008. *Pseudomonas pseudoalcaligenes* KF707 upon biofilm formation on a polystyrene surface acquire a strong antibiotic resistance with minor changes in their tolerance to metal cations and metalloids oxyanions. *Archives of Microbiology* **190**:29-39.
20. **Rong X, Gurel FB, Meulia T, McSpadden Gardener BB.** 2012. Draft genome sequences of the *Pseudomonas fluorescens* biocontrol strains Wayne1R and Wood1R. *Journal of Bacteriology* **194**:724-725.
21. **Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson Iii LS, Thomashow LS, Loper JE.** 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nature Biotechnology* **23**:873-878.
22. **Workentine ML, Chang L, Ceri H, Turner RJ.** 2009. The GacS-GacA two-component regulatory system of *Pseudomonas fluorescens*: a bacterial two-hybrid analysis. *FEMS Microbiology Letters* **292**:50-56.
23. **Mailloux RJ, Lemire J, Appanna VD.** 2010. Metabolic networks to combat oxidative stress in *Pseudomonas fluorescens*. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*:1-10.
24. **Booth SC, Workentine ML, Wen J, Shaykhtudinov R, Vogel HJ, Ceri H, Turner RJ, Weljie AM.** 2011. Differences in Metabolism between the Biofilm and Planktonic Response to Metal Stress. *Journal of Proteome Research* **10**:3190-3199.
25. **Lemire J, Mailloux R, Auger C, Whalen D, Appanna VD.** 2010. *Pseudomonas fluorescens* orchestrates a fine metabolic-balancing act to counter aluminium toxicity. *Environmental Microbiology* **12**:1384-1390.
26. **Dupont CL, Grass G, Rensing C.** 2011. Copper toxicity and the origin of bacterial resistance - New insights and applications. *Metallomics* **3**:1109-1118.
27. **Piña RG, Cervantes C.** 1996. Microbial interactions with aluminium. *BioMetals* **9**:311-316.
28. **Exley C.** 2009. Darwin, natural selection and the biological essentiality of aluminium and silicon. *Trends in Biochemical Sciences* **34**:589-593.
29. **Harrison JJ, Ceri H, Stremick CA, Turner RJ.** 2004. Biofilm susceptibility to metal toxicity. *Environmental Microbiology* **6**:1220-1227.
30. **Sandrin TR, Maier RM.** 2003. Impact of metals on the biodegradation of organic pollutants. *Environmental Health Perspectives* **111**:1093-1101.
31. **Campbell PGC.** 1995. Interactions between trace metals and aquatic organisms: a critique of the free-ion activity model. *Metal speciation and bioavailability in aquatic systems*, p

- 45-102. *In* Tessier A, Turner DR (ed), Metal Speciation and Bioavailability in Aquatic Systems. John Wiley & Sons, Chichester.
32. **O'Toole G, Kaplan HB, Kolter R.** 2000. Biofilm formation as microbial development, vol 54, p 49-79.
 33. **Hernandez-Delgadillo R, Velasco-Arias D, Diaz D, Arevalo-Niño K, Garza-Enriquez M, De la Garza-Ramos MA, Cabral-Romero C.** 2012. Zerovalent bismuth nanoparticles inhibit *Streptococcus mutans* growth and formation of biofilm. *International journal of nanomedicine* **7**:2109-2113.
 34. **Kalishwaralal K, BarathManiKanth S, Pandian SRK, Deepak V, Gurunathan S.** 2010. Silver nanoparticles impede the biofilm formation by *Pseudomonas aeruginosa* and *Staphylococcus epidermidis*. *Colloids and Surfaces B: Biointerfaces* **79**:340-344.
 35. **Schue M, Fekete A, Ortet P, Brutesco C, Heulin T, Schmitt-Kopplin P, Achouak W, Santaella C.** 2011. Modulation of metabolism and switching to biofilm prevail over exopolysaccharide production in the response of *Rhizobium alamii* to cadmium. *PLoS ONE* **6**:e26771.
 36. **Nybroe O, Brandt KK, Ibrahim YM, Tom-Petersen A, Holm PE.** 2008. Differential bioavailability of copper complexes to bioluminescent *Pseudomonas fluorescens* reporter strains. *Environmental Toxicology and Chemistry* **27**:2246-2252.
 37. **Sezonov G, Joseleau-Petit D, D'Ari R.** 2007. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology* **189**:8746-8749.
 38. **Borghese R, Cicerano S, Zannoni D.** 2011. Fructose increases the resistance of *Rhodobacter capsulatus* to the toxic oxyanion tellurite through repression of acetate permease (ActP). *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **100**:655-658.
 39. **Bohrer D, do Nascimento PC, Mendonça JKA, Polli VG, de Carvalho LM.** 2004. Interaction of aluminium ions with some amino acids present in human blood. *Amino Acids* **27**:75-83.
 40. **Lemire J, Kumar P, Mailloux R, Cossar K, Appanna VD.** 2008. Metabolic adaptation and oxaloacetate homeostasis in *P. fluorescens* exposed to aluminum toxicity. *Journal of Basic Microbiology* **48**:252-259.
 41. **Furukawa K, Fujihara H.** 2008. Microbial degradation of polychlorinated biphenyls: Biochemical and molecular features. *Journal of Bioscience and Bioengineering* **105**:433-449.
 42. **Nies DH.** 1999. Microbial heavy-metal resistance. *Applied Microbiology and Biotechnology* **51**:730-750.
 43. **Park JH, Bolan N, Megharaj M, Naidu R.** 2011. Isolation of phosphate solubilizing bacteria and their potential for lead immobilization in soil. *Journal of Hazardous Materials* **185**:829-836.
 44. **Meador JP.** 1991. The interaction of pH, dissolved organic carbon, and total copper in the determination of ionic copper and toxicity. *Aquatic Toxicology* **19**:13-32.
 45. **L. Eriksson EJ, N. Kettaneh-Wold and S. Wold.** 2001. *Multi- and Megavariate Data Analysis: Principles and Applications*. Umetrics AB, Umea, Sweden.
 46. **Casey WH.** 2006. Large aqueous aluminum hydroxide molecules. *Chemical Reviews* **106**:1-16.

47. **Zhang P, Hahn HH, Hoffmann E, Zeng G.** 2004. Influence of some additives to aluminium species distribution in aluminium coagulants. *Chemosphere* **57**:1489-1494.
48. **Casey WH, Phillips BL, Furrer G.** 2001. Aqueous aluminum polynuclear complexes and nanoclusters: A review. *Reviews in Mineralogy and Geochemistry* **44**:167-190.

Chapter Three: Preface

In this chapter I present a metabolomics study that provides direct biochemical evidence of how metal toxicity specifically affects bacteria during the degradation of an organic pollutant. Using my earlier work that determined the concentrations of metals that *P. pseudoalcaligenes* KF707 could tolerate while growing on biphenyl and simple carbon sources I sought to determine what differences there were inside cells grown on these different nutrients but exposed to the same toxic metals. I improved upon our past metabolomics protocols both at the bench and for data analysis, then proceeded to apply them to examining how copper and aluminium affected cultures grown on either biphenyl or succinate. By identifying and quantifying metabolites using gas-chromatography mass-spectrometry under each combination of possible conditions, both from within the cells and from the spent growth media I was able to gain great insight into how metal toxicity differed in cells grown on biphenyl. I identified many different metabolites and associated pathways that were highly indicative of oxidative stress being exacerbated by the metals. This key finding provides a plausible, nigh-universal explanation for why metals inhibit the degradation of organic pollutants. I designed the experiments, analyzed the data and wrote the manuscript with assistance from my co-authors, particularly Dr. Weljie for the experimental and data analysis aspects of the metabolomics.

Chapter Three: **Metabolomics reveals differences of metal toxicity in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on different carbon sources**

Sean C. Booth, Aalim M. Weljie and Raymond J. Turner

Published in *Frontiers in Microbiology*, August 2015, 6:827

Copyright permissions for the reproduction of this manuscript can be found in Appendix E

3.1 Abstract

Co-contamination of metals and organic pollutants is a global problem as metals interfere with the metabolism of complex organics by bacteria. Based on a prior observation that metal tolerance was altered by the sole carbon source being used for growth, we sought to understand how metal toxicity specifically affects bacteria using an organic pollutant as their sole carbon source. To this end metabolomics was used to compare cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl or succinate as the sole carbon source in the presence of either aluminum or copper. Using multivariate statistical analysis it was found that the metals caused perturbations to more cellular processes in the cultures grown on biphenyl than those grown on succinate. Aluminum induced many changes that were indicative of increased oxidative stress as metabolites involved in DNA damage and protection, the Krebs cycle and anti-oxidant production were altered. Copper also caused metabolic changes that were indicative of similar stress, as well as appearing to disrupt other key enzymes such as fumarase. Additionally, both metals caused the accumulation of biphenyl degradation intermediates indicating that they interfered with biphenyl metabolism. Together these results provide a basic understanding of how metal toxicity specifically affects bacteria at a biochemical level during the degradation of an organic pollutant and implicate the catabolism of this carbon source as a major factor that exacerbates metal toxicity.

3.2 Introduction

Anthropogenic pollution in the form of organic compounds and metal elements is widespread around the globe (1, 2). Bioremediation is the process of using living organisms to either degrade organic pollutants into innocuous end-products or immobilize metals and is an excellent method for cleaning up pollution and preventing ecological damage (3). While many sites contaminated with just organic pollutants have been successfully remediated, co-contamination of metals has been found to interfere with bacterial degradation of organic pollutants (4). This is especially problematic as a large proportion of sites (e.g. 40% of U.S. E.P.A. Superfund sites) are contaminated with both types of pollutants (5). Despite this issue, little of the research on understanding mechanisms of metal toxicity in bacteria has focused on this problem (6). Additionally, applied research has emphasized only the characterization of situations where various metals inhibited the degradation of different pollutants (7) while the underlying physiological effects of these metals on bacteria have yet to be investigated thoroughly.

Past work on metal toxicity in bacteria was mainly focused on determining the concentration of metals that inhibited growth, but these concentrations were found to vary widely depending on the growth medium as this determines the speciation and therefore bioavailability of metals (5). While investigating this issue of how media composition affects metal toxicity, we observed that tolerance to metals also differed depending on the sole carbon source provided for growth, all other media components being equivalent (8). As cells growing on different carbon sources make use of different metabolic pathways, it was postulated that the cellular targets of metal toxicity could differ based on the carbon source being used. Previously we used metabolomics to uncover differences between how surface attached biofilms of *P. fluorescens* and free-swimming planktonic cultures were affected by copper exposure (9). This systems biology

technique enables the identification and quantification of the low-molecular weight compounds within a sample, thereby providing a metabolic profile. As metabolomics has been successfully applied to understanding metal toxicity in a variety of prokaryotic and eukaryotic systems (10), here we sought to use this approach to characterize the differential effects of metal toxicity in bacterial cultures grown on either a simple carbon source or a model pollutant.

Pseudomonas pseudoalcaligenes KF707 is a bacterium that was first studied due to its ability to degrade biphenyl and polychlorinated biphenyls (11, 12) and has since been studied with regards to its metal-resistance capabilities (13, 14), chemotaxis towards biphenyl (15) and recently had its genome sequenced (16). Compared to other metals and bacteria, *P. pseudoalcaligenes* KF707 was found to be more sensitive to copper (Cu) and aluminum (Al) (15) but was able to tolerate higher concentrations of these metals when grown on succinate compared to biphenyl (8). Both of these metals have been found in co-contamination with polycyclic aromatic hydrocarbons and polychlorinated biphenyls (17–22), especially in electronic waste (23–27), their mechanisms of toxicity have been characterized in other systems (28–31), and are physicochemically very distinct from one-another (6). As such, these metals were selected to determine if the physiological effects of metal toxicity were the same in cultures grown on different carbon sources. To this end we used gas-chromatography mass-spectrometry (GC-MS) metabolomics to characterize cultures and spent media of *P. pseudoalcaligenes* KF707 grown on either succinate or biphenyl, in the presence of the same, sub-inhibitory concentrations of either Al or Cu. By comparing metabolic profiles using multi-variate statistical techniques, differences were discovered in how the carbon source being used for growth influenced the effects of Al and Cu. To our knowledge this provides the first systems-wide characterization of the combined effects of metal toxicity and growth on an aromatic carbon source. Our results indicate that biphenyl catabolism is both affected by and

exacerbates metal toxicity as multiple metabolic pathways were altered in response to this combined stress. These insights into the physiological effects of metal toxicity in an environmentally isolated bacterium should provide a basis for further investigations into the biochemical mechanisms of how metal toxicity disrupts the metabolism of complex aromatic substrates.

3.3 Experimental Procedures

3.3.1 Culture Growth

Pseudomonas pseudoalcaligenes KF707 was routinely cultured in minimal salts medium (MSM) consisting of (in g/L) K_2HPO_4 , 4.4; KH_2PO_4 , 1.7; $(NH_4)_2SO_4$, 2.6; $MgSO_4 \cdot 7H_2O$, 0.4; $CaSO_4 \cdot 2H_2O$, 0.0031; $MnSO_4 \cdot H_2O$, 0.05; $FeSO_4 \cdot 7H_2O$, 0.1 (15). Trace metals were filter sterilized using a 0.2 μm filter and added as a 20X stock directly to each culture flask. 8% dimethyl sulfoxide (DMSO) frozen stocks were used to inoculate 5 mL subcultures which were grown overnight. 50 μL (succinate) or 1mL (biphenyl) was then used to inoculate 250 mL flasks containing 50 mL of MSM with either 5mM succinate or 0.39g of sterile biphenyl as this compound is insoluble in water. At the same time as inoculation Al was added in the form of $Al_2(SO_4)_3$ or Cu as $CuSO_4$ to a final concentration of 3 mM or 60 μM (respectively). One of each culture type was grown simultaneously 5 separate times for 5 replicates of each condition. For growth curves, every 8 hours two aliquots of 50 μL were removed and separately serially diluted 1/10 down to 10^{-7} . 20 μL spots were plated on LB agar and counted after 24h of incubation at 30 °C. For pH determination, separate cultures were grown and every 8 hours 1mL of culture was removed, centrifuged for 5 min at 10,000 RPM and the pH determined using a Beckman 720 pH meter and probe (Beckman, Pasadena, USA).

3.3.2 Collection of Samples and Extraction of Metabolites

All cultures were treated identically for sample collection and metabolite extraction, except that biphenyl-grown cultures were poured through a coarse filter to strain out residual biphenyl particles. After 24 hours of growth cultures were harvested by rapid centrifugation at 4 °C for 5 minutes at 5,000 RPM in a Sorvall RC5B-plus using the SLA-1500 rotor (Thermo Scientific, Waltham, USA). Supernatant was collected as ‘Spent Media’ and all samples were immediately frozen in liquid nitrogen for storage at -80 °C. After all samples were collected, metabolites were extracted. The extraction solution (900 µL 2:1 methanol:chloroform) was added directly to frozen cell pellets. After homogenization by pipetting the sample was transferred to a 2 mL MpBio FastDNA Spin kit vial. Cells were lysed via bead-beating according to the manufacturer’s instructions (40 s at 6.0 power level) (MP Biomedicals, Santa Ana, USA). Samples were returned to ice immediately and processed as previously (9). Briefly, 300 µL water and chloroform were added to all samples. After hand-mixing, samples were centrifuged for 7min at 14,000 RPM and the aqueous phase was transferred to a fresh tube. These steps were repeated twice to obtain pure aqueous samples which were dried down in a vacuum concentrator at room temperature and stored until analysis at -80°C. For spent media samples, 1.5mL thawed sample was dried down at room temperature in a vacuum concentrator. After resuspension in 50 µL ddH₂O, 900 µL 2:1 methanol:chloroform was added and samples were treated identically to cellular samples post bead beating. For extraction/derivitization controls, clean, empty vials were treated identically to true samples.

3.3.3 Derivatization and Analysis by GC-MS

Sample derivatization and GC-MS analysis was performed identically to our previous work (9, 32) according to (33). 50 µL of 20 mg/mL methoxylamine in pyridine was added to samples

and which were mixed and incubated for 2 h at 37 °C, 200 RPM. Next, 50 µL of N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA, Sigma Aldrich, St. Louis, USA) was added to all samples which were incubated identically for 45 min. Samples were then diluted with hexane and centrifuged for 7 min at 14,000 RPM to remove particulates. 150 µL was transferred to gastight vials for analysis on a Waters GCT premier mass spectrometer. Helium was used as the carrier gas at a constant flow of 1.2 ml min⁻¹. 1 µL derivatized sample was injected into a DB5-MS column (splitless, 30 m × 0.25 mm ID × 0.25 µm) at an injector temperature of 275 °C. Initial column temperature was 80 °C which was held for 1 minute and then ramped at 12 °C min⁻¹ to 320 °C and held for 8 minutes. The MS was operated in a range of 50-800 m/z.

3.3.4 Identification of Metabolites

Mass spectral deconvolution, calibration, identification and analysis was performed using Automated Mass Spectral Deconvolution and Identification Software (AMDIS) (34). Data were first calibrated for retention time shifts using a set of alkane standards (C10-C30). The GOLM Metabolome Database (GMD) (35) VAR5 library was imported into the NIST MS Search program (36) and components were identified and confirmed manually using both the GMD library and the NIST11 Mass Spectral library. A custom library of reproducible but unidentifiable analytes was also generated from components extracted from representative samples using AMDIS. For details see supplementary material.

3.3.5 Removal of derivatization artifacts

Any analytes found in the extraction/derivatization controls that contained no sample were considered artifactual and were excluded from quantification. These compounds originated either from derivatization reactions occurring between the plastics of the sample vessel, solvents, or from the GC column. The most biologically relevant compounds that were removed from quantification

were uracil, decanoic, dodecanoic, hexadecanoic, heptadecanoic and octadecanoic acid. For a detailed list of compounds see the supplementary material.

3.3.6 Quantification of Metabolites

Concurrent to the generation of a library of mass spectra for the identification of components, ions were selected from each compound as representatives for quantification. These ions were manually selected to ensure that they were unique to the retention index window of the analyte and of high intensity relative to all other fragmentation ions. The common trimethyl-silyl (TMS) ions 73, and 147 as well as ions with high background such as 121, 266, 285 and 299 were excluded from selection to ensure that the ions being quantified were truly representative of the analyte in question. Peak quantification was performed using this ion retention time list using MET-IDEA (37). Peak selection parameters were manually tuned (for specifics, see supplementary material) to ensure that quantification was representative of individual analytes. After quantification the validity of each analyte's ions as representative of that analyte was determined. Using a custom R script (R-Project for Statistical Computing, CRAN.R-project.org) the correlation across all samples between all pairs of ions as well as the sum of all quantified ions was determined for each analyte, for details see supplementary material. Any ion with a correlation to the sum of all ions from that analyte below 0.8 was removed from subsequent analysis. This process excluded some analytes from further analysis as they had no ions that passed this threshold. Most notable was cysteine 3TMS. The remaining ions were summed for each analyte. Any analytes that represented the same metabolite (i.e. aspartate 2TMS and aspartate 3TMS) were summed. This gave 269 metabolites quantified, of which 89 were identified and used for subsequent analysis.

3.3.7 Statistical Analysis: Pre-processing

Data were first processed to enable their proper downstream analysis. A noise threshold was determined by calculating the mean intensity of analytes in the extraction control samples. Any value below this mean was interpreted as noise and was thus set to zero. Data were then normalized by probabilistic quotient normalization (PQN) (38) in order to account for any variation in cellular material collected, despite cell densities being highly similar between control and treated samples. This kind of normalization was preferred as past experiences attempting to normalize to sample wet-weight or total integral normalization were inferior (9). For details on the normalization procedure, see the supplementary material. After normalization data were log transformed and each analyte was mean-centered and scaled to unit-variance. These transformations allow variables with disparate dynamic ranges and means to be compared on the same scale. Comparison of sample clustering between raw, log transformed data (Supplementary Figure 1) and normalized data (Supplementary Figure 2) indicated that the normalization did not unduly skew the dataset as clustering remained similar.

3.3.8 Statistical Analysis

Statistical analyses were performed using SIMCA P+ v13.0 (UMETRICS) and R (R-Project for Statistical Computing, CRAN.R-project.org). Principal component analysis (PCA) was performed on all samples together as well as individual analyses for each class. The separate analysis was used to remove any outlying samples that were not highly similar to the remaining members of that class. Only a single sample was removed by this process, for details see the supplementary material. After this trimming process PCA and orthogonal partial least squares discriminant analysis (OPLS-DA) was performed on the entire dataset. OPLS-DA models were

then generated for each combination of control and treated samples (e.g. biphenyl control and biphenyl Al exposed, succinate control spent media and succinate copper exposed spent media etc.) in order to minimize the amount of variation being examined in any one model, thereby maximizing the interpretability of each model. For each of these pairwise OPLS-DA models the R^2Y , Q^2 and CV-ANOVA p values were used to assess model quality; only models with CV-ANOVA p values below 0.05 were accepted as statistically significant. From each significant model the variable influence on projection (VIP) and p(corr) values were exported for further interpretation. Shared and unique structures plots were used as this type of plot simplifies analysis of metabolomics data while maintaining the depth of complexity within the dataset (39).

3.3.9 Identification of Unknown Analytes

Biological interpretation of the data implied the possibility of several metabolites that had not been identified, and could not be identified due to their lack of standards available in either Golm or NIST libraries. Using manual fragmentation to predict the mass spectra, all the intermediates of the two possible catechol-degrading pathways (starting either with catechol 1,2-dioxygenase or catechol 2,3-dioxygenase) were searched for, as well as several other metabolites that were predicted to be present. Only two compounds with unknown peaks that matched to a high enough threshold were found. From these predicted analytes their functional groups were used to predict the Kováts' retention index (40). This approach confidently identified two unknown analytes: 2-hydroxymuconic semialdehyde and 2-phosphoglycolic acid. For details of the identification procedure see the supplementary material.

3.3.10 Pathway Enrichment Analysis

mBROLE was used to determine which metabolic pathways were being affected by metal toxicity in cultures grown on the two carbon sources. By calculating the number of metabolites

affected that occur in a particular pathway mBROLE can determine which metabolic pathways were most affected under a condition (41). From the OPLS-DA models comparing control and metal exposed cells for each carbon source and metal, metabolites with a VIP>0.8 were submitted using their Kyoto Encyclopedia of Genes and Genomes (KEGG) IDs (42). Pathways that were enriched with a false discovery rate adjusted p-value <0.05 were accepted as affected by metal exposure. Metabolic pathways that are commonly found by this type of analysis due to overbroad interpretation of KEGG pathways were removed. For details see supplementary material.

3.4 Supplementary Material

Author's Note: This portion was originally published as Supplementary Material. Most of the supplementary figures and tables have been integrated with the main text for this thesis, except those presented here. The Results and Discussion begin on page 87.

3.4.1 Supplementary Methods

3.4.1.1 Removal of poor samples and outliers

GC-MS spectra were initially inspected to identify any samples where excessively low signal was detected. Several samples were removed from this inspection though they all occurred in different sample types leaving at least 4 replicates in all samples. Class based PCA was later used to check for outliers (Table 3-1). Only one model had significant statistics and the outlying sample was removed (Figure 3-1).

3.4.1.2 Identification of Metabolites

Data files were converted into netCDF format by MassLynxDatabridge (Waters, Milford, USA). For component selection, AMDIS parameters were set to: Component width: 11, adjacent peak subtraction: Two, Resolution: Low, Sensitivity: High, Shape requirements: High. The GOLM Metabolome Database (GMD) VAR5 library was imported into the NIST MS Search program and

Table 3-1: Model statistics of PCA-Class models showing intraclass variation of GC-MS metabolomics quantification of samples from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl or succinate as the sole carbon source and exposed to either control, Aluminum or copper.

Model	A	N	R ²	Q ²
PCA-Class(Bp)	1	5	0.335	-0.1
PCA-Class(Bp Al)	1	4	0.586	0.0114
PCA-Class(Bp Cu)	1	5	0.464	-0.0057
PCA-Class(M Bp)	1	5	0.377	-0.0924
PCA-Class(M Bp Al)	2	5	0.686	-0.21
PCA-Class(M Bp Cu)	1	5	0.447	-0.0642
PCA-Class(M Sc)	1	5	0.408	0.0078
PCA-Class(M Sc Al)	1	4	0.545	-0.1
PCA-Class(M Sc Cu)	1	4	0.425	-0.1
PCA-Class(Sc)	1	5	0.397	-0.1
PCA-Class(Sc Al)	1	5	0.417	-0.1
PCA-Class(Sc Cu)	1	5	0.723	0.581

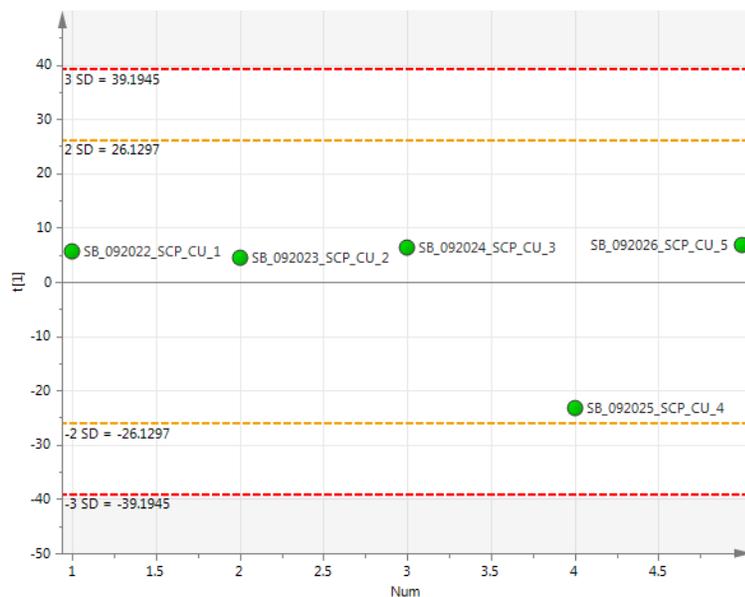


Figure 3-1: PCA Scores plot from PCA-Class model of GC-MS metabolomic quantification of samples from cells of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate as the sole carbon source and exposed copper.

components were identified using both the GMD library and the NIST11 Mass Spectral library. Positively identified components were confirmed manually by inspecting the library fragmentation and the fragmentation of representative components extracted from representative chromatograms from each type of sample. This was important as some experimental fragmentation patterns did not match the library fragmentation patterns to a high enough threshold (MS Search probability >90%) to be accepted as positive identifications, despite the retention indices matching exceptionally well (99% similar). In these cases, manual inspection allowed fragmentation ions that were causing the decrease in match similarity to be confirmed as true fragmentation products of that compound. For example, malic acid 3TMS has only a small peak for the 117 ion in both the NIST and GMD library, however this ion was the dominant ion in experimental samples. Since this ion corresponds to a carboxylic acid group with a TMS group replacing the removed proton, of which malic acid can produce two such ions it was concluded that this component was indeed malic acid 3TMS. Hexose-6-phosphate was the name given to an ambiguous metabolite that may have been glucose-6-phosphate or mannose-6-phosphate as this difference cannot be resolved by GC-MS. While confirming all positive identifications a custom library of fragmentation patterns was generated that included all identified compounds from all sample types as well as reproducible but unknown components. At completion this library contained 489 spectra including both identified and unidentified true and artefactual compounds.

3.4.1.3 Removal of Derivatization Artefacts

Artefacts were removed as described in the main text. For a complete list of artefacts, their retention times and intensities, see Table 3-2.

3.4.1.4 Quantification of Metabolites

MET-IDEA parameters: Chromatography: GC, Average peak width: 0.05, Minimum peak width: 1, Maximum peak width: 3, Peak start/stop slope, Adjusted retention time accuracy: 1, Peak overload factor: 0.9. These parameters were determined manually to ensure that adjacent peaks would not contribute to peak quantification.

3.4.1.5 Confirmation of Analyte Quantification Quality

To ensure that the ions that had been quantified were representative of that analyte, the sum of all ions that had been pre-selected for each analyte was determined. The correlation between the ions that were part of this sum and the sum was then determined across all samples. If the mean correlation was below 0.8 then these ions were determined to not be representative enough of that analyte and so quantification of the analyte was deemed inaccurate and was excluded from further analysis. This ensured that the sum of quantified ions that was used to represent each analyte was robust and only analytes that were present in enough samples to generate a strong correlation were used.

3.4.1.6 Data Analysis: Pre-processing

Probabilistic quotient normalization (PQN) procedure: the data were divided into four blocks, one for each type of sample. These blocks were biphenyl cells, succinate cells, biphenyl spent media and succinate spent media. Block based normalization allows for the data to be normalized more accurately as the correction factor is specific for each of the disparate sample types. Within each block an 'average' spectrum was determined by taking the median of each analyte. For each sample the difference ratio between the 'average' spectrum was determined for each analyte. The median of these difference ratios was then determined and used as correction factor for all analytes in that sample. The median was used instead of the mean as it is less sensitive

to outlying high or low values. After normalization the blocks were recombined, data were log transformed and each analyte was mean-centered and scaled to unit-variance. These transformations allow variables with disparate dynamic ranges and means to be compared on the same scale.

3.4.1.7 Identification of Unknown Analytes

Biological interpretation of the data implied the possibility of several metabolites that had not been identified, and could not be identified due to their lack of standards available in either Golm or NIST libraries. Predicted compounds were drawn in Mass Spec Calculator Pro V 4.09 (ChemSW, Inc/Quadtech Associates, Ringoes, USA) and hydroxyl and amine groups had their free hydrogens replaced with trimethylsilyl (TMS) groups. The previously generated library of unknown analytes was then searched for analytes which had their largest m/z at calculated m/z of the unfragmented derivitized possible analyte. Additionally an m/z at $M-15$ of the mass ion was used as a pre-screening ion to identify unknowns that could possibly be the derivitized compound of interest as the loss of a single methyl group is a common fragmentation. Unknown analytes which had m/z peaks that matched the M and $M-15$ masses of a compound being searched for were further checked for m/z fragmentation peaks that matched possible fragmentation outcomes of the known, derivitized compound drawn in Mass Spec Pro. By this process only 3 unknown analytes were given probable identifications. The analytes that these unknowns were hypothesized to be were 2-hydroxymuconic semialdehyde (2TMS) and 2-phosphoglycolic acid (3TMS). After confirming the hypothetical fragmentation matched the observed m/z peaks of these unknown analytes, the Kovát's retention index of these compounds were calculated and compared to the unknown analyte's. By this process Unknown RT:11.37 was identified as 2-hydroxymuconic semialdehyde 2TMS. Unknown RT:11.41 which was immediately adjacent to this peak, with an

almost identical fragmentation pattern is likely the same analyte. Unknown RT:11.58 was identified unequivocally as phosphoglycolic acid 3TMS. See the supplementary identification of unknowns for the mass spec fragmentation patterns.

3.4.1.8 Use of p(corr) to determine correlation between metabolites and sample type

p(corr) indicates the degree of correlation between a metabolite and a component. In all pairwise models that were generated in this study, there was only a single predicted component, upon which samples of different types were consistently separated. This means that the correlation with the predicted component corresponds to a correlation with a sample type. The particular sample type for each model was determined using the corresponding scores plots for that model as it is arbitrary whether the control class is assigned to the positive or negative side of a component when the OPLS-DA model is generated (Supplementary Figures 4 and 5). Thus p(corr) was used as an indication of whether a metabolite was correlated with control or metal exposed samples.

3.4.1.9 Generation and Interpretation of Secreted Metabolites SUS plots

Metabolites were quantified from the spent media of cultures, but the same metabolites that were quantified from within cells were also found in the spent media. To determine which metabolites were actually altered in the spent media not just due to changes in the cells spilling over to the spent media, VIP and p(corr) values from both models comparing control and exposed cells and spent media were considered together. Metabolites were considered to be secreted only if their VIP was above 0.8 in the spent media model (indicating a significant difference between the control and exposed samples). From this reduced list of metabolites only those with either a VIP below 0.8 in the cells (indicating no significant difference between the control and exposed cells meaning the difference in the spent media must be due to changes in secretion) or if the p(corr) in the spent media model was opposite that of the cells model. Metabolites that increased

in the spent media but decreased in the cells in response to metal exposure must have been secreted more in response to the metal whereas those that decreased in the spent media but increased in the cells in response to metal exposure must have been secreted less in response to metal exposure. By indirectly comparing the data in this fashion it precludes the need to directly compare the cells to the spent media samples which cannot be done accurately as only a portion of the spent media was analyzed while all of the cells from a sample were collected and analyzed.

3.4.1.10 Pathway Enrichment Analysis

Pseudomonas aeruginosa PA7 was used to generate the background set of metabolites as according to phylogenetic analysis (Triscari-Barberi, unpublished results) and BLAST searches of tRNA synthetases this is *P. pseudoalcaligenes* KF707's closest relative that was available in the database. Metabolic pathways that are commonly found by this type of analysis due to the inclusion of large numbers of metabolites in their pathway (metabolic pathways, ABC transporters, phospho-transfer system, two component system, cyanoaminoacid metabolism) as well as pathways that were obviously erroneously enriched (carbazole degradation, toluene and xylene degradation, methane metabolism and biosynthesis of siderophore group non-ribosomal peptides) were excluded. Pathways were then further screened for relevance to ensure that the affected metabolites that were implicating a pathway would actually be involved in that pathway in *Pseudomonas pseudoalcaligenes* KF707. By this screening, glutathione metabolism (implicated metabolites are only involved in trypanosomatid glutathione metabolism), oxidative phosphorylation (this pathway only includes 9 metabolites making its enrichment guaranteed by only succinate and fumarate being affected), lysine degradation (this pathway would only be expected in a culture growing on lysine or peptides containing lysine, not on a defined media with ammonia as the sole nitrogen source and carbon coming from only a single source, additionally,

the metabolites that implicated this pathway were mostly disjointed from one another), phenylalanine metabolism (implicated by benzoic acid and salicylic acid, which can be derived from phenylalanine but in the biphenyl grown cultures that this pathway was implicated in these compounds obviously were derived from biphenyl degradation not phenylalanine), butanoate metabolism (implicated mainly by Krebs' cycle intermediates present in this pathway) and pyrimidine metabolism (only Cu, both β -alanine and propanedioic acid implicated but these are only uracil degradation products in other organisms).

Table 3-2: Analytes present in derivitization control samples. Retention time (RT), retention index (RI) and peak area are mean of all derivitization control samples run.

Name	RT	RI	Area
Low MW RT:5.0	5.004	981.7	266375
Unknown Low MW RT: 5.1	5.1137	991.4	3573868
Low MW RT:5.14	5.1456	994.3	332613.5
Low MW RT:5.17	5.1692	996.4	142543
Tetrasiloxane	5.3549	1012.9	14647
Unknown: Bp_2 RT:5.51	5.4008	1016.9	16864
2-Hydroxy pyridine	5.4331	1019.8	22978
Unknown RT:5.58	5.5786	1032.7	35919
N,N-Diethylcarbamate 1TMS	5.6755	1041.3	104711.5
Unknown: Bp_3 RT:5.74	5.7397	1047	11248
Unknown RT:5.96	5.9577	1066.4	352625.5
Unknown RT: 6.0	6.0097	1071	272351.5
Unknown RT:6.17	6.1778	1085.9	146841
Unknown RT:6.2	6.2264	1090.3	286432
Unknown RT:6.24	6.2424	1091.7	170877.5
Unknown RT:6.31	6.308	1097.5	58428
Unknown RT:6.5	6.487	1113.4	291267
Unknown Silane RT:6.8	6.6897	1131.4	1331628
Unknown RT:6.78	6.7797	1139.4	66098.5
Pentasiloxane	6.9637	1155.7	14305
Unknown: Bp_M_3_R RT:7.60	7.0553	1163.9	4447
Unknown RT:7.17	7.1658	1173.7	3210
Dodecane	7.474	1201.1	8887
Unknown: Bp_2 RT:5.33	7.5199	1205.1	980
2-siloxyltetrasiloxane	7.65	1216.7	2174

Silanamine 3TMS	7.7668	1227.1	39694
Unknown RT:7.841	7.8412	1233.7	31394
Unknown RT:7.94	7.9423	1242.7	112864.5
Pentadecane	8.347	1278.6	73379.5
Unknown RT:8.349	8.347	1278.6	55980.5
Unknown RT:8.450	8.4495	1287.7	8010
Unknown RT:8.65	8.6511	1305.6	480136.5
Unknown RT:8.85	8.8458	1322.9	369826
Unknown RT:8.89	8.8827	1326.2	122777.5
Unknown: (Similar to Lumichrome)	8.9492	1332.1	16036
Uracil 2TMS	9.0173	1338.1	7111
Unknown RT:9.09	9.2064	1354.9	25942
Unknown RT:9.77	9.7706	1405	1513
Unknown RT:9.96	9.9518	1421.1	358892
Unknown RT:10.26	10.2603	1448.5	19447
Unknown RT:10.37	10.3658	1457.9	44617
Unknown RT:10.43	10.4359	1464.1	1743
Unknown RT:10.46	10.4559	1465.9	64633
Unknown RT:10.56	10.5566	1474.8	107840
Unknown RT:13.40	10.5749	1476.5	18212
Unknown: ScP_M_1_No_D RT:10.60	10.5965	1478.4	17589
Unknown RT:10.71	10.7055	1488.1	27184
Silanamine	10.841	1500.1	14155
Unknown RT:11.10	11.0977	1527.6	83237
Similar to Tertbutylphenol RT:11.2	11.1959	1538.1	523882.5
Unknown RT:11.28	11.2782	1546.9	98424
Unknown RT:11.38	11.3793	1557.7	111028
Unknown RT:11.49	11.4924	1569.8	11403
Unknown RT:11.59	11.592	1580.4	27401
Dodecanoic acid TMS	12.3225	1658.5	15939
Unknown RT:12.42	12.4181	1668.7	75076
Unknown RT:12.79	12.7909	1708.6	26676
Unknown RT:13.06	13.062	1737.6	64977
Unknown RT:13.15	13.145	1746.5	40132
Unknown RT:13.20	13.1951	1751.8	47849
Unknown RT:13.283	13.2825	1761.2	38927
Unknown RT:13.32	13.3231	1765.5	122705
Unknown RT:13.51	13.3976	1773.5	142681
Unknown RT:13.46	13.4547	1779.6	11698
Unknown RT:15.01	13.5668	1791.6	6147
Unknown RT:13.61	13.6118	1796.4	68969
Unknown RT:13.76	13.7553	1811.7	2588
Unknown RT:14.76	14.7564	1922.2	229053

Unknown RT:14.94	14.945	1946	5975
Hexadecanoic acid TMS	15.7499	2047.7	3543529
Heptadecanoic acid TMS	16.5057	2143.1	100858
Octadecanoic acid TMS	17.2506	2243	8267357
Unknown RT:18.63	18.6279	2444.1	66502.5
Unknown RT:18.99	18.986	2496.3	171614
Unknown High MW RT:21.3	21.2651	2865.4	441333
Unknown RT:21.27	21.2656	2865.5	259353
Unknown RT:22.89	22.8897	3140.2	2069168
Unknown RT:28.07	28.0598	4014.6	132886.5

3.4.1.11 Identification of Unknown Metabolites

3.4.1.11.1 Identification of Unknown RT 11.58 as 2-Phosphoglycolic acid (3TMS)

Table 3-3: Predicted fragmentation of 2-Phosphoglycolic acid (3TMS)

Ion	Mass	Loss	Loss Identity
Mass	372	0	-
Methyl loss	357	15	Methyl
O-silyl and =O loss	267	105	O-silyl and =O
O-Methyl (-H) phosphate 2TMS	255	117	Carboxylic TMS
O-Methyl (-H) phosphate 2TMS -=O	239	133	Carboxylic TMS and =O
Phosphate 2TMS -O	225	147	Glycolic acid (-H) TMS
O-Methyl (+H) phosphate (-O) 1TMS	167	205	Carboxylic TMS and O-silyl
O-Methyl phosphate (-O) 1TMS - methyl	151	221	Carboxylic TMS and O-silyl and methyl
2TMS OR Glycolic acid(-H) TMS	147	225	Mass -2TMS OR Phosphate 2TMS -O
1TMS	73	299	Mass -TMS

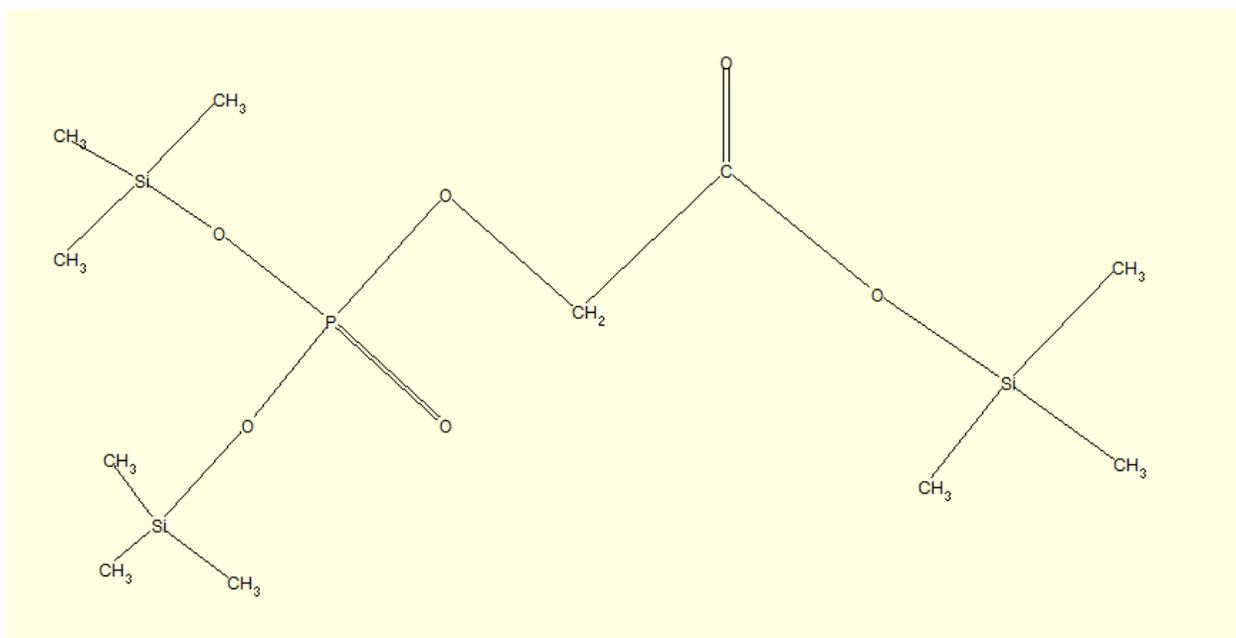


Figure 3-2: Structure of 2-phosphoglycolic acid 3TMS.

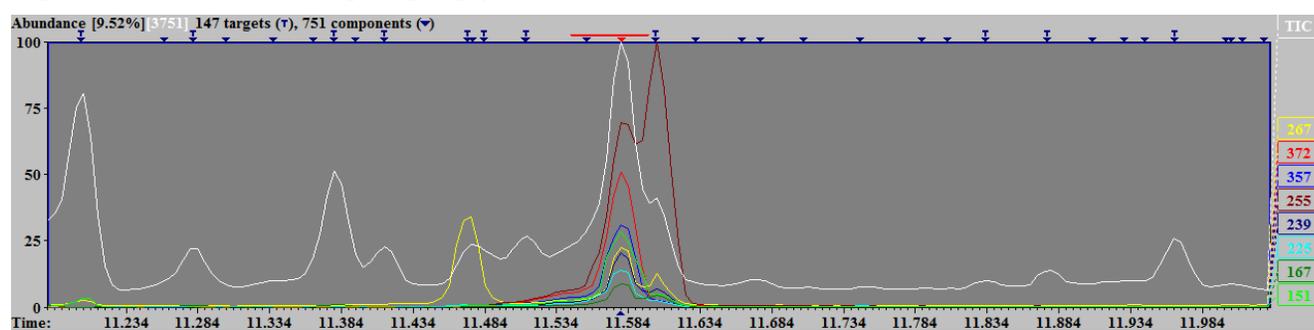


Figure 3-3: Gas chromatogram showing peak at RT:11.5787 (RI:1579). Peak is composed of ions 267 (yellow), 372 (red), 357 (blue), 255 (dark red), 239 (dark blue), 225 (cyan), 167 (dark green), 151 (light green).

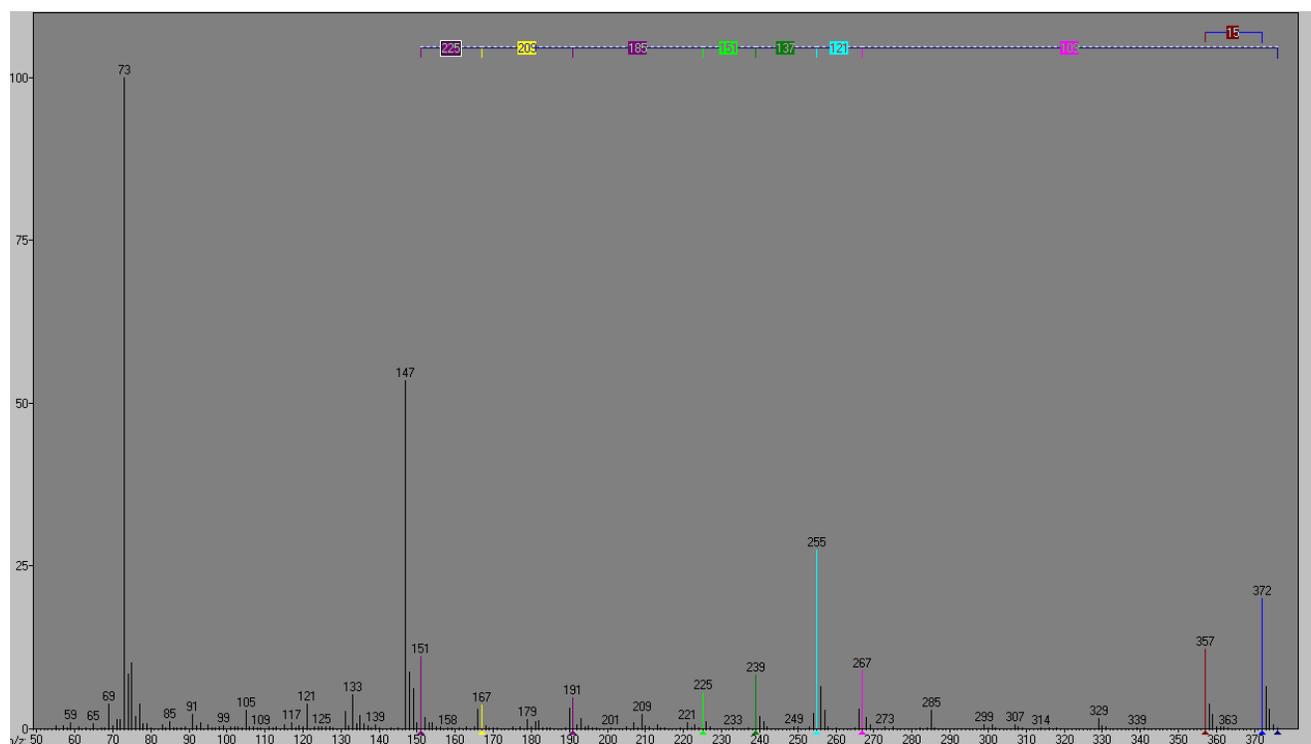


Figure 3-4: Mass spectrum of unknown RT 11.58.

Table 3-4: Functional group analysis to determine retention index of 2-phosphoglycolic acid (3TMS). Carboxyl group was interpreted as one >CO group and one -O- group.

Group	Number of instances	Value	Sum
-CH3	9	112	1008
>CH2	1	99	99
>CO	1	253	253
>Si<	3	-115	-345
>PO-	1	246	246
-O-	4	75	300
Total	x	x	1561

Table 3-5: Functional group analysis to determine retention index of 2-phosphoglycolic acid (3TMS). Carboxyl group was interpreted as one –CO-O-.

Group	Number of instances	Value	Sum
-CH3	9	112	1008
>CH2	1	99	99
-CO-O-	1	266	266
>Si<	3	-115	-345
>PO-	1	246	246
-O-	3	75	225
Total	x	x	1499

2-Phosphoglycolic acid was drawn and derivitized to 2-phosphoglycolic acid (3TMS) (Figure 3-2) and fragmentation predicted (Table 3-3). Deconvoluted, calibrated GC spectra were searched for peaks that had all 8 predicted fragmentation ions. Unknown RT:11.58 was identified as containing all predicted fragments (Figures 3.3 and 3.4). The Kovát's retention index (40) was predicted for 2-phosphoglycolic acid (3TMS) using two possible interpretations of the functional groups present (Tables 3.4 and 3.5). The mean of these two estimates was RI: 1530. This is very similar to the retention index of Unknown RT:11.58 of 1579, only differing by 49 units. As a comparison 2-hydroxyglutaric acid (3TMS) has an estimated RI of 1489 but the NIST library value is 1572, a difference of 83 units. Between this very good match and the excellent coherence between the predicted and actual fragmentation patterns unknown RT:11.58 was concluded to be positively identified as 2-phosphoglycolic acid (3TMS).

3.4.1.11.2 Identification of Unknown RT:11.37 as 2-Hydroxyomuconic semialdehyde (2TMS)

Table 3-6: Predicted fragmentation of 2-hydroxyomuconic semialdehyde(2TMS)

Ion	Mass	Loss	Loss Identity
Mass	286		
Mass –methyl	271	15	Methyl
Mass –methyl – CH=O (+H)	243	43	Methyl and CH=O (+H)
Mass –methyl –methyl –CH=O	227	59	2Methyl and CH=O
Mass –TMS –H	196	90	TMS+H
Mass –carboxylic TMS	169	117	Carboxylic TMS
Mass –carboxylic TMS –=O	153	133	Carboxylic TMS and =O
C-C=CH-O-TMS(-methyl)	111	175	O-TMS and =O and O=CH-CH=CH and methyl
CH=C-OC-O=O	85	201	2TMS O=CH-CH=CH
CH-CH=CH-CH=C-C (-H)	75	211	2 O-TMS and 2=O
2TMS	147	139	Mass -2TMS
1TMS	73	213	Mass-1TMS

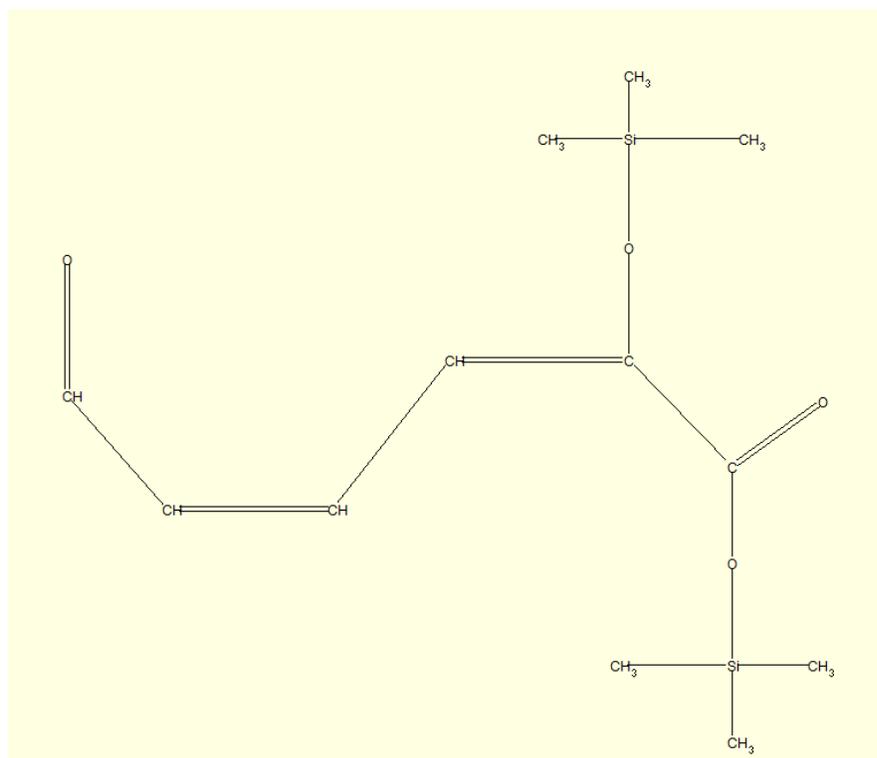


Figure 3-5: Structure of 2-hydroxyomuconic semialdehyde.

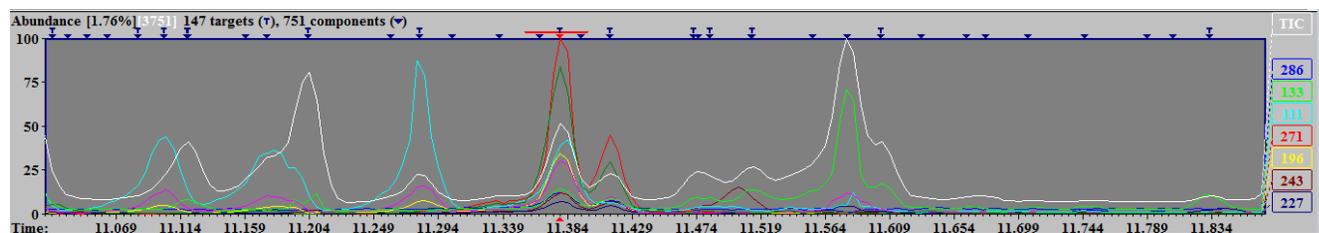


Figure 3-6: Gas chromatogram showing peak at RT:11.37 (RI:1557). Peak is composed of ions 286 (blue), 133 (light green), 111 (cyan), 271 (red), 196 (yellow), 243 (dark red), 227 (dark blue).

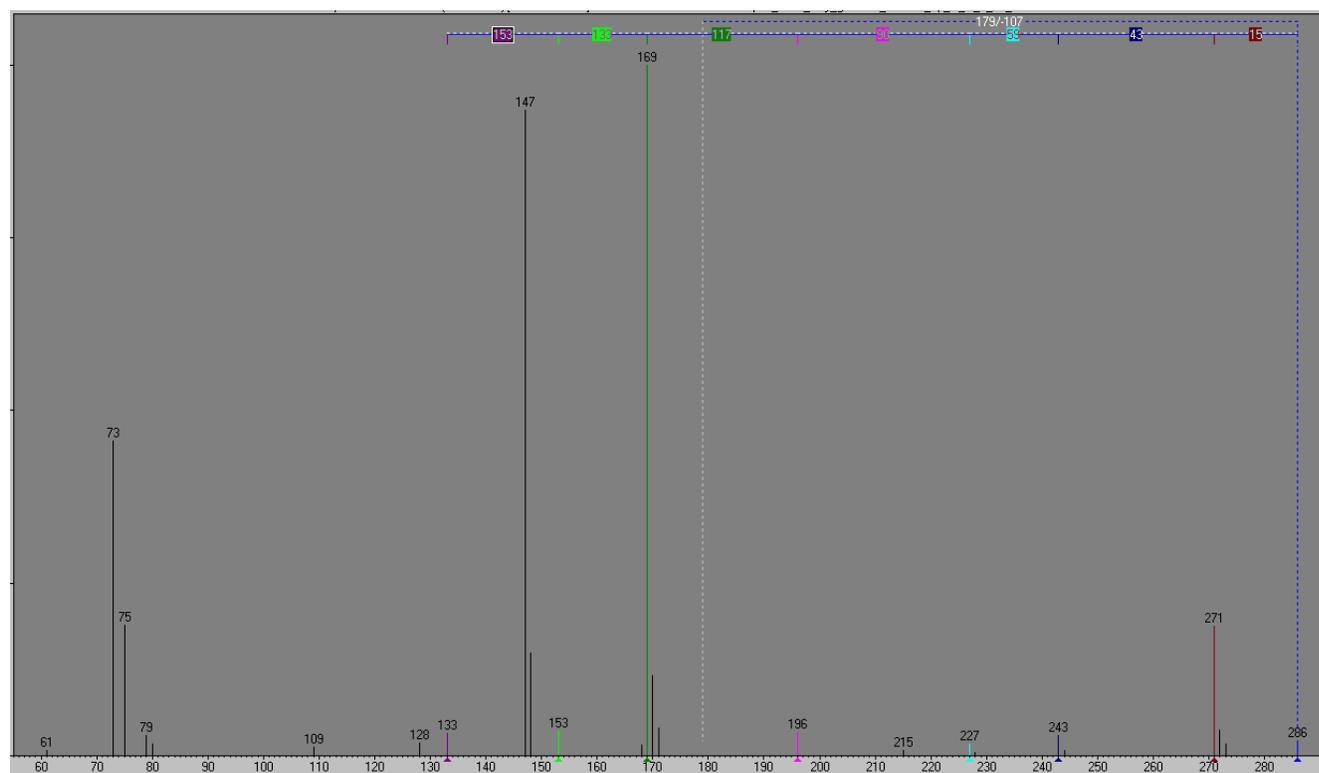


Figure 3-7: Mass spectrum of unknown RT 11.37.

Table 3-7: Functional group analysis to determine retention index of 2-hydroxymuconic semialdehyde (2TMS). Carboxyl group was interpreted as one –CO-O-.

Group	Number of instances	Value	Sum
-CH3	6	112	672
>Si<	2	-115	-230
=CH-	3	102	306
-HCO	1	235	235
=C<	1	67	67
-CO-O-	1	266	266
-O-	1	75	75
Total	x	x	1391

Table 3-8: Functional group analysis to determine retention index of 2-hydroxymuconic semialdehyde (2TMS). Carboxyl group was interpreted as one >CO group and one –O– group.

Group	Number of instances	Value	Sum
-CH3	6	112	672
>Si<	2	-115	-230
=CH-	3	102	306
-HCO	1	235	235
=C<	1	67	67
>CO	1	235	235
-O-	2	75	150
Total	x	x	1435

2-Hydroxymuconic semialdehyde was drawn and derivatized to 2-hydroxymuconic semialdehyde (2TMS) (Figure 3-5) and its fragmentation was predicted (Table 3-6). Deconvoluted, calibrated GC spectra were searched for peaks that had all 8 predicted fragmentation ions above >100. Unknown RT:11.37 was identified as containing all predicted fragments, though the adjacent, smaller peak at 11.41 also contained these ions (Figures 3.6 and 3.7). The Kovát's retention index (40) was predicted for 2-hydroxymuconic semialdehyde (2TMS) using two possible interpretations of the functional groups present (Tables 3.7 and 3.8). The mean of these two estimates was RI:1413. This is similar to the retention index of Unknown RT:11.37 of 1557, differing by 144 units. While this is not as good a match as was observed with 2-phosphoglycolic acid, it is within range of variation of the difference between predicted and actual retention indices of compounds in the NIST library. As the mass fragmentation patterns of unknown RT:11.37 and 11.41 were very similar both were treated as possible analytes of 2-hydroxymuconic semialdehyde. Still the similarity between the predicted mass spectra of 2-hydroxymuconic semialdehyde (2TMS) and unknown RT11.37 and the predicted and actual retention index led to the conclusion that unknown RT:11.37 is 2-hydroxymuconic semialdehyde (2TMS).

3.5 Results and Discussion

3.5.1 Growth of *P. pseudoalcaligenes* KF707 in the presence of metals on succinate and biphenyl

The minimum concentrations of Al and Cu that inhibited the growth of *P. pseudoalcaligenes* KF707 using succinate or biphenyl as the sole carbon source were previously determined using high-throughput microtitre plate assays (8). To confirm that these concentrations were relevant in the larger cultures needed for metabolomics, culture growth in the presence of these metals was quantified over time. Based on our prior work (8), 3mM Al and 60 μ M Cu were selected as metal concentrations that would elicit a phenotype but not inhibit growth. The higher concentration of Al was used to overcome the lack of bioavailability of Al caused by the phosphate in the medium. Phosphate was used as a buffer despite its well-characterized property of chelating aluminum (43). However, we could not use organic buffering agents such as MOPS as this would have negated the single carbon source nature of the study. To confirm that these concentrations did not inhibit growth in 250mL flasks, cultures were grown for 32 hours in the presence of either 3mM Al or 60 μ M Cu. After 24h of growth, the number of viable cells from metal-exposed cultures were found to be similar to their control counterparts for both carbon sources (Figure 3-8, A). As comparable cell densities and growth period were desirable to make comparisons between control and metal exposed cultures as similar as possible, 24 hours was thus selected as the time point for metabolomics harvest.

Culture pH at each time point was also determined. Over 32h, the pH of the culture medium from biphenyl-grown cells decreased about 1 unit for Al exposed cultures and 0.5 units for control and Cu exposed (Figure 3-8, B). The pH of succinate-grown cultures did not change in such a manner. During GC-MS metabolomic characterization, large quantities of benzoic acid were found

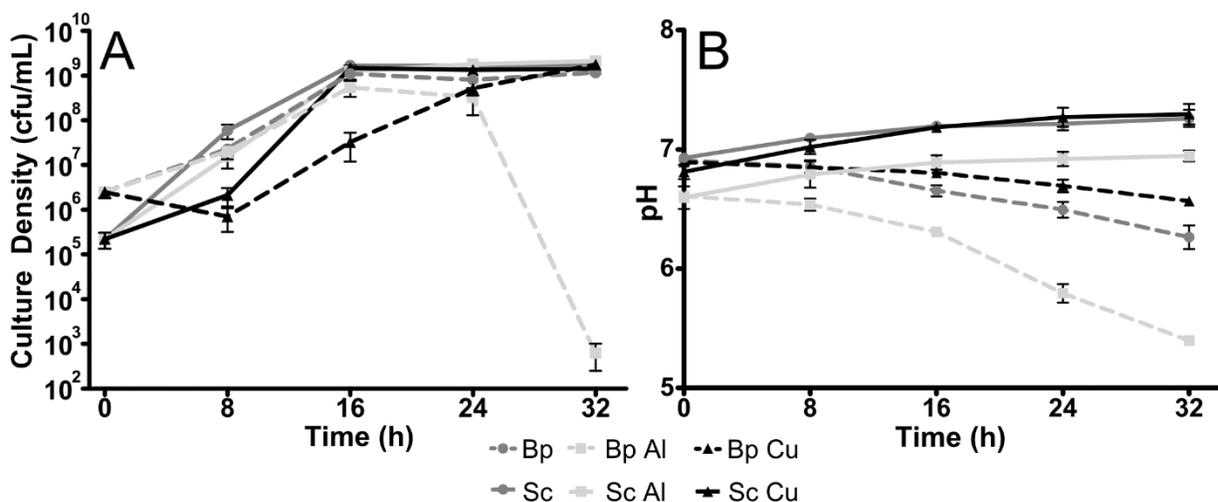


Figure 3-8: Growth (A) and culture pH (B) of *Pseudomonas pseudoalcaligenes* KF707 grown over 32 hours in minimal salts medium with either succinate (Sc) or biphenyl (Bp) as the sole carbon source, with either nothing, 3mM Al₂(SO₄)₃ (Al) or 60μM CuSO₄ (Cu). Points denote the mean of 3 biological replicates, error bars indicate SEM.

in the cells and spent media of all samples grown on biphenyl. The amounts were so great that it could not be quantified comparably to other metabolites that were detected as the GC-MS detector was saturated. The only other metabolite that was saturated was phosphate, the buffering agent from the medium. Benzoate is produced during catabolism of biphenyl (44), making it an unavoidable byproduct. The decrease in pH observed in biphenyl grown cultures (Figure 1B) was likely due to this acid being produced. As pH decreases from neutral, Al is known to increase in solubility, which is considered one of the main problems of acid rain (45). The biphenyl-grown Al³⁺ exposed cultures' viability decreased sharply from 24 to 32 hours as the pH dropped from about 6 to 5.5. This low of a pH would have increased the bioavailability of Al, evidently to the point of lethal toxicity. As Al³⁺ is prevalent throughout the Earth's crust (46), the degradation of organic compounds in sites with low pH would be expected to be more difficult as increased bioavailability of Al would result in greater stress to bacteria.

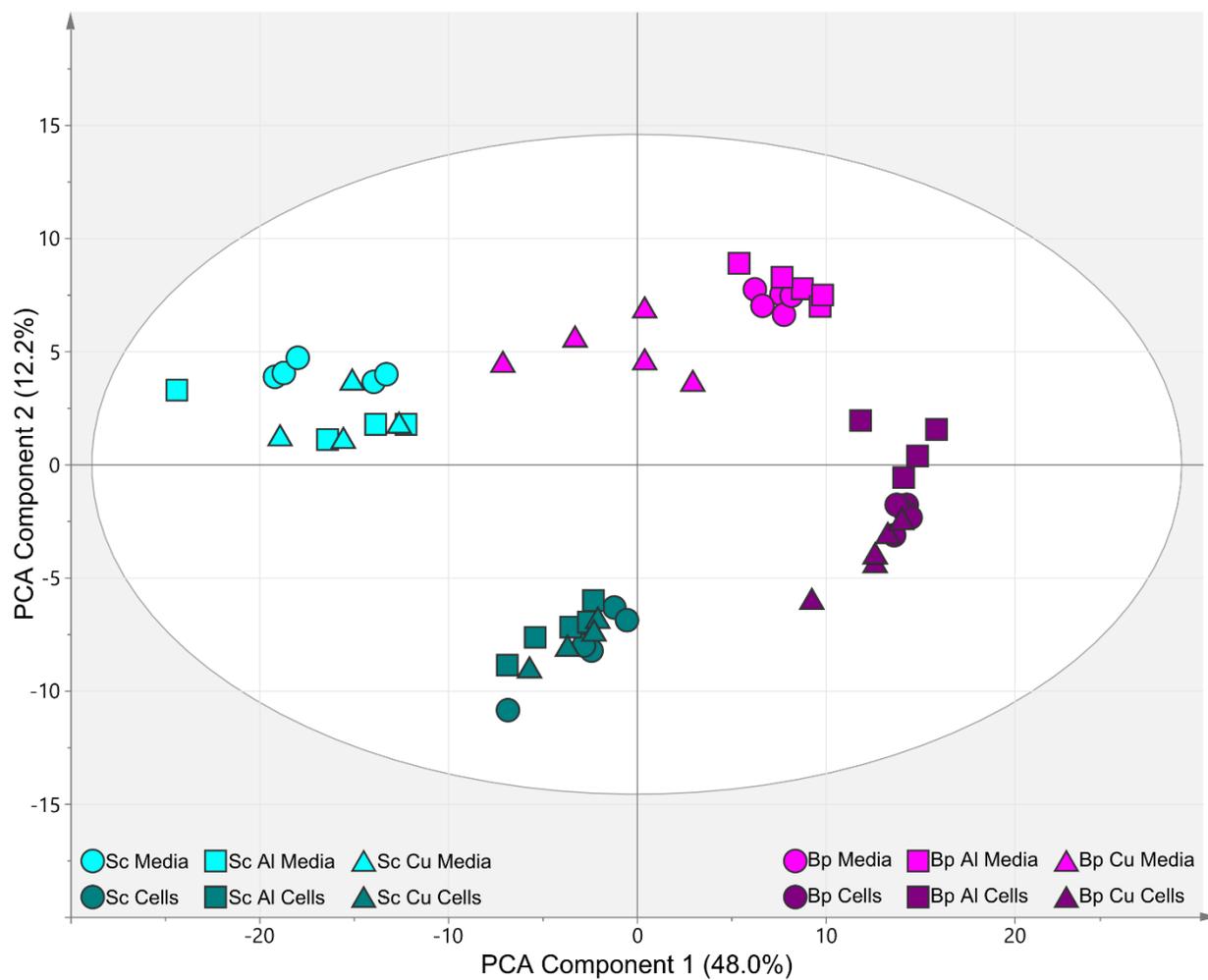


Figure 3-9: Principal component analysis (PCA) scores plot of GC-MS metabolite profiles of cells and spent media from cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl (Bp, purple) or succinate (Sc, teal) as the sole carbon source and exposed to either control (circles), 3mM Al (squares) or 60µM Cu (triangles). Results were normalized, scaled and centred before analysis.

3.5.2 Metabolomic characterization of cultures

To understand how exposure to Al or Cu affected bacterial cultures, untreated samples were compared to those grown in the presence of each metal. GC-MS metabolic profiles were obtained from cells and spent media from cultures grown either on succinate or biphenyl as the sole carbon source and exposed separately to each metal. After exclusion of low quality analytes and artifactual compounds derived from reactions between the derivatization agents and plastics of the sample vessels, 269 metabolites were quantified, of which 89 were identified. These data were analyzed by the unsupervised statistical techniques hierarchical clustering analysis (HCL) and principal component analysis (PCA). Inspection of the PCA scores plot (Figure 3-9) revealed that samples separated first by carbon source and next by sample type (cells or spent media). This model had a good R^2 (0.741, variance explained) and Q^2 (0.643, goodness of fit) (Table 3-9), indicating that close to 75% of the variation in the dataset could be explained by carbon source and sample type. These overall trends were confirmed by HCL of the raw, log-transformed data (Figure 3-10) and normalized, scaled data (Figure 3-11). Both techniques showed that after being separated into groups of the same sample type and carbon source, there were still differences between control and metal exposed samples. While these analyses demonstrated that the metabolic profiles of samples varied based on the treatments applied, discerning specific changes to metabolites based on these treatments was non-trivial. In order to better understand the relation of specific metabolites to the altered conditions, we further extended the multivariate analysis using supervised techniques.

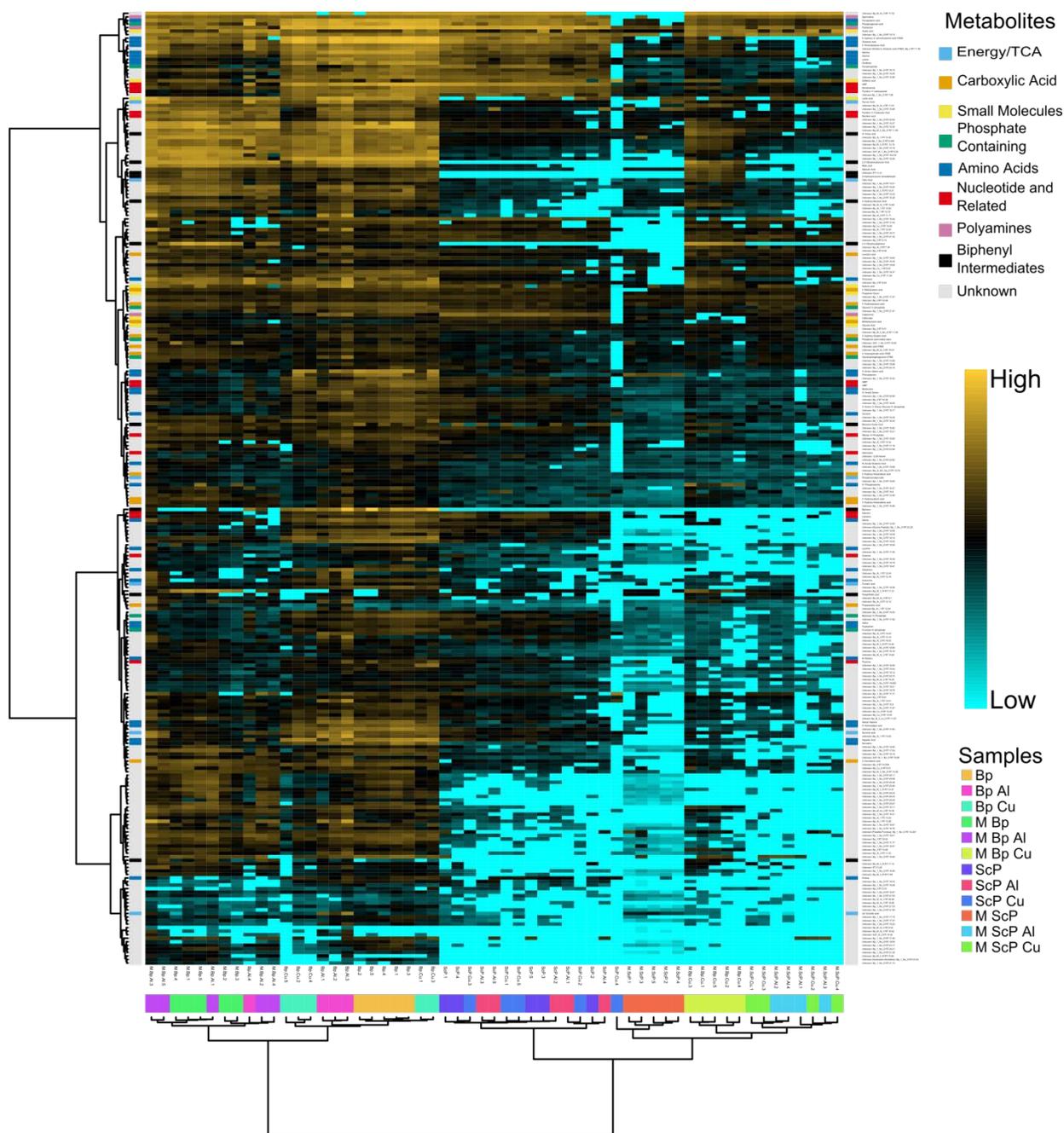


Figure 3-10: Hierarchical clustering analysis and heatmap of GC-MS quantified metabolites from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate or biphenyl as the sole carbon source under control, Aluminum or Copper exposure. Data were only log transformed. Gold indicates high concentrations of metabolite while cyan indicates low, black being average (across ALL samples) for that metabolite. Distances between samples and variables were determined using euclidean distance and clusters (both of metabolites and samples) were determined by Ward's method. The color bars at the end of the sample dendrogram indicate the sample type. Color bars at the end of the metabolite dendrogram and beside the metabolite names indicate general metabolic class.

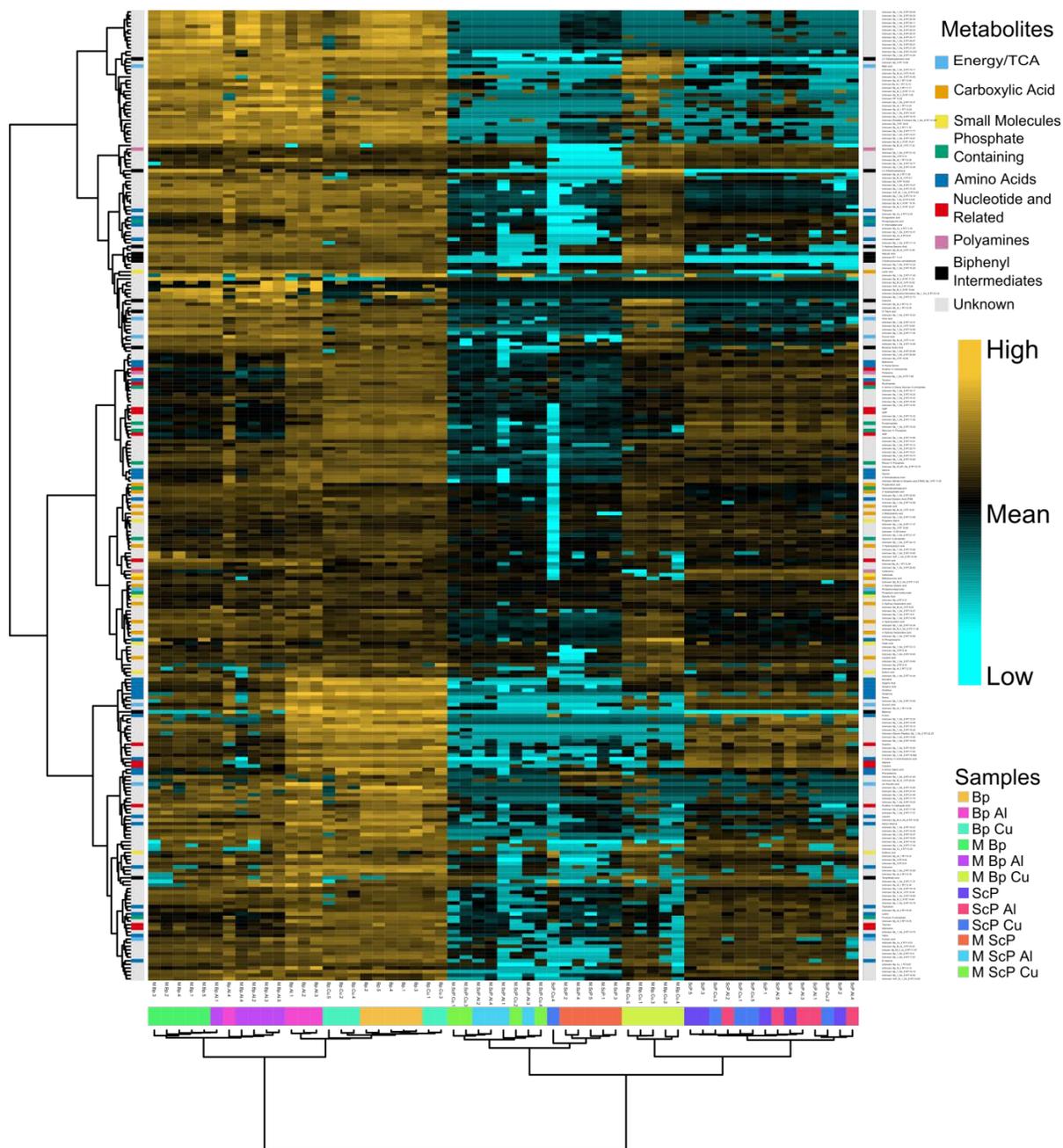


Figure 3-11: Hierarchical clustering analysis and heatmap of GC-MS quantified metabolites from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate or biphenyl as the sole carbon source under control, Aluminum or Copper exposure. Data were normalized, mean-centered and scaled to unit variance to allow comparison of metabolites with varying dynamic ranges and variances. Gold indicates high concentrations of metabolite while cyan indicates low, black being average (across ALL samples) for that metabolite. Distances between samples and variables were determined using euclidean distance and clusters (both of metabolites and samples) were determined by Ward's method. The color bars at the end of the sample dendrogram indicate the sample type. Color bars at the end of the metabolite dendrogram and beside the metabolite names indicate general metabolic class.

3.5.3 Supervised Statistical Analysis

Pairwise orthogonal partial least squares discriminant analysis (OPLS-DA) models were used to identify exactly which metabolites were being altered by metal exposure and how. These models compared just the control and one type of metal exposed sample in each carbon source for each sample type, allowing for the relative concentrations of metabolites to be compared between the control and treated samples. This eased interpretability and avoided confounding influences caused by carbon source and sample type. Examination of the statistics for each of these models revealed that they all accounted for most of the variation between samples ($R^2 \geq 82\%$) as well as predicted the vast majority of variation (all but two models $Q^2 \geq 89\%$) (Table 3-9, Figures 3.13, 3.14). Additionally, the 7-fold cross-validation analysis of variance (CV-ANOVA) p-value, which essentially indicates the probability that such a model would be generated by chance, was <0.05 for all models. As in past studies (9, 32) the pairwise OPLS-DA models showed significant differences between control and treated samples, so from these models the variable influence on projection (VIP) and $p(\text{corr})$ were extracted. These values respectively indicate the importance of a metabolite in distinguishing the sample classes (i.e. control from metal exposed) and whether it is correlated with the control or metal exposed samples (for exact details on $p(\text{corr})$, see the supplementary material). These data were subsequently used to produce shared and unique structures plots (39) for intracellular metabolites (Figure 3-15) as well as those that were found within the spent medium (Figure 3-16) to determine similarities and differences between how metabolites were altered in each carbon source.

Table 3-9: Model statistics from principal component analysis and pairwise orthogonal partial least squares discriminant analysis (OPLS-DA) models comparing normalized, centered and scaled metabolite abundances in cells and spent media from *Pseudomonas pseudoalcaligenes* KF707 grown on either succinate or biphenyl with or without aluminum or copper. Components indicates number of predictive and orthogonal components present in the model, R^2 indicates cumulative variance accounted for by the model, Q^2 cumulative variance predicted by the model and CV-ANOVA p-value was obtained from 7-fold cross-validation analysis of variance. Models with a p-value <0.05 were considered significant.

Model	Type	Components	R^2	Q^2	CV-ANOVA p-value
All	PCA	5	0.741	0.643	NA
All	OPLS-DA	7+0	0.577	0.404	<0.001
Biphenyl Al	OPLS-DA	1+1	0.993	0.931	0.014
Biphenyl Cu	OPLS-DA	1+1	0.999	0.945	0.002
Biphenyl Al Media	OPLS-DA	1+0	0.821	0.647	0.026
Biphenyl Cu Media	OPLS-DA	1+1	0.989	0.952	0.002
Succinate Al	OPLS-DA	1+1	0.996	0.895	0.012
Succinate Cu	OPLS-DA	1+0	0.944	0.667	0.037
Succinate Al Media	OPLS-DA	1+1	0.989	0.897	0.029
Succinate Cu Media	OPLS-DA	1+1	0.996	0.916	0.020

3.5.4 Pathway Enrichment Analysis

Pathway enrichment analysis was used to identify metabolic pathways that were affected under each condition. Lists of metabolites that were identified as changing significantly between control and metal exposed samples (based on their VIP) from each model were separately submitted to mBROLE (41). This tool uses the annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (42) to determine which metabolic pathways a metabolite is involved in. Pathways are assigned p-values based on the probability of enough closely connected metabolites from the same pathway being altered only by random chance. After exclusion of spurious pathways (47), many pathways were found to be affected by Al and Cu exposure in cultures of *P. pseudoalcaligenes* KF707 grown on biphenyl (Table 3-10). Fewer pathways were affected in the cultures grown on succinate. The most pathways were affected in biphenyl-grown cultures exposed to Cu, whereas succinate-grown cultures had the least number of pathways affected. This

confirmed our expectation that metal toxicity would affect cultures differently depending on the carbon source being used. The pathways identified by mBROLE were subsequently used to contextualize the meaning of changes to individual metabolites and interpret why these metabolites were altered under each condition.

Table 3-10: Metabolic pathways affected by metal toxicity in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either biphenyl (Bp) or succinate (Sc) as the sole carbon source in the presence of Al or Cu as determined by mBROLE. Presented here are modified p-values (i.e. multiple-testing corrected) indicating the probability that a pathway was affected under a particular condition. All reported pathways had an unmodified p-value <0.05 in at least one sample type. Metabolites were selected based on their VIP >0.8 as determined from OPLS-DA models comparing control to metal exposed cultures and were submitted to mBROLE. Only metabolites with a known KEGG ID were used. p-values are coloured according to significance: unmodified >0.05, not significant (black), unmodified <0.05 but modified >0.05 (grey), unmodified and modified <0.05 (white). Pathways are grouped into carbon metabolism (white), purine/pyrimidine metabolism (light grey), amino acid metabolism (grey) and sulfur containing metabolism (dark grey).

Pathway	Bp Al	Sc Al	Bp Cu	Sc Cu
Benzoate degradation (via hydroxylation)	<0.01	NA	<0.01	NA
C5-Branched dibasic acid metabolism	0.02	NA	0.06	0.05
(Glycolysis)/Gluconeogenesis	NA	NA	0.06	0.05
Pentose phosphate pathway	0.09	NA	NA	0.05
Citrate cycle (Krebs cycle)	<0.01	0.01	<0.01	<0.01
Pyruvate metabolism	NA	0.02	<0.01	<0.01
Glyoxylate/dicarboxylate metabolism	<0.01	NA	<0.01	0.02
Pantothenate/CoA-biosynthesis	NA	NA	0.06	0.05
Purine metabolism	<0.01	<0.01	0.03	0.02
Nicotinate/nicotinamide metabolism	0.03	NA	0.02	NA
beta-Alanine metabolism	0.09	0.08	<0.01	0.01
Alanine, aspartate and glutamate metabolism	<0.01	0.06	<0.01	<0.01
Arginine/proline metabolism	<0.01	NA	0.00	NA
Glycine, serine and threonine metabolism	0.01	0.03	0.03	0.09
Valine, leucine and isoleucine biosynthesis	0.01	0.07	<0.01	0.05
Cysteine/methionine metabolism	0.01	0.04	0.04	0.11
Sulfur metabolism	0.04	0.04	0.03	NA

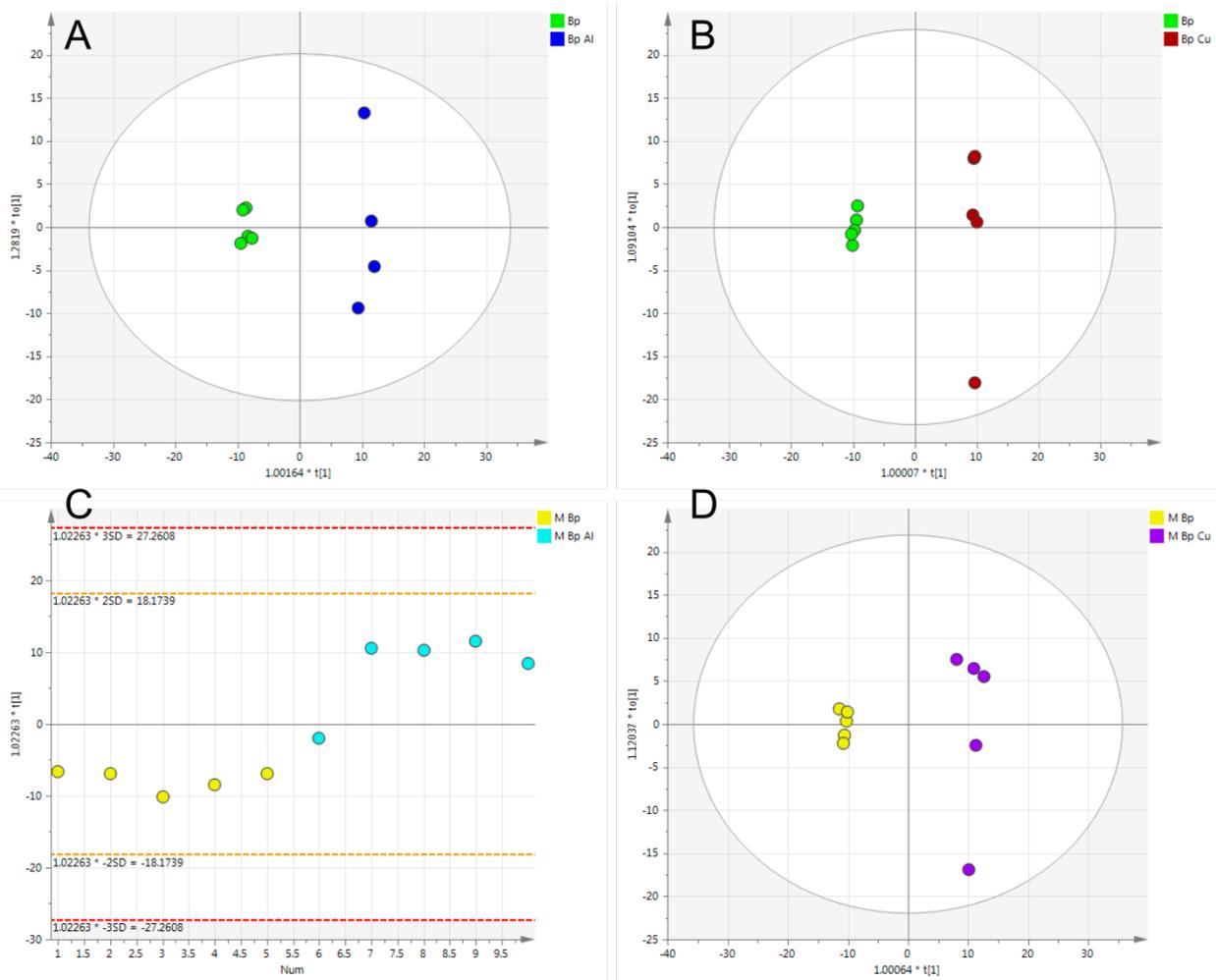


Figure 3-13: OPLS-DA Scores Plots from pairwise models comparing GC-MS metabolomic quantification of samples from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as the sole carbon source and exposed to either control, aluminum or copper.

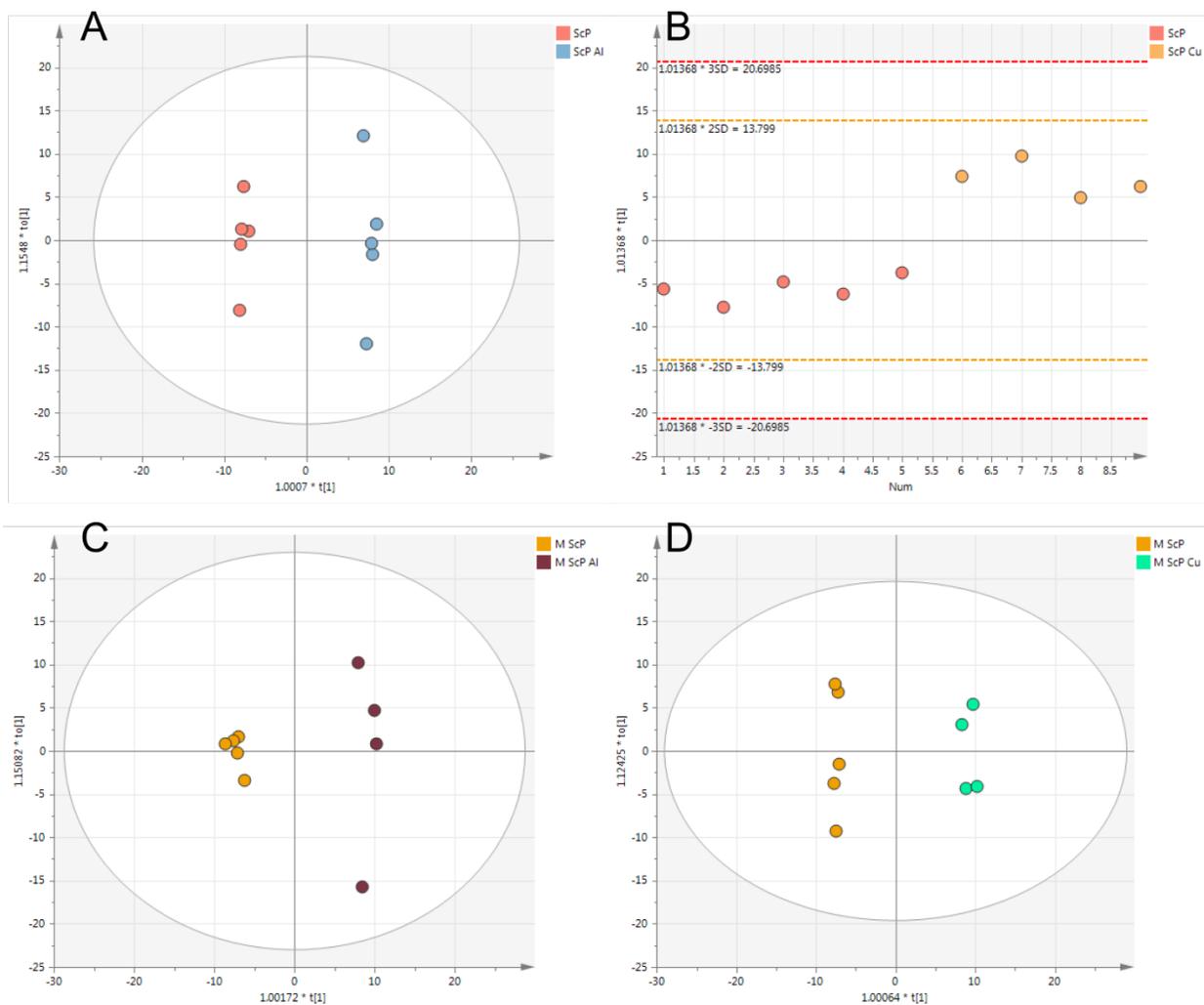


Figure 3-14: OPLS-DA Scores Plots from pairwise models comparing GC-MS metabolomic quantification of samples from cells and spent media of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on succinate as the sole carbon source and exposed to either control, Aluminum or copper.

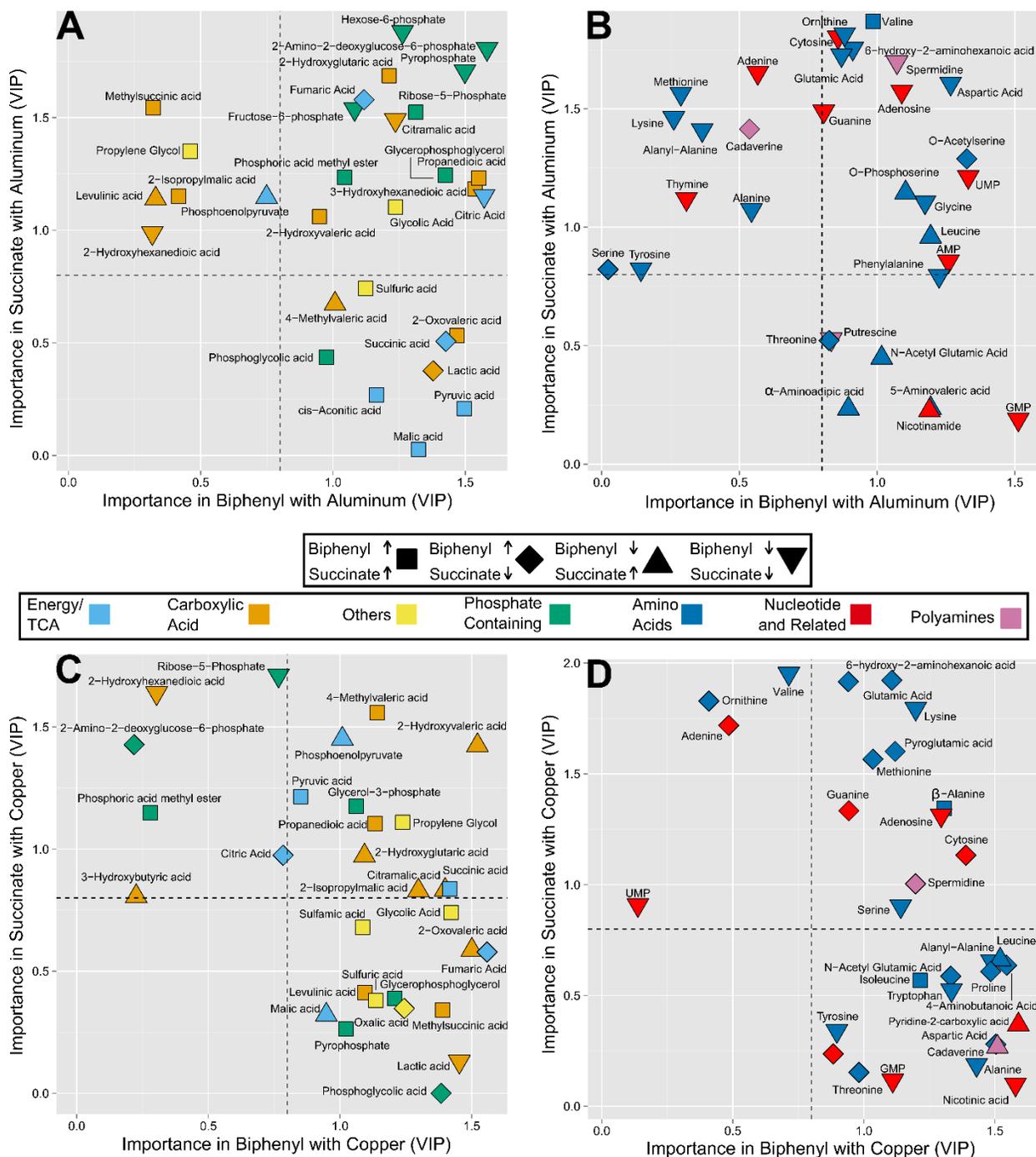


Figure 3-15: Shared and unique structures plots showing comparison of changes to intracellular metabolites caused by Al (A,B) and Cu (C,D) in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either succinate or bipheryl as the sole carbon source. Coordinates were determined by the variable influence on projection (VIP) of each metabolite, as obtained from OPLS-DA models comparing control and metal exposed samples for either succinate (y-axes) or bipheryl (x-axes). Metabolites with a $VIP \geq 0.8$ (dashed lines) indicate a significant change occurred in the metal exposed samples, those that were below in both cases were omitted. The

association of each metabolite with control or metal exposed samples was determined using $p(\text{corr})$, which indicates the degree of correlation of the metabolite with a sample type. Shapes were assigned that indicate how the metabolite was altered by metal exposure: increased in both biphenyl and succinate (squares), increased in biphenyl but decreased in succinate (diamonds), increased in succinate but decreased in biphenyl (triangle up) and decreased in both (triangle down).

3.5.5 Toxicity effects of aluminum

Aluminum exposure caused similar alterations to many phosphate containing metabolites in cells grown on both carbon sources (Figure 3-15, A, green symbols). Phosphate is a strong chelator of Al (43). In *Rhizobium* species, increased production of extracellular polymeric substances was correlated with increased tolerance to Al (48). The decrease of phosphorylated sugars observed here could indicate their use in generating EPS with functional groups for binding Al, in a manner similar to our past observation that metabolites involved in EPS production were increased in biofilm cultures of *P. fluorescens* exposed to Cu (9). Alternatively, the sugar-phosphates could be being used in lipopolysaccharide (LPS) synthesis as these outer membrane molecules could prevent Al entry into the cell by chelation (49). Our results here indicate that phosphate containing EPS or LPS mediated protection of cells from metal stress may be used both by planktonic and biofilm cultures of *Pseudomonas*.

Ribose-5-phosphate (R5P) was increased in both carbon sources under Al stress. When exposed to Al, *P. fluorescens* was previously observed to increase NADPH production, partially via overexpression of glucose-6-phosphate (G6P) dehydrogenase (50). This enzyme catalyzes the first step in the pentose-phosphate-pathway (PPP) of which the oxidative portion uses ATP to generate NADPH and ends with R5P (51). The PPP was identified by mBROLE, but was not considered significant (Table 3-10). Still, the observed accumulation of R5P, and depletion of fructose-6-phosphate and hexose-6-phosphate (representative of G6P, see supplementary material)

thus suggests that the oxidative portion of the PPP was being used to generate NADPH in response to oxidative stress being caused by Al. Conversely to R5P, malic acid and pyruvic acid were only accumulated in biphenyl-grown cultures (Figure 3-15, A, light blue symbols). These two metabolites were previously observed to be increased when *P. fluorescens* was subject to oxidative stress from menadione as part of a metabolic network aimed at converting NADH to NADPH (52). In this network, pyruvate was increased to generate oxaloacetic acid, which was converted to malic acid in order to oxidize NADH to NAD⁺ and the malic acid was cleaved to produce pyruvic acid and reduce NADP to NADPH. The metabolomic results found here indicate that a similar metabolic network was thus likely active in *P. pseudoalcaligenes* KF707 cultures grown on biphenyl and exposed to Al, but not those grown on succinate. Additionally, pyruvate accumulation could be due to the ability of pyruvate to react and detoxify hydrogen peroxide (53), making its accumulation a potentially useful anti-oxidant strategy. Pyruvate and malic acid were not affected by Al stress when cultures were grown on succinate, implying that the PPP sufficient for anti-oxidant production under these conditions.

Glycolic acid was increased in both carbon sources (Figure 3-15, A, yellow square). This was an unexpected metabolite as it is produced by very few metabolic reactions (42). One reaction is the dephosphorylation of phosphoglycolic acid. As this metabolite was not present in the libraries used for identification, it was manually identified from the unknown metabolites (See Supplementary Material for details). Thus it was observed that 2-phosphoglycolic acid was increased only in biphenyl-grown cultures (Figure 3-15A, green square), which was surprising as this metabolite also has few precursors and is associated generally with carbon fixation (54). More pertinently, when hydroxyl radicals react with the 4' carbon of a ribose moiety of DNA, repair of this oxidative damage results in the production of 2-phosphoglycolic acid (55). Subsequent

cleavage by phosphoglycolate phosphatase allows the salvage of the phosphate and two-carbon glycolic acid (56). Thus the accumulation of these two unexpected metabolites implies that Al is exerting toxicity by oxidatively damaging DNA, especially under biphenyl degrading conditions. Al has been well characterized as a pro-oxidant (57) while normal aerobic metabolism produces the reactive oxygen species (ROS) superoxide (O_2^*), peroxide (H_2O_2) and hydroxyl radicals (OH^*) by incidental reactions between molecular oxygen (O_2) and electron transport chain components (58). It has been hypothesized that hydrated Aluminum complexes stabilize superoxide radicals (which has received recent support (59)), and this complex can then reduce Fe(III) to Fe(II), regenerating the active Al(III)-superoxide complex. Fe(II) undergoes the Fenton reaction with H_2O_2 generated from aerobic metabolism to produce 2 OH^* radicals (60). These radicals could then go on to react with DNA, causing the aforementioned accumulation of metabolites. These mechanisms were likely active and responsible for DNA damage which was repaired to produce 2-phosphoglycolic acid. *P. pseudoalcaligenes* KF707 possess a phosphoglycolic acid phosphatase, however, based on a BLAST search it surprisingly does not have any of the genes encoding for any subunits of glycolate oxidase (16). This explains the accumulation of glycolic acid as it was generated from oxidative DNA damage but cannot be re-assimilated into central carbon metabolism. Further indicating that oxidative stress caused DNA damage, nucleobases and nucleotides were decreased in response to Al, the specifics depending on carbon source (Figure 3-15, B, red symbols) and purine metabolism was implicated by mBROLE (Table 3-10). In addition to oxidative stress that can damage the ribose moiety, the bases of DNA can be affected by ROS (61). Repairing this damage requires all four nucleotides, though only AMP and GMP were detected. The individual bases were also detected and decreased, implying that they were being used up generating nucleotides for use in repairing DNA.

Further similarities were observed between *P. pseudoalcaligenes* KF707 and *P. fluorescens* exposed to Al. Under Al stress *P. fluorescens* also modifies its Krebs cycle to produce less NADH and more NADPH by using the glyoxylate cycle to shunt carbon from isocitrate to succinyl-CoA (62). This metabolic pathway produces both oxalate and glyoxylate, of which only the former was detected in this experiment. As in *P. fluorescens*, oxalate may have been secreted to chelate Al, but such an increase in secretion was only detected in succinate grown cultures (Figure 6A). Alternatively, this pathway may not have been a viable option in *P. pseudoalcaligenes* growing on biphenyl as the accumulation of *cis*-Aconitate acid (Figure 3-15, A) indicates that aconitase was dysfunctional. *cis*-Aconitate is an unexpected metabolite as it is only an intermediate in the isomerization of citrate to isocitrate. Given that aconitase has a [4Fe-4S] cluster in its active site, which is sensitive to decomposition by oxidative attack it is likely that ROS affected the function of aconitase, a phenomenon which has previously been observed in *P. fluorescens* (63). Other metals have been found to cause similar stress reactions in *Pseudomonas*. Exposure to high concentrations of zinc caused a shift in ATP production from oxidative to substrate-level phosphorylation and a simultaneous decrease of NADH and increase of NADPH production (64). Vanadium toxicity was also linked to the Krebs cycle as mutations to the *idh* (coding for isocitrate dehydrogenase) and *acnD* (coding for an aconitase) genes increased resistance to this metal, presumably due to a change in expression to less metal-sensitive isozymes (65). Overall our results indicate that the metabolic changes in *P. pseudoalcaligenes* KF707 were similar to those observed in *P. fluorescens*, but more pronounced when growing on biphenyl indicating that growth on this carbon source exacerbates stress caused by Al.

o-Acetylserine is an intermediate in cysteine biosynthesis, and was increased with biphenyl but decreased with succinate (Figure 3-15, B). Apart from being the assimilation point of inorganic

sulfur and being used to synthesize all other sulfur-containing metabolites, cysteine is the amino acid that allows for the formation of disulfide bonds and iron-sulfur clusters in proteins as well as serving as an intermediate in the biosynthesis of glutathione, the main antioxidant within the cell (66). While cysteine could not be quantified, the only role of o-acetylserine is in the biosynthesis of cysteine. Cysteine, methionine and sulfur metabolism were implicated by mBROLE (Table 3-10) in both carbon sources, implying that Al toxicity caused alterations to anti-oxidant production pathways.

All three polyamines that were detected, spermidine, putrescine and cadaverine, have been implicated in oxidative stress resistance in *E. coli* as they are able to scavenge free radicals (67, 68). All of these compounds were decreased under Al exposure, cadaverine being specific to succinate and putrescine specific to biphenyl (Figure 3-15, B, light purple symbols). Al induced oxidative stress could be expected to cause polyamine levels to increase, but since their free levels are normally very low as most cellular polyamines are complexed with nucleic acids (67) an increase in ROS would rapidly deplete these free levels. Polyamines also induced the expression of acid resistance genes in *E. coli* that resulted in the secretion of 4-aminobutyric acid (69), which was secreted more in response to Al stress in biphenyl-grown cultures (Figure 3-16A). In *E. coli* this secretion was due to the action of a glutamate/4-aminobutyric acid antiporter, of which *P. pseudoalcaligenes* KF707 does not possess a homologue based on a BLAST search (16). These cultures experienced a decrease in pH due to benzoic acid production, as well as a decrease in intracellular glutamate, but as the growth medium did not contain exogenous glutamate a system comparable to *E. coli* would not have functioned anyway. *P. pseudoalcaligenes* could still have been using this secreted 4-aminobutyric acid as a proton sink similarly to *E. coli* (68). In succinate-grown cells exposed to Al, secretion of spermidine increased (Figure 3-16). Spermidine was found

to be associated with the outer membrane of *P. aeruginosa* and protected against exogenous oxidative stress, though the role of secreted spermidine was less clear (70). In the present study the spermidine accumulated in the spent medium was likely derived from spermidine being exported for such a purpose. Intracellular putrescine has also been associated with increasing resistance to oxidative stress in *Burkholderia cenocepacia* (71). Putrescine was decreased intracellularly in biphenyl-grown cells exposed to Al (Figure 3-15, B) implicating it as an additional anti-oxidant polyamine. Polyamines were also implicated in resistance to oxidative stress induced by chromium in a number of environmental isolates (72). Ornithine was decreased in Al exposed cells grown on both carbon sources (Figure 3-15, B). Decreased levels of this amino acid precursor to putrescine (which is the precursor to spermidine) and the above further supports the use of polyamines to assist with resistance to oxidative stress caused by Al in *P. pseudoalcaligenes* KF707.

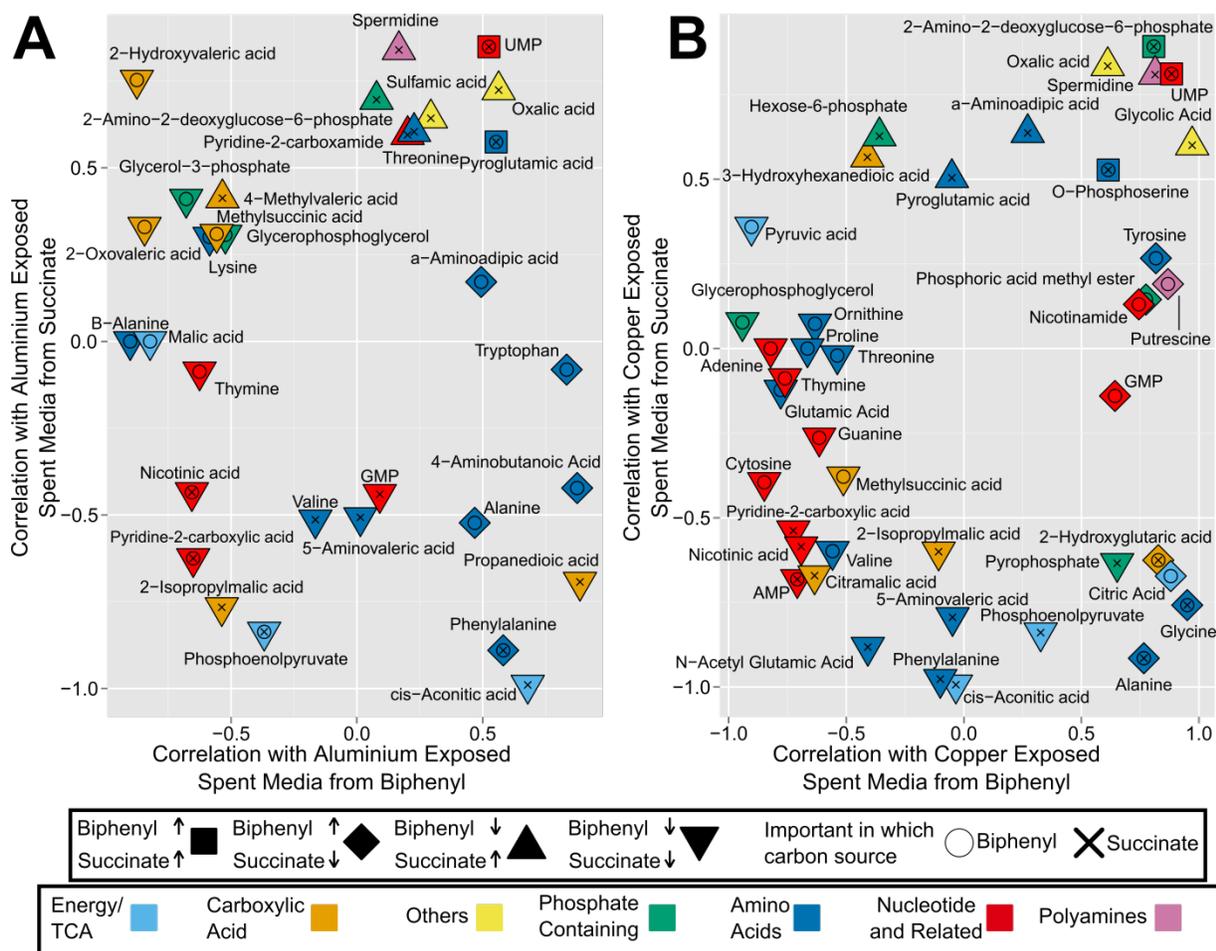


Figure 3-16: Shared and unique structures plots comparing extracellular metabolic changes of identified metabolites caused by Aluminum (A) and Copper (B) in cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on either succinate or biphenyl as the sole carbon source. The data plotted here were derived from pairwise OPLS-DA models that compared control to metal exposed samples for each carbon source. Coordinates were determined by the $p(\text{corr})$ of each metabolite from each model, only those with a $\text{VIP} > 0.8$ in at least one media type were plotted. Metabolites with a $\text{VIP} > 0.8$ in biphenyl models are marked with a circle, in succinate an X. The larger the $p(\text{corr})$ the greater the correlation with metal exposure for that media type. To identify metabolites that were being secreted, data from the OPLS-DA models comparing the control and metal exposed cells samples for each media type were also used. Metabolites were determined to be secreted if the VIP was above 0.8 in the spent media models and either the VIP was below 0.8 in the cells models or the $p(\text{corr})$ was opposite what was observed in the spent media. I.E. if a metabolite was found to change significantly in both the cells and spent media and was correlated with the metal exposed samples in the media and the control in the cells, this metabolite was secreted in response to metal exposure. Conversely, if the opposite was true, i.e. the metabolite was correlated with the control samples in the spent media and the metal exposed cells, this metabolite's secretion decreased in response to metal exposure. The association of each metabolite with control or metal exposed cell samples was determined using $p(\text{corr})$, which indicates the

degree of correlation of the metabolite with a sample type. Actual values were not used, only the directionality (i.e. above or below zero) as metabolites with high (>0.8) VIP scores tended to have high p(corr) absolute values. Using these values metabolites were assigned a shape that indicates how secretion of the metabolite was altered by metal exposure: squares were secreted more in both biphenyl and succinate, diamonds were secreted more in biphenyl but less in succinate, upwards pointed triangles were secreted more in succinate but less in biphenyl and downwards pointed triangles were secreted less in both in response to metal exposure. Metabolites were broadly classified according to chemical structure and/or biological function: those from the TCA cycle and energy producing pathways (light blue), other carboxylic acids (orange), small molecules (yellow), phosphate containing (green) (A and C), amino acids (dark blue), nucleotides and related pyridine/pyrimidine compounds (red) and polyamines (purple) (B and D). The top two panels show the effects of Aluminum (A and B) and the bottom two copper (C and D).

3.5.6 Toxicity effects of copper

Metabolic profiles indicated that Cu exerted its toxicity through some different mechanisms than Al. Surprisingly, Cu appeared to cause some metabolic changes similar to Al that indicated that Cu was exerting oxidative stress, despite this possibility being definitively ruled out in *E. coli* (31). As with Al, Cu induced the accumulation of glycolic acid and phosphoglycolic acid in biphenyl-grown cultures (Figure 3-15, C) and increased secretion of glycolic acid in succinate-grown cultures (Figure 3-16). While Cu alone may not cause oxidative stress, growth on biphenyl has the potential to generate far more ROS than succinate as complete catabolism of biphenyl requires four dioxygenases that use O₂ to activate the conjugated carbons (44). In *E. coli* Cu was no more toxic and prevented toxicity when exogenous hydrogen peroxide was added (31). In *P. pseudoalcaligenes* KF707, phosphoglycolic acid was accumulated indicating that the Cu added was not preventing oxidative stress but rather contributing to it. Transcriptional profiling of *P. aeruginosa* exposed to Cu during log phase showed changes to gene expression that were indicative of oxidative stress, which did not occur in cultures grown in the presence of Cu (73). Genes for active efflux of Cu were a main component of the response to Cu stress in both cultures indicating a non-metabolic response, especially compared to the elaborate metabolic re-

configuring observed in *P. fluorescens* responding to Al (29). While oxidative stress was not implicated in gene expression profiles of *P. aeruginosa* grown in the presence of Cu, the observed buildup of glycolic and phosphoglycolic acid in *P. pseudoalcaligenes* KF707 indicate that ROS were present in biphenyl-grown cultures and were causing the same DNA damage discussed in the Al treated cultures. Oxalic acid was accumulated intracellularly with biphenyl (Figure 3-15, C) and extracellularly with succinate (Figure 3-16). This indicates that oxidative stress was occurring under Cu exposure as the glyoxylate/oxalate shunt induced in *P. fluorescens* under oxidative stress causes oxalic acid accumulation (62). Changes to polyamines were also observed in Cu exposed cultures. Spermidine secretion was again increased under Cu stress in succinate grown cultures (Figure 3-16), however intracellular levels increased with biphenyl (Figure 3-15, D, purple symbols). Putrescine was not affected in either carbon source and instead cadaverine was decreased with biphenyl, indicating a possible shift in polyamine use for mitigating oxidative stress. The secretion of spermidine in succinate-grown cultures is of further interest as Cu surfaces have been demonstrated to kill bacteria via ROS mediated lipid peroxidation resulting in membrane destruction (74, 75). The polyamine pre-cursor ornithine was decreased only with succinate, thus together with the other observed changes this indicates that polyamines may not play as an important role in biphenyl-grown cultures under Cu stress compared to Al.

In *E. coli*, Cu disrupts Fe-S clusters of dehydratases, such as those involved in the synthesis of branched chain amino acids (30). Here we observed that metabolites associated with this pathway, citramalic and isopropylmalic acid (Figure 3-15, C, orange symbols) were affected by Cu, as well as the end-products valine, leucine and isoleucine (Figure 3-15, D, blue symbols) though not in an expected manner. In succinate-grown cultures valine was decreased while with biphenyl leucine was decreased but isoleucine increased. Isoleucine is synthesized by converting

aspartate to threonine, which is then deaminated to make 2-oxobutanoate (42). Citramalic acid was decreased in biphenyl grown cultures but was increased with succinate, as well as 2-isopropylmalic acid which is used to synthesize leucine. 2-Isopropylmalic acid is synthesized from 2-oxoisovalerate through a dehydratase-mediated reaction, which is susceptible to inhibition by Cu (30). This enzyme, dihydroxy acid-dehydratase is also used in isoleucine synthesis. As there was no consistent response to Cu stress (i.e. depletion of end-products and buildup of intermediates) no general conclusion can be made about the effect of Cu on branched-chain amino acid synthesis of *P. pseudoalcaligenes* grown on either carbon source. As these metabolites were definitely affected by Cu, quantification of enzyme activity and/or gene expression would help elucidate how Cu affected this pathway. Based on KEGG annotation (and BLAST searches), citramalic acid (both the R and S enantiomers) appears to be a dead-end metabolite in *P. pseudoalcaligenes* KF707 and cannot be synthesized from pyruvate and acetyl-CoA as in methanogens. This poses the question of why this metabolite was present in any cultures as well as why its amounts were affected by metal presence. It may have some uncharacterized role in metabolism when *Pseudomonas spp.* are grown on minimal media and so warrants further investigation, especially given *P. fluorescens*' tendency to alter its Krebs' cycle under when stressed (29).

Fumarase, which converts fumarate to malate during normal functioning of the Krebs cycle, was also found to be inhibited by Cu toxicity in *E. coli* (30). In biphenyl-grown cultures exposed to Cu fumaric acid levels increased while malic acid decreased (Figure 3-15, C, light blue symbols), indicating a similar disruption to fumarase occurred in *P. pseudoalcaligenes* KF707. In *P. fluorescens* under Al or gallium stress expression of FumA/B isozymes was decreased while FumC increased as this isozyme does not require the use of an Fe-S cluster (76). Based on BLAST searches of the KF707 genome, it does not possess a *fumC* gene and so would be unable overcome

any inhibition to FumA/B (77). Interestingly, *E. coli* grown on succinate was more susceptible to Cu toxicity acting through the inhibition of fumarase than when it was grown on glucose (30). Here it appears that *P. pseudoalcaligenes* KF707 is less susceptible when grown on succinate than when grown on biphenyl. Together these results emphasize the importance of considering how a bacterium is assimilating carbon for determining the effects of metal toxicity, a notion that is highly relevant for developing solutions for the bioremediation of co-contaminated sites.

Changes to various metabolites with possible roles as intracellular chelators were observed in Cu exposed cultures. Under Cu stress, biphenyl-grown cultures accumulated glycerol-3-phosphate, pyrophosphate and glycerophosphoglycerol (Figure 3-15, C, green symbols). This metabolite's secretion was also decreased while in succinate-grown cultures only glycerol-3-phosphate was accumulated and secretion of pyrophosphate decreased (Figure 3-16). As phosphate was in excess due to its use as the buffering agent in the medium, it could not be quantified along with the other metabolites. In *E. coli* addition of Cu induced intracellular poly-phosphate degradation and export of phosphate (78). Polyphosphate has also been implicated in resistance to oxidative stress (79). Thus combined stress from Cu and ROS generated by growing on biphenyl would likely create conflicting signals for poly-P accumulation or degradation whereas succinate-grown cells would only have incentive to degrade poly-P. This dichotomy could explain why only one phosphate containing metabolite, glycerophosphoglycerol, changed in the same fashion in succinate and biphenyl grown cells. Given the observed changes in phosphate containing metabolites further investigation into the possible role of poly-P in mitigating the combined toxicity of metals and organic pollutant catabolism is warranted, especially given the role phosphate can play in metal speciation in soils (80).

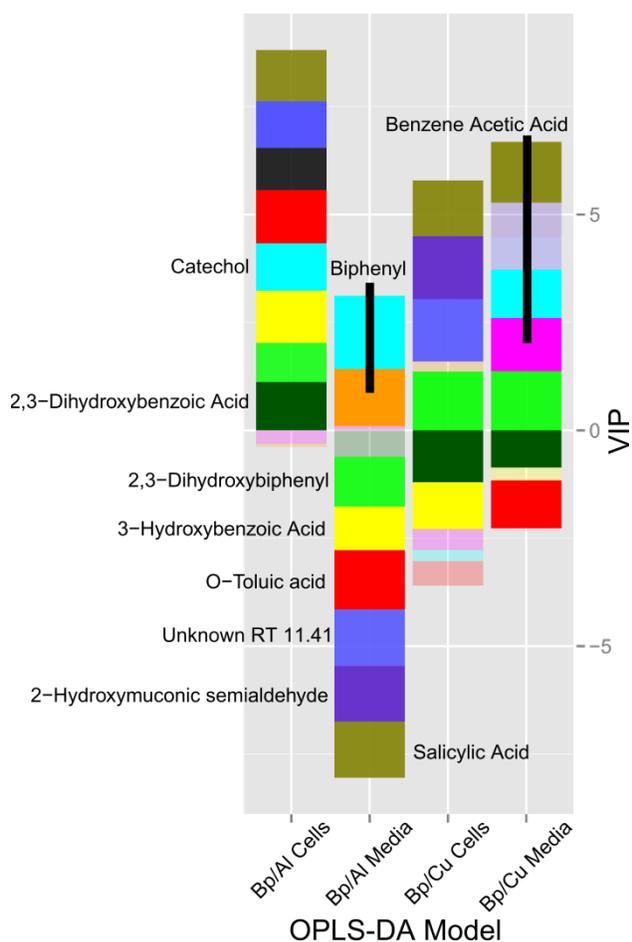


Figure 3-17: Variable Influence on Projection (VIP) of metabolites related to degradation of biphenyl from OPLS-DA models comparing control to metal exposed cells and spent media of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as a sole carbon source. Positive VIP values indicate a positive $p(\text{corr})$ value meaning an increase in metal exposed samples, negative a decrease. Transparent bars indicate a $\text{VIP} < 0.8$, not significantly changed. Metabolites categorized as relating to biphenyl degradation are: biphenyl (orange), 2,3-dihydroxybiphenyl (light green), 2,3-dihydroxybenzoic acid (dark green), 3-hydroxybenzoic acid (yellow), salicylic acid (brown), benzene acetic acid (pink), o-toluic acid (red), catechol (cyan), 2-hydroxymuconic semialdehyde (purple), unknown RT:11.41 (light purple).

Non-phosphate containing metabolites were also affected. Methionine was found to be a key intracellular chelator of Cu (I) in *E. coli*; when it was not present in the growth medium under anaerobic conditions free Cu(I) accumulated in the cytoplasm and interfered with Fe-S cluster assembly proteins (81). Here, methionine was increased in biphenyl grown cultures but decreased

with succinate potentially indicating its similar use as an intracellular chelator (Figure 3-15, D, blue symbol). Citric acid was also accumulated in the spent media of Cu exposed cultures grown on biphenyl (Figure 3-16), a phenomenon that was previously observed to occur in *P. putida* grown on glucose but not aromatic substrates (82). Interestingly, this accumulation was only extracellular and was not observed in Al exposed cultures where it would have been a logical response as citrate chelates Al (83). Oxalate also strongly chelates Al and is generated under Al exposure by *P. fluorescens* (62). Here the secretion of this metabolite was increased by both metals in succinate-grown cultures, indicating its potential role as a general metal scavenger in less-stressed cells (Figure 3-16). Glutamate has been observed to be accumulated as an osmoprotectant in *P. aeruginosa* (84). In biphenyl-grown, Cu exposed cultures glutamate was increased and its secretion was decreased (Figure 3-15, D, Figure 3-16B), indicating that Cu toxicity elicited a response similar to osmotic stress. These results thus indicate that Cu toxicity causes changes to the secretion of multiple metabolites, an interesting result given that in *P. aeruginosa* Cu exposure induced many genes involved in its efflux (73). These observations, and the aforementioned Cu surface mediated lipid peroxidation (74, 75) indicate that further investigations into Cu toxicity should consider its interactions with both the inner and outer membrane and associated proteins.

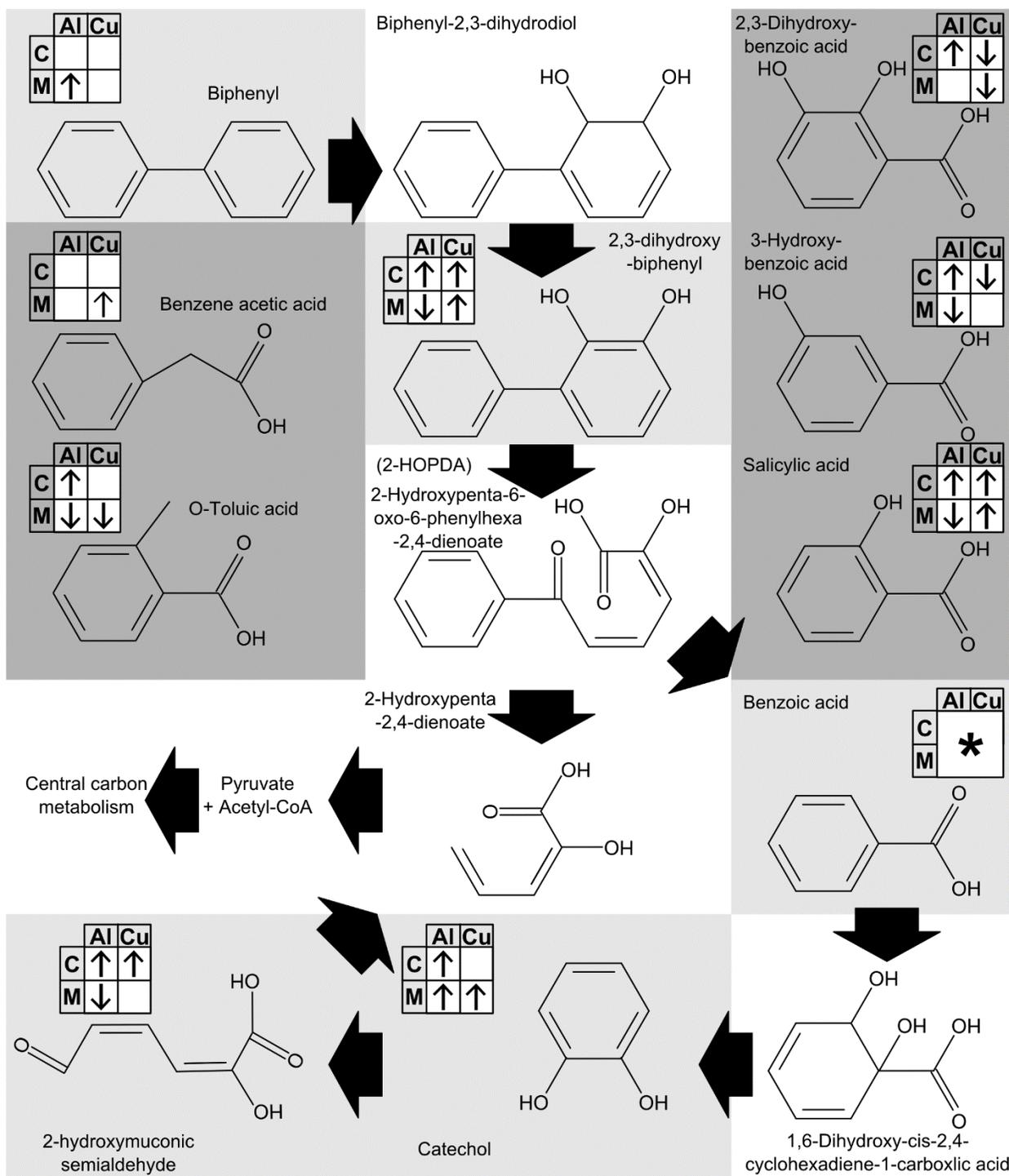


Figure 3-18: Changes to metabolites derived from biphenyl degradation in cells (C) and spent media (M) of cultures of *Pseudomonas pseudoalcaligenes* KF707 grown on biphenyl as the sole carbon source and exposed to Al or Cu. Arrows indicate whether a metabolite was increased (↑) or decreased (↓) in response to each metal. Values were derived from VIP and p(corr) from OPLS-DA models comparing control to metal exposed samples. Empty boxes mean there was no

significant change (VIP <0.8). For details, see Figure 3-16. Benzoic acid was detected but could not be accurately quantitated due to elevated concentrations resulting in detector saturation (*). Metabolite background shading denotes whether it is a canonical biphenyl degradation product and was detected (light grey) or not detected (white). Several metabolites that are not involved in biphenyl degradation but are structurally related were also detected (dark grey).

β -alanine and alanine/aspartate/glutamate metabolism were implicated by pathway enrichment analysis in Cu exposed cultures grown on either carbon source (Table 3-10). Compared to other 4-carbon molecules being used as the sole carbon source, aspartate was previously observed to greatly increase the tolerance of *P. fluorescens* to copper (8). Aspartate and β -alanine can be interconverted and are basal intermediates in many biosynthetic pathways including pantothenate and CoA biosynthesis (42), which were affected in Cu exposed cultures though they were just over the p-value cutoff (Table 3-10). Acetyl-CoA synthetase, which catalyzes the key reaction of adding the acetyl group to coenzyme A was found to be inhibited by Cu in wastewater treatment bacteria (85). β -alanine was increased in both carbon sources in response to Cu exposure and aspartate was increased in biphenyl grown cells. This could be an indication of a similar inhibition of acetyl-CoA synthetase by Cu occurring in *P. pseudoalcaligenes* KF707 and occurring more drastically when grown on biphenyl resulting in the additional accumulation of aspartate. Both β -alanine and acetyl-CoA synthesis were implicated in toxic effects of phenanthrene in *Sinorhizobium* (86) indicating possibly that Cu was exacerbating the toxicity of biphenyl. Inhibition of acetyl-CoA synthetase would be devastating to the cell and may explain our previous observation that Cu tolerance was lower in biphenyl-grown cultures than those grown on succinate (8).

The results presented here indicate that Cu toxicity elicits multiple metabolic changes that are suggestive of oxidative stress, a phenomenon that was comprehensively ruled out in *E. coli* (31). These *E. coli* cultures were grown in rich or amino acid containing medium and were resistant

to millimolar levels of Cu compared to the 60 μ M used here. Our past work showed that the tolerance of *P. pseudoalcaligenes* and *P. fluorescens* to Cu decreased 100-fold from LB medium to minimal salts medium, which was used in the present study (8). Contrary to *E. coli*, when the plant pathogen *Xanthomonas campestris* was exposed to Cu it became more susceptible to hydrogen peroxide and also upregulated ROS-detoxification genes (87). Thus it seems possible that Cu can exert toxicity through ROS-mediated mechanisms under particular conditions in some organisms.

3.5.7 Effects of metal toxicity on biphenyl metabolism

Under exposure to both metals, changes to biphenyl-degradation intermediates were observed, as well as changes to several structurally related metabolites (Figure 3-17, summarized in Figure 3-18). Also, benzoate degradation was implicated by mBROLE with both metals (Table 3-10). Biphenyl itself was only changed in the spent media from Al exposed cultures, and was increased. Biphenyl is normally insoluble in water, so this is of interest as it could indicate that bacterial activity, potentially the decrease in pH, increased solubility. No metabolites with obvious surfactant properties were identified, though they could be present within the unknown compounds. This warrants further investigation as increasing the solubility of pollutants by biosurfactants has been found to improve degradation (88). In *P. pseudoalcaligenes* KF707 biphenyl metabolism begins with dioxygenation to produce 2,3-dihydroxybiphenyl as originally characterized by Furukawa and Miyazaki (11), which was increased in Al exposed cells but decreased in the media and increased in both sample types with Cu (Figure 3-18). These increases indicate inhibition of the next step of biphenyl metabolism, ring opening by 2,3-dihydroxybiphenyl 1,2-dioxygenase. Both biphenyl and 2,3-dihydroxy biphenyl dioxygenase, are heteromultimeric Rieske-type non-heme oxygenases that contain catalytically active [2Fe-2S] clusters (89). As was

previously noted, both Al and Cu have been characterized as damaging iron-sulfur clusters, and thus these results indicate that the iron-sulfur clusters of aromatic oxygenases are likely targets of metal toxicity. After 2,3-dihydroxybiphenyl is cleaved to produce benzoate and 2-hydroxy-2,4-pentadienoate, the benzoate is dioxygenated and decarboxylated to catechol, which was accumulated under all conditions except Cu exposed cells (Figure 3-18). Benzoic acid also accumulated to an unquantifiable, GC-MS detector saturating level in all samples. This indicates that both benzoic acid 1,2-dioxygenase and catechol dioxygenase may have been inhibited by both metals. Catechol can be further metabolized either via the β -keto-adipate pathway (starting with catechol-1,2-dioxygenase) or the pyruvate/acetyl-CoA pathway (starting with catechol-2,3-dioxygenase). From genomic characterization, both pathways appear to be present in *P. pseudoalcaligenes* KF707 (16). An unknown analyte was identified as 2-hydroxymuconic semialdehyde (See supplementary material) indicating the use of catechol-2,3-dioxygenase, which was confirmed spectroscopically using a cell-free enzymatic assay (results not shown). Unexpectedly, given the accumulation of the upstream metabolites, 2-hydroxymuconic semialdehyde was also accumulated, though only in cells, it was decreased in the spent media of Al exposed samples. As all intermediates detected were increased by both metals, it could be surmised that the stress induced created a greater demand for carbon and energy, resulting in greater quantities of all intermediates. The alternative hypothesis of iron-sulfur clusters of aromatic dioxygenase being damaged by two different metals, which would mediate this damage in very different ways, makes for an intriguingly broad explanation for why metals inhibit organic pollutant catabolism. Identification of the exact mechanisms by which Al and Cu, as well as other metals, inhibit aromatic oxygenases would enable the development of possible ways to prevent this inhibition and improve bioremediation of co-contaminated sites.

Several unexpected metabolites similar to benzoic acid were identified and quantified: salicylic, 3-hydroxybenzoic, 2,3-dihydroxybenzoic, benzeneacetic and *o*-toluic acid (Figure 3-18). While these metabolites have been found in other biological systems, they are not intermediates of biphenyl degradation. Their presence here implies possible non-specific action of oxygenases from biphenyl catabolism or reactions between biphenyl degradation intermediates and ROS. The two hydroxybenzoic acids that were observed could have been formed by OH[•] reaction with benzoic acid, which is favoured at the 2-position (90) and subsequent deprotonation and ring closure. 2,3-dihydroxybenzoic acid could have been formed by an additional such reaction. Alternatively, 2,3-dihydroxybenzoic acid may have been formed by catechol 2,3-dioxygenase in the presence of excess benzoic acid substrate from biphenyl cleavage. A final possible explanation for this compound's presence is its intentional production for use in siderophores (91) for iron acquisition or as a protective measure from the metals that were added. Phenylacetic and *o*-toluic acid are the most unexpected metabolites as these cannot be formed directly from ROS interactions with biphenyl degradation intermediates and so were most likely derived from incorrect degradation of biphenyl. All of these unexpected metabolites were altered by the presence of metals, indicating that their formation had some relation to Al and Cu. As these metals appeared to be inhibiting the aromatic oxygenases, it could be possible that the interaction between the metals and the oxygenases was causing non-specific reactions to occur. The generation of dead-end metabolites due to metals thus presents a possible mechanism of enhanced toxicity.

3.5.8 Metabolic changes and implication of oxidative stress

The metabolic changes caused by Al and Cu in cultures grown on biphenyl repeatedly implied that oxidative stress was increased compared to succinate grown cultures. Many other

studies have found that organic pollutants cause oxidative stress in bacteria. 1,3-Dichloroprop-1-ene induced oxidative stress during its degradation in *P. pavonaceae* (92) which resulted in the accumulation of the antioxidant NADPH, a metabolite whose production was implicated in oxidative stress by others (29). In *Pseudomonas* sp. strain As1 overexpression of ROS detoxifying enzymes enhanced the degradation of naphthalene (93). In *B. xenovorans* LB400 growth on biphenyl induced expression of alkyl hydroperoxide reductase, which is also expressed under hydrogen peroxide treatment (94). Supplementation of this strain with the antioxidant α -tocopherol decreased the amount of time needed to degrade poly-chlorinated biphenyls in soil (95). In another *Pseudomonas* spp. (strain B4) growth on 2-chlorobiphenyl induced ROS generation (96). Other metabolomic studies have also found evidence for ROS-based toxicity: the pesticide 2,4-dichlorophenoxyacetic acid (2,4-D) was found to cause oxidative stress in *E. coli* and *R. leguminosarum* (32, 97), as well as nicotine in a *Pseudomonas* species (98) and phenanthrene in *Sinorhizobium* (86). This phenomenon is not restricted to Gram negative proteobacteria; when *Rhodococcus aetherivorans* I24 was grown in polychlorinated biphenyl contaminated soil microarray analysis showed that the expression of many stress related genes such as chaperones and oxidative stress protection genes were increased (99). As complete catabolism of biphenyl by *P. pseudoalcaligenes* KF707 requires four O₂-using dioxygenases (11, 12), aerobic metabolism causes the generation of ROS in all organisms (100), and following from the above studies, oxidative stress caused by biphenyl metabolism is thus the most likely explanation for the observed metabolic differences caused by Al and Cu.

3.6 Conclusions

Succinate and biphenyl are assimilated into central metabolism completely differently, with succinate as a component of the Krebs cycle whereas complete catabolism of biphenyl

requires ten catabolic enzymes and six intermediate steps to enter central metabolism. Here we have shown that these differences in assimilation cause the toxicity of Al and Cu to be exacerbated when biphenyl is the carbon source. This was manifested as changes to more metabolites involved in a wider variety of functions than those affected in cultures grown on succinate. As the concentrations of metals were the same for both carbon sources, we have thus demonstrated that metal toxicity is physiologically more pronounced when bacteria are using a complex aromatic substrate as their sole carbon source. Though the affected metabolites were diverse in their role within the cell, they are united by their implication in oxidative stress from studies in other systems. This makes the understanding how organic pollutants induce oxidative stress, especially in conjunction with metals, a topic of paramount importance for further laboratory and applied studies on the problem of environmental co-contamination.

3.7 References

1. **Zalasiewicz J, Williams M, Haywood A, Ellis M.** 2011. The Anthropocene: a new epoch of geological time? *Philos Trans A Math Phys Eng Sci* **369**:835–841.
2. **Darnault C, Rockne K, Stevens a, Mansoori G a, Sturchio N.** 2005. Fate of environmental pollutants. *Water Environ Res* **77**:2576–2658.
3. **Shukla KP, Singh NK, Sharma S.** 2010. Bioremediation: Developments, current practices and perspectives. *Genet Eng Biotechnol J* **2010**:1–20.
4. **Sandrin TR, Hoffman DR.** 2007. Bioremediation of organic and metal co-contaminated environments: Effects of metal toxicity, speciation, and bioavailability on biodegradation, p. 1–34. *In* Singh, S, Tripathi, R (eds.), *Environmental Bioremediation Technologies*. Springer Berlin Heidelberg.
5. **Sandrin TR, Chech AM, Maier RM.** 2000. A rhamnolipid biosurfactant reduces cadmium toxicity during naphthalene biodegradation. *Appl Environ Microbiol* **66**:4585–8.
6. **Lemire J a, Harrison JJ, Turner RJ.** 2013. Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nat Rev Microbiol* **11**:371–84.
7. **Olaniran AO, Balgobind A, Pillay B.** 2013. Bioavailability of heavy metals in soil: Impact on microbial biodegradation of organic compounds and possible improvement strategies. *Int J Mol Sci* **14**:10197–10228.
8. **Booth SC, George IFS, Zannoni D, Cappelletti M, Duggan GE, Ceri H, Turner RJ.** 2013. Effect of aluminium and copper on biofilm development of *Pseudomonas pseudoalcaligenes* KF707 and *P. fluorescens* as a function of different media compositions. *Metallomics* **5**:723–35.
9. **Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, Ceri H, Turner RJ,**

- Weljie AM.** 2011. Differences in metabolism between the biofilm and planktonic response to metal stress. *J Proteome Res* **10**:3190–3199.
10. **Booth SC, Workentine ML, Weljie AM, Turner RJ.** 2011. Metabolomics and its application to studying metal toxicity. *Metallomics* **3**:1142.
 11. **Furukawa K, Miyazaki T.** 1986. Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. *J Bacteriol* **166**:392–398.
 12. **Taira K, Hirose J, Hayashida S, Furukawa K.** 1992. Analysis of bph operon from the Polychlorinated Biphenyl-degrading strain of *pseudomonas pseudoalcaligenes* KF707 **267**:4844–4853.
 13. **Tremaroli V, Fedi S, Turner RJ, Ceri H, Zannoni D.** 2008. *Pseudomonas pseudoalcaligenes* KF707 upon biofilm formation on a polystyrene surface acquire a strong antibiotic resistance with minor changes in their tolerance to metal cations and metalloids. *Arch Microbiol* **190**:29–39.
 14. **Tremaroli V, Workentine ML, Weljie AM, Vogel HJ, Ceri H, Viti C, Tatti E, Zhang P, Hynes AP, Turner RJ, Zannoni D.** 2009. Metabolomic investigation of the bacterial response to a metal challenge. *Appl Environ Microbiol* **75**:719–728.
 15. **Tremaroli V, Suzzi CV, Fedi S, Ceri H, Zannoni D, Turner RJ.** 2010. Tolerance of *Pseudomonas pseudoalcaligenes* KF707 to metals, polychlorobiphenyls and chlorobenzoates: Effects on chemotaxis-, biofilm- and planktonic-grown cells. *FEMS Microbiol Ecol* **74**:291–301.
 16. **Triscari-Barberi T, Simone D, Calabrese FM, Attimonelli M, Hahn KR, Amoako KK, Turner RJ, Fedi S, Zannoni D.** 2012. Genome sequence of the polychlorinated-biphenyl degrader *pseudomonas pseudoalcaligenes* KF707. *J Bacteriol* **194**:4426–4427.

17. **Allen EW.** 2008. Process water treatment in Canada's oil sands industry: I. Target pollutants and treatment objectives. *J Environ Eng Sci* **7**:123–138.
18. **Annicchiarico C, Buonocore M, Cardellicchio N, Di Leo a., Giandomenico S, Spada L.** 2011. PCBs, PAHs and metal contamination and quality index in marine sediments of the Taranto Gulf. *Chem Ecol* **27**:21–32.
19. **Renzi M, Perra G, Guerranti C, Mariottini M, Baroni D, Volterrani M, Graziosi M, Specchiulli a, Focardi S.** 2009. Assessment of environmental pollutants in ten southern Italy harbor sediments. *Toxicol Ind Health* **25**:351–363.
20. **Burgess RM, Terletskaia A V., Milyukin M V., Povolotskii M, Demchenko VY, Bogoslavskaya T a., Topkin Y V., Vorobyova T V., Petrov AN, Lyashenko A, Ho KT.** 2009. Concentration and distribution of hydrophobic organic contaminants and metals in the estuaries of Ukraine. *Mar Pollut Bull* **58**:1103–1115.
21. **Jartun M, Pettersen A.** 2010. Contaminants in urban runoff to Norwegian fjords. *J Soils Sediments* **10**:155–161.
22. **Hassanvand MS, Naddafi K, Faridi S, Nabizadeh R, Sowlat MH, Momeniha F, Gholampour A, Arhami M, Kashani H, Zare A, Niazi S, Rastkari N, Nazmara S, Ghani M, Yunesian M.** 2015. Characterization of PAHs and metals in indoor/outdoor PM10/PM2.5/PM1 in a retirement home and a school dormitory. *Sci Total Environ* **527–528**:100–10.
23. **Fornalczyk A, Willner J, Francuz K, Cebulski J.** 2013. E-waste as a source of valuable metals. *Arch Mater Sci Eng* **63**:87–92.
24. **Itai T, Otsuka M, Asante KA, Muto M, Opoku-Ankomah Y, Ansa-Asare OD, Tanabe S.** 2014. Variation and distribution of metals and metalloids in soil/ash mixtures

- from Agbogbloshie e-waste recycling site in Accra, Ghana. *Sci Total Environ* **470–471**:707–16.
25. **Robinson BH.** 2009. E-waste: an assessment of global production and environmental impacts. *Sci Total Environ* **408**:183–91.
 26. **Liu M, Huang B, Bi X, Ren Z, Sheng G, Fu J.** 2013. Heavy metals and organic compounds contamination in soil from an e-waste region in South China. *Environ Sci Process Impacts* **15**:919–29.
 27. **Pradhan JK, Kumar S.** 2014. Informal e-waste recycling: environmental risk assessment of heavy metal contamination in Mandoli industrial area, Delhi, India. *Environ Sci Pollut Res Int* **21**:7913–28.
 28. **Lemire J, Mailloux R, Auger C, Whalen D, Appanna VD.** 2010. *Pseudomonas fluorescens* orchestrates a fine metabolic-balancing act to counter aluminium toxicity. *Environ Microbiol* **12**:1384–1390.
 29. **Mailloux RJ, Lemire J, Appanna VD.** 2011. Metabolic networks to combat oxidative stress in *Pseudomonas fluorescens*. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **99**:433–442.
 30. **Macomber L, Imlay J a.** 2009. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc Natl Acad Sci U S A* **106**:8344–8349.
 31. **Macomber L, Rensing C, Imlay J a.** 2007. Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. *J Bacteriol* **189**:1616–1626.
 32. **Bhat S V, Booth SC, Vantomme EA, Afroj S, Yost CK, Dahms TE.** 2015. Oxidative stress and metabolic perturbations in *Escherichia coli* exposed to sublethal levels of 2,4-dichlorophenoxyacetic acid. *Chemosphere*.

33. **Gullberg J, Jonsson P, Nordström A, Sjöström M, Moritz T.** 2004. Design of experiments: An efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal Biochem* **331**:283–295.
34. **Stein SE.** 1999. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J Am Soc Mass Spectrom* **10**:770–781.
35. **Hummel J, Selbig J, Walther D, Kopka J.** 2007. The Golm Metabolome Database : a database for GC-MS based metabolite profiling, p. 75–95. *In* Topics in Current Genetics.
36. **Stein SE.** 2011. NIST Standard Reference Database 1A. U.S. Department of Commerce, Gaithersburg, MD.
37. **Broeckling CD, Reddy IR, Duran AL, Zhao X, Sumner LW, Division PB, Roberts S, Foundation N, Box PO.** 2006. MET-IDEA: Data extraction tool for mass spectrometry-based metabolomics. *Anal Chem* **78**:4334–4341.
38. **Dieterle F, Ross A, Schlotterbeck G, Senn H.** 2006. Probabilistic quotient normalization as a robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabonomics. *Anal Chem* **78**:4281–90.
39. **Wiklund S, Johansson E, Sjöström L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J.** 2008. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal Chem* **80**:115–122.
40. **Stein SE, Babushok VI, Brown RL, Linstrom PJ.** 2007. Estimation of Kováts retention indices using group contributions. *J Chem Inf Model* **47**:975–80.

41. **Chagoyen M, Pazos F.** 2011. MBRole: Enrichment analysis of metabolomic data. *Bioinformatics* **27**:730–731.
42. **Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M.** 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **27**:29–34.
43. **Berthon G.** 2002. Aluminium speciation in relation to aluminium bioavailability, metabolism and toxicity. *Coord Chem Rev* **228**:319–341.
44. **Furukawa K, Fujihara H.** 2008. Microbial degradation of polychlorinated biphenyls: biochemical and molecular features. *J Biosci Bioeng* **105**:433–449.
45. **Macdonald TL, Bruce Martin R.** 1988. Aluminum ion in biological systems. *Trends Biochem Sci* **13**:15–19.
46. **Exley C.** 2009. Darwin, natural selection and the biological essentiality of aluminium and silicon. *Trends Biochem Sci* **34**:589–593.
47. **Booth SC, Weljie AM, Turner RJ.** 2013. Computational Tools for the Secondary Analysis of Metabolomics Experiments. *Comput Struct Biotechnol J* **4**:e201301003.
48. **Ferreira PAA, Bomfeti CA, Soares BL, de Souza Moreira FM.** 2012. Efficient nitrogen-fixing *Rhizobium* strains isolated from amazonian soils are highly tolerant to acidity and aluminium. *World J Microbiol Biotechnol* **28**:1947–1959.
49. **Silipo A, Molinaro A.** 2010. The diversity of the core oligosaccharide in lipopolysaccharides. *Subcell Biochem* **53**:69–99.
50. **Singh R, Beriault R, Middaugh J, Hamel R, Chenier D, Appanna VD, Kalyuzhnyi S.** 2005. Aluminum-tolerant *Pseudomonas fluorescens*: ROS toxicity and enhanced NADPH production. *Extremophiles* **9**:367–373.
51. **Wood T.** 1986. Distribution of the pentose phosphate pathway in living organisms. *Cell*

- Biochem Funct **4**:235–40.
52. **Singh R, Lemire J, Mailloux RJ, Appanna VD.** 2008. A novel strategy involved anti-oxidative defense: The conversion of NADH into NADPH by a metabolic network. *PLoS One* **3**.
 53. **Giandomenico AR, Cerniglia GE, Biaglow JE, Stevens CW, Koch CJ.** 1997. The Importance of Sodium Pyruvate in Assessing Damage Produced by Hydrogen Peroxide. *Free Radic Biol Med* **23**:426–434.
 54. **Shively JM, van Keulen G, Meijer WG.** 1998. Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annu Rev Microbiol* **52**:191–230.
 55. **Kuznetsova a a, Knorre DG, Fedorova OS.** 2009. Oxidation of DNA and its components with reactive oxygen species. *Russ Chem Rev* **78**:659–678.
 56. **Pellicer MT, Aguilar J, Badia J, Baldoma L.** 2003. Role of 2-Phosphoglycolate Phosphatase of *Escherichia coli* in metabolism of the 2-phosphoglycolate formed in DNA repair. *J Bacteriol* **185**:5815–5821.
 57. **Exley C.** 2004. The pro-oxidant activity of aluminum. *Free Radic Biol Med* **36**:380–387.
 58. **Imlay J a.** 2003. Pathways of oxidative damage. *Annu Rev Microbiol* **57**:395–418.
 59. **Mujika JI, Rezabal E, Mercero JM, Ruipérez F, Costa D, Ugalde JM, Lopez X.** 2014. Aluminium in biological environments: a computational approach. *Comput Struct Biotechnol J* **9**:e201403002.
 60. **Stohs SJ, Bagchi D.** 1995. Oxidative mechanisms in the toxicity of metal ions. *Free Radic Biol Med* **18**:321–336.
 61. **Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P.** 2002. DNA Repair *Molecular Biology of the Cell*, 4th ed. Garland Science, New York.

62. **Singh R, Lemire J, Mailloux RJ, Chénier D, Hamel R, Appanna VD.** 2009. An ATP and oxalate generating variant tricarboxylic acid cycle counters aluminum toxicity in *Pseudomonas fluorescens*. *PLoS One* **4**.
63. **Middaugh J, Hamel R, Jean-Baptiste G, Beriault R, Chenier D, Appanna VD.** 2005. Aluminum triggers decreased aconitase activity via Fe-S cluster disruption and the overexpression of isocitrate dehydrogenase and isocitrate lyase: A metabolic network mediating cellular survival. *J Biol Chem* **280**:3159–3165.
64. **Alhasawi A, Auger C, Appanna VP, Chahma M, Appanna VD.** 2014. Zinc toxicity and ATP production in *Pseudomonas fluorescens*. *J Appl Microbiol* **117**:65–73.
65. **Denayer S, Matthijs S, Cornelis P.** 2006. Resistance to vanadium in *Pseudomonas fluorescens* ATCC 17400 caused by mutations in TCA cycle enzymes. *FEMS Microbiol Lett* **264**:59–64.
66. **Brosnan JT, Brosnan ME.** 2006. The Sulfur-Containing Amino Acids: An Overview. *J Nutr* **136**:1636S–1640.
67. **Wortham BW, Patel CN, Oliveira MA.** Mechanisms.
68. **Chattopadhyay MK, Tabor CW, Tabor H.** 2003. Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen. *Proc Natl Acad Sci U S A* **100**:2261–5.
69. **Chattopadhyay MK, Tabor H.** 2013. Polyamines are critical for the induction of the glutamate decarboxylase-dependent acid resistance system in *Escherichia coli*. *J Biol Chem* **288**:33559–70.
70. **Johnson L, Mulcahy H, Kanevets U, Shi Y, Lewenza S.** 2012. Surface-localized spermidine protects the *Pseudomonas aeruginosa*: Outer membrane from antibiotic treatment and oxidative stress. *J Bacteriol* **194**:813–826.

71. **El-Halfawy OM, Valvano MA.** 2014. Putrescine reduces antibiotic-induced oxidative stress as a mechanism of modulation of antibiotic resistance in *Burkholderia cenocepacia*. *Antimicrob Agents Chemother* **58**:4162–71.
72. **Joutey NT, Sayel H, Bahafid W, Ghachtouli N El.** 2014. Effet des polyamines sur la réduction du chrome hexavalent par des souches bactériennes et leur résistance. Base.
73. **Teitzel GM, Geddie A, De Long SK, Kirisits MJ, Whiteley M, Parsek MR.** 2006. Survival and growth in the presence of elevated copper: Transcriptional profiling of copper-stressed *Pseudomonas aeruginosa*. *J Bacteriol* **188**:7242–7256.
74. **Warnes SL, Caves V, Keevil CW.** 2012. Mechanism of copper surface toxicity in *Escherichia coli* O157:H7 and *Salmonella* involves immediate membrane depolarization followed by slower rate of DNA destruction which differs from that observed for Gram-positive bacteria. *Environ Microbiol* **14**:1730–1743.
75. **Hong R, Kang TY, Michels C a., Gadura N.** 2012. Membrane lipid peroxidation in copper alloy-mediated contact killing of *Escherichia coli*. *Appl Environ Microbiol* **78**:1776–1784.
76. **Chenier D, Beriault R, Mailloux R, Baquie M, Abramia G, Lemire J, Appanna V.** 2008. Involvement of fumarase C and NADH oxidase in metabolic adaptation of *Pseudomonas fluorescens* cells evoked by aluminum and gallium toxicity. *Appl Environ Microbiol* **74**:3977–3984.
77. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–10.
78. **Grillo-Puertas M, Schurig-Briccio LA, Rodríguez-Montelongo L, Rintoul MR, Rapisarda VA.** 2014. Copper tolerance mediated by polyphosphate degradation and low-

- affinity inorganic phosphate transport system in *Escherichia coli*. *BMC Microbiol* **14**:72.
79. **Gray MJ, Jakob U.** 2015. Oxidative stress protection by polyphosphate-new roles for an old player. *Curr Opin Microbiol* **24C**:1–6.
80. **Shahid M, Xiong T, Masood N, Leveque T, Quenea K, Austruy A, Foucault Y, Dumat C.** 2013. Influence of plant species and phosphorus amendments on metal speciation and bioavailability in a smelter impacted soil: a case study of food-chain contamination. *J Soils Sediments* **14**:655–665.
81. **Fung DKC, Lau WY, Chan WT, Yan A.** 2013. Copper efflux is induced during anaerobic amino acid limitation in *Escherichia coli* to protect iron-sulfur cluster enzymes and biogenesis. *J Bacteriol* **195**:4556–68.
82. **Basu A, Das D, Bapat P, Wangikar PP, Phale PS.** 2009. Sequential utilization of substrates by *Pseudomonas putida* CSV86: signatures of intermediate metabolites and online measurements. *Microbiol Res* **164**:429–37.
83. **Hue N V., Craddock GR, Adams F.** 1986. Effect of Organic Acids on Aluminum Toxicity in Subsoils1. *Soil Sci Soc Am J* **50**:28.
84. **Behrends V, Ryall B, Wang X, Bundy JG, Williams HD.** 2010. Metabolic profiling of *Pseudomonas aeruginosa* demonstrates that the anti-sigma factor MucA modulates osmotic stress tolerance. *Mol Biosyst* **6**:562–569.
85. **Tsai Y-P, Tzeng H, Lin JJ-WJ-YJ, Lu M, Lin JJ-WJ-YJ.** 2013. Verification of enzymes deterioration due to Cu(II) presence in an enhanced biological phosphorus removal system. *Chemosphere* **91**:602–7.
86. **Keum YS, Seo JS, Li QX, Kim JH.** 2008. Comparative metabolomic analysis of *Sinorhizobium* sp. C4 during the degradation of phenanthrene. *Appl Microbiol Biotechnol*

- 80:863–72.
87. **Sornchuer P, Namchaiw P, Kerdwong J, Charoenlap N, Mongkolsuk S, Vattanaviboon P.** 2014. Copper chloride induces antioxidant gene expression but reduces ability to mediate H₂O₂ toxicity in *Xanthomonas campestris*. *Microbiology* **160**:458–66.
 88. **Cameotra SS, Makkar RS, Kaur J, Mehta SK.** 2010. Synthesis of Biosurfactants and Their Advantages to Microorganisms and Mankind, p. 261–280. *In* Sen, R (ed.), *Biosurfactants*. Springer New York, New York, NY.
 89. **Gibson DT, Parales RE.** 2000. Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr Opin Biotechnol* **11**:236–243.
 90. **Tanaka N.** 2013. Density Functional Theory Studies on the Addition and Abstraction Reactions of OH Radical with Benzoate Anion. *Open J Phys Chem* **3**:7–13.
 91. **O'Brien IG, Cox GB, Gibson F.** 1970. Biologically active compounds containing 2,3-dihydroxybenzoic acid and serine formed by *Escherichia coli*. *Biochim Biophys Acta - Gen Subj* **201**:453–460.
 92. **Nikel PI, Pérez-Pantoja D, de Lorenzo V.** 2013. Why are chlorinated pollutants so difficult to degrade aerobically? Redox stress limits 1,3-dichloroprop-1-ene metabolism by *Pseudomonas pavonaceae*. *Philos Trans R Soc Lond B Biol Sci* **368**:20120377.
 93. **Kang Y-SS, Lee Y, Jung H, Jeon CO, Madsen EL, Park W.** 2007. Overexpressing antioxidant enzymes enhances naphthalene biodegradation in *Pseudomonas* sp. strain As1. *Microbiology* **153**:3246–3254.
 94. **Agulló L, Cámara B, Martínez P, Latorre V, Seeger M.** 2007. Response to (chloro)biphenyls of the polychlorobiphenyl-degrader *Burkholderia xenovorans* LB400 involves stress proteins also induced by heat shock and oxidative stress. *FEMS Microbiol*

- Lett **267**:167–175.
95. **Ponce BL, Latorre VK, González M, Seeger M.** 2011. Antioxidant compounds improved PCB-degradation by Burkholderia xenovorans strain LB400. *Enzyme Microb Technol* **49**:509–16.
 96. **Chávez FP, Lünsdorf H, Jerez C a.** 2004. Growth of polychlorinated-biphenyl-degrading bacteria in the presence of biphenyl and chlorobiphenyls generates oxidative stress and massive accumulation of inorganic polyphosphate. *Appl Environ Microbiol* **70**:3064–3072.
 97. **Bhat S V, Booth SC, McGrath SGK, Dahms TES.** 2014. Rhizobium leguminosarum bv. viciae 3841 Adapts to 2,4-Dichlorophenoxyacetic Acid with “Auxin-Like” Morphological Changes, Cell Envelope Remodeling and Upregulation of Central Metabolic Pathways. *PLoS One* **10**:e0123813.
 98. **Ye Y, Wang X, Zhang L, Lu Z, Yan X.** 2012. Unraveling the concentration-dependent metabolic response of Pseudomonas sp. HF-1 to nicotine stress by ¹H NMR-based metabolomics. *Ecotoxicology* **21**:1314–24.
 99. **Puglisi E, Cahill MJ, Lessard P a., Capri E, Sinskey AJ, Archer J a C, Boccazzi P.** 2010. Transcriptional Response of Rhodococcus aetherivorans I24 to Polychlorinated Biphenyl-Contaminated Sediments. *Microb Ecol* **60**:505–515.
 100. **Asad NR, Asad LMBO, de Almeida CEB, Felzenszwalb I, Cabral-Neto JB, Leitão AC.** 2004. Several pathways of hydrogen peroxide action that damage the E. coli genome. *Genet Mol Biol* **27**:291–303.

Chapter Four Preface

Chapter 4 describes a combined bioinformatic and molecular biology study aimed at characterizing how the gene *aer* and others are involved in aerotaxis. This work began as an attempt to rule out energy-taxis as the method by which *P. pseudoalcaligenes* KF707 swims towards biphenyl. When I observed that it had three homologs of *aer*, and the type species *P. aeruginosa* PAO1 had only one I became interested in understanding the phylogenetic relationship of these genes in the *Pseudomonas* genus. I also became interested in two other receptors, Aer-2 and CttP when I discovered that very few *Pseudomonas* species had these genes. I used bioinformatics to characterize the Aer family and discovered that these homologs were variably distributed in the genus and had very likely been transferred horizontally. Using molecular biology tools to generate deletion mutants in *P. pseudoalcaligenes* KF707, I discovered that in this species one Aer homolog in particular was key to energy-taxis, but surprisingly Aer-2 and CttP were also involved. This was a novel characterization of the function of these receptors, which along with the definition of the Aer family presents a salient contribution to the field of *Pseudomonas* chemotaxis. I designed the experiments, analyzed the data and wrote the manuscript in consultation with my supervisor Dr. Turner.

Chapter Four: **Aer is a family of Energy-Taxis Receptors in *Pseudomonas* which also involves Aer-2 and CttP**

Sean C. Booth and Raymond J. Turner. 2016. Aer is a family of Energy-Taxis Receptors in *Pseudomonas* which also involves Aer-2 and CttP. bioRxiv.

Submitted to *Molecular Microbiology* October, 2016

4.1 Abstract

Energy-taxis by *Pseudomonas* species allows swimming towards optimal environments for generating cellular energy, precluding the need to sense carbon sources and oxygen. Aerotaxis, which has been taken to mean swimming towards oxygen or a source of metabolizable carbon has been attributed to the methyl-accepting chemotaxis protein Aer. In *P. aeruginosa* McpB (Aer-2) also contributes and may directly sense oxygen whereas in *P. putida* the second of three Aer energy-taxis sensor homologs, 'Aer2' is the key receptor. Here we sought to disentangle the ambiguity between aerotaxis and energy-taxis and between McpB (Aer-2) and Aer2. Phylogenetic characterization of sequences from a wide range of *Pseudomonas* species revealed that Aer homologs have been duplicated and horizontally transferred within the genus but can be divided into 5 groups. Phenotypic characterization in *P. pseudoalcaligenes* KF707, which has 3 Aer homologs and McpB showed that all homologs as well as McpA (CttP) and McpB (Aer-2) contribute to energy-taxis, though the most ancestral Aer homolog is key but not essential to this behaviour. As McpB (Aer-2) is rare in the genus, these results indicate that energy-taxis and aerotaxis are linked but distinct behaviours and that energy-taxis in *Pseudomonas* involves more receptors than previously thought.

4.2 Introduction

Chemotaxis, the ability to sense and swim along chemical gradients, is a widespread and important behaviour in bacteria (1). Canonically it functions by extracellular compounds binding to membrane-bound methyl-accepting chemotaxis proteins (MCPs) causing a phosphorylation signal cascade through CheA and CheY to alter the direction of flagellar rotation (1), allowing the cell to direct swimming through concentration gradients. The first MCPs were characterized in *E. coli* but now many more have been described, particularly in *Pseudomonas* (2). While receptors for specific ligands are being identified, many through a clever high-throughput approaches (3), our understanding of some ‘characterized’ MCPs remains incomplete. The receptors in *Pseudomonas* involved in aerotaxis and energy-taxis have been identified, but questions remain about their exact function and whether the two processes are actually distinct.

Chemotaxis is generally used as a catch-all to describe any kind of directed bacterial motion towards or away from a chemical, but for the purposes of this study, chemotaxis, aerotaxis and energy-taxis must be more strictly defined. Chemotaxis refers to directed swimming mediated by binding of a ligand either directly or through a periplasmic binding protein to an MCP. Distinguishing between aerotaxis and energy-taxis is more important as these have been traditionally conflated, as energy-taxis is a relatively new term. Here aerotaxis will be defined as directed swimming towards oxygen through direct sensing of O₂, and energy-taxis as directed swimming towards a more favourable environment through indirect sensing of energy production within the cell. Aerotaxis was among the first bacterial behaviours ever described as both Engelmann and Beijerinck described the aggregation of bacterial cells near sources of oxygen (4, 5). When the MCP responsible for this phenomenon was discovered in *E. coli*, it was given the name Aer (6). While it does indeed mediate taxis towards oxygen, the mechanism was later found

not to involve direct binding of oxygen (7). Instead, Aer has an FAD cofactor which when reduced allows the bacterium to sense energy generation within the cell, meaning Aer is actually an energy-taxis receptor. The distinction between energy and aerotaxis became important when two genes in *P. aeruginosa* were found to be involved in ‘aerotaxis’, one a homolog of Aer, the second a distinct cytoplasmic MCP which was given the name Aer-2 (8), though it is also still called McpB. Further issues were raised when 3 homologs of Aer were found in *P. putida* (9), and the most abundant and only functional homolog was given the name ‘Aer2’. Biochemical characterization of Aer-2 (also called McpB or TlpG) showed that it binds gases including O₂ but functional reports disputed its role in ‘aerotaxis’ (10). Thus despite multiple investigations it remains unclear which genes are responsible for aerotaxis and energy-taxis in *Pseudomonas*, and whether they are distinct processes. Here we sought answers to these questions using a combination of comparative genomics and phenotypic characterization of mutants.

In *Pseudomonas*, Aer was first characterized in *P. putida* PRS2000 (11), then in *P. aeruginosa* PA01 (8), then in *P. putida* KT2440 (9) and F1 (12) as an ‘aerotaxis’ receptor, using a variety of methods. Unfortunately, it is difficult to discern a difference between energy-taxis from aerotaxis from the experimental approach used in such studies. Despite the observation that *P. putida* had 3 homologs whose functionality had to be determined by process of elimination, compared to *P. aeruginosa*’s 1, no comparison of the genes or gene products was made. This raises the question of how homologs are distributed in the genus, which ones are functional in other species, and most importantly why do they possess multiple similar copies of the same chemoreceptor? Aer-2 was originally implicated with Aer as an aerotaxis receptor and while the experimental evidence characterizing CttP as a tetrachloroethylene (TCE) receptor is clear (13), yet it seems unlikely that this is the natural function of the protein. Due to its appearance as the

first gene in the *che2* operon, which also contains *mcpB* (*aer-2*) we hypothesized that both may have a role in energy-taxis. To investigate these questions, Aer sequences were obtained from 65 *Pseudomonas* species as well as the presence/absence of Aer-2 and CttP. A phylogeny of Aer was built providing insight into the distribution in the genus and allowing the ‘Aer’ family to be defined. Only one species, *P. pseudoalcaligenes* KF707 had the surprising feature of possessing 3 Aer homologs, Aer-2 (McpB) and CttP making it the ideal candidate to investigate if these receptors had related functions. Deletion mutants were generated revealing that in KF707, all 5 of these genes contributed to energy-taxis, though the ancestral Aer homolog was pivotal. Together these results provide a definition of the Aer energy-taxis receptor as a family with varied distribution in *Pseudomonas* and implicate Aer-2 and CttP in this behaviour.

4.3 Materials and Methods

4.3.1 Protein Sequences

All sequences were obtained using NCBI databases and tools (14) in May 2016. BLAST searches (15) (pBLAST for draft and completed genomes, tBLASTn for whole-genome shotgun genomes) were performed using *P. aeruginosa* PA01 Aer (NP_0250252.1) as a query sequence. From each species, all hits with >95% sequence coverage, no matter how low the sequence identity, were selected for inclusion. BLAST did not return any results with coverage values between 67% and 95%, indicating that all included sequences were true Aer homologs. Expect values were always below 1×10^{-100} . Two sequences were removed as they were redundant entries resulting from incorrect start site annotations resulting in two proteins with the same C-terminus but slightly different N-termini. Two more sequences were removed for similar reasons. Sequence accession numbers were thus obtained from the international nucleotide sequence database

collaboration (INSDC) (16). NCBI Entrez was used to obtain FASTA formatted sequences which were then compiled into a single file. A sequence to be used as a root was also acquired by BLASTing NP_250252.1 against the NCBI database, but excluding any *Pseudomonas* sequences from the results. WP_0027981445.1 from an unknown gamma proteobacterium L18 was thus obtained.

4.3.2 Alignment

Two alignments were made, with and without the root, using COBALT (17). The default settings were used as adjustment did not improve the alignment noticeably. As this alignment had numerous unnecessary gaps introduced, these were manually removed by adjusting the alignment in Jalview (18), which was then re-aligned using MUSCLE (19). Names were cleaned up using a custom script in R (20). The alignments were input into the Mobylye web platform (21) and maximum likelihood phylogenies were constructed using PHYML (22). PHYML options: Amino-Acids, 100 bootstrap replicates, JTT amino-acid substitution model, BEST tree topology search operation, tree topology, branch length and rate parameters were optimized. Consense was then used to generate a consensus tree from the bootstrap replicates using the majority rule (extended), and only treating the tree as rooted when the non-pseudomonas root was included.

4.3.3 Sequence Harmony and Multi-Relief Analysis to Determine Groups

The multi-Harmony server was used to apply sequence harmony and multi-relief (SHMR) to validate groupings made based on the ML tree and alignment. Groups were manually decided based on the tree topology and corresponding alignment, initially making for 7 groups. These seven groups were compared in pair-wise fashion using SHMR (23). Thus for each pair of groups a score was calculated for each amino acid that indicated how conserved it was within groups and how divergent it was between groups, a score of 1 indicating perfect conservation within and

perfect divergence between. Empirical cumulative distribution functions (ECDF) were plotted for each comparison and the percent of AAs above the 0.8 cutoff recommended by the SHMR authors was determined. As most comparisons resulted in 25% of AAs reaching this threshold, the two pairs of groups that, when compared, only had 10% of AAs above the cutoff, were deemed incorrect group assignments. These groups were merged and the process repeated to produce 5 groups that all had ~25% of AAs above the cutoff threshold. Group 5 was consistently excluded as it only contains 2 highly divergent sequences.

4.3.4 Comparison of Groups

Unique regions of each group of Aer homologs were identified by comparing the SHMR scores for each pair-wise comparison with the overall conservation score of each AA and the domain architecture of the Aer protein. Conservation scores were obtained from Jalview (18), and along with the SHMR scores were smoothed and plotted using ggplot2 (24) in R. Smoothing was performed by calculating the average of the 3 preceding and following AAs for each position. The domain architecture of Aer was obtained from the conserved domain database (CDD) (25).

4.3.5 Distribution of Groups

Based on the ML tree, the number of times a strain appeared in each group was counted. This matrix was then inverted (now indicating number of Aer homologs per group for each strain). A hierarchically clustered heatmap using Bray-Curtis distance and average clustering was made in R. The presence of Aer-2 (NP_248866.1) and CttP (WP_003106690.1) were determined using BLAST searches specifically against the strains. Pseudogenes were detected by BLASTing the nucleotide sequence of Aer.

4.3.6 Detection of Evidence of Horizontal Gene Transfer

Graphical representations of the Aer homologs nucleotide sequences were manually inspected on NCBI. The first two genes upstream and downstream of the *aer* homolog were noted, along with any mobile elements (transposase, integrase and inverted repeats) within 5 kb. For each different upstream and downstream gene their frequency of occurrence was calculated for each homolog group. The frequency of occurrence of each type of mobile element was also calculated, as well as for the complete set of homologs.

4.3.7 Culture Growth

For molecular biology, cultures were routinely cultured in lysogeny broth (LB, 5 g/L yeast extract, 10 g/L Tryptone, 10 g/L NaCl). For energy-taxis experiments *P. pseudoalcaligenes* KF707 strains were grown overnight (16 h) in minimal salts media containing 10 mM pyruvate or succinate. Minimal salts media contained (in g/L) K₂HPO₄, 3; NaH₂PO₄, 1.15; NH₄Cl, 1; KCl, 0.15; MgSO₄, 0.15; CaCl₂, 0.01; FeSO₄, 0.0025. The latter four were sterile filtered and added after autoclaving.

4.3.8 Generation of Deletion Constructs and Mutants

Nucleotide sequences for Aer.g1, Aer.g2, Aer.g4, Aer2 and CttP were obtained from the draft genome sequence of *P. pseudoalcaligenes* KF707 (26). Primers for a ~500 bp region up and downstream of each region were generated using Primer BLAST (27). Benchling was used to alter the primers, adding BamHI or HindIII restriction sites to the outer primers and the reverse-complement of the other inner primer to each of the internal primers. Genomic DNA was isolated by the phenol/chloroform method (28). In separate PCR reactions the upstream and downstream fragments were amplified using Hi-Fidelity (HF) Enzyme mix (Fisher Scientific, USA). Fragments were purified by gel extraction using an EZDNA kit (Omega Bio-Tek, USA) then pooled and used

as the template for the second PCR reaction using only the outer primers. The pG19II vector (29), purified using an EZDNA plasmid mini kit II (Omega Bio-Tek, USA), and insert were digested using BamHI and HindIII (Invitrogen, USA). Digestion products were purified then ligated together using T4 ligase (Invitrogen, USA). Ligations were transformed either directly into *E. coli* Top10F' or first into DH5 α chemical competent cells using standard methods (28). White colonies were picked from LB 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) gentamycin (20 μ g/mL) plates and their plasmids isolated and screened for a ~1kb bandshift. Those with the appropriate shift were sequenced (Eurofins, USA) to confirm the correct insert sequence. To delete the genes from the *P. pseudoalcaligenes* KF707 genome, the deletion construct containing plasmids were introduced by conjugation. Cultures of *E. coli* HB101 carrying the helper plasmid pRK2013 (30) and *E. coli* Top10F' carrying the deletion construct in pG19II were grown to early log-phase (OD ~0.3) along with the KF707 wild-type, or later, deletion mutants. Donor, helper and recipients were mixed and plated, grown overnight then the cell mass was collected and spread on AB glucose plates (5 g/L glucose, minimal salts media) containing 20 μ g/mL gentamycin for 48 h at 30°C. Transconjugant colonies were picked off the AB glucose Gm plates into LB no salt, LB 10% sucrose and LB + 20 μ g/mL Gm. Colonies that were able to grow with Gm but not (at all) with sucrose were selected for continued use. The LB no salt overnight culture was used to inoculate LB 10% sucrose. After 4h, the culture was plated on LB 10% sucrose and colonies were screened to determine if the deletion occurred. Colony PCR was used to find those that had ONLY a band at ~1K bp which were sent for sequencing (Eurofins, USA) to confirm the deletion.

4.3.9 Energy-taxis Swim Plates

Swim plates were made by making minimal salts medium with 0.3% agar and 50 mM succinate or pyruvate. Strains were grown overnight in minimal salts media containing 10 mM of the appropriate carbon source. An inoculation needle was sterilized by ethanol and flaming then dipped into the overnight culture and carefully stabbed into the swim plate. The needle was re-inoculated for each stab and was re-sterilized for each strain. The diameter of growth for each strain was measured at 24 and 48 h either manually using a ruler or digitally using a photograph and ImageJ (31). The experiment was repeated at least 3 times for all strains in each medium, always including at least one wildtype as a normalizing control. Collected data were processed in R to normalize the size of the growth diameter to the corresponding wildtype size at 24 and 48 h. The data were analyzed in a number of ways and it was determined that pooling the normalized data from 24 and 48 h to produce at least 6 replicates resulted in the smallest error bars. Tukey's Honest Significant Differences test was used to determine if the differences between strains were significant for each carbon source. As the TukeyHSD() function in R (20) tests ALL pair-wise comparisons, only those comparing each strain to the wild-type are presented here. However, this makes these results more robust as the false positive correction for a confidence level of 0.95 was applied to all 325 comparisons which were tested.

4.3.10 Chemotaxis Swim Plates

Strains were grown up overnight as before, then 1 mL was pelleted, washed once with 1 mL minimal salts medium (no carbon source) then resuspended in 100 μ L minimal salts media (no carbon source). 20 μ L was spotted at the edge of a minimal salts media plate containing no carbon source and 0.3% agar. Either an agar plug containing 50 mM carbon source or small amount of crystals was placed in the centre and plates were incubated overnight at 30°C. Plates were

photographed and positive chemotaxis was interpreted as an arc of cells nearer to the centre of the plate than the spot had been placed.

4.4 Results

4.4.1 Bioinformatics Results

65 *Pseudomonas* species were selected for this analysis. Species with completely sequenced genomes were selected, but for those with many strains only a few representatives that have been highly studied were included. Species with incomplete (draft) sequences were also included in an attempt to ensure representation from all major *Pseudomonas* clades, based on works by Bodilas *et al* and Gomila *et al* (32, 33). 144 Protein sequences were obtained from the NCBI database using the *P. aeruginosa* PA01 Aer sequence (NP_250252.1) as a BLAST (15) query sequence. All hits with >95% sequence coverage were included. These sequences were aligned using COBALT (17) as this alignment algorithm ensures that conserved domains are aligned despite a lack of similarity elsewhere in the sequence. This was ideal for the Aer sequences as, like other MCPs, they are conserved in the C-terminal signaling region but less conserved in the N-terminal transmembrane and ligand binding region. Inspection of the alignment revealed numerous unnecessary gaps, which could not be remedied by changing the alignment settings. Instead, these gaps were removed and the regions re-aligned manually. This altered alignment was then re-aligned using MUSCLE (19) to remedy any small incongruities introduced by the manual process. This produced a suitable alignment for determining the phylogeny of the sequences (Figure 4-2). The Mobyly web platform (21) was used to construct a maximum likelihood phylogeny using PHYML (22). Two phylogenies were generated, one using only the *Pseudomonas* sequences and one which included the closest non-*Pseudomonas* sequence in the NCBI database to be used as a root (WP_027981445.1, unknown gamma proteobacterium L18, (34)).

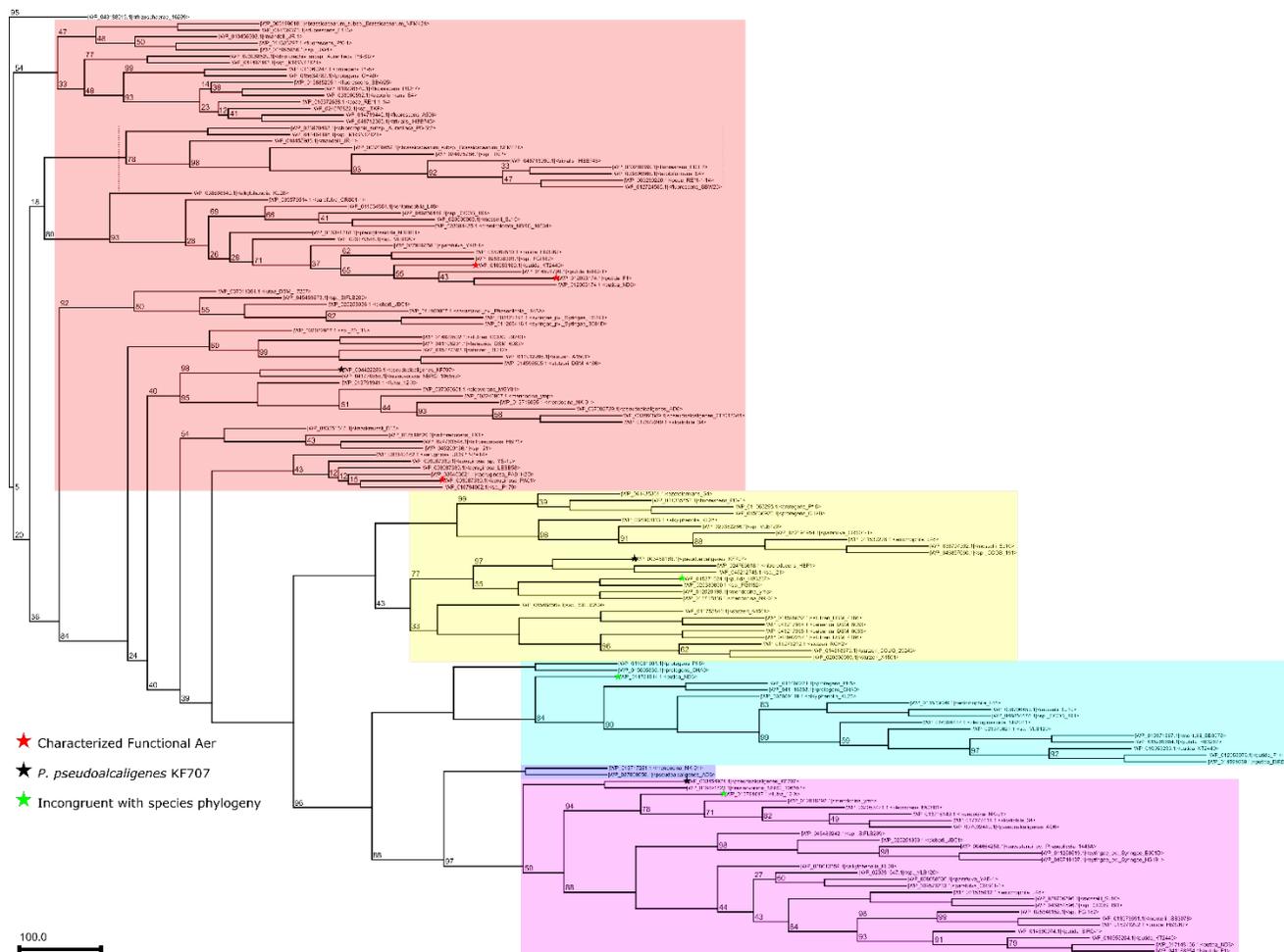


Figure 4-2: Maximum likelihood consensus tree showing phylogenetic relationship between Aer homolog protein sequences from select *Pseudomonas* species. Colours indicate homolog group: Aer.g1 (red), Aer.g2 (yellow), Aer.g3 (cyan), Aer.g4 (purple) and Aer.g5 (blue). Groups were based on inspection of the alignment used to generate the phylogeny and other analyses, see the text for details. Tree was generated unrooted, then the root placed based on a similar tree that was rooted to the closest non-*Pseudomonas* sequence available. Branch lengths indicate number of AA substitutions per site. Numbers at nodes indicate bootstrap support values from 100 replicates. Dotted box indicates original separation of Aer.g1 into two groups.

This rooted phylogeny (Figure 4-1) was examined and the location of the root was used to place the root on the unrooted tree. This revealed that the Aer homolog from *P. rhizosphaerae* 16299 (WP_043188019.1) was more similar to the non-*Pseudomonas* root than to any *Pseudomonas* sequence.

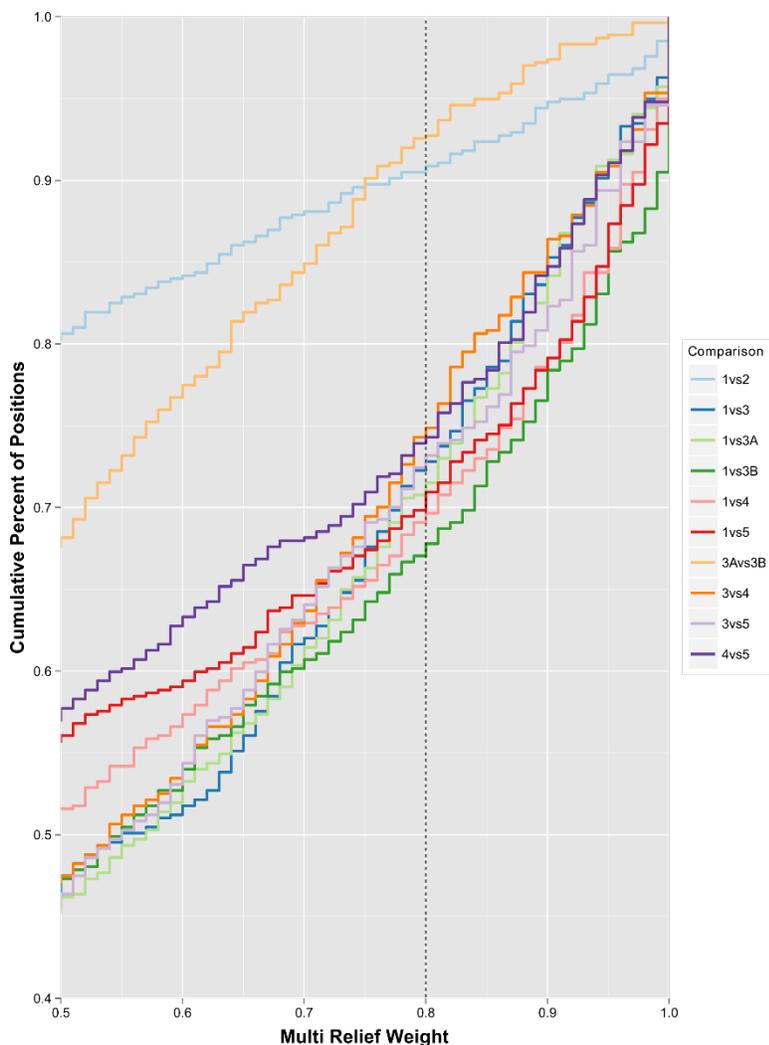


Figure 4-3: Empirical cumulative distribution function of amino acid multi-relief scores for pairwise comparisons of preliminary Aer homolog groups. Only scores above 0.5 are shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of amino acid positions for each comparison are above this cutoff. E.g. only 10% of amino acid positions in the 1vs2 comparison and even less in the 3Avs3B comparison are above this cutoff. Conversely, there is between 25% and 35% of amino acid positions above the cutoff for all other comparisons.

4.4.1.1 Grouping of Aer Homologs

Inspection of the unrooted, phylogenetically organized alignment implied that there are several sub-families of Aer homologs (Figure 4-2). Initially, 7 groups were identified based on the alignment and tree branching. These groups were aligned in pair-wise fashion then analyzed to determine if the group assignments were accurate. Manual inspection and Sequence Harmony / Multi-Relief (SHMR) demonstrated that there were actually only 5 groups. SHMR is a pair of algorithms that takes as input a pair of alignments and determines, for each AA position, whether that AA is conserved within each group, and whether it is divergent between the two groups, giving each AA position a score (23). Each pair of groups was compared in this fashion, then their results compared by determining what percent of AAs were conserved within a group but divergent between the groups (Figure 4-3). AAs with a multi-relief weight score >0.8 were accepted as fitting these criteria (cutoff is based on the recommendations of the SHMR authors (23)). Comparisons of groups where only 10% of AAs fit into this category indicated that these two groups were too similar as most comparisons indicated that 25-35% of AAs should be above the distinction cutoff. Through this process, groups 1 and 2 were combined and groups 3A and 3B were combined. Re-analysis using SHMR showed that all inter-group comparisons had between 25-35% of AAs distinguishing the two groups (Figure 4-4). Throughout this process, group 5 was consistently excluded as it only contains 2 sequences, which are highly divergent from all others. When the genomic context of each *aer* homolog gene was later examined (Figure 4), showing that the genes from the same group shared synteny. This supports the conclusion that there are 5 groups within the Aer family, and they are defined by shared amino acid sequence features their associated genes.

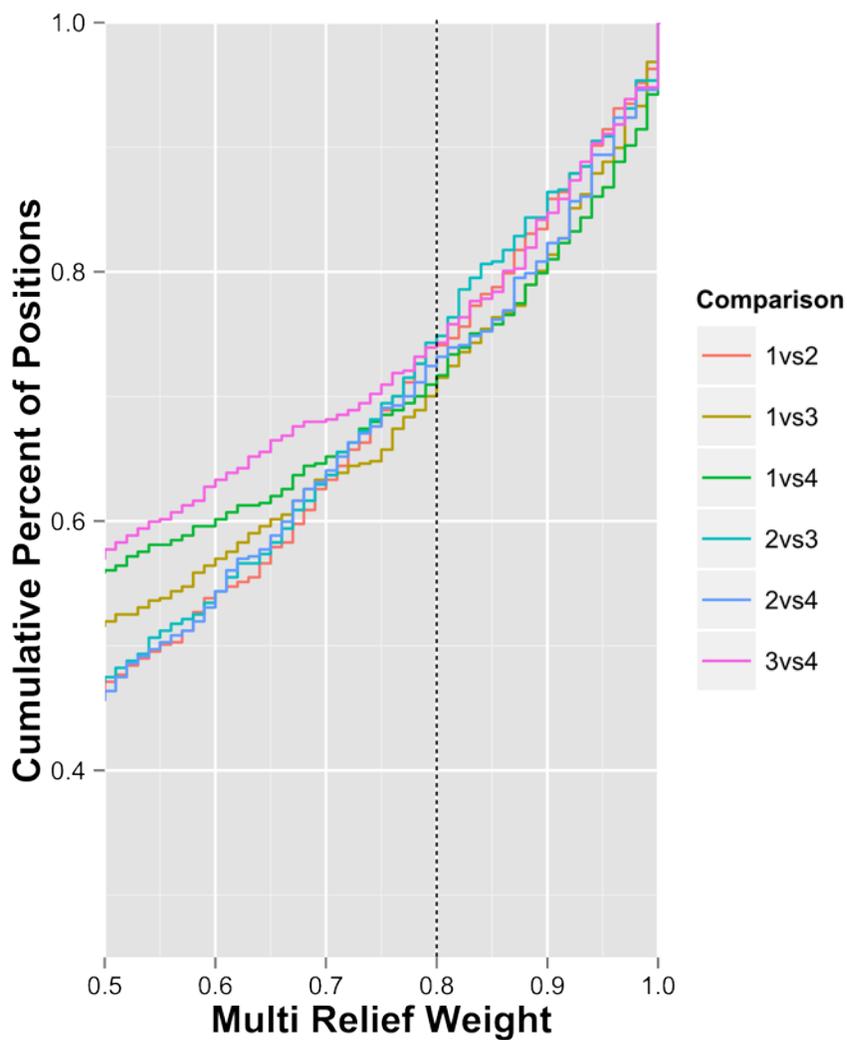


Figure 4-4: Empirical cumulative distribution function of amino acid multi-relief scores for pairwise comparisons of adjusted Aer homolog groups. Only scores above 0.5 are shown, and accumulation above 40%. The dotted line at 0.8 allows to easily see what percent of amino acid positions for each comparison are above this cutoff. All comparisons show that between 25% and 35% of amino acid positions are above the cutoff.

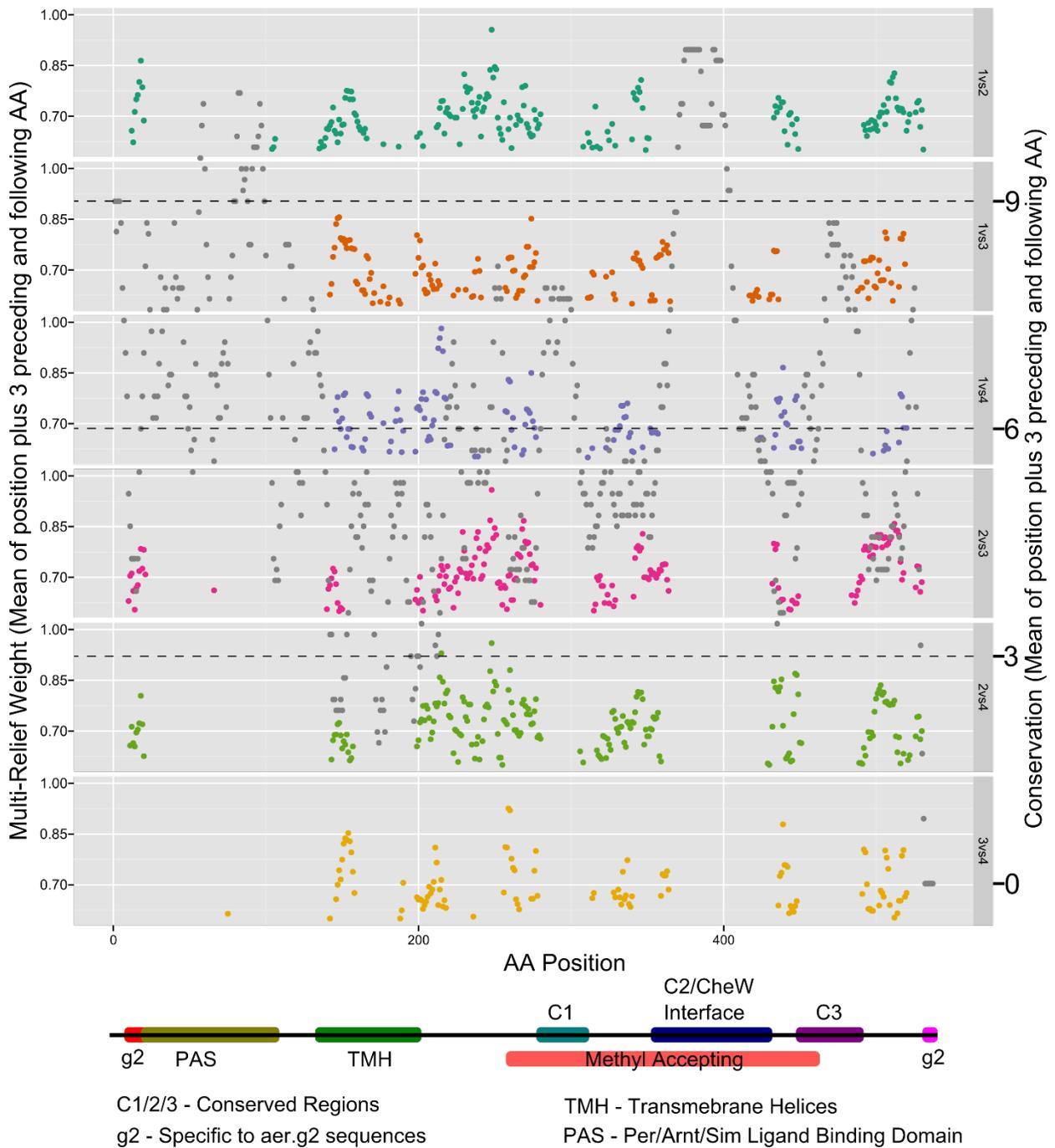


Figure 4-5: Amino acid conservation and multi-relief weight scores for all positions of Aer and domain architecture of Aer. Conservation was calculated across the entire alignment of all 144 Aer homolog sequences. Multi-relief scores were calculated for each pair-wise comparison of homolog groups. Both values have been smoothed by taking the mean of the 3 preceding and following amino acids for each position. Domain architecture was obtained from the conserved domain database. Regions have been colour coded based on known features, and from features derived from the current study.

4.4.1.2 Comparison of Groups

Results from the SHMR analysis for each group comparison were plotted against the entire length of the Aer protein and compared to the overall conservation at each AA position (Figure 4-5). Using a smoothing function, the regions that were varying between groups were identified and compared to the domain architecture of the Aer protein. These comparisons showed that the regions determining the specificity of each group were the same for all groups, (except for two regions at each extremity of the protein that were specific to Aer.g2) and that these regions were poorly conserved in the overall alignment. As expected, the CheW interface region of the methyl-accepting domain and the Per/Arnt/Sim (PAS) ligand binding domain were well conserved and did not differ between groups, though the CheW interface was far more conserved than the PAS domain. Two additional conserved regions (deemed C1 and C3) were found before and after the CheW interface region, each separated by an unconserved, group specific region. The predicted transmembrane domains (AAs ~130-200) were the least conserved region and were not conserved within groups except for the initial portion of these domains (AA ~150). Between this region and the C1 conserved region of the methyl-accepting domain was another group specific region, though part of this region was conserved between groups 1,3 and 4 (AAs ~220-260) making this a third region specific to Aer.g2. Finally, the C-terminal region after the last conserved C3 region was specific for each group, especially Aer.g2 as it has an extension that the other groups do not. Thus within the Aer family, aer.g2 is distinguished by its N-terminal region, central region and C-terminal extension while aer.g5 is notable in its complete differentiation the entire length of the protein. The other groups, Aer.g1, Aer.g3 and Aer.g4 can be distinguished from one another based on the composition of the variable regions (AAs ~155-165, ~195-225, ~250-275, ~305-360 and ~395-410).

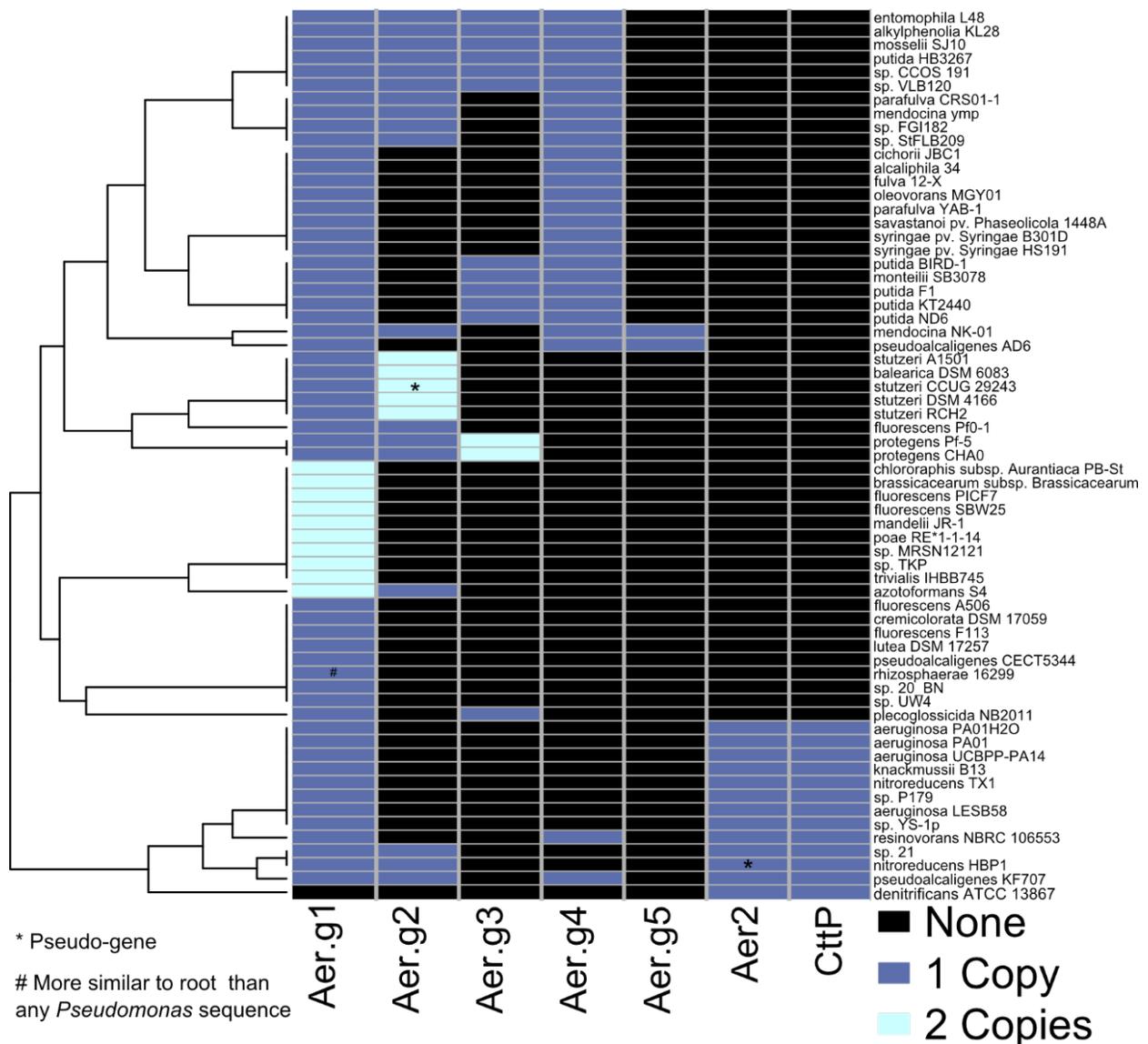


Figure 4-6: Hierarchically clustered heatmap showing the presence and number of Aer homologs, Aer2 and CttP in select *Pseudomonas* species. Species were clustered using the Bray-Curtis distance metric and average linkages. The number of homologs that each strain possess is indicated by the box colour: zero (black), one (blue), two (cyan). Pseudogenes are marked by asterisks (*). The homolog of *P. rhizosphearae* is more similar to non-*Pseudomonas* sequences and is marked by a pound (#).

4.4.1.3 Distribution of Groups within *Pseudomonas*

To determine the prevalence of each Aer homolog group within the *Pseudomonas* genus, the number of homologs from each group that each species possessed was counted and a hierarchically clustered heatmap was generated (Figure 4-6). Aer-2 and CttP were also counted

and included. This analysis showed that all included species, except *P. dentrificans* ATCC13867, had an Aer.g1 homolog. Only species related to *P. aeruginosa* had Aer-2 and CttP, which were always found together. *P. sp 21*, *P. nitroreducens* HBP1, *P. resinovorans* NBRC 106553 and *P. pseudoalcaligenes* KF707 were the only species that had Aer-2/CttP and additional Aer homologs, the first two possessing an Aer.g2 homolog, *P. resinovorans* having an Aer.g4 homolog and KF707 having both. Only genes from Aer.g1, Aer.g2 and Aer.g3 were duplicated, and duplications were rarer than possession of homologs from multiple groups. Duplications of Aer.g1 were restricted to *P. fluorescens* and related subgroups (1-8) (33). Duplications of Aer.g2 occurred only (and always) in *P. stutzeri* and the closely related *P. balearica*. Unlike the other duplications, these were tandem duplications. Duplications of Aer.g3 only occurred in *P. protegens* CHA0 and Pf-5. No single species had a representative from all groups, though *P. mendocina* NK-01 was only missing Aer.g2 and Aer-2/CttP and *P. pseudoalcaligenes* KF707 was only missing Aer.g3 and Aer.g5.

Overall, the tree clustering *Pseudomonas* species based on possession of Aer homologs roughly matched the known species phylogeny, though there were several noteworthy instances of variability at the strain level. The presence of Aer-2/CttP created a major separation, however this separation matches that of the aeruginosa subgroup. Only 1/2 *P. parafulva* strains had an Aer.g2 homolog and only 2/4 *P. fluorescens* strains had the same Aer homolog profile. 3/4 of *P. putida* strains were the same, though *P. putida* HB3267 had a homolog from all four groups. *P. mendocina* ymp and NK-01 differed as the latter possesses the rare Aer.g5. The other strain with this homolog, *P. pseudoalcaligenes* AD6, was quite different than KF707 as it also did not have an Aer.g2 homolog nor Aer-2/CttP. This was the species with the most variation as the third strain

included in the analysis, *P. pseudoalcaligenes* CECT5344, only possessed a single *Aer.g1* homolog.

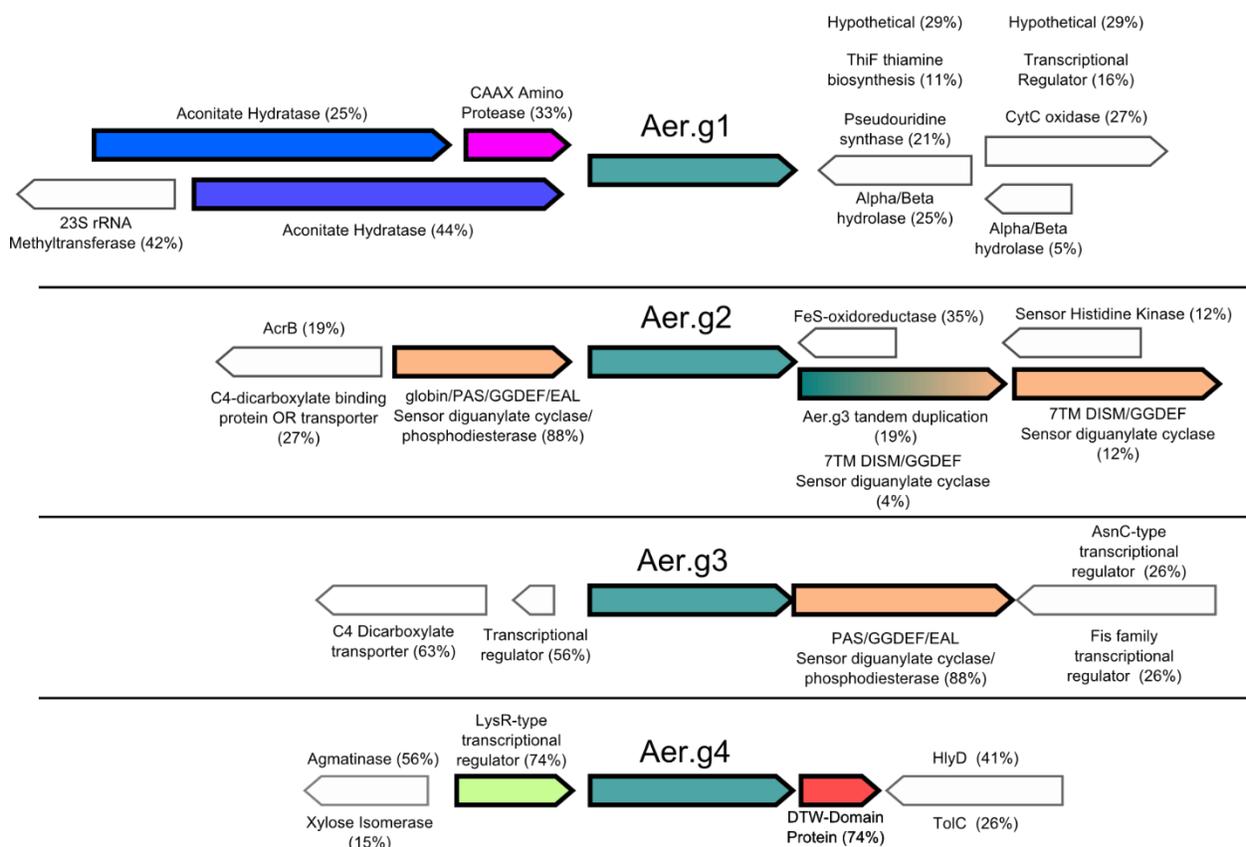


Figure 4-7: Frequency of occurrence and orientation of genes upstream and downstream from *aer* homologs from select *Pseudomonas* species. Coloured and bolded genes are expected to be co-transcribed based on their orientation and high frequency of occurrence with *aer*. Numbers beside gene functions indicate the frequency that they were found within the 144 sequences examined. Gene lengths are approximate, *aer* is about 1.5kb long.

Table 4-1: Frequency of association of mobile elements within 5kb of *aer* genes.

Aer Group	Repeats	Transposase	Integrase
<i>aer.g1</i>	20/73	2/73	0/73
<i>aer.g2</i>	4/26	1/26	0/26
<i>aer.g3</i>	1/16	1/16	1/16
<i>aer.g4</i>	4/27	2/27	1/27
<i>aer.g5</i>	0/2	0/2	0/2
Total	29/144	6/144	2/144

4.4.1.4 Evidence for Horizontal Gene Transfer of Aer Homologs

Based on the varied distribution of Aer homologs within the genus *Pseudomonas* it was hypothesized that horizontal gene transfer (HGT) could have caused this. To examine this possibility, the Aer phylogenetic tree (Figure 4-2) was compared to the known *Pseudomonas* species phylogeny (33). Additionally, upstream and downstream regions of the *aer* genes were examined to determine whether the genomic context was consistent across species (Figure 4-7) and whether they were associated with mobile elements (Table 4-1). Inspection of the phylogenetic tree revealed 3 sequences that were highly divergent from the expected species phylogeny (Figure 4-2, green stars). These 3 sequences (WP_015271024.1, *P. putida* HB3267; WP_014754514.1, *P. putida* ND6; and WP_013791017.1, *P. fulva* 12-X) were each found clustered with sequences from unrelated species whereas other Aer sequences were consistently clustered with sequences from closely related species. Of these 3 genes only the *aer.g3* gene from *P. putida* ND6 was associated with mobile elements.

The first two upstream and downstream genes of each *aer* gene were identified, as well as any nearby mobile elements (inverted repeats, transposases, integrases) (Table 4-2). The frequency of each of the associated genes and the general genomic context for each Aer homolog were examined to identify the most commonly associated genes (summarized in Figure 4-7). Each homolog was part of an apparently unique operon: Aer.g1 with an aconitate hydratase (69%) and a CAAX aminoprotease (33%), Aer.g2 with a PAS-containing diguanylate cyclase/phosphodiesterase (88%), Aer.g3 with a different PAS-containing diguanylate cyclase/phosphodiesterase (88%) and Aer.g4 with a LysR-type transcriptional regulator (64%) and a DTW-domain containing protein (74%). Beyond these, the second upstream and downstream genes tended to vary more widely, though some were clearly conserved: Aer.g1, 23S rRNA methyl transferase (42%); Aer.g3, C4

dicarboxylate transporter (63%); Aer.g4, agmatinase (56%). Still this indicates clear variability in the genomic context of homologs within each group. Also of note were two other diguanylate cyclases, occurring at much lower frequencies than those mentioned above (Table 4-2). As Aer.g5 had only two members and their genomic contexts were identical, they are not presented here.

Mobile elements were found within 5Kb of *aer* homolog genes (Table 4-1). Inverted repeats were most common (20%), followed by transposases (4%) and integrases (1%). Aer.g5 did not have any mobile elements associated, Aer.g1 was the least common, then Aer.g3, then Aer.g2 and Aer.g4 had the highest frequency of association with mobile elements. When mobile elements were found, they were often located immediately up and/or downstream of the *aer* homolog and its associated gene(s) described in the above section.

Table 4-2: Frequency of occurrence of genes neighbouring *aer* homologs in *Pseudomonas*.

Homolog Group	UpStream.1	Occurrences Within Group	Frequency Within Group (%)
1	aconitate hydratase	32	44
1	CAAX aminoprotease	24	33
1	Hypothetical	5	7
1	hypothetical	3	4
1	protein involved in RimO-mediated methylthiolation of ribosomal protein S12	2	3
1	4-hydroxy-tetrahydrodipicolinate synthase	1	1
1	GDP-mannose-6-dehydrogenase	1	1
1	GntR-Family Transcriptional Regulator	1	1
1	haloacid dehydrogenase	1	1
1	mannose-6-phosphate isomerase	1	1
1	metal ABC transporter substrate-binding protein	1	1
1	PEP synthase regulatory protein	1	1
2	diguanylate cyclase (version 4)	7	27
2	RND transporter MFP subunit	5	19

2	C4 dicarboxylate binding protein	4	15
2	hypothetical	4	15
2	aer	3	12
2	amino acid transporter substrate binding protein	1	4
2	membrane protein	1	4
2	transposase	1	4
3	Transcriptional Regulator	10	63
3	GNAT acetyl transferase	2	13
3	C4 dicarboxylate transporter	1	6
3	C4 dicarboxylate transporter	1	6
3	hypothetical protein	1	6
3	Transposase (2)	1	6
4	LysR-type transcriptional regulator	20	74
4	Topoisomerase	3	11
4	hypothetical	1	4
4	Hypothetical	1	4
4	membrane protein	1	4
4	TonB dependent receptor	1	4
5	ribulose-phosphate 3-epimerase	2	100

Homolog Group	UpStream.2	Occurences Within Group	Frequency Within Group (%)
1	methyltransferase	31	42
1	aconitate hydratase	18	25
1	cystathionine beta-lyase	4	5
1	DMT transport permease	3	4
1	Hypothetical	3	4
1	multidrug DMT transporter	2	3
1	aldehyde dehydrogenase	1	1
1	amidase	1	1
1	bifunctional isocitrate dehydrogenase kinase/phosphatase	1	1
1	dimethylglycine dehydrogenase	1	1
1	glycosyltransferase Alg8	1	1
1	haloacid dehydrogenase	1	1
1	hypothetical	1	1
1	lipase	1	1

1	membrane protein	1	1
1	PEP synthase	1	1
1	sulfurtransferase TusA	1	1
1	zinc ABC transporter permease	1	1
2	acriflavine resistance protein B	5	19
2	C4 dicarboxylate transporter	4	15
2	C4 dicarboxylate binding protein	3	12
2	Hypothetical	3	12
2	AraC family transcriptional regulator	2	8
2	hypothetical	2	8
2	FMN-dependent NADH-azoreductase	1	4
2	GntR family transcriptional regulator (different)	1	4
2	multidrug resistance protein MdtA	1	4
2	porin	1	4
2	RND transporter MFP subunit	1	4
2	sodium dependent phosphate transporter	1	4
2	transposase	1	4
3	C4 dicarboxylate transporter DctA	7	44
3	4-hydroxy-tetrahydrodipicolinate synthase	2	13
3	C4 dicarboxylate transporter	2	13
3	extradiol oxygenase	1	6
3	Glyoxylase	1	6
3	GntR-Family Transcriptional Regulator	1	6
3	GntR Transcriptional regulator	1	6
3	Integrase	1	6
4	agmatinase	15	56
4	xylose isomerase	4	15
4	LysM	3	11
4	diguanylate cyclase (version 1)	1	4
4	DMT transport permease	1	4
4	histidine kinase	1	4
4	Hypothetical	1	4
4	transcriptional regulator	1	4
5	2-dehydropantoate 2-reductase	2	100

Homolog Group	DownStream.1	Occurences Within Group	Frequency Within Group
1	Hypothetical	21	29
1	alpha/beta hydrolase	18	25
1	pseudouridine synthase	15	21
1	ThiF	8	11
1	LysE lysine transporter	2	3
1	acetyltransferase	1	1
1	glutamine amidotransferase	1	1
1	glyoxylase	1	1
1	hydrolase	1	1
1	lysozyme	1	1
1	peptide release factor H	1	1
1	thioesterase	1	1
1	transposase	1	1
1	Transposase	1	1
2	Fe-S oxidoreductase	9	35
2	aer	5	19
2	Hypothetical	5	19
2	sensor histidine kinase	2	8
2	Type IV secretion system protein Vask	2	8
2	diguanylate cyclase	1	4
2	glutamine synthetase	1	4
2	NADH:flavin oxidoreductase	1	4
3	Diguanylate cyclase (version 2)	14	88
3	LysE	2	13
4	DTW	20	74
4	Hypothetical	2	7
4	OM transporter barrel	2	7
4	Pseudo-gene	1	4
4	transposase	1	4
4	Type IV secretion system protein	1	4
5	polyamine ABC transporter substrate-binding protein	2	100

Homolog Group	DownStream.2	Occurences Within Group	Frequency Within Group
1	CytC oxidase	20	27
1	Hypothetical	15	21
1	transcriptional regulator	12	16
1	hypothetical	5	7
1	alpha/beta hydrolase	4	5
1	adenylate kinase	1	1
1	amino acid transporter	1	1
1	damage-inducible protein	1	1
1	diguanylate phosphodiesterase	1	1
1	glyoxylase	1	1
1	GNAT-family acetyltransferase	1	1
1	haloacid dehydrogenase	1	1
1	hydrolase	1	1
1	hypothetical	1	1
1	methyltransferase	1	1
1	MFS transporter	1	1
1	O-methyltransferase	1	1
1	Outer membrane autotransporter barrel	1	1
1	phenylalanine permease	1	1
1	RNA ligase RctB-family	1	1
1	transposase	1	1
1	Transposase	1	1
2	Hypothetical	5	19
2	diguanylate cyclase (version 3)	3	12
2	membrane protein	3	12
2	sensor histidine kinase	3	12
2	DNA binding response regulator	2	8
2	peroxiredoxin	2	8
2	Type IV secretion system protein	2	8
2	CheW	1	4
2	Cro/C1 family transcriptional regulator	1	4
2	Fused spore maturation proteins A/B	1	4
2	glutamine synthase	1	4
2	GNAT-family acetyltransferase	1	4
2	transcriptional regulator	1	4
3	hypothetical	4	25
3	AsnC transcriptional regulator	3	19

3	sigma-54-dependent Fis family transcriptional regulator	3	19
3	amino acid ABC transporter substrate-binding protein	2	13
3	AraC	2	13
3	K ⁺ Transporting ATPase subunit F	1	6
3	leucine-responsive regulatory protein	1	6
4	HlyD	11	41
4	ToIC	7	26
4	serine protease	2	7
4	adhesin lapA membrane protein component	1	4
4	aldehyde dehydrogenase	1	4
4	DTW	1	4
4	hypothetical	1	4
4	Integrase	1	4
4	LysR (different)	1	4
4	quercetin 2,3-dioxygenase	1	4
5	aldehyde dehydrogenase	2	100

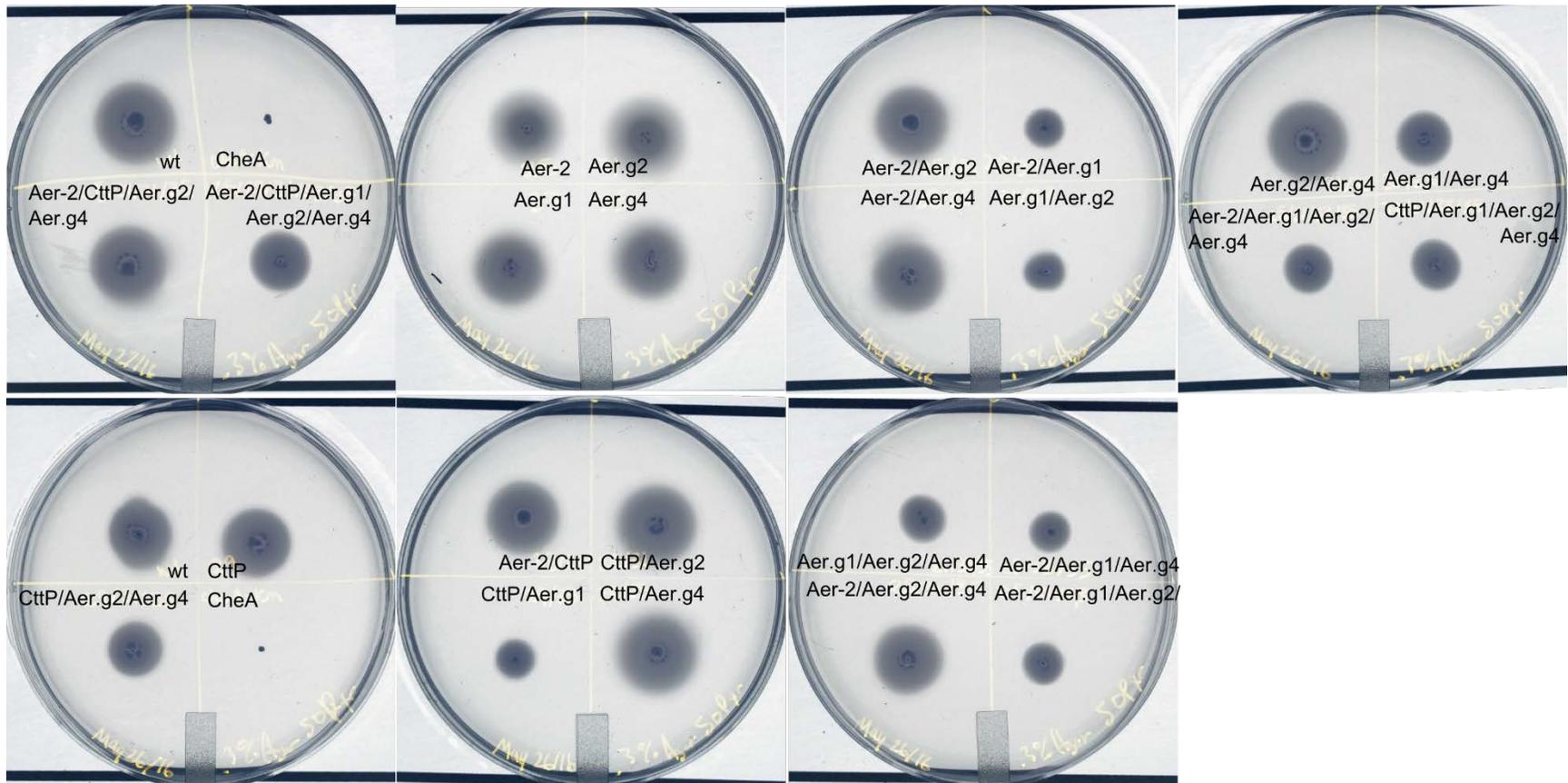


Figure 4-8: Photographs of energy-taxis swim plates after 24h of growth at 30°C in 50mM pyruvate. Energy-taxis swim plates of *P. pseudoaeruginosa* strains with deletions in *aer-2*, *cttP* and *aer* homologs after 24h growth at 30°C in 50mM pyruvate.

4.4.2 Phenotypic Results

To investigate the functions of Aer homologs, CttP and Aer-2, and to see whether they were linked, deletion mutants were constructed in the one species of *Pseudomonas* that possessed 3 Aer homologs Aer-2 and CttP: *P. pseudoalcaligenes* KF707. Single and combination deletions were constructed using 2-step homologous recombination, the suicide vector pG19II, and SacB sucrose counter-selection (35). Energy-taxis was tested using soft agar swim plates (36) (sometimes called ‘swarm’ plates, though the motility is definitely swimming not swarming). In this assay, a small amount of cells are inoculated into a plate by stabbing a needle covered in culture into an agar plate containing a low (0.3%) percentage of agar and a high amount of carbon source (50 mM) in minimal salts medium. As the culture grows and consumes the carbon source at the point of inoculation it will produce a ring as energy-tactic cells seek out a better place to grow. Non-motile or non-energy tactic cells will produce a smaller ring (Figure 4-8). All 24 mutants generated in this study, along with the wild-type and a CheA::KmR insertional inactivation non-motile mutant (37) were tested in triplicate in medium containing pyruvate (Figure 4-9) or succinate (Figure 4-10). Ring diameters were normalized to the wild-type, then tested for differences using Tukey’s Honest Significant Differences test (Table 4-3).

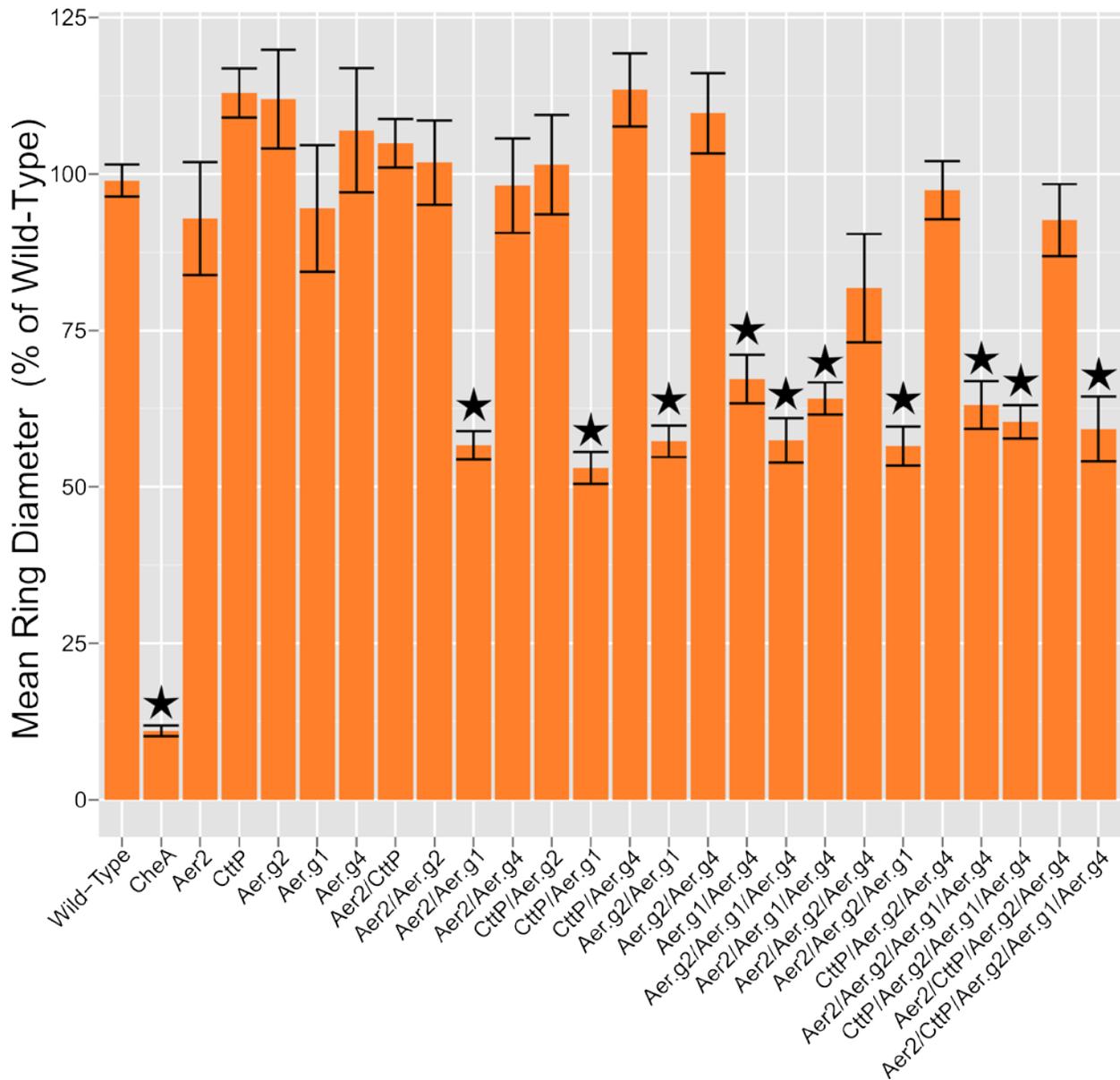


Figure 4-9: Normalized energy-taxis growth diameters in 50mM pyruvate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer2*, *c ttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey’s Honest Significant Differences test with a confidence value of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present.

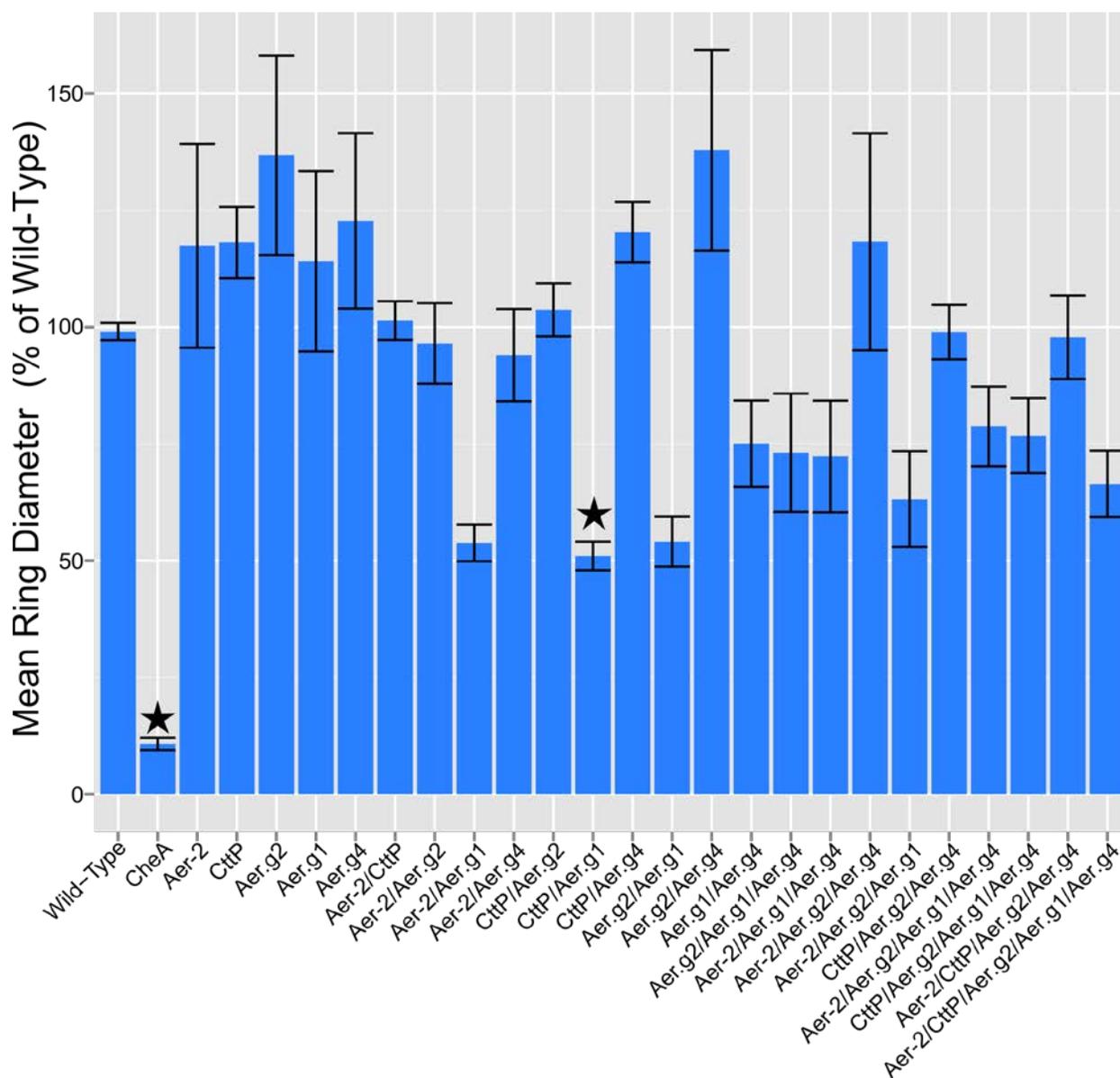


Figure 4-10: Normalized energy-taxis growth diameters in 50mM succinate of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2*, *c ttP* and *aer* homologs. Bars indicate the average growth diameter, normalized to the wild-type, at both 24h and 48h from at least 3 experimental replicates. Wild-type strains were normalized to the mean of technical replicates within each experiment. Error bars indicate standard error. Stars indicate significant differences from the wild-type based on Tukey's Honest Significant Differences test with a confidence value of 0.95. The *cheA::KmR* mutant was not grown with antibiotic present.

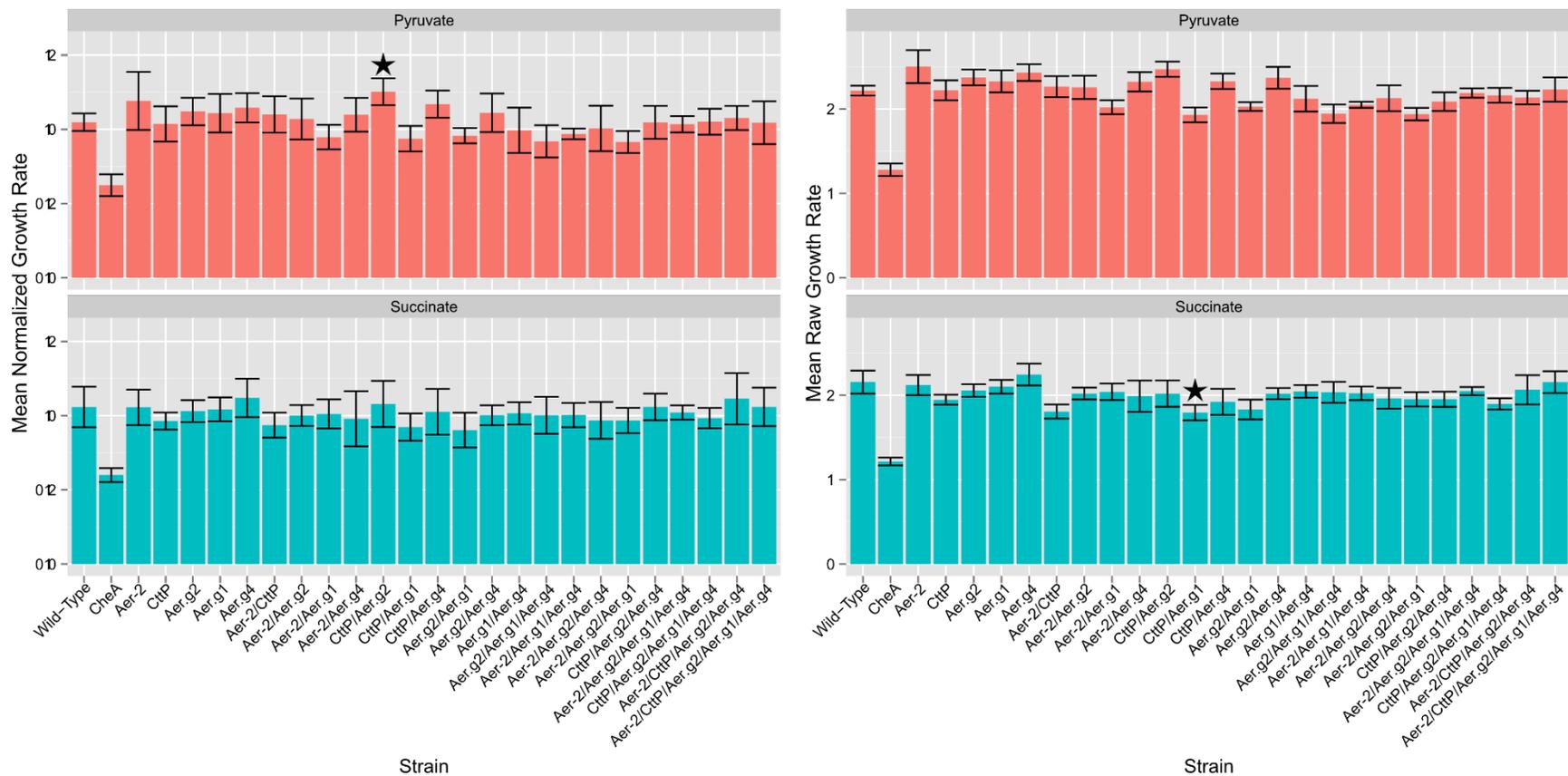


Figure 4-11: Mean raw and normalized growth rates of energy-taxis swim-growth-diameters of strains of *P. pseudoalcaligenes* KF707 with deletions of *aer-2*, *cttP* and *aer* homologs. Growth rates were calculated by dividing the diameter at 48h by the diameter at 24h. Normalized rates were calculated from diameters that were normalized to the wild-type for each experiment at each time point. Stars indicate significant differences from the corresponding wild-type according to Tukey's Honest Significant Differences test.

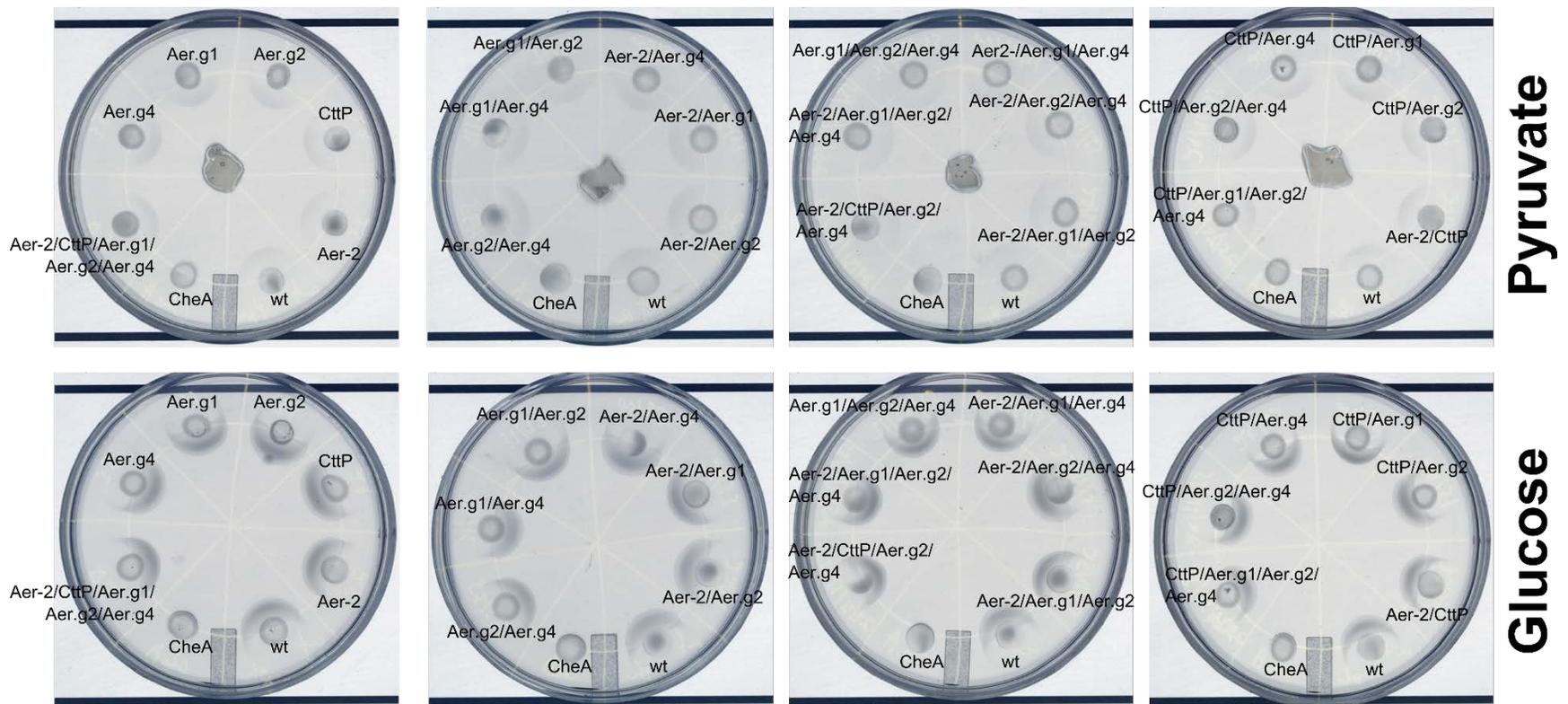


Figure 4-12: Chemotaxis swim assays of *P. pseudoalcaligenes* strains with deletions in *aer-2*, *cttP* and *aer* homologs. Strains were grown overnight, concentrated then spotted on minimal salts plates containing 0.3% agar. Either 50mM pyruvate in 1.5% agar or crystals of glucose were placed in the centre of the plates. Photographs were taken after 24h.

The diameter of the chemotactic negative CheA mutant was about 10% of the wild-type, whereas energy-taxis negative mutants with diameters about 60% smaller were significantly different from the wild-type (Figure 4-9). All the singular deletions of Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP had no effect on energy-taxis; only strains that had Aer.g1 deleted in combination with at least one other receptor were energy-taxis negative. This included both homologs, Aer-2 and CttP. Conversely, deletion of any combination of these receptors, including the quadruple Aer-2/CttP/Aer.g2/Aer.g4 mutant did not adversely affect energy-taxis. The Aer-2/Aer.g2/Aer.g4 triple mutant had a diminished diameter, about 80% of the wild-type, but this was not significant. These results from plates containing pyruvate were similar to those obtained using succinate (Figure 4-10), however those swim diameters were far more variable resulting in larger error bars and lack of significant differences. The observed differences in swim diameters were not due to differences in growth rate as the rate of growth of the swim diameters was similar between all strains (Figure 4-11, Table 4-4). All strains were also tested for chemotaxis towards pyruvate and glucose (Figure 4-12) using traditional swim plate assays, and all except the CheA::KmR mutant were able to swim towards, pyruvate and glucose.

Table 4-3: Tukey Honest Significant Differences results comparing differences in normalized energy-taxi diameter in pyruvate or succinate plates. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant.

Comparison	Pyruvate Difference	Pyruvate Adjusted p Value	Succinate Difference	Succinate Adjusted p Value
Wild-Type-Aer.g1	0.044557106	1	-0.15022	1
Wild-Type-Aer.g1/Aer.g4	0.316929667	0.001584542	0.240355	0.995568
Wild-Type-Aer.g2	-0.13026019	0.994474052	-0.37755	0.470077
Wild-Type-Aer.g2/Aer.g1	0.416623963	6.99E-07	0.449853	0.123956
Wild-Type-Aer.g2/Aer.g1/Aer.g4	0.41531338	7.83E-07	0.259256	0.985363
Wild-Type-Aer.g2/Aer.g4	-0.10765718	0.999855723	-0.38801	0.404151
Wild-Type-Aer.g4	-0.08027211	0.999999891	-0.23695	0.996521
Wild-Type-Aer2	0.06079056	1	-0.18355	0.99998
Wild-Type-Aer2/Aer.g1	0.423034811	3.99E-07	0.452405	0.116907
Wild-Type-Aer2/Aer.g1/Aer.g4	0.348397359	0.000171831	0.267403	0.977193
Wild-Type-Aer2/Aer.g2	-0.02874334	1	0.025494	1
Wild-Type-Aer2/Aer.g2/Aer.g1	0.424269068	3.58E-07	0.35864	0.594415
Wild-Type-Aer2/Aer.g2/Aer.g1/Aer.g4	0.358356796	8.13E-05	0.203102	0.999816
Wild-Type-Aer2/Aer.g2/Aer.g4	0.172011243	0.810645941	-0.19209	0.999944
Wild-Type-Aer2/Aer.g4	0.00831534	1	0.050637	1
Wild-Type-Aer2/CttP	-0.05951297	0.999999999	-0.02338	1
Wild-Type-Aer2/CttP/Aer.g2/Aer.g1/Aer.g4	0.396996279	3.72E-06	0.326159	0.792546

Wild-Type- Aer2/CttP/Aer.g2/Aer.g4	0.06307225 7	0.999999997	0.012354	1
Wild-Type-CheA	0.87948192	0	0.88287	3.16E-10
Wild-Type-CttP	- 0.13987730	0.983128982	-0.19063	0.999953
Wild-Type-CttP/Aer.g1	0.45933186 3	4.31E-11	0.480318	0.011177
Wild-Type-CttP/Aer.g2	- 0.02519549	1	-0.04643	1
Wild-Type- CttP/Aer.g2/Aer.g1/Aer.g4	0.38554092	1.60E-07	0.22296	0.991638
Wild-Type-CttP/Aer.g2/Aer.g4	0.01557938 7	1	0.001138	1
Wild-Type-CttP/Aer.g4	- 0.14504144	0.935787346	-0.21263	0.998431

Table 4-4: P values from Tukey HSD results comparing differences in raw and normalized energy-taxis diameter growth rates in pyruvate or succinate plates. Growth rates were obtained by dividing the raw or normalized diameter at 48h by the value at 24h. HSD test compared all pairs of strains, only comparisons to the wild-type are presented here. P values were computed using a 0.95 confidence value, those below 0.05 were taken as significant.

Comparison	Pyruvate Normalized	Succinate Normalized	Pyruvate Raw	Succinate Raw
Wild-Type-Aer.g1	1	1	1	1
Wild-Type-Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer.g2	1	1	0.999	1
Wild-Type-Aer.g2/Aer.g1	0.997992	0.194316	0.96148	0.37878
Wild-Type-Aer.g2/Aer.g1/Aer.g4	0.826779	1	0.332046	1
Wild-Type-Aer.g2/Aer.g4	1	1	0.999447	1
Wild-Type-Aer.g4	0.999305	0.999923	0.894703	0.999999
Wild-Type-Aer2	0.791526	1	0.305462	1
Wild-Type-Aer2/Aer.g1	0.990515	1	0.937616	1
Wild-Type-Aer2/Aer.g1/Aer.g4	0.999856	1	0.99263	1
Wild-Type-Aer2/Aer.g2	1	1	1	1
Wild-Type-Aer2/Aer.g2/Aer.g1	0.776794	0.994759	0.290882	0.997895
Wild-Type-Aer2/Aer.g2/Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer2/Aer.g2/Aer.g4	1	0.995619	1	0.999439
Wild-Type-Aer2/Aer.g4	1	0.99969	1	0.999984
Wild-Type-Aer2/CttP	1	0.516674	1	0.066568
Wild-Type-Aer2/CttP/Aer.g2/Aer.g1/Aer.g4	1	1	1	1
Wild-Type-Aer2/CttP/Aer.g2/Aer.g4	1	0.999829	1	1
Wild-Type-CheA	0	0	0	0
Wild-Type-CttP	1	0.990882	1	0.996631
Wild-Type-CttP/Aer.g1	0.866421	0.259124	0.070324	0.039548
Wild-Type-CttP/Aer.g2	0.03075	1	0.416662	1
Wild-Type-CttP/Aer.g2/Aer.g1/Aer.g4	1	0.999082	1	0.712737
Wild-Type-CttP/Aer.g2/Aer.g4	1	1	0.999329	0.986805
Wild-Type-CttP/Aer.g4	0.926885	1	0.999998	0.939527

4.4.3 Discussion

In *Pseudomonas* the MCP Aer has been characterized as a receptor for the conflated processes of aerotaxis and energy-taxis. The presence of multiple homologs in some species, of which only one is functional, has led to confusion with a putative oxygen receptor as the former is named Aer2 and the latter Aer-2. Here we combined a bioinformatics analysis of Aer within the genus *Pseudomonas* with a molecular biology characterization of Aer homologs in the one species (*P. pseudoalcaligenes* KF707) that possessed 3 Aer homologs, Aer-2 and the putative tetrachloroethylene receptor CttP, revealing that energy-taxis and aerotaxis are linked but distinct processes and that Aer is a family of patchily distributed receptors in *Pseudomonas*.

4.4.3.1 Sequence and genetics classify Aer homologs into 5 groups

The phylogeny and alignment of Aer sequences implied there were up to 7 different groups, but comparison of sequence differences using SHMR (23) showed there were actually 5. All groups varied in the same regions of the protein, except Aer.g2 which had unique regions at the N and C termini and between the transmembrane helices and conserved methyl-accepting domain. The most conserved domains were unsurprisingly the PAS ligand binding domain and CheW interface region which confer the necessary functions of FAD cofactor binding/signal reception (38) and signal transduction, respectively. Two other conserved regions were identified, on either side of the CheW interface indicating their importance for the core function of the protein. The conserved regions in the signaling region were better conserved than the predicted PAS domain, though these domains are known to differ at the sequence level (38) so it is likely that proteins from the different groups function similarly. Sequence based group assignments were supported by the genomic context of the corresponding genes; there tended to be shared synteny in the immediate vicinity of the *aer* genes from each group, though this also supported the existence of

additional groups. Including the branch of Aer.g1 which was initially classified as its own group (Figure 1, dotted box), within Aer.g1 was supported by the SHMR analysis however these genes did not share synteny with the rest of Aer.g1. These sequences represent the only duplications of Aer.g1 and only occurred in species which had no other Aer homologs (except *P. azotoformans* S4). As the protein sequences do not differ as much as the other groups, and the genomic context within this group differed, this may represent a recent duplication event within these lineages. Phenotypic characterization of the two Aer receptors in these species could be particularly enlightening to understanding why only some *Pseudomonas* species have extra *aer* genes.

4.4.3.2 Aer.g1 is ancestral, prevalent and functional

The Aer.g1 homolog group was the closest to the root, with the exception of Aer from *P. rhizospherae*, indicating this group is ancestral to the other Aer groups. It was also the most prevalent as all the *Pseudomonas* species included in this study had at least one homolog from this group except *P. denitrificans* ATCC13867. This means that Aer is not part of the *Pseudomonas* core genome, though it is likely ancestral and was lost in *P. denitrificans*. Interestingly this also implies that the oxygen receptor Aer-2 is sufficient for the lifestyle of this organism, named for its ability to use a terminal electron acceptor other than oxygen. All of the previously characterized functional Aer homologs, from *P. aeruginosa* PA01 (8), *P. putida* F1 (39) and *P. putida* KT2440 (9) are part of the homolog group Aer.g1, which was demonstrated to be pivotal in *P. pseudoalcaligenes* KF707. The functional Aer homolog from *P. putida* PRS2000 (11) was not included in this study as this species genome has not been fully sequenced. The receptors from *P. putida* F1 and KT2440 were both named Aer2 but the phylogeny presented here indicates they belong to the Aer.g1 homolog group. As KF707 is the only strain where the other Aer homologs

have been shown to have semi-redundant functionality with Aer.g1, it is safe to predict that the Aer.g1 homolog from other species will be the receptor responsible for energy-taxis.

Almost all *aer.g1* genes were associated with an aconitase gene, indicating that they are likely co-transcribed. In *Helicobacter pylori*, TlpD, which controls tactic behaviour in low-energy conditions has been found to interact with aconitase (40). Our results thus indicate that association with aconitase is a common feature of energy-taxis receptors. While Aer cannot obtain electrons for reducing the FAD co-factors from aconitase as the isomerization of citrate to isocitrate is a non-redox process, this suggests that isocitrate dehydrogenase may be the step of the citric acid cycle that is being sensed for energy-taxis.

A gene encoding a predicted CAAX amino protease was found adjacent to one third of Aer.g1 homologs. This type of membrane-bound protease aids in prenylating proteins to ease their membrane localization (41). The associated Aer.g1 proteins did not have the expected CAAX motif though they had high sequence similarity to each other. There were no obvious features at the sequence level that distinguished these sequences completely from other Aer homologs indicating that the genetically associated CAAX protease may not necessarily be involved in their functioning.

4.4.3.3 Other Aer groups likely have unique functions

As the 5 characterized Aer.g1 homologs appear to be broad-purpose energy-taxis sensors, this implies that the other homolog groups may have a related, but different or more specific function. The three Aer homologs of *P. putida* KT2440 were found to be differentially expressed (9), supporting this hypothesis. In *E. coli* Aer is also a thermosensor (42) indicating another possible role or they could aid in tuning the energy-taxis response similarly to how Tsr and Tar from *E. coli* enable taxis towards the ideal pH from both higher and lower initial pHs (43). The

function of the additional Aer homologs could also be related to biofilm formation as Aer.g2 and Aer.g3 were associated with their own PAS-domain containing diguanylate cyclase/phosphodiesterase. Deletion of the diguanylate cyclase/phosphodiesterase associated with Aer.g3 in *P. putida* KT2440 caused a general defect in motility (9) and was later shown to be a bifunctional cyclic di-GMP phosphodiesterase and diguanylate cyclase formation (44). These two homolog groups may thus be involved in energy-sensing behaviour in biofilms as c-di-GMP is an important regulator in the transition from planktonic to sessile growth modes, though it will be interesting to see if the putative diguanylate cyclase/phosphodiesterase associated with Aer.g2 is also bifunctional and whether they both respond to the same signal. Aer.g4 could also have its own distinct function, possibly transducing its signal not only to the *che* pathway but also to the transcriptional regulator adjacent to it allowing for gene regulation in response to energy-sensing. The function of Aer.g5 may be even more distant as it is the least similar from the others, though it is likely a specific niche function since only two of the included species possessed copies. This receptor definitely evolved in *Pseudomonas* though as all its closest relatives (by BLAST) are *Pseudomonas* proteins. These results thus raise many questions about energy sensing and how it is integrated into other cellular responses than flagellar swimming.

4.4.3.4 *aer* has been horizontally transferred

Multiple lines of evidence indicate that genes from the *aer* group have been horizontally transferred. Foremost is the observed variation at the strain level as there were several noted examples of differences in the number of homologs possessed by strains of the same species. Three specific sequences, from *P. putida* HB3267, ND6 and *P. fulva* 12-X clustered beside Aer sequences from unrelated species indicating they may have been obtained from those species. The observation that many strains did not have the same complement of Aer homologs also supports

this notion. For example, of the 3 *P. pseudoalcaligenes* strains examined here, each one had a drastically different complement of genes (CECT5344, *aer.g1*; AD6, *aer.g1/aer.g4/aer.g5*; KF707, *aer.g1/aer.g2/aer.g4/aer-2/cttP*). This demonstrates the ancestrality of *aer.g1* and the variability of other members, but also makes KF707 an outlier from the species. That strain KF707 has *aer-2* and *cttP* indicates either that these have also been transferred horizontally or that KF707 is more closely related to *P. aeruginosa* than the other strains. As only the *aeruginosa* sub-group has those two genes they were also likely horizontally transferred, albeit less recently than the various *aer* homologs.

The observed variation in the distribution of *Aer* homologs is unsurprising as even within the *P. aeruginosa* pan-genome from 7 strains there are 2,000 accessory genes compared to 5,000 core (45). The fact that *aer.g1* almost qualifies as part of the genus core genome (*P. denitrificans* being the only outlier) indicates that is a very useful gene in the varied environments inhabited by *Pseudomonas*. The existence of the various homologs speaks to their utility as they each may confer a more specific function. As many *Pseudomonas* species live in the soil and rhizosphere, which promotes horizontal gene transfer (46), there would be ample opportunity for these *aer* homolog genes to spread. Interestingly they appear to be restricted to *Pseudomonas* still as reverse BLAST searches consistently turned up *Pseudomonas* sequences as the closest hits. Even more compelling are the associated genes which were transferred alongside. The bifunctional diguanylate cyclase/phosphodiesterase adjacent to *aer.g3* in *P. putida* KT2440 influences motility (9) but this is only 1 of 4 c-di-GMP processing gene products found near *aer* homologs. The function of these genes and the transcriptional regulator and DTW-domain protein beside *aer.g4* are unknown and likely represent interesting connections of energy-sensing to other cellular processes.

4.4.3.5 Aer2 and CttP contribute to energy-taxis in *P. pseudoaligenes* KF707

CttP was characterized as a receptor for positive chemotaxis towards chlorinated ethylenes (13), but this function may be adventitious, similar to how Tar and Tsr from *E. coli* can mediate attractant and repellent responses to phenol through its direct interaction with the transmembrane helices or cytoplasmic HAMP signaling domains (47). Here we showed that co-deletion of *cttP* and *aer.g1* in *P. pseudoalcaligenes* KF707 abolished energy-taxis indicating that it is also an energy-taxis receptor. This is likely untrue though as prediction of the domain architecture of CttP indicates that it begins with two-transmembrane helices and does not have an N-terminal ligand binding domain like other MCPs. Instead it has the unusual feature of a long (>100 AA) C-terminal extension after the CheW interface region. It could thus be possible that CttP does not directly sense any signal, but assists in forming MCP receptor complexes and/or linking them with the Che complex. Overexpression of CttP (then called McpA) abolished all chemotaxis in both *P. aeruginosa* and *E. coli* (36) further implicating it as general chemotaxis protein. MCPs have been proposed to amplify signals by propagating ligand-binding induced changes to adjacent MCPs (48). CttP could be performing such a function as it is expressed in stationary phase (49) and overexpression induces a smooth-swimming phenotype (36).

The same abolishment of chemotaxis was observed when Aer2 (McpB) was overexpressed in *E. coli* and *P. aeruginosa* (36). This could be related to its function of being necessary for the complexation of the other Che proteins located in the same operon *che2* (49). Along with this necessity, Aer2 has a C-terminal extension that specifically allows CheR2, and only CheR2, to methylate it (50), making it appear that the entire Che2 pathway exists to transduce the signal sensed by Aer2. In the current study we found that deletion of either *cttP* (*mcpA*) or *aer2* (*mcpB*) in combination with *aer.g1* abolished energy-taxis indicating that there is a connection between

the receptors from the *che2* operon and energy-taxis. As these receptors are only present in a narrow subsection of *Pseudomonas* species (the ‘aeruginosa’ subgroup), which varied widely in their number of Aer homologs, it is unlikely that this is a direct connection. The *in vitro* biochemical studies of Aer2 indicating that it binds oxygen and the demonstration that its expression in *E. coli* causes a repellent response to oxygen (10), along with the original characterization of Aer2 as an aerotaxis receptor (8), strongly indicate that it is a true aerotaxis receptor that binds oxygen. This fits with the observations we made in *P. pseudoalcaligenes* KF707 where the deletion of *aer.g1* and *aer2* resulted in an energy-tactic negative phenotype. In the soft agar energy-taxis swim plate assay used in this study, oxygen as well as metabolizable carbon source would be diminished at the center of the growth ring. This would allow the energy-sensing of Aer and oxygen sensing of Aer2 to work together, e.g. in the *Aer.g1* and *Aer.g2/Aer.g4* mutants that did not display diminished energy-taxis.

4.5 Conclusions

In *Pseudomonas*, Aer does not simply refer to a single receptor for energy-taxis, but to a family of receptors of mostly unknown function. The most ancestral Aer, here named *Aer.g1*, is indeed an energy-taxis receptor, though other proteins may contribute to its function, including other Aer family members. In *P. pseudoalcaligenes* KF707, unique among *Pseudomonas* species for having 3 Aer homologs, *Aer-2* and *CttP*, also uses these other receptors for energy-taxis. As most species do not have these receptors and other evidence indicates their true purpose, this may be a minor function. Apart from *Aer.g1*, Aer family members, *Aer-2* and *CttP* have likely been horizontally transferred within the genus indicating that all these receptors play important roles in *Pseudomonas* physiology.

4.6 References

1. **Wadhams GH, Armitage JP.** 2004. Making sense of it all: Bacterial chemotaxis.
2. **Sampedro I, Parales RE, Krell T, Hill JE.** 2015. Pseudomonas chemotaxis. *FEMS Microbiol Rev* **39**:17–46.
3. **Krell T.** 2015. Tackling the bottleneck in bacterial signal transduction research: high-throughput identification of signal molecules. *Mol Microbiol* **96**:685–8.
4. **Beijerinck MW.** 1893. Ueber atungsfiguren beweglicher bakterien. *Zentrabl Bakteriol Parasitenkd* **14**:827–845.
5. **Engelmann TW.** 1881. Neue Methode zur Untersuchung der Sauerstoffausscheidung pflanzlicher und thierischer Organismen. *Pflüger, Arch für die Gesamte Physiol des Menschen und der Thiere* **25**:285–292.
6. **Bibikov SI, Biran R, Rudd KE, Parkinson JS.** 1997. A signal transducer for aerotaxis in *Escherichia coli*. *J Bacteriol* **179**:4075–4079.
7. **Rebbapragada A, Johnson MS, Harding GP, Zuccarelli AJ, Fletcher HM, Zhulin IB, Taylor BL.** 1997. The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for *Escherichia coli* behavior. *Proc Natl Acad Sci U S A* **94**:10541–6.
8. **Hong CS, Shitashiro M, Kuroda A, Ikeda T, Takiguchi N, Ohtake H, Kato J.** 2004. Chemotaxis proteins and transducers for aerotaxis in *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* **231**:247–252.
9. **Sarand I, Österberg S, Holmqvist S, Holmfeldt P, Skärfstad E, Parales RE, Shingler V.** 2008. Metabolism-dependent taxis towards (methyl)phenols is coupled through the most abundant of three polar localized Aer-like proteins of *Pseudomonas putida*. *Environ*

- Microbiol **10**:1320–1334.
10. **Watts KJ, Taylor BL, Johnson MS.** 2011. PAS/poly-HAMP signalling in Aer-2, a soluble haem-based sensor. *Mol Microbiol* **79**:686–99.
 11. **Nichols NN, Harwood CS.** 2000. An aerotaxis transducer gene from *Pseudomonas putida*. *FEMS Microbiol Lett* **182**:177–183.
 12. **Parales RE, Luu R a., Chen GY, Liu X, Wu V, Lin P, Hughes JG, Nesteryuk V, Parales J V., Ditty JL.** 2013. *Pseudomonas putida* F1 has multiple chemoreceptors with overlapping specificity for organic acids. *Microbiol (United Kingdom)* **159**:1086–1096.
 13. **Shitashiro M, Tanaka H, Hong CS, Kuroda A, Takiguchi N, Ohtake H, Kato J.** 2005. Identification of chemosensory proteins for trichloroethylene in *Pseudomonas aeruginosa*. *J Biosci Bioeng* **99**:396–402.
 14. **Tatusova T, Ciufu S, Fedorov B, O’Neill K, Tolstoy I.** 2014. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res* **42**:D553-9.
 15. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–10.
 16. **Cochrane G, Karsch-Mizrachi I, Takagi T, International Nucleotide Sequence Database Collaboration IN.** 2016. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res* **44**:D48-50.
 17. **Papadopoulos JS, Agarwala R.** 2007. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**:1073–1079.
 18. **Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ.** 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–91.

19. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797.
20. **R Core Team.** 2014. R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria 2014.
21. **Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C.** 2009. Mobyle: a new full web bioinformatics framework. *Bioinformatics* **25**:3005–3011.
22. **Guindon S, Gascuel O.** 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* **52**:696–704.
23. **Brandt BW, Feenstra KA, Heringa J.** 2010. Multi-Harmony: detecting functional specificity from sequence alignment. *Nucleic Acids Res* **38**:W35-40.
24. **Wickham H.** 2011. ggplot2. *Wiley Interdiscip Rev Comput Stat* **3**:180–185.
25. **Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH.** 2015. CDD: NCBI’s conserved domain database. *Nucleic Acids Res* **43**:D222-6.
26. **Triscari-Barberi T, Simone D, Calabrese FM, Attimonelli M, Hahn KR, Amoako KK, Turner RJ, Fedi S, Zannoni D.** 2012. Genome sequence of the polychlorinated-biphenyl degrader *Pseudomonas pseudoalcaligenes* KF707. *J Bacteriol* **194**:4426–4427.
27. **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL.** 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**:134.
28. **Sambrook J, Fritsch EF, Maniatis T.** 1989. No Title. *Mol Biol A Lab Man.*

29. **Maseda H, Sawada I, Saito K, Uchiyama H, Nakae T, Nomura N.** 2004. Enhancement of the mexAB-oprM efflux pump expression by a quorum-sensing autoinducer and its cancellation by a regulator, MexT, of the mexEF-oprN efflux pump operon in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **48**:1320–8.
30. **Figurski DH, Helinski DR.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**:1648–52.
31. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**:671–675.
32. **Bodilis J, Nsigue Meilo S, Cornelis P, De Vos P, Barray S.** 2011. A long-branch attraction artifact reveals an adaptive radiation in *Pseudomonas*. *Mol Biol Evol* **28**:2723–6.
33. **Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E.** 2015. Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* **6**:214.
34. **Livermore JA, Emrich SJ, Tan J, Jones SE.** 2014. Freshwater bacterial lifestyles inferred from comparative genomics. *Environ Microbiol* **16**:746–758.
35. **Hmelo LR, Borlee BR, Almblad H, Love ME, Randall TE, Tseng BS, Lin C, Irie Y, Storek KM, Yang JJ, Siehnel RJ, Howell PL, Singh PK, Tolker-Nielsen T, Parsek MR, Schweizer HP, Harrison JJ.** 2015. Precision-engineering the *Pseudomonas aeruginosa* genome with two-step allelic exchange. *Nat Protoc* **10**:1820–1841.
36. **Ferrandez A, Hawkins AC, Summerfield DT, Harwood CS.** 2002. Cluster II che Genes from *Pseudomonas aeruginosa* Are Required for an Optimal Chemotactic Response. *J Bacteriol* **184**:4374–4383.

37. **Tremaroli V, Fedi S, Tamburini S, Viti C, Tatti E, Ceri H, Turner RJ, Zannoni D.** 2011. A histidine-kinase cheA gene of *Pseudomonas pseudoalcaligenes* KF707 not only has a key role in chemotaxis but also affects biofilm formation and cell metabolism. *Biofouling* **27**:33–46.
38. **Henry JT, Crosson S.** 2011. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu Rev Microbiol* **65**:261–286.
39. **Luu R a., Schneider BJ, Ho CC, Nesteryuk V, Ngwesse SE, Liu X, Parales J V., Ditty JL, Parales RE.** 2013. Taxis of *Pseudomonas putida* F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. *Appl Environ Microbiol* **79**:2416–2423.
40. **Behrens W, Schweinitzer T, McMurry JL, Loewen PC, Buettner FFR, Menz S, Josenhans C.** 2016. Localisation and protein-protein interactions of the *Helicobacter pylori* taxis sensor TlpD and their connection to metabolic functions. *Sci Rep* **6**:23582.
41. **Pei J, Mitchell DA, Dixon JE, Grishin N V.** 2011. Expansion of Type II CAAX Proteases Reveals Evolutionary Origin of γ -Secretase Subunit APH-1 *Journal of Molecular Biology*.
42. **Nishiyama S, Ohno S, Ohta N, Inoue Y, Fukuoka H, Ishijima A, Kawagishi I.** 2010. Thermosensing function of the *Escherichia coli* redox sensor Aer. *J Bacteriol* **192**:1740–3.
43. **Yang Y, Sourjik V.** 2012. Opposite responses by different chemoreceptors set a tunable preference point in *Escherichia coli* pH taxis. *Mol Microbiol* **86**:1482–1489.
44. **Österberg S, Åberg A, Herrera Seitz MK, Wolf-Watz M, Shingler V.** 2013. Genetic dissection of a motility-associated c-di-GMP signalling protein of *Pseudomonas putida*. *Environ Microbiol Rep* **5**:556–65.
45. **Qiu X, Kulasekara BR, Lory S.** 2009. Role of Horizontal Gene Transfer in the Evolution

- of *Pseudomonas aeruginosa* Virulence. *Genome Dyn* **6**:126–39.
46. **Sengeløv G, Kristensen KJ, Sørensen AH, Kroer N, Sørensen SJ.** 2001. Effect of Genomic Location on Horizontal Transfer of a Recombinant Gene Cassette Between *Pseudomonas* Strains in the Rhizosphere and Spherosphere of Barley Seedlings. *Curr Microbiol* **42**:160–167.
 47. **Pham HT, Parkinson JS.** 2011. Phenol sensing by *Escherichia coli* chemoreceptors: a nonclassical mechanism. *J Bacteriol* **193**:6597–604.
 48. **Wolanin PM, Stock JB.** 2004. Bacterial Chemosensing: Cooperative Molecular Logic. *Curr Biol* **14**:R486–R487.
 49. **Güvener ZT, Tifrea DF, Harwood CS.** 2006. Two different *Pseudomonas aeruginosa* chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. *Mol Microbiol* **61**:106–18.
 50. **García-Fontana C, Corral Lugo A, Krell T.** 2014. Specificity of the CheR2 methyltransferase in *Pseudomonas aeruginosa* is directed by a C-terminal pentapeptide in the McpB chemoreceptor. *Sci Signal* **7**:ra34.

Chapter 5 Preface

This chapter addresses the goals and specific aims set out in the introduction and integrates the conclusions from the other chapters together in the grander scheme of microbiological research. It begins with follow-up results from *Chapter 4*, regarding whether energy-taxis is involved with chemotaxis towards biphenyl in *P. pseudoalcaligenes* KF707. These results are then discussed in the general context of chemotaxis research. The implications of the results from *Chapter 2* are then discussed with relation to other microbiological studies involving adjusting media composition. *Chapter 3* represented a fundamental science approach to an applied research problem so its results are discussed in the context of applying them to improve bioremediation of co-contaminated environments. Finally, all the results from this thesis are discussed in the broader context of scientific research and how they can be used to move forward.

Chapter Five: Overall Discussion and Conclusions

5.1 Chemotaxis

5.1.1 Energy Taxis in *Pseudomonas pseudoalcaligenes* KF707

P. pseudoalcaligenes KF707 was observed to be chemotactic towards biphenyl (78). It was hypothesized that this taxis was due to energy-taxis as it can use biphenyl as a sole source of carbon and energy. To rule this out, an energy-tactic deficient strain was made by deleting genes involved in energy-taxis, presented in *Chapter Four*. All of the strains of KF707 with deletions of *aer* homologs, *aer-2* and *cttP* were also tested for chemotaxis towards biphenyl (Figure 5-1). All strains, including those that were energy-taxis deficient such as the quintuple *cttP/aer-2/aer.g1/aer.g2/aer.g4* mutant were still capable of swimming towards biphenyl. Only the non-chemotactic *cheA* mutant was unable to swim towards biphenyl, as expected. This indicates that chemotaxis towards biphenyl by *P. pseudoalcaligenes* KF707 is not solely due to energy-taxis.

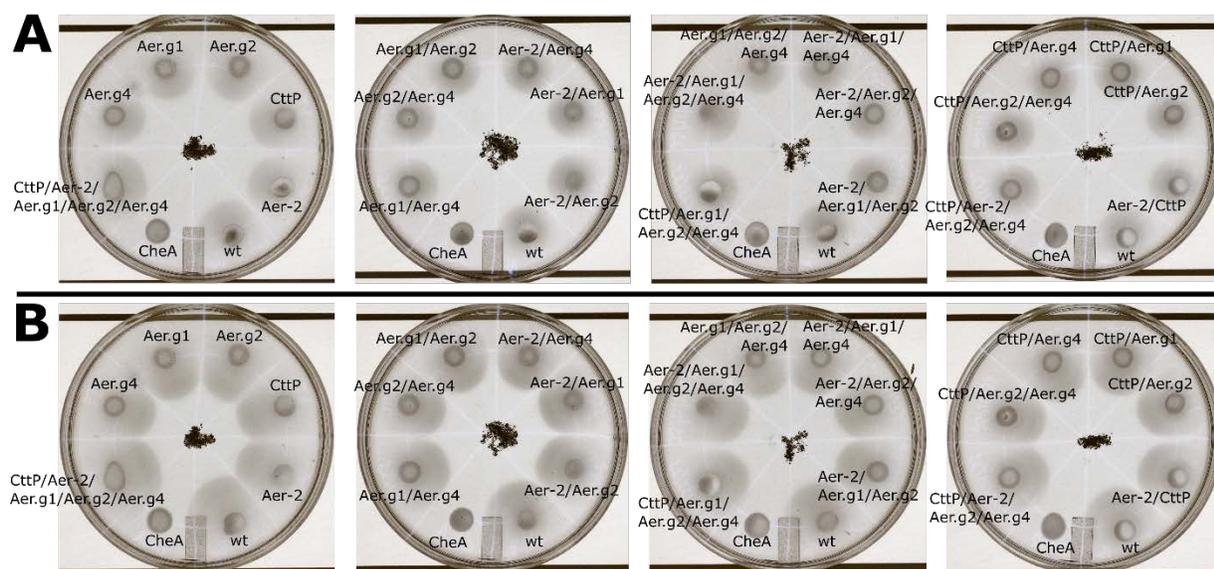


Figure 5-1: Chemotaxis swim plates showing taxis of *P. pseudoalcaligenes* KF707 wild-type and deletion mutants towards biphenyl after 24h (A) and 48h (B).

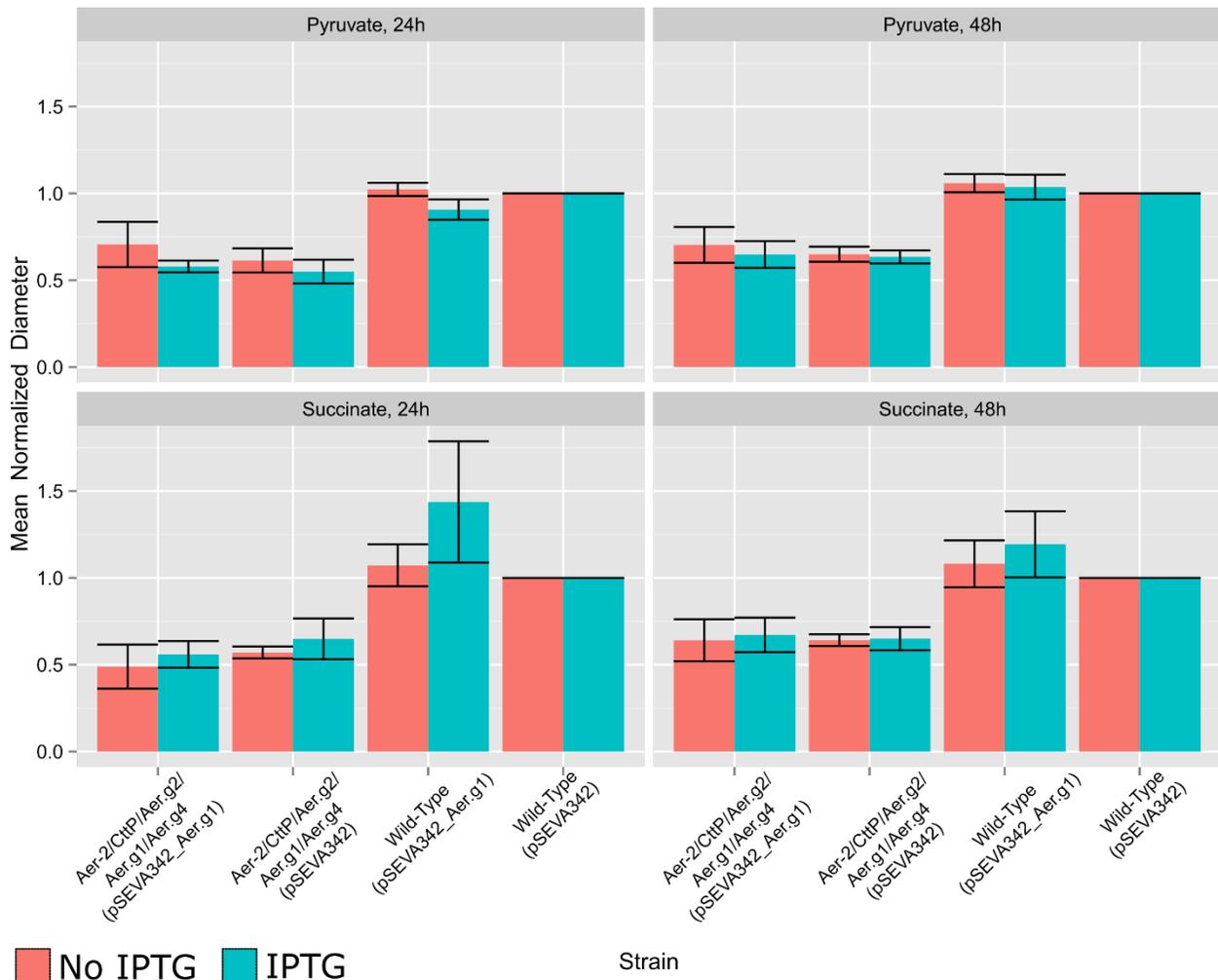


Figure 5-2: Normalized energy-taxis growth diameters in 50mM pyruvate or succinate strains of *P. pseudoalcaligenes* KF707 wild-type and quintuple *cttP/aer-2/aer.g1/aer.g2/aer.g4* mutant complemented with pSEVA342 or pSEVA342_Aer.g1. Bars indicate the average growth diameter, normalized to the wild-type, from 3 experimental replicates. Error bars indicate standard error.

This investigation into energy-taxis also revealed that *Aer.g1* is the most important *aer* homolog in *P. pseudoalcaligenes* KF707 as deletion of it in combination with any other resulted in an energy-tactic negative phenotype. It was surprising to observe that the single deletion of *aer.g1*, and the deletion of all other tested genes except *aer.g1* also produced an energy-tactic negative phenotype. This implied that *Aer.g1* by itself was sufficient to enable energy-taxis. To confirm this, the *aer.g1* gene was cloned into the complementation vector, pSEVA342 (113) which

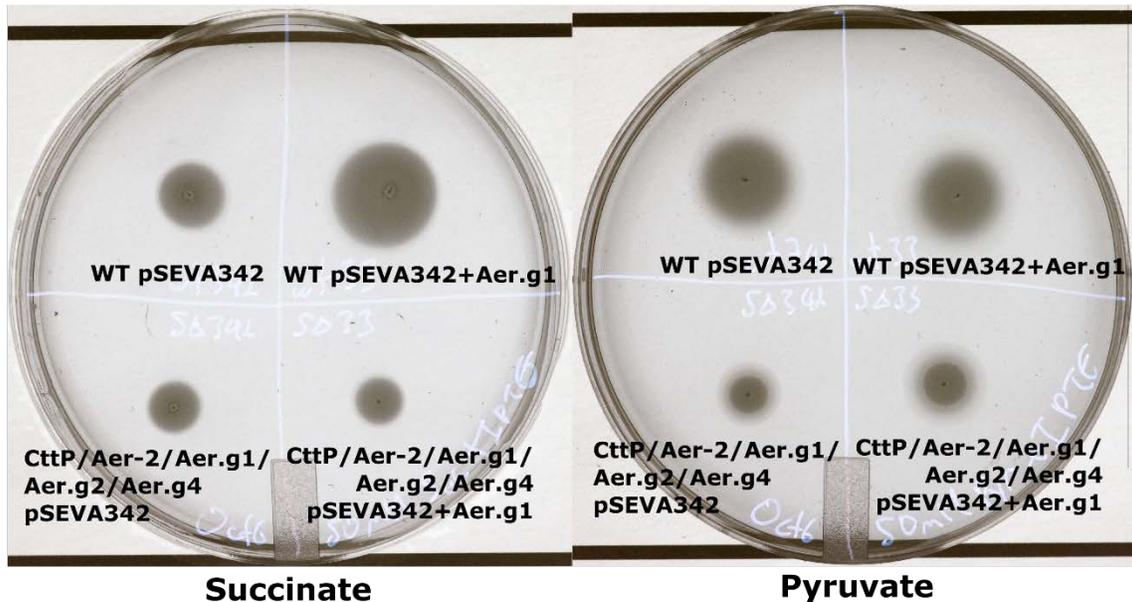


Figure 5-3: Energy-taxis swim plates showing exemplary results of *P. pseudoalcaligenes* KF707 wild-type and quintuple *cttP/aer-2/aer.g1/aer.g2/aer.g4* mutant complemented with pSEVA342 or pSEVA342_Aer.g1 in 50mM pyruvate with 1mM IPTG after 48h.

was transformed into both the wild-type and quintuple *cttP/aer-2/aer.g1/aer.g2/aer.g4* mutant (See additional methods in Appendix D). These complemented strains were tested the same as the deletion strains in energy-taxis swim plates, but the complemented strains did not have consistently larger diameters than the empty vector controls (Figure 5-2). Addition of IPTG at 1mM to induce expression of *aer.g1* also had no apparent effect. Results where the complemented strains had clearly larger swim diameters were obtainable (Figure 5-3) but not consistently reproducible. This was somewhat expected as complementation rarely produces a 100% phenotypic rescue, and it is not atypical to see less than 60% rescue with a plasmid approach. Although such an experiment has been routine in molecular microbiology, it is becoming less relevant in the sense that the complementation of the gene on a variable copy number plasmid, under the control of a non-native promoter is a far less controlled system. The precision genomic deletions that were performed to obtain the energy-taxis negative phenotype are less prone to secondary mutations/genetic

regulatory issues, unlike an insertional inactivation mutation, the deletions were carefully planned to minimally disrupt the upstream and downstream genes and were confirmed by colony-PCR and sequencing. It may be argued that the observed phenotypes could have been caused by some other, unknown, secondary effect somewhere else in the genome, but the observed phenotypes were expected given the known functions of the deleted genes.

5.1.2 Chemotaxis towards Biphenyl

One of the questions addressed in this thesis was: does *P. pseudoalcaligenes* KF707 use energy-taxis to swim towards biphenyl? The answer, is “No”. This indicates that there may be a methyl-accepting chemotaxis protein (MCP) specific for biphenyl. The NahY receptor from *P. putida* G7, which mediates taxis towards naphthalene also enabled taxis towards biphenyl despite this being unusable as a growth substrate for this species (114). A BLAST search of NahY against the *P. pseudoalcaligenes* KF707 genome revealed no close homologs (Table 1.2), indicating that chemotaxis towards biphenyl in this organism must involve a different receptor. In *Comamonas testosteroni* CNB-1 an MCP that binds all the major TCA cycle intermediates was found to enable its chemotaxis towards a range of monoaromatic compounds (115). This suggests it is possible that KF707 uses an MCP for TCA cycle intermediates for chemotaxis towards biphenyl. Based on BLAST searches (Table 1.2): it has a likely homolog of the malate receptor from *P. aeruginosa* PAO1 PA2652 (92); the malate and succinate receptor of *P. fluorescens* Pf0-1 Pfl01_3768 (96); and two possible homologs of McpQRS/McfQRS from *P. putida* (94). If deletion of these genes from *P. pseudoalcaligenes* KF707 abolished taxis towards biphenyl, then this hypothesis would be confirmed. It also remains possible that it has an MCP that is specific to biphenyl or one of its direct metabolites. Bacterial chemotaxis towards organic pollutants has been identified as a potential way to improve bioremediation (116) so finding this putative biphenyl MCP could lead

to its use in enhancing bioremediation efforts. It would be interesting to see how PCBs bind to the biphenyl MCP as *Pseudomonas* sp. B4 was found to be chemotactic towards some mono- and one di-chlorinated congener (118). The existence of a biphenyl MCP would also raise the question of how such a receptor evolved. MCPs for pollutants exist such as NahY, NbaY which mediates taxis towards 2-nitrobenzoic acid and McpT for toluene but it remains unclear what the function of these receptors are in natural ecosystems as their cognate ligands would not be expected concentrated except in environments impacted by human activities.

5.1.3 Chemotaxis in the Natural Environment

Bacterial chemotaxis and motility in natural environments are subject to far different influences than controlled setups in the laboratory. Liquid and semi-solid media used for chemotaxis are not representative of the natural environment where there will be a mixture of solids and liquids, a gradient between fully liquid and fully solid materials and these environments are also characterized by altered physical parameters, such as anisotropy which will affect bacterial motility (119). This characteristic, where a material's properties depend on the direction they're measured in, can alter swimming behaviour. When a medium is too solid for flagellar swimming to work, bacteria including *Pseudomonas* can resort to twitching motility where a type IV pilus is extended, attached to the substrate and then retracted, pulling the cell along (81). While not fully characterized, it is hypothesized that *Pseudomonas* can use this type of motility for chemotaxis using a specific set of *che* analogs to modulate the response. It is unclear whether such a system would have its own cognate MCPs or whether they can be shared between the two systems. Cultures of *P. aeruginosa* experimentally evolved to specialize in either twitching or swimming motility became deficient in the opposite type indicating a genetic trade-off between behaviours (120). If some MCPs were specific to twitching motility this could help explain why *Pseudomonas*

species have so many (>25) of these genes (81). Even if 10 MCPs were specific to twitching motility this would mean that a cell in a natural environment would still have to integrate the signals from 15 MCPs, each measuring the concentration of a different attractant or repellent. In *P. putida* KT2440 the expression of some MCPs is induced or repressed by their cognate ligand or have their expression altered based on growth phase (121). Still, many MCPs for different compounds are co-expressed. This would pose a problem to the cell as it could be driven in different directions by gradients of different attractants/repellents. Understanding how such conflicting signals could lead to an optimized response remains an unexplored field.

5.1.4 Function of Chemoreceptors in Pseudomonas

Pseudomonas species have been found to be chemotactic towards over 100 different specific compounds ranging from simple, like molecular oxygen and acetate, to complex molecules like disaccharides and oligopeptides (81). While the functions of over 30 MCPs have been determined, in the single species with the most known receptors, *P. putida* KT2440 there are still over 10 (of 24 total) MCPs with unknown ligands. Progress has been made using a high-throughput *in vitro* screening technique where the ligand binding regions have been heterologously expressed and purified (122), but this technique will likely not succeed in determining the function of all genes that are predicted to be MCPs in this species. Additionally, it is currently unknown how many different MCPs exist in the *Pseudomonas* pangenome that would also have to be screened. It bears thinking: what is the function of these unknown MCPs, in *P. putida* KT2440, and the rest of the genus? The approach taken by Krell *et al*, where they exogenously expressed, purified and screened for ligand interactions in high-throughput, has been successful, but may soon run out of low-hanging fruit. First, only MCPs with easily predictable periplasmic binding domains can be investigated this way. This makes up a large percentage (~80%) of MCPs (110), but the function

of transmembrane MCPs like Aer and CtpP, and cytosolic ones like Aer-2 have interesting functions that would not be detected by this screen. It is possible that one of these non-standard MCPs could be a sensor for oxidative stress in the cell, allowing it to swim away from dangerous environments, like a repellent form of energy-taxis. Second, if ligand-binding requires a periplasmic-binding protein, such as with the phosphate receptor CtpL, then its function will not be determinable in an *in vitro* system devoid of the unknown periplasmic binding proteins. A likely function for one of these unknown MCPs is as an iron receptor, but it may actually interact with a siderophore-bound form only. Finally, these screens are limited by the availability of pure chemicals to test. *Pseudomonas* species have been found to also swim towards many uncharacterized extracts, such as root exudates, meaning the cognate ligand of the MCP may not be present in any chemical library. An MCP for quorum-sensing molecules could also exist as this would allow cells to swim towards each other to reach the critical concentration sooner, but these molecules may not be known or exist in pure forms for *in vitro* testing.

The number of MCPs varies substantially between species, e.g. *P. pseudoalcaligenes* KF707 has 35 MCPs compared to *P. putida* KT2440's 24. While there appears to be a somewhat conserved core of amino acid, TCA cycle intermediate, and energy-taxis receptors, there is no clear indication how many different MCPs are present in the genus. I showed that Aer is a family, that has been subject to horizontal transfer and of which species vary in how many homologs they possess. Are there other MCPs with similar phylogenetic distributions? TCA cycle intermediates have the most characterized MCPs, followed by amino acids. Both are problematic from a bioinformatics perspective as there is little sequence homology between these receptors with extremely similar functions. For example, the malate/fumarate receptor of *P. putida* KT2440, McpR, has at best 47% identity over 61% coverage to the next closest MCP that also recognizes

malate, McpS. Still when these proteins are used as query sequences to search the genome of other *Pseudomonas* species there is generally either a single hit with close to 100% coverage and >60% identity or no hits with sufficient sequence coverage. If a few more MCPs with homologous functions are identified in other species, an estimate of how much sequence identity is needed to execute the same function could be obtained. All predicted MCPs in sequenced *Pseudomonas* species could then be collected and sorted into subfamilies, providing a solid estimate of the number of different MCPs in the pangenome. Having associated genomic context data, similarly to my Aer study, would enable 'guilt-by-association' functional predictions for MCPs consistently associated with particular metabolic pathways. Functions could also be predicted based on differential chemotaxis profiles between species and easily confirmed by transferring the MCP gene from one species to another.

5.1.5 Environmental Relevance of Chemotaxis and Energy Taxis

A theme of this thesis was the interaction between organisms and their environment. For bacteria, chemotaxis is one of their most important interactions with their environment as it is the only behaviour that allows them to optimize their immediate environment in response to local concentrations of nutrients and stressors. Non-chemotactic mutants have been found to have a competitive disadvantage for colonizing environments such as the rhizosphere and while infecting higher organisms (81). This is not surprising as these mutants are completely subject to the whims of their environment and cannot make use of motility to improve their circumstances. In this thesis, I focused on energy-taxis, a form of chemotaxis that allows cells to sense their own generation of cellular energy and move to locations that optimize this generation. In *Pseudomonas*, I showed that the receptor for this behaviour, Aer, is actually a family of related receptors with varied distribution in the genus. At a glance, energy-taxis seems superior to chemotaxis as it requires only

a single receptor to mediate swimming towards any metabolizable carbon source. This could explain why there are different versions of Aer as each one could be expressed differentially and have different sensitivities allowing for more control of energy-taxis. This does not, however, fit with observations that *Pseudomonas* species have many different MCPs, each specific for different attractants as their existence implies that energy-taxis is not a sufficiently competitive behaviour by itself. This is likely due to the much lower concentrations of substrate that chemotaxis can direct swimming towards (123). The combined presence of chemotaxis and energy-taxis in *Pseudomonas* suggests that they are actually complementary behaviours that significantly enhance competitiveness in the varied environments *Pseudomonas* inhabits.

Chemotaxis is a widespread behaviour in bacteria. Energy-taxis may also be but it has been characterized in only a few organisms, including *Azospirillum brasilense*, *Rhodobacter sphaeroides*, *Helicobacter pylori* and *E. coli* (123). This is likely because energy-taxis cannot be easily distinguished from chemotaxis; observations of directed swimming in most chemotaxis assays provide no indication whether the behaviour is metabolism-dependent. This is a prerequisite for separating energy-taxis from chemotaxis as the latter only requires binding of a ligand to an MCP whereas energy-taxis requires the cognate receptor to measure some portion of the cell's energy-generating pathway (123). Bacteria have evolved different ways to monitor energy production in the cell. Aer from *E. coli* and AerC from *A. brasilense* both have PAS domains that bind flavin adenine dinucleotide (FAD) co-factors which are necessary for the function of the receptor though it is unclear if they are sensing the reduction of FAD to FADH₂ while it is bound and what partner is mediating this reaction or if they are sensing the global ratio in the cell of FAD to FADH₂ (123). *E. coli* has an additional receptor, Tsr, which appears to be able sense the proton motive force across the inner membrane through its transmembrane domains (124). TlpD of *H.*

pylori mediates energy-taxis through an unknown mechanism, however it physically interacts with aconitase of the TCA cycle (125). This diversity of mechanisms and species with energy-taxis receptors, along with observations of its importance in *H. pylori* infection strongly indicate that energy-taxis is an important ecological behaviour. While this has also been demonstrated in *Ralstonia solanacearum* (126) and *A. brasilense* (127) the existence of multiple Aer homologs in *Pseudomonas* indicates that it is likely a highly useful behaviour for multiple situations in this genus. Interestingly, *P. aeruginosa* which can live in varied environments and act as a pathogen, only has one homolog indicating that it is likely not important for pathogenicity. Determining the ecological significance of energy-taxis receptors is also difficult as their broad-purpose utility makes them useful for colonizing any niche. Conversely, MCPs for specific substrates, such as Cca of *Campylobacter jejuni* which detects aspartate and is key for invading the intestine (128, 129), can be key for colonizing specific niches.

Metagenomics has enabled the comparison of environmental communities without the need for culturing the members (130, 131). The functionality of these communities is determined based mostly on known sequences for metabolic enzymes. This makes sense as we have very good knowledge of a huge array of different metabolic pathways and the genes that enable each individual step. Compared to metabolic genes, the function of barely any MCPs are known. If more MCP genes were characterized they could be used as an additional level of information about the substrates that uncultivated bacteria deem worthy of swimming towards. This will likely be more difficult as MCPs are poorly conserved in the LBR. For example, PctA from *P. aeruginosa* (99), Tsr from *E. coli* (132) and VfcA from *Vibrio fischeri* all mediate responses towards serine, but Tsr has only ~30% identity over ~55% coverage with the other receptors and PctA and VfcA

have only 33% identity over 100% coverage. Still MCPs present an untapped potential resource for understanding the preferred substrates of (uncultivated) environmental organisms.

5.1.6 Chemotaxis, a Key Interaction between Bacteria and their Environment

Chemotaxis control by MCPs represent one of the fastest ways bacteria can respond to their environment. When a compound with a cognate MCP is encountered by a bacterium its behaviour is immediately modulated without any delay for changes in gene expression. This makes chemotaxis and MCPs an ideal lens through which to view the interactions between bacteria and their environment as behavioural changes in response to nutrients and/or stressors are immediate. Because the cause and the effects of chemotaxis have such a direct effect on the cell we can expect that the systems controlling this behaviour will have evolved to a very high degree of precision and that they closely reflect the ecological preferences of the organism. This has been demonstrated in *E. coli* as the metabolic preference for amino acids correlated with their strength as chemoattractants whereas in *B. subtilis* there was no correlation (133). Along with other differences in their chemotactic preferences this reflects differences in the ecology of the two organisms, demonstrating how observations about chemotaxis could be used to understand a bacterium's environmental role. Understanding other tactic behaviours like energy-taxis, pH-taxis, electron acceptor taxis (aerotaxis) and taxis towards/away from non-carbon source compounds could be extremely revealing about an organism's ecological function. Knowing what carbon-sources attract an organism should be as useful as knowing whether it can metabolize them. Overall chemotaxis and chemoreceptors represent an open field for developing a better understanding of bacteria, both in the lab and environmental samples.

5.2 Growth Media: Bacterial Environments in the Lab

Growth of bacteria in the laboratory enables total control over their environment. In *Chapter 2*, I demonstrated how changes to the growth medium altered how metal toxicity was perceived by *Pseudomonas* species. Differences were observed between rich/undefined and minimal media as well as between minimal media with different carbon sources. Much of these observed differences were due to modifications of metal bioavailability in the different media. Aluminum chelation by phosphate reduced the amount of free aluminium that could interact with bacteria and inhibit their growth. Other differences, like the large increase in copper tolerance in *P. fluorescens* grown on aspartate were more difficult to explain. These observations reinforce the notions that interactions between bacteria and nutrients, and between bacteria and stressors are not independent, they influence each other, and that interactions between nutrients/stressors and the environment in turn affect how they are perceived by bacteria. Taking these effects into account and designing experiments so that their influences can be understood is key to investigating interactions between bacteria and their environment. Although our lab and others now routinely see differences in metal tolerance under different media conditions, this study was the first to systematically study this.

5.2.1 LB: Lysogeny Broth

One of the most common laboratory media used for growing bacteria is LB, which should be an acronym of Lysogeny Broth because of its original use for growing phage, but is also called Luria-Bertani broth or Luria Broth based on the scientists that developed it and popularized its use (134). Easy to make, and even easier to buy pre-made, LB has permeated the field of microbiology. This is problematic because it has been found to be a poor choice for physiological studies (134) as it mostly contains amino acids, which are a poor energy source in *E. coli* and are used sequentially, its composition can easily vary and it contains no divalent cations (e.g. Mg^{2+}) (135).

Despite these limitations and the warnings of this seminal paper, it continues to be used widely. This is not entirely problematic as the function of many genes have been determined using it, for example many of the studies on mechanisms of metal toxicity I discussed in *Chapter One*. Still, this is a high-profile example of how important it is to consider the composition of growth media.

In *Chapter 2*, I used LB to determine how much aluminum and copper *P. pseudoalcaligenes* KF707 and *P. fluorescens* could tolerate despite its failings. This was because many previous metal resistance studies have been performed using this medium, providing a useful foundation for interpreting the results and to enable comparisons. One of these studies was the systematic challenging of *P. fluorescens* with representative elements from almost all groups of the periodic table (136). Physicochemical data about the elements were then used to determine correlations between properties and toxicity. While correlations could be made, especially for metals, the use of other media would have provided additional data to improve these correlations. In particular, defined media with different concentrations of metal binding compounds, similar to what I used, would have allowed the exact physicochemical properties that determine toxicity to be more precisely determined. While this prevalent use of LB and other rich media has generated a useful background of literature to assist in interpreting results, the conclusions derived from these studies may not be directly applicable in actual bacterial environments.

5.2.2 Growth Media: Mimicking an Organism's Environment in the Lab

One of the biggest problems in environmental microbiology is the cultivation of 'unculturable' bacteria (137). This term 'unculturable' is representative of the problem; a more accurate description would be 'uncultured' (in pure culture). The discovery of syntrophs, organisms whose metabolism obligately depends on a partner organism (138), indicates that it may be impossible to culture all bacteria/archaea as pure isolates. Still there is expected to be a large

number of uncultured bacteria that could be grown in the laboratory given the appropriate conditions (139). The systematic modification of media conditions, informed by chemical and physical measurements of an environment, is expected to yield results in this regard. These kinds of investigations have led to improved cultivation of environmental organisms, the successful laboratory growth of ‘unculturables’ and an improved understanding of microbial physiology. These works highlight the importance of media composition in microbiological research.

Many reviews have emphasized the importance of culture conditions for improving cultivation efficiency of environmental isolates (139–142). Traditional culturing techniques select for fast-growing species that can rapidly adjust to, and exploit high concentrations of nutrients (140). This has led to the common use of genera such as *Pseudomonas* as they can easily be isolated and repeatedly cultured from a variety of environmental samples. Species with specific nutrient requirements, low growth rates and recalcitrance to cultivation on agar solidified media require the use of alternative cultivation techniques (140). This includes alternative gelling agents, cultivation in microcolonies or diffusion chambers but especially manipulation of media composition. Some bacteria require specific nutrients (such as amino acids) they are unable to synthesize themselves, others require low concentrations of nutrients and some require a mix of conditions that explicitly favours their specific metabolism. This has led to the creation of MediaDB, an under-appreciated database aimed at collecting details media conditions that allow the growth of specific organisms (141). This is the kind of resource that should enable improved cultivation of bacterial species by centralizing data about the importance of media composition.

Laboratory studies where media composition was meticulously altered have revealed insights into microbial physiology. Carbon catabolite repression, the regulation of catabolic metabolism to allow selective preference of carbon sources, is perhaps the best example of

manipulating media conditions to understand bacterial behaviour (143). Providing bacteria with a mixture of carbon sources, then observing the order in which they are consumed, allowed for the discovery of many genetic regulatory mechanisms. Observations of *E. coli* revealed that its growth rate is highly dependent on the medium it is in, and the cellular physiology can differ between different media, despite having the same growth rate (144). In these cases it was observed that cultures provided with exogenous amino acids had less ribosomes. A large-scale study of *E. coli* showed that any change in culture conditions, be it carbon/energy source, growth rate or presence of stress resulted in significant changes to the proteome (145). These are just some examples of the utility of experimentation with modified media compositions, it has also been useful with regards to metal toxicity. By adding copper chelators, geochemical modeling its speciation and assessing toxicity it was clearly demonstrated that total copper does not determine toxicity (146). This study influenced my own work which supported the notion that free copper and other specific mediate its toxicity towards bacteria.

Integration of the approaches described above would provide a complete view of the parameters that affect metal toxicity in environmental bacteria. I tested a combined biofilm and planktonic growth system, but cultivation on solid media, in microcolonies or using alternate biofilm growth systems could better represent the natural environment and would likely influence metal tolerance. As carbon source affected metal tolerance, growth in an environment with a mixture of carbon sources would likely have even more influence. For example, would aspartate still increase the tolerance of *P. fluorescens* to copper if there were other carbon sources present? I only quantified total growth at a set time point, but media composition affects growth rate. Closer monitoring of the combined effects of metals and carbon source on growth kinetics would provide insight into metal toxicity. As I demonstrated in *Chapter 3*, metabolomics can be used to

understand metal toxicity, but other approaches such as proteomics would allow confirmation of effects implied by metabolic changes. While these investigations would be informative about metal toxicity in bacteria, my work on manipulating media composition to understand its effect on metal toxicity demonstrated the importance of speciation and interplay between environmental nutrients and stressors and enabled further delving using metabolomics.

5.3 Basic Research on how Metal Toxicity Affects Organic Pollutant Bioremediation

The problem of environmental contamination with pollutants requires practical applied solutions. Developing solutions through empirical observation is the domain of engineering, but improved solutions can also be informed through fundamental research. In *Chapter 3*, I presented a purely fundamental study aimed at understanding why metals inhibit organic pollutant degradation. To gain insight into this problem I reduced the problem to a simplified representation of the situation by examining a single species of bacteria grown in liquid media. I examined the independent effects of two metals and compared their effects between a simple carbon source, succinate and a single complex substrate, biphenyl, each at only one concentration. These experiments provided a strong suggestion of how metal toxicity interferes with organic pollutant degradation: oxidative stress caused by aromatic substrate metabolism exacerbates metal toxicity. This raises many questions about what the exact mechanisms are, what concentrations are needed to initiate the effect and how other metals and pollutants are influenced but also provides an immediate direction for practical applications.

5.3.1 Further Investigations into the Effects of Xenobiotic Metabolism and Metal Toxicity

The mechanisms of oxidative stress have been well characterized in model bacterial systems (28) but remain an important topic in microbiology. For example, metal nanoparticles,

which have been proposed as a solution to rising antimicrobial resistance, kill bacteria through reactive oxygen species (ROS) mediated mechanisms (147). Methods for more accurately measuring ROS production in cells are also improving, mostly through the development of improved fluorescent probes (148). These probes, such as dichlorodihydrofluorescein, work by diffusing into the cell then reacting with ROS to produce a fluorescent signal that can be measured (149). While different probes have different specificities and may be interfered with by other cellular components, many studies have successfully used such probes to measure oxidative stress in bacteria. These methods could be applied to understanding the synergistic increase of oxidative stress observed in xenobiotic degrading bacteria exposed to metal toxicity. It would be important to first determine baseline levels and concentration dependence for cultures only degrading xenobiotics, then test a matrix of carbon source and metal concentrations. This would rapidly increase the number of experiments but is necessary for confirming the synergism, establishing a dose-response relationship and determining if it is linear or if there is a threshold that when breached results in a rapid increase of ROS. Modeling of oxidative stress in unicellular organisms implies that such a threshold exists between ROS production and irreparable damage (150) so a similar relationship may exist between organic pollutant and metal concentration. The use of modeling techniques could also help with the exponential growth of experimental conditions to test as there are many different organic pollutant/metal combinations in contaminated sites (15) that would require laboratory experimentation. It would be important to test metals with different known mechanisms of toxicity as my data showed that copper, which was demonstrated not to induce ROS mediated oxidative stress in bacteria (37), in combination with biphenyl exacerbated oxidative stress. Testing these situations in the laboratory should help determine the most

problematic pollutant/metal combinations and concentrations, as well as provide direction for mitigating their negative effects in real bioremediation applications.

5.3.1.1 Validation and specific follow-up experiments

The metabolomics experiments presented in *Chapter Three* generated a large swath of data with many possible hypotheses to pursue, many which could be directly examined. The observed changes to relative metabolite concentrations of TCA cycle intermediates indicate that expression levels of these genes should be quantified by reverse transcriptase quantitative PCR (RT-qPCR) as this would be far simpler than quantifying protein expression levels. Activity assays of cell lysates could also be used to assess whether the proteins are functional as aconitase and fumarase were expected to have decreased activities in the respectively aluminum and copper exposed cultures. The expression and activity of phosphoglycolic phosphatase should also be tested as it is expected to be increased in expression under oxidative stress conditions. Control tests using cultures exposed to various concentrations of hydrogen peroxide, combined with fluorescent dye quantification of ROS would allow for a link to be established between this gene and oxidative stress. Further comparisons of ROS levels in peroxide treated cultures, metal exposed cultures and PCB exposed cultures grown on either succinate or biphenyl would provide definitive evidence that metal toxicity exacerbates oxidative stress.

5.3.2 Connecting Fundamental and Applied Research

Bioremediation is known to be far superior in effectiveness and cost to physical or chemical methods of remediation, but cleaning up a polluted site in a timely fashion requires more than just leaving the bacteria to do their work (151). Some combination of aeration, mixing, addition of nutrients (biostimulation) or specific bacteria (bioaugmentation) is generally needed. Directed, specific adjustment of these techniques should lead to better outcomes for metal and organic

pollutant co-contaminated sites. Metal bioavailability is highly dependent on pH (152) so adjusting a site's pH becomes even more important when it is co-contaminated. Oxygen availability is often a rate-limiting factor for bioremediation (153), however in co-contaminated sites excess oxygen may contribute to excessive ROS generation. This will require experimentation to determine optimum oxygen concentrations or further research on anaerobic processes. While bioremediation of organic pollutants generally requires their catabolism as a carbon source, if their metabolism generates enough ROS that is exacerbated by co-contaminating metals then they will not be degraded. Addition of simpler carbon sources may be necessary as this increases microbial biomass (151), though they may just be used preferentially. Again, experimentation with different carbon sources and concentrations will be necessary. Other nutrients are often also depleted in contaminated soils, biostimulation with a source of nitrogen and phosphorus greatly increases pollutant degradation (153). For co-contaminated soils even higher concentrations may be needed to help the resident bacteria withstand the combined stresses. Bioaugmentation, the addition of exogenous bacteria to a contaminated site, is often ineffective due to the inability of the new bacteria to compete with the indigenous community (154). This would likely be no different for co-contaminated sites, however bioaugmentation in *ex situ* reactors is generally far more effective so this may be the necessary route for co-contamination. Anaerobic bioreactors have long been a staple of wastewater treatment (155), and the lack of oxygen would prevent the formation of metal toxicity exacerbating ROS, perhaps making this an ideal solution for co-contaminated waste. This will require further research as few studies have examined metal toxicity in anaerobic bacteria. This is fundamental work that needs to be done in a controlled laboratory setting, but could prove highly useful for field applications, emphasizing the importance of communication between fundamental and applied research.

5.3.2.1 A Comment on Publishing and Reaching the Intended Audience

I initially submitted my metabolomics study on the effects of metals on biphenyl metabolism to *Microbial Biotechnology*, but it was turned away due to not being relevant enough to applied situations. Instead it was published in *Frontiers in Microbiology*, an open-access broad-scope microbiology journal. This was disappointing and frustrating as a young scientist trying to make a difference with my research. Rejections are an expected part of academic publishing, but the reasoning bothered me. While it is true that it was not an applied study, I think the main finding should inform future applied studies. I'm concerned that because it was published in a general microbiology journal, researchers working on real-world applications will not read it. I read papers based on their appearance in Scopus searches, not the journal. I hope that other researchers act similarly and that its open access status will allow anyone interested to read it. Still I think it is important as basic science researchers to strive to bridge this disconnect between fundamental and applied research as both are needed to develop evidence-based solutions to the many problems facing humanity today.

Publishing in *Frontiers* was an interesting experience though due to their novel approach to peer review. Their system, and others that have open peer review where the reviewer's names are published are improvements to the old system I think. Pre-print servers for pre-submission review are also a good idea, which is why I deposited my energy-taxis work in the bioRxiv before submitting it to *Molecular Microbiology*. It has not received any comments but has been downloaded 50 times. If I had promoted it better on social media, which I think is an important place for scientists to be active, it would have likely garnered more attention though my network is still fairly small.

Publishing with *Frontiers* also gave me the opportunity to become a ‘review editor’, though I was removed from the position when I visited their booth at the American Society of Microbiology meeting and they found out I was still a student, something they don’t allow for the position. Ironically, despite this position essentially being a frequently contacted potential reviewer, I ended up reviewing more papers for *Microbial Biotechnology*, receiving the requests from the editor that rejected my submission. Regardless of these experiences I’m still excited to be starting my scientific career while these shakeups of how science is reviewed and published are occurring.

5.4 Final comments

Successful research will always raise more questions than it answers. At the beginning of my PhD I set out to understand how media composition and carbon source affect metal toxicity, what the biochemical effects of metal toxicity are on xenobiotic degradation and how *P. pseudoalcaligenes* KF707 swims towards biphenyl. I found what I consider expected results, that media composition affects metal bioavailability that in turn determines toxicity, but also unexpected differences between similar organisms and carbon sources. My metabolomics work demonstrated that metal toxicity is synergistically enhanced by oxidative stress caused by xenobiotic metabolism. Following this up to understand exactly how this happens at the biochemical level, and how to mitigate it for bioremediation applications will require a great deal of further work. By deleting the genes for energy-taxis I was able to determine that this was not how *P. pseudoalcaligenes* KF707 swims towards biphenyl, but I did not find the receptor that does enable this behaviour. Instead I discovered that two other genes can affect energy-taxis and that the main receptor, Aer, is actually a family with variable distribution in *Pseudomonas*. Determining the function of these alternative members, and how other chemoreceptors are

phylogenetically related are further avenues of research. These, and the other avenues above were opened by my steps forward made during my research journey aimed towards understanding how bacteria interact with their environment.

The chelation of metals by growth media, the use of chemotaxis to choose a habitat, and the exacerbation of metal toxicity by xenobiotic metabolism were all modulated by chemical, physiological and metabolic interactions. *Pseudomonas pseudoalcaligenes* KF707 was an excellent model organism for these studies as it has multiple energy-taxis receptors that allow it to detect environmental nutrients whose metabolism, in the case of biphenyl, produces reactive oxygen species that make metal toxicity worse. Depending on the chemical composition of the medium used for its growth, differing amounts of metal will be able to reach the cell, changing how much it can tolerate. Knowing how these interactions work allows us to understand the lifestyle of bacteria and how they can survive, and thrive in variable conditions.

Appendix A Preface

Appendix A is a review of the literature (up to 2011) of studies that applied metabolomics to studying metal toxicity. This review was invited by *Metallomics* and was published as part of their themed collection on Metal Toxicity. At the time metabolomics was still a new technique and few researchers had applied it in any organism. This review aimed to educate readers on how metabolomics experiments are performed and how they can contribute to understanding metal toxicity in a variety of organisms. The available studies on plants, model eukaryotes (earthworms and rats) and bacteria are summarized then a discussion of the benefits of using metabolomics to study metal toxicity is provided. I wrote the majority of this review with guidance and editorial support from my co-authors.

APPENDIX A: METABOLOMICS AND ITS APPLICATION TO STUDYING METAL TOXICITY

Sean C. Booth, Matthew L. Workentine, Aalim M. Weljie and Raymond J. Turner

Published in *Metallomics*, November 2011, 3:11 pp1142-1152

Copyright permissions for the reproduction of this manuscript can be found in Appendix E

A.1. Summary

Here we explain the omics approach of metabolomics and how it can be applied to study a physiological response to toxic metal exposure. This review aims to educate the metallomics field to the tool of metabolomics. Metabolomics is becoming an increasingly used tool to compare natural and challenged states of various organisms, from disease states in humans to toxin exposure to environmental systems. This approach is key to understanding and identifying the cellular or biochemical targets of metals and the underlying physiological response. Metabolomics steps are described and overviews of its application to metal toxicity to organisms are given. As this approach is very new there are yet only a small number of total studies and therefore only a brief overview of some metal metabolomics studies is described. A frank critical evaluation of the approach is given to provide newcomers to the method a clear idea of the challenges and the rewards of applying metabolomics to their research.

A.2. Introduction

The challenge of understanding metal toxicity is as broad and complex as the effects of metals on the cell. While some recent studies have begun to elucidate the mechanisms behind the actual toxic effects metals exert upon specific microbes and higher organisms, a complete understanding of these phenomena is far from being realized. Fortunately, modern technologies are providing new tools to delve deeper, and generate a more comprehensive understanding of the

interaction between cells and metals. ‘Omics’ technologies provide a system-wide view of alterations in cell physiology upon being stressed. Metabolomics is one such technology with great potential for studying metal toxicity and providing a unique perspective on metal-induced changes in cellular metabolic architecture. Here we overview the methods of metabolomics and provide a brief overview of some of the studies that have begun to use this approach. As this approach is still in its infancy we will also critically evaluate the pros and cons of this approach as well as the necessities for moving this method forward.

A.2.1. Metabolomics: A Systems Biology View of Metabolism

Understanding physiology at the whole cell level is a daunting task. However, since the advent of the genomics age, ‘omic’ technologies have been providing unprecedented access to the underlying structure and function of biological networks. Various layers of the cellular architecture are revealed through genomics, transcriptomics, and proteomics, yet in order to make a clear link between genotype and phenotype another layer of information is required¹. The promise of metabolomics is to fill this gap and provide quantitative information about levels of intracellular metabolites which represent the highest level of functional components within cellular processes². It is the cell’s complement of small, low molecular weight compounds, the metabolites, required for growth, function and maintenance that are defined as the metabolome. The systematic identification and quantification of these compounds is the goal of metabolomics²⁻³.

Metabolomics presents a unique challenge given the breadth of chemical diversity inherent to cellular metabolites. The broad classes of metabolites include amino acids, nucleotides, carbohydrates and lipids. An analysis of the *Escherichia coli* metabolome classified the 745 metabolites present in the EcoCyc database⁴ into 57 different structural classes⁵. Estimates of the *E. coli* metabolome size using both bioinformatic and experimental approaches, range from 625 to

1195 metabolites but it is thought the actual size could be much larger ⁶. This diversity presents a significant challenge both in terms of extracting the metabolites and the analytical tools required to analyze them. The reality of today's technology is that no single extraction procedure or analytical platform can cover all the different chemical classes of metabolites ⁷. This obstacle can be potentially hurdled by building a workflow that contains multiple extraction methodologies and multiple analytical platforms. For example, van der Werf and coworkers have developed a platform that utilizes six different analytical separation methods in an attempt to obtain comprehensive coverage of microbial metabolomes ^{7a}. Using this platform, they were able to identify 380 of the 399 commercially available *E. coli* metabolites^{7a} and they also applied it to a study of *Pseudomonas putida* grown on four different carbon sources⁸. However, for many researchers the analytical expertise and large expense required to run multiple analytical platforms are significant deterrents. As for sampling procedures that rapidly quench metabolism and allow for accurate and reproducible metabolite extraction, there is no conclusive best method for all biological samples in the literature, with many opposing techniques purported to be superior, which will be discussed later ^{7b, 9}. Since this is a developing field with no consensus or 'gold-standard' quenching and extraction method, future studies will need to keep up to date with the state of the art. The functional alternative is to optimize a single extraction method and analytical technique to detect as many metabolites as possible and accept as a caveat that the coverage will not be comprehensive (but this can be in part overcome through statistical and bioinformatic approaches; see below). One way to improve upon metabolite diversity coverage is to analyze a single sample using more than one technique, such as nuclear magnetic resonance (NMR) and gas chromatography coupled to a mass spectrometer (GC-MS) which allows better access to a large part of the metabolome than a single analytical technique can provide ¹⁰.

Metabolite diversity is only one challenge among many to be faced by those who wish to access the information promised by metabolomics. Metabolite turnover can be extremely rapid, less than 1 second in some cases ^{3a, 11}, a concern that must also be addressed. Metabolism quenching and metabolite extraction, which will be discussed later, is a very important consideration when doing metabolomics experiments. So given these technical challenges as well as the expense of advanced analytical instrumentation, what is the motivation for doing metabolomics? As was pointed out earlier, metabolomics promises to provide the missing level of information in the hierarchy of cell processes. The draw for many is the potential of metabolic profiling to identify previously unknown gene functions, which is particularly attractive to plant functional genomics^{2, 3b}. Powerful examples of this concept have been demonstrated in both yeast and plants where the analysis of metabolic networks was used to identify the phenotype of previously silent mutants ¹².

The profiling of microbial metabolomes is also of significant interest for several reasons. The coupling of metagenomics with metabolomics has been proposed as a powerful way to elucidate the function of microbial communities¹³. Others have used metabolomics as a way to analyze the differences in a bacteria's growth between similar carbon sources ⁸. Metabolic and genetic engineering also stands to benefit tremendously from metabolomics, particularly for those who are working to improve upon the production of metabolites for industrial and commercial processes¹⁴. As for the field of metal toxicity, there is substantial evidence to suggest that certain toxic metals, such as aluminum, evoke significant metabolic shifts in the cell ¹⁵. Transcriptomic and proteomic studies on metal exposed microbes demonstrate significant changes in cell physiology, including metabolic pathways, suggesting that metal stress induces significant metabolic changes in the cells of microorganisms. An example of which is seen in a study of the

yeast *Candida* where a range of metals were found to affect mycelia formation and biofilm differentiation at sub-inhibitory concentrations¹⁶. Furthermore, metabolomics has begun to be used to examine bacterial biofilms¹⁷, which are known for their increased resistance to antibiotics and biocides. Similarly, differences in metal toxicity levels between planktonic and biofilm growth states have been found¹⁸.

A.3. Experimental Overview of Metabolomics

Metabolomics experiments involve a series of steps: collection of organisms, metabolism quenching and sample collection, metabolite extraction, direct detection or chromatographic separation coupled to detection and finally data analysis. In the following section a brief summarization of the techniques for each of these steps will be described. For an excellent practical reviews of setting up metabolomics experiments, see also¹⁹.

A.3.1. Collection of Organisms

For the study of microorganisms, cultures can be grown in the lab under tightly controlled conditions of growth and metal exposure which ensures that any observed effects can be ascribed to known variables. The same can be said for laboratory grown plants and higher animals. Studies of organisms collected from the field however are subject to the confounding influence of a variable environment, which can result in a wide range of ‘control’ metabolic states²⁰. Still, the use of metabolomics to examine differences between toxin exposed organisms, including earthworms exposed to metals, have been successful in overcoming this inherent metabolic variation²¹. For practical purposes, directly sampling organisms from their native environment rather than importing them into the lab decreases metabolic variability²⁰.

A.3.2. Metabolism Quenching and Sampling

As the Highest level of functionality within the cell, metabolism is subject to extremely rapid turnover, making sample handling an important concern ²². As such, metabolism must be quenched as quickly as possible without any opportunity for the cells to react to the sampling procedure, and without causing any chemical change to metabolites or leakage of metabolites from within the cell. For plants and animals this process is generally achieved by flash-freezing either whole organisms or dissected tissue in liquid nitrogen, then subsequently grinding up samples for metabolite extraction. Ideally a quenching procedure perfectly freezes the cell as a metabolic ‘snap shot’ completely halting all enzymatic activity, though this is an extremely difficult task when working with microorganisms^{19a}. The issue of metabolite leakage is complex though it has been overcome by one well-cited study where cultures of *Saccromyces cerevisiae* were fed with ¹³C labeled carbon source which allowed accurate quantification of intracellular metabolites by using isotopomer ratio analysis to compare a known amount of labeled cells to unlabeled samples²³. This technique has not seen widespread application so one of the major issues with microbial metabolomics is still the collection of cells without significantly altering metabolism. This can be achieved using centrifugation or filtration (the most popular/common methods) though both take time and provide the opportunity for metabolites to leak from the cell, which is problematic to future data analysis. A method where cells are grown directly on a membrane filter with nutrients being provided by agar-loaded media underneath has been developed ²⁴. While very useful for rapid sample collection, this method does not seem appropriate for metal exposure as the metal must diffuse across the membrane thereby decreasing the concentration perceived by the cells.

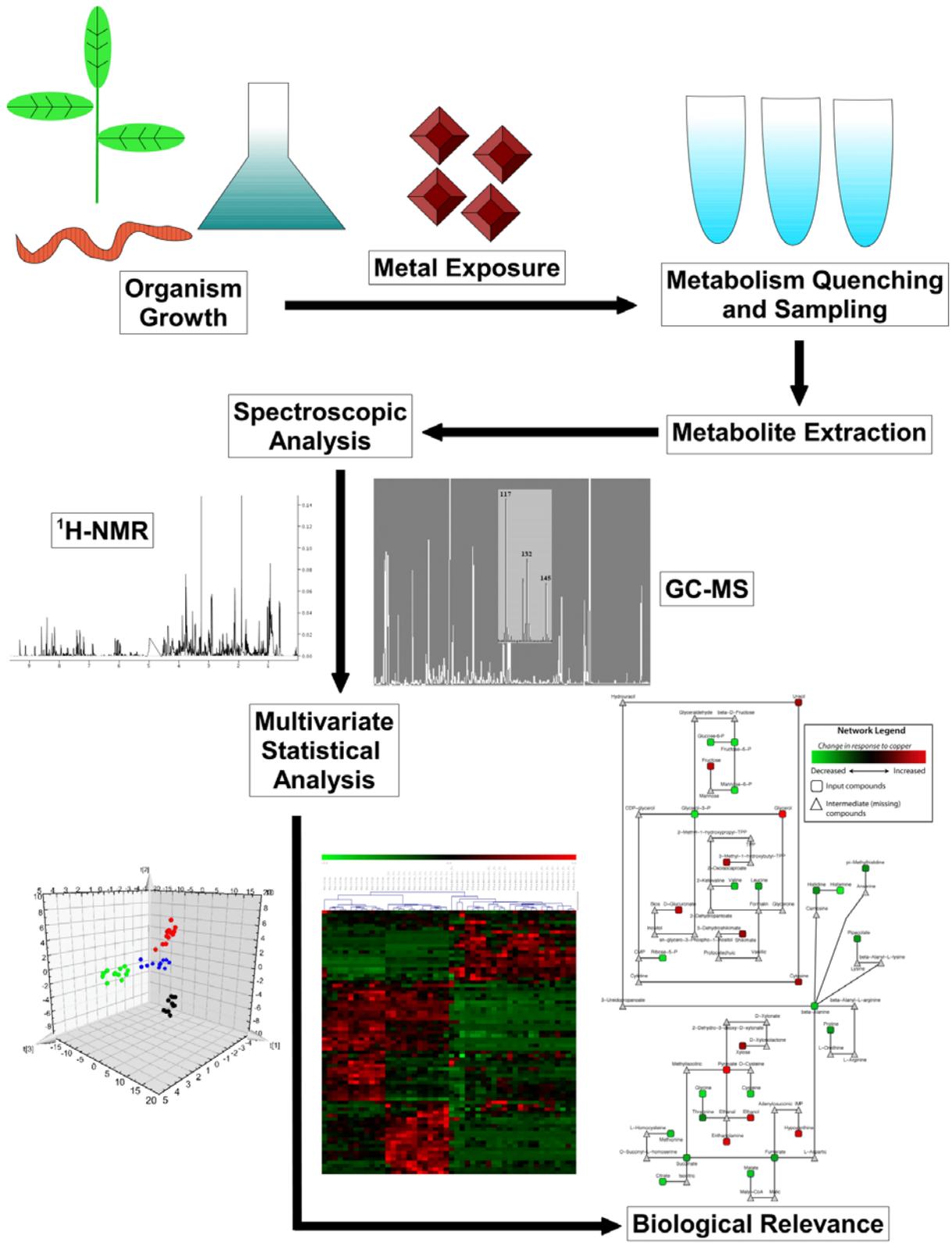


Figure A-A-1: Schematic Overview of Metabolomics Workflow Metal Toxicity Experiments. First organisms are grown and exposed to the metal(s). Next metabolism is quenched and samples are collected. Cells are lysed and metabolites are extracted prior to preparation for the analytical platform to be used. Samples are then analyzed by the platform(s) of choice, here only ¹H-NMR (proton nuclear magnetic resonance spectroscopy) and GC-MS (gas chromatography mass spectrometry) are shown as they are most common. The data acquired in this fashion must then first be deconvoluted before being interpreted using various multivariate statistical and bioinformatic tools in order to generate conclusions of biological relevance.

Additionally, growth on membrane like this can be described neither as planktonic nor as a biofilm making data obtained in this fashion difficult to contextualize. Sample quenching and collection is one of the major points of contention in microbial metabolomics: separate articles concluding that either filtration or centrifugation is the superior technique have both been published ^{7b, 9a, b}. Choice of quenching solution is also hotly debated, though it is widely accepted that a cold (~-40°C) solution is best ^{9b, 22}. While commonly used for the past decade, quenching in cold 60% methanol in water is contended to cause severe metabolite leakage ^{9b}. Proposed alternatives though, such as cold glycerol-saline are also problematic as glycerol, which is difficult to remove from the sample, will show up during the detection procedure ^{7b, 25}. As in any field, new developments continue to be touted as the solution, such as cold 40:40:20 acetonitrile:methanol:water for *E. coli* and rapid filtration then quenching in *Staphylococcus aureus* ^{7b, 25}. Since the issue is still in contention, as reviewed recently in ²² and ^{9b}, solace can be taken in that regardless which method is chosen, reproducible sample handling technique and the presence of control samples will always allow comparative analysis of relative concentrations.

After metabolism has been quenched and the sample has been collected, a method for collecting and concentrating all the metabolites must be undertaken ^{19a}. Due to the chemical diversity of metabolites this is an ultimately impossible process to achieve perfectly, however protocols that obtain a large portion of metabolites have been developed. Additionally, the

extraction procedure should remove larger ‘contaminant’ molecules such as DNA, proteins, and eventually the solvent used in the extraction. Metabolite extraction generally involves the addition of a mixture of organic solvents, the lysis of cells, a pelleting step to separate out dissolved metabolites from contaminant debris and the subsequent evaporation of the solvent leaving a dried sample to be prepared for whatever analytical platform(s) will be used. Since organic solvents are used it is possible to separate polar from non-polar metabolites and then subsequently analyze both fractions. Contrary to quenching, a mixture of water:methanol:chloroform has generally been accepted as a comprehensive extraction solution ^{19a}. Once extracted, evaporation concentrated samples can be frozen and stored until ready for analysis. The further steps of sample preparation depend on the analytical tool to be used. For ¹H-NMR, samples must be dissolved in a deuterated solvent as any solvent containing hydrogen (i.e. water, chloroform, methanol) produces a strong (and unavoidable) peak. For some methods of analyzing the NMR spectra (such as targeted profiling, to be discussed shortly), an internal chemical shift standard of known concentration with very particular, unique NMR peaks such as 3-(trimethylsilyl) propionate-2-2-3-3-d₄ (TMSP) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) must be added to allow quantitation and in order to properly align the NMR spectra. For GC-MS analysis of the non-volatile components of the metabolome, samples must be derivatized into volatile analogs that will migrate through the GC column by using a derivitizing agent which adds functional groups that allow for migration through the gas chromatography column.

A.3.3. Analytical Techniques

Metabolite analysis requires a high throughput tool capable of identifying the structure of many chemically diverse substances simultaneously. As such, proton nuclear magnetic resonance spectroscopy (¹H-NMR) and mass spectrometry (MS) have emerged as the two main technologies

used in metabolomics²⁶. ¹H-NMR functions based on the behavior of atomic nuclei in the presence of a magnetic field; each proton has a distinct resonant frequency dependent on the local chemical environment of that molecule. This allows for the simultaneous, non-biased detection and quantification of all the metabolites within a sample at one time without any separation or modification of the sample²⁷. The problem with the simultaneous detection of all compounds at once is that signals from different compounds tend to overlap creating a complex convoluted spectrum. In the past, methods such as spectral binning have been used in an attempt to overcome this issue but now with the development of targeted profiling identification and quantification of many compounds from a single complex spectrum is much easier^{10c}. This differs significantly from MS analysis as the sample can either be directly injected into the mass spectrometer (DIMS) or first be separated along either a gas or liquid chromatography (GC or LC) column^{10b}. After separation the exact treatment of the molecules depends on the type of instrument, though the identification of molecules is always through the determination of the mass to charge ratio (m/z). Both ¹H-NMR and MS have pros and cons, one of the key differences being sensitivity. ¹H-NMR has an approximate lower limit of 1-5 μ M and requires around 500 μ L of sample whereas LC-MS can detect picograms of a compound²⁶⁻²⁷. MS does have its drawbacks as GC requires a biased derivitization step and the ionization step necessary for detection can also be selective for polar molecules²⁶. While LC is an extremely powerful tool for separating metabolites, libraries for use in identifying compounds are currently limited. However, modern developments in both fields such as larger NMR magnets (900MHz) and improved LC-MS protocols, instrumentation and software will serve to diminish the drawbacks of both methods in the future²⁶⁻²⁷. Fourier transform ion cyclotron mass spectrometry (FTICR-MS) is also a powerful MS technique whose further application to metabolomics will be beneficial to the field²⁸. Unfortunately the high-end

developments of all these techniques are unlikely to affect the average experimenter due to the relatively high startup cost and more likely instrument availability will be the deciding factor in which method is used. The plus side is that $^1\text{H-NMR}$ and any variety of MS can both provide extremely rich data sets on their own, which if available, together can be used to examine the same sample in order to strengthen analytical acuity.

A.3.4. Data Processing

The dataset for both a single $^1\text{H-NMR}$ or MS experiment is an extremely complicated spectrum that must be correctly deconvoluted in order to identify and quantify the metabolites present in the sample. For NMR experiments this can mean up to 5000 distinct resonance peaks in a spectrum²⁷, each corresponding to a proton within the sample. The challenge is thus to determine which compounds are present in the sample and at what concentrations, which can be quite difficult as it is possible for protons from separate molecules to have indistinguishable resonances. Fortunately the field of quantitative metabolomics has emerged, also called targeted profiling, which uses a reference library of spectra obtained from pure compounds in order to identify molecules in a complex spectra^{10c}. While this requires manual examination and processing of each sample's spectrum using a software package such as Chenomx NMR Suite (Chenomx Inc., Edmonton AB, Canada), it allows for a highly accurate determination of the content and concentrations of metabolites. Identification of metabolites from MS experiments is similarly made easier through comparison to standards found within libraries. GC-MS spectra require significant processing in order to deconvolute spectral peaks and identify metabolites from libraries such as the GOLM Metabolome Database²⁹, or the NIST Library³⁰. For the processing of the spectra there a number of freely available tools such as TargetSearch³¹, MET-IDEA (Metabolomics Ion-based Data Extraction Algorithm)³² and MetaboliteDetector³³. Depending on

how the original experiment was set up and analyzed the eventual output is a list of metabolites and either their concentrations or relative abundances in each sample.

A.3.5. Data Interpretation

Due the inherent size and complexity of metabolomics data, similar to datasets obtained from other omics technologies, the use of computer-assisted statistical analysis is extremely useful in deriving biologically relevant information from an experiment. One of the major causes of this issue is the fact that in a given experiment there will be far more variables measured (IE metabolites) than samples. While it is entirely possible to garner relevant information merely by investigating the individual metabolites that were observed to change in treated groups, analysis of the dataset as a whole is much more involved. Fortunately, multivariate statistical tools, some which have been developed for use with comparable datasets such as gene arrays, are available that greatly aid in the reduction of complexity and visualization of metabolomics data. These can be divided into two types: unsupervised and supervised analysis²⁶. Unsupervised techniques such as principal component analysis (PCA) and hierarchical clustering seek to distinguish inherent variation in the data and cluster similar samples together. PCA also reduces the long list of metabolite concentrations into components of variation, which allow for direct comparison of samples in 2D or 3D space. In supervised analyses such as Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), the sample classes (such as metal exposed vs. unexposed) are provided allowing for the directed clustering of samples. Techniques such as OPLS-DA also provide information as to which metabolites are causing the most variation between samples, thereby providing a starting point for biological interpretation.

A.3.6. Bioinformatic Tools

Again similar to genomics analysis, metabolomics relies on an ever growing number of bioinformatic tools and data repositories in order to be understood and interpreted. Once significant metabolites have been identified using multivariate statistics, their relationship towards one another and to the experiment needs to be determined. This can be done using the Tool for automatic Interpretation of Compound List (TICL)³⁴. This tool takes a list of metabolites and polls the Kyoto Encyclopedia of Genes and Genomes³⁵, while allowing for the selection of any organism's genome within the database, and tries to integrate them into biologically pertinent sub-networks while also determining which metabolic pathways are most represented by the list. This allows for the relationship between metabolites as well as their relevance to whole cell metabolism to be determined simultaneously. MetExplore is a similar tool which is still adding more organisms but also has many reaction and enzyme filtration options in order to prevent the identification of circular networks and over representing common co-factor reactions³⁶. Both of these tools produce visualizations of metabolic networks without accounting for the concentrations of metabolites determined in the experiment. Network analysis and visualization tools such as Cytoscape can be used to integrate this kind of data³⁷.

A.4. Metabolomics of Metal Toxicity

The use of metabolomics to study metal toxicity in any organism remains a novel undertaking, with most studies having been published in the past several years. The field of environmental metabolomics, under which the sub-discipline of environmental toxicometabolomics falls is itself a very young field and has only recently been reviewed^{20, 38}. A cursory review of the use of metabolomics to study metal toxicity shows several studies involving the effects of cadmium in various plants³⁹, how two *Pseudomonas* species respond to metal stress

^{17a, 40}, the effects of several metals on different tissues of salt-water clams ⁴¹, how metal contaminated sites affect earthworms^{21, 42}, the effects of several metals in rodents⁴³ and the model organisms *Caenorhabditis elegans*⁴⁵, *Daphnia magna*⁴⁶ and *Gasterosteus aculeatus*⁴⁴. Some of these studies will be discussed in depth below in order to provide a sampling of how metabolomics can be applied to study metal toxicity in different systems. While there are obvious differences in experimenting on macroscopic organisms compared to microbes, in metabolomics the core experiment and variables remain the same though there are some advantages and challenges with each organism type. First, a metal and its concentration must be chosen. Metal choice is potentially important when working with ¹H-NMR metabolomics as paramagnetic ions will affect the spectra ⁴⁸, though this can be dealt with during data processing. The concentration of metal, and exposure conditions are also important parameters. These can be chosen based on the purpose of the experiment; for example, while studying the marine clams environmentally relevant metal concentrations of common polluting metals were picked ⁴¹. For microorganisms though this may not be the best way to choose a concentration as environmentally relevant levels may be high enough to inhibit growth or too low to illicit a sufficient response. Additionally, bioavailable metal will be affected by media components making preliminary experiments important. As for exposure conditions, in both macro and microorganisms time and adaptation can be expected to affect cellular metabolism. More difficult in microorganisms, enough sample needs to be collected in order to surpass the lower detection limit of the analytical platform. While simple for macro organisms or when using a high sensitivity platform such as GC-MS, acquiring enough microbial cell mass, especially from a slow-growing bacterium could be challenging. When working with animals the choice of tissue to analyze is also important as different responses were observed in the various tissues of metal exposed clams ⁴¹. For plant metal toxicity metabolomics, the use of

LC-MS may be highly desirable as it can identify phytochelators, small peptides that bind metals
39a, 47.

A.4.1. Environmental Monitoring Organisms

While the field of metal toxicity metabolomics is fairly sparsely populated, the successes of the few studies that have been performed demonstrate the promise of undertaking this kind of enterprise. Studies on the response of the Manila clam (*Ruditapes philippinarum*) to mercury and copper and of the green mussel *Perna viridis* to cadmium have been recently published⁴¹. The first of these studies examined the metabolic effects of mercury in different pedigrees of Manila clam^{41a, b}. Using ¹H-NMR metabolomics of adductor muscle tissue the clam that showed the most discernable response was identified and suggested for use as a sentinel species^{41a}. The same result was obtained using gill tissue^{41b}. This demonstrates the ability of metabolomics to discern differences in closely related species for a very environmentally relevant purpose. Metabolomics studies also frequently show results which support previously identified metabolic changes, just one example from this study is that disturbances to the osmotic regulator metabolites glycine and taurine were observed in all clam pedigrees, and it is known that mercury causes osmotic stress^{41b}. In a similar study, the effect of two different concentrations of copper (10 and 40 µg/L), and 3 different exposure times (24, 48 and 96 hour) in a single pedigree of clam was examined^{41d}. Their results showed that in the low dose samples, some metabolic networks that were initially affected such as branch-chained amino acids, succinate and citrate returned to control levels after 4 days of acclimatization, however the osmotic regulator metabolites betaine and taurine remained changed. Separate and distinct differences were also observed in the high dose samples over time from the acute to chronic exposure. This study shows the elegant power of metabolomics to identify the varied effects of metal toxicity and understand their complexity. The final study by this group

involved a much longer (2 and 4 week) exposure of green mussels to cadmium^{41c}. As in the previous studies, which involved different metals, concentrations, exposure conditions and species, changes to metabolites involved in osmotic regulation and energy metabolism were again identified. The high reproducibility of these effects shows how metabolomics can be used to correlate similar effects of metal toxicity between different experimental conditions.

A.4.2. Model Organisms

The effects of metal toxicity on model invertebrate and vertebrate organisms have also been studied using metabolomics. Recently, the use of metabolomics to study ecotoxicity, including some work on metal toxicity in earthworms has been reviewed⁴⁹ and so the use of metabolomics to study metal toxicity in these organisms will not be discussed here. *C. elegans*, *D. magna* and *G. aculeatus* are all model organisms who have been very well studied, making it no surprise that they have been used to examine metal toxicity using metabolomics. In *C. elegans* the effect of metallothionein (metal binding proteins) knockout mutations on exposure to sub-lethal cadmium was examined using a combined NMR and ultra performance (UP)LC-MS approach⁴⁵. Surprisingly, it was found that the absence of metallothionein synthase genes did not affect the metabolic profile in both control and exposed organisms, thus implicating a different metal resistance mechanism. Based on an NMR observed decrease in cystathionine and UPLC-MS observed increase in phytochelatins in exposed organisms it was concluded that metabolism was shifted in order to produce these metal-chelating peptides. This study shows how the use of multiple analytical techniques is synergistic and how metabolomics can identify mechanisms of changes that could not be observed using proteomics or transcriptomics. The effects of copper and cadmium on the water flea *D. magna* have been studied using FTICR-MS⁴⁶. In response to both metals, perturbations in amino acid levels were attributed to increased synthesis of stress-related

proteins, whereas a decrease in fatty acid levels in response to cadmium and an increase of N-acetylspermidine in response to copper were respectively attributed to decreased nutrient uptake and as evidence of a systemic stress response. These studies thus highlight how metabolomics can be used to identify similarities and differences between the toxic actions of relatively similar toxic metals. *G. aculeatus*, the model stickleback fish, has also been the subject of a metabolomic metal toxicity study⁴⁴. In this study, ¹H-NMR metabolomic data was mostly used as a corollary to transcriptomic data, strengthening their conclusions that copper induces a shift to anaerobic metabolism. Together, these studies demonstrate some of the strengths of the use of metabolomics in model organisms and set the stage for future discoveries of mechanisms of metal toxicity to be discovered through this kind of approach.

Laboratory rats have been studied as model mammalian organisms. The hepato- and nephrotoxic effects of rare-earth metals^{43c, 43e} and metal nano particles^{43d, 43g} as well as the acute and chronic toxicity effects of cadmium^{43a, b} have been examined using metabolomics. The kidney and liver are responsible for most detoxification mechanisms in mammals, making them a good target for understanding toxicity mechanisms of metals. Metabolic changes in these organs in response to exposure to the rare-earth elements gadolinium and cerium were studied using solution and tissue-state¹H-NMR metabolomics^{43c, 43e}. Both methods found that both metals caused a decrease in glycogen and an increase in lactate in the liver suggesting they were causing acute liver damage. In the kidneys, levels of renal osmolytes such as betaine which aid in the balancing of osmotic pressure between extra- and intracellular fluids were found to be decreased. Interestingly, there was no histological evidence of kidney impairment in response to gadolinium demonstrating how metabolomics can be used to find evidence of toxicity not observable by conventional means. The toxic effects of ingesting metal nano-particles have also been studied using ¹H-NMR

metabolomics. Copper nano-particles were found to have a similar effect as soluble ions as evidenced by increased lactate, 3-hydroxybutyrate and acetate indicating a shift away from aerobic glycolytic energy production^{43d}. Titanium oxide nano particles showed a similar toxicity effect but also a perturbation to amino acid metabolism as taurine levels were found to be increased without associated liver necrosis (which could have been expected to cause taurine leakage)^{43g}. Both studies thus concluded that metabolomic analysis greatly added to their understanding of the previously understudied toxic effects of nano particles.

A.4.3. Plants

Several studies have investigated the effects of metal toxicity in various plants^{39, 50}. Cadmium is a popular metal to study in regards to plant metal toxicity as it is frequent contaminant of farmland due to a variety of industrial agriculture practices⁵¹. As such both crop plants (i.e. tomatoes^{39b, 39e} and rice⁴⁷) and model systems (*Arabidopsis thaliana*^{39a, 39c}) have been examined using metabolomics. In the most recent study of the effects of cadmium on *A. thaliana* whole plants, GC-MS metabolomics was used to identify a number of interesting metabolite changes^{39c}. Several compounds with compatible solute properties such as alanine, proline, glycerol and trehalose were found to increase as well as antioxidant molecules such as campesterol and isoflavone in order to respectively counteract the osmotic and oxidative stress induced by cadmium exposure. In the older *Arabidopsis* cell culture study, LC-MS was used for metabolite analysis. As such, metabolites were not identified until after multivariate statistical analysis had picked the spectral peaks that were contributing the most to the variation between the control and cadmium exposed samples. This kind of analysis was very useful in this study as it allowed for the identification of many phytochelatins that were being produced in the exposed samples^{39a}. As for crop plants, two studies have examined the short and long-term effects of cadmium on tomato

plants using $^1\text{H-NMR}$ metabolomics^{39b, 39e}. Their results show that metabolite responses depend on the tissue (i.e. leaves vs. roots) and tissue age, exposure concentration and that there are similarities and differences between acute and chronic exposure. Citrate was found to be increased in both roots and leaves though it was postulated that this accumulation was due to photosynthetic activity in the leaves and as a chelating agent in the roots^{39e}. A decrease in carotenoid metabolites was only observed in mature leaves in response to cadmium stress^{39b}. Chronic exposure (90 days) induced an increase in sucrose but a decrease in fructose and glucose levels in leaves hypothesized to be due to the inhibition of invertase^{39b}, alterations which were not observed after acute exposure (10 days)^{39e} which suggests a more long-term metabolic change. A different study on rice roots used both GC-MS and $^1\text{H-NMR}$ metabolomics in conjunction with in depth transcriptomic analysis in order to elucidate the effects of chromium (VI) toxicity⁴⁷. Since a non-biased quantitative approach was used, the metabolomic data obtained was more in depth about cellular metabolism. The proline and its plant precursor ornithine were observed to increase in treated plants likely to protect against osmotic stress as proline is a compatible osmolyte⁴⁷. Interestingly, corresponding changes to the transcriptome were not observed but a specific protein regulatory mechanism was not proposed. Increased levels of linoleic acid were also observed, an expected result as chromium stress is known to increase the levels of unsaturated fatty acids in plants. The metabolomic data was also integrated with the microarray dataset and co-analyzed to find that the sucrose degradation pathway was down regulated whereas fermentative pathways were activated as a rescue mechanism⁴⁷. These results exemplify several of the features that make metabolomics such a powerful technique. Since metabolites are the basal level of functionality within the cell, their quantitation can be taken as an accurate snapshot of the state of the cell. The observed proline changes would have been missed had only transcriptomics been performed, instead a known stress

response mechanism was observed. Similarly, the linoleic acid changes corresponded to a known toxicity mechanism, which grants the metabolomic data validity and thereby gives confidence to the novel finding that fermentative pathways are activated in response to chromium stress. This kind of result where the metabolomic data matches past conventional studies, which was also observed in the clam studies, is frequent in metabolomics studies that serve to support the validity of the metabolomic experiment. This validation allows for the remaining unexplained metabolomic data to be used for hypothesis generation, one of the main strengths of omic analyses. For the purposes of metal toxicity the utility of this kind of data is obvious as the global view of the cell provided by metabolomics can be used to propose, refute and explain mechanisms of metal toxicity without being tainted by questions about expression and regulation.

A.4.4. Bacteria

Evaluation of metal interactions with microorganisms has been underway since the 1970's. The focus over the decades have had two general research fields, that of geomicrobiology and medical microbiology. The work, therefore, was split between evaluating organisms involved in biogeochemistry and that of the evaluation of specific genetic determinants that are responsible for resistance towards metal ions. This work has lead to an understanding of metal bioleaching and geochemical cycling of metals in the environment and a good understanding of the prevalence of metal resistance determinants and their genetics and biochemistry. However, examination of the literature reveals an evaluation of toxic metal resistance and tolerance but little attention has been given to understanding how metals generate these toxic effects. Microbes make up ~50% of the carbon biomass of the planet and they are a key foundation of food chains via the higher organisms feeding on free microbes as well as in periphyton biofilms. Thus metal bioaccumulation and level

of resistance and or metal species bioconversion becomes an important consideration for biosphere health and toxic metal cycling.

Over the past few years our group has produced studies examining metal toxicity in bacteria using metabolomics. The first examined the differences between wild-type and a mutant that is hyper resistant to the highly toxic metalloid oxyanion, tellurite ^{40a}. A second study made use of metabolomics to distinguish metabolic differences between two phenotypic variants and the ancestral strain that showed altered metal tolerances ^{40b}. A recent study examined the difference between the effects of copper toxicity in free swimming planktonic populations and surface-attached biofilms ^{17a}. Quantitative ¹H-NMR metabolomics was used in all the studies while a combined approach also using GC-MS in the latter study. While the questions that were studied were quite similar, the metabolomics approach provided a breadth of novel results. In *Pseudomonas pseudoalcaligenes* KF707 cultures, it was found that a tellurite hyper resistance mutant had a comparable metabolic profile to the tellurite pre-exposed wild-type. This indicated that it was the mutant's metabolic priming by upregulating the oxidative stress metabolites glutathione and branched chain amino acids and the osmoprotectant betaine that was conferring the resistance advantage. Certain global regulatory mutants of *Pseudomonas fluorescens* had been found to produce more colony variants than the wildtype strain. Analysis of two of these variants, the mutant strain and the wild-type showed that all four had significant metabolic differences. While this demonstrates the ability of metabolomics to distinguish the phenotypes of both genetically different and identical strains, the metabolite changes were also shown to correlate to differences between the strains ability to withstand the toxic metals copper and silver. For example, as would be expected, glutathione was correlated with sensitivity to these metals. Surface attached assemblages of bacteria, called biofilms, have been found to be able to withstand the toxic effects

of metals far better than free swimming planktonic populations of the same species^{18a}. With improved techniques of bacterial biofilm growth, metabolite extraction, and data analysis the differences between biofilm and planktonic cultures exposed to copper was able to go beyond the clustering results. This demonstrated that metabolism was different between all the cultures of planktonic and biofilm and their response to copper^{17a}. Since in addition to NMR, GC-MS was also used in this study, it was possible to obtain a much larger list of metabolites, which allowed for the use of TICL analysis. This showed the metabolic pathways represented by the lists of metabolites that were changing most in response to copper in each culture type. The response was markedly different in biofilms compared to planktonics as oxidative stress related metabolism was prevalent in the free swimming population whereas sugar and polymer metabolism (related to biofilm slime production) was more affected in the biofilms. Based on hierarchical clustering analysis it was also observed that the metabolism changes in the biofilms were more concerted than in the planktonic cultures indicating that the biofilms were responding to the stress rather than reacting.

A.5. Concluding comments

The above examples demonstrate the power of metabolomics for revealing features of cellular metal toxicity that would be difficult to elucidate with other approaches. We expect that given further development of these methods, metabolomics will become an important tool for those investigating metal toxicity. As this field is still developing, especially when compared to more established 'omics' techniques such as transcriptomics, we are only beginning to learn some of the true potentials as well as downfalls of this approach. Below we highlight the pros and cons of metabolomics and provide a critical evaluation weighing the cost and effort required as a function of the useful information gained.

For metal toxicity studies, metabolomics is a potential route for identifying novel detoxification mechanisms in a variety of organisms. Also, it may be used to identify changes in metabolic pathways that would be missed by other ‘omics’ techniques. However, transcriptomic and proteomic studies have the potential to miss a great deal of information. Levels of proteins and mRNA in the cell do not necessarily translate to changes in metabolic pathways and cannot detect changes in metabolic flux. Here, metabolomics promises to fill the gap and provide information not accessible through other techniques.

Detailing the small-molecule complement of the cell is very important and has already revealed important information regarding microbial metal toxicity^{17a}. However, there are some very important aspects of metabolomics that need to be considered for these types of studies. One of the more significant challenges is identifying which metabolic pathways are altered when using a knowledge-discovery approach or metabolome survey. These types of approaches aim to measure as many metabolites as possible in a sample, or set of samples and use this information to generate hypothesis for further targeted work. In this way novel information can be obtain about the process of interest. Our own group has used this approach quite successfully to identify a number of significant metabolic shifts in metal-exposed microbial samples^{17a, 40}, as have the group working with metal exposed molluscs⁴¹. However, a key issue for these studies, and others, is pinpointing the pathways that are altered, particularly for central metabolic pathways where any one metabolite may be involved in a myriad of pathways. As mentioned previously, tools are available which attempt to identify the altered pathways through an enrichment analysis and to build networks to help interpret the data. Even so identifying the key pathways for follow-up work remains a significant challenge. Isotopic labelling a substrate and following this label through various metabolic pathways is one way around this challenge⁵², however this requires prior

knowledge of target pathways and may limit metabolism by the use of a single carbon source. Beginning with a broad survey and subsequent focus on key pathways of interest with a targeted analysis will likely yield the most useful information.

Many metabolomic studies to date have been focused on the generation and interpretation of metabolite data only. Furthermore, microbial metabolomics, including our studies, have been primarily proof-of-principle studies demonstrating the usefulness of metabolomics for metal toxicity and resistance^{17a, 40a}, identification of bacterial species⁵³, host-pathogen interactions⁵⁴, and establishment and testing of analytical platforms^{7a, 8, 10b}. Although all of these studies have provided valuable information, in order for metabolomics data to be most useful it will also need to be used in conjunction with other systems biology approaches⁵⁵ as well as becoming an integrated part of physiological analysis, much in the way that transcriptomics and proteomics provide essential information but are not the sole focus of a study. A good example of this is the integration of metabolomics with temporal gene expression analysis to identify biofilm-induced global metabolic shifts towards stress tolerance and gluconeogenesis in *Salmonella*^{17b}. These types of studies will lead to a more complete picture of cellular metabolism, not attainable with either technique alone.

Overall metabolomics holds tremendous promise for revealing a layer of information not given with other 'omics' techniques. Metabolite profiling can reveal the "real-time" physiology of cell, not just the potential. Furthermore, metabolomics can, at least in theory, give you a true global picture of the metabolic process in a cell and can provide target pathways for downstream analysis. However, even though metabolomics is developing as an independent field and much progress has been made, there are still significant challenges to be addressed. Technically, the metabolomics approaches are demanding and can be expensive. This is primarily due to the large

chemical diversity present in the metabolome as well as rapid metabolite turnover, which can also lead to difficulties with reproducibility. Regardless, the approach can still be quite high throughput. Other issues that must be addressed when considering metabolomic experiments are data management and analysis; however, new tools to assist researchers with these challenges are beginning to emerge. Despite these challenges, metabolomics holds tremendous potential to reveal novel and important information for metal toxicity and other areas. With good experimental design, effective analysis methods, and integration with other approaches, metabolomics can provide a wealth of information not attainable any other way.

A.6. References

1. D. A. Fell, Beyond genomics. *Trends in Genetics* 2001, 17. 680-682.
2. O. Fiehn, Metabolomics - The link between genotypes and phenotypes. *Plant Molecular Biology* 2002, 48. 155-171.
3. (a) M. R. Mashego, K. Rumbold, M. Mey, E. Vandamme, W. Soetaert, J. J. Heijnen, Microbial metabolomics: past, present and future methodologies. *Biotechnology Letters* 2006, 29. 1-16, DOI: 10.1007/s10529-006-9218-0;(b) D. Ryan, K. Robards, Metabolomics: The greatest omics of them all? *Analytical Chemistry* 2006, 78. 7954-7958.
4. I. M. Keseler, C. Bonavides-Martínez, J. Collado-Vides, S. Gama-Castro, R. P. Gunsalus, D. A. Johnson, M. Krummenacker, L. M. Nolan, S. Paley, I. T. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, A. G. Shearer, P. D. Karp, EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Research* 2009, 37. D464-D470.
5. I. Nobeli, H. Ponstingl, E. B. Krissinel, J. M. Thornton, A structure-based anatomy of the *E. coli* metabolome. *Journal of Molecular Biology* 2003, 334. 697-719.
6. M. Robert, T. Soga, M. Tomita. 2007, vol. 18, pp 189-234.
7. (a) M. J. v. d. Werf, K. M. Overkamp, B. Muilwijk, L. Coulier, T. Hankemeier, Microbial metabolomics: Toward a platform with full metabolome coverage. *Analytical Biochemistry* 2007, 370. 17-25;(b) X. H. Wu, H. L. Yu, Z. Y. Ba, J. Y. Chen, H. G. Sun, B. Z. Han, Sampling methods for NMR-based metabolomics of *Staphylococcus aureus*. *Biotechnology Journal* 2010, 5. 75-84.
8. M. J. Van Der Werf, K. M. Overkamp, B. Muilwijk, M. M. Koek, B. J. C. Van Der Werff-Van Der Vat, R. H. Jellema, L. Coulier, T. Hankemeier, Comprehensive analysis of the metabolome of *Pseudomonas putida* S12 grown on different carbon sources. *Molecular BioSystems* 2008, 4. 315-327.

9. (a) C. L. Winder, W. B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G. M. Stephens, R. Goodacre, Global metabolic profiling of *Escherichia coli* cultures: An evaluation of methods for quenching and extraction of intracellular metabolites. *Analytical Chemistry* 2008, 80. 2939-2948;(b) W. M. van Gulik, Fast sampling for quantitative microbial metabolomics. *Current Opinion in Biotechnology* 2010, 21. 27-34;(c) R. Prasad Maharjan, T. Ferenci, Global metabolite analysis: The influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Analytical Biochemistry* 2003, 313. 145-154.
10. (a) W. B. Dunn, N. J. C. Bailey, H. E. Johnson, Measuring the metabolome: Current analytical technologies. *Analyst* 2005, 130. 606-625;(b) M. M. Koek, B. Muilwijk, M. J. Van Der Werf, T. Hankemeier, Microbial metabolomics with gas chromatography/mass spectrometry. *Analytical Chemistry* 2006, 78. 1272-1281;(c) A. M. Weljie, J. Newton, P. Mercier, E. Carlson, C. M. Slupsky, Targeted Profiling: Quantitative Analysis of ¹H NMR Metabolomics Data. *Analytical Chemistry* 2006, 78. 4430-4442, DOI: 10.1021/ac060209g.
11. W. De Koning, K. Van Dam, A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Analytical Biochemistry* 1992, 204. 118-123.
12. (a) L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. Van Dam, S. G. Oliver, A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology* 2001, 19. 45-50;(b) W. Weckwerth, M. E. Loureiro, K. Wenzel, O. Fiehn, Differential metabolic networks unravel the effects of silent plant phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101. 7809-7814.

13. P. J. Turnbaugh, J. I. Gordon, An Invitation to the marriage of metagenomics and metabolomics. *Cell* 2008, 134. 708-713.
14. S. Vaidyanathan, Profiling microbial metabolomes: what do we stand to gain? *Metabolomics* 2005, 1. 17-28, DOI: 10.1007/s11306-005-1104-6.
15. J. Lemire, P. Kumar, R. Mailloux, K. Cossar, V. D. Appanna, Metabolic adaptation and oxaloacetate homeostasis in *P. fluorescens* exposed to aluminum toxicity. *Journal of Basic Microbiology* 2008, 48. 252-259, DOI: 10.1002/jobm.200800007.
16. J. J. Harrison, H. Ceri, J. Yerly, M. Rabiei, Y. Hu, R. Martinuzzi, R. J. Turner, Metal ions may suppress or enhance cellular differentiation in *Candida albicans* and *Candida tropicalis* biofilms. *Applied and Environmental Microbiology* 2007, 73. 4940-9, DOI: 10.1128/AEM.02711-06.
17. (a) S. C. Booth, M. L. Workentine, J. Wen, R. Shaykhutdinov, H. J. Vogel, H. Ceri, R. J. Turner, A. M. Weljie, Differences in Metabolism between the Biofilm and Planktonic Response to Metal Stress. *Journal of Proteome Research* 2011. null-null, DOI: 10.1021/pr2002353;(b) A. P. White, A. M. Weljie, D. Apel, P. Zhang, R. Shaykhutdinov, H. J. Vogel, M. G. Surette, A global metabolic shift is linked to *Salmonella* multicellular development. *PLoS ONE* 2010, 5.
18. (a) J. J. Harrison, H. Ceri, R. J. Turner, Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology* 2007, 5. 928-938, DOI: 10.1038/nrmicro1774;(b) M. L. Workentine, J. J. Harrison, P. U. Stenroos, H. Ceri, R. J. Turner, *Pseudomonas fluorescens*' view of the periodic table. *Environmental Microbiology* 2008, 10. 238-50, DOI: 10.1111/j.1462-2920.2007.01448.x.
19. (a) B. Álvarez-Sánchez, F. Priego-Capote, M. D. L. d. Castro, Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC - Trends in Analytical Chemistry* 2010,

29. 120-127;(b) B. Álvarez-Sánchez, F. Priego-Capote, M. D. Luque de Castro, *Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. TrAC - Trends in Analytical Chemistry* 2010, 29. 111-119.
20. M. R. Viant, Recent developments in environmental metabolomics. *Molecular BioSystems* 2008, 4. 980-986.
21. J. G. Bundy, H. C. Keun, J. K. Sidhu, D. J. Spurgeon, C. Svendsen, P. Kille, A. J. Morgan, Metabolic profile biomarkers of metal contamination in a sentinel terrestrial species are applicable across multiple sites. *Environmental Science and Technology* 2007, 41. 4458-4464.
22. M. L. Reaves, J. D. Rabinowitz, Metabolomics in systems microbiology. *Current Opinion in Biotechnology* 2011, 22. 17-25.
23. M. R. Mashego, L. Wu, J. C. Van Dam, C. Ras, J. L. Vinke, W. A. Van Winden, W. M. Van Gulik, J. J. Heijnen, MIRACLE: Mass Isotopomer Ratio Analysis of U-13C-Labeled Extracts. A New Method for Accurate Quantification of Changes in Concentrations of Intracellular Metabolites. *Biotechnology and Bioengineering* 2004, 85. 620-628.
24. B. D. Bennett, J. Yuan, E. H. Kimball, J. D. Rabinowitz, Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols* 2008, 3. 1299-1311.
25. S. G. Villas-Bôas, P. Bruheim, Cold glycerol-saline: The promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Analytical Biochemistry* 2007, 370. 87-97.
26. J. L. Spratlin, N. J. Serkova, S. G. Eckhardt, Clinical applications of metabolomics in oncology: A review. *Clinical Cancer Research* 2009, 15. 431-440.

27. D. S. Wishart, Quantitative metabolomics using NMR. *Trac-Trends Anal. Chem.* 2008, 27. 228-237, DOI: 10.1016/j.trac.2007.12.001.
28. S. C. Brown, G. Kruppa, J. L. Dasseux, Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrometry Reviews* 2005, 24. 223-231.
29. J. Hummel, J. Selbig, D. Walther, J. Kopka. 2007, vol. 18, pp 75-95.
30. J. Hummel, N. Strehmel, J. Selbig, D. Walther, J. Kopka, Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics* 2010, 6. 322-333.
31. Á. Cuadros-Inostroza, C. Caldana, H. Redestig, M. Kusano, J. Lisec, H. Peña-Cortés, L. Willmitzer, M. A. Hannah, TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics* 2009, 10.
32. C. D. Broeckling, I. R. Reddy, A. L. Duran, X. Zhao, L. W. Sumner, MET-IDEA: Data extraction tool for mass spectrometry-based metabolomics. *Analytical Chemistry* 2006, 78. 4334-4341.
33. K. Hiller, J. Hangebrauk, C. Jäger, J. Spura, K. Schreiber, D. Schomburg, Metabolite detector: Comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Analytical Chemistry* 2009, 81. 3429-3439.
34. A. V. Antonov, S. Dietmann, P. Wong, H. W. Mewes, TICL - a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. *Febs Journal* 2009, 276. 2084-2094, DOI: 10.1111/j.1742-4658.2009.06943.x.
35. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 2000, 28. 27-30.

36. L. Cottret, D. Wildridge, F. Vinson, M. P. Barrett, H. Charles, M. F. Sagot, F. Jourdan, MetExplore: A web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Research* 2010, 38. W132-W137.
37. M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, T. Ideker, Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 2011, 27. 431-432.
38. (a) J. G. Bundy, M. P. Davey, M. R. Viant, Environmental metabolomics: A critical review and future perspectives. *Metabolomics* 2009, 5. 3-21;(b) C. Y. Lin, M. R. Viant, R. S. Tjeerdema, Metabolomics: Methodologies and applications in the environmental sciences. *Journal of Pesticide Science* 2006, 31. 245-251.
39. (a) J. E. Sarry, L. Kuhn, C. Ducruix, A. Lafaye, C. Junot, V. Hugouvieux, A. Jourdain, O. Bastien, J. B. Fievet, D. Vailhen, B. Amekraz, C. Moulin, E. Ezan, J. Garin, J. Bourguignon, The early responses of *Arabidopsis thaliana* cells to cadmium exposure explored by protein and metabolite profiling analyses. *Proteomics* 2006, 6. 2180-2198;(b) H. Hédiji, W. Djebali, C. Cabasson, M. Maucourt, P. Baldet, A. Bertrand, L. Boulila Zoghlami, C. Deborde, A. Moing, R. Brouquisse, W. Chaïbi, P. Gallusci, Effects of long-term cadmium exposure on growth and metabolomic profile of tomato plants. *Ecotoxicology and Environmental Safety* 2010, 73. 1965-1974;(c) X. Sun, J. Zhang, H. Zhang, Y. Ni, Q. Zhang, J. Chen, Y. Guan, The responses of *Arabidopsis thaliana* to cadmium exposure explored via metabolite profiling. *Chemosphere* 2010, 78. 840-845;(d) X. Liu, C. Yang, L. Zhang, L. Li, S. Liu, J. Yu, L. You, D. Zhou, C. Xia, J. Zhao, H. Wu, Metabolic profiling of cadmium-induced effects in one pioneer intertidal halophyte *Suaeda salsa* by NMR-based metabolomics. *Ecotoxicology* 2011. 1-10;(e) L. B. Zoghlami, W. Djebali, Z. Abbes, H. Hediji, M. Maucourt, A. Moing, R. Brouquisse, W. Chaïbi, Metabolite modifications in

Solanum lycopersicum roots and leaves under cadmium stress. *African Journal of Biotechnology* 2011, 10. 567-579.

40. (a) V. Tremaroli, M. L. Workentine, A. M. Weljie, H. J. Vogel, H. Ceri, C. Viti, E. Tatti, P. Zhang, A. P. Hynes, R. J. Turner, D. Zannoni, Metabolomic investigation of the bacterial response to a metal challenge. *Applied and Environmental Microbiology* 2009, 75. 719-28, DOI: 10.1128/AEM.01771-08;(b) M. L. Workentine, J. J. Harrison, A. M. Weljie, V. A. Tran, P. U. Stenroos, V. Tremaroli, H. J. Vogel, H. Ceri, R. J. Turner, Phenotypic and metabolic profiling of colony morphology variants evolved from *Pseudomonas fluorescens* biofilms. *Environmental Microbiology* 2010, 12. 1565-1577, DOI: 10.1111/j.1462-2920.2010.02185.x.

41. (a) X. Liu, L. Zhang, L. You, M. Cong, J. Zhao, H. Wu, C. Li, D. Liu, J. Yu, Toxicological responses to acute mercury exposure for three species of Manila clam *Ruditapes philippinarum* by NMR-based metabolomics. *Environmental Toxicology and Pharmacology* 2011, 31. 323-332;(b) X. Liu, L. Zhang, L. You, J. Yu, J. Zhao, L. Li, Q. Wang, F. Li, C. Li, D. Liu, H. Wu, Differential toxicological effects induced by mercury in gills from three pedigrees of Manila clam *Ruditapes philippinarum* by NMR-based metabolomics. *Ecotoxicology* 2011, 20. 177-186;(c) H. Wu, W. X. Wang, Tissue-specific toxicological effects of cadmium in green mussels (*Perna viridis*): Nuclear magnetic resonance-based metabolomics study. *Environmental Toxicology and Chemistry* 2011, 30. 806-812;(d) L. Zhang, X. Liu, L. You, D. Zhou, H. Wu, L. Li, J. Zhao, J. Feng, J. Yu, Metabolic responses in gills of Manila clam *Ruditapes philippinarum* exposed to copper using NMR-based metabolomics. *Marine Environmental Research*.

42. (a) J. G. Bundy, D. J. Spurgeon, C. Svendsen, P. K. Hankard, J. M. Weeks, D. Osborn, J. C. Lindon, J. K. Nicholson, Environmental metabonomics: Applying combination biomarker analysis in earthworms at a metal contaminated site. *Ecotoxicology* 2004, 13. 797-806;(b) J. G.

Bundy, J. K. Sidhu, F. Rana, D. J. Spurgeon, C. Svendsen, J. F. Wren, S. R. Stürzenbaum, A. J. Morgan, P. Kille, 'Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm *Lumbricus rubellus*. *BMC Biology* 2008, 6;(c) Q. Guo, J. K. Sidhu, T. M. D. Ebbels, F. Rana, D. J. Spurgeon, C. Svendsen, S. R. Stürzenbaum, P. Kille, A. J. Morgan, J. G. Bundy, Validation of metabolomics for toxic mechanism of action screening with the earthworm *Lumbricus rubellus*. *Metabolomics* 2009, 5. 72-83.

43. (a) J. L. Griffin, L. A. Walker, R. F. Shore, J. K. Nicholson, Metabolic profiling of chronic cadmium exposure in the rat. *Chemical Research in Toxicology* 2001, 14. 1428-1434;(b) Y. Sugiura, M. Kashiba, K. Maruyama, K. Hoshikawa, R. Sasaki, K. Saito, H. Kimura, N. Goda, M. Suematsu, Cadmium exposure alters metabolomics of sulfur-containing amino acids in rat testes. *Antioxidants and Redox Signaling* 2005, 7. 781-787;(c) H. Wu, X. Zhang, X. Li, Z. Li, P. Liao, W. Li, Y. Wu, F. Pei, Investigation on acute biochemical effects of $Ce(NO_3)_3$ on liver and kidney tissues by MAS 1H NMR spectroscopic-based metabonomic approach. *Journal of Rare Earths* 2006, 24. 357-363;(d) R. Lei, C. Wu, B. Yang, H. Ma, C. Shi, Q. Wang, Y. Yuan, M. Liao, Integrated metabolomic analysis of the nano-sized copper particle-induced hepatotoxicity and nephrotoxicity in rats: A rapid in vivo screening method for nanotoxicity. *Toxicology and Applied Pharmacology* 2008, 232. 292-301;(e) P. Liao, L. Wei, H. Wu, W. Li, Y. Wu, X. Li, J. Ni, F. Pei, Biochemical effects of gadolinium chloride in rats liver and kidney studied by 1H NMR metabolomics. *Journal of Rare Earths* 2009, 27. 280-287;(f) J. L. Griffin, L. Walker, R. F. Shore, J. K. Nicholson, High-resolution magic angle spinning 1H-NMR spectroscopy studies on the renal biochemistry in the bank vole (*Clethrionomys glareolus*) and the effects of arsenic (As^{3+}) toxicity. *Xenobiotica* 2001, 31. 377-385;(g) Q. Bu, G. Yan, P. Deng, F. Peng, H. Lin, Y. Xu, Z. Cao, T.

- Zhou, A. Xue, Y. Wang, X. Cen, Y. L. Zhao, NMR-based metabonomic study of the sub-acute toxicity of titanium dioxide nanoparticles in rats after oral administration. *Nanotechnology* 2010, 21.
44. E. M. Santos, J. S. Ball, T. D. Williams, H. Wu, F. Ortega, R. Van Aerle, I. Katsiadaki, F. Falciani, M. R. Viant, J. K. Chipman, C. R. Tyler, Identifying health impacts of exposure to copper using transcriptomics and metabolomics in a fish model. *Environmental Science and Technology* 2010, 44. 820-826.
45. S. L. Hughes, J. G. Bundy, E. J. Want, P. Kille, S. R. Stürzenbaum, The metabolomic responses of *Caenorhabditis elegans* to cadmium are largely independent of metallothionein status, but dominated by changes in cystathionine and phytochelatins. *Journal of Proteome Research* 2009, 8. 3512-3519.
46. (a) H. C. Poynton, N. S. Taylor, J. Hicks, K. Colson, S. Chan, C. Clark, L. Scanlan, A. V. Loguinov, C. Vulpe, M. R. Viant, Metabolomics of microliter hemolymph samples enables an improved understanding of the combined metabolic and transcriptional responses of *Daphnia magna* to cadmium. *Environmental Science and Technology* 2011, 45. 3710-3717;(b) N. S. Taylor, R. J. M. Weber, A. D. Southam, T. G. Payne, O. Hrydziusko, T. N. Arvanitis, M. R. Viant, A new approach to toxicity testing in *Daphnia magna*: Application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* 2009, 5. 44-58.
47. S. Dubey, P. Misra, S. Dwivedi, S. Chatterjee, S. K. Bag, S. Mantri, M. H. Asif, A. Rai, S. Kumar, M. Shri, P. Tripathi, R. D. Tripathi, P. K. Trivedi, D. Chakrabarty, R. Tuli, Transcriptomic and metabolomic shifts in rice roots in response to Cr (VI) stress. *BMC Genomics* 2010, 11.
48. G. Otting. 2010, vol. 39, pp 387-405.

49. M. J. Simpson, J. R. McKelvie, Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. *Analytical and Bioanalytical Chemistry* 2009, 394. 137-149.
50. N. J. C. Bailey, M. Oven, E. Holmes, J. K. Nicholson, M. H. Zenk, Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ¹H NMR spectroscopy and chemometrics. *Phytochemistry* 2003, 62. 851-858.
51. L. Sanità Di Toppi, R. Gabbrielli, Response to cadmium in higher plants. *Environmental and Experimental Botany* 1999, 41. 105-130.
52. C. L. Winder, W. B. Dunn, R. Goodacre, TARDIS-based microbial metabolomics: time and relative differences in systems. *Trends in Microbiology* 2011. 1-8.
53. J. G. Bundy, T. L. Willey, R. S. Castell, D. J. Ellar, K. M. Brindle, Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomic profiling. *FEMS Microbiology Letters* 2005, 242. 127-136.
54. (a) C. M. Slupsky, A. Cheypesh, D. V. Chao, H. Fu, K. N. Rankin, T. J. Marrie, P. Lacy, *Streptococcus pneumoniae* and *Staphylococcus aureus* pneumonia induce distinct metabolic responses. *Journal of Proteome Research* 2009, 8. 3029-3036;(b) C. M. Slupsky, K. N. Rankin, H. Fu, D. Chang, B. H. Rowe, P. G. P. Charles, A. McGeer, D. Low, R. Long, D. Kunimoto, M. B. Sawyer, R. N. Fedorak, D. J. Adamko, E. J. Saude, S. L. Shah, T. J. Marrie, Pneumococcal pneumonia: Potential for diagnosis through a urinary metabolic profile. *Journal of Proteome Research* 2009, 8. 5550-5558.
55. W. Zhang, F. Li, L. Nie, Integrating multiple 'omics' analysis for microbial biology: Application and methodologies. *Microbiology* 2010, 156. 287-301.

Appendix B Preface

This invited review from *Computational and Structural Biology Journal* aimed to provide a practical summary of the currently available tools for analyzing metabolomics experiments. As these kinds of studies produce metabolite concentrations for dozens to hundreds of compounds, understanding the meaning of these concentrations is a difficult task that requires computational assistance. At the time, many different groups had developed tools for performing what I called ‘secondary analysis’ where these large complex datasets were distilled into digestible figures and tables. As I had recently run into the issue of needing these tools and was expecting to need them again, I set out to determine the pros and cons of the available tools. This review provides a guide on how to perform secondary analysis and the benefits provided by each of the different tools. I investigated all the tools and wrote the manuscript with guidance and editorial assistance from my co-authors.

APPENDIX B: COMPUTATIONAL TOOLS FOR THE SECONDARY ANALYSIS OF METABOLOMICS EXPERIMENTS

Sean C. Booth, Aalim M. Weljie and Raymond J. Turner

Published in *Computational and Structural Biology Journal*, January 2013, 4:5 e201301003

Copyright permissions for the reproduction of this manuscript can be found in Appendix E

B.1. Abstract

Metabolomics experiments have become commonplace in a wide variety of disciplines. By identifying and quantifying metabolites researchers can achieve a systems level understanding of metabolism. These studies produce vast swaths of data which are often only lightly interpreted due to the overwhelmingly large amount of variables that are measured. Recently, a number of computational tools have been developed which enable much deeper analysis of metabolomics data. These data have been difficult to interpret as understanding the connections between dozens of altered metabolites has often relied on the biochemical knowledge of researchers and their speculations. Modern biochemical databases provide information about the interconnectivity of metabolism which can be automatically polled using metabolomics secondary analysis tools. Starting with lists of altered metabolites, there are two main types of analysis: enrichment analysis computes which metabolic pathways have been significantly altered whereas metabolite mapping contextualizes the abundances and significances of measured metabolites into network visualizations. Many different tools have been developed for one or both of these applications. In this review the functionality and use of these software is discussed. Together these novel secondary

analysis tools will enable metabolomics researchers to plumb the depths of their data and produce farther reaching biological conclusions than ever before.

B.2. Introduction

Over the past decade metabolomics has emerged as a powerful tool used in a variety of quite diverse fields for hypothesis development, to elaborate unknown gene functions, biomarker discovery and to complement proteomic and transcriptomic experiments. While considerable progress has been made, the datasets obtained from metabolomics experiments still remain extremely large and dense and thus subsequently a challenge to interpret and derive biological meaning. This challenge lies in the difficulty of understanding how dozens of chemically diverse compounds, a small subset of the hundreds to thousands of metabolites present within cells, are functionally related to each other and the perturbed condition of the experiment. While it is possible, and common, for experimenters to intuitively interpret these results using their knowledge of metabolism and the tested conditions, or manually map them onto known metabolic pathways, computational analysis allows for more comprehensive interpretation. As metabolomics remains a developing field, bioinformatic tools designed to perform this task continue to be developed and released by various groups using diverse algorithms. While many databases, tools and projects such as the human metabolome database [1] have focused on creating tools specifically for interpreting human metabolomics experiments, the options for more diverse organism metabolomics are somewhat limited. This review seeks to introduce the problems faced when interpreting metabolomics results and describe the most current approaches to solving these problems in various model and experimental systems without a human centric bias.

B.3. Background

The central dogma of molecular biology delineated the basic transfer of biological information as moving from DNA to RNA to protein [2]. While many proteins interact with each other and the nucleic acids, the real metabolic function of the cell relies on the enzymatic interconversion of the various small, low molecular weight compounds, termed metabolites [3]. These metabolites represent the actual functional phenotype of the cell that when systematically identified and quantified, the process of metabolomics, will show an accurate snapshot of the cell's physiological state [4]. A relative newcomer to the 'omics' field compared to proteomics and transcriptomics, the technologies and techniques behind metabolomics have been evolving rapidly to even the point where commercial kits are available for common clinical samples [5]. While still a developing field, excellent reviews of topics in designing a metabolomics experiment from sample selection and preparation [6,7], analytical techniques [4,8] to data processing [9,10] and statistics [11] are available. The frequent final product of the metabolomics pipeline is the generation of a list of metabolites whose concentrations have been (significantly) altered which must be interpreted in order to derive biological meaning. While tools designed for this function exist, the development of many of these tools have been driven by the application of metabolomics to human pathologies such as kidney [12], heart [13], and neurological [14] disease and especially cancer [15,16] leaving more broadly applicable tools lagging somewhat behind. Additionally, there is no widely accepted standard for the computational interpretation of metabolite data whereas the interpretation of protein and transcript expression datasets is much more mature [17]. To fill these voids a number of tools have recently been developed with fresh ideas, providing new releases constantly, as this field emerges out of its adolescence. A challenge though is that as of yet, none have emerged as a standardized approach. Here, current solutions for metabolomic data

interpretation will be described with reference to studies that have taken advantage of these new methods will be presented. Throughout, tools with a focus on those which can be broadly applied to any organism will be highlighted.

Metabolomics requires many steps and choices before getting to the point of data interpretation which will affect how this process is undertaken. The main decisions are analytical platform (likely GC/LC-MS or NMR as they are the most common), each with their own advantages and disadvantages though the choice will more likely be dictated by instrument availability and analytical method (chemometric or quantitative) determined by the scope of the experiment. GC-MS is an extremely common metabolomics platform, resulting in a high frequency of tools which allow for the direct input of GC-MS spectra. The popularity of GC-MS is due to its relatively high sensitivity, broad range of detectable metabolites, existence of well-established identification libraries and ease of automation [18]. Even with its popularity, separation-coupled MS data requires much processing and careful handling to ensure the information it contains is not artifactual [19]. While scientists have been quantifying metabolite levels for over 50 years through targeted analysis, the focus here will be on untargeted metabolomics as the problem of interpretation mainly needs to be dealt with for this kind of experiment. Untargeted metabolomics aims to identify and quantify as many of the metabolites in a sample as possible then determine which are important, rather than focusing on identifying and quantifying a specific set of metabolites which are expected to be important (the targeted approach) [20]. When this approach is undertaken with one of the three most common instruments (GC/LC-MS or NMR) metabolites are identified by using pure reference spectra (plus chromatographic information if applicable) which also allow for quantification [10]. These techniques, and others were reviewed by Zhang et al. [8].

Advances in instrumentation and technical treatment of samples as well as data preprocessing and development of improved databases have been arriving rapidly in metabolomics leading to ever increasing numbers of metabolites identified and accuracy of their quantification [8]. With these improvements, one would expect the results of metabolomics to have a profound effect on the questions they're being applied to. Indeed, metabolomics approaches have shown many successes in identifying potential therapeutic targets and also assigning function to unknown genes/proteins [20], thus effectively connecting to the field of functional genomics. Phenotype characterization studies however, such as in environmental metabolomics, often tend to be limited to speculating cause/effect relationships based on prior knowledge [21]. Many studies results' are discussed in terms of 'suggestions', 'correlations', or the individual metabolites changing are not even discussed, just the fact that discerning metabolic patterns are identifiable [22]. This process of comparing metabolic profiles and only looking for differences is more exaggerated when metabolomics is used for biomarker discovery. This process of identifying specific metabolites that are altered in a disease state, as well as general metabolic differences is common in metabolomics studies of human pathologies [12,15,23]. Putative biomarkers are often then confirmed using a second dataset and/or by confirmatory experiments examining the metabolite in cell cultures [12]. While these methods have obvious and well-realized implications in the clinical field, current metabolomics interpretations, especially outside of human medicine, are generally over-reliant on additional research for explanations as well as providing underwhelming conclusions for data that purports to represent the basal level of functionality within the cell culture, tissue or organism. Fortunately, tools designed to better mine and interpret metabolomics data have been under rapid development recently. Indeed, this step has been called a 'bottleneck' in the metabolomics pipeline [24,25,26].

B.4. Key Issues

It is in the final steps of interpretation where the most potential remains to improving the quality of information obtained from metabolomics [27,28,29,30]. By this point though prior steps have created several problems which must be overcome when interpreting metabolomics data: 1) All of the metabolites within a system cannot be identified with any one analytical method due to chemical heterogeneity, which will cause downstream issues as all metabolites in a pathway have not been quantified; 2) not all metabolites have been identified and characterized and so do not exist in the standards libraries, leading to large number of unannotated and/or unknown metabolites of interest; 3) organism specific metabolic databases/networks only exist for the highest use model organisms making contextual interpretations difficult for many researchers; 4) interpreting the huge datasets of metabolite concentrations under various conditions with biological context is an inherently complex problem requiring extremely in depth knowledge of metabolism. There is also one final problem, the issue of determining which metabolites are actually important in the experimental system in question. While there is no standardized method for this, there are many statistical tests and tools available to researchers to pick out statistically significant metabolites from noise [11]. The remaining issues have fortunately already been and will continue to be addressed to varying degrees as advances in technologies and method developments rapidly evolve. The first three problems will generally be solved/alleviated over time as advances in instrumentation and their combined use as well as the continued curation and community development of databases allows for more metabolites to be identified in a more contextual fashion. The final problem, which is the main topic of this review, will only be solved as our understanding of systems biology evolves and tools to tap this knowledge keep up. The

current generation of platforms, which are at the cutting edge of the field have generally been built upon the foundations laid by the large biochemical databases.

B.5. Bioinformatic Basis

With the advent of the genomic age, the amount of biochemical knowledge has exploded in the last two decades which has necessitated its storage in large databases. A variety of top-down (gene to protein to metabolite) and bottom-up (chemical entity to biological function) approaches have been taken resulting in a rich expanse of metabolic knowledge bases available to query. These databases provide the contextual biochemical basis for metabolomics data interpretation. By supplying information about metabolites, such as defining which enzymatic reactions consume or produce them, and which pathways they're involved in, researchers can use them to interpret their experiments to higher levels. An excellent review of these (top-down) types of databases is available in [31], while a more expanded review of databases is available in [32] and more recently in [33]. Also the Metabolomics Society website provides an excellent resource (www.metabolomicssociety.org/database). Additionally, and more specific to the development of metabolomics, mass spectral databases like the Golm Metabolome Database (GMD), which link mass spectrum and chromatographic retention time to specific compounds have been developed for use in the identification stages of metabolomics [34,35]. Some tools designed for higher level metabolomic analysis can take GC-MS spectra as input and so have integrated select databases into their platform. The human metabolome database (HMDB)[1] warrants mention here as while it is highly specific, it contains integrated information from spectra (multiple NMR, GC-MS) to clinical relevance. As a result it has been integrated into several platforms. By far the major database that has been integrated into metabolomics interpretation platforms is the Kyoto Encyclopedia of Genes and genomes(KEGG), which is divided into several sub-databases with

LIGAND, REACTION PAIR and PATHWAY being the most relevant to metabolomics [36]. These databases have been undergoing continuous updating and annotation for close to 20 years and so contain a great deal of valuable information. KEGG and MetaCyc are currently the largest (most number of organisms) and most in depth comprehensive (i.e. contains linked information from metabolite to gene) databases available, and so have been frequently integrated into interpretation platforms. The most commonly integrated databases have been summarized in Table A-B-1. This leaves other databases (further reviewed in [31]), such as Reactome [37] (human), KNApSAcK [38] (plants), Model SEED [39] (diverse), and BiG [40] (6 model organisms), somewhat overshadowed, though they do have their own tools for use in metabolomic analysis, and can be more useful than the large databases if a specific organism is desired. The KEGG and MetaCyc databases each contain a generalized ‘conserved’ set of pathways based on metabolic pathways that are more or less the same throughout life in general. For KEGG, organism specific annotations are available to query while for MetaCyc, individual ‘Cyc’ databases have been generated for a number of organisms, some just computationally, others extensively manually curated such as AraCyc for Arabidopsis [41]. A more recent development are the cheminformatic databases like PubChem [42] and ChEBI [43], which provide a chemically ontological approach to cataloguing the ill-defined category of ‘small molecules’ active in biological systems. These types of databases can provide additional non-biology specific information as well alternative formatting options for datasets. Finally, it is important to note that the few databases discussed here are by no-means exhaustive and that these databases are cross-referenced and linked to each other as well as against more widely known databases such as the well-known Chemical Abstract Service (CAS) [44] among many others.

Table A-B-1: Selected Biochemical Databases

	KEGG[36]	MetaCyc [45]	PubChem [42]	ChEBI[43]	GMD[34]	HMDB[1]
Link	www.genome.jp/kegg/	www.genome.jp/kegg/	pubchem.ncbi.nlm.nih.gov/	www.ebi.ac.uk/chebi/	gmd.mpimp-golm.mpg.de/	www.hmdb.ca/
Type	Comprehensive	Comprehensive	Chemical	Chemical	Mass Spectral	Mass Spectral Comprehensive
Database Features	Genomes, genes, proteins, metabolites, drugs, diseases, pathways, visualizations	Genes, proteins, metabolites, pathways, interactive visualizations	Compound	Compound	Metabolites	Metabolites
Specificity	Generalized annotations, 2260 organism semi-specific annotations	Generalized annotations, 1939 organism specific annotations	Broad	Broad	Broad, plant heavy	Human

B.6. Metabolomics Secondary Analysis: Enrichment Analysis and Metabolite Mapping

Biochemical databases provide an excellent backdrop of information for metabolomic analysis tools to query. Like many techniques in metabolomics, the algorithms for using these databases for interpretation evolved from methods developed for transcriptomic and proteomic analysis, such as Gene Set Enrichment Analysis (GSEA) [46]. This landmark technique has been the clear influence for several recent metabolomic tools, namely PAPI [24], MBRole [27], MSEA [30,47] (as implemented by two different groups) and MPEA [29]. While each tool is unique in its algorithm, the general idea of enrichment analysis is used by all. Enrichment analysis depends on meta-data being associated with metabolites as biochemical entities. As such they can be annotated with various classifiers such as chemical family or which metabolic pathways it is involved in. Enrichment analysis can then take a list of metabolites, and with some tools their relative abundances (including positive/negative changes), and calculate based on some metric whether any particular pathway(s) (or some other classifier such as chemical family) is (statistically) more

represented than any other, based on all possibilities. The assumption is then that this particular pathway is being more perturbed by the experimental condition than others, hence the observed significance and alterations to concentrations in the input metabolites. This method of secondary analysis has evolved alongside the complementary technique of metabolite mapping of which available non-specific network visualization tools have been reviewed in [48,49,50]. These generic network tools allow for integration of multiple ‘omics datasets, as well as more user controlled flexibility. Metabolomics specific network mapping tools also exist, some of which are components of databases such as the KEGG pathway databases (KEGG Atlas) and MetaCyc’s Pathway Tools [51,52]. Other explicitly designed tools are also available, some of which have been summarized in Table A-B-2. CytoScape [53] is a highly used/integrated stand-alone networking program for ‘omics datasets which even has plugins like MetScape[54] designed for viewing human metabolic data. The principal idea behind pathway mapping is the contextual visualization of metabolomics data. On these networks, nodes represent metabolites and edges (connecting lines) represent enzymatic conversions. By highlighting the significantly changed metabolites (with or without magnitudes) on organism specific (if available) or life-general metabolic pathways a researcher is provided with an interpretable visualized representation of their data. Biological inferences can then be made by manually inspecting these figures, while some platforms provide network topology analysis tools. The complex subject of computational representation and analysis of metabolic networks has been reviewed in [55]. Between visualization and enrichment analysis secondary analysis is becoming an important step in biological interpretation of metabolomics experiments.

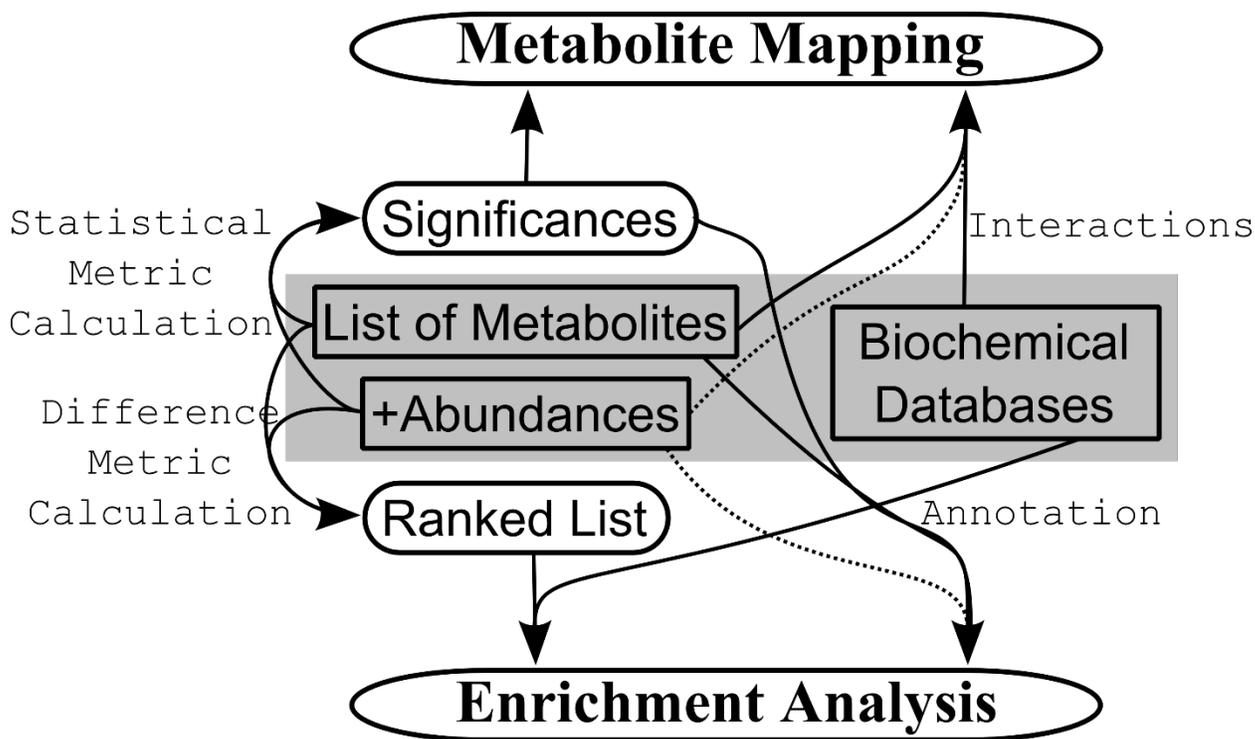


Figure A-B-1: Flow chart showing possibilities of metabolomics secondary analysis. Beginning with a list of metabolites, and also in some cases associated with relative abundances or a comparable metric the data can be analyzed two different ways which may include intermediary steps. Also needed is a biochemical database to be used in annotating the biological (and/or chemical function) of the listed metabolites. The list of metabolites plus abundances can undergo statistical analysis in order to pre-screen for metabolites having significant differences between sample classes, or the abundances can be used to calculate how differently they're expressed between each sample class (difference metric) which is then used to rank the list from most different to most similar. Significances can be used for both metabolite mapping and enrichment analysis. Metabolite mapping is the visual attribution of specified metabolites within known, pre-defined metabolic pathways which can include further information like significance and abundances as node attributes such as size and colour. Enrichment analysis is a statistical calculation that uses biological annotation to attempt to discern out the input metabolites which higher level functional properties (pathways) are being affected. This can take the form of searching for particular annotations at the top/bottom of the ranked list or examining whether a particular set is over-represented in the significant list.

Metabolomics secondary analysis tools have been developed by a number of groups with diverse implementations, however there are many commonalities. One of the major benefits of many of these tools is their implementations of user-friendly GUIs, allowing greater accessibility and precluding the necessity of learning the complicated tools they're based on, most prevalently,

R (The R Project for Statistical Computing, www.R-project.org). Before continuing, enrichment analysis will be used as a synonym for over-representation analysis, which some tools prefer to use. For enrichment analysis two objects are needed, a (ranked) list of items (i.e. genes or metabolites) provided by the experiment and a background set of annotations, derived from biochemical databases, computationally, through manual curation or some combination thereof. Before continuing, enrichment analysis will be used as a synonym for over-representation analysis, which some tools prefer to use. For enrichment analysis two objects are needed, a (ranked) list of items (i.e. genes or metabolites) provided by the experiment and a background set of annotations, derived from biochemical databases, computationally, through manual curation or some combination thereof. The list of metabolites can be ranked based on some metric indicating how different the abundances of each metabolite is between two sample classes, which can be calculated a number of different ways. As such the list will then show the metabolites with the most different abundances at the top of the list and the most similar at the bottom. The background set should contain all known metabolic pathways in an organism, each pathway including all the involved metabolites. For example, the ‘TCA cycle’ contains the metabolites succinate, oxaloacetate, isocitrate etc. Compounds can occur multiple times as they are parts of many pathways, such as oxaloacetate which also appears in glyoxylate metabolism, among other pathways.

A danger with KEGG is that it includes pathways such as ‘metabolic pathways’ and ‘microbial metabolism in diverse environments’ which contain huge numbers of metabolites, and so as such are relatively meaningless when found as enrichment analysis hits. ‘Aminoacyl-tRNA biosynthesis’ is also a common hit to be taken with a grain of salt, as it is often highlighted when several amino acids are identified as significant. Careful scrutinization of metabolic pathways to

ensure that they are logical is an important step in analyzing results produced by any platform. Other problems can arise when a dataset contains inordinate representation of certain pathways (either very few or very many). When many metabolites from one pathway are found in a dataset this pathway may be found to be significant mainly due to the large number of metabolites. Also the converse can happen if only a few metabolites are present in the dataset, but they changed significantly between classes, the pathway may not be found to be important due to the low number of metabolites [56]. Another issue occurs when querying the general (non-species specific) KEGG database as pathways that are non-existent in the experimental organism arise as significant. Sometimes this is obvious as with ‘synthesis of plant secondary metabolites’, though other times it may be difficult to know especially since well-curated metabolism databases exist for a scant few organisms. This is problematic for checking not only whether the pathway exists, but whether the annotation is accurate. Finally, even with well-annotated organisms there will metabolites identified that have not been assigned to any reaction whereas in poorly annotated organisms metabolites may be in metabolic pathways differently than expected from the canonical databases. Thus it is of the utmost importance for researchers to carefully regard the results produced by any secondary analysis tool and to understand how each piece of software works to ensure that the biological interpretation of the data is not skewed by some computational artifact. Cross-validating results through the use of multiple tools or multiple users producing the same result with a given platform is time consuming but would buttress the confidence in a result. Ultimately the best form of validation is a follow-up experiment however finding support from the literature for a result will also boost confidence.

These issues, and KEGG’s issues are somewhat alleviated by the BioCyc[45] series of organism specific databases, which if one has not already been generated, researchers with fully

sequenced organisms can automatically produce such a customized database using the powerful Pathway Tools software [52]. This tool takes a sequenced, annotated genome and determines which metabolic reactions exist by comparing against the MetaCyc database of ‘all’ known metabolic reactions. A rudimentary metabolic network is then generated which must be manually curated using actual experimental knowledge to ensure that the computational model is actually accurate. These models have shown useful to many researchers, however their use is less prevalent among enrichment analysis tools.

B.7. Overviews of Metabolomics Secondary Analysis Tools

As with the rest of the metabolomics field, the sub-field of secondary analysis is rapidly evolving. Many tools for metabolite mapping and enrichment have been recently developed and are available for use. Generally, these tools can be divided into two categories: enrichment analysis and metabolite mapping. Enrichment analysis aims to provide higher level information about metabolism from a list of metabolite abundances in different sample classes. Metabolite mapping provides a visual representation of metabolomic data by showing the identified metabolites (and their abundances) on a network graph, often obtained from a biochemical database. Some tools provide other functions as well, or can perform both simultaneously. Additionally, the option to integrate other ‘omics data is becoming more prevalent. MetaboAnalyst and MeltDB are two platforms that warrant special mention as they provide a comprehensive environment to analyze metabolomics data from raw spectra all the way to secondary analysis. Finally it should be noted that the platforms discussed here are by no means an exhaustive list, merely a representative set of the most used and promising tools at this time.

Table A-B-2: Select Platforms for Metabolomics Analysis and Interpretation

Name	Link	Access	Input	Databases Used	Functions	Comments
MetExplore [28]	http://metexplore.toulouse.inra.fr	Web-based	Compound IDs, Mass IDs	Generally BioCyc related	Compound mapping, graph analysis of metabolism maps.	Choice of organism database, filtering options, multiple graph analysis tools, Cytoscape integration.
PAPi [24]	http://www.4shared.com/file/0v5zSobM/PAPi_10.html	R Package	KEGG Compound IDs	KEGG	Compares activity of metabolic pathways between sample types.	Non organism specific, more difficult/powerful command line R interface. Usable with spent media results.
MBRole [27]	http://csbg.cnb.csic.es/mbrole/	Web-based	Compound IDs	KEGG, HMDB, PubChem, ChEBI, SMILES	Enrichment analysis of metabolites' annotations.	Background set from known organisms or custom set. Metabolite ID converter.
MetaboAnalyst [57]	http://www.metaboanalyst.ca/MetaboAnalyst/	Web-based.	Raw Spectra (GC and LC MS) , peak lists and spectral bins (MS and NMR)	Custom, KEGG, HMDB	Full processing, Statistical Analysis	Comprehensive metabolomics analysis platform with easy interface, tutorials, help. Human focused though some model organisms or custom metabolite set option.
MetaboAnalyst (MSEA) [58]	http://www.metaboanalyst.ca/MetaboAnalyst/	Web-based.	Compound IDs and abundances	Custom, KEGG, HMDB	Enrichment Analysis	Comprehensive metabolomics analysis platform with easy interface, tutorials, help. Human focused

	boAnalyst/					though some model organisms or custom metabolite set option.
MetaboAnalyst (MetPa [59])	http://www.metaboanalyst.ca/MetaboAnalyst/	Web-based	Compound IDs and abundances	KEGG	Pathway Analysis	Select model organisms. Network topology analysis. Intuitive network visualization.
MPEA [29]	http://ekhidna.biocenter.helsinki.fi/poxo/mpea/	Web-based	Compound IDs, GC-MS Spectrum as ranked list	KEGG, GMD, SMPDB	Pathway enrichment analysis.	Optional background set. Limited to top-down/bottom-up analysis.
MeltDB (MSEA) [30,60]	http://www.cebit.ec.uni-bielefeld.de/groups/brf/software/meltdb_info/	Web-based, login required	Raw GC/LC-MS spectra, processed spectra, compound IDs and abundances	GMD, KEGG, ChEBI, CAS	Comprehensive preprocessing, statistical analysis and metabolite mapping, enrichment analysis.	Integrated comprehensive online system, accessible by multiple users. Many statistical tools, custom metrics and sets for enrichment analysis.
Meta P-server [61]	http://metabolomics.helmholtz-muenchen.de/metap2/	Web-based	Compound IDs, sample meta-data	KEGG, HMDB, LipidMaps, PubChem	Data quality control, statistical analysis, hypothesis testing.	No use of organismal databases. Focus mainly on global statistical analysis.

MassTriX [62,63]	http://metabolomics.helmholtz-muenchen.de/masstrix2/	Web-based	MS spectra	KEGG, HMDB, LipidMaps	Compound mapping	Choice of KEGG organism. Optional background set. Color-coding.
BioCyc (Pathway Tools) [52]	http://biocyc.org/	Installation required	Annotated genome, 'omics data	MetaCyc	Network exploration, genome annotation, 'omics data painting.	Comprehensive systems biology network analysis.
Pathos [64]	http://motif.gla.ac.uk/Pathos/index.html	Web-based	Simple m/z values, Compound IDs	KEGG	Compound mapping	Choice of limited organism databases.
PaintOmics [65]	http://www.paintomics.org	Web-based	KEGG formatted metabolites and/or genes	KEGG	Compound mapping	Choice of 100 hundred top species. Colours pathway metabolites and genes according to increase/decrease.
IMPala [66]	http://impala.molgen.mpg.de/	Web-based	Gene IDs and/or Compound IDs	KEGG, HMDB, CAS, ChEBI, PubChem, Reactome, Wikipathways	Enrichment Analysis	Combined analysis with proteins or transcripts. Organism independent. Optional background set.
MetaMapp [25]	http://uranus.fiehnlab.ucdavis.edu:	Web-based	Compound IDs	KEGG	Metabolite networking	Organism independent. Network construction based on chemical similarity.

	8080/ Meta Mapp /hom ePage					
VANTED [67]	http://vanted.ipk-gatersleben.de/	Installation required	Compound abundances	KEGG	Metabolite networking, compound mapping, statistical analysis	Combined analysis with proteins and transcripts. Organism independent. Direct visualization of results on networks. Time course analysis. Statistical analysis.
TICL [68]	http://mips.helmholtz-muenchen.de/projects/home.html	Web-based	Compound IDs	KEGG	Enrichment analysis.	No choice of organism. Currently non-functional

B.8. Comprehensive Platforms

B.8.1. *MetaboAnalyst* [57].

MetaboAnalyst provides a suite of utilities allowing comprehensive analysis from raw spectral data to pathway analysis within one platform. Also included are tutorials and example datasets that can easily be loaded to practice analysis. Five main choices are available: statistical, enrichment and pathway analysis as well as time course analysis. A number of other utilities including data quality checking (useful for batch effects) and a metabolite ID converter among others are also included. If beginning from raw GC or LC-MS data MetaboAnalyst uses XCMS [69] for peak fitting, identification etc. Once at the peak list (NMR or MS) stage, various preprocessing options such as data-filtering and missing value estimation can be used. Next a

number of normalization, transformation and scaling operations can be performed. At this point the dataset is entered and can be subjected to MetaboAnalyst's entire suite of statistical analyses including metabolomics standards like PCA, PLS-DA and hierarchically clustered heatmaps, among many other options. While all these tools are useful and highly convenient, they can similarly be performed by many other platforms, albeit often with less accessibility making MetaboAnalyst a good option for those new to the field. It is the secondary analysis tools MSEA and MetPa (accessible as enrichment and pathway analysis) however which are of interest to this review.

The Enrichment Analysis tool of MetaboAnalyst was one of the earliest implementations of GSEA for metabolomics datasets. As it stands, it is quite biased towards human metabolism as except for the custom option, all the available background sets for enrichment analysis are of various mammalian derived human-centric sets including blood, urine and disease associated metabolite sets. It is however possible to provide a custom background set thereby allowing any organism to be studied. This implementation of MSEA provides three options for input: a single column list of compounds (Over Representation Analysis, ORA), a two column list of compounds AND abundances (Single Sample Profiling, SSP) and a multi-column table of compound abundances in classed samples (Quantitative Enrichment Analysis, QEA). Each option can provide different information. ORA will calculate whether a particular set of metabolites is statistically significantly higher in the input list than a random list, which can be used to examine ranked or threshold cut-off lists. SSP is aimed at determining whether any metabolites are above the normal range for common human biofluids. QEA is the most canonical and will determine which metabolite sets are enriched within the provided class labels, while providing a correlation value and p-value. MetaboAnalyst's MSEA has been used for a number of applications including aiding

in characterizing the metabolic basis of Fragile X syndrome [70], understanding how various environmental pollutants affect goldfish tissues differently [71], and in identifying metabolic changes that occur as mice age [72]. Generally the results provided by MSEA were used to contextualize the observed changes in individual metabolites.

The Pathway Analysis tool of MetaboAnalyst, MetPa, performs somewhat similarly to MSEA, however it performs pathway enrichment and network topology analysis. It also provides broader options for organism databases including 17 common model organisms such as *C. elegans*, *A. thaliana*, *E. coli*, *M. musculus* among others, as well as a custom option. Input options are the same as MSEA. Once the data is loaded, the background database selected, the test method (Fisher's Exact or Hypergeometric) and network topology metric must also be chosen. Output from MetPa is an interactive set of graphs. One graph plots the p-values vs pathway impact for the computed metabolic pathways. This graph allows one to discover the highest significantly impacted pathways for further exploration. Clicking data points on this graph will cause the network to be displayed in an adjacent view, with the input metabolites highlighted. Clicking these metabolites will show a box whiskers plot for each class allowing one to visualize increases and decreases. Through these functions MetPa can be used to visualize metabolomic data within known metabolic pathways, along with calculating which pathways are significantly affected. This ability makes it a powerful tool for secondary analysis. It has been used to understand the diabetes-contextual effects of a high-fructose diet in rats [73]. The results were used along with corroboration from TICL [68] to provide a starting point for analysis. MetPa has also been used to understand metabolomic results of renal injury in heart failure patients [74]. Finally it was used to identify perturbations to leucine and cysteine amino acid metabolism as well as energy metabolism

in Dupuytren's disease of fibroblasts from the palm of the hand [75]. This presence of a specific metabolic phenotype will aid in the pursuit of the cause of the disease.

B.8.2. MeltDB [30].

MeltDB is another comprehensive suite for metabolomics data analysis designed explicitly as a free and platform independent integrated project management and analytical pipeline that takes raw GC or LC-MS data through spectral preprocessing, data normalization, statistical analysis and recently integrated, enrichment analysis all within the same system. Their registration required web-based implementation allows multiple users (with various privileges such as view-only accounts) to work in parallel on the same projects. Also available are features for experiment meta-data description such as test conditions, extraction method and analytical parameters. MeltDB allows for data to be input at any level of processing, IE raw spectra can be imported and have their peaks detected, identified and quantified (with a choice from a variety of methods) or previously pre-processed data can be imported. This includes access to their implementation of MSEA [60]. Similar to MetaboAnalyst, MeltDB provides a wealth of options and utilities for analyzing MS metabolomics data including a large number of platforms (also accessible as stand-alone applications eg. XCMS [69]) for chromatographic MS processing. It is also able to import preprocessed data from common (non open-source) vendor-specific software. Further in the pipeline and similar to MetaboAnalyst, MeltDB can be used to perform statistical analysis like PCA and generate figures such as hierarchically clustered heatmaps which again is very useful for new comers. These features have been used in demonstrating how SEC is a superior sample collection method for *Corynebacterium glutanicum* and in the metabolic characterization of different parts of the grain during the highly important process of industrial barley malting [76].

MSEA takes a ranked list of compounds and determines whether a particular pathway is enriched towards the top or bottom of this list, however it provides the highly-convenient option of being able to natively rank the list based on a number of metrics. In their analysis used to test this new tool it was found that the use of a highly specific background set, CglCyc (from their sequenced *C. glutanicum*), which was automatically produced then manually curated produced better results than the use of the KEGG database. Two main reasons were provided. First, KEGG pathways are much larger and interconnected than their CglCyc pathways resulting in the obstruction of information when there are opposite fluxes in different parts of the pathway (the provided occurrence was in opposite abundances in the upper and lower regions of gluconeogenesis). Second, as previously noted, KEGG annotated pathways do not exist in all organisms. Further testing was performed on datasets obtained from a number of mutant lysine production strains, which generally showed the expected results of alterations to lysine and threonine metabolic pathways as their mutations targeted this split of branched chain amino acid metabolism. This new tool appears to be quite powerful, though also very reliant on the quality of the background set.

B.9. Enrichment Analysis

B.9.1. PAPI [24].

Pathway Activity Profiling is an R-based tool designed specifically for secondary analysis of metabolomic data. As input it takes a list with abundances (normalized and scaled) and working on the assumptions that the detection (IE presence in the list) of more metabolites in a pathway and that lower abundances of those metabolites indicates higher flux and therefore higher pathway activity PAPI calculates an activity score (AS) for each pathway. The metabolic pathways are

taken from the general KEGG database and the AS indicates the probability of this pathway being active in the cell. These scores can then be used to compare experimental and control conditions by performing ANOVA or a t-test to compare two sample types. As such, PAPI is a classic implementation of metabolomics secondary analysis, allowing users to derive higher level information from a simple list of metabolites. It has been used to show the similarity between genetic and environmental perturbation of yeast strains, which was in agreement with the previously published conclusions. It has also been used to show that sound caused frequency dependent metabolic alterations[77] and that different biological interpretations will be made in microbial metabolomics based on the extraction methodology [78]. While PAPI's assumptions may not be universally accurate (TCA cycle intermediates can have high abundance even when flux through the reactions in this pathway is also high) and the interface is more difficult than other platforms, it still provides an excellent option for enrichment analysis.

B.9.2. MBRole [27].

Metabolic Biological Role is another classic implementation of enrichment analysis. Taking as input a list of significantly changing metabolites (IE statistically processed already) MBRole calculates which pathways and chemical groups are enriched either against a pre-compiled (from KEGG) or user supplied background set. Output is a table of metabolic pathways with significance p-values and the pathways hyper-linked to KEGG metabolism maps. MBRole is an easy to use yet powerful tool as it can take input under many different database formats and compute the enrichment based on any of the available annotations. Also the use of any of the organism-specific KEGG annotations makes the investigation of diverse organisms easy. It has been used as a starting point to interpret steatotic liver tissue metabolomic data. Results were interpreted in the context of the identified enriched pathways that were altered in steatotic tissue with prior knowledge and

direct examination of the metabolite pools [79]. While the flaws of KEGG annotations remain present, MBRole provides an excellent simple implementation of enrichment analysis for the average user.

B.9.3. MPEA [29].

Metabolite Pathway Enrichment Analysis is a stand-alone tool that takes a ranked list of metabolites (either KEGG IDs or mass spectra with retention index) and determines if a particular known metabolic pathway (as annotated in the background set) tends to appear more towards the top or bottom of that list. Output is a table of metabolic pathways (linked to KEGG) with p-values (among other data) indicating whether the pathway was significantly enriched. The default settings are human-biased as KEGG and the SMPDB[80] (a curated set of human pathways) are queried however a custom background set option is also available. The list can be ranked by any metric, such as significance to a model or t-tests of concentration. One of the main differences between MPEA and other tools is the ability to work with ambiguously identified compounds, especially useful when working from mass spectra. Mass spectra are first identified using the GMD, then ambiguous identifications resolved within the pathway enrichment analysis. This tool has been used by the group that developed it to make a minor contribution in studying Alzheimer's progression, showing that pentose-phosphate pathway was altered in patients that were developing dementia [81].

B.9.4. TICL [68].

The Tool for automatic Interpretation of a Compound List is an early example of metabolomics secondary analysis. It was designed to take a list of (significantly) changed metabolites from an experiment and calculate whether they are biologically related, according to KEGG pathways. Taking a list of KEGG IDs as input, TICL outputs a list of pathways with p-

values indicating the probability of this pathway appearing by chance. A relatively underused tool, TICL has been used to demonstrate differences between the biofilm and planktonic response to metal stress[82], and to supplement/compare MetPa results in studying the effects of a high-fructose diet on rats[73]. An early pioneer in the field with a sound premise, at the time of writing TICL was not functional.

B.9.5. IMPaLA [66].

Integrated Molecular Pathway-Level Analysis is a tool designed to perform enrichment analysis on both metabolomic and proteomic or transcriptomic datasets simultaneously. Taking as input a list of metabolites plus a list of genes/proteins if available (not necessary) IMPaLA can calculate pathway enrichment using one of two methods. Enrichment is computed either against a user-provided background set or against the whole set chosen from the available input format databases (KEGG, HMDB, ChEBI etc). The input can either be preselected for significance by some other analysis or can include abundance information between two different classes. In either case, output is a table of pathways hyper-linked to the database it was found in along with a p-value indicating significance. For the purposes of combined analyses there is a p-value calculated separately for genes and metabolites as well as a combined value. This makes IMPaLA a good tool for analyzing combined datasets however the limited outlinking with pathways and lack of visualization means that there are potentially better options for just metabolomics enrichment analysis.

B.10. Metabolite Mapping

B.10.1. MetaMapp [25].

MetaMapp presents a novel approach to metabolic mapping which uses chemical similarity of compounds in order to overcome the difficulties of missing, unknown and unannotated

metabolites prevalent in metabolomics data. Development of this platform was due to a dissatisfaction with other available metabolic mapping tools, generally due to the above being more or less addressed, depending on the particular tool. Hence MetaMapp was developed on the premise that since biochemistry is the interconversion of chemically similar entities, compounds can be clustered solely by their chemical similarity. While this was found to be highly beneficial for metabolites without reaction annotation, chemical similarity mis-clustered some obviously biologically-related metabolites. As such MetaMapp uses both chemical similarity and KEGG reactant pair data. Finally, the problem of unknown compounds was addressed by adding the possibility to map metabolites based on their mass spectral similarity. While the resultant graphs are somewhat busy, especially when statistical information such as significance or fold-change is applied to node attributes such as size and colour, this novel approach can provide much needed contextual information about unannotated and unknown metabolites. The function of this tool was demonstrated using GC-MS metabolomic data from three tissues involved in fetal exposure to tobacco smoke: maternal plasma and lungs and fetal lungs. Using MetaMapp, an identical network of the 179 identified metabolites (excluding unknowns) was generated for each tissue, with various biologically and chemically related clusters clearly visible. For each graph, only significant metabolites were labeled, with color representing up or down regulation (compared to the unexposed control) and size representing fold-change. Aligning these three graphs allowed for a visual inspection of the metabolomic data which made interpretation pleasantly obvious. These results clearly showed that the fetal lungs were most affected, with fatty acids being the most dysregulated. Also present were alterations to several amino acids. These results show the promise of this novel technique in interpreting metabolomics experiments. One of the most exciting features, which was not involved in the confirmatory results is the ability to map unknown

metabolites. This possibility will likely be very useful in discovering novel metabolic pathways in the future.

B.10.2. MassTrix [62].

MassTrix is a platform for automatically identifying high precision spectra and mapping data in the context of organism-specific KEGG pathways. It is one of the oldest tools discussed and has been well-utilized. Developed by the same group as Meta P-server, the ability to integrate raw transcriptomic data was recently added [63]. This ability, plus the identification of compounds previously annotated within an organism (from KEGG) differentiates MassTrix from other platforms. The identification procedure is based on comparing the masses of input ions to known metabolites obtained from their multi-integrated database including options for adducts and isotopes though it may be by-passed by entering previously identified KEGG IDs. Once data has been uploaded and analyzed, two sets of results are provided. The Compound section shows all of the annotated compounds with mass, formula, identity and which database the ID was acquired from. This section can be examined for ambiguity issues and compounds are clickable to find their pathway annotations linked from KEGG. The Pathway section of results allows pathways of interest (those which include identified compounds) to be visualized with ID'd compounds highlighted as well as transcriptomic data applied. One drawback here is the inability to assign metabolite abundances. Indeed, MassTrix is somewhat limited compared to more modern tools, though the added integration of transcriptomic data has great potential. Additionally, MassTrix has an excellent track record of use for a wide variety of applications. It has been used to study the effects of dry-bean consumption on carcinogenesis in rats [83], to explore the wide dynamic range of the human metabolome in healthy individuals[84] and also quite interestingly to study the 'metabolome' of organic matter in sea-spray [85] among many other successful applications

B.10.3. PaintOmics [65].

PaintOmics is an ‘omics mapping web-tool that takes metabolite and transcript abundances and significances and maps them onto organism-specific KEGG maps. Taking either or both types of data, PaintOmics will produce a series of KEGG pathway maps with the data highlighted on the networks, as well as providing an enrichment analysis p-value for each pathway. It is capable of coloring objects (metabolites or transcripts) for each condition provided in the input. Given that PaintOmics will include any pathway with at least one entry the enrichment analysis or prior knowledge will be needed to assist in interpretation. While an improvement over the combined mapping abilities of MassTrix, PaintOmics suffers from the same drawbacks of other KEGG based utilities, namely quality of annotation and size of pathways, both of which are addressed by MetaMapp. Still the ability to seamlessly integrate both transcriptomic and metabolomic datasets, as well as display the results of multiple classes in one visualization make it a useful tool. Additionally as KEGG annotations continue to improve and include more species it will only increase in utility. Thusfar it has be used to map transcriptomic data from differently cultured hepatocellular carcinoma cells [86].

B.10.4. VANTED [67].

The tool for the Visualization and Analysis of Networks with related Experimental Data (VANTED) is another tool capable of mapping ‘omics data onto custom and KEGG derived networks with additional visualization and analysis options. Contrary to most other programs, VANTED must be installed on the user’s computer. Taking any combination of data (in the form of relative or absolute concentrations in different samples), it will present the data upon the relevant biological networks. This allows users to see the concentrations of metabolites in sample classes AND their connection to other metabolites and/or genes and/or proteins together. This works best

when many linked metabolites have been quantified, which is unfortunately often not the case in metabolomics experiments. Statistical tests indicating whether metabolite concentrations are significantly different from the control can also be automatically performed and their result appended to the visualization. VANTED provides numerous options for how the networks are generated including downloaded organism-specific KEGG maps as well as correlation-based mapping using various metrics. These statistically oriented maps, along with the convenient presentation of metabolite abundance data make VANTED a powerful tool for metabolomic secondary analysis. Since its release it has been updated frequently and has been highly used in the field. VANTED has been used to interpret metabolomics results in a wide variety of studies including the effects of drought response on wheat leaves [87], the effects of pyruvate for treatment of mitochondrial disease [88] and understanding how glucose starvation affects *Staphylococcus aureus* [89], among many others.

B.10.5. Pathos [64].

Pathos is a metabolite mapping tool designed in response to MassTriX's limitations. Specifically it was made to include the ability to map data from different experimental conditions and compare their degree of change. Apart from this difference, Pathos identification functions similarly to MassTriX taking mass/charge values (or previously identified compound IDs) and identifies them using an organism-specific KEGG database, then displays the KEGG pathways with the input metabolites highlighted. Different than MassTriX though, a p-value for each pathway is not provided. Output is a list of pathways with the number (out of the total) of identified metabolites which are clickable to show the mapped pathway. On the visualization identified metabolites can be clicked to show a column plot comparing the abundances under each condition. Generally, this

tool is relatively comparable to the many other metabolite mapping tools. It has been used in conjunction with Ingenuity Pathway Analysis to monitor stem cells in regenerative medicine [90].

B.10.6. ProMeTra [91].

ProMeTra is an ‘omics viewing web-tool designed to visualize any kind of ‘omics data not only on KEGG database derived metabolic pathways but also on user supplied pathways. Its visualization system was designed to take advantage of the Scalable Vector Graphics (SVG) format allowing easy coloring (eg by abundance differences), extra annotation and even the production of animations. These features allow for the easy generation of clear, visually appealing multi-class annotated pathway maps for use in biological interpretation. Regulons can also be visualized, which when annotated with transcriptomic can clearly show biological effects. ProMeTra’s main draw compared to other mapping tools is the use of SVG graphics which allow for infinite zooming, output at any resolution and easy manipulation in SVG capable drawing programs. Even so it is an underused tool, perhaps due to the login-based (but not required) system or the less intuitive UI.

B.11. Other

B.11.1. MetExplore [28].

MetExplore is a metabolism exploration suite which can analyze metabolic networks without metabolomic data, though it also has a tool which will identify all the pathways individual metabolites can be involved in. This implementation was designed to overcome MetaCyc’s shortcomings of mapping compounds iteratively onto each relevant pathway, instead MetExplore aims for one single representation of each metabolite. Using the MetaCyc/BioCyc series of

databases there is a relatively wide choice of organism databases. MetExplore's main tools are Metabolome Mapping and a series of computational analysis tools. These tools do not involve data input, they just provide a variety of methods to analyze MetaCyc derived metabolic networks. Choke point analysis can identify reactions/metabolites that are unique within the network whereas scope and precursor analysis allow the investigation of what metabolites are required/are possible to produce the other metabolites in the network. Such analyses can be used to identify a minimal set of media or whether a particular metabolite can be generated given a defined media. These tools have been used to work on understanding the symbiotic relationship between Buchnera and its aphid host [92]. For all types of analysis filters are available to restrict artifacts and adjust the analysis. MetExplore's metabolome mapping tool is somewhat more limited than other comparably named tools. It can take as input a list of masses or identified metabolites, but does not output a visualization. Instead it provides a table view which indicates for each metabolites which metabolic pathways they are involved in and also topological information IE the number of reactions that produce/consume it, ranging from none to many for each direction. While this has its uses, the same information can generally be obtained from other mapping tools, however the computational analysis tools provided by MetExplore could be quite useful to researchers working on organisms that have a -Cyc database.

B.11.2. Meta P-server [61].

Meta P-server is a metabolomics exploration tool specifically designed to work with multi-class experiments. Taking as input a metabolite quantitation table and a sample description matrix, a number of statistical tests are automatically performed which can then be viewed and colored according to any identifiers in the sample description matrix. This allows the quick and easy checking for batch effects, outlying samples and also overall data quality. The two main statistical

outputs are PCA plots and hypothesis testing. The generated PCA plots can be colored by each possible class identifier allowing the most important classifier to be quickly found. Hypothesis testing of whether metabolite concentrations are different is performed for each possible class division, generating a series of boxplots with significant differences highlighted. For example if the data are classified by sampling day (Day 2, 3 and 5) and by drug dosage (none, high) two sets of hypothesis test results will be produced, one showing whether there are differences between concentration between each day for a given dosage, and the other showing differences between each dosage for a given day. This is performed for each metabolite. The other statistical result generated is a heatmap of correlations for any given numerical classifiers. Also included an option for direct import of Biocrates AbsoluteIDQ [5] kit derived data. While not providing any secondary analysis options, Meta P-server provides a quick and easy method for statistical analysis of multi-class experiments.

B.12. Commercial Software

All of the above software are completely free to use, with most not even requiring registration of an account. There are also a number of commercial pieces of software available. These software are generally designed to be comprehensive solutions for use in multi-omics experiments including a number of integrated pathway analysis and contextual visualization choices. Additionally, these software are more geared towards human/mammalian model (i.e. mouse, rat) disease and drug investigations and use manually curated proprietary databases.

B.12.1. IPA.

Ingenuity Systems Inc. (Redwood, CA), offers a data analysis suite deemed Ingenuity Pathway Analysis (www.ingenuity.com). Using their Ingenuity Knowledge Base, metabolomic data (among other 'omics data types) can be mapped onto networks and enrichment analysis can

also be performed. These features are among a large suite of systems biology analysis tools that are designed to allow biologically contextual representation of data. These data can come from a large variety of different types of experiments ranging from small-scale drug target experiments to combined transcriptomic and metabolomic studies. While many types of data can be used, IPA can take just a list of altered metabolites as input and use literature-characterized signaling and metabolic pathways to identify the biologically relevant effects of an experiment. This approach has been used to aid in the understanding of diverse metabolomics experiments, including colorectal cancer [93] and detoxifying processes in traditional Chinese medicine [94].

B.12.2. MetaCore.

GeneGo Inc. (Carlsbad, CA), of Thompson Reuters, along with MetaMiner and MetaDrug is a series of pathway analysis and data mining tools highly geared towards human disease investigations. Among these tools are features for mapping multi-omics experiments, drug target prediction, and pathway perturbation analysis for toxicity studies. These tools all work against their proprietary manually curated databases, which for some features extend into common model and pathology-relevant organisms. Metacore has been used to perform over representation and network analysis on datasets combining metabolomic and either transcriptomic or proteomic data to understand toxin mode of actions [95] and biomarkers in colorectal cancer [96], respectively.

B.12.3. GeneSpring.

Agilent Technologies (Santa Clara, CA) provides a comparable suite while also providing features for MS analysis of raw data by integrating with their Mass Profiler Professional, however this program has mostly been used for transcriptomic analysis.

B.13. Summary and Outlook

Metabolomics represents the apical step in the paradigm of systems biology. Its rapid development has provided unparalleled understanding of metabolic processes to a plethora of different fields. The secondary analysis of metabolomics data is a recent addition which will provide researchers with much more power in finding biological interpretation. Currently though, few researchers tend to be using enrichment tools and the provided results are rarely heavily discussed in their manuscripts. While the results from this type of analysis should not be the sole source of information for biological interpretation of metabolomics experiment, the provided results are highly useful in giving metabolic context for many metabolites at once, without having to search through databases one at a time. Still, a number of researchers have shown they provide an excellent springboard for diving into the depths of metabolomics data.

As the field of metabolomics secondary analysis evolves, a number of challenges remain. Beyond the ever-changing processes of data pre-processing and statistical analysis for metabolite significance, the contextual interpretation of metabolomics results will also need improving. While the software described in this review are a good beginning, future analyses will need to be highly tailored towards organism-specific metabolic reconstructions. The MetaCyc derived series of databases have begun to fill this role, however as genome annotation remains a marginally accurate process, the models generated from such data are affected equally. Community driven confirmation and elucidation of genetic and metabolic annotation of databases like EcoCyc have shown that it is only a matter of time and effort for reasonable computational models to be built. For other organisms, comparable tools will be developed as the research community deems them important. Another major improvement to be made would be decreasing the number of inaccurate and meaningless pathway hits made in enrichment analysis. This may be difficult though as

metabolites will only become annotated into more metabolic pathways as their connections are elucidated. Great steps have been made in metabolite mapping techniques, especially with the chemical similarity connections provided by MetaMapp, however the interpretability of all maps remains difficult. Computational improvements in the graphical presentation and ease of producing legible maps will make metabolite mapping better for metabolomic secondary analysis.

Achieving good biological interpretations of metabolomics data is easier in medical studies of humans and generally when using highly studied model organisms due to the preponderance and quality of databases associated with these subjects. While it is generally true in science that studying organisms with a wealth of literature makes data interpretation easier, this effect is greatly amplified in metabolomics. Thus *Homo sapiens*, *Escherichia coli*, and *Arabidopsis thaliana* are much more convenient to perform metabolomics secondary analysis upon compared to, for example, a freshly isolated environmental bacterial strain. Again MetaCyc and its associated tools present the solution though producing an organism-specific –Cyc database from an annotated genome remains a complex time consuming endeavour. Fortunately, the documentation continues to improve and SRI International is well-engaged with the research community for educational and software feedback purposes.

Regarding the current set of available tools, they all have strengths and weaknesses and it should not come to the use of one over the other. Once a metabolomics dataset has reached the point of secondary analysis, applying any of the above tools is not a hugely time-consuming process and so it may be wise to use multiple tools and take the consensus results. Definitely the use of at least one enrichment analysis and one visualization/mapping tool is recommended. Given the complexity of metabolomic data, it is also important to carefully regard the results from secondary analysis as it is possible for enrichment analysis to produce significant pathway hits

from only one or two metabolites in a pathway. As such, careful scrutinization and logical biological interpretation of the data must be undertaken. With this in mind metabolomics researchers should strive to integrate secondary analysis into their studies as these highly useful results can be obtained very rapidly. Clearly the field of secondary analysis is coming into its own and its continuing development will only serve to improve the success of the metabolomics approach.

B.14. References

1. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, et al. (2007) HMDB: The human metabolome database. *Nucleic Acids Research* 35: D521-D526.
2. Crick F (1970) Central dogma of molecular biology. *Nature* 227: 561-563.
3. Fiehn O (2002) Metabolomics - The link between genotypes and phenotypes. *Plant Molecular Biology* 48: 155-171.
4. Dunn WB, Bailey NJC, Johnson HE (2005) Measuring the metabolome: Current analytical technologies. *Analyst* 130: 606-625.
5. Bogumil R, Rohring C, Prehn C, Romisch-Margl W (2009) A High-Throughput Method for Targeted Metabolomics Analysis of Different Tissue Samples using the AbsoluteIDQ™ Kit.
6. Álvarez-Sánchez B, Priego-Capote F, Castro MDLd (2010) Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC - Trends in Analytical Chemistry* 29: 120-127.
7. Álvarez-Sánchez B, Priego-Capote F, Luque de Castro MD (2010) Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *TrAC - Trends in Analytical Chemistry* 29: 111-119.
8. Zhang A, Sun H, Wang P, Han Y, Wang X (2012) Modern analytical techniques in metabolomics analysis. *The Analyst* 137: 293.
9. Hendriks MMWB, Eeuwijk FAv, Jellema RH, Westerhuis JA, Reijmers TH, et al. (2011) Data-processing strategies for metabolomics studies. *TrAC Trends in Analytical Chemistry* 30: 1685-1698.
10. Wishart DS (2010) Computational approaches to metabolomics. *Methods in molecular biology* (Clifton, NJ) 593: 283-313.

11. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD (2009) Analytical and statistical approaches to metabolomics research. *Journal of Separation Science* 32: 2183-2199.
12. Weiss RH, Kim K (2012) Metabolomics in the study of kidney diseases. *Nature Reviews Nephrology* 8: 22-33.
13. Grainger DJ, Mosedale DE (2012) Metabolomics in coronary heart disease. *Heart and Metabolism*: 8-12.
14. Hassan-Smith G, Wallace GR, Douglas MR, Sinclair AJ (2012) The role of metabolomics in neurological disease. *Journal of Neuroimmunology* 248: 48-52.
15. Spratlin JL, Serkova NJ, Eckhardt SG (2009) Clinical applications of metabolomics in oncology: A review. *Clinical Cancer Research* 15: 431-440.
16. Ma Y, Zhang P, Yang Y, Wang F, Qin H (2012) Metabolomics in the fields of oncology: A review of recent research. *Molecular Biology Reports* 39: 7505-7511.
17. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* 37: 1-13.
18. Dunn WB (2008) Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical Biology* 5: 011001.
19. Want E, Masson P (2011) Processing and analysis of GC/LC-MS-based metabolomics data. *Methods in molecular biology (Clifton, NJ)* 708: 277-298.
20. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* 13: 263-269.
21. Viant MR (2008) Recent developments in environmental metabolomics. *Molecular BioSystems* 4: 980-986.

22. Bundy JG, Davey MP, Viant MR (2009) Environmental metabolomics: A critical review and future perspectives. *Metabolomics* 5: 3-21.
23. Beger RD, Colatsky T (2012) Metabolomics data and the biomarker qualification process. *Metabolomics* 8: 2-7.
24. Aggio RBM, Ruggiero K, Villas-Boas SG (2010) Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. *Bioinformatics* 26: 2969-2976.
25. Barupal DK, Haldiya PK, Wohlgemuth G, Kind T, Kothari SL, et al. (2012) MetaMapp: Mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 13: 99.
26. Çakir T, Patil KR, Önsan ZI, Ülgen KÖ, Kirdar B, et al. (2006) Integration of metabolome data with metabolic networks reveals reporter reactions. *Molecular Systems Biology* 2: 50.
27. Chagoyen M, Pazos F (2011) MBRole: Enrichment analysis of metabolomic data. *Bioinformatics* 27: 730-731.
28. Cottret L, Wildridge D, Vinson F, Barrett MP, Charles H, et al. (2010) MetExplore: A web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Research* 38: W132-W137.
29. Kankainen M, Gopalacharyulu P, Holm L, Orešič M (2011) MPEA-metabolite pathway enrichment analysis. *Bioinformatics* 27: 1878-1879.
30. Persicke M, Rückert C, Plassmeier J, Stutz LJ, Kessler N, et al. (2012) MSEA: Metabolite set enrichment analysis in the MeltDB metabolomics software platform: Metabolic profiling of *Corynebacterium glutamicum* as an example. *Metabolomics* 8: 310-322.
31. Karp PD, Caspi R (2011) A survey of metabolic databases emphasizing the MetaCyc family. *Archives of Toxicology*: 1-19.

32. Go EP (2010) Database resources in metabolomics: An overview. *Journal of Neuroimmune Pharmacology* 5: 18-30.
33. Fiehn O, Barupal DK, Kind T (2011) Extending biochemical databases by metabolomic surveys. *Journal of Biological Chemistry* 286: 23637-23643.
34. Hummel J, Selbig J, Walther D, Kopka J (2007) The golm metabolome database: A database for GC-MS based metabolite profiling. pp. 75-95.
35. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, et al. (2009) FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry* 81: 10038-10048.
36. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27-30.
37. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, et al. (2011) Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Research* 39: D691-D697.
38. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, et al. (2012) KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53.
39. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* 33: 5691-5702.
40. Schellenberger J, Park JO, Conrad TM, Palsson BT (2010) BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11.
41. Mueller LA, Zhang P, Rhee SY (2003) AraCyc: A biochemical pathway database for Arabidopsis. *Plant Physiology* 132: 453-460.

42. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, et al. (2009) PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research* 37: W623-W633.
43. Degtyarenko K, De matos P, Ennis M, Hastings J, Zbinden M, et al. (2008) ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Research* 36: D344-D350.
44. Whitley KM (2002) Analysis of SciFinder scholar and web of science citation searches. *Journal of the American Society for Information Science and Technology* 53: 1210-1215.
45. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* 38: D473-D479.
46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15545-15550.
47. Xia J, Wishart DS (2010) MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research* 38: W71-W77.
48. Merico D, Gfeller D, Bader GD (2009) How to visually interpret biological data using networks. *Nature Biotechnology* 27: 921-924.
49. Suderman M, Hallett M (2007) Tools for visually exploring biological networks. *Bioinformatics* 23: 2651-2659.
50. Pavlopoulos GA, Wegener AL, Schneider R (2008) A survey of visualization tools for biological network analysis. *BioData Mining* 1: 12.
51. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 40: D109-D114.

52. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, et al. (2010) Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* 11: 40-79.
53. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27: 431-432.
54. Gao J, Tarcea VG, Karnovsky A, Mirel BR, Weymouth TE, et al. (2010) Metscape: A Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. *Bioinformatics* 26: 971-973.
55. Lacroix V, Cottret L, Thébault P, Sagot MF (2008) An introduction to metabolic networks and their structural analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 5: 594-617.
56. Nam H, Lee J, Lee D (2009) Computational identification of altered metabolism using gene expression and metabolic pathways. *Biotechnology and Bioengineering* 103: 835-843.
57. Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* 37: W652-W660.
58. Xia J, Wishart DS (2011) Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nature Protocols* 6: 743-760.
59. Xia J, Wishart DS (2010) MetPA: A web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics* 26: 2342-2344.
60. Neuweger H, Albaum SP, Dondrup M, Persicke M, Watt T, et al. (2008) MeltDB: A software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics* 24: 2726-2732.

61. Suhre K, Kastenmüller G, Römisch-Margl W, Wägele B, Altmaier E (2011) Meta P-server: A web-based metabolomics data analysis tool. *Journal of Biomedicine and Biotechnology* 2011.
62. Suhre K, Schmitt-Kopplin P (2008) MassTRIX: mass translator into pathways. *Nucleic Acids Research* 36: W481-484.
63. Wägele B, Witting M, Schmitt-Kopplin P, Suhre K (2012) Masstrix reloaded: Combined analysis and visualization of tran-scriptome and metabolome data. *PLoS ONE* 7.
64. Leader DP, Burgess K, Creek D, Barrett MP (2011) Pathos: A web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. *Rapid Communications in Mass Spectrometry* 25: 3422-3426.
65. García-Alcalde F, García-López F, Dopazo J, Conesa A (2011) Paintomics: A web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics* 27: 137-139.
66. Kamburov A, Cavill R, Ebbels TMD, Herwig R, Keun HC (2011) Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics* 27: 2917-2918.
67. Junker BH, Klukas C, Schreiber F (2006) Vanted: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7.
68. Antonov AV, Dietmann S, Wong P, Mewes HW (2009) TICL - a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. *Febs Journal* 276: 2084-2094.
69. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* 78: 779-787.

70. Davidovic L, Navratil V, Bonaccorso CM, Catania MV, Bardoni B, et al. (2011) A metabolomic and systems biology perspective on the brain of the Fragile X syndrome mouse model. *Genome Research* 21: 2190-2202.
71. Jordan J, Zare A, Jackson LJ, Habibi HR, Weljie AM (2012) Environmental contaminant mixtures at ambient concentrations invoke a metabolic stress response in goldfish not predicted from exposure to individual compounds alone. *Journal of Proteome Research* 11: 1133-1143.
72. Houtkooper RH, Argmann C, Houten SM, Cantó C, Jenning EH, et al. (2011) The metabolic footprint of aging in mice. *Scientific Reports* 1.
73. Lin S, Yang Z, Liu H, Tang L, Cai Z (2011) Beyond glucose: Metabolic shifts in responses to the effects of the oral glucose tolerance test and the high-fructose diet in rats. *Molecular BioSystems* 7: 1537-1548.
74. Diercks DB, Owen K, Tolstikov V, Sutter M (2012) Urinary metabolomic analysis for the identification of renal injury in patients with acute heart failure. *Academic Emergency Medicine* 19: 18-23.
75. Rehman S, Xu Y, Dunn WB, Day PJR, Westerhoff HV, et al. (2012) Dupuytren's disease metabolite analyses reveals alterations following initial short-term fibroblast culturing. *Molecular BioSystems* 8: 2274-2288.
76. Persicke M, Plassmeier J, Neuweger H, Rückert C, Pühler A, et al. (2011) Size exclusion chromatography-An improved method to harvest *Corynebacterium glutamicum* cells for the analysis of cytosolic metabolites. *Journal of Biotechnology* 154: 171-178.
77. Aggio RBM, Obolonkin V, Villas-Bôas SG (2012) Sonic vibration affects the metabolism of yeast cells growing in liquid culture: A metabolomic study. *Metabolomics* 8: 670-678.

78. Duportet X, Aggio RBM, Carneiro S, Villas-Bôas SG (2011) The biological interpretation of metabolomic data can be misled by the extraction method used. *Metabolomics*: 1-12.
79. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A (2011) A comprehensive untargeted metabolomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. *Journal of Proteome Research* 10: 4825-4834.
80. Frolkis A, Knox C, Lim E, Jewison T, Law V, et al. (2010) SMPDB: The small molecule pathway database. *Nucleic Acids Research* 38: D480-D487.
81. Orešič M, Hyötyläinen T, Herukka SK, Sysi-Aho M, Mattila I, et al. (2011) Metabolome in progression to Alzheimer's disease. *Translational Psychiatry* 1.
82. Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, et al. (2011) Differences in Metabolism between the Biofilm and Planktonic Response to Metal Stress. *Journal of Proteome Research* 10: 3190-3199.
83. Mensack MM, McGinley JN, Thompson HJ (2012) Metabolomic analysis of the effects of edible dry beans (*Phaseolus vulgaris* L.) on tissue lipid metabolism and carcinogenesis in rats. *British Journal of Nutrition* 108: S155-S165.
84. Krug S, Kastenmüller G, Stückler F, Rist MJ, Skurk T, et al. (2012) The dynamic range of the human metabolome revealed by challenges. *FASEB Journal* 26: 2607-2619.
85. Schmitt-Kopplin P, Liger-Belair G, Koch BP, Flerus R, Kattner G, et al. (2012) Dissolved organic matter in sea spray: A transfer study from marine surface water to aerosols. *Biogeosciences* 9: 1571-1582.

86. Cheng S, Prot JM, Leclerc E, Bois FY (2012) Zonation related function and ubiquitination regulation in human hepatocellular carcinoma cells in dynamic vs. static culture conditions. *BMC Genomics* 13.
87. Bowne JB, Erwin TA, Juttner J, Schnurbusch T, Langridge P, et al. (2012) Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Molecular Plant* 5: 418-429.
88. Kami K, Fujita Y, Igarashi S, Koike S, Sugawara S, et al. (2012) Metabolomic profiling rationalized pyruvate efficacy in cybrid cells harboring MELAS mitochondrial DNA mutations. *Mitochondrion* 12: 644-653.
89. Liebeke M, Dörries K, Zühlke D, Bernhardt J, Fuchs S, et al. (2011) A metabolomics and proteomics study of the adaptation of *Staphylococcus aureus* to glucose starvation. *Molecular BioSystems* 7: 1241-1253.
90. McNamara LE, Sjöström T, Meek RMD, Oreffo ROC, Su B, et al. (2012) Metabolomics: A valuable tool for stem cell monitoring in regenerative medicine. *Journal of the Royal Society Interface* 9: 1713-1724.
91. Neuweger H, Persicke M, Albaum SP, Bekel T, Dondrup M, et al. (2009) Visualizing post genomics data-sets on customized pathway maps by ProMeTra - Aeration-dependent gene expression and metabolism of *Corynebacterium glutamicum* as an example. *BMC Systems Biology* 3: 82.
92. Charles H, Balmand S, Lamelas A, Cottret L, Pérez-Brocal V, et al. (2011) A genomic reappraisal of symbiotic function in the aphid/buchnera symbiosis: Reduced transporter sets and variable membrane organisations. *PLoS ONE* 6.

93. Farshidfar F, Weljie AM, Kopciuk K, Buie WD, MacLean A, et al. (2012) Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome Medicine* 4.
94. Wang X, Wang H, Zhang A, Lu X, Sun H, et al. (2012) Metabolomics study on the toxicity of aconite root and its processed products using ultraperformance liquid-chromatography/electrospray-ionization synapt high-definition mass spectrometry coupled with pattern recognition approach and ingenuity pathways analysis. *Journal of Proteome Research* 11: 1284-1301.
95. Jennen D, Ruiz-Aracama A, Magkoufopoulou C, Peijnenburg A, Lommen A, et al. (2011) Integrating transcriptomics and metabonomics to unravel modes-of-action of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in HepG2 cells. *BMC Systems Biology* 5.
96. Ma Y, Zhang P, Wang F, Liu W, Yang J, et al. (2012) An integrated proteomics and metabolomics approach for defining oncofetal biomarkers in the colorectal cancer. *Annals of Surgery* 255: 720-730.

Appendix C Preface

This appendix contains a book chapter published in the Encyclopedia of Magnetic Resonance (*eMagRes*). It describes the many uses of nuclear magnetic resonance for addressing questions in environmental microbiology. I wrote it with assistance from my co-authors, especially Dr. Weljie for the section on NMR methodology.

APPENDIX C: METABOLOMICS IN ENVIRONMENTAL MICROBIOLOGY

Sean C. Booth, Raymond J. Turner and Aalim M. Weljie

Published in *eMagRes*, December 2013, 2:4

Copyright permissions for the reproduction of this manuscript can be found in Appendix E

C.1. Abstract

Environmental microbiology is a diverse field which stands to gain greatly from the application of metabolomics. Studies on biofilms, metal resistance, and responses to environmental stress have all benefited from the application of NMR based metabolomics. Here we present an overview of how metabolomics can be used to address questions in environmental microbiology. Briefly, experimental factors as well as NMR parameters and data analysis are discussed. Next, metabolomics investigations by the authors that used NMR are described providing details on how it can provide real answers to real questions in environmental microbiology. We, as well as other researchers whose work is discussed have found that NMR metabolomics can provide both hypothesis-generating and hypothesis supporting results. While these results were based upon a relatively low number of metabolites, their accurate quantitation provided by NMR can provide both metabolic profile based differences between treatments as well as potential biochemical explanations for these differences. Despite these successes, NMR has been relatively under-utilized in environmental microbiology. Future improvements to NMR technologies should not remain un-exploited as metabolomics has great potential to aid in understanding physiological issues in environmental microbiology.

C.2. Introduction

The field of environmental microbiology explores how microbes interact with and influence their environment. It encompasses a wide breadth of important endeavors from attempts at cultivating bacteria from lakes below the Antarctic ice (1) to the isolation of novel biotechnology enzymes from geothermal hot springs (2). Indeed much of the recent work in the field has focused on studying these so-called ‘extremophiles’(3).How do organisms such as *Deinococcus radiodurans* withstand huge doses of radiation (4) or how does *Cupriavidus metallidurans* thrive in the midst of toxic heavy metals (5)? Researchers also seek to understand *how these microbes interface with their environment*. In the post-genomic era there is an abundance of ‘systems biology’ tools (so-called ‘omics’ approaches) which allow for the high through-put analysis of an organism’s functionality. Genome sequencing can give an explanation of all the potential gene products an organism *can* make. Next, transcriptomics describes what genes an organism *are* actually expressing. Proteomics takes this a step further by seeking to reveal all the proteins that are present, hence what the cell *could* be doing. Finally, at the pinnacle is metabolomics, which provides a snapshot of exactly what the cell *is* doing. While the former three tools have important uses in environmental microbiology, especially for culture-free techniques, metabolomics is an ideal tool for evaluating the physiological state of a microbe and thus how it is functioning in response to their environment. (For higher organisms *see* emrstm1350 and emrstm1342.)

C.2.1. Questions in environmental microbiology

The current investigations in environmental microbiology are as diverse as the habitats the organisms come from. Much of the focus of the field in recent years has been directed towards understanding previously uncharacterized and/or unknown metabolic pathways. Another direction

is the evaluation of microbes under severe environmental stress. For example, due to rising levels of anthropogenic pollution, microbes capable of degrading organic xenobiotics such as pesticides and plastics as well as those able to survive and potentially immobilize or biotransform the speciation of toxic heavy metals have gained economic value (*see* emrstm1347 and emrstm1346). Thus, researchers are interested in how the hyper-resistors and poison-eaters are capable of such useful metabolic feats.

While much of the field of environmental microbiology involves working with samples from the field, many issues can be addressed with wholly laboratory based experiments. One of the most important revelations in microbiology has been the acceptance that bacteria naturally grow as biofilms, not as free-swimming planktonic individuals (6). A biofilm can be defined as an assemblage of cells (single or multi-species) growing attached to a surface in a self-secreted layer of extracellular polymeric substance (7). Biofilms have been found to be very physiologically different compared to genetically identical planktonic counterparts (8). Due to their tolerant properties, and differences from planktonic cultures (which have classically been used to determine inhibitory concentrations of anti-bacterials) the study of microbial biofilms has evolved into an important subject unto itself. For example, a major reason that biofilms have reached a prominent position in microbiology, and even in society at large is their much higher tolerance and resistance to anti-microbials of any kind (9). With respect to environmental microbiology, the ability to withstand much higher concentrations of toxins such as organic pollutants as well as heavy metals is potentially beneficial since these abilities can be put to use for bioremediation. This is the process of using living organisms to diminish (completely) the concentration and toxicity of anthropogenic pollutants that have been released into the environment. For organic pollutants this entails the metabolic conversion of pollutants into innocuous end-products

(preferably carbon dioxide and water). Since metals cannot be broken down their remediation must either involve sequestration or chemical alteration into a less toxic form. Thus understanding changes in physiology of bioremediation capable organisms under the stress of the pollutant is important for optimization of the process leading to downstream economic benefits.

C.2.2. Metabolomics

Each specific question in microbiology that requires an understanding of the physiological state would be aided by the use of metabolomics (*see* emrstm1256). The choice of the metabolomics method however is key to success. One can evaluate preferential metabolic pathways by simply evaluating the growth of the organism under different reductant, oxidant, and carbon sources. A popular approach is the BioLog Phenotype Microarray technology (10), which is a high throughput growth screen method. Although this gives a good overview of metabolic possibilities it does not provide specific information on pathways used or build up of intermediates. Evaluation of metabolites present is a good way to obtain such information. In metabolomics, one of the first questions is considering what metabolites and biomolecules one wants to investigate. Many cell extraction protocols are available (11), but it is clear this step needs to be optimized for each microbial system and question being asked.

Once a metabolite extraction method is worked out, the next decision is the separation technology combined with the detection method one wishes to use to identify and quantify the metabolites. The most common competing technologies are that of mass spectrometry (MS) and NMR based approaches. Chromatography-coupled MS has the advantage of separating hundreds of compounds out and detection with high sensitivity, yet identification remains a significant

challenge (12). NMR is considered a stronger method for quantitation and identification, yet will yield a much smaller number of metabolites.

As with any approach the specific use of NMR for metabolomics must be refined to particular cases. This is due to the relatively high concentrations required to generate a detectable NMR signal, a non-trivial consideration for the lower cell density in environmental microbiology. It is worth noting though that NMR can be applied to problems in microbiology in a myriad of useful ways, of which metabolomics is just one (13). For laboratory based work where ease of sample prep and unbiased quantification is of high priority, NMR metabolomics occupies a highly useful laboratory niche. This includes so-called ‘footprinting’ or ‘fingerprinting’ applications where a reproducible and reliable overall pattern of change provides phenotypic information (14). Specific metabolites of interest can be followed up using additional analytical tools such as mass spectrometry.

C.2.3. Methods

In order to perform a metabolomics experiment using NMR, the bacteria must be cultured, and the resulting metabolite sample mixture placed into the magnet. Between these steps a number of important considerations must be made (*see* emrstm1352). An excellent set of reviews for developing metabolomics experiments are available (15, 16). Metabolomics inherently works best as a comparative platform, so the relevant control and experimental treatments must be decided. Here factors such as growth time, method, media, amount of stress etc. must be carefully controlled to ensure that the metabolic response that is to be observed is relevant to these parameters. For example, to test the effect of a stress, an undue effect upon the doubling time of an organism must be considered as it may confound the analysis. Additionally, with regards to stresses such as toxic

metals, ideally a concentration which is toxic but not lethal should be used (17). Sampling methods tend to vary widely and are a source of some controversy as there is no well-accepted standard within the field (18). The main concern is that a sampling method should stop metabolism and any residual enzymatic activity, thereby providing an accurate snapshot of the metabolic state, without altering the sample in anyway. This is difficult as metabolite turnover can be below the order of seconds so most techniques rely upon cooling. Techniques used in the studies described below ranged from none (19, 20), to cold 60% methanol (21-23), to liquid nitrogen (24, 25). Additionally, the method (if any) used for extracting metabolites from a sample has the potential to bias the types of compounds quantified (11). Compared to other tools used in metabolomics NMR has the advantage of being able to directly analyze a sample. When working with microbial samples this can be used to examine the spent media from a culture. This will provide the researcher with information about the so-called 'exo-metabolome', which is to be contrasted with the intracellular metabolites, deemed the 'endometabolome'. Analysis of each has its benefits. While pooling samples is an option in efforts to increase sensitivity in the NMR, further statistical analysis relies on the variance of each sample, and as such it is preferable to obtain as many biological replicates as possible. We recommend 5-10 per biological grouping or condition.

Once a suitable sample is obtained, NMR spectra are typically acquired for each sample of interest. The choice of NMR acquisition and processing parameters should be considered carefully as they have a noticeable impact on the resulting data (26). While a number of NMR nuclei may be and have been considered, we will focus our discussion on the most common proton 1D experiment, although 2D experiments are useful both for validation and as an experimental approach in their own right. Furthermore, quantitative metabolomics work is increasingly

important, and as such specific considerations will be provided in this regard. The readers is also referred to excellent more extensive discussion of these considerations for further information (27).

C.2.4. Acquisition parameters and spectral processing

We have shown previously that the pulse sequence chosen has a small, but measurable, impact on the quantitative aspects of the overall analysis (28). The most commonly employed sequence is a 1D ^1H spectrum, which is analogous to the first slice of a NOESY experiment with presaturation of the solvent (water) resonance during both the relaxation delay and mixing time. Water suppression properties of the typical NOESY-preset pulse sequence may not always be as ‘clean’ as other sequences which use gradient or shaped pulses (such as WET, Watergate, Excitation Sculpting, etc). Worth noting, is that each sequence will provide a characteristic excitation profile across the spectrum, which will differ slightly from each other. If comparison to existing data or acquired spectra is desirable, then the NOESY-preset sequence with a 90° pulse is recommended. This is particularly true for quantitative applications where the chemical shift reference is at the edge of the spectrum (such as DSS or TSP) as the distortions are greatest at the spectral extremities. Furthermore, it is often not possible to acquire spectra in which the recycle time is 3-5 times the longest T_1 of the analyte resonances. In this case a correction factor can be applied such as the integral ratio to a reference peak. In addition, the sweep width of acquisition should be considered carefully in order to provide some allowance for baseline correction. At least 2 to 5 ppm on each side of the range of resonances of interested should be allowed. Finally, the number of acquired points must be high enough to provide resolution to define the multitude of peaks, which will be present in the complex mixture of bacterial metabolites. Typical experiments

range from 32 scans to several thousand with total experiments taking between 5 and 30 minutes per samples depending on desired signal to noise.

On a practical note, the sample properties are often variable depending on the type of analysis being undertaken, a consideration for tuning and matching of the probe electronic circuitry. For example, analysis of bacterial spent media and pellets from a single experiment likely require adjustment of the probe due to differences in ionic concentrations and content. Samples that are dried and reconstituted in a similar solvent tend to behave more homogeneously, albeit with experimental variance due to the extra processing steps. Also, it should be ensured that prior to acquisition each sample is appropriately shimmed. In our hands, during the course of an automated experiment, automatic gradient shimming can maintain a linewidth of <1.0 at halfheight as measured on an internal standard of DSS or TSP at fields of 500 MHz or greater. Appropriate quality control checks are critical to ensure that the data being collected is consistent and useful.

Several strategies to improve the signal-to-noise or spectral resolution are often employed in the processing stage for metabolomics experiments by NMR. Zero-filling is highly recommended, and a suitable size is an under-appreciated element of quantitative data processing. In high resolution NMR, spectral features may be subject to registration errors due to slight peak shifts from sample to sample due to insufficient digitization. As such, we recommend that a typical 1D proton spectrum be zero filled to a minimum of 128k points. The choice of window function is dependent on the particular demands of the application. Typical values for line broadening are 0.1 to 0.5 Hz, although it is worth noting that in very complex mixtures at high resolution the upper bounds will cause a perceptible loss of resolution. Once these parameters are determined for a given sample matrix, they should be applied consistently throughout the dataset. In contrast, the optimal settings for phase-shifts and baseline correction are less clear and not always applicable

to each sample. One will face a tradeoff between automation for large datasets and the need for fine manual adjustments. This is particularly true in the event of distortions from water suppression artifacts and/or large dynamic range issues.

C.2.5. Data pre-processing

Analysis of the spectral data needs to be converted into a format amenable to multivariate statistical analysis as described below. Several strategies have been employed which can be summarized in three broad categories: 1) spectral binning; 2) high-resolution analysis, and 3) peak deconvolution. Spectral binning was chronologically the first approach used in data analysis, and is conceptually the simplest. The spectrum is divided into discrete integral bins, or buckets, typically ranging between 0.01 to 0.04 ppm. This allows for peak shifts due to pH and ionic strength variation to be captured, and each resonance registered into the same ‘bin’ from sample to sample. In practice this is a very rapid approach, however it suffers from several drawbacks. The first is the so-called edge-case when a particular resonance is on the border of two bins causing the peak to be registered in different variables from sample to sample. Another drawback comes from peak overlap, as low intensity peaks will be completely masked by the total binned integral of larger peaks within the same integration region. To reduce the artifacts, the group from Imperial College employed a strategy in which the whole spectrum is analyzed at high resolution (29). This approach has been employed with great success by their group and others, and has the advantage of maintaining spectroscopic characteristics that allow for spectral interpretation of exact peak locations, and the ability to visualize specific peaks by statistical parameters (30). The major downside is that the data interpretation remains quantitative only in a relative sense, and that significant

NMR and statistical analysis skills are required for successful data interpretation. This strategy also suffers from overlap restrictions masking weak peaks (31).

The final category of spectra processing entails conversion of spectral information to simplified peak and/or metabolite information through some kind of peak deconvolution. The simplest of these consist of conversion of spectral peaks into discrete variables such as line spectra or Lorentzian peaks. More sophisticated approaches take into account information about related resonances from a given molecular structure, depending on a library of metabolites which can be overlaid to reconstruct the mixture spectrum. This latter set of methods has been shown to be quantitatively useful through approaches such as ‘Targeted Profiling’ (32), however manual input is needed (see Figure-A-C-1). Newer approaches which use Bayesian methods are promising as a means to automate this process, although are yet unproven for quantitative work. Also spectral ordering is a promising technique in which information from one spectrum can be tracked through an entire dataset for quantitative information (33). Once quantified, the data must undergo further processing in order to make comparisons between samples. Normalization allows for the comparison of samples of different initial concentration, essentially converting concentrations to a relative measure thus ensuring that any observed differences between samples were not caused by mere sample size/volume/mass etc. Regardless of the quantification method, the data can be normalized based on inherent properties of the dataset, or using externally measured properties. Measuring the dry weight, number of cells or total protein concentration of a sample will allow normalization to this metric, providing metabolite concentrations as a per gram of dry weight or protein. Statistical methods are also available which only rely on the inherent properties of the dataset. Total integral normalization is a common method whereby the concentration of any one compound (or peak) is divided by the sum of ALL others, which is performed for each sample.

Other more complex methods exist as well such as median fold-change normalization which computes an ‘average’ (either by mean or median) spectrum then calculates the difference in concentration for each feature in a sample and normalizes all the features in that sample by the median of the fold-change differences from the ‘average’ sample. Using a proper normalization procedure is important for making comparisons between samples, especially in situations where the dynamic range of concentrations is very wide.

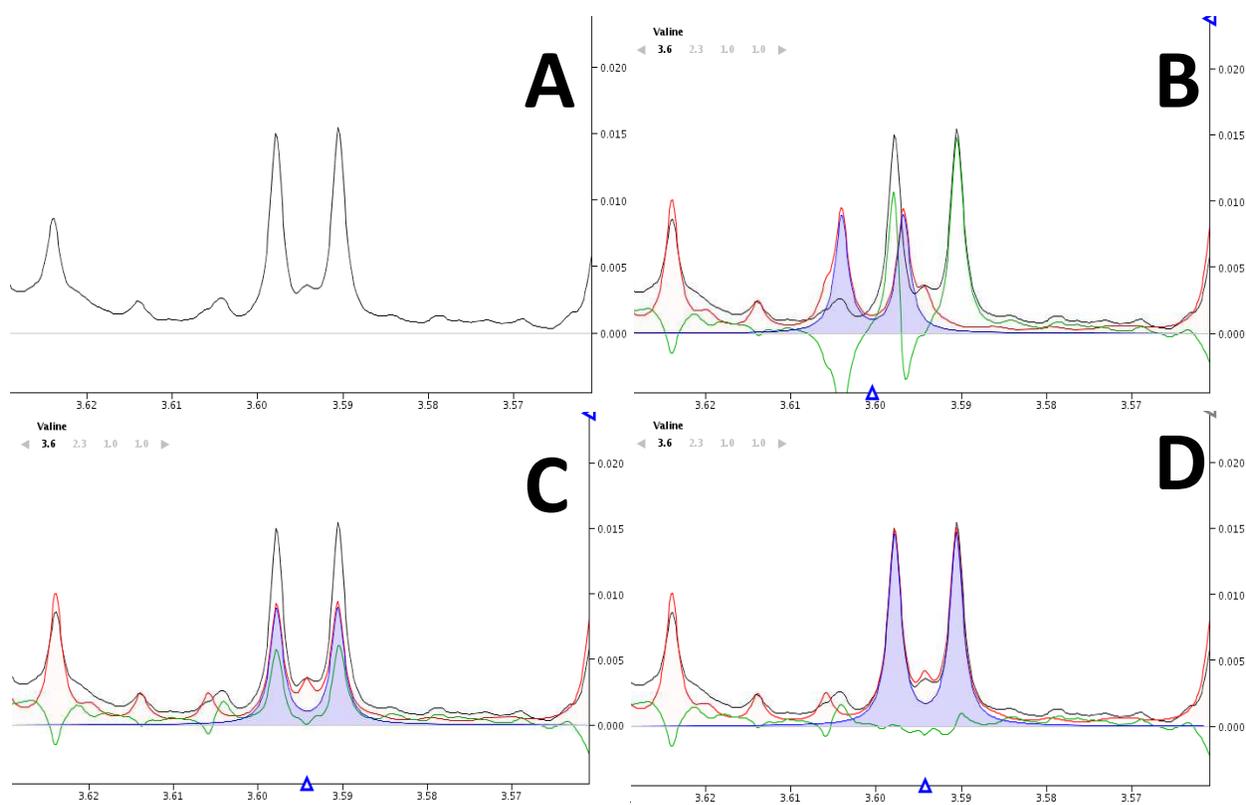


Figure A-C-1: Targeted Profiling in Chenomx™. After phase and baseline correction, NMR spectra can be viewed in the Chenomx™ Profiler (28). Initially, just the original spectrum is visible (A), as a black line. Using the provided library of reference compounds, peaks at the appropriate chemical shift can be added (B), in blue. An additive spectrum of all profiled compound’s peaks can then be viewed, as a red line. A subtractive spectrum, as a green line, is also provided which removes the profiled spectrum from the original to ensure that peaks have been properly identified and quantified. Profiled compounds are then matched up to peaks in the original spectrum (C) and manipulated until they match in amplitude to provide quantitative information (D).

C.2.6. Statistical Analysis

Powerful statistical methods have been a necessity for deriving useful biological information from metabolomics datasets (34). This is the case as cells can contain thousands of different metabolites but only some will be indicative of changes caused by perturbations due to the environmental stress under investigation. Since there are likely more variables than samples (especially with spectral binning) as well as numerous confounding factors not the least of which is biological variability and issues of co-variance between variables, univariate statistical techniques such as t-tests and ANOVA fall short in understanding differences between sample classes. Hence researchers have tended to use tools such as PCA, PLS-DA and clustering. These techniques are all similar in that they take a collection of samples (i.e. replicates and treatments) and a collection of observations (metabolites or bins/spectra) together and determine the overall relationships and variances within the entire dataset. Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) are modeling techniques whereby all the observations for a single sample are condensed and used to calculate its value in a new variation-describing space. Similarly, for each observation, the values in all samples are compressed into a single value, which describes that variable's (metabolite) relationship to the variance between samples. For both techniques this process is iterated in components, whereby the first component attempts to account for as much variation as possible in the dataset and subsequent components

Box 1: Statistical Terminology

(O)PLS-DA: Partial Least Squares Discriminant Analysis. A supervised multivariate modeling technique where the differences between the input classes are maximized and within class variance is minimized. The Orthogonal version computes all variation in the data that is not related to the provided classes in a separate, orthogonal component portion of the model which makes the model easier to interpret. The discriminant analysis implies that two sample classes are used for discriminating between one-another.

Coefficient: A value calculated in (O)PLS-DA models for every variable (metabolite) which indicates its relative abundance in each class.

Class: A group of one type of samples, e.g. control or mutant.

Model: A statistical projection of a large dataset with multiple variables. This type of analysis allows for the simultaneous comparison of many variables at once, even in cases where there are more variables measured than samples.

VIP: Variable Influence on Projection. A value calculated in (O)PLS-DA models for every variable (metabolite) which indicates its importance in distinguishing the classes in the model.

Supervised Analysis: Statistical technique where the differences between defined classes are determined. Compare with unsupervised analysis where the inherent variation within a dataset is determined.

SUS Plot: Shared and Unique Structures. A plot that shows the similar and distinct features of two different models. Variables in the top right and bottom left quadrants have similar values in both models. Those in the top left are unique in the y-axis model and those in the bottom right are unique in the x-axis model.

are calculated in the same fashion until the remainder of variation cannot be modeled. The difference between the two techniques is that PCA computes the variation inherent within the model whereas PLS-DA takes class designates (e.g. control, mutant, treatment etc.) and attempts to maximize variation between these groups. These algorithms are implemented in various general mathematical and software packages such as MATLAB, R, Stata, SPSS, but also in specific chemometrics packages such as SIMCA-P (Umetrics) and CAMO.

C.2.7. Biological Interpretation

Once metabolites have been identified, quantified and those that were significantly altered determined the biological meaning can be determined. Most interpretations rely on comparisons between control and treated samples, using knowledge of the treatment to make sense of particular metabolite concentration changes. Metabolomics is very powerful to establish the foundation of hypothesis based research in this regard. Here NMR is particularly strong, as its accurate quantitation allows for reliable interpretation. Deriving biological meaning can also go beyond manual interpretation of metabolite associations which can be aided by NMR specific databases such as the HMDB (35), BMRB (36) or SDBS (37). Numerous computational approaches have been developed that take a list of metabolites and automatically query biochemical databases for associations between these compounds (38). This provides higher level information on which metabolic pathways are likely being affected. Again the quantitative power of NMR becomes useful as metabolic flux through these pathways can be observed by mapping concentrations onto the pathways. These computations have increasing power when more metabolites are input, which will not always be the case using an NMR approach.

C.3. Case Studies

The power of NMR based metabolomics will be highlighted using key environmental examples. Here we briefly give some background for such studies.

C.3.1. Effect of tellurium oxyanion

Tellurium (Te) is of industrial interest due to its semiconductor properties(39). Pure Te is insoluble and thus non-toxic. The oxyanion forms of tellurite (TeO_3^{2-}) and tellurate (TeO_4^{2-}) are highly toxic to microorganisms (40) at concentrations much lower than other metals (41). Even with considerable effort, a complete picture of the mechanisms of toxicity and resistance is still lacking (39).

Pseudomonas pseudoalcaligenes KF707 is a model organism capable of degrading polychlorinated biphenyls which has also been used to study metal resistance (22). Its natural resistance to tellurite has led to its use in understanding resistance mechanisms to this metalloid (42). Tremaroli *et al.* began investigating tellurite resistance mechanisms by screening a library of mutants generated by transposon insertion for strains that were hyper-resistant to tellurite (22). Through this process a particular mutant, deemed T5, was isolated which was 2-3 times more resistant than the wild-type and also able to grow without a lag phase in the presence of tellurite. These similarities, along with the increased resistance made T5 a good candidate for use in understanding tellurite resistance. Genetic analysis demonstrated that the transposon had inserted into the *cheA* chemotaxis gene. It was unclear how this was responsible for the observed tellurite resistance, thus obfuscating the nature of the mutant. This is exactly the type of situation where metabolomics can be applied to aid in understanding the effect of an unknown mutation. The hyper-resistant strain was thus subjected to metabolomic analysis in order to shed light on its tellurite resistance mechanisms.

Control and Te exposed cultures were compared using NMR metabolomics. In order to optimize the metabolic differences due to metal exposure between the two strains at the point of observation, the amount of tellurite and exposure time was carefully considered in order to cause maximal damage in the wild-type but before the T5 mutant began to show any effect of the tellurite.

The experiment identified and quantified 28 metabolites, which accounted for most of the peaks in any single NMR spectrum. To understand the similarities and differences of how tellurite was affecting each strain, several OPLS-DA models were generated. These models respectively compared the wild-type to T5, and independently each strain control and exposed to tellurite. This allowed the effect of the mutation alone to be separately examined as well as the different effects of tellurite in each strain.

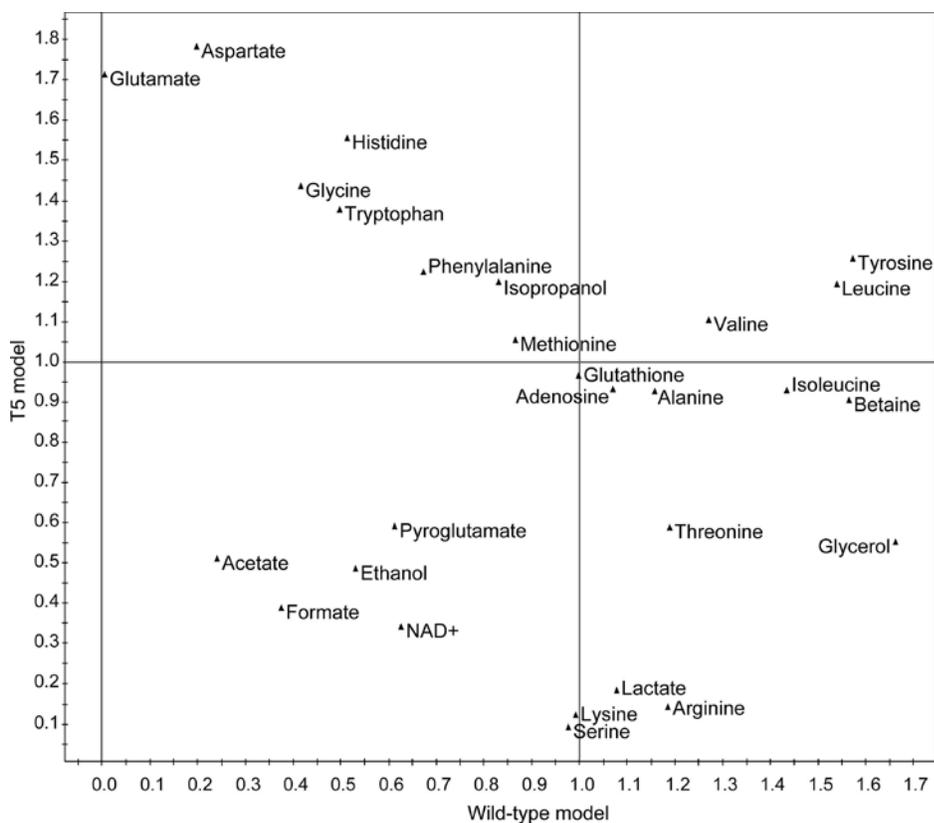


Figure-A-C-2: SUS Plot of VIP values obtained from models comparing *P. pseudoalcaligenes* KF707 wild-type with and without tellurite and the T5 hyperresistant mutant with and without tellurite. The x-axis shows the metabolites that are important in the wild-type; the y-axis shows those that are important in the T5. Metabolites with a VIP > 1 were considered significant in distinguishing the control from tellurite exposed samples in each strain. Metabolites in the bottom right were important for the wild-type, in the top left for T5. Those in the top right were important in both and those the bottom left neither. (Figure adopted from (22), with permission)

As was expected, the effect of tellurite was different between the two strains. A VIP-SUS plot highlighted how there were only a few metabolites that were important in control/exposed models (Figure-A-C-2). Contrarily, there were many metabolites important independently in each model. These differences demonstrated that the T5 mutant had a strong effect on the metabolic response towards tellurite. By examining the coefficients of the important metabolites in each case, it was found that the effect of the T5 mutation was complex. Decreases in glutamate and aspartate suggested their conversion into citric acid cycle intermediates, but corresponding increases in α -ketoglutarate and oxaloacetate were not observed as these metabolites were not present in the NMR spectra, perhaps due to insufficient sample volumes. This highlights an importance consideration the NMR approach, that the experimental design is key to success. It is thus vital to ensure that sufficient quantities of culture are used that lead to concentrations of metabolites that are high enough to be detected by a particular NMR set-up. However, through comparing the control samples for each strain it was found that the T5 hyper-resistant mutant showed significant differences in metabolite levels from the wild-type. This was an interesting result as it indicated that the mutation caused a systematic change, not a reactionary change that was dependent upon the exposure to tellurite. Additionally, when both cultures were exposed to tellurite there were no significant differences between them, as well as between the wild-type exposed to tellurite and the unexposed T5. Together these results indicated that the T5 mutant was able to withstand tellurite better through metabolic 'priming'. Specifically, glutathione was observed to be at much higher level in the mutant compared to the wild-type. Tellurite has previously been observed to cause oxidative damage(43), so this observation along with increased levels of branched chain amino acids (which have been linked to oxidative stress (44)) led to the conclusion that the T5 hyper-resistance was due to its superior anti-oxidant abilities. This conclusion was supported by prior

observations of anti-oxidant reactions increasing bacterial tellurite resistance. Another metabolite that was associated with priming resistance to tellurite was betaine. This small molecule is used as an osmoprotectant (45). Increases to this metabolite correlated with the observation that the membrane potential of T5 cells was less perturbed than that of wild-type. Since tellurite interferes with the electron transport chain, stabilization of the membrane was postulated to be mitigating this damage.

This study demonstrated how careful comparison of a few conditions, even when only measuring a relatively low number of metabolites, can lead to a good understanding of a small genetic difference. Examining the metabolites that changed in each strain under tellurite stress did provide some information about why the T5 mutant was hyper-resistant, however it was by comparing the control samples of both strains as well as the comparison with less differences of exposed wild-type to control T5 which really demonstrated what metabolic factors were responsible for the increased resistance. While it would have been beneficial to use larger sample volumes in order to identify more relevant metabolites, such as the missing citric acid cycle intermediates, these observations were not completely necessary to explain the hyper-resistant phenotype. Finally, the certitude in concentrations provided by NMR gave increased confidence on those metabolites that were found to be important such as betaine and glutathione.

C.3.2. Evaluation of phenotypic variants arising from a global regulator mutation.

Bacterial biofilms have been found to contain phenotypically different sub-populations(23). While genetically identical to their parents, these phenotypic variants (which are generally described by their colony morphology on solid media) have been found to display a number of altered traits (46). In *P.aeruginosa* this phenomenon has been implicated in the lung

colonization of cystic fibrosis patients and also in *P. fluorescens* rhizospheric interactions (46, 47). These phenotypic variants can be found at higher frequencies from cultures under stress, which along with the observed increases in antimicrobial resistance in some variants indicates that they may be an important component of resistance mechanisms in biofilms (48). *P. fluorescens* is a soil bacterium which has been used for many biofilm and metal resistance studies (49). Strains in this organism harboring an inactivating mutation in the global activator of cyanide biosynthesis/regulator of secondary metabolism (*gac/rsm*) signal transduction pathway have been found to produce phenotypic variants more frequently (23).

To better understand the differences between phenotypic variants, Workentine *et al.* investigated two colony morphology variants from a $\Delta gacS$ strain of *P. fluorescens* (23). The GacS/GacA two-component regulatory system is an important moderator in a number of physiological processes, including biofilm formation (50). The *P. fluorescens* $\Delta gacS$ mutant became a good candidate for studying metal resistance mechanisms in biofilms, as it was found that variant cells in biofilms is a factor of metal tolerance (51). To this end, Workentine *et al.* isolated two phenotypic variants from cultures of *P. fluorescens* $\Delta gacS$ exposed to non-lethal concentrations of metal ions. The small colony variant (SCV) and ‘wrinkly spreader’ (WS) were separated based both on their appearance as well as dye-binding properties. Apart from morphology only a few other properties distinguished these phenotypic variants, which could be stably propagated even when cultured planktonically.

In order to further differentiate these phenotypic variants and also to find a biochemical basis for the other observed traits, the two phenotypic variants as well as the parental $\Delta GacS$ mutant strain and wild-type *P. fluorescens* were subjected to metabolomic analysis (23). For these purposes, the methods developed for use with *P. pseudoalcaligenes* were again used to quench

metabolism with cold methanol before centrifuging to collect enough cell mass for detection of metabolites with NMR. This time, 32 metabolites were identified and quantified. Using PLS-DA (Figure-A-C-3), it was shown that all four cultures were metabolically distinct from one another.

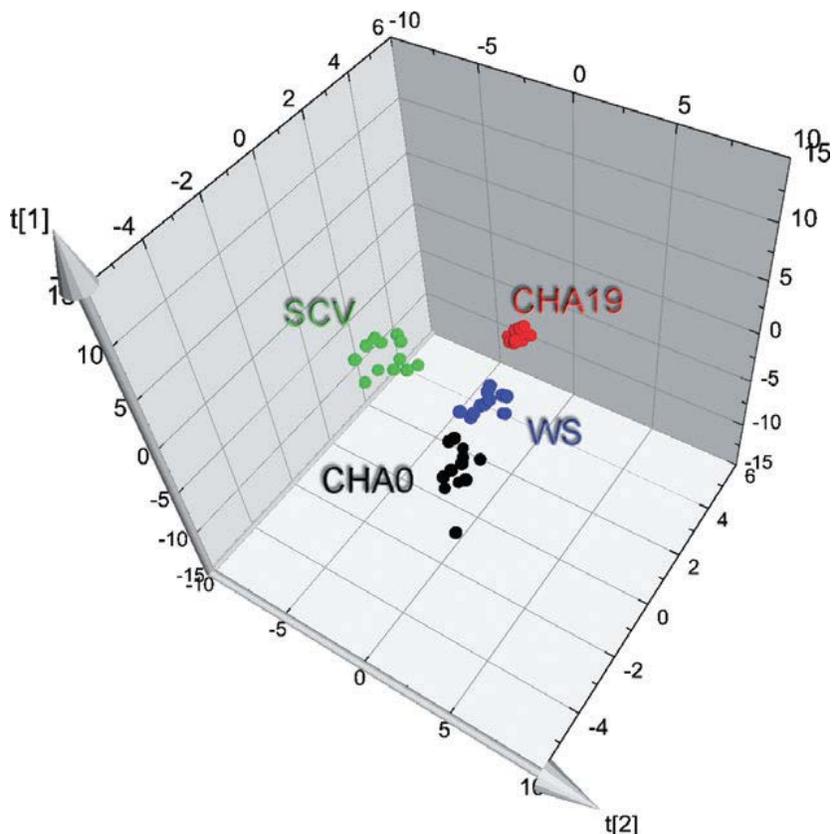


Figure-A-C-3: Partial least squares discriminate analysis (PLS-DA) 3D scores plot from a model comparing *P. fluorescens* wild-type (CHA0), $\Delta gacS$ (CHA19) and the colony morphology phenotypic variants wrinkly spreader (WS) and small colony variant (SCV) obtained from the parental $\Delta gacS$ strain. Each data point represents metabolites extracted from a single sample. Each axis represents a component of variation that separates the four classes based on the concentrations of 32 metabolite concentrations. The clustering of all samples in each class indicates the similarity of the samples and the separation along each axis indicates the difference between classes. (Figure adapted from (23), with permission)

Pair-wise comparisons for all combinations of strains were then performed using OPLS, which further demonstrated the metabolic differences. Metabolite VIPs were extracted from these models and used to produce SUS-plots, which showed which metabolites distinguished the variants from the parental $\Delta gacS$ and wild-type strains. Comparing and contrasting these plots showed that each

variant had a distinct set of metabolites that distinguished it from both ancestral strains. Valine, phenylalanine and glycine were important for the WS but acetate, pyruvate, aspartate, proline and glutamate were important for the SCV. The metabolites implicated in both the WS and SCV phenotypes matched well with prior proteomic observations of a ‘large spreader, wrinkly spreader’ (LSWS) phenotypic variant (52). This study demonstrated upregulated catabolic and transport pathways for all the important metabolites from both phenotypes, which helped define the level of similarity between the variants. The identification of these differences in metabolite concentrations made it clear that global metabolic adaptations are part of the physiology of colony morphology variants leading to an understanding of increased survival in stressed environments.

C.3.3. Evaluation of metal resistance in biofilms

Understanding the differences between biofilm and planktonic cultures has important implications medically but also industrially as biofilms are generally more resistant to toxic antimicrobials (48). This capability has been attributed to a number of factors inherent to cultures growing as a biofilm including the afore-mentioned phenotypic heterogeneity, quorum-sensing mediated gene regulation, as well as metal chelation by the extracellular polymeric substance (EPS) all biofilms produce (48). The phenotypic variability within biofilms is caused by physical heterogeneity within the biofilm micro-environment (53). This heterogeneity is considered to be caused by gradients of nutrients, waste products, pH and oxygen. These features are only found in biofilms and are not present in shake-flask planktonic cultures. All these factors are known to have an influence on cellular metabolism and therefore it was expected that the metabolism occurring within a biofilm is different than that of a planktonic culture. Thus it was reasoned that metabolism may be an additional factor responsible for the differences in metal tolerance found in microbial biofilms (21).

To investigate this possibility, Booth *et al.* expanded upon past studies by using a combined NMR and GC-MS metabolomics approach to examine the differences between biofilm and planktonic cultures of *P. fluorescens* exposed to copper (21). This combined approach was also exploited by another group to understand medically relevant biofilms of *Salmonella* (54). Copper, a metal that has been used by humans for millennia, is responsible for much environmental contamination due to mining activities but is also seeing medical uses as an anti-microbial surface (55) and for water sterilization (56). In both cases an in depth understanding of copper toxicity and resistance in microbial biofilms is important as their metal tolerating traits could be used in mitigating environmental contamination, i.e. bioremediation (57).

P. fluorescens was grown both as a biofilm and planktonically up to stationary phase then exposed to copper sulfate. After a period of copper exposure, metabolism was quenched and cells were harvested. A final concentration of 1.5mM $\text{Cu}(\text{SO}_4)_2$ was used, which is about half the minimum inhibitory concentration (49) as this amount of metal was expected to cause toxicity effects without being lethal as cell death would confound the metabolic results. The biofilms were grown on the inside of silicone tubing by pumping media through for 48 hours to allow for a biofilm to form with enough cells for the extraction. Cell mass was harvested by draining the tube then extruding it through a metal clamp. 30cm of tubing was used to ensure there was enough sample to generate a strong NMR signal. The same metabolite extraction procedure was used as in prior work, only slightly modified to divide the sample for both NMR and GC-MS processing. In this study 48 metabolites were identified and quantified by NMR and several hundred by GC-MS. Due to the inherent noisiness of GC-MS data, significance analysis of microarrays (SAM) and PLS-DA were used to reduce the number of metabolites to only those that were significantly different. This resulted in 200 metabolites, which included most of those found by NMR. The

relative abundances of these metabolites were determined as an average from the unit-variance scaled and mean-centered values from each method.

Unsupervised statistical analysis showed that there were significant metabolic differences between the control and metal exposed biofilm cultures but not the planktonic samples. Supervised analysis however demonstrated that there were significant differences between the control and metal exposed planktonic cultures, in addition to major differences between the planktonic and biofilm samples of each type. The goal of this study was to understand how biofilms are resistant to toxic metals so individual OPLS-DA models were generated for each culture type. Comparison of these models allowed for the effect of metal stress under each growth condition to be understood, and contrasted. The first indication that the response to copper was very different between culture types was a VIP SUS-plot that showed that only NAD, glutathione and phosphoric acid were important in both models. This was further evidenced by a coefficient SUS-plot, which showed that metabolites with a large magnitude of change in on culture type tended to have a much lower magnitude in the other.

In order to understand the actual metabolic differences between toxicity responses in each culture type, hierarchical clustering and enrichment analysis were used. Hierarchical clustering is a technique that organizes a dataset such where the most similar samples are placed adjacent to one another while the most similarly varying metabolites are simultaneously organized in the same way. A heatmap produced using this technique showed a clear division between the biofilm and planktonic samples, as well as the control and exposed biofilms. The planktonic control and exposed samples were also separated, but not as well as the biofilm. As for the metabolites, they clearly grouped into three groups, those that were altered in the planktonic cultures, those in the biofilm and those altered in both. These lists of metabolites were used as input for the enrichment

analysis. Enrichment analysis is a technique which takes a list of metabolites and, using annotation information from a large biochemical database such as KEGG (58) computes which metabolic pathways are likely being perturbed (38). Through this analysis, a number of different metabolic pathways were found to be altered in each culture type (Table-A-C-1). The metabolic pathways altered in the biofilms in response to copper stress appeared mostly to be involved in EPS production. The EPS is composed of both proteins and polysaccharides. Alterations to starch, sucrose and galactose metabolism indicated changes were being made to polysaccharide production. Changes were also observed in a large number of amino acid metabolic pathways which suggested wholesale alterations in protein production. The EPS of biofilms has been found to chelate metals, protecting the cells inside. It was thus concluded that the observed changes were to thicken the EPS to chelate the copper and prevent it from poisoning the biofilm cells (59). In the planktonic cultures exposed to copper, changes to many of the same amino acid metabolic pathways were also observed. Since planktonic cells do not produce EPS, it was suspected that these changes were due to greater protein turn over as those damaged by copper were recycled and stress proteins synthesized. Additionally, the TCA cycle, pyruvate and nicotinate/nicotinamide metabolism were perturbed in planktonic cultures.

These observations corresponded to similar physiological changes produced in *P. fluorescens* cultures exposed to aluminium (60). These studies demonstrated that central metabolism was re-arranged to deal with the oxidative stress caused by aluminium. Well the metabolites altered by copper exposure did not match exactly with those affected by aluminium, it was concluded that copper induces a similar stress response. The metabolomics investigation from the Booth *et al.* study (21) demonstrated for the first time a clear physiological difference between the planktonic and biofilm modes of growth. Furthermore, the study also showed the

drastic difference between biofilm and planktonic culture's responses to metal stress. While it was known that biofilms are more resistant to toxic metals than planktonic cultures, it was not known that biofilms use modified metabolism to enable this resistance.

Table-A-C-1. Metabolic pathways enriched in response to copper exposure determined by enrichment analysis. Only pathways with a p-value <0.05 were included, and life-general metabolic pathways (such as aminoacyl-tRNA synthesis) and plant specific pathways were excluded. (reproduced from (21) with permission)

Planktonic		Both		Biofilm	
Metabolic Pathway	p-Value	Metabolic Pathway	p-Value	Metabolic Pathway	p-Value
Nicotinate and Nicotinamide Metabolism	0.01	Glycine, Serine and Threonine Metabolism	0.01	Arginine and Proline Metabolism	.01
Pyruvate Metabolism	0.01	Cyanoamino Acid Metabolism	0.01	Phosphotransferase System	.01
Glycolysis/Gluconeogenesis	0.01	Nitrogen Metabolism	0.01	Lysine Degradation	0.01
Purine Metabolism	0.01	Phenylalanine Metabolism	0.01	Galactose Metabolism	0.01
TCA Cycle	0.01	Glutathione Metabolism	0.01	Glutathione Metabolism	0.01
Alanine, Aspartate and Glutamate Metabolism	0.01	Alanine, Aspartate and Glutamate Metabolism	0.01	Starch and Sucrose Metabolism	0.015
Valine, Leucine and Isoleucine Biosynthesis	0.01	Valine, Leucine and Isoleucine Biosynthesis	0.01		
		Phenylalanine, Tyrosine and	0.035		

C.3.4. Other Representative applications

Other groups have used NMR metabolomics to address a number of other issues in environmental microbiology. Bundy *et al.* have examined osmotic stress in both *Burkholderia cenocepacia* and *P. aeruginosa* (19, 20). While these organisms are mostly studied from a clinical standpoint, neither are obligate pathogens making an understanding of their natural growth relevant.

P. aeruginosa is a well-studied opportunistic pathogen. Isolates obtained from long-term infections have been found to over produce alginate, a polysaccharide whose exact environmental function is unknown. Strains with a mutated *MucA* gene overproduce this compound, which is not normally an important component of wild-type biofilms. To gain a better understanding of the mutant *MucA* phenotype, a temporal analysis of spent media from wild-type and *mucA22* mutant *P. aeruginosa* as well as an end-point analysis of both cells and spent media under osmotic stress were performed using NMR metabolomics (20). The initial screen comparing exogenous metabolites from the two strains indicated differences in the concentrations over time of valine, methionine, glycine-betaine and trehalose. These last two metabolites are known bacterial osmolytes suggesting that the *mucA22* mutant may have an effect on osmotic stress. This metabolomics derived hypothesis was confirmed by growing the two strains under a range of salt concentrations, with the wild-type consistently outgrowing the mutant. At this point the researchers returned to metabolomics in order to identify which metabolites might be responsible for the differences in osmotolerance. This time however both strains were either exposed to various

salt concentrations and an end-point measurement taken or a single high salt concentration and samples were taken over the entire course of growth. Four metabolites were found to increase inside the cells of both strains when salt concentration was increased. N-acetylglutaminyglutamine amide (NAGGN), glutamate, glycine-betaine and trehalose were found at comparable levels in both strains at most concentrations of salt. At the highest concentrations tested (0.7, 0.8M), the wild-type showed significantly higher levels of all these osmolytes. When analyzing the differences between the two strains at a single concentration of salt (0.4M) over time, PCA analysis showed that the main difference between samples was growth time however the second component separated the control from salt exposed samples. This was surprising as these data did not support the previously observed differences between the wild-type and *mucA22* mutant. As such a more sophisticated analysis was undertaken which integrated the observations over time for each treatment. The hierarchical PCA that was then obtained showed clear differences between the strains, as well as between the control and exposed samples. This emphasizes the importance of understanding when metabolic differences will be maximized in a metabolomics experiment as merely observing an endpoint in this case would not have demonstrated the effect of the *mucA22* mutation on osmotic stress response. Also from this analysis, differences in the uptake of several metabolites were observed. Relevant to the osmotic stress hypothesis, when under salt stress both glycine-betaine and glutamate were taken up later by the mutant compared to the wild-type. These results, along with the data from the salt concentration experiments clearly showed that *mucA* has a role in the osmotic stress response. This study demonstrated how NMR metabolomics, even when only quantifying a small number of metabolites but under a wide variety of conditions, can generate and subsequently support testable hypotheses.

Strains of *B. cenocepacia* have been found to be plant-growth promoting in the rhizosphere, an environment that may be prone to osmotic stress. To understand the differences in osmotic stress tolerance of several strains, including two highly pathogenic strains, both intracellular and extracellular metabolites were quantified under control and hyper-saline conditions using NMR metabolomics (19). Of the metabolites that were detected, five were deemed osmoresponsive in at least some strains. Glycine-betaine and trehalose are well known bacterial osmolytes, the other three were the amino acids alanine, glutamate and phenylalanine. Contrary to expectations, the concentrations of these metabolites were not altered under increased salt in the same way in the highly pathogenic strains. In fact, three separate responses were observed. Three strains, including one of the highly pathogenic strains, showed minimally altered levels of osmolytes, but increased levels of the three amino acids. These strains were the most impaired by hypersalinity and so it was postulated that these changes were due to a general stress response. The remaining two strains, one which was highly pathogenic displayed different metabolite alterations which granted some protection. One strain's resistance was attributed to its constitutively high levels of glycine-betaine whereas the other, a highly pathogenic strain showed marked increases in all five metabolites, including trehalose and glycine-betaine. Results from the quantification of these five metabolites in the spent media were less compelling. All strains showed decreased usage of the amino acids, likely due to decreased growth, while the osmolytes showed only slight increases in use in the corresponding strains. Still this study demonstrated how diverse the osmotic stress response can be, even among five closely related strains. Additionally, the lack of congruity between the osmoprotective strategies of the highly pathogenic strains as well as the differences between the other strains support the importance of environmental stress responses in clinically relevant bacteria.

Coral reef ecosystems have become endangered due to human activities (61). Normally corals live symbiotically with bacteria such as *Vibrio coralliilyticus* but this interaction can turn pathogenic as temperatures above 24°C induces virulent coral bleaching and lysis (62). Boroujerdi *et al.* set out to understand this temperature dependent pathogenicity by comparing the metabolic profiles of *V. coralliilyticus* cultures grown at 24°C and 27°C using NMR metabolomics (25). While much of this work focused on understanding inter and intra-batch effects caused by natural biological variation, PCA identified a number of osmoprotectants that distinguished the high temperature grown pathogenic cultures from the controls. Additionally, the observed dissimilarity between metabolic profiles obtained at each temperature has led to continued research into this phenomenon (44).

Staphylococcus epidermis is whose biofilms have been associated with nosocomial infections (63). Since the transition to growth as a biofilm requires a large modification of physiology and hence gene regulation, the methods by which bacteria sense, interpret and respond to signals in their environment must be understood. The TCA cycle is central to a correctly metabolically functional cell, which, due to the propensity of its enzymes to be affected by a wide variety of environmental and nutritional factors led to its implication in transducing signals interpreted from beyond the cell (64). To further investigate this function, Somerville and Powers *et al.* used NMR metabolomics to first compare the independent effects of iron limitation and ethanol stress on cultures of *S. epidermis* (65) then followed up upon this work by examining and contrasting these results with the metabolic effects of an aconitase mutant as well as the effects of salt, glucose and sub-inhibitory levels of antibiotics (66). Using PCA, the initial results showed that cultures stressed either with iron limitation or ethanol were metabolically similar to the unstressed aconitase mutant strain. While the ethanol stress caused some additional effects, this

similarity clearly linked the two stresses with TCA cycle inhibition. To control for a generalized stress response, comparably treated cultures of a *sigB* mutant (a general stress response transcription factor) were also compared. PCA showed that these cultures were different from the control and stressed wild-type cultures as well as the aconitase mutant, indicating that the TCA inhibition was independent from the *sigB* controlled stress response. In the stressed cultures and the aconitase mutant glucose-6-phosphate as well as lactate, acetate, acetaldehyde and acetyl phosphate were increased in concentration relative to the wild-type. These metabolites are indicative of fermentative metabolism, glucose-6-phosphate as an upstream inducer, the rest as downstream end-products. The inhibited TCA cycle associated with these changes was hypothesized to be inducing the production of the biofilm molecule polysaccharide intracellular adhesion (PIA). Using the metabolomic data as a starting point, this was confirmed using appropriate regulatory mutants. Thus it was confirmed that the TCA cycle was being used to regulate responses to environmental stress.

This was followed up by comparing the metabolic effects of additional glucose, ethanol and iron-limitation, all of which were expected to be sensed by the TCA cycle (66). Cultures of both the wild-type and aconitase mutant were exposed separately to each of the stressors and harvested during and after exponential phase making for 12 different experimental conditions. Using spectral binning, concentrations of 55 metabolites were altered under these various conditions. PCA, OPLS-DA and clustering demonstrated that except for the wild-type control, after 6h of growth (post exponential phase) the metabolomes of all these cultures were highly similar. Since the cultures of the stress exposed TCA inactivated aconitase mutants did not differ from their unexposed control, this further supported the notion that the TCA cycle interprets multiple stresses and modulates them into a common response. The concentrations of the amino

acids aspartate, asparagine, glutamate and glutamine, which are biosynthetically derived from the TCA cycle, were found to be lower in stressed cultures as would be expected from an inhibited TCA cycle. Phosphoenolpyruvate, acetaldehyde and fructose-6-phosphate however were found to be increased which was taken to be indicative of active glycolytic metabolism, which would be necessary to produce ATP in the absence of the TCA cycle. Increased levels of several amino sugars were observed including UDP-N-Acetylglucosamine that is a precursor to PIA again linking TCA cycle modulation with biofilm formation. Overall this study clearly demonstrated how different environmental stresses can affect the TCA cycle.

C.4. Conclusions and Prospects

NMR metabolomics has been successfully applied to a number of problems in environmental microbiology of diverse nature, but overall it is an underutilized technique. This may be due to the perception of NMR as a structural tool or due to the comparatively low sensitivity. Improvements in NMR technology, such as increasingly accessible cryo/cold probes, improved detector technology, dynamic nuclear polarization and hyphenation with tools such as solid-phase extraction will serve to overcome some limitations inherent to the technique. Another possible reason for the lack of NMR-based metabolomics studies in this field is the lack of consensus in quenching metabolism and extracting metabolites for microbial metabolomics (18). This issue only serves to further confound environmental microbiologists already working with field samples of unknown concentration. Additionally, many researchers are already fully engaged in keeping up with the rapid advances in genetic sequencing of whole communities, i.e. metagenomics, and optimizing their analysis pipeline from DNA extraction to functional annotation (67). Compared to other metabolomics technologies, analysis of NMR data requires

considerably more expert chemical knowledge, however various efforts towards automating metabolite identification are underway (33). In situations where identification of compounds isn't necessary, such as in finger or footprinting analysis though NMR can be quite powerful. Spent media and filtered aquatic samples can essentially be directly analyzed allowing for rapid screening for metabolic differences between treatments or environments. The use of such metabolic profiling (as in non-identification centered) for ecological metabolomics of higher organisms has been touted as a rapid method for classifying samples (68). While such an approach would be useful for preliminary research, the use of NMR would also be beneficial for elucidating structures from samples containing unknown compounds (69). It is well established by now that NMR and MS are complementary tools for high-throughput applications. For example, compounds from GC-MS are confounded by derivatizing agents and potentially different fragmentation between samples, but detected with greater sensitivity. Tandem MS and tandem separation technologies (LC/LC, LC/GC, or GC/GC) would also be an excellent tool for this type of application, although are not unbiased to the same extent as NMR. For compounds with known structures, NMR is superior for distinguishing ^{13}C isotopomers (70). Combined with its simple sample preparation and quantitative strength this makes NMR an ideal tool for monitoring carbon flux (71), for example in a chemostat (72) or even a bioreactor. Furthermore, the ability of NMR to be applied to other NMR-active nuclei can lead to convenient biodegradation studies of fluorinated compounds (73). While MS has its advantages and will remain a useful tool for metabolomics, NMR has great potential to be further exploited in the field of environmental microbiology. With continuing developments to NMR hardware and software for data analysis, the potential for this technology to help solve the myriad of questions in environmental microbiology will only rise.

C.5. References

1. Murray AE, Kenig F, Fritsen CH, McKay CP, Cawley KM, Edwards R, Kuhn E, McKnight DM, Ostrom NE, Peng V, Ponce A, Priscu JC, Samarkin V, Townsend AT, Wagh P, Young SA, Yung PT, Doran PT. 2012. Microbial life at -13°C in the brine of an ice-sealed Antarctic lake. *Proceedings of the National Academy of Sciences of the United States of America* 109:20626-20631.
2. Fuciños P, González R, Atanes E, Sestelo ABF, Pérez-Guerra N, Pastrana L, Rúa ML. 2012. Lipases and esterases from extremophiles: Overview and case example of the production and purification of an esterase from *thermus thermophilus* HB27, vol 861, p 239-266.
3. Canganella F, Wiegel J. 2011. Extremophiles: From abyssal to terrestrial ecosystems and possibly beyond. *Naturwissenschaften* 98:253-279.
4. Cox MM, Battista JR. 2005. *Deinococcus radiodurans* - The consummate survivor. *Nature Reviews Microbiology* 3:882-892.
5. Nies DH. 2000. Heavy metal-resistant bacteria as extremophiles: Molecular physiology and biotechnological use of *Ralstonia* sp. CH34. *Extremophiles* 4:77-82.
6. Jain A, Gupta Y, Agrawal R, Khare P, Jain SK. 2007. Biofilms - A microbial life perspective: A critical review. *Critical Reviews in Therapeutic Drug Carrier Systems* 24:393-443.
7. López D, Vlamakis H, Kolter R. 2010. Biofilms. *Cold Spring Harbor perspectives in biology* 2:a000398.
8. Spormann AM. 2008. Physiology of microbes in biofilms, vol 322, p 17-36.
9. Ceri H, Olson ME, Turner RJ. 2010. Needed, new paradigms in antibiotic development. *Expert Opinion on Pharmacotherapy* 11:1233-1237.

10. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research* 11:1246-1255.
11. Werf MJvd, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. 2007. Microbial metabolomics: Toward a platform with full metabolome coverage. *Analytical Biochemistry* 370:17-25.
12. Want E, Masson P. 2011. Processing and analysis of GC/LC-MS-based metabolomics data. *Methods in molecular biology (Clifton, NJ)* 708:277-298.
13. Grivet JP, Delort AM. 2009. NMR for microbiology: In vivo and in situ applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* 54:1-53.
14. Resmer KL, White RL. 2011. Metabolic footprinting of the anaerobic bacterium *Fusobacterium varium* using ¹H NMR spectroscopy. *Molecular BioSystems* 7:2220-2227.
15. Álvarez-Sánchez B, Priego-Capote F, Castro MDLd. 2010. Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC - Trends in Analytical Chemistry* 29:120-127.
16. Álvarez-Sánchez B, Priego-Capote F, Luque de Castro MD. 2010. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *TrAC - Trends in Analytical Chemistry* 29:111-119.
17. Booth SC, Workentine ML, Weljie AM, Turner RJ. 2011. Metabolomics and its application to studying metal toxicity. *Metallomics*.
18. van Gulik WM. 2010. Fast sampling for quantitative microbial metabolomics. *Current Opinion in Biotechnology* 21:27-34.

19. Behrends V, Bundy JG, Williams HD. 2011. Differences in strategies to combat osmotic stress in *Burkholderia cenocepacia* elucidated by NMR-based metabolic profiling. *Letters in Applied Microbiology* 52:619-625.
20. Behrends V, Ryall B, Wang X, Bundy JG, Williams HD. 2010. Metabolic profiling of *Pseudomonas aeruginosa* demonstrates that the anti-sigma factor MucA modulates osmotic stress tolerance. *Molecular BioSystems* 6:562-569.
21. Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, Ceri H, Turner RJ, Weljie AM. 2011. Differences in Metabolism between the Biofilm and Planktonic Response to Metal Stress. *Journal of Proteome Research* 10:3190-3199.
22. Tremaroli V, Workentine ML, Weljie AM, Vogel HJ, Ceri H, Viti C, Tatti E, Zhang P, Hynes AP, Turner RJ, Zannoni D. 2009. Metabolomic investigation of the bacterial response to a metal challenge. *Applied and Environmental Microbiology* 75:719-728.
23. Workentine ML, Harrison JJ, Weljie AM, Tran VA, Stenroos PU, Tremaroli V, Vogel HJ, Ceri H, Turner RJ. 2010. Phenotypic and metabolic profiling of colony morphology variants evolved from *Pseudomonas fluorescens* biofilms. *Environmental Microbiology* 12:1565-1577.
24. Boroujerdi AFB, Jones SS, Bearden DW. 2012. NMR analysis of metabolic responses to extreme conditions of the temperature-dependent coral pathogen *Vibrio coralliilyticus*. *Letters in Applied Microbiology* 54:209-216.
25. Boroujerdi AFB, Vizcaino MI, Meyers A, Pollock EC, Huynh SL, Schock TB, Morris PJ, Bearden DW. 2009. NMR-based microbial metabolomics and the temperature-dependent coral pathogen *Vibrio coralliilyticus*. *Environmental Science and Technology* 43:7658-7664.

26. Kaiser KA, Merrywell CE, Fang F, Larive CK. 2008. Chapter 5 - Metabolic Profiling, p 233-267. *In* Ulrike H, Iwona W, Bernd DiehlA2 - Ulrike Holzgrabe IW, Bernd D (ed), *NMR Spectroscopy in Pharmaceutical Analysis*. Elsevier, Amsterdam.
27. Powers R. 2009. NMRmetabolomics and drug discovery. *Magnetic Resonance in Chemistry* 47:S2-S11.
28. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. 2006. Targeted Profiling: Quantitative Analysis of ¹H NMR Metabolomics Data. *Analytical Chemistry* 78:4430-4442.
29. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, Nicholson JK, Holmes E. 2005. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ¹H NMR spectroscopic metabonomic studies. *Analytical Chemistry* 77:517-526.
30. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson J. 2005. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic ¹H NMR data sets. *Analytical Chemistry* 77:1282-1289.
31. Weljie AM, Newton J, Jirik FR, Vogel HJ. 2008. Evaluating low-intensity unknown signals in quantitative proton NMR mixture analysis. *Analytical Chemistry* 80:8956-8965.
32. Chang D, Weljie A, Newton J. Leveraging latent information in NMR spectra for robust predictive models, p 115-126. *In* (ed),
33. Alm E, Slagbrand T, Åberg KM, Wahlström E, Gustafsson I, Lindberg J. 2012. Automated annotation and quantification of metabolites in ¹H NMR data of biological origin. *Analytical and Bioanalytical Chemistry* 403:443-445.

34. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. 2009. Analytical and statistical approaches to metabolomics research. *Journal of Separation Science* 32:2183-2199.
35. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. 2007. HMDB: The human metabolome database. *Nucleic Acids Research* 35:D521-D526.
36. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalnia HR, Sussman MR, Markley JL. 2008. Metabolite identification via the Madison Metabolomics Consortium Database [3]. *Nature Biotechnology* 26:162-164.
37. Technology NIOAISA. 2013. SDBS Web. Accessed March 18.
38. Booth SC, Weljie AM, Turner RJ. 2013. Computational Tools for the Secondary Analysis of Metabolomics Experiments. *Computational and Structural Biotechnology Journal* 4.
39. Turner RJ, Borghese R, Zannoni D. 2012. Microbial processing of tellurium as a tool in biotechnology. *Biotechnology Advances* 30:954-963.
40. Chasteen TG, Fuentes DE, Tantaleán JC, Vásquez CC. 2009. Tellurite: History, oxidative stress, and molecular mechanisms of resistance: Review article. *FEMS Microbiology Reviews* 33:820-832.
41. Zannoni D, Borsetti F, Harrison JJ, Turner RJ. 2007. The Bacterial Response to the Chalcogen Metalloids Se and Te, vol 53, p 1-71,312.

42. Di Tomaso G, Fedi S, Carnevali M, Manegatti M, Taddei C, Zannoni D. 2002. The membrane-bound respiratory chain of *Pseudomonas pseudoalcaligenes* KF707 cells grown in the presence or absence of potassium tellurite. *Microbiology* 148:1699-1708.
43. Tremaroli V, Fedi S, Zannoni D. 2006. Evidence for a tellurite-dependent generation of reactive oxygen species and absence of a tellurite-mediated adaptive response to oxidative stress in cells of *Pseudomonas pseudoalcaligenes* KF707. *Archives of Microbiology* 187:127-135.
44. Tweeddale H, Notley-McRobb L, Ferenci T. 1999. Assessing the effect of reactive oxygen species on *Escherichia coli* using a metabolome approach. *Redox Report* 4:237-241.
45. Roeßler M, Müller V. 2001. Osmoadaptation in bacteria and archaea: Common principles and differences. *Environmental microbiology* 3:743-754.
46. Starkey M, Hickman JH, Ma L, Zhang N, De Long S, Hinz A, Palacios S, Manoil C, Kirisits MJ, Starner TD, Wozniak DJ, Harwood CS, Parsek MR. 2009. *Pseudomonas aeruginosa* Rugose small-colony variants have adaptations that likely promote persistence in the cystic fibrosis lung. *Journal of Bacteriology* 191:3492-3503.
47. Sánchez-Contreras M, Martín M, Villaceros M, O'Gara F, Bonilla I, Rivilla R. 2002. Phenotypic selection and phase variation occur during alfalfa root colonization by *Pseudomonas fluorescens* F113. *Journal of Bacteriology* 184:1587-1596.
48. Harrison JJ, Ceri H, Turner RJ. 2007. Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology* 5:928-938.
49. Workentine ML, Harrison JJ, Stenroos PU, Ceri H, Turner RJ. 2008. *Pseudomonas fluorescens*' view of the periodic table. *Environmental Microbiology* 10:238-250.

50. Workentine ML, Chang L, Ceri H, Turner RJ. 2009. The GacS-GacA two-component regulatory system of *Pseudomonas fluorescens*: a bacterial two-hybrid analysis. *FEMS Microbiology Letters* 292:50-56.
51. Harrison JJ, Ceri H, Stremick CA, Turner RJ. 2004. Biofilm susceptibility to metal toxicity. *Environmental Microbiology* 6:1220-1227.
52. Knight CG, Zitzmann N, Prabhakar S, Antrobus R, Dwek R, Hebestreit H, Rainey PB. 2006. Unraveling adaptive evolution: How a single point mutation affects the protein coregulation network. *Nature Genetics* 38:1015-1022.
53. Stewart PS, Franklin MJ. 2008. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* 6:199-210.
54. White AP, Weljie AM, Apel D, Zhang P, Shaykhutdinov R, Vogel HJ, Surette MG. 2010. A global metabolic shift is linked to *Salmonella* multicellular development. *PLoS ONE* 5.
55. Elguindi J, Hao X, Lin Y, Alwathnani HA, Wei G, Rensing C. 2011. Advantages and challenges of increased antimicrobial copper use and copper mining. *Applied Microbiology and Biotechnology* 91:237-249.
56. Martínez SS, Gallegos AA, Martínez E. 2004. Electrolytically generated silver and copper ions to treat cooling water: An environmentally friendly novel alternative. *International Journal of Hydrogen Energy* 29:921-932.
57. Singh R, Paul D, Jain R. 2006. Biofilms: implications in bioremediation. *Trends in Microbiology* 14:389-397.
58. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28:27-30.

59. Flemming HC, Neu TR, Wozniak DJ. 2007. The EPS matrix: The "House of Biofilm Cells". *Journal of Bacteriology* 189:7945-7947.
60. Lemire J, Mailloux R, Auger C, Whalen D, Appanna VD. 2010. *Pseudomonas fluorescens* orchestrates a fine metabolic-balancing act to counter aluminium toxicity. *Environmental Microbiology* 12:1384-1390.
61. Hoegh-Guldberg O. 2011. Coral reef ecosystems and anthropogenic climate change. *Regional Environmental Change* 11:215-227.
62. Ben-Haim Y, Zicherman-Keren M, Rosenberg E. 2003. Temperature-regulated bleaching and lysis of the coral *Pocillopora damicornis* by the novel pathogen *Vibrio coralliilyticus*. *Applied and Environmental Microbiology* 69:4236-4242.
63. Von Eiff C, Peters G, Heilmann C. 2002. Pathogenesis of infections due to coagulase-negative staphylococci. *Lancet Infectious Diseases* 2:677-685.
64. Somerville GA, Proctor RA. 2009. At the crossroads of bacterial metabolism and virulence factor synthesis in staphylococci. *Microbiology and Molecular Biology Reviews* 73:233-248.
65. Sadykov MR, Zhang B, Halouska S, Nelson JL, Kreimer LW, Zhu Y, Powers R, Somerville GA. 2010. Using NMR metabolomics to investigate tricarboxylic acid cycle-dependent signal transduction in *Staphylococcus epidermidis*. *Journal of Biological Chemistry* 285:36616-36624.
66. Zhang B, Halouska S, Schiaffo CE, Sadykov MR, Somerville GA, Powers R. 2011. NMR analysis of a stress response metabolic signaling network. *Journal of Proteome Research* 10:3743-3754.
67. Chmolewska D. 2013. A practical introduction to microbial community sequencing. *Central European Journal of Biology* 8:399-409.

68. Sardans J, Peñuelas J, Rivas-Ubach A. 2011. Ecological metabolomics: overview of current developments and future challenges. *Chemoecology*:1-35.
69. Leiss KA, Choi YH, Verpoorte R, Klinkhamer PGL. 2011. An overview of NMR-based metabolomics to identify secondary plant compounds involved in host plant resistance. *Phytochemistry Reviews* 10:205-216.
70. Lane AN, Fan TWM, Higashi RM. 2008. Isotopomer-Based Metabolomic Analysis by NMR and Mass Spectrometry, vol 84, p 541-588.
71. Winder CL, Dunn WB, Goodacre R. 2011. TARDIS-based microbial metabolomics: time and relative differences in systems. *Trends in Microbiology*.
72. Beste DJV, Bonde B, Hawkins N, Ward JL, Beale MH, Noack S, Nöh K, Kruger NJ, Ratcliffe RG, McFadden J. 2011. ¹³C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation. *PLoS Pathogens* 7.
73. Boersma MG, Solyanikova IP, Van Berkel WJH, Vervoort J, Golovleva L, Rietjens IMCM. 2001. ¹⁹F NMR metabolomics for the elucidation of microbial degradation pathways of fluorophenols. *Journal of Industrial Microbiology and Biotechnology* 26:22-34.

Preface

This chapter contains detailed molecular biological methods for how I made deletion mutants in *P. pseudoalcaligenes* KF707.

APPENDIX D: SUPPLEMENTARY METHODS FOR ENERGY TAXIS EXPERIMENTS

D.1. Detailed Molecular Biology Methods

D.1.1. PCR

Nucleotide sequences for Aer.g1, Aer.g2, Aer.g4, Aer2 and CttP were obtained from the draft genome sequence of *P. pseudoalcaligenes* KF707 (1). Primers for a ~500bp region up and downstream of each region were generated using Primer BLAST (2). Benchling was used to alter the primers, adding BamHI or HindIII restriction sites to the outer primers and the reverse-complement of the other inner primer to each of the internal primers. Genomic DNA was isolated by the phenol/chloroform method (3). In separate PCR reactions the upstream and downstream fragments were amplified using Hi-Fidelity (HF) Enzyme mix (Fisher Scientific, USA). The mixture contained in μL HF buffer, 2; 1.25mM dNTPS, 2; HF enzyme, 0.2; gDNA, 0.1; 50mM MgCl_2 , 1.6; forward primer, 2; reverse primer, 2; PCR enhancer (Invitrogen, USA), 0/1/2/4; nuclease-free water to 20 μL . PCR parameters: 95°C, 5min; 30 cycles of: 95°C, 40s; 60°C, 45s; 72°C, 180s, then final extension for 10min at 60°C. Fragments were purified by gel extraction using an EZDNA kit (Omega Bio-Tek, USA). These were then pooled and used as the template for the second PCR reaction using only the outer primers. The SOE PCR mixture contained in μL HF buffer, 4; PCR enhancer (Invitrogen, USA), 4; 1.25mM dNTPs, 6.4; 50mM MgCl_2 , 3.2; HF enzyme, 0.2; forward primer, 4; reverse primer, 4; gel purified mixture of upstream and downstream fragments, 10. PCR parameters: 95°C, 3min; 30 cycles of: 95°C, 40s; 60°C, 45s; 72°C, 240s, then final extension for 10min at 60°C.

D.1.2. Cloning

The pG19II vector(4), purified using an EZDNA plasmid mini kit II (Omega Bio-Tek, USA), and insert were digested using BamHI and HindIII (Invitrogen, USA) (3 μ L Buffer K, 1 μ L BamHI, 1 μ L HindIII, 1 μ g DNA, NF water to 30 μ L) overnight at 37°C. Digestion products were purified by precipitating with 3 μ L 3M (NH₄)₂SO₄ and 70 μ L -20°C ethanol, then washing twice with 70% ethanol and resuspending the dried pellet in 20 μ L nuclease-free water. Inserts were ligated into pG19II using T4 ligase (Invitrogen, USA), overnight at 16°C. Ligation mixture contained 4 μ L ligase buffer, 0.1 μ L ligase, 1-30fmol plasmid, 30-90fmol plasmid (total DNA was kept below 1 μ g) and NF water to 20 μ L. Ligations were transformed either directly into *E. coli* Top10F' or first into DH5 α competent cells. Competent cells were generated using the rubidium chloride method (REF). Briefly cultures were grown to early-log phase (OD ~0.3) in LB media, chilled on ice for 15min, then pelleted by centrifuging for 5min time at 2000x g. The pellet was resuspended in 10mL Transformation Buffer I (in g/L RbCl, 12; MnCl₂•4H₂O, 9.9; KCH₃COO, 2.9; CaCl₂ 1.13; glycerol, 150. pH adjusted to 5.8 with acetic acid) and chilled on ice for 15 minutes. Cells were then pelleted again and resuspended in 8mL Transformation Buffer II (in g/L RbCl, 1.2; CaCl₂ 8.3; glycerol, 150; MOPS, 21), aliquoted and flash frozen in -80°C ethanol and stored until later use. For transformations, ligations or purified, uncut pG19II as a positive control, was added to 100 μ L thawed competent cells. Cells were chilled on ice for 15 minutes, heat shocked at 42°C for 90s then 1mL LB was added and they were incubated for 1h at 37°C. Cells were then pelleted and spread on LB plates containing 20 μ g/mL gentamycin and 1mM X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside). After incubation for 24h, white colonies were picked and plasmids purified by alkaline lysis (3). Briefly, 1mL of overnight culture was pelleted and

resuspended in 100 μ L Tris-EDTA buffer (10mM TrisHCl, 1mM EDTA) with 1 μ L RNase (Omega Bio-Tek, USA). 200 μ L fresh 0.1% SDS, 1M NaOH was added and samples were incubated on ice for 5 minutes. 150 μ L 3M potassium acetate was added and samples were chilled another 5min. Samples were then spun for 15min at 14,000RPM in a benchtop centrifuge and the supernatant was transferred to a fresh tube. 1mL cold (-20°C) ethanol was added and DNA was precipitated for 30min. Samples were spun again, washed twice with 70% ethanol then spun a final time. The supernatant was removed, samples were dried for 15min at 37°C then resuspended in 20 μ L nuclease-free water. Purified plasmids were run on a 1% agarose gel and those with a band shift of 1Kbp more than pG19II were sequenced using M13F and M13R primers (Eurofins, USA). Plasmid sequences were compared to the expected sequence in Benchling and those with the correct sequence were used to generate the deletion mutants.

D.1.3. Conjugation

To delete the genes from the *P. pseudoalcaligenes* KF707 genome, the deletion construct containing plasmids were introduced by conjugation. Cultures of *E. coli* HB101 carrying the helper plasmid pRK2013 (5) and *E. coli* Top10F' carrying the deletion construct in pG19II were grown to early log-phase (OD ~0.3) along with the KF707 wild-type, or later, deletion mutants. 2mL of donor and recipient and 1mL of helper were spun down, resuspended in 100 μ L LB, mixed, then spun down to 100 μ L. This mixture was spotted on LB plates and incubated overnight at 30°C. The conjugation smear was then collected into 0.9% saline, serially diluted and plated on AB glucose plates (5g/L glucose, minimal salts media) containing 20 μ g/mL gentamycin. Plates were incubated for 48h at 30°C.

D.1.4. Counter Selection

Transconjugant colonies were picked off the AB glucose Gm plates into 5mL LB no salt, LB 10% sucrose and LB + 20µg/mL Gm. Colonies that were able to grow with Gm but not (at all) with sucrose were selected for continued use. From the LB no salt overnight culture 50µL was inoculated into 5mL LB 10% sucrose. After exactly 4h, 1mL was spun down, resuspended in 100µL LB and plated on LB 10% sucrose.

D.1.5. Identification of Deletion Mutants

After overnight growth at 30°C colonies were picked for colony PCR. Briefly, colonies were picked onto an index plate and into 20µL lysis buffer (0.01% Triton X-100, 0.1% SDS, 0.1mM TrisHCl, 0.002mM EDTA). After boiling for 5minutes, samples were spun down and 1µL was added to the 24µL PCR mix (in µL: Taq Buffer, 25; 1.25mM dNTPs, 40; nuclease-free water, 120; MgCl₂, 7.5; DMSO, 20; Taq Polymerase (Invitrogen, USA), 1). PCR parameters: 95°C, 3min; 30 cycles of: 95°C, 40s; 60°C, 40s; 72°C, 120s, then final extension for 10min at 60°C. Samples were run on a 1% agarose gel and colonies that showed ONLY a band at ~1Kbp were sent for sequencing (Eurofins, USA) to confirm the deletion. Confirmed deletions were grown overnight in LB then mixed 2:1 with LB 24% DMSO to make freezer stocks.

D.1.6. Complementation

To generate a complementation construct of Aer.g1, an additional primer was produced annealing with the start codon directly in the centre. A HinDIII site was added to this primer and the opposite outer primer from the deletion construct was used to amplify the Aer.g1 gene. The PCR mixture contained in µL HF buffer, 2; 1.25mM dNTPS, 2; HF enzyme, 0.2; gDNA, 0.1; 50mM MgCl₂, 1.6; forward primer, 2; reverse primer, 2; PCR enhancer (Invitrogen, USA), 2; nuclease-free water to 20µL. PCR parameters: 95°C, 5min; 30 cycles of: 95°C, 40s; 60°C, 45s; 72°C, 240s, then final extension for 10min at 60°C. A pre-existing BamHI site 36nt from the end

of the gene was taken advantage of when the gel purified (EZDNA gel extraction kit, Omega Bio-Tek, USA) product was digested alongside pSEVA_342 overnight at 37°C, (3μL Buffer K, 1μL BamHI, 1μL HinDIII, 1μg DNA, NF water to 30μL). Digested products were purified, ligated and transformed into Top10F' the same as the deletion constructs. After confirming the insert by sequencing (Eurofins, USA) the complementation and empty vector were transformed into the wildtype and quintuple deletion mutant as before, only 30μg/mL chloramphenicol was used to select for transformants. Resistant colonies of *P. pseudoalcaligenes* KF707 were picked into LB to make stocks. For swim plate assays, normal plates and plates containing 100mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) were used.

Table A-D-1: Strains used for energy-taxis experiments in *Pseudomonas pseudoalcaligenes* KF707.

Bacterial Strain	Description	Reference
<i>P. pseudoalcaligenes</i>		
KF707	Wild-type	(6)
<i>cheA::Km</i>	<i>cheA::km</i> , Km ^R	(7)
Δ <i>aer-2</i>		This Study
Δ <i>cttP</i>		This Study
Δ <i>aer.g1</i>		This Study
Δ <i>aer.g2</i>		This Study
Δ <i>aer.g4</i>		This Study
Δ <i>aer-2/\Delta</i> <i>cttP</i>		This Study
Δ <i>aer-2/\Delta</i> <i>aer.g1</i>		This Study
Δ <i>aer-2/\Delta</i> <i>aer.g2</i>		This Study
Δ <i>aer-2/\Delta</i> <i>aer.g4</i>		This Study
Δ <i>cttP/\Delta</i> <i>aer.g1</i>		This Study
Δ <i>cttP/\Delta</i> <i>aer.g2</i>		This Study
Δ <i>cttP/\Delta</i> <i>aer.g4</i>		This Study
Δ <i>aer.g1/\Delta</i> <i>aer.g2</i>		This Study
Δ <i>aer.g1/\Delta</i> <i>aer.g4</i>		This Study
Δ <i>aer.g2/\Delta</i> <i>aer.g4</i>		This Study

$\Delta aer-2/\Delta aer.g1/\Delta aer.g2$		This Study
$\Delta aer-2/\Delta aer.g1/\Delta aer.g4$		This Study
$\Delta aer-2/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta aer.g1/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta cttP/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta aer-2/\Delta aer.g1/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta cttP/\Delta aer.g1/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta aer-2/\Delta cttP/\Delta aer.g2/\Delta aer.g4$		This Study
$\Delta aer-2/\Delta cttP/\Delta aer.g1/\Delta aer.g2/\Delta aer.g4$		This Study
<i>E. coli</i>		
DH5 α	<i>supE44hsdR17recA1 gyrA96thi1relA1</i>	(8)
Top10F'	F' { <i>lacIq</i> , Tn10(TetR)} <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80/ <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1 araD139</i> Δ (<i>ara leu</i>) 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen™
HB101	<i>recA thi pro leu hsdR</i> , Sm ^R	(9)

Table A-D-2: Plasmids used for energy-taxis experiments in *Pseudomonas pseudoalcaligenes* KF707.

Plasmid	Description	Reference
pRK2013	KmR ori <i>colE1</i> RK2-Mob ⁺ RK2-Tra ⁺	(5)
pG19II	Gm ^R , <i>sacB</i> , <i>lacZ</i> , cloning vector, conjugative plasmid	(4)
pG19II- $\Delta aer-2$	Gm ^R , <i>sacB</i> , <i>lacZ</i> , <i>aer-2</i> deletion construct	This Study
pG19II- $\Delta cttP$	Gm ^R , <i>sacB</i> , <i>lacZ</i> , <i>cttP</i> deletion construct	This Study
pG19II- $\Delta aer.g1$	Gm ^R , <i>sacB</i> , <i>lacZ</i> , <i>aer.g1</i> deletion construct	This Study
pG19II- $\Delta aer.g2$	Gm ^R , <i>sacB</i> , <i>lacZ</i> , <i>aer.g2</i> deletion construct	This Study
pG19II- $\Delta aer.g4$	Gm ^R , <i>sacB</i> , <i>lacZ</i> , <i>aer.g4</i> deletion construct	This Study
pSEVA324	Cm ^R , pR01600/ <i>ColE1</i> , <i>lacZ</i> α -pUC19	(10)
pSEVA324- <i>aer.g1</i>	Cm ^R , pSEVA342 with <i>aer.g1</i> in the MCS	This Study

Table A-D-3: Primers used to amplify upstream and downstream regions to generate deletion constructs for deletion of Aer.g1, Aer.g2, Aer.g4, Aer-2 and CttP from *Pseudomonas pseudoalcaligenes* KF707. Restriction enzyme sites are in bold, annealing regions underlined.

	Outer Forward (HindIII)	Inner Reverse	Inner Forward	Outer Reverse (BamHI)
Aer-2	ATATA AAGCTT <u>TTCC</u> <u>AGCGGGTAGATGC</u> C	CGCAGCCTGACAC GGAAG <u>AGGAATGGGAAGATTCT</u> <u>GAGG</u>	CCTCAGAACTCTTCCCATT CTCTCCGTGTCAGGCTGC <u>G</u>	ATAT GGATCC <u>GAG</u> <u>TACCTGACCTTCAC</u> <u>CCTC</u>
CttP	ATATA AAGCTT <u>CGAA</u> <u>GAACACCTGGAGG</u> <u>AACTG</u>	GACGATCACTGGCGATTTG ACAAC GCTTCTACACCTAA <u>G</u> <u>GCATGG</u>	CCATGCTTAGGTGTAGAA GCGTT G TCAAATCGCCAGT <u>GATCGTC</u>	ATAT GGATCC <u>GAG</u> <u>TGGCTACGCGAAT</u> <u>GGAG</u>
Aer.g2	ATATA AAGCTT <u>CGGT</u> <u>GATGAGTCCCTACG</u> <u>AC</u>	CTACGGCTTGTACGCACCA <u>CTGAAATAAAGGTCTAGCA</u> <u>CGCTG</u>	CAGCGTGCTAGACCTTTAT TTCAGTGGTGCCTACAAGC <u>CGTAG</u>	ATAT GGATCC <u>GCT</u> <u>TCGAGTGTGGACA</u> <u>AGG</u>
Aer.g1	ATATA AAGCTT <u>TTIGC</u> <u>AGAAGACCCTCACG</u> C	CTGGATGCCGATAGATGCT GTA ACTGGATCAGCCCAGG <u>ATCAG</u>	CTGATCCTGGGCTGATCCA <u>GTTACAGCATCTATCGGCA</u> <u>TCCAG</u>	ATAT GGATCC <u>GCA</u> <u>AGACCCTCAACCT</u> <u>CACAG</u>
Aer.g4	ATATA AAGCTT <u>GCTG</u> <u>GTGGTGACGAAGA</u> <u>AGAAC</u>	CAAGAACAACAACAGGGA <u>AGCACGGTGGGAAGGAGGC</u> <u>CGTTAGC</u>	GCTAACGGCCTCCTTCCAC <u>CGTGCTCCCTGTTGTTGT</u> <u>TCTTG</u>	ATAT GGATCC <u>AGA</u> <u>TCACCGAAAGCCT</u> <u>GCTG</u>

Table A-D-4: Primers used to amplify Aer.g1 for complementation. PCR amplification product is 2228nt. After digestion with BamHI and HindIII a 1623nt product is produced that contains only the Aer.g1 gene. Restriction enzyme sites are in bold, annealing regions underlined.

	Outer FW (HindIII)	Outer Reverse for Complementation (HindIII)
Aer.g1	ATATAAGCTTTTGCAGAAGACCC	ATATAAGCTTGACACCATGAGAAACAAC
Complementation	<u>TCACGC</u>	<u>CAACC</u>

D.2. References

1. **Triscari-Barberi T, Simone D, Calabrese FM, Attimonelli M, Hahn KR, Amoako KK, Turner RJ, Fedi S, Zannoni D.** 2012. Genome sequence of the polychlorinated-biphenyl degrader pseudomonas pseudoalcaligenes KF707. *J Bacteriol* **194**:4426–4427.
2. **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL.** 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**:134.
3. **Sambrook J, Fritsch EF, Maniatis T.** 1989. No Title. *Mol Biol A Lab Man.*
4. **Maseda H, Sawada I, Saito K, Uchiyama H, Nakae T, Nomura N.** 2004. Enhancement of the mexAB-oprM efflux pump expression by a quorum-sensing autoinducer and its cancellation by a regulator, MexT, of the mexEF-oprN efflux pump operon in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **48**:1320–8.
5. **Figurski DH, Helinski DR.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**:1648–52.
6. **Furukawa K, Miyazaki T.** 1986. Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. *J Bacteriol* **166**:392–398.
7. **Tremaroli V, Fedi S, Tamburini S, Viti C, Tatti E, Ceri H, Turner RJ, Zannoni D.** 2011. A histidine-kinase cheA gene of *Pseudomonas pseudoalcaligenes* KF707 not only has a key role in chemotaxis but also affects biofilm formation and cell metabolism. *Biofouling* **27**:33–46.
8. **Hanahan D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**:557–580.
9. **Boyer HW, Roulland-dussoix D.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* **41**:459–472.
10. **Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, Páez-Espino AD, Durante-Rodríguez G, Kim J, Nickel PI, Platero R, De Lorenzo V.** 2013. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* **41**:D666–D675.

APPENDIX E: COPYRIGHT PERMISSIONS

Chapter Two, Appendix One, published in *Metallomics*.

“Author reusing their own work published by the Royal Society of Chemistry

You do not need to request permission to reuse your own figures, diagrams, etc, that were originally published in a Royal Society of Chemistry publication. However, permission should be requested for use of the whole article or chapter **except if reusing it in a thesis**. If you are including an article or book chapter published by us in your thesis please ensure that your co-authors are aware of this.” (emphasis added)

All co-authors were notified of the use of these articles in my thesis.

Appendix Two, published in *Computational and Structural Biology Journal*.

“Copyright Notice

Copyright for articles published in this journal is retained by the authors, with first publication rights granted to the *Computational and Structural Biotechnology Journal*. All journal content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license. This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. If the original authors and source are cited, **No permission is required from the authors or the publishers** for anyone to download, reuse, reprint, modify, distribute, and/or copy articles in this journal.”

Chapter Three, published in *Frontiers in Microbiology*

“Copyright Statement

Under the Frontiers Conditions for Website Use and the Frontiers General Conditions for Authors, authors of articles published in Frontiers journals retain copyright on their articles, except for any third-party images and other materials added by Frontiers, which are subject to

copyright of their respective owners. Authors are therefore free to disseminate and re-publish their articles, subject to any requirements of third-party copyright owners and subject to the original publication being fully cited. Visitors may also download and forward articles subject to the citation requirements and subject to any fees Frontiers may charge for downloading licenses. The ability to copy, download, forward or otherwise distribute any materials is always subject to any copyright notices displayed. Copyright notices must be displayed prominently and may not be obliterated, deleted or hidden, totally or partially.”

Appendix Three, published in *eMagRes*

From the Wiley-Blackwell Copyright Transfer Agreement

“C. PERMITTED USES BY CONTRIBUTOR

1. Submitted Version.

Wiley-Blackwell licenses back the following rights to the Contributor in the version of the Contribution as originally submitted for publication:

a. After publication of the final article, the right to self-archive on the Contributor’s personal intranet page or in the Contributor’s institution’s/ employer’s institutional intranet repository or archive. The Contributor may not update the submission version or replace it with the published Contribution. The version posted must contain a legend as follows: This is the pre-peer reviewed version of the following article: FULL CITE, which has been published in final form at [Link to final article].

b. The right to transmit, print and share copies with colleagues.

...

3. Final Published Version.

Wiley-Blackwell hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution:

a. Copies for colleagues. The personal right of the Contributor only to send or transmit individual copies of the final published version to colleagues upon their specific request provided no fee is charged, and further-provided that there is no systematic distribution of the Contribution, e.g. posting on a listserve, website or automated delivery. For those Contributors who wish to send high-quality e-prints, purchase reprints, or who wish to distribute copies more broadly than allowed hereunder (e.g. to groups of colleagues or mailing lists), please contact the publishing office.

b. Re-use in other publications. The right to re-use the final Contribution or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications should be accurately noted.”

Dear Sean Booth

Thank you for your request.

Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements (author, title of material, title of book/journal, ourselves as publisher) and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially. You must also duplicate the copyright notice that appears in the Wiley publication in your use of the Material; this can be found on the copyright page if the material is a book or within the article if it is a journal.

Permission is granted solely for use in conjunction with the thesis, and the material may not be posted online separately.

Any third party material is expressly excluded from this permission. If any of the material you wish to use appears within our work with credit to another source, authorisation from that source must be obtained.

Best wishes,

Aimee Masheter

Permissions Coordinator

John Wiley & Sons Ltd

The Atrium

Southern Gate, Chichester

West Sussex, PO19 8SQ

UK

References

1. **Zalasiewicz J, Williams M, Haywood A, Ellis M.** 2011. The Anthropocene: a new epoch of geological time? *Philos Trans A Math Phys Eng Sci* **369**:835–841.
2. **Darnault C, Rockne K, Stevens A, Mansoori G, Sturchio N.** 2005. Fate of environmental pollutants. *Water Environ Res* **77**:2576–2658.
3. **Erickson MD, Kaley RG.** 2011. Applications of polychlorinated biphenyls. *Environ Sci Pollut Res* **18**:135–151.
4. **Schell LM, Burnitz KK, Lathrop PW.** 2010. Pollution and human biology. *Ann Hum Biol* **37**:347–366.
5. **Dirinck E, Jorens PG, Covaci A, Geens T, Roosens L, Neels H, Mertens I, Van Gaal L.** 2011. Obesity and persistent organic pollutants: possible obesogenic effect of organochlorine pesticides and polychlorinated biphenyls. *Obesity (Silver Spring)* **19**:709–714.
6. **ATSDR.** 2011. Toxicological profile for polychlorinated biphenyls.
7. **Aoki Y.** 2001. Polychlorinated biphenyls, polychlorinated dibenzo-p-dioxins, and polychlorinated dibenzofurans as endocrine disrupters--what we have learned from Yusho disease. *Environ Res* **86**:2–11.
8. **Mayes BA, McConnell EE, Neal BH, Brunner MJ, Hamilton SB, Sullivan TM, Peters AC, Ryan MJ, Toft JD, Singer AW, Brown JF, Menton RG, Moore JA.** 1998. Comparative carcinogenicity in Sprague-Dawley rats of the polychlorinated biphenyl mixtures Aroclors 1016, 1242, 1254, and 1260. *Toxicol Sci* **41**:62–76.
9. **Malik A.** 2004. Metal bioremediation through growing cells. *Environ Int.*
10. **Gadd GM, White C.** 1993. Microbial treatment of metal pollution - a working

- biotechnology? *Trends Biotechnol* **11**:353–359.
11. **Shukla KP, Singh NK, Sharma S.** 2010. Bioremediation: Developments, current practices and perspectives. *Genet Eng Biotechnol J* **2010**:1–20.
 12. **Megharaj M, Ramakrishnan B, Venkateswarlu K, Sethunathan N, Naidu R.** 2011. Bioremediation approaches for organic pollutants: A critical perspective. *Environ Int* **37**:1362–1375.
 13. **Sinha S, Chattopadhyay P, Pan I, Chatterjee S, Chanda P, Bandyopadhyay D, Das K, Sen SK.** 2009. Microbial transformation of xenobiotics for environmental bioremediation. *African J Biotechnol* **8**:6016–6027.
 14. **Lemire JA, Harrison JJ, Turner RJ.** 2013. Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nat Rev Microbiol* **11**:371–84.
 15. **Sandrin TR, Hoffman DR.** 2007. Bioremediation of organic and metal co-contaminated environments: Effects of metal toxicity, speciation, and bioavailability on biodegradation, p. 1–34. *In* Singh, S, Tripathi, R (eds.), *Environmental Bioremediation Technologies*. Springer Berlin Heidelberg.
 16. **Sandrin TR, Maier RM.** 2003. Impact of metals on the biodegradation of organic pollutants. *Environ Health Perspect* **111**:1093–1101.
 17. **Olaniran AO, Balgobind A, Pillay B.** 2013. Bioavailability of heavy metals in soil: Impact on microbial biodegradation of organic compounds and possible improvement strategies. *Int J Mol Sci* **14**:10197–10228.
 18. **Liu R, Jiang H, Xu P, Qiao C, Zhou Q, Yang C.** 2014. Engineering chlorpyrifos-degrading *Stenotrophomonas* sp. YC-1 for heavy metal accumulation and enhanced chlorpyrifos degradation. *Biodegradation* **25**:903–10.

19. **Ontañon OM, González PS, Agostini E.** 2015. Optimization of simultaneous removal of Cr (VI) and phenol by a native bacterial consortium: Its use for bioaugmentation of co-polluted effluents. *J Appl Microbiol* **119**:1011–1022.
20. **Bhattacharya A, Gupta A, Kaur A, Malik D.** 2014. Efficacy of *Acinetobacter* sp. B9 for simultaneous removal of phenol and hexavalent chromium from co-contaminated system. *Appl Microbiol Biotechnol* **98**:9829–9841.
21. **Weyens N, Beckers B, Schellingen K, Ceulemans R, van der Lelie D, Newman L, Taghavi S, Carleer R, Vangronsveld J.** 2015. The Potential of the Ni-Resistant TCE-Degrading *Pseudomonas putida* W619-TCE to Reduce Phytotoxicity and Improve Phytoremediation Efficiency of Poplar Cuttings on A Ni-TCE Co-Contamination. *Int J Phytoremediation* **17**:40–48.
22. **Arjoon A, Olaniran AO, Pillay B.** 2015. Kinetics of heavy metal inhibition of 1,2-dichloroethane biodegradation in co-contaminated water. *J Basic Microbiol* **55**:277–284.
23. **Arjoon A, Olaniran AO, Pillay B.** 2013. Enhanced 1,2-dichloroethane degradation in heavy metal co-contaminated wastewater undergoing biostimulation and bioaugmentation. *Chemosphere* **93**:1826–1834.
24. **Thavamani P, Megharaj M, Naidu R.** 2015. Metal-tolerant PAH-degrading bacteria: development of suitable test medium and effect of cadmium and its availability on PAH biodegradation. *Environ Sci Pollut Res* **22**:8957–8968.
25. **Obuekwe IS, Semple KT.** 2013. Impact of Zn and Cu on the development of phenanthrene catabolism in soil. *J Soils Sediments* **13**:10039–47.
26. **Oyetibo GO, Ilori MO, Obayori OS, Amund OO.** 2013. Biodegradation of petroleum hydrocarbons in the presence of nickel and cobalt. *J Basic Microbiol* **53**:917–927.

27. **Pornwongthong P, Mulchandani A, Gedalanga PB, Mahendra S.** 2014. Transition metals and organic ligands influence biodegradation of 1,4-dioxane. *Appl Biochem Biotechnol* **173**:291–306.
28. **Imlay JA.** 2003. Pathways of oxidative damage. *Annu Rev Microbiol* **57**:395–418.
29. **Stohs SJ, Bagchi D.** 1995. Oxidative mechanisms in the toxicity of metal ions. *Free Radic Biol Med* **18**:321–336.
30. **Stulke J, Hillen W.** 1999. Carbon catabolite repression in bacteria. *Curr Opin Microbiol* **2**:195–201.
31. **Macomber L, Elsey SP, Hausinger RP.** 2011. Fructose-1,6-bisphosphate aldolase (class II) is the primary site of nickel toxicity in *Escherichia coli*. *Mol Microbiol* **82**:1291–1300.
32. **Bogachev A V., Verkhovsky MI.** 2005. Na⁺-translocating NADH: Quinone oxidoreductase: Progress achieved and prospects of investigations. *Biochem* **70**:143–149.
33. **Fadeeva MS, Bertsova Y V, Euro L, Bogachev a V.** 2011. Cys377 residue in NqrF subunit confers Ag(+) sensitivity of Na⁺-translocating NADH:quinone oxidoreductase from *Vibrio harveyi*. *Biochemistry (Mosc)* **76**:186–195.
34. **Dibrov P, Dzioba J, Gosink KK, Häse CC, Ha CC.** 2002. Chemiosmotic Mechanism of Antimicrobial Activity of Ag⁺ in *Vibrio cholerae* Chemiosmotic Mechanism of Antimicrobial Activity of Ag² in *Vibrio cholerae*. *Society* **46**:8–11.
35. **Xu FF, Imlay JA.** 2012. Silver(I), mercury(II), cadmium(II), and zinc(II) target exposed enzymic iron-sulfur clusters when they toxify *Escherichia coli*. *Appl Environ Microbiol* **78**:3614–3621.
36. **Calderón IL, Elías AO, Fuentes EL, Pradenas GA, Castro ME, Arenas FA, Pérez JM, Vásquez CC.** 2009. Tellurite-mediated disabling of [4Fe-4S] clusters of *Escherichia coli*

- dehydratases. *Microbiology* **155**:1840–1846.
37. **Macomber L, Imlay JA.** 2009. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc Natl Acad Sci U S A* **106**:8344–8349.
 38. **Middaugh J, Hamel R, Jean-Baptiste G, Beriault R, Chenier D, Appanna VD.** 2005. Aluminum triggers decreased aconitase activity via Fe-S cluster disruption and the overexpression of isocitrate dehydrogenase and isocitrate lyase: A metabolic network mediating cellular survival. *J Biol Chem* **280**:3159–3165.
 39. **Ogunseitan OA, Yang S, Ericson J.** 2000. Microbial α -aminolevulinic acid dehydratase as a biosensor of lead bioavailability in contaminated environments. *Soil Biol Biochem* **32**:1899–1906.
 40. **Erskine PT, Senior N, Awan S, Lambert R, Lewis G, Tickle IJ, Sarwar M, Spencer P, Thomas P, Warren MJ, Shoolingin-Jordan PM, Wood SP, Cooper JB.** 1997. X-ray structure of 5-aminolaevulinic acid dehydratase, a hybrid aldolase. *Nat Struct Biol* **4**:1025–1031.
 41. **Hong R, Kang TY, Michels CA, Gadura N.** 2012. Membrane lipid peroxidation in copper alloy-mediated contact killing of *Escherichia coli*. *Appl Environ Microbiol* **78**:1776–1784.
 42. **Li WR, Xie XB, Shi QS, Zeng HY, Ou-Yang YS, Chen Y Ben.** 2010. Antibacterial activity and mechanism of silver nanoparticles on *Escherichia coli*. *Appl Microbiol Biotechnol* **85**:1115–1122.
 43. **Santo CE, Quaranta D, Grass G, Esp C, Quaranta D, Grass G.** 2012. Antimicrobial metallic copper surfaces kill *Staphylococcus haemolyticus* via membrane damage. *Microbiologyopen* **1**:46–52.
 44. **Janero DR.** 1990. Malondialdehyde and thiobarbituric acid-reactivity as diagnostic indices

- of lipid peroxidation and peroxidative tissue injury. *Free Radic Biol Med* **9**:515–540.
45. **Warnes SL, Caves V, Keevil CW.** 2012. Mechanism of copper surface toxicity in *Escherichia coli* O157:H7 and *Salmonella* involves immediate membrane depolarization followed by slower rate of DNA destruction which differs from that observed for Gram-positive bacteria. *Environ Microbiol* **14**:1730–1743.
 46. **Xu H, Qu F, Xu H, Lai W, Wang YA, Aguilar ZP, Wei H.** 2012. Role of reactive oxygen species in the antibacterial mechanism of silver nanoparticles on *Escherichia coli* O157:H7. *BioMetals* **25**:45–53.
 47. **Oves M, Khan MS, Zaidi A, Ahmed AS, Ahmed F, Ahmad E, Sherwani A, Owais M, Azam A.** 2013. Antibacterial and Cytotoxic Efficacy of Extracellular Silver Nanoparticles Biofabricated from Chromium Reducing Novel OS4 Strain of *Stenotrophomonas maltophilia*. *PLoS One* **8**.
 48. **Macomber L, Hausinger RP.** 2011. Mechanisms of nickel toxicity in microorganisms. *Metallomics* **3**:1153.
 49. **Chen YP, Dilworth MJ, Glenn AR.** 1984. Aromatic metabolism in *Rhizobium trifolii* - protocatechuate 3,4-dioxygenase. *Arch Microbiol* **138**:187–190.
 50. **Chen YP, Glenn AR, Dilworth MJ.** 1985. Aromatic metabolism in *Rhizobium trifolii* - catechol 1,2-dioxygenase. *Arch Microbiol* **141**:225–228.
 51. **Guzik U, Hupert-Kocurek K, Salek K, Wojcieszńska D.** 2013. Influence of metal ions on bioremediation activity of protocatechuate 3,4-dioxygenase from *Stenotrophomonas maltophilia* KB2. *World J Microbiol Biotechnol* **29**:267–273.
 52. **Helbig K, Grosse C, Nies DH.** 2008. Cadmium toxicity in glutathione mutants of *Escherichia coli*. *J Bacteriol* **190**:5439–5454.

53. **Kaneko Y, Thoendel M, Olakanmi O, Britigan BE, Singh PK.** 2007. The transition metal gallium disrupts *Pseudomonas aeruginosa* iron metabolism and has antimicrobial and antibiofilm activity. *J Clin Invest* **117**:877–888.
54. **Barras F, Fontecave M.** 2011. Cobalt stress in *Escherichia coli* and *Salmonella enterica*: molecular bases for toxicity and resistance. *Metallomics* **3**:1130.
55. **Ratledge C, Dover LG.** 2000. Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* **54**:881–941.
56. **Djoko KY, Mcewan AG.** 2013. Antimicrobial Action of Copper Is Amplified via Inhibition of Heme Biosynthesis.
57. **Valko M, Morris H, Cronin MTD.** 2005. Metals, Toxicity and Oxidative Stress. *Curr Top Med Chem* **12**:1161–1208.
58. **Andrews SC, Robinson AK, Rodríguez-Quñones F.** 2003. Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**:215–237.
59. **Pérez JM, Calderón IL, Arenas FA, Fuentes DE, Pradenas GA, Fuentes EL, Sandoval JM, Castro ME, Elías AO, Vásquez CC.** 2007. Bacterial toxicity of potassium tellurite: Unveiling an ancient enigma. *PLoS One* **2**.
60. **Strlič M, Kolar J, Šelih VS, Kočar D, Pihlar B.** 2003. A comparative study of several transition metals in fenton-like reaction systems at circum-neutral pH. *Acta Chim Slov* **50**:619–632.
61. **Macomber L, Rensing C, Imlay JA.** 2007. Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. *J Bacteriol* **189**:1616–1626.
62. **Nies DH.** 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* **27**:313–339.

63. **Asakura K, Satoh H, Chiba M, Okamoto M, Serizawa K, Nakano M, Omae K.** 2009. Genotoxicity studies of heavy metals: Lead, bismuth, indium, silver and antimony. *J Occup Health* **51**:498–512.
64. **Helbig K, Bleuel C, Krauss GJ, Nies DH.** 2008. Glutathione and transition-metal homeostasis in *Escherichia coli*. *J Bacteriol* **190**:5431–5438.
65. **Harrison JJ, Tremaroli V, Stan MA, Chan CS, Vacchi-Suzzi C, Heyne BJ, Parsek MR, Ceri H, Turner RJ.** 2009. Chromosomal antioxidant genes have metal ion-specific roles as determinants of bacterial metal tolerance. *Environ Microbiol* **11**:2491–2509.
66. **Furukawa K, Miyazaki T.** 1986. Cloning of a gene cluster encoding biphenyl and chlorobiphenyl degradation in *Pseudomonas pseudoalcaligenes*. *J Bacteriol* **166**:392–398.
67. **Martínez P, Agulló L, Hernández M, Seeger M.** 2007. Chlorobenzoate inhibits growth and induces stress proteins in the PCB-degrading bacterium *Burkholderia xenovorans* LB400. *Arch Microbiol* **188**:289–297.
68. **Ohtsubo Y, Kudo T, Tsuda M, Nagata Y.** 2004. Strategies for bioremediation of polychlorinated biphenyls. *Appl Microbiol Biotechnol* **65**:250–258.
69. **Pieper DH, Seeger M.** 2008. Bacterial metabolism of polychlorinated biphenyls. *J Mol Microbiol Biotechnol* **15**:121–138.
70. **Murphy CD, Quirke S, Balogun O.** 2008. Degradation of fluorobiphenyl by *Pseudomonas pseudoalcaligenes* KF707. *FEMS Microbiol Lett* **286**:45–49.
71. **Tremaroli V, Fedi S, Zannoni D.** 2007. Evidence for a tellurite-dependent generation of reactive oxygen species and absence of a tellurite-mediated adaptive response to oxidative stress in cells of *Pseudomonas pseudoalcaligenes* KF707. *Arch Microbiol* **187**:127–135.
72. **Di Tomaso G, Fedi S, Carnevali M, Manegatti M, Taddei C, Zannoni D.** 2002. The

- membrane-bound respiratory chain of *Pseudomonas pseudoalcaligenes* KF707 cells grown in the presence or absence of potassium tellurite. *Microbiology* **148**:1699–1708.
73. **Zanaroli G, Fedi S, Carnevali M, Fava F, Zannoni D.** 2002. Use of potassium tellurite for testing the survival and viability of *Pseudomonas pseudoalcaligenes* KF707 in soil microcosms contaminated with polychlorinated biphenyls. *Res Microbiol* **153**:353–360.
 74. **Jain A, Gupta Y, Agrawal R, Khare P, Jain SK.** 2007. Biofilms--a microbial life perspective: a critical review. *Critical reviews in therapeutic drug carrier systems*.
 75. **Tremaroli V, Fedi S, Turner RJ, Ceri H, Zannoni D.** 2008. *Pseudomonas pseudoalcaligenes* KF707 upon biofilm formation on a polystyrene surface acquire a strong antibiotic resistance with minor changes in their tolerance to metal cations and metalloids. *Arch Microbiol* **190**:29–39.
 76. **Tremaroli V, Workentine ML, Weljie AM, Vogel HJ, Ceri H, Viti C, Tatti E, Zhang P, Hynes AP, Turner RJ, Zannoni D.** 2009. Metabolomic investigation of the bacterial response to a metal challenge. *Appl Environ Microbiol* **75**:719–728.
 77. **Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, Ceri H, Turner RJ, Weljie AM.** 2011. Differences in metabolism between the biofilm and planktonic response to metal stress. *J Proteome Res* **10**:3190–3199.
 78. **Tremaroli V, Suzzi CV, Fedi S, Ceri H, Zannoni D, Turner RJ.** 2010. Tolerance of *Pseudomonas pseudoalcaligenes* KF707 to metals, polychlorobiphenyls and chlorobenzoates: Effects on chemotaxis-, biofilm- and planktonic-grown cells. *FEMS Microbiol Ecol* **74**:291–301.
 79. **Tremaroli V, Fedi S, Tamburini S, Viti C, Tatti E, Ceri H, Turner RJ, Zannoni D.** 2011. A histidine-kinase *cheA* gene of *Pseudomonas pseudoalcaligenes* KF707 not only has

- a key role in chemotaxis but also affects biofilm formation and cell metabolism. *Biofouling* **27**:33–46.
80. **Wadhams GH, Armitage JP.** 2004. Making sense of it all: Bacterial chemotaxis. *Nat Rev Microbiol* **5**:1024–1037.
 81. **Sampedro I, Parales RE, Krell T, Hill JE.** 2015. Pseudomonas chemotaxis. *FEMS Microbiol Rev* **39**:17–46.
 82. **Ortega DR, Zhulin IB.** 2016. Evolutionary Genomics Suggests That CheV Is an Additional Adaptor for Accommodating Specific Chemoreceptors within the Chemotaxis Signaling Complex. *PLoS Comput Biol* **12**:e1004723.
 83. **Rico-Jiménez M, Reyes-Darias JA, Ortega Á, Díez Peña AI, Morel B, Krell T.** 2016. Two different mechanisms mediate chemotaxis to inorganic phosphate in *Pseudomonas aeruginosa*. *Sci Rep* **6**:28967.
 84. **Wu H, Kato J, Kuroda A, Ikeda T.** 2000. Identification and Characterization of Two Chemotactic Transducers for Inorganic Phosphate in *Pseudomonas aeruginosa* Identification and Characterization of Two Chemotactic Transducers for Inorganic Phosphate in *Pseudomonas aeruginosa*.
 85. **Vangnai AS, Takeuchi K, Oku S, Kataoka N, Nitisakulkan T, Tajima T, Kato J.** 2013. Identification of CtpL as a chromosomally encoded chemoreceptor for 4-chloroaniline and catechol in *Pseudomonas aeruginosa* PAO1. *Appl Environ Microbiol* **79**:7241–8.
 86. **Hong CS, Shitashiro M, Kuroda A, Ikeda T, Takiguchi N, Ohtake H, Kato J.** 2004. Chemotaxis proteins and transducers for aerotaxis in *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* **231**:247–252.
 87. **Airola M V, Huh D, Sukomon N, Widom J, Sircar R, Borbat PP, Freed JH, Watts KJ,**

- Crane BR.** 2013. Architecture of the soluble receptor Aer2 indicates an in-line mechanism for PAS and HAMP domain signaling. *J Mol Biol* **425**:886–901.
88. **Nichols NN, Harwood CS.** 2000. An aerotaxis transducer gene from *Pseudomonas putida*. *FEMS Microbiol Lett* **182**:177–183.
89. **Sarand I, Österberg S, Holmqvist S, Holmfeldt P, Skärfstad E, Parales RE, Shingler V.** 2008. Metabolism-dependent taxis towards (methyl)phenols is coupled through the most abundant of three polar localized Aer-like proteins of *Pseudomonas putida*. *Environ Microbiol* **10**:1320–1334.
90. **Luu R a., Schneider BJ, Ho CC, Nesteryuk V, Ngwesse SE, Liu X, Parales J V., Ditty JL, Parales RE.** 2013. Taxis of *Pseudomonas putida* F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. *Appl Environ Microbiol* **79**:2416–2423.
91. **García V, Reyes-Darias J-A, Martín-Mora D, Morel B, Matilla MA, Krell T.** 2015. Identification of a Chemoreceptor for C2 and C3 Carboxylic Acids. *Appl Environ Microbiol* **81**:5449–57.
92. **Alvarez-Ortega C, Harwood CS.** 2007. Identification of a malate chemoreceptor in *Pseudomonas aeruginosa* by screening for chemotaxis defects in an energy taxis-deficient mutant. *Appl Environ Microbiol* **73**:7793–7795.
93. **Martín-Mora D, Reyes-Darias J-A, Ortega Á, Corral-Lugo A, Matilla MA, Krell T.** 2016. McpQ is a specific citrate chemoreceptor that responds preferentially to citrate/metal ion complexes. *Environ Microbiol* **18**:3284–3295.
94. **Parales RE, Luu R a., Chen GY, Liu X, Wu V, Lin P, Hughes JG, Nesteryuk V, Parales J V., Ditty JL.** 2013. *Pseudomonas putida* F1 has multiple chemoreceptors with overlapping specificity for organic acids. *Microbiol (United Kingdom)* **159**:1086–1096.

95. **Lacal J, Alfonso C, Liu X, Parales RE, Morel B, Conejero-Lara F, Rivas G, Duque E, Ramos JL, Krell T.** 2010. Identification of a chemoreceptor for tricarboxylic acid cycle intermediates: Differential chemotactic response towards receptor ligands. *J Biol Chem* **285**:23126–23136.
96. **Oku S, Komatsu A, Nakashimada Y, Tajima T, Kato J.** 2014. Identification of *Pseudomonas fluorescens* chemotaxis sensory proteins for malate, succinate, and fumarate, and their involvement in root colonization. *Microbes Environ* **29**:413–9.
97. **Oku S, Komatsu A, Tajima T, Nakashimada Y, Kato J.** 2012. Identification of chemotaxis sensory proteins for amino acids in *Pseudomonas fluorescens* Pf0-1 and their involvement in chemotaxis to tomato root exudate and root colonization. *Microbes Environ* **27**:462–469.
98. **McKellar JLO, Minnell JJ, Gerth ML.** 2015. A high-throughput screen for ligand binding reveals the specificities of three amino acid chemoreceptors from *Pseudomonas syringae* pv. *actinidiae*. *Mol Microbiol* **96**:694–707.
99. **Kuroda A, Kumano T, Taguchi K, Nikata T, Kato J, Ohtake H.** 1995. Molecular cloning and characterization of a chemotactic transducer gene in *Pseudomonas aeruginosa*. *J Bacteriol* **177**:7019–7025.
100. **Taguchi K, Fukutomi H, Kuroda A, Kato J, Ohtake H.** 1997. Genetic identification of chemotactic transducers for amino acids in *Pseudomonas aeruginosa*. *Microbiology* **143** (Pt 1):3223–9.
101. **Reyes-Darias JA, García V, Rico-Jiménez M, Corral-Lugo A, Lesouhaitier O, Juárez-Hernández D, Yang Y, Bi S, Feuilloley M, Muñoz-Rojas J, Sourjik V, Krell T.** 2015. Specific gamma-aminobutyrate chemotaxis in pseudomonads with different lifestyle. *Mol*

- Microbiol **97**:488–501.
102. **Liu X, Wood PL, Parales J V., Parales RE.** 2009. Chemotaxis to pyrimidines and identification of a cytosine chemoreceptor in *Pseudomonas putida*. *J Bacteriol* **191**:2909–2916.
 103. **Parales RE, Nesteryuk V, Hughes JG, Luu RA, Ditty JL.** 2014. Cytosine chemoreceptor McpC in *Pseudomonas putida* F1 also detects nicotinic acid. *Microbiol (United Kingdom)* **160**:2661–2669.
 104. **Fernández M, Morel B, Corral-Lugo A, Krell T.** 2016. Identification of a chemoreceptor that specifically mediates chemotaxis toward metabolizable purine derivatives. *Mol Microbiol* **99**:34–42.
 105. **Luu RA, Kootstra JD, Nesteryuk V, Brunton CN, Parales J V., Ditty JL, Parales RE.** 2015. Integration of chemotaxis, transport and catabolism in *Pseudomonas putida* and identification of the aromatic acid chemoreceptor PcaY. *Mol Microbiol* **96**:134–47.
 106. **Iwaki H, Muraki T, Ishihara S, Hasegawa Y, Rankin KN, Sulea T, Boyd J, Lau PCK.** 2007. Characterization of a pseudomonad 2-nitrobenzoate nitroreductase and its catabolic pathway-associated 2-hydroxylaminobenzoate mutase and a chemoreceptor involved in 2-nitrobenzoate chemotaxis. *J Bacteriol* **189**:3502–3514.
 107. **Grimm AC, Harwood CS.** 1999. Nahy, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol* **181**:3310–3316.
 108. **Lacal J, Muñoz-Martínez F, Reyes-Darías JA, Duque E, Matilla M, Segura A, Calvo JJO, Jiménez-Sánchez C, Krell T, Ramos JL.** 2011. Bacterial chemotaxis towards aromatic hydrocarbons in *Pseudomonas*. *Environ Microbiol* **13**:1733–1744.

109. **Kim HE, Shitashiro M, Kuroda A, Takiguchi N, Ohtake H, Kato J.** 2006. Identification and characterization of the chemotactic transducer in *Pseudomonas aeruginosa* PAO1 for positive chemotaxis to trichloroethylene. *J Bacteriol* **188**:6700–6702.
110. **Lacal J, García-Fontana C, Muñoz-Martínez F, Ramos JL, Krell T.** 2010. Sensing of environmental signals: Classification of chemoreceptors according to the size of their ligand binding regions. *Environ Microbiol* **12**:2873–2884.
111. **Armitage JP.** 1999. Bacterial tactic responses. *Advances in microbial physiology*.
112. **Henry JT, Crosson S.** 2011. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annu Rev Microbiol* **65**:261–286.
113. **Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, Páez-Espino AD, Durante-Rodríguez G, Kim J, Nickel PI, Platero R, De Lorenzo V.** 2013. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* **41**:D666–D675.
114. **Grimm A, Harwood C.** 1997. Chemotaxis of *Pseudomonas* spp. to the polyaromatic hydrocarbon naphthalene. *Appl Envir Microbiol* **63**:4111–4115.
115. **Ni B, Huang Z, Fan Z, Jiang C, Liu S.** 2013. *Comamonas testosteroni* uses a chemoreceptor for tricarboxylic acid cycle intermediates to trigger chemotactic responses towards aromatic compounds **90**:813–823.
116. **Lacal J, Reyes-Darias J a., García-Fontana C, Ramos JL, Krell T.** 2013. Tactic responses to pollutants and their potential to increase biodegradation efficiency. *J Appl Microbiol* **114**:923–933.
117. **Furukawa K, Fujihara H.** 2008. Microbial degradation of polychlorinated biphenyls:

- biochemical and molecular features. *J Biosci Bioeng* **105**:433–449.
118. **Gordillo F, Chávez FP, Jerez CA.** 2007. Motility and chemotaxis of *Pseudomonas* sp. B4 towards polychlorobiphenyls and chlorobenzoates. *FEMS Microbiol Ecol* **60**:322–328.
119. **Kumar A, Galstian T, Pattanayek SK, Rainville S.** 2013. The Motility of Bacteria in an Anisotropic Liquid Environment. *Mol Cryst Liq Cryst* **574**:33–39.
120. **Taylor TB, Buckling A.** 2011. Selection experiments reveal trade-offs between swimming and twitching motilities in *Pseudomonas aeruginosa*. *Evolution (N Y)* **65**:3060–3069.
121. **López-Farfán D, Reyes-Darias JA, Krell T.** 2016. The expression of many chemoreceptor genes depends on the cognate chemoeffector as well as on the growth medium and phase. *Curr Genet* **Epub ahead**.
122. **Krell T.** 2015. Tackling the bottleneck in bacterial signal transduction research: high-throughput identification of signal molecules. *Mol Microbiol* **96**:685–8.
123. **Alexandre G.** 2010. Coupling metabolism and chemotaxis-dependent behaviours by energy taxis receptors. *Microbiology* **156**:2283–93.
124. **Edwards JC, Johnson MS, Taylor BL.** 2006. Differentiation between electron transport sensing and proton motive force sensing by the Aer and Tsr receptors for aerotaxis. *Mol Microbiol* **62**:823–837.
125. **Behrens W, Schweinitzer T, McMurry JL, Loewen PC, Buettner FFRR, Menz S, Josenhans C.** 2016. Localisation and protein-protein interactions of the *Helicobacter pylori* taxis sensor TlpD and their connection to metabolic functions. *Sci Rep* **6**:23582.
126. **Yao J, Allen C.** 2007. The plant pathogen *Ralstonia solanacearum* needs aerotaxis for normal biofilm formation and interactions with its tomato host. *J Bacteriol* **189**:6415–6424.
127. **Greer-Phillips SE, Stephens BB, Alexandre G.** 2004. An energy taxis transducer

- promotes root colonization by *Azospirillum brasilense*. *J Bacteriol* **186**:6595–6604.
128. **Vegge CS, Brøndsted L, Li Y-P, Bang DD, Ingmer H.** 2009. Energy taxis drives *Campylobacter jejuni* toward the most favorable conditions for growth. *Appl Environ Microbiol* **75**:5308–5314.
 129. **Hartley-Tassell LE, Shewell LK, Day CJ, Wilson JC, Sandhu R, Ketley JM, Korolik V.** 2010. Identification and characterization of the aspartate chemosensory receptor of *Campylobacter jejuni*. *Mol Microbiol* **75**:710–730.
 130. **Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM.** 2005. Comparative metagenomics of microbial communities. *Science (80-)* **308**:554–557.
 131. **Handelsman J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**:669–85.
 132. **Boyd A, Kendall K, Simon MI.** 1983. Structure of the serine chemoreceptor in *Escherichia coli*. *Nature* **301**:623–626.
 133. **Yang Y, M Pollard A, Höfler C, Poschet G, Wirtz M, Hell R, Sourjik V.** 2015. Relation between chemotaxis and consumption of amino acids in bacteria. *Mol Microbiol* **96**:1272–82.
 134. **Hiroshi Nikaido.** 2009. The Limitations of LB Medium. *Small Things Considered*.
 135. **Sezonov G, Joseleau-Petit D, D’Ari R.** 2007. *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol* **189**:8746–8749.
 136. **Workentine ML, Harrison JJ, Stenroos PU, Ceri H, Turner RJ.** 2008. *Pseudomonas fluorescens*’ view of the periodic table. *Environ Microbiol* **10**:238–250.
 137. **Dewi Puspita I, Kamagata Y, Tanaka M, Asano K, Nakatsu CH.** 2012. Are Uncultivated

- Bacteria Really Uncultivable? *Microbes Environ* **27**:356–366.
138. **Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C.** 2013. Microbial syntrophy: Interaction for the common good. *FEMS Microbiol Rev* **37**:384–406.
139. **Pham VHT, Kim J.** 2012. Cultivation of unculturable soil bacteria. *Trends Biotechnol* **30**:475–484.
140. **Prakash O, Shouche Y, Jangid K, Kostka JE.** 2013. Microbial cultivation and the role of microbial resource centers in the omics era. *Appl Microbiol Biotechnol* **97**:51–62.
141. **Richards MA, Cassen V, Heavner BD, Ajami NE, Herrmann A, Simeonidis E, Price ND.** 2014. MediaDB: A database of microbial growth conditions in defined media. *PLoS One* **9**.
142. **Pham VHT, Kim J.** 2016. Improvement for isolation of soil bacteria by using common culture media. *J Pure Appl Microbiol* **10**:49–59.
143. **Görke B, Stülke J.** 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**:613–624.
144. **Ehrenberg M, Bremer H, Dennis PP.** 2013. Medium-dependent control of the bacterial growth rate. *Biochimie* **95**:643–658.
145. **Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M.** 2015. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* **34**:104–110.
146. **Nybroe O, Brandt KK, Ibrahim YM, Tom-Petersen A, Holm PE.** 2008. Differential bioavailability of copper complexes to bioluminescent *Pseudomonas fluorescens* reporter strains. *Environ Toxicol Chem* **27**:2246–2252.
147. **Stankic S, Suman S, Haque F, Vidic J.** 2016. Pure and multi metal oxide nanoparticles:

- synthesis, antibacterial and cytotoxic properties. *J Nanobiotechnology* **14**:73.
148. **Winterbourn CC.** 2014. The challenges of using fluorescent probes to detect and quantify specific reactive oxygen species in living cells. *Biochim Biophys Acta - Gen Subj* **1840**:730–738.
 149. **Wardman P.** 2007. Fluorescent and luminescent probes for measurement of oxidative and nitrosative species in cells and tissues: Progress, pitfalls, and prospects. *Free Radic Biol Med* **43**:995–1022.
 150. **Klanjscek T, Muller EB, Nisbet RM.** 2016. Feedbacks and tipping points in organismal response to oxidative stress. *J Theor Biol* **404**:361–374.
 151. **Chen M, Xu P, Zeng G, Yang C, Huang D, Zhang J.** 2015. Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: Applications, microbes and future research needs. *Biotechnol Adv* **33**:745–755.
 152. **Burton GA.** 2010. Metal Bioavailability and Toxicity in Sediments. *Crit Rev Environ Sci Technol* **40**:852–907.
 153. **Ghosal D, Ghosh S, Dutta TK, Ahn Y.** 2016. Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): A review. *Front Microbiol* **7**.
 154. **El Fantroussi S, Agathos SN.** 2005. Is bioaugmentation a feasible strategy for pollutant removal and site remediation? *Curr Opin Microbiol* **8**:268–275.
 155. **Chan YJ, Chong MF, Law CL, Hassell DG.** 2009. A review on anaerobic-aerobic treatment of industrial and municipal wastewater. *Chem Eng J* **155**:1–18.