

Northumbria Research Link

Citation: Thompson, Jonathan (2019) Iron uptake and homeostasis in the veterinary pathogen *Rhodococcus equi*: an integrated omics approach. Doctoral thesis, Northumbria University.

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/43358/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>



**Northumbria
University**
NEWCASTLE



UniversityLibrary

**IRON UPTAKE AND HOMEOSTASIS IN
THE VETERINARY PATHOGEN
RHODOCOCCLUS EQUI: AN
INTEGRATED OMICS APPROACH**

J R THOMPSON

PhD

2019

**IRON UPTAKE AND HOMEOSTASIS
IN THE VETERINARY PATHOGEN
RHODOCOCCUS EQUI: AN
INTEGRATED OMICS APPROACH**

JONATHAN ROBERT THOMPSON

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of
Doctor of Philosophy

Research undertaken in the
Faculty of Health and Life Sciences

December 2019

ABSTRACT

Rhodococcus equi, a veterinary pathogen that causes pyogranulomatous pneumonia, can secrete low molecular weight chelators called siderophores to scavenge iron when its bioavailability is limited. When iron is plentiful, synthesis of siderophores and ferri-siderophore transport systems are repressed.

Current literature on bacterial iron regulation and homeostasis indicates two distinct protein families of global iron-dependent transcriptional repressor: Fur and DtxR. Gram-negative bacteria produce Fur to regulate iron uptake genes and the biosynthesis of siderophores in response to the iron level in the cell. However, the Gram-positive *Corynebacteriaceae* produce DtxR-like proteins to regulate analogous genes.

Much remains undefined with respect to rhodococcal siderophore biosynthesis and uptake. Detailed analysis of the *R. equi* 103S genome for genes related to iron homeostasis identified two potential metal regulatory genes each from the Fur and DtxR families: iron dependent regulatory protein (IdeR), Diphtheria toxin repressor (DtxR), Ferric uptake regulator A (FurA) and Ferric uptake regulator B (FurB). Bioinformatic analysis confirmed that this complement of genes was conserved throughout *Rhodococcus* and the *Corynebacteriaceae* in general. To investigate their individual roles in metal homeostasis, molecular cloning and gene expression was performed, to facilitate analysis of regulator-metal specificities. Each gene was cloned but over-expression for functional analysis could only be achieved for *ideR*; thus, a thorough systematic analysis could not be achieved. In order to address their individual roles, homology-based protein modelling was used, and comparisons made with characterised homologues from *M. tuberculosis*. The geometrical conservation of key ligand amino acid residues strongly suggests *R. equi* utilises *ideR* as an iron regulator; *furB* as a zinc regulator, *dtxR* as a manganese regulator and *furA* as an oxidative stress response protein.

Most bacteria generate an exaggerated response to iron limitation *in vitro*, however *R. equi* produces very small siderophore yields, which has complicated their characterisation. In-frame deletion of the putative metal regulator genes *ideR*, *dtxR*, *furA* and *furB* was attempted in order to address the hypothesis that de-repression might generate greater yields. All genes were deleted individually; a marked phenotypic difference was noted only for *R. equi-ΔfurA*, which significantly upregulated the catalase encoded by the neighbouring gene and was coincidentally hyper-resistant to hydrogen peroxide. Surprisingly, analysis of siderophore production in the mutants indicated no increase in yield. The thesis discusses the relevance of this observation to microbial ecology. The availability of these mutants in combination with their predicted metal specificities facilitates the design of experiments to define their individual roles in metal homeostasis beyond the scope of this thesis.

The combination of 'omic' analyses was attempted here to initiate the ultimate definition of the complex molecular network associated with iron uptake. The genomic investigation informed hypothesis building for the other omic analyses. It suggested *R. equi* is capable of synthesising two siderophores, rhequibactin and rhequichelin; up to three had previously been postulated in the

literature. Culture optimisation was required to deliver a robust experimental design to impose iron limitation in isolation from other stresses. Once medium composition and biomarker-indicated harvesting criteria were established, biomass and associated secretomes were produced *en masse* for integrated omics analysis.

A comparative untargeted metabolomics study demonstrated an adapted iron-starved metabolome; strong siderophore candidates were then investigated using a targeted strategy. A strong candidate metabolite was identified by mass that appeared to be responsible for a heterobactin-like chromophore, however further biochemical characterisation has been elusive. Interestingly, the metabolite readily precipitates on complexation with iron, an observation also made for heterobactins.

Secondly, a transcriptomic study was attempted to study the global gene expression under iron starvation, and the impact of the loss of the IdeR in the deletion mutant generated in this work. However, the RNA extraction proved particularly challenging likely due to difficulties arising from lysis of the mycolic acid-containing cell wall. In the absence of a high-quality transcriptome sample, the study did not advance further and other aspects of the study were prioritised.

Finally, a comparative proteomic analysis into iron regulatory mechanisms associated with the rhodococcal cell wall was performed. Current literature deliberates how *R. equi* uses a range of strategies to overcome iron limitation through proposed uptake mechanisms associated with translocation across the cytoplasmic membrane *via* ABC transport systems, while no consideration has yet been made with regards to transport across the mycolic acid-containing cell wall structure. In this study no obvious candidate proteins for ferri-siderophore transport across the mycolate region were identified, therefore it is possible that *R. equi* utilises facilitated diffusion *via* a porin for entry of ferri-siderophore complexes into the pseudoperiplasm, where a substrate-binding lipoprotein may act as the primary receptor to facilitate cytosolic transfer through an ABC transport system.

Table of Contents

1	Introduction.....	1
1.1	Biology and bioavailability of iron	2
1.1.1	Understanding the link between iron and pathogenicity.....	3
1.2	Bacterial acquisition and exportation of metal ions.....	3
1.2.1	Iron uptake in Gram-negative bacteria.....	4
1.2.2	Iron uptake in Gram-positive bacteria	9
1.3	Siderophores.....	11
1.3.1	Siderophore selectivity and ligand structures	11
1.4	Release of iron from siderophore complex	15
1.4.1	Redox chemistry of siderophores.....	15
1.4.2	Hypothesis for iron release – redox method	16
1.4.3	Alternative hypothesis for iron release – protonation method.....	16
1.4.4	Lipophilicity in iron release	17
1.4.5	Iron regulatory circuits: mediating siderophore production	17
1.4.6	Non-ribosomal peptide synthetases	18
1.4.7	Synthesis of siderophoric compounds within NRPS mechanisms.....	19
1.5	Introduction to the <i>Rhodococcus</i> genus	21
1.6	<i>Rhodococcus equi</i> : the virulent zoonotic pathogen.....	23
1.6.1	Equine <i>R. equi</i> infections	24
1.6.2	Human <i>R. equi</i> infections.....	26
1.6.3	Reclassification of <i>R. equi</i>	26
1.6.4	Pathogenicity and virulence of <i>R. equi</i>	29
1.6.5	Project Aims.....	30
2	Materials and Methods.....	32
2.1	Bacterial strains and plasmid vectors.....	33
2.1.1	Cell culture and morphology testing.....	33
2.1.2	Preparation of Vectors	33
2.2	Iron limitation methods.....	34
2.2.1	Media production as a method of iron limitation.....	34
2.2.2	Modifications to impose iron limitation #1.....	34
2.2.3	Modifications to impose iron limitation #2.....	35
2.2.4	Modifications to impose iron limitation #3.....	35
2.3	DNA methods	35
2.3.1	Genomic DNA extraction	35
2.3.2	Plasmid DNA extraction	35
2.3.3	Polymerase Chain Reaction methods.....	36
2.3.4	Quality control using PCR.....	36

2.3.5	DNA analysis by electrophoresis.....	36
2.3.6	DNA recovery from agarose gel.....	37
2.3.7	Purification of PCR products in preparation for cloning.....	37
2.3.8	Drop dialysis of purified product.....	37
2.4	Oligonucleotide primers.....	38
2.4.1	Quality control primers.....	38
2.4.2	Metalloregulatory gene primers.....	38
2.4.3	Generation of <i>R. erythropolis</i> PR4 GroESL amplicon.....	38
2.4.4	Sequencing primers.....	38
2.4.5	Oligonucleotide primers for fXa-hybrid gene insertions.....	39
2.4.6	Oligonucleotide primers for mutagenic constructs.....	40
2.5	Transformation protocols.....	42
2.5.1	Preparation of competent <i>E. coli</i> cells.....	42
2.5.2	Transformation of competent <i>E. coli</i>	42
2.6	Electroporation protocols.....	43
2.6.1	Preparation of electrocompetent rhodococcal cells.....	43
2.6.2	Electroporation of rhodococcal cells.....	43
2.7	Molecular cloning methods.....	44
2.7.1	Blunt-ended molecular cloning.....	44
2.7.2	Cohesive-ended molecular cloning.....	45
2.7.3	NEBuilder HiFi DNA assembly.....	46
2.8	DNA sequencing.....	47
2.9	Protein methods.....	47
2.9.1	Gene expression.....	47
2.9.2	Percoll phase separation - Cell wall associated protein fractionation.....	48
2.9.3	Triton X114 hydrophobic protein solubilisation.....	48
2.9.4	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).....	49
2.9.5	Non-denaturing polyacrylamide gel electrophoresis (NATIVE-PAGE).....	50
2.10	Chromatography methods.....	51
2.10.1	Affinity chromatography.....	51
2.10.2	Adsorption chromatography.....	52
2.10.3	Gel Filtration chromatography.....	52
2.10.4	Thin layer chromatography.....	52
2.11	Biochemical analysis methods.....	53
2.11.1	Zone of inhibition.....	53
2.11.2	Chromazurol S Agar.....	53
2.11.3	Modified Chromazurol S Assay.....	54
2.12	<i>in silico</i> homology-based protein-modelling methods.....	55
2.12.1	Phyre2 protein modelling pipeline.....	55

2.13	Proteomic methods.....	56
2.13.1	In-gel digestion	56
2.13.2	Liquid chromatography and mass spectrometry	57
2.13.3	Protein Identification by MS/MS ion searching	58
2.13.4	Differential protein analysis.....	58
2.13.5	Data analysis methods – Protein localisation pipeline.....	60
2.14	Metabolomic methods.....	61
2.14.1	Liquid chromatography and mass spectrometry	61
2.14.2	Data alignment and pre-processing.....	62
2.14.3	Data analysis	63
2.15	Transcriptomic methods.....	64
2.15.1	RNA isolation - RiboPure™-Bacteria Kit	64
3	Genomic survey of iron homeostasis: A Bioinformatic analysis of iron acquisition, targets for mutagenesis and cloning in <i>Rhodococcus equi</i> 103S	66
3.1	Introduction.....	67
3.2	A bioinformatic investigation into transcriptional regulation.....	67
3.2.1	Alignment of clustered regulators.....	72
3.2.2	Multispecies identification of iron-dependent transcriptional regulators using BLASTP 74	
3.3	A review of literature regarding <i>Rhodococcus equi</i> siderophores	84
3.3.1	Rhequibactin	84
3.3.2	Rhequichelin	84
3.3.3	Potential for a mycobactin-like non-diffusible siderophore	85
3.4	Siderophore biosynthesis	86
3.4.1	NRPS – A putative method of siderophore production in <i>R. equi</i>	86
3.5	Predicted structures of secondary metabolites	87
3.5.1	HMM analysis.....	87
3.5.2	Antibiotics and Secondary Metabolite Analysis SHell (antiSMASH) analysis	87
3.5.3	Comparison of rhequichelin cluster prediction tools	88
3.5.4	Comparison of rhequibactin cluster prediction tools	96
3.5.5	Predictive structures of <i>R. equi</i> siderophores.....	102
3.6	Iron transport in <i>R. equi</i>	105
3.7	A bioinformatic analysis into transcriptional control of iron-responsive genes	111
3.7.1	Genomic analysis of DtxR(-like) co-repression.....	111
3.8	Summary of the bioinformatic investigation	119
4	Predicted Metal-binding capacities for the putative metal-dependent regulators of <i>Rhodococcus equi</i> 103S	121
4.1	Introduction to metalloregulation.....	122
4.2	A brief review of iron-dependent regulatory protein families	122

4.2.1	Fur protein superfamily.....	122
4.2.2	DtxR protein family	125
4.2.3	An overview of homologous metal regulatory systems.....	127
4.3	Putative gene selections	129
4.4	Strategy 1 – Molecular cloning, overexpression and <i>in vitro</i> characterisation of metal regulatory proteins	129
4.4.1	Amplification of metalloregulatory genes	129
4.4.2	Molecular cloning and sequencing analysis.....	130
4.4.3	Recombinant gene expression and protein production	133
4.4.4	Affinity chromatography using HiTrap Heparin HP	135
4.5	Strategy 2 - Fusion of metal regulatory proteins with Glutathione-S transferase.....	137
4.5.1	Generation of GST fusion constructs.....	137
4.5.2	Vector preparation for GST-tagged protein purification	139
4.5.3	Assembly of pGEXPCR with fXa hybrid amplicons.....	142
4.5.4	Recombinant gene expression.....	143
4.5.5	GSTrap Purification of IdeR-GST	145
4.6	Strategy 3 - <i>In silico</i> analysis of the putative metal regulators of <i>R. equi</i> 103S by protein scaffolding.....	147
4.6.1	Importance of <i>in silico</i> structural and physicochemical analysis of proteins	147
4.6.2	<i>in silico</i> homology-based protein-modelling	147
4.6.3	Prediction of IdeR structure	149
4.6.4	Prediction of DtxR structure	156
4.6.5	Prediction of FurA structure	162
4.6.6	Prediction of FurB structure.....	168
4.7	Concluding remarks for this chapter.....	173
5	Generation and <i>in vitro</i> characterisation of <i>Rhodococcus equi</i> 103S mutants in four putative metal-dependent transcriptional regulators: <i>ideR</i>, <i>dtxR</i>, <i>furA</i> and <i>furB</i>.	177
5.1	The construction of unmarked in-frame deletion mutants in <i>Rhodococcus equi</i>	178
5.1.1	Counter selection of double-crossover mutants by concomitant expression of cytosine deaminase and uracil phosphoribosyltransferase	179
5.2	Chapter aims	182
5.3	Generation of mutagenic constructs.....	182
5.3.1	Generation of mutagenic gene amplicons <i>via</i> PCR	182
5.4	Assembly of mutagenic constructs into pSelAct	184
5.4.1	Preparation of vector and PCR amplicons for assembly.....	184
5.4.2	Molecular cloning - NEBuilder HiFi assembly	184
5.4.3	SpeI digest of mutagenic constructs.....	185
5.4.4	Sequencing of mutagenic constructs.....	185
5.5	Transfer of mutagenic constructs into <i>R. equi</i> 103S	186
5.6	Selection of the double crossover 5-FC resistant phenotype	186

5.6.1	Production of a new promoter to improve selectivity.....	187
5.7	pLongJon - Generation of a plasmid for improved 5-FC selection	188
5.7.1	NEBuilder HIFI assembly of pGroESL with pSelAct.....	188
5.7.2	Confirmation of pLongJon- GroESL incorporation.....	191
5.7.3	<i>In silico</i> – <i>in vitro</i> digestion comparison of plasmids	192
5.8	Incorporation of mutagenic constructs into <i>R. equi</i> 103S (V2)	193
5.9	Selection of the double crossover 5-FC resistant phenotype (V2).....	193
5.9.1	Diagnostic PCR as a molecular screen and optimisation.....	193
5.9.2	Identification of a flawed methodology	194
5.9.3	Optimising the 5-FC selection methodology	195
5.10	Molecular confirmation of the generation of <i>R. equi</i> 103S Δ metal regulatory genes.....	196
5.10.1	Selection of the mutant <i>R. equi</i> - Δ ideR	196
5.10.2	Selection of the mutant <i>R. equi</i> - Δ txR	197
5.10.3	Selection of the mutant <i>R. equi</i> - Δ furA	198
5.10.4	Selection of the mutant <i>R. equi</i> - Δ furB	199
5.11	Colony morphology and phenotypic characterisation	201
5.11.1	Cellular morphology	201
5.12	Protein profiling of <i>R. equi</i> mutants.....	201
5.13	<i>R. equi</i> - Δ furA – De-repression of an essential catalase gene	202
5.13.1	Proteomic analysis	202
5.13.2	Analysis of phenotypic effects on sensitivity to oxidants.....	203
5.14	Concluding remarks for this chapter.....	205
6	An analysis of the <i>Rhodococcus equi</i> 103S metabolome under iron limitation: An insight into siderophore mediated iron metabolism.....	208
6.1	Introduction.....	209
6.1.1	Introduction to metabolomics	209
6.2	Chapter aims and hypothesis.....	211
6.3	Media and growth conditions for <i>R. equi</i> 103S	212
6.3.1	Medium optimisation 1	213
6.3.2	Medium optimisation 2 - Fiss Minimal media.....	215
6.3.3	Medium optimisation 3 – Modified lactate minimal media.....	215
6.4	Untargeted metabolite profiling using MS1 Analyses.....	218
6.4.1	Chemometric analysis.....	220
6.4.2	Data analysis – Initial candidates.....	221
6.4.3	Feature selection – metabolite ranking by statistical significance.....	227
6.4.4	Potential biomarker identification for a targeted analysis	231
6.5	A targeted approach to characterise the <i>Rhodococcus equi</i> 103 siderophores	231
6.5.1	Large scale production of <i>R. equi</i>	231
6.5.2	Isolation of the first candidate – M0872	232

6.5.3	Targeted identification of other potential siderophore candidates.....	239
6.6	Concluding remarks	246
7	A comparative proteomic analysis of the composition of the rhodococcal cell wall under iron limitation.....	248
7.1	Introduction.....	249
7.1.1	Sample preparation for LC-MS analysis.....	249
7.1.2	Protein separation.....	250
7.1.3	Protein identification.....	250
7.2	Project workflow.....	251
7.3	Rhodococcal cell wall.....	252
7.4	Hypothesis.....	253
7.5	Experimental design.....	255
7.6	Results.....	257
7.6.1	Initial unfractionated proteomics analysis	257
7.6.2	Proteomic analysis of cell wall material fractionated by density gradient centrifugation	261
7.6.3	Triton X-114 extraction of fractionated cell wall	264
7.6.4	Consensus.....	266
7.6.5	Large scale proteomic survey	273
7.7	Discussion	283
7.7.1	Outcomes of the Proteomic study	283
7.7.2	Project limitations	284
7.7.3	Development of an appropriate hypothesis for iron uptake mechanisms in <i>R. equi</i>	285
8	Method optimisation for RNA isolation for use in RNA-Seq analysis of <i>Rhodococcus equi</i> 103S global gene expression under iron starvation.....	288
8.1	Introduction.....	289
8.1.1	Bacterial transcriptomics.....	290
8.2	Aims of the chapter	291
8.3	Experimental Design.....	291
8.4	Results.....	293
8.4.1	Cell culture and media modifications	293
8.4.2	Initial RNA isolation using RiboPure Bacteria Kit.....	293
8.5	Future methodology	296
8.5.1	Identification of methodology issues	296
8.5.2	Method optimisation	297
8.5.3	Importance of stressed growth conditions for RNA isolation methodology	298
8.6	Concluding remarks	298
9	- General discussion	300

9.1	The importance of a Multi-Omics experimental approach	301
9.1.1	Designing Experiments Suitable for Multi-Omics Integration	301
9.1.2	Methods of Multi-omics Data Integration	302
9.2	Initial Project Aims	303
9.3	Genomic survey of iron homeostasis with <i>Rhodococcus equi</i> 103S	304
9.3.1	Iron sequestration.....	305
9.4	Predicted Metal-binding capacities for the putative metal-dependent regulators of <i>Rhodococcus equi</i> 103S	306
9.5	Generation and <i>in vitro</i> characterisation of <i>Rhodococcus equi</i> 103S mutants for four putative metal regulatory genes: <i>ideR</i> , <i>dtxR</i> , <i>furA</i> and <i>furB</i>	309
9.6	An analysis of the <i>Rhodococcus equi</i> 103S metabolome under iron limitation: An insight into siderophore mediated iron metabolism.....	311
9.7	A comparative proteomic analysis of the composition of the rhodococcal cell wall under iron limitation.....	314
9.8	Method optimisation for RNA isolation for use in RNA-Seq analysis of <i>Rhodococcus equi</i> 103S global gene expression under iron starvation.....	316
9.9	Data integration of the multi-omics study.....	317
9.10	Future work.....	320
9.10.1	Metal regulatory proteins.....	320
9.10.2	Mutagenesis project and further candidates.....	320
9.10.3	Metabolomic analysis of siderophore candidates	321
9.10.4	Transcriptomics.....	321
10	References.....	322

List of Figures

Figure 1-1: An example of an iron (III) transport system utilised by <i>E. coli</i> and other Enterobacteriaceae.	4
Figure 1-2: A cartoon representation of the outer membrane receptor FecA (1KMO) (ligand free) (Krewulak and Vogel, 2008).....	5
Figure 1-3: The TonB-mediated transport mechanism employed to facilitate transition metal translocation through the cell envelope in Gram-negative bacteria (Klein and Lewinson, 2011).....	6
Figure 1-4 A ribbon structure of typical ABC-type transporter.....	7
Figure 1-5 Models of the Gram-positive siderophore-shuttle mechanism and displacement mechanism of YxeB (Fukushima et al., 2014).....	10
Figure 1-6: An enantiomeric isomer pair, highlighting the optical differences between the two classes (Hider and Kong, 2010).	12
Figure 1-7 Representative examples of different siderophores and their natural producers.	13
Figure 1-8 Three major ligand groups employed by siderophores, and their complexation with iron (III).	14
Figure 1-9: Examples of apo- and ferric-Siderophore complex formations.	15
Figure 1-10 The domain arrangement of bacterial non-ribosomal synthetases and associated catalytic reactions (Süssmuth, 2017).	19
Figure 1-11 KEGG Pathway for biosynthesis of bacterial catecholate siderophores using phenylalanine, tryptophan and tyrosine precursor amino acids.	20
Figure 1-12 Comparison of mycolic acid structures in the suborder <i>Corynebacteriales</i>	22
Figure 1-13: A lung from a foal suffering from purulent pneumonia with numerous abscesses caused by <i>R. equi</i> infection (Vazquez-Boland et al., 2013).....	24
Figure 1-14: A phylogenetic tree from 400 universal proteins highlighting the differentiating genomic relationship between <i>R. equi</i> and other related species (Sangal et al., 2019).....	28
Figure 3-1 Conserved domains prediction for putative DtxR type metal sensitive regulators	70
Figure 3-2 Conserved domains prediction for putative Fur type metal sensitive regulators	71
Figure 3-3: Sequence alignment of Fur-type transcriptional regulators of <i>Rhodococcus equi</i>	72
Figure 3-4: Sequence alignment of DtxR-type transcriptional regulators of <i>Rhodococcus equi</i>	73
Figure 3-5: Evolutionary analysis of multispecies iron-dependent transcriptional regulators FurA and FurB by Maximum Likelihood.	78
Figure 3-6: Evolutionary analysis of multispecies iron-dependent transcriptional regulators DtxR and IdeR by Maximum Likelihood.	82
Figure 3-7 NRPS prediction tool using a HMM model to predict potential amino acids used to create the secondary metabolite.	89
Figure 3-8: A Non-continuous extended motif alignment for adenylation domain 1 binding pocket of the putative rhequichelin in <i>R. equi</i> 103S.	90
Figure 3-9: A Non-continuous extended motif alignment to identify the binding pocket residues of Adenylation domain 3 of the NRPS system encoding the putative siderophore rhequichelin.	91
Figure 3-10: A Non-continuous extended motif alignment using an alternative NRPS adenylation domain binding pocket residue for the NRPS system encoding the putative siderophore rhequichelin.	92
Figure 3-11: antiSMASH KnownClusterBlast search identifying that Erythrochelin has 57 % similarity to the Query sequence Rhequichelin.	93

Figure 3-12: Biosynthetic gene cluster Region 3 (rhequichelin) for <i>R. equi</i> 103S identified using antiSMASH.....	93
Figure 3-13: The detailed domain annotation for the Biosynthetic gene cluster Region 3 encoded by the NRPS gene REQ_07630.....	94
Figure 3-14: NRPS prediction tool using a HMM model to predict amino acyl components of secondary metabolites.....	98
Figure 3-15: antiSMASH KnownClusterBlast search identifying Heterobactin as 54 % similarity to the Query sequence Rhequibactin.....	99
Figure 3-16: Biosynthetic gene cluster Region 4 (rhequibactin) prediction for <i>R. equi</i> 103S identified using antiSMASH.....	99
Figure 3-17: The detailed domain annotation for the Biosynthetic gene cluster Region 4 encoded by the precursor genes REQ_08110, REQ_08120, NRPS genes REQ_08140 and REQ_08160.....	100
Figure 3-18: Predictive structure of the tetrapeptide rhequichelin.....	102
Figure 3-19: Rhequibactin predicted structure.....	103
Figure 3-20: Siderophore structures analogous to the predicted structure of rhequibactin.....	104
Figure 3-21: Conserved domains for operon iupABC.....	106
Figure 3-22: Proposed siderophore mediated uptake and transport of iron employed by <i>S. coelicolor</i> using an ABC-transporter system – ChhCDEF.....	108
Figure 3-23: Multiple sequence alignment of putative iron-box operators in the <i>R. equi</i> 103S chromosome, using the experimentally verified iron-box from <i>fxbA</i> in <i>Mycobacterium smegmatis</i> as a template.....	112
Figure 3-24: The basic structure of prokaryotic transcriptional de-repression.....	113
Figure 3-25: An <i>R. equi</i> iron homeostasis network map generated from the preliminary genomic survey and current literature.....	120
Figure 4-1: Model comparisons to explain the Fur box consensus sequence.....	125
Figure 4-2: PCR amplification of four putative <i>R. equi</i> iron regulatory genes.....	130
Figure 4-3: Strategy for cloning of <i>R. equi</i> metal regulators in pET23b.....	131
Figure 4-4: Colony PCR confirming amplicon incorporation for <i>furA</i> and <i>furB</i>	132
Figure 4-5: Colony PCR confirming amplicon incorporation for <i>ideR</i>	133
Figure 4-6 Colony PCR confirming amplicon incorporation for <i>dtxR</i>	133
Figure 4-7: 12 % SDS-PAGE gel for the expression study of pET23b-IdeR.....	134
Figure 4-8: Attempted purification of <i>Rhodococcus equi</i> IdeR on Heparin-agarose.....	135
Figure 4-9: 12 % SDS-PAGE analysis of chelex-treated IdeR clarified lysate flow-through.....	136
Figure 4-10 :2 % SDS-PAGE analysis of chelex-treated IdeR clarified lysate fractions 30-38.....	136
Figure 4-11: Experiment methodology for producing pGEXJT-fXa hybrid plasmids.....	138
Figure 4-12: PCR amplification for four putative iron regulatory genes with incorporation of a Factor Xa region.....	139
Figure 4-13: GST vector pGEX-6P-1 showing the correct reading frame, and cutting region of PreScission Protease denoted by the arrow.....	139
Figure 4-14: GST vector pGEX-5X-1 showing the correct reading frame and cutting region of Factor Xa Protease denoted by the arrow.....	140
Figure 4-15: Elimination of the PreScission protease cut site using the restriction enzyme <i>AloI</i> in the vector pGEX-6P-1.....	140

Figure 4-16: 1 % agarose gel electrophoretic analysis of the plasmid pGEX-6P-1 when cut with the restriction enzyme <i>AloI</i> .	141
Figure 4-17: Sequencing results with preScission site identified	141
Figure 4-18 Production of the amplicon pGEXPCR via PCR	142
Figure 4-19: Experiment methodology for producing pGEXPCR-fXa hybrid plasmids, the example above uses the gene <i>ideR</i> .	143
Figure 4-20: Electrophoretic analysis of heterologous protein over production from clarified lysates.	144
Figure 4-21 FPLC chromatogram of GSTrap purification of IdeR-GST	145
Figure 4-22 Electrophoretic analysis of FPLC flowthrough fractions for purification of IdeR-GST	146
Figure 4-23: Electrophoretic analysis of FPLC peak fractions for the purification of IdeR-GST.	146
Figure 4-24:Flowchart detailing the basic principles of <i>in silico</i> protein-modelling.	148
Figure 4-25: A comparative analysis of known and predicted IdeR structures.	150
Figure 4-26: Chain Sequence View for the Entity PDB 1FX7.	151
Figure 4-27: Sequence alignment between IdeR proteins from <i>R. equi</i> and <i>M. tuberculosis</i> .	152
Figure 4-28: Metal binding site 1 of 1FX7, <i>M. tuberculosis</i> IdeR in complex with Co(II).	152
Figure 4-29: Hypothesised metal binding site 1 of the homology-based model of IdeR from <i>R. equi</i> , in apo-form.	153
Figure 4-30: A Polar pocket occupied by a water molecule adjacent to metal binding site 1 of IdeR (1FX7).	153
Figure 4-31: Metal binding site 2 of <i>M. tuberculosis</i> IdeR in complex with Co(II).	154
Figure 4-32: Hypothesised metal binding site 2 of the homology-based model of IdeR from <i>R. equi</i> , in apo-form.	155
Figure 4-33: A comparative analysis of known and predicted DtxR/MntR structures.	157
Figure 4-34: Chain Sequence View for the Entity PDB 5ZR6.	158
Figure 4-35: Sequence alignment between DtxR and MntR proteins from <i>R. equi</i> and <i>M. tuberculosis</i> respectively.	158
Figure 4-36: Metal binding site 1 <i>M. tuberculosis</i> MntR 5zr6 complexed with Mn(II).	159
Figure 4-37: Predicted metal binding site 1 of the homology-based model for DtxR in <i>R. equi</i> , in apo-form.	160
Figure 4-38: Metal binding site 2 <i>M. tuberculosis</i> MntR 5zr6 complexed with Mn(II).	160
Figure 4-39: Predicted metal binding site 2 of the homology-based model for DtxR in <i>R. equi</i> , in apo-form.	161
Figure 4-40: A comparative analysis of known and predicted PerR/FurA structures.	163
Figure 4-41: Chain Sequence View for the Entity PDB 2FE3.	164
Figure 4-42 Sequence alignment between PerR and two variations of FurA proteins from <i>B.subtilis</i> , <i>R. equi</i> and <i>M. tuberculosis</i> respectively.	165
Figure 4-43: Structural metal binding site of PerR (2FE3) from <i>B. subtilis</i> occupied by a Zn (II) atom.	165
Figure 4-44: Predicted structural metal binding site of the homology-based model of FurA from <i>R. equi</i> in apo-form.	166
Figure 4-45: The regulatory metal binding site of PerR (2FE3) from <i>B. subtilis</i> in apo-form.	166

Figure 4-46: Predicted regulatory metal binding site of the homology based model of FurA from <i>R. equi</i> in apo-form.....	167
Figure 4-47: A comparative analysis of known and predicted FurB structures.....	168
Figure 4-48: Chain Sequence View for the Entity PDB 2o03.	169
Figure 4-49: Sequence alignment of FurB proteins from <i>Rhodococcus equi</i> (<i>R. equi</i>) and <i>Mycobacterium tuberculosis</i> (<i>M.tb</i>) respectively.....	170
Figure 4-50 Metal binding site 1 of FurB 2o03 from <i>M. tuberculosis</i> occupied by a Zn(II) ion. ..	170
Figure 4-51 Putative metal binding site 1 of the homology-based model of FurB in <i>R. equi</i> , in apo-form.....	171
Figure 4-52: Metal binding site 2 of FurB 2o03 from <i>M. tuberculosis</i> occupied by a Zn(II) ion. .	172
Figure 4-53: Putative metal binding site 2 of the homology-based model of FurB in <i>R. equi</i> , in apo-form.....	172
Figure 4-54: The updated <i>R. equi</i> iron homeostasis network map (after chapter 4).....	176
Figure 5-1: Gibson assembly style methodology utilised to produce gene mutations in the vector pSelAct.....	178
Figure 5-2: An example of homologous recombination to incorporate an in-frame silent gene mutation in <i>Rhodococcus equi</i> using the vector pSelAct.....	181
Figure 5-3: PCR products generated for <i>R. equi</i> 103S mutagenesis.....	183
Figure 5-4 Δ <i>FurA</i> -downstream (2) optimisation.....	184
Figure 5-5 Purified mutagenic amplicons and vector required for HiFi assembly.....	184
Figure 5-6: Diagnostic restriction digest of plasmids using <i>SpeI</i>	185
Figure 5-7 Blastn alignment of <i>groESL</i> from <i>R. equi</i> 103S to <i>R. erythropolis</i> PR4 for production of the chaperonin to increase the efficiency of the 5-FC selection process.	187
Figure 5-8: PCR amplicon for generation of PR4 <i>groESL</i>	188
Figure 5-9: pLongJon vector map.....	189
Figure 5-10 A flowchart detailing the molecular strategy employed to produce pLongJon deletion mutants.....	190
Figure 5-11: Colony PCR of pLongJon plasmids to identify incorporation of pGroESL promoter	191
Figure 5-12: <i>In silico</i> digestion of pLongJon and mutant plasmids compared to <i>in vitro</i> digestions.	192
Figure 5-13: An example of replica plating for <i>R. equi</i> - Δ <i>ideR</i> on LB agar and LBapr80 using colonies from a 100 μ g/ml 5-FC plate.....	193
Figure 5-14: Colony PCR to identify secondary homologous recombinant mutants	194
Figure 5-15: Optimised PCR screening to identify both wild type and truncate gene variants.....	194
Figure 5-16: Replica plating for <i>R. equi</i> - Δ <i>ideR</i> on LB agar and LBapr80 using colonies from a 200 μ g/ml 5-FC plate after 3 days growth.....	195
Figure 5-17: Replica plating for <i>R. equi</i> - Δ <i>dtxR</i> on LB agar and LBapr80 using colonies from a 200 μ g/ml 5-FC plate after 3 days growth.....	196
Figure 5-18: Colony PCR for <i>R. equi</i> - Δ <i>ideR</i> to identify secondary homologous recombination...	197
Figure 5-19: Colony PCR for <i>R. equi</i> - Δ <i>dtxR</i> to identify secondary homologous recombination...	198
Figure 5-20: Colony PCR for <i>R. equi</i> - Δ <i>furA</i> to identify secondary homologous recombination...	199
Figure 5-21: Colony PCR for <i>R. equi</i> - Δ <i>furB</i> to identify secondary homologous recombination...	200

Figure 5-22: BHI agar plates showing colony morphology of <i>R. equi</i> 103S wildtype and mutant variations.....	201
Figure 5-23 A 12 % SDS-PAGE total protein load for <i>R. equi</i> 103S wildtype compared to <i>R. equi</i> 103S mutants.....	202
Figure 5-24: Protein identification of the hyper-expressed band in <i>R. equi</i> 103S-FurA.....	203
Figure 5-25: A zone of inhibition assay for breakdown of hydrogen peroxide by catalase.....	204
Figure 5-26 An updated <i>R. equi</i> iron homeostasis network map (after chapter 5).....	207
Figure 6-1: Identification of optimal growth conditions to induce an iron starvation response in <i>R. equi</i> 103S.....	213
Figure 6-2: Refined growth conditions in <i>R. equi</i> 103S imposing different iron regimens to elicit an iron starvation response.....	214
Figure 6-3: A microtitre CAS assay to detect siderophore production in <i>R. equi</i> 103S.....	215
Figure 6-4: Visual inspection of <i>R. equi</i> 103S cultures after 120 hours growth under different iron regimes.....	216
Figure 6-5: Visual inspection of <i>R. equi</i> 103S culture supernatants after 120 hours growth under different iron regimes.....	216
Figure 6-6: Visual inspection of <i>R. equi</i> 103S culture supernatants aliquots under different iron regimes with additional iron supplementation.....	217
Figure 6-7: Metabolite profiles of <i>R. equi</i> 103S using RP (C18) chromatography, positive ionisation mode.....	218
Figure 6-8: Metabolite profiles of <i>R. equi</i> 103S using HILIC (amide) chromatography, positive ionisation mode.....	218
Figure 6-9: Box plots and kernel density plots before and after normalization.....	219
Figure 6-10: 2D Principal component analysis for unsupervised variance analysis between sample classes of <i>R. equi</i> 103S.....	220
Figure 6-11: Loading values of mass spectral feature M0537.....	222
Figure 6-12: Loading values of mass spectral feature M0872.....	223
Figure 6-13: Loading values of mass spectral feature M1067.....	223
Figure 6-14: Extracted ion chromatographs for M0537.....	225
Figure 6-15: Extracted ion chromatographs for M0872.....	226
Figure 6-16: Extracted ion chromatographs for M1067.....	227
Figure 6-17: Significant features identified by SAM.....	229
Figure 6-18: Visual stages of the large-scale production of iron limited <i>R. equi</i> 103S.....	232
Figure 6-19: Chromatograph of FPLC purification of bacterial supernatant using XAD-2 polymeric resin.....	233
Figure 6-20: Fractions eluted from the XAD-2 resin purification of iron limited <i>R. equi</i> 103S.....	233
Figure 6-21: Microtitre CAS assay for fractions of interest from the XAD-2 resin purification of iron limited <i>R. equi</i> 103S.....	234
Figure 6-22: Visual examination of Fraction 8 from the XAD-2 resin purification of iron limited <i>R. equi</i> 103S.....	234
Figure 6-23: Attempted solvent extractions for hypothetical ferri- and des-ferri siderophore fractions.....	235
Figure 6-24: Total ion chromatograph for LC/MS analysis of Fraction 8.....	236

Figure 6-25: Extracted ion chromatograph and MS spectra in Fraction 8.....	237
Figure 6-26: Total Ion Chromatograph comparison of Fraction 8 and Fraction 8+FeCl ₃	238
Figure 6-27: XAD-2 column load, wash and elute for secondary candidate identification.....	240
Figure 6-28: Microtitre CAS assay for the second candidate identification.....	240
Figure 6-29: CAS agar test for siderophore detection	241
Figure 6-30: P2 Microtitre CAS assay for siderophore detection of Gel filtration fractions.....	241
Figure 6-31: UV-visible spectral analysis of Gel filtration fractions.....	242
Figure 6-32: UV-visible spectral analysis of Gel filtration fractions using CAS assay solution....	242
Figure 6-33: TLC analysis of XAD-2 CAS positive pool and gel filtration fractions 41 and 42....	243
Figure 6-34: Analysis of a gel filtration Fraction 41 concentrate on CAS agar for siderophore detection.....	243
Figure 6-35: Loading values of Alternative candidates - mass spectral feature M0646, M0667, M0780, M0852, M0968 and M1039.....	245
Figure 6-36: An updated <i>R. equi</i> iron homeostasis network map (after Chapter 6)	247
Figure 7-1: A spider diagram exploring the field of proteomics (Graves and Haystead, 2002)....	249
Figure 7-2: An overview of a Bottom-up proteomic analysis.	250
Figure 7-3: Cell envelope architecture of Gram-positive, acid-fast and Gram-negative bacteria (Maitra et al., 2019).....	252
Figure 7-4: BlastP search for the mycobacterial HupB protein against the <i>R. equi</i> 103S chromosome.....	254
Figure 7-5: Flowchart detailing comparison of planned fractionation methods.....	256
Figure 7-6 Methodology comparison analysis needed prior to final study.....	256
Figure 7-7: 12 % SDS-PAGE gel for initial proteomic analysis of <i>R. equi</i> 103S cellular material under iron limitation.....	257
Figure 7-8: 12 % SDS-PAGE gel for cell wall material fractionated by density gradient centrifugation of <i>Rhodococcus equi</i> 103S.	262
Figure 7-9: 12 % SDS-PAGE gel for proteomic analysis of Triton X-114 extracted <i>R. equi</i> 103S cellular material under iron limitation.	265
Figure 7-10: Venn diagram detailing the quantity of identified proteins for each given fractionation method.....	267
Figure 7-11: STRING analysis for the hypothetical protein CBH50452.1 encoded by the gene REQ_44910.....	268
Figure 7-12: STRING analysis for the hypothetical protein CBH50452.1 encoded by the gene REQ_17680.....	269
Figure 7-13: Venn diagram detailing the quantity of identified proteins for each given fractionation method, reinforced by the in-silico prediction methodology.....	270
Figure 7-14: STRING analysis for the putative ABC transporter substrate binding lipoprotein (CBH48875.1) encoded by the gene REQ_28560.....	271
Figure 7-15: Principle Component Analysis for total features (peptides) identified in the large-scale proteomic survey. Total feature count: 72687.	274
Figure 7-16: Principal Component Analysis generated by Progenesis for significant and differentially expression proteins in the large-scale proteomic survey.....	275
Figure 7-17: Principal Component Analysis and Standardised Expression profiles generated by Progenesis for significant and upregulated proteins under iron limited conditions in in the large-	

scale proteomic survey.....	275
Figure 7-18: Evidence based STRING analysis for the multiple non-cytoplasmic proteins identified by the large-scale proteomic survey and <i>in silico</i> correlation methods.	279
Figure 7-19: Gene co-occurrence between the candidate list generated in Table 8-9 from <i>R. equi</i> against other closely related bacterium from the order Corynebacteriales.	282
Figure 7-20: An updated <i>R. equi</i> iron homeostasis network map (after Chapter 7)	287
Figure 8-1: Electropherogram analysis of RNA integrity.....	295
Figure 8-2: Quality control flowchart for RNA integrity analysis. Adapted from (Mueller et al., 2004)	296
Figure 9-1-: An <i>R. equi</i> iron homeostasis network map generated from the results in this thesis..	319

List of Tables

Table 1-1 Clustering of periplasmic binding proteins based on ligand specificity.....	6
Table 2-1: List of bacterial strains and plasmid vectors	33
Table 2-2 Composition of lactate minimal medium for iron limitation in <i>Rhodococcus equi</i>	34
Table 2-3: Composition of Vishniac and Santer Trace elements solution.....	34
Table 2-4: PCR reaction setup using Q5 DNA polymerase.....	36
Table 2-5 : Typical thermocycling conditions for a Q5 DNA polymerase PCR reaction	36
Table 2-6: Oligonucleotides primers designed for quality control PCR reactions specific for <i>R. equi</i> and VapA.	38
Table 2-7: Oligonucleotides primers designed for metal regulatory gene amplicons.	38
Table 2-8 Generation of the <i>R. erythropolis</i> PR4 GroESL fragment.....	38
Table 2-9 Oligonucleotide primers used in Sanger sequencing methods	38
Table 2-10 Primers for fXa-hybrids.....	39
Table 2-11: IdeR gene knockout insertion (REQ_20130) primer design	40
Table 2-12: DtxR gene knockout insertion (REQ_19260) primer design	40
Table 2-13: FurA gene knockout insertion (REQ_04740) primer design	41
Table 2-14: FurB gene knockout insertion (REQ_29130) primer design.....	41
Table 2-15 An example of blunt-ended vector preparation, by digestion of 1 µg of plasmid with a restriction enzyme	44
Table 2-16 An example of a ligation using T4 ligase (a 3:1 molar ratio is shown).....	44
Table 2-17 An example of plasmid preparation for cohesive-ended cloning, by digestion of 1 µg of plasmid with two restriction enzymes.....	45
Table 2-18 NEBuilder HiFi DNA assembly	46
Table 2-19 Optimisation of conditions for gene expression	47
Table 2-20 SDS-PAGE buffer compositions	49
Table 2-21 Composition of PAGE resolving gel at various acrylamide concentrations	49
Table 2-22 Composition of PAGE stacking gel.....	50
Table 3-1: Abridged genomic survey of iron regulatory genes in <i>R. equi</i> 103S.....	67
Table 3-2 Specific hits generated by conserved domain database searching for candidate iron-responsive transcriptional regulators	69
Table 3-3: BLASTP ranked table identifying conservation of the Fur-family proteins restricted to <i>Rhodococcus</i>	76
Table 3-4: Multispecies identification of the iron-dependent transcriptional regulators FurA and FurB using BLASTP.....	77
Table 3-5 BLASTP ranked table identifying conservation of the DtxR-family proteins restricted to <i>Rhodococcus</i>	80
Table 3-6 Multispecies identification of the iron-dependent transcriptional regulators DtxR and IderR using BLASTP.....	81
Table 3-7: NRPS found within the <i>R. equi</i> 103S chromosomal genome.....	86
Table 3-8 A NRPS-refined cluster analysis from record FN563149.1 <i>R. equi</i> 103S genome showing	

the biosynthetic gene cluster location and the most similar known cluster.	88
Table 3-9: Biosynthetic gene cluster Region 3 (rhequichelin) features tabulated using Snappgene, .	94
Table 3-10: Consensus NPRS monomer prediction output as detailed by antiSMASH v5.0.0rc1 ..	95
Table 3-11: Biosynthetic gene cluster Region 4 (rhequibactin)	100
Table 3-12: Consensus NPRS monomer prediction output as detailed by antiSMASH v4.2.1	101
Table 3-13: Potential variations in acyl groups for rhequichelin, C1-6 and associated predicted masses	103
Table 3-14: BlastP results for NP_624811.1 iron-siderophore binding lipoprotein [<i>Streptomyces coelicolor</i> A3(2)] restricted to TxID: <i>Rhodococcus equi</i> 103S	109
Table 3-15: BlastP results for NP_830222.1 ferrichrome-binding protein [<i>Bacillus cereus</i> ATCC 14579] restricted to TxID: <i>Rhodococcus equi</i> 103S	110
Table 3-16: Identification of putative iron boxes in the <i>R. equi</i> 103S chromosome with consideration of genetic context.	116
Table 4-1: Metal-dependent regulatory genes of interest	130
Table 4-2: Metalloregulatory genes of interest with given gene and fXa-amplicon lengths	139
Table 5-1 pSelAct- Δ mutant sequencing results	186
Table 5-2: pSelAct- Δ mutant sequencing results	191
Table 5-3: MASCOT protein hit output detailing the top 5 protein families, and their associated members.	202
Table 6-1: Peak intensity table for the metabolite profiling of <i>R. equi</i> 103S using RP LC-MS methodology; ranked by average intensity in LMM conditions.	221
Table 6-2: Peak intensity table for the metabolite profiling of <i>R. equi</i> 103S using RP LC-MS methodology; significant features ranked by SAM.	230
Table 6-3: Siderophore candidate list from Fraction 41.	244
Table 7-1: Abridged <i>in silico</i> protein localisation results for the initial unfractionated proteomic analysis	259
Table 7-2: Abridged <i>in silico</i> protein localisation results for the proteomic analysis of the cell wall material fractionated by density gradient centrifugation	263
Table 7-3: Abridged <i>in silico</i> protein localisation results for the Triton X-114 extracted cell wall material	266
Table 7-4: Conserved proteins between all 3 methodologies.	268
Table 7-5: Conserved non-cytoplasmic proteins between all 3 study methods, as predicted using a consensus approach to bioinformatic subcellular localisation	270
Table 7-6: Abridged results for protein localisation of in the large-scale proteomic survey	277
Table 7-7: Putative ABC transporter integral membrane subunits.	280
Table 7-8 Putative ABC transporter substrate binding.	280
Table 7-9: A list of candidates identified to contain a secretory signal or lipid anchor signal	281
Table 8-1: Prokaryote Total RNA Nano Series II Assay	294

Acknowledgements

First and foremost, I would like to thank my principal supervisor Dr Lynn G. Dover for the opportunity to conduct this research. The guidance and support that he provided over the course of the past four years has enabled me to develop my confidence and skills to become an independent researcher. I would also like to thank my secondary supervisor Professor Iain C. Sutcliffe for his opinions and expertise throughout the project.

I would also like to thank Dr Jonathon Brown for his friendship and support in the laboratory early on in the project and appreciate his role in developing my technical proficiencies. My thanks also go to the technicians and scientific officers from the Ellison building laboratories for their friendship and advice as well as regular loaning of both their time and equipment.

A special thank you is needed to all of the past and present PhD students that I have had the pleasure of working with at Northumbria University, their company and encouragement in the office and laboratory was instrumental to maintaining a positive outlook even in the most stressful of times.

I must also thank my family for the support throughout this long process and the understanding that my desire to complete this task often took precedence over time spent together. Finally, I hereby dedicate the work contained within this thesis to my parents Russell and Sandra, and my grandparents Joseph and Margaret, who have always said “Stick at it, Jonathan” right from First School through to completion of this Doctoral programme.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Name: Jonathan Thompson

Signature:

A handwritten signature in black ink, appearing to read 'J Thompson', followed by a period.

Date: 29th November 2019

1 Introduction

1.1 Biology and bioavailability of iron

Iron is one of the Earth's most abundant elements contributing greater than 30 % of the Earth's mass. It can be detected in the atmosphere, biosphere, lithosphere, and hydrosphere (Morgan and Anders, 1980) and, given its abundance, it is not surprising that most eukaryotic and prokaryotic organisms have evolved to utilise iron as a critical component for cellular processes and metabolic pathways. Despite its abundance in the environment, the bioavailability of iron can be limited due to the notable chemical property of the ability to mediate electron transfer to change between a reduced ferrous iron [Fe (II) / Fe²⁺] and an oxidised ferric iron [Fe (III) / Fe³⁺] form among other oxidation-reduction potentials. Throughout evolution, and due to the increase of oxygen tension, the redox properties of iron became extremely desirable. While ferrous iron is soluble in anoxic conditions, it readily insolubilises in the presence of oxygen, becoming the oxidised ferric form, that can be reduced to the ferrous again through acidification (Miller and Berner, 1989). Generally, two mechanisms are utilised by organisms to sequester iron (Silva and Faustino, 2015). The first involves the acidification of the environment in order to firstly solubilise the readily available ferric iron and to stimulate reduction to the ferrous form, that can be translocated intracellularly. The second methodology often used (and is discussed in detail later) is the synthesis and secretion of low-molecular weight chelatory peptide known as siderophores.

Furthermore, the wide range of metabolic reactions that utilise iron clearly highlights its utility as a biological catalyst; the versatile nature of iron is somewhat due to the capability to adopt several electron configurations, and as such is listed as a transition metal, due to incomplete d orbitals. The oxidative states of iron range from Fe²⁺ through to Fe⁶⁺, with the principal states, being the aforementioned 2⁺ (d6) and 3⁺ (d5) stable valences unlike a number of iron-dependent monooxygenases that produce high valence iron as a reactive intermediate during catalytic cycles (Outten and Theil, 2009). Iron (III) is a hard acid utilising hard oxygen ligand such as hydroxyl, carboxyl and other oxygen containing groups, whereas iron (II) is an intermediate of hard / soft acid, accommodating oxygen-based ligands but preferring nitrogen and sulphur ligands (Foster et al., 2014). The redox transition between Fe²⁺ and Fe³⁺ ions and ligand-donor atoms are dependent on the chemical bond strength formed. Furthermore, the suitability of iron for biocatalysis is a result of

extreme variability of the $\text{Fe}^{2+}/\text{Fe}^{3+}$ redox potential, which can be modified by appropriate ligands, so that iron complexes can encompass almost the entire biologically relevant range of redox potentials. Interestingly, the distribution of donor atoms which ligate the metal, and their geometry, will therefore control the functional properties (Nies, 2016).

Typically, the free Fe^{3+} concentration in most environments is severely limited (10^{-18} M), as a result of $\text{Fe}(\text{OH})_3$ insolubility. This is significantly lower than the required iron concentration needed to support life. The low bioavailability of iron for mammalian cells is further exacerbated by ferric iron sequestration that is necessary for production of haemoglobin synthesis (Andrews, 1999) with up to 70% of the human body iron pool comprised of haem iron in erythrocytes.

When an aqueous solution, iron is monopolised by the Fe^{2+} and Fe^{3+} valence forms as the complexes are able to readily undergo electron transfer and acid–base reactions, illuminating the extensive diversity of catalytic functions of which the element is able to perform, further underlining the importance in biological systems (Silva and Faustino, 2015).

1.1.1 Understanding the link between iron and pathogenicity

With the bioavailability of iron in a mammalian host (typically 10^{-18} M) far below the concentrations require for growth (Dussurget and Smith, 1998), many pathogens have developed a range of intricate systems in order to obtain sufficient iron to survive and replicate (Mietzner and Morse, 1994). Some of the systems employed include siderophore-mediated acquisition (Andrews et al., 2003), and iron acquisition mechanisms from haem and haem-containing proteins through either direct uptake of haem and or through the use of haemophores followed secretion of haemolysins to lyse red blood cells and release haemoglobin (Caza and Kronstad, 2013).

1.2 Bacterial acquisition and exportation of metal ions

The acquisition and the intracellular import of metal ions including iron into the cytosol of bacterial cells must predominantly be performed by transporter systems, as transition metal cations, oxyanions or metallo-complexes are unable to diffuse through the bacterial membranes or associated porin channels.

1.2.1 Iron uptake in Gram-negative bacteria

Gram-negative bacteria often utilise high-affinity outer membrane receptors to acquire and transport iron through the outer membrane into the cytoplasm by binding to a range of glycoproteins including transferrin, lactoferrin and ferritin; iron specific chelatory peptide molecules known as siderophores or other small molecule chelators such as citrate, or salicylic acid (Caza and Kronstad, 2013; Ratledge and Dover, 2000; Wandersman and Delepelaire, 2004). The transport of iron is facilitated by attachment of the complex with a specific outer membrane receptor; that are high-affinity to facilitate scavenging from the environment and allow transfer into the periplasm, with subsequent periplasmic and cytoplasmic translocation as depicted by the example in Figure 1-1.

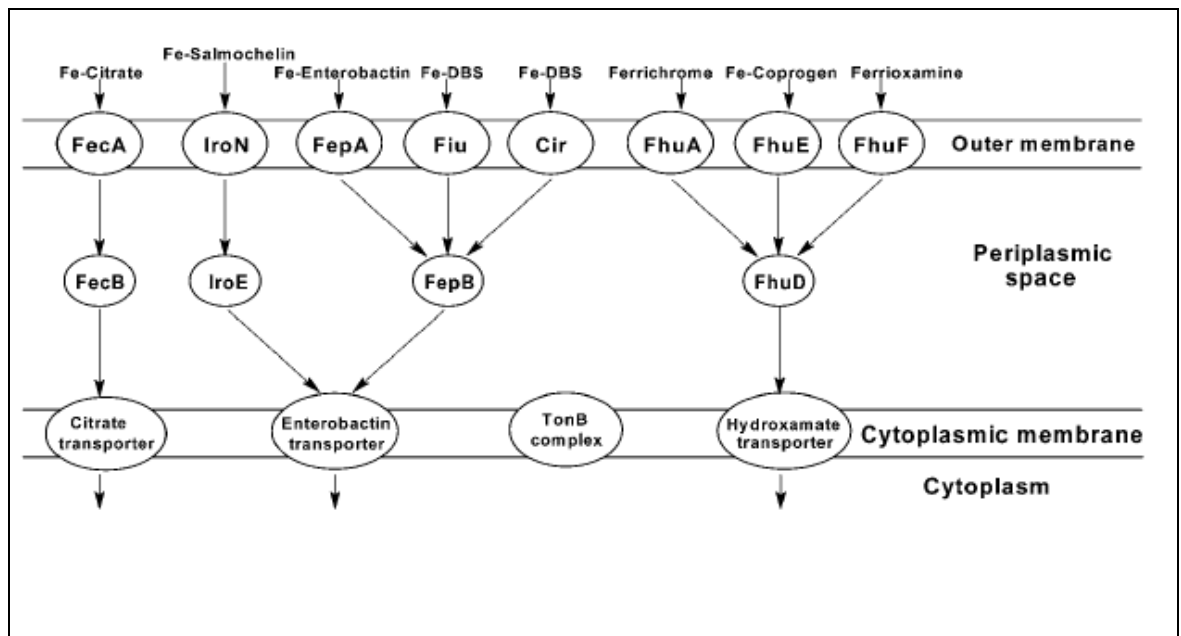


Figure 1-1: An example of an iron (III) transport system utilised by *E. coli* and other Enterobacteriaceae.

The range of mechanisms employed used to transport iron-siderophore complexes through the periplasm and associated membranes into the cytoplasm. Cytoplasmic iron-siderophore transporters utilise a standard ABC-type transporter mechanism, pushed by ATP-hydrolysis (Hider and Kong, 2010).

1.2.1.1 Outer-membrane receptors

These outer-membrane receptors are employed as a result of the ferric-complexes exceeding the maximum size for translocation using a channel-forming porin and therefore require a specific ferric-uptake mechanism. Ferric-siderophore complex associated uptake is often mediated in the form of a conserved 22-stranded antiparallel β -barrelled receptor with a *N*-terminal plug domain as seen in Figure 1-2.

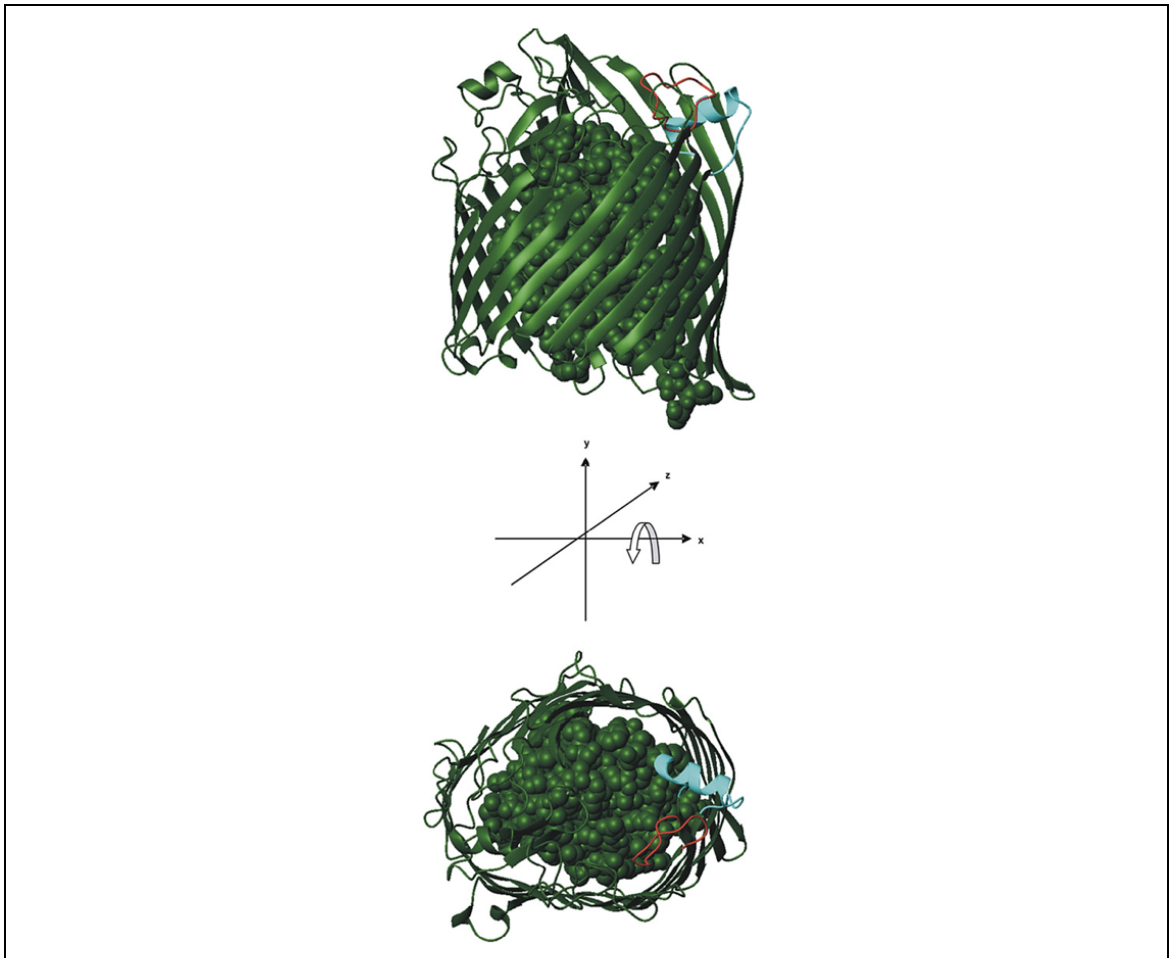


Figure 1-2: A cartoon representation of the outer membrane receptor FecA (1KMO) (ligand free) (Krewulak and Vogel, 2008).

The outer membrane receptor is displayed the form of a ribbon structure, showing the 22- β strand barrel (ribbon) and N-terminal plug domain (space filling).

Importantly, the β -barrels extend beyond the Gram-negative lipid bilayer, and regardless of sequence similarity, these transporters are structurally similar when the $C\alpha$ backbones of the β -barrels are compared, and to maintain correct protein folding and insertion into the outer-membrane there is a highly conserved C-terminal phenylalanine (or tryptophan) residue (Struyve et al., 1991).

The transportation of the iron-complexes through outer membrane is driven by the proton motive force at the cytoplasmic membrane mediated by the TonB, ExbB and ExbD protein complex, this is due to the physical separation of the outer membrane from the plasma membrane, and the necessity for substrate translocation (Krewulak and Vogel, 2011). ExbB and ExbD form the proton channel that energizes uptake through TonB, which facilitates the interaction between the *N*-terminal plug of the outer-membrane receptor and TonB causing a conformational change to release the substrate into the periplasmic space (Moeck and Coulton, 1998).

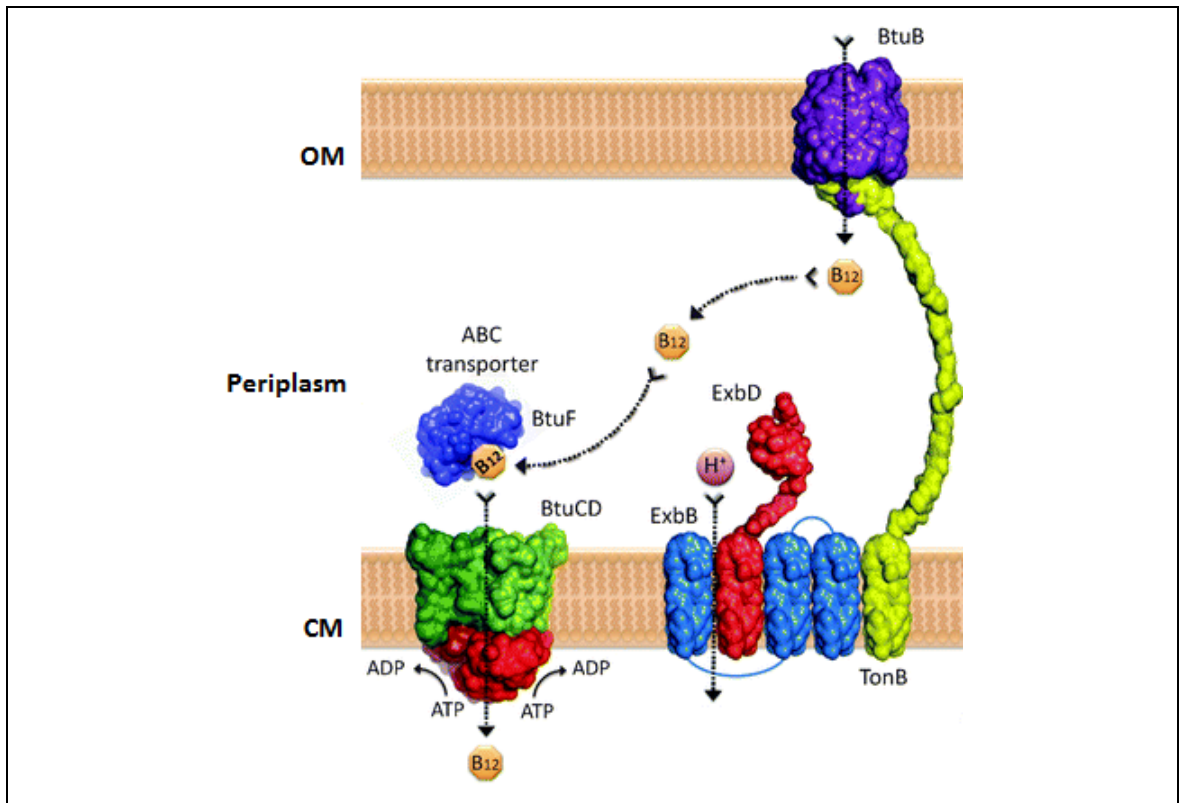


Figure 1-3: The TonB-mediated transport mechanism employed to facilitate transition metal translocation through the cell envelope in Gram-negative bacteria (Klein and Lewinson, 2011).

1.2.1.2 Periplasmic binding proteins

From the periplasmic space, the iron-complexes are further bound by specific binding proteins and translocated to the cytoplasm, a mechanism enabled by ATP-binding cassette (ABC-type) transporter proteins (Clarke et al., 2001; Hider and Kong, 2010). These periplasmic binding proteins can be classified into groups based upon their sequence similarity, and subsequently their ability to bind to a specific type of ligand (Krewulak and Vogel, 2008). The classifications are listed in the Table 1-1.

Table 1-1 Clustering of periplasmic binding proteins based on ligand specificity.

Adapted from papers by Claverys (2001); Tam and Saier (1993), with particular interest in cluster 8, for siderophore ligand binding proteins.

Cluster	Specific Ligand Binding Proteins
1	Malto-Oligosaccharides, Multiple Sugars, α -Glycerol, Phosphate and Iron
2	Hexoses and Pentoses
3	Polar Amino Acids
4	Aliphatic Amino Acids and Amides
5	Peptides and Nickel
6	Inorganic Polyanions
7	Organic Polyanions
8	Iron Complexes (and Vitamin B12)
9	Manganese and Zinc

Cluster 8 of the periplasmic binding proteins is responsible for shuttling iron-complexed molecules, including iron-siderophore complexes; interestingly, each siderophore-ligand group appears to have a defined periplasmic binding protein responsible for across the periplasm. For example, *E. coli* utilises the proteins FhuD and FepB to bind the siderophore-ligands with hydroxamate and catecholate moieties respectively as shown Figure 1-1 (Hider and Kong, 2010).

As with most documented iron-uptake pathways, once transported into the periplasm, the ferri-siderophore complexes with the specific siderophore-periplasmic-binding protein component of the ABC-type transporter, an example structure is depicted in Figure 1-4 (Ma et al., 2009). The resultant effect caused by permease-component interaction is to import the complex into the cytoplasm facilitated by ATP hydrolysis.

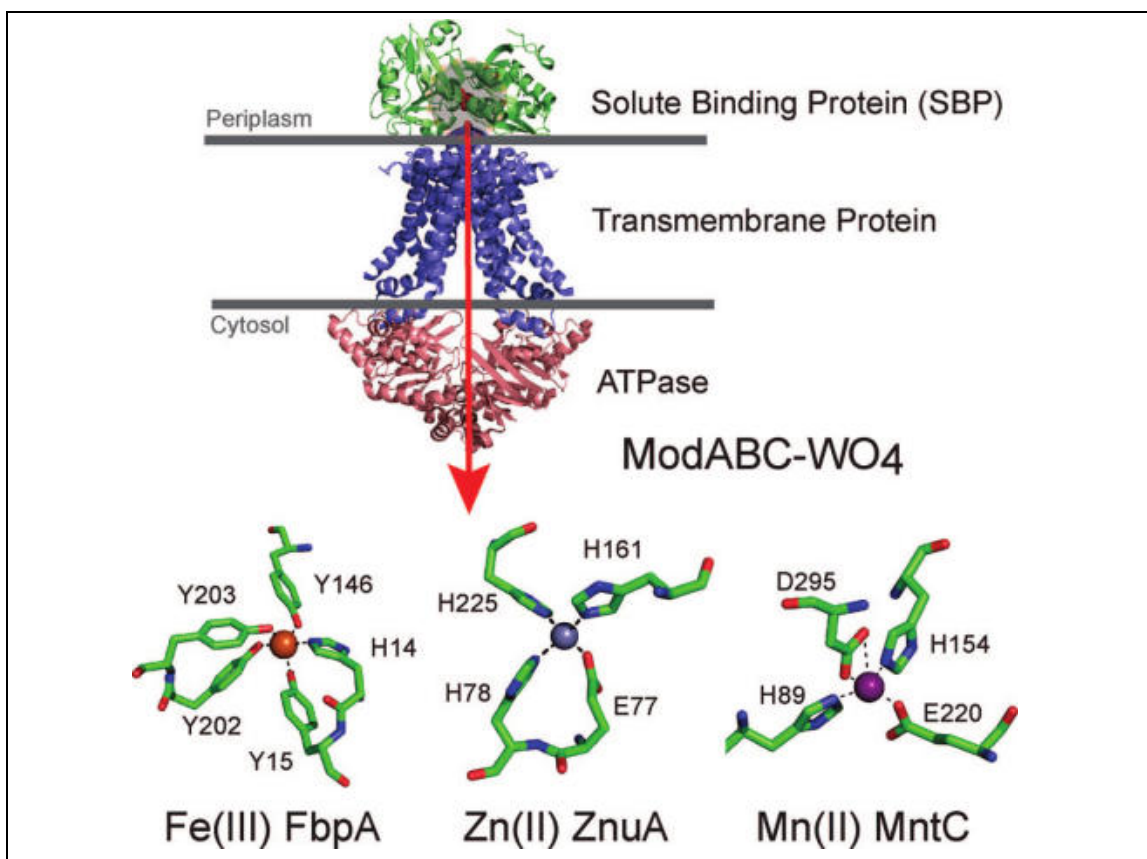


Figure 1-4 A ribbon structure of typical ABC-type transporter.

E. coli ModABC shown above, Solute binding protein ModA (shown in green). This example shows transport of tungstate (WO_4^{2-}) anion (shown in red, with grey background shading, within ModA) (Hollenstein et al., 2007). Other solute-binding-proteins utilise a universal structure with differing metal coordination sites, as shown in the lower section of the figure by Fe(III)-specific solute binding protein *Campylobacter jejuni* FbpA (1Y4T) (Tom-Yew et al., 2005), as well as other relevant metal ion binding sites for Zn(II) and Mn(II).

1.2.2 Iron uptake in Gram-positive bacteria

Unlike in Gram-negative bacteria, there is a significant gap in the literature regarding iron regulation in Gram-positive bacteria. While siderophore compounds have been both well described and characterised in these organisms, the transport mechanisms associated are still widely postulated among the scientific community, especially in the mycolic acid-containing group of actinobacteria.

As archetypal Gram-positive bacteria lack an outer membrane, the mechanisms utilised by the Gram-negative bacteria are likely to be redundant. Nonetheless, Gram-positive organisms have been shown to produce both lipoproteins and ABC-type transporter proteins analogous to those present in Gram-negative organisms (Hider and Kong, 2010). However the 30-50 kDa soluble proteins that are found in the periplasm of Gram-negative bacteria are accessory proteins covalently anchored by lipoproteins to outer leaflet of the plasma membrane (Wilkins, 2015). Interestingly, the soluble section previously mentioned, folds in an identical manner to Gram-negative periplasmic binding proteins, and both anchor onto the transmembrane channel of the transporter, highlighting an analogous method of metal transport (Miethke and Marahiel, 2007).

While it is well established that Gram-positive bacteria have import mechanisms to facilitate uptake of ferri-siderophores, this has recently extended to include transport of apo-siderophores (Fukushima et al., 2013). A study of the protein YxeB of the Gram-positive organism *Bacillus cereus* provided the first direct experimental evidence of a Gram-positive siderophore-shuttle mechanism, in which iron exchanges from a ferri-siderophore complex to an apo-siderophore bound to the substrate binding protein without intermediate metal ion reduction (Fukushima et al., 2014), suggesting that *Bacillus cereus* utilises a novel method of iron importation. The model described by Fukushima et al. (2013) highlights that YxeB may have a two-pathway importation system, preferentially using a siderophore shuttle system, with a secondary displacement strategy (Figure 1-5).

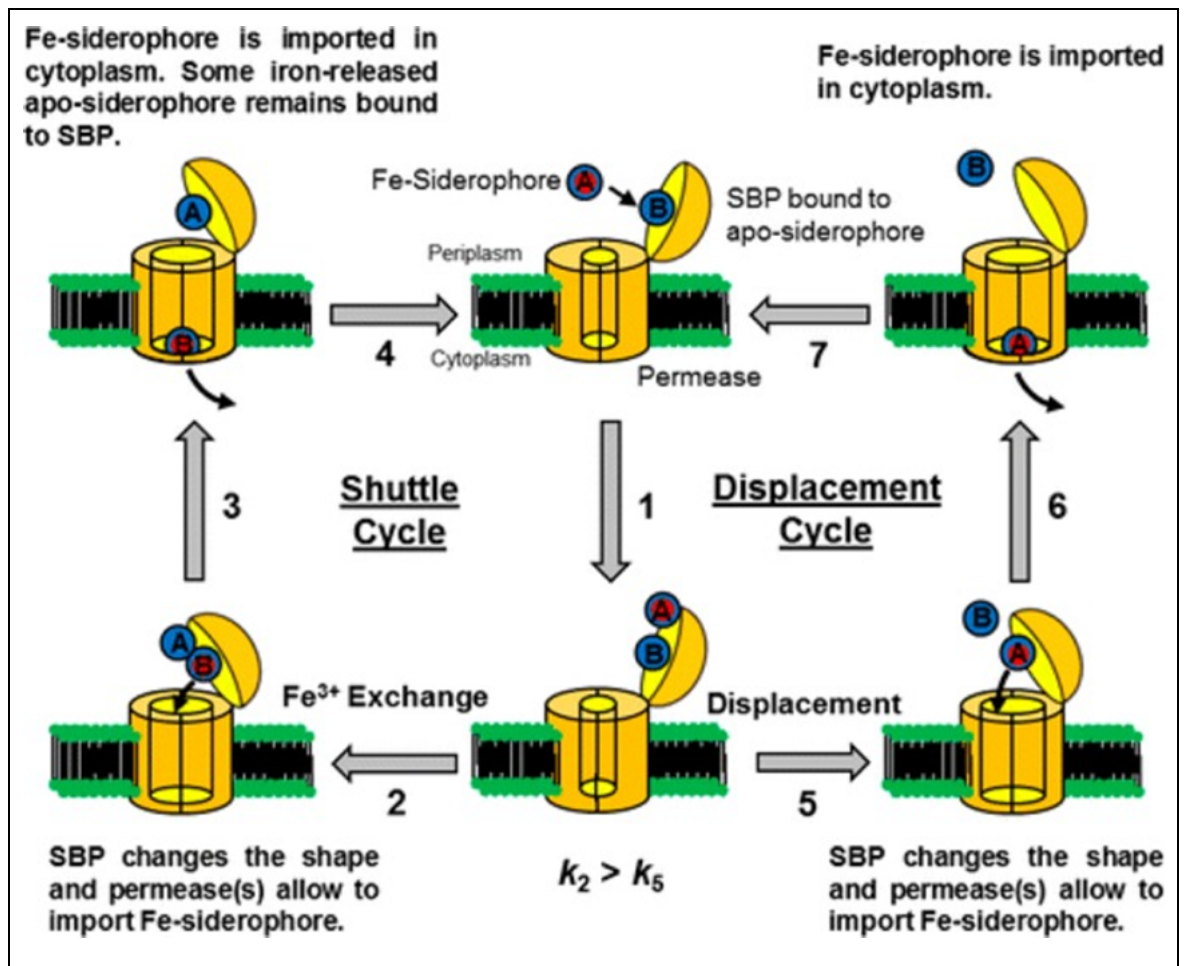


Figure 1-5 Models of the Gram-positive siderophore-shuttle mechanism and displacement mechanism of YxeB (Fukushima et al., 2014).

Both pathways operate in Gram-positive organisms, but the shuttle pathway is preferred. (1) The iron bound siderophore approaches the YxeB-apo-siderophore complex and rests near the binding pocket, with two potential routes of uptake. Steps 2–4 provide the shuttle pathway, Steps 5–7 provide displacement pathway. (2) Iron exchanges from the Fe–siderophore to the apo-siderophore in the binding pocket. The YxeB protein catalyses this step by increasing the local concentration of the entering ligand and the ferric complex. (3) The newly transferred Fe–siderophore complex (B) is transported across the membrane and the original (now apo-) siderophore (A) remains bound to the YxeB protein. (4) The receptor is bound to an apo-siderophore, and the process can repeat. (5) The iron–siderophore complex (A) displaces the apo-siderophore (B) and occupies the binding pocket. (6) The original iron–siderophore (A) is transported across the membrane. (7) The siderophore-binding-protein is bound to an apo-siderophore (B)

1.3 Siderophores

One of the most described methods of iron acquisition is through the use of siderophores. These secreted chelatory compounds are typically of low-molecular weight have a natural high affinity for iron (III) and are employed by microbes to scavenge iron from iron-binding proteins. Once bound, the ferri-siderophore complexes are transported through the use of specific membrane-bound iron-siderophore receptors and associated import systems. After transportation, the iron is dissociated from the siderophore principally through redox-mediated reactions (Pluháček, 2016).

Siderophore biosynthesis is usually regulated by the amount of environmental iron available to the microbe, with siderophore secretion occurring, quite often to outcompete other less-well adapted organisms (Emery, 1982). As a result, the ability to sequester iron, and thereby survive over less well-developed organisms, can change and impact the ecology. Furthermore, it has been documented that a range of plasmids have siderophore biosynthesis genes encoded, and as a result can confer horizontal gene transfer thereby associating secretion of siderophores with both a diverse nature of pathogenicity and virulence (Bullen et al., 1999; Holden and Bachman, 2015; Naka et al., 2013).

1.3.1 Siderophore selectivity and ligand structures

As previously mentioned, siderophores are a group of secreted high-affinity chelators preferential to iron (III), this selectivity relates to the ligand specificity of iron (II), as the nature of the divalent d-block metal cations such as zinc, copper, manganese, and nickel are crucial for many biological processes. Unlike the divalent cations, the trivalent forms are generally inessential, thereby providing a route for iron (III) selectivity with subsequent release occurring *via* reduction (Hider and Kong, 2010), a reaction not possible for other trivalent metals.

Typically, iron-siderophore complexes are formed using oxygen donor atoms presenting an octahedral geometry, with the binding sites located often in specific orientations, which are can be pivotal in ferric-siderophore-complex recognition for cellular import (Roosenberg et al., 2000). The formation of the octahedral complex is generally favourable for binding with iron (III) due to the thermodynamically stable, high spin nature of iron (III), frequently employing a three bidentate-ligand arrangement, to facilitate a hexadentate structure for binding to the iron atom with minimal

ligand repulsion (Hider and Kong, 2010).

As described earlier, iron (III) ions are almost always bound by oxygen-based ligands but can accommodate nitrogen and sulphur ligands, and as such siderophores may have an interrupted geometry due to inclusion of these other donor atoms, thereby potentially impairing the affinity (Roosenberg et al., 2000).

The stereochemistry of ferri-siderophore complexes can vary depending on the ligand structure; for example, a tris-chelate of catechol, (a symmetrical bidentate ligand) can exist as an enantiomeric pair as shown in Figure 1-6, whereas other asymmetrical ligand structures such as hydroxamates can form Δ -*cis*, Δ -*trans*, Λ -*cis* and Λ -*trans* geometric and optical (enantiomer) isomers.

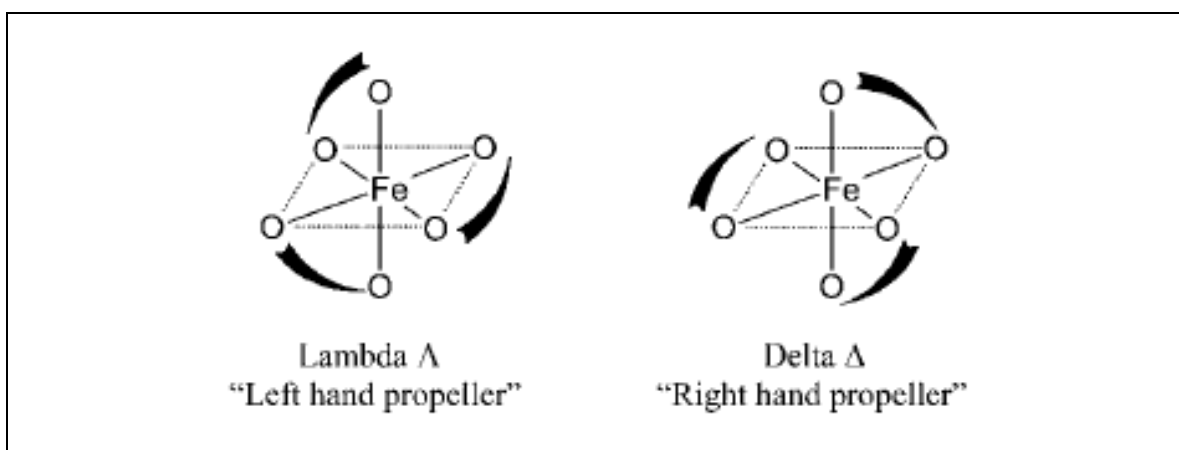


Figure 1-6: An enantiomeric isomer pair, highlighting the optical differences between the two classes (Hider and Kong, 2010).

One of the enantiomers rotates the light in one direction, the other rotates the light in the opposite direction but by the same amount. In all other aspects, including boiling point, refractive index, density and viscosity, the two optical isomers are identical.

Recent reviews indicate that upwards of 500 siderophores have been isolated and characterised, of which they can be classified by the composition and topology of the iron-binding moieties, including catecholates (including phenolates), hydroxamates, and carboxylates, however an ever-emerging group of mixed-type ligands are being discovered (Miethke and Marahiel, 2007). Siderophores can also be classified by their backbone structure or family (Roosenberg et al., 2000). A selection of characterised siderophores are shown in Figure 1-7, grouped by the iron-binding moiety present (Miethke and Marahiel, 2007).

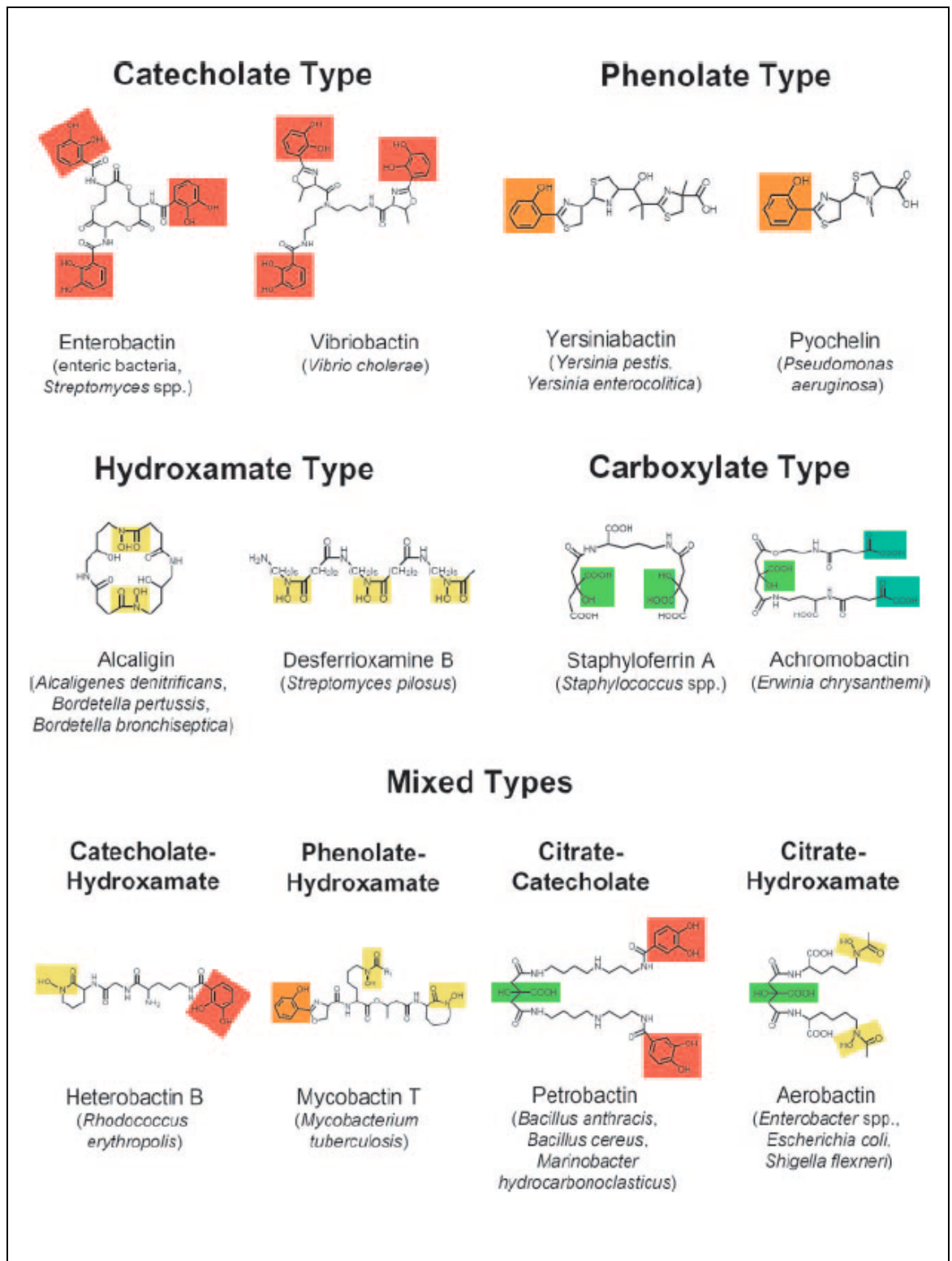


Figure 1-7 Representative examples of different siderophores and their natural producers. Moieties involved in iron coordination are highlighted as follows: catecholates are in red, phenolates are in orange, hydroxamates are in pale yellow, α -hydroxy-carboxylates (deriving from citrate units) are in light green, and α -keto-carboxylates (deriving from 2-oxo-glutarate units) are in blue-green (Miethke and Marahiel, 2007).

The three major classifications of siderophores, catecholates, hydroxamates and carboxylates are shown in Figure 1-8, which highlights the complexation that occurs when the hydroxyl groups are reduced upon iron binding (Gorska et al., 2014).

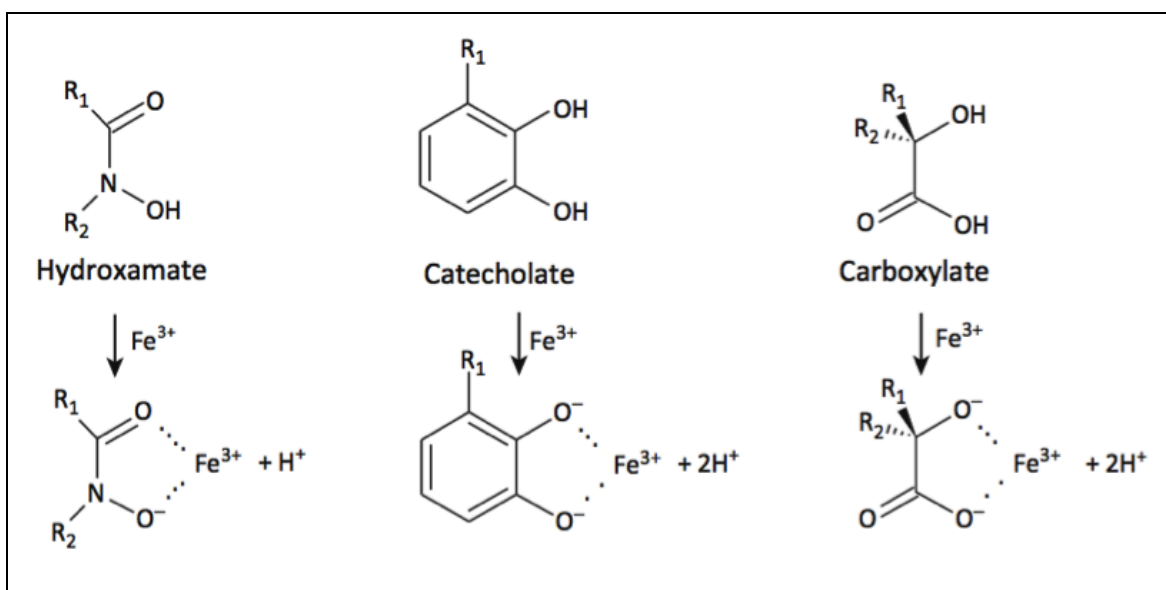


Figure 1-8 Three major ligand groups employed by siderophores, and their complexation with iron (III).

The R groups present on each of the moieties are representative of variable backbone groups.

Interestingly, both metal-binding and siderophore chirality are influential for conformation and configuration of ferric-siderophore complexes and as a result can affect protein–siderophore interactions including the biologically significant mechanism of cellular transport (Raymond et al., 2015).

Prior to iron-binding, siderophore structures are flexible, with the octahedral geometry only present when stabilised by the metal-ligand interaction Figure 1-9. This stability is a result of the strict ligand denticity, and hard acid-base interactions between the iron (III) and siderophore respectively (Miethke and Marahiel, 2007).

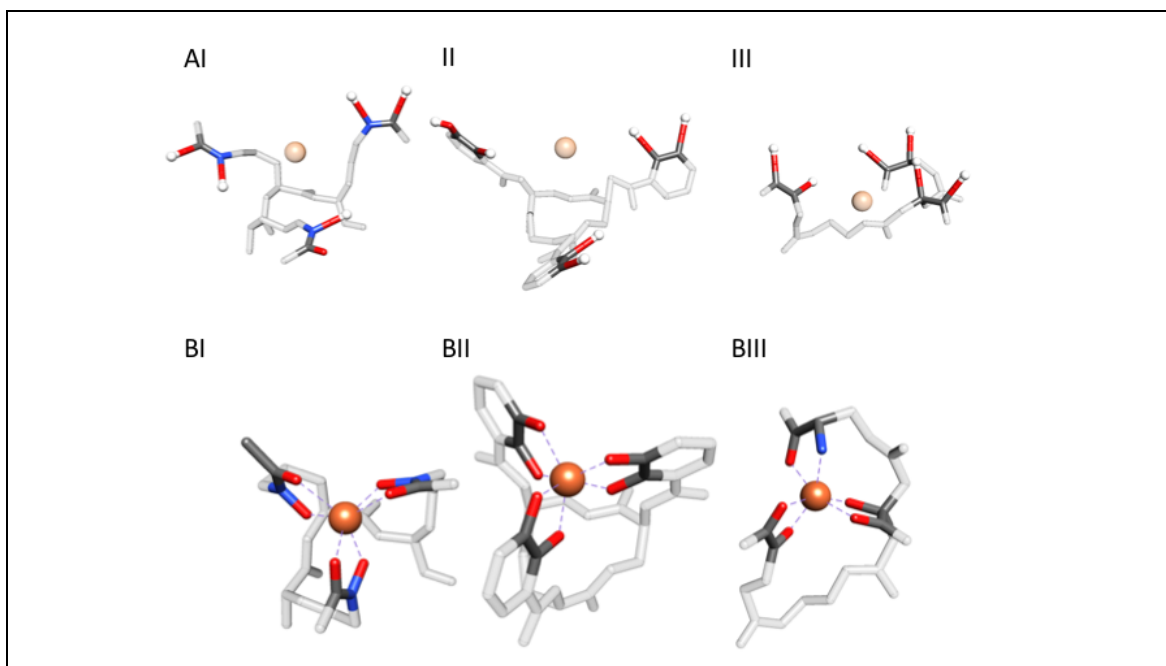


Figure 1-9: Examples of apo- and ferric-Siderophore complex formations.

A) Siderophores unbound to iron highlighting molecule flexibility, B) Siderophores bound to iron producing a stable octahedral geometry in a hexadentate formation. Coloured ligands highlight siderophore ligand groups, grey sections highlight variable peptide backbone structures. I) A hydroxamate ligand from ferrichrome(Locher et al., 1998) utilised by *E. coli*. II) A catechol ligand from enterobactin (Peuckert et al., 2011) utilised by *B. subtilis*. III) A carboxylate ligand from staphyloferrin (Grigg et al., 2010) utilised by Staphylococci.

1.4 Release of iron from siderophore complex

1.4.1 Redox chemistry of siderophores

Siderophores by nature, have a universally predominant affinity for the ferric form over ferrous iron, a structure-function principle essential for ensuring selectivity over other abundant metal ions in the environment. Once ferri-siderophore chelation is achieved, the molecular recognition can occur, and the receptor can facilitate capture, transport and eventual release of the metal ions from the siderophore.

Complexation with the ferric form of iron will prevent the formation of hydroxyl radicals due to the highly negative redox potential, thereby preventing cellular damage. (Hider and Kong, 2010). This damage prevention is achieved as the siderophore ligand-donor groups have a higher selectivity of iron(III) over iron (II) as a result of binding of the negatively-charged oxygen donor groups in to the high-spin Fe^{3+} metal centre, severely reducing the redox potential to highly negative values (Harrington and Crumbliss, 2009).

The redox potential values for siderophore classes can vary, and typically catechol-derived siderophores occupy the most negative redox potential comparative to other groups, while also hexadentate siderophore complexes have a greater negative potential over tetradentate and bi-dentate siderophore complexes. These factors affecting redox potential appear to share a significant relationship with the iron binding capacity of a siderophore – pFe^{III} value, where the greater the pFe^{III} value the more negative the redox potential (Hider and Kong, 2010; Miethke and Marahiel, 2007).

1.4.2 Hypothesis for iron release – redox method

The most commonly accepted hypothesis for iron release is through reduction of the bound iron (III) ion to iron (II), facilitated by a decrease in thermodynamic and kinetic stability promoting ligand exchange. The reduction of the Fe^{3+} to the lower charge density Fe^{2+} ion is likely to promote dissociation from the ligand binding sites to target locations within the cells (Dhungana and Crumbliss, 2005; Hider and Kong, 2010).

The reduction is predicted to occur through changes in cellular environment, where the biological reducing agents such as NADH with electron donor co-factors can be utilised (Dhungana and Crumbliss, 2005). The siderophore denticity is likely to influence the mechanisms required to promote release of iron into the cell as lower negative redox potential siderophores such as α -hydroxycarboxylates and hydroxamate-ligand siderophores are liable to be reduced upon interaction with such reducing agents. However, siderophores with greater negative redox values such as hexadentate tris-catechol-ligand siderophores may require alternative mechanisms or a combination of reduction mechanisms such as protonation.

1.4.3 Alternative hypothesis for iron release – protonation method

The redox potential of iron-siderophore complexes can be further influenced by protonation of the aqueous environment, as competition between Fe^{3+} and H^+ for siderophore ligand groups already occurs. Crumbliss and colleagues have previously shown that a reduction in pH can release Fe^{3+} from thermodynamically-stable siderophore complexes using competitive protonation (Albrecht-Gary and Crumbliss, 1998; Boukhalfa and Crumbliss, 2002) While the environmental pH change may not be sufficient for the H^+ ions to compete with Fe^{3+} for protonation of the negatively-charged donor

groups, the speciation of iron(III)-complexes can be manipulated, where the protonation of ligand donor groups can occur in iron(II)-complexes using proton-driven dissociation and eventual release of the bound iron.

1.4.4 Lipophilicity in iron release

Prokaryotic organisms often allow importation of iron-bound siderophore complexes using lipophilic cell membrane bound receptors including FhuA, and yet other mechanisms may also be involved in an iron exchange process from a hydrophilic extracellular siderophore system to a membrane-bound siderophore system even with an exceptionally negative redox potential as described by Dhungana et al. (2003), highlighting how an environmental lipophilicity caused shift in redox potential may facilitate a cascade event of redox methods to facilitate ligand exchange (Ratledge and Dover, 2000).

1.4.5 Iron regulatory circuits: mediating siderophore production

Research on regulation of iron uptake pathways and associated expression of genes is somewhat limited, with the majority of studies have focussed on the Ferric uptake regulator (Fur) protein family (discussed in detail later) which globally coordinate responses to intracellular iron levels; iron-complexed Fur proteins bind to promoters to directly repress transcription or act *via* alternative indirect activation of alternative regulons (Crosa and Walsh, 2002). Although Fur has been identified as the as the main iron-responsive regulator in a vast range of bacteria, the Diphtheria toxin regulator protein (DtxR) has been identified as the global iron regulator in many high G+C content Gram-positive bacteria (Boland and Meijer, 2000) (also discussed later in the thesis), and may also contribute to iron regulation through other indirect regulatory methods including expression of the diphtheria toxin, haem-associated iron regulation, siderophore synthesis and efflux as well as siderophore dependent-uptake of iron (Brune et al., 2006). Importantly, DtxR and Fur have no significant sequence homology, but both proteins function as Fe²⁺-dependent transcriptional repressors.

As well as coordination chemistry-based ligand-based classification, siderophores can be classified predominantly into two groups based on their biosynthetic production methods, these include production by 1) a non-ribosomal peptide synthetase (NRPS) pathway or 2) NRPS-independent

pathway.

1.4.6 Non-ribosomal peptide synthetases

Non-ribosomal peptide synthetases are multi-modular enzymes responsible for an assembly-line production of specific peptides without the need for an RNA template. Functioning in a similar manner to the fatty-acid synthetase and polyketide synthetase enzymes (Crosa and Walsh, 2002), the NRPS link together predominantly nonproteinogenic amino acids *via* thioester intermediates; the proteins are modular with each module being responsible for the extension of the product of the previous module by a single specific amino acid. Each module consists of an (i) adenylation domain, which selects and activates the extending amino acid *via* adenylation, (ii) a peptidyl carrier protein domain, which maintains the product as a enzyme-linked thioester and a (iii) condensation domain which forms the peptide bond. The product may be passed to successive NRPS proteins for further extension and the production line often is terminated by a thioesterase domain which releases and often cyclises the final product.

The extension of the peptide product is colinear, with the *N*-terminal amino acids of the product being condensed at the first module on the *N*-terminal region of the NRPS and additions being made to the *C*-terminus as the peptide progresses towards the *C*-terminus of the NRPS. The thiol groups that tether the growing product at the peptidyl carrier domains are 4-phosphopantethinyl groups introduced as post-translational modifications of a conserved serine residue that are long and flexible enough to promote the transfer of the intermediates to the spatially distinct reaction centres (Mootz et al., 2001) in the condensation domains of neighbouring modules. It is during this thioester intermediate stage that any additional editing such as epimerisation or N-methylation can be performed.

The chain elongation occurs upon peptide bond formation catalysed by a condensation domain which facilitates the connection of the free amine group of a downstream PCP-bound acceptor amino acid with the activated upstream thioester PCP-bound donor (Crosa and Walsh, 2002).

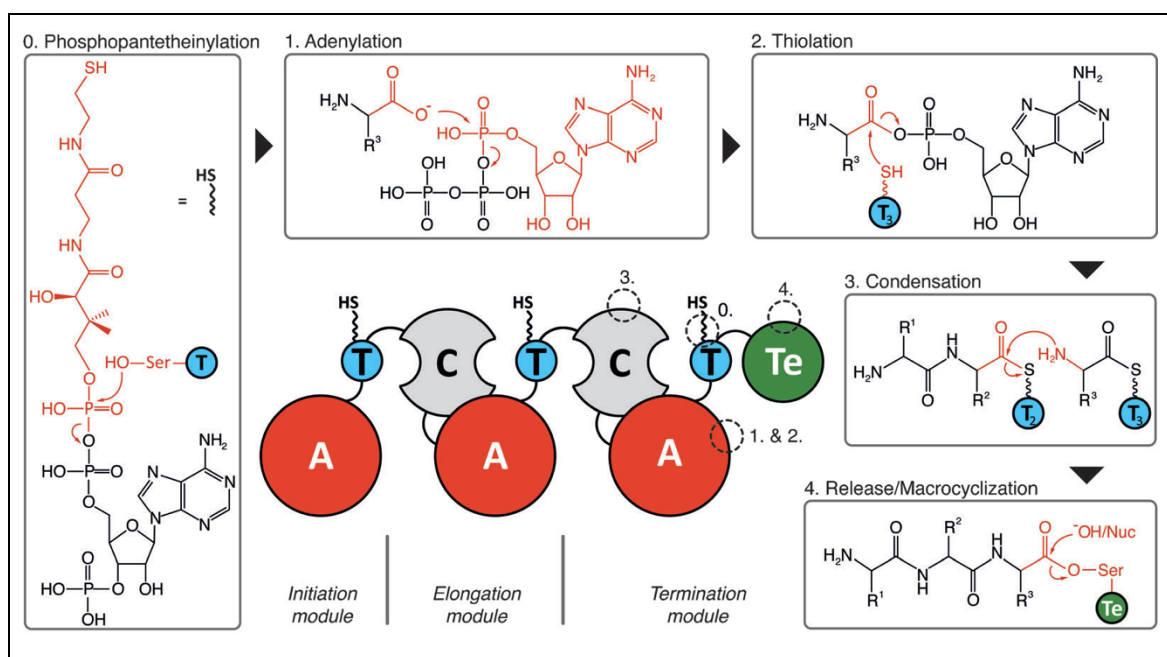


Figure 1-10 The domain arrangement of bacterial non-ribosomal synthetases and associated catalytic reactions (Süssmuth, 2017).

0) NRPS priming: phosphopantetheinyl transferase-mediated installation of phosphopantetheine at a conserved serine of the apo-T domain. 1) Selection and adenylation of the amino acid by the A domain to produce an aminoacyl-AMP species and pyrophosphate. 2) Thiolation of the activated amino acid and release of AMP to produce an aminoacyl-thioester attached to the phosphopantetheine of the holo-T domain, catalysed by the A domain. 3) Peptide bond formation by the Condensation domain couples the activated amino acid to the amino acid attached to the upstream module. 4) Oligopeptide release is achieved by formation of an intermediate ester bond between the C terminus of the peptide and a conserved serine of the Te domain. Hydrolysis or intramolecular attack of a nucleophilic moiety yields a linear or macrocyclic product, respectively. Reaction products given in red. (Abbreviation : Nuc = nucleophile).

1.4.7 Synthesis of siderophoric compounds within NRPS mechanisms

As previously described, siderophores are usually classified by the type of functional ligand architecture present used to bind the extracellular iron. The catechol and phenolic ligand groups typically utilise 2,3-dihydroxybenzoic acid and salicylate respectively, which are synthesised using phenylalanine, tryptophan and tyrosine precursor amino acids (KEGG Pathway, Figure 1-11), and as such a isochorismate synthetase gene is usually present within the siderophore biosynthetic gene cluster (Lamb, 2015). Hydroxamate ligand-bearing siderophores, however, are produced within the NRPS mechanism by modification of lysine, ornithine or polyamine residues, when a free side chain undergoes hydroxylation and formylation of the nitrogen molecule (Lamb, 2015). The production of these ligands is usually catalysed by a mono-oxygenase (Hider and Kong, 2010), while other groups produced including carboxylate ligands, are occasionally attached as preformed compounds such as citrate.

BIOSYNTHESIS OF SIDEROPHORE GROUP NONRIBOSOMAL PEPTIDES

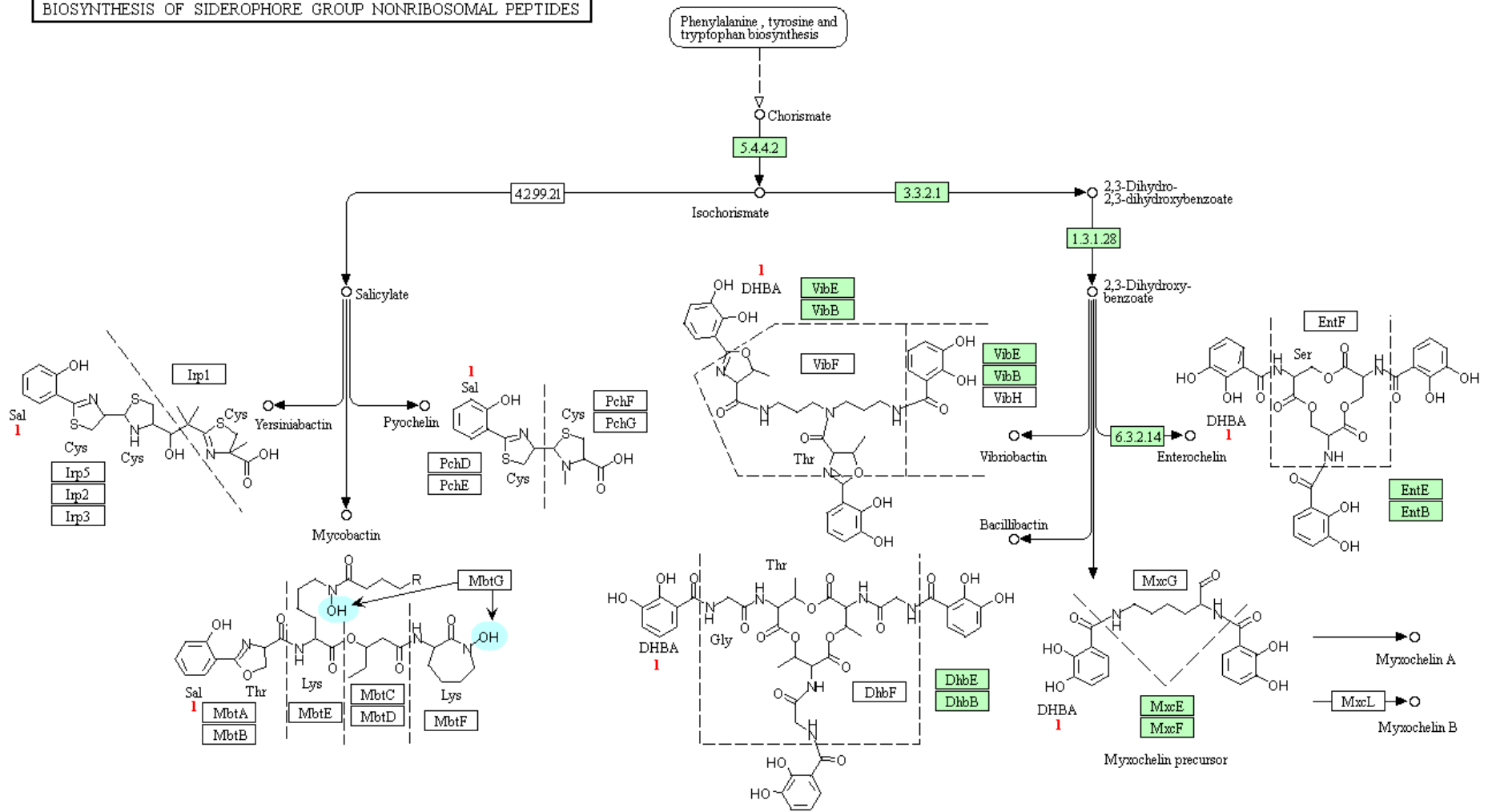


Figure 1-11 KEGG Pathway for biosynthesis of bacterial catechol siderophores using phenylalanine, tryptophan and tyrosine precursor amino acids.

1.5 Introduction to the *Rhodococcus* genus

The genus *Rhodococcus* was first proposed in 1891 by Zopf, as a reclassification of *Micrococcus erythromyxa* and *Micrococcus rhodochrous* (Overbeck, 1891; Zopf, 1891) based on the production of red pigment, with the species reclassified a number of times further (Buchanan, 1915). Due to the lack of suitable characteristics for the genus description, the strains were, once again, included in the genus *Micrococcus* within Bergey's *Manual of Determinative Bacteriology: Fifth and Sixth Editions*. The genus *Rhodococcus* was re-established by Goodfellow and Alderson in 1977, by identification of nine species that could be described to a greater extent through the use of a more robust advanced numerical taxonomic study (Goodfellow and Alderson, 1977). As of 2013, 33 unique species were assigned to the genus *Rhodococcus* (Jones et al., 2013a).

Characterised as a nocardioform actinomycetes, rhodococci, like other closed related genera are defined by their mycolic acid-containing cell wall and G+C-rich genomic content (Bell et al., 1998; Zhi et al., 2009). The genus can be identified as Gram-positive, non-motile, obligate aerobes. The morphological variations within *Rhodococcus* are diverse; species are often pleomorphic and may produce filaments (Alvarez, 2010) which can readily lead to a misidentification of analogous genera within the order *Actinomycetales* (Savini et al., 2012).

Classification within the genus *Rhodococcus* is further validated by presence of a distinctive cell envelope feature, the arabinogalactan layer, a covalently attached cell wall polysaccharide (Sutcliffe et al., 2010) that links mycolic acids to peptidoglycan, this mycolate-containing cell envelope region is a defining feature that facilitates survival under tough environmental conditions. The complexity and size of these 2-alkyl branched 3-hydroxy long chain fatty acids can vary within *Rhodococcus* and are comparable to other mycolate-containing bacteria such as *Mycobacterium* and *Corynebacterium* (Figure 1-12). Similar to *M. tuberculosis*, the high hydrophobicity of the rhodococcal cell envelope frequently functions as a barrier to prevent entry to the cell including antibiotic treatments (de Carvalho et al., 2014). Transport across the mycolate-layer can be achieved if the molecule is hydrophobic, otherwise active transports systems are typically required. Literature regarding membrane transport is typically focussed on translocation of the cytoplasmic membrane as for typical gram-positive bacterium, without consideration of the mycolic acid-containing cell

wall region.

R. equi has a cell wall chemotype IV defined by the presence of arabinogalactan, galactose and *meso*-2,6-diaminopimelic acid to provide a scaffold for covalent anchoring of the mycolates (Schaechter, 2009), with A1 γ peptidoglycan also promoting stability within the cell wall. Although other similar genera from Family *Corynebacteriaceae* have a chemotype IV cell wall (Jones and Goodfellow, 2012), rhodococci can be readily distinguished through chemotaxonomic characteristics including 10-methyl tuberculostearic-branched fatty acid chains; a phospholipid profile containing di-phosphatidylglycerol, phosphatidylinositol, phosphatidylethanolamine, and phosphatidylinositol mannosides, and non-cyclised dehydrogenated menaquinones (Bell et al., 1998; de Carvalho et al., 2014).

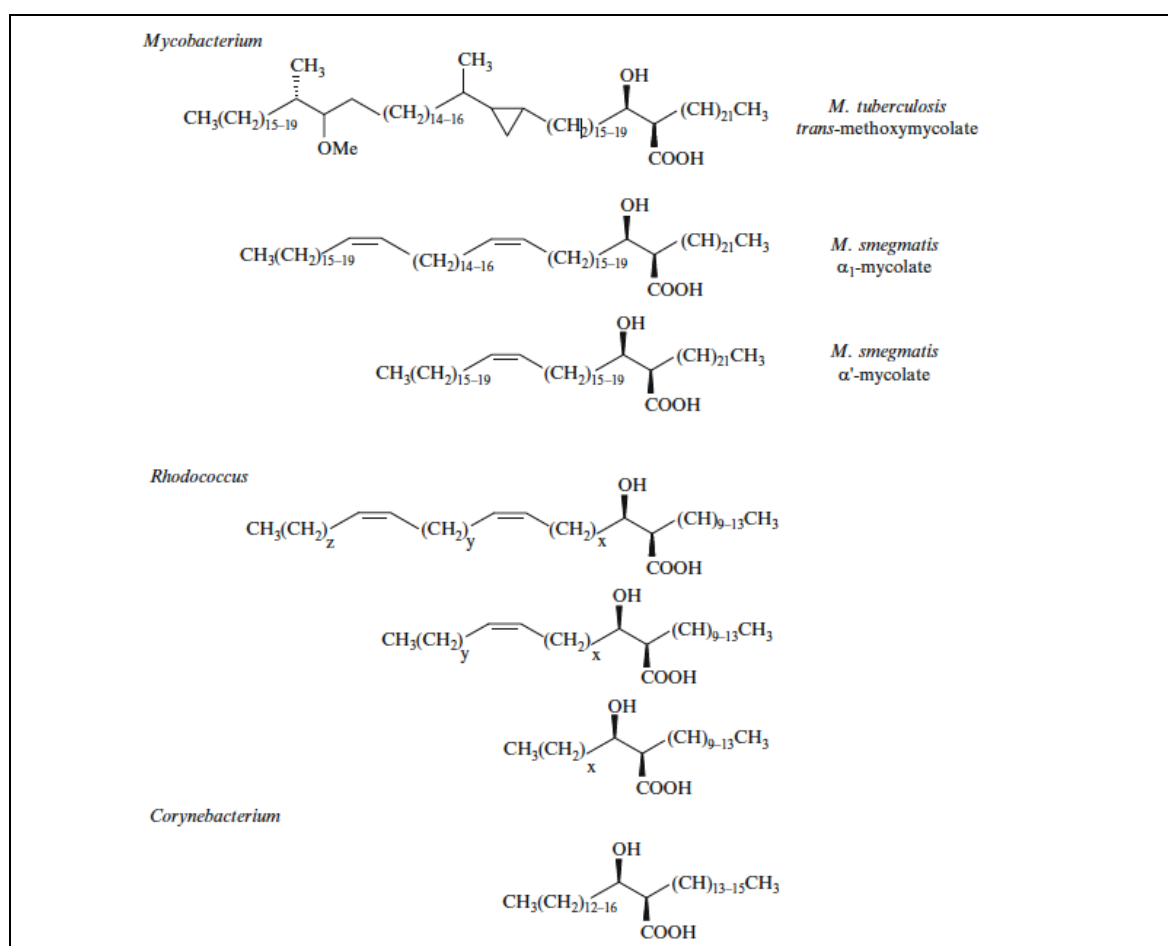


Figure 1-12 Comparison of mycolic acid structures in the suborder *Corynebacteriales*.

This comparison emphasises contrasting complexities. Typically, rhodococcal mycolates (ester of a mycolic acid) are intermediary sized displaying an aliphatic 2-alkyl chain, with 12-16 carbon molecules whilst the 3-hydroxyl meromycolate typically contains 18–40 carbons denoted by x, y and z; comparative to the complex *M. tuberculosis* methoxymycolate ranging between 18-40 carbons. *M. smegmatis* mycolates encompass either double or single unsaturations whilst retaining a longer chain length, comparative to rhodomeromycolates that contain simple modifications for example, four unsaturations (Sutcliffe et al., 2010).

Over the years, rhodococci have been identified as ubiquitous, with isolation occurring in a wide range of environments (Kuyukina and Ivshina, 2010), including extreme conditions such as an oil polluted desert in Kuwait (Radwan et al., 1995), and Antarctic soil (de Carvalho et al., 2014). The majority of isolations emanate from contaminated materials including aquatic sediment, ground water, soils, and manure (Bell et al., 1998), from which persistent xenobiotic contamination or uptake of environmental organic compounds are often used as a sole carbon source (Gürtler and Seviour, 2010). The degradation of xenobiotics and recalcitrant contaminants such as petroleum hydrocarbons (Kuyukina and Ivshina, 2010) by rhodococcal species is of particular commercial interest, as typically they are not naturally degraded in the environment, highlighting the significance of using rhodococcal species as a bioremediation tool. Other commercial uses for *Rhodococcus* species include the production of acrylamide from a purified recombinant nitrilase enzyme originally encoded by *Rhodococcus rhodochrous* (Yamada and Kobayashi, 1996).

The genus *Rhodococcus* contains only a handful of pathogenic species, such as *Rhodococcus fascians*, a phytopathogen causing leafy gall disease in plants (Putnam and Miller, 2006) and *Rhodococcus equi*, the opportunistic pathogen promoting infection in animals and immunocompromised human patients (Jones et al., 2013a; Savini et al., 2012; Topino et al., 2010).

1.6 *Rhodococcus equi*: the virulent zoonotic pathogen

Rhodococcus equi was first isolated in 1923 from pyogranulomatous bronchopneumonia encountered in Swedish foals (Magnusson, 1923). The majority of research regarding this pathogen has centred on the ability to cause pyogranulomatous infections in equine species (Vazquez-Boland et al., 2009). However, *R. equi* is a well-known multi-host pathogen infecting a wide array of other animals including cattle, swine, cats, dogs, wild birds and humans (Makrai et al., 2002; Muscatello et al., 2007; Takai et al., 2003; Topino et al., 2010). *R. equi* is a facultative intracellular pathogen commonly isolated from animal dung, soils and dust particles, and is primarily found in environments such as horse breeding farms that contain these elements.

The pathogen can be spread by aerosol inoculation, through either dust inhalation or droplet transmission between hosts, providing potential infection routes for humans and animals alike (Mosser and Hondalus, 1996).

1.6.1 Equine *R. equi* infections

R. equi infections typically occur in foals when weaning, as maternal antibodies from colostral milk are reduced. Furthermore, the foal immune system is not fully developed, leading to a severely increased infection rate between 6 and 12 weeks. The most common clinical manifestation of infection is chronic pyogranulomatous bronchopneumonia in conjunction with abscessation (Giguere and Prescott, 1997) (Figure 1-13).



Figure 1-13: A lung from a foal suffering from purulent pneumonia with numerous abscesses caused by *R. equi* infection (Vazquez-Boland et al., 2013).

Early symptoms begin with mild fevers and slight cough, and can progress with an irregular respiratory rate however, this may not be discernible early on unless exposed to stress factors or exercise. As a result, lung function can deteriorate over time and may not present symptomatically (Vazquez-Boland et al., 2013).

These signs, if missed can allow for disease progression, therefore identification of chronic conditions can appear as acute, termed ‘acute on chronic’ (Giguere and Prescott, 1997). Clinical signs may include lethargy with decreased appetite, fever or increased effort when breathing. Other clinical manifestations of *R. equi* can include intestinal infections including ulcerative colitis and typhlitis with granulomatous or suppurative mesenteric lymphadenitis (Vazquez-Boland et al., 2013).

In nearly all infected foals, a surface antigen, VapA encoded on an 85–90-kbp virulence plasmid, mediates the virulence of *R. equi*, (discussed in 1.6.4).

Although *R. equi* disease is common in foals, it is unusual disease in adult horses with occasional

reports of lung and lymph node infections. A rare example of an *R. equi* infection in an adult horse included a septicaemia and lung abscessation in conjunction with acquired immunodeficiency (Freestone et al., 1987).

Importantly, *R. equi* can be cultured from almost all surfaces in horse breeding environments, it is interesting that the disease in some farms can be endemic, whilst others remain disease-free (Giguère et al., 2011). The costs associated with endemic *R. equi* diseases can be great therefore appropriate steps towards disease management and treatment should be taken.

Treatment of *R. equi* infections can be troublesome due to a range of natural antimicrobial resistance mechanisms the pathogen has developed, as *R. equi* is typically isolated from soil, the organism's genome has evolved to contain a number of chromosomal resistance mechanisms including resistance to oxidative stress and low pH in the soil (Pei et al., 2007), which can reduce the efficacy of natural soil-borne antimicrobials (Vazquez-Boland et al., 2013). Additionally, the cell wall of the pathogen, specifically the hydrophobic mycolic acids may provide intrinsic resistance preventing internalisation of the drugs.

The administration of rifampin and erythromycin for *R. equi* infections was the preliminary drug choice in the 1980's (Sweeney et al., 1987; Vazquez-Boland et al., 2013) established from *in vitro* susceptibility and pharmacokinetic statistics (Giguère et al., 2011). The combination therapy of rifampin and erythromycin has an active bacteriostatic effect on *R. equi* infections, subsequently requiring the continuation of treatment until the bacterial infection can be cleared by the host immune system (Giguère et al., 2012; Vazquez-Boland et al., 2013). Further to this, a case study by Gurel and colleagues presented adjusted treatment regimens to include Streptolysin-O, an oxygen-labile haemolytic endotoxin in addition to the standard combination therapy of rifampin and erythromycin. This resulted in a superior treatment by characterised by lowered bacterial counts in the lungs, providing longer survival time in foals, through modulation of site-specific expression of extracellular matrix and inflammatory response genes (Gurel et al., 2013).

The combination of new generations of macrolides including clarithromycin and azithromycin with rifampin has proven to be an effective treatment. Unresolved macrolide and rifampin-resistant strains have been treated using antimicrobial agents including gentamicin and doxycycline (Burton et al.,

2013; Vazquez-Boland et al., 2013). The treatment method using rifampin and macrolides together has a synergistic effect both *in vitro* and *in vivo* and reduces the risk of resistance to *R. equi* emerging for either drug classes (Giguère et al., 2012).

1.6.2 Human *R. equi* infections

Although *R. equi* has primarily been associated with the animal populations, opportunistic infection can occur in immunocompromised humans exposed to *R. equi*-contaminated areas and animals (Prescott, 1991). First reported in 1967, *R. equi* was cultured from an immunocompromised male presenting with a fever and apparent cavity pneumonia and a subsequent abscess (Golub et al., 1967). Treatment consisted of erythromycin for 8 weeks, followed by a further 6 weeks upon discovery of abscesses, resolving the symptoms and infection.

Over the years, the incidence of *R. equi* infections has increased coincident with the immunodeficiency associated with cancer therapy, organ transplantation and infection with the Human Immunodeficiency Virus, and subsequent acquired immune deficiency syndrome (Weinstock and Brown, 2002). As a result of their compromised immune function, these patients are at greater risk of *R. equi* infections, with approximately only 10 – 15 % of infections occurring in immunocompetent hosts (Kedlaya et al., 2001). Typically 80 % of total *R. equi* infections affect the pulmonary system and can present with symptoms including nodular infiltrates or a pneumonic consolidations and pulmonary cavitation often leading to the misdiagnosis of pulmonary tuberculosis (Vazquez-Boland et al., 2013; Weinstock and Brown, 2002). Misdiagnosis appears to be common especially in developing countries that have inferior laboratory facilities, this results in the application of ineffective treatment regimes, facilitating the emergence of drug resistance (Gray et al., 2000).

1.6.3 Reclassification of *R. equi*

The taxonomic position of *R. equi* has been recently subject to review, as a study performed by Jones and colleagues provided presented evidence that *R. equi* is reasonably differentiated from other rhodococcal species (Jones et al., 2013a), while another taxonomic revision identified *Rhodococcus hoagii* as a validly published synonym of *R. equi* (Kampfer et al., 2014).

The use of chemotaxonomic, molecular systematic and numerical taxonomic data provided sufficient evidence to reclassify *R. equi* into a new proposed genus: *Prescottia*, type species *Prescottia equi* (Jones et al., 2013a) however the genus *Prescottia* was deemed illegitimate due to previous use within the family Orchidaceae. As a result the new genus name *Prescottella*, type strain *Prescottella equi* was proposed (Jones et al., 2013b).

The proposal of the new bacterial taxon *Prescottella equi* was supported by an investigation performed by Sangal and colleagues, that analysed the phylogenetic relationship between *R. equi* strains, other rhodococcal species and nocardial species (Sangal et al., 2019; Sangal et al., 2014). Sangal's series of phylogenetic trees (Figure 1-14) strongly agrees for the differentiation of *R. equi* species from other Corynebacteriales, supporting reclassification of *R. equi* as *Prescottella equi*.

However, the genus name *Prescottella* cannot be validated until the Judicial Commission reports on whether the species epithet *equi* should be conserved over the nomenclatural variant *hoagii*. (Garrity, 2014).

1.6.4 Pathogenicity and virulence of *R. equi*

Virulent isolates of *R. equi* from equine pulmonary infections harbour an 85 – 90 kb plasmid that carries a pathogenicity island (PAI) encoding virulence-associated proteins (Letek et al., 2010; Meijer and Prescott, 2004; Vazquez-Boland et al., 2013). Although some virulence mechanisms are not characterised, a significant component for pathogenicity is the PAI plasmid. Plasmid-deficient strains are unable to replicate in macrophages *in vitro* or in mice (Meijer and Prescott, 2004). Containing 69 open reading frames, the PAI plasmid comprises four regions; devoted to replication, conjugation, and pathogenicity, while the function of the fourth is unknown.

The series of *vap* genes present in the PAI, encode virulence associated proteins A, C, D, E, F, G, H (Meijer and Prescott, 2004). VapA is a 17.4-kDa surface lipoprotein antigen (Byrne et al., 2001; Letek et al., 2008) and is thought to contribute to scaffolding within a protein complex responsible for interaction with host macrophage components (von Bargen et al., 2009). The deletion of *vapA* diminishes the virulence of the pathogen, further highlighting the contribution of VapA in *R. equi* pathogenicity. A study by Ocampo-Sosa and colleagues developed a useful PCR-based typing method, demonstrating clear associations between plasmid types and animal hosts (Ocampo-Sosa et al., 2007). VapB, isolated from a plasmid variant associated with swine, shares a 78 % amino acid sequence identity with VapA, while a bovine equivalent VapN has also been isolated (Valero-Rello et al., 2015), and are thought to be allelic variations of one locus, that have evolved to produce different Vap-containing plasmids (Ocampo-Sosa et al., 2007). The zoonotic nature is highlighted by the presence of VapA VapB and VapN in *R. equi* strains commonly isolated from immunocompromised human patients. Ocampo-Sosa proposed the hypothesis that host tropism is determined by the plasmid variant harboured by the bacterium, supported by a further study demonstrating a conserved backbone but different *vap* pathogenicity island components (Letek et al., 2008).

1.6.5 Project Aims

Rhodococcus equi is a pathogenic representative of the mycolate-containing actinobacteria. Little is understood regarding the transport processes shuttling materials across their important outer membrane structure. Our over-arching hypothesis is that characterisation of key transmycolate transport processes will identify new transporter archetypes that will illuminate the functional biology of this membrane. This study aimed to further characterise iron acquisition systems as these are considered critical in establishing infection and are worthy of further detailed study in their own right.

The main aim for the research project was to identify and characterise the molecular components of siderophore-mediated iron acquisition in *Rhodococcus equi*. This requires mapping the complex iron regulatory network in the bacterium, primarily by focussing on the iron homeostasis mechanisms, as well as attempting to identify the associated transmycolate cell wall transport mechanisms utilised for iron uptake by their co-ordinated over expression with the siderophores. Currently, the literature suggests *R. equi* synthesises two siderophores under iron starvation; a catecholate-hydroxamate mixed type-siderophore termed rhequibactin (Miranda-CasoLuengo et al., 2008) and a hydroxamate siderophore termed rhequichelin (Miranda-CasoLuengo et al., 2012), however both regulation and uptake of these siderophores are uncharacterised. Genes encoding ABC importers that would mediate transport into the cytoplasm are evident in siderophore gene clusters but candidates for transmycolate uptake pathways have not been identified.

Canonically, bacterial metal ion homeostasis is regulated *via* a series of metal-binding transcriptional repressor proteins. The initial project objectives were identification of any metalloregulatory proteins that might regulate transcription of essential virulence factors including the biosynthetic machinery that synthesise siderophores and associated transport mechanisms. Complementary approaches would be explored to characterise their associated regulons and metallic co-regulators. Overexpression would allow protein purification and functional analysis such as DNA-binding and metal-binding studies. Furthermore, unlike most bacteria that generate an exaggerated response to iron limitation *in vitro*, *R. equi* appears to produce a very small yield of its siderophores, which has complicated their characterisation. Therefore, in order to address the hypothesis that complete

deregulation might generate a greater siderophore yield, production of in-frame deletion mutations of the putative metal regulator genes may facilitate this and also might reveal co-regulated uptake components *via* integrated post-genomics analysis of the mutants

Ultimately, analysis of data from combined studies of mutant and iron-starved wild-type cells would allow, 1) identification and characterisation of *R. equi* siderophores, using a metabolomics approach. 2) identification of mycolate-associated proteins associated with the transmycolate uptake of ferri-siderophores using a proteomic approach and. 3) mapping of the global gene expression of *R. equi* under iron limitation and overlaps with the regulons of transcriptional regulators to identify candidates for these and further roles in *R. equi* iron homeostasis.

2 Materials and Methods

2.1 Bacterial strains and plasmid vectors

The bacterial strains and plasmid vectors used in this thesis are listed in **Error! Reference source not found.** Plasmid maps are available in the Appendices for Chapter 2.

Table 2-1: List of bacterial strains and plasmid vectors

Strain/Vector (supplier)	Genotype	Supplier
<i>Escherichia coli</i>		
XL-10 GOLD	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac The $\Delta(\text{mcrA})183$ $\Delta(\text{mcrCB-hsdSMR-mrr})173 \text{ tet}^R \text{ F}'[\text{proAB lacI}^q\text{Z}\Delta\text{M15}$ Tn10(Tet ^R Amy Cm ^R)]	Stratagene
Ecloni 10G	F- mcrA $\Delta(\text{mrr-hsdRMS-mcrBC})$ endA1 recA1 $\Phi 80\text{dlacZ}\Delta\text{M15}$ $\Delta\text{lacX74 araD139 } \Delta(\text{ara,leu})7697 \text{ galU galK rpsL nupG } \lambda\text{- tonA}$	Lucigen
BL21 (DE3)	B F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)</i> $\lambda(\text{DE3 } [\text{lacI lacUV5-}$ <i>T7p07 ind1 sam7 nin5])</i> [<i>malB</i> ⁺] _{K-12} (λ^S)	Invitrogen
<i>Rhodococcus</i>		
<i>Rhodococcus equi</i> 103S		Prof I. Sutcliffe
<i>Rhodococcus erythropolis</i> PR4 (=NBRC 100887)		NBRC
Plasmids		
pUC18	Bacterial cloning, vector Synthetic construct (L09136) oriColE1 Amp ^R lacZ α M13mp18 polylinker	Norranders et al. (1983)
pET23b	Bacterial expression vector, T7 tag, His-tag MCS Amp ^R	Novagene
pSelAct	Bacterial vector <i>lacZ</i> , <i>codA:upp</i> , Apr ^R	van der Geize et al. (2008)
pLongJon	Derived from pSelAct, contains <i>pGroES</i> insertion.	This work
pGEX-6P-1	Bacterial expression vector, (U78872) MCS Amp ^R GST-tag lacI	GE Healthcare

2.1.1 Cell culture and morphology testing

R. equi strains were grown in complex broth (LB or BHI) for 48 hours at 37 °C with shaking at 180 rpm. Cultures were then heat-fixed to a microscope slide, stained using a standard Gram staining protocol and their morphology visualised by oil immersion microscopy at 100 x objective magnification.

2.1.2 Preparation of Vectors

Competent *E. coli* cells were prepared, and vectors transformed as previously described, purified using QIAprep spin Miniprep Kit to produce pure plasmid. The vectors were linearised using restriction enzymes, ready for amplicon assembly.

2.2 Iron limitation methods

2.2.1 Media production as a method of iron limitation

Lactate minimal medium was identified as an appropriate medium for culture and imposing iron limitation (Table 2-2). The iron limiting conditions were achieved by omission of iron chloride from the trace element solution. Cultures were grown at 37 °C until growth limitation was achieved, monitored by optical density at 600 nm. 1 ml aliquots were centrifuged at 13,000 x g; supernatants and pellets were frozen for further analysis.

Table 2-2 Composition of lactate minimal medium for iron limitation in *Rhodococcus equi*

Component	Amount (g) per 1 L
K ₂ HPO ₄	5
NaH ₂ PO ₄	1.5
MgSO ₄ ·7H ₂ O	0.2
(NH ₄) ₂ SO ₄	1
Trace element solution	0.2 ml
Thiamine hydrochloride	0.034
Sodium L-lactate	2.24

2.2.1.1 Trace elements solution

The trace element solution was prepared according to methods described by Vishniac and Santer (1957) in Table 2-3, later modified by addition of 0.5 M MES buffer to maintain stability. A secondary modified trace element solution was also prepared without the addition of iron for use in iron limited cultures.

Table 2-3: Composition of Vishniac and Santer Trace elements solution

Component	Amount (g) per 1 L
Na ₂ -EDTA	50.00
ZnSO ₄ ·7H ₂ O	22.00
CaCl ₂ ·2H ₂ O	5.54
MnCl ₂ ·4H ₂ O	5.06
FeSO ₄ ·7H ₂ O	5.00
(NH ₄) ₆ Mo ₇ O ₂₄ ·4H ₂ O	1.10
CuSO ₄ ·5H ₂ O	1.57
CoCl ₂ ·6H ₂ O	1.61

2.2.2 Modifications to impose iron limitation #1

2.2.2.1 Glassware cleaning

Glassware was cleaned by submersion in 10 % (v/v) Nitric acid overnight to remove mineral traces

and was rinsed twice with MilliQ H₂O and Acetone.

2.2.2.2 Media modification

The lactate minimal media was modified to determine growth limiting factors, including doubled carbon source, doubled nitrogen source and 10 mg/ml Riboflavin supplementation. All cultures were performed in triplicate, and differences were monitored via spectrophotometry.

2.2.3 Modifications to impose iron limitation #2

Fiss-glucose minimal media was utilised based upon previous literature for successful siderophore production in *Rhodococcus erythropolis* (Vellore 2001) that appears to be adapted from Minimal medium (MM) described by Ratledge that was used as the iron-limiting medium for mycobacteria (Hall and Ratledge, 1982).

2.2.4 Modifications to impose iron limitation #3

Lactate-Minimal media was utilised as described in 2.2.1, with additional iron limiting parameters including use of acid-washed glassware (2.2.2.1) and passage of liquid medium through a chromatography column packed with Chelex-100 resin, prior to sterilisation.

2.3 DNA methods

2.3.1 Genomic DNA extraction

Genomic DNA was prepared by suspending a colony of bacteria in microcentrifuge tube containing 50 µl 18.2 MΩ•cm H₂O. The suspension was then held at 100 °C for 5 minutes and centrifuged for 1 minute at 13,000 RPM. The supernatant (genomic DNA) was recovered and stored at -20 °C for future use.

2.3.2 Plasmid DNA extraction

Plasmid DNA was extracted from overnight cultures of *E. coli* using a QIAprep spin mini-prep kit (Qiagen) according to the manufacturers' protocol.

2.3.3 Polymerase Chain Reaction methods

Polymerase chain reaction (PCR) was performed using the ultra-high fidelity Q5 DNA polymerase (New England Biolabs Inc.), which are thermostable with 3'→5' exonuclease activity. The reactions were setup on ice (Table 2-4), and quickly transferred to a thermocycler for amplification (Table 2-5).

Table 2-4: PCR reaction setup using Q5 DNA polymerase

Component	25 µl Reaction	Final Concentration
5X Q5 Reaction Buffer	5 µl	1X
10 mM dNTPs	0.5 µl	200 µM
10 µM Forward Primer	1.25 µl	0.5 µM
10 µM Reverse Primer	1.25 µl	0.5 µM
Template DNA	variable	< 1,000 ng
Q5 High-Fidelity DNA Polymerase	0.25 µl	0.02 U/µl
5X Q5 High GC Enhancer	5 µl	1X
Nuclease-Free Water	to 25 µl	

Table 2-5 : Typical thermocycling conditions for a Q5 DNA polymerase PCR reaction

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
Annealing	98°C	5–10 seconds
	*°C	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	

2.3.4 Quality control using PCR

PCR was performed as a diagnostic tool to rapidly confirm species identify and presence of virulence plasmid using appropriate primers. Standard protocol and typical cycling conditions, and PCR optimisation methods were performed as described in standard operating procedures.

2.3.5 DNA analysis by electrophoresis

Agarose gel was prepared at varying concentrations dependent on expected PCR product size (typically between 0.8 and 1.5 % (w/v)) by dissolving the agarose powder in TAE which was heated until dissolved. After cooling to around 50°C, Sybr safe was added to give a final 1x working solution; the gel was then poured into a casting tray with a suitable well-forming comb inserted. Once set, the casting tray was transferred to an electrophoresis chamber and submerged in TAE. The comb was removed, and samples supplemented with 6x DNA loading dye (NE Biolabs, UK), and

molecular standard were pipetted into the wells. Samples were electrophoresed for ~ 40 minutes using a constant voltage of 130 volts. The gel was then transferred to the Gel Doc scanner (BioRad) and visualised using Quantity One software, molecular size of DNA in bands was estimated by comparison to an appropriate molecular standard.

2.3.6 DNA recovery from agarose gel

When necessary, samples containing PCR products were pooled, using taped combs to produce wider wells. Samples were then electrophoresed to recover DNA as in 2.3.5. The bands of interest identified by reference to the molecular standard were excised from the gel, using a transilluminator and scalpel. Excised bands were then weighed and purified using the QIAquick Gel Extraction kit (QIAGEN) or NZYgelpure (NZYtech), according to the manufacturer's protocol.

2.3.7 Purification of PCR products in preparation for cloning

To improve efficiency of subsequent molecular cloning techniques, the PCR reaction mixtures were purified using the NZYgelpure (NZYtech) kit with a PCR-clean up protocol to remove any additional primer-dimers or contaminating nucleotides.

2.3.8 Drop dialysis of purified product

Millipore MF-Millipore™ VSWP plain white mixed cellulose ester membrane filters were utilized for microdialysis of DNA. Extraction kit eluents were pipetted into the centre of the membrane, inside a petri dish containing 18.2 MΩ•cm H₂O. Once dialysed for up to 30 minutes, the desalted purified product was recovered and stored in a clean microcentrifuge tube.

2.4 Oligonucleotide primers

2.4.1 Quality control primers

Table 2-6: Oligonucleotides primers designed for quality control PCR reactions specific for *R. equi* and VapA.

Primer	Sequence (5'-3')	T _m (°C)
equispecificup	TCCAGAAGCGGGATGAGGATTC	62.1
equispecificlo	TGGTGTGATGGCGGAAGATC	59.4
VapAup	GAGCAAGCGATACCGCCGG	63.1
VapAlo	CTGGATATGGCCGAGGAAGC	61.4

2.4.2 Metalloregulatory gene primers

Table 2-7: Oligonucleotides primers designed for metal regulatory gene amplicons.

Recognition sites introduced for restriction endonucleases upstream and downstream of the target sequence to facilitate cloning into vectors. Restriction sites are bold highlighted, with the associated endonuclease given in the adjacent cell.

Primer	Sequence (5'-3')	Restriction endonuclease	T _m (°C)
23REQ20130_IdeRf	GCTAGCTACATATGAAGGATCTGGTCGACACC	<i>NdeI</i>	69.5
23REQ20130_IdeRr	TAATAATAGCGGCCGCTCAGACCTGCTTCACCTGGA	<i>NotI</i>	72.9
23REQ19260_DtxRf	GCTAGCTACATATGGCCACGCAGAAATCAGACGCAC	<i>NdeI</i>	72.9
23REQ19260_DtxRr	TAATAATAGCGGCCGCTAGACCAGCCAGATGGCTT	<i>NotI</i>	72.9
23REQ04740_FurAf	GCTAGCTACATATGCAACAAGGAGAGCACGACTTCG	<i>NdeI</i>	71.7
23REQ04740_FurAr	TAATAATAGCGGCCGCTCATTGTGCGAAGCCTCCTT	<i>NotI</i>	71.7
23REQ29120_FurBf	GCTAGCTACATATGACCGAGAACGTGACCGACCCGA	<i>NdeI</i>	74
23REQ29120_FurBr	TAATAATAGCGGCCGCTCAGCGAGCTTGC GCGCAGT	<i>NotI</i>	>75

2.4.3 Generation of *R. erythropolis* PR4 GroESL amplicon

Table 2-8 Generation of the *R. erythropolis* PR4 GroESL fragment

Name	Sequence 5' -3'-	n	GC	T _m
PR4GroEpromup	GCGAGGCTGGCGGGAACCTCGAAGAAGTGAACAATAGGTG	40	55%	80°C
PR4GroEpromlo	CTAAAGCGTTATTTCGACACCATGGTGAGCCCTCCACTTTGGG	42	52%	72°C

2.4.4 Sequencing primers

Table 2-9 Oligonucleotide primers used in Sanger sequencing methods

Name	Sequence 5' -3'-
T7minus1	AATACGACTCACTATAGGG
M13 FP	TGTAACACGACGGCCAGT
M13 RP	CAGGAAACAGCTATGACC
pSelAct FP	AGGGTTTTCCAGTCACGAC

2.4.5 Oligonucleotide primers for fXa-hybrid gene insertions

Table 2-10 Primers for fXa-hybrids

Primer name	Sequence	n	%GC	Therm. T _m	%GC T _m
Upstream vector overlap	GGCGACCATCCTCCAAAATC				
Factor Xa recognition site	AATCGAAGGTCGT				
Downstream vector overlap	GACCCGGGAATTCCGGGGAT				
ideR					
Original ideR forward primer	ATGAAGGATCTGGTCGACACC	21	52.4	52.7	49.2
Original ideR reverse primer (rc)	TCAGACCTGCTTCACCTGGA	20	55.0	52.9	48.7
Complete fXa-ideR hybrid forward primer	GGCGACCATCCTCCAAAATCAATCGAAGGTCGTATGAAGGATCTGGTCGACACC	54	51.9	84.2	68.7
Complete fXa-ideR hybrid reverse primer (rc)	GACCCGGGAATTCCGGGGATTCAGACCTGCTTCACCTGGA	40	60.0	82.1	67.6
dtxR					
Original dtxR forward primer	ATGGCCACGCAGAAATCAGACGCAC	25	56.0	66.6	55.9
Original dtxR reverse primer (rc)	CTAGACCAGCCAGATGGCTT	20	55.0	51.0	48.7
Complete fXa-dtxR hybrid forward primer	GGCGACCATCCTCCAAAATCAATCGAAGGTCGTATGGCCACGCAGAAATCAGACGCAC	58	53.4	87.7	70.2
Complete fXa-dtxR hybrid reverse primer (rc)	GACCCGGGAATTCCGGGGATCTAGACCAGCCAGATGGCTT	40	60.0	80.7	67.6
furA					
Original furA forward primer	ATGCAACAAGGAGAGCACGACTTCG	25	52.0	62.4	54.2
Original furA reverse primer (rc)	TCATTGTGCGAAGCCTCCTT	20	50.0	54.7	46.7
Complete fXa-furA hybrid forward primer	GGCGACCATCCTCCAAAATCAATCGAAGGTCGTATGCAACAAGGAGAGCACGACTTCG	58	51.7	86.0	69.5
Complete fXa-furA hybrid reverse primer (rc)	GACCCGGGAATTCCGGGGATTCATTGTGCGAAGCCTCCTT	40	57.5	81.9	66.6
furB					
Original furB forward primer	ATGACCGAGAACGTGACCGACCCGA	25	60.0	68.0	57.5
Original furB reverse primer	TCAGCGAGCTTGCGCGCAGT	20	65.0	63.9	52.8
Complete fXa-furB hybrid forward primer	GGCGACCATCCTCCAAAATCAATCGAAGGTCGTATGACCGAGAACGTGACCGACCCGA	58	55.2	88.3	70.9
Complete fXa-furB hybrid reverse primer	GACCCGGGAATTCCGGGGATTCAGCGAGCTTGCGCGCAGT	40	65.0	86.0	69.7

2.4.6 Oligonucleotide primers for mutagenic constructs

Table 2-11: IdeR gene knockout insertion (REQ_20130) primer design

Fragment	name	component	sequence	n	%GC	%GC Tm	Therm Tm
upstrm	ideRupup	homologous	CGGAAAGTTCCGCGCGCGTT	20	65.0	52.8	65.8
		vector overlap	CGCGGTGGCGGCCGCTCTAGAAGTAGT	27	76.2	59.0	69.7
		complete	CGCGGTGGCGGCCGCTCTAGAAGTAGTCGGAAAGTTCCGCGCGCGTT	47	66.0	72.6	88.5
	ideRuplo	homologous	CGTGGTGTGACAGATCCT	20	60.0	50.8	54.5
		dwnstrm overlap	ACCTGGACGGCGTGGGCCAT	20	70.0	54.9	65.1
		complete	ACCTGGACGGCGTGGGCCATCGTGGTGTGACAGATCCT	40	65.0	69.7	85.1
dwnstrm	ideRdwnup	homologous	ATGGCCCACGCCGTCCAGGT	20	70.0	54.9	65.1
		upstrm overlap	AGGATCTGGTCGACACCACG	20	60.0	50.8	54.5
		complete	AGGATCTGGTCGACACCACGATGGCCCACGCCGTCCAGGT	40	65.0	69.7	85.1
	ideRdwnlo	homologous	CGCTACCGATCCGGTCAGCT	20	65.0	52.8	58.8
		vector overlap	TTCCTGCAGCCCAGGGGATCCACTAGT	27	63.0	60.7	70.8
		complete	TTCCTGCAGCCCAGGGGATCCACTAGTCGCTACCGATCCGGTCAGCT	47	63.8	71.7	86.4

Table 2-12: DtxR gene knockout insertion (REQ_19260) primer design

Fragment	name	component	sequence	n	%GC	%GC Tm	Therm Tm
upstrm	dtxRupup	homologous	AGACGCTCGATGAGCGAGCG	20	65.0	52.8	60.2
		vector overlap	CGCGGTGGCGGCCGCTCTAGAAGTAGT	27	76.2	59.0	69.7
		complete	CGCGGTGGCGGCCGCTCTAGAAGTAGTAGACGCTCGATGAGCGAGCG	47	66.0	72.6	86.4
	dtxRuplo	homologous	TGCGGGTGCCTGATTCT	20	55.0	48.7	57.8
		dwnstrm overlap	ATGGCTTGTGCCGAGGGTT	20	60.0	50.8	61.1
		complete	ATGGCTTGTGCCGAGGGTTGCGGGTGCCTGATTCT	40	57.5	66.6	83.2
dwnstrm	dtxRdwnup	homologous	AACCCTGCGGCACAAGCCAT	20	60.0	50.8	61.1
		upstrm overlap	AGAAATCAGACGCACCCGCA	20	55.0	48.7	57.8
		complete	AGAAATCAGACGCACCCGCA AACCCTGCGGCACAAGCCAT	40	57.5	66.6	83.2
	dtxRdwnlo	homologous	CGCCGCTCGCGAGTTTCGAC	20	70.0	54.9	64.6
		Vector overlap	TTCCTGCAGCCCAGGGGATCCACTAGT	27	63.0	60.7	70.8
		complete	TTCCTGCAGCCCAGGGGATCCACTAGTCGCCGCTCGCGAGTTTCGAC	47	66.0	72.6	88.6

Table 2-13: FurA gene knockout insertion (REQ_04740) primer design

Fragment	name	component	sequence	n	%GC	%GC Tm	Therm Tm
upstrm	furAupup	homologous	CGGTCGCCCTGTGCGCGCAG	20	80.0	59.0	70.4
		vector overlap	CGCGGTGGCGGCCGCTCTAGAAGTACTAGT	27	66.7	62.2	71.3
		complete	CGCGGTGGCGGCCGCTCTAGAAGTACTAGT CGGTCGCCCTGTGCGCGCAG	47	72.3	75.2	91.0
	furAuplo	homologous	GGGGTCGAAGTCGTGCTCTC	20	65.0	52.8	55.9
		dwnstrm overlap	CCTCCTTGGCGCTGACGGGC	20	75.0	56.9	65.5
		complete	CCTCCTTGGCGCTGACGGGC GGGGTCGAAGTCGTGCTCTC	40	70.0	86.2	71.7
dwnstrm	furAdwnup	homologous	GCCCGTCAGCGCCAAGGAGG	20	75.0	56.9	65.5
		upstrm overlap	GAGAGCACGACTTCGACCCC	20	65.0	52.8	55.9
		complete	GAGAGCACGACTTCGACCCC GCCCGTCAGCGCCAAGGAGG	40	70.0	86.2	71.7
	furAdwnlo	homologous	GTCGAGATTCGACGGGCCCCG	20	70.0	54.9	63.3
		Vector overlap	TTCCTGCAGCCCCGGGGGATCCACTAGT	27	63.0	60.7	70.8
		complete	TTCCTGCAGCCCCGGGGGATCCACTAGT GTCGAGATTCGACGGGCCCCG	47	66.0	72.6	88.6
furAdwnlo2	homologous	GCCCTCGACCCCCGCGCTCGT	20	80.0	59.0	69.0	
	Vector overlap	TTCCTGCAGCCCCGGGGGATCCACTAGT	27	63.0	60.7	70.8	
		complete	TTCCTGCAGCCCCGGGGGATCCACTAGT GCCCTCGACCCCCGCGCTCGT	47	70.2	74.3	90.7

Table 2-14: FurB gene knockout insertion (REQ_29130) primer design

Fragment	name	component	sequence	n	%GC	%GC Tm	Therm Tm
upstrm	furBupup	homologous	ACGTCTGGTCGAAGTAGCTCAGGTC	25	56.0	55.9	58.9
		vector overlap	CGCGGTGGCGGCCGCTCTAGAAGTACTAGT	27	66.7	62.2	71.3
		complete	CGCGGTGGCGGCCGCTCTAGAAGTACTAGT ACGTCTGGTCGAAGTAGCTCAGGTC	52	61.5	72.2	84.3
	furBuplo	homologous	ATCGGTTCGGGTCGGTCACGT	20	65.0	52.8	60.9
		dwnstrm overlap	GCGCAGTCCCGACAGGTACC	20	70.0	54.9	59.0
		complete	GCGCAGTCCCGACAGGTACC ATCGGTTCGGGTCGGTCACGT	40	67.5	70.7	85.1
dwnstrm	furBdwnup	homologous	GGTACCTGTCGGGACTGCGC	20	70.0	54.9	59.0
		upstrm overlap	ACGTGACCGACCCGACCGAT	20	65.0	52.8	60.9
		complete	ACGTGACCGACCCGACCGAT GGTACCTGTCGGGACTGCGC	40	67.5	70.7	85.1
	furBdwnlo	homologous	ACGGCCGGGACCAGGACGGT	20	75.0	56.9	66.2
Vector overlap		TTCCTGCAGCCCCGGGGGATCCACTAGT	27	63.0	60.7	70.8	
		complete	TTCCTGCAGCCCCGGGGGATCCACTAGT ACGGCCGGGACCAGGACGGT	47	68.1	73.5	88.4

2.5 Transformation protocols

2.5.1 Preparation of competent *E. coli* cells.

To prepare the competent cells, 5 ml LB media was inoculated with a single colony of *E. coli* and incubated overnight at 37 °C. 1 ml of the overnight culture was inoculated into a sterile conical flask containing 50 ml LB broth. The flask was then incubated at 37 °C with 200 RPM in an orbital incubator, until a cell density 0.6 – 0.7 was achieved. Cell density was determined by taking a 1 ml aliquot and reading using a spectrophotometer at 600nm.

Once the desired optical density was achieved, the culture was chilled on ice for 10 minutes, then dispensed into 50 ml falcon tubes and centrifuged at 4000 x g at 4 °C for 5 minutes. The supernatant was discarded, and the pellet resuspended in 25 ml 0.1M CaCl₂. The suspension was placed on ice for 30 minutes before centrifugation again at 4000 x g at 4 °C for 5 minutes. Supernatant was discarded and pellet gently resuspended in 2.5 ml 0.1M CaCl₂.

The fresh competent cells are ready for transformation, and will remain competent for up to 24 hours at 4 °C. However, for long term storage, 2.5 ml glycerol (50 % v/v) should be added, then dispensed into 100 µl aliquots and stored at -80 °C until needed.

2.5.2 Transformation of competent *E. coli*

If in long-term storage, thaw cells slowly on ice. 1 µl plasmid DNA (or all of ligation mixture) was added to a 50 µl aliquot of competent cells, then incubated on ice for 30 minutes. The mixture was then heat shocked using a water bath at 42 °C for 45 – 90 seconds and transferred back on ice for 2 minutes. Recovery was mediated with the addition of 200 µl of non-selective media (typically LB or BHI) and incubated at 37 °C for 45 minutes – 1 hour using an orbital incubator, allowing the cells to express the antibiotic resistance marker.

Cells were then spread (~100 µl/plate) on to selective media and incubated overnight at 37 °C.

2.6 Electroporation protocols

2.6.1 Preparation of electrocompetent rhodococcal cells

To prepare electrocompetent rhodococcal cells, a single colony from an agar plate was inoculated into 50 ml of non-selective complex media (typically Brain-Heart Infusion or Luria-Bertoni broth). The culture was grown for 48 hours at 28 – 37 °C depending on the strain. Once an optical density of between 2 – 4 was achieved at 600 nm the cultures were cooled on ice to prevent further growth or degradation. The culture was transferred to a pre-cooled sterile falcon tube and centrifuged at 2880 x g for 10 minutes to collect cells. The supernatant was discarded at this stage, and the cell pellet was resuspended in 30 ml sterile ice-cold 10 % glycerol (v/v). The suspension was centrifuged as above, and the supernatant discarded, the washed cell pellet was then resuspended in 15 ml sterile ice-cold 10 % glycerol (v/v). The centrifugation and resuspension in glycerol was repeated twice more, with resuspensions in 2.5 ml and then a final 600 µl glycerol.

The now-competent cells were aliquot into 100 µl amounts and kept cold on ice until electroporation. Alternatively, cells were stored at -80 °C until needed, but efficiency may be reduced.

2.6.2 Electroporation of rhodococcal cells

Electroporation of rhodococcal cells was achieved by adding 50 ng of the required recombinant plasmid into a 100 µl aliquot of cells prepared earlier. A negative control of 100 µl cells was also prepared at this stage. The plasmid-cells mixture was added to a pre-cooled sterile electroporation cuvette (0.2 mm gap) without generating bubbles. The cuvette was added to the Gene Pulser System (Bio-Rad) and pulsed using parameters set at 25 µF, 2500 V, and 600 Ω. Recovery was mediated by the addition of 1 ml pre-warmed complex media, and transferred into a sterile 1.5 ml microcentrifuge tube, which was then incubated at 30 °C for between 2 – 5 hours depending on the strain.

100 µl of the recovered culture was then spread onto agar containing the appropriate antibiotic, and incubated for 2 – 3 days at 30 °C.

2.7 Molecular cloning methods

2.7.1 Blunt-ended molecular cloning

2.7.1.1 Vector preparation

Shuttle vectors such as pUC18/19 contain a multiple cloning site that facilitates blunt ended ligation. The multiple cloning site of these vectors contains a recognition site for restriction enzymes such as including *EcoRV*, *SmaI*, that when digested produce linearised plasmid with blunt ends. An example of plasmid digestion for blunt-ended cloning is shown in Table 2-15, where the reaction is incubated for between 15 minutes to 1 hour at the appropriate temperature for the enzyme described by the manufacturer.

Table 2-15 An example of blunt-ended vector preparation, by digestion of 1 µg of plasmid with a restriction enzyme

Component	Volume (µl)
Vector	1 µg
Buffer (x10)	5
Restriction enzyme	1
Milli Q H ₂ O	(up to 50 µl)

2.7.1.2 Ligation methods

Unlike traditional cloning methods, blunt-ended cloning does not require the insert DNA to be digested, therefore can be directly ligated with the digested blunt-ended vector. The ligations were performed using ATP-dependant T4-Ligase. An example of plasmid-vector ligation reaction is shown in Table 2-16.

Table 2-16 An example of a ligation using T4 ligase (a 3:1 molar ratio is shown)

Component	Volume (µl)
T4 DNA Ligase Buffer (10X)	2 µl
Vector DNA (4 kb)	50 ng (0.020 pmol)
Insert DNA (1 kb)	37.5 ng (0.060 pmol)
Nuclease-free water	to 20 µl
T4 Ligase	1 µl

Once ligated, the recombinant vector is useful intermediate stage that provides an opportunity to sequence the PCR amplicon before further downstream processing.

2.7.1.3 Screening bacterial colonies Using X-gal and IPTG: α-Complementation

To detect positive transformants, a blue white screen was performed as a screening method, activity is detected by X-gal and cleaved to produce 5-bromo-4-chloro-indoxyl that consequently results in

the characteristic blue colour. The inhibition of β -galactosidase activity prevents the cleavage, with the colony remaining white indicating a successful ligation and transformation. The screening process was performed by either spreading the IPTG and Xgal on top of the plates before use, or by incorporating the IPTG and Xgal into media before pouring plates.

For addition of IPTG and Xgal on top of premade agar plates 40 μ l of IPTG and X-gal was pipetted directly on top of each premade agar plate as needed and spread around using a plate-spreader. The plates were dried for 15 minutes before use. Alternatively, incorporation of IPTG and Xgal into agar plates was performed by preparation of the agar of choice autoclaved as per standard operating procedure and allowed to cool in a 55 °C water bath. After cooling, antibiotic was added if required, and IPTG and X-gal were added to a final concentration of 0.1 mM and 40 μ g/ml respectively. The agar was then mixed and poured into sterile petri-dishes, ensuring no bubbles were present.

2.7.2 Cohesive-ended molecular cloning

2.7.2.1 Vector preparation

Cloning vectors such as the pET vector range contain a multiple cloning site that facilitates traditional cohesive ended ligation. The multiple cloning site of these vectors contains a recognition site for restriction enzymes such as *NdeI*, *NotI*, *XhoI* that when digested produce linearised plasmid with cohesive overhanging ends. Double digestion using restriction enzymes facilitates a configuration at the cleavage site that is compatible with the ends of the amplicon. An example of plasmid digestion for cohesive-ended cloning is shown in Table 2-17, where the reaction is typically incubated for between 15 minutes to 1 hour at the appropriate temperature for the enzymes described by the manufacturer.

Table 2-17 An example of plasmid preparation for cohesive-ended cloning, by digestion of 1 μ g of plasmid with two restriction enzymes

Component	Volume (μ l)
Vector	1 μ g
Buffer (x10)	5
Restriction enzyme 1	1
Restriction enzyme 2	1
Milli Q H ₂ O	(up to 50 μ l)

2.7.2.2 Amplicon preparation

Similar to the cloning vector, the purified PCR amplicon of interest need to be digested with matching restriction enzymes to generate cohesive ends to produce a configuration site that is compatible with the ends of the linearised vector.

2.7.2.3 Ligation methods

The ligations were performed using ATP-dependant T4-Ligase, as previously described for the blunt ended ligations, with modifications of using the complementary digested vector and amplicon.

2.7.3 NEBuilder HiFi DNA assembly

The plasmid was assembled according to the supplier protocol for assembly of 2-3 fragments (Table 2-18). Optimal assembly efficiency was achieved by producing ≥ 20 bp overlap regions between each fragment with equimolarity.

The picomolar quantity of fragments was calculated using Equation 1. Optimised cloning efficiency for HiFi assembly is based on 50–100 ng of vector with 2-fold excess of inserts, using the picomolar values previously calculated. The reactions were incubated in a thermocycler at 50°C for 15 minutes and were subsequently stored at -20 °C, on ice for transformation. The competent cell transformation was performed using 2 μ l assembled product according to standard protocol, plated onto an appropriate antibiotic agar plate.

Table 2-18 NEBuilder HiFi DNA assembly

Recommended Amount of Fragments Used for 2–3 Fragment Assembly*	
Recommended DNA Molar Ratio	vector:insert = 1:2
Total Amount of Fragments	0.03–0.2 pmols*
	X μ l
NEBuilder	10 μ l
HiFi DNA Assembly Master Mix	
Deionized H ₂ O	10-X μ l
Total Volume	20 μ l ^{††}

$$\mu g DNA \times \frac{pmol}{660pg} \times \frac{10^6}{1\mu g} \times \frac{1}{N} = pmol DNA$$

Equation 1: Calculation for required pmol DNA to be used in the HiFi DNA assembly

2.8 DNA sequencing

Recombinant plasmids were sent to GATC Biotech / Eurofins Genomics for sequence analysis by Sanger sequencing. The service employed – SupremeRUN, required 20 µl of previously purified plasmid DNA at an approximate concentration of 50 ng / µl.

The trace data results (available in .fasta and .ab1 format) were viewed using the program FinchTV chromatogram viewer (Geospiza, Inc.) and processed using Blastn (NCBI) for sequence confirmation.

2.9 Protein methods

2.9.1 Gene expression

2.9.1.1 Optimisation of gene expression

Gene expression was performed after transformation of the recombinant plasmids into an expression host (described earlier), this was achieved using chemically competent *E. coli* BL21 (DE3) cells (unless otherwise stated). After transformation, the culture was plated on complex agar (supplemented with the appropriate antibiotic) and incubated at 37 °C overnight. Antibody-resistant colonies were selected and grown in LB broth (antibiotic supplemented) overnight at 37 °C with shaking at 200 RPM.

For optimisation, 1 mL culture aliquots were transferred into different Erlenmeyer flasks containing 50 mL sterile LB broth (antibiotic supplemented) and grown at 37 °C at 180 RPM to an OD₆₀₀ of 0.5–0.7. At which point the cultures were supplemented with Isopropyl β-D-thiogalactopyranoside (IPTG) and incubated at different temperatures before harvesting as shown in Table 2-19.

Table 2-19 Optimisation of conditions for gene expression

Flask no.	IPTG concentration	Temperature	Time
1	1 mM	37 °C	4 hours
2	1 mM	30 °C	4 hours
3	1 mM	25 °C	4 hours
4	1 mM	20 °C	4 hours
5	1 mM	16 °C	Overnight

2.9.1.2 Large scale expression

For large scale expression, an antibody resistant colony was selected and grown in 10 mL LB broth (antibiotic supplemented) overnight at 37°C with shaking at 200 RPM. The overnight culture was used to inoculate one litre of LB media (antibiotic supplemented) and grown to an OD₆₀₀ of 0.5 – 0.7 at 180 RPM, at 37°C. Protein production was induced by the addition of IPTG and incubated at the temperature that indicated following previous optimisation.

2.9.1.3 Lysate production

After expression, the cells were harvested by centrifuging at 5000 x g for 15 min at 4°C. Pelleted cells were re-suspended to 10 ml/g in basic lysis buffer (0.05 M Tris.HCl; 0.5M NaCl pH7.4) and lysed by sonicating resuspended cells to 7 cycles of 10 s sonication/ 10 s cooling with amplitude setting at 10 microns using the Soniprep 150 plus ultrasonic disintegrator. The lysate was clarified by centrifuging at 27,000 x g for 30 min at 4°C.

2.9.2 Percoll phase separation - Cell wall associated protein fractionation

After lysate centrifugation, cellular fractionation was performed on the lysate debris pellet by addition of Percoll (Amersham, Pharmacia Biotech, Sweden) to yield a 60 % solution prepared in phosphate-buffered saline (PBS) to isolate integral cell-wall associated proteins by density gradient centrifugation.

The pellet was resuspended in the 60 % (v/v) Percoll solution and centrifuged at 27,000 x g for 1 hour, the subsequent buoyant fraction was harvested and washed using PBS and centrifuged at 27,000 x g for 1 hour. The washes were repeated until a pellet free from Percoll was acquired. The cell wall fractionated pellet was resuspended into in 12.5 mM Tris.HCl pH 6.8 containing 2 % SDS, stirred overnight at 4°C followed by centrifugation at 27,000 x g for 30 min at 4°C for clarification. The cell wall fraction was analysed by SDS-PAGE.

2.9.3 Triton X114 hydrophobic protein solubilisation

To isolate and recover amphiphilic proteins from the cell envelope region, Percoll phase separation was performed as previously described in 2.9.2, until the final resuspension stage of the cell wall

fractionated pellet, upon which, the buoyant cell wall containing fraction was resuspended in 2 % Triton X-114 (v/v) by shaking overnight at 4 °C, 150 rpm adapted from methods listed by Tan et al. (1995). Insoluble material was removed by centrifugation at 27,000 x g, 4 °C for 15 minutes, followed by initial incubation of the supernatant at 37 °C for 10 minutes to facilitate the biphasic separation that occurred upon centrifugation at 14,400 x g at 25 °C for 15 minutes. The cell wall fraction-contained detergent phase was extracted and stored separately, with the aqueous phase resuspended in PBS and phase partitioned as aforementioned, before pooling the extracted detergent phases.

The phase partitioned triton protein extract was precipitated by addition of 10 volumes of acetone overnight at -20 °C. The extract was further centrifuged at 27,000 x g at 4 °C to pellet proteins with subsequent removal of the supernatant. The extracted pellet was dried by centrifugal evaporation and resuspended in Tris-HCl pH 7.4, before addition of SDS-loading buffer and SDS-PAGE analysis.

2.9.4 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Denatured protein samples were separated based on their electrophoretic mobility using Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) in a mini-gel format (Miniprotean® II, Bio-Rad). Buffer formulation shown in Table 2-20.

Table 2-20 SDS-PAGE buffer compositions

Compounds	Resolving buffer (pH 8.8)	Stacking buffer (pH 6.8)
Tris base	46.75 g	15.125 g
SDS	1 g	1 g
H ₂ O	250 ml	250 ml

Gel preparation was performed by addition of resolving gel (Table 2-21) between the casting plates (0.75 mm depth) and overlaid with isopropanol.

Table 2-21 Composition of PAGE resolving gel at various acrylamide concentrations

Components	8 %	10 %	12 %	15 %
Acrylamide/bisacrylamide (37.5:1; 40 %)	2.00 ml	2.50 ml	3.00 ml	3.75 ml
H ₂ O	5.50 ml	4.75 ml	4.50 ml	3.75 ml
Resolving buffer	2.50 ml	2.50 ml	2.50 ml	2.50 ml
10 % Ammonium persulphate (APS)	50 µl	50 µl	50 µl	50 µl
N,N,N',N'- tetramethylethylene-diamine (TEMED)	10 µl	10 µl	10 µl	10 µl

Once set, the isopropanol was discarded and stacking gel (Table 2-22) was added above the resolving gel; a well-forming comb was inserted and the gel was left to set. Once set, the gels were clamped into the electrophoresis chamber and anode buffer region filled with 1x SDS-running buffer. The cathode buffer region was also filled with 1x SDS-running buffer, appropriate for the number of gels ran. The comb was removed and 10 µl protein sample or molecular marker was loaded per well, protein separation occurred by electrophoresis at 200V for 1 hour. Once adequate separation was achieved, the gel was removed and immersed in an appropriate staining solution.

Table 2-22 Composition of PAGE stacking gel

Components	Stacking gel
Acrylamide/bisacrylamide (37.5:1; 40 %)	0.50 ml
H₂O	2.50 ml
Stacking buffer	1.00 ml
10 % Ammonium persulphate (APS)	30 µl
N,N,N',N'- tetramethylethylene-diamine (TEMED)	10 µl

2.9.4.1 SDS protein loading

6x SDS-loading buffer (375 mM Tris-HCl, 9 % SDS, 50 % glycerol, 9 % betamercatoethanol and 0.03 % bromophenol blue) was added to protein lysate to give 1x final concentration, and boiled for 10 minutes at 100 °C, before loading onto the gel.

2.9.4.2 Coomassie blue staining and imaging

Typical protein staining for SDS-PAGE was achieved using Coomassie blue R-250 staining by immersion for 20 minutes, before thorough rinsing with water and overnight de-staining using Coomassie de-stain. Gel images were captured using a G:Box system (SYNGene) with GenSys imaging software.

2.9.5 Non-denaturing polyacrylamide gel electrophoresis (NATIVE-PAGE)

Similarly, to SDS-PAGE, Native gels were performed using the mini-gel format (Miniprotean® II, Bio-Rad). Gel compositions varied as both resolving and stacking gel buffers did not contain SDS to prevent protein denaturation, all other components were identical.

2.9.5.1 Native protein loading

4x Native-loading buffer (400 mM Tris HCl pH 8.6, 40% Glycerol and 0.02% Bromophenol blue.)

was added to protein lysate to give a 1x final concentration and loaded directly onto the gel.

2.9.5.2 Native gel staining and imaging

Native-gel staining was achieved using Coomassie blue R-250 staining as performed for SDS-PAGE. Specific Native-gel staining methods are described in detail in the appropriate chapter methods sections. All gel images were captured using a G:Box system (SYNGene) with GeneSys imaging software.

2.10 Chromatography methods

2.10.1 Affinity chromatography

2.10.1.1 Heparin HP HiTrap

The clarified lysate was subjected to group specific affinity chromatography for DNA binding as the first step in the purification of the recombinant proteins. A 1 ml Heparin Sepharose high performance resin column (GE healthcare) fixed to an automated AKTA purifier Fast Protein Liquid Chromatography (FPLC) system (Amersham Pharmacia). The column was equilibrated with 10 column volumes of equilibration buffer (10 mM sodium phosphate, pH 7) at a flow rate of 1 ml / min. The sample was loaded onto the column at a flow rate of 0.5 ml / min and washed with 10 column volumes of equilibration buffer at a flow rate of 1 ml/min. Elution of bound protein was performed using linear gradient with elution buffer (10 mM sodium phosphate, 1–2 M NaCl, pH 7) using equilibration buffer as the diluent at a flow rate of 2 ml / min. Eluate was collected in 5 ml fractions. Fractions of interest identified by UV absorbance by FPLC were analysed by SDS-PAGE.

2.10.1.2 GSTrap FF HiTrap

The clarified lysate was subjected to GST-tagged protein purification as the first step in the purification of the recombinant proteins. A 1 ml GSTrap FF column prepacked with Glutathione Sepharose Fast Flow resin (GE healthcare) was fixed to an automated AKTA purifier Fast Protein Liquid Chromatography (FPLC) system (Amersham Pharmacia). The column was equilibrated with 5 column volumes of equilibration buffer (PBS, pH 7.3) at a flow rate of 2 ml / min. The sample was loaded onto the column at a flow rate of 0.5 ml / min and washed with 10 column volumes of

equilibration buffer at a flow rate of 2 ml/min. Elution of bound protein was performed using isocratic gradient with 100 % elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 7.9) at a flow rate of 1 ml / min. Eluate was collected in 4 ml fractions. Fractions of interest identified by UV absorbance by FPLC were analysed by SDS-PAGE.

2.10.2 Adsorption chromatography

2.10.2.1 XAD-2 resin

As a first stage purification from complex culture conditions, the bacterial supernatant was complexed with 1 % Ferric chloride and adjusted to pH 2.0 to aid column retention. The acidified bacterial supernatant was then loaded on to Amberlite XAD-2 adsorption resin pre-equilibrated with H₂O at a rate of 4.0 ml / min. The column was washed copiously with H₂O to remove any unbound material and flow through collected in 50 ml fractions. After sufficient washing, the column was eluted using 50 / 50 % (v/v) methanol / water followed by 100 % methanol.

2.10.3 Gel Filtration chromatography

2.10.3.1 P2 Biogel

As a secondary purification measure, fractions of interest were pooled together and applied to the top of a 1-meter BioRad chromatography column packed with P2 Biogel size exclusion resin, that was previously equilibrated in 50 % methanol running buffer to ensure that a change in buffer did not affect solubility of the sample. The sample was left to run through the column until resin was visible before addition of 10 ml running buffer with addition of the buffer adapter. The column was set up to run at a flowrate of 0.4 ml/min and to collect fraction volumes of 4.8 ml using an attached fraction collector. Fractions were tested via UV-vis spectral analysis (280 nm – 700 nm) both with and without the addition of the modified CAS solution (2.11.3).

2.10.4 Thin layer chromatography

Normal phase TLC was used to detect the presence of siderophores. Concentrated fractions were spotted one inch from the bottom of 10 x 20 or 5 x 10 silica gel plates and allowed to dry. The plates were then placed in a sealed glass chamber containing a n-butanol:acetic acid:dH₂O (12:3:5) solvent

and allowed to run until the solvent front came within approximately one inch of the top of the plate. The plate was then dried and developed by spraying with 0.1 M FeCl₃ in 0.1 N HCl or by looking directly under UV light.

2.11 Biochemical analysis methods

2.11.1 Zone of inhibition

2.11.1.1 Hydrogen peroxide

A modified disc inhibition assay (Rosner, 1993) was used to determine the relative sensitivities of the bacterium. An overnight bacterial starter culture was prepared in complex media (LB or BHI) of which 200 µl of inoculum was spread onto BHI agar. A filter disk was then applied to surface of plate, inoculated with 10 µl reagent. The plate was then incubated for two days to allow growth of a bacterial lawn. The zone of inhibition was calculated by an average diameter measurement of the inhibited area in three directions.

2.11.2 Chromazurol S Agar

Chromazurol S agar plates were prepared in 3 steps: (1) preparation of CAS indicator solution; (2) preparation of basal agar medium; and (3) preparation of CAS agar plates.

2.11.2.1 Preparation of CAS Indicator Solution

To prepare 100 ml of CAS indicator solution, 60.5 mg chromeazurol S was dissolved in 50 ml of distilled H₂O, to which 10 ml of ferric chloride solution was added (prepared by addition of 27 mg FeCl₃.6H₂O and 83.3 µl concentrated HCl in 100 ml distilled H₂O). To the above solution, 72.9 mg hexadecyltrimethyl ammonium bromide (HDTMA) dissolved in 40 ml distilled H₂O was added slowly along with stirring. The resultant dark blue CAS indicator solution was autoclaved.

2.11.2.2 Preparation of Basal Agar Medium

To prepare 100 ml of basal agar medium, 3 g 3-(N-morpholino) propane sulfonic acid (MOPS), 0.05 g sodium chloride, 0.03 g potassium phosphate, 0.01 g ammonium chloride, 0.5 g L-asparagine were dissolved in 83 ml H₂O. The pH was adjusted to 6.8 by addition of sodium hydroxide and the total

volume brought up to 88 ml using H₂O. Ultimately, 1.5 g of agar was added to the above solution was autoclaved.

2.11.2.3 Preparation of CAS Agar Plates

To prepare CAS agar plates, the autoclaved basal agar medium was cooled to 50 °C in a water bath along with the CAS indicator solution and a 50 % glucose solution. After adequate cooling, 2 ml of the 50% glucose solution was added to the basal agar medium followed by addition of 10 ml CAS indicator solution slowly down the container walls, with slight agitation. The CAS Blue agar was poured aseptically into sterile plastic petri dishes and stored at 4 °C until required.

2.11.3 Modified Chromazurol S Assay

2.11.3.1 Preparation of the modified CAS solution

To prepare the solution, 21.9 mg HDTMA was dissolved in 25 ml water while stirring constantly over low heat. In a separate container 1.5 ml of 1 mM FeCl₃.6H₂O (prepared in 10 mM HCl) was mixed with 7.5 ml of 2 mM CAS. This solution was slowly added to the HDTMA solution while stirring, and the mixture was transferred to a 100-ml volumetric flask.

An alternate buffer solution was prepared by dissolving 9.76 g MES in 50 ml H₂O, to prevent dye precipitation. The pH was adjusted to 5.6 using 50% KOH, and the buffer solution was then added to the dye-containing volumetric flask, the volume was finally increased up to 100 ml using H₂O.

2.11.3.2 Microtitre assay to detect siderophore production

To identify siderophore production, bacterial culture aliquots were taken at appropriate time points and cells were pelleted by centrifugation at 13,000 x g for 1 minute, the supernatant was then filtered using a 0.2 µM membrane filter and used promptly or stored at -20 °C until needed.

The bacterial supernatants of interested were transferred in 100 µl aliquots to a microtitre plate, to which 100 µl of the modified CAS assay solution was added. The reaction was controlled by incubation at 37 °C for 30 – 45 minutes. Absorbance was measured at 630 nm with a microplate reader, with a blank performed using media and CAS solution. Alternatively, an identification could be made based on visual inspection of colour, comparative to a blank.

2.12 *in silico* homology-based protein-modelling methods

2.12.1 Phyre2 protein modelling pipeline

The protein structure prediction tool of choice was the Protein Homology/analogy Recognition Engine V2.0 (Phyre2) server (Kelley et al., 2015a) (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>). Single sequences were modelled using a 4-stage automated pipeline: 1) Gathering of homologous sequences, 2) Fold library scanning, 3) Loop modelling and 4) Sidechain placement.

2.12.1.1 Advanced Phyre2 - One to one threading

2.12.1.2 Crystallographic Object-Oriented Toolkit (*Coot*)

The molecular-graphics application *Coot* was used for superposition of structures for sequence and fold comparison (Emsley et al., 2010). The method of choice was Secondary Structure Matching (SSM) superposition, an automatic method of protein superposing performed by alignment of secondary structure elements from one protein to another, without dependence on sequence identity.

2.12.1.3 PyMol viewer

Structures were viewed using the open source molecular visualisation tool PyMOL (<https://pymol.org/>) to provide a visual comparison of query and subject structures. The structures were shown as a *cartoon* and metal binding sites were shown using *sticks*. Thesis images were captured using the Draw/Ray tool, with a transparent background and CMYK colouring for future publications.

2.12.1.4 DALI protein structure comparison server

To identify the root-mean-square deviation (RMSD) of atomic positions of superimposed proteins, a pairwise structure comparison was performed using the DALI server (Holm and Laakso, 2016) (<http://ekhidna2.biocenter.helsinki.fi/dali>). The results output is given as a summary report, and a detailed pairwise structural alignment detailing three-state secondary structure definitions by the algorithm *Define Secondary Structure of Proteins* (DSSP).

2.13 Proteomic methods

2.13.1 In-gel digestion

The stained protein bands were excised from the gel with a scalpel, cut into 1 mm³ segments and transferred to a LoBind microcentrifuge tube. To each tube 200 µl of 100 mM ammonium bicarbonate (NH₄HCO₃) and 120 µl ACN was added and incubated for 15 min at room temperature with shaking, to remove the stain, this step was repeated twice. The gel pieces were then dehydrated by addition of 200 µl 100% ACN. Once dehydrated, the ACN was removed and discarded. The protein-containing gel pieces were then reduced by addition of 100 µl 20 mM DTT for 30 min at 56°C, with shaking. The excess liquid was removed, and the gel pieces dehydrated as above. The protein-containing gel pieces were then subject to alkylation by addition of 100 µl 55 mM IAA for 20 min in the dark, room temperature. Excess liquid was discarded, and the gel pieces washed twice with 100 µl 100 mM NH₄HCO₃, before dehydrating with ACN as above. The excess was discarded, and gel pieces dried by evaporation using a vacuum centrifuge. The protein-containing gel pieces were then digested by complete saturation with (30 µl) 20 µg/ml Trypsin/Lys- C mix for 20 min on ice. After which, 50 µl of 50 mM NH₄HCO₃ was added to completely cover the gel pieces, and incubated up to 18 hours at 37 °C.

2.13.1.1 Extraction of peptides

To extract digested peptides from the protein-gel 50 µl 50% (v/v) ACN / 5% (v/v) formic acid (FA); was added and shaken for 30 minutes. The peptides were collected in a new LoBind microcentrifuge tube. The extraction was repeated using 50µl of 83% (v/v) ACN / 0.2% (v/v) FA, from which the peptides was removed and pooled with the previous extraction.

The pooled peptide solution was snap-frozen using liquid nitrogen, and the LoBind microcentrifuge tube was pierced to facilitate freeze drying. The solutions were lyophilised overnight and stored at -80 °C until needed for analysis. Once ready for LC-MS analysis, the lyophilised samples were re-suspended in 20 µL of 5% (v/v) ACN, 0.1% (v/v) TFA.

2.13.2 Liquid chromatography and mass spectrometry

2.13.2.1 System information

The peptide characterisations were performed on a Nanoflow Dionex™ 3000 RSLC (Dionex, Sunnyvale, CA) linked to a Q-Exactive Plus (Thermo, Hemel Hempstead, UK). high resolution mass spectrometry system using C18 EasySpray column, in a data dependant acquisition (DDA). The system was maintained by Dr W. Cheung.

2.13.2.2 LC Instrument settings

The instrumental system settings for nanoflow liquid chromatographic separation used a binary buffer system for the peptide separation: Buffer A (95 % ultrapure water / 5 % ACN with 0.1 % formic acid), Buffer B (95 % ACN / 5% ultrapure water with 0.1 % formic acid), the loading and transport buffers (95 % ultrapure water/ 5% ACN with 0.1 % Tetrafluoroacetic acid (TFA)). The sample injection was set load 5 µl; flow rate was set to 0.3 µl / minute. The trap column used was Acclaim™ PepMap™ 100 C18 LC column (Thermo Scientific™), (5 µm particle size; pore size 100 Å), maintained at 45 °C.

2.13.2.3 LC gradient elution

The liquid chromatographic profile was performed as a gradient listed below.

Starting condition 96% buffer A, 4% buffer B, 0 min (4% buffer B/ 96% buffer A), 3 min (8% buffer B/ 92% buffer A), 93 min (30% buffer B/ 70% buffer A), 98 min (80% buffer B/ 20% buffer A), held for additional 10 minutes and then returned to starting condition with 20 min allowed for column equilibration.

2.13.2.4 MS Instrument settings

Full scan MS scanning was performed at 70,000 MS resolution with an automatic gain control (AGC) of $1e^6$ and injection time of 100 ms, the scan range was set to 375 to 1400 m/z. For data-dependant-MS², acquisition was performed at 35,000 with an AGC of $1e^5$ with a maximum injection time of 100 ms. The isolation window was set to 1.3 m/z, with a underfilled ratio of 0.4 %, Dynamic exclusion was set to 15 seconds, and the top 10 most abundant ions were selected for MS/MS with a

normalized Collision energy (NCE) level of 10, 30 and 50.

2.13.3 Protein Identification by MS/MS ion searching

Protein identification was performed by searching raw uninterpreted tandem mass spectral ion peak lists against the relevant database. For analysis, the thermo .RAW files were converted to mascot generic format (.mgf) using RawConverter (He et al., 2015). This analysis process was facilitated using the MS/MS ion search function of the in-house MASCOT server (Matrix Science).

The following parameters were used: (a) database: *Rhodococcus equi* 103S, (b) enzyme: Trypsin, (c) missed cleavages: allow up to one, (d) fixed modifications: carbamidomethyl (C), (e) variable modifications: oxidative (M), (f) peptide tolerance: 25 ppm, (g) MS / MS tolerance: 50 ppm, (h) peptide charge: 2+, 3+ and 4+, (i) mass value: monoisotopic, (j) data format: Mascot generic, (k) instrument: ESI-TRAP.

The resulting peptide score distribution identifies the probability that the observed match is a random event, further identifying the quantity of peptide matches above the identity or homology thresholds. Furthermore, each analysis predicts an average score from which greater than the value indicates identity or extensive homology ($p < 0.05$). Results were viewed using either the Protein family summary or report builder view, ranked by score, from which the peptide coverage and fragmentation pattern can be viewed.

2.13.4 Differential protein analysis

To facilitate identification of biologically relevant changes in protein expression, a proteomic workflow using the software Progenesis LCMS (Nonlinear Dynamics) and MASCOT (Matrix Science) was utilised.

2.13.4.1 Data alignment and pre-processing

The raw chromatography data files (thermo. RAW) were imported into the Progenesis LCMS software, and the chromatographic profiles were automatically aligned, with additional vectors manually applied for unaligned regions in each sample. Sample filtering was performed using the default peak picking parameters, from which peptide ions were further filtered based on charge,

excluding all peptides not exhibiting a 2+, 3+ or 4+ charge.

Prior to further analysis, the software automatically selects a normalisation reference, to apply a normalisation factor for future statistical analysis, calculated by finding the mean of the log abundance ratios of the peptide ions that fall within the robust estimated limits. Peptide ions outside these limits are considered to be outliers and therefore will not affect the normalisation.

2.13.4.2 Experimental design and Protein identification

The experimental design was setup to analyse runs using a ‘between-subject design’ grouping samples based upon appearance in a given condition. Next, the validation, and review of peptides was performed by selection filtering to exclusively scrutinise statistically significant features (peptides with a fold change of ≥ 2 and ANOVA p-value ≤ 0.05). Principal component analysis (PCA) and standardised expression profiles are produced to review selected peptide ions if necessary.

With the given parameters applied, the peptide ion peak list table was exported into the mascot generic format to facilitate protein identification through MASCOT MS/MS ion searching as described in 2.13.3. After the appropriate search, the results were exported from MASCOT in the extensible markup language file format (.xml) and imported into Progenesis LCMS assigning identified peptides to the significant features. Peptide identifications were refined based upon an acceptance criterion using the batch detection option, excluding results that have: (a) a peptide score of < 40 ; and/or (b) a peptide hit count of < 2 . The peptide search results were further validated at the protein level by resolving peptide conflicts, where conflicts occur between peptides, manual interpretation of results were performed, based upon peptide score, peptide hits and the given mass error (ppm). After which, the acceptance criterion detailed above was reapplied to the dataset.

2.13.4.3 Protein filtering and statistical analysis

Using the Review Proteins tab, proteins were reviewed, tagged and filtered in a similar manner as the peptide analysis, with selection filtering performed to exclusively scrutinise statistically significant proteins (with a fold change of ≥ 2 and ANOVA p-value ≤ 0.05). Ranking proteins by highest mean also facilitated identification, tagging and filtering for proteins upregulated in a given condition. Protein statistics could also be performed at this stage, with PCA plots and standardised

expression profiles given, dependant on the filters applied. Report outputs of the given results were used in subsequent protein localisation data analysis methods.

2.13.5 Data analysis methods – Protein localisation pipeline

Using the report produced in 2.13.4 differentially expressed proteins were further examined with regards to cellular localisation. For large quantities of identified proteins, a batch analysis methodology was produced, where Gi accession numbers were isolated and entered into Batch Entrez (<https://www.ncbi.nlm.nih.gov/sites/batchentrez>) against the Protein database. Retrieved batch records were compiled into a .FASTA format, ordered by Accession number for subsequent analysis.

Protein localisation was determined using a consensus strategy for prediction of transmembrane topology and signal peptides using the online bioinformatic tools TOPCONS (Tsirigos et al., 2015), PredLIPO (Bagos et al., 2008), SignalP (Almagro Armenteros et al., 2019) and Phobius (Kall et al., 2007). The compiled protein .FASTA file of interest was uploaded to the desired webserver, with the Gram-positive organism option selected (if available). The analysis outputs were assembled into a single spreadsheet and consensus established based upon majority.

2.14 Metabolomic methods

2.14.1 Liquid chromatography and mass spectrometry

2.14.1.1 System information

The chemical analysis were performed on a Dionex UltiMate 3000 UHPLC system (Dionex, Sunnyvale, CA) connected to a ThermoScientific Q-Exactive mass spectrometer system (Thermo, Hemel Hempstead, UK).

2.14.1.2 LC instrument settings - C18 Reverse Phase

Chromatographic separations were achieved using a Waters T3 high strength silica (HSS) C18 UHPLC column (150 x 1.8 mm, 1.7 μm) (Waters, Elstree, UK) with a flow rate of 0.4 mL/min operating at 45 °C and a 5 μL injection volume. Samples were held at 4 °C within the autosampler module prior and following injections.

2.14.1.3 LC gradient elution - C18 Reverse Phase

The LC buffers were a binary solvent system consisting of Buffer A (Ultrapure water 18.2 Ω collected from a Millipore Deionizer (MODEL) with ≤ 2 ppb TOC and 0.1% formic acid). Buffer B consisted of LC-MS Optima grade acetonitrile with 0.1% formic acid. The LC profile was as follows: 0 min (5% B) hold for 1 min proceeding to a linear gradient to 100% B at 12 mins, held for further 2 min (wash period) and returned to starting condition at 14 min with a column stabilization time of another 4 mins. The total run time per analytical run was approximately 20 min when including needle wash cycles.

2.14.1.4 MS instrument settings - C18 Reverse Phase

For the heated electrospray ionization (HESI) introduction source, the capillary temperature and voltage were maintained at 325°C and 3.8 KV (Positive mode)/ 3.5KV (Negative mode) respectively. (N_2) sheath flow was set to 45 and an auxiliary flow was set to 15 (all arbitrary units). The radio frequency of the S-lens was set to 50. For MS1 profiling the mass spectrometer was operating at 17.5K mass resolution with a scan rate of 13.2 scan/ s^{-1} with automatic gain control (AGC) at $1\text{e}6$ and a maximum injection time of 100ms.

For MS2 profiling the mass spectrometer was operating at 35K mass resolution with a scan rate of 8 scan/s⁻¹ with automatic gain control (AGC) set to 5e5 and a maximum injection time of 50ms. The mass ranges were set to 85-900 for positive ionisation and 115-950 for negative respectively, unless independently stated. Positive and Negative polarity data sets were acquired independently using the identical cinematographic profile as described above.

2.14.1.5 LC instrument settings – HILIC

Chromatographic separations were performed on a Water Acquity Ethylene Bridge Hybrid (BEH) Amide analytical column (2.1 x 150mm) with particle size of 1.7micron at a flow rate of 400µL/min; column temperature set to 45°C.

2.14.1.6 LC gradient elution – HILIC

The Binary buffer system was as follows: Buffer A was MilliQ water and Buffer B was ACN, both with 10mM ammonium formate adjusted to pH 3.5 using formic acid. The LC profile was as follows: T:0 min: 90%(B), T: 2min 60% (B) T: 5min 40%(B), T:7.5 min 40%(B), T:7.6min 90%(B), T:10 min 90% (B).

2.14.1.7 MS instrument settings – HILIC

The Heated spray ionization source (HESI) was set to the following parameters: Sheath gas flow rate of 50, the Aux gas flow rate was set to 13 and the sweep gas flow rate was 3. The Spray voltage of set to 3.5kV with a Capillary temperature of 275°C. The Aux gas heater temperature was adjusted to 425°C. The system was primed with a minimum of 10 sequential injections of pooled QC to stabilise the HESI source and to check for chromatographic stability before initialising batch analysis.

The MS1 mass acquisition range was as follows: 75-1000 m/z units at a mass resolution of 35,000 at approximately 7.6 scans per second, microscan: 1, lock mass: off. The AGC was set to 1e6 and the ion injection time was 100mS⁻¹. The data was acquired on Positive mode polarity only.

2.14.2 Data alignment and pre-processing

2.14.2.1 Compound Discoverer 2.0

Post data acquisition processing and alignment was performed using Thermo Scientific Compound

Discoverer 2.0 software suite (Thermo Fisher Scientific, UK). The Data files were parsed into its respective polarities and further groups by cultivar and additional truss positions. Pooled QC samples and sample blanks were also included and grouped accordingly in order to assess and evaluate system stability and tracked potential carry over effect throughout the entire batch analysis. Chromatographic alignment window was set to 0.15 min/ second, with mass tolerance of 5 ppm using an adaptive curve algorithm. The minimum peak intensity was set to 500,000 counts with minimum signal/noise threshold of 3/1 with protonated adducts (M+H) preferred, the gap filling protocol was also turned, and missing values were replaced based on a predefined experimental class design. Positive and Negative polarity data sets were processed independently.

2.14.3 Data analysis

2.14.3.1 MetaboAnalyst 4.0

The compound list generated by Compound Discoverer was firstly prepared prior to analysis using MetaboAnalyst. Firstly, the compound list was ranked lowest to highest based upon column retention time with each metabolite assigned a metabolite number (M). Next each condition was assigned a class value, and samples renamed S1-S9. For software compatibility, molecular weight and retention time columns were removed and the spreadsheet saved as a .csv file.

Analysis of the dataset is performed by using the Statistical analysis node of MetaboAnalyst 4.0, from which the dataset is uploaded, and the parameters set to Peak Integrity Table with the format Sample in rows (unpaired). The dataset integrity step was optionally skipped due to the gap filling protocol utilised in 2.14.2.1. Data filtering was performed using Non-parametric relative standard deviation (MAD/median), with normalisation achieved by log transformation and pareto scaling. The resulting analysis path facilitated both Univariate and Chemometric analysis as well as clustering analysis.

2.15 Transcriptomic methods

2.15.1 RNA isolation - RiboPure™-Bacteria Kit

2.15.1.1 Cell disruption and RNA purification

For RNA isolation of *R. equi* 103S, 10 ml was aliquot into a sterile RNase-free falcon tube, the bacterial cells were harvested by centrifugation for 1 minute at 13,000 rpm and supernatant discarded. 350 µL RNA-WIZ was then added to the cell pellet and resuspended by vigorous vortexing.

The cells in RNA-WIZ were added to approximately 250 µl of ice-cold Zirconia Beads in a 0.5 ml screw cap tube, and bacterial cells lysed by vigorous vortexing at maximum speed for 10 minutes, modified to 30 seconds pulsing using a FastPrep 120 at 6.5 speed, followed by 30 seconds cooling on ice, repeated three times.

The Zirconia beads were pelleted by centrifugation at 13,000 rpm for 5 minutes at 4 °C, and the bacterial lysate was aliquot into a fresh 1.5 ml microcentrifuge tube. Lysate quantity was estimated, and 0.2 volumes of chloroform was added, mixed thoroughly and incubated at room temperature for 10 minutes. Aqueous and organic phases were separated by centrifugation for 5 minutes at 13,000 rpm at 4 °C, the RNA-containing aqueous layer was extracted and aliquot into a fresh 1.5 ml microcentrifuge tube.

2.15.1.2 Final RNA purification

The quantity of aqueous phase extracted was estimated when pipetting, and 0.5 volumes of 100% ethanol was added, and mixed thoroughly. The sample was then transferred into a filter cartridge and centrifuged at 13,000 rpm for 1 minute to bind RNA to the filter, flow through supernatant was discarded. The filter was washed by addition of 700 µl Wash Solution 1 and subject to centrifugation at 13,000 rpm, with flow through discarded. The wash step was repeated twice more using 500 µl Wash Solution 2/3, with centrifugation and supernatant discarded as performed previously. The filter cartridges were then centrifuged for an additional minute to remove excess wash solution, before transferring the filter cartridge to a fresh 2 ml collection tube.

Elution of RNA was achieved by addition of 50 µl Elution buffer (preheated to 95 – 100 °C) to the centre of the filter, and centrifugation for 1 minute at 13,000 rpm.

2.15.1.3 DNase I treatment

To ensure any traces of genomic DNA was removed from the eluted RNA, 1/9th RNA volume of 10X DNase buffer and 4 µl DNase I was added. The sample was incubated at 37 °C for 30 minutes to induce digestion of genomic DNA, after which the process was inactivated by addition of DNase Inactivation Reagent equivalent to 0.2 volumes of initial RNA elution volume. The sample was then thoroughly mixed by vortexing. The sample was then incubated at room temperature for 2 minutes, with intermittent flicking of the tube to resuspend the DNase Inactivation Reagent.

2.15.1.4 Quality Assurance of RNA integrity

The isolated RNA was quantified using a Qubit Fluorometer and quality checked using an Agilent Bioanalyzer. An RNA integrity number (RIN) of > 7 was required to proceed before depletion. QA Prior to library preparation, the depleted RNA was quantified using a Qubit Fluorometer and quality checked using an Agilent Bioanalyzer. An RNA integrity number (RIN) of > 7 was required to proceed with the library preparation.

3 Genomic survey of iron homeostasis: A
Bioinformatic analysis of iron acquisition, targets
for mutagenesis and cloning in *Rhodococcus equi*

103S

3.1 Introduction

The aim of this chapter was to audit the requirements for siderophore-mediated iron-acquisition by *R. equi* 103S assigning gene identities, where possible, to each expected component. This survey would inform transcriptome and proteome data analysis and *vice versa*. The requirements are classified by function: regulation, biosynthesis and export, and ferri-siderophore uptake. Candidate genes were initially assembled *via* a text-based analysis of the *R. equi* genome annotations using key search terms (identified in each section).

3.2 A bioinformatic investigation into transcriptional regulation

A manual genomic data mining survey was performed to identify genes of interest that may be involved with regulation of iron homeostasis within the *R. equi* 103S chromosome. An abridged version of the search terms with the predicted products is shown in Table 3-1.

Table 3-1: Abridged genomic survey of iron regulatory genes in *R. equi* 103S.

Search terms were derived from logical association with previously identified iron regulatory terms and mechanisms.

Search terms	locus tag	predicted product
bactin	REQ_24100	iron-enterobactin transporter ATP-binding protein
	REQ_03460	Fe ²⁺ -enterobactin ABC transporter substrate-binding protein
chel	REQ_47020	enterochelin ABC transporter permease
	REQ_22510	ferrochelataase
dtxR	REQ_19260	putative DtxR-like transcriptional regulator
	REQ_20130	DtxR-like transcriptional regulator
fur	REQ_04740	ferric uptake regulator FurA
	REQ_29130	ferric uptake regulator FurB
heme*	REQ_10250	heme ABC transporter ATP-binding protein
	REQ_22210	heme A synthase
	REQ_22880	heme ABC transporter ATP-binding protein
	REQ_26300	hemerythrin
iron	REQ_03140	iron-regulated <i>lsr2</i> protein
	REQ_08040	iron ABC transporter permease
	REQ_11550	iron ABC transporter
	REQ_12610	iron transporter
	REQ_24080	iron ABC transporter permease
	REQ_24090	iron ABC transporter
	REQ_24100	iron-enterobactin transporter ATP-binding product
	REQ_29940	iron siderophore-binding protein
	REQ_31890	iron siderophore-binding protein
	REQ_47010	iron ABC transporter permease
metal	REQ_18320	zinc metalloprotease
	REQ_44570	metal-binding protein
siderophore	REQ_08190	putative siderophore binding protein
	REQ_18150	putative siderophore binding protein
	REQ_31980	putative siderophore binding protein
	REQ_37010	putative siderophore binding protein

*Haem also used with no further returns

The bioinformatic investigation revealed the presence of four genes that might encode iron regulator proteins: *ideR* (REQ_20130), *dtxR* (REQ_19260), *furA* (REQ_04740), *furB* (REQ_29130).

Analysis of their sequence *via* the conserved domains database at NCBI revealed that all were predicted to be metal-dependent transcriptional regulators (Table 3-2 and Figure 3-1, Figure 3-2). This analysis confirms the validity of the outcome of the automated gene assignment / annotation, however, the presence of four potentially iron-dependent transcriptional regulators requires careful analysis. The best characterised organism from this taxon is *M. tuberculosis* H37Rv, which also appears to have a complement of four such regulators (Rodriguez and Smith, 2003). Here a bioinformatic analysis was used to determine whether this complement of regulators has arisen through recent gene duplication and whether it is a feature of this strain of *R. equi*, is found more broadly across the genus or the nocardioform actinomycetes in general.

Table 3-2 Specific hits generated by conserved domain database searching for candidate iron-responsive transcriptional regulators

Candidate <i>R equi</i> gene		CDD Domain hits						
Locus ID	Assignment	Name	Accession	Description	Query length	Sequence coverage	E value	Graphical representation
REQ_20130	<i>ideR</i>	MntR	COG1321	Mn-dependent transcriptional regulator, DtxR family	230	4-134	3.9 x 10 ⁻⁴⁹	Figure 3-1A
		HTH_DTXR	smart00529	Helix-turn-helix diphtheria tox regulatory element; iron dependent repressor		29-122	1.33 x 10 ⁻³⁷	
		Fe_dep_repr_C	pfam02742	Iron dependent repressor, metal binding and dimerisation domain		67-136	6.81 x 10 ⁻³³	
		FeoA	smart00899	Represents the core domain of the ferrous iron (Fe2+) transport protein FeoA found in bacteria		153-228	4.73x 10 ⁻⁸	
		FeoA	pfam04023	FeoA domain; This family includes FeoA a small protein, probably involved in Fe2+ transport		153-228	2.92 x 10 ⁻⁶	
REQ_19260	<i>dtxR</i>	MntR	COG1321	Mn-dependent transcriptional regulator, DtxR family;	244	25-176	2.36 x 10 ⁻⁵²	Figure 3-1B
		Fe_dep_repr_C	pfam02742	Iron dependent repressor, metal binding and dimerisation domain		88-157	1.93 x 10 ⁻³⁸	
		HTH_DTXR	smart00529	Helix-turn-helix diphtheria tox regulatory element; iron dependent repressor		50-144	1.13 x 10 ⁻³⁵	
		FeoA	pfam04023	FeoA domain; This family includes FeoA a small protein, probably involved in Fe2+ transport		170-242	5.41 x 10 ⁻¹⁴	
		FeoA	smart00899	Represents the core domain of the ferrous iron (Fe2+) transport protein FeoA found in bacteria		170-242	6.18 x 10 ⁻¹²	
REQ_04740	<i>furA</i>	Fur	COG0735	Fe ²⁺ or Zn ²⁺ uptake regulation protein [Inorganic ion transport and metabolism];	184	9-146	3.40 x 10 ⁻³⁵	Figure 3-2A
		Fur_like	cd07153	Ferric uptake regulator(Fur) and related metalloregulatory proteins;		24-137	4.39 x 10 ⁻³¹	
REQ_29130	<i>furB</i>	Fur	COG0735	Fe ²⁺ or Zn ²⁺ uptake regulation protein [Inorganic ion transport and metabolism];	149	21-45	2.75 x 10 ⁻⁴⁵	Figure 3-2B
		FUR	pfam01475	Ferric uptake regulator family		21-138	6.63 x 10 ⁻⁴¹	
		Fur_like	cd07153	Ferric uptake regulator(Fur) and related metalloregulatory proteins;		26-141	8.93 x 10 ⁻⁴¹	

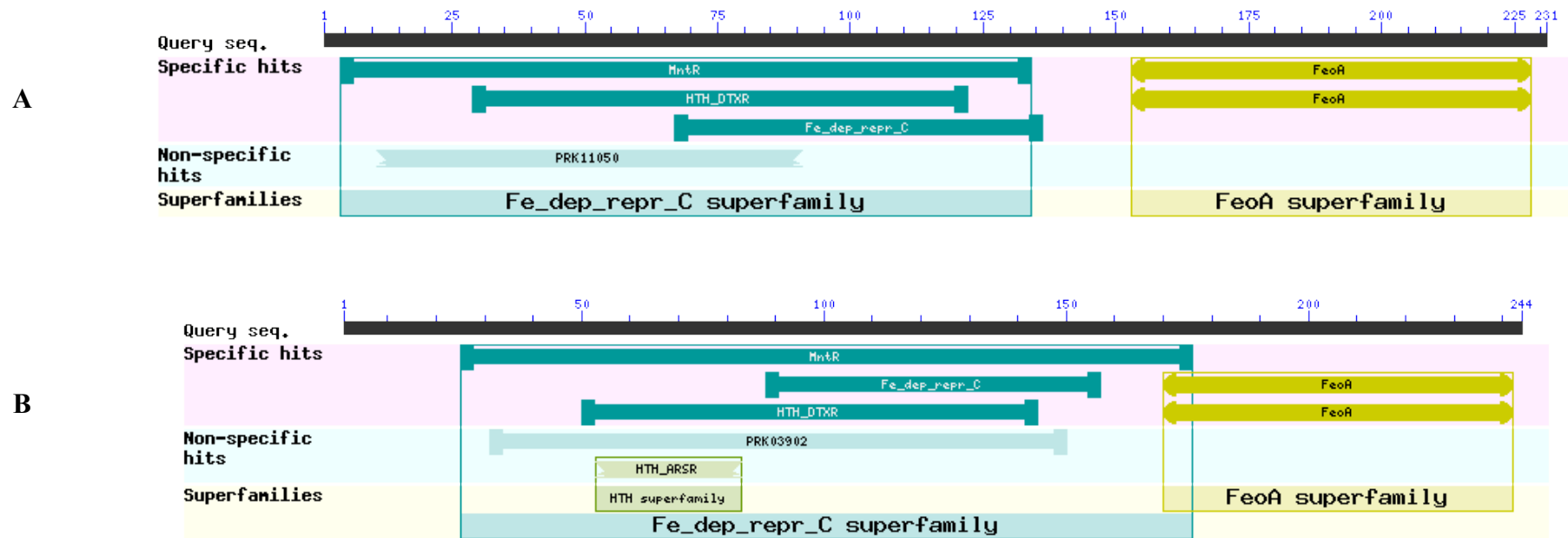


Figure 3-1 Conserved domains prediction for putative DtxR type metal sensitive regulators

The predicted primary structures for (A) REQ_20130, and (B) REQ_19260, were used as the query in a search of the Conserved Domain Database. Database v 3.16 was used with composition-based adjustment rather than the low complexity filter, and with an E-value threshold of 0.01 and a maximum of 500 hits prescribed.

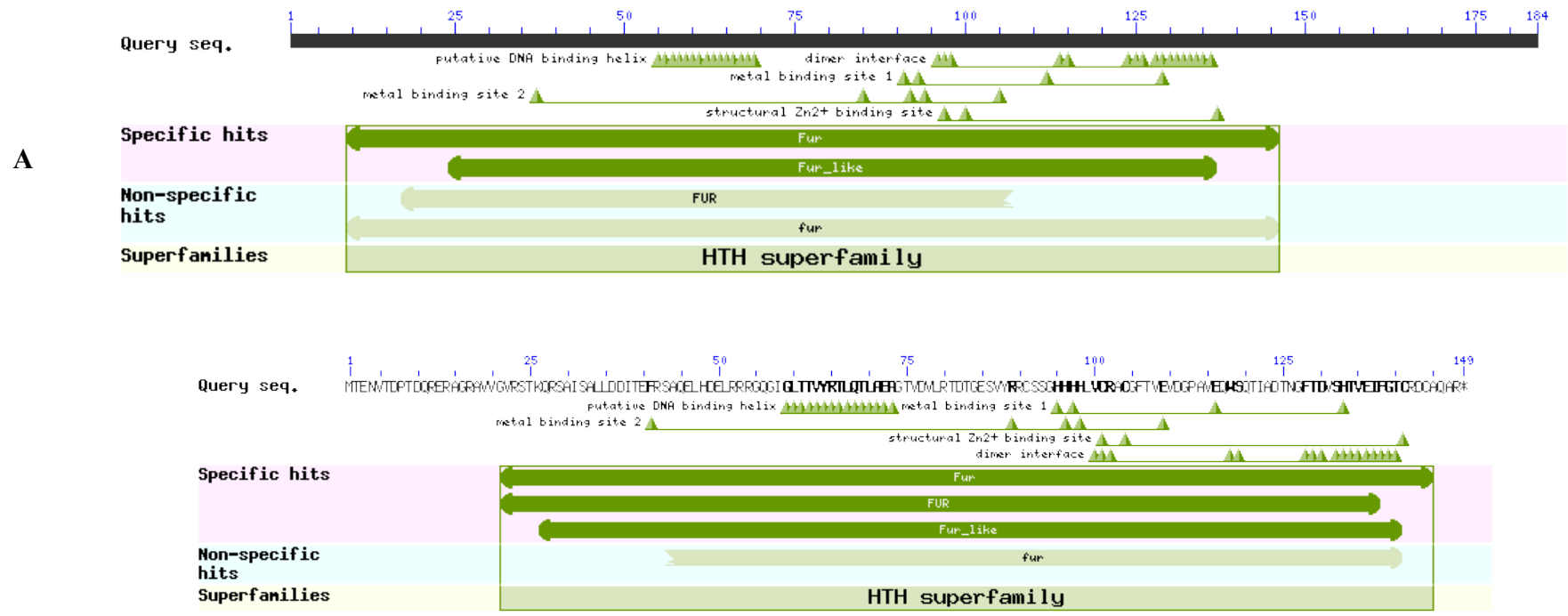


Figure 3-2 Conserved domains prediction for putative Fur type metal sensitive regulators

The predicted primary structures for (A) REQ_04740, and (B) REQ_29130 were used as the query in a search of the Conserved Domain Database. Database v 3.16 was used with composition-based adjustment rather than the low complexity filter, and with an E-value threshold of 0.01 and a maximum of 500hits prescribed.

apparent redundancy is not the consequence of a recent gene duplication event and likely indicates divergent function.

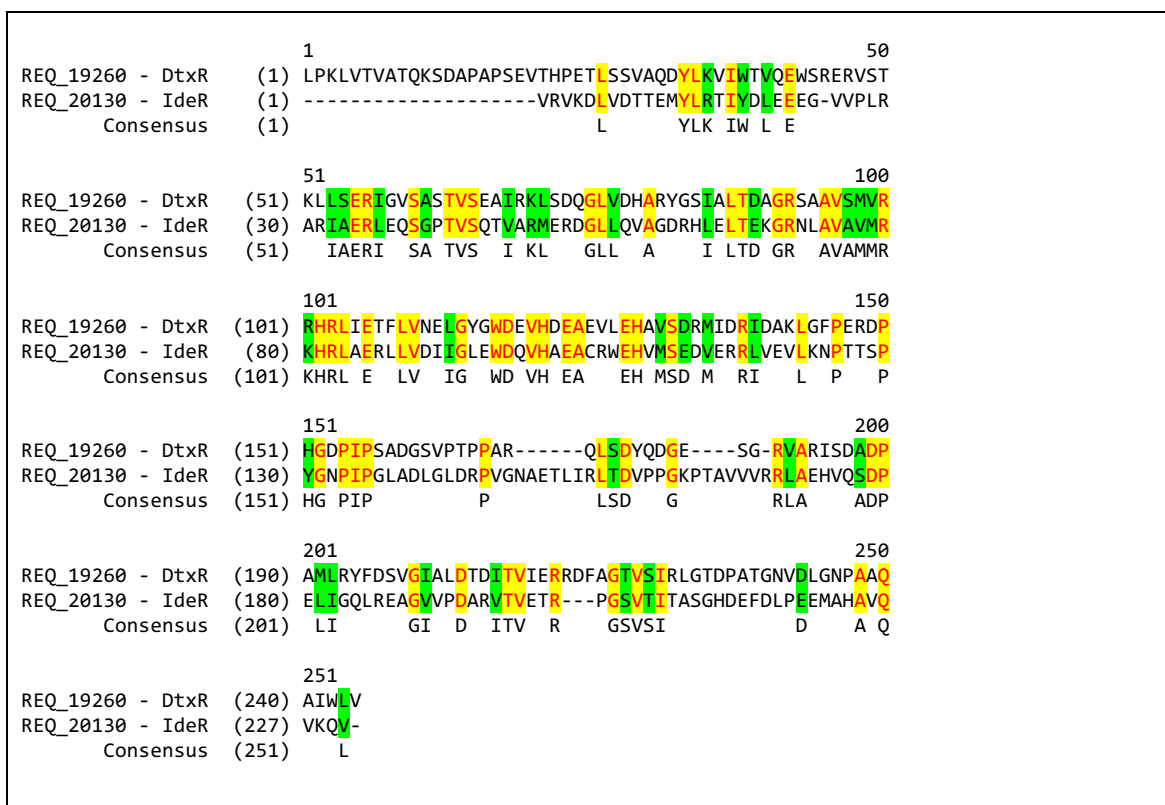


Figure 3-4: Sequence alignment of DtxR-type transcriptional regulators of *Rhodococcus equi*. Residues highlighted in yellow/red are identical and those highlighted in green/black are similar.

3.2.2 Multispecies identification of iron-dependent transcriptional regulators using BLASTP

In order to determine whether this apparent redundancy in iron-dependent transcriptional regulators was a conserved feature of genus *Rhodococcus* or an unusual feature of *R. equi* 103S, each candidate gene product sequence was used as a query in BLASTP (v2.6.1+) with outputs being restricted to *Rhodococcus* (Taxon ID 1827). The large number of rhodococcal genomes represented in Genbank complicates analysis as multiple entries have been deposited for several species. The Organism Reports associated with the BLASTP analyses were useful, facilitating a structured analysis of gene content and allele diversity within species groups.

For simplicity, the analysis focussed on the multispecies Refseq entries indicated by BLASTP. Entries for each set of hits were ranked (using E) and entries that were identified in other searches were labelled with the rank number assigned to the other protein for easy cross referencing.

3.2.2.1 Multispecies analysis of Fur regulatory proteins within *Rhodococcus*

The Fur family was considered first. Representative bacterial genomes were selected from the Refseq descriptors for FurA hits and were retained for the analyses of each of the four proteins. The 27 genomes chosen represent 6 of 7 clades of *Rhodococcus* identified by Creason *et al.* (2014) using a whole genome sequence approach (clade I is not represented). In the descriptive tables that follow, the ordering of these genomes (which reflects the similarity of their FurA homologue to that of *R. equi* 103S) is retained throughout for ease of comparison.

For the Fur family, this approach identified credible hits for both FurA and FurB in most genomes of this panel. Some of the Refseq entries appeared in both FurA and FurB hit lists. Such matches are the expected consequences of comparing similar sequences and demonstrate the appropriate sensitivity of the methodology in seeking homologues; *i.e.*, if the sequences can be used to detect each other they should detect intermediate orthologous sequences. Poor quality hits have been shaded in Table 3-3 and Table 3-4 for ease of reference; the criteria used to isolate these was a marked increase in E-value in the ranking and/or cross-matching with the partner data set. In some cases, namely WP_042575680, WP_037185410, and WP_019665665, these ‘poor hits’ were the best hits within the indicated genome for FurA and they did not appear among the FurB hits, simply indicating

significant sequence divergence of FurA over evolutionary time within *Rhodococcus*. In some cases, hits were identified with either the A or B subfamily but were secondary hits for that genome. Sometimes proteins were identified in both searches, but these may even have represented tertiary hits within that genome and might represent other regulatory proteins. In one case, WP_064075786 was found as a hit for both FurA and FurB in the *Rhodococcus qingshengii* JCM 15477 genome suggesting that only a single regulator of the Fur lineage was encoded, bearing greatest resemblance to FurA of *R. equi* 103S rather than FurB (based on comparison of respective E-values only).

In summary, this search strategy identified discrete FurA and FurB homologues in 26 of the 27 strains selected for this analysis (on the basis of their possession of FurA). It is clear that the possession of two members of this protein family is not an unusual feature of *R. equi* 103S but is a common feature of rhodococci.

The validity of these assignments was interrogated by generation of a maximum-likelihood phylogenetic tree (Figure 3-5) which considered the evolutionary relationships inferred by the sequence variations in the group of gene candidates. The Jones Taylor Thornton matrix-based algorithm differentiated two major clusters which effectively separated FurA and FurB candidates as assigned in Table 3-4. This second sequenced-based approach is wholly consistent with these representing two orthologous proteins families executing discrete physiological roles.

Table 3-3: BLASTP ranked table identifying conservation of the Fur-family proteins restricted to *Rhodococcus*.

Poor quality hits are indicated by shading, cross-matched hits are also included but identified.

ranking	FurA				FurB			
	Refseq accession	Score	E value	match	Refseq accession	Score	E value	match
1	WP_005515947	348	3.00x10 ⁻¹²³		WP_013416459	280	2.00x10 ⁻⁹⁷	
2	WP_054187028	231	5.00x10 ⁻⁷⁷		WP_003944514	231	2.00x10 ⁻⁷⁸	
3	WP_003943517	231	5.00x10 ⁻⁷⁷		WP_033233185	231	5.00x10 ⁻⁷⁸	
4	WP_019746588	227	2.00x10 ⁻⁷⁵		WP_020969953	230	6.00x10 ⁻⁷⁸	
5	WP_033235908	218	2.00x10 ⁻⁷²		WP_020908177	230	6.00x10 ⁻⁷⁸	
6	WP_005248503	208	2.00x10 ⁻⁶⁸		WP_006943911	229	1.00x10 ⁻⁷⁷	
7	WP_011596752	207	6.00x10 ⁻⁶⁸		WP_010594270	229	2.00x10 ⁻⁷⁷	
8	WP_024102806	206	7.00x10 ⁻⁶⁸		WP_005261219	228	5.00x10 ⁻⁷⁷	
9	WP_005256074	205	2.00x10 ⁻⁶⁷		WP_005247888	228	5.00x10 ⁻⁷⁷	
10	WP_029547199	201	9.00x10 ⁻⁶⁶		WP_011594307	227	1.00x10 ⁻⁷⁶	
11	WP_010594993	195	2.00x10 ⁻⁶³		WP_032368170	222	1.00x10 ⁻⁷⁴	
12	WP_068054043	183	1.00x10 ⁻⁵⁸		WP_027497490	222	1.00x10 ⁻⁷⁴	
13	WP_019665665	152	1.00x10 ⁻⁴⁶		WP_037193030	219	2.00x10 ⁻⁷³	
14	WP_043781216	149	2.00x10 ⁻⁴⁵		WP_020111630	213	4.00x10 ⁻⁷¹	
15	WP_005243590	149	2.00x10 ⁻⁴⁵		WP_016691347	211	2.00x10 ⁻⁷⁰	
16	WP_037185410	147	8.00x10 ⁻⁴⁵		WP_016935071	209	9.00x10 ⁻⁷⁰	
17	WP_037129600	147	2.00x10 ⁻⁴⁴	B24	WP_006554468	209	9.00x10 ⁻⁷⁰	
18	WP_042575680	146	2.00x10 ⁻⁴⁴		WP_042572828	207	4.00x10 ⁻⁶⁹	
19	WP_081557467	147	3.00x10 ⁻⁴⁴	B28	WP_027505728	206	1.00x10 ⁻⁶⁸	
20	WP_064075786	147	3.00x10 ⁻⁴⁴	B27	WP_068100239	206	1.00x10 ⁻⁶⁸	
21	WP_003941459	147	3.00x10 ⁻⁴⁴	B26	WP_026137834	59.3	1.00x10 ⁻¹⁰	
22					WP_010595420	58.2	3.00x10 ⁻¹⁰	
23					WP_020906505	56.2	2.00x10 ⁻⁰⁹	
24					WP_037129600	55.8	3.00x10 ⁻⁰⁹	A17
25					WP_019747524	55.8	3.00x10 ⁻⁰⁹	
26					WP_003941459	55.8	3.00x10 ⁻⁰⁹	A21
27					WP_064075786	55.5	5.00x10 ⁻⁰⁹	A20
28					WP_081557467	55.1	7.00x10 ⁻⁰⁹	A19
29					WP_029546787	54.7	7.00x10 ⁻⁰⁹	

Table 3-4: Multispecies identification of the iron-dependent transcriptional regulators FurA and FurB using BLASTP.

Hits are cross-correlated with Table 3-3, poor quality hits are indicated by grey shading. Black shading indicates a single Fur family transcriptional regulator.

Species identifier	Clade	Fur family	Ref seq accession	
			Primary hit	Secondary / tertiary hits
<i>equi</i> 103S	V	A	WP_005515947	
		B	WP_013416459	
<i>equi</i> C7	V	A	WP_005515947	
		B	WP_013416459	
sp Br-6		A	WP_005515947	
		B	WP_013416459	
sp YH3-3		A	WP_054187028	WP_081557467
		B	WP_003944514	WP_081557467
sp ADH		A	WP_054187028	WP_003941459
		B	WP_003944514	WP_003941459
<i>erythropolis</i> SK121	III	A	WP_003943517	WP_037129600
		B	WP_003944514	WP_037129600
<i>enclensis</i> NIO-1009		A	WP_003943517	WP_037129600
		B	WP_003944514	WP_037129600
<i>qingshengii</i> CW25	III	A	WP_003943517	WP_003941459
		B	WP_003944514	WP_003941459
<i>erythropolis</i> PR4	III	A	WP_019746588	
		B	WP_020908177	WP_020906505
<i>opacus</i> NRRL B-24011	IV	A	WP_019746588	
		B	WP_020969953	
<i>globerulus</i> NBRC 14531		A	WP_033235908	
		B	WP_033233185	
<i>opacus</i> PD630	IV	A	WP_005248503	
		B	WP_005247888	
<i>jostii</i> RHA1	IV	A	WP_011596752	WP_043781216
		B	WP_011594307	
<i>pyridinivorans</i> SB3094	VII	A	WP_024102806	
		B	WP_006554468	
<i>rhodochrous</i> NBRC 16069	III/VI/VII	A	WP_024102806	
		B	WP_016691347	
<i>imtechensis</i> RKJ300	IV	A	WP_005256074	WP_005243590
		B	WP_005261219	
<i>wratislaviensis</i> IFP 2016	IV	A	WP_005256074	WP_005243590
		B	WP_005247888	
<i>aetherivorans</i> BCP1		A	WP_029547199	
		B	WP_006943911	WP_029546787
<i>ruber</i> NBRC 15591	VI	A	WP_010594993	
		B	WP_010594270	
sp. EPR-147		A	WP_068054043	
		B	WP_037193030	
<i>fascians</i> A73a	II	A	WP_019665665	
		B	WP_020111630	
sp. Leaf225		A	WP_037185410	
		B	WP_027505728	
sp. MEB064		A	WP_042575680	
		B	WP_042572828	
sp. Leaf247		A	WP_042575680	
		B	WP_042572828	
<i>qingshengii</i> JCM 15477	III	A	WP_064075786	
		B	WP_064075786	
<i>kroppenstedtii</i> DSM 44908		A	WP_068362965	
		B	WP_068100239	

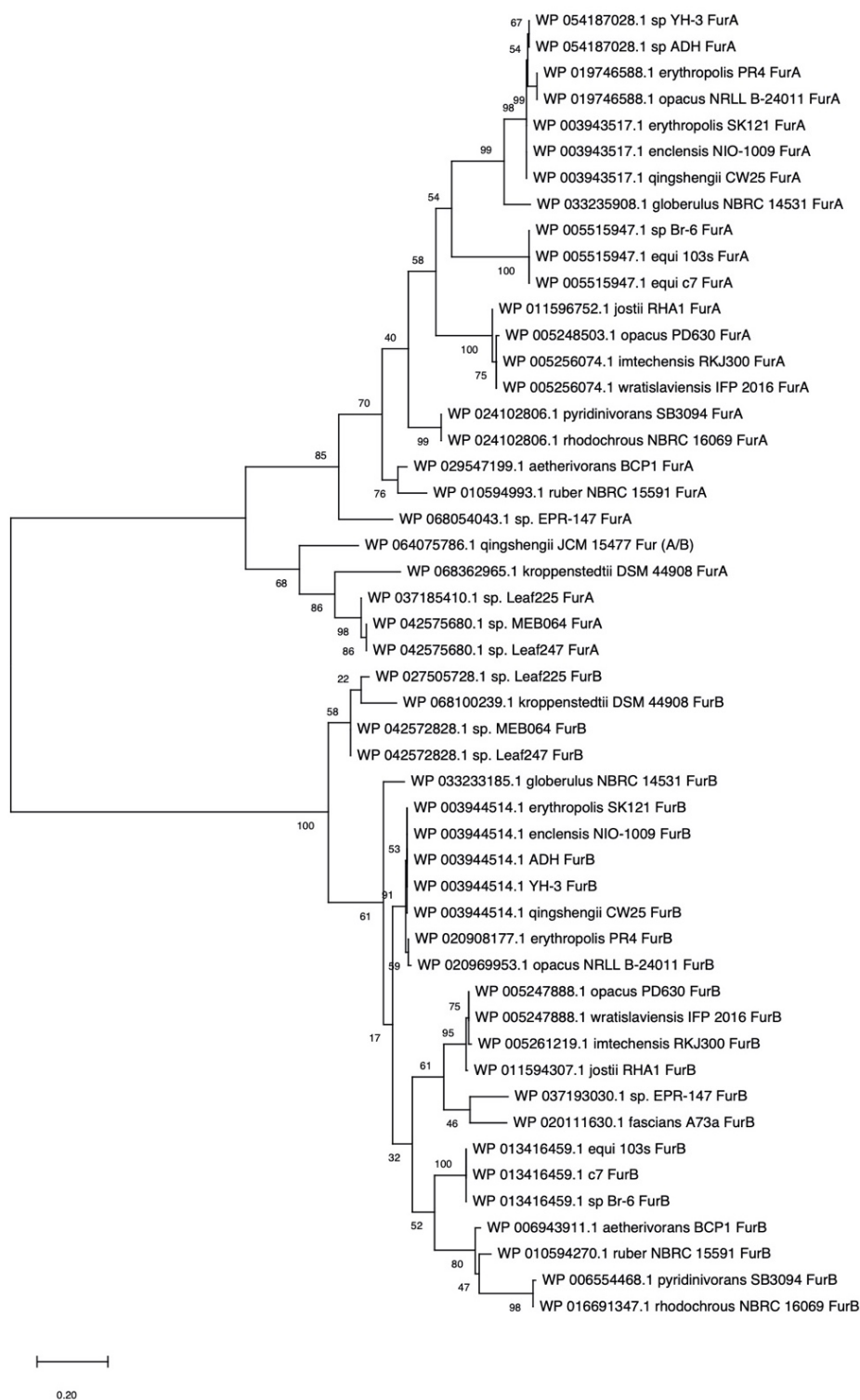


Figure 3-5: Evolutionary analysis of multispecies iron-dependent transcriptional regulators FurA and FurB by Maximum Likelihood.

Evolutionary history was inferred using the maximum-likelihood method and Jones Taylor Thornton matrix-based approach. Phylogeny was inferred utilitising a bootstrap value of 1000. The tree with the highest log likelihood (-3680.87) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 50 amino acids sequences. There were a total of 200 positions in the final dataset. Evolutionary analyses were conducted in MEGAX.

3.2.2.2 Multispecies analysis of DtxR(-like) regulatory proteins within *Rhodococcus*

Consideration of BLASTP hits arising from searches in which *R. equi* 103S DtxR and IdeR as queries generated similar outcomes with discrete hits being indicated for DtxR and IdeR in 26 of the 27 genomes of the panel with *Rhodococcus* sp. MEB064 appearing to encode a single member of this family which more closely resembles DtxR of *R. equi* 103S.

In summary, of the 27 genomes investigated here, 25 appear to possess homologues of all four putative iron-dependent transcriptional regulator proteins with none possessing less than three of them. It is unlikely that all of these proteins participate in iron homeostasis. Members of each of the Fur and DtxR/IdeR families have been demonstrated to coordinate responses to variations in other ions (Capdevila et al., 2017); this may well be the case in *R. equi* 103S. More detailed bioinformatic analysis provided more insight.

Similarly, the validity of these assignments was interrogated by generation of a maximum-likelihood phylogenetic tree (Figure 3-6) which considered the evolutionary relationships inferred by the sequence variations in the group of gene candidates. The Jones Taylor Thornton matrix-based algorithm differentiated two major clusters which effectively separated DtxR and IdeR candidates as assigned in Table 3-6. This second sequenced-based approach is wholly consistent with these representing two orthologous proteins families executing discrete physiological roles.

Table 3-5 BLASTP ranked table identifying conservation of the DtxR-family proteins restricted to *Rhodococcus*.

Poor quality hits are indicated by shading, cross-matched hits are also included but identified.

ranking	DtxR				IdeR			
	Refseq accession	Score	E value	match	Refseq accession	Score	E value	match
1	WP_013415758	482	3.00x10 ⁻¹⁷⁴		WP_005513651	449	6.00x10 ⁻¹⁶²	
2	WP_005240680	382	6.00x10 ⁻¹³⁵	IdeR22	WP_005240804	409	9.00x10 ⁻¹⁴⁶	
3	WP_011598478	381	2.00x10 ⁻¹³⁴	IdeR24	WP_015890477	408	1.00x10 ⁻¹⁴⁵	
4	WP_051636871	377	7.00x10 ⁻¹³³		WP_027499530	407	5.00x10 ⁻¹⁴⁵	
5	WP_019748070	374	1.00x10 ⁻¹³¹	IdeR21	WP_010592606	393	1.00x10 ⁻¹³⁹	
6	WP_033234902	373	4.00x10 ⁻¹³¹	IdeR20	WP_006551330	393	1.00x10 ⁻¹³⁹	
7	WP_030535607	373	4.00x10 ⁻¹³¹	IdeR25	WP_016692727	392	4.00x10 ⁻¹³⁹	
8	WP_003942116	373	4.00x10 ⁻¹³¹	IdeR26	WP_006935008	392	4.00x10 ⁻¹³⁹	
9	WP_050655153	372	1.00x10 ⁻¹³⁰	IdeR27	WP_033234793	390	2.00x10 ⁻¹³⁸	
10	WP_029543295	362	5.00x10 ⁻¹²⁷		WP_003942150	390	2.00x10 ⁻¹³⁸	
11	WP_006551268	360	2.00x10 ⁻¹²⁶		WP_019664070	386	6.00x10 ⁻¹³⁷	
12	WP_016693574	360	3.00x10 ⁻¹²⁶		WP_032397915	384	3.00x10 ⁻¹³⁶	
13	WP_040271917	356	5.00x10 ⁻¹²⁵		WP_032377729	384	7.00x10 ⁻¹³⁶	
14	WP_032397843	352	4.00x10 ⁻¹²³		WP_027496875	383	1.00x10 ⁻¹³⁵	
15	WP_027496759	350	2.00x10 ⁻¹²²		WP_068050657	383	1.00x10 ⁻¹³⁵	
16	WP_068050758	350	3.00x10 ⁻¹²²		WP_027505168	375	1.00x10 ⁻¹³²	
17	WP_019664180	348	1.00x10 ⁻¹²¹		WP_068105572	369	3.00x10 ⁻¹³⁰	
18	WP_056074537	310	2.00x10 ⁻¹⁰⁶	IdeR23	WP_042572644	369	4.00x10 ⁻¹³⁰	
19	WP_012686883	102	3.00x10 ⁻²⁵		WP_051730506	213	2.00x10 ⁻⁶⁸	
20					WP_033234902	96.3	8.00x10 ⁻²³	DtxR6
21					WP_019748070	93.2	1.00x10 ⁻²¹	DtxR5
22					WP_005240680	92.8	1.00x10 ⁻²¹	DtxR2
23					WP_056074537	92	2.00x10 ⁻²¹	DtxR18
24					WP_011598478	92	3.00x10 ⁻²¹	DtxR3
25					WP_030535607	92	3.00x10 ⁻²¹	DtxR7
26					WP_003942116	92	3.00x10 ⁻²¹	Dtx8
27					WP_050655153	91.7	4.00x10 ⁻²¹	DtxR9

Table 3-6 Multispecies identification of the iron-dependent transcriptional regulators DtxR and IderR using BLASTP.

Hits are cross correlated with Table 3-3, poor quality hits are indicated by grey shading. Black shading indicates a single DtxR-like family transcriptional regulator.

Species identifier	family	Ref seq accession	
		Primary hit	Secondary hit
<i>equi</i> 103S	DtxR	WP_013415758	
	IdeR	WP_005513651	
<i>equi</i> C7	DtxR	WP_013415758	
	IdeR	WP_005513651	
sp Br-6	DtxR	WP_013415758	
	IdeR	WP_005513651	
sp YH3-3	DtxR	WP_030535607	
	IdeR	WP_003942150	
sp ADH	DtxR	WP_030535607	
	IdeR	WP_003942150	
<i>erythropolis</i> SK121	DtxR	WP_003942116	
	IdeR	WP_003942150	
<i>enclensis</i> NIO-1009	DtxR	WP_003942116	
	IdeR	WP_003942150	
<i>qingshengii</i> CW25	DtxR	WP_003942116	
	IdeR	WP_003942150	
<i>erythropolis</i> PR4	DtxR	WP_019748070	
	IdeR	WP_003942150	
<i>opacus</i> NRRL B-24011	DtxR	WP_019748070	
	IdeR	WP_003942150	
<i>globerulus</i> NBRC 14531	DtxR	WP_033234902	
	IdeR	WP_033234793	
<i>opacus</i> PD630	DtxR	WP_005240680	
	IdeR	WP_005240804	
<i>jostii</i> RHA1	DtxR	WP_011598478	
	IdeR	WP_005240804	
<i>pyridinivorans</i> SB3094	DtxR	AHD19433	
	IdeR	WP_006551330	
<i>rhodochrous</i> NBRC 16069	DtxR	WP_059381216	
	IdeR	WP_016692727	
<i>imtechensis</i> RKJ300	DtxR	WP_005240680	WP_012686883
	IdeR	WP_005240804	
<i>wratislaviensis</i> IFP 2016	DtxR	WP_005240680	
	IdeR	WP_005240804	
<i>aetherivorans</i> BCP1	DtxR	WP_029543295	
	IdeR	WP_029543389	
<i>ruber</i> NBRC 15591	DtxR	WP_040271917	
	IdeR	WP_010592606	
sp. EPR-147	DtxR	WP_068050758	
	IdeR	WP_068050657	
<i>fascians</i> A73a	DtxR	WP_019664180	
	IdeR	WP_019664070	
sp. Leaf225	DtxR	KQU34775	
	IdeR	WP_027505168	
sp. MEB064	DtxR	KIQ20630	
	IdeR	KIQ20630	
sp. Leaf247	DtxR	WP_056074537	
	IdeR	WP_042572644	
<i>qingshengii</i> JCM 15477	DtxR	WP_003942116	
	IdeR	WP_003942150	
<i>kroppenstedtii</i> DSM 44908	DtxR	SFA60262	
	IdeR	WP_068105572	

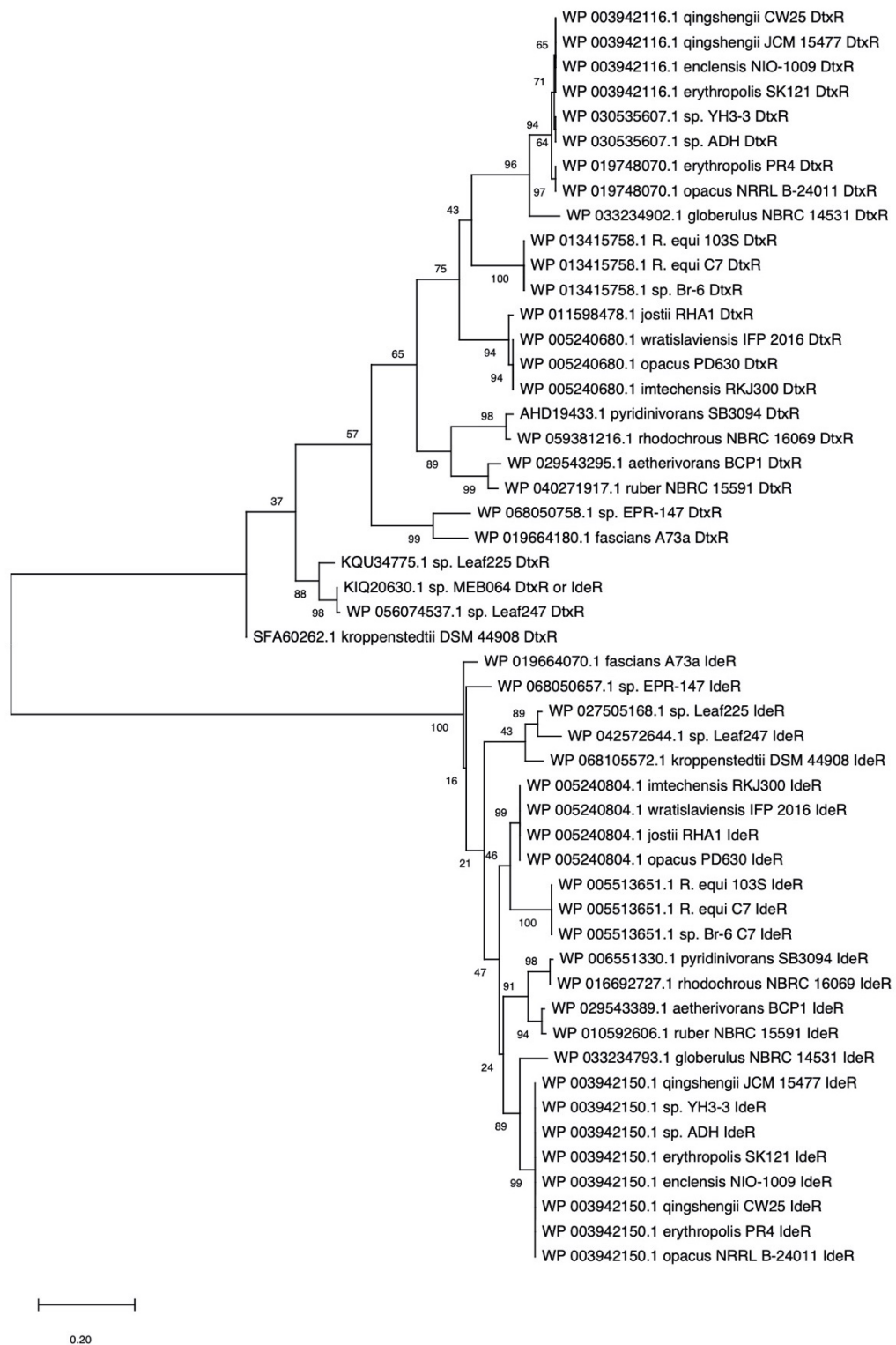


Figure 3-6: Evolutionary analysis of multispecies iron-dependent transcriptional regulators DtxR and IdeR by Maximum Likelihood.

Evolutionary history was inferred using the maximum-likelihood method and Jones Taylor Thornton matrix-based approach. Phylogeny was inferred utilising a bootstrap value of 1000. The tree with the highest log likelihood (-4319.37) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 51 amino acids sequences. There were a total of 272 positions in the final dataset. Evolutionary analyses were conducted in MEGAX.

Next, the genetic context of the regulators was addressed in comparison with more distant relatives, the pathogenic mycobacteria *M. tuberculosis* H37Rv and *M. leprae* TN. Like *R. equi*, both of these pathogens respectively transit through or reside in intracellular locations during the disease process (Sibley et al., 1987; Smith, 2003). The obligate intracellular pathogen *M. leprae* has a very degenerate genome (Young and Robertson, 2001) and is a particularly useful comparison, as it is widely recognised as having retained a minimal gene set to achieve the requirements of an intracellular mycobacterial pathogen (Vissa and Brennan, 2001). The roles of conserved gene products are usefully considered in this context.

In order to consider genetic synteny, text-based queries were used to interrogate the Database of prokaryotic Operons (Door², url: <http://csbl.bmb.uga.edu/DOOR/index.php>). Identities of genes were confirmed using a reciprocal BlastP protocol, *i.e.* the relationships between hits and the original queries were interrogated in the opposing direction considering those that provided mutual best hits as good evidence of functional orthology.

The text search for FurA revealed 36 operons; those of the nocardioform actinomycetes (n=34) appear to have a well-conserved genetic structure. Several entries including Operon ID 766677 (*R. equi* 103S), indicate that *furA* is located in a multi-gene operon of between 6 and 2 components. In other cases, a single gene operon is predicted. However, in each case *furA* precedes a gene encoding a catalase-peroxidase, a common component of the bacterial response to oxidative stress. This gene is particularly important in tuberculosis chemotherapy as it activates the major bactericidal drug, isoniazid (Zhao et al., 2006). The corresponding gene in *M. leprae*, ML2008, is a pseudogene as is the catalase-peroxidase. Although the latter might be perceived as indicating that these genes are not necessary for intracellular growth, *M. leprae* shows a particular tropism for Schwann cells (Ooi and Srinivasan, 2004) and this might represent an environment where they can thrive due to a lack of selective pressure; they then engineer these infected Schwann cells to become stem-like cells, which likely inhibits clearance and promotes dissemination (Masaki et al., 2013). In contrast, both genes appear to be essential in *Mycobacterium tuberculosis* mouse infection models (Pym et al., 2001). When Pym *et al.* investigated complementation of a *furA katG* double mutant, the absence of FurA caused the catalase-peroxidase to be upregulated and restored full virulence. It appears likely, therefore, that *furA* genes of *M. tuberculosis* and *R. equi* are orthologous.

3.3 A review of literature regarding *Rhodococcus equi* siderophores

Currently, there are two partially-identified siderophores produced by *R. equi*, a catecholate ligand-group siderophore – rhequibactin and a hydroxamate siderophore – rhequichelin. Interestingly both of these potential siderophores were identified by bioinformatic analysis of the genome, as after previous characterisation of *Rhodococcus equi* as a siderophore-negative bacterial species (Fiss and Brooks, 1991; Hall and Ratledge, 1986).

3.3.1 Rhequibactin

Miranda-CasoLuengo and colleagues predicted that *R. equi* produces rhequibactin; a diffusible catecholate-containing siderophore during growth in iron-limited conditions (Miranda-CasoLuengo et al., 2008). Catecholate-containing siderophores are typically produced using a NRPS pathway, confirmed for rhequibactin through *R. equi* genome analysis. Seven NRPS-encoding gene clusters were initially identified, with two hypothesised to be involved in siderophore biosynthesis.

While *R. equi* produces a catecholate-containing siderophore, rhequibactin and a non-diffusible siderophore required for saprophytic growth under low iron conditions, *in vivo* proliferation of *R. equi* mutants SID1 (*iupU*::pSID1K, derivative of ATCC 33701) and SID3 (*iupS*::pSID3K, derivative of ATCC 33701) results did not significantly differ compared to the wild-type, with a failure to attenuate in the presence of the bidentate chelating ligand 2,2 dipyridyl suggesting that *R. equi* utilises an additional method of iron acquisition (Miranda-CasoLuengo et al., 2008), later identified as a hydroxamate-containing siderophore, rhequichelin (Miranda-CasoLuengo et al., 2012).

3.3.2 Rhequichelin

A bioinformatic study performed by Letek and colleagues identified the gene cluster *rhbABCDE* (REQ07610 – REQ07650) (*rhb* - rhequichelin biosynthesis), with *rhbBCDE* forming a four-cistron operon likely encoding biosynthesis components (Letek et al., 2010; Miranda-CasoLuengo et al., 2012). Bioinformatic analysis of the gene cluster supported involvement in siderophore biosynthesis (Herbst et al., 2013).

REQ07630 was shown to encode RhbC, a protein homologous to the exochelin-producing NRPS

enzymes of *Mycobacterium smegmatis*, FxB and FxB (Miranda-CasoLuengo et al., 2012; Zhu et al., 1998). RhbC was predicted to synthesise a tetrapeptide-hydroxamate siderophore (Dhungana et al., 2007; Miranda-CasoLuengo et al., 2008). Transcription of *RhbC* was determined to be influenced by iron levels present in the medium, as had previously been observed with *iupS* and *iupT* encoding rhequibactin. Genes flanking *RhbC* were predicted to encode proteins involved with amino acid modification, as well as activation and release of the RhbC hydroxamate-siderophore product (Miranda-CasoLuengo et al., 2012).

Mutagenesis confirmed the requirement of the *rhbCD* genes to produce rhequichelin. The gene deletions prevented growth of *R. equi* SID2 (deletion of *rhbCD*) in low concentrations of 2,2-dipyridyl, and in macrophages; the former could be recovered by feeding with wild-type supernatant (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2008). The lack of growth of this mutant in macrophages is in contrast to Miranda-CasoLuengo's previous work with rhequibactin which was not required for virulence. Importantly, the rhequichelin biosynthesis dependency exhibited by *R. equi* is reminiscent of the requirement of (carboxy)mycobactin for macrophage proliferation in *M. tuberculosis* (Luo et al., 2005).

3.3.3 Potential for a mycobactin-like non-diffusible siderophore

In addition to the identification of the rhequibactin biosynthetic gene cluster, Miranda-CasoLuengo et al. (2008) hypothesised that the gene *iupU* (REQ_23810) may be involved with *R. equi* siderophore biosynthesis, given the proximal location of the *iupABC* siderophore transport system and a *mbtH* gene linkage (Lautru et al., 2007). Notably, *mbtH-like* genes have previously been identified as an essential component for production of some NRPS-dependent natural products (Lautru et al., 2007). Transcriptional analysis identified that *iupU* did not appear to be regulated by iron, with comparable levels under iron starvation. Remarkably, mutagenic disruption of *iupU*: SID1 (*iupU*::pSID1K, derivative of ATCC 33701) inhibited *R. equi* growth under iron limitation, and unlike *R. equi* mutant SID3 (*iupS*::pSID3K, derivative of ATCC 33701), SID1 could not be crossed by the wild-type strain, indicating that the product of *IupU* either is not secreted or is not diffusible in water. If accurate, production of a non-diffusible siderophore would be reminiscent of a cell-wall bound mycobactin produced by *M. tuberculosis*.

Recently, Chu et al. (2016) characterised a selection of humimycins from the *Rhodococcus* genera, that function by inhibiting lipid II flippase and potentiate β -lactam activity against methicillin-resistant *Staphylococcus aureus* in mice. This investigation into MRSA-active antibiotics, identified that the NRPS encoded by REQ_23810 (*iupU*) synthesises humimycin A, rather than a cell-wall bound mycobactin-like siderophore.

3.4 Siderophore biosynthesis

Bacterial iron metabolism is regulated depending on intracellular availability, to prevent oversupply and the overwhelming formation of reactive oxygen species (Fenton's reaction), produced as a byproduct of redox stress (Andrews et al., 2003). Siderophore biosynthesis and transport systems can be repressed in many bacteria when under iron-replete conditions, to prevent such toxicity. Often, regulation is mediated by the Fur protein; acting as a negative regulator, transcription is repressed upon interaction with Fe^{2+} and in its absence is de-repressed, (Andrews et al., 2003; Hantke, 2001) and is discussed in detail later in Chapter 4.

3.4.1 NRPS – A putative method of siderophore production in *R. equi*

Siderophores are known to be synthesized by both non-ribosomal peptide synthetase (NRPS)-dependent and NRPS-independent pathways. Therefore, the *R. equi* 103S chromosomal genome was searched for NRPS proteins (Table 3-7) by mining the FN563149.1 genome annotation text for associated terms, such as NRPS and non-ribosomal peptide synthetase.

Table 3-7: NRPS found within the *R. equi* 103S chromosomal genome.

NRPS	locus tag (former tag)	Product annotation	Refseq protein ID
1	REQ_RS03725 (REQ_07630)	non-ribosomal peptide synthetase	WP_049799404.1
2	REQ_RS03985 (REQ_08140)	non-ribosomal peptide synthetase	WP_013414937.1
3	REQ_RS03995 (REQ_08160)	non-ribosomal peptide synthetase	WP_013414938.1
4	REQ_RS11660 (REQ_23810)	non-ribosomal peptide synthetase	WP_013416064.1
5	REQ_RS13285 (REQ_27140)	non-ribosomal peptide synthetase	WP_041674289.1
6	REQ_RS17625 (REQ_35940)	non-ribosomal peptide synthetase	WP_013416894.1
7	REQ_RS17630 (REQ_35950)	non-ribosomal peptide synthetase	WP_041674405.1
8	REQ_RS17635 (REQ_35960)	non-ribosomal peptide synthetase	WP_041674406.1
9	REQ_RS19895 (REQ_40560)	non-ribosomal peptide synthetase	WP_013417196.1
10	REQ_RS20875 (REQ_42660)	non-ribosomal peptide synthetase	WP_013417352.1

3.5 Predicted structures of secondary metabolites

3.5.1 HMM analysis

In silico secondary metabolite structure predictions were produced based on nonribosomal peptide biosynthetic pathways retrieved from *R. equi* 103S DNA sequence data by manual data mining. A Hidden Markov Model (HMM) was used to predict domain identity for the ten multi-modular enzymes identified (Bachmann and Ravel, 2009).

Predicted operon structures were identified using the online DOOR² operon database (Mao et al., 2009), and were tabulated along with predicted structures generated using the predicted amino acids from the corresponding NRPS (Chapter 3 Appendices - 1). Where no predictions were generated for the adenylation (specificity-imparting) domain of a particular module, the product was assembled using an alanine scaffold containing an undefined ‘R’ group.

Apart from the previously identified operon (gene ID: REQ_07630/rhbC) encoding the putative siderophore rhequichelin (Miranda-CasoLuengo et al., 2012) the prediction tools used to identify secondary metabolites cannot solely predict based upon between secondary metabolite produced by NRPS pathways, and therefore cannot predict any contributions made by NRPS-independent siderophore (NIS) synthetase pathways, polyketide synthases, terpenes, alkaloids or saccharides, therefore a more robust tool for analysis of biosynthetic gene clusters was required.

3.5.2 Antibiotics and Secondary Metabolite Analysis SHell (antiSMASH) analysis

The bioinformatic tool antibiotics and Secondary Metabolite Analysis SHell (antiSMASH) was identified as a suitable alternative to manual annotation of NRPS operons, by providing a “rapid genome-wide identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial genomes” (Weber et al., 2015).

antiSMASH is comprehensive pipeline capable of automatically datamining genomes of interest to identify putative secondary metabolite biosynthetic gene clusters. Essentially the software aligns identified regions at the gene cluster level to their nearest relatives from a gene cluster database and

facilitates integration of specific secondary-metabolite predictions tools to function as a complementary tool for the identification of novel metabolites.

Therefore the *R. equi* 103S genome (Genbank ID: FN563149.1) was analysed using antiSMASH with ClusterFinder BCG detection addition, with ClusterBlast and Whole-genome PFAM analysis. Full genome cluster analysis available in Chapter 3 Appendices - 2. A refined NRPS cluster analysis is shown in Table 3-8.

Table 3-8 A NRPS-refined cluster analysis from record FN563149.1 *R. equi* 103S genome

Region	Type	From	To	Most similar known cluster	Similarity
3	NRPS	770,105	825,773	Erythrochelin	57%
4	NRPS	831,286	892,258	Heterobactin	54%
8	NRPS	2,488,176	2,551,002	Skyllamycin	4%
9	NRPS	2,840,029	2,895,666	Thailanstatin	10%
12	NRPS	3,780,516	3,880,782	Fengycin	20%
13	NRPS	4,281,932	4,337,935	N/A	
14	NRPS	4,503,597	4,577,253	Rimosamide	14%

showing the biosynthetic gene cluster location and the most similar known cluster.

From the refined analysis, two clusters appear analogous to known siderophore biosynthetic gene clusters (highlighted in yellow), a further two are analogous to antibiotic/ antifungal biosynthetic gene clusters, two are similar to anti-cancer agents and one is uncharacterised. The identification of two putative siderophore biosynthetic gene clusters would corroborate with the previous literature for *R. equi* (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2008), with biosynthetic gene cluster regions 3 producing rhequichelin, and cluster 4 responsible for production of rhequibactin.

These putative siderophore clusters were also identified in the original prediction made in 3.5, with the first identified NRPS from section 3.5 correlating with rhequichelin and the Erythrochelin-like cluster NRPS region 3 from the antiSMASH analysis in Table 3-8; with the second and third NRPS identified from section 3.5 correlating with rhequibactin and the Heterobactin-like cluster region 4 from the antiSMASH analysis.

3.5.3 Comparison of rhequichelin cluster prediction tools

The first analysis was performed using the PKS/NRPS analysis web tool (Bachmann and Ravel, 2009) for *R. equi* 103S (full results available in Chapter 3 Appendices - 3). Analysis of NRPS 1

predicted a four-module pathway for incorporation of amino acids into a peptide assembly mechanism (Figure 3-7). Based on the the predicted specificity for each aminoacyl adenylation domain the product is predicted to be N^5 -formyl-hydroxyornithinyl-X-threoninyl-X (Figure 3-7).

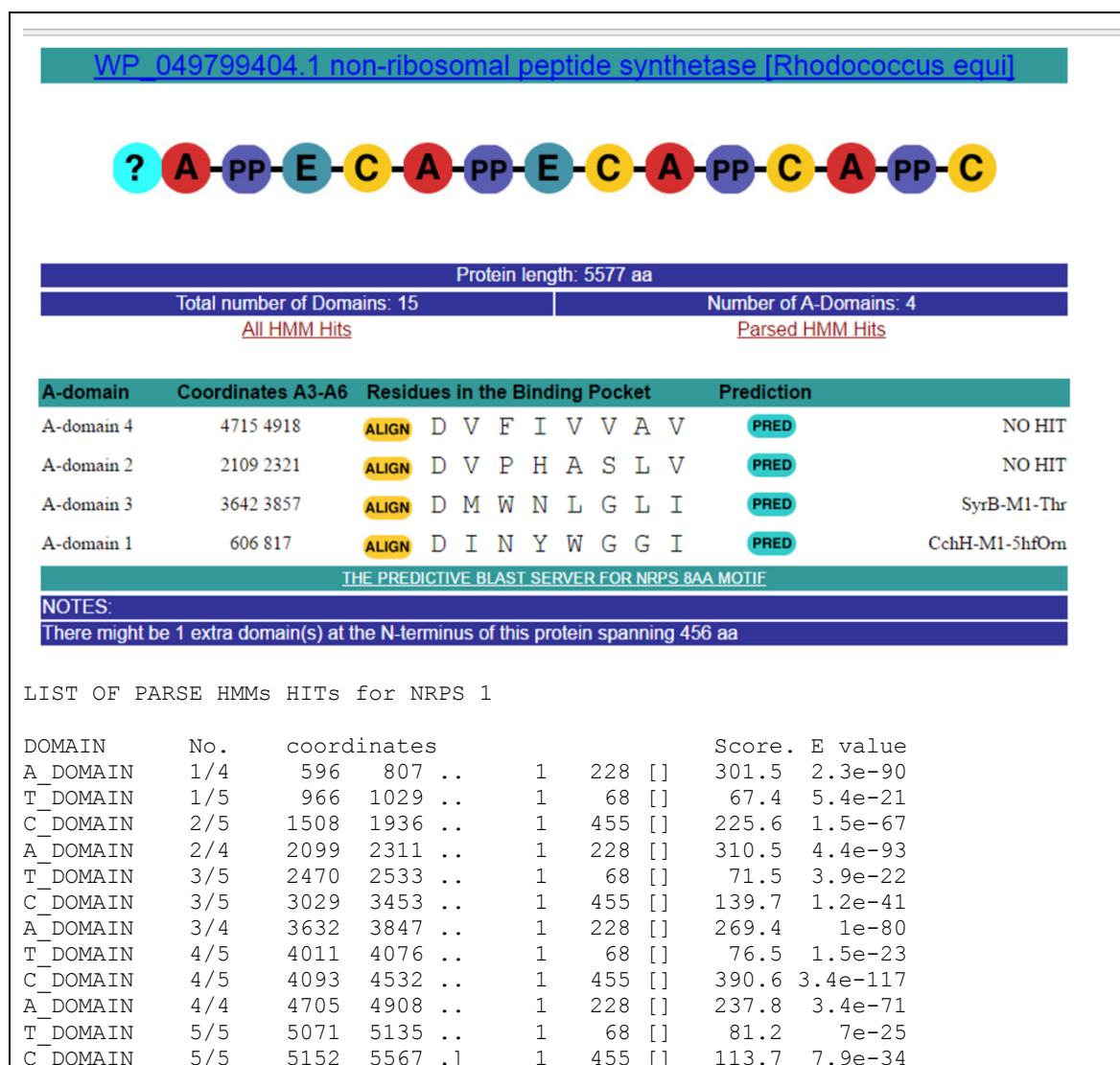


Figure 3-7 NRPS prediction tool using a HMM model to predict potential amino acids used to create the secondary metabolite.

NRPS 1 contains 4 Adenylation domains, thereby producing a metabolite containing 4 linked amino acid structures. The parsed HMM domain hits are given below the graphic, detailing the domain types and numbers, coordinates, score and E value probability.

The predictions associated with the clusters associated with the *R. equi* siderophores were of great interest. The virulence-associated siderophore rhequichelin was predicted as a tetrapeptide of the form N^5 -formyl-hydroxyornithinyl-X-X-X. The prediction of N^5 -formyl-hydroxyornithine at position 1 is considered likely accurate due to the conservation of key residues forming the adenylation domain binding pocket with respect to that of the characterised FxbB in *Mycobacterium smegmatis* (exochelin MS) and CchH in *Streptomyces coelicolor* (Coelichelin) (Figure 3-8).

Query= AD1		
Sequences producing significant alignments:	Score (bits)	E Value
gi 5763943 CchH-M1-5hfOrn Coelichelin synthetase	20	0.008
gi 3560506 gb AAC82549.1 FxbB-M1-5hfOrn Exocheline synthetase	20	0.008
>gi 5763943 CchH-M1-5hfOrn Coelichelin synthetase		
Length = 8		
Score = 20.4 bits (41), Expect = 0.008, Method: Compositional matrix adjust.		
Identities = 8/8 (100%), Positives = 8/8 (100%)		
Query: 1	DINYWGGI	8
	DINYWGGI	
Sbjct: 1	DINYWGGI	8
>gi 3560506 gb AAC82549.1 FxbB-M1-5hfOrn Exocheline synthetase		
Length = 8		
Score = 20.4 bits (41), Expect = 0.008, Method: Compositional matrix adjust.		
Identities = 8/8 (100%), Positives = 8/8 (100%)		
Query: 1	DINYWGGI	8
	DINYWGGI	
Sbjct: 1	DINYWGGI	8

Figure 3-8: A Non-continuous extended motif alignment for adenylation domain 1 binding pocket of the putative rhequichelin in *R. equi* 103S.

Binding site search performed using eightball.pep database; conserved domains are highlighted in yellow.

However, the adenylation domain 3 binding pocket residues identified by the HMM prediction tool differs by one amino acid, from those identified by Miranda-CasoLuengo and colleagues. The conflicting residue identified by the NRPSpredictor2 tool, Tryptophan₃₆₉₃, is contiguously located to the alternatively predicted Glutamic acid₃₆₉₄. The inclusion of Tryptophan₃₆₉₃ in the binding pocket produced significant alignments given in Figure 3-9, with a maximum of 62 % residue identities matching, and Tryptophan₃₆₉₃ being consistent in all alignments. Therefore, an alternative NRPS adenylation domain binding pocket blast search was performed to challenge the robustness of the initial prediction using Glutamic acid₃₆₉₄ rather than Tryptophan₃₆₉₃, the results shown in Figure 3-10 demonstrated a marked increase in identity match (up to 100 %) and a probability significance (E value) increase from 0.082 to 0.024. Given both the reliability score and genetic similarities between rhodococci and mycobacteria, as well as the contiguous locations of the residues it appears plausible that the third adenylation domain is responsible for synthesising *N*⁵-hydroxyornithine as predicted by Miranda-CasoLuengo *et al.* rather than threonine, as initially indicated in this research using the PKS/NRPS analysis web tool.


```

Query= AD3
Database: eightball.pep
Sequences producing significant alignments:
Score      E
(bits) Value
gi|837256|gb|AAA85160.1|SyrB-M1-Thr|syringomycin synthetase 2      17  0.082
gi|9715734|gb|CAC01604.1|AdpB-M3-Thr|Anabaenopeptilide synthetase B  17  0.10
gi|9715733|gb|CAC01603.1|AdpA-M2-Thr|Anabaenopeptilide synthetas...  17  0.10
gi|5763943|CchH-M2-Thr|Coelichelin synthetase                    17  0.14
gi||genome sequencing|PvdD-M8-Thr|Pyoverdin synthetase          17  0.14

>gi|837256|gb|AAA85160.1|SyrB-M1-Thr|syringomycin synthetase 2
Length = 8

Score = 17.3 bits (33), Expect = 0.082, Method: Compositional matrix adjust.
Identities = 5/8 (62%), Positives = 7/8 (87%)
Query: 1  DMWNLGLI 8
          D WNLG++
Sbjct: 1  DFWNLGMV 8
>gi|9715734|gb|CAC01604.1|AdpB-M3-Thr|Anabaenopeptilide
synthetase B
Length = 8

Score = 16.9 bits (32), Expect = 0.10, Method: Compositional matrix adjust.
Identities = 5/8 (62%), Positives = 7/8 (87%)
Query: 1  DMWNLGLI 8
          D WN+G+I
Sbjct: 1  DFWNIGMI 8
>gi|9715733|gb|CAC01603.1|AdpA-M2-Thr|Anabaenopeptilide
synthetase A
Length = 8

Score = 16.9 bits (32), Expect = 0.10, Method: Compositional matrix adjust.
Identities = 5/8 (62%), Positives = 7/8 (87%)
Query: 1  DMWNLGLI 8
          D WN+G+I
Sbjct: 1  DFWNIGMI 8
>gi|5763943|CchH-M2-Thr|Coelichelin synthetase
Length = 8

Score = 16.5 bits (31), Expect = 0.14, Method: Compositional matrix adjust.
Identities = 4/8 (50%), Positives = 7/8 (87%)
Query: 1  DMWNLGLI 8
          D WN+G++
Sbjct: 1  DFWNIGMV 8
>gi||genome sequencing|PvdD-M8-Thr|Pyoverdin synthetase
Length = 8

Score = 16.5 bits (31), Expect = 0.14, Method: Compositional matrix adjust.
Identities = 4/8 (50%), Positives = 7/8 (87%)
Query: 1  DMWNLGLI 8
          D WN+G++
Sbjct: 1  DFWNIGMV 8

```

Figure 3-9: A Non-continuous extended motif alignment to identify the binding pocket residues of Adenylation domain 3 of the NRPS system encoding the putative siderophore rhequichelin.

Binding site search performed using eightball.pep database; matching residue identities are highlighted in yellow, with Tryptophan₃₂₉₃ highlighted in red throughout.

```

Query= >Alternative_hypothesis: DMENLGLI
Database: eightball.pep

Sequences producing significant alignments:
Score      E
(bits) Value
gi|5763943|CchH-M3-5hOrn|Coelichelin synthetase      19  0.024
gi|3560507|gb|AAC82550.1|FxbC-M3-5hOrn|Exochelin synthetase      19  0.024
gi|3560507|gb|AAC82550.1|FxbC-M1-5hOrn|Exochelin synthetase      19  0.024
gi|837256|gb|AAA85160.1|SyrE-M8-Asp/hAsp/Glu|syringomycin synthe... 15  0.44
gi|9715734|gb|CAC01604.1|AdpB-M2-Ahp|Anabaenopeptilide synthetase B 15  0.49

>gi|5763943|CchH-M3-5hOrn|Coelichelin synthetase
Length = 8

Score = 18.9 bits (37), Expect = 0.024, Method: Compositional matrix adjust.
Identities = 8/8 (100%), Positives = 8/8 (100%)
Query: 1  DMENLGLI 8
          DMENLGLI
Sbjct: 1  DMENLGLI 8

>gi|3560507|gb|AAC82550.1|FxbC-M3-5hOrn|Exochelin synthetase
Length = 8

Score = 18.9 bits (37), Expect = 0.024, Method: Compositional matrix adjust.
Identities = 8/8 (100%), Positives = 8/8 (100%)
Query: 1  DMENLGLI 8
          DMENLGLI
Sbjct: 1  DMENLGLI 8

>gi|3560507|gb|AAC82550.1|FxbC-M1-5hOrn|Exochelin synthetase
Length = 8

Score = 18.9 bits (37), Expect = 0.024, Method: Compositional matrix adjust.
Identities = 8/8 (100%), Positives = 8/8 (100%)
Query: 1  DMENLGLI 8
          DMENLGLI
Sbjct: 1  DMENLGLI 8

>gi|837256|gb|AAA85160.1|SyrE-M8-Asp/hAsp/Glu|syringomycin
synthetase
Length = 8

Score = 14.6 bits (26), Expect = 0.44, Method: Compositional matrix adjust.
Identities = 4/8 (50%), Positives = 8/8 (100%)
Query: 1  DMENLGLI 8
          DM++LG++
Sbjct: 1  DMKDLGMV 8

>gi|9715734|gb|CAC01604.1|AdpB-M2-Ahp|Anabaenopeptilide
synthetase B
Length = 8

Score = 14.6 bits (26), Expect = 0.49, Method: Compositional matrix adjust.
Identities = 4/8 (50%), Positives = 7/8 (87%)

Query: 1  DMENLGLI 8
          D+EN G++
Sbjct: 1  DVENAGVV 8

```

Figure 3-10: A Non-continuous extended motif alignment using an alternative NRPS adenylation domain binding pocket residue for the NRPS system encoding the putative siderophore rhequichelin..

Binding site search performed using eightball.pep database; matching residues are highlighted in yellow and conserved Glutamic acid₃₆₉₄ residues are highlighted in green. Glutamic acid₃₆₉₄ used as a variant in candidate residue choice over Tryptophan₃₆₉₃, to predict alternative peptide synthesis by the adenylation domain 3.

When the cluster encoding rhequichelin was further analysed using antiSMASH, a secondary metabolite genome-mining pipeline that identifies and aligns regions likely to generate secondary metabolites at the gene cluster level to their nearest relatives from a database containing all other known gene clusters, and integrates or cross-links all previously available secondary-metabolite

Table 3-9: Biosynthetic gene cluster Region 3 (rhequichelin) features tabulated using Snappene,
Identified by feature names, location, size, colour, directionality and product type.

Feature	Location	Size	Color	Directionality	Type
mce2A	1 .. 1179	1179 bp	Grey	→	CDS putative Mce family protein Mce2A
mce2B	1181 .. 2200	1020 bp	Grey	→	CDS putative Mce family protein Mce2B
mce2C	2197 .. 3255	1059 bp	Grey	→	CDS putative Mce family protein Mce2C
mce2D	3252 .. 4631	1380 bp	Grey	→	CDS putative Mce family protein Mce2D
mce2E	4628 .. 5863	1236 bp	Grey	→	CDS putative Mce family lipoprotein Mce2E
mce2F	5860 .. 7263	1404 bp	Grey	→	CDS putative Mce family protein Mce2F
mas2A	7260 .. 8048	789 bp	Grey	→	CDS putative Mce associated protein Mas2A
mas2B	8045 .. 851	507 bp	Grey	→	CDS putative Mce associated protein Mas2B
REQ_07530	8623 .. 10,152	1530 bp	Grey	←	CDS alpha,alpha-trehalose-phosphate synthase
REQ_07540	10,186 .. 10,386	201 bp	Grey	←	CDS conserved hypothetical protein
REQ_07550	10,495 .. 12,192	1698 bp	Blue	→	CDS putative ABC transporter integral membrane subunit
REQ_07560	12,189 .. 13,967	1779 bp	Blue	→	CDS putative ABC transporter integral membrane subunit
REQ_07570	13,964 .. 14,482	519 bp	Grey	←	CDS hypothetical protein
REQ_07580	14,511 .. 15,338	828 bp	Grey	←	CDS putative secreted tyrosine/serine phosphatase
REQ_07590	15,698 .. 17,347	1650 bp	Grey	→	CDS putative secreted calcineurin-like phosphoesterase
REQ_07600	17,455 .. 17,901	447 bp	Grey	←	CDS hypothetical protein
REQ_07610	17,994 .. 19,064	1071 bp	Grey	←	CDS putative methionyl-tRNA formyltransferase
REQ_07620	19,030 .. 19,845	816 bp	Red	→	CDS alpha/beta hydrolase
REQ_07630	19,975 .. 36,678	16,704 bp	Red	→	CDS putative non-ribosomal peptide synthetase
REQ_07640	36,799 .. 38,154	1356 bp	Red	→	CDS putative lysine/ornithine N-monooxygenase
REQ_07650	38,151 .. 38,384	234 bp	Red	→	CDS MbTH-like protein
REQ_07660	38,399 .. 39,853	1455 bp	Grey	←	CDS putative integral membrane protein
REQ_07670	39,855 .. 40,235	381 bp	Grey	←	CDS conserved hypothetical protein
REQ_07680	41,151 .. 41,756	606 bp	Grey	←	CDS putative membrane protein
REQ_07690	42,153 .. 42,338	186 bp	Grey	←	CDS hypothetical protein
REQ_07700	42,571 .. 42,756	186 bp	Grey	←	CDS hypothetical protein
REQ_07710	42,882 .. 43,520	639 bp	Red	←	CDS putative SAM dependent methyltransferase
REQ_07720	43,517 .. 44,530	1014 bp	Red	←	CDS putative oxidoreductase
REQ_07730	44,583 .. 45,176	594 bp	Grey	→	CDS putative transcriptional regulator
REQ_07750	45,558 .. 46,274	717 bp	Grey	→	CDS putative carboxymethylglutaminase
REQ_07760	46,245 .. 47,090	846 bp	Grey	←	CDS putative secreted protein
REQ_07770	47,194 .. 47,508	315 bp	Grey	←	CDS putative small multidrug resistance protein
phoP	47,737 .. 48,411	675 bp	Green	→	CDS two component system response regulator PhoP
phoR	48,420 .. 49,814	1395 bp	Green	→	CDS two component system sensor kinase PhoR
REQ_07800	49,819 .. 50,274	456 bp	Grey	←	CDS conserved hypothetical protein
REQ_07810	50,425 .. 51,087	663 bp	Grey	→	CDS putative integral membrane protein
REQ_07820	51,219 .. 52,457	1239 bp	Red	→	CDS putative GtrA-like glycosyl transferase
REQ_07830	52,454 .. 54,397	1944 bp	Grey	→	CDS putative dolichyl-phosphate-mannose-protein mannosyltransferase
REQ_07840	54,445 .. 54,687	243 bp	Grey	→	CDS conserved hypothetical protein
REQ_07850	54,770 .. 55,669	900 bp	Grey	→	CDS putative ectoine hydroxylase

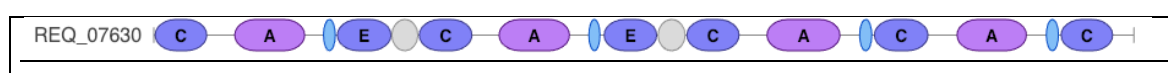


Figure 3-13: The detailed domain annotation for the Biosynthetic gene cluster Region 3 encoded by the NRPS gene REQ_07630.

Domains - C: condensation; A: Adenylation, E- Epimerization; Pale blue circle: Peptide Carrier Protein with attached 4'-phospho-pantetheine, Grey circle: TIGR01720 identified for post-condensation modification.

The antiSMASH monomer predictions for the siderophore rhequichelin are orn - ser - orn - haorn (full predictions results shown in Table 3-10. where a range of bioinformatic prediction tools are pipelined by antiSMASH to predict substrate specificities of non-ribosomal peptide synthetase adenylation domains.)

The main output predictions are based on the NRSPredictor2; SVM and Stachelhaus code predictions tools, which are then collated to present a consensus (Blin et al., 2019).

Table 3-10: Consensus NPRS monomer prediction output as detailed by antiSMASH v5.0.0rc1

REQ_07630: orn - ser - orn - haorn				
Consensus:	AMP-binding (460..857): orn	AMP-binding (1961..2361): ser	AMP-binding (3483..3897): orn	AMP-binding (4562..4956): haorn
NRSPredictor2	orn	ser	orn	haorn
SVM prediction details:				
Predicted physicochemical class:	hydrophilic	hydrophobic-aliphatic	hydrophilic	N/A
Large clusters prediction:	orn, lys, arg	ser, thr, dhp, hpg	orn, lys, arg	N/A
Small clusters prediction:	orn, horn	ser	N/A	N/A
Single AA prediction:	orn	ser	N/A	N/A
Stachelhaus prediction details:				
Stachelhaus sequence:	DINYWGGIGK	DVpHaSLVeK	DMwNLGLINK	DVFiLGAVNK
Nearest Stachelhaus code:	orn	ser	orn	orn
Stachelhaus code match:	100 % (strong)	70 % (weak)	90 % (moderate)	90 % (moderate)

As previously mentioned Miranda-CasoLuengo and colleagues proved the biosynthetic gene cluster *rhbABCDE* (REQ_07610-50) was responsible for production of a putative tetrapeptide hydroxamate siderophore, rhequichelin (Miranda-CasoLuengo et al., 2012). They predicted that the NRPS modules 1 and 3 of REQ_07630 are responsible for assembly of *N*^δ-formyl-*N*^δ-hydroxyornithine and *N*^δ-hydroxyornithine respectively based on binding pocket homology identified by Challis et al. (2000), whereas modules 2 and 4 were predicted to be serine and *N*^δ-acyl-*N*^δ-hydroxyornithine using the bioinformatic tool NRSPredictor2.

The bioinformatic tools used do not take into consideration potential tailoring reactions, that may alter the structure of the assembled peptide, analysis of neighbouring genes include: a putative lysine/ornithine N-monooxygenase, probably to facilitate hydroxylation of L-ornithine to *N*^δ-hydroxyornithine to form a ferric ligands, and a putative methionyl-tRNA formyltransferase which

is likely to catalyse the formylation of N^5 -hydroxyornithine to produce to N^5 -formyl- N^5 -hydroxyornithine (Miranda-CasoLuengo et al., 2012).

Interestingly, a putative secreted tyrosine/serine phosphatase flanks the NRPS encoding rhequichelin, which may prove to be an excellent candidate for future mutagenesis studies to explore any relationship between protein/peptide phosphorylation and iron acquisition. Recently a systematic global analysis of genes encoding protein phosphatases in *Aspergillus fumigatus* identified, via mutagenesis, that a selection of phosphatases are involved directly or indirectly in post-translational modifications affecting the response to iron assimilation, impacting not only the transcription of several genes involved in siderophore biosynthesis (Winkelstroter et al., 2015). However, the apparent extracytoplasmic location of this protein, would indicate a direct functional role in siderophore-mediated iron acquisition in either export or import. Finally, as both the PKS/NRPS analysis web tool (Bachmann and Ravel, 2009) and antiSMASH pipeline (Blin et al., 2017) were not able to detect a canonical type I thioesterase domain to cleave the fully assembled peptide from the peptidyl carrier protein (PCP) domain in the final module, it is reasonable to predict that the product of the upstream gene (REQ_07620), which encodes an α/β -hydrolase fold protein, functions as an external stand-alone type II thioesterase that could release the product from the NRPS synthetase after assembly (Linne et al., 2004). The distinct lack of a functional thioesterase, may be the result of a liberation event that occurs due to by an intramolecular nucleophilic substitution, as described in the paper that recently identified the mixed-ligand siderophores qinichelin produced by *Streptomyces* sp. MBT76. Gubbens et al. (2017) speculates that the condensation domain in the final module catalyses both α -amidation of hOrn-6 to finalise polypeptide and δ -amidation to self-cyclise the last hydroxyornithine to release product.

3.5.4 Comparison of rhequibactin cluster prediction tools

Previously, Miranda-CasoLuengo et al. (2008) identified two NRPS genes of interest, *iupS* (REQ_08140) and *iupT* (REQ_08160) clustered in a six cistron operon and separated by a gene encoding a putative substrate-binding lipoprotein (subsequently discussed in detail). The cluster was based on neighbouring, divergently transcribed operons annotated as encoding putative siderophore transport components as well as genes involved in the production of the siderophore ferric ligand-

bearing precursor 2,3-dihydroxybenzoic acid. This precursor molecule is also present in the closely related rhodococcal species *Rhodococcus erythropolis*, *Rhodococcus rhodochrous*, and *Rhodococcus jostii*, all of which can produce mixed-type catecholate-hydroxamate siderophores: heterobactin (Carran et al., 2001), rhodobactin (Dhungana et al., 2007), and rhodochelin (Bosello et al., 2011) respectively.

While Miranda-CasoLuengo et al. (2008) discussed the presence of the biosynthetic gene cluster likely to encode rhequibactin, along with experimental evidence of a catechol peptide as well as a chromophore, it is surprising that the chemical structure and characterisation of rhequibactin has proven to be elusive even after a decade. Therefore, the NRPS genes listed above (listed as NRPS 2 and NRPS 3 in this analysis) were submitted to the PKS/NRPS analysis web tool (Bachmann and Ravel, 2009), as an independent attempt to predict the siderophore structure, however the tool was unable to accurately predict the adenylation domains specificities for the NRPS (Figure 3-14).

[tr|E4WDC5|E4WDC5_RHOE1 Putative non-ribosomal peptide synthetase](#)
 OS=Rhodococcus equi (strain 103S) GN=REQ_08140 PE=4 SV=1



Protein length: 2580 aa

Total number of Domains: 7

Number of A-Domains: 2

[All HMM Hits](#)

[Parsed HMM Hits](#)

A-domain	Coordinates A3-A6	Residues in the Binding Pocket	Prediction
A-domain 2	1658 1867	ALIGN D S - D L V M V	PRED NO HIT
A-domain 1	615 825	ALIGN D A W E G G F V	PRED NO HIT

THE PREDICTIVE BLAST SERVER FOR NRPS 8AA MOTIF

LIST OF PARSE HMMs HITs for NRPS 2 (tr|E4WDC5|E4WDC5_RHOE1 Putative non-ribosomal peptide synthetase OS=Rhodococcus equi (strain 103S) GN=REQ_08140 PE=4 SV=1)

C_DOMAIN	1/2	8	439 ..	1	455 []	101.7	3.1e-30	
A_DOMAIN	1/2	615	825 ..	1	228 []	220.4	5.9e-66	
T_DOMAIN	1/3	985	1047 ..	1	66 []	71.0	5.1e-22	
C_DOMAIN	2/2	1071	1500 ..	1	455 []	312.8	8.7e-94	
A_DOMAIN	2/2	1658	1867 ..	1	228 []	240.3	6.1e-72	
T_DOMAIN	3/3	2027	2090 ..	1	68 []	71.2	4.7e-22	

[tr|E4WDC7|E4WDC7_RHOE1 Putative non-ribosomal peptide synthetase](#)
 OS=Rhodococcus equi (strain 103S) GN=REQ_08160 PE=4 SV=1



Protein length: 3054 aa

Total number of Domains: 8

Number of A-Domains: 2

[All HMM Hits](#)

[Parsed HMM Hits](#)

A-domain	Coordinates A3-A6	Residues in the Binding Pocket	Prediction
A-domain 2	2152 2364	ALIGN D A W E G G L V	PRED TycC-M2-Gln
A-domain 1	615 825	ALIGN D S A D X X C V	PRED NO HIT

THE PREDICTIVE BLAST SERVER FOR NRPS 8AA MOTIF

LIST OF PARSE HMMs HITs for NRPS 3 (tr|E4WDC7|E4WDC7_RHOE1 Putative non-ribosomal peptide synthetase OS=Rhodococcus equi (strain 103S) GN=REQ_08160 PE=4 SV=1)

C_DOMAIN	1/2	5	440 ..	1	455 []	226.3	9.8e-68	
A_DOMAIN	1/2	615	825 ..	1	228 []	242.7	1.2e-72	
T_DOMAIN	1/2	980	1043 ..	1	68 []	60.2	5.8e-19	
C_DOMAIN	2/2	1544	1971 ..	1	455 []	208.7	1.9e-62	
A_DOMAIN	2/2	2152	2364 ..	1	228 []	228.2	2.6e-68	
T_DOMAIN	2/2	2520	2583 ..	1	68 []	53.6	4.1e-17	

Figure 3-14: NRPS prediction tool using a HMM model to predict amino acyl components of secondary metabolites.

NRPS 2 (top) contains 2 unpredicted Adenylation domains, NRPS 3 (bottom) contains 2 Adenylation domains; one predicted Glutamine and the other unknown. The parsed HMM domain hits are given below the graphic, detailing the domain types and numbers, coordinates, score and E value probability.

Given the lack of identifiable adenylation domains substrate specificities predicted by the PKS/NRPS analysis web tool, the *R. equi* genome was further analysed using the antiSMASH pipeline, the software identified a gene cluster prediction with 54 % homology to the Heterobactin BGC of *R. erythropolis* (Figure 3-15) as well as other homologous gene cluster for both previously identified

pathways and unknown analogous gene clusters.

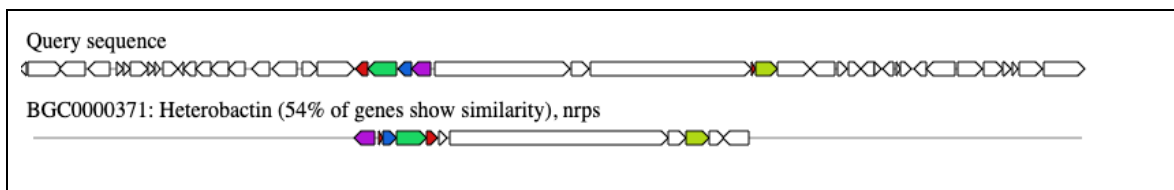


Figure 3-15: antiSMASH KnownClusterBlast search identifying Heterobactin as 54 % similarity to the Query sequence Rhequibactin.

Homologous genes include biosynthetic machinery for production of the precursor peptide 2,3-dihydroxybenzoic acid and a MFS transporter.

Interestingly, the homology between sequences appears to relate to the production of 2,3-dihydroxybenzoic acid as a siderophore-precursor forming catecholate moiety, thereby highlighting a functional similarity. The composition and organisation of this gene cluster is described in Figure 3-16 and Table 3-11, while the annotation of the peptide assembly is shown in Figure 3-17.

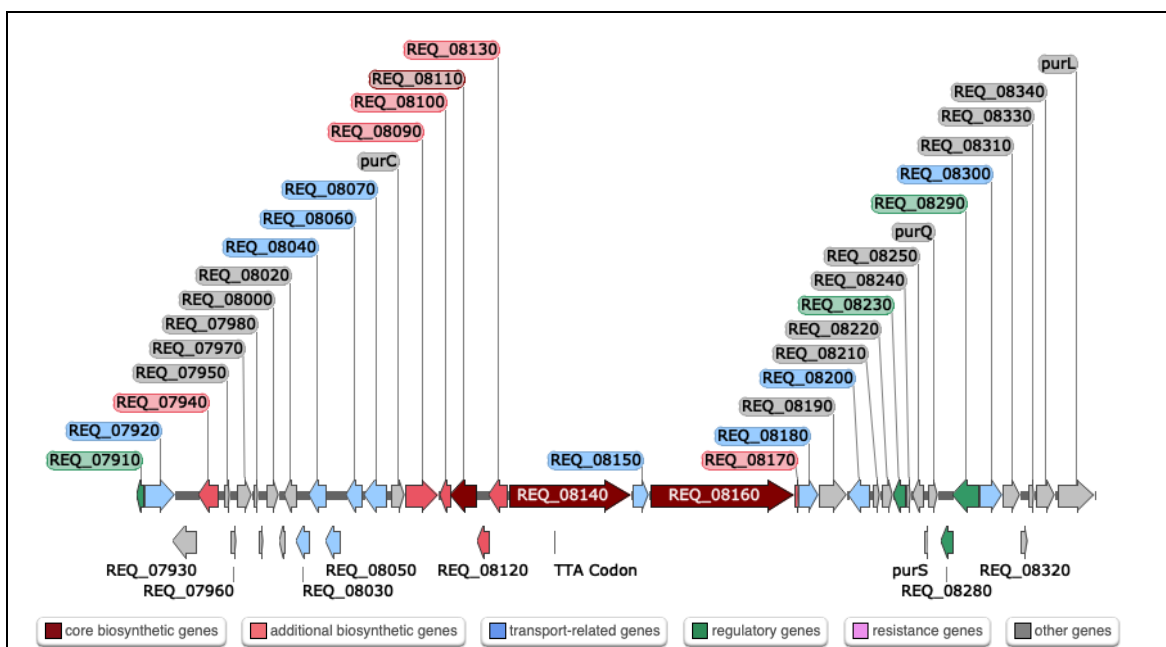


Figure 3-16: Biosynthetic gene cluster Region 4 (rhequibactin) prediction for *R. equi* 103S identified using antiSMASH.

The cluster is displayed using SnapGene, with gene functions colours described in the legend.

Similar to the biosynthetic gene cluster encoding rhequichelin, the antiSMASH pipeline was not able to detect a canonical type I thioesterase domain to cleave the fully assembled peptide from the PCP domain in the final module (Figure 3-17), so an external stand-alone type II thioesterase may be employed to release the product from the NRPS synthetase after assembly (Linne et al., 2004).

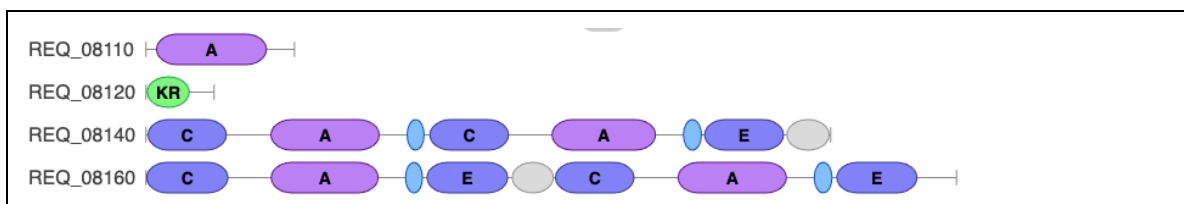


Figure 3-17: The detailed domain annotation for the Biosynthetic gene cluster Region 4 encoded by the precursor genes REQ_08110, REQ_08120, NRPS genes REQ_08140 and REQ_08160.

Domains - C: condensation; A: Adenylation, E- Epimerization KR- Keto-reductase ; Pale blue circle: Peptide Carrier Protein with attached 4'-phospho-pantetheine, Grey circle: TIGR01720 identified for post-condensation modification.

Table 3-11: Biosynthetic gene cluster Region 4 (rhequibactin)

features tabulated using Snappgene, identifying feature names, location, size, colour, directionality and product type.

Feature	Location	Size	Color	Directionality	Type	Product
REQ_07910	1 .. 435	435 bp	Green	←	CDS	putative MarR family transcriptional regulator
REQ_07920	524 .. 2473	1950 bp	Blue	→	CDS	putative ABC transporter ATPase subunit
REQ_07930	2324 .. 3802	1479 bp	Grey	←	CDS	putative fumarate reductase/succinate dehydrogenase
REQ_07940	3927 .. 5153	1227 bp	Red	←	CDS	putative cytochrome P450 monooxygenase
REQ_07950	5562 .. 5945	384 bp	Grey	→	CDS	putative secreted protein
REQ_07960	5957 .. 6385	429 bp	Grey	→	CDS	conserved hypothetical protein
REQ_07970	6378 .. 7394	1017 bp	Grey	→	CDS	putative 1-aminocyclopropane-1-carboxylate deaminase
REQ_07980	7414 .. 7779	366 bp	Grey	→	CDS	putative small multidrug resistance protein
REQ_07990	7776 .. 8105	330 bp	Grey	→	CDS	putative small multidrug resistance protein
REQ_08000	8275 .. 9066	792 bp	Grey	→	CDS	secreted steroid dehydrogenase
REQ_08010	9106 .. 9396	291 bp	Grey	←	CDS	conserved hypothetical protein
REQ_08020	9393 .. 10,145	753 bp	Grey	←	CDS	putative ferredoxin-like protein
REQ_08030	10,142 .. 10,954	813 bp	Blue	←	CDS	putative ABC transporter ATPase subunit
REQ_08040	10,951 .. 12,009	1059 bp	Blue	←	CDS	putative ABC transporter integral membrane subunit
REQ_08050	12,006 .. 12,950	945 bp	Blue	←	CDS	putative ABC transporter substrate binding lipoprotein
REQ_08060	13,305 .. 14,363	1059 bp	Blue	←	CDS	putative ABC transporter substrate binding lipoprotein
REQ_08070	14,487 .. 15,905	1419 bp	Blue	←	CDS	putative MFS transporter
purC	16,199 .. 17,083	885 bp	Grey	→	CDS	phosphoribosylaminoimidazole-succinocarboxamide synthase PurC
REQ_08090	17,091 .. 19,214	2124 bp	Red	→	CDS	putative serine peptidase
REQ_08100	19,283 .. 19,933	651 bp	Red	←	CDS	putative isochorismatase
REQ_08110	19,941 .. 21,623	1683 bp	Red	←	CDS	putative 2,3-dihydroxybenzoate-AMP ligase
REQ_08120	21,674 .. 22,447	774 bp	Red	←	CDS	putative short chain dehydrogenase
REQ_08130	22,444 .. 23,589	1146 bp	Red	←	CDS	putative isochorismate synthase
REQ_08140	23,767 .. 31,509	7743 bp	Red	→	CDS	putative non-ribosomal peptide synthetase
TTA Codon	26,620 .. 26,622	3 bp	White	→	misc_feature	
REQ_08150	31,619 .. 32,638	1020 bp	Blue	→	CDS	putative substrate binding lipoprotein
REQ_08160	32,730 .. 41,894	9165 bp	Red	→	CDS	putative non-ribosomal peptide synthetase
REQ_08170	41,931 .. 42,155	225 bp	Red	→	CDS	putative isochorismatase
REQ_08180	42,163 .. 43,437	1275 bp	Blue	→	CDS	putative MFS transporter
REQ_08190	43,453 .. 45,276	1824 bp	Grey	→	CDS	putative siderophore binding protein
REQ_08200	45,326 .. 46,678	1353 bp	Blue	←	CDS	putative MFS transporter
REQ_08210	46,869 .. 47,363	495 bp	Grey	→	CDS	putative thiol peroxidase
REQ_08220	47,464 .. 48,168	705 bp	Grey	→	CDS	conserved hypothetical protein
REQ_08230	48,179 .. 48,925	747 bp	Green	←	CDS	putative TetR family transcriptional regulator
REQ_08240	48,998 .. 49,387	390 bp	Grey	→	CDS	putative glyoxalase family protein
REQ_08250	49,401 .. 50,048	648 bp	Grey	←	CDS	metallo-beta-lactamase superfamily protein
purS	50,190 .. 50,426	237 bp	Grey	→	CDS	phosphoribosylformylglycinamide synthase PurS
purQ	50,423 .. 51,100	678 bp	Grey	→	CDS	phosphoribosylformylglycinamide synthase I PurQ
REQ_08280	51,229 .. 51,939	711 bp	Green	←	CDS	two component system response regulator
REQ_08290	51,936 .. 53,573	1638 bp	Green	←	CDS	putative two component system sensor kinase
REQ_08300	53,722 .. 55,149	1428 bp	Blue	→	CDS	putative sodium/dicarboxylate symporter
REQ_08310	55,193 .. 56,329	1137 bp	Grey	→	CDS	conserved hypothetical protein
REQ_08320	56,326 .. 56,805	480 bp	Grey	→	CDS	conserved hypothetical protein
REQ_08330	56,838 .. 57,188	351 bp	Grey	→	CDS	putative glyoxalase family protein
REQ_08340	57,251 .. 58,552	1302 bp	Grey	→	CDS	putative metallopeptidase
purL	58,694 .. 60,973	2280 bp	Grey	→	CDS	phosphoribosylformylglycinamide synthase II PurL

The antiSMASH v5.00rc1 polymer prediction for the siderophore rhequibactin are (d**hb**) + (X - X) + (X - X), where d**hb** is the precursor 2,3-dihydroxybenzoic acid, the catecholate moiety, produced by REQ_08110, identified as a consensus using NRPSPredictor2; SVM and Stachelhaus code predictions. However, the other monomer predictions are unclassified (X), likely due to the uncertainty of AMP-binding domains prediction algorithms employed in antiSMASH v5.00rc1 prediction tools. Upon updating from v4.2.1 to v5.00rc1, the source code was refactored to improve runtime and program stability, however as a result some of the externally-contributed modules (SANDPUMA, trans-AT PKS comparisons, terpene PrediCAT) were temporarily removed while contributors prepare v5.x compliant versions (Blin et al., 2019).

Utilising a stable version of antiSMASH v4.2.1 facilitated a consensus prediction method based using the SANDPUMA ensemble, as this was proven to out-perform each of the other methods individually (Blin et al., 2017), the underlying algorithm predictions as well as SVM, profileHMM and PrediCAT predictions are still shown and are used in forming the consensus. The NRPS prediction output using antiSMASH 4.2.1 given in Table 3-12 suggest a polypeptide consisting of the monomers (d**hb**|sal) + (lys-orn) + (orn-[d-cyclo-horn]).

Table 3-12: Consensus NPRS monomer prediction output as detailed by antiSMASH v4.2.1

Monomers prediction:					
(dhb sal) + (lys-orn) + (orn-[d-cyclo-horn])					
Gene	REQ_08110	REQ_08140		REQ_08160	
Monomer	1	2	3	4	5
Stachelhaus code:	d hb sal	N/A	N/A	N/A	d-cyclo-horn/orn
NRPSPredictor3	dhb	N/A	N/A	N/A	N/A
SVM:	dhb	lys	orn	orn	orn
pHMM:	dhb	lys	orn	orn	orn
PrediCAT:	N/A-N/A	Q45R83_A1-8--lys	N/A-N/A	N/A-N/A	N/A-N/A
SANDPUMA ensemble:	d hb sal	ala	orn	orn	orn
%ID to nearest neighbour:	62.781	47.325	47.561	49.18	57.32
SNN score:	0.0157247776787	0.0	0.0	0.0	0.0

3.5.5 Predictive structures of *R. equi* siderophores

The lack of experimental characterisation of the *R. equi* siderophores by previous studies, is indicative that *R. equi* does not regulate production of its siderophores in a conventional rate analogous to other iron starved bacteria. Therefore, the bioinformatic analysis of biosynthetic gene clusters was used to produce putative structures and associated masses that can be used in targeted approaches to siderophore characterisation.

3.5.5.1 Rhequichelin putative structure

Following the previous analysis of the rhequichelin biosynthetic gene cluster, and the predicted specificities of the NRPS adenylation domains in the context of the the previous literature (Miranda-CasoLuengo et al., 2012), it is hypothesised that the hydroxamate siderophore rhequichelin is comprised of the peptide monomers: N^5 -formyl- N^5 -hydroxyornithine -serine - N^5 -hydroxyornithine - N^5 -acyl- N^5 -hydroxyornithine.

Assembly of the hypothetical structure was performed with support from Dr. D. Tetard, following the orthodox peptide assembly parameters. Using the previous in-depth predictions made by Miranda-CasoLuengo, modifications have been considered for this putative structure as shown in Figure 3-18.

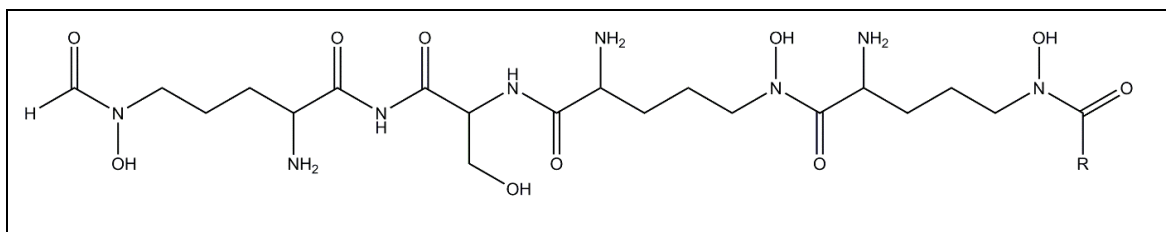


Figure 3-18: Predictive structure of the tetrapeptide rhequichelin.

Due the prediction of a N^5 -acyl- N^5 -hydroxyornithine peptide monomer, deduction of an accurate mass clearly becomes more challenging. The predicted amphiphilic nature of this siderophore suggests that this undefined acylation does not directly contribute to metal coordination, but the presumed fatty-acid attachment may vary in length, degree of unsaturation and functionality (Kem and Butler, 2015). A non-exclusive selection of potential acylation groups and their predicted masses were compiled (Table 3-13), that might support the identification of candidate molecules in a targeted metabolomic approach to the molecular characterisation of rhequichelin.

Table 3-13: Potential variations in acyl groups for rhequichelin, C1-6 and associated predicted masses

R =	C1:0	C2:0	C3:0	C4:0	C5:0	C6:0
Chemical Formula:	C ₂₁ H ₄₀ N ₈ O ₁₀	C ₂₂ H ₄₂ N ₈ O ₁₀	C ₂₃ H ₄₄ N ₈ O ₁₀	C ₂₄ H ₄₆ N ₈ O ₁₀	C ₂₅ H ₄₈ N ₈ O ₁₀	C ₂₆ H ₅₀ N ₈ O ₁₀
Exact Mass:	564.29	578.30	592.32	606.33	620.35	634.36
Molecular Weight:	564.60	578.62	592.65	606.6	620.71	634.73

3.5.5.2 Rhequibactin putative structure

After the analysis of the rhequibactin biosynthetic gene cluster, combined with the predicted specificities of NRPS adenylation domains in the context of the previous literature, it is almost certain that rhequibactin utilises 2,3-hydroxybenzoic acid residues for iron coordination, given the proximity of identification of the upstream 2,3-DHB biosynthetic machinery, previous experimental verification of catecholates moieties (Miranda-CasoLuengo et al., 2008). Additionally, two siderophores from the *Rhodococcus* genus, heterobactin A and rhodobactin isolated from *Rhodococcus erythropolis* IGTS8 and *Rhodococcus rhodochrous* OFS respectively, (Bosello et al., 2013; Dhungana et al., 2007) both contain 2,3-dihydroxybenzoic acid.

As with rhequichelin, assembly of the hypothetical structure rhequibactin was performed with direction from Dr. D. Tetard, following the orthodox linear peptide assembly, a catecholate-hydroxamate mixed ligand model of the siderophore rhequibactin was constructed (Figure 3-19) using the peptide monomers; 2,3-dihydroxybenzoic acid, lysine, ornithine (x2), and a cyclised δ -N-hydroxy-ornithine.

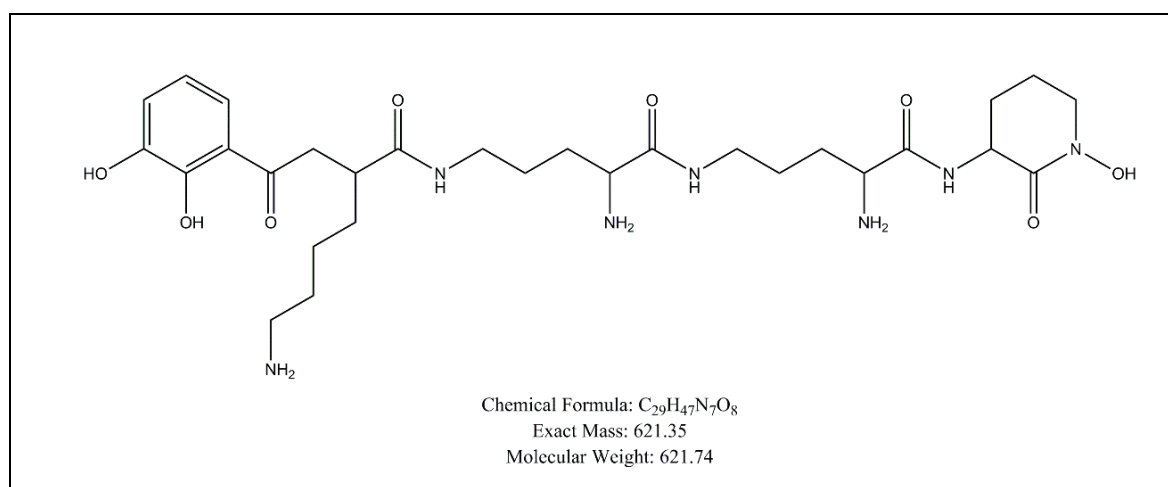


Figure 3-19: Rhequibactin predicted structure

Intriguingly, the predicted structure of rhequibactin is reminiscent of the recently identified mixed-ligand siderophores qinichelin produced by *Streptomyces* sp. MBT76 (Figure 3-20a) (Gubbens et al., 2017) and the heterobactins produced by *Rhodococcus erythropolis* PR4 (Bosello et al., 2013). Assembly of qinichelin follows an orthodox linear peptide assembly logic, where the catecholate-hexapeptide backbone (Ser-1, Orn-2, Ser-3, Ser-4, hOrn-5, and hOrn-6), is attached to the precursor 2,3-DHB and are sequentially bound and converted to aminoacyl adenylates.

Assembly of heterobactin B also follows an orthodox linear peptide assembly logic, with a catecholate tripeptide backbone (Arg-1, Gly-2 and hOrn-3) attached to a precursor 2,3-DHB. The presence of 2,3-DHB and modified ornithine residues in not only the heterobactins, but both rhodochelin and rhodobactin, implies that these shared features are indicative of a conserved motif within the *Rhodococcus* genus (Bosello et al., 2013).

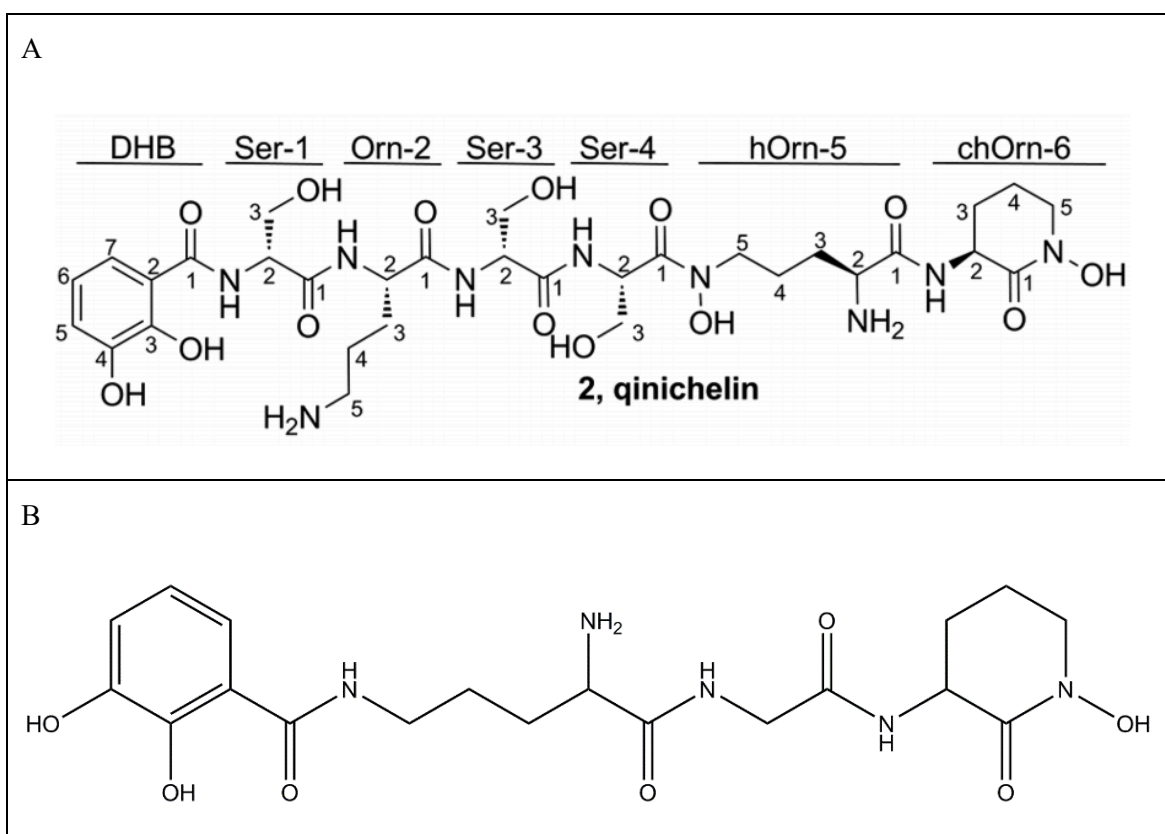


Figure 3-20: Siderophore structures analogous to the predicted structure of rhequibactin.
a) qinichelin. b) heterobactin B

When considering peptidic modifications, rhequibactin appears to utilise three epimerisation domains on monomers 3, 4 and 5; resulting in assembly of ornithine D-enantiomers. Furthermore, without experimental evidence, accurate bioinformatic identification of ornithine tailoring can be challenging, especially between unmodified ornithine (Orn), δ -N-hydroxyl and ornithine (hOrn)

residues, however although not identified as part of the rhequibactin biosynthetic gene clusters, the putative lysine/ornithine N-monooxygenase involved in rhequichelin synthesis may be utilised in rhequibactin synthesis, as functional crosstalk can occur between biosynthetic gene clusters.

It remains questionable as to how liberation of rhequibactin occurs, however given the lack of canonical C-terminal thioester domain, it hypothetically could occur in an analogous manner to qinichelin by self-cyclisation of the last hydroxyornithine to release the compound.

3.6 Iron transport in *R. equi*

As previously described in chapter one, iron is essential for life, therefore pathogenic microbes must have associated virulence mechanisms to ensure enough is available intracellularly to survive. Siderophores, as previously alluded are examples of such a virulence factor, however the mechanisms associated with cellular transport are poorly described. *R. equi* appears to have a putative ABC siderophore transport system IupABC (REQ_24080 – REQ_24100, Door² operon: 767082) (Miranda-CasoLuengo et al., 2008), homologous to the FxuABC transport system in *M. smegmatis*; CchDCE system of *S. coelicolor* and belongs to the binding-protein-dependent transport system permease family – FecCD (Accession: pfam01032) (Staudenmaier et al., 1989). The conserved domain analysis of IupABC is described in Figure 3-21 (Marchler-Bauer et al., 2017).

The similarities between the transport system for the hydroxamate siderophore coelichelin (*cchCDE*) and the siderophore transport system encoded by *iupABC* in *R. equi* further suggest that the transport system is likely responsible for cellular iron transport (Miranda-CasoLuengo et al., 2008). However, as *R. equi* is predicted to synthesise two siderophores, it is logical to assume that there is at least one additional transport system for cellular transfer of iron, and that both siderophore systems do not rely on the single operon *iupABC*. Other putative iron transport systems identified by manual *R. equi* genome mining include; a canonical ABC transport operon containing a substrate-binding lipoprotein, two permeases and an ATPase (REQ_4700-47440), a heme oxygenase and transport system (REQ_22880-28910), a ferredoxin transport system (REQ_08010-08050) and a EfeB exogenous heme recovery system (REQ_12590-12610).

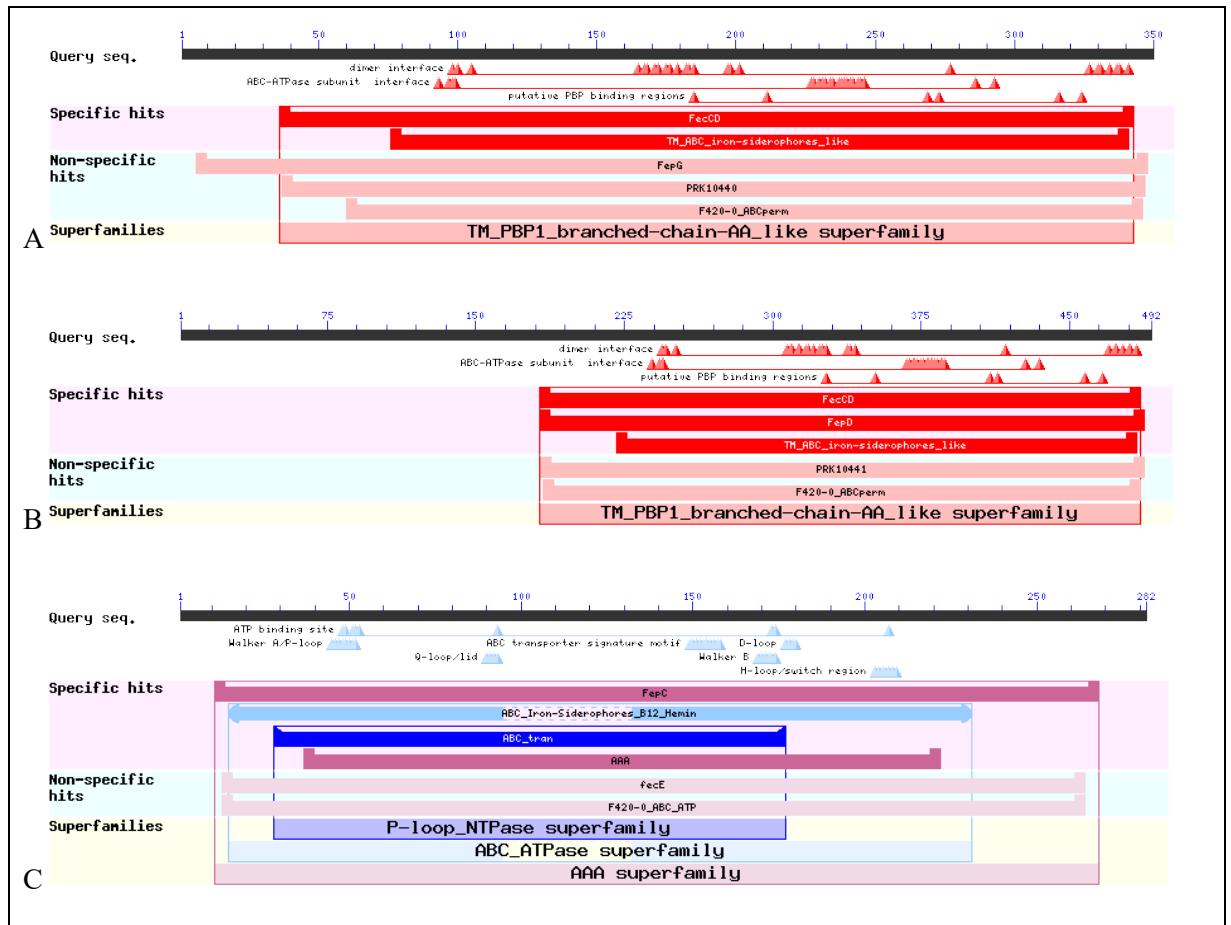


Figure 3-21: Conserved domains for operon iupABC.

(A) putative ABC transporter integral membrane subunit IupA [*Rhodococcus equi* 103S] Domain hits identified include: FecCD, accession: pfam01032, interval: 192-486, E-value: 2.77 e-87, FepD, accession: COG0609, interval: 182-488, E-value: 1.94 e-83. (B) putative ABC transporter integral membrane subunit IupB [*Rhodococcus equi* 103S] Domain hits identified include: FecCD, accession: pfam01032, interval: 36-343, E-value: 7.06 e-67, FepG, accession: COG4779, interval: 6-348, E-value: 7.24 e-78. (C) putative ABC transporter ATPase subunit IupC [*Rhodococcus equi* 103S] Domain hits identified include: FepC, accession: COG1120, interval: 11-268, E-value: 4.50 e-132. Database: CDSEARCH/cdd Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 (Marchler-Bauer et al., 2017)

Currently, the methods of rhodococcal iron import have not been well characterised, however, other biologically analogous cytoplasmic transport systems have been hypothesised.

Challis (2008) predicted that the transport proteins for the *Streptomyces coelicolor* siderophore - coelichelin encoded by the *cchCDEF* genes are functionally similar to other known ferric-siderophore systems, including use of a lipoprotein receptor, permease and ATPase components as previously described by Miethke and Marahiel (2007). Interestingly, this transport system appears to utilise an iron-siderophore binding lipoprotein receptor encoded by the gene *cchF*, the hypothesis proposed is that after initial capture of the ferri-siderophore complex by CchF, the transfer of the iron-bound siderophore through the membrane to the cytoplasm is mediated by the membrane-

spanning permeases and an ATPase as shown in Figure 3-22 (Challis, 2008), before release of ferrous iron by reduction as described in chapter one.

Similarly, *Bacillus cereus* utilises a siderophore-shuttle mechanism for iron uptake (Fukushima et al., 2014), using YxeB - a lipoprotein siderophore-binding protein encoded by *BC_0383* that is located directly upstream of putative permease genes *BC_0381* and *BC_0382*.

However, unlike *S. coelicolor* and *B. cereus*, *R. equi* does not appear to have a locally encoded lipoprotein receptor, therefore the candidate lipoprotein homologous to CchF may be an orphan gene located elsewhere in the chromosome. A BlastN search for CchF NP_624811.1 with taxa ID restricted to *Rhodococcus equi* 103S (Table 3-14) highlighted nine candidate lipoproteins, associated with iron-uptake proteins, from which 4 could be excluded due to their association with coherent transport operons (greyed-out tabulated results). Likewise, a BlastP search for YxeB NP_830222.1 with taxa ID restricted to *Rhodococcus equi* 103S (Table 3-15) highlighted 5 candidate lipoproteins, four of which were previously identified in the CchF BlastP search above. Where 3 of the candidates were deemed to be suitable; one candidate, the ABC transporter substrate-binding protein [*Rhodococcus hoagii*] encoded by the gene REQ_08150 is located between the two-NRPS system predicted to synthesise rhequibactin. While no further elucidations can be made without experimental evidence, the genomic analysis in combination with literature reviews for Gram-positive transport systems provide a valuable insight to candidates for future characterisation of iron-regulatory transport systems within *R. equi*, which are currently undefined.

The mechanism around cellular exportation of siderophores to sequester the iron for survival is currently not well described, with literature (mainly described in Gram-negative bacteria) elucidating transport is likely to occur using members of the major facilitator superfamily (MFS)-type transporter and are typically transcriptionally regulated by the ferric uptake repressor Fur.

Rhequibactin, like other siderophores, potentially features aromatic rings and side-chains capable of binding ferric iron, these characteristics are likely to prevent passive diffusion through the lipid domain of the membrane and suggest that cellular export is an active process. The membrane protein with homology to the MFS-class efflux pump suggests that the protein encoded by REQ_08180 functions as a rhequibactin exporter similar to EntS for enterobactin.

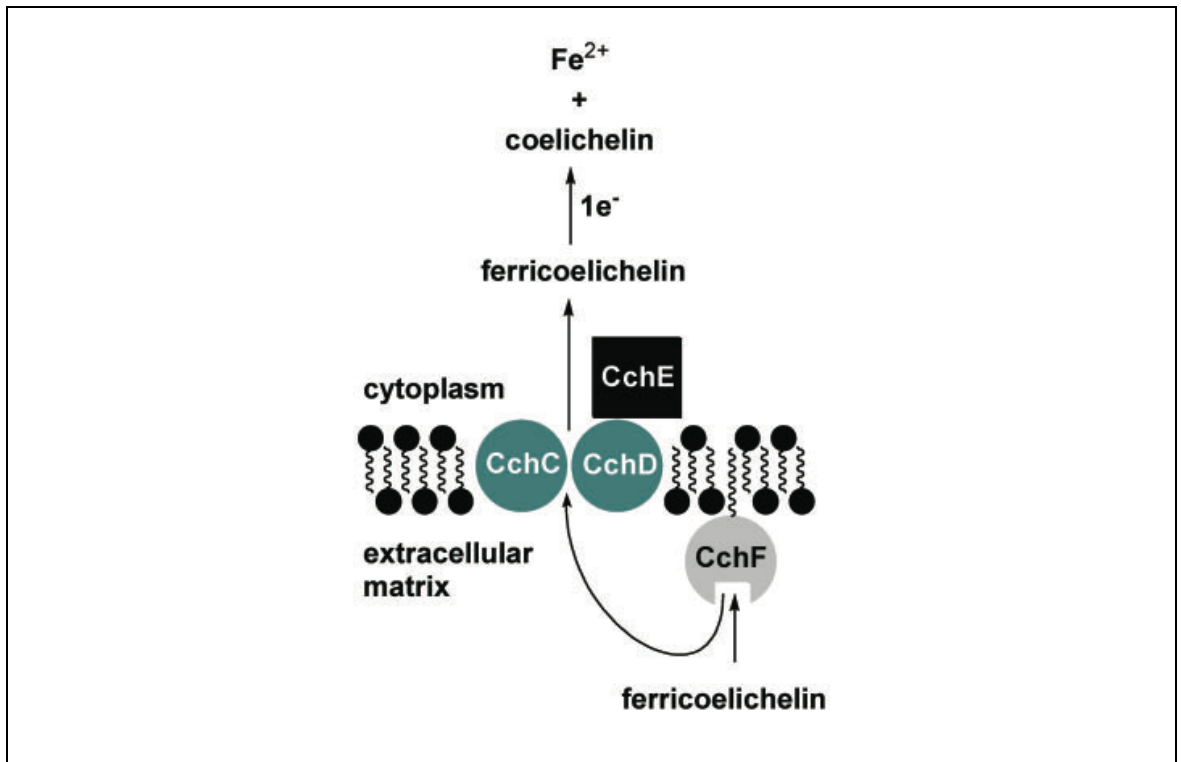


Figure 3-22: Proposed siderophore mediated uptake and transport of iron employed by *S. coelicolor* using an ABC-transporter system – CchCDEF.

CchF - lipoprotein receptor, CchCD – membrane spanning permeases, CchE – ATPase. Ferrous iron is predicted to be released intracellularly from the ferric-siderophore complex by reduction (Challis, 2008).

Table 3-14: BlastP results for NP_624811.1 iron-siderophore binding lipoprotein [*Streptomyces coelicolor* A3(2)] restricted to TxID: *Rhodococcus equi* 103S
 Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects, Lipoprotein prediction were also calculated using PRED-LIPO and reliability scores given (Bagos et al., 2008). Greyed out results indicated incompatibility for orphan gene prediction.

Description	Score	Query cover	E value	Ident	Accession	Gene ID	Associated genes in operon	Lipo signal	Lipo reliability score
MULTISPECIES: iron-siderophore ABC transporter substrate-binding protein [Rhodococcus]	155	95%	4e-45	35%	WP_013415831.1	REQ_20510	none	Yes	1.000
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	154	95%	7e-45	35%	WP_013414625.1	REQ_03550	fmn flavoprotein	Yes	0.995
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	150	82%	2e-43	34%	WP_013416507.1	REQ_29940	puromycin N-acetyltransferase, S9 family peptidase	Yes	0.967
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	114	86%	1e-29	30%	WP_013417129.1	REQ_39540	none	Yes	0.997
MULTISPECIES: iron-siderophore ABC transporter substrate-binding protein [Rhodococcus]	89.4	77%	1e-20	29%	WP_005515932.1	REQ_04870	none	Yes	0.996
Fe ²⁺ -enterobactin ABC transporter substrate-binding protein [Rhodococcus hoagii]	80.9	90%	7e-18	29%	WP_013414617.1	REQ_03460	none	Yes	0.493
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	75.1	79%	7e-16	27%	WP_013416640.1	REQ_31890	none	Yes	0.998
iron ABC transporter substrate-binding protein [Rhodococcus hoagii]	33.5	82%	0.024	24%	WP_013417667.1	REQ_47000	None, but upstream of two iron ABC permeases and a iron ABC transporter ATP-binding protein	Yes	0.995
putative ABC transporter substrate binding lipoprotein [Rhodococcus hoagii 103S]	25.4	28%	9.8	23%	CBH48336.1	REQ_22900	Heme oxygenase and associated permease and ATP-binding protein.	Yes	0.993

Table 3-15: BlastP results for NP_830222.1 ferrichrome-binding protein [*Bacillus cereus* ATCC 14579] restricted to TxID: *Rhodococcus equi* 103S

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects, Lipoprotein prediction were also calculated using PRED-LIPO and reliability scores given (Bagos et al., 2008). Greyed out results indicated incompatibility for orphan gene prediction. Yellow highlighted result indicated a significant result of interest.

Description	Score	Query cover	E value	Ident	Accession	Gene ID	Associated genes in operon	Lipo signal	Lipo reliability score	CchF search
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	65.5	83%	7e-13	27%	WP_013416640.1	REQ_31890	none	Yes	0.998	Yes
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	58.5	51%	2e-10	32%	WP_013414625.1	REQ_03550	fmm flavoprotein	Yes	0.995	Yes
Fe2+-enterobactin ABC transporter substrate-binding protein [Rhodococcus hoagii]	54.7	82%	3e-09	24%	WP_013414617.1	REQ_03460	none	Yes	0.493	Yes
ABC transporter substrate-binding protein [Rhodococcus hoagii]	52.4	58%	2e-08	23%	WP_005515427.1	REQ_08150	none situated between dual NRPS system encoding rhequibactin	Yes	0.998	No
iron-siderophore ABC transporter substrate-binding protein [Rhodococcus hoagii]	49.7	78%	1e-07	26%	WP_013416507.1	REQ_29940	puromycin N-acetyltransferase, S9 family peptidase	Yes	0.967	Yes

3.7 A bioinformatic analysis into transcriptional control of iron-responsive genes

Bacterial iron acquisition (as previously discussed) is typically regulated by iron-responsive proteins belonging to the metalloregulator families Fur and DtxR. These transcriptional regulators sense the appropriate metal ions and upon binding to the cognate metal, can either activate or prevent DNA operator binding. Metalloregulatory co-repressors are of particular interest in managing metal-homeostasis, as the corepressor-operator association prevents RNA polymerase binding thereby inhibiting the initiation of gene transcription.

3.7.1 Genomic analysis of DtxR(-like) co-repression

The protein IdeR, appears to be well conserved within rhodococci (demonstrated earlier), as well as being essential for mycobacteria (Rodriguez et al., 2002a). It is both a structural and functional homologue of DtxR from *Corynebacterium diphtheriae* (Kunkle and Schmitt, 2003) and in *Mycobacterium smegmatis* functions as a co-repressor responsible for iron-dependent repression of siderophore biosynthesis (Dussurget et al., 1999).

IdeR, like Fur and DtxR, has been experimentally confirmed to allosterically bind iron and interact with a defined sequence (known as the iron-box) in the operator regions of iron-regulated genes to control their transcription. Previously, Rodriguez et al. (2002a) performed global expression profiling with DNA microarrays comparing wild-type *M. tuberculosis* H37Rv with an *ideR* mutant derivative to compile a list of gene candidates repressed by IdeR. This definition of the IdeR regulome facilitated a comparison of known IdeR operator sequences from *M. tuberculosis* and the proposal of a 19-bp consensus iron box sequence.

3.7.1.1 Identification of *R. equi* 103S co-repressor recognition sequences

Given the inherent similarities between rhodococcal and mycobacterial regulatory systems, it was feasible to hypothesise that rhodococci would utilise an orthologous iron-regulatory mechanism through allosteric co-repression. The IdeR operator sequence of the *Mycobacterium smegmatis* *fxbA* gene was initially used as a template to identify any putative rhodococcal iron-boxes.

Although no identical sequence matches were identified, allowing up to 4 nucleotide mismatches facilitated a list of candidates to be produced, the sequences were then compared using ClustalΩ 1.2.4 multiple sequence alignment to identify conserved nucleotides and produce a putative *R. equi* iron-box consensus sequence Figure 3-23, as well as a nucleotide logo comparison between a consensus of identified ideR binding sites and the canonical iron box consensus, from which there was minimal difference.

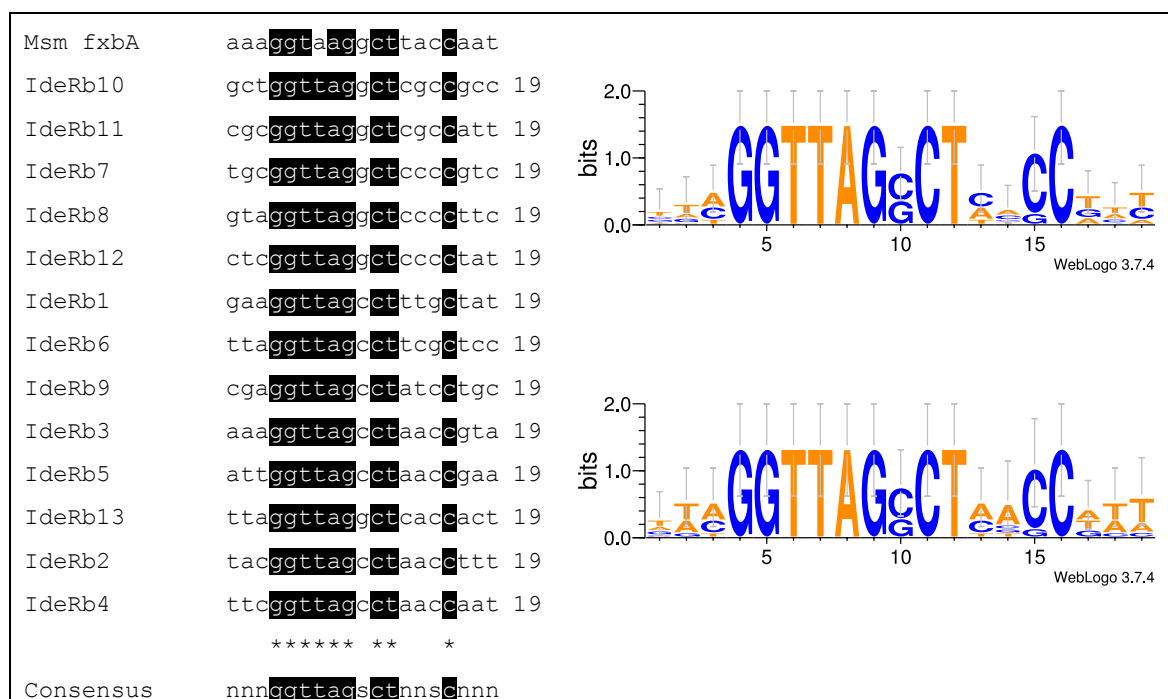


Figure 3-23: Multiple sequence alignment of putative iron-box operators in the *R. equi* 103S chromosome, using the experimentally verified iron-box from *fxbA* in *Mycobacterium smegmatis* as a template.

N = A, C, G, or T; W = A or T; S = C or G; V = A, C, or G. Solid boxes (black) indicate a predominance of one nucleotide in the IdeR binding site sequences.

To identify if the recognition sequences function as an iron-responsive operator, the genomic context of the putative IdeR binding sites was also considered (Table 3-16). To be considered a canonical operator sequence, the recognition sequence should be present on the DNA strand upstream of the gene of interest, to allow the RNA polymerase to drive transcription as shown in the example, Figure 3-24.

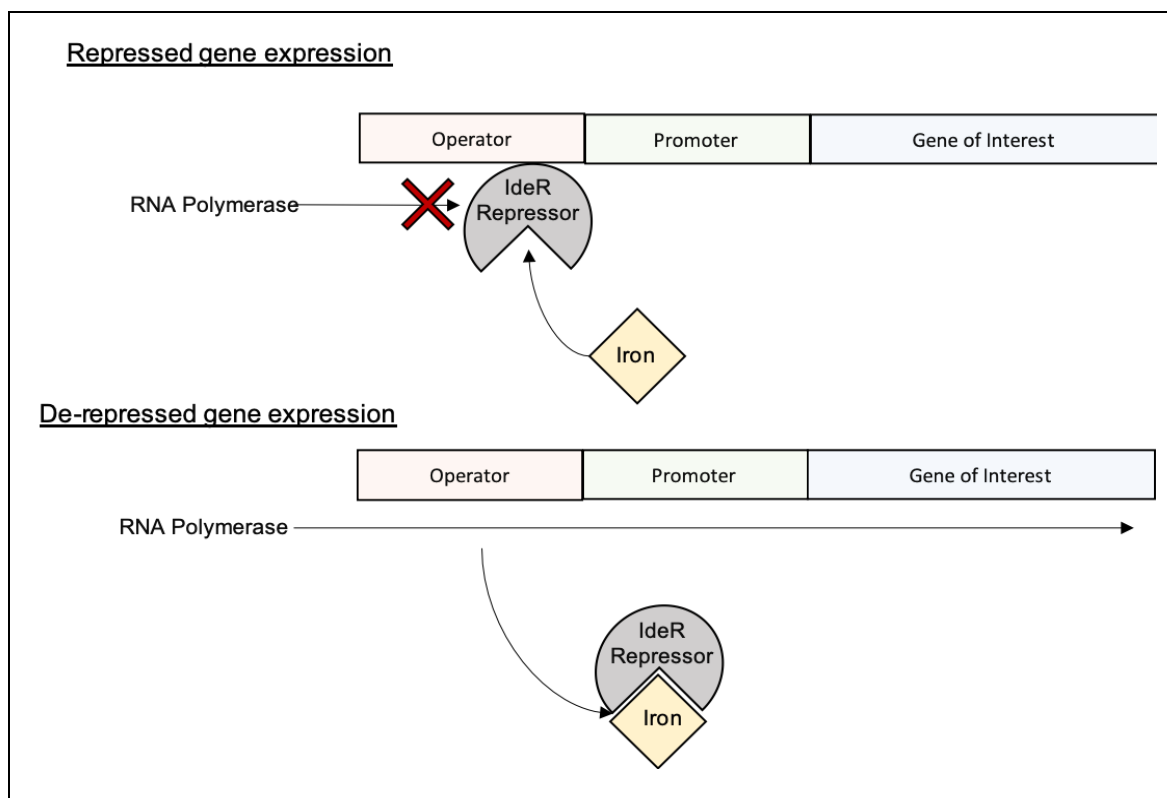


Figure 3-24: The basic structure of prokaryotic transcriptional de-repression.

The binding of RNA polymerase to the promoter is prevented by the repressor binding to the DNA at the operator. De-repression occurs when an inducer (given as iron above) binds to the repressor. The operator-DNA binding affinity is reduced, facilitating transcription by allowing RNA polymerase access to the promoter driving gene expression.

Table 3-16 lists the putative iron boxes, further identifying chromosomal coordinates, the recognised sequence, genomic location and context, as well as associated gene ID and annotations. The SnapGene sequence visualiser and search function allowed for rapid identification of sequence context, and as such the candidate list could be further categorically refined by: Canonical de-repression, non-canonical and potentially random encounters.

Firstly, the canonical de-repression operons identified include IdeRB1 upstream of a putative lysine/ornithine N-monooxygenase (REQ_07640), a gene predicted to facilitate hydroxylation of L-ornithine to *N*⁵-hydroxyornithine in the production of the siderophore rhequichelin.

Interestingly, IdeRB2/3 utilises a perfectly palindromic recognition sequence 5'- ggtagcctaacc - 3' on both DNA strands to potentially inhibit transcription in both directions, this would be functionally useful as downstream genes in either direction is directly relevant to bacterial iron sequestration employing the putative siderophore rhequibactin. Downstream of the IdeRB2 site is the non-ribosomal peptide synthetase (REQ_08140) predicted to synthesise rhequibactin,

furthermore Downstream of IdeRB3 the operon (REQ_08130, REQ_08120, REQ_08110) is predicted to produce the rhequibactin precursor catechol moiety 2,3-dihydroxybenzoic acid.

IdeRB5 and IdeRB8 recognition sites are upstream of genes encoding a substrate-binding lipoprotein (REQ_08150) and an ABC transporter substrate-binding protein (REQ_20510) respectively, likely to be involved with the uptake of a ferri-siderophore complex across the cytoplasmic membrane, and is consistent with the predictive functional analysis performed earlier on iron transport methods in *R. equi*.

Additionally, IdeRB11 is another example of a canonical iron-box recognition site upstream of a gene predicted to express a putative glutamate dehydrogenase (REQ_30320), which is noteworthy as most bacteria require a large intracellular pool of glutamate to provide amine groups for amino acid synthesis (Miethke et al., 2006), and consequently, a range enzymes in amino acid biosynthesis pathways are iron-dependent therefore iron starvation may consequently instigate amino acid starvation.

The final canonical iron-box recognition site IdeRB13 is located upstream of an operon predicted to encode an ABC transporter system consisting of a substrate-binding protein, two transmembrane proteins and an ATPase (REQ_47000-47030). The localisation of the iron-box upstream of a complete ABC transport system suggests that this may be one of the established mechanisms of cellular importation of iron utilised by *R. equi*, however this would need to be verified experimentally.


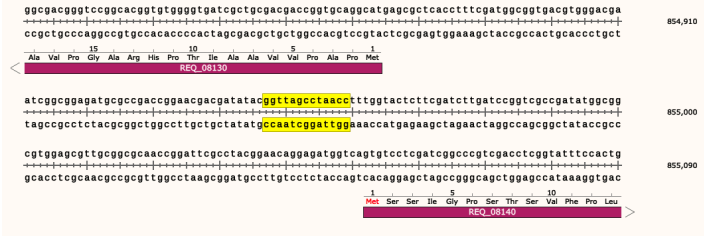

Unlike the aforementioned iron-boxes sequences, the *R. equi* chromosome also appears to contain a series of non-canonically located recognition sites for transcription repression. Considering the 9 strictly conserved bases of the recognition sequence (Figure 3-23) and assuming random sequence distributions, then one would expect to encounter the sequence every 4^9 nucleotides and thus find 19 instances on the chromosome.

IdeRB9 is a non-canonical iron-box recognition sequence located in the promoter for a heme oxygenase and ABC transport operon on the sense strand, however the operon ORF is divergently transcribed on the anti-sense DNA strand. Therefore, possible answers may include the theory mentioned about regarding repression of both DNA stands, or, potentially an alternative recognition

sequence may be utilised. This could be further explored experimentally, using a chIP on chip or chIP-seq analysis methodologies to explore the significance of transcription factor binding sites (Mundade et al., 2014; Pillai and Chellappan, 2015).

Table 3-16: Identification of putative iron boxes in the *R. equi* 103S chromosome with consideration of genetic context.

The chromosomal coordinates, the recognised sequence, genomic location and context, as well as associated gene ID and annotations are identified in respective columns.

IdeR Binding site	Gene ID	Coordinates	Recognition sequence 5' - 3'	Location	Gene annotation / function
ideR b1	REQ_07640	806,825 806,837	.. ggtagccttgc	Intergenic recognition site present upstream of gene	 <p>putative lysine/ornithine N-monoxygenase</p>
ideR b2/3	REQ_08130 REQ_08120 REQ_08110 REQ_08140	854,947 854,959	.. ggtagcctaacc	Palindromic intergenic recognition site present upstream of both genes	 <p>isochorismate synthase short chain dehydrogenase 2,3-dihydroxybenzoate-AMP ligase non-ribosomal peptide synthetase</p>
ideR b4/5	REQ_08150	862,836 862,848	.. ggtagcctaacc	Palindromic, intergenic recognition site present upstream of sense-strand encoding gene. Non-intergenic recognition site present on antisense strand.	 <p>substrate-binding lipoprotein</p>

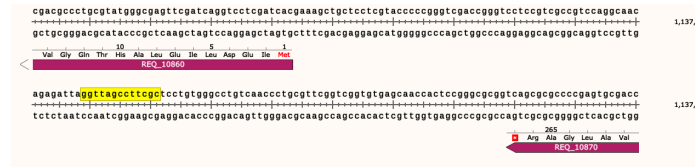
ideR b6

N/A

1,137,709
1,137,721

.. ggtagccttcgc

Non-intergenic-
recognition site
present present on
sense-strand



N/A

ideR b7

REQ_15910

1,662,459
1,662,447

.. ggtaggctcccc

Non-intergenic
recognition site
present, near end of
gene, prior to stop
codon. Also
upstream of
additional gene.



MFS transporter

REQ_15900

Hypothetical protein

ideR b8

REQ20510

2,162,443
2,162,431

.. ggtaggctcccc

intergenic
recognition site
present upstream of
encoding gene.



ABC transporter substrate-
binding protein

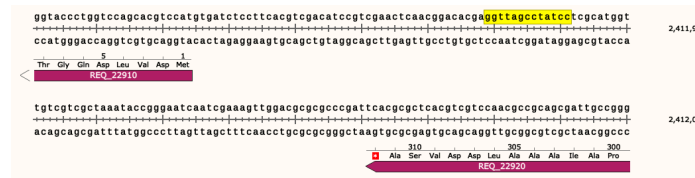
ideR b9

REQ_22910

2,411,889
2,411,901

.. ggtagcctatcc

Non-intergenic
recognition site
present on sense
strand, with both up
and downstream
genes read from
antisense strand.



heme oxygenase

REQ_22900

ABC transporter substrate-
binding protein

REQ_22890

ABC transporter
transmembrane protein

REQ_22880

ABC transporter ATPase

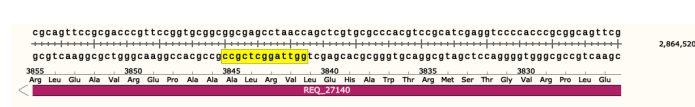
ideR b10

REQ_27140

2,864,472
2,864,460

.. ggtaggctcgcc

Non-intergenic
recognition site
present in the middle
of gene



non-ribosomal peptide
synthetase

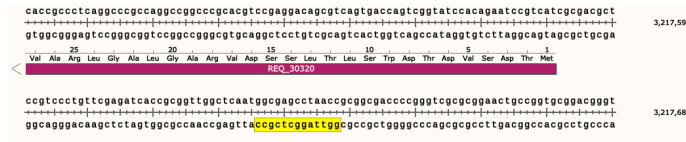
ideR b11

REQ_30320

3,217,638
3,217,626

.. ggtaggctcgcc

intergenic recognition site present upstream of gene on antisense strand.



putative dehydrogenase glutamate

ideR b12

REQ_33100

3,521,446
3,521,434

.. ggtaggctcccc

Non-intergenic recognition site present on antisense strand, with both up and downstream genes read from sense strand.



putative diacylglycerol kinase family protein

ideR b13

REQ_47000
REQ_47010
REQ_47020
REQ_47030

4,995,347
4,995,359

.. ggtaggctcacc

intergenic recognition site present upstream of gene on sense strand.



ABC transporter substrate-binding protein
ABC transporter transmembrane protein
ABC transporter transmembrane protein
ABC transporter ATPase

3.8 Summary of the bioinformatic investigation

While very little is known regarding *R. equi* iron regulation, previously literature highlights that these iron regulatory proteins are responsible for controlling siderophore biosynthesis and transport. The working hypothesis and currently known literature regarding iron homeostasis is summarised in Figure 3-25, detailing the core functions predicted to regulate intracellular iron levels within *R. equi*. Firstly, a cellular response to iron starvation; according to literature, *R. equi* is unusual in that it does not produce sufficient siderophore under iron limitation to be detected in a Chromazurol S assay, however it undoubtedly has the capacity to produce multiple siderophoric compounds. Secondly, *R. equi* is likely to have a mechanism responsible for prevention of oxidative stress, through a variety of control parameters likely to include intracellular iron storage, repression of iron uptake mechanisms and detoxification of reactive oxygen species. These themes are explored in detail throughout this thesis.

The bioinformatics analysis has established that the possession of two each of the Fur and IdeR/DtxR families of repressors is common in rhodococci and related genera. It is unlikely that all of these proteins co-repress their regulons with ferrous iron and interactions with other metals and discrete physiological roles have been described in other bacteria (see later in Chapter 4). Each of these four regulators was considered as a target for molecular cloning and expression techniques, in order to facilitate regulator-binding site analysis with pull-down assays or physio-chemical analysis such as ICP-OES analysis to confirm metalloprotein binding site specificities. Additionally, each of the regulator genes was targeted for unmarked in-frame gene deletion, in order to define their influence individually on rhodococcal physiology. Ultimately, transcriptomes could be defined relating to *R. equi* response to iron limitation and the definition of individual repressor regulons in the mutants.

Analysis of the biosynthetic gene clusters predicted to synthesise the siderophores, also aids experimental analysis, including use of the putative structures in a targeted metabolomic approach for siderophore identification and characterisation.

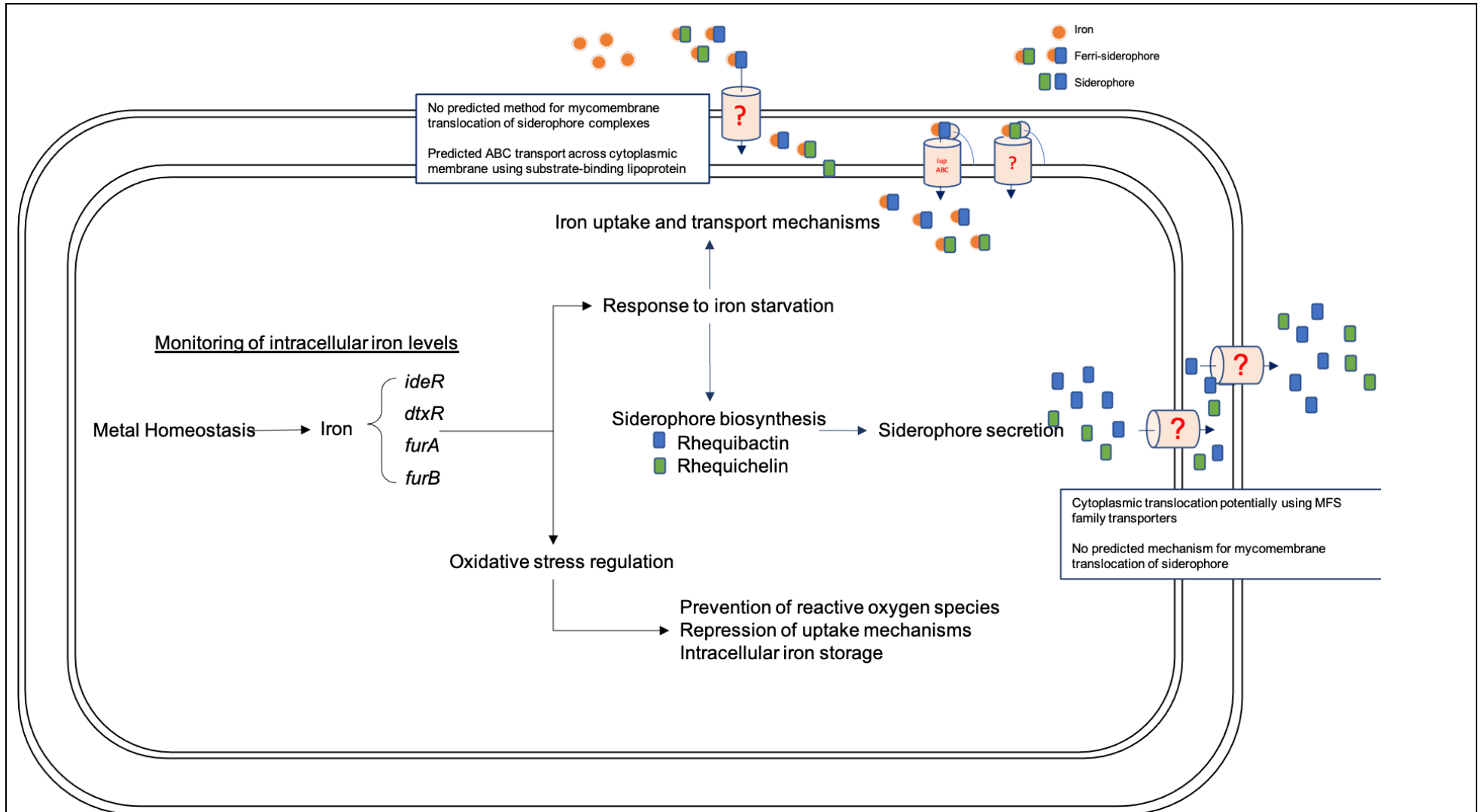


Figure 3-25: An *R. equi* iron homeostasis network map generated from the preliminary genomic survey and current literature.

4 Predicted Metal-binding capacities for the putative
metal-dependent regulators of *Rhodococcus equi*

103S

4.1 Introduction to metalloregulation

The first row of *d*-block ions, especially manganese, iron, cobalt, nickel, copper, and zinc are essential to cell function, whether it be for stabilising biomolecular structure, their role in metal-sensing signal transduction systems, or acting as cofactors for metalloenzymes. Despite their essential roles, if present to excess they can exert a toxic effect (Ma et al., 2009).

Bacteria have evolved transition metal homeostatic systems that can employ metalloregulatory proteins to effectively sense intracellular bioavailability and either scavenge the metals when they are in poor supply and may sequester or export them oversupplied (Higgins and Giedroc, 2013). Additionally, there are also metallochaperone components which manage their intracellular transit to the site of use (O'Halloran and Culotta, 2000).

Most bacteria utilise an array of DNA-binding transcriptional regulators to manage the homeostasis of transition metals. Typically, this type of regulator forms a coordination complex with a specific metal, to activate or inhibit binding to a DNA operator, limiting access to, and activity of RNA polymerase thereby regulating expression of downstream genes such as metal transporters, chaperones and metallo-storage proteins (Capdevila et al., 2017).

4.2 A brief review of iron-dependent regulatory protein families

4.2.1 Fur protein superfamily

The Fur protein was originally described in *E. coli* as an iron-sensing repressor responsible for regulation of gene expression relating to iron acquisition (siderophore) and transport methods (Stojiljkovic et al., 1994). Fur is frequently described as a global transcriptional regulator responsible for over 90 genes encoding both proteins and noncoding RNAs and involved in regulation of iron and prevention of oxidative stress, acid tolerance and production of toxins and other virulence factors (Litwin and Calderwood, 1993). Often described as a positive metal-dependent repressor, it represses transcription upon coordination with the metal co-repressor Fe (II) (although Fur is able to bind other related transition metals to a lesser effect *in vitro*) and is de-repressed in its absence by blocking of target gene transcription.

Typically, bacteria utilise members of the Fur superfamily in regulation of iron and zinc metabolism using transcriptional repression with Fur (ferric uptake regulator) or Zur (Zinc uptake regulator) proteins respectively (Fillat, 2014; Hantke, 2001). However, other members of the Fur superfamily have been functionally characterised, that are responsible for sensing and regulating homeostatic control in a diverse range of metals (Lee and Helmann, 2007) including Mur (manganese uptake regulator) (Diaz-Mireles et al., 2005), Nur (nickel uptake regulator) (An et al., 2009) as well as Irr (a haem-responsive regulator) (Kobayashi et al., 2016).

An additional member of the Fur superfamily – PerR, a peroxide stress response repressor is often found in Gram-positive organisms and due to the profound relationship between iron metabolism, free-radical production and the regulation of genes associated in response to oxidative stress it is quite common to find interactivity between PerR and Fur regulators, as well as extensive overlapping between both associated regulons (Fillat, 2014).

4.2.1.1 Understanding the features of the Fur family proteins

The canonical mechanism utilised for regulation by Fur proteins is typically by binding as a dimer to palindromic A/T-rich sequences found in the promoter region of the target gene using a metal co-factor for co-repression (Bagg and Neilands, 1987). This metal coordination facilitates a conformational change of the metalloregulatory protein to both stabilise and promote DNA interaction (Fillat, 2014).

The average length of a Fur superfamily protein is approximately 120 amino acids, and the common structural fold associated Fur proteins is comprised of an *N*-terminus winged-helix DNA-binding domain, and mechanistic regulatory located in the *C*-terminal region that are specific to metal regulator nature; while also facilitating dimerisation. The superfamily typically features a conserved region rich in histidine and cysteine residues represented by the motif **HHHXHX₂CX₂C** that is located at the beginning of the dimerisation domain, after a hinge region present between the DNA-binding domain and the *C*-terminus (Capdevila et al., 2017).

Often described as homodimeric DNA-binding proteins, members of the Fur superfamily typically comprise up to three metal binding sites; a regulatory metal binding site, a structural zinc binding site and occasionally a secondary metal site that utilises ligands derived from the *C*-terminal

regulatory domain and is interlinked with metal ligands that coordinate with the primary metal ion (Capdevila et al., 2017).

4.2.1.2 Fur protein superfamily: metal-coordination

The primary regulatory metal binding site of Fur family proteins is located at the start of the dimerisation domain. In all known cases the primary binding occurs in a high-affinity metal-sensing site located in the hinge region between the DNA-binding domain and the dimerisation domain. The metal coordination in this primary site is believed to engage amino acids from both domains of the regulator to stabilise a ‘closed calliper’ formation of the dimer that in turn can stabilise its interaction with the DNA operator (Capdevila et al., 2017). Further to this, the Fur superfamily of proteins are a great example of how transition metal selectivity can be regulated by discrete modifications of coordinating ligands to facilitate an allosteric regulatory mechanism for the desired primary metal ion (Capdevila et al., 2017; Ji et al., 2018). This allosteric mechanism employed by Fur/PerR proteins to bind Fe (II) or Mn (II) is facilitated in a penta-coordination using three Histidine and two aspartic acid/glutamic acid residues comparative to Zur proteins utilising a tetra-coordination, with two Histidine residues, one aspartic acid/glutamic acid residue, and one cysteine residue to bind Zn (II) (Fillat, 2014; Ji et al., 2018). The primary/regulatory metal coordination is likely required to stabilise the dimer at the hinge region and allow formation of a DNA-binding calliper form, increasing the affinity approximately 1000-fold thereby accounting for Fur activation as described by Deng et al. (2015).

The structural zinc (II) binding site mentioned previously, is highly conserved among the Fur superfamily, the coordination groups are consistently comprised of four cysteine residues (also known as a Cys₄ zinc finger) (Fillat, 2014), readily identifiable within amino acid sequences by a CXXC motif and is essential to promote stability within the C-terminal dimerization domain as zinc-binding motifs are stable configurations and seldom produce conformational changes upon binding (Tang et al., 2014). While the use of two CXXC motifs is widely accepted to be responsible for coordination of structural sites, there are a few exceptions to the rule, notably *S. coelicolor*'s Nur, which was found to lack this zinc-coordinated structural site, even under reducing conditions for crystal growth to prevent oxidation of the cysteine residues (An et al., 2009).

4.2.1.3 Understanding the activation of Fur: The Fur box

The iron (II)-Fur protein-complex model has been well established to regulate activation through recognition and binding to the upstream promoter region of Fur-repressed genes.

Fur targets a characteristic 19-bp AT-rich palindromic sequence, known as the ‘Fur box’. A 19-bp consensus sequence was derived from analysis of promoter regions of Fur-regulated genes (Escolar et al., 1999). Only one promoter matched the consensus sequence identically, with a match of 14-15 bp being common, with a minimum of 11 matched bases being required for recognition (Baichoo et al., 2002). A comparison of models was performed to explain the Fur box consensus sequence is given in Figure 4-1.

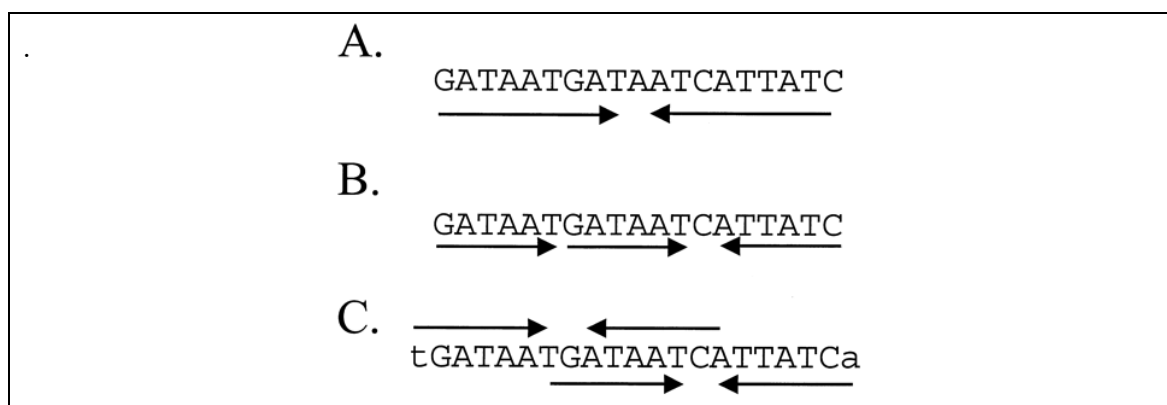


Figure 4-1: Model comparisons to explain the Fur box consensus sequence

(A) The classical 19-bp inverted repeat sequence, originally thought to bind a single Fur dimer. (B) An alternative model comprised of three GATAAT motifs in a head-to-head-to-tail (6-6-1-6) array, suggesting that Fur binds to a repeated array of the hexamer GATAAT. (C) Baichoo’s model suggesting that 19-bp Fur box results from two overlapping heptamer inverted repeats [(7-1-7)2] that together define a 21-bp sequence. (Baichoo et al., 2002).

Interestingly, Calderwood and Mekalanos (1988) identified that insertion of this Fur box recognition sequence into an operator site of a promoter upstream of a non-iron-regulated gene, inclusion led to transcriptional repression that was relieved under iron limitation.

4.2.2 DtxR protein family

The DtxR family of metalloregulatory proteins includes two major subfamilies that act as sensors of Fe(II) and Mn(II) and global regulators in Gram-positive organisms. The prototype DtxR is associated with the iron-dependent regulation of the diphtheria toxin in *Corynebacterium diphtheriae* (Tao et al., 1994). Due to their prominence as global pathogens and the association with iron

regulation of important virulence factors *C. diphtheriae* DtxR and a *Mycobacterium tuberculosis* homologue IdeR have been extensively studied; both are responsible for regulating gene expression associated with bacterial iron transport and storage (Rodriguez et al., 2002b; Tao et al., 1994).

4.2.2.1 DtxR protein family: domain architecture and metal coordination

Although the DtxR transcriptional response is highly specific for iron *in vivo* (Schmitt et al., 1995), *in vitro* analyses demonstrated that the metal coordination sites can often accommodate other divalent cations such as cobalt (II) or nickel (II), and subsequent DNA-binding can often occur (Qiu et al., 1995). Most of the resolved crystal structures for DtxR and associated homologues are characterised using an alternative cation such as cobalt or nickel. Surprisingly however, there have been no crystal structures resolved containing the cognate ferrous ion, this is likely due to the order relative stabilities of complexes formed by transition metal known as the Irving-Williams series, in which the stability of complexes formed by divalent first-row transition metal ions typically increase following the order of $\text{Mn(II)} < \text{Fe(II)} < \text{Co(II)} < \text{Ni(II)} < \text{Cu(II)} > \text{Zn(II)}$ (Irving and Williams, 1953).

The general structure of the DtxR-family of proteins consist of firstly, an N-terminal DtxR-type helix-turn-helix domain, that is comprised of approximately 65 amino acids conserved within the DtxR-like protein family. This winged helix motif comprises two wings, three α -helices and three β -sheets. Functionally this serves, to facilitate DNA-binding as the DNA-recognition helix makes sequence-specific DNA interactions with the major groove of DNA, while the wings make complimentary DNA contacts, often within the minor grooves or with the DNA backbone (Mitchell et al., 2019). In addition to this the DtxR-family also contains a dimerization domain in the C-terminal region that includes two metal-coordination sites, and an SH3-like domain extension (Feese et al., 2001) thought to enrich the DNA-binding affinity by stabilisation of protein-protein interactions (Mitchell et al., 2019).

The two metal binding sites described for the DtxR-family proteins have previously been characterised in *Corynebacterium diphtheriae* DtxR and *M. tuberculosis* IdeR, and can be differentiated by their function, firstly the primary regulatory site and a structural ancillary site.

The ancillary site has previously been observed to be occupied in all high-resolution crystal structures of DtxR determined in the presence of divalent metal ions, and was found to be tetrahedrally coordinated by side chain ligands of His79, Glu83, His98, and a sulphate ion with equivalent ligand groups identified for DtxR homologues, suggesting a conservation within the superfamily (Pohl et al., 1999). Coordination within this ancillary binding site is hypothesised to increase allosteric activation of the primary metal sensing site (Capdevila et al., 2017).

In wild type DtxR, the primary metal-coordinating site has only been observed to be partially occupied in Cd-DtxR and Mn-DtxR structures. In both of those cases, the metal appears to be coordinated by Glu105, His106 side chains, the main chain carbonyl oxygen of Cys102, and a water molecule (Pohl et al., 1998). Site-directed mutagenesis studies were performed based upon reports that lack of metal coordination may be due to the oxidation or other modification of Cys102 (Qiu et al., 1995). Structures of mutant C102D DtxR proteins were resolved in holo-form in the presence of NiCl₂ from which six metal-coordinating ligands included Met10, Asp102, Glu105, His106, the main chain of Asp102 and a water molecule, form an octahedral geometry.

Interestingly these coordinating ligands are fully conserved in the DtxR homologue IdeR from *M. tuberculosis*, from which a crystal structure (1FX7) was resolved with a primary Co(II) cofactor under pentavalent coordination with distorted trigonal bipyramidal geometry (Feese et al., 2001), this was likely due to the , the ionic radii of Co (II) and Fe (II) are distinctly analogous and although the native ligand for DtxR-like proteins is Fe(II), other divalent cations including nickel, cobalt, manganese, and zinc have also been shown to bind to and activate the regulators *in vitro*.

However, some DtxR-like proteins have been identified to control de-repression of two metal ion transport systems in response to manganese(II) but not iron(II) and as such coordinate the metal ligand using a binuclear geometry (Cong et al., 2018; Que and Helmann, 2000).

4.2.3 An overview of homologous metal regulatory systems

4.2.3.1 Divalent metal homeostasis in *Bacillus subtilis*

A review by Helmann (2014) discusses the bacterial response to metal ion sufficiency, in that *B. subtilis* utilises MntR to regulate manganese homeostasis, as well as two Fur-family paralogs Fur

and Zur to principally regulate intracellular iron and zinc levels respectively, while an additional Fur-family paralog PerR, is responsible for regulation of peroxide response genes .

Functioning as a metal-dependent peroxide sensor by transcriptional repression in *B. subtilis*, PerR has the ability to coordinate with Mn(II) or Fe(II) to facilitate DNA-binding both *in vitro* and *in vivo*.

The protein PerR in *B. subtilis* is an archetypal example of the family of metal-containing peroxide sensors that are found in an extensive range of bacteria. (Ma, 2013)(Ma, 2013)The global regulon of PerR is diverse in nature, containing itself, a catalase gene *KatA*, an alkyl hydroperoxidase *AhpCF*, haem biosynthetic genes, as well as the iron-storage protein *MrgA*, a Zn(II) uptake transport gene *ZosA*, and the global iron homeostasis regulator *Fur*. This diverse nature of PerR regulated genes, advocates a deep-rooted relationship between management of the cellular peroxide stress response and the cellular homeostasis of divalent metal cations.

The transcriptional response to changes in iron availability by *B. subtilis*, is coordinated using Fur to sense intracellular iron availability. The *B. subtilis* Fur regulon has been well characterised over the years and includes over 50 genes responsible for iron uptake systems, siderophore biosynthesis and transport mechanisms as well as xenosiderophore equivalents (Pi and Helmann, 2017). Conventionally, Fe(II)-coordinated Fur binds the Fur-box DNA region with high affinity and represses the transcription of these genes to maintain adequate iron homeostasis, to prevent cellular toxicity or oxidative effects.

4.2.3.2 Divalent metal homeostasis in *M. tuberculosis*

A review by Zondervan discussed a trio of virulence strategies employed by *M. tuberculosis* including: immune modulation, dormancy and phagosomal rupture, specifically discussing the role of divalent metals in the associated regulatory mechanisms (Zondervan et al., 2018). The review highlights the use of the divalent metals iron, zinc and manganese, and how they are regulated by IdeR (Pandey and Rodriguez, 2014) , Zur (Maciag et al., 2007) and MntR (Papp-Wallace and Maguire, 2006) respectively. The binding of co-factor iron (II) into the regulatory metal site of IdeR and zinc (II) into the structural metal site, stabilises dimer formation increasing binding capacity to sites involved in gene suppression.

4.3 Putative gene selections

At the outset, the aims relating to the Fur and DtxR superfamily members of *R. equi* 103S identified in Chapter 3 were to characterise these proteins by a combination of molecular cloning to support *in vitro* analysis of metal and DNA binding and unmarked in-frame deletion mutagenesis to establish their regulons.

4.4 Strategy 1 – Molecular cloning, overexpression and *in vitro* characterisation of metal regulatory proteins

4.4.1 Amplification of metalloregulatory genes

The nucleotide sequence data for the open reading frames (ORF) that encode the metal regulatory genes of interest (*ideR*, *dtxR*, *furA* and *furB*) was retrieved from the *Rhodococcus equi* 103S entry (accession: FN563149.1) in the National Centre for Biotechnology Information (NCBI) nucleotide database (<https://www.ncbi.nlm.nih.gov>). The aim of the molecular cloning was to produce native proteins in a heterologous host (*E. coli*) and to attempt to purify these based using Heparin agarose; a common approach taken for DNA-binding proteins. The use of artificial tags to promote purification was avoided in case such modifications influenced protein function; the use of a hexahistidine-tags was particularly avoided due to the likelihood of their interference with downstream metal binding analyses.

The oligonucleotide primers (section 2.4.2) were designed in pairs to amplify the gene region of interest, the length of the primers was kept between 32 and 36 base pairs, with differences in primer melting temperatures being limited to a maximum 5 °C difference for each pair. Restriction endonuclease recognition sites were introduced to facilitate ligation with the vector of choice. The regions of interest were amplified by polymerase chain reaction (PCR) using Q5 polymerase (New England Biolabs) using *Rhodococcus equi* 103S genomic DNA as a template (as described in Section 2.3.3). The PCR procedure was performed without an annealing step given that the primer melting temperatures exceeded the PCR extension step temperature of 72 °C. The gene of interest and amplicons sizes produced via PCR are given in Table 4-1.

Table 4-1: Metal-dependent regulatory genes of interest

Gene of interest (locus ID)	Length (base pairs)
<i>ideR</i> (REQ_20130)	687
<i>dtxR</i> (REQ_19260)	717
<i>furA</i> (REQ_04740)	555
<i>furB</i> (REQ_29120)	447

The reaction products were analysed using agarose gel electrophoresis (Figure 4-2); and each appeared to produce single DNA bands consistent with the expected amplicon size (Table 4-1). However, due to the presence of potential primer dimers formed, the bands of interest were excised from the gel and the DNA purified as previously described in section 2.3.6.

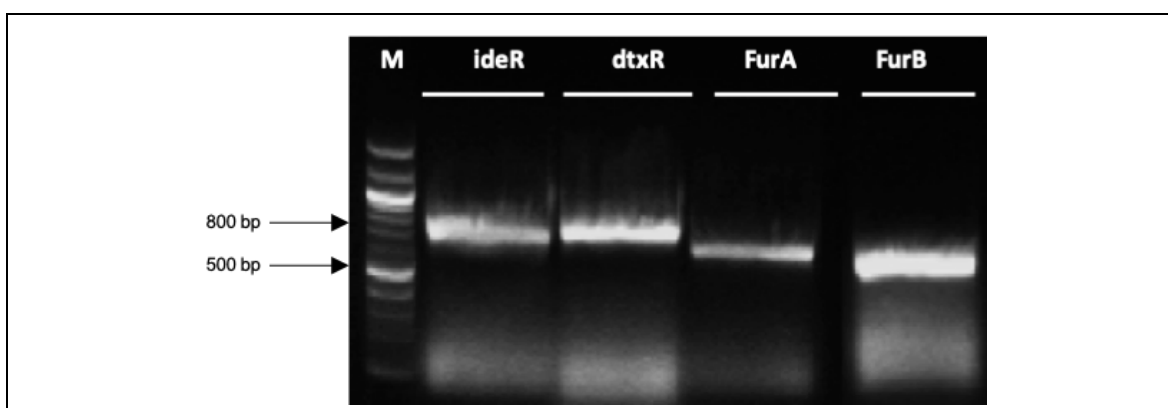


Figure 4-2: PCR amplification of four putative *R. equi* iron regulatory genes. M = NEB 100 bp ladder. Analysed on a 1 % agarose electrophoresis gel with Sybr safe DNA stain and imaged using Quantity One (Bio-Rad) imaging software.

4.4.2 Molecular cloning and sequencing analysis

The purified DNA amplicons and the vector pET23b(+) were prepared for cloning by digestion using *NdeI/NotI* restriction endonucleases and subsequent ligation as previously described in section 2.7.2.

The cloning strategy is depicted in Figure 4-3 and available in Chapter 4 Appendices.

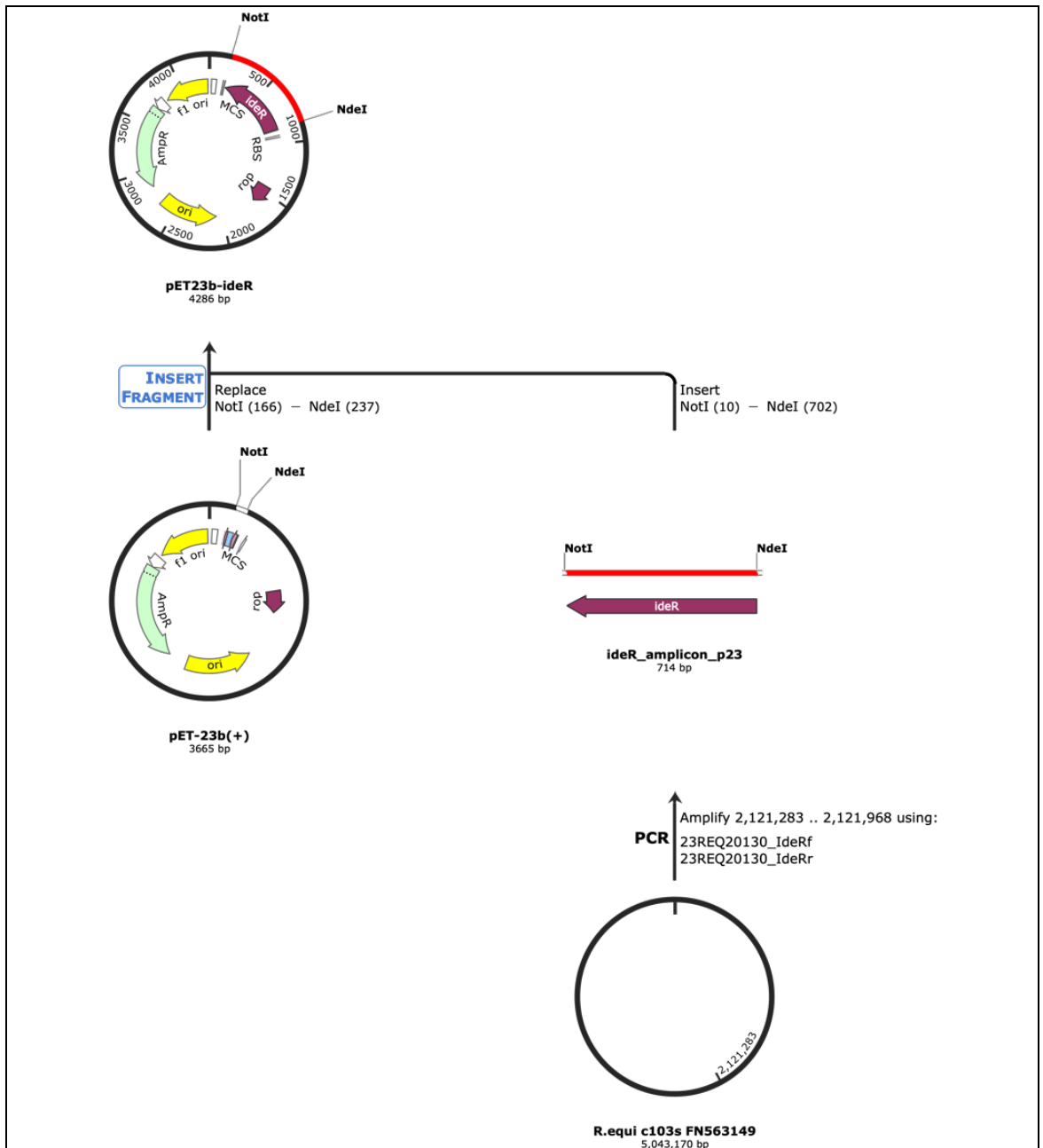


Figure 4-3: Strategy for cloning of *R. equi* metal regulators in pET23b.

The same approach was taken for each gene and here the *ideR* amplicon is used as an illustrative example. Each amplicon was generated by PCR using primers listed in Table 2-7 and *R. equi* 103S genomic DNA as a template. Amplicons and plasmid DNA were prepared for cloning by digestion with restriction endonucleases *NdeI* and *NotI* before ligation with T4 DNA ligase.

Recombinant plasmids were selected by transformation of competent *E. coli* XL-10 Gold cells, where transformants acquired ampicillin resistance; transformant clones were examined using colony PCR to detect incorporation of the PCR amplicon into the vector.

Positive colony PCR results detecting amplicon incorporation into the vector pET23b was initially achieved for genes *furA* and *furB* (Figure 4-4) and subsequently both *ideR* (Figure 4-5) and *dtxR* (Figure 4-6). These samples were sent to GATC biotech for Sanger sequencing confirmation using the T7-minus1 sequencing primer, which returned at least one plasmid isolate of each type without nucleotide insertion or mutation.

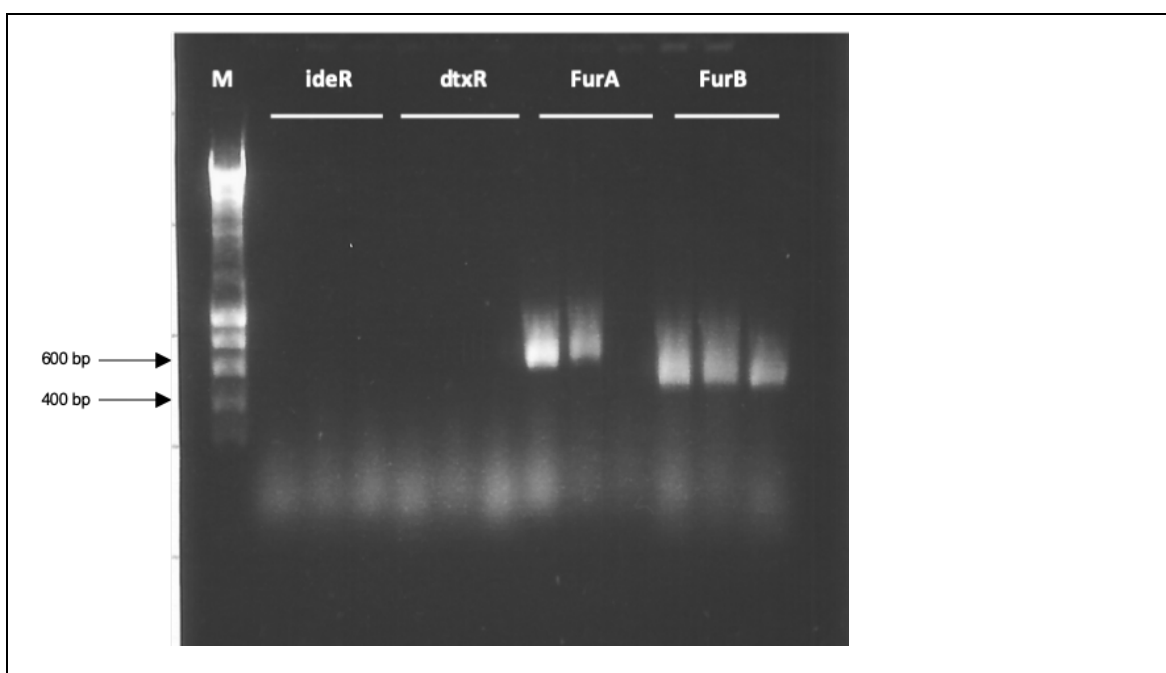


Figure 4-4: Colony PCR confirming amplicon incorporation for *furA* and *furB*. Lane M = Hyperladder 1kb plus molecular marker. Analysed by 1 % agarose gel electrophoresis, DNA stained using SYBR Safe; imaged using Quantity One (Bio-Rad) imaging software.

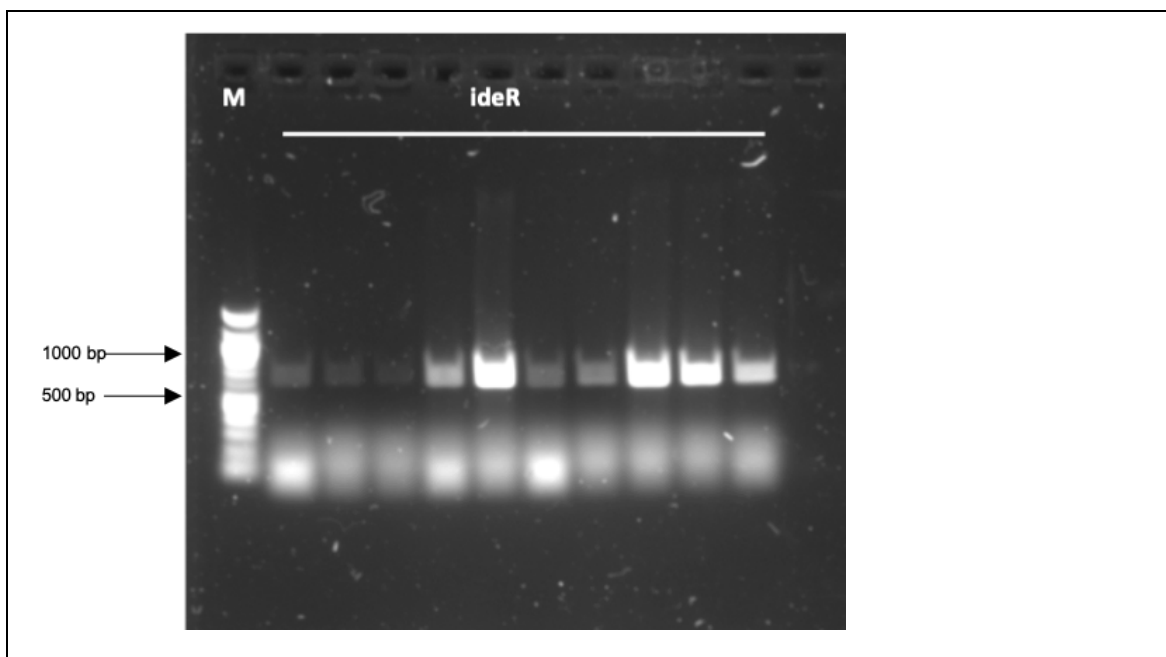


Figure 4-5: Colony PCR confirming amplicon incorporation for *ideR*.

Lane M = Hyperladder 1kb plus molecular marker. Analysed by 1 % agarose gel electrophoresis, DNA stained using SYBR Safe; imaged using Quantity One (Bio-Rad) imaging software.

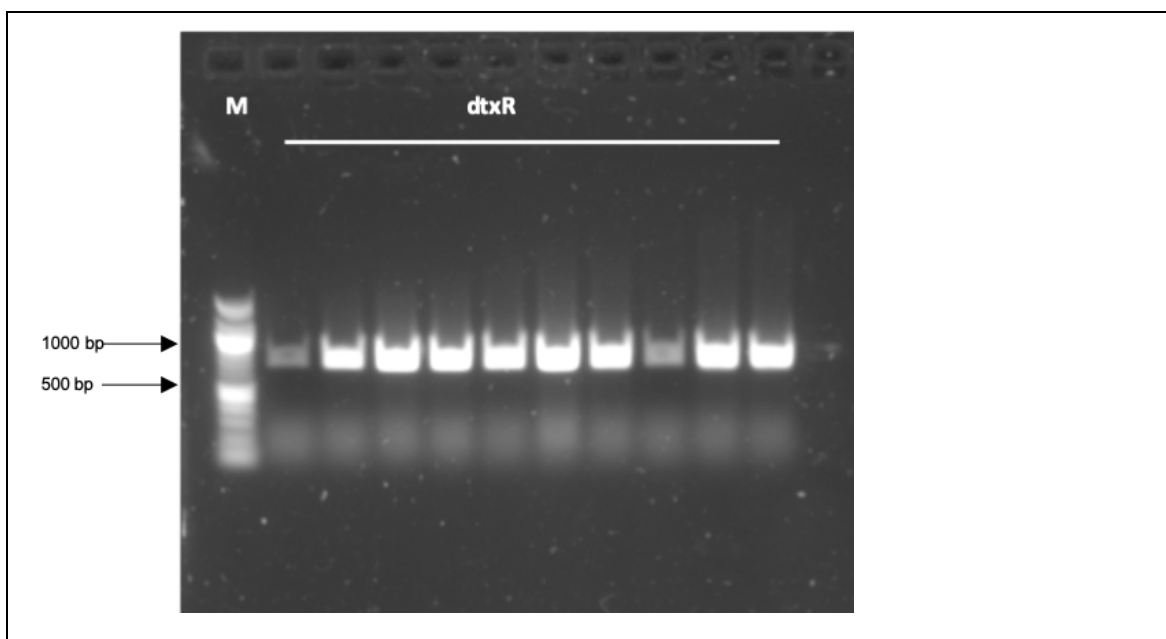


Figure 4-6 Colony PCR confirming amplicon incorporation for *dtxR*.

Lane M= Hyperladder 1kb plus molecular marker. Analysed by 1 % agarose gel electrophoresis, DNA stained using SYBR Safe; imaged using Quantity One (Bio-Rad) imaging software.

4.4.3 Recombinant gene expression and protein production

To facilitate gene expression, *E. coli* BL21 (DE3) were transformed using the recombinant plasmids. Both recombinant Fur-family plasmids (pET23b-*furA*; pET23b-*furB*) were generated transformants. Interestingly however, transformation of this strain with the DtxR-family constructs was unsuccessful on multiple occasions. This failure to generate transformants indicated a toxicity

related issue, therefore both DtxR-family recombinant plasmids (pET23b-ideR; pET23b-dtxR) were used to transform Overexpress C41 (DE3), a strain that was phenotypically selected for conferring tolerance to toxic proteins. Using this approach transformants were successfully isolated.

Induction of recombinant gene expression using isopropyl β -D-1-thiogalactopyranoside (IPTG) was attempted. Optimal production of soluble recombinant protein was sought using by addition of 1mM IPTG at mid-log cellular growth, with subsequent incubation at varying temperatures as previously described (section 2.9.1.1).

Analysis of clarified lysates using SDS-PAGE, identified heterologous protein overproduction for IdeR based upon an estimated protein size of 25.5 kDa (Figure 4-7), the overproduction of protein was considerably larger in the IPTG-induced cultures comparative to the non-induced culture, for which some expression of the protein occurs likely due to leaky expression.

The other regulatory genes of interest *dtxR*, *furA* or *furB*, were not successful in heterologous protein production, under any IPTG induction conditions. Lysates were also denatured using 8 M Urea sample buffer without detection of the desired heterologous protein. In order to make timely progress on the project, the successfully overproduced soluble IdeR protein was trialled in the downstream processing before further optimisation of the other products.

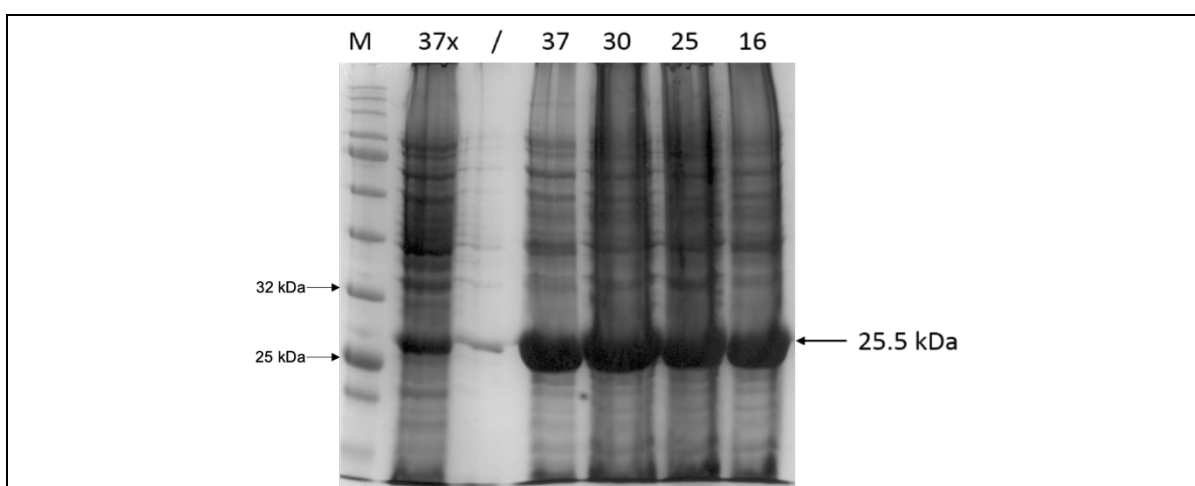


Figure 4-7: 12 % SDS-PAGE gel for the expression study of pET23b-IdeR.

The recombinant cells were grown at 37°C until mid-log phase. Once the desired optical density was achieved, expression was induced by addition of IPTG to a final concentration of 1mM. Expression was monitored at 4 different temperatures (16 [overnight], 25, 30 and 37 °C [4 hours]). (37x): - A control performed at 37 °C without induction.

The recombinant protein production for IdeR was scaled to a culture of 1 litre using the 1mM IPTG; 37 °C induction for four hours. Cellular material was harvested; lysed in 0.05 M Tris-HCl; 0.5 M NaCl, pH 7.4 and clarified by centrifugation as described in (section 2.9.1.3).

4.4.4 Affinity chromatography using HiTrap Heparin HP

The clarified lysate was subjected to affinity chromatography as the first step in the purification of the recombinant proteins. A HiTrap Heparin HP column prepacked with High Performance Heparin Sepharose was fixed to an automated AKTA purifier Fast Protein Liquid Chromatography (FPLC) system (Amersham Pharmacia) to identify DNA-binding proteins including the recombinant protein IdeR. After the clarified lysate loading phase and extensive column washing, the chromatogram an increase in eluate UV absorbance indicated an elution of protein on initiation of a linear gradient of NaCl concentration. After several fractions, the UV absorbance returned to baseline and no further elution of protein was indicated.(Figure 4-8).

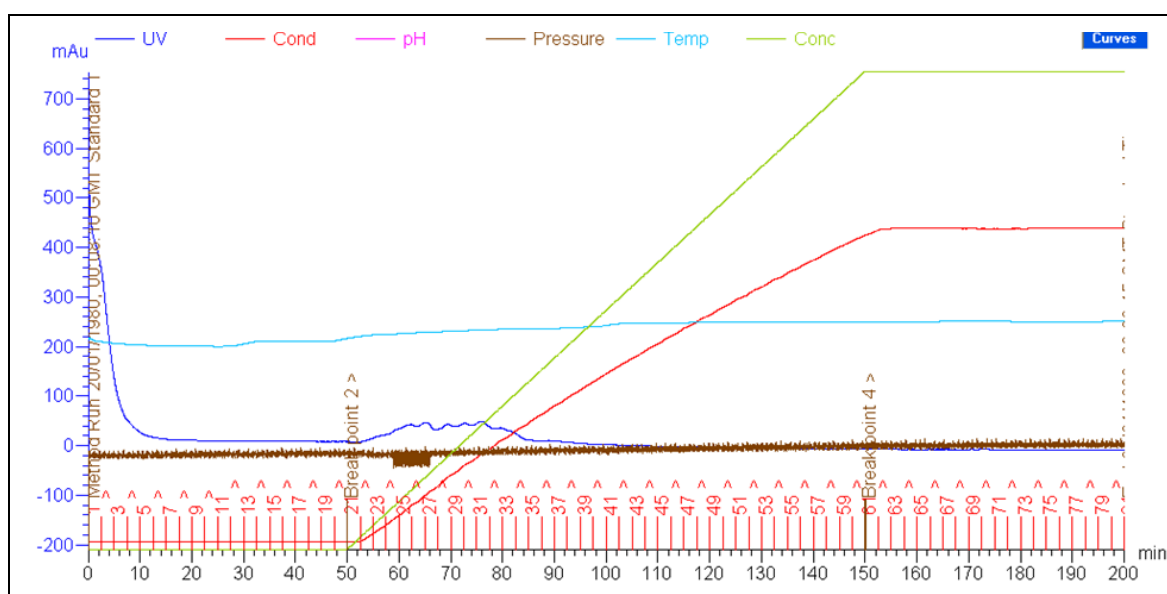


Figure 4-8: Attempted purification of *Rhodococcus equi* IdeR on Heparin-agarose.

After loading of the lysate and extensive washing, elution of any bound IdeR was attempted using a linear gradient of NaCl (0 - 2 M) in 10 mM Sodium phosphate over 100 ml at a flow rate of 2 mL/min. Eluate was collected in 5 mL fractions. The composition of each fraction was analysed on SDS-PAGE to identify those containing protein of interest.

Analysis of the fractions with increased UV absorbance was performed by SDS-PAGE; this analysis provided no evidence for the retention of IdeR on the heparin-agarose column.

It was considered that metal ions represented in the lysate might influence protein behaviour and another lysate was treated with Chelex-100 resin before chromatography on heparin-agarose; the

resulting chromatogram resembled that one seen previously. Both the non-adsorbed flow-through fractions and early eluting protein fractions were analysed via SDS-PAGE (Figure 4-9 and Figure 4-10 respectively).

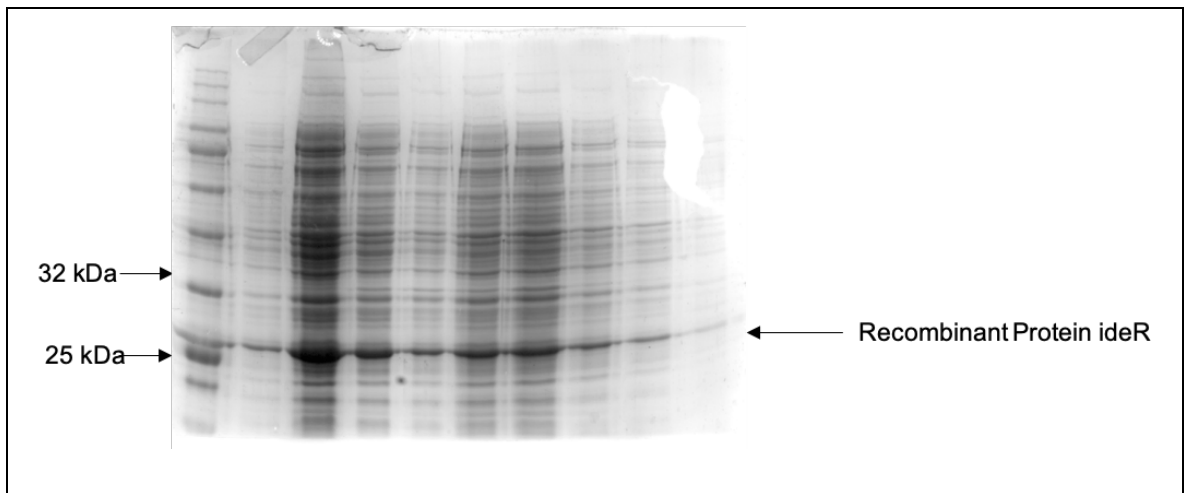


Figure 4-9: 12 % SDS-PAGE analysis of chelex-treated IdeR clarified lysate flow-through

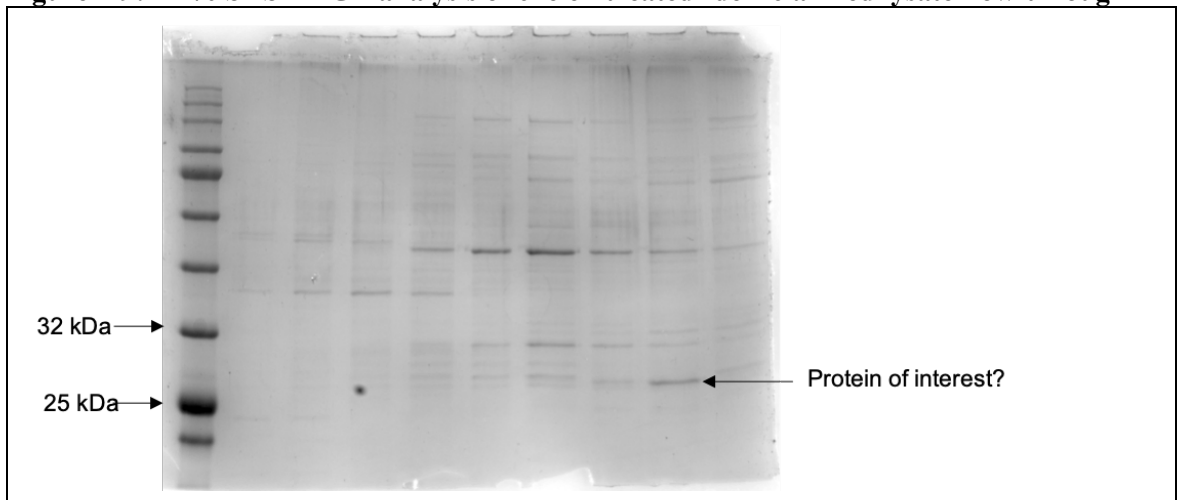


Figure 4-10 :2 % SDS-PAGE analysis of chelex-treated IdeR clarified lysate fractions 30-38.

Analysis of the flow-through fractions indicated that significant quantities of a major protein consistent with the expected mobility of IdeR had not bound to the column.

Analysis of fractions 30-38 yielded a selection of faint protein bands, one in particular of interest in fraction 37 that correlated with the appropriate band size of the recombinant protein, however given the low yield, an alternative approach to protein purification was sought.

4.5 Strategy 2 - Fusion of metal regulatory proteins with Glutathione-S transferase

Due to the unsuccessful attempts at protein purification using Heparin columns, it was hypothesised that the regulators might be produced in higher yield and could be readily purified by producing them as Glutathione-S transferase (GST) fusion proteins. The GST fusion partner facilitates protein recovery, on-column proteolytic cleavage, allowing elution of purified protein, without likely interference with the metal-binding activities of the protein of interest.

4.5.1 Generation of GST fusion constructs

GST has a history of successful use as a fusion partner however, it is possible and even likely that fusion of the metal-dependent regulators with the enzyme would generate a steric hindrance that interfered with DNA binding activity. The cloning strategy used allowed for the proteolytic cleavage of the fusion partner using the specific protease Factor Xa. Oligonucleotide primers were designed (section 2.4.5) to incorporate a Factor Xa recognition sequence adjacent to the genes of interest, and to facilitate their subsequent assembly with a modified pGEX- Δ 6P-1 vector, to form the pGEXJT-fXa-[gene] plasmid series (Figure 4-11).

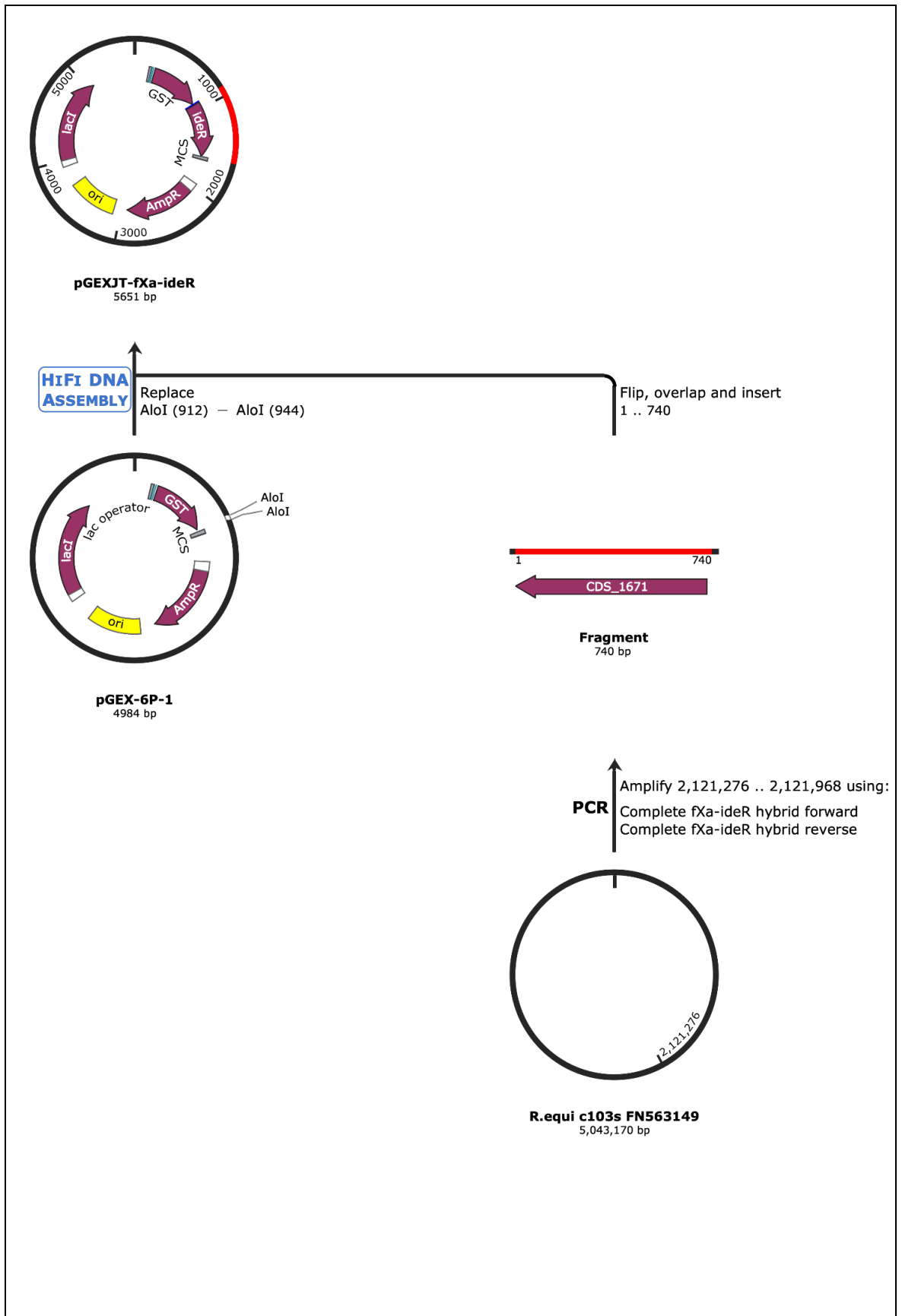


Figure 4-11: Experiment methodology for producing pGEXJT-fXa hybrid plasmids.
 The example above uses the gene *ideR* isolated from *R. equi* 103S chromosomal DNA assembled with linearised pGEX-6P-1 using NEB HIFI assembly kit

The regions of interest were amplified by polymerase chain reaction (PCR) using Q5 polymerase (New England Biolabs) using *Rhodococcus equi* 103S genomic DNA as a template using a analogous methodology to section 4.4.1 performed earlier. Amplicon lengths are given in Table 4-2.

Table 4-2: Metalloregulatory genes of interest with given gene and fXa-amplicon lengths

Gene of interest (locus ID)	Gene Length (base pairs)	fXa-Amplicon
ideR (REQ_20130)	686	740
dtxR (REQ_19260)	716	770
FurA (REQ_04740)	555	608
FurB (REQ_29120)	446	500

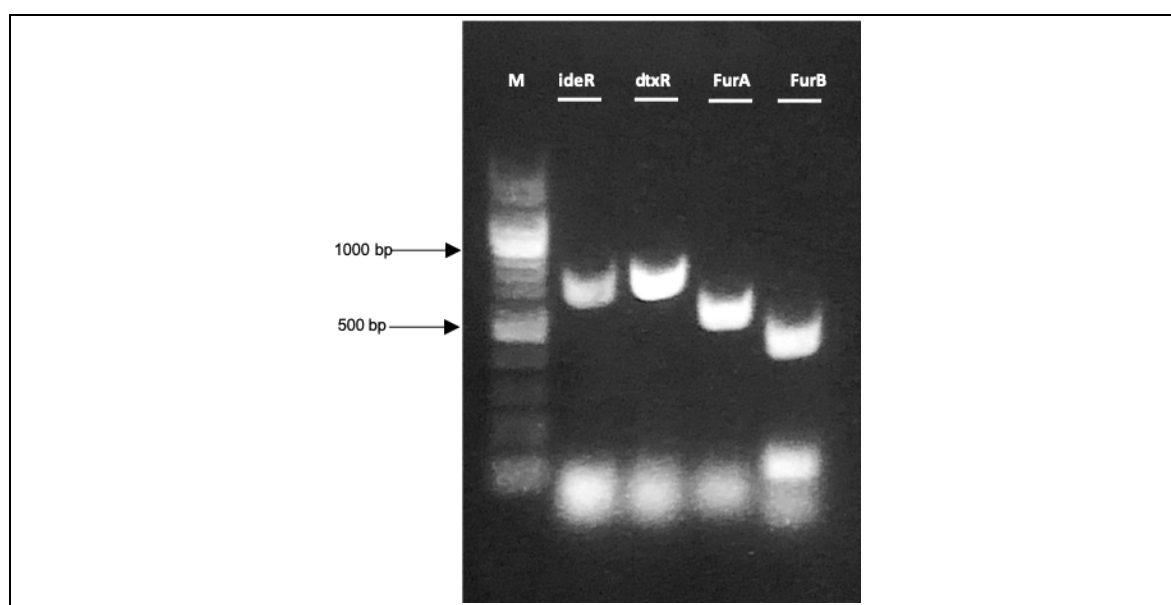


Figure 4-12: PCR amplification for four putative iron regulatory genes with incorporation of a Factor Xa region.

M = Hyperladder 1kb (Bioline). Analysed on a 1 % agarose electrophoresis gel, imaged using Quantity One (Bio-Rad) imaging software.

4.5.2 Vector preparation for GST-tagged protein purification

The pGEX-6P-1 vector stocks were acquired from Professor Black's research group at Northumbria University. The vector pGEX-6P-1 encodes the recognition sequence for site-specific cleavage by PreScission Protease between the GST domain and the MCS (Figure 4-13).

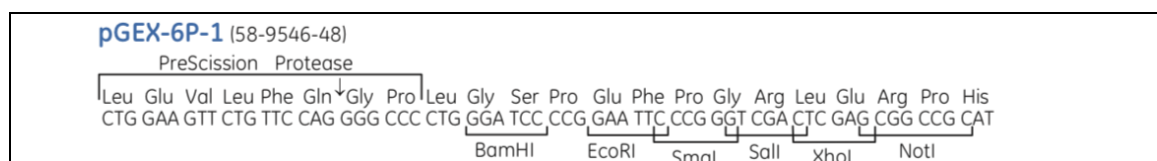


Figure 4-13: GST vector pGEX-6P-1 showing the correct reading frame, and cutting region of PreScission Protease denoted by the arrow.

While pGEX-6P-1 offered the most efficient method of cleavage and purification of GST-tagged

proteins, the proteolytic processing of the fusion protein would leave two non-native amino acid residues (Gly-Pro), which may affect the downstream DNA binding capabilities of the metal-dependent regulatory proteins.

4.5.2.1 Modification of pGEX-6P1

In order to preserve the functionality of the C-terminal DNA binding domain of the regulatory proteins, an alternative protease was selected. Factor Xa protease facilitates cleavage directly after the recognition sequence.

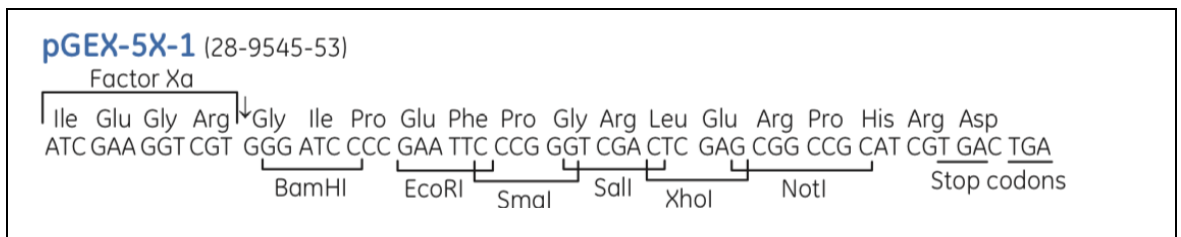


Figure 4-14: GST vector pGEX-5X-1 showing the correct reading frame and cutting region of Factor Xa Protease denoted by the arrow.

Modification of pGEX-6P-1 was selected over purchase of pGEX-5X-1, as the protease replacement could be performed concurrently with regulatory gene insertion. The removal of the PreScission protease site (Figure 4-15) was performed by digestion using the restriction enzyme *AloI* recognising $\wedge(7/12-13)GAAC(N)_6TCC(12-13/7)\wedge$ sites.

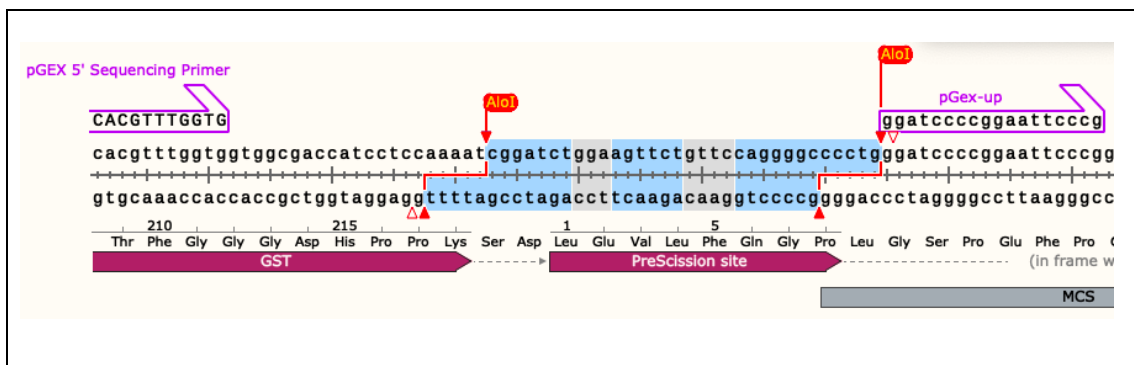


Figure 4-15: Elimination of the PreScission protease cut site using the restriction enzyme *AloI* in the vector pGEX-6P-1.

Unfortunately the enzyme *AloI* appeared to inefficiently cut the vector pGEX-6P-1 producing multiple DNA bands (Figure 4-16), therefore a linearised band of cut vector consistent with the desired size was purified via electrophoresis and subsequent extraction from excised agarose.

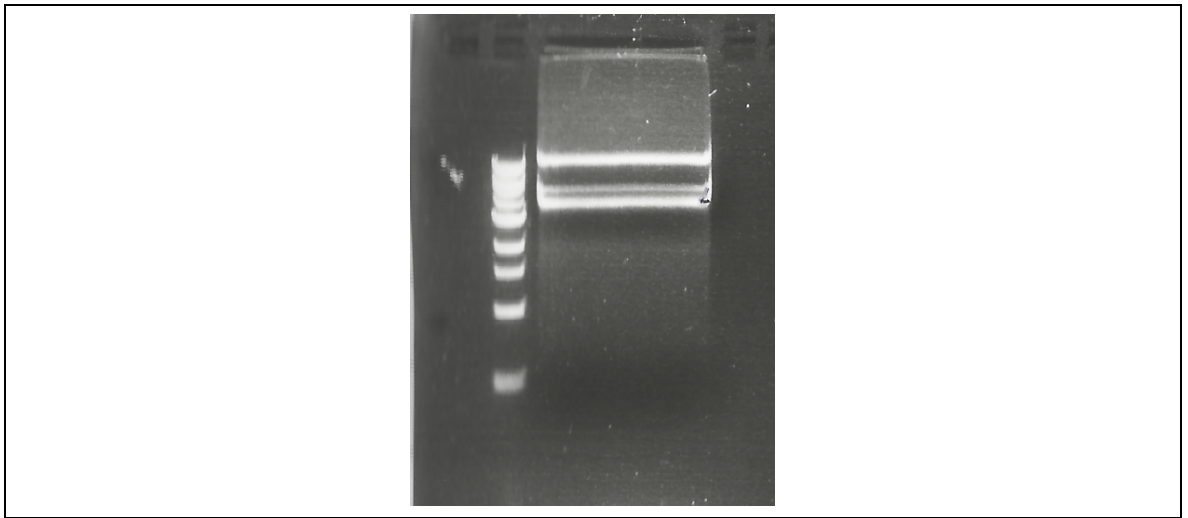


Figure 4-16: 1 % agarose gel electrophoretic analysis of the plasmid pGEX-6P-1 when cut with the restriction enzyme *AloI*.

M = Hyperladder 1kb molecular marker

4.5.2.2 Molecular cloning – NEBuilder HiFi Assembly and sequencing analysis

The linearised vector and fXa-gene PCR amplicons were assembled using NEBuilder HiFi assembly kit as described in (section 2.7.3), Transformants of *E. cloni*[®]10G (Lucigen) were plated on ampicillin-containing agar plates. Single colonies were selected, and liquid cultures grown overnight to extract and purify plasmid DNA for nucleotide sequence verification.

The sequencing results identified that the assembly process was unsuccessful, or rather the enzyme *AloI* was inefficient at digestion of the vector as the sequencing data identified an intact PreScission protease recognition site (Figure 4-17).

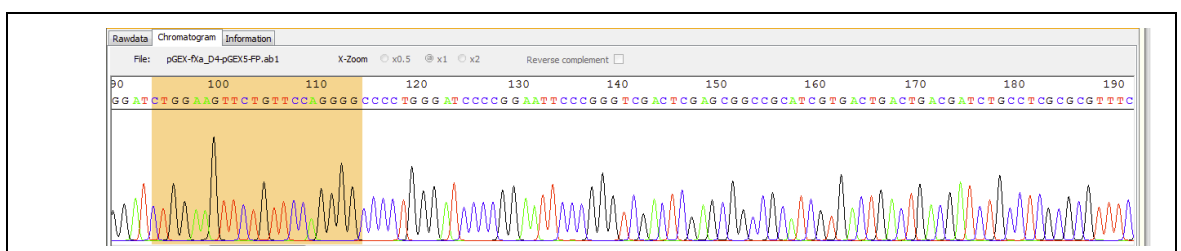


Figure 4-17: Sequencing results with preScission site identified

This inefficiency could potentially be the result of *AloI* remaining bound to the cleaved DNA altering the electrophoretic mobility, effecting the attainment of the appropriate DNA during excision.

4.5.2.3 Alternative vector preparation – pGEXPCR

Due to the previous unsuccessful vector preparations using the restriction enzyme *AloI*, an alternative strategy was implemented, to produce a compatible vector preparation *via* polymerase chain reaction using Q5 DNA polymerase and the plasmid pGEX-6P-1 as DNA template (diluted 1:10 with distilled

H₂O) (Figure 4-18a). The thermocycling conditions were based on a standard 3-step and 30 cycle protocol; annealing at 70 °C for 30 seconds. The PCR amplicon size of 4947 base pairs was analysed by agarose gel electrophoresis (Figure 4-18b) against a molecular marker standard for indication of product size, before processing with a PCR purification kit and subsequent quantification of 158.0 ng/μl and purity analysis of 1.87 (260/280 nm absorbance) using a Nanodrop1000 spectrophotometer.

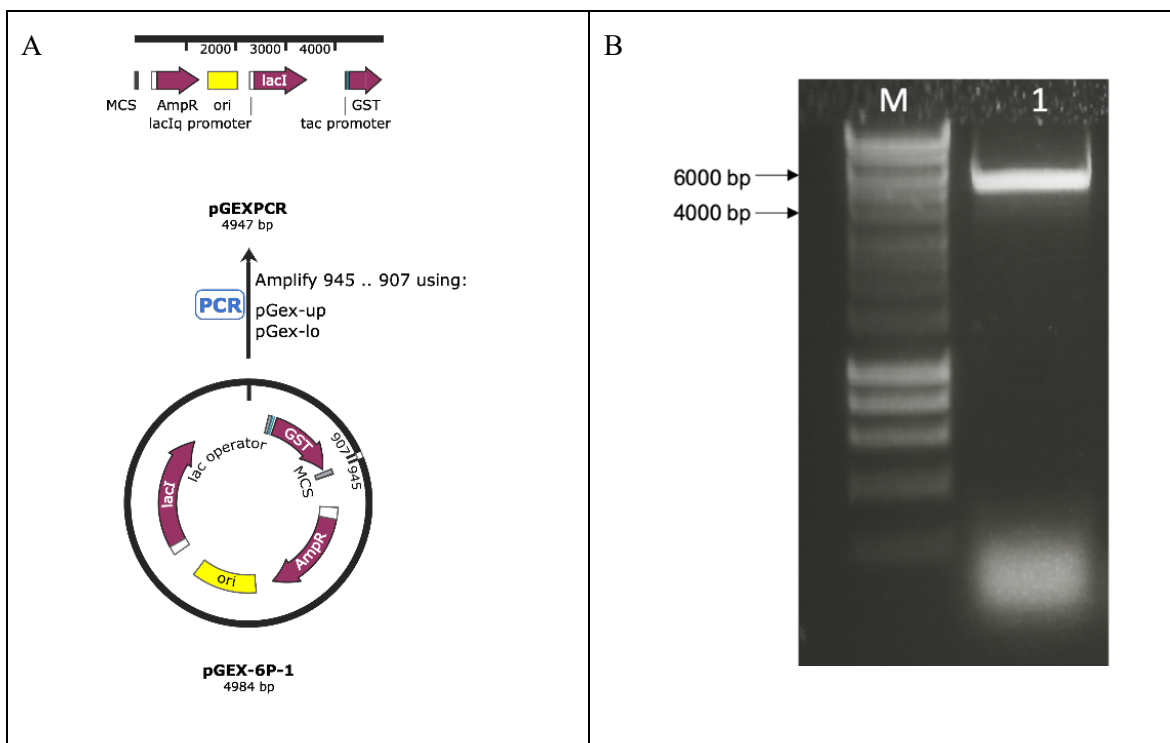


Figure 4-18 Production of the amplicon pGEXPCR via PCR

A) Experimental methodology highlighting the region for amplification. B) 1) 1% agarose gel electrophoretic analysis of the amplicon pGEXPCR. M – Hyperladder 1Kb molecular marker.

4.5.3 Assembly of pGEXPCR with fXa hybrid amplicons

The newly produced pGEXPCR vector and fXa-gene PCR amplicons were assembled using NEBuilder HiFi assembly kit as previously described (experimental strategy shown in Figure 4-19 and in Chapter 4 Appendices). Transformant clones of *E. coli*[®] 10G cells (Lucigen) were plated on ampicillin-containing agar plates. Single colonies were selected, and liquid cultures grown overnight to extract and purify plasmid DNA for nucleotide sequence verification. Vector insertion regions were sequenced to confirm no mutations had occurred during the amplification and assembly stages.

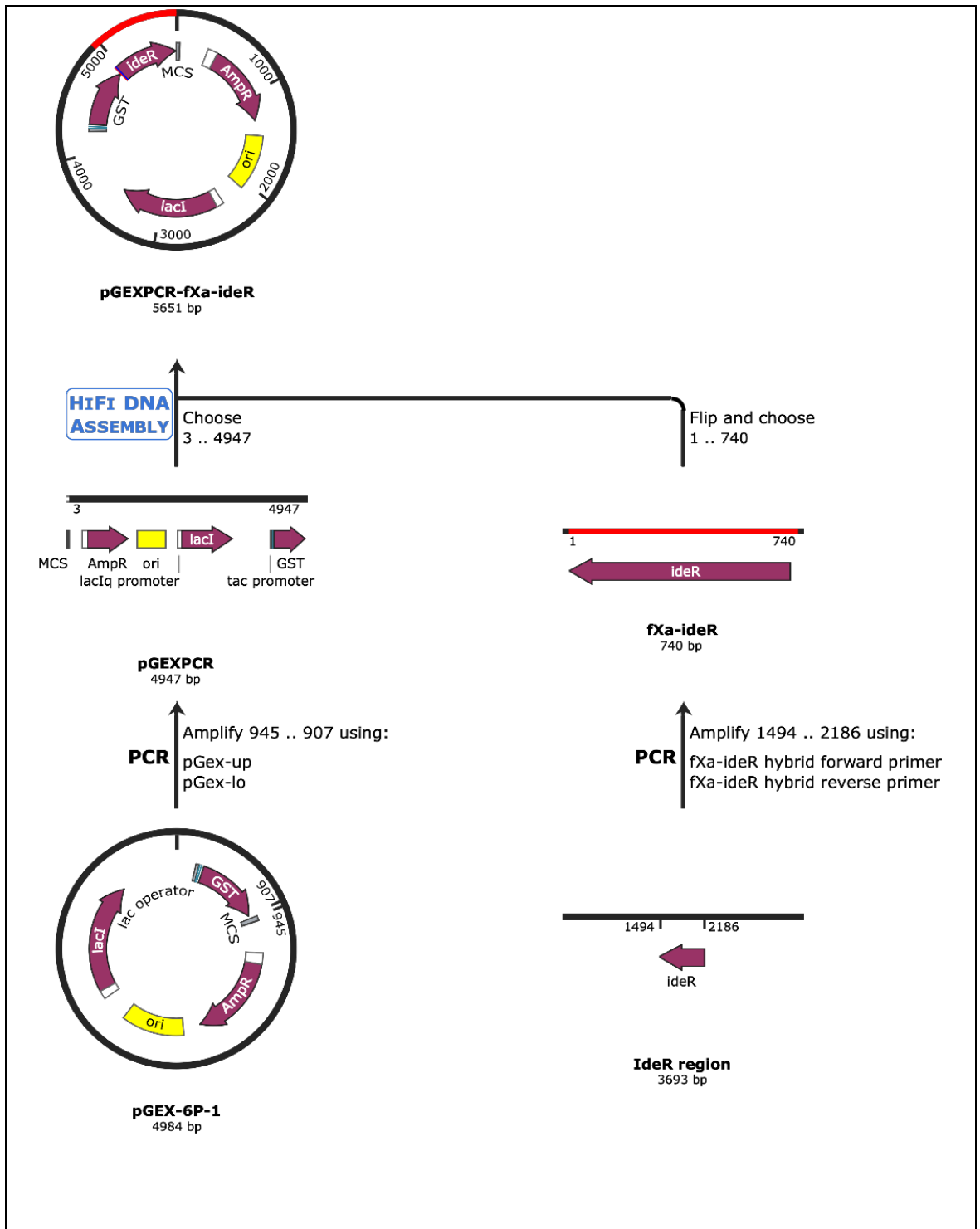


Figure 4-19: Experiment methodology for producing pGEXPCR-fXa hybrid plasmids, the example above uses the gene *ideR*.

4.5.4 Recombinant gene expression

The sequence-confirmed plasmids pGEXPCR-fXa-*ideR*, pGEXPCR-fXa-*dtxR* and pGEXPCR-fXa-*furA* were transformed using *E. coli* BL21 (DE3) competent cells to facilitate heterologous protein overproduction, however pGEXPCR-fXa-*furB* was not due to an unsuccessful plasmid assembly. Overnight cultures were setup using LB media, from which 1 ml aliquots were sub-cultured into

autoinduction medium supplemented with trace elements. The cultures were grown at 30 °C for two days, to facilitate cellular growth and complete glucose utilisation, which subsequently initiates automatic protein induction upon utilisation of lactose in the media.

The cellular material was harvested, resuspended and lysed as described previously, before centrifugal clarification. The clarified lysates were analysed using SDS-PAGE for protein overproduction (Figure 4-20), from which it appeared that the recombinant fusion protein IdeR-GST was both soluble and overproduced, but *dtxR*-GST and *furA*-GST did not appear to facilitate overexpression. In order to make timely progress on this project, it was decided to proceed with purification of IdeR-GST, as the others would need further optimisation.

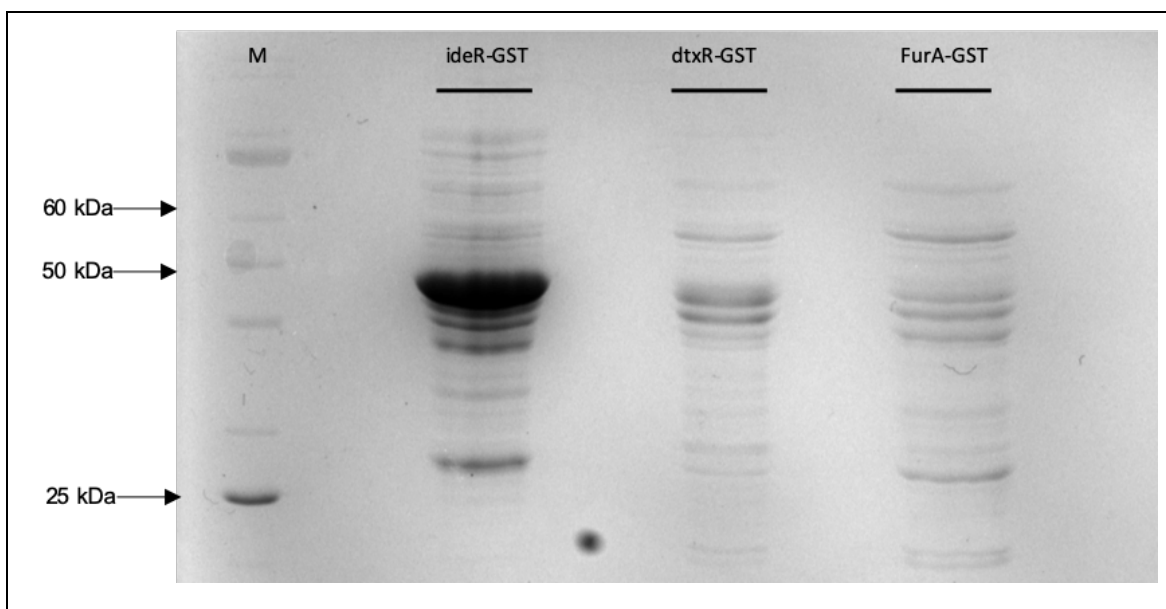


Figure 4-20: Electrophoretic analysis of heterologous protein over production from clarified lysates.

Analysis performed by 12 % SDS-PAGE; stained using coomassie blue. M – NEB broadrange prestained protein marker, Samples were clarified lysates of the GST-fusion proteins.

4.5.5 GStrap Purification of IdeR-GST

The clarified lysate of pGEXPCR-fXa-*ideR* was subjected to affinity chromatography using a prepacked 1 ml GStrap FF HiTrap column (GE Healthcare) as described in section 2.10.1.2. Column bound proteins were eluted under non-denaturing conditions using an isocratic gradient of 100 % elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 7.9) at a flow rate of 1 ml / min. Eluate was collected in 4 ml fractions; fractions of interest identified by UV absorbance by FPLC (Figure 4-21) were analysed by SDS-PAGE shown in Figure 4-22 and Figure 4-23.

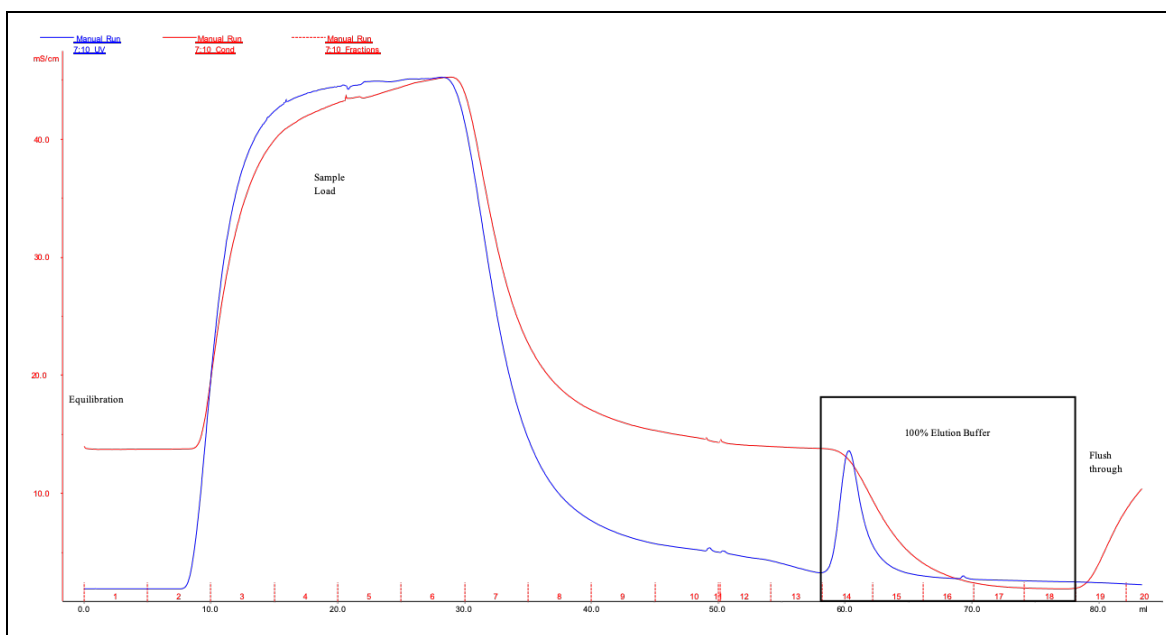


Figure 4-21 FPLC chromatogram of GStrap purification of IdeR-GST

Column was equilibrated using buffer PBS, pH 7.3; sample applied and washed with 10 column volumes of equilibration buffer and eluted using an elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 7.9)

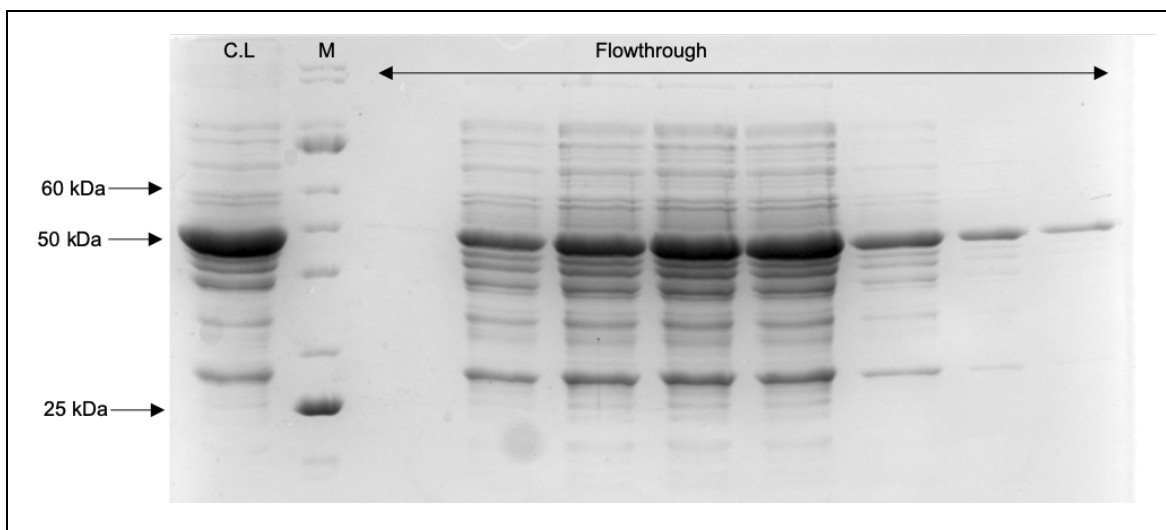


Figure 4-22 Electrophoretic analysis of FPLC flowthrough fractions for purification of IdeR-GST

Analysis performed by 12 % SDS-PAGE; stained using coomassie blue. M – NEB broadrange prestained protein marker, C.L – clarified lysate from Figure 4-20; Flowthrough fractions 2 – 9. .

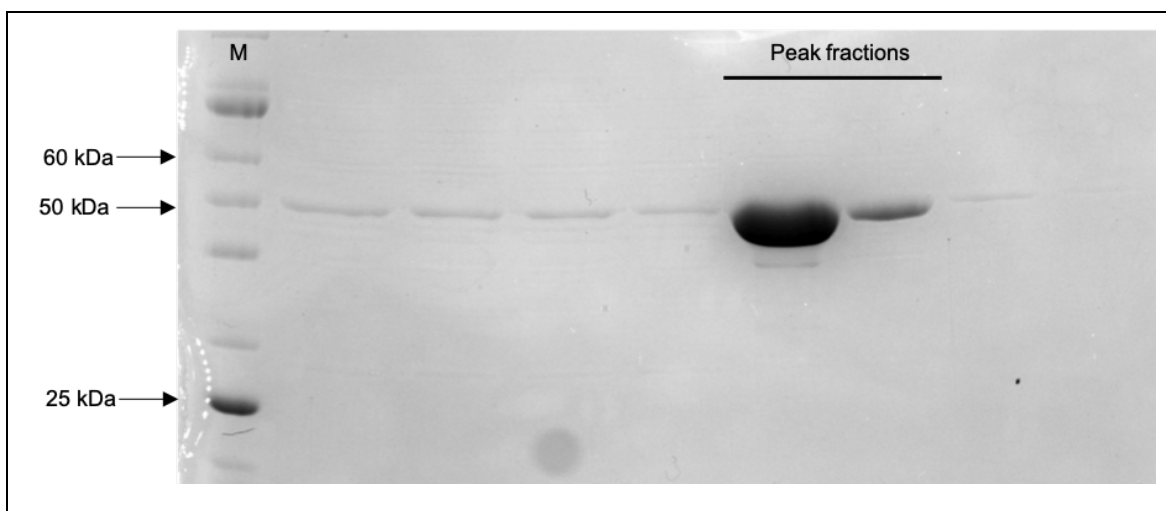


Figure 4-23: Electrophoretic analysis of FPLC peak fractions for the purification of IdeR-GST

Analysis performed by 12 % SDS-PAGE; stained using coomassie blue. M – NEB broadrange prestained protein marker, fractions 10 – 17 tested with peak fractions 14 and 15 identified.

Purification of the protein using a one-step elution under non-denaturing conditions provided the option to identify the native metal binding capacity using an ICP-OES. To perform this analysis, first the fusion-tag was to be cleaved using ImmunoPure Factor Xa purified from Bovine plasma.

While measurement of the IdeR protein metal-binding capacity using ICP-OES would potentially infer a preference for a range of different divalent cations, analysis without the other three metalloregulatory proteins would limit the scope of the project. Therefore, the IdeR-GST protein sample was stored at -20 °C until heterologous gene expression for the other metalloregulatory-GST constructs was successful.

4.6 Strategy 3 - *In silico* analysis of the putative metal regulators of *R. equi* 103S by protein scaffolding

The initial experimental design was to analyse the native metal binding sites of the metal regulatory proteins; however, the project was refocussed to facilitate an *in-silico* metal binding analysis was performed alongside metalloregulatory gene mutagenesis and -omic analysis.

4.6.1 Importance of *in silico* structural and physicochemical analysis of proteins

Due to the ever-emerging development of new bioinformatics tools and online databases in the Digital era, protein analysis has developed to facilitate in-depth physicochemical analysis and structure elucidation through *in silico* protein modelling, as well as prediction of protein function and protein-protein interactions based upon homology to existing structures. The considerable advantage of *in silico* studies over experimental studies (X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy (EM)) is the reduction in costs associated with labour and time but relies on having a pre-existing homologous database entry. The ability to perform high-throughput sequence analysis *in silico* is advantageous to rapidly assign putative identities, contributing to identification of novel properties and drug screening.

4.6.2 *in silico* homology-based protein-modelling

The most successful method for *in silico* protein modelling is using a homology-based approach, in which the three-dimensional structure of the query sequence is predicted by scaffolding to an existing structure of an evolutionarily-related template protein structure, an overview is given in Figure 4-24 (Sanchez and Sali, 1997).

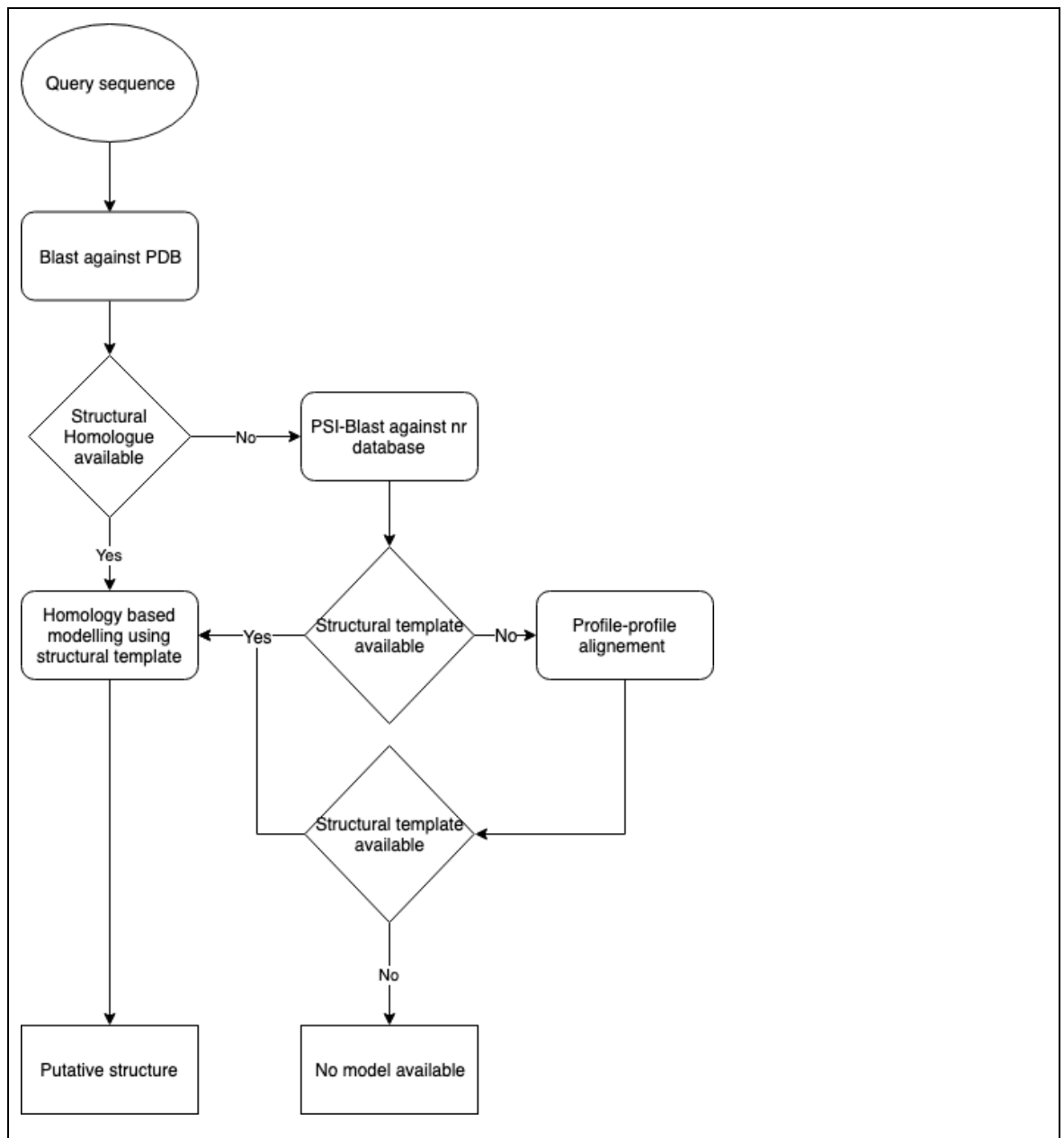


Figure 4-24:Flowchart detailing the basic principles of *in silico* protein-modelling

The pipeline of *in silico* homology-based modelling is typically performed in four stages, first identification of the best template typically using a BlastP search; from which resolved crystal structural homologues in the protein data bank will be selected preferentially. Secondly, a target-template sequence alignment will be performed, followed by structure building, and finally model evaluation (Nikolaev et al., 2018).

With the perpetually developing field of bioinformatics, there are now a range of tools that can be utilised to streamline and automate the process. Phyre2 is a protein homology/analogy recognition engine designed to predict a query structure by using a library of known protein structures taken from the Structural Classification of Proteins database strengthened with newer depositions in the Protein

Data Bank (PDB) (Kelley et al., 2015b). The protein sequences of each of these structures is cross-referenced against the nonredundant sequence database and a profile constructed, from which known and predicted secondary structures are also compiled. The query sequence is searched against the nonredundant sequence database, and a profile is created, as well as multiple PSI-Blast searches used to gather both close and remote sequence homologs, which are compiled to produce a multiple pairwise sequence alignment. The secondary structure is typically predicted using Psi-Pred, SSPro and JNet tools to identify alpha helices, beta strands and coils. A profile-profile alignment is performed for both the profile and secondary structure to rank alignments, from which the top ten alignments are used to produce full three-dimensional models of the query.

By using the Phyre2 prediction software, it was anticipated that the aim of the chapter, that the roles of the four metalloregulators could be distinguished based upon the structural geometry of their metal coordinating sites when cross-referenced with the RCSB Protein Data Bank.

4.6.3 Prediction of IdeR structure

The IdeR amino acid sequence was submitted to the Phyre2 protein-modelling tool (Kelley et al., 2015b). The output produced a hypothetical protein structure (100 % residues modelled at > 90% confidence) using the crystal structure **1FX7-C** (iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis*, chain C, resolution: 2.00 Å) (Feese et al., 2001). The template was ranked 1 based on raw alignment score, taking into consideration both sequence and secondary structure similarities. The probability of sequence-template homology is given as 100 % confidence, and percentage identity between the sequence and template is 78 % suggesting an extremely high model accuracy.

The predicted output model dimensions (Å) were: X:46.174 Y:47.643 Z:53.182. The *R. equi* IdeR model was imported to the programme Coot v0.8.9.1 (CCP4i2 shell) and superposed based upon protein topology using SSM superposition to 1FX7-C for structural comparison, visualised using PyMol (Figure 4-25).

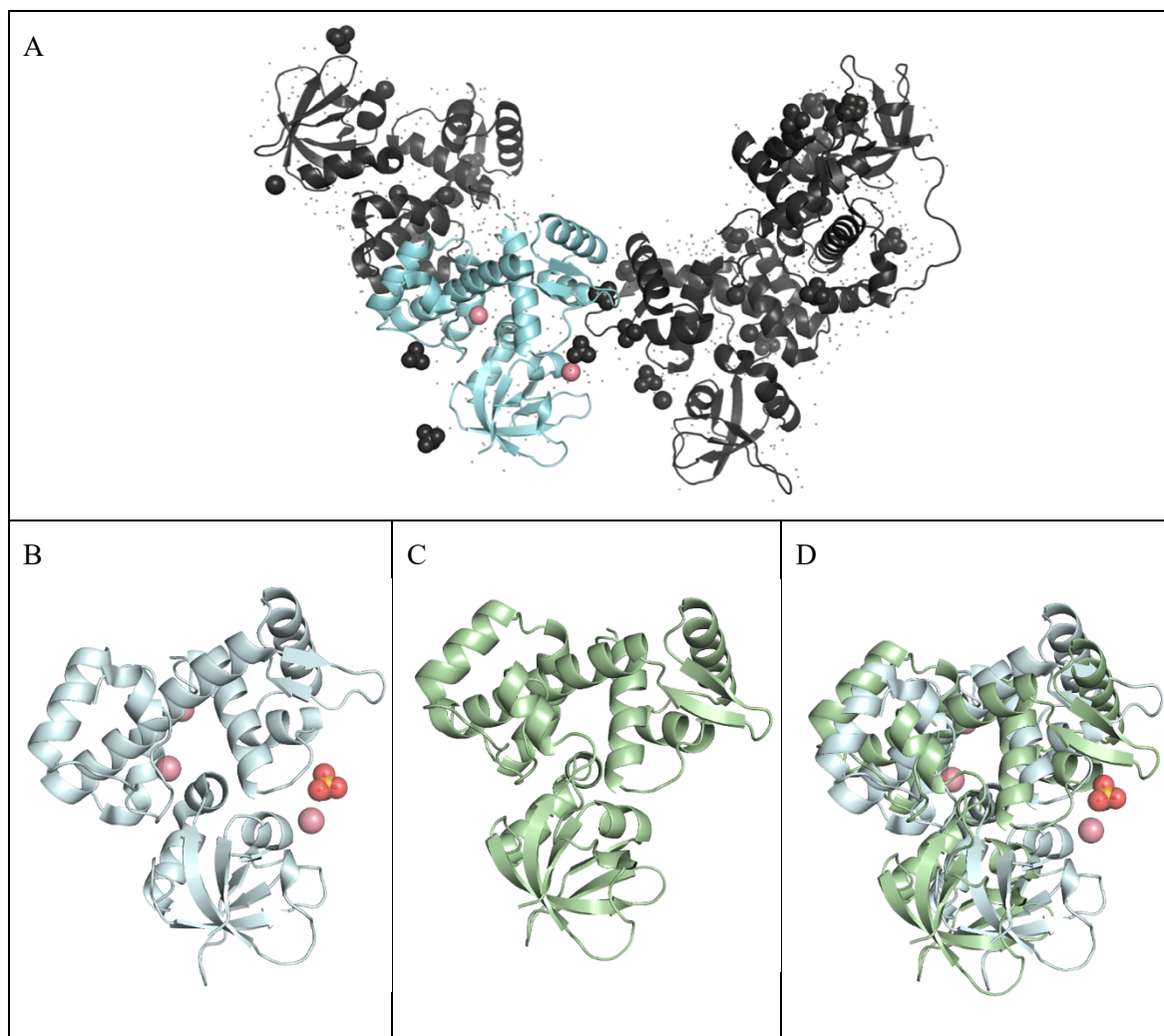


Figure 4-25: A comparative analysis of known and predicted IdeR structures.

A) 2.0 Å crystal structure of wild type *M. tuberculosis* IdeR in complex with Co(II). B) Crystal structure 1FX7; Chain C. A 2.0 Å crystal structure of wild type *M. tuberculosis* IdeR in complex with Co(II). C) Structure prediction of wildtype *R. equi* 103S IdeR using Chain C of 1FX7 as the template. Secondary structure elucidation predicted using Phyre2, aligned using Coot SSM supposition. D) Alignment of known IdeR structure from *M. tuberculosis* (blue) and predicted IdeR structure from *R. equi* 103S (green).

When submitted to the DALI server (Holm, 2019), a pairwise alignment between 1FX7-C and *R. equi* IdeR model produced a Z-score of 36.5 and a rmsd value of 0.5, further illustrating the high degree of similarity; it is important to note that a further degree of variance is to be expected due to the comparison of holo- and apo- structures, and that the protein modification in the form of metal binding may produce a degree of molecular flexion.

Given the predicted high model accuracy, the predicted metal coordinating sites of IdeR from *R. equi* are reminiscent of those observed in IdeR from *M. tuberculosis*; (pdb no: 1FX7; sequence chain view shown in Figure 4-26). Upon performing a sequence alignment (Figure 4-27), the three metal coordinating sites appear to be inherently conserved. Before in depth discussion regarding the metal

binding coordination, it is important to note that although the crystal structure of 1FX7 was resolved with binding of cobalt, the ionic radii of Co (II) and Fe (II) are distinctly analogous even in differing coordination states. Furthermore, a range of transition metal ions [ranging from Fe(II), Co(II), Ni(II), Zn(II), Mn(II), or Cd(II)] can act as corepressors with DtxR *in vitro* (Schmitt and Holmes, 1993).

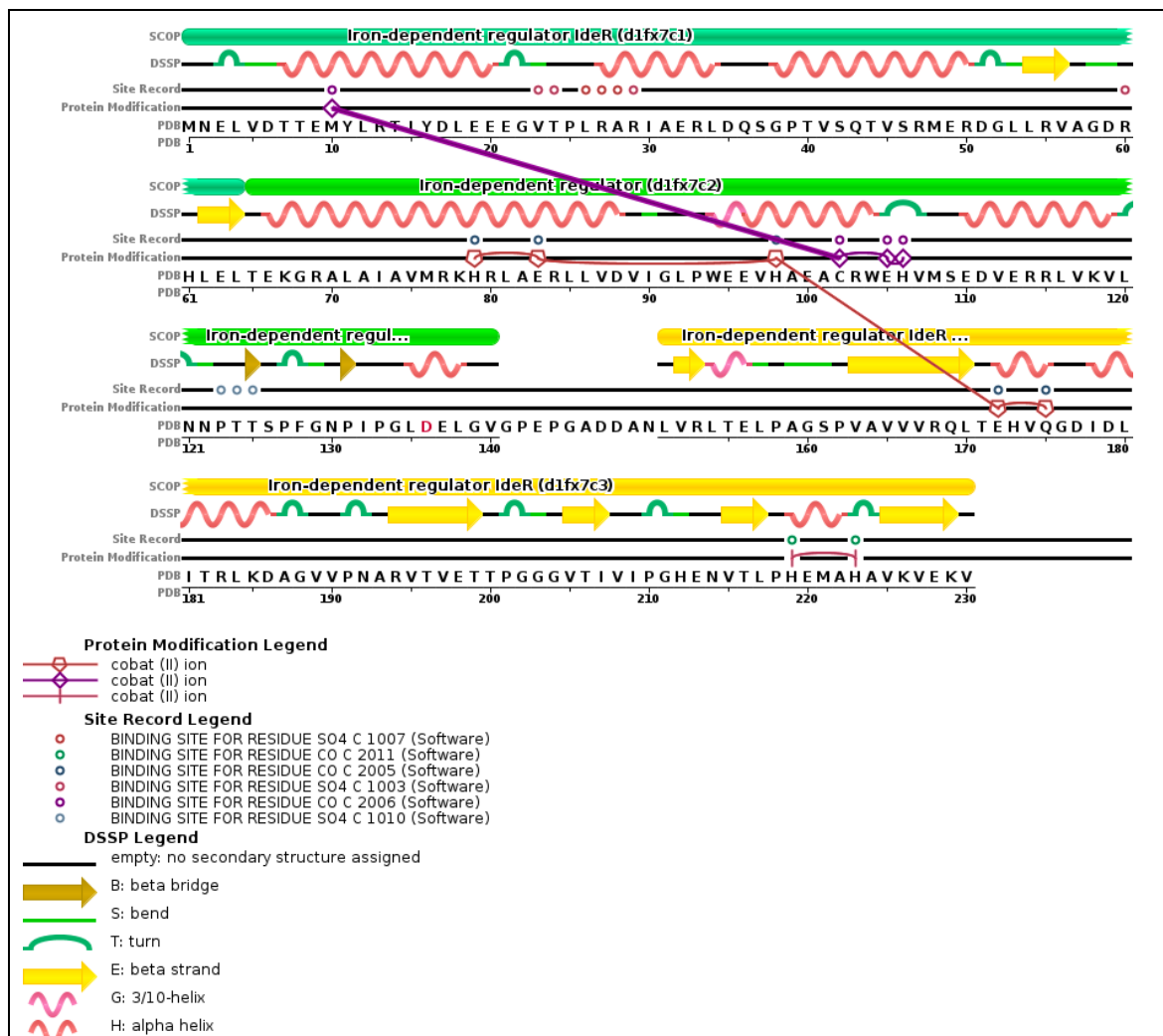


Figure 4-26: Chain Sequence View for the Entity PDB 1FX7.

The three domains are identified by both colour and d1fx7c(x) variation given by SCOP (Murzin et al., 1995). Protein secondary structure was identified by DSSP and variations are listed in the DSSP legend (Kabsch and Sander, 1983). Structural features including ligand binding sites are listed on by Site Record and Protein modification and variations are listed in the Legend (Golovin et al., 2005; Montecchi-Palazzi et al., 2008).

IdeR(m.tb)	MNELVDTEYLRTIYDLEEEGVPLRARIAERLDQSGPTVSQTVSRMERDGLLRVAGDR	60
ideR(r.eq)	MKDLVDTEYLRTIYDLEEEGVPLRARIAERLEQSGPTVSQTVARMERDGLLQVAGDR	60
	*.:*****:*****:*****:*****:*****:*****:*****:*****:*****	
IdeR(m.tb)	HLELTEKGRALAIAMRKHRLAERLLVDVIGLPWEEVHAEACRWEHVMSDEVERRLVKVL	120
ideR(r.eq)	HLELTEKGRNLAVAMRKHRLAERLLVDIIGLEWDQVHAEACRWEHVMSDEVERRLVEVL	120
	***** **.:*****:*****:*** *:*****:*****:*****:*****:*****	
IdeR(m.tb)	NNPTTSPFGNPIGPLVELGVGPEPGADDANLRLTELPAGSPVAVVVRQLTEHVQGDIDL	180
ideR(r.eq)	KNPTTSPYGNPIGPLADLGLDRPVG-NAETLIRLTDVPPGKPTAVVVRRLAEHVQSDPEL	179
	.*****:*****:.*:. * : .*:***:* *.*.*****:*****:*.:*	
IdeR(m.tb)	ITRLKDAGVVPNARVTVETTPGGGVTVIPGHENVTLPEMAHAVKVEKV	230
ideR(r.eq)	IGQLREAGVVPDARVTVETRPV-SVTITASGHDEFDLPPEMAHAVQVKQV	228
	*.:*****:***** ** .*** **.:. **.:*****:.*:*	

Figure 4-27: Sequence alignment between IdeR proteins from *R. equi* and *M. tuberculosis*. Metal binding sites identified for IdeR(m.tb) were annotated using relevant binding domain analyses from Figure 4-26 and IdeR(R.eq) binding sites were extrapolated accordingly. Metal binding coordinating groups are clustered by colour.

Analysis of the first metal binding site for IdeR (1FX7) identifies that a Cobalt (II) atom is pentavalently coordinated (Feese et al., 2001) by directly using protein side chain-ligands His₇₉, Glu₈₃, His₈₉, Glu₁₇₂ and Gln₁₇₅ (Figure 4-28) with distorted trigonal bipyramidal geometry (a common geometric configuration for first-row transition metals given the ionic radii as previously mentioned).

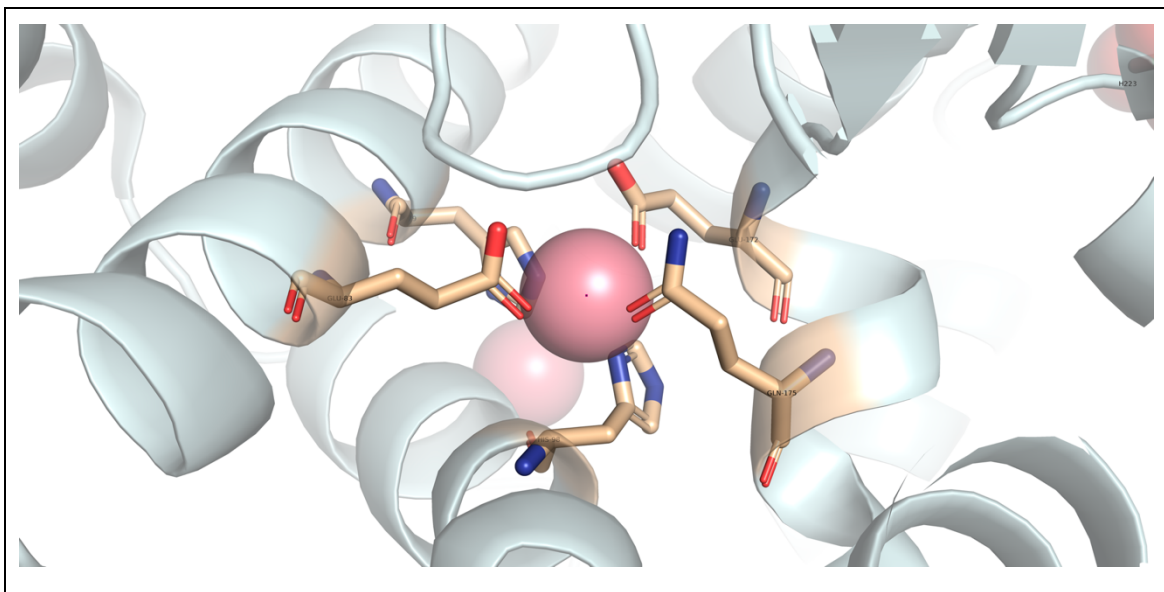


Figure 4-28: Metal binding site 1 of 1FX7, *M. tuberculosis* IdeR in complex with Co(II).

Given the significant sequence similarities, it is plausible to hypothesise that the protein IdeR produced by *R. equi* coordinates metal by means of an analogous strategy utilising protein side chain-ligands His₇₉, Glu₈₃, His₈₉, Glu₁₇₁ and Gln₁₇₄.

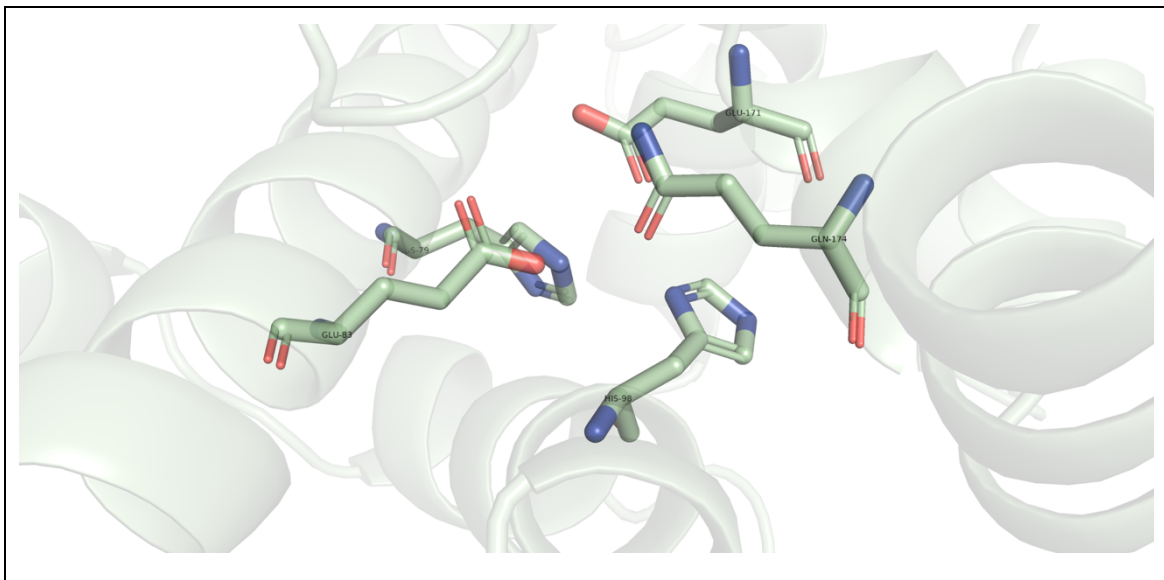


Figure 4-29: Hypothesised metal binding site 1 of the homology-based model of IdeR from *R. equi*, in apo-form.

Interestingly, when characterising the crystal 1FX7, Feese and colleagues identified a water molecule (1) residing within a polar pocket generated by the side chains of Arg₈₀, Glu₈₃, Ser₁₂₆, and Asn₁₃₀ and Glu₁₇₂, and Gln₁₇₅. Although the water molecule does not directly coordinate with the metal, it may be involved in metal-dependent activation of IdeR, based upon electrostatic interactions that can bridge the two metal binding ligands of domain 3 with domain 2 (Feese et al., 2001). This polar pocket appears to be present within the model of IdeR from *R. equi* given that the amino acids involved are also conserved.

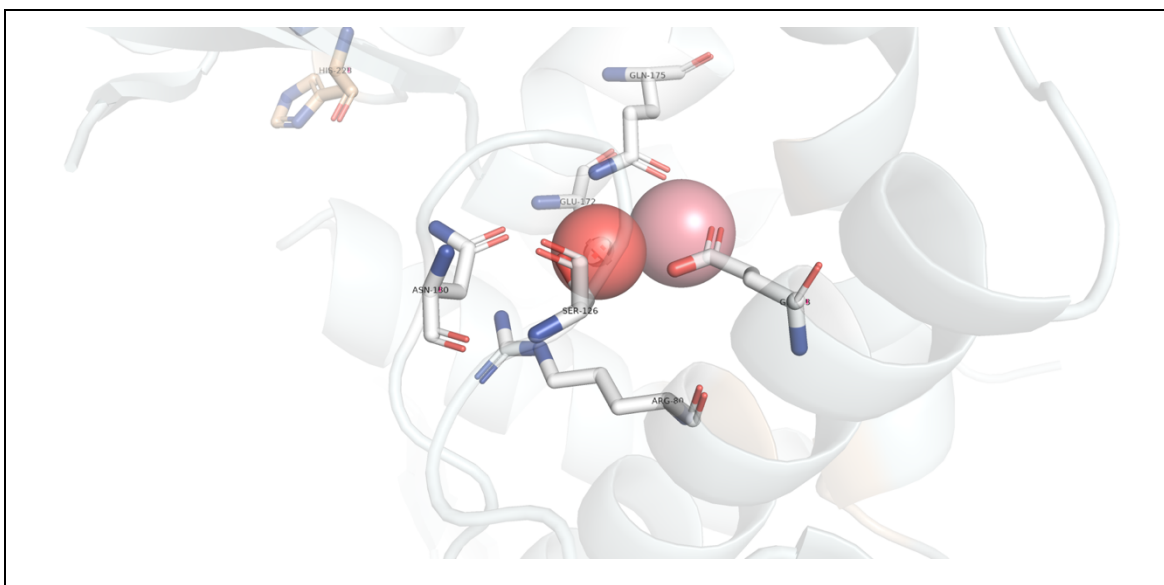


Figure 4-30: A Polar pocket occupied by a water molecule adjacent to metal binding site 1 of IdeR (1FX7).

Analysis of the second metal binding domain in the IdeR crystal structure 1FX7 yields a Cobalt (II) atom in an octahedral coordination complex (Feese et al., 2001), liganded by four protein side chains,

Met₁₀, Cys₁₀₂, Glu₁₀₅ and His₁₀₆ (Figure 4-31). Critically, the Cysteine₁₀₂ residue operates as a bidentate ligand utilising both the sulphur and carbonyl oxygen atoms adjacently. The final ligand required to complete the octahedral geometry was identified as a solvent group (likely a water molecule (2) or a hydroxyl ion), which unlike that in the first metal binding site, directly coordinates with the metal as well as bridging to the main chain carbonyl oxygen of Leu₄ and as such is postulated to facilitate metal-dependent repressor activity. Stabilisation of the *N*-terminal residues through electrostatic and hydrophobic interactions is understood to be vital for orientation and maintenance of the DNA-binding helices to bind DNA.

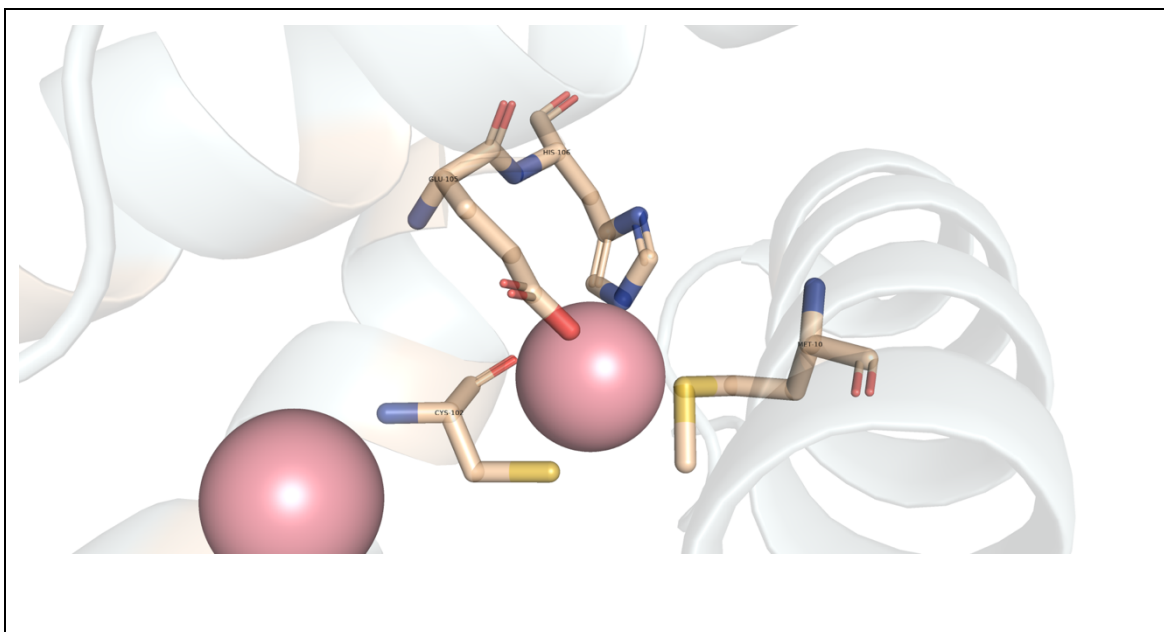


Figure 4-31: Metal binding site 2 of *M. tuberculosis* IdeR in complex with Co(II).

As for the first, the second metal binding site of *R. equi* IdeR is also likely to operate in a similar manner to the mycobacterial IdeR counterpart, conforming in an octahedral coordination complex. The conservation of the binding pocket using Met₁₀, Cys₁₀₂, Glu₁₀₅ and His₁₀₆ amino acid side-chains (Figure 4-32) illustrates that a transition metal is likely to coordinate in a corresponding manner including both the bidentate Cys₁₀₂, and senary solvent ligand interface.

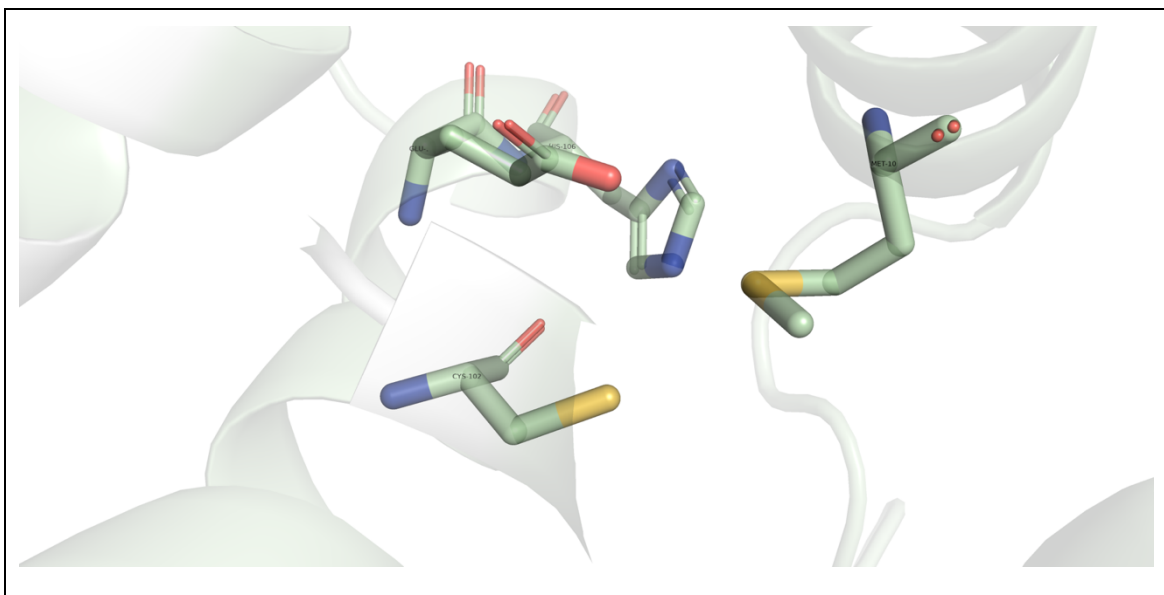


Figure 4-32: Hypothesised metal binding site 2 of the homology-based model of IdeR from *R. equi*, in apo-form.

The crystal structure 1FX7 for IdeR also was identified to have a third metal-coordinating site, where another cobalt(II) atom was found complexed with two histidine ligands His219, His 223 and four coordinating water molecules. The close proximal location of the histidine residues is characteristic of a metal coordinating site denoted by a **HEX₂H** motif and is likely that a single α -helix turn between the two histidine's facilitate the residues to function as adjacent ligands in an octahedral coordination.

Unusually, two coordinating water molecules appear to bridge the cobalt (II) metal atom to a sulphate ion in a non-canonical metal binding site for DtxR-like structures, it was therefore postulated by the authors that this third metal binding domain was likely a construct of the crystallisation process, given the high concentrations of cobalt and lithium sulphate available

This is further corroborated, when contrasted with the *R. equi* IdeR sequence alignment, which identified an amino acid substitution in the **HEX₂H** motif, that has been postulated to alter the metal-binding specificity by instead utilising glutamic acid rather than histidine for location 219 (**EEX₂H**) (Taylor et al., 2005). This, however, reduces our confidence in the metal binding analysis, conferring that this third metal binding site is not essential, and as previously predicted, a convenient by-product of the crystallisation process.

4.6.4 Prediction of DtxR structure

The DtxR amino acid sequence from *R. equi* was submitted to the Phyre2 protein-modelling tool. The initial output produced a hypothetical structure using the crystal structure 5ZR4 (manganese regulator – MntR, *Mycobacterium tuberculosis*, chain B) as a template. The template was a 70 % identity match with 100 % confidence.

The crystal structure for MntR from the closely related species *Mycobacterium tuberculosis* was only recently resolved by Cong et al. (2018). Although the study of DtxR-like metalloregulators within *Mycobacterium tuberculosis* has previously been explored with structural studies of IdeR, Cong and colleagues succeeded in producing crystal structures for MntR in both apo- and –holo- forms (5ZR4 and 5ZR6 respectively). Given the gaps in secondary structure annotation by DSSP for the apo-form of MntR, it was hypothesised that the holo-form may be a better structural scaffold for DtxR to further clarifying the putative structure by *in silico* methods, from which a one-to-one threading of the crystal structure 5ZR6 yielded a 73 % identity match with 100 % confidence.

The *R. equi* DtxR model was imported to the programme Coot v0.8.9.1 EL (CCP4i2 shell) and superposed based upon protein topology using SSM superposition to 5ZR6-A and 5ZR4-A for structural comparison, visualised using PyMol (Figure 4-33)

When submitted to the DALI server (Holm, 2019), a pairwise alignment between 5ZR6-A and *R. equi* DtxR model produced a Z-score of 35.4 and a rmsd value of 0.4.

Given the predicted high model accuracy, the anticipated metal coordinating sites of DtxR from *R. equi* are reminiscent of those observed in MntR from *M. tuberculosis*; (pdb 5ZR6; sequence chain view shown in Figure 4-34). Upon performing a sequence alignment (Figure 4-35), the two metal coordinating sites appear to be inherently conserved.

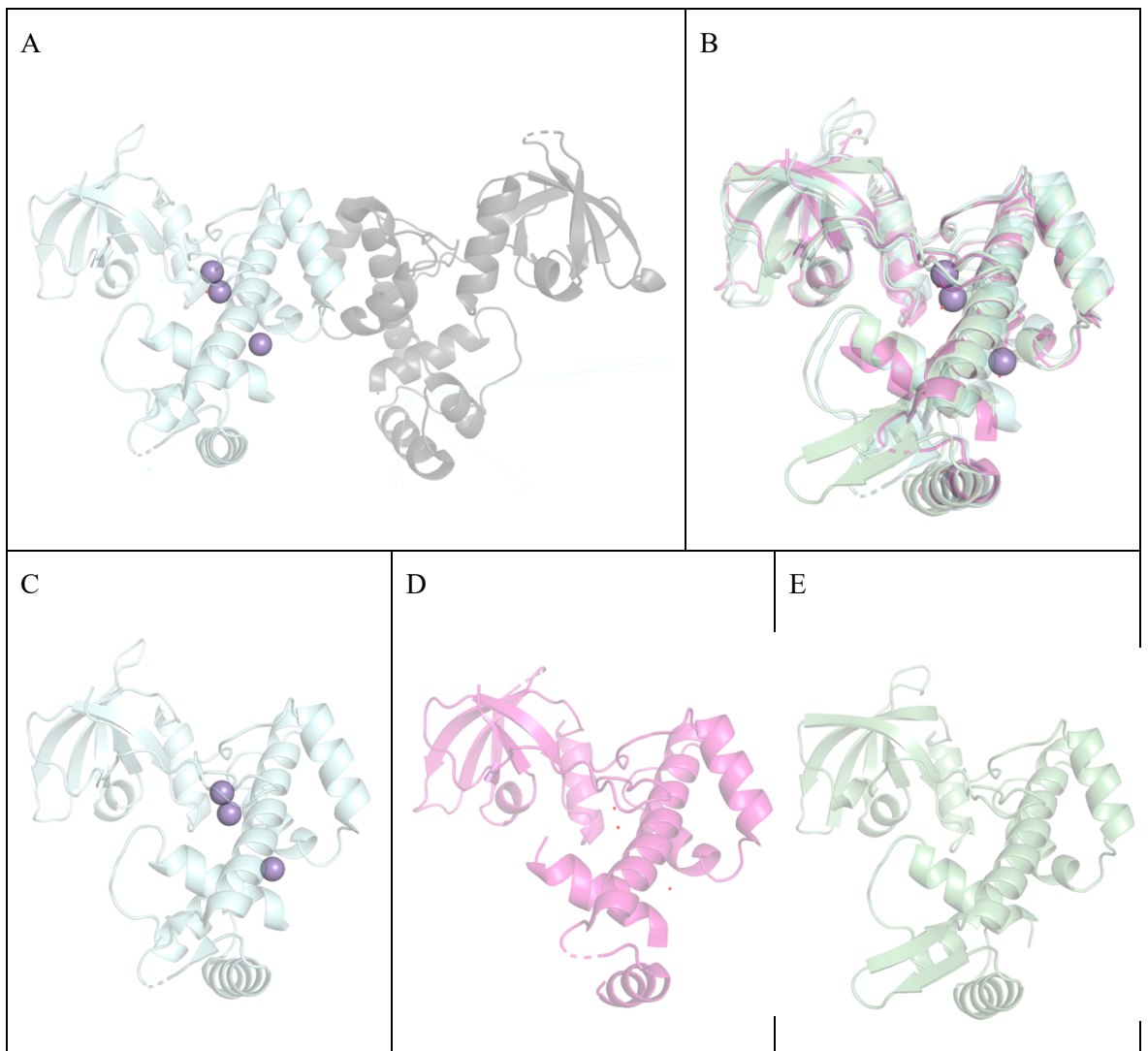


Figure 4-33: A comparative analysis of known and predicted DtxR/MntR structures.

A) 3.0 Å crystal structure of wild type *M. tuberculosis* MntR in complex with Mn(II). B) Structural superposition of 5ZR6, 5ZR4 and DtxR (*R. equi*) C) Crystal structure 5ZR6; Chain A. A 3.0 Å crystal structure of wild type *M. tuberculosis* DtxR in complex with Mn(II). D) Crystal structure 5ZR4; Chain A. A 3.1 Å crystal structure of wild type *M. tuberculosis* DtxR in apo-form. E) Structure prediction of wildtype *R. equi* 103S DtxR using Chain A of 5ZR6 as the template. Secondary structure elucidation predicted using Phyre2, aligned using Coot SSM superposition.

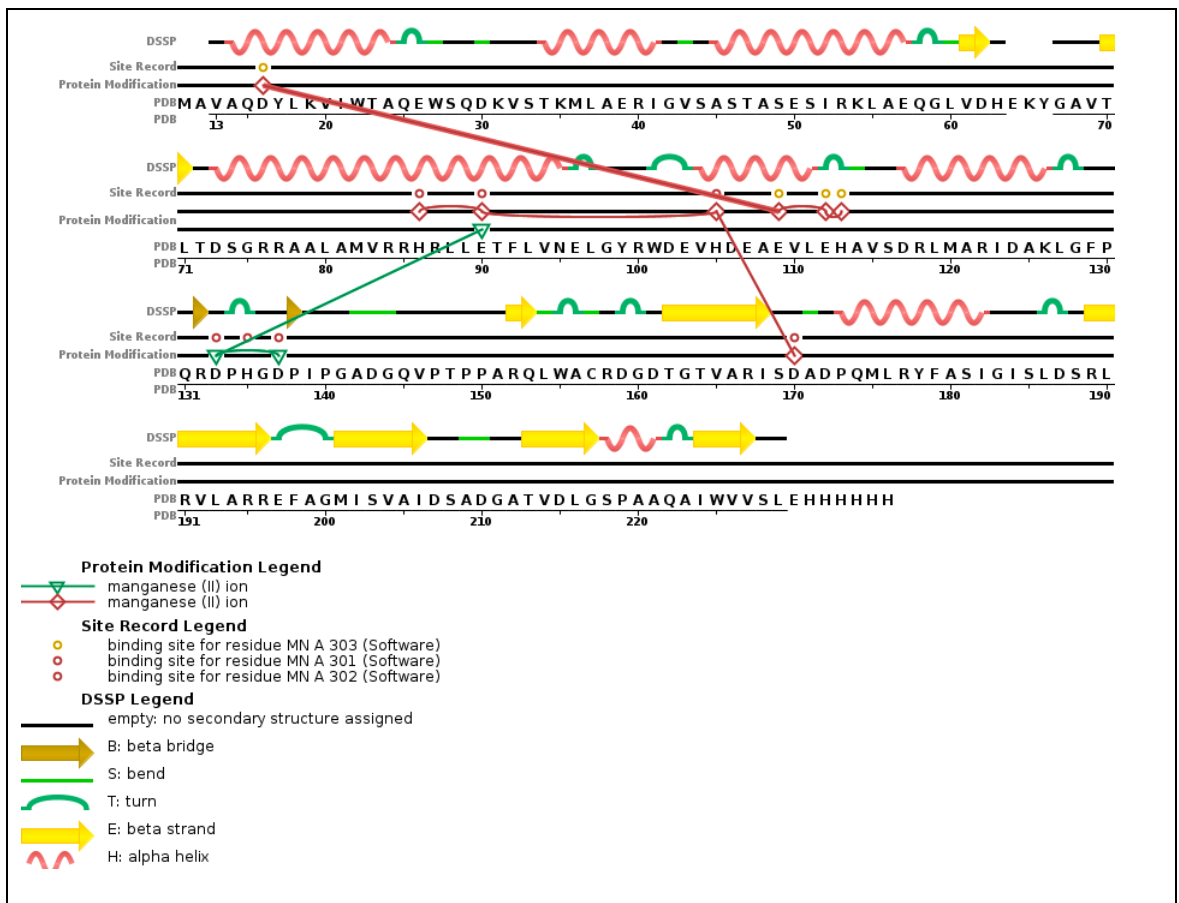


Figure 4-34: Chain Sequence View for the Entity PDB 5ZR6.

Protein secondary structure was identified by DSSP and variations are listed in the DSSP legend (Kabsch and Sander, 1983). Structural features including ligand binding sites are listed on by Site Record and Protein modification and variations are listed in the Legend (Golovin et al., 2005; Montecchi-Palazzi et al., 2008).

MntR	-----MRADEEPGDL\$AVAQDYLVKIWTAQEWSQDKVSTKMLAERIGVS	44
DtxR	MPKLVTVATQKSDAPAPSEVTHPETLSSVAQDYLVKIWTVQEWSRERVSTKLLSERIGVS	60
	* **.******.*****::****.*:*****	
MntR	ASTASESIRKLAEQGLVDHEKYGAVTLTDSGRRALAMVRRRLITFLVNELGYRWDEV	104
DtxR	ASTVSEAIRKLSQGLVDHARYGSIALTDAGRSAAVSMVRRRLITFLVNELGYGWDEV	120
	.*:.:*****.*:***:***.*:*****:***** *****	
MntR	HDEAEVLEHAVSDRLMARIDAKLGFPERPHGPIPGADGQVPTPPARQLWACRDGDTGT	164
DtxR	HDEAEVLEHAVSDRMIDRIDAKLGFPERPHGPIPSADGSVPTPPARQLSDYQDGESGR	180
	*****:*****:*****.*:*****.*:***.*	
MntR	VARISDAPQMLRYFASIGISLDSRLRVLARREFAGMISVATDS-ADGATVDLGSPPAAQA	223
DtxR	VARISDAPAMLYFDSVGIALDITVIERRDFAGTVSIRLGTDPATGNVDLGNPAAQA	240
	***** *****.*:***:***.:*:*:***:***.*:..****.*****	
MntR	IWVVS 228	
DtxR	IWLV- 244	
	**:*	

Figure 4-35: Sequence alignment between DtxR and MntR proteins from *R. equi* and *M. tuberculosis* respectively.

Metal binding sites identified for MntR were annotated using relevant binding domain analyses from Figure 4-33 and DtxR binding sites were extrapolated accordingly. Metal binding coordinating groups are clustered by colour. Black highlighted nucleotides were deleted in for crystal constructs to improve stability.

Analysis of the first metal binding site for MntR (5ZR6) identified an Mn(II) atom in an octahedral coordination configuration located between the helix-turn-helix domain and dimerization domain (Cong et al., 2018). The metal ion is coordinated using protein side chain-ligands Asp₁₆, Glu₁₀₉, Glu₁₁₂, His₁₁₃; pivotally, the Glu₁₀₉ residue operates as a bidentate ligand utilising both the side-chain and carbonyl oxygen atoms adjacently, as well as a senary solvent ligand to complete the octahedral coordinating interface (Figure 4-36; water group not shown).

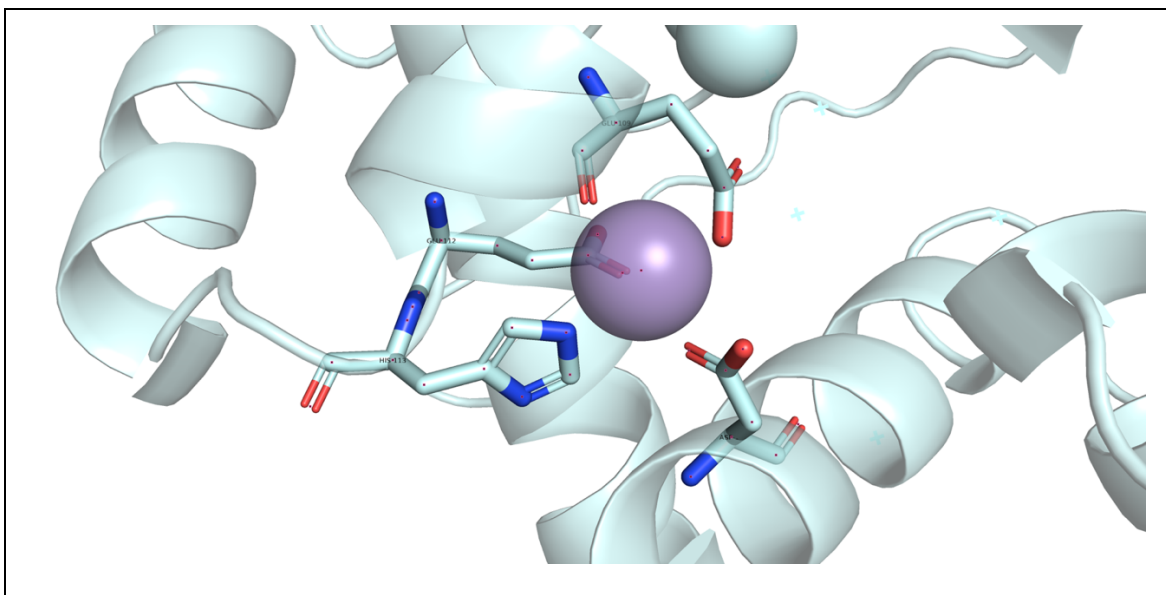


Figure 4-36: Metal binding site 1 *M. tuberculosis* MntR 5zr6 complexed with Mn(II)

Given the significant sequence similarities, it is conceivable that the protein DtxR produced by *R. equi* coordinates metal ions by analogously utilising protein side chain-ligands Asp₃₂, Glu₁₂₈, His₁₂₉, as well as Glu₁₂₅ functioning as a bidentate ligand, and a senary solvent molecule ligand (Figure 4-37) as described for MntR above.

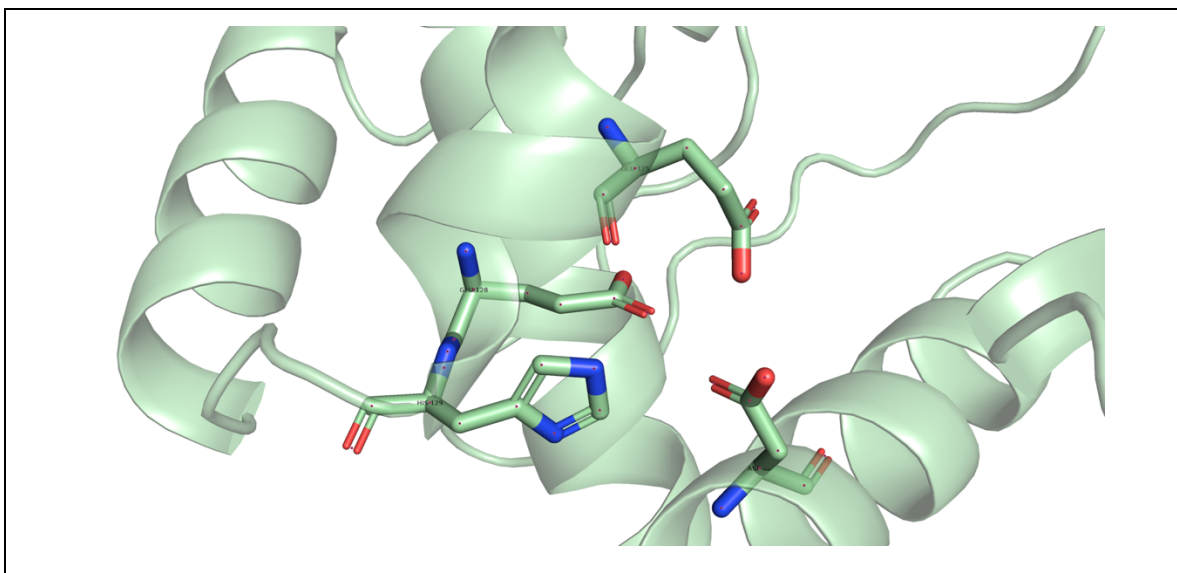


Figure 4-37: Predicted metal binding site 1 of the homology-based model for DtxR in *R. equi*, in apo-form.

Analysis of the secondary metal binding site identified an unusual pentavalent bi-nuclear metal ion cluster, with two manganese ions present, bridged by one water molecule (Cong et al., 2018). The first manganese ion is coordinated using protein side chain-ligands His₈₆, His₁₀₅, Asp₁₇₀, while the second is coordinated using protein side chain-ligands, Asp₁₃₃, His₁₃₅ and Asp₁₃₇, the configuration is completed by bridging the manganese ions using a water molecule and the side chain of Glu₉₀ (Figure 4-38).

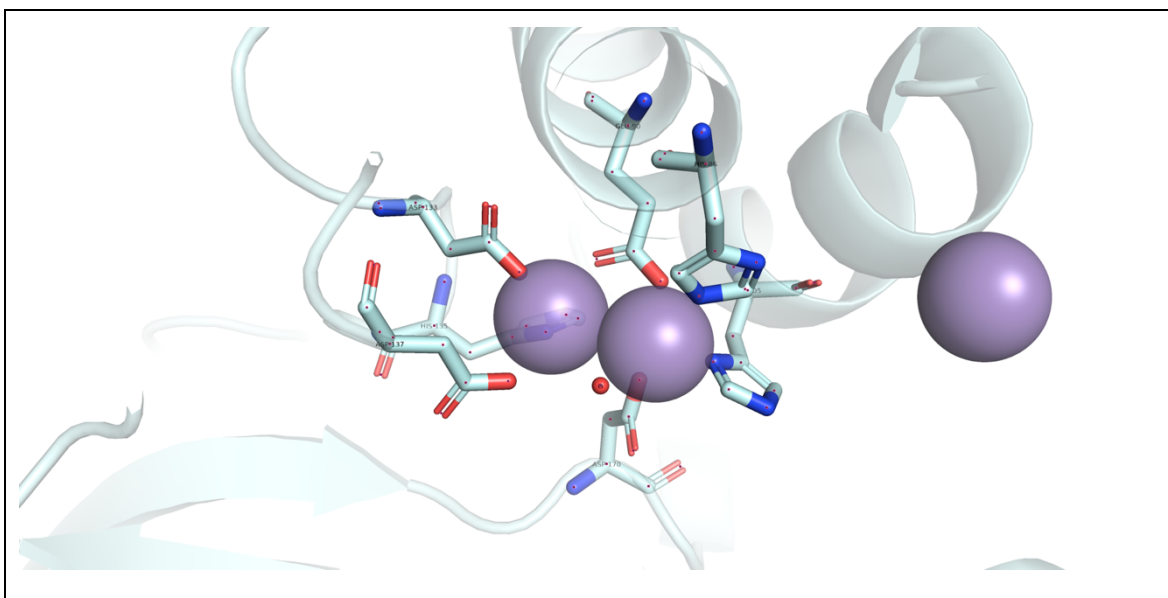


Figure 4-38: Metal binding site 2 *M. tuberculosis* MntR 5zr6 complexed with Mn(II)

The intrinsic nature of the sequence similarities between MntR and DtxR from *M. tuberculosis* and *R. equi* respectively, advocates a deep-rooted functional protein similarity. The conservation of the amino acids involved in this unusual pentavalent bi-nuclear metal ion cluster for the second metal

binding site shows minimal variation when the structures are superposed, further supporting the hypothesis that DtxR operates as a manganese-dependent transcriptional repressor with analogous binding sites to that of MntR. The secondary metal coordinating site includes protein side-chain ligands His₁₀₂, Glu₁₀₆, His₁₂₁, Asp₁₈₆, likely bridged by a water molecule and the side-chain of Glu₁₀₆ to the second ion that is coordinated by side-chain ligands of Asp₁₄₉, His₁₅₁ and Asp₁₅₃ (Figure 4-39).

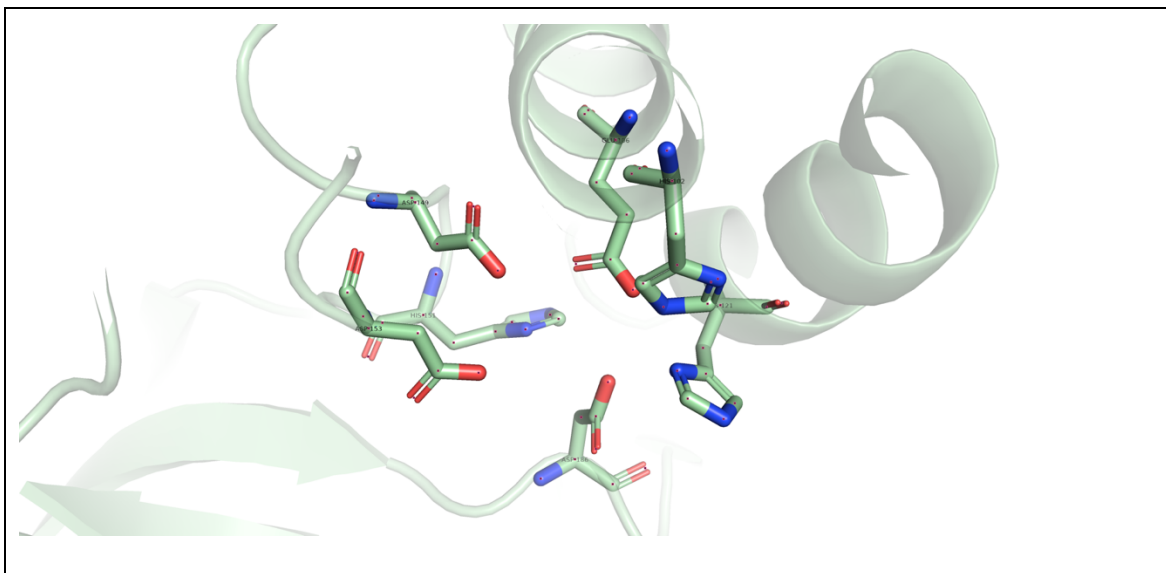


Figure 4-39: Predicted metal binding site 2 of the homology-based model for DtxR in *R. equi*, in apo-form.

4.6.5 Prediction of FurA structure

The FurA amino acid sequence was submitted to the Phyre2 protein-modelling tool (Kelley et al., 2015b). The output produced a hypothetical protein structure (74 % residues modelled at 100 % confidence) using the crystal structure **2FE3-B** (peroxide operon regulator from *Bacillus subtilis*, chain B, resolution: 1.75 Å). The template was ranked 1 based on raw alignment score, taking into consideration both sequence and secondary structure similarities. The probability of sequence-template homology is given as 100 % confidence, and percentage identity between the sequence and template is 24 % indicating perhaps a core structural premise.

The *R. equi* FurA model was imported to the programme Coot v0.8.9.1 (CCP4i2 shell) and superposed based upon protein topology using SSM superposition to the 2FE3-B for structural comparison, visualised using PyMol (Figure 4-25).

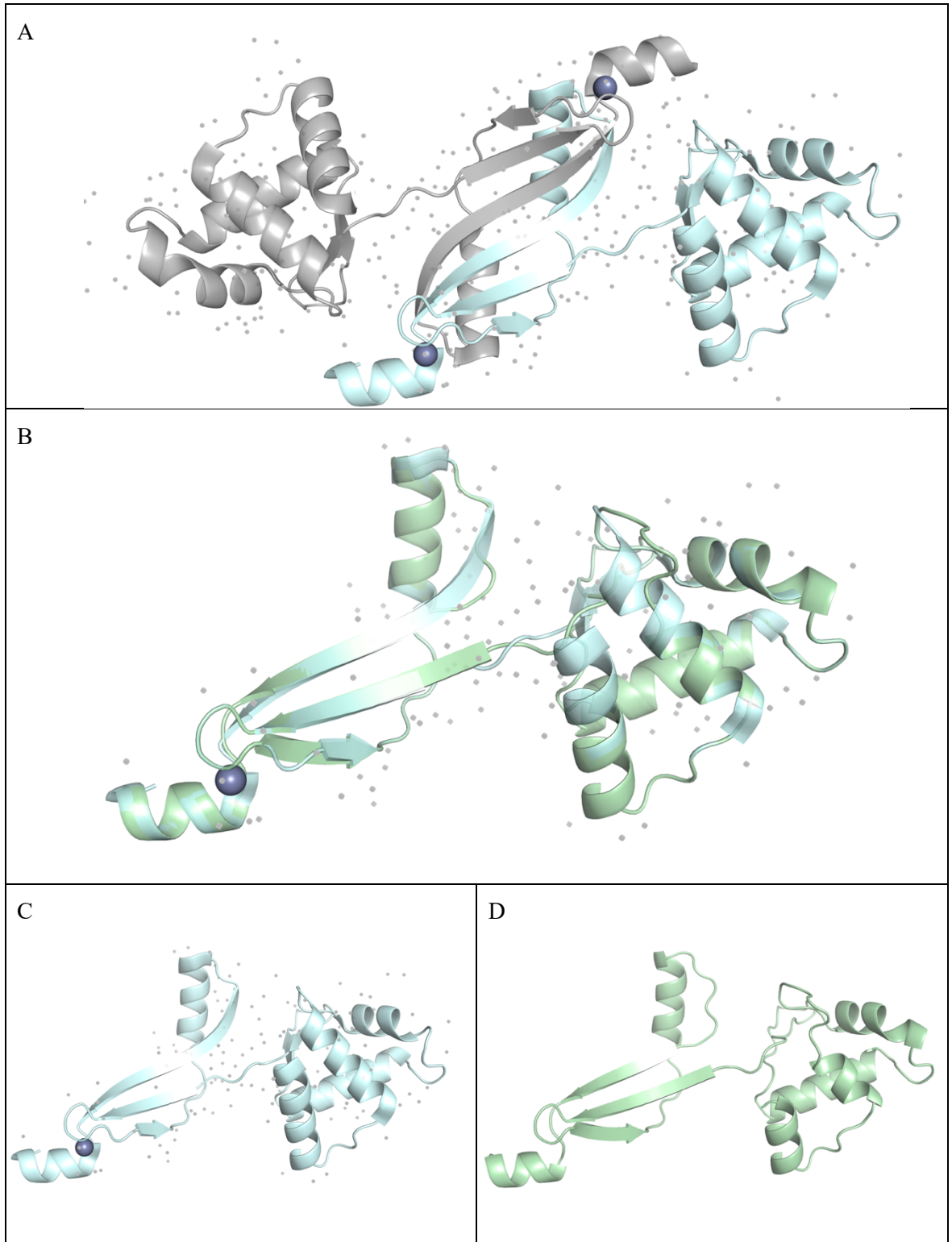


Figure 4-40: A comparative analysis of known and predicted PerR/FurA structures.

A) 1.75 Å crystal structure of wild type *B. subtilis* PerR in a structural complex with Zn(II), the functional regulatory site remains in apo-form. B) Structural superposition of 2FE3 and FurA (*R. equi*) C) Crystal structure 2FE3; Chain B. A 1.75 Å crystal structure of wild type *B. subtilis* PerR in a structural complex with Zn(II), the functional regulatory site remains in apo-form. D) Structure prediction of wildtype *R. equi* 103S FurA using Chain B of 2FE3 as the template. Secondary structure elucidation predicted using Phyre2, aligned using Coot SSM superposition.

When submitted to the DALI server (Holm, 2019), a pairwise alignment between 2FE3-B and *R. equi* FurA model produced a Z-score of 19.4 and a rmsd value of 0.6.

Previous studies identified a comparable sequence homology between a Fur protein from *Pseudomonas aeruginosa* and PerR from *B. subtilis*, yielding a ID score of 26 %, while presenting structural similarities in both the DNA-binding and dimerisation domains. The *in-silico* model of *R. equi* FurA appears to correspond well with these findings, given the supposition shown in Figure 4-40.

Interestingly, the anticipated metal coordinating sites of FurA from *R. equi* also appear to be conserved between analogous species *B. subtilis* and *M. tuberculosis* (Figure 4-41 and Figure 4-42). The sequence alignment revealed a conservation of two metal coordinating sites, that are characteristic of these regulatory proteins.

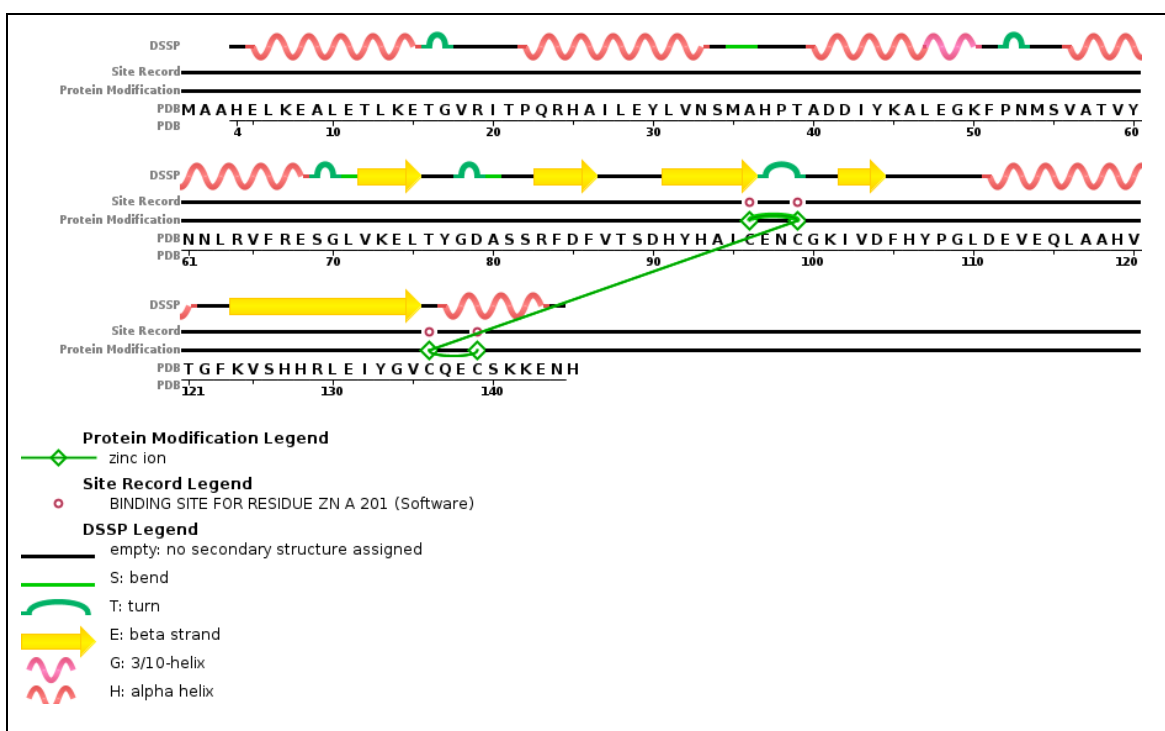


Figure 4-41: Chain Sequence View for the Entity PDB 2FE3.

Protein secondary structure was identified by DSSP and variations are listed in the DSSP legend (Kabsch and Sander, 1983). Structural features including ligand binding sites are listed on by Site Record and Protein modification and variations are listed in the Legend (Golovin et al., 2005; Montecchi-Palazzi et al., 2008)

PerR_(B.subtilis)	-MAAHELKEALETLLKETGVRITPQRHAILEYLVNSMAHPTADDIYKALEGKFPNMSVATV	59
FurA_(R.equi)	MQQGEHDFDPRAEALRGAGLRVTAPRVAVLNT -VAANPHSDADQVATEVRRQLGSVSTQAV	59
FurA_(M.tuberculosis)	---MSSIPDYAEQLRTADLRVTRPRVAVLEA -VNAHPHADTETIFGAVRFALPDVSRQAV	56
	: * : : : * * * * : * : * : : : : : : * : *	
PerR_(B.subtilis)	YNNLRVRFRESGLVKELTYGDASSRFD FVT -SDHYHAI CENCGKIVDFHYPGLDEVEQLAA	118
FurA_(R.equi)	YDVLGACVRVGLLRRIEPAGSPARYE TRTADNHHHLVCRSGTVDVDCVGHAPCLEPS	119
FurA_(M.tuberculosis)	YDVLHALTAAGLVKIQPSGVSARYE SRVGDNHHHIVCRSGVIADVDCAVGEAPCLTAS	116
	* : * . * : : : : : : : * : : . . : * : * : * : * : * : * : * : * : * : *	
PerR_(B.subtilis)	HVTGFKVSHHRLIYGVQEQESKKNH-----	145
FurA_(R.equi)	SNHGFIDEAEVFWGLCPDRHDSAKTGAQSVTSSQNQDDVPGSGGSITSKTATARQEQ	179
FurA_(M.tuberculosis)	DHNGFLLDEAEVIYWGLPDSISDTSRSHP-----	147
	** : . . : : * : * : * : *	
PerR_(B.subtilis)	-----	145
FurA_(R.equi)	GGFAQ	184
FurA_(M.tuberculosis)	-----	147

Figure 4-42 Sequence alignment between PerR and two variations of FurA proteins from *B. subtilis*, *R. equi* and *M. tuberculosis* respectively.

Metal binding sites identified for PerR were annotated using relevant binding domain analyses from Figure 4-41 and in-text analysis of the regulatory binding site that remains in apo-form. FurA binding sites for *R. equi* were extrapolated accordingly and cross compared with FurA from *M. tuberculosis*. Metal binding coordinating groups are clustered by colour. Green highlighted text indicates ligand groups for the Zn(II) structural domain. Yellow highlighted text indicates ligand groups for metal-ion regulatory binding site. Red text highlights amino acid variation.

Analysis of the structural metal binding site in PerR identified a Zn(II) atom in a tetrahedral coordination configuration using a canonical $(CXaa_2C)_2$ motif to stabilise the dimeric protein by ‘locking in’ three β -strands of each PerR monomer, to form the two sections of the dimer β -sheet (Traoré et al., 2006). As such, the metal ion (in this case, zinc (II)) is coordinated using the sulphhydryl side-chain ligands of Cys₉₆, Cys₉₉ and Cys₁₃₆, Cys₁₃₉ (Figure 4-43).

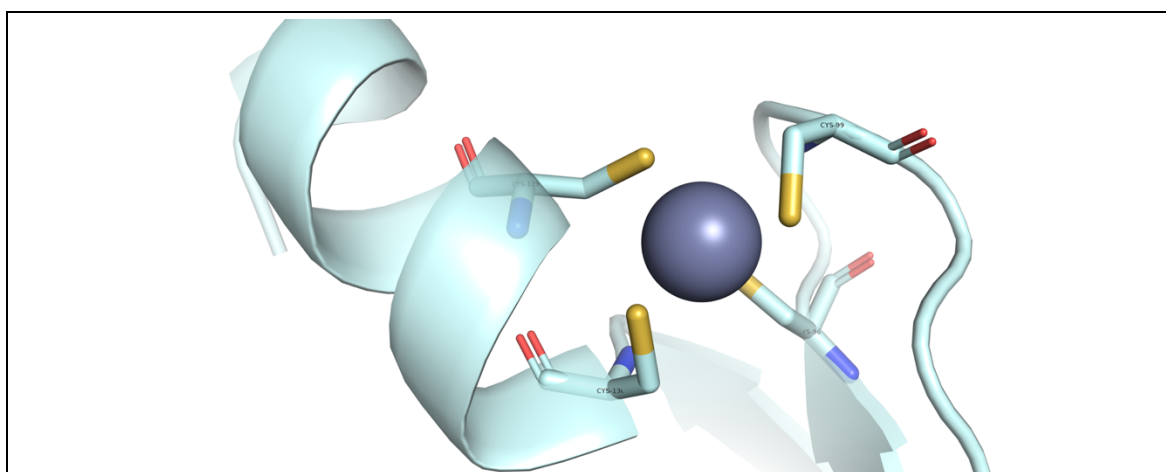


Figure 4-43: Structural metal binding site of PerR (2FE3) from *B. subtilis* occupied by a Zn (II) atom.

As previously alluded with discussion of the sequence alignments in Figure 4-42, it is extremely plausible that FurA from *R. equi* utilises an analogous structural zinc binding domain coordinated using the sulphhydryl side-chain ligands of Cys₉₇, Cys₁₀₀ and Cys₁₃₇, Cys₁₄₀ (Figure 4-44), and

presumably will function in dimer stabilisation through β -sheet formation.

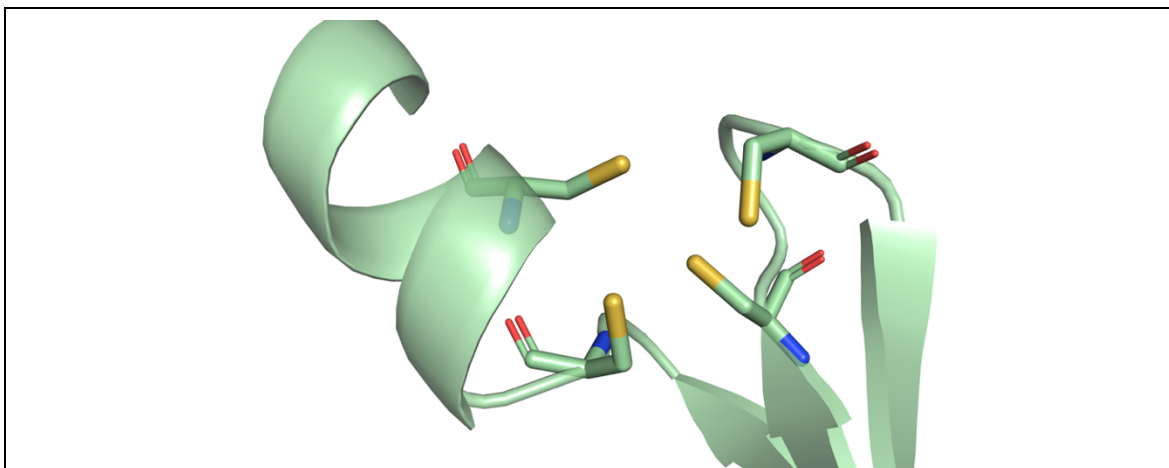


Figure 4-44: Predicted structural metal binding site of the homology-based model of FurA from *R. equi* in apo-form.

Analysis of the PerR crystal structure 2FE3 recognised the regulatory metal binding site in apo-form (Figure 4-45). Traoré et al. (2006) identified that a transition metal may bind in a pentavalent coordination configuration, likely by means of distorted trigonal bipyramidal geometry similar to that described for the primary IdeR metal binding site earlier.

Critically, the ion of the holo-PerR protein form appears to be coordinated using a *C*-terminal ligand His₃₇ and the *N*-terminal ligands His₉₁, His₉₃, Asp₈₅ and Asp₁₀₄, and all 5 of these residues are well-conserved within the PerR regulatory family to facilitate a high-affinity metal binding site and are essential for repressor functionality *in vivo* for Fe(II) or Mn(II) cofactors (Lee and Helmann, 2006). Pivotaly, PerR activation occurs by wrapping around the metal ion connecting the *N*- and *C*- termini to facilitate DNA binding.

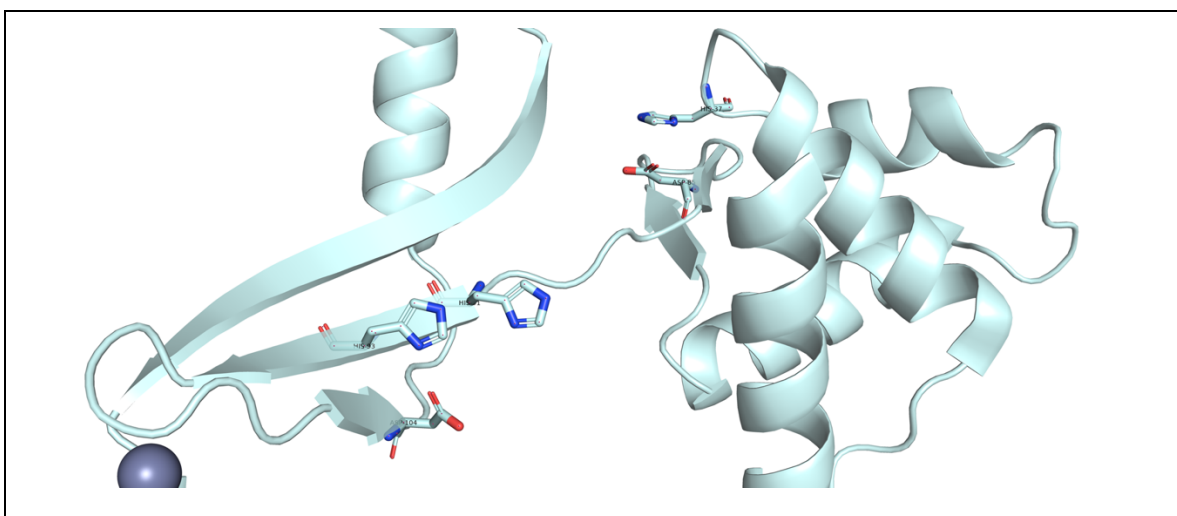


Figure 4-45: The regulatory metal binding site of PerR (2FE3) from *B. subtilis* in apo-form.

Analysis of the metal binding site in the putative FurA model from *R. equi* implies that metal coordination is achieved in a comparable mechanism of pentavalent coordination utilising protein side chain-ligands His₃₇, Glu₈₅, His₉₂, His₉₄ and Asp₁₀₅ (Figure 4-46).

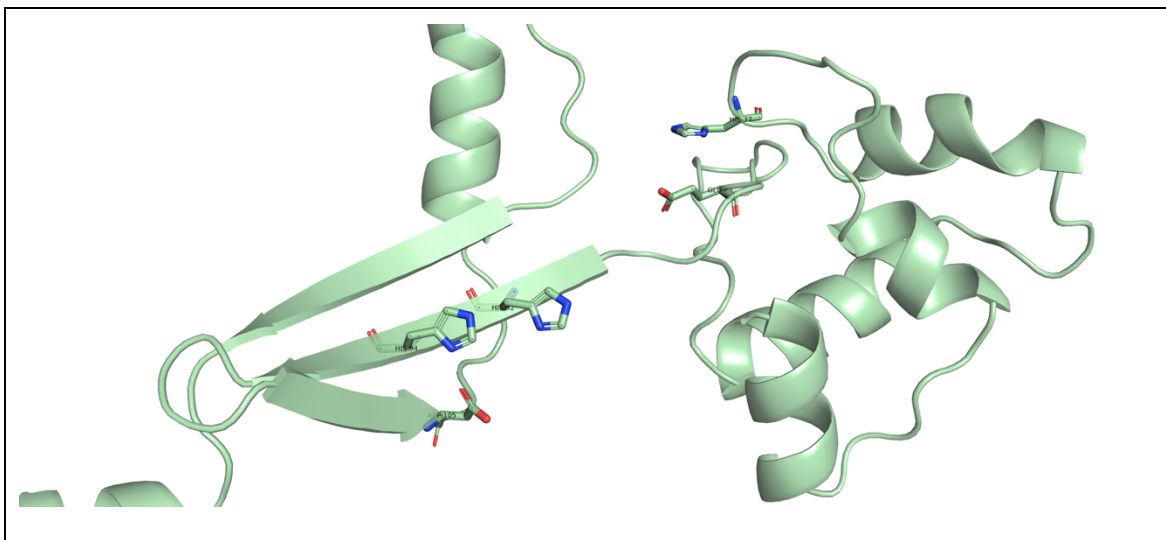


Figure 4-46: Predicted regulatory metal binding site of the homology based model of FurA from *R. equi* in apo-form.

Interestingly, an amino acid substitution is observed for metal-binding pocket ligand Glu₈₅; in both the FurA sequences of *R. equi* and *M. tuberculosis* an Asp residue, likely a result of the phylogenetic divergence. Substitution of Asp with Glu is an frequently observed conservative replacement, due to only having an Epstein's coefficient difference of 0.03 (Epstein, 1967); of which the insignificant variance is a result of an amino group present *in lieu* of one of the oxygen atoms, thereby avoiding induction of a negative charge.

More significantly though, the *C*-terminal ligand His₃₇, which is believed to permit DNA-binding by aligning the *C*- and *N*- termini when the metal ion is bound creating the binding pocket, is conserved in the *R. equi* FurA model. It is therefore conceivable that mutation of the His₃₇ residue would prevent activation and subsequent DNA binding.

4.6.6 Prediction of FurB structure

The FurB amino acid sequence was submitted to the Phyre2 protein-modelling tool (Kelley et al., 2015b). The output produced a hypothetical protein structure (100 % residues modelled at >90 % confidence) using the crystal structure **2o03-A** (zinc uptake regulator from *M. tuberculosis*, chain A, resolution: 2.70 Å). The template was ranked 7th based on raw alignment score, taking into consideration both sequence and secondary structure similarities. The probability of sequence-template homology is given as 100 % confidence, and percentage identity between the sequence and template is 69 % suggesting a high model accuracy and was selected over then other higher-ranking alignments due to their significantly lower sequence identity conservation.

The *R. equi* FurB model was imported to the programme Coot v0.8.9.1 (CCP4i2 shell) and superposed based upon protein topology using SSM superposition to the 2o03-A for structural comparison, visualised using PyMol (Figure 4-47).

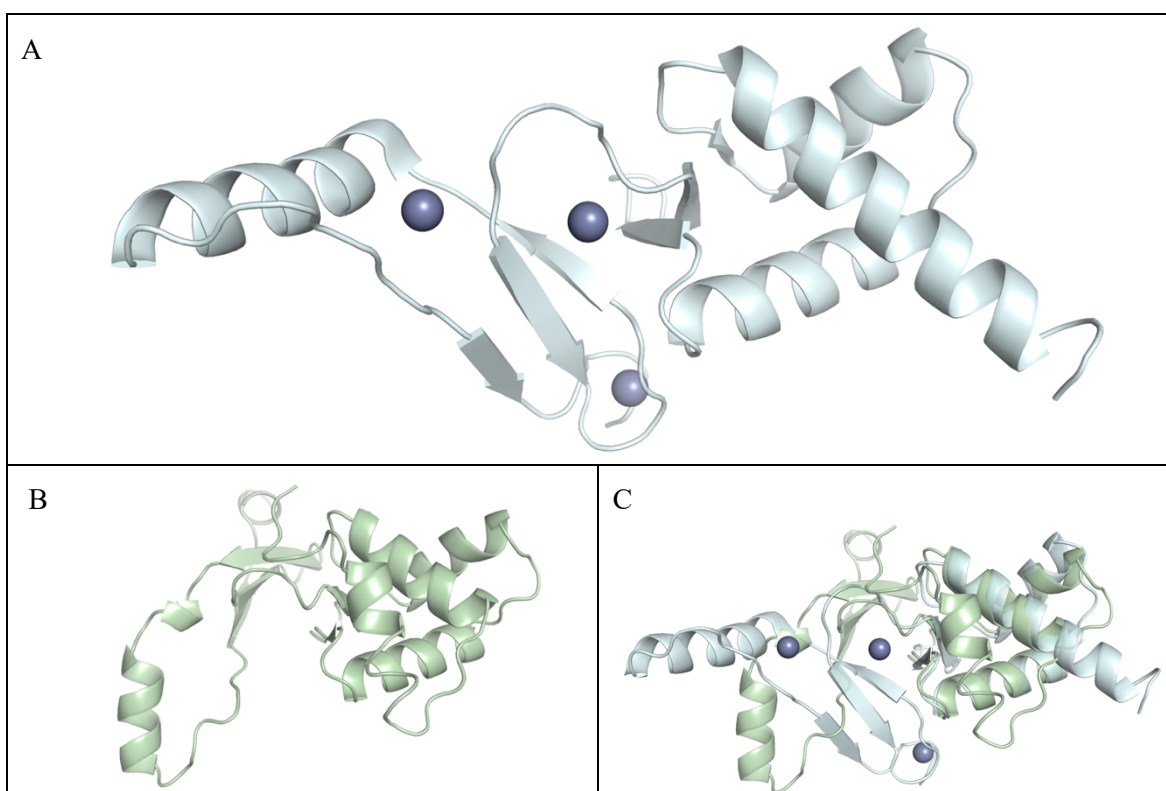


Figure 4-47: A comparative analysis of known and predicted FurB structures.

A) 2.70 Å crystal structure of wild type *M. tuberculosis* FurB complexed with Zn(II). B) Structure prediction of wildtype *R. equi* 103S FurB using Chain B of 2o03 as the template. Secondary structure elucidation predicted using Phyre2, aligned using Coot SSM supposition. C) Structural supposition of 2o03-B and FurB (*R. equi*)

When submitted to the DALI server (Holm, 2019), a pairwise alignment between 2o03-B and *R. equi* FurB model produced a Z-score of 9.6 and a rmsd value of 3.3.

The crystal structure 2o03 of *M. tuberculosis* FurB identified a flexible two-domain arrangement, comprised of an *N*-terminal canonical three winged helix and two-stranded antiparallel β -sheet DNA binding domain and the *C*-terminal dimerisation domain that contains three antiparallel β -strands and metal binding sites (Lucarelli et al., 2007). Interestingly, the overall fold of this *M. tuberculosis* FurB protein is reminiscent of Fur from *P. aeruginosa*, in respect of the domain architecture.

As the *in silico* model was predicted with high model accuracy, it is reasonable to hypothesise that the predicted metal coordinating sites of FurB from *R. equi* are reminiscent of those observed in FurB from *M. tuberculosis*; (pdb no:2o03; sequence chain view shown in Figure 4-48). Upon performing a sequence alignment (Figure 4-27), the three metal coordinating sites appear to be inherently conserved.

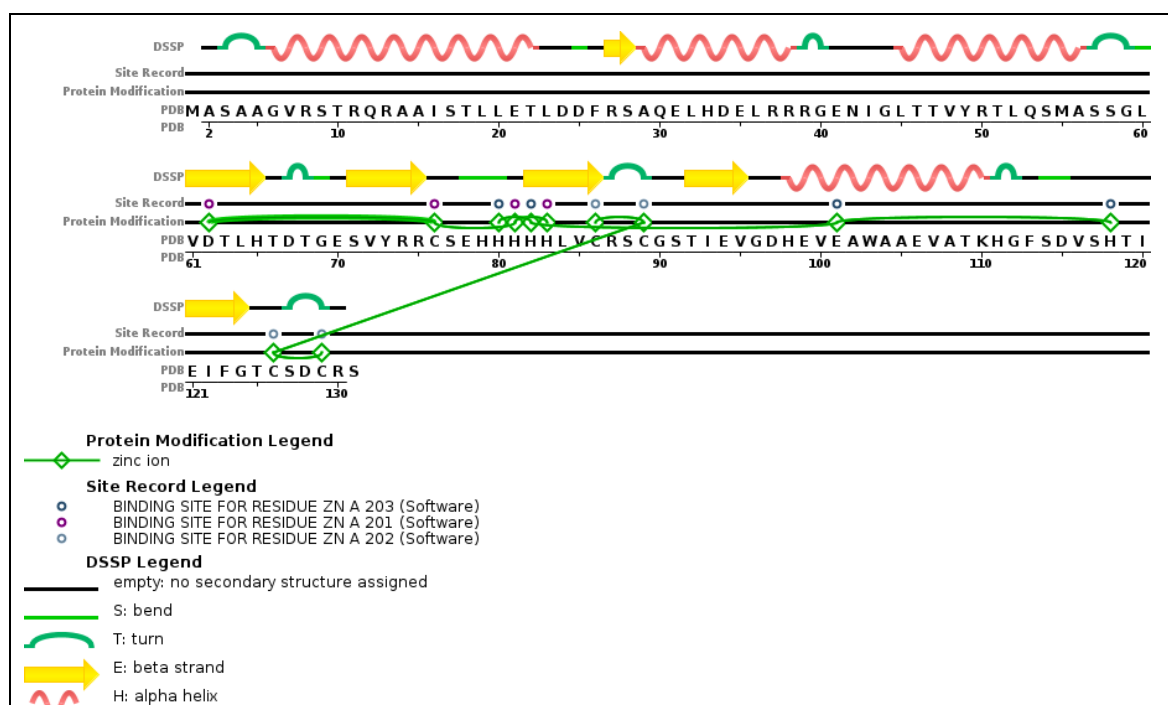


Figure 4-48: Chain Sequence View for the Entity PDB 2o03.

Protein secondary structure was identified by DSSP and variations are listed in the DSSP legend (Kabsch and Sander, 1983). Structural features including ligand binding sites are listed on by Site Record and Protein modification and variations are listed in the Legend (Golovin et al., 2005; Montecchi-Palazzi et al., 2008)

FurB(<i>R. equi</i>)	MTENVTDPTDQRRERAGRAVVGVRSTKQRS AISALLDDITEFRSAQELHDELRRRGQIGL	60
FurB(<i>M. tb</i>)	-----MSAAGVRSTRQRAAISTLLETLDLDFRSAQELHDELRRRGENIGL	44
	:. . *****. *. *. *. *. *. *. : : *****. *. *. *. *. *. *. *	
FurB(<i>R. equi</i>)	TTVYRTLQTLAEAGTVDVLRTDTGESVYRRCSSGHHHLLVCRAIGFTVEVDGPAVEQWSQ	120
FurB(<i>M. tb</i>)	TTVYRTLQSMASSGLVDTLHTDTGESVYRRCSEH HHHLLVCRSIGSTIEVDGHEVEAWAA	104
	*****. *. *. *. *. *. *. *****. *. *. *. *. *. *. *	
FurB(<i>R. equi</i>)	TIADTNGFTDVSHTVEIFGTCRDCAQAR	148
FurB(<i>M. tb</i>)	EVATKHGFSDVDHTIEIFGTCSDCRS--	130
	.* :. *. *. *****. ***** ** .	

Figure 4-49: Sequence alignment of FurB proteins from *Rhodococcus equi* (*R. equi*) and *Mycobacterium tuberculosis* (*M. tb*) respectively.

Metal binding sites identified for FurB were annotated using relevant binding domain analyses from Figure 4-48 and FurB binding sites were extrapolated accordingly. Metal binding coordinating groups are clustered by colour.

Analysis of the first metal binding site for FurB (2o03) identified a zinc(II) atom in a tetrahedral coordination configuration located in the hinge region between the dimerisation and the DNA-binding domain (Lucarelli et al., 2007). The metal ion is coordinated using protein side-chain ligands of the *N*-terminal DNA binding domain Asp62 and Cys76, as well as the *C*-terminal dimerisation domain ligands His₈₁, and His₈₃ (Figure 4-50).

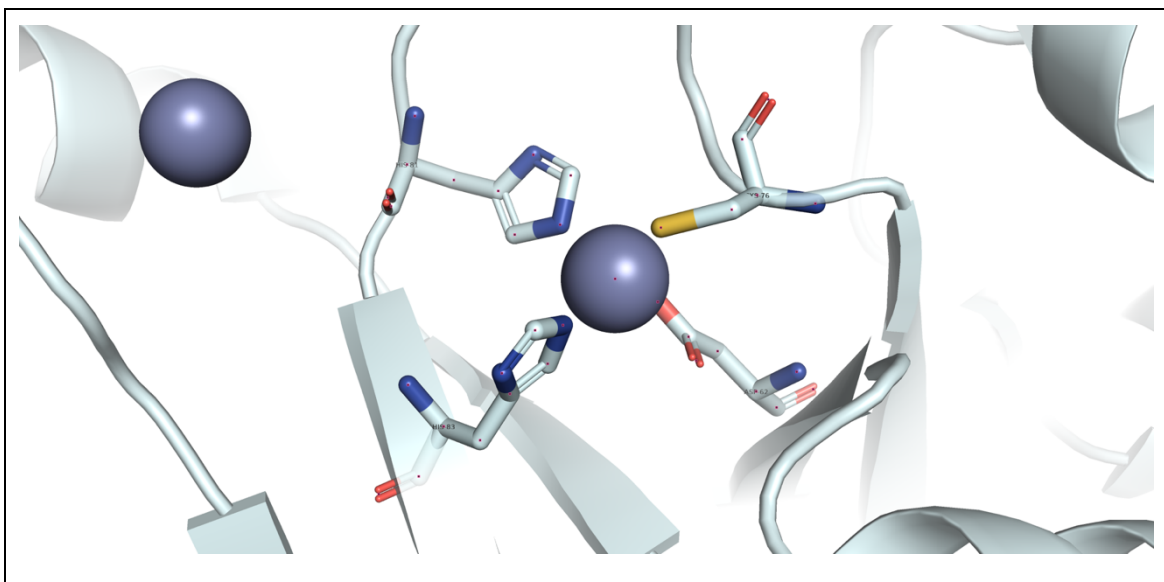


Figure 4-50 Metal binding site 1 of FurB 2o03 from *M. tuberculosis* occupied by a Zn(II) ion.

It is therefore logical to assume that this metal binding site is essential for protein activation, as the hinge is likely to provide the flexibility required to bind in a tetragonal coordination forming the binding pocket. The metal selectivity of each binding site is anticipated to be based around the coordination, with a preference for zinc arising from the preference for tetragonal geometry, further

confirmed by metal exchange with Co(II) ions that possess the same geometric binding preferences, unlike Fe(II) which typically adopts an octahedral coordination. Although the DNA-binding interactions of this protein are currently uncharacterised, it would be rational to hypothesise that protein activation occurs by wrapping around the metal ion connecting the *N*- and *C*- termini to facilitate DNA binding as previously identified for PerR from *B. subtilis* (Traoré et al., 2006).

Analysis of the first metal binding site for the protein FurB in *R. equi* appears to coordinate a metal ion in an analogous mechanism to the site observed earlier for the *M. tuberculosis* FurB utilising protein side chain-ligands Asp₇₇, Cys₉₁, His₉₆, and His₉₈ (Figure 4-51). However, as the FurB predicted model was constructed without integration of metal co-factors, the orientation of the two domains appears to be significantly relaxed as a result of the hinged binding pocket being unoccupied, and is further highlighted in the structural comparisons of Figure 4-47C . The flexion observed is likely to account for the high rmsd value identified for FurB due to the comparison of apo- and holo-protein homologues.

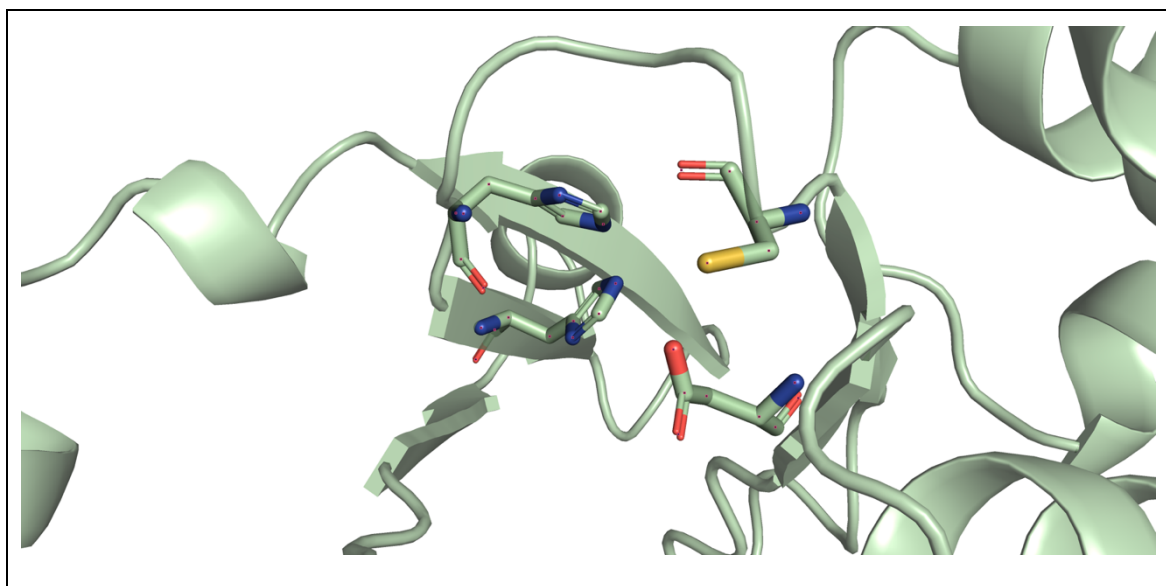


Figure 4-51 Putative metal binding site 1 of the homology-based model of FurB in *R. equi*, in apo-form.

Analysis of the secondary metal binding site in FurB identified a Zn(II) ion coordinated in a tetrahedral configuration using two canonical CXaa₂C motifs for protein stability (Lucarelli et al., 2007). The metal ion is coordinated using the sulphhydryl side-chain ligands of Cys₈₆, Cys₈₉ and Cys₁₂₆ and Cys₁₂₉ to connect the *N*- and *C*- termini (Figure 4-52).

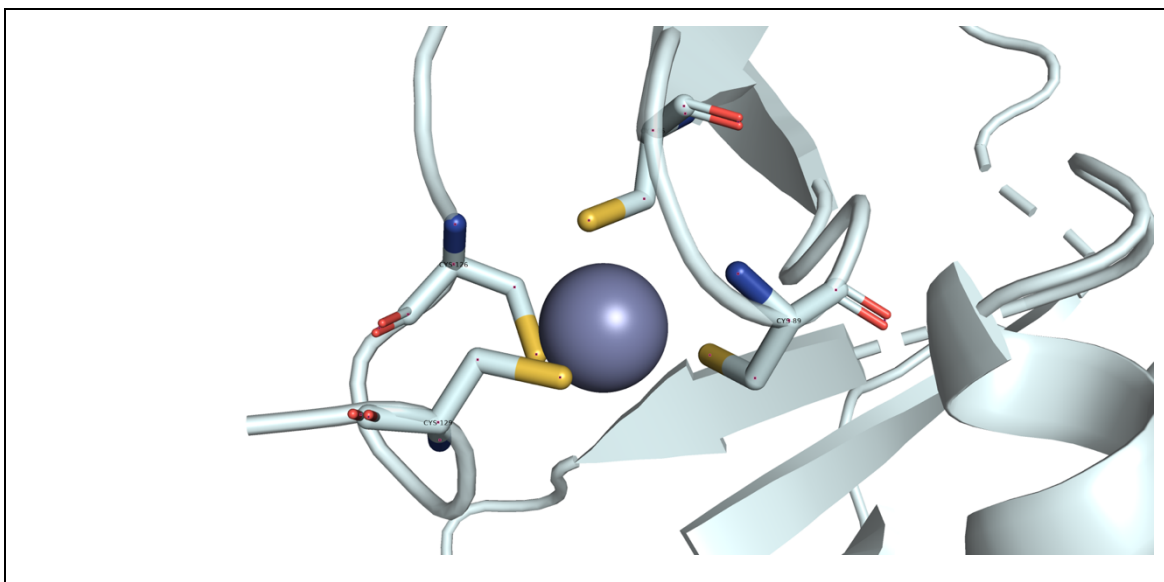


Figure 4-52: Metal binding site 2 of FurB 2o03 from *M. tuberculosis* occupied by a Zn(II) ion. Analysis of the secondary metal binding site in the putative FurB model from *R. equi* implies that metal coordination is achieved in a comparable mechanism of tetrahedral coordination utilising sulphhydryl side chain ligands Cys₁₀₁, Cys₁₀₄, Cys₁₄₀ and Cys₁₄₅ (Figure 4-53). Although this domain does not superpose with the FurB from crystal structure 2o03, the architecture is analogous, and variation is likely due to the comparison of apo- and holo- protein forms, therefore without co-factor binding in the primary metal binding site, the hinge region is not orientated appropriately. Given the sequence conservation between the FurB proteins, and the preservation of the CXXC binding motif, it is likely that this secondary binding site functions in protein stability.

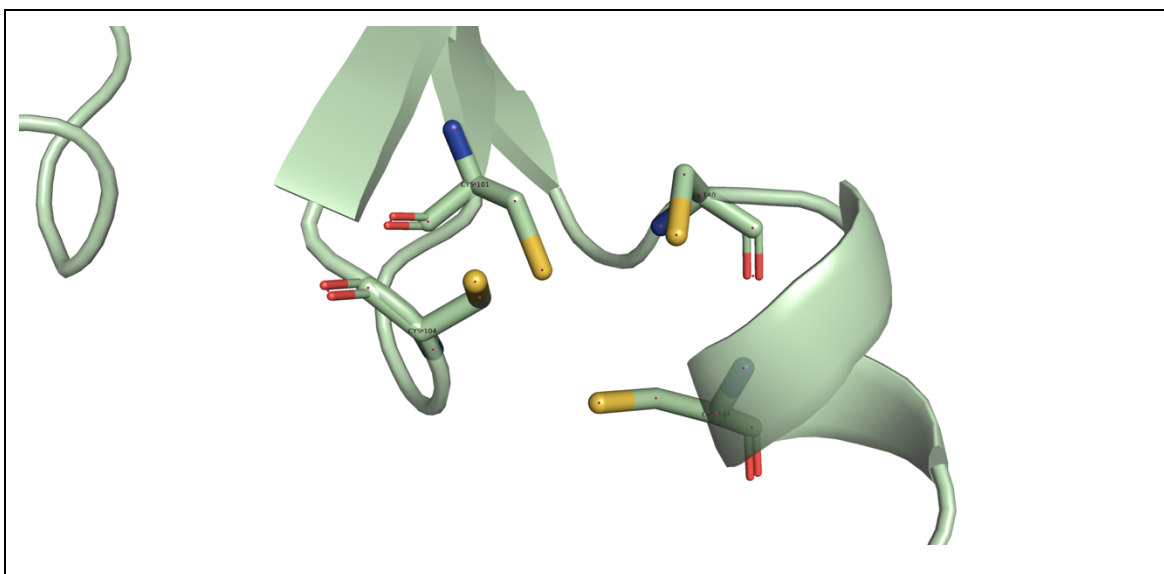


Figure 4-53: Putative metal binding site 2 of the homology-based model of FurB in *R. equi*, in apo-form.

The crystal structure 2o03 for FurB also was identified to contain third metal-coordinating site, where

an additional zinc ion was complexed using side chain ligands of three histidine residues (His80, His82 and His118) and a glutamate residue (Glu101). Concealed within the dimerisation domain, it is plausible that this metal binding site functions as an additional structural coordination site, supported by previously structural zinc binding site studies (Auld, 2001). Alternatively, the site and bound metal could be a by-product of the crystallisation process, but interestingly the conservation between *R. equi* and *M. tuberculosis* FurB sequences also includes this third metal binding pocket. Without further analysis we can only make inferences on these metal binding sites, and their significance.

4.7 Concluding remarks for this chapter

The initial aims of this chapter were to build on the bioinformatic analysis performed in Chapter 3 to ideally characterise the native metal-binding sites of the metalloregulatory proteins of interest.

The project initially started by taking the four metal regulatory genes of interest and designing constructs that could be used for heterologous gene expression in *E. coli* to facilitate physico-chemical binding analysis using ICP-OES.

Originally, protein purification was attempted by using heparin affinity chromatography to retain DNA-binding proteins, without success. Upon reflection this may have been unsuccessful due to having inactive apo-protein forms, as often co-factor binding can stabilise DNA-binding domain architecture.

The project was re-designed to incorporate a GST-fusion tag to facilitate a fast, convenient one-step purification of the metalloregulatory proteins of interest. This process was achieved for IdeR-GST, for both heterologous protein production and subsequent purification, however the other three proteins require further expression optimisation.

Due to a change in circumstances at this stage, the research project aims were altered to focus on -omic analysis of *R. equi* under iron limitation (consequently discussed later in the thesis), and mutagenic studies of the metalloregulatory genes. To provide suitable project closure for this chapter, a compromise of performing predictive analyses of the putative metal regulators of *R. equi* 103S by *in silico* protein scaffolding, providing an indication to the metallo-preference for the binding sites

of interest based upon homology studies with other closely related structures.

The *in-silico* predictions for the *R. equi* metalloregulatory proteins show homology to proteins responsible for different transition metal coordination. IdeR does appear to coordinate Fe(II) ions based on the octahedral coordination geometry, with a very significant sequence similarity to the IdeR protein from *M. tuberculosis*. These results also corroborate with the findings of the genomic survey and the Iron-box sequences identified in the operator sequence of iron-responsive genes including siderophore biosynthesis, transport and iron storage genes, which are also prevalent in mycobacterial species.

The *in-silico* prediction for DtxR however appears to coordinate Mn(II) ions based upon the conservation of the liganding residues and high overall sequence and structure conservation to that of MntR in *M. tuberculosis*.

Canonically, Fur proteins are characterised as global iron regulators, however, as IdeR has been predicted to function in this capacity, it seems redundant to have Fur proteins function in the same manner. The *R. equi* FurA protein sequence has significant homology to the FurA from *M. tuberculosis* and PerR in *B. subtilis*, both Gram-positive organisms, which unlike Gram-negative bacteria utilise DtxR-family proteins to regulate intracellular iron. Both of the Gram-positive homologues appear to function as catalase/peroxide repressors; *R. equi furA* is located in a gene cluster with a catalase gene which would be consistent with a function as a metal-dependent peroxide sensor accordingly. This is further explored in subsequent mutagenic studies later in the thesis.

R. equi FurB has significant sequence homology to the FurB protein of *M. tuberculosis* with the conservation of the coordination ligands in a tetragonal geometric pattern characteristically favourable to zinc ions, supported by the crystal structure 2o03 with multiple zinc ions bound. Furthermore, it has been previously identified that the *furB* gene is co-transcribed with an additional zinc-dependant regulator, this appears to also be preserved in the rhodococcal genome as *furB* shares a predicted operon with REQ_29120. This additionally co-transcribed gene share 73 % sequence homology with the mycobacterial co-transcribed equivalent Rv2358, therefore it is logical to presume that *furB* operates as a zinc uptake regulator.

As previously alluded, it would be biologically redundant to maintain a multitude of genes encoding

proteins that have the same function, given that the gene annotations were performed autonomously, the analyses performed in this chapter develop the working hypothesis for the complex divalent metal regulatory systems employed by *Rhodococcus equi* as shown in Figure 4-54, that assigns regulatory roles for multiple divalent cations based upon homologous metal-binding capacities. Specifically, the protein IdeR can be assigned a role for regulating the response to iron starvation, and the protein FurA can be assigned a role for regulating oxidative stress. The other candidate proteins DtxR and FurB are likely to be involved with coordination with manganese and zinc respectively.

Going forward, this in-silico driven hypothesis could be advanced by completion of the IdeR metallo-preference by ICP-OES as well as optimisation of the expression and the subsequent binding analysis for the other proteins. It would also be interesting to perform crystallisation studies on the four regulatory proteins and compare between the in-silico scaffolding models produced here; and the associated homologous protein templates used. Furthermore, site-directed mutagenesis on the key liganding residues of these metallo-regulatory proteins could identify if either metal-coordination or ultimately DNA-binding can be impaired, thereby altering the characteristics of metal homeostasis in the bacterium.

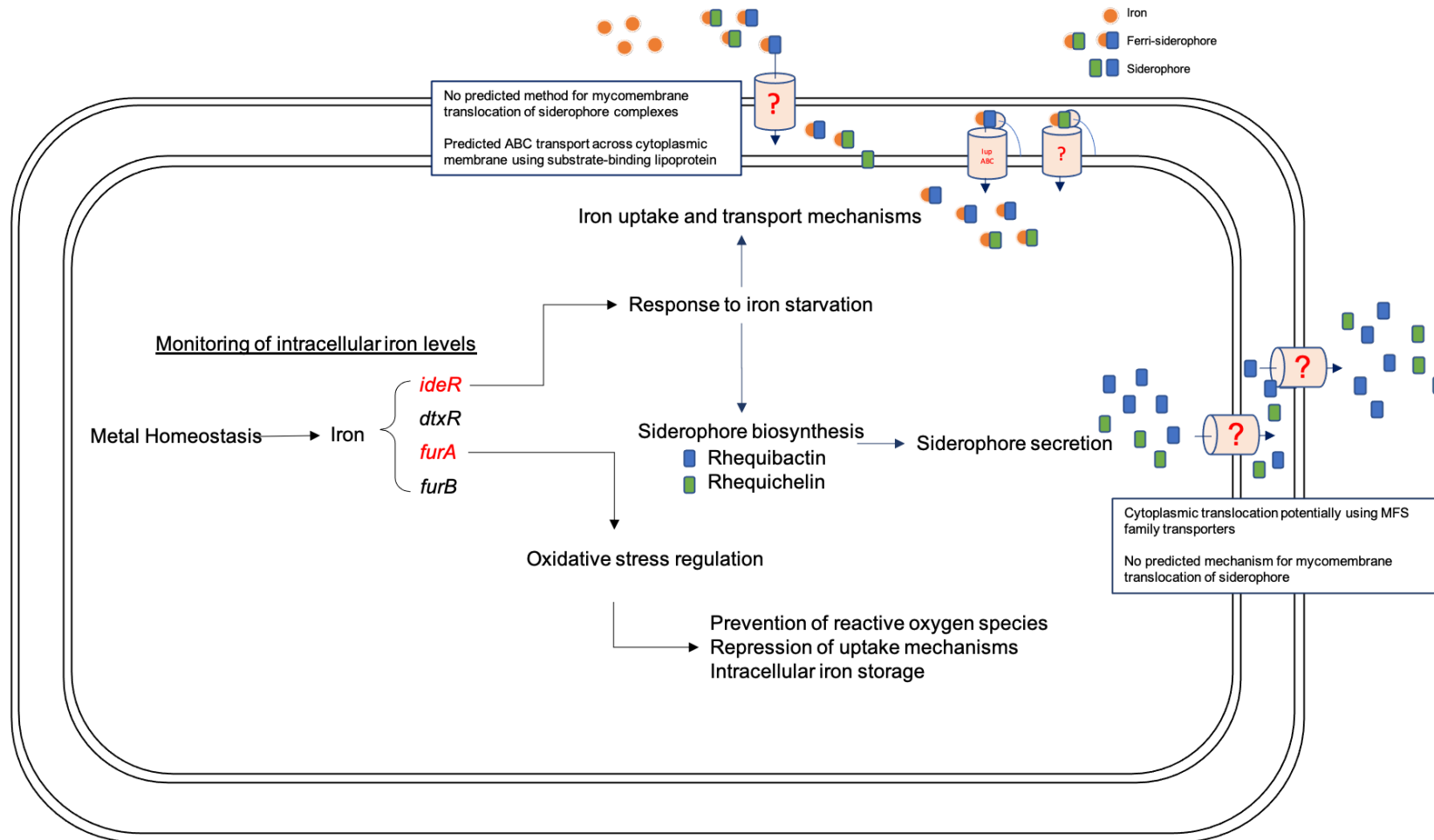


Figure 4-54: The updated *R. equi* iron homeostasis network map (after chapter 4)

Data generated from the preliminary genomic survey and comparison of *in silico* metalloregulatory protein metal binding capacities. Here functional candidates are shown in red that warrant further investigation of the predicted roles.

5 Generation and *in vitro* characterisation of
Rhodococcus equi 103S mutants in four putative
metal-dependent transcriptional regulators: *ideR*,
dtxR, *furA* and *furB*.

5.1 The construction of unmarked in-frame deletion mutants in *Rhodococcus equi*

A strategy was established to investigate the regulatory influence of each of the four metal-dependent transcription regulators *via* generation of unmarked in-frame gene deletion using the pSelAct suicide vector developed by van der Geize *et al.* (2008). The system produces isogenic mutants varying by carefully defined nucleotide deletions without alteration of circumambient genes or their expression. Mutagenesis is achieved by constructing a vastly truncated ORF within pSelAct flanked by 1.5kb of DNA directly copied from the chromosomal context of the modified gene in both the upstream and downstream directions. Ultimately this truncated DNA sequence replaces the native DNA sequence in the chromosome. These pSelAct constructs are assembled and replicated within *E. coli* (methodology described in Figure 5-1).

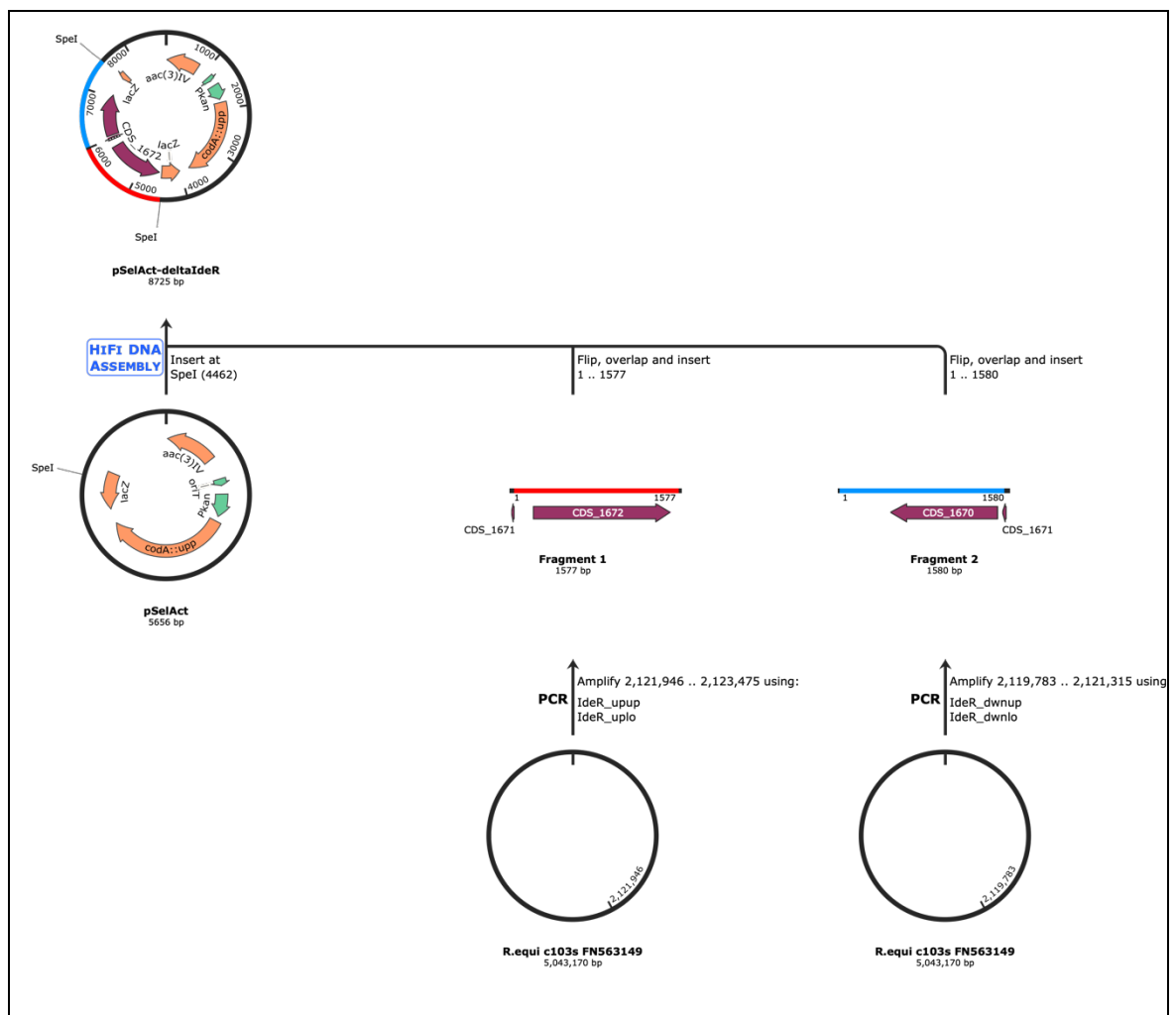
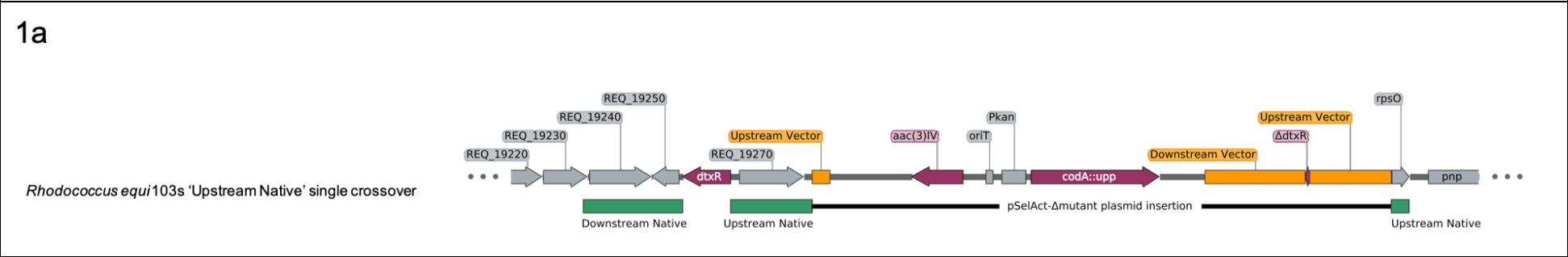
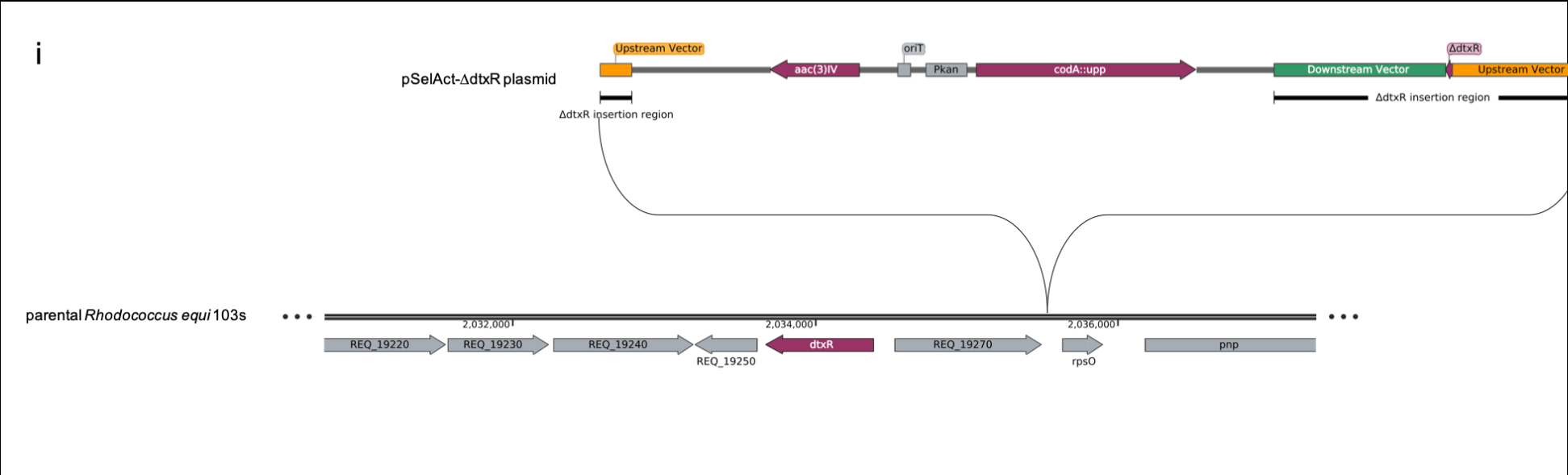


Figure 5-1: Gibson assembly style methodology utilised to produce gene mutations in the vector pSelAct.

The pSelAct plasmid lacks a valid origin of replication for *Rhodococcus* and is unable to replicate within these bacteria. Stable transformation of rhodococci with the pSelAct-based deletion constructs will produce clones that can be selectively identified based on apramycin resistance encoded by *AAC(3)-IV* (van der Geize et al., 2008), as a result of homologous recombination in either of the 1.5kb upstream or downstream regions relative to the gene of interest that are defined during plasmid assembly (Figure 5-2(1a)).

5.1.1 Counter selection of double-crossover mutants by concomitant expression of cytosine deaminase and uracil phosphoribosyltransferase

The generation of the mutant requires two independent homologous recombination events to take place; one within each of the two DNA sequences defined by the upstream and downstream PCRs used in constructing the mutagenic plasmid (Figure 5-2(1b)). Growing the strain carrying the integrated vector without selective pressure allows a further round of homologous recombination to occur. Cells undergoing such an event will lose the integrated plasmid, which also contains a counter-selection mechanism based on the metabolic activation of 5-fluorocytosine (5-FC). The counter-selective agent 5-FC is not toxic to *R. equi* but, when acted upon by cytosine deaminase (CD) and uracil phosphoribosyltransferase (UPRT) enzymes, the 5-FC is converted to 5-fluoro-deoxyuridiny monophosphate which exerts a strong cytotoxic effect (van der Geize et al., 2008). Here 5-FC permits only the growth of bacteria that have lost the pSelAct vector and its counter selection gene *coda::upp* encoding the fusion protein of CD/UPRT. If the second homologous combination happens in the same flanking region as the insertional event, then a wild-type organism is regenerated. However, the occurrence of the second homologous recombination in the opposite flank generates the gene deletion, unmarked through simultaneous *AAC(3)-IV* loss, conveying apramycin sensitivity. The nature of counter-selection survivors was ascertained through a diagnostic PCR to distinguish between the native gene and the truncate generated by the mutagenesis process (Figure 5-2(2a/2b)).



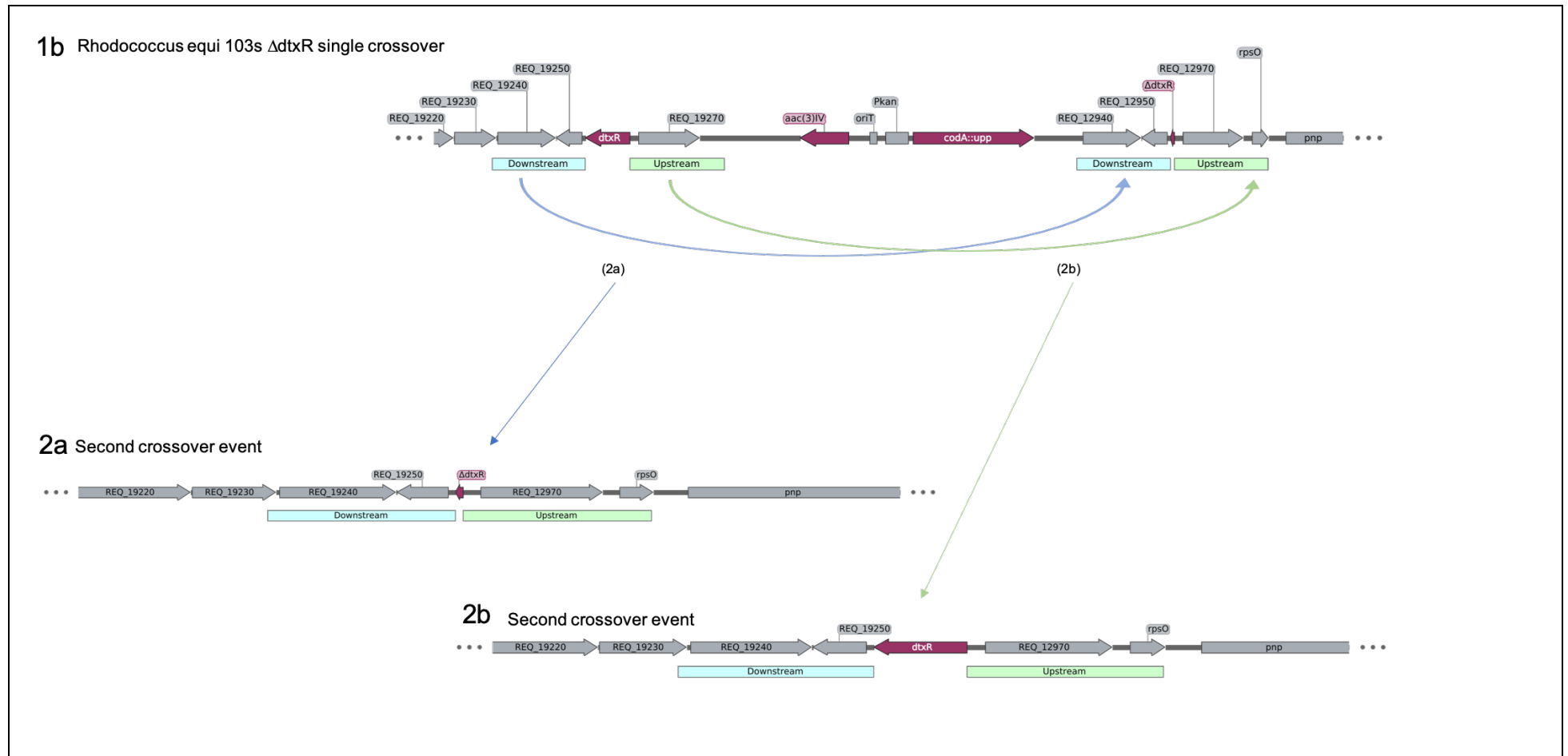


Figure 5-2: An example of homologous recombination to incorporate an in-frame silent gene mutation in *Rhodococcus equi* using the vector pSelAct.
 i) Incorporation of the pSelAct-mutant containing plasmid into the *R. equi* 103S chromosome. 1a) A representation of a single-crossover homologous recombination event. 1b) Identification of the two potential secondary crossover homologous recombination steps that may occur. 2a) an unmarked in-frame deletion mutant. 2b) A wildtype gene revertant.

5.2 Chapter aims

As previously alluded, the main objective of this study is to construct in-frame deletion mutants for the iron-responsive regulatory proteins in *R. equi* 103S, IdeR, DtxR, FurA and FurB that were previously considered in the genomic survey and the subsequent *in silico* metal binding analysis in Chapters 3 and 4 respectively. Production of aforementioned mutants will ideally help further define the regulatory mechanisms associated with divalent metal cations in this understudied veterinary pathogen, through the identification of any phenotypic changes, or subsequent effect in gene transcription.

5.3 Generation of mutagenic constructs

Gene knockout constructs were produced by ligation of upstream and downstream PCR products using a Gibson assembly style methodology. The PCR products were produced by amplification of an overlapping region (10 in-frame codons and stop codon) and 1.5 kb either side of the target gene. Both the start of the upstream and the end of the downstream product contained additional nucleotides for a *SmaI* restriction enzyme site to facilitate pSelAct vector ligation.

5.3.1 Generation of mutagenic gene amplicons *via* PCR

The PCR products required to form the mutagenic constructs were produced by 2-step PCR using standard conditions for Q5 polymerase and oligonucleotide primers listed in sections 2.3.3 and 2.4.6 respectively. All of the amplicons were successfully produced with exception to *ideR*-upstream and *furA*-downstream. The amplification of *ideR*-upstream was subsequently achieved using a longer PCR extension period (Figure 5-3).

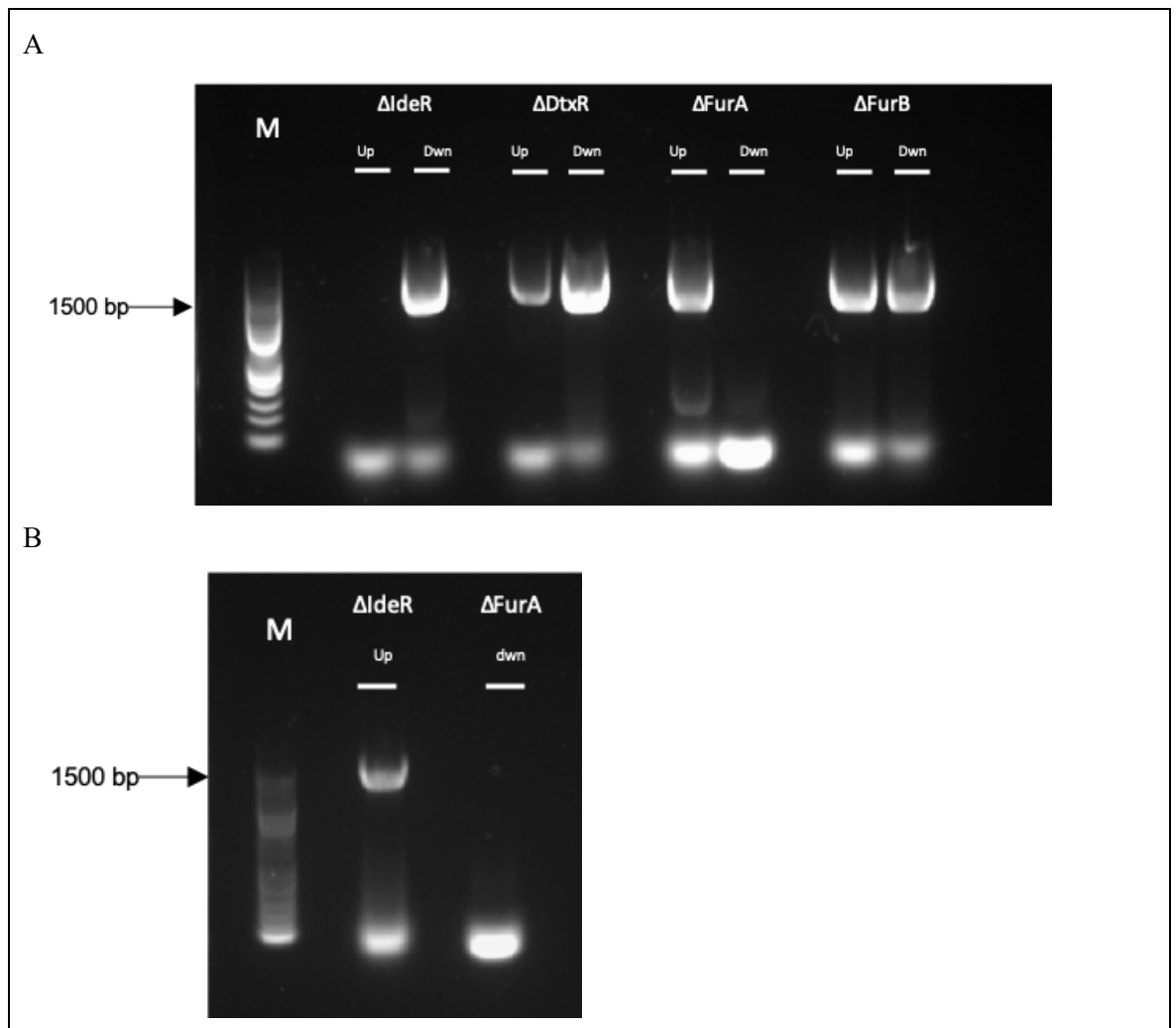


Figure 5-3: PCR products generated for *R. equi* 103S mutagenesis.

All product sizes were expected to be 1500 bp. A) The PCR reaction featured two-step cycling without a primer annealing temperature, using NEB Q5 Polymerase. B) The PCR reaction featured an extended 72 °C primer extension temperature, using NEB Q5 Polymerase. Samples were analysed by electrophoresis on a TAE-buffered 1.0 % agarose gel and their migration compared to M [NEB 100 bp molecular marker].

5.3.1.1 Alternative *furA*-downstream design

The *furA*-downstream oligonucleotide primer was re-designed due to many unsuccessful PCR optimisation attempts. The failure to produce this amplicon was hypothesised to be the result of either a hairpin structure formation or a potential self-dimer within the *furA*-dwnlo primer. Consequently, the *furA*-dwnlo primer was modified to incorporate an extra 500 base pairs (section 2.4.6, highlighted grey), and PCR optimised (Figure 5-4).

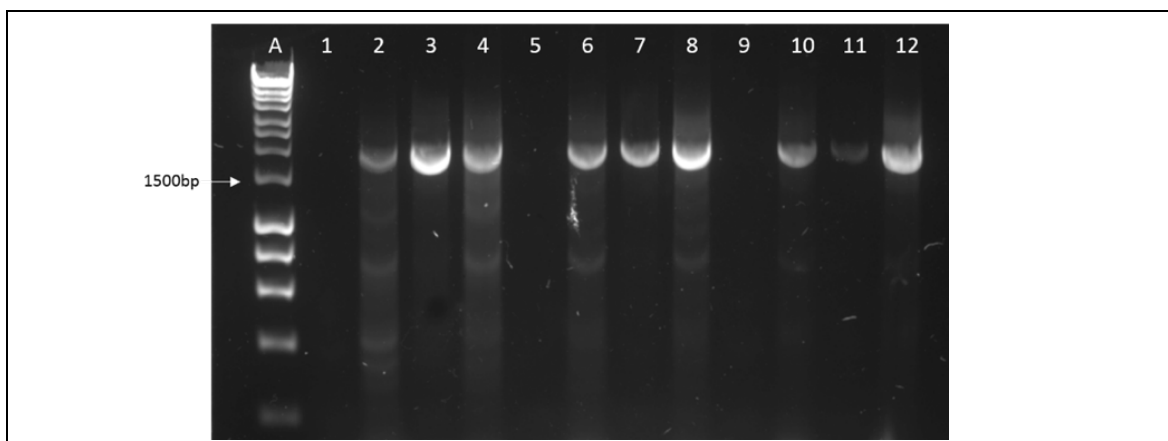


Figure 5-4 $\Delta FurA$ -downstream (2) optimisation

The PCR reaction featured three-step cycling using a range of primer annealing temperature, and supplementation with $MgCl_2$. Samples were analysed by electrophoresis on a TAE-buffered 1.0 % agarose gel and their migration compared to A [Hyperladder 1kb molecular marker].

5.4 Assembly of mutagenic constructs into pSelAct

5.4.1 Preparation of vector and PCR amplicons for assembly

The successful PCR amplicons and the linearised vector pSelAct were purified by gel extraction and drop dialysis methods (section 2.3) to remove any contaminants that may impair the assembly process. The purified vector and amplicons were analysed using agarose gel electrophoresis (Figure 5-5). The respective DNA was quantified by analysis on a Nanodrop microvolume spectrophotometer.

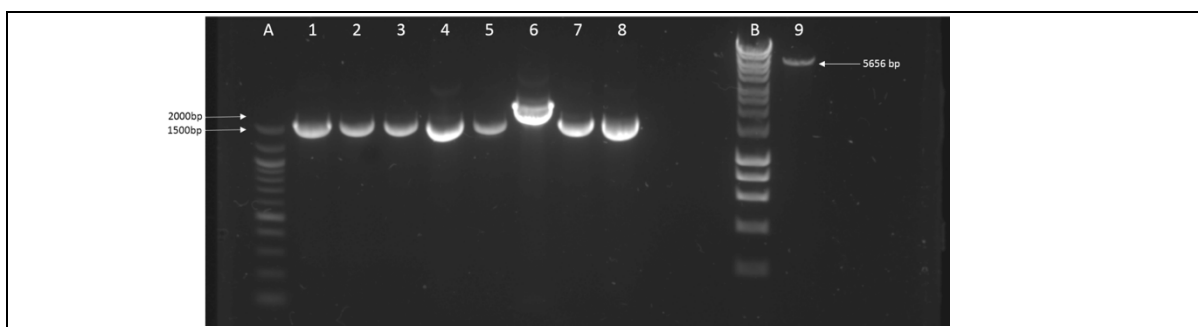


Figure 5-5 Purified mutagenic amplicons and vector required for HiFi assembly.

A: 100 bp ladder (NEB), B 1kb ladder (Bioline). 1: $\Delta ideR$ -upstream, 2: $\Delta ideR$ -downstream, 3: $\Delta dtxR$ -upstream, 4: $\Delta dtxR$ -downstream, 5: $\Delta furA$ -upstream, 6: $\Delta furA$ -downstream(2), 7: $\Delta furB$ -upstream, 8: $\Delta furB$ -downstream, 9: pSelAct vector (SmaI cut). Samples were analysed by electrophoresis on a TAE-buffered 1.0 % agarose gel and their migration compared to A Hyperladder 1kb molecular marker.

5.4.2 Molecular cloning - NEBuilder HiFi assembly

The Gibson assembly style of plasmid construction was performed using a NEBuilder® HiFi DNA

Assembly kit to ligate the linearised vector and mutagenic PCR amplicons together as described in (section 2.7.3). The assembly mixture was subsequently transformed into competent *E. coli*®10G cells (Lucigen) and plated on antibiotic-containing agar plates. Four overnight cultures of each recombinant *E. coli* were setup, plasmid DNA was isolated using a Plasmid MiniPrep kit (QIAGEN) and DNA concentrations were also quantified by Nanodrop analysis.

5.4.3 *SpeI* digest of mutagenic constructs

As a diagnostic method of DNA analysis, the recombinant plasmids were digested with *SpeI* according the manufacturer's instruction to excise the 3/3.5 kb insert containing the truncate of the gene of interest, analysed by agarose gel electrophoresis (Figure 5-6) prior to confirmation by Sanger sequencing, These preliminary results indicated at least one successful assembly per mutagenic plasmid construct.

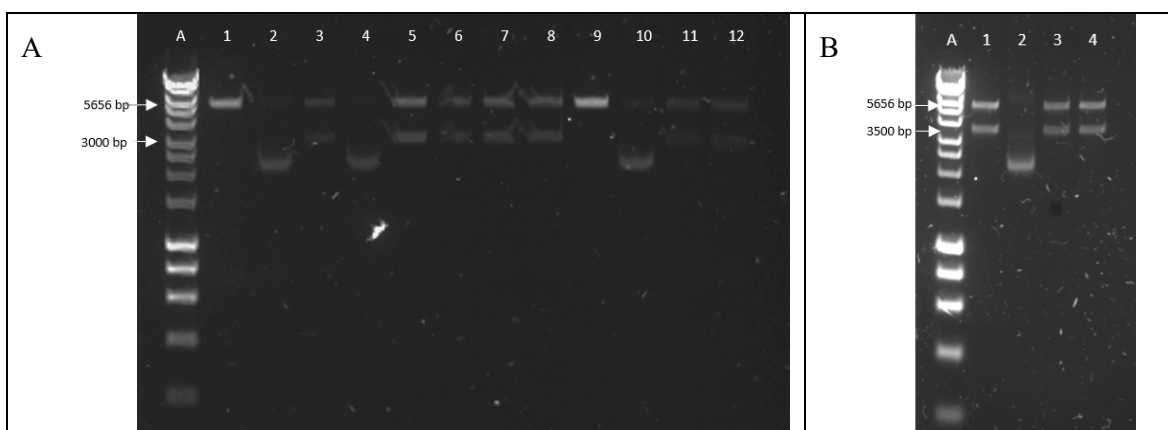


Figure 5-6: Diagnostic restriction digest of plasmids using *SpeI*

Successful digestion indicated in isolate 3 of pSelAct- Δ *ideR* (Gel A well 3), isolates 1:4 of pSelAct- Δ *dtxR* (Gel A wells 5-8), isolates 1, 3 and 4 of pSelAct- Δ *furA2.0* (Gel B wells 1, 3 and 4) and isolates 3 and 4 of pSelAct- Δ *furB* (Gel A well 11 and 12). Samples were analysed by electrophoresis on a TAE-buffered 1.0 % agarose gel and their migration compared to A Hyperladder 1kb molecular marker.

5.4.4 Sequencing of mutagenic constructs

To confirm successful assembly of the desired gene truncation into the pSelAct vector, 20 μ l of each plasmid previously used in the *SpeI* restriction digest was sent for Sanger sequencing (GATC biotech) against the pSelAct primers (section 2.4.4) which were designed to identify nucleotides present in the insertion region.

5.4.4.1 Sequencing results

All 16 plasmids that were used in the *SpeI* digest were sent for Sanger sequencing (GATC Biotech), using pSelAct primers, the chromatograms are summarised in Table 5-1.

Table 5-1 pSelAct- Δ mutant sequencing results

Table identified the plasmid isolates were sent, if sequencing was done, and/or reverse complement sequencing for confirmation and if subsequent transformation was performed.

Isolate	<i>ideR</i>				<i>dtxR</i>				<i>furA</i>				<i>furB</i>			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
pSelAct_for	vec	poor	✓	poor	✓	✓	✓	✓	✓	poor	✓	✓	vec	poor	✓	✓
pSelAct_rev	Nd	nd	✓	nd	✓	nd	nd	nd	✓	nd	nd	nd	nd	nd	✓	✓
Transform			✓		✓				✓						✓	
Complete seq			✓		✓				✓						✓	

5.5 Transfer of mutagenic constructs into *R. equi* 103S

The recombinant plasmids (pSelAct- Δ *ideR*, pSelAct- Δ *dtxR*, pSelAct- Δ *furA* and pSelAct- Δ *furB*) were used to transform *R. equi* 103S chromosomal DNA. Positive selection for single crossover transformants was based upon growth on 80 μ g/ μ l Apramycin LB agar plates. *R. equi* 103S without plasmid incorporation was used as a negative control due to a lack of apramycin resistance.

5.6 Selection of the double crossover 5-FC resistant phenotype

Single crossover transformants were grown without selection overnight, 100 μ l culture aliquots were serial diluted from 1:10 – 1:1000 using acetate minimal media (MMAc) adjusted to 10 g acetate per litre as recommended for increased selectivity (Van de Geize, R., personal communication) and plated onto MMAc agar supplemented with 100 μ g/ml 5-FC and incubated for 3 days at 37 °C.

When selecting for the 5-FC resistant phenotype, every colony was recovered on both non-selective media as well as 80 μ g/ μ l Apramycin LB agar plates indicating that all of colonies retained apramycin resistance. Several other groups had indicated that this counterselection method had proven difficult, even with acetate media amendments. Our interpretation of the vector synthesis (van der Geize et al., 2008), suggested that the crucial *codA::upp* gene fusion that converts 5-FC to 5-FU, the counter selective active, was associated with an *E. coli* derived promoter which might be poorly active in rhodococci. A strategy based on replacing this promoter with a strong constitutive rhodococcal promoter (*pgroES*) to drive production of the critical CD/UPRT fusion.

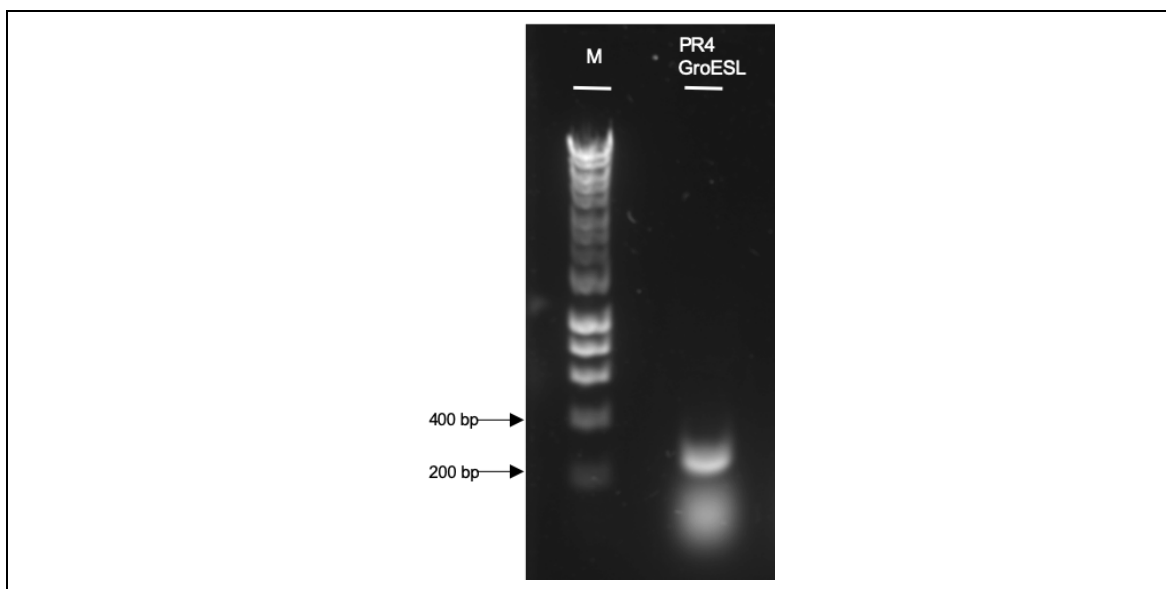


Figure 5-8: PCR amplicon for generation of PR4 *groESL*

The PCR was designed to amplify up to 2 bases away from the ATG start codon of the native gene and incorporates the CC dinucleotide required to introduce a *NcoI* restriction site, while preserving the distance between the ribosome binding site and the start codon. To confirm the promoter was amplified correctly, the PCR product was ligated into the shuttle vector pUC18 (as previously described), transformed into *E. Cloni* 10G competent cells and screened for α -complementation of β -galactosidase before being confirmed *via* Sanger sequencing (GATC Biotech) corroborated using NCBI Blastn alignment tool with *R. erythropolis* PR4 genome (data not shown).

5.7 pLongJon - Generation of a plasmid for improved 5-FC selection

5.7.1 NEBuilder HIFI assembly of pGroESL with pSelAct

To facilitate the incorporation of the newly designed pGroESL promoter, the vector pSelAct and mutant equivalents, were cut using the restriction enzymes *BstBI* and *NcoI* and combined together with pGroESL promoter amplicon using the NEBuilder® HiFi DNA Assembly kit. The assembly mixtures were subsequently transformed into competent *E. cloni*® 10G cells (Lucigen) and plated on antibiotic-containing agar plates as previously performed.

The new plasmid pLongJon (map shown in Figure 5-9) was produced to promote the recovery of mutants that have undergone the secondary homologous recombination step by improving protein expression of the downstream 5-FC selection marker. The pLongJon-mutant assembly is shown in Figure 5-10.

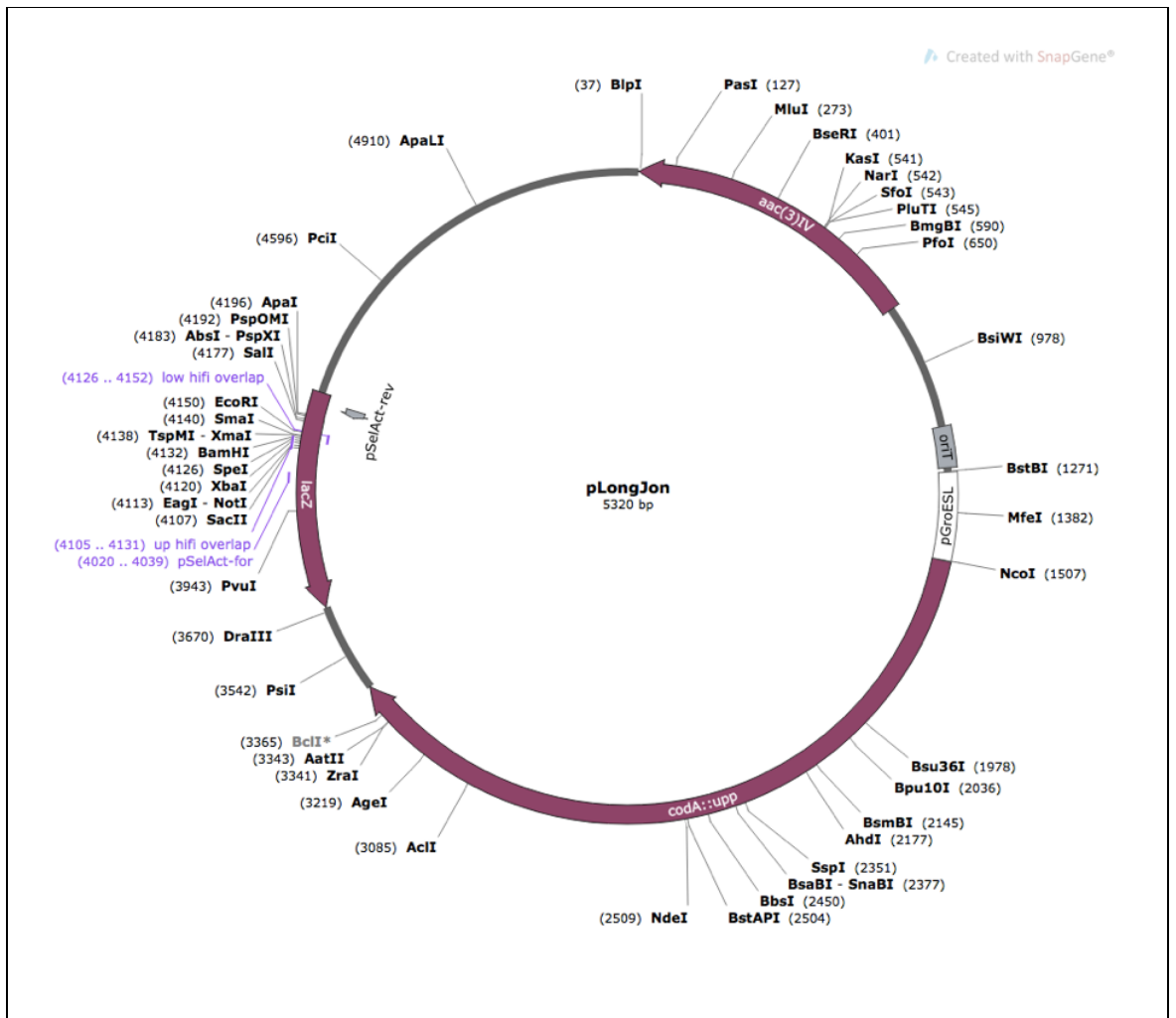


Figure 5-9: pLongJon vector map

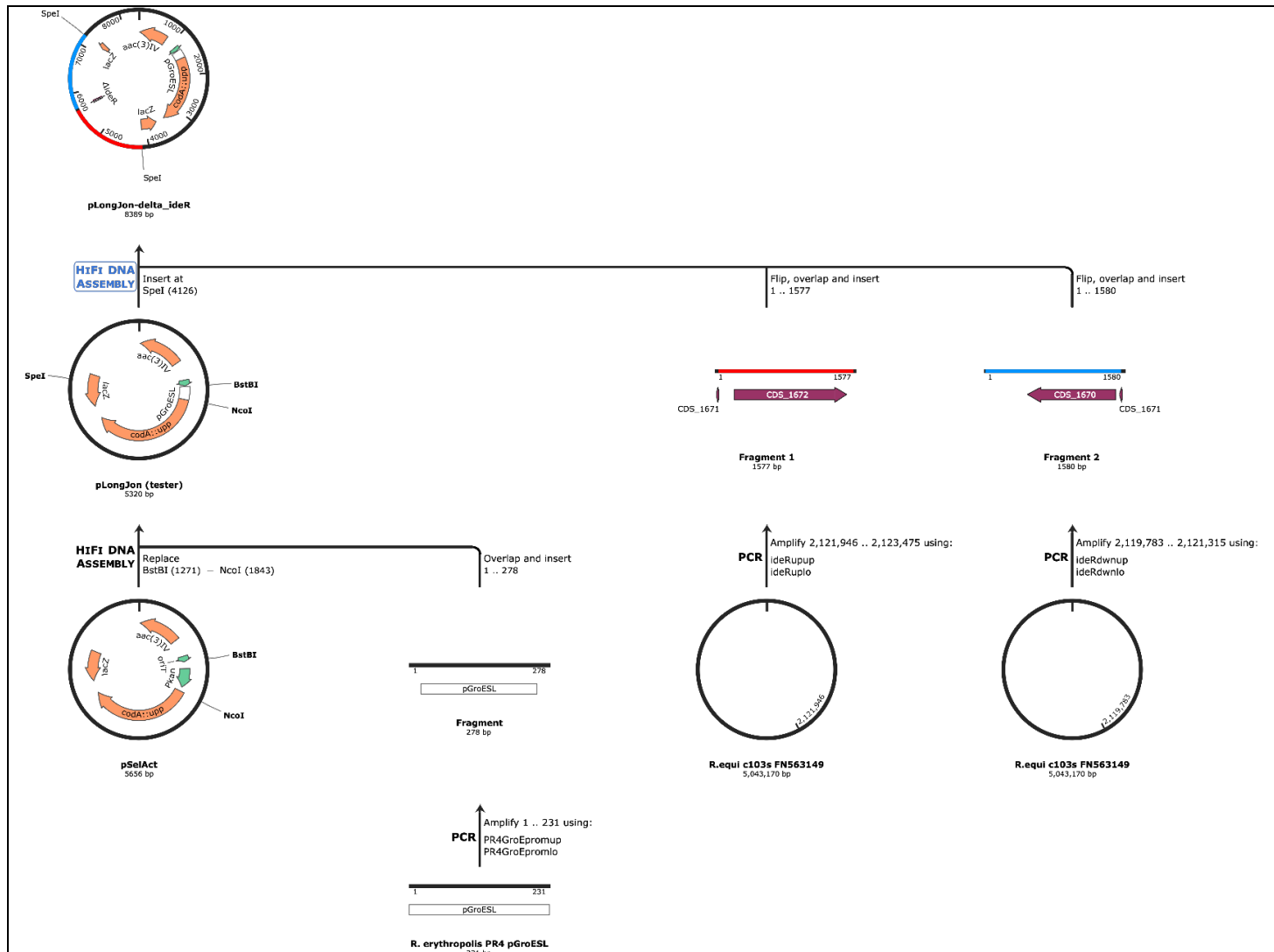


Figure 5-10 A flowchart detailing the molecular strategy employed to produce pLongJon deletion mutants

5.7.2 Confirmation of pLongJon- GroESL incorporation

5.7.2.1 Diagnostic colony PCR

Due to the potential for incorrect plasmid assembly, the mutagenic plasmids with the newly incorporated GroESL promoter were examined. The transformed clones in *E. coli* 10G cells were used as template DNA in a diagnostic PCR to amplify the PR4 GroESL promoter. The results indicated presence of the promoter in all but one of the plasmids as pLongJon-*ΔfurB* did not generate bands indicative of pGroESL incorporation (Figure 5-11), therefore to be time-efficient the project advanced using the other plasmids, the GroESL diagnostic amplicons ligated into pUC18 and sequenced for confirmation.

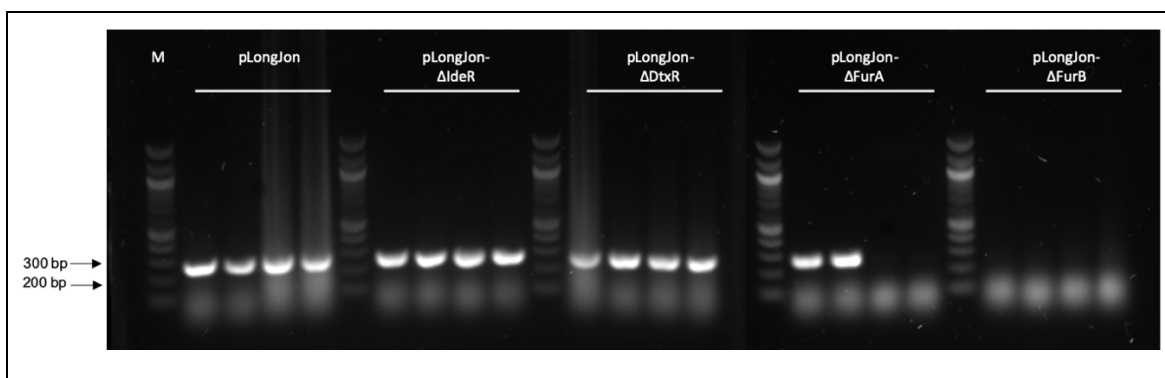


Figure 5-11: Colony PCR of pLongJon plasmids to identify incorporation of pGroESL promoter

5.7.2.2 pUC18-GroESL plasmid sequencing

The diagnostic pUC18-GroESL plasmids generated above were sequenced (Table 5-2) as previously described and queried using the Blastn alignment tool. The plasmids, if assembled correctly, were subsequently transformed into *R. equi* 103S as described for pSelAct above.

Table 5-2: pSelAct- Δ mutant sequencing results

showing the plasmid isolates sent, if sequencing was done, confirmation of Blastn results, and if subsequent transformation was performed. The sequencing primer used M13-FP is also listed.

	pLongJon				pLongJon- Δ ideR				pLongJon- Δ dtxR				pLongJon- Δ furA				pLongJon- Δ furB			
isolate	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
M13-FP	✓	✓	nd	nd	✓	✓	nd	nd	✓	poor	nd	nd	✓	✓	nd	nd	nd	nd	nd	nd
Blastn	✓	✓	nd	nd	✓	✓	nd	nd	✓	nd	nd	nd	✓	✓	nd	nd	nd	nd	nd	nd
confirm																				
<i>R. equi</i>	-	-	-	-	✓	nd	nd	nd	✓	nd	nd	nd	✓	nd	nd	nd	nd	nd	nd	nd
103S																				
transform																				
M13-FP																				

TGTA AACGACG GCCAGT

5.7.3 *In silico* – *in vitro* digestion comparison of plasmids

Further to the promoter analysis, a multi-enzyme restriction digest was performed on the successfully transformed plasmids, which could be compared to an *in-silico* digestion of the respective plasmid. This detailed analysis further indicated correct plasmid assembly of both pLongJon and the mutagenic inserts selected, based on corresponding band sizes.

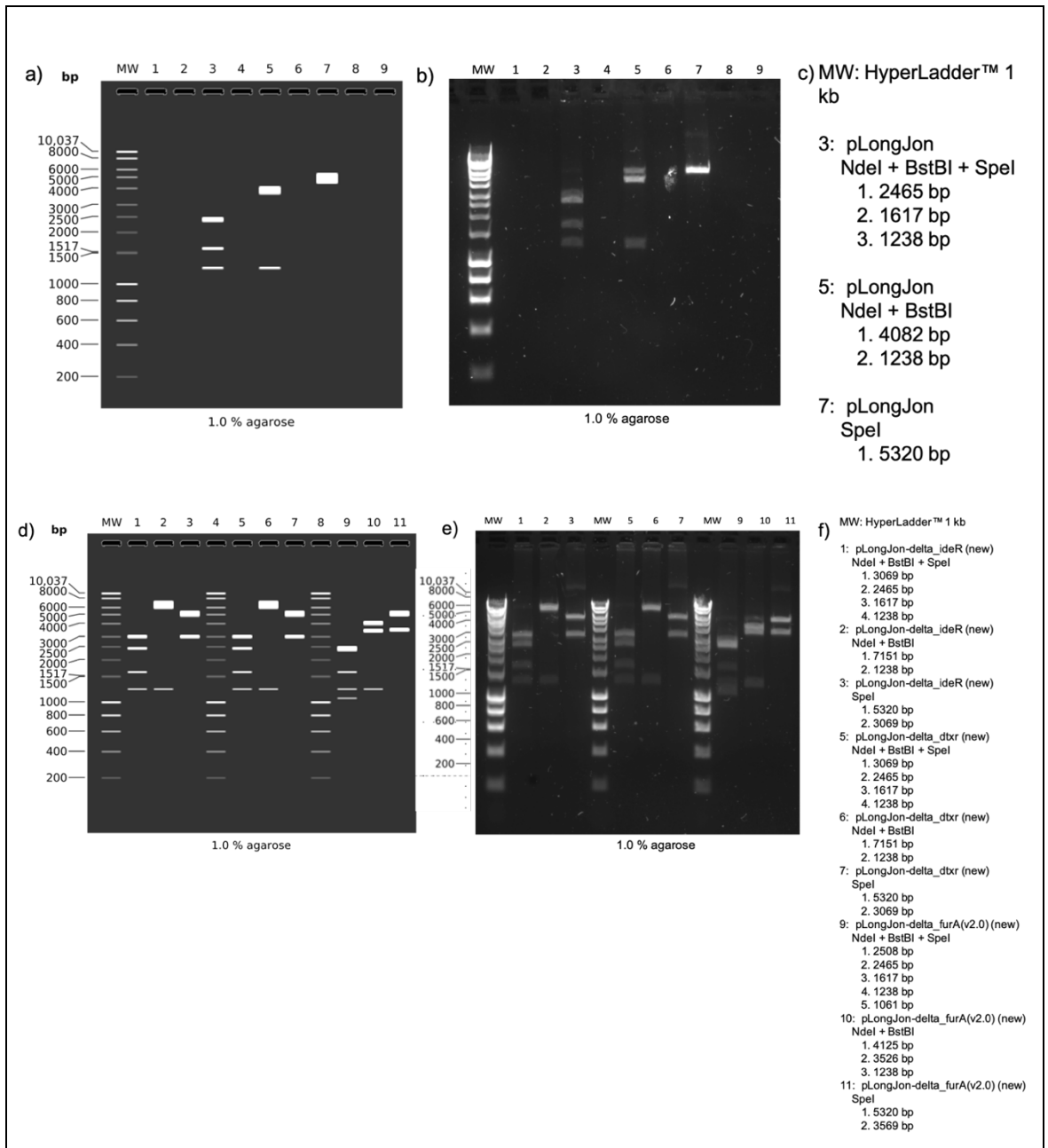


Figure 5-12: *In silico* digestion of pLongJon and mutant plasmids compared to *in vitro* digestions.

5.8 Incorporation of mutagenic constructs into *R. equi* 103S (V2)

The recombinant plasmids (pLongJon- Δ *ideR*, pLongJon- Δ *dtxR* and pLongJon- Δ *furA*) were used to transform *R. equi* 103S by electroporation. Positive selection for single crossover transformants was based upon growth on 80 μ g/ μ l Apramycin LB agar plates. *R. equi* 103S without plasmid incorporation was used as a negative control due to a lack of apramycin resistance.

5.9 Selection of the double crossover 5-FC resistant phenotype (V2)

As before, single crossover transformants (this time all four pSelAct mutants and the three pLongJon mutants) were grown without selection overnight, 100 μ l culture aliquots were serial diluted from 1:10 – 1:1000 using acetate minimal media (MMAc) adjusted to 10 g acetate per litre as recommended for increased selectivity (Van de Geize, R., personal communication) and plated onto MMAc agar supplemented with 100 μ g/ml 5-FC and incubated for 3 days at 37 °C.

Even after attempting to optimise the selection process by incorporation of a new promoter, selection of the 5-FC resistant phenotype, in excess of 1500 colonies were screened by replica plating on both non-selective media as well as on the 80 μ g/ μ l Apramycin LB agar plates indicating that all of colonies retained apramycin resistance.

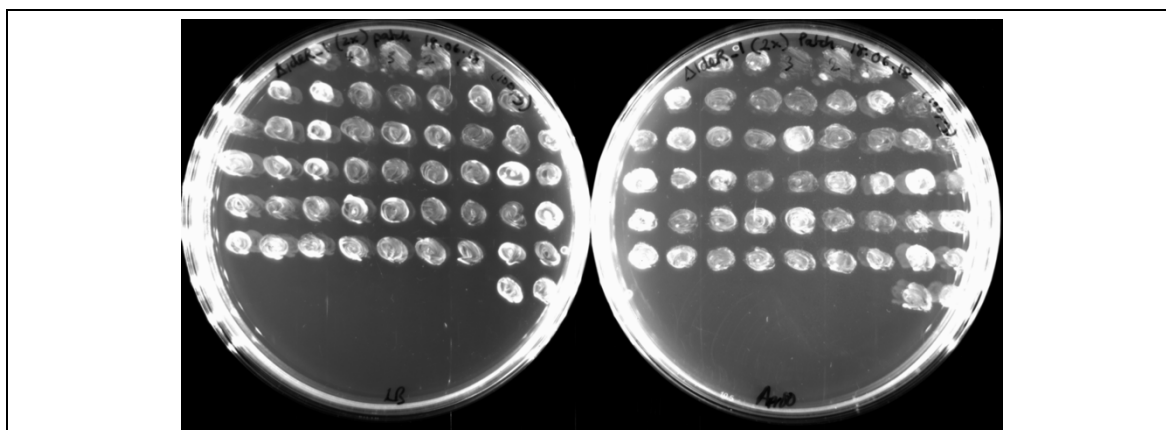


Figure 5-13: An example of replica plating for *R.equi- Δ ideR* on LB agar and LBapr80 using colonies from a 100 μ g/ml 5-FC plate.

5.9.1 Diagnostic PCR as a molecular screen and optimisation

The non-selective replica streaks were screened using colony PCR for mutants, however when analysed the gel, the results were difficult to interpret due to an inconsistent bias of amplicon bands (Figure 5-14). Ideally the PCR should show both wildtype and truncate copies of the amplicon in the

single cross over and the appropriate amplicon for the double crossover dependent on route of homologous recombination.

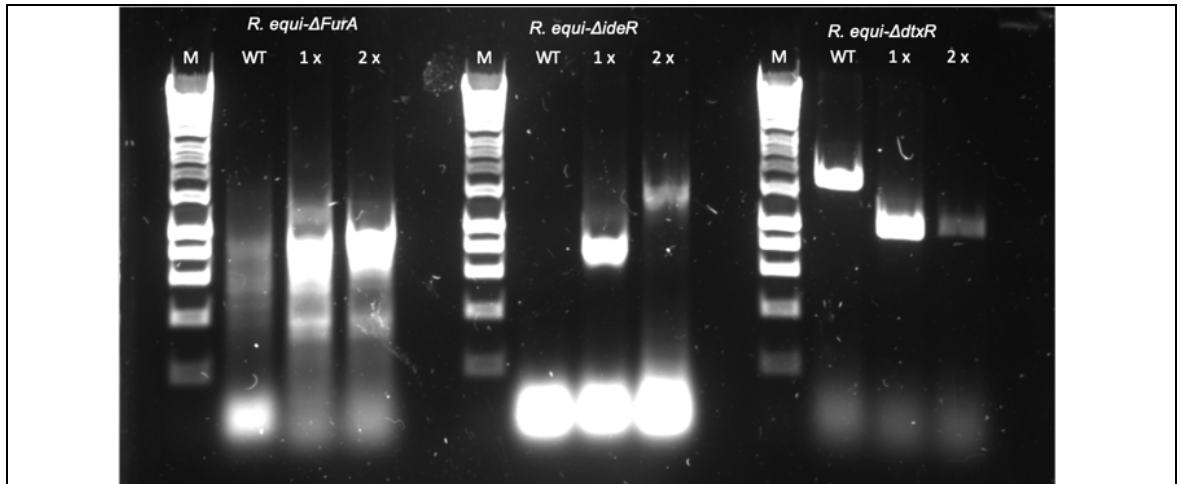


Figure 5-14: Colony PCR to identify secondary homologous recombinant mutants

The PCR was eventually optimised by alteration of PCR extension times from 30 seconds up to 60 seconds. This modification produced a vastly superior screening method, of which the robustness was further tested by amplification of at least eight single-crossover colonies that were proven to amplify both wildtype and truncate variants (Figure 5-15).

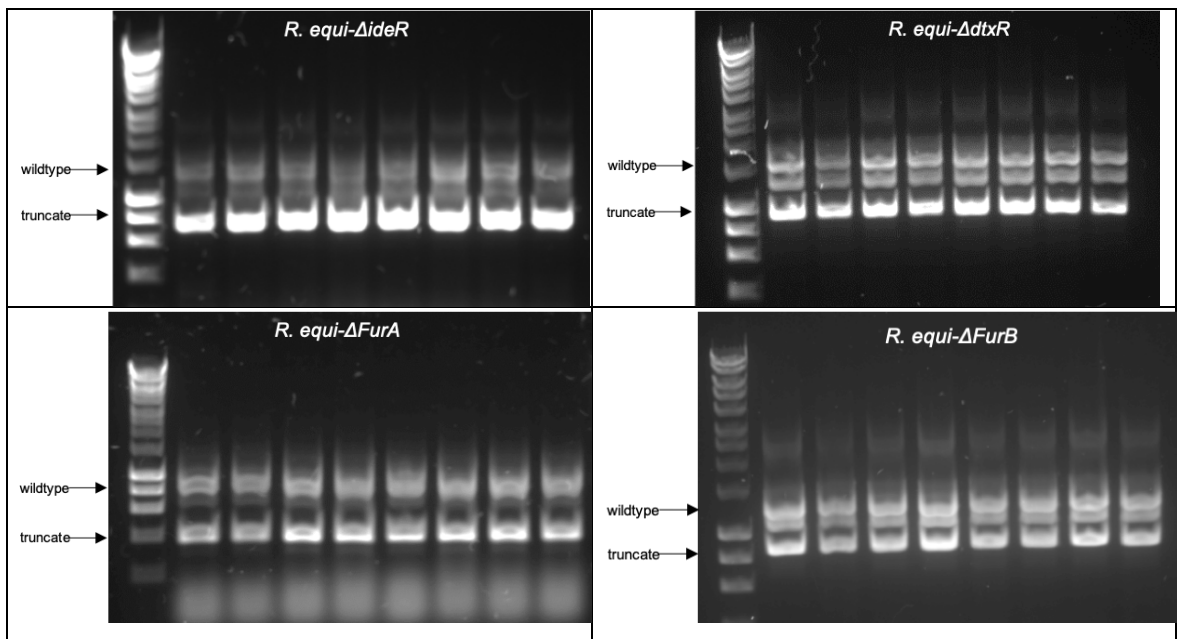


Figure 5-15: Optimised PCR screening to identify both wild type and truncate gene variants.

5.9.2 Identification of a flawed methodology

Given the number of colonies screened, it became apparent that the current methodology for selection of double-crossover mutants using 5-FC as a negative selection marker was severely ineffective. Significantly, a paper by Dubeau et al. (2009) describes the use of cytosine deaminase as a negative

selection marker in actinobacteria and in particular how resistance to CD in actinobacteria is common, with MIC ranges of in excess of 800 $\mu\text{g/ml}$. Although *R. equi* was not explicitly mentioned within the scope of the work, other closely related bacteria were detailed as needing concentrations between 400 and 800 $\mu\text{g/ml}$ to inhibit growth on minimal media, and therefore it was hypothesised that using a concentration of 100 $\mu\text{g/ml}$ in *R. equi* may be too low, and could be the reason for inefficient counter selection.

5.9.3 Optimising the 5-FC selection methodology

With a new methodology in mind, single crossover transformants using the pSelAct strategy were grown again overnight, and 100 μl culture aliquots were serially diluted from 1:10 – 1:1000 using acetate minimal media (MMAc) as before, and plated onto MMAc agar supplemented with the increased 200 $\mu\text{g/ml}$ 5-FC and incubated for 3 days at 37 °C. Single colonies were selected and replica plated onto both non-selective media as well as on 80 $\mu\text{g/ml}$ Apramycin LB agar plates. Remarkably, the replica plating indicated an increased efficiency due to a reduction in colony formation on the Apramycin LB agar plates (Figure 5-16).

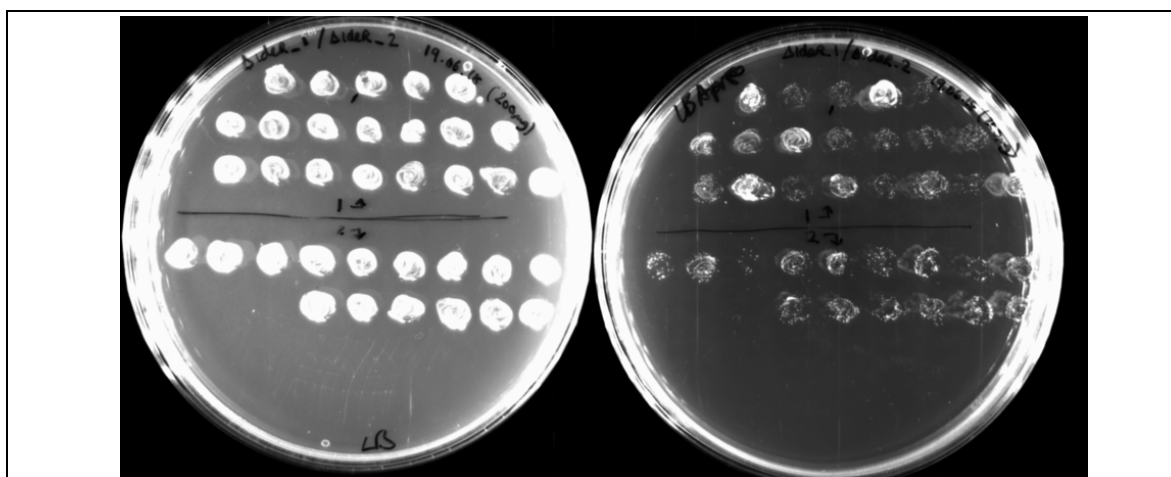


Figure 5-16: Replica plating for *R. equi-ΔideR* on LB agar and LBapr80 using colonies from a 200 $\mu\text{g/ml}$ 5-FC plate after 3 days growth

However even with this refined methodology, some of the double crossover's PCR lanes still show both PCR variant bands, this could be due to a mixed colony collection of single and double crossover recombinant mutants. As a result, the selection process was further repeated with an increased 5-FC concentration of 400 $\mu\text{g/ml}$; as expected, the increase in 5-FC concentration resulted in a significantly lower number of single colonies observed, even at lower cell density dilutions. Colonies that did

successfully grow were streaked onto the replica plates as before, with a remarked improvement, showing absolutely zero growth on Apramycin LB agar for both *R.equi-ΔideR*, *R.equi-ΔfurA* and *R.equi-ΔdtxR* (Figure 5-17), and at least one streak of and *R.equi-ΔfurB* with no growth.

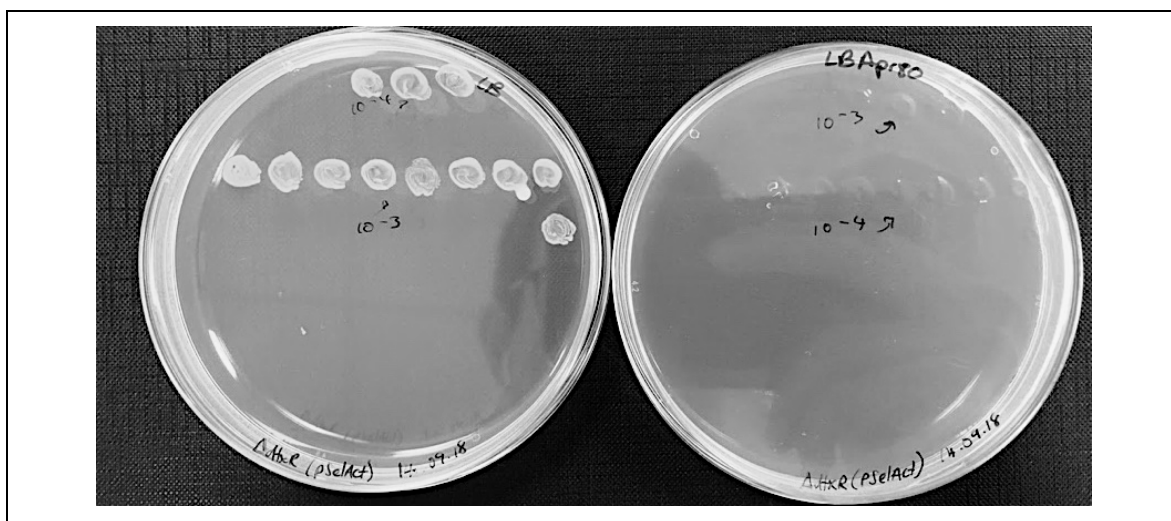


Figure 5-17: Replica plating for *R. equi-ΔdtxR* on LB agar and LBapr80 using colonies from a 200 µg/ml 5-FC plate after 3 days growth

5.10 Molecular confirmation of the generation of *R. equi* 103S Δ metal regulatory genes

With the refined methodology now producing consistent counter-selection, indicated using apramycin sensitivity, the replica plated colonies of interest were screened using the diagnostic PCR screen optimised earlier.

5.10.1 Selection of the mutant *R. equi-ΔideR*

To identify the route of the secondary homologous recombination event in *R. equi-ΔideR*, ten of the colonies that showed apramycin sensitivity were screened by colony PCR, in parallel with a single-crossover mutant of *R. equi-ΔideR*.

The modification of the diagnostic PCR earlier provided a robust method of confirming both gene variants for the single crossover, producing a wildtype band of *ideR* at 1413 base pairs and a truncated *ΔideR* band at 783 base pairs whilst also providing satisfactory conditions to identify single bands produced by double-crossover mutants in a single thermocycling run. Significantly, the PCR screen in Figure 5-18 indicates that all of the apramycin sensitive colonies tested underwent a secondary

homologous recombination event, of which, seven reverted to wildtype and three contained only the truncated gene mutation.

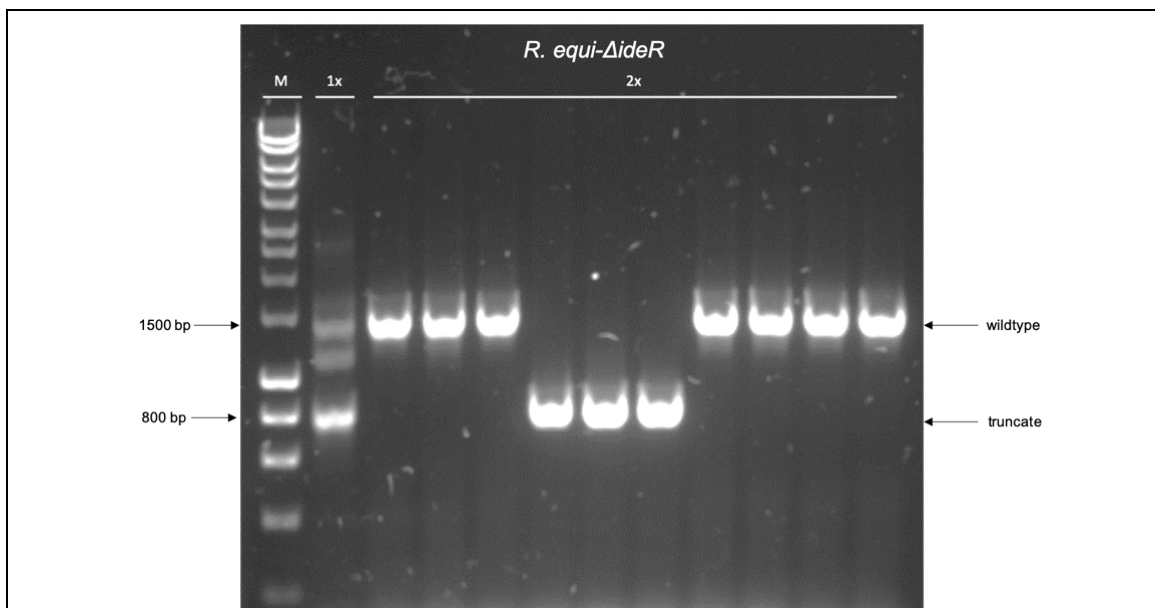


Figure 5-18: Colony PCR for *R. equi-ΔideR* to identify secondary homologous recombination

5.10.2 Selection of the mutant *R. equi-ΔdtxR*

To identify the route of the secondary homologous recombination event in *R. equi-ΔdtxR*, only three out of twelve colonies showed absolute apramycin sensitivity, these were screened by colony PCR, parallel with a single-crossover mutant of *R. equi-ΔdtxR*.

As previously mentioned, diagnostic PCR provided a robust method of confirming both gene variants for the single crossover, producing a wildtype band of *dtxR* at 1517 base pairs and a truncated *ΔdtxR* band at 917 base pairs whilst also providing satisfactory conditions to identify single bands produced by double-crossover mutants in a single thermocycling run.

Similar to the *R. equi-ΔideR* mutants, Figure 5-19 indicates that all of the apramycin sensitive colonies tested underwent a secondary homologous recombination event, of which, two reverted to wildtype and one contained only the truncated gene mutation having lost the wildtype chromosomal copy of the gene.

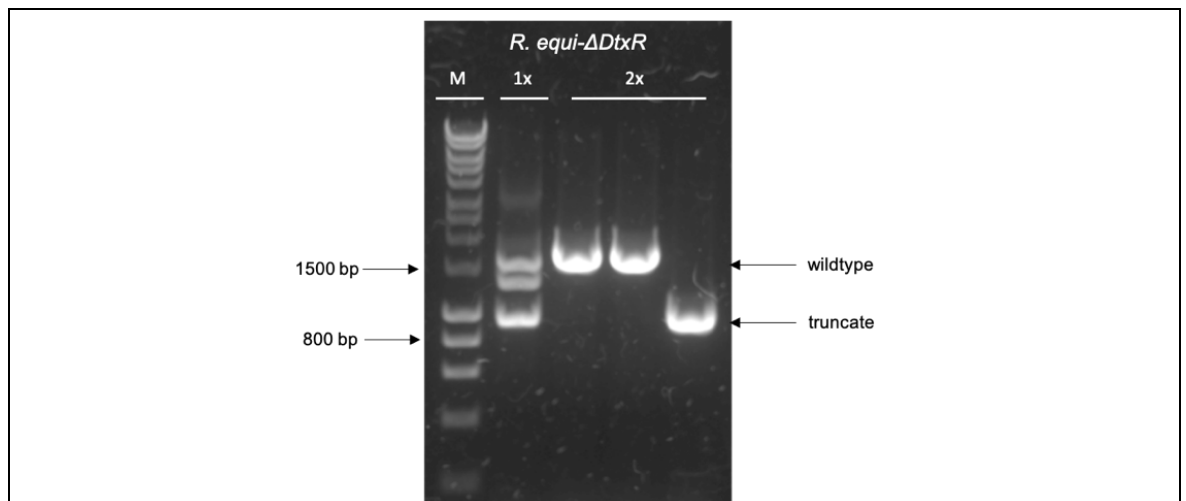


Figure 5-19: Colony PCR for *R. equi-ΔdtxR* to identify secondary homologous recombination

5.10.3 Selection of the mutant *R. equi-ΔfurA*

To identify the route of the secondary homologous recombination event in *R. equi-ΔfurA*, over thirty colonies were identified to have apparent absolute apramycin sensitivity, these were screened by colony PCR, as described previously for the other mutants.

As previously mentioned, diagnostic PCR provided a robust method of confirming both gene variants for the single crossover, producing a wildtype band of *furA* at 892 base pairs and a truncated *ΔfurA* band at 400 base pairs whilst also providing satisfactory conditions to identify single bands produced by double-crossover mutants in a single thermocycling run (Figure 5-20).

Similar to the other mutants, under the refined methodology a significant number of colonies appeared to undergo a secondary homologous recombination event, and yet some remained as single-crossover mutants even with apramycin sensitivity. This may have been as the result of a 5-FC resistant phenotype from an infrequent recombination event excluding *coda::upp* but not *aac(3)IV*, or through acquiring apramycin resistance, this theory was further explored by previous researchers (L. Dover, pers. comms) working on in-frame mutants in *R. equi*.

However, due to the improved methodology associated with counter-selection and screening, twenty-three of the colonies appear to have undergone a secondary homologous recombination event resolving to wildtype, and only one truncated gene mutation having lost the wildtype chromosomal copy of the gene (Figure 5-20).

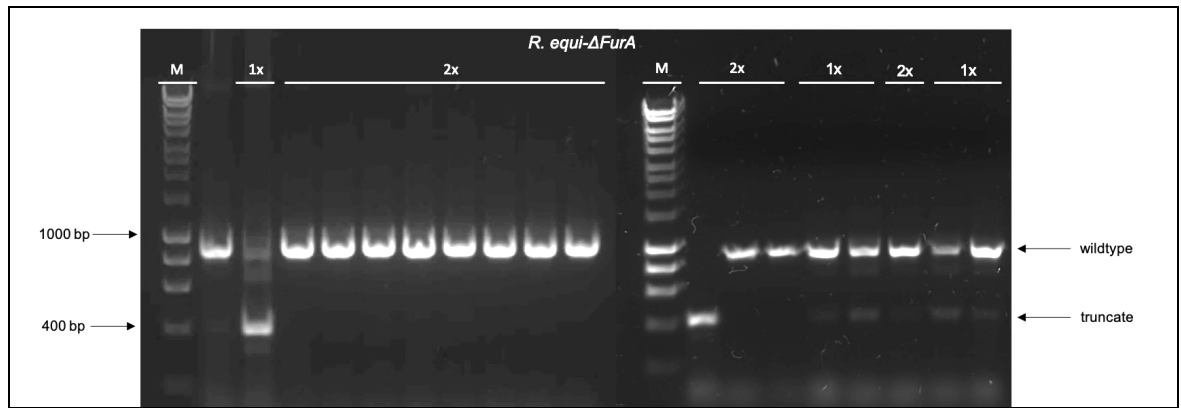


Figure 5-20: Colony PCR for *R. equi-ΔFurA* to identify secondary homologous recombination
 Considering there is no selection process for the in-frame deletion gene over the wildtype revertant secondary homologous recombination event, it seems logical that the selection may be driven by the biological significance of the gene. The recombination ratio of *R. equi-ΔFurA* appears preferentially favoured towards resolving by wildtype and presents the question is this as a result of conditional survivability, with regards to a lack of a functional oxidative stress response.

In the bacterial homologue *M. tuberculosis*, it is clear that the *furA* gene equivalent is responsible for regulation of both mycobacterial physiology and intracellular survival controlling oxidative stress response (Zahrt et al., 2001).

The genetic locality of the *furA* gene may harbour the desired answers; sharing an operon with a catalase gene, it is likely that the operon functions in a homologous way to that of *M. tuberculosis*, which has a *furA* gene responsible for negative regulation of the catalase-peroxidase gene *KatG*, suggesting that the *furA* operon in *R. equi* may be responsible for regulation of an oxidative stress response.

5.10.4 Selection of the mutant *R. equi-ΔfurB*

To identify the route of the secondary homologous recombination event in *R. equi-ΔfurB*, only two out of twenty-eight colonies showed absolute apramycin sensitivity, these were screened by colony PCR, parallel with a single-crossover mutant of *R. equi-ΔfurB*. The previously optimised diagnostic PCR confirmed both gene variants for the single crossover, producing a wildtype band of *furB* at 1235 base pairs and a truncated *ΔfurB* band at 853 base pairs whilst also providing satisfactory conditions to identify single bands produced by double-crossover mutants in a single thermocycling run.

When screened using the diagnostic PCR method, the results indicate that the apramycin sensitive colonies underwent a secondary homologous recombination event, of which, one reverted to wildtype and one contained only the truncated gene mutation

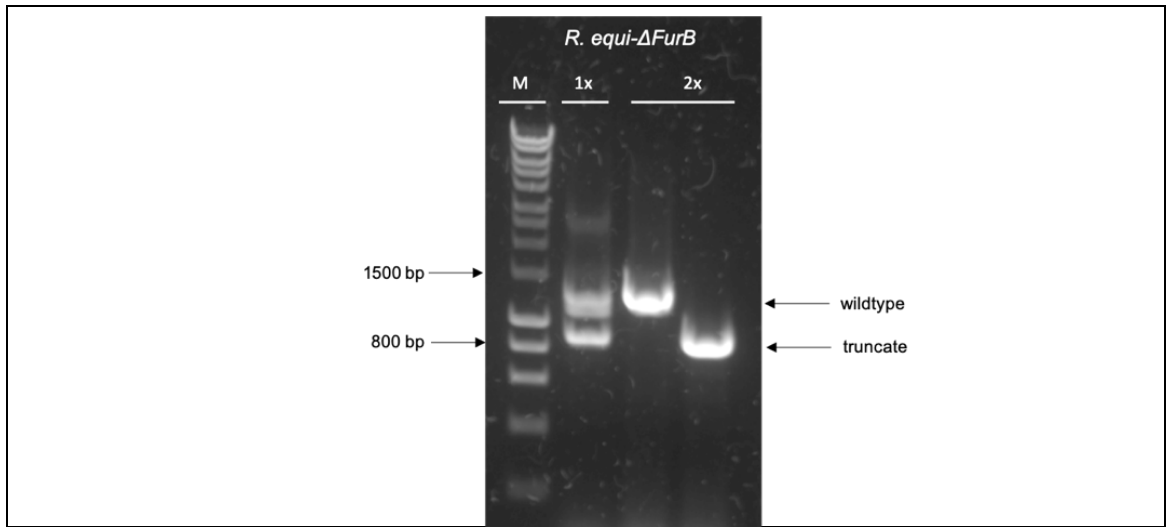


Figure 5-21: Colony PCR for *R. equi-ΔfurB* to identify secondary homologous recombination

5.11 Colony morphology and phenotypic characterisation

The mutant phenotypic appearances and growth of *R. equi* 103S- $\Delta ideR$; $\Delta dtxR$; and $\Delta furB$ appear normal comparative to the wildtype strain, displaying salmon pink circular mucoid colonies, with further orange pigment developed over a series of days. Comparatively, the mutant phenotypic of *R. equi* 103S- $\Delta furA$ varies after 24 hours growth, as the rate of growth is reduced, and while the colony shape and surface are similar, the pigment colour is a darker orange-brown (Figure 5-22).

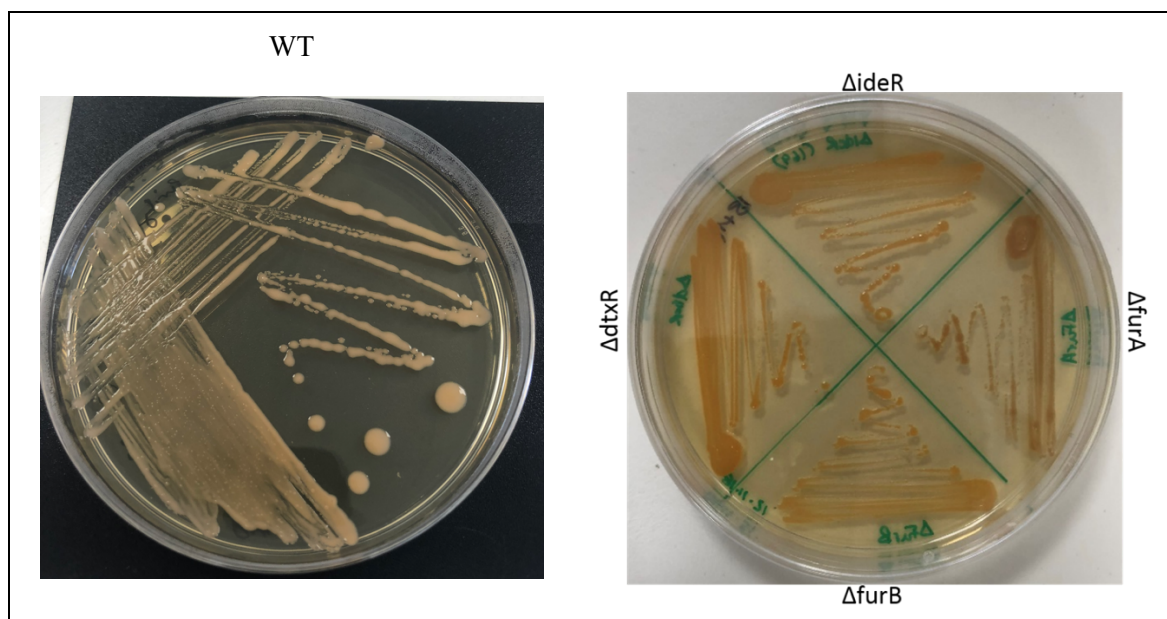


Figure 5-22: BHI agar plates showing colony morphology of *R. equi* 103S wildtype and mutant variations.

Mutant strains $\Delta ideR$; $\Delta dtxR$; and $\Delta furB$ show no observable phenotypic variation compared to the wildtype, strain $\Delta furA$ showed an impaired growth and developed with a darker pigment.

5.11.1 Cellular morphology

No changes in cellular morphology was observed for any of the four mutants, comparative to the wild-type strain, with all strains consistently showing bacillus morphology when cultivated in broth, as previously described for *R. equi* (Goodfellow et al., 1990).

5.12 Protein profiling of *R. equi* mutants

To understand the proteomic effect of mutagenesis on the *R. equi* genome, total protein profiles (Figure 5-23) were compared using a 12 % SDS-PAGE gel as described in section 2.9.4. The profiles generally look consistent to the wildtype, apart from a highly overexpressed band present in *R. equi* 103S- $\Delta furA$ between the 50 – 60 kDa range when compared with the molecular standards (Figure

5-23).

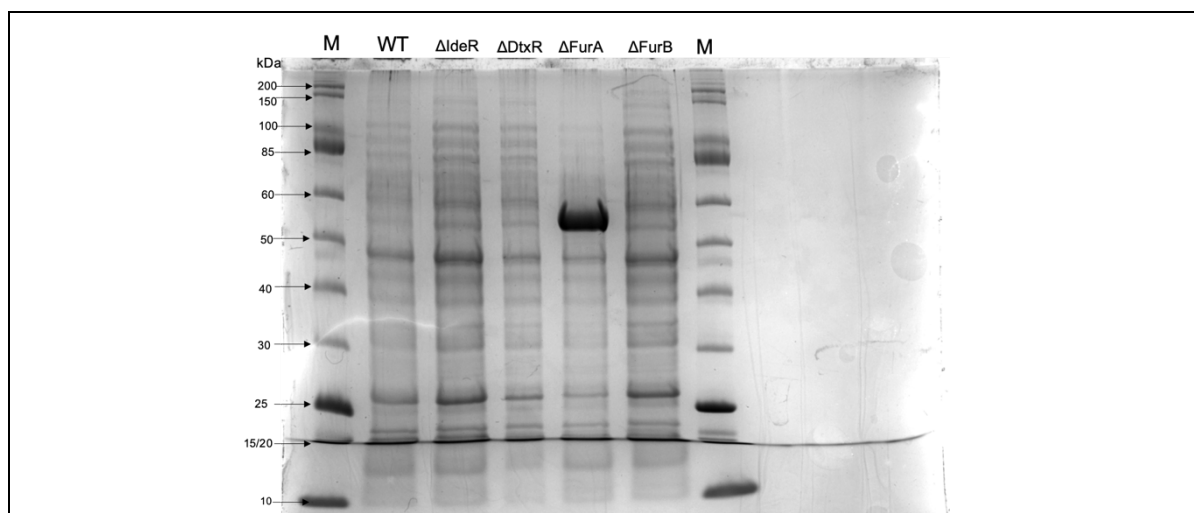


Figure 5-23 A 12 % SDS-PAGE total protein load for *R. equi* 103S wildtype compared to *R. equi* 103S mutants.

5.13 *R. equi*- $\Delta furA$ – De-repression of an essential catalase gene

5.13.1 Proteomic analysis

The protein band of interest from *R. equi*- $\Delta furA$ was excised (Figure 5-24A), subject to trypsinolysis and analysed *via* LC-MS for identification. The data file was processed using a MS/MS ion search on the local MASCOT server (methods section 2.13.3), identifying the most significant protein match as a catalase protein (Table 5-3, highlighted yellow), with a corresponding mass of ~55 kDa, consistent with the migration of the band in SDS-PAGE. A protein score of 16340 was attained. The peptide score distribution ion scores > 16 indicates identity or significant homology ($p < 0.05$) (Figure 5-24C), supported by a protein sequence coverage of 77 % by matched peptides (Figure 5-24B).

Table 5-3: MASCOT protein hit output detailing the top 5 protein families, and their associated members.

Table is ranked on total family significance, incorporating protein score, matches and distinct sequences. Protein identifications given in the description.

<u>Family</u>	<u>M</u>	<u>Accession</u>	<u>Score</u>	<u>Mass</u>	<u>Matches</u>	<u>Match(sig)</u>	<u>Sequences</u>	<u>Seq(sig)</u>	<u>emPAI</u>	<u>Description</u>
<u>1</u>	1	gi311887293	16340	55009	554	507	37	33	27.24	catalase
<u>1</u>	2	gi311891095	928	57487	64	57	4	2	0.18	catalase
<u>2</u>	1	gi311890529	3337	56506	95	86	26	26	7.76	chaperonin GroEL2
<u>2</u>	2	gi311890174	3317	55868	86	78	26	22	5.41	chaperonin GroEL
<u>3</u>	1	gi311890841	3177	60416	117	104	29	28	17.01	urocanate hydratase
<u>4</u>	1	gi311888517	3170	52193	98	83	23	21	8.58	ATP synthase beta chain AtpD

As previous alluded, and in conjunction with this new proteomic evidence, it seems highly likely that *R. equi* FurA functions as a homolog to its mycobacterial counterpart by negatively regulating the upstream catalase gene *Cat*.

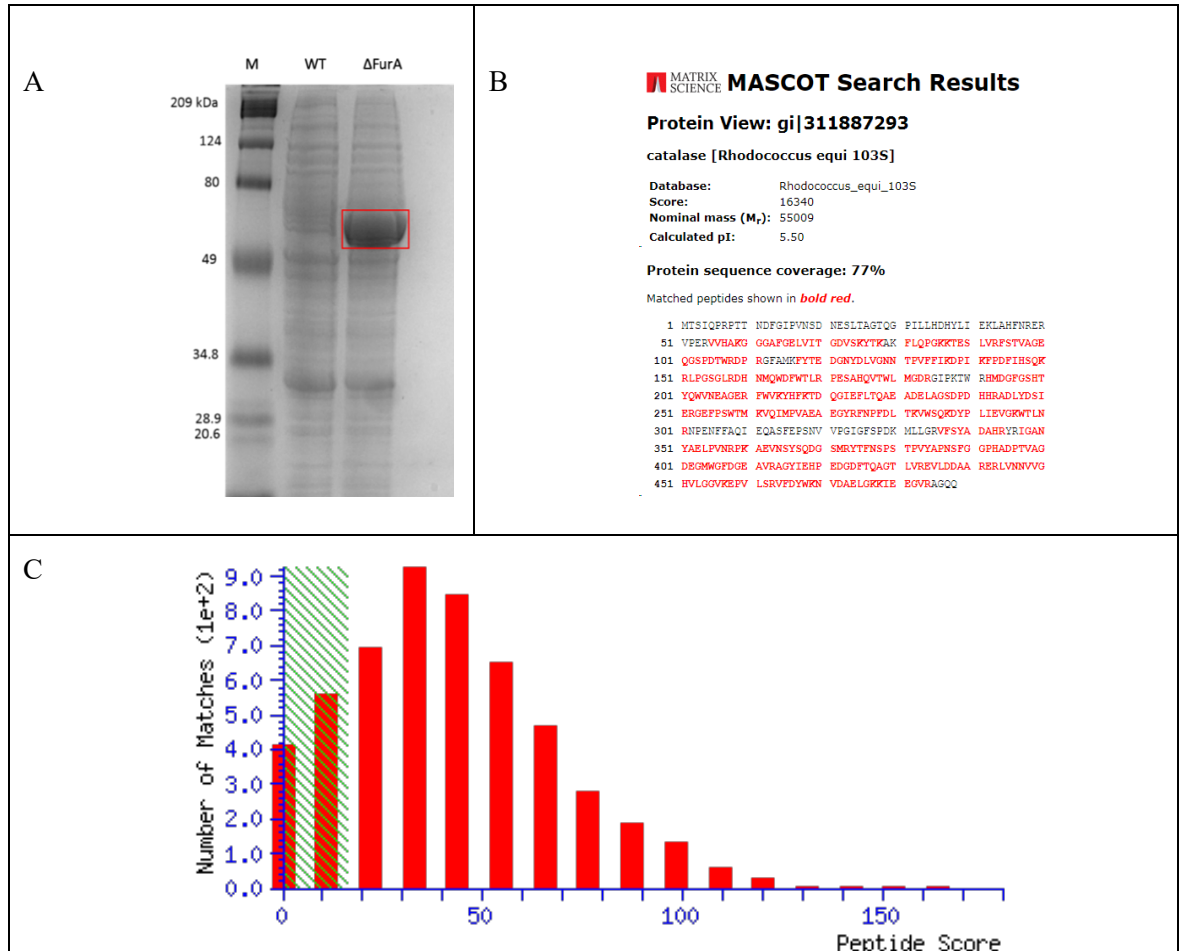


Figure 5-24: Protein identification of the hyper-expressed band in *R. equi* 103S-FurA.

A) detailed excision site for the band of interest. B) Protein sequence coverage. C) Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event. There are 4,229 peptide matches above identity threshold and 4,278 matches above homology threshold for 24,615 queries. Histogram score range is (0, 164). On average, individual ions scores > 16 (beyond green shading) indicate identity or extensive homology ($p < 0.05$).

5.13.2 Analysis of phenotypic effects on sensitivity to oxidants

The ubiquitous clustering of *furA* and a downstream catalase-oxidase gene *katG* and their role in coordination of an oxidative stress response has been well documented in mycobacteria (Milano et al., 2001; Pagan-Ramos et al., 1998; Pym et al., 2001; Zahrt et al., 2001), leading us to hypothesise that a homologous oxidative stress response is similarly coordinated in *R. equi*.

To determine if the in-frame deletion mutant *R. equi*- Δ *furA* exhibited a phenotype typical of

oxidative stress, susceptibility testing with hydrogen peroxide (H₂O₂) was carried out. Given the previous literature on the effect of mycobacterial *furA* disruption on peroxide susceptibility, the results for *R. equi-ΔfurA* clearly conform to a homologous regulatory system, proving to be significantly more resistant to H₂O₂ in comparison to the parental wildtype strain *R. equi* 103S (Figure 5-25) using one-way ANOVA with Dunnett's Multiple Comparison Test ($P \leq 0.001$). This further highlighted that FurA is responsible for negative regulation of a catalase for detoxification of hydrogen peroxide, in an attempt to prevent formation of reactive oxygen species.

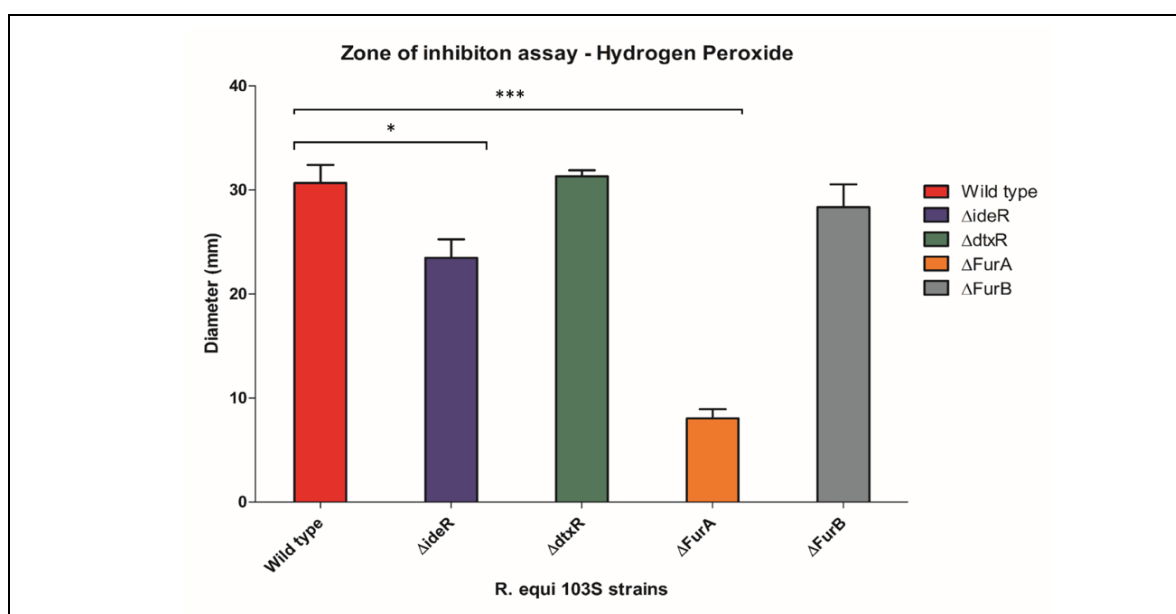


Figure 5-25: A zone of inhibition assay for breakdown of hydrogen peroxide by catalase. Zones of inhibition ($n = 3$) were measured after 72 hours. Data was plotted as mean \pm SEM. Statistical significance was determined using one-way ANOVA with Dunnett's Multiple Comparison Test. Wildtype vs. *ΔideR* : * $p \leq 0.05$, Wildtype vs. *ΔfurA* : *** $p \leq 0.001$.

Interestingly, the H₂O₂ sensitivity testing also highlighted a significant difference for the *R. equi-ΔideR* mutant when compared to the parental strain ($P \leq 0.05$). This could likely arise due to the inherent association of oxidative stress and iron metabolism; this result however, is contradictory to a similar study in *M. smegmatis*, where inactivation of *ideR* resulted in an increased peroxide sensitivity and de-repression of siderophore synthesis under iron sufficiency (Dussurget et al., 1996).

Given the interlinked relationship between oxidative stress and iron metabolism, and the sequence similarities to other *fur* genes, the *R. equi* mutants were studied to identify whether gene inactivation by in-frame deletion affected siderophore production, in comparison to wildtype. The wildtype and mutant strains of *R. equi* were grown in lactate minimal media both with and without iron supplementation; there was no observable difference regarding siderophore secretion, even when

bacterial supernatant was added to CAS agar plates.

Although this is consistent with the wildtype *R. equi* displaying CAS-negative results, it is contraindicated to our hypothesis, and surprising that impairment of the iron regulatory system by mutation of the gene *ideR* does not promote overproduction of siderophores, as predicted with the genomic survey of iron regulatory mechanisms (as described in earlier chapters). This could suggest a much more complex iron regulatory network that may rely on other methods of iron sequestration. Furthermore, given the complex regulation and control of oxidative stress response genes, it is highly unlikely that *FurA* is involved in siderophore secretion within *R. equi* and rather, is almost certainly an essential component in redox sensing, consistent with other bacterial *Fur* homologues.

5.14 Concluding remarks for this chapter

The aim of this project was to produce a selection of in-frame gene deletion mutants predicted to encode iron-responsive regulatory proteins in *R. equi* 103S, *IdeR*, *DtxR*, *FurA* and *FurB*. The mutagenesis process proved particularly challenging and frustrating given the methodology had been previously optimised. Even with substitution of a superior promoter the selection was not adequate, resulting in an optimisation of the cytosine deamination selection by increasing the 5-FC concentration. Having previously replica plated over 1500 colonies without successful counter-selection, this alteration to the methodology provided a significant improvement in selecting cells that had undergone the secondary homologous recombination event and achieving in-frame gene deletion mutants for all four targets.

The success of the regulator mutagenesis experiment further develops the working hypothesis regarding the iron regulatory mechanisms within *R. equi*, as described in Figure 5-26. The subsequent phenotyping of mutants identified a noteworthy change for *R. equi-ΔfurA*, which displayed a remarked resistance to peroxides, leading us to hypothesise that it is responsible for coordinating an oxidative stress response in *R. equi* by transcriptional repression of catalase. The other mutants did not produce any identifiable phenotypic changes even under iron limitation, and it would be valuable to identify transcriptomics changes comparative to the parental wild type strain. Contrary to our working hypothesis, the production of *R. equi-ΔideR* did not lead to a hyperexpression of siderophores, suggesting that this bacterium has a much stronger control over the iron starvation

response mechanisms than previously hypothesised.

The genomic survey and the subsequent *in silico* metal binding analysis completed earlier, associated each of the remaining mutants with a different divalent cation; going forward it would be useful to impose metal limitations for the mutants on *R. equi* 103S, *R. equi-ΔdtxR* and *R. equi-ΔfurB* as well as subsequent phenotyping and transcriptomic analysis under these conditions

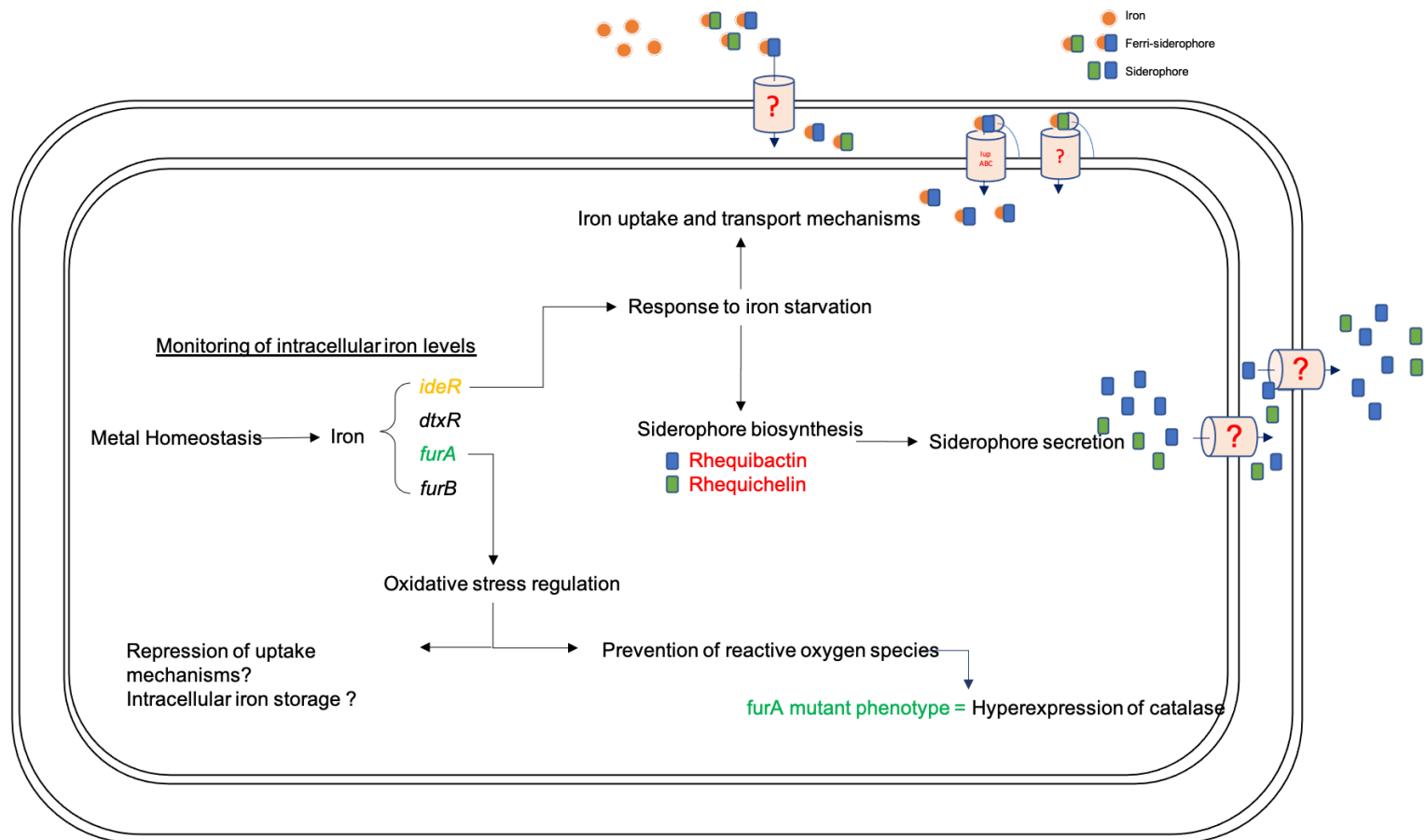


Figure 5-26 An updated *R. equi* iron homeostasis network map (after chapter 5)

Data generated from the preliminary genomic survey, developed using in silico metal binding capacities, and production of metal regulator mutants. The FurA candidate is displayed in green to indicate successful prediction with the mutant phenotyping identifying an oxidative stress response. The IdeR candidate is displayed in orange, to indicate incomplete results regarding iron regulatory mechanisms that are to be further explored in future work. Siderophore candidates at this stage are given in red as at this predictions have yet to be made.

6 An analysis of the *Rhodococcus equi* 103S
metabolome under iron limitation: An insight into
siderophore mediated iron metabolism

6.1 Introduction

6.1.1 Introduction to metabolomics

Metabolomics is the extensive analysis of small molecules (also known as metabolites) in a biological sample, of which metabolites and their interactions within a biological system are collectively recognised as the metabolome. The term ‘metabolomics’ was first coined by Fiehn (2002) to describe the synthesis of end-products manufactured by cellular regulatory processors, however the concept of a metabolic profile was first described as early as the 1940’s (Gates and Sweeley, 1978). Critically, the study of a metabolome and as such, the creation of a metabolic profile can provide a snapshot of the cellular processes occurring at a given time in response to the physiological conditions.

Often produced as a by-product of cellular metabolism, metabolites can include hydrophobic lipids, hydrophilic carbohydrates, ketones, amino and non-amino organic acids, and complex natural compounds such as antibiotics, pigments, non-ribosomal peptides, cofactors (often termed secondary metabolites) that can influence the way a cell can transfer energy, or sequester important minerals to survive (Mussap et al., 2018; Tang, 2011). Changes in metabolite composition and abundance within a biological system can occur in response to either therapeutic treatments or physiological stresses.

Current metabolomic technologies vastly exceed the capacity of standard biochemical methods and consequently are proficient in the accurate analyses of thousands of metabolites. Therefore, metabolomic studies can often deliver comprehensive characterisation of metabolic phenotypes, characterisation of metabolic disorders that can often be the fundamental cause of disease, as well as discovery of biomarkers and therapeutic targets (Clish, 2015).

Metabolomics is often termed as a top-down analytical technique, designed to produce broad and integrative results to identify effects in a given biological system; there is a drive within the systems biology to collaborate with the customary bottom-up methodology techniques (genomics, transcriptomics and proteomics) to assemble a more complete picture of the biological processes occurring and associated cellular regulation (Tang, 2011). Integration of metabolomic data is considered to be key in deciphering the pathways involved in cellular signalling, as metabolites can

act as intermediates of these biochemical reactions.

In contrast to the aforementioned bottom-up analyses (genomics, transcriptomics and proteomics), compounds produced from metabolic reactions can vary significantly in their chemical structures and associated properties thereby making determination of a given metabolome under a specific physiological condition exceptionally challenging (Garcia et al., 2008).

The experimental methodologies for metabolomics primarily use liquid or gas chromatography coupled to a mass spectrometer or High-resolution ^1H nuclear magnetic resonance spectroscopy for identification. Both of these methods are well established in investigation of complex analytes for measurement of an array metabolites in a single run (Rochfort, 2005).

The application of metabolomics can be performed using either targeted or untargeted approaches, each possessing their own advantages. Firstly, untargeted metabolomics can be classified as a comprehensive investigation of all analytes within a sample under a given condition (often termed a snapshot) including chemically undetermined metabolites that might represent potential novel biomarkers (Ribbenstedt et al., 2018; Roberts et al., 2012). Due to the size and complexity of untargeted metabolomic datasets, it is frequently paired with chemometric analysis methods including multivariate analysis to moderate the large datasets generated into more manageable subgroups, often to further investigate statistically significant variation in metabolite abundance.

Comparatively, targeted metabolomics often termed metabolite profiling is the collection and investigation of a particular group of metabolites and can often involve quantitative or semi-quantitative analysis to focus on the effects of external factors or environmental conditions on specific pathways or biosynthesis routes of the metabolites of interest. Furthermore, optimisation of sample preparation can also be advantageous for targeted analysis by isolation and purification techniques, alternatively depletion techniques can be utilised to reduce the pre-eminence of highly abundant metabolites.

6.2 Chapter aims and hypothesis

This chapter is apportioned to addresses three fundamental aims:

(1) To establish the growth conditions for *R. equi* that can elicit a defined a biological response to iron starvation. It was hypothesised that optimisation of media conditions to induce iron starvation would produce a physiological response including the secretion of two potential *R. equi* siderophores consistent with the bioinformatic analyses described here and elsewhere (Chapter 3). Additionally, these growth conditions would subsequently be used consistently in a range of -omics studies in this thesis; including this comparative analysis of secreted metabolites (metabolomics), a comparative analysis of cell-wall associated iron-regulated protein mechanisms (proteomics) and characterisation of global regulation of gene expression under iron starvation (transcriptomics).

(2) An untargeted metabolomics study to compare the *R. equi* 103S metabolome under iron limitation. It is hypothesised that a list of candidate metabolites secreted exclusively under iron starvation would be produced, of which some candidates may correlate with the predicted mass sizes of the siderophores predicted earlier.

(3) A targeted metabolomics approach to identify *R. equi* 103S siderophores, using the candidate list produced in chapter aim 2. It was hypothesised that if candidates were identified in (2), it would be possible to use chromatography to isolate metabolites of interest and subsequently characterise by MS fragmentation methods.

6.3 Media and growth conditions for *R. equi* 103S

Previous literature describing growth of *R. equi* under iron starvation is limited, but it has been consistently reported that iron is essential for growth (Jordan et al., 2003; Miranda-CasoLuengo et al., 2008). *R. equi* strains were maintained on Luria-Bertani (LB) or Brain Heart Infusion (BHI) medium or on minimal medium supplemented with 20 mM lactate (LMM) at 37 °C as previously described (Kelly et al., 2002; Miranda-CasoLuengo et al., 2005). Iron starvation was induced using a variant of the trace element solution devoid of ferrous sulphate as described in Section 2.2.

Interestingly, previous work performed by Miranda-CasoLuengo et al. (2008) identified that after growth in low iron conditions *R. equi* produces a chromophore when supplemented with an iron-containing trace elements solution. This chromophore also occurred during growth of *R. equi* $\alpha 5$, a strain containing a transposome insertion in *iupABC*, encoding a siderophore uptake system in iron sufficient conditions. This chromophore was hypothesised to be as a result of formation of an Fe³⁺-siderophore complex given the red pigmentation of the medium supernatant.

According to Miranda-CasoLuengo, the observed chromophore was not present when analysing growth of wild-type strain in iron sufficient medium or when the trace element solution lacking FeSO₄ was added as a supplement to the iron deplete cultures.

The initial experiment attempted growth of *R. equi* 103S under iron limitation using LMM-Fe media, of which the culture reached stationary phase after 24 hours (data not shown). Unlike the experiment performed by Miranda-CasoLuengo the chromophore was not observed after 72 hours growth even after supplementation of the supernatant with an ferric chloride solution. The cultures were also aliquot onto a Chromazurol S containing agar plate to detect siderophore production through a dye-assisted chelatory mechanism, however the the cultures failed to grow likely due to toxicity associated with hexadecyltrimethylammonium bromide noted in Gram-positive organisms (Louden et al., 2011). An alternative method was also attempted using culture supernatant aliquot into wells within a Chromazurol S agar plate, without development of a halo indicative of siderophore presence. This result was surprising given siderophore-producing bacteria typically exhibit an exaggerated response *in vitro*.

6.3.1 Medium optimisation 1

To ensure that the growth limiting factor for this series of experiments was the iron concentration, a variety of media modifications was performed by supplementation of LMM (section 2.2.2) with either Riboflavin, a double concentration of the carbon or nitrogen sources, , as well as substitution of the sodium L-lactate for sodium acetate.

The growth curve (Figure 6-1) indicated that *R. equi* cultured in LMM without supplementation grew slower than other supplemented media conditions but still achieved a roughly similar stationary phase optical density. The media supplemented with double of the nitrogen source ammonium sulphate appeared to show the best growth rate overall, and the double carbon source LMM exhibited exceptional growth in the first 24 hours in comparison to others. Furthermore, it was determined that to ensure the results were comparable, the acetate supplemented media should not be used for further studies.

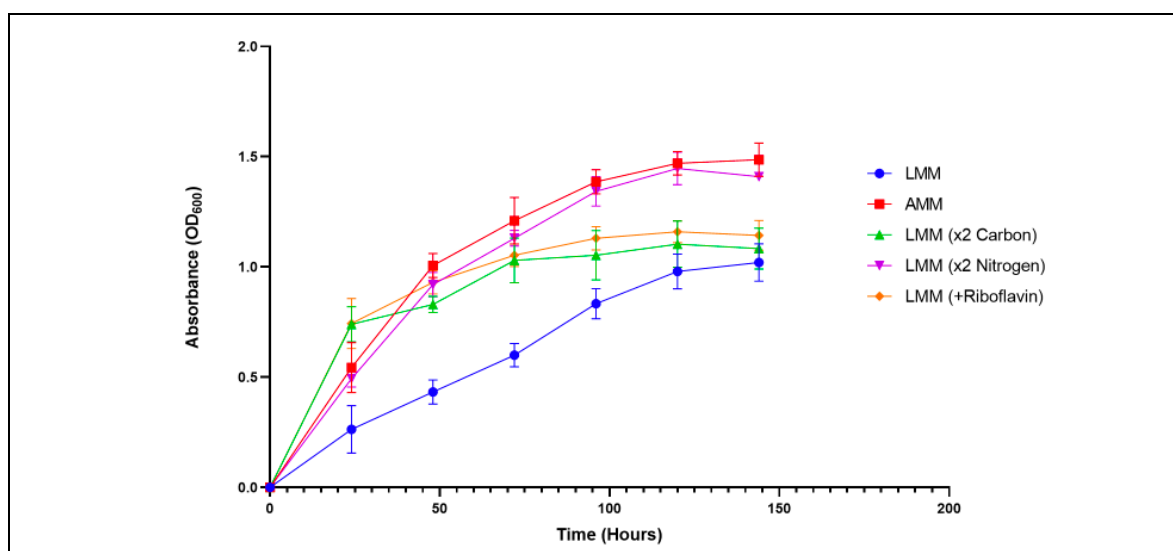


Figure 6-1: Identification of optimal growth conditions to induce an iron starvation response in *R. equi* 103S.

Growth was performed at 37 °C, 180 RPM. Optical density measured every twenty-four hours. LMM = Lactate minimal media;

Growth analyses were repeated using two supplemented LMM conditions (x2 Carbon, and x2 Nitrogen), but imposing different iron regimens to elicit iron starvation (LMM-Fe) and iron-replete (LMM+Fe) phenotypes (Figure 6-2). Iron-replete growth of *R. equi* was greater than iron-starved growth, but remarkably the final yields (as estimated by optical density) were comparable and consistent with previous studies (Miranda-CasoLuengo et al., 2008).

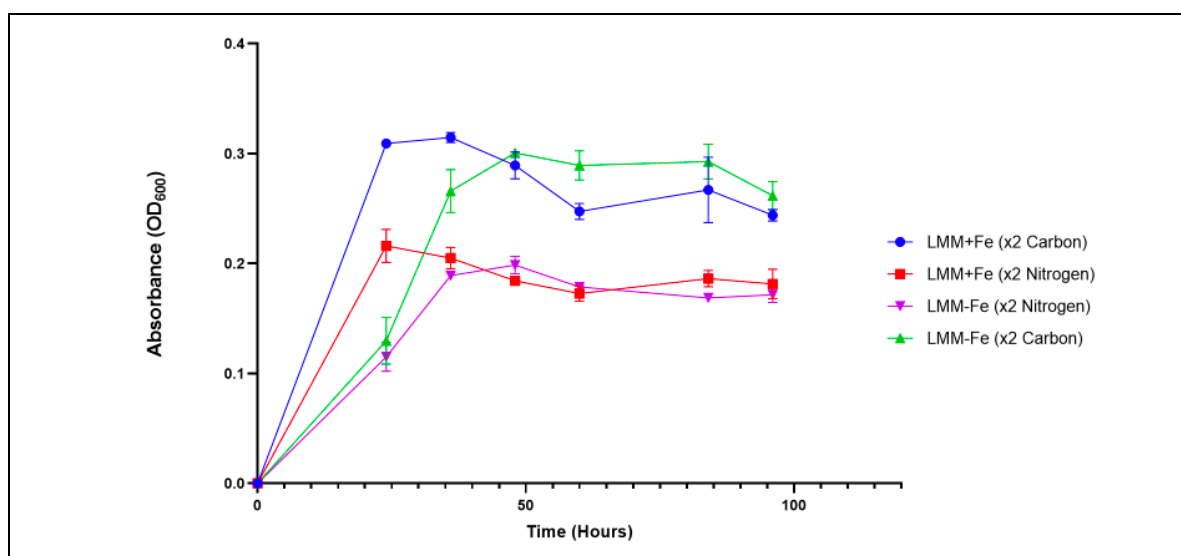


Figure 6-2: Refined growth conditions in *R. equi* 103S imposing different iron regimens to elicit an iron starvation response.

Growth was performed at 37 °C, 180 RPM. Optical density measured every twelve hours. LMM = Lactate minimal media; +/- Fe = iron supplemented/iron deficient conditions.

6.3.1.1 Modified CAS agar to detect siderophore production

Previous attempts to identify siderophore production using CAS agar had been unsuccessful, likely due to the inherent toxicity of hexadecyltrimethylammonium bromide (HDTMA) preventing growth of Gram-positive organisms on the agar (Louden et al., 2011). However, this toxicity should not prevent siderophore detection of an iron-starved culture supernatant; however, the colour change associated with extraction of iron from the dye complex was not detected consistent with the findings of Fiss and Brooks (1991) that identified the genus *Rhodococcus* as having variable siderophore activity, and *R. equi* as siderophore negative.. This suggested that *R. equi* does not generate the exaggerated siderophore yields commonly observed with many microorganisms.

A more-sensitive modified microtitre CAS method (Arora and Verma, 2017) was assessed and significant colourimetric changes were generated in culture supernatants derived from iron depleted and iron supplemented bacteria (Figure 6-3). This outcome indicated that under iron starvation *R. equi* does synthesise and secrete iron-chelating compounds, given the complete sequestration of the metal from the dye (Figure 6-3, columns 1:3; 7:9). The partial extraction of iron from the dye under iron supplemented conditions (Figure 6-3, columns 4:6; 10:12) may indicate that the iron was becoming depleted at this time point.

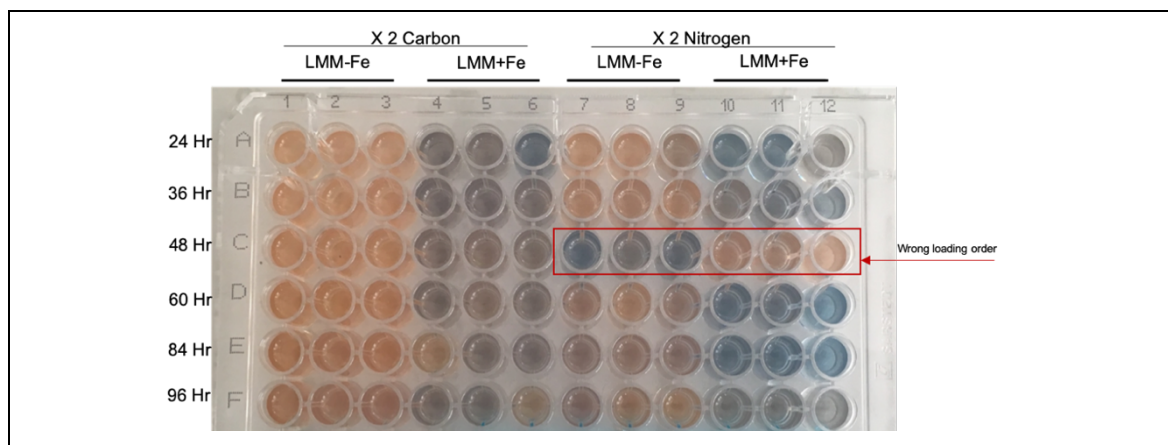


Figure 6-3: A microtitre CAS assay to detect siderophore production in *R. equi* 103S.

In a 96 well plate 100 μ l culture supernatant aliquots from Figure 6-2 were mixed with 100 μ l modified CAS solution and left to incubate for 3 hours before analysis. Colour change from blue to orange is indicative of siderophore production

6.3.2 Medium optimisation 2 - Fiss Minimal media

As Fiss minimal media (FMM) had previously been successfully used for siderophore identification in other rhodococci (Vellore, 2001) a similar trial design to that described above was employed using FMM. In these trials, three different medium compositions, including FMM containing no iron, FMM containing low iron (0.5 μ M FeSO₄) and FMM containing high iron (20 μ M FeSO₄).

However, marked inconsistency in yields indicated that FMM medium should not be considered as the foundation for ongoing experiments. *R. equi* appears to preferentially use carbon sources such as lactate and acetate over glucose, as the latter was recently shown to be both inefficient and to lead to variable outcomes (Anastasi et al., 2016; Vázquez-Boland and Meijer, 2019) which is consistent with the findings here.

6.3.3 Medium optimisation 3 – Modified lactate minimal media

With the inconsistencies of FMM, the decision was made to return back to using lactate minimal media, for direct comparisons with previously published *R. equi* studies (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2005; Miranda-CasoLuengo et al., 2008). Additional measures were made to more reproducibly impose iron limitation, including using acid-washed glassware and Chelex-100 treatment of the liquid medium to sequester traces of iron.

The initial experiment was repeated using these modified LMM (+/- Fe) conditions, in an attempt to replicate the chromophore observed by Miranda-CasoLuengo et al. (2008). After 120 hours, a

noticeable red pigment developed in the iron-limited culture only, with the iron supplemented culture remaining an off-white colour and the uninoculated negative control displaying no growth at all (Figure 6-4). This is unlikely to be the chromophore reported by Miranda-CasoLuengo et al. (2008) as no metal supplement was required to generate the colour. The development of such a colour based on binding of surplus iron in the medium is inconsistent with the elaboration of siderophores.

An additional set of otherwise identical cultures were incubated further and harvested after 300 hours, to monitor the effect of longer starvation periods. All cultures were streaked on LB agar to confirm culture purity, observed colony morphology was entirely consistent with *R. equi*.

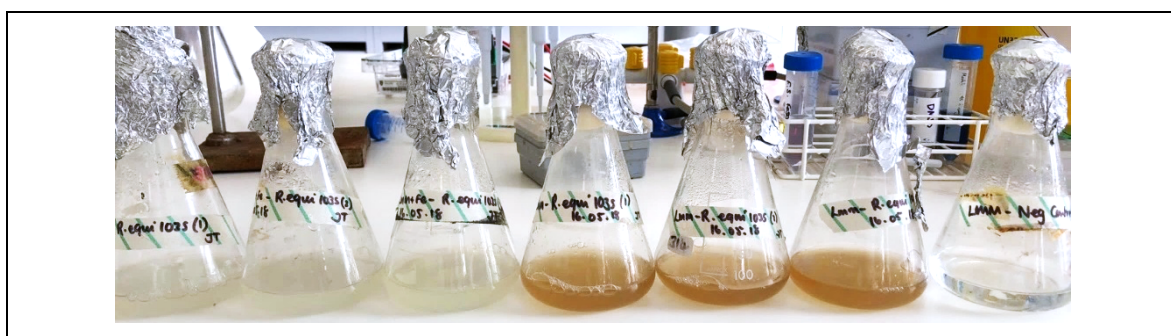


Figure 6-4: Visual inspection of *R. equi* 103S cultures after 120 hours growth under different iron regimes.

Conditions - Left: 3x iron sufficient LMM(+Fe), Right: 3x iron starved LMM (-Fe), and an uninoculated LMM flask.

The cellular material and supernatant were separated by centrifugation, the red pigment was observed in both supernatant and pellets of iron-deficient samples (Figure 6-5).

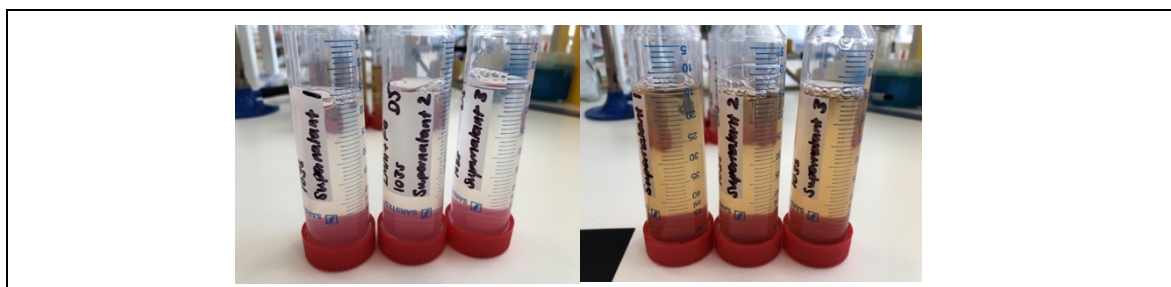


Figure 6-5: Visual inspection of *R. equi* 103S culture supernatants after 120 hours growth under different iron regimes.

Conditions - Left: 3x iron sufficient LMM(+Fe), Right: 3x iron starved LMM (-Fe).

Aliquots of the bacterial supernatants were taken and complexed with 1 % ferric chloride (Figure 6-6AB) to promote formation of ferri-siderophore complexes that might be detected colourimetrically or by mass spectrometry. Surprisingly, upon complexation a precipitate formed in both sets of cultures, more noticeably in the iron-limited supernatants; the precipitates were then collected by centrifugation. Significantly, the red pigment associated with the precipitate pellet,

leaving a notably paler supernatant (Figure 6-6CD). Therefore, the analysis was expanded to profile the metabolites from both the original iron-starved and iron sufficient conditions, and also to investigate the effect of ferric chloride supplementation prior to LC-MS analysis and identify if production of siderophores was evident.

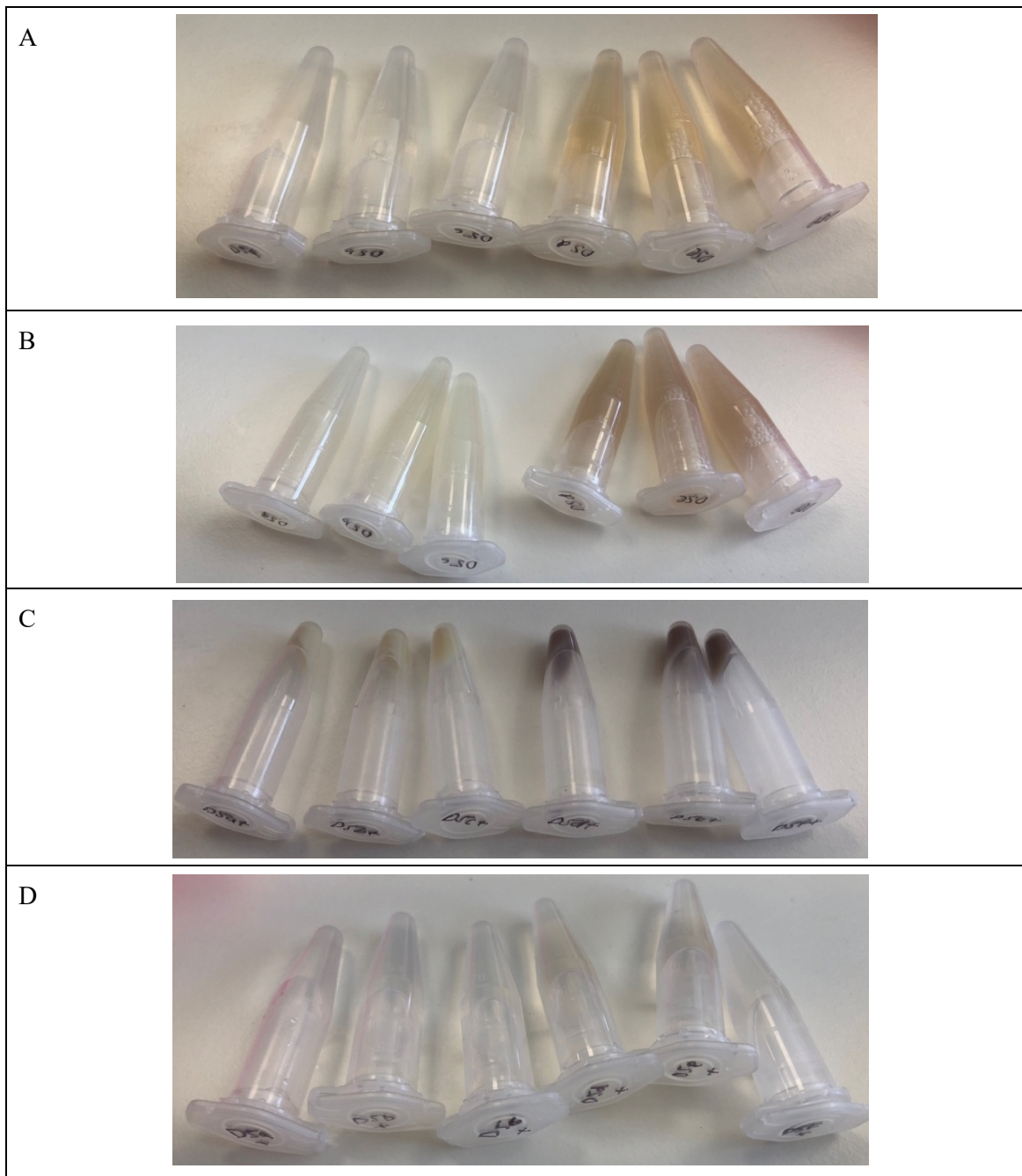


Figure 6-6: Visual inspection of *R. equi* 103S culture supernatants aliquots under different iron regimes with additional iron supplementation.

A) 1 ml supernatant aliquots of *R. equi* 103S: Left: 3x iron sufficient LMM(+Fe), Right: 3x iron starved LMM (-Fe). B) 1 ml supernatant aliquots of *R. equi* 103S: Left: 3x iron sufficient LMM(+Fe) supplemented with 1 % FeCl₃, Right: 3x iron starved LMM (-Fe) supplemented with 1 % FeCl₃. C) Precipitation from B collected by centrifugation. D) Culture supernatant counterparts to the precipitation observed in C.

6.4 Untargeted metabolite profiling using MS1 Analyses

The bacterial supernatant samples were prepared, and separated using both a Reversed Phased (RP) methodology with a C18 column and a HILIC methodology using an amide analytical column linked to an UHPLC-MS, producing chromatograms shown in Figure 6-7 and Figure 6-8 respectively.

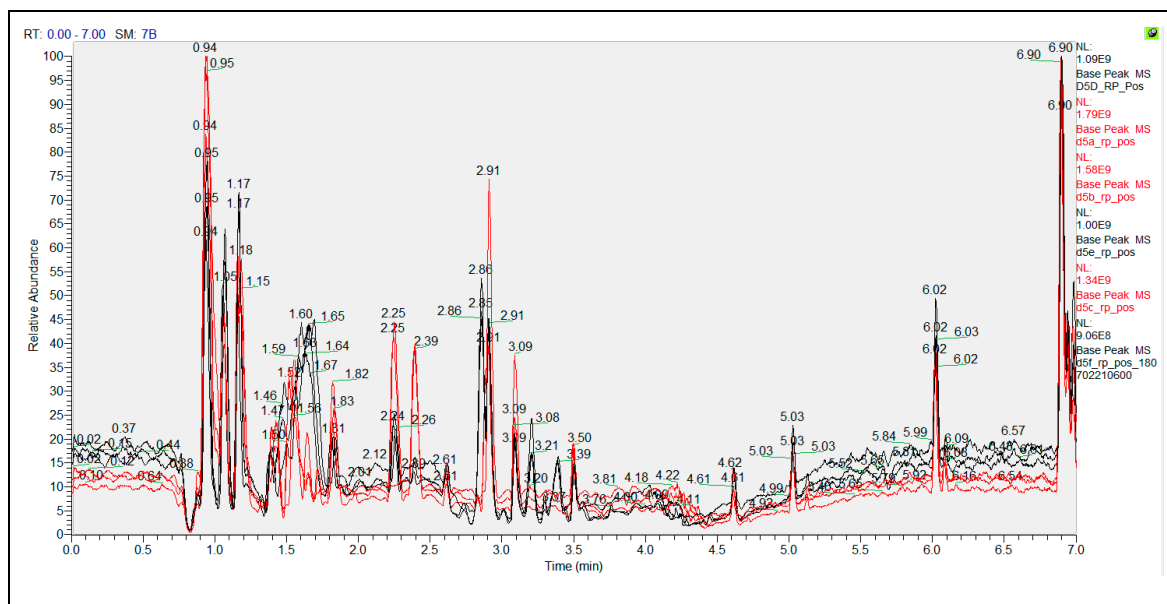


Figure 6-7: Metabolite profiles of *R. equi* 103S using RP (C18) chromatography, positive ionisation mode.

Conditions used to elicit a biological response were iron-limited lactate minimal media (LMM-Fe) (black; D5D, D5E and D5F) and iron-sufficient lactate minimal media (LMM+Fe) (red; D5A, D5B and D5C)

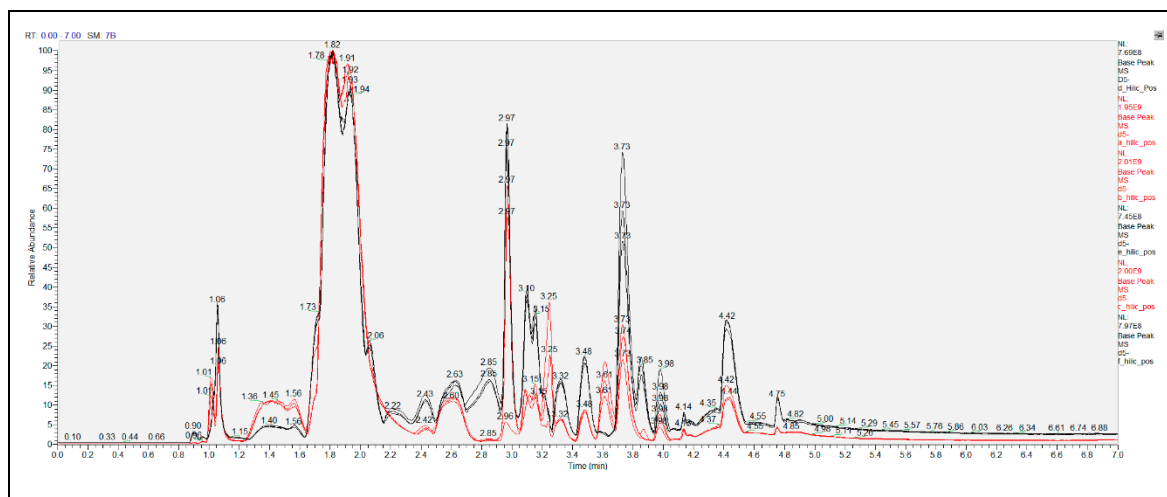


Figure 6-8: Metabolite profiles of *R. equi* 103S using HILIC (amide) chromatography, positive ionisation mode.

Conditions used to elicit a biological response were iron-limited lactate minimal media (LMM-Fe) (black; D5D, D5E and D5F) and iron-sufficient lactate minimal media (LMM+Fe) (red; D5A, D5B and D5C)

The RP chromatograph was analysed further using Compound Discoverer 3.1 as described in section 2.14.2. The dataset peak intensity tables were organised into rows of unpaired samples with the aim of identifying the metabolomic profile of each of the samples and how significantly they vary; to do this the “Statistical Analysis” module of MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca>) was utilised. Data uploading was performed as described in section 2.14.3, including missing values, filtering, normalisation, transformation and scaling of the datasets. The data normalisation was essential to reduce systematic bias within the data set and to improve overall data consistency so that meaningful biological comparisons can be made (Figure 6-9) (Mizuno et al., 2017).

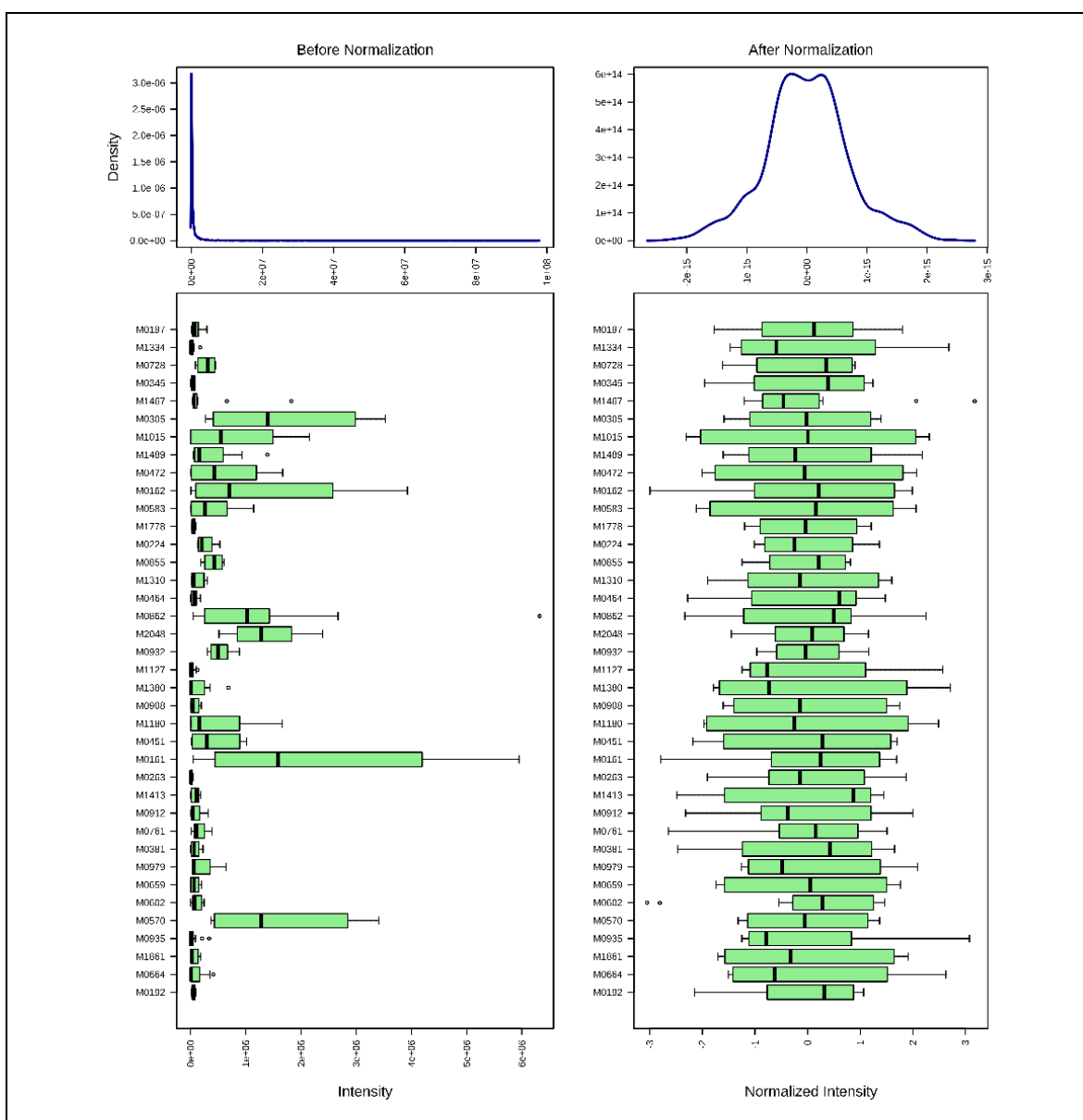


Figure 6-9: Box plots and kernel density plots before and after normalization. The boxplots show at most 50 features due to space limit. The density plots are based on all samples. Selected methods: Data transformation: Log Normalisation; Data scaling: Pareto Scaling.

6.4.1 Chemometric analysis

The Principle Component Analysis algorithm was used as an unsupervised method to visually determine if a distinct class separation was achieved in the multivariate data (Figure 6-10).

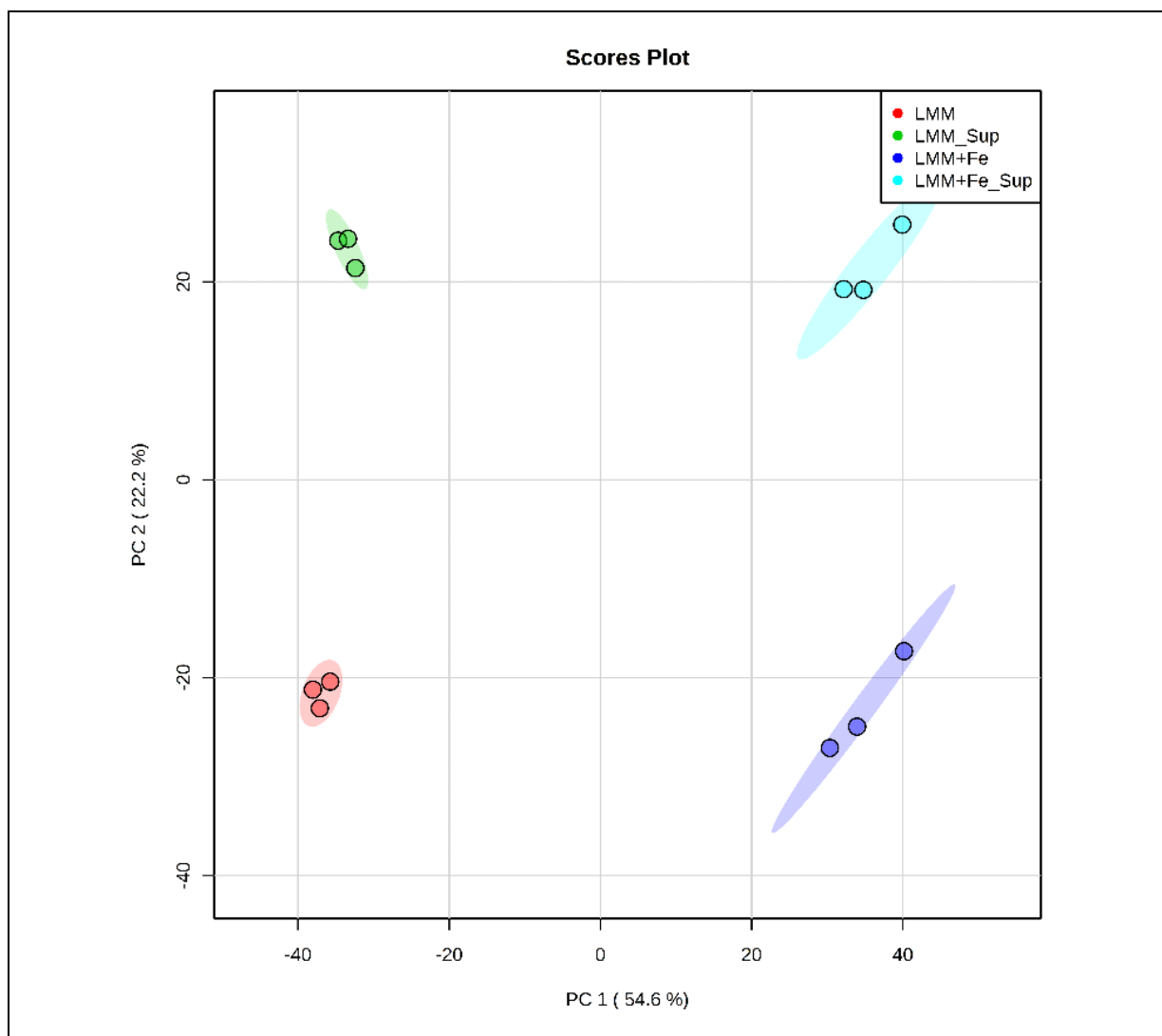


Figure 6-10: 2D Principal component analysis for unsupervised variance analysis between sample classes of *R. equi* 103S

Red: iron limited samples; Green: iron limited samples with FeCl₃ supplementation Dark Blue: iron sufficient samples; Light Blue: iron sufficient samples with FeCl₃ supplementation.

It was clear that even using an unsupervised method that there was a distinct metabolic variation between the four sample classes, with the first three principal components accounting for 89.3 % of variance within the dataset. As expected, the PC1 accounts for 54.6 % variance observed, and is likely related to the biological differences occurring during growth in the different medium conditions. PC2 however, accounts for 22.2 % of variance observed and appears to be derived from the addition of FeCl₃ to the bacterial supernatants and the associated molecular interactions and substance precipitations (refer to Figure 6-6).

6.4.2 Data analysis – Initial candidates

Prior to any further statistical analysis, a rationale was established to scout the dataset to identify if any metabolites conform to expectations for a siderophore-based biological response, specifically providing evidence for a secreted compound generated under iron-limited conditions, that is not present under iron-replete conditions, that after addition of iron the metabolite abundance changes due to metal complexation and as a result a mass change (56 Da) occurs.

This initial manual data analysis was performed by sorting the average peak intensity table by abundance in the iron-deplete conditions (LMM) where we would expect to see the metabolites of interest; special interest was paid to masses made earlier in the thesis regarding putative rhequibactin and rhequichelin structures.

Table 6-1 details the top 25 most abundant metabolites detected in spent LMM; of these 18 do not follow expected trends and were discarded. The remaining 7 metabolites (highlighted in Table 6-1) all appear to be more abundant in the LMM conditions than in LMM+Fe.

Table 6-1: Peak intensity table for the metabolite profiling of *R. equi* 103S using RP LC-MS methodology; ranked by average intensity in LMM conditions.

Molecular Weight			Sample Class	AVERAGE LMM+Fe_Sup	AVERAGE LMM+Fe	AVERAGE LMM_Sup	AVERAGE LMM
M	[M+H]	RT [min]	Metabolite				
562.21514	563.22242	1.626	M0537	16,454,485	33,335,196	160,510,104	180,853,188
273.26706	274.27434	6.895	M1758	123,295,570	54,458,475	104,097,665	48,209,431
166.08566	167.09294	0.945	M0154	97,261,581	101,874,029	51,455,463	39,819,122
562.21526	563.22254	1.485	M0496	1,881,814	1,961,948	28,729,896	35,209,407
434.12633	435.13361	2.857	M0872	2,701	4,063	1,549,688	32,262,119
286.14194	287.14922	6.981	M1770	24,810,184	37,393,016	27,132,311	31,566,796
180.1013	181.10858	1.068	M0222	26,596,747	40,127,412	31,573,921	29,910,706
201.1731	202.18038	1.165	M0288	20,379,552	39,898,627	23,892,009	27,849,386
563.21788	564.22516	1.632	M0539	1,412,771	2,703,450	15,670,614	23,900,299
301.18903	302.19631	2.903	M0898	14,740,921	34,370,931	15,712,678	23,368,754
317.29246	318.29974	6.942	M1762	39,039,465	31,351,414	36,551,378	22,089,356
245.23574	246.24302	6.021	M1681	24,026,918	26,767,137	22,614,555	20,490,974
299.18492	300.1922	2.904	M0899	8,394,651	14,207,560	9,700,305	19,702,877
253.15415	254.16143	1.553	M0518	28,522,533	31,779,363	15,659,620	18,025,740
562.21555	563.22283	1.558	M0520	1,051,625	3,134,061	39,783,881	17,484,282
234.11177	235.11905	1.165	M0289	8,840,640	10,898,204	14,328,157	16,891,531
148.05268	149.05996	2.254	M0706	30,995,327	38,594,736	14,897,822	16,529,527
165.0792	166.08648	2.256	M0709	30,737,316	38,296,265	14,761,273	16,370,176
136.0387	137.04598	1.068	M0223	34,525,718	26,056,880	24,751,873	15,706,430
212.11628	213.12356	1.235	M0335	14,681,444	19,932,152	20,663,397	14,941,861
262.08922	263.0965	1.817	M0585	27,688,971	32,791,190	15,669,046	14,786,760
264.10485	265.11213	3.381	M1169	2,287,912	2,371,950	10,476,412	14,472,399
484.35202	485.3593	3.083	M0990	16,468,703	30,759,197	12,789,806	13,934,894
462.13268	463.13996	3.199	M1067	7,425	6,421	19,264,297	11,413,071
219.11084	220.11812	2.615	M0766	13,489,118	14,624,645	9,748,333	10,144,174

Interestingly, metabolites M0537 (Figure 6-11), M0496 and M0520 are likely to be the same compound being detected at different retention times (i.e. exhibiting a broad elution profile) given the minor difference in observed masses (< 1 ppm). These generally follow the inclusion criteria in that they are detected at 5.4-, 17.9- and 5.6-fold higher abundance in LMM over LMM+Fe. Additionally, a similar observed metabolite M0539, has an 8.8-fold increase in abundance under the same conditions, however it exhibits a mass increase of 1.00233, that is likely to have arisen from formation of an [M+2H] adduct.

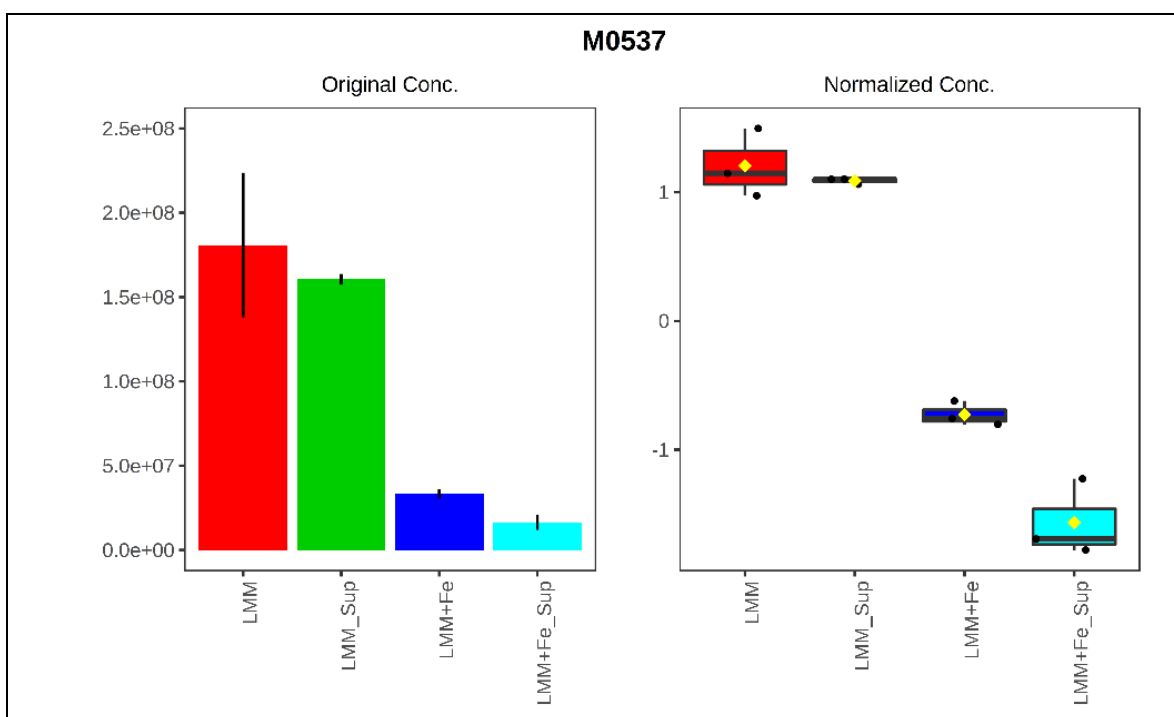


Figure 6-11: Loading values of mass spectral feature M0537.

The bar plots on the left show the original values (mean +/- SD). The box and whisker plots on the right summarise the normalised values. This feature was upregulated in both iron limited conditions.

Another metabolite that meets the inclusion criteria is M1169 which exhibits a 6.1-fold increase for LMM over LMM+Fe, however given the low mass observed (265.11213 [M+H]), it is unlikely to represent a siderophore. Lastly, metabolites M0872 (Figure 6-12) and M1067 (Figure 6-13) both represent clear siderophore candidates given their marked increase in abundance in LMM over LMM+Fe, being approximately 7940 and 1777-fold respectively.

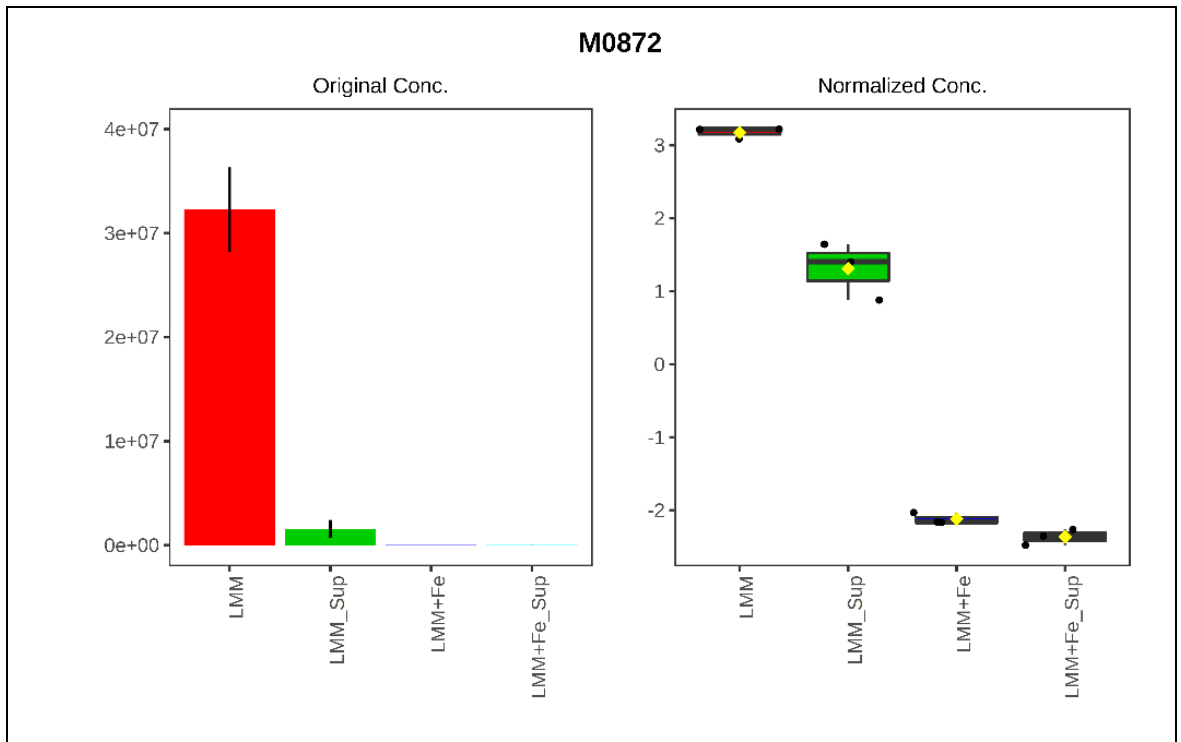


Figure 6-12: Loading values of mass spectral feature M0872.

The bar plots on the left show the original values (mean +/- SD). The box and whisker plots on the right summarise the normalised values. This feature was upregulated predominantly in iron limited conditions without supplementation.

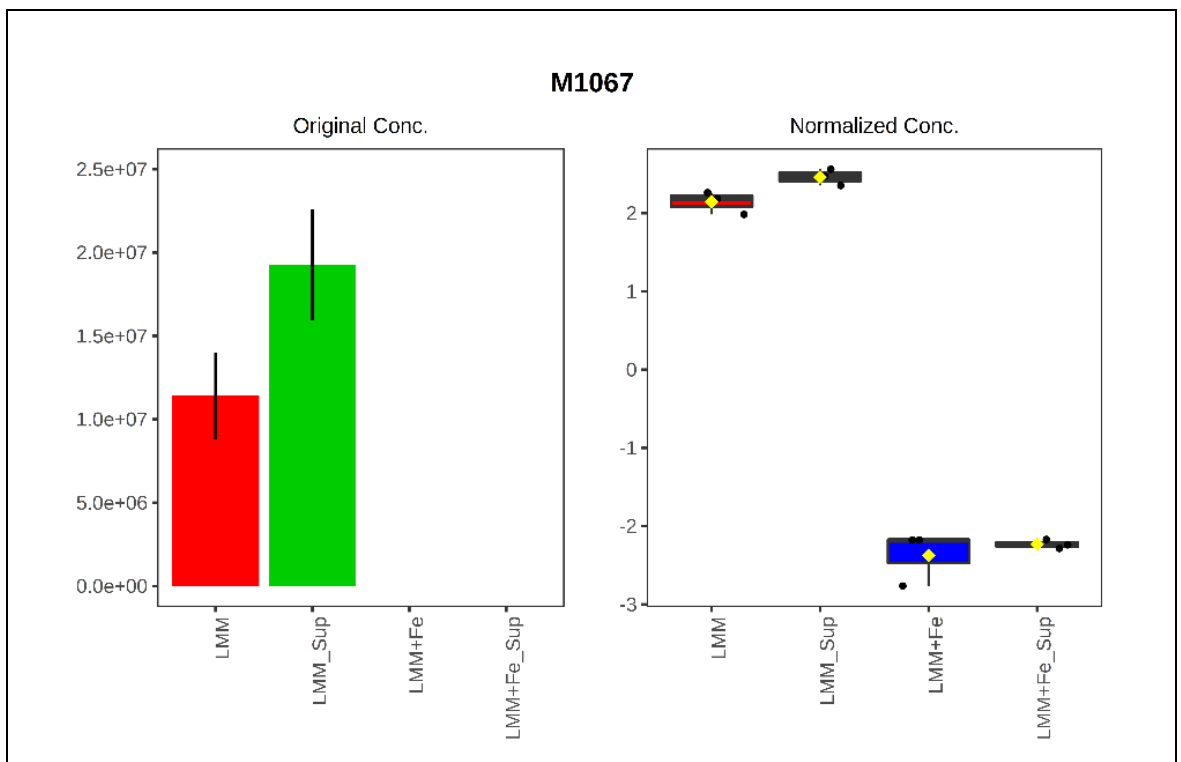


Figure 6-13: Loading values of mass spectral feature M1067.

The bar plots on the left show the original values (mean +/- SD). The box and whisker plots on the right summarise the normalised values. This feature was upregulated in both iron limited conditions.

6.4.2.1 Extracted ion chromatographs

Extracted ion chromatographs (XIC) were generated for each of these siderophore candidates using the Compound Discoverer 3.1 output to confirm their legitimacy.

The first analysis indicated that M0496, M0520 and M0537 are the same metabolite given the similar column retention times and masses as shown in Figure 6-14. Similarly, M0539 appears to be an additional compound due to the increased mass size of 1.00322, however analysis of the XIC indicates that this metabolite is likely the same compound (as M0537), differentiated due to different adducts formed relating to the mass to charge (m/z) ratio observed on-column for a different isotopic variation.

Although, metabolite M1169 appeared to follow the expectations of a siderophore, this was also excluded given a misrepresentation of data as an additional compound of equivalent mass was present at a different retention time displaying increased abundance levels in iron-sufficient conditions (data not shown).

The metabolite M0872 was explored further, and also appeared to follow expectations for a siderophore; the XIC is consistent with the secretion of this metabolite under iron limited conditions, and is not present under replete conditions (Figure 6-15). Furthermore, the abundance of this metabolite appears to diminish upon addition of ferric chloride, this could be accounted to either ferri-siderophore complexation that would be expected to be seen elsewhere on the chromatograph or the precipitation observed.

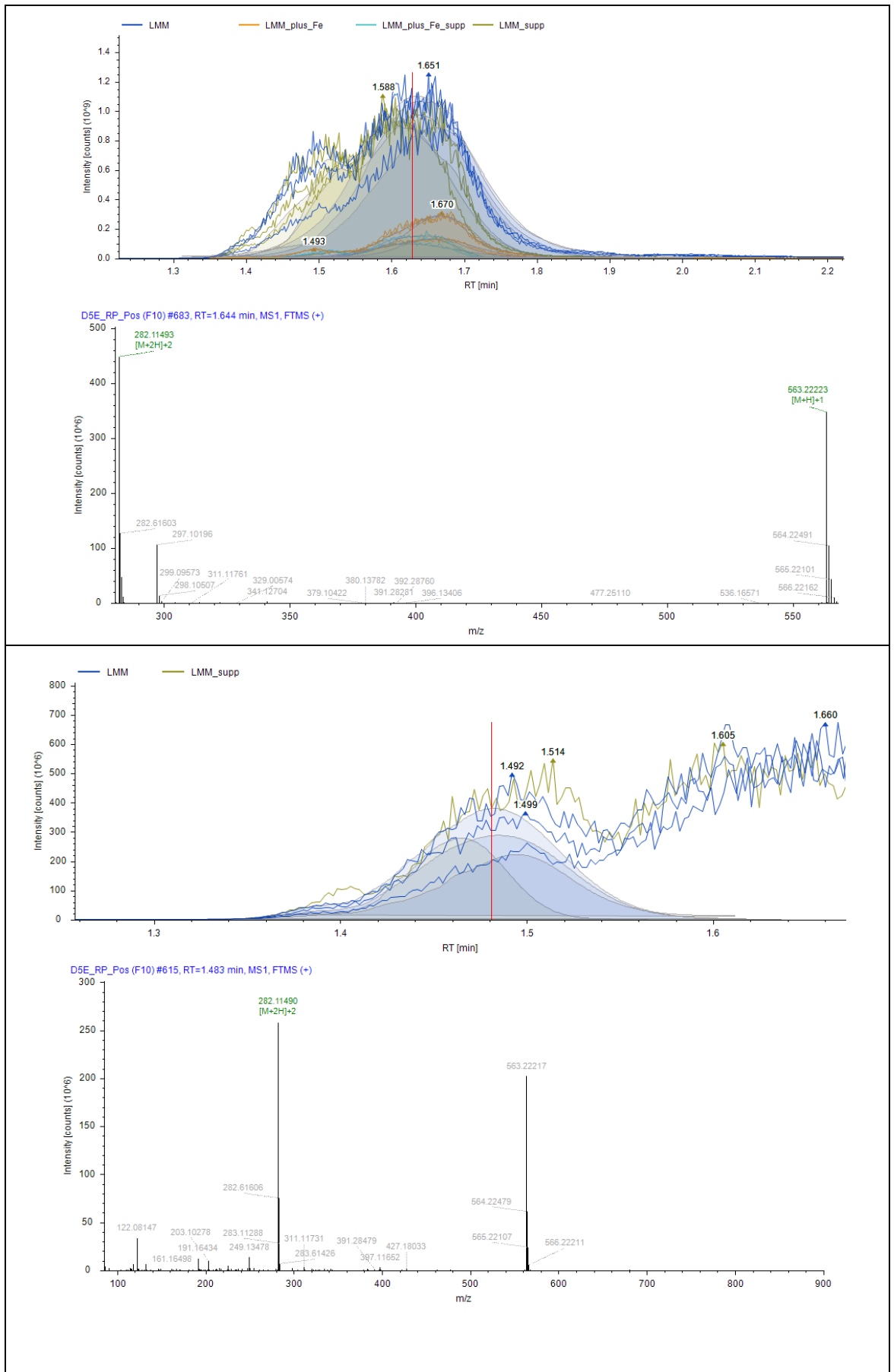


Figure 6-14: Extracted ion chromatographs for M0537
Spectral peaks identified for mass 563.22223 [M+H]⁺ at different m/z ratios at a retention time of 1.4 and 1.6 minutes for the top and bottom spectral analysis respectively.

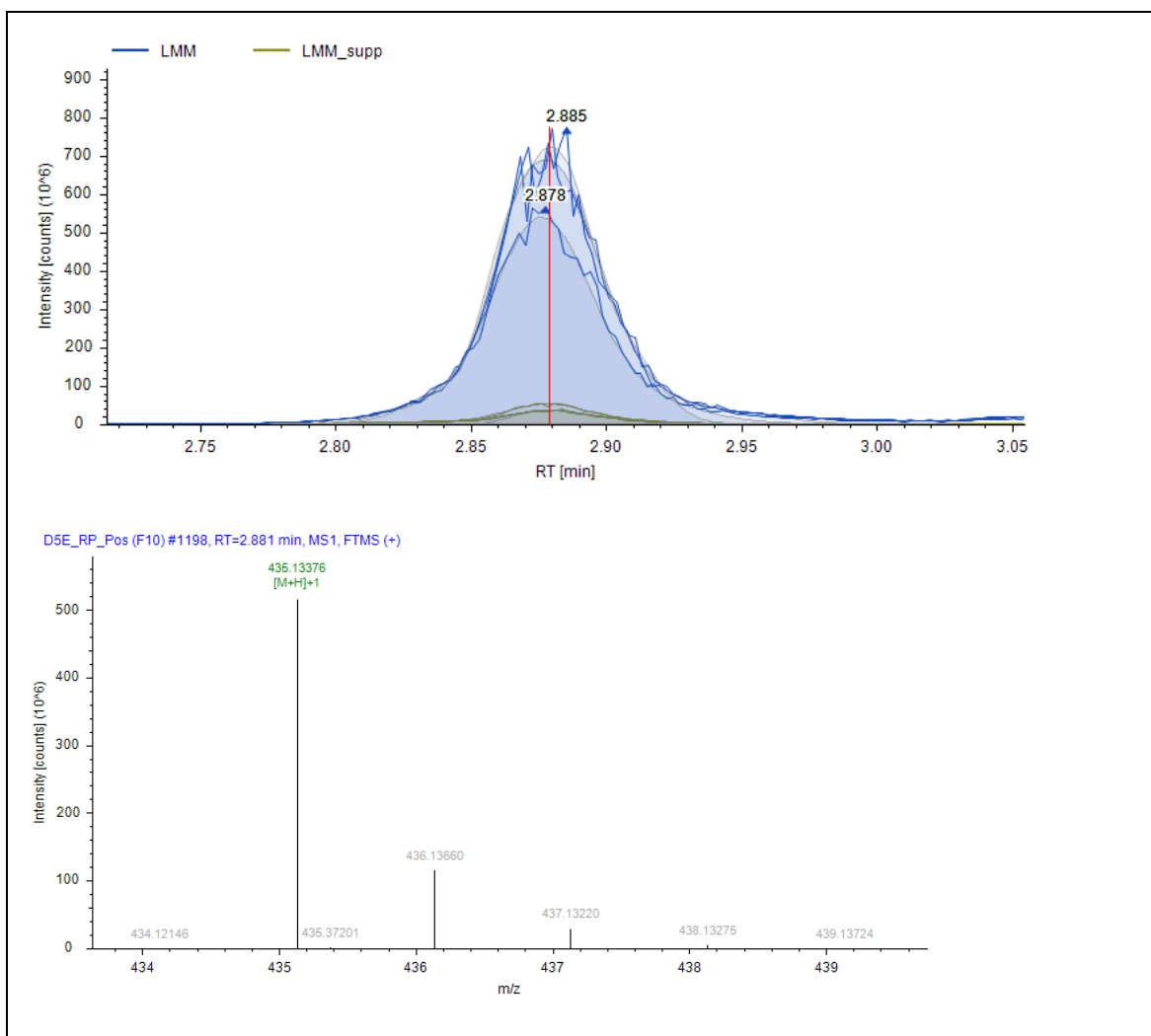


Figure 6-15: Extracted ion chromatographs for M0872

Spectral peak identified for mass 435.13376 [M+H] under iron limited conditions, at a retention time of 2.86 minutes.

A similar observation was made for M1067; it appeared to be secreted under iron limited conditions and not present under replete conditions. However, unlike M0872, M1067 appears to be unresponsive to the ferric chloride supplementation. The difference in observed masses between M0872 and M1067 is 28.00635, which does not correspond to any known positive-mode ionisation adducts but could be consistent with a biosynthetic relationship and modification of M1067 by an acyl chain extended by inclusion of an additional [CH₂]₂. It could be postulated that this may improve metabolite solubility allowing it to avoid precipitation.

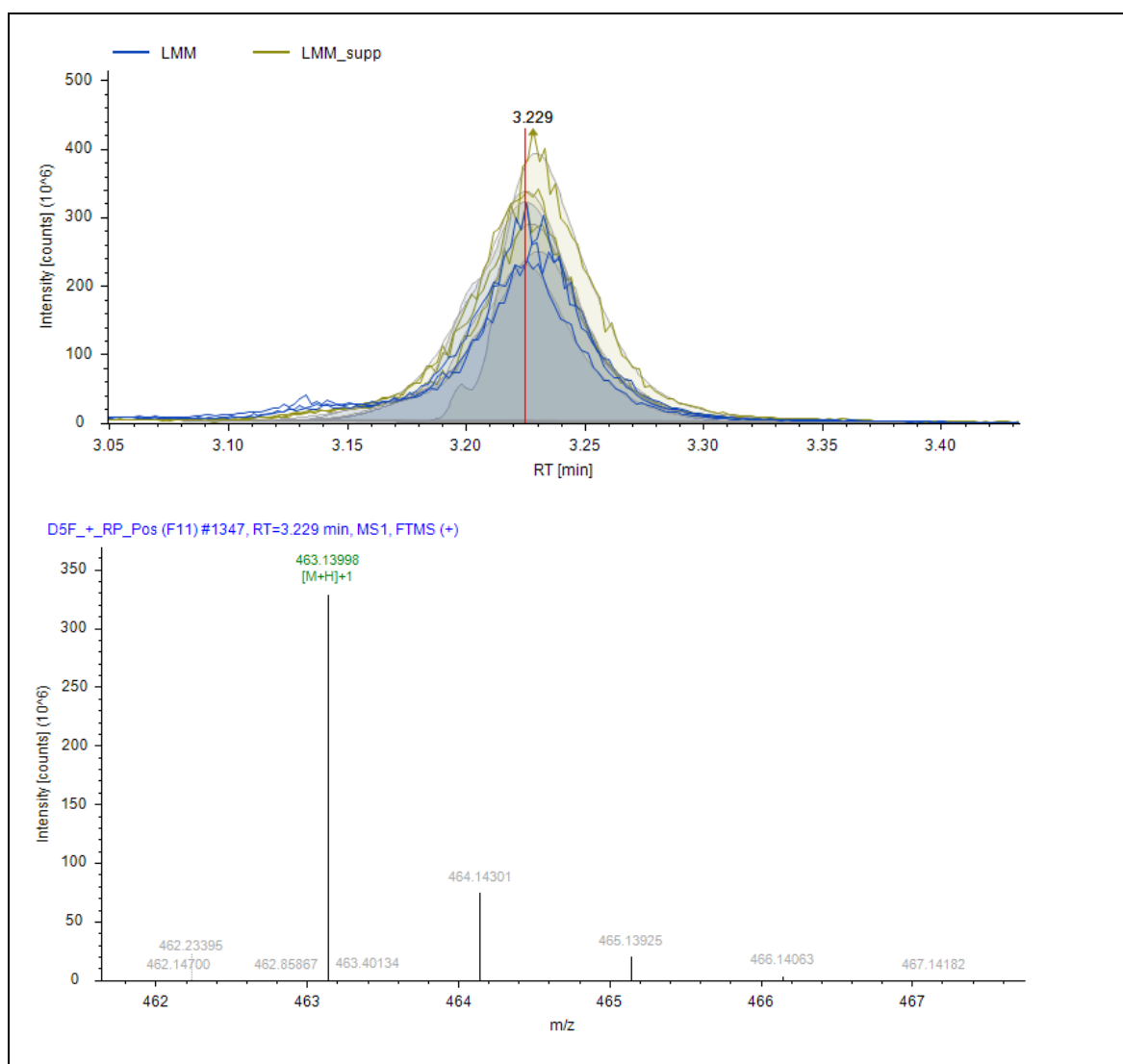


Figure 6-16: Extracted ion chromatographs for M1067
Spectral peak identified for mass 463.13998 [M+H]⁺ under both iron limited conditions, at a retention time of 3.229 minutes.

6.4.3 Feature selection – metabolite ranking by statistical significance

With over 2000 metabolites detected; it was logical to reduce the complexity to facilitate further analysis, this was performed by identifying the most significant features and corroborating that the results follow the same variance trend.

6.4.3.1 Significance Analysis of Microarray (& Metabolites)

The Significance Analysis of Microarray (SAM) test is a well-established statistical test to identify differentially expressed genes, or in this case – metabolites. Although LC-MS analyses predictably differ vastly from the DNA microarray investigations for which it was developed, some statistical methods for data analysis including SAM can be used in analogous approaches (Roxas and Li, 2008;

Xia et al., 2009). For differential expression determination, the metabolite abundance data can be analysed identically to both protein and gene abundance data.

The SAM plot was developed to address sample false discovery rate issues when performing multiple tests on high-dimensional data. Firstly, a significance score is assigned to each variable based on its change relative to the standard deviation of repeated measurements, then the algorithm chooses variables with scores greater than the adjustable 'Delta' threshold and compares their relative difference to the distribution estimated by random permutations of the class labels. For each defined threshold value, a certain amount of the variables in the permutation set will be found to be significant by chance (Xia et al., 2009; Xia and Wishart, 2016). Furthermore, SAM plots can be used in multi-class analyses, where metabolites will be indicated if there are significant differences in abundance across some combination of the groups, with scores being based on F-statistics (Tusher et al., 2001).

An extremely conservative delta value selected ($\Delta = 1.2$) was based upon parameter adjustment to minimise the degree of false discovery rates given per dataset (data not shown) as suggested by Xia and Wishart (2016), where the given delta value indicates 409 significant metabolites with a false discovery rate of 0.001, with 1 (1.01) false positive. The metabolites were ranked according to their SAM d value correlating with the coordinates displayed in Figure 6-17, of which the top 50 metabolites are represented in Table 6-2.

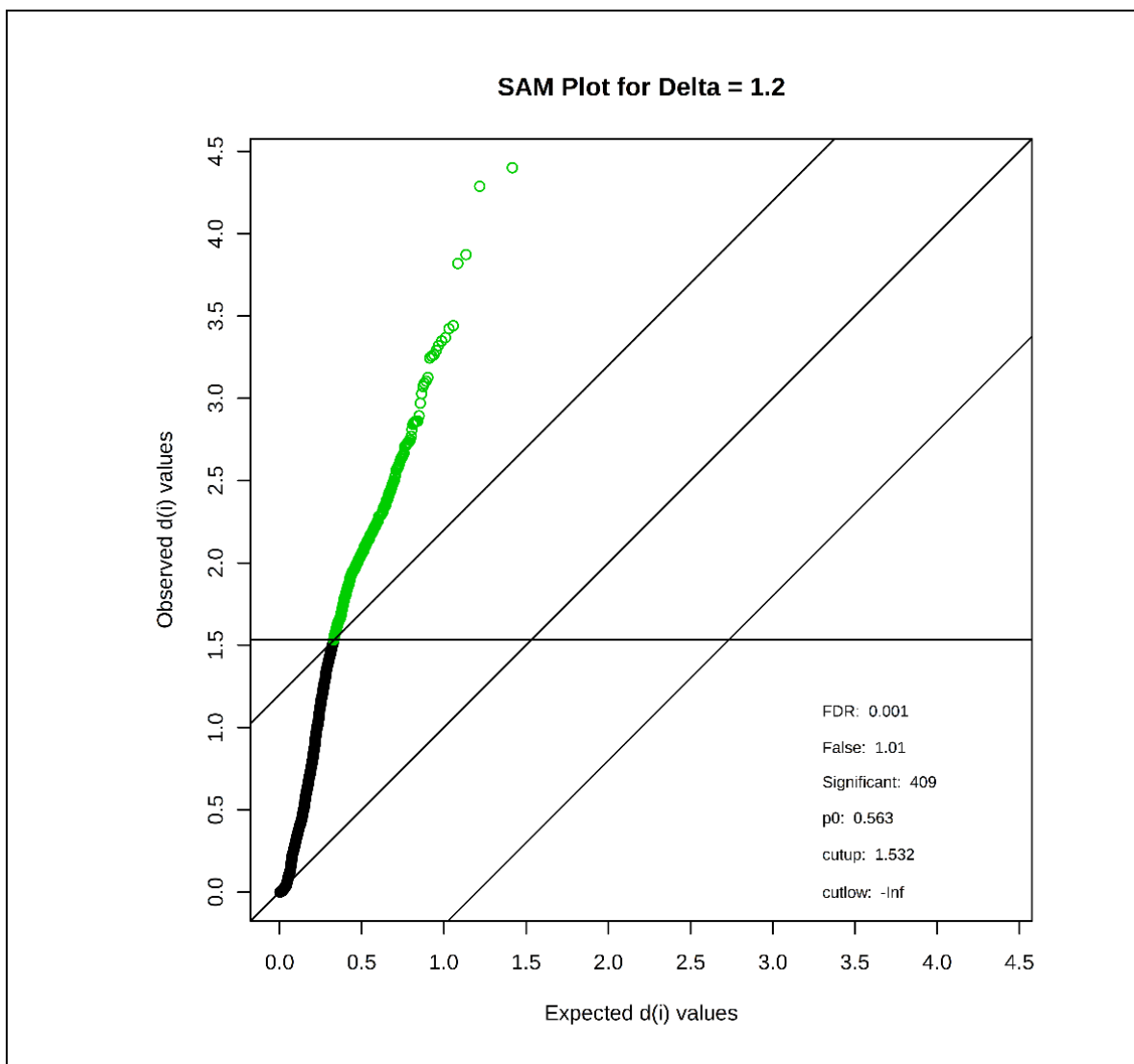


Figure 6-17: Significant features identified by SAM.

The green circles represent features that exceed the specified threshold, where the more a variable deviate from the observed = expected line (centre diagonal line) the more likely it is to be significant.

The most significant features identified by SAM appeared to correspond to the initial candidates identified, of which the most significant appear to be M0872 (Figure 6-12) and M1067 (Figure 6-13) ranking first and second respectively. Additionally, ranking metabolites 3, 4 and 5 appear to be isotopic variations with slightly differing column retention times, but importantly all follow the expected trends, being more abundant in deplete media conditions over replete media conditions.

While no other metabolites of interest are common between the two datasets, corresponding masses may be present for SAM-ranked metabolites 29 (M1483, m/z 564.15563 [M+H], RT : 4.117) and 45 (M1507, m/z 563.19251 [M+H], RT : 1.198) to the metabolites M0539 (m/z 564.22516 [M+H], RT : 1.632) and M0537 (m/z 563.22242 [M+H], RT : 1.626) which are the most abundant metabolite overall in LMM media conditions that may be eluting broadly.

Other statistically ranked metabolites of interest identified here that do not correlate with the initial peak intensity table (Table 6-1) potentially include M0994 (SAM ranked 6th), M1350 (SAM ranked 10th), M0924 (SAM ranked 19th), M0667 (SAM ranked 31st) and M0968 (SAM ranked 46th), however these metabolites appear to exhibit have considerably lower fold changes for observed expression levels between LMM and LMM+Fe than the main candidates M0872 and M1067.

Table 6-2: Peak intensity table for the metabolite profiling of *R. equi* 103S using RP LC-MS methodology; significant features ranked by SAM.

Top 50 significant features and average abundance of samples presented due to space limitations.

SAM		Molecular Weight		RT [min]	Metabolite	Sample	AV (S1:3)	AV (S4:6)	AV (S7:9)	AV (S10:12)
Rank	d value	M	[M+H]			Class	LMM+Fe_Sup	LMM+Fe	LMM_Sup	LMM
1	4.401	434.12633	435.13361	2.857	M0872	2,701	4,063	1,549,688	32,262,119	
2	4.2878	462.13268	463.13996	3.199	M1067	7,425	6,421	19,264,297	11,413,071	
3	3.8728	463.11321	464.12049	3.111	M1006	934	1,368	892,302	1,612,502	
4	3.8191	462.13262	463.1399	2.754	M0816	3,310	3,844	3,488,342	3,113,432	
5	3.441	463.11525	464.12253	3.131	M1015	2,842	3,623	1,147,146	1,889,884	
6	3.4227	445.10599	446.11327	3.088	M0994	6,442	9,499	4,477,684	2,967,142	
7	3.3683	511.49714	512.50442	10.341	M2076	3,431	2,620,710	3,569	808,815	
8	3.348	376.14931	377.15659	1.342	M0400	4,831,635	7,486,824	4,234	188,396	
9	3.3198	644.1872	645.19448	1.434	M0469	2,789,779	2,700,425	6,421	7,594	
10	3.2907	760.25693	761.26421	3.745	M1350	7,147	8,467	2,934,581	2,607,648	
11	3.2641	439.16818	440.17546	4.311	M1539	1,120	632,278	1,143	301,094	
12	3.2558	229.13296	230.14024	2.707	M0792	1,139,552	1,497,370	3,186	3,270	
13	3.2446	411.13653	412.14381	4.232	M1519	1,168	1,407,978	639	42,145	
14	3.1263	481.14273	482.15001	4.117	M1484	2,780	422,293	1,347	612,374	
15	3.1041	362.14133	363.14861	2.982	M0934	2,929	3,309	780,573	822,849	
16	3.0908	357.11223	358.11951	2.079	M0668	3,629	4,763	590,277	1,984,953	
17	3.0719	230.11696	231.12424	3.006	M0947	16,061	17,284	3,883,520	4,161,851	
18	3.0276	554.14786	555.15514	3.392	M1180	3,490	3,722	351,590	1,503,123	
19	2.9699	654.20424	655.21152	2.959	M0924	8,661	10,273	2,707,843	1,260,378	
20	2.8936	435.13672	436.144	3.703	M1330	693,567	3,350	1,107,352	7,557	
21	2.8618	789.40633	790.41361	4.875	M1595	3,641	2,478	538,549	581,595	
22	2.861	205.0853	206.09258	1.578	M0525	7,276	9,311	1,236,405	1,635,812	
23	2.8607	419.12644	420.13372	3.005	M0945	3,563	6,953	504,076	1,219,413	
24	2.8564	872.30554	873.31282	4.111	M1479	1,447	94,312	732	331,588	
25	2.8525	538.15303	539.16031	3.961	M1427	2,523	3,017	537,669	371,128	
26	2.8418	576.20122	577.2085	4.547	M1574	1,264	1,223	181,932	223,028	
27	2.8101	453.14485	454.15213	4.121	M1485	569	7,574	261	197,588	
28	2.7691	606.22447	607.23175	3.031	M0956	603,662	728,735	4,200	4,265	
29	2.7474	563.14835	564.15563	4.117	M1483	2,708	3,108	90,683	1,038,898	
30	2.7454	632.22008	633.22736	2.322	M0718	3,355	3,465	356,372	654,888	
31	2.7326	714.22505	715.23233	2.079	M0667	7,338	9,630	558,905	2,281,629	
32	2.7302	467.43443	468.44171	9.444	M1984	3,982	632,514	2,362	294,316	
33	2.7233	436.15276	437.16004	4.113	M1480	2,520	125,824	1,223	398,481	
34	2.7202	337.11718	338.12446	4.528	M1570	822	743	352,889	4,621	
35	2.7104	698.24139	699.24867	2.827	M0852	7,566	9,587	825,877	1,545,828	
36	2.7102	469.14248	470.14976	3.417	M1191	487,957	3,815	913,188	8,126	
37	2.7041	261.18461	262.19189	3.245	M1089	426,936	577,701	4,190	3,772	
38	2.666	237.05723	238.06451	1.239	M0338	1,222,677	7,349	309,475	4,604	
39	2.6594	437.13714	438.14442	1.356	M0409	6,550	7,267	793,968	801,302	
40	2.6479	219.10093	220.10821	2.513	M0746	6,704	1,360,744	2,666	108,236	
41	2.6467	414.2042	415.21148	7.611	M1854	4,982	1,526,833	28,007	1,319,901	
42	2.6362	396.10994	397.11722	1.505	M0504	13,442	6,650	2,126,672	345,803	
43	2.6321	413.19527	414.20255	4.768	M1591	110,741	114,471	1,058	915	
44	2.6206	724.14381	725.15109	1.437	M0472	1,165,007	1,249,503	10,746	10,522	
45	2.6051	562.18523	563.19251	4.198	M1507	1,781	2,249	183,173	267,775	
46	2.5958	699.22561	700.23289	3.055	M0968	21,809	59,802	3,431,617	3,620,427	
47	2.5854	208.09615	209.10343	1.171	M0297	573,468	16,645	2,682,273	11,804	
48	2.5837	387.63384	388.64112	3.526	M1240	7,944	8,447	1,066,314	619,251	
49	2.5767	610.18511	611.19239	4.575	M1575	838	1,026	86,164	102,855	
50	2.5684	437.15254	438.15982	4.122	M1488	2,515	109,862	1,161	238,562	

6.4.4 Potential biomarker identification for a targeted analysis

The aforementioned metabolites have potential to be used as biomarkers when utilising a targeted approach to identifying and characterising the *R. equi* siderophores. Both peak intensity tables would serve as a candidate screening list for confirmation of successful chromatographic separations.

Of particular interest, is the metabolite M0872 (Figure 6-12) as the reduced peak intensities after iron addition appears to correlate with the precipitation observed in Figure 6-6C, reminiscent of the heterobactins that displayed very low aqueous solubility and were initially isolated in their native des-ferri forms (Carrano et al., 2001). These heterobactin characteristics are resonant with M0872, given the observed precipitation event occurring upon addition of ferric chloride, further implying this candidate to be the potential catecholate siderophore observed by Miranda-CasoLuengo et al. (2008). Moreover, given that this metabolite has an unusual red chromophore, it is logical to assume that it would be possible to track the metabolite through various purification stages, if adequately concentrated, with subsequent MS confirmation. This is an unusual feature of a potential siderophore, as they would be expected to develop a red colour on iron complexation only. The red colour of the more soluble, uncomplexed molecule would likely facilitate progress.

6.5 A targeted approach to characterise the *Rhodococcus equi* 103 siderophores

6.5.1 Large scale production of *R. equi*

As previously noted, the level of siderophores synthesised by *R. equi* was not detectable by conventional assay methods but was achieved with the application of the microtitre CAS assay. Consequently, given the tenuous siderophore production it was clear that large-scale production of iron-limited culture would be necessary for subsequent characterisation.

Therefore, the iron-limited conditions (LMM-Fe) identified earlier were replicated and scaled up accordingly, to produce 1 litre of bacterial culture. The culture was harvested after 5 days, when the red chromophore was apparent (Figure 6-18). Cellular material was separated by centrifugation and discarded, while the supernatant was retained and concentrated to 50 ml under vacuum *via* rotary

evaporation; this concentrated supernatant was then retested via microtitre CAS assay (Figure 6-18B and C respectively).

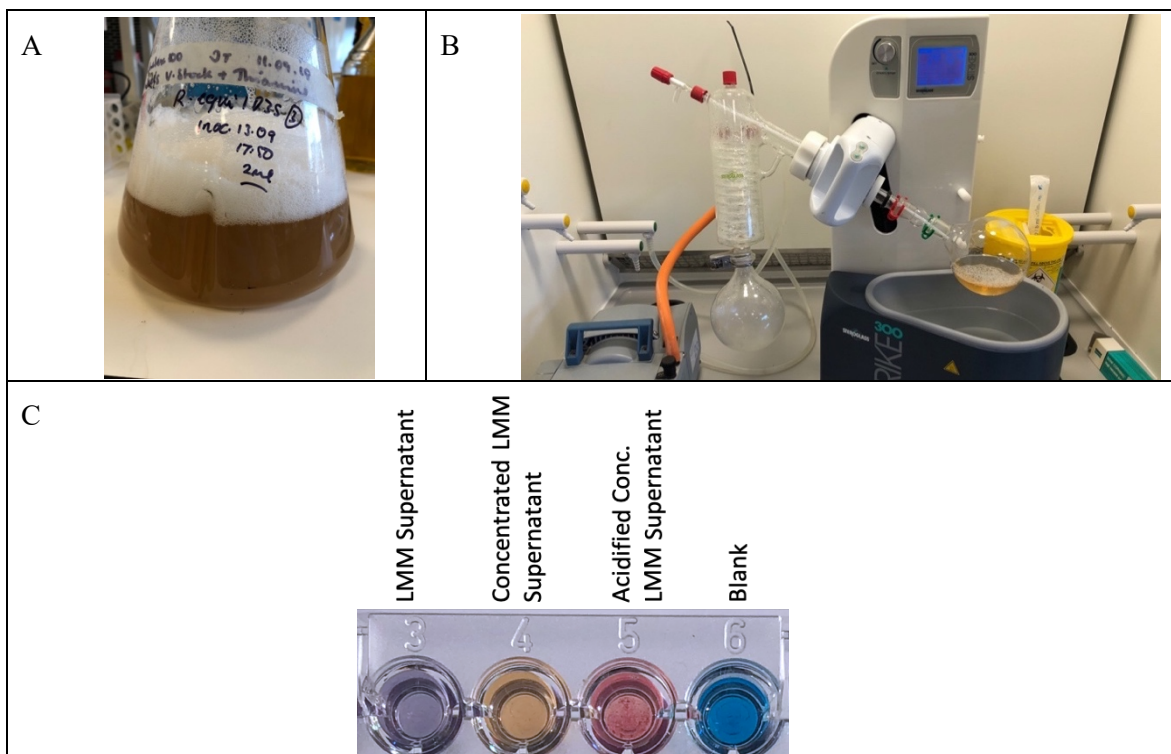


Figure 6-18: Visual stages of the large-scale production of iron limited *R. equi* 103S

A) 1 litre of LMM-Fe culture ready for harvesting indicated by the red pigment. B) Rotary evaporation of the iron-limited supernatant. C) Modified CAS assay of supernatants before and after concentration and acidification.

The microtitre CAS assay indicates that the method of supernatant concentration was effective in improving detection of siderophore production. The addition of hydrochloric acid to the concentrate produced a red pigment that impairs the detection; however, acidification was performed to facilitate downstream column retention, and the pH level was altered upon column washing and elution.

6.5.2 Isolation of the first candidate – M0872

As a first stage purification, the *R. equi* 103S supernatant was adjusted to pH 2.0 to aid column retention. The acidified supernatant was loaded on to Amberlite XAD-2 adsorption resin pre-equilibrated with H₂O at a rate of 4.0 ml / min. The column was washed copiously with H₂O to remove any unbound material and the flow through was collected in 50 ml fractions. After sufficient washing, retained materials were eluted using 50 / 50 % (v/v) methanol/water followed by 100 % methanol.

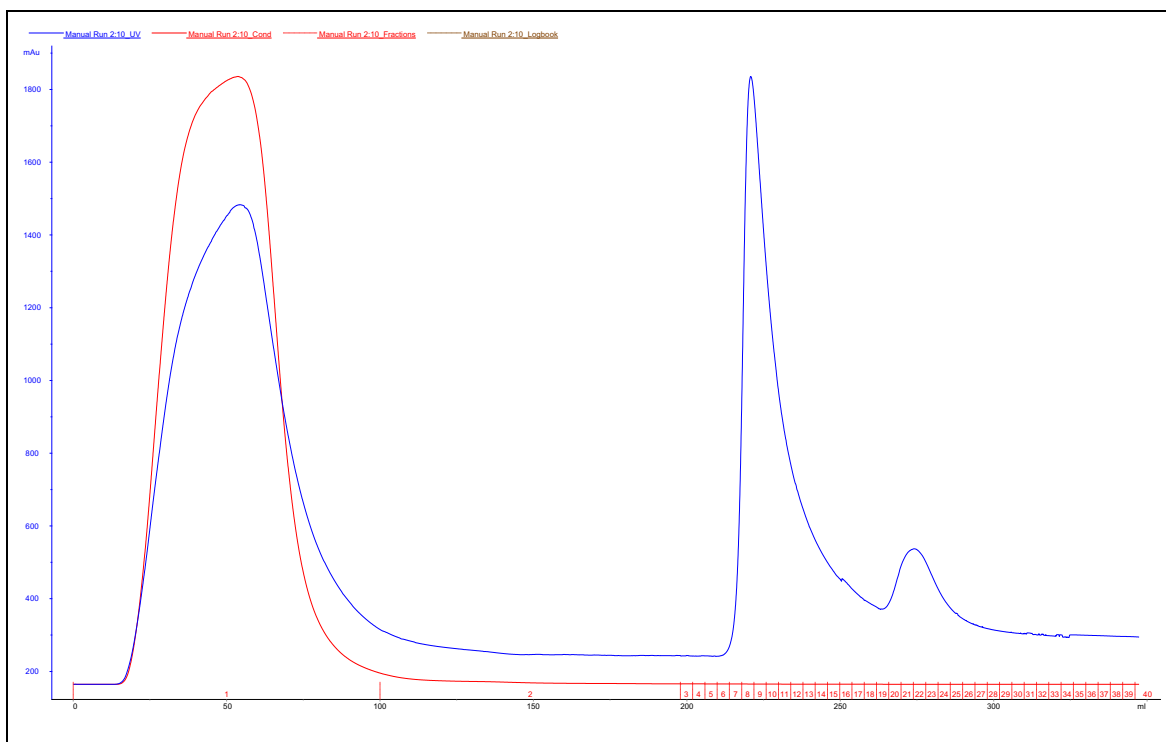


Figure 6-19: Chromatograph of FPLC purification of bacterial supernatant using XAD-2 polymeric resin.

Displaying the UV and conductivity for the sample loading, elution and corresponding fractions.

Interestingly, during the loading of the column, the resin appeared to retain the pigment, that was eluted in 50 % MeOH along with UV absorbing material (Figure 6-19). Fraction 8 onwards suggest an elution of target compounds, including the apparent chromophore of interest given the pigment transfer (Figure 6-20).

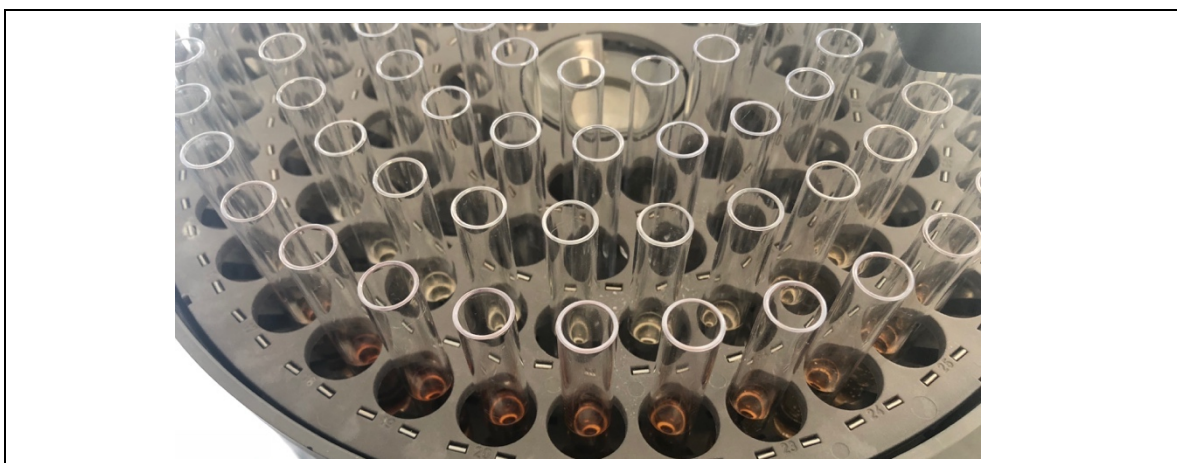


Figure 6-20: Fractions eluted from the XAD-2 resin purification of iron limited *R. equi* 103S
 Fractions retaining the red pigment correlating with the peak observed in Figure 6-19

Thus, the fractions of interest were tested using the microtitre CAS assay (Figure 6-21), of which fractions 8 - 13 displayed a clear positive reaction in iron extraction from the dye that in fractions associated with corroborate with the primary UV absorbing eluants observed in Figure 6-19. It is

also noteworthy that the secondary peak observed in fractions 22-25 also exhibit a reaction when complexed with CAS solution, therefore these fractions required further scrutiny.

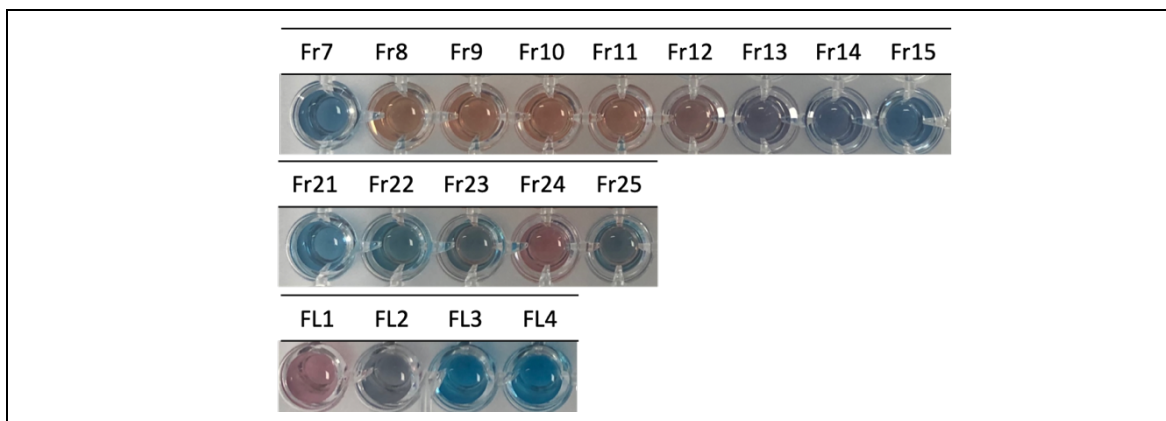


Figure 6-21: Microtitre CAS assay for fractions of interest from the XAD-2 resin purification of iron limited *R. equi* 103S

A visual examination of fraction 8 prior to and after supplementation with 1/10th volume of 1% FeCl₃, suggested the formation of a ferri-siderophore complex, as a further dark-grey pigment became evident that also produced a precipitant (Figure 6-22). Next, an aliquot of both fraction 8 and the iron supplemented fraction 8 were resuspended in a variety of solvents to identify if the pigments and thus the siderophore metabolites, could be extracted in the solvent phases (Rane et al., 2005), without success. The composition of both samples was examined using LC-MS to identify siderophore candidates in both ferri- and desferri- forms.

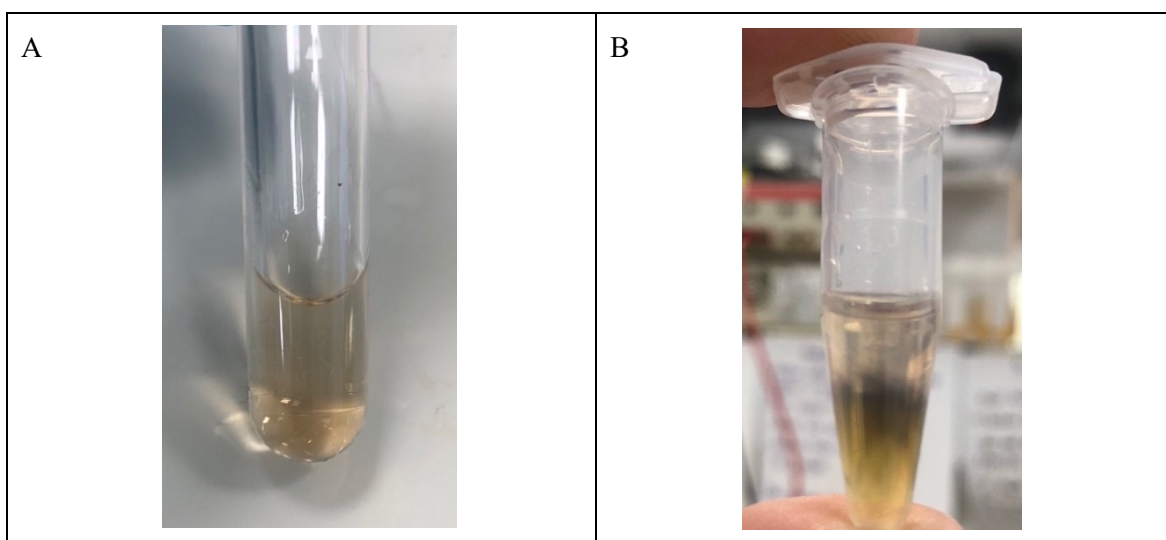


Figure 6-22: Visual examination of Fraction 8 from the XAD-2 resin purification of iron limited *R. equi* 103S

A) Fraction 8. B) Fraction 8 + 1/10th volume of 1% FeCl₃,

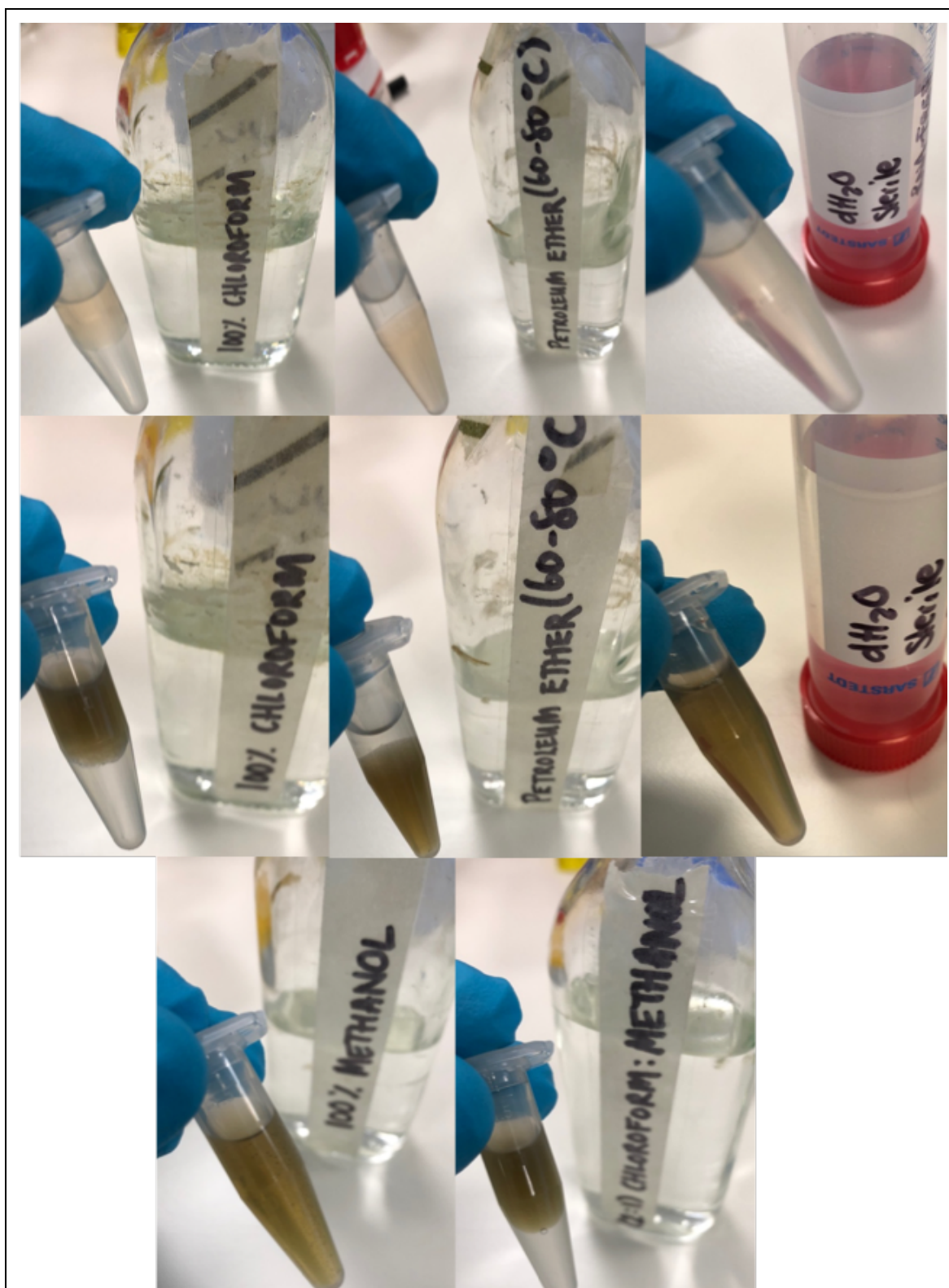


Figure 6-23: Attempted solvent extractions for hypothetical ferri- and des-ferri siderophore fractions.

Solvent extractions performed combining 0.5 ml solvent phase with 0.5 ml fraction of interest. Solvents used include chloroform, methanol, petroleum ether (60-80°C boiling point), H₂O and 2:1 chloroform: methanol.

LC-MS analysis of the bacterial supernatants was performed, including the LMM supernatant, the concentrated LMM supernatant and concentrated acidified LMM supernatant. The positive ionisation mode profiles appeared to be consistent throughout, including the acidified supernatants which were analysed to identify if the acidification process altered the chromatography, however no major peak changes were observed.

The flow through fractions were analysed to identify any compounds of interest that did not bind to XAD-2 resin; these mainly consisted of low molecular weight metabolites outside of the range of interest (data not shown).

The LC-MS analysis of Fraction 8 identifies one predominant peak that had a column retention time of ~2.8 minutes, and a corresponding mass of 435.1346 $[M+H]^+$ (Figure 6-25), this ion is found in subsequent fractions in decreasing abundance, which correlates with the siderophore activity in the microtitre assay in Figure 6-21. Rationally, given the retention of this mass in the fraction of interest, the pigment transfer and CAS activity for this fraction, it is likely that this metabolite previously identified as the candidate M0872 acts as a siderophore compound.

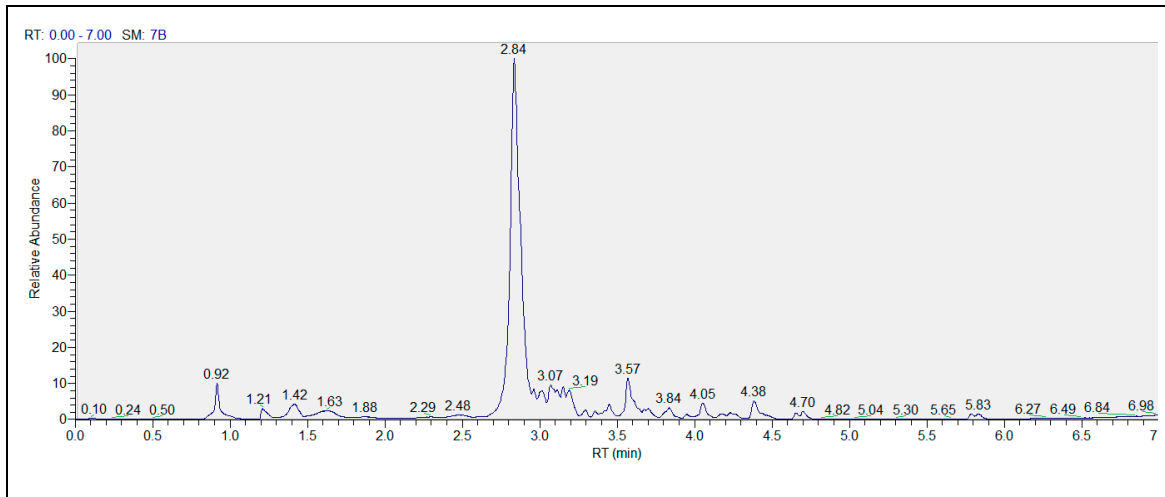


Figure 6-24: Total ion chromatograph for LC/MS analysis of Fraction 8.

TIC fixed normalised abundance level to the highest chromatograph for comparison Peak of interest detected at the retention time 2.84 minutes

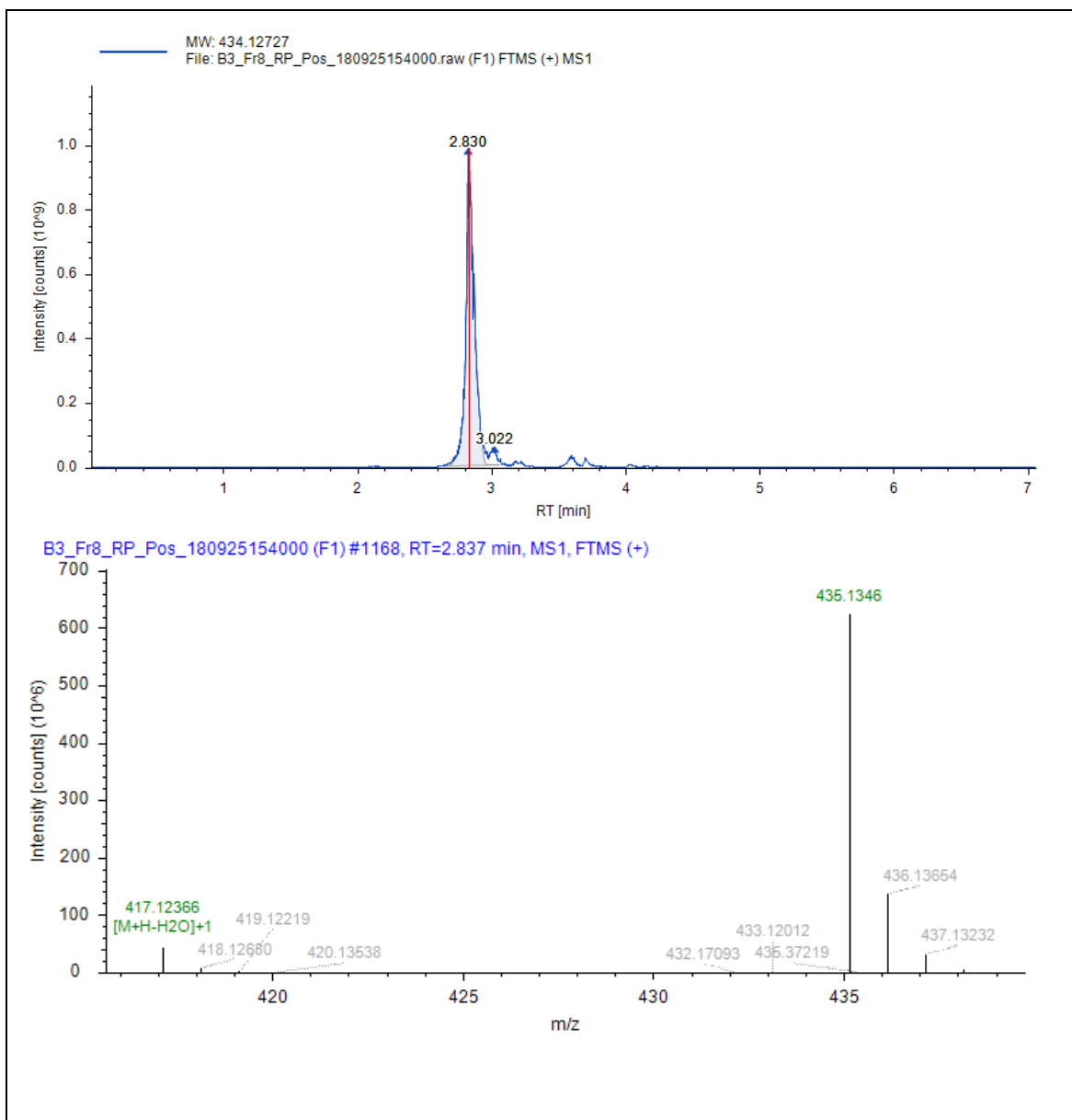


Figure 6-25: Extracted ion chromatogram and MS spectra in Fraction 8.

Peak of interest detected at the retention time 2.83 minutes and mass detection of 435.1346 $[M+H]^+$ and 417.12366 $[M+H-H_2O]^+$

When $FeCl_3$ was added to a sample of Fraction 8, the LC-MS chromatography profile was altered by removal of the significant peak at 2.84 mins, previous identified to have the mass of 435.1346 $[M+H]^+$ (Figure 6-26). This outcome of this experiment is wholly consistent with that described in Chapter 4.

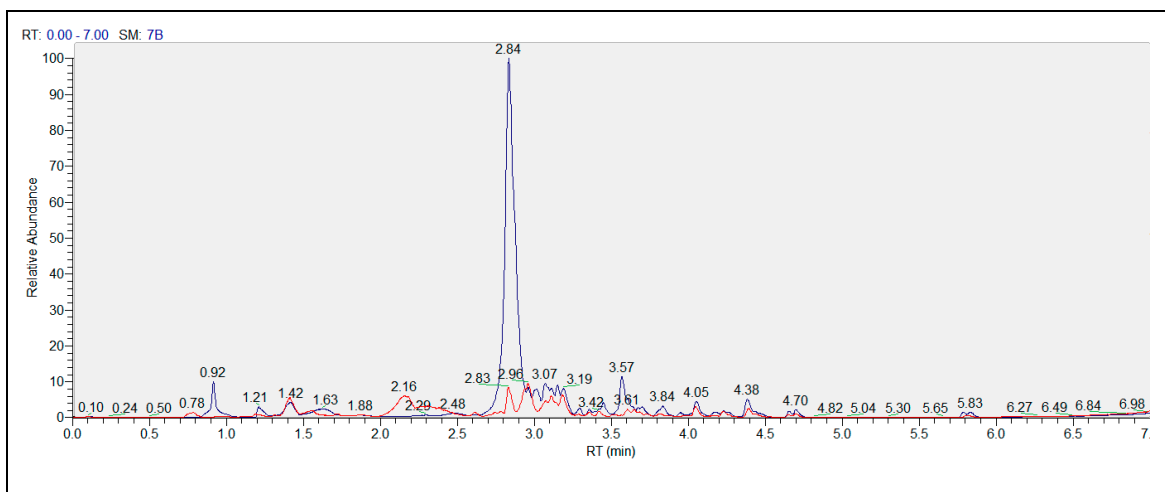


Figure 6-26: Total Ion Chromatograph comparison of Fraction 8 and Fraction 8+FeCl₃. TIC fixed normalised abundance level to the highest chromatograph for comparison. Fr8 shown in Blue, Fr8+ shown in Red.

Although it is unconventional for a siderophore candidate to precipitate upon complexing of iron, however this metabolite should not be excluded solely on this basis given that it has been previously established that the heterobactin siderophores have very low aqueous solubility for iron complexes at neutral to acidic pH (Carrano et al., 2001). Furthermore, addition of an iron solution at this concentration *in vitro* is not representative of the conditions observed in a natural microbial environment.

Further exploration of this candidate included MS² fragmentation pattern analysis, but this experiment did not produce useable data. While the compound achieved sufficient chromatographic separation for MS² analysis, the analysis was restricted due to limitations of the LC-MS system utilising an outdated version of software. As a result, identification and characterisation methods of these metabolites were not considered feasible, as the software available performs fragmentation using ThermoScientific's proprietary normalised collision energy (NCE) ionisation setting rather than the more widely utilised electron volts (eV) methodology, therefore analyte fragmentation would predominantly not be comparable to fragmentation of other compounds in the more widely used spectral databases online which is considered standard practice (Kind and Fiehn, 2010). This is significant as NCE fragmentation takes into account the increased ability of larger molecules to absorb energy when compared to those with lower molecular weights, whereas a low energy eV methodology would result in a greater observed fragmentation for a small molecular weight metabolite compared to a large molecular weight compound.

Rather than database comparisons, other attempts for fragmentation pattern analysis were considered, including the use of a *de-novo* interpretation software for metabolite mass spectra, SIRIUS (Böcker et al., 2008); this software seeks to find the sum formula of the molecule, whose isotopic pattern best matches the fragmentation input. Interestingly however, the authors (Böcker et al., 2008) comment that “this method cannot be used as-is to identify peptides or amino acid compositions, because certain sum formulas correspond to multiple peptides”.

Going forward, it may be useful to collaborate with other departments familiar with definition of siderophore structures that would be able to help decipher the MS² fragmentation patterns of future candidates.

6.5.3 Targeted identification of other potential siderophore candidates

In order to identify any other potential siderophore candidates the large-scale iron-limited culture production was performed as described earlier, until the red chromophore was detected. Cellular biomass was discarded after centrifugation, and supernatant was concentrated as previously described. At this stage, the concentrated supernatant was complexed with 1% FeCl₃ acidified with HCl then clarified to remove precipitants.

The rationale for complexation at this stage was to partly saturate the siderophore candidates, in an attempt to preferentially remove the metabolite M0872 that was previously identified to have somewhat low aqueous solubility when complexed with an iron solution.

The complexed supernatant was subsequently loaded on the AKTA FPLC controlled Amberlite XAD-2 column as previously described, using the same wash and elution procedures, the chromatogram is shown in Figure 6-27.

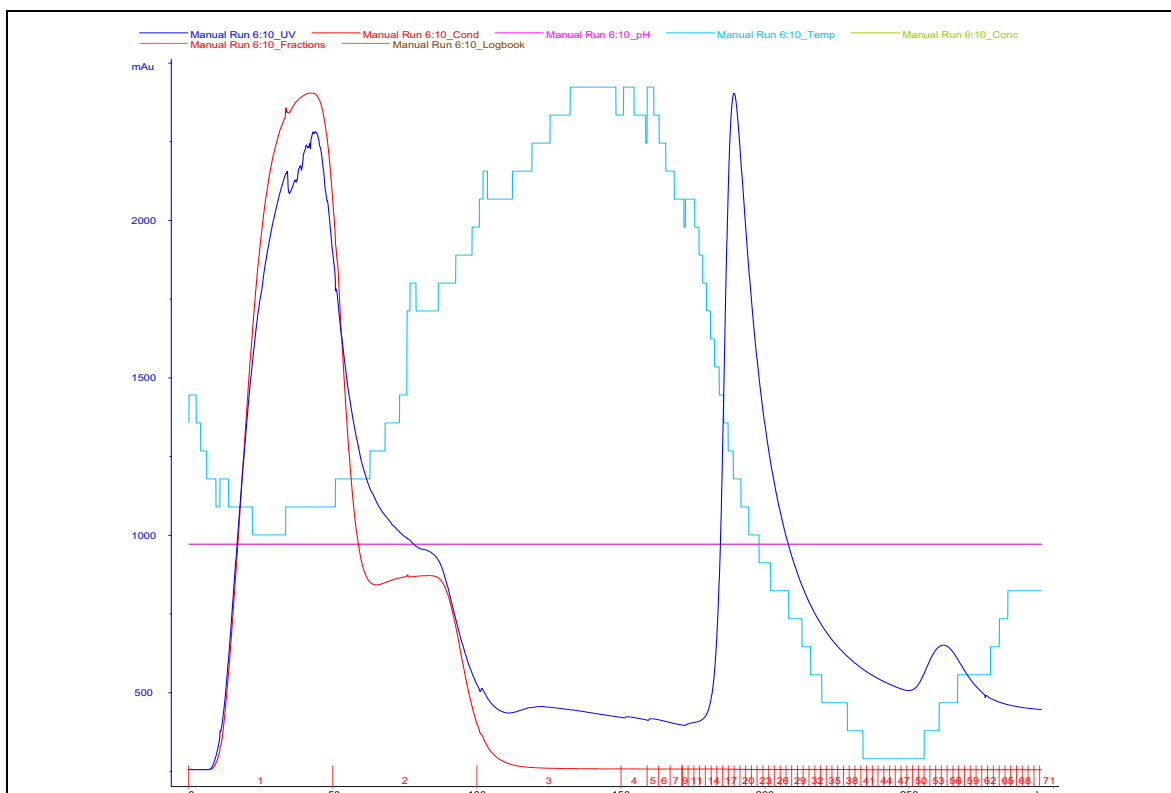


Figure 6-27: XAD-2 column load, wash and elute for secondary candidate identification

Every eluted fraction from the purification was tested via microtitre CAS assay to determine siderophore activity. Interestingly, enough activity appears to correlate with chromatogram UV absorbance peak (Figure 6-27). Furthermore, even with previous partial iron-complexation to hypothetically precipitate M0872, there appeared to be sufficient siderophore activity to facilitate a positive colorimetric change in this sensitive microtitre form of the CAS assay (Figure 6-28) .

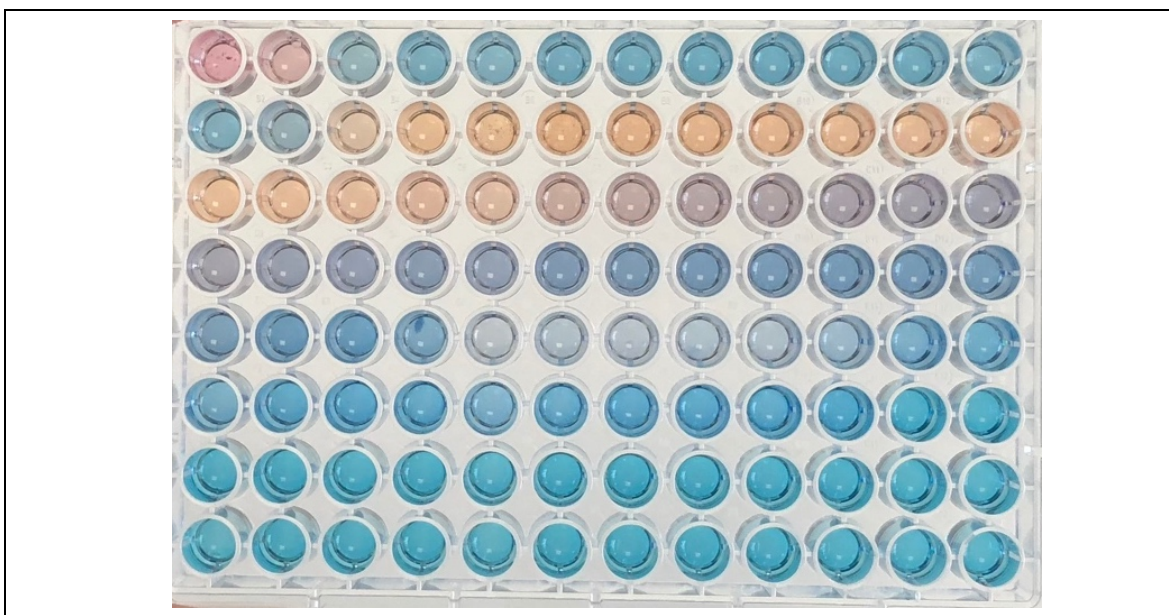


Figure 6-28: Microtitre CAS assay for the second candidate identification

Assay-positive fractions were pooled together, and tested using the less sensitive CAS agar method,

however this yield only a very slight colour change. To improve detection of potential siderophore compounds, the pooled sample was concentrated *via* rotary evaporation as described earlier. The pooled fractions and pool concentrate were re-tested via CAS agar, producing a significant detectable colorimetric change Figure 6-29 that had not previously been observed. This also likely confirms our previous intention of only partial saturation of iron to facilitate removal of the M0872 metabolite.

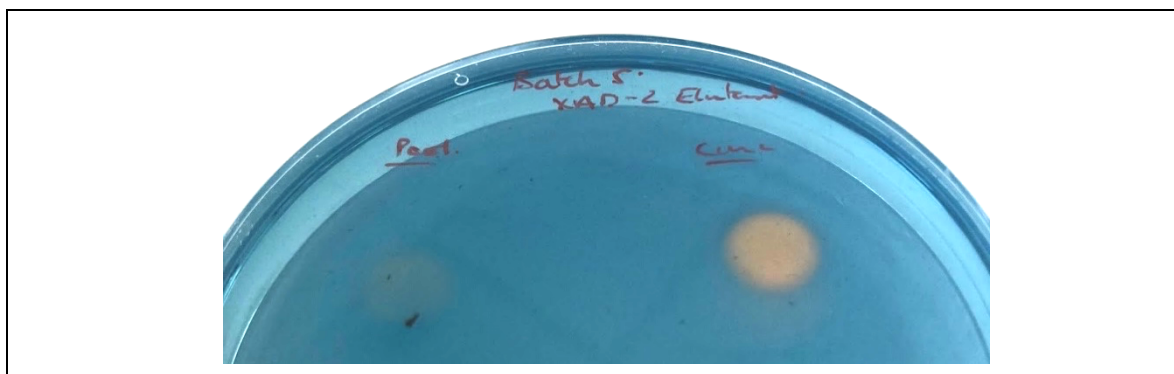


Figure 6-29: CAS agar test for siderophore detection

Left: Pooled microtitre CAS positive fractions, Right: Concentrated pooled microtitre CAS positive fractions

The concentrate was chromatographically separated further using P2 BioGel Gel Filtration column equilibrated with water; a flow rate of 0.3 ml/min was applied, collecting eluates in 4.5 ml fractions. Each of the fractions was then tested by microtitre CAS assay from which two fractions (Fr41/42) appeared to indicate siderophore activity.

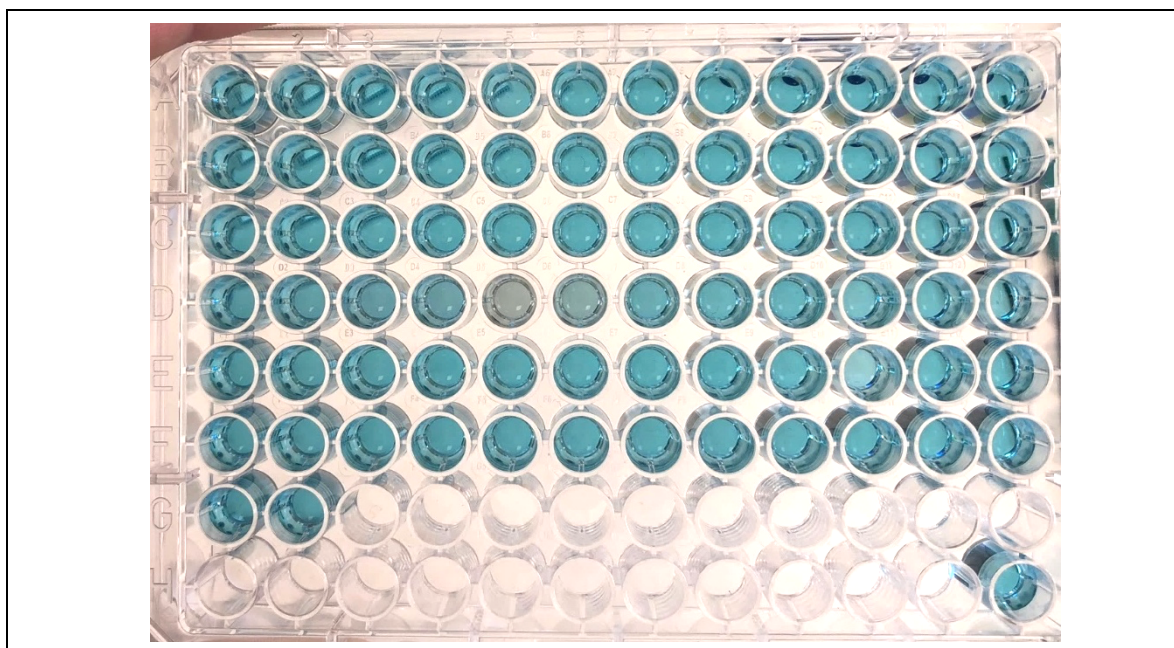


Figure 6-30: P2 Microtitre CAS assay for siderophore detection of Gel filtration fractions.

The microtitre plate of fractions was subject to a UV-Vis absorbance spectrum analysis between 200-800 nm before and after addition of CAS for confirmation (Figure 6-31 and Figure 6-32 respectively).

Both graphs clearly demonstrate an increase in absorbance at 395nm for fractions 40-42, with 41 generating the greatest change, with the CAS assay further confirming iron extraction from the blue dye indicated by a marked reduction in absorbance between 650 and 700 nm.

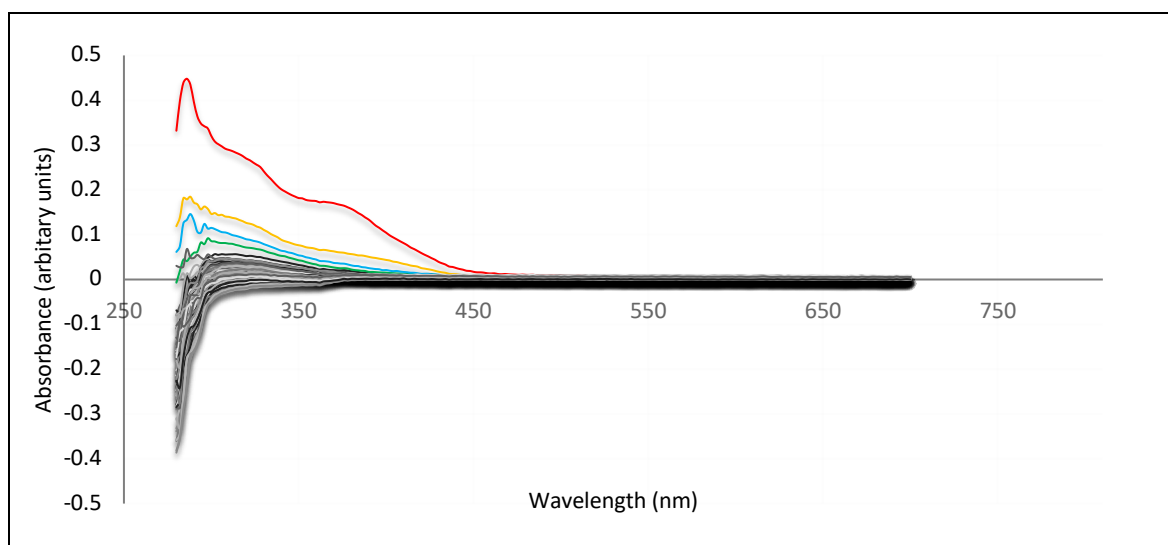


Figure 6-31: UV-visible spectral analysis of Gel filtration fractions. Spectral range 280 nm – 700 nm. Blank reduced. Fraction 41 is shown in red.

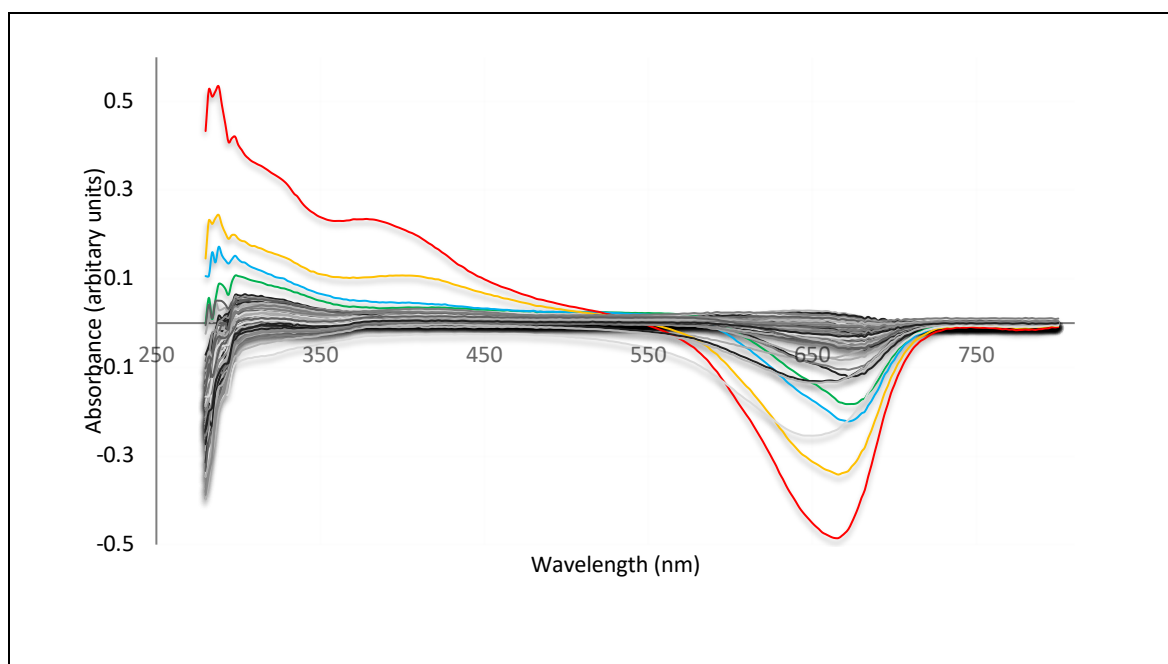


Figure 6-32: UV-visible spectral analysis of Gel filtration fractions using CAS assay solution. Spectral range 280 nm – 700 nm. Blank reduced. Fraction 41 is shown in red.

The supernatant and fractions of interest from all stages of purification were tested *via* thin layer chromatography using an n-butanol: acetic acid: water (12:3:5) solvent mix. Interestingly, between the XAD-2 concentrated pool sample and Fraction 41/42, there appears to be a conservation of bands detected using UV light highlighted in red (Figure 6-33). For siderophore detection, the TLC silica

plate was dried and then sprayed using a 0.1N Ferric chloride solution, that upon complexation with siderophore typically develops a wine colour. While development of a greyish pigment can be observed for the XAD-2 pool concentrate, was not replicated in the fractions of interest. This may be as a result of the sample dilution caused by the gel filtration chromatography.

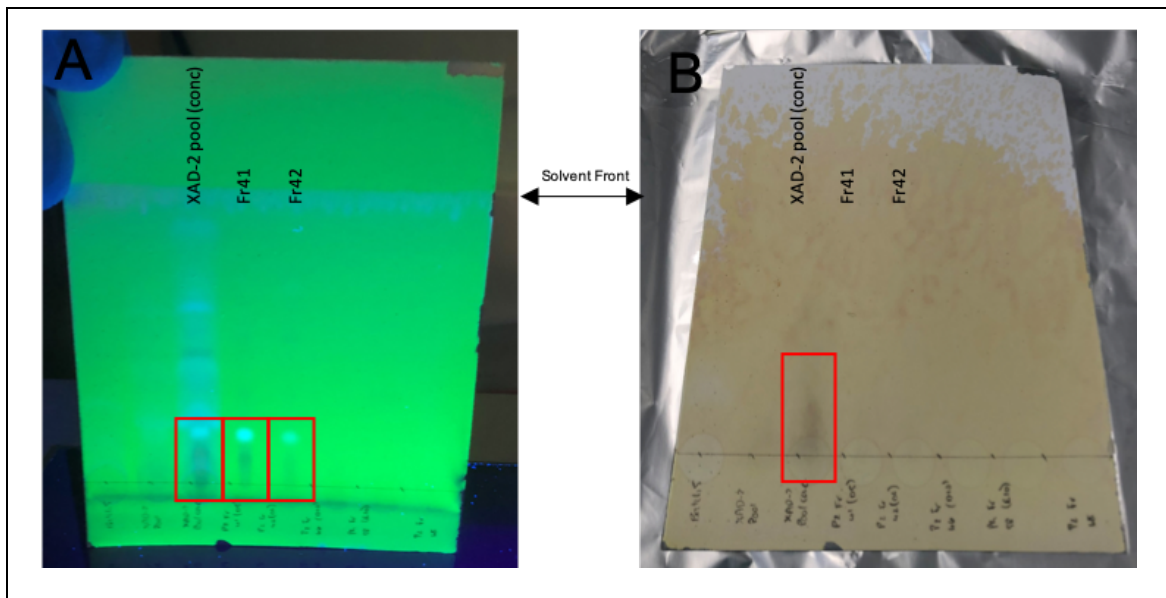


Figure 6-33: TLC analysis of XAD-2 CAS positive pool and gel filtration fractions 41 and 42. Ran using n-butanol: acetic acid: water (12:3:5) solvent mix and visualised under UV light (left) and developed with 0.1N Ferric chloride solution.

Consequently, a 1 ml of Fraction 41 was lyophilised, and resuspended in 50 μ l H₂O checked for siderophore activity on the less sensitive CAS agar, with a significant positive reaction (Figure 6-34), implying that the metabolites were successfully followed through the chromatography to reduce sample complexity, and identify other siderophore candidates.

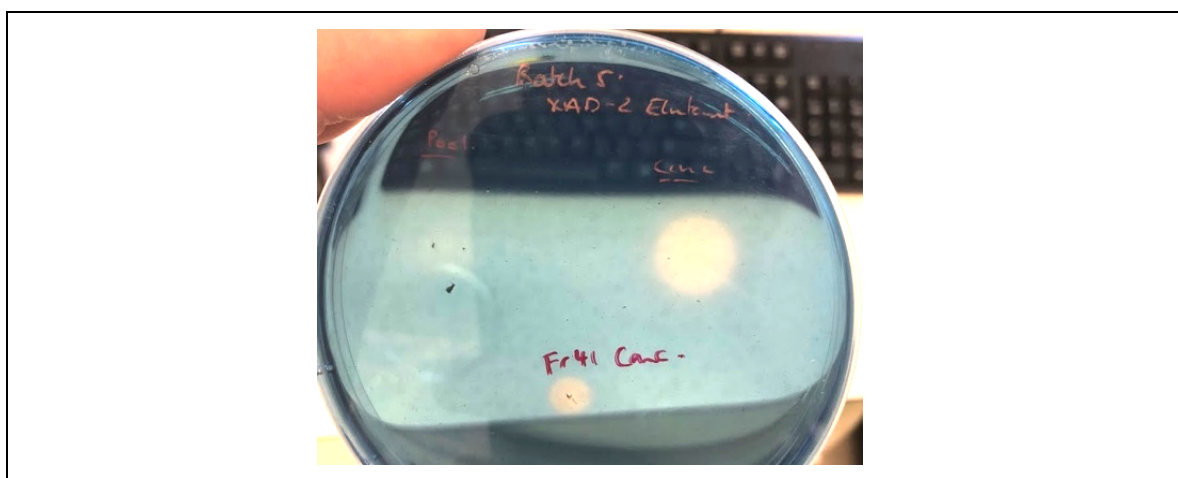


Figure 6-34: Analysis of a gel filtration Fraction 41 concentrate on CAS agar for siderophore detection.

Each of the samples of interest were subsequently analysed using LC-MS as previously described

using RP (C18) with both positive and negative ionisation modes. Candidates from the CAS positive concentrated Fraction 41 sample were selected based on approximate mass pairing and expressed under iron-limited conditions when cross correlated with the metabolite profile of *R. equi* 103S using RP LC-MS result table.

Table 6-3: Siderophore candidate list from Fraction 41.

Corresponding Metabolite from initial Experiment	Molecular Weight	RT [min]
M0872	434.12434	3.154
M1067	462.13078	3.498
M0537	562.21478	2.985
M0537	562.21483	2.191
M0780	658.23714	2.997
M0646	660.21339	3.082
M01039	682.19653	3.364
M01039	682.19771	3.116
M0852	698.24006	3.053
M0968	699.22319	3.279
M0667	714.22359	2.804
M0667	714.22359	3.129

The candidate list for alternative siderophores produced a collection of compounds that were previously identified as over-produced under iron-limited conditions. While the attempts to remove the candidate M0872, that previously dominated the mass spectra, by precipitation were largely successful, it is apparent that not all of the compound was removed, therefore it would be impossible at this stage to differentiate assume that the other candidates are siderophoric compounds without further chromatographic separation.

Interestingly however, other candidates namely M1067 and M0537 are both present in this list as well as selected as siderophore candidates in the initial candidate screening experiment, further refining the evidence that these compounds may function as siderophores.

Other newly identified candidates include M0646, M0667, M0780, M0852, M0968 and M1039, all of which follow the general trend of significantly greater metabolite abundance in iron-limited culture supernatants (Figure 6-35), while also appearing roughly within the expected molecular compound mass range of the siderophores predicted for the biosynthetic gene clusters identified in Chapter 3.

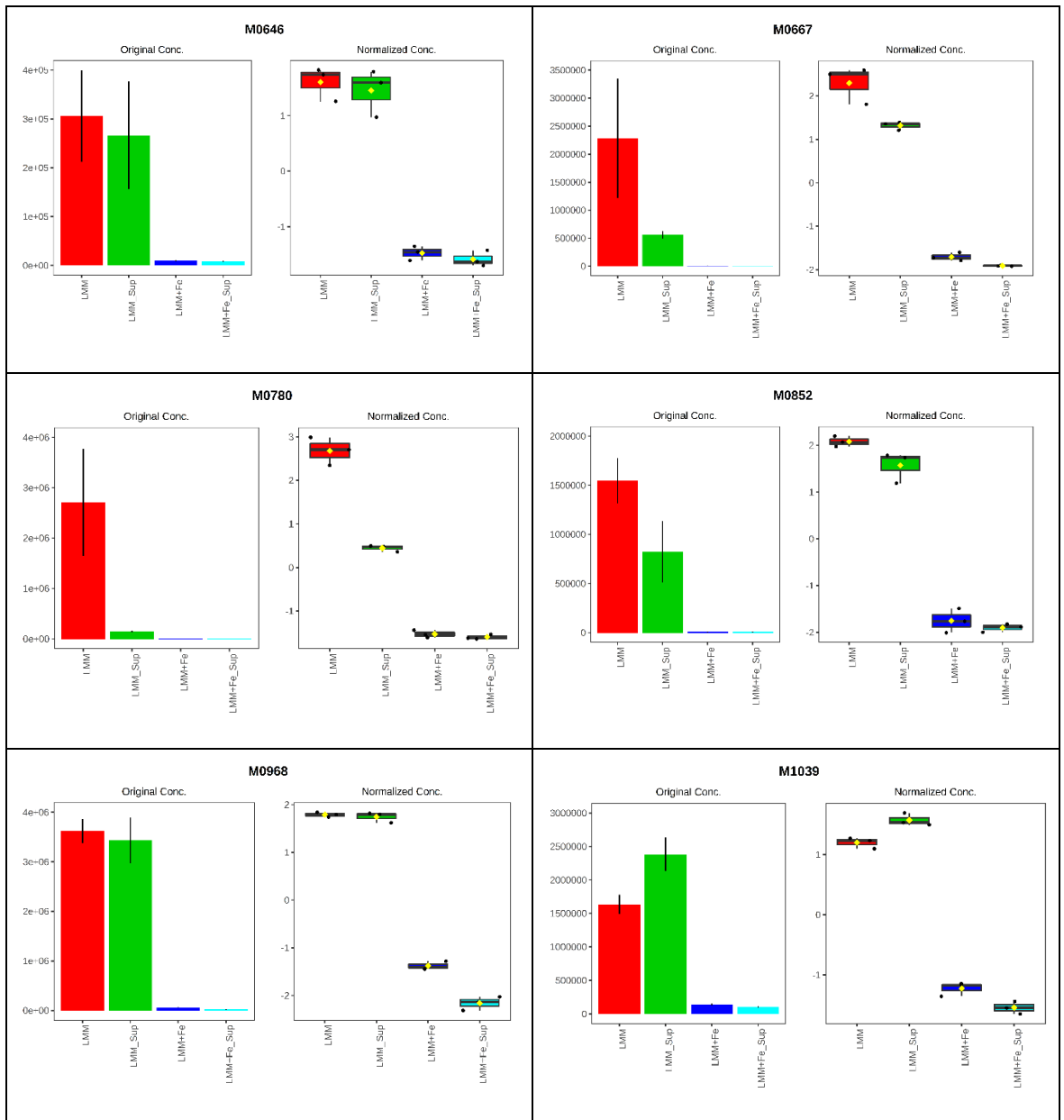


Figure 6-35: Loading values of Alternative candidates - mass spectral feature M0646, M0667, M0780, M0852, M0968 and M1039.

The bar plots on the left show the original values (mean \pm SD). The box and whisker plots on the right summarise the normalised values. These feature was selected based upon their upregulation under iron limited conditions.

6.6 Concluding remarks

The primary goal was to establish a reliable method of imposing iron limitation on *R. equi*. Ultimately, modified LMM was chosen and the visual cue of the developing red-chromophore adopted as a biomarker for iron-limitation. An untargeted metabolomics study clearly demonstrated an adaptation in the metabolome in response to iron-starvation, suggested two strong siderophore candidates (M0872 and M1067) and expedited a targeted strategy to characterise the *R. equi* siderophores that have been suggested in previous literature (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2008), yet remain uncharacterised. The behaviour of M0872 and its apparent iron complex is reminiscent of the heterobactins of *Rhodococcus erythropolis* (Bosello et al., 2013).

Moreover, these same culture parameters were used for studies into changes that occur in the proteome and transcriptome profiles planned to further define the iron-limitation response of *R. equi*.

The elaboration of the red chromophore is puzzling as it develops without introduction of additional iron, which is not compatible with the findings of Miranda-CasoLuengo et al. (2008) and the expectations of most siderophores. The continued overproduction of a siderophore when significant quantities of a ferri-siderophore are available in the culture supernatant would be counterintuitive. Moreover, the chromatography employed here has not been able to separate the chromophore and the siderophore-like activity exhibited in the CAS-based assay. Coloured desferri-siderophores have been observed previously, notably the yellow-green pyoverdine of *Pseudomonas fluorescens* which adopts a red-brown colouration on iron complexation (Meyer and Abdallah, 1978).

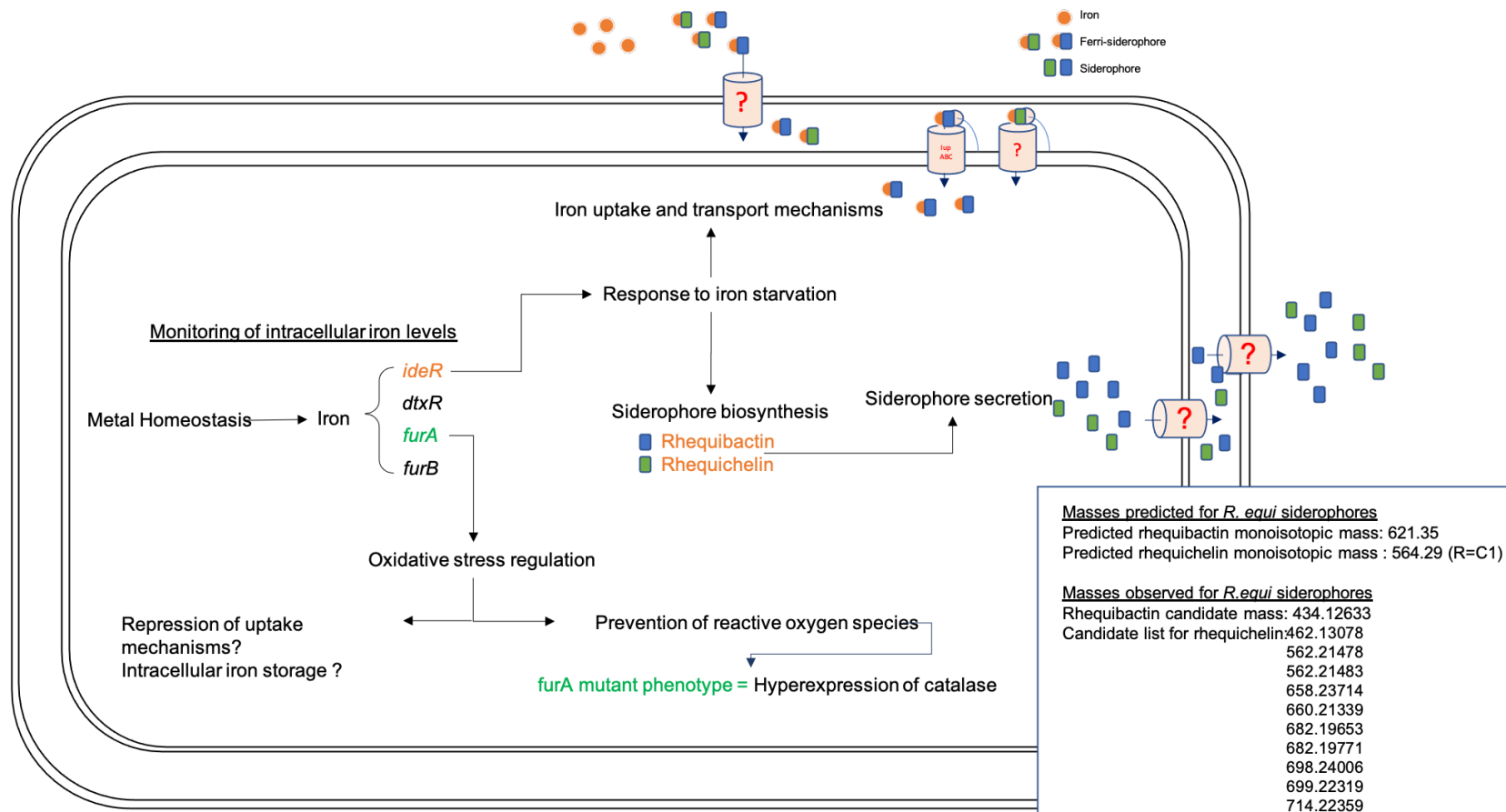


Figure 6-36: An updated *R. equi* iron homeostasis network map (after Chapter 6)

Data generated from the preliminary genomic survey, developed using *in silico* metal binding capacities, and production of metal regulator mutants. Siderophore candidates were identified *via* untargeted and subsequent targeted metabolomic analyses. The FurA candidate is displayed in green to indicate successful prediction with the mutant phenotyping identifying an oxidative stress response. The IdeR candidate is displayed in orange, to indicate incomplete results regarding iron regulatory mechanisms that are to be further explored in future work. Siderophore candidates at this stage are given in orange as at this predictions have been made, but further characterisation is needed.

7 A comparative proteomic analysis of the
composition of the rhodococcal cell wall under iron
limitation

7.1 Introduction

Proteomics, was first defined in 1995 as the large-scale characterisation of the entire protein complement of a given cell-line or organism, termed the *proteome* (Wilkins et al., 1996). The field has since developed to accommodate a variety of proteomics: Firstly, most project objectively look to characterise the proteome of a given organism for which will reflect the immediate environment in which it is studied, with the aim of not only to identify the complement of proteins, but also to perform quantitative differences between two or more physiological states of a biological system (Graves and Haystead, 2002).

Alternatively, a branch of Structural proteomics and Functional proteomics focus on analyses of the associated biological protein function, regulation and expression, protein-protein interaction and protein modifications (Figure 7-1).

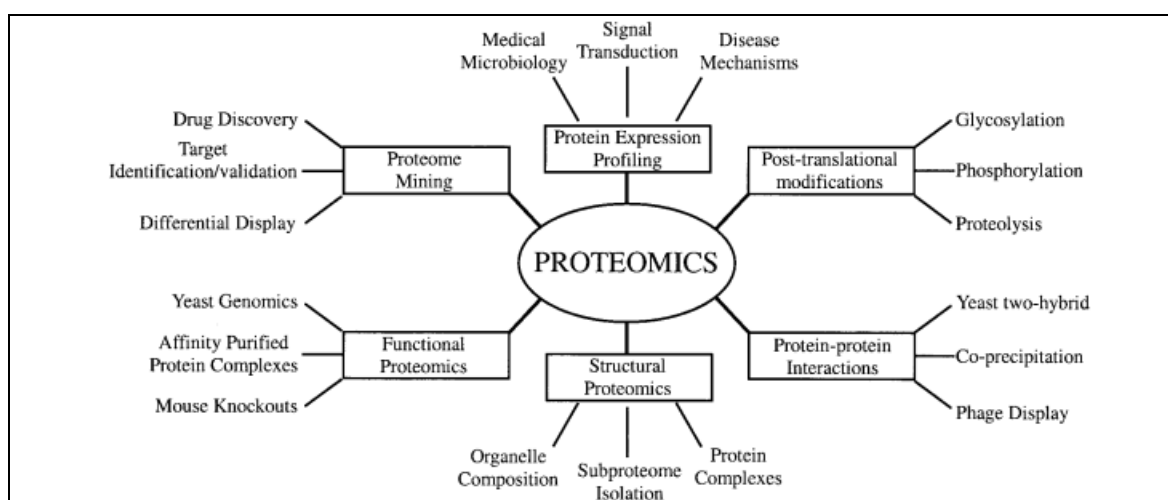


Figure 7-1: A spider diagram exploring the field of proteomics (Graves and Haystead, 2002).

7.1.1 Sample preparation for LC-MS analysis

Due to the inherent complexity of a proteome, there are multiple sample preparation methods that can be utilised for mass spectrometry analysis. These can vary depending on experimental aims, sample types and the analytical equipment available. Multiple factors should be considered when designing proteomic studies including sample complexity, protein abundance, physical properties and subcellular localisation. Preparation of biological samples prior to LC-MS/MS analysis can often include enhanced cell lysis, subcellular fractionation techniques for isolation of proteins in specific cellular locations, as well as depletion and or protein enrichment steps.

7.1.2 Protein separation

Some of the main challenges associated with proteomic analysis include ensuring sample complexity is within the detection rate of analyser, and the subsequent ability to assign protein identity once adequately separated. Typically, these techniques are performed using polyacrylamide gel electrophoresis and liquid chromatography, the result, being an analyte separated into manageable fractions that each can be investigated, thereby reducing sample complexity and increasing the chance of detection by mass spectrometry.

The strategy employed is often termed as a Bottom-up Proteomics approach, referring to the characterisation of proteins by analysis of peptides digested by proteolysis and subject to LC-MS/MS analysis. The protein samples prior to analysis are reduced, alkylated and digested *in situ*, often using the endoproteinase trypsin, which specifically cleaves at arginine and lysine amino acid residues (Paulo, 2016). The peptides are then extracted from the gel and prepared for LC-MS/MS analysis (Zhang et al., 2013) (Figure 7-2).

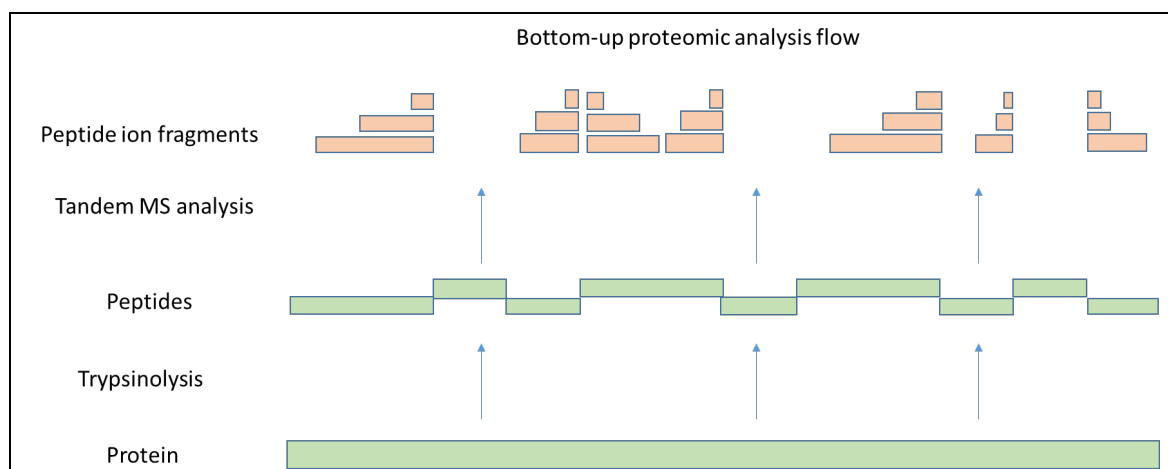


Figure 7-2: An overview of a Bottom-up proteomic analysis.

7.1.3 Protein identification

For decades mass spectrometry has been the gold-standard for small molecule analysis. Its accurate mass measurements for charged ions, differentiating closely related species to generate a peptide mass fingerprint. Further isolation and fragmentation of the molecular ions can be achieved using tandem mass spectrometry to reveal potential structural information (Patterson and Aebersold, 2003).

To accurately measure mass / charge ratio using a mass spectrometer, the analyte undergoes

ionisation before entering the vacuum system of the mass analyser. The ionisation process utilised within this project was an electrospray type. Given the usage as an analytical method, major advantages of using this soft electrospray ionisation include investigation of biological analytes that may be defined by non-covalent interactions, and without major fragmentation of molecular ions. The use of an electrospray ionisation (ESI) quadrupole mass analyser further facilitates structure deduction with reasonable accuracy, and given the excellent sensitivity of the apparatus, is a well-established method in both qualitative and quantitative measurements.

Peptide identification is attained by comparing the tandem mass spectra from the extracted peptide with theoretical tandem mass spectra generated using *in silico* digestion located in an online protein database.

7.2 Project workflow

To understand the effect of iron limitation on the proteome, the project workflow involves harvesting cells and utilising a GeLC-MS/MS approach, where the sample is manual fractionated after SDS-PAGE analysis, with digested peptides processed by reversed phased UPLC chromatography and identification performed using Tandem-Mass Spectrometry (). Peptide detection is performed using label-free quantification based upon relative protein quantities observed in multiple conditions. This can be attained by combining quadruple precursor ion selection with high-resolution, accurate-mass (HRAM) Orbitrap detection. This arrangement can then facilitate traced tracing peptides by retention time and allows reconstruction of a chromatographic elution profile of the monoisotopic peptide mass. The extracted peptide signals can be mapped across multiple LC-MS runs using mass-to-charge and retention-time coordinates to facilitate label-free quantification (Bridges et al., 2007).

Data analysis was performed using bottom-up analysis using MASCOT to generate un-interpreted MS/MS data and was used Progenesis LC-MS software to identify statistically significantly differences in protein expression between iron-deplete and iron-replete cultures of *Rhodococcus equi* 103S.

7.3 Rhodococcal cell wall

Bacterial cells have variable cell wall architectures, often relating to the depth of peptidoglycan layer,

which facilitates a method of differential variation based upon the cell penetrating Gram-staining technique. The rhodococcal cell wall, due to the mycolate-containing cell wall resembles aspects of both Gram positive and Gram-negative cell envelopes, due to the inherent thickness of the rhodococcal peptidoglycan layer, in conjunction with the hydrophobic mycolate layers, and must be further distinguished using an acid-fast staining technique (Wanger et al., 2017). The complex rhodococcal cell envelope is important for survival within challenging environments, providing some protection from noxious chemicals such as antibiotics and the killing mechanisms of macrophages (Riess et al., 2003). Primarily comprised of lipids, glycolipids and secreted proteins, the rhodococcal cell wall is defined by the presence of mycolic acids covalently attached to the underlying arabinogalactan component of the arabinogalactan–peptidoglycan layer (Figure 7-3) (Schaechter, 2009) that form an outer wall permeability barrier often described as a mycomembrane (Bayan et al., 2003).

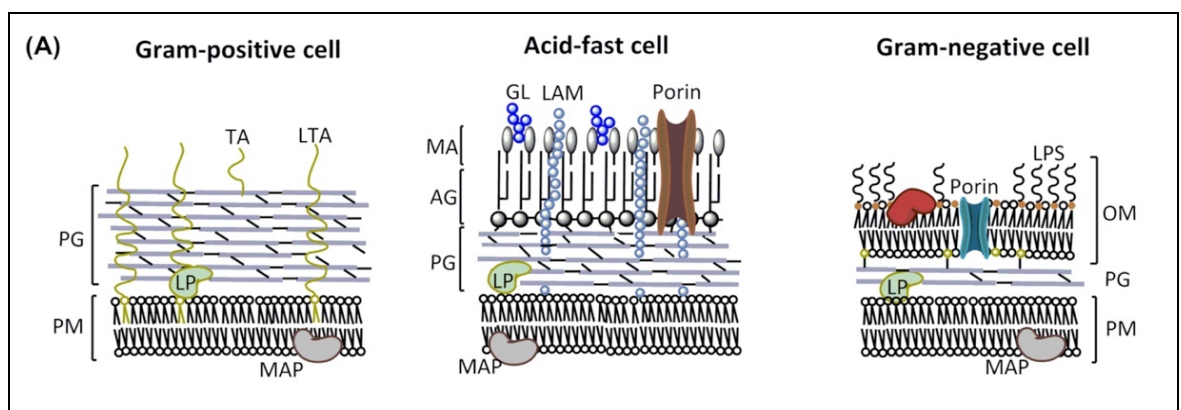


Figure 7-3: Cell envelope architecture of Gram-positive, acid-fast and Gram-negative bacteria (Maitra et al., 2019).

AG, arabinogalactan; GL, glycolipid; LAM, lipoarabinomannan; LP, lipoprotein; LPS, lipopolysaccharide; LTA, lipoteichoic acid; MA, mycolic acid; MAP, membrane-associated protein; OM, outer membrane; PG, peptidoglycan; PM, plasma membrane; TA, teichoic acid.

While ABC transporter systems have been considered to facilitate cytoplasmic translocation, mechanisms associated with translocation of the hydrophobic mycolate layer is greatly underexplored. A small number of proteins are hypothesised to interact with the mycolic acid layer of the rhodococcal cell wall, the first of which is a porin, which would provide pathways for hydrophilic solutes to traverse the mycolic acid layer (Maitra et al., 2019; Singh et al., 2015). Additionally, lipoproteins may be anchored to the outer mycolic acid layer (Takai et al., 2000), producing a surface-exposed protein perhaps in a complex with other bacterial factors or surface glycolipids contained within the cell wall.

7.4 Hypothesis

While the extant literature considers how *R. equi* uses a range of strategies to overcome iron limitation through proposed mechanisms associated with siderophore production and ferri-siderophore import across the cytoplasmic membrane *via* ABC transport systems (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2005; Miranda-CasoLuengo et al., 2008), no consideration has been made with regards to the mycomembrane.

The proposed mechanism of iron transport in *M. tuberculosis*, is that the iron from ferri-carboxymycobactin is transferred to mycobactin in the cell envelope region (Ratledge, 2004), with the protein HupB functioning as an iron transporter (based on its surface localisation; the ability to bind Fe^{3+} , and the potential to interact with both ferricarboxymycobactin and ferrimycobactin). At this stage, the iron is transferred from the ferrimycobactin to a ABC transporter to facilitate transfer across the cytoplasmic membrane (Sritharan, 2016).

Unlike *M. tuberculosis*, the *R. equi* genome does not appear to have the required biosynthetic machinery to produce siderophores analogous to the mycobactins, however it does contain a gene predicted to encode for a histone-like bacterial DNA-binding protein (HupB) identified using a BlastP search for the mycobacterial HupB protein against the *R. equi* 103S genome (Figure 7-4).

Sequences producing significant alignments:					
Description	Max Score	Total Query Score	cover	E Value	Per. Ident Accession
histone-like bacterial DNA-binding protein HupB [Rhodococcus...	174	174	49%	2e-55	85.71 CBH49137.1
Alignments:					
>histone-like bacterial DNA-binding protein HupB [Rhodococcus hoagii 103S]					
Sequence ID: CBH49137.1 Length: 242					
Range 1: 1 to 105					
Score:174 bits(440), Expect:2e-55, Method:Compositional matrix adjust., Identities:90/105(86%), Positives:100/105(95%), Gaps:0/105(0%)					
Query	1	MNKAELIDVLTQKLGSDRRQATAAVENVVDTIVRAVHKGDSVTITGFGVFEQRRRAARVA	60		
		MNKAELIDVLT+KLG+DRR A+ AVE+VVDITIVRAVH G+SVTITGFGVFEQRRRAARVA			
Sbjct	1	MNKAELIDVLTTEKLGDRRTASEAVEHVVDITIVRAVHAGESVTITGFGVFEQRRRAARVA	60		
Query	61	RNPRTGETVKVKPTSVPAPFRPGAQFKAVVSGAQRRLPAEGPAVKRG	105		
		RNPRTGETVKVKPTSVPAPFRPGAQFKA+++G Q+LP+ GPAVKRG			
Sbjct	61	RNPRTGETVKVKPTSVPAPFRPGAQFKALIAGGQKLPSSGPAVKRG	105		

Figure 7-4: BlastP search for the mycobacterial HupB protein against the *R. equi* 103S chromosome

However, HupB also has a cytoplasmic role as a nucleoid-associated protein with an ability to modulate chromatin structure and is proposed to work in a growth phase-dependent manner (Gupta et al., 2014), interestingly it has also been shown to influence (carboxy)mycobactin biogenesis positively (Pandey et al., 2014). The multiple roles and locations of HupB define it as one of the emerging groups of ‘moonlighting’ proteins (Huberts and van der Klei, 2010). To date, 399 moonlighting proteins from both prokaryotes and eukaryotes have been described in the dedicated database Moon Prot (Chen et al., 2018) (<http://www.moonlightingproteins.org/proteins/>). The process for extracellular localisation of such proteins is poorly understood and cannot be predicted. As the external function its homologue plays in *M. tuberculosis* may not be relevant in *R. equi*, then the possession of a HupB homologue should not be over interpreted.

To further explore the mechanisms utilised by *R. equi* to facilitate iron transport across the cell envelope, the experiment should focus on two key questions: 1) “Is it possible to identify proteins associated with isolated cell wall material from *R. equi*?” and 2) “from this dataset can we identify any potential outer membrane receptors that may facilitate transport of iron across the mycomembrane”. Production of a candidate list will allow for further characterisation studies to be performed in the future.

7.5 Experimental design

To perform a comparative analysis into iron regulated protein profile changes associated with the rhodococcal cell wall, first the methodology for isolation of cell wall material was to be optimised. The experimental methodology required a design that would best isolate components associated with the cellular envelope as well as allow for further characterisation of predicted mechanisms such as putative siderophore transport mechanisms.

Traditionally, isolation of the mycomembrane envelope region was performed using density gradient centrifugation, however development of bioinformatics prediction tools over the last decade has facilitated a process of performing a consensus prediction of membrane protein topology signal peptides.

Three methods were used and compared here (Figure 7-5). In each case, cells were lysed by sonication and particulate material was recovered by centrifugation, thus the cytoplasmic fraction and a proportion of the cytoplasmic membrane fraction would be removed in the supernatant and mycomembrane would be enriched in the deposited particulate material. Proteins were extracted from this material either directly using SDS or after density gradient centrifugation in Percoll at 60 % (v/v) to differentiate between intact cells and lysed cellular wall material after centrifugation at 27,000 x g. Proteins associated with the buoyant phase of cell wall-containing material were then solubilised by SDS or the non-ionic detergent Triton X-114

To determine the most optimal methodology, a comparison these techniques was performed for identification of upregulated non-cytoplasmic proteins ; with the methodology with the greatest impact being experimentally upscaled for the main analysis (Figure 7-6).

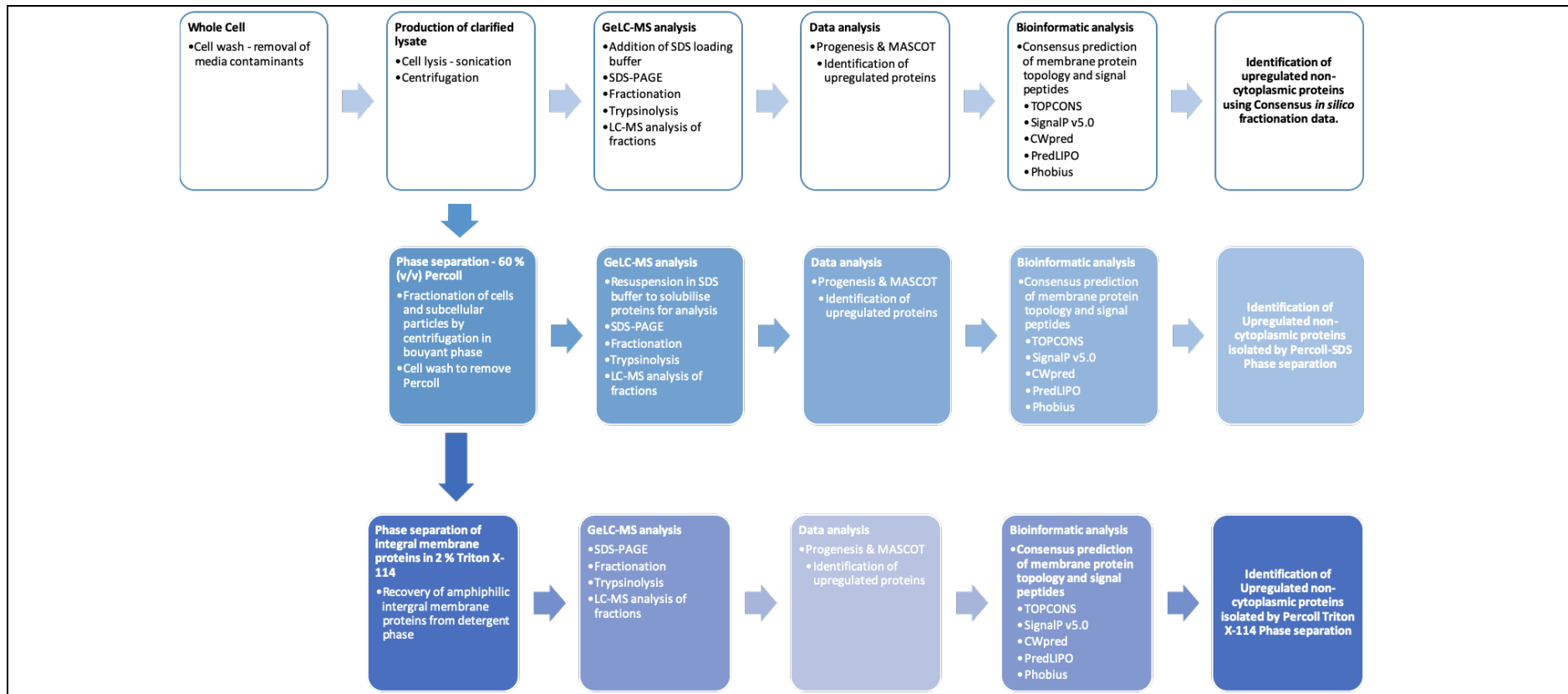


Figure 7-5: Flowchart detailing comparison of planned fractionation methods.



Figure 7-6 Methodology comparison analysis needed prior to final study.

7.6 Results

7.6.1 Initial unfractionated proteomics analysis

Initially, a small-scale comparison was performed to identify any differences at the proteome level between the previously optimised lactate minimal media growth conditions for iron limitation, identified in Chapter 6. The proteomic analysis was performed on *Rhodococcus equi* 103S cellular material harvested from the centrifugation step in Section 6.3.3, facilitating a multi-omics analysis on the effect of iron limitation. Cellular material isolated from 10 ml culture aliquots were re-suspended in 30 μ l 6x SDS-loading buffer, of which 10 μ l was analysed via SDS-PAGE (Figure 7-7), fractionated and treated with trypsin prior to LC-MS/MS analysis.

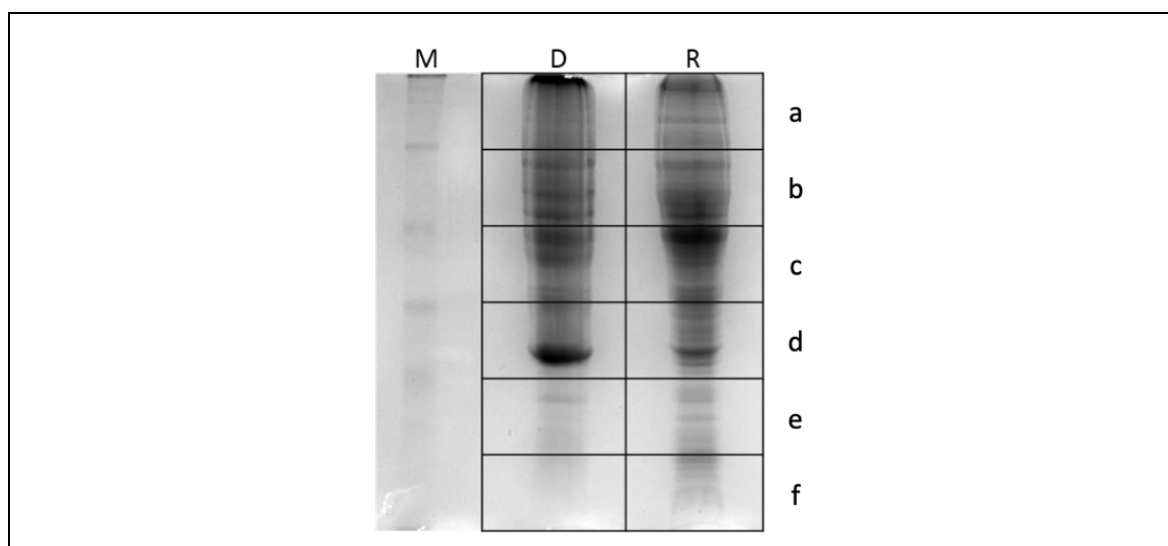


Figure 7-7: 12 % SDS-PAGE gel for initial proteomic analysis of *R. equi* 103S cellular material under iron limitation.

M = NEB broad range protein ladder, D = *R. equi* 103S grown using modified lactate minimal media without iron supplementation. R = *R. equi* 103S grown using modified lactate minimal media with iron supplementation. Gel fractionation performed according the grid pattern provided.

The proteomic workflow for identification of biologically relevant changes in protein abundance was performed using the software Progenesis LCMS (Nonlinear Dynamics) and MASCOT (Matrix Science) as described in Sections 2.13.3 and 2.13.4.

Given the crude sample complexity, the GeLC-MS approach incorporated gel fractionation and independent horizontal comparisons. The six partial fraction analyses were recombined *in-silico* using Progenesis LCMS (Nonlinear Dynamics), with the output taking into account the experimental design of ‘iron deplete vs iron replete’ conditions; for the initial analyses the ANOVA (p) values

were excluded due to the sample numbers for each condition ($n = 1$).

The protein statistics for initial unfractionated analysis identified a total of 1693 proteins, which represents 36 % of the entire genome with *R. equi* 103S having 4598 coding sequence (CDS) locations manually annotated. Importantly mid-analysis using Progenesis LC-MS, peptides are screened based upon their reliability score with < 40 being excluded, this value is generally considered exceptionally reliable for reducing incorrect identification, therefore the observed protein count is likely resulting from the high sensitivity of the mass spectrometer. Further to this, refinement of the protein report indicated that differential expression was observed for 1021 proteins having met the threshold of >2 peptide matches and ≥ 2 max fold change.

With regards to the aim of this study, 523 proteins were identified as meeting the desired threshold of having > 2 peptide matches; ≥ 2 max fold change and 'iron-deplete' highest mean value to be considered upregulated in iron-deplete conditions. The report produced (Chapter 8 Appendices) identifying differentially expressed proteins under iron starvation was further examined, exploring the putative cellular localisation of each candidate.

For each detected protein, its expected cellular location was considered using a consensus strategy involving prediction of transmembrane topology and signal peptides using the online bioinformatic tools TOPCONS (Tsirigos et al., 2015), PredLIPO (Bagos et al., 2008), SignalP (Almagro Armenteros et al., 2019) and Phobius (Kall et al., 2007). The outputs of these analyses were assembled into a single spreadsheet and a consensus was established based upon majority. An abridged 'non-cytoplasmic protein' subset is shown in Table 7-1.

Table 7-1: Abridged in silico protein localisation results for the initial unfractionated proteomic analysis

Accession	Peptides	Fold	Description	Consensus
CBH46933.1	2	148.94	putative MFS transporter	Membrane
CBH46930.1	7	119.11	putative substrate binding lipoprotein	Lipoprotein
CBH46486.1	16	114.99	putative substrate binding lipoprotein	Lipoprotein
CBH50056.1	2	99.25	putative ABC transporter integral membrane subunit	Membrane
CBH48875.1	10	73.31	putative ABC transporter substrate binding lipoprotein	Secreted
CBH50642.1	13	64.93	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH47947.1	13	62.84	putative integral membrane protein	membrane
CBH48109.1	14	53.84	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH47948.1	7	50.76	putative membrane protein	membrane
CBH49190.1	8	38.56	putative substrate binding lipoprotein	Lipoprotein
CBH48366.1	2	31.42	putative lipoprotein	Lipoprotein
CBH46613.1	6	15.72	putative substrate binding lipoprotein	Lipoprotein
CBH46973.1	2	13.42	putative secreted protein	Lipoprotein
CBH49471.1	6	12.58	putative cation transporter ATPase P-type	membrane
CBH47928.1	5	12.42	putative beta-lactamase	membrane
CBH49489.1	2	10.67	putative fasciclin-like lipoprotein	Lipoprotein
CBH48506.1	10	9.04	putative secreted protein	Membrane
CBH47544.1	2	8.97	putative ABC-type glycine betaine transporter substrate binding lipoprotein	Lipoprotein
CBH50002.1	7	8.57	putative branched-chain amino acid ABC transporter substrate binding lipoprotein	Lipoprotein
CBH49009.1	8	8.52	putative substrate binding lipoprotein	Secreted
CBH46872.1	8	8.43	putative ABC transporter integral membrane subunit	Membrane
CBH46625.1	15	8.34	putative secreted short chain dehydrogenase	Secreted
CBH47529.1	4	8.06	putative twin-arginine translocation protein TataA	membrane
CBH47276.1	3	7.79	putative integral membrane protein	Lipoprotein
CBH48545.1	2	7.07	putative lipoprotein	Lipoprotein
WP_041674135.1	10	6.59	signal peptidase I	membrane
CBH48968.1	15 (14)	5.67	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1	Lipoprotein
CBH46351.1	7	5.63	putative secreted cutinase	Secreted
CBH47587.1	2	5.54	putative lipoprotein	Lipoprotein
CBH46743.1	2	5.49	putative secreted protein	Membrane
CBH48307.1	14	5.45	putative secreted protein	Secreted
CBH48336.1	6	5.33	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH49708.1	4	4.72	putative secreted protein	Lipoprotein
CBH46276.1	19	4.43	secreted short chain dehydrogenase	Secreted
CBH48258.1	5	4.28	putative secreted tyrosine phosphatase	Secreted
CBH48798.1	10	4.21	UDP-muramoylpentapeptide beta-N-acetylglucosaminyltransferase MurG	membrane
CBH46208.1	2	4.16	putative secreted cutinase	Secreted
WP_041673883.1	2	4	membrane protein [Rhodococcus equi]	membrane
CBH46459.1	7	3.99	putative secreted protein	Membrane
CBH49231.1	5	3.98	putative secreted protein	Membrane
CBH46518.1	2	3.97	putative secreted protein	Secreted
CBH48555.1	2	3.92	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH48883.1	3	3.91	secreted alpha/beta hydrolase	Secreted
CBH50085.1	16	3.78	putative membrane protein	Membrane
CBH46341.1	14	3.74	putative mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase LytB	Secreted
CBH46210.1	3	3.67	putative secreted protein	Secreted
CBH47452.1	2	3.64	putative secreted phosphoglycerate mutase family protein	Secreted
CBH49330.1	8	3.62	putative lipoprotein LpqB	Lipoprotein
CBH46357.1	2	3.56	putative indolylacetyltransferase	Membrane
CBH49202.1	4	3.5	putative lipoprotein	Lipoprotein
CBH48658.1	2	3.44	putative integral membrane protein	Membrane
CBH48154.1	4	3.4	putative secreted peptidyl prolyl isomerase	Membrane
CBH46464.1	6	3.39	putative integral membrane protein MmpL1	membrane
CBH48786.1	2	3.16	putative lipoprotein	Lipoprotein
CBH47932.1	8	3.15	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH48148.1	5	3.11	putative preprotein translocase subunit SecD	Lipoprotein
CBH48926.1	3	3.09	putative membrane protein	Membrane
CBH47412.1	4	3.07	putative lipoprotein	Secreted
CBH46270.1	4	3.06	putative copper transporter	membrane
CBH48796.1	4	2.96	cell division protein FtsQ	Lipoprotein
CBH46555.1	3	2.86	putative integral membrane protein	membrane
CBH48374.1	4	2.85	putative 20S proteasome beta subunit PreB	Secreted
CBH46350.1	3	2.83	putative lipoprotein	Membrane
CBH47545.1	6	2.76	putative ABC-type glycine betaine transporter substrate binding lipoprotein	Lipoprotein
CBH49026.1	25	2.68	putative aminopeptidase	Lipoprotein
CBH50448.1	3	2.65	putative substrate binding protein	Secreted
CBH47957.1	5	2.63	putative secreted protein	Secreted
CBH48153.1	4	2.61	putative secreted peptidyl prolyl isomerase	Lipoprotein
CBH48491.1	2	2.61	putative membrane protein	Membrane
CBH49638.1	3	2.61	putative secreted protein	Secreted
CBH48182.1	2	2.58	putative secreted metalloproteinase	Secreted
CBH48489.1	4	2.56	putative membrane protein	Lipoprotein
CBH46207.1	5	2.53	putative secreted protein	Secreted
CBH47492.1	2	2.5	putative high affinity substrate binding lipoprotein	Lipoprotein

Accession	Peptides	Fold	Description	Consensus
CBH48150.1	6	2.48	putative high affinity substrate binding lipoprotein	Lipoprotein
CBH48965.1	2	2.47	sulfate/thiosulfate ABC transporter integral membrane subunit CysT	membrane
CBH46206.1	2	2.44	putative secreted protein	Secreted
CBH49639.1	3	2.42	putative lipoprotein	Lipoprotein
CBH50256.1	2	2.3	putative high affinity substrate binding lipoprotein	Lipoprotein
CBH48899.1	5	2.28	putative integral membrane protein	membrane
CBH47306.1	7	2.27	putative high affinity substrate binding lipoprotein	Membrane
CBH49243.1	6	2.26	cell division integral membrane protein FtsX	membrane
CBH46868.1	2	2.25	putative Mce associated protein Mas2B	Membrane
CBH46920.1	5	2.22	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH49418.1	2	2.16	putative secreted metalloproteinase	Lipoprotein
CBH50664.1	4	2.08	putative molybdate ABC transporter substrate binding lipoprotein	Lipoprotein
CBH46976.1	7	2.07	putative phosphate transporter substrate binding lipoprotein	Lipoprotein
CBH48149.1	4	2.04	putative preprotein translocase subunit SecF	Membrane
CBH48813.1	2	2.03	putative membrane protein	Membrane

7.6.1.1 Analysis of initial unfractionated results

The initial results identified a diverse range of proteins that were upregulated under iron limitation which provided a sound foundation for analysis under these conditions. Preliminary identification of 1693 proteins (36 % of *R. equi* 103S CDS regions) suggest that the gel fractionation and subsequent peptide extraction, fragmentation and identification were both successful and sufficient. However, this number of proteins represents a very large proportion of potential genome products to be associated with the cell envelope fraction. It likely reflects both the sensitivity of the instrumentation and the obvious scope for cross-contamination of the mycolyl-arabinogalactan-peptidoglycan matrix with materials from other cellular compartments on lysis that can persist through subsequent manipulations. Further refinement using statistical analysis identified that identities ascribed to 523 proteins were both statistically sound and these were upregulated under iron limitation. Analysis of these upregulated proteins indicates that the iron-limiting conditions selected for analysis are suitable given the prevalence of predicted iron-regulated proteins (identified earlier in Chapter 6) associated with siderophore synthesis and siderophore-related functions. These include non-ribosomal peptide synthetases, isochorismatase, 2,3-dihydroxybenzoate-AMP-ligase, numerous putative siderophore binding proteins and HupB an iron-regulated protein in *M. tuberculosis* that arbitrates transfer from the iron-bound carboxymycobactin to the mycobactin (Yeruva et al., 2006).

Additional data refinement was performed to exclude likely cytoplasmic contamination, as the project aims relate to identification of wall located iron-regulated proteins that may participate in trans-mycolate siderophore transport mechanisms. The bioinformatic protein localisation pipeline applied indicated that 89 proteins with statistically significant IDs were non-cytoplasmic and upregulated in iron-limited conditions.

As expected, the most highly upregulated proteins identified appear to be transport components that would typically be associated with siderophore import/export mechanisms. The presence of multiple ABC transporter components including substrate-binding lipoproteins and integral membrane subunits also appears to suggest multiple complex transport mechanisms utilised simultaneously by this bacterial pathogen. The only research to date regarding siderophore transport in *R. equi* identified that the ABC transporter system *iupABC* was required for growth under iron limitation (Miranda-CasoLuengo et al., 2005). The *iupABC* system closely resembles the hydroxamate siderophore ABC transport system *cchCDE* utilised by *S. coelicolor* (discussed earlier in Section 3.6) therefore is likely to be associated with internalisation of the hydroxamate siderophore ferri-rhequichelin. Given that only one siderophore transport system has previously been discussed for *R. equi*, it is logical to assume that additional transport systems may be employed to facilitate multiple ferri-siderophore cytoplasmic translocation. The presence of multiple gene clusters encoding for putative ABC transporter substrate binding lipoproteins and putative ABC transporter integral membrane subunits is encouraging and may be associated with internalisation of ferri-rhequichelin and ferri-rhequibactin relating to xenosiderophore piracy, a common feature of bacterial iron-stress responses that reflects their natural existence in competitive and complex communities (Arias et al., 2015; Challis and Hopwood, 2003).

In vitro fractionation is also explored to compare the best strategy for isolation and characterisation of cell-wall associated proteins and may be useful in identification of novel protein architecture for understanding structure-function relationships in the gram-positive cell envelope region, a typically understudied area of microbial physiology.

7.6.2 Proteomic analysis of cell wall material fractionated by density gradient centrifugation

In order to explore whether data quality could be improved by attempting to remove potentially intact cells from the particulate fraction analysed, density gradient centrifugation in 60% Percoll was introduced. Here cell wall fragments would be buoyant but intact cells (containing cytoplasmic and cytoplasmic membrane contaminants) would be deposited as a pellet. In practice, the vast majority

of the cell wall material was buoyant indicating a very effective lysis.

Removal of contaminating cytoplasmic material was performed with multiple wash steps throughout the protocol before suspension in an SDS extraction buffer. After removing insoluble material by centrifugation, proteins extracted from the isolated cell-wall fraction were analysed by SDS-PAGE (Figure 7-8a), with the loading adjusted for comparison between conditions. The subsequent GeLC-MS approach incorporated a brief electrophoresis to ensure migration into the resolving gel, fractionation (Figure 7-8b) and independent comparisons as previously described. The two partial fraction analyses were recombined *in-silico* using Progenesis LCMS (Nonlinear Dynamics), with the output taking into account the experimental design of ‘iron-deplete vs iron-replete’ conditions; as before the ANOVA (p) values were excluded due to the sample numbers for each condition (n = 1).

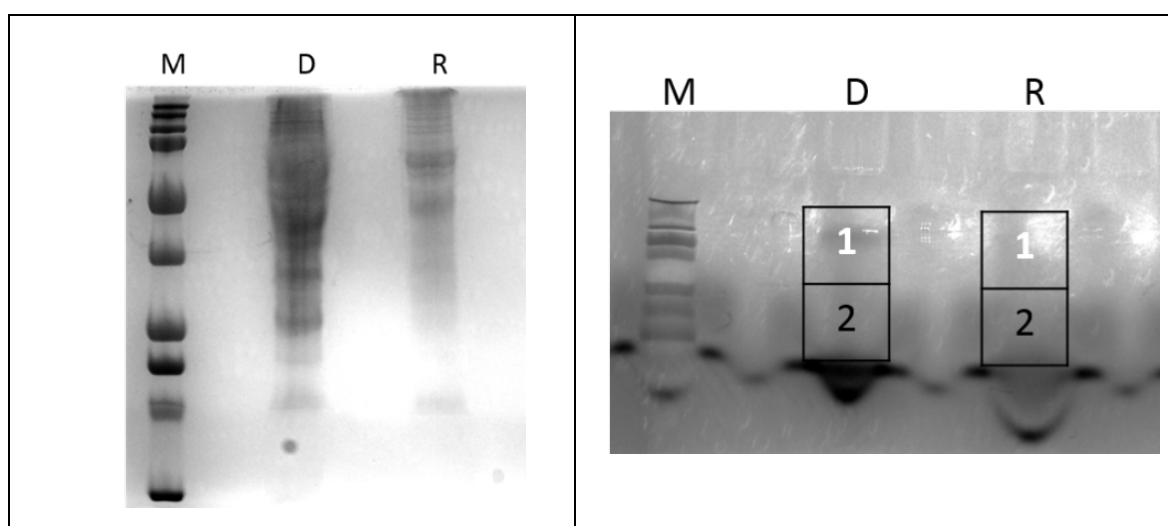


Figure 7-8: 12 % SDS-PAGE gel for cell wall material fractionated by density gradient centrifugation of *Rhodococcus equi* 103S.

Fractionation of the unstained gel (right) was performed according the grid pattern indicated. M = NEB broad range protein ladder, D = LMM-Fe. R = LMM+Fe.

This analysis identified 1104 proteins, of which 295 were identified as meeting the desired thresholds[±] to be considered upregulated in iron-deplete conditions. The report identifying differentially expressed proteins under iron starvation indicates a reduction of sample complexity by 43.8 %. As there was no evidence for the survival of intact cells then this decrease in complexity is likely related to the extra washing steps.

Further examination of this dataset by *in silico* exploration of the putative cellular location of each

[±] > 2 peptide matches; ≥ 2 max fold change and ‘iron deplete’ highest mean value.

candidate, using the localisation pipeline described earlier, revealed that the density gradient separation was not as effective as initially projected, as only 53 out of 294 proteins identified were predicted to be non-cytoplasmic proteins. These proteins of interest were compiled into an abridged ‘non-cytoplasmic protein’ index listing the accession number, peptide count, fold change, description and consensus is shown in Table 7-2.

Table 7-2: Abridged *in silico* protein localisation results for the proteomic analysis of the cell wall material fractionated by density gradient centrifugation

Accession	Peptides	Fold	Description	Consensus
CBH49161.1	3	85.95	putative integral membrane protein	Membrane
CBH47679.1	2	81.04	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH50218.1	6	31.78	putative secreted osmotically inducible protein	Secreted
CBH50021.1	4	9.8	putative integral membrane protein	Secreted
CBH49731.1	3	8.22	cytochrome c assembly protein	Lipoprotein
CBH49471.1	2	5.25	putative cation transporter ATPase P-type	Membrane
CBH47403.1	4	5.18	putative secreted protein	Lipoprotein
CBH48374.1	3	5.09	putative 20S proteasome beta subunit PrcB	Membrane
CBH49272.1	2	5.09	putative secreted protein	secreted
CBH46625.1	12	4.9	putative secreted short chain dehydrogenase	Secreted
CBH48477.1	2	4.63	putative secreted protein	Membrane
CBH49661.1	2	4.63	putative Mce family lipoprotein Mce4E	Lipoprotein
CBH49848.1	2	4.58	putative secreted protein	Secreted
CBH46208.1	10	4.38	putative secreted cutinase	Secreted
CBH50604.1	2	4.37	PadR family transcriptional regulator	Secreted
CBH48968.1	16	3.88	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1	Lipoprotein
CBH48875.1	10	3.78	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH46176.1	4	3.65	putative secreted lipase	Secreted
CBH46703.1	2	3.65	cell division related ATP-dependent protease FtsH	Membrane
CBH49663.1	4	3.5	putative Mce family protein Mce4C	Membrane
CBH48366.1	2	3.47	putative lipoprotein	Lipoprotein
CBH48350.1	6	3.41	putative secreted chitinase	Secreted
CBH48109.1	3	3.3	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH48891.1	3	3.22	sodium/dicarboxylate symporter	Membrane
CBH47054.1	3	3.12	putative integral membrane protein	Membrane
CBH48926.1	2	3.12	putative membrane protein	Membrane
CBH50055.1	4	3.06	putative ABC transporter substrate binding subunit	Membrane
CBH47021.1	2	3.05	putative high affinity substrate binding lipoprotein	Lipoprotein
CBH47957.1	4	3.01	putative secreted protein	Secreted
CBH48906.1	4	2.93	putative secreted protein	Membrane
WP_005514445.1	2	2.93	membrane protein [Rhodococcus equi]	Membrane
CBH47057.1	4	2.82	putative resuscitation-promoting factor rpfA	Secreted
CBH50015.1	3	2.8	putative membrane protein	Membrane
CBH49211.1	2	2.61	putative purine catabolism regulatory protein	Secreted
CBH48885.1	2	2.58	putative lipoprotein	Lipoprotein
CBH47276.1	2	2.54	putative integral membrane protein	Membrane
CBH47586.1	2	2.48	putative ABC transporter integral membrane subunit	Membrane
CBH50017.1	4	2.47	conserved hypothetical protein	Secreted
CBH48090.1	4	2.43	putative integral membrane protein	Membrane
CBH46868.1	2	2.42	putative Mce associated protein Mas2B	Membrane
CBH46459.1	5	2.33	putative secreted protein	secreted
CBH47948.1	15	2.33	putative membrane protein	Secreted
CBH48096.1	3	2.3	putative potassium binding protein	secreted
CBH47529.1	2	2.27	putative twin-arginine translocation protein TatA	Membrane
CBH50002.1	2	2.27	putative branched-chain amino acid ABC transporter substrate binding lipoprotein	Lipoprotein
CBH49938.1	2	2.25	putative substrate binding lipoprotein	Lipoprotein
CBH47928.1	3	2.17	putative beta-lactamase	Lipoprotein
CBH48418.1	3	2.03	putative secreted protein	Secreted

7.6.2.1 Discussion of cell wall material fractionated proteomic analysis

Although the introduction of density gradient fractionation using 60 % (v/v) Percoll reduced the complexity of the sample, the number of cytoplasmic proteins identified still exceeds the number of non-cytoplasmic proteins identified. The post-analysis *in silico* filtering of results to remove

cytoplasmic proteins significantly reduces the index from 295 to 53 proteins.

As hypothesised and similar to the previous experiment, the most highly upregulated proteins identified proteins observed are predicted to be transporter components. The significant upregulation of a putative integral membrane protein CBH49161.1 (encoded by REQ_31600) and putative ABC transporter substrate-binding lipoprotein CBH47679.1 (encoded by REQ_16020) at 85.95- and 81.04-fold respectively was observed. Both are representatives of protein families expected to be significantly upregulated under the iron-deplete conditions as they represent routes for iron importation (Henderson and Payne, 1994).

Compared to the previous experiment, the fold changes of upregulated proteins observed here is substantially lower with exception of the examples above, this is likely due to the additional experimental procedures for isolation of the cell-wall associated material and associated wash steps.

7.6.3 Triton X-114 extraction of fractionated cell wall

In this trial, 2 % Triton X-114 (v/v) was used in place of SDS for extraction and biphasic separation of proteins based upon hydrophobicity adapted from methods listed by Tan et al. (1995). A solution of this non-ionic detergent is homogeneous at 0°C but a bi-phase develops with distinct aqueous and detergent phases above 20°C. Proteins partition during this phase separation; hydrophilic proteins are found exclusively in the aqueous phase, and integral membrane proteins in the detergent phase (Bordier, 1981). This phase separation was adopted to determine if cytoplasmic containments could be more adequately excluded. The isolated cell wall proteins were subject to acetone precipitation and resuspended in SDS-loading buffer before electrophoretic analysis.

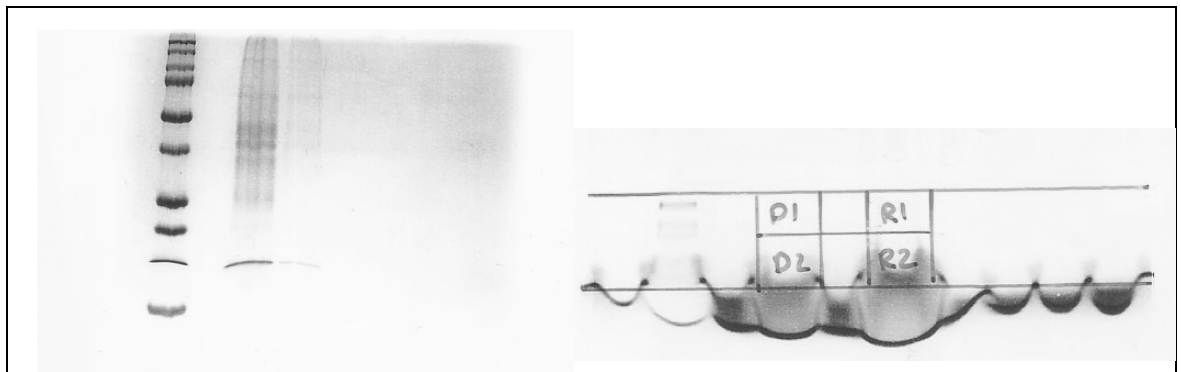


Figure 7-9: 12 % SDS-PAGE gel for proteomic analysis of Triton X-114 extracted *R. equi* 103S cellular material under iron limitation.

M = NEB broad range protein ladder, D = LMM-Fe. R = LMM+Fe. Clear differences in protein loading were seen for initial analysis (left), protein load of R was adjusted to compensate in the unstained peptide extraction gel (right). Fractionation of the unstained gel (right) was performed according to the grid pattern indicated.

The subsequent GeLC-MS approach incorporated partial electrophoresis to ensure migration into the resolving gel, fractionation and independent comparisons as described for the previous experiment. The two partial fraction analyses were recombined *in-silico* using Progenesis LCMS (Nonlinear Dynamics), with the output taking into account the experimental design of ‘iron deplete vs iron replete’ conditions; as before the ANOVA (p) values were excluded due to the sample numbers for each condition (n = 1).

This approach identified a total of 644 proteins, of which 145 were identified as meeting the desired thresholds of > 2 peptide matches; ≥ 2 max fold change and ‘iron deplete’ highest mean value to be considered upregulated in iron deplete conditions. The report predicting differentially expressed proteins under iron starvation indicated a reduction of sample complexity by 72.3 % and 50.7 % compared to the unfractionated analysis and the cell wall material fractionated by density gradient centrifugation respectively.

To further explore the efficacy of the isolation technique used, the putative cellular localisation of each candidate protein was predicted using the localisation pipeline as previously described. As for seen in the previous experiment, it was apparent that while sample complexity was significantly reduced, the percentage of non-cytoplasmic to cytoplasmic proteins comprised just 10 % of proteins isolated using Triton X-114 extracted cell wall methods. While the pipeline indicated substantial cytoplasmic contamination, 15 proteins were compiled into an abridged ‘non-cytoplasmic protein’ index (Table 7-3).

Table 7-3: Abridged *in silico* protein localisation results for the Triton X-114 extracted cell wall material

Accession	Peptides	Fold	Description	Consensus
CBH48968.1	14	5.96	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1	lipoprotein
CBH47948.1	12	5.22	putative membrane protein	membrane
CBH50021.1	4	4.59	putative integral membrane protein	secreted
CBH46276.1	6	4.29	secreted short chain dehydrogenase	Secreted
CBH47531.1	3	4.26	putative secreted protein	secreted
CBH48366.1	2	4.2	putative lipoprotein	lipoprotein
WP_041674135.1	5	3.97	signal peptidase I	membrane
CBH49471.1	2	3.89	putative cation transporter ATPase P-type	membrane
CBH48875.1	5	3.27	putative ABC transporter substrate binding lipoprotein	lipoprotein
CBH50557.1	10	3.04	putative secreted cutinase	membrane
CBH46625.1	10	2.73	putative secreted short chain dehydrogenase	Membrane
CBH47021.1	2	2.67	putative high affinity substrate binding lipoprotein	lipoprotein
CBH46208.1	5	2.62	putative secreted cutinase	Secreted
CBH49061.1	2	2.26	putative secreted protein	secreted
CBH47947.1	10	2.17	putative integral membrane protein	membrane

7.6.3.1 Discussion of Triton X-114 extracted cell wall proteomic results

The adaptation to the methodology described above appears to generate similar experimental shortcomings as the previous experiment regarding cytoplasmic contamination, with the quantity cytoplasmic proteins identified greatly outweighing the number of non-cytoplasmic proteins identified. The use of the post-processing *in silico* filtering stage to remove cytoplasmic proteins significantly reduces the index from 145 to 15 proteins.

Analysis of the 15 remaining non-cytoplasmic proteins, yielded a selection of putative membrane proteins, lipoproteins and secreted proteins that may be of interest. Interestingly, from this dataset the highest differential protein expression appeared to be a 5.96-fold increase of sulfate/thiosulfate ABC transporter substrate binding lipoprotein (CysP1), while this protein might not be directly related to iron-regulatory mechanisms, it does highlight a significant reduction in both protein appearance and abundance, likely as a result of the stringent experimental methodology associated with isolation of amphiphilic cell envelope proteins.

The outcomes of all three initial methodological trials should be considered to identify the optimal method for detection of cell wall-associated proteins, before initiating the large-scale analysis.

7.6.4 Consensus

A comparison of the three preliminary extraction methodologies described in this chapter was performed to identify the strategy that would be used for a more rigorous analysis of iron-regulated

envelope-associated proteins. Firstly, a Venn diagram was produced to summarise the candidate lists (Figure 7-13), of which 30 proteins were consistently detected (Table 7-4).

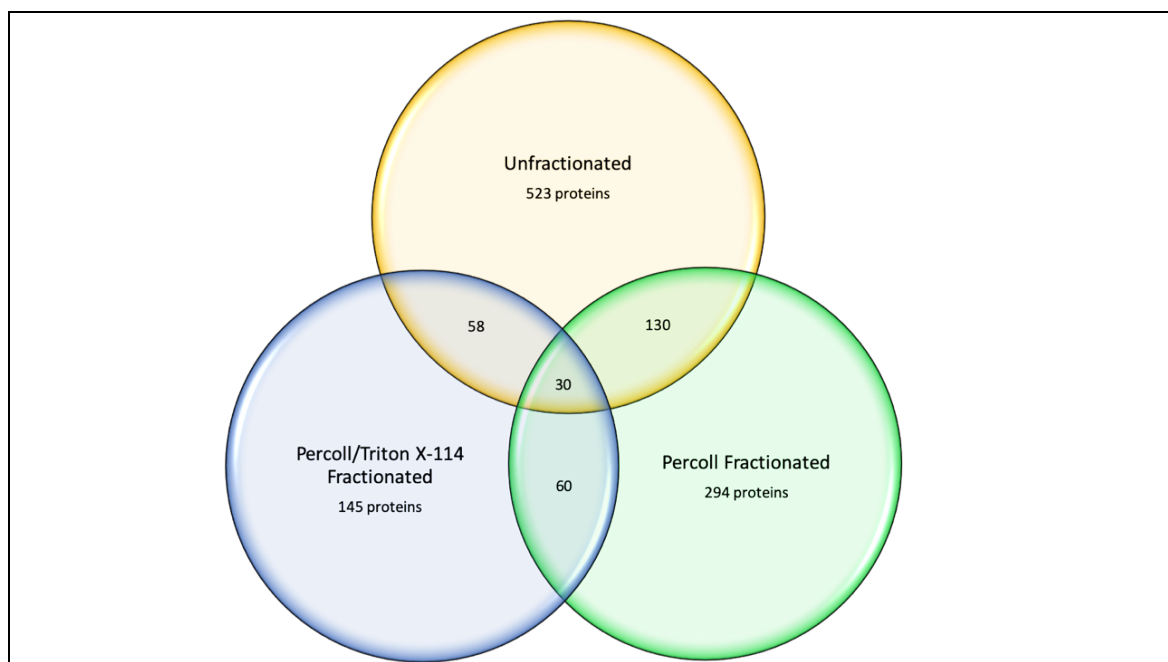


Figure 7-10: Venn diagram detailing the quantity of identified proteins for each given fractionation method.

Analysis of the isolated proteins common to all trials (Table 7-4) emphasised the presence of cell envelope-associated proteins that would be characteristically produced under iron-limited condition, including putative ABC transporter substrate-binding lipoproteins, putative membrane proteins, however there is also clear evidence for considerable cytoplasmic contamination by the presence of putative siderophore binding proteins, and siderophore production components such as putative 2,3-dihydroxybenzoate-AMP ligase and putative isochorismatase.

Table 7-4: Conserved proteins between all 3 methodologies.

Accession	Protein Description
CBH46934.1	putative siderophore binding protein
CBH48968.1	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1
CBH46970.1	thiosulfate sulfurtransferase
CBH47948.1	putative membrane protein
CBH46625.1	putative secreted short chain dehydrogenase
CBH50218.1	putative secreted osmotically inducible protein
CBH46927.1	putative short chain dehydrogenase
CBH46925.1	putative isochorismatase
CBH46208.1	putative secreted cutinase
CBH48875.1	putative ABC transporter substrate binding lipoprotein
CBH48005.1	secreted metallo-beta-lactamase superfamily protein
CBH48881.1	glutamate-ammonia-ligase adenylyltransferase GlnE
CBH46926.1	putative 2,3-dihydroxybenzoate-AMP ligase
WP_041674435.1	acyl-CoA synthetase
CBH48871.1	putative alkanesulfonate monooxygenase
CBH46285.1	putative PadR family transcriptional regulator
CBH49482.1	putative IMP dehydrogenase/GMP reductase
CBH50452.1	conserved hypothetical protein
CBH48986.1	gamma-glutamyl kinase-GP-reductase multienzyme complex ProA
CBH47350.1	ribose-phosphate diphosphokinase PrsA
CBH49471.1	putative cation transporter ATPase P-type
CBH49605.1	aldehyde dehydrogenase
CBH47837.1	conserved hypothetical protein
CBH49733.1	glutamate-1-semialdehyde 2,1-aminomutase HemL
CBH48366.1	putative lipoprotein
CBH48550.1	putative dephospho-CoA kinase
CBH50234.1	acyl-CoA dehydrogenase
CBH48119.1	putative chlorite dismutase
CBH49907.1	putative FMN-dependent monooxygenase
CBH48156.1	histidine-tRNA ligase HisS

Of greatest interest were proteins without detailed annotations, principally secreted hypothetical proteins, that may represent a novel class of mycolate-associated proteins. An interesting example from the set was the hypothetical protein CBH50452.1 encoded by the gene REQ_44910, however upon exploration of the protein-protein interaction network and functional analysis using STRING Version 11 (Szklarczyk et al., 2019), this appeared to be clustered with a range of ribosomal proteins and therefore is unlikely to contribute to any cell envelope-related mechanisms.

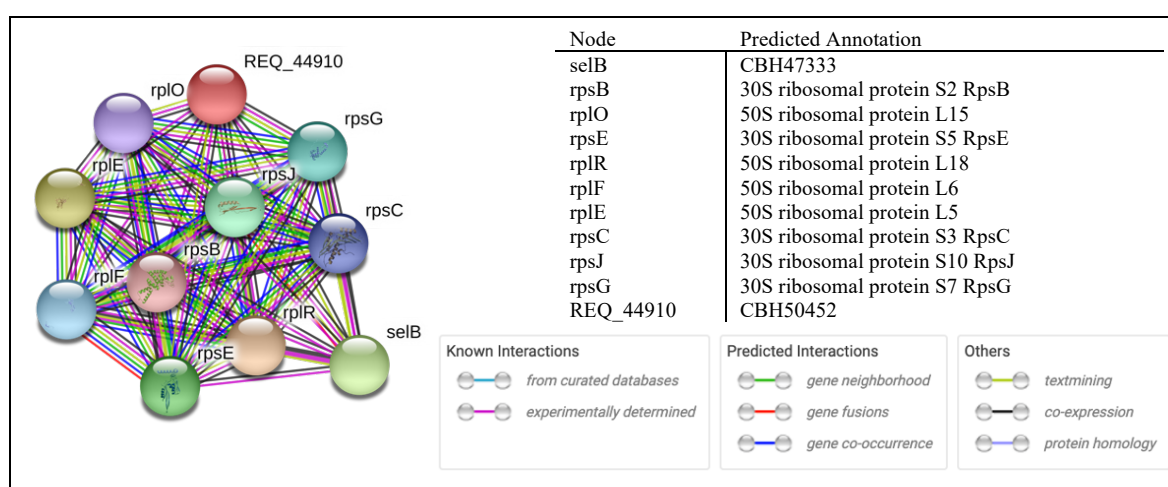


Figure 7-11: STRING analysis for the hypothetical protein CBH50452.1 encoded by the gene REQ_44910.

Annotations provided in the included table, with node interactions identified by colour.

Also, another protein of interest was the hypothetical protein CBH47837.1 encoded by the gene REQ_17680. Exploration of the protein-protein interaction network and functional analysis using STRING Version 11 (Figure 7-12) identified potential co-expression with a DNA gyrase subunit B protein and co-occurrence with a group of integral membrane proteins. Further investigation of this protein by considering conserved domain architecture suggested that it is likely a coiled-coil protein given the prediction of coiled-coil regions by the COILS program. While coiled-coil proteins can often have functions relating to membrane sensors, channels and regulation of influx/export as well as protein trafficking and quality control (Rose et al., 2005), in this instance it is likely that the hypothetical protein CBH47837.1 is involved in DNA repair given the likely co-expression with the DNA gyrase subunit B protein, comparative to the co-occurrence with the integral membrane proteins.

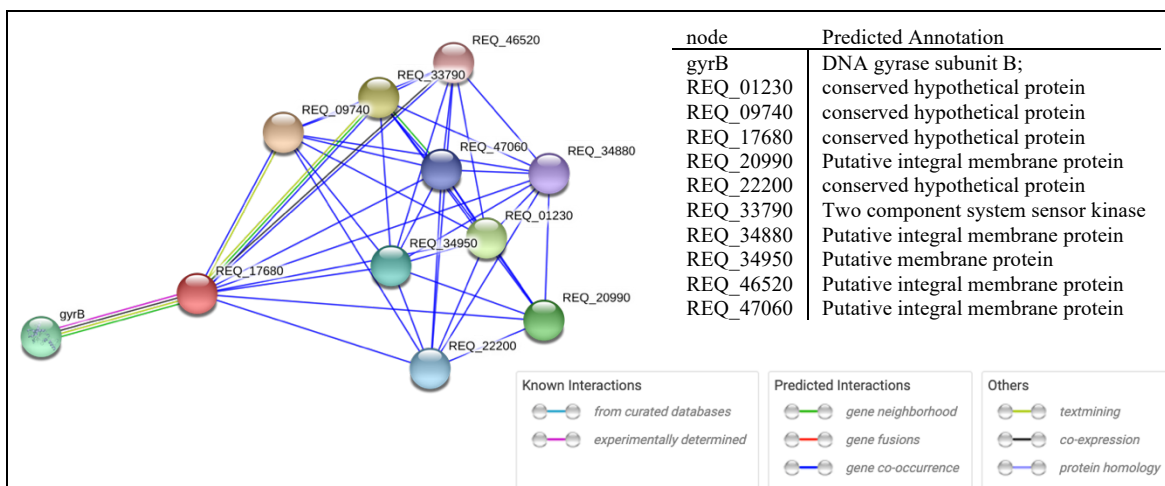


Figure 7-12: STRING analysis for the hypothetical protein CBH50452.1 encoded by the gene REQ_17680.

Annotations provided in the included table, with node interactions identified by colour.

While the complexity of the samples appears to reduce through the *in vitro* methodologies, implementation of the *in silico* subcellular localisation suggests that the ratio of cytoplasmic to non-cytoplasmic does not change as initially predicted, as 89 non-cytoplasmic (17% of 523 total) proteins were reduced to 53 (18% of 295 total) and 15 (10% of 145 total) for unfractionated compared to Percoll-SDS and Percoll-Triton X-114 methods respectively. To visually compare the combination of *in vitro* fractionation and the additional *in silico* protein localisation methods, an additional Venn diagram was produced to compare the candidate lists (Figure 7-13), that emphasises that only 7 non-cytoplasmic proteins appear to have been common between methods.

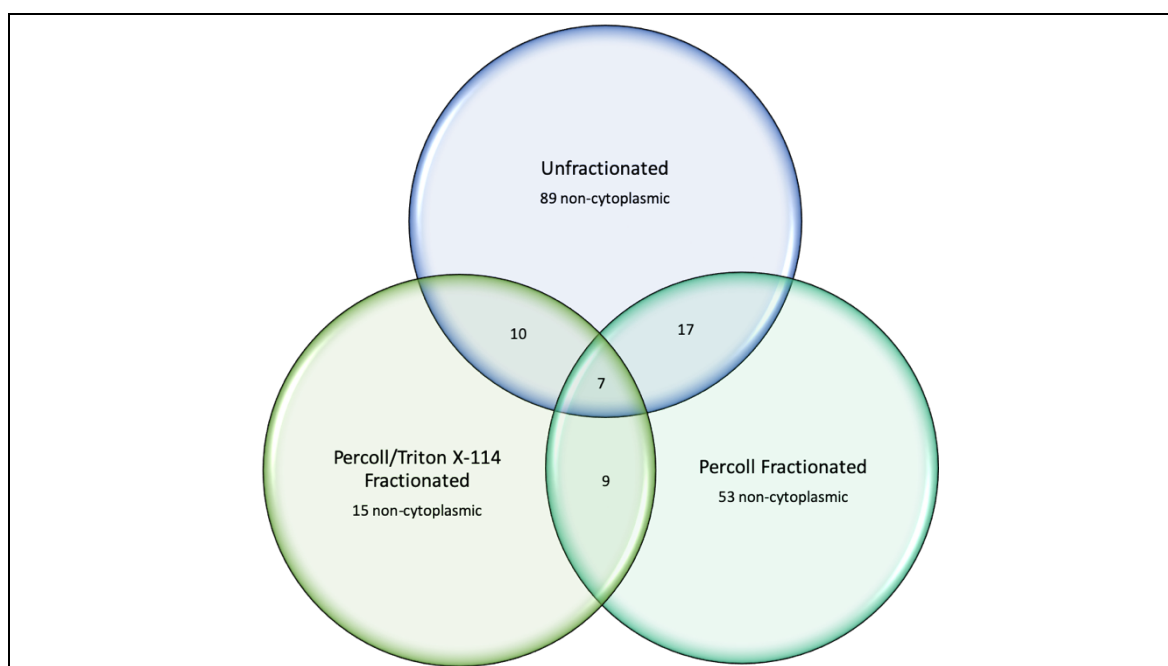


Figure 7-13: Venn diagram detailing the quantity of identified proteins for each given fractionation method, reinforced by the in-silico prediction methodology.

An in-depth analysis of the 7 non-cytoplasmic proteins (shown in Table 7-5), suggested conservation of: 1) a putative secreted short chain dehydrogenase (CBH46625.1) that may be functionally involved with proteins including KasA, a fatty acid synthase and other short chain dehydrogenases.

Table 7-5: Conserved non-cytoplasmic proteins between all 3 study methods, as predicted using a consensus approach to bioinformatic subcellular localisation

Accession	Unfractionated		Percoll		Triton X-114		Description	location
	peptide	fold	peptide	fold	peptide	fold		
CBH46625.1	15	8.34	12	4.9	10	2.73	putative secreted short chain dehydrogenase	Membrane
CBH47948.1	7	50.76	15	2.33	12	5.22	putative membrane protein	Membrane
CBH48968.1	15	5.67	16	3.88	14	5.96	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1	Lipoprotein
CBH48875.1	10	73.31	10	3.78	5	3.27	putative ABC transporter substrate binding lipoprotein	Lipoprotein
CBH46208.1	2	4.16	10	4.38	5	2.62	putative secreted cutinase	Secreted
CBH48366.1	2	31.42	2	3.47	2	4.2	putative lipoprotein	Lipoprotein
CBH49471.1	6	12.58	2	5.25	2	3.89	putative cation transporter ATPase P-type	Membrane

- 2) a putative membrane protein (CBH47948.1), that appears to be functionally uncharacterised.
- 3) a sulfate/thiosulfate ABC transporter substrate binding lipoprotein that is likely produced to regulate the rhodococcal sulphur metabolism pathway, including transport of extracellular sulphate across the cell envelope (KEGG pathway: req00920).
- 4) A putative ABC transporter substrate binding lipoprotein CBH48875.1 (encoded by REQ_28560) that is likely to be attached to the cytoplasmic membrane to facilitate transfer to a nearby transport system, or perhaps the outer leaflet of the rhodococcal cell wall, however the intricacies regarding interaction with the mycolate-layer have not yet been characterised. The STRING functional analysis

network (Figure 7-14) implies a deep-rooted functional relationship with predicted iron transport system IupABC as well as a host of other ABC transport systems.

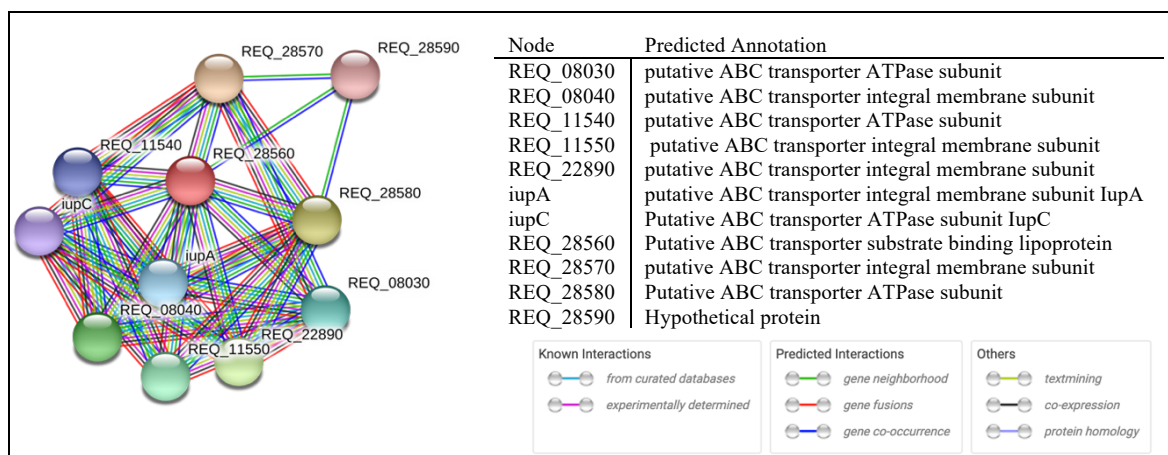


Figure 7-14: STRING analysis for the putative ABC transporter substrate binding lipoprotein (CBH48875.1) encoded by the gene REQ_28560.

Annotations provided in the included table, with node interactions identified by colour.

5) a putative secreted cutinase (CBH46208.1), although there is currently little information regarding bacterial cutinases (Chen et al., 2008; Nikolaivits et al., 2018), the enzyme family containing cutinases are well established and have been characterised within the fungal community for a number of years, being responsible for hydrolytic activity and degradation of plant cell walls. The detection of this cutinase may be of significance as recently Verma et al. (2015) identified a range of mycobacterial cutinases, of which Cut5b was found to be present in the cell wall, membrane and cytosol subcellular fractions, with on-going studies to evaluate its role in mycobacterial physiology and pathogenesis. Given the similarity between these related genera, there may also be of significance within *R. equi*, interestingly, the ubiquitous detection of the cutinase in mycobacterial cellular fractions emphasises the technical issues relating to the accurate fractionation of the components of the *Corynebacteriaceae*.

6) a putative lipoprotein (CBH48366.1) that appears to be functionally uncharacterised.

7) a putative cation transporter ATPase P-type (CBH49471.1) that may function as a mechanism for regulation of metal homeostasis and prevention of metabolic stress (Waldron et al., 2009) by facilitating transport of a range of transition or heavy metals (Smith et al., 2014). This example of a P_{1B}-ATPase may be functionally relevant for regulation of oxidative stress, given that the STRING functional analysis network predicted that the ATP-ase was associated with a secreted ferroxidase and putative heavy-metal detoxification protein, further exploration of these two linked proteins

identified conservation of a multicopper oxidase domain and a Copper chaperone domain respectively. Given the biological conditions of iron-limitation in these experiments it would be unusual to encounter an efflux pump mechanism for exportation of iron, therefore perhaps a biological stress response to metal nutrient limitation include efflux of surplus metals to prevent nonspecific binding intracellularly, promoting a more efficient metal utilisation system.

7.6.4.1 Method optimisation

Comparison of the three methodologies employed here highlighted a range of positive and negative attributes that were considered for the experimental design of the large-scale proteomic study.

It is apparent at this stage that significant cytoplasmic/cytoplasmic membrane contamination is unavoidable. It appeared that the most appropriate methodology to utilise for the large-scale study, was a simple recovery of buffer-washed particulate material after lysis as the density gradient centrifugation analyses had demonstrated there to be no significant recovery of unbroken cells. Application of inclusion criteria relating to at least 2-fold up-regulation in iron-deplete conditions and to have been identified by detection of multiple peptides, in combination with the *in-silico* cellular location consensus methodologies outlined earlier would be expected to provide an inclusive candidate list that avoids significant losses of proteins as false negatives during the more extensive processing.

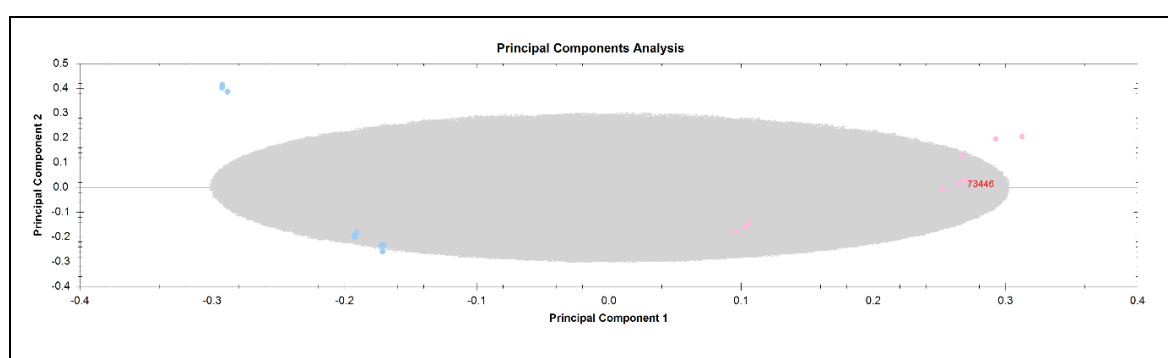
7.6.5 Large scale proteomic survey

The experiment was designed to incorporate experimental robustness by including technical and biological replication in triplicate. Using the consensus methodology for protein fractionation, the cellular material was harvested and processed according to the clarified lysate analysis described in Figure 7-5. Due to limitations regarding polyacrylamide gel peptide extraction (discussed later) and to take into account the crude sample complexity, the GeLC-MS approach was utilised by performing gel fractionation. Prior to LC-MS analysis the fractions were combined and resuspended in the loading buffer, this was advised due to the high performance and excellent resolving power of the LC-MS/MS system, this also permitted technical replicates of each sample to be performed.

The raw data was processed using Progenesis LC-MS (Nonlinear Dynamics), with the output taking into account the experimental design of ‘iron deplete vs iron replete’ conditions; with each condition represented by biological and technical replicates in triplicate. Data analysis was performed using bottom up proteomic analysis using MASCOT to generate un-interpreted MS/MS data in tandem with the Progenesis LC-MS software to identify statistically significant differences in protein abundance between iron-deplete and iron-replete cultures of *Rhodococcus equi* 103S.

7.6.5.1 Peptide analysis

Initially, features (peptides) identified were un-interpreted, with the total count calculated at 72,687 features, Figure 7-15 demonstrates the variability between sample classes using principal component analysis. Peptide refinement[‡] was performed to screen for statistically significant differences at the peptide level, resulting in a reduction of complexity by 61.8%. Incorporation of statistical refinement at this stage improved assurance that differential expression could be determined at the protein level.



[‡] Peptide statistical refinements: ANOVA p-value ≤ 0.05 and Max fold change ≥ 2

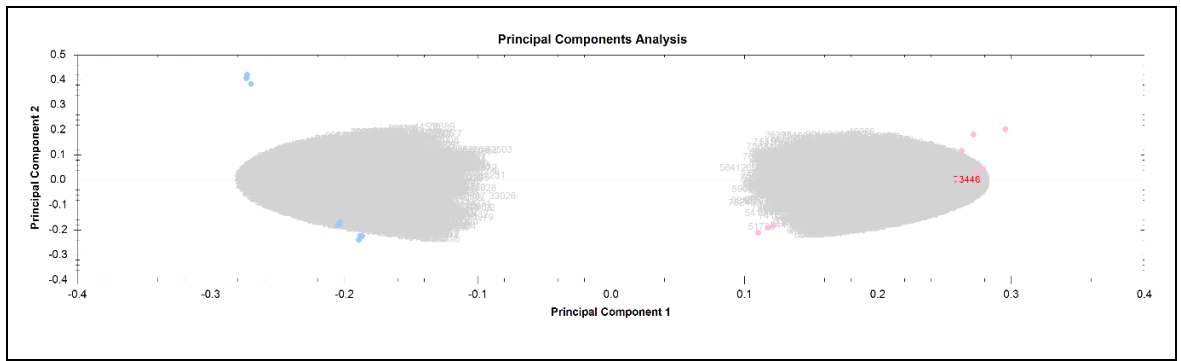


Figure 7-15: Principle Component Analysis for total features (peptides) identified in the large-scale proteomic survey. Total feature count: 72687.

Top panel: Unfiltered data. Lower panel: Features filtered by ANOVA p-value ≤ 0.05 and Max fold change ≥ 2 , Filtered feature count: 27797 Protein Accession numbers are listed in grey. Protein conditions are coloured blue (iron replete) and pink (iron deplete).

7.6.5.2 Protein analysis

Refined peptides features were mapped at the protein level, with conflicting peptide matches resolved by manual interpretation, from which the protein statistics identified a total of 680 statistically significant differentially expressed proteins (Figure 7-16). The differentially expressed proteins are distinctively clustered relative to the samples which as expected are grouped tightly in threes, with biological replication offering the greatest degree of variability within sample groupings, and much tighter grouping within technically replicated samples. The principal component analysis revealed that 88.28 % of the total variance likely resulted from the effect of different medium composition, and that the secondary dimension accounted for just 5.64 % of the variance, likely arising from the biological replications performed.

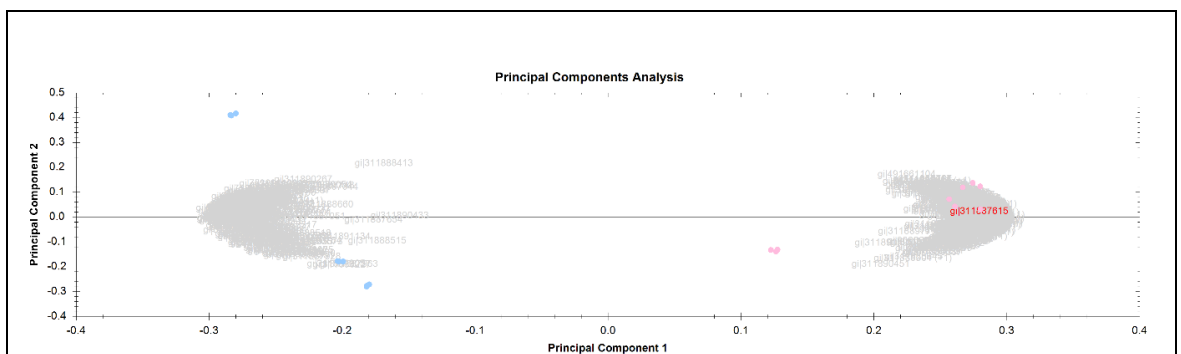


Figure 7-16: Principal Component Analysis generated by Progenesis for significant and differentially expression proteins in the large-scale proteomic survey.

Proteins filtered by ANOVA p-value ≤ 0.05 and Max fold change ≥ 2 and peptide count > 2 .

Further analysis indicated that 445 proteins achieved the desired thresholds[±] to be considered upregulated in iron deplete conditions, from which the dataset in Figure 7-17 visibly isolates the differentially expressed proteins of interest as well as presenting the standardised expression profiles for the selected group, highlighting evident upregulation under iron limited conditions.

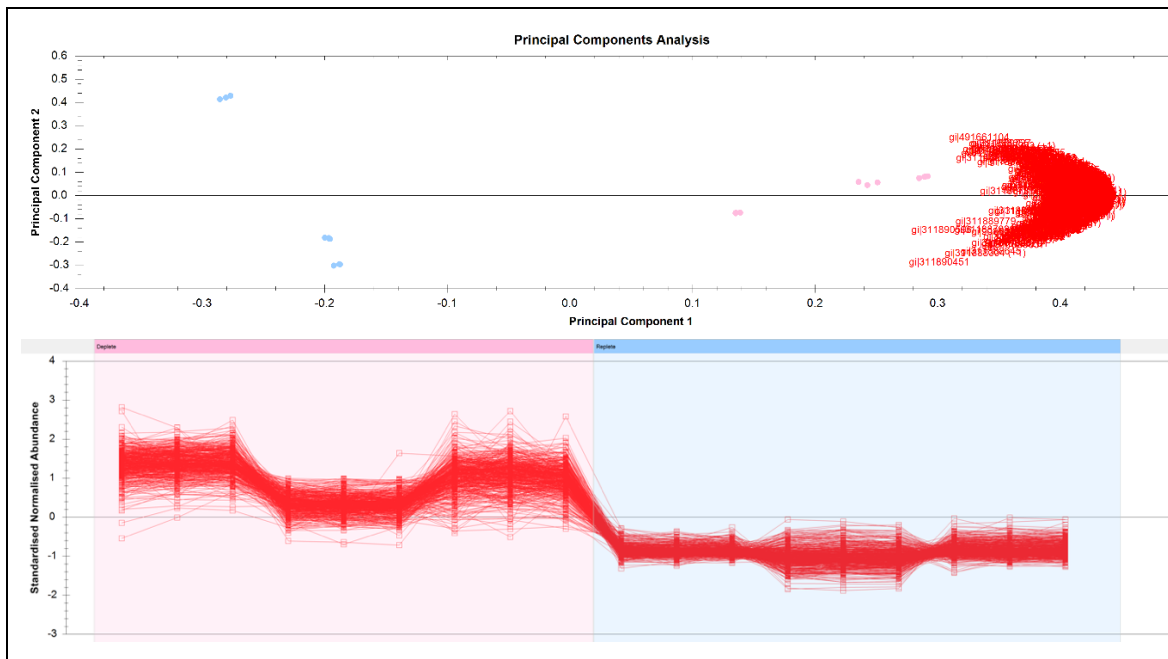


Figure 7-17: Principal Component Analysis and Standardised Expression profiles generated by Progenesis for significant and upregulated proteins under iron limited conditions in the large-scale proteomic survey.

Proteins filtered by ANOVA p-value ≤ 0.05 and Max fold change ≥ 2 and peptide count > 2 .

[±] Protein statistical refinements: ANOVA (p < 0.05), > 2 peptide matches; ≥ 2 max fold change and 'iron deplete' highest mean value.

The report detailing *R. equi* proteins upregulated under iron limitation was exported and further scrutinised, exploring the putative cellular localisation of each candidate using the previously described *in silico* pipeline for prediction of transmembrane topology and detection of signal peptides. The analysis outputs were assembled into a single spreadsheet and a consensus was established based upon majority. An abridged ‘non-cytoplasmic protein’ version is displayed in Table 7-6.

Table 7-6: Abridged results for protein localisation of in the large-scale proteomic survey

Gene	Accession	Peptides	Fold	Description	Consensus
REQ_18820	CBH47948.1	12	75.95	putative membrane protein	Membrane
REQ_28560	CBH48875.1	13	58.56	putative ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_18810	CBH47947.1	11	41.31	putative integral membrane protein	Membrane
REQ_39540	CBH49938.1	5	22.75	putative substrate binding lipoprotein	Lipoprotein
REQ_08150	CBH46930.1	10	22.47	putative substrate binding lipoprotein	Lipoprotein
REQ_12100	CBH47307.1	3	21.52	putative cation transporter ATPase P-type	Membrane
REQ_31890	CBH49190.1	12	18.96	putative substrate binding lipoprotein	Lipoprotein
REQ_03550	CBH46494.1	7	16.39	putative substrate binding lipoprotein	Lipoprotein
REQ_03460	CBH46486.1	15	13.81	putative substrate binding lipoprotein	Lipoprotein
REQ_28570	CBH48876.1	2	13.39	putative ABC transporter integral membrane subunit	Membrane
REQ_12110	CBH47308.1	5	13.16	putative secreted protein	Secreted
REQ_22900	CBH48336.1	6	10.93	putative ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_21450	CBH48198.1	2	10.67	putative oligopeptide/dipeptide ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_31600	CBH49161.1	2	9.62	putative integral membrane protein	Membrane
REQ_07560	CBH46872.1	4	9.17	putative ABC transporter integral membrane subunit	Membrane
cysP1	CBH48968.1	15	8.72	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP1	Lipoprotein
REQ_47000	CBH50642.1	12	7.99	putative ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_40830	CBH50055.1	4	7.78	putative ABC transporter substrate binding subunit	membrane
REQ_34800	CBH49471.1	7	7.74	putative cation transporter ATPase P-type	Membrane
REQ_40020	CBH49983.1	2	7.74	putative branched-chain amino acid ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_22890	CBH48335.1	2	7.62	putative ABC transporter integral membrane subunit	Lipoprotein
iupA	CBH48450.1	2	7.33	putative ABC transporter integral membrane subunit IupA	Membrane
REQ_23200	CBH48366.1	3	7.18	putative lipoprotein	Lipoprotein
REQ_40440	CBH50021.1	3	6.62	putative integral membrane protein	Secreted
REQ_00480	CBH46208.1	8	6.25	putative secreted cutinase	Secreted
REQ_16020	CBH47679.1	3	5.21	putative ABC transporter substrate binding lipoprotein	Lipoprotein
iupB	CBH48451.1	2	5.12	putative ABC transporter integral membrane subunit IupB	Membrane
REQ_01910	CBH46341.1	3	4.96	putative mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase LytB	Secreted
REQ_03191	CBH46459.1	5	4.24	putative secreted protein	Membrane
cysP2	CBH48966.1	11	4.04	sulfate/thiosulfate ABC transporter substrate binding lipoprotein CysP2	Lipoprotein
ftsI	CBH48804.1	4	3.93	penicillin-binding protein FtsI	secreted
REQ_28170	CBH48836.1	3	3.57	putative secreted peptidase	Secreted
REQ_18620	CBH47928.1	4	3.44	putative beta-lactamase	Lipoprotein
mce4C	CBH49663.1	6	3.39	putative Mce family protein Mce4C	membrane
mce2C	CBH46863.1	2	3.33	putative Mce family protein Mce2C	Membrane
gpdA2	CBH49133.1	5	3.24	NAD-dependent glycerol-3-phosphate dehydrogenase GpdA2	Secreted
REQ_20510	CBH48109.1	8	3.17	putative ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_34780	CBH49469.1	2	3.16	hypothetical protein REQ_34780	Membrane
REQ_29940	CBH49009.1	7	3.15	putative substrate binding lipoprotein	Lipoprotein
mce4E	CBH49661.1	3	3.15	putative Mce family lipoprotein Mce4E	Lipoprotein
REQ_40430	CBH50020.1	2	3.02	putative secreted peptidase	Secreted
REQ_23040	CBH48350.1	4	2.94	putative secreted chitinase	secreted
REQ_00150	CBH46176.1	5	2.93	putative secreted lipase	Secreted
REQ_28660	CBH48885.1	2	2.89	putative lipoprotein	Lipoprotein
mce2D	CBH46864.1	2	2.86	putative Mce family protein Mce2D	Membrane
REQ_37330	CBH49720.1	3	2.76	putative secreted protein	Secreted
REQ_09520	CBH47057.1	3	2.69	putative resuscitation-promoting factor rpfA	Secreted
REQ_23870	CBH48429.1	4	2.62	putative secreted lipase	Secreted
REQ_07760	CBH46891.1	2	2.58	putative secreted protein	Secreted
mce2E	CBH46865.1	6	2.56	putative Mce family lipoprotein Mce2E	Lipoprotein
mce4B	CBH49664.1	4	2.49	putative Mce family protein Mce4B	Secreted
REQ_40400	CBH50017.1	2	2.48	conserved hypothetical protein	membrane
REQ_27950	CBH48813.1	2	2.47	putative membrane protein	Lipoprotein
REQ_14510	CBH47531.1	2	2.44	putative secreted protein	Secreted
REQ_46060	CBH50557.1	6	2.43	putative secreted cutinase	membrane
REQ_01250	CBH46276.1	8	2.4	secreted short chain dehydrogenase	Secreted
REQ_13230	CBH47412.1	2	2.36	putative lipoprotein	Lipoprotein
REQ_18910	CBH47957.1	2	2.36	putative secreted protein	secreted
mce4D	CBH49662.1	4	2.35	putative Mce family protein Mce4D	membrane
REQ_04950	CBH46621.1	5	2.31	putative secreted protein	Secreted
REQ_18060	CBH47874.1	5	2.26	putative serine peptidase	Membrane
cysT	CBH48965.1	3	2.19	sulfate/thiosulfate ABC transporter integral membrane subunit CysT	Membrane
REQ_18660	CBH47932.1	3	2.16	putative ABC transporter substrate binding lipoprotein	Lipoprotein
REQ_33700	CBH49366.1	2	2.15	putative LytR family transcriptional regulator	membrane

7.6.5.3 Discussion of results from the large-scale proteomic survey

The *in-silico* analysis of the final dataset was performed to remove cytoplasmic contamination as the project aims were relevant to identification of iron-regulated proteins associated with the rhodococcal cell wall that might be involved in aspects of trans-mycolate transport. The bioinformatic protein localisation pipeline indicated that 64 proteins were likely to occupy a non-cytoplasmic location, further reducing the dataset complexity by 86 %. Examination of the proteins listed in Table 7-6 yielded an assortment of proteins that was submitted to STRING v11.0 for a multi-protein analysis to predict networking based upon current interactions within the dataset (Figure 7-18).

The non-cytoplasmic proteins isolated produce a number of visible clusters; the most extensive being an apparent iron-siderophore importation system that displays a tight bundle of putative ABC transporter integral membrane subunits that include the IupA and IupB proteins connected with a number of putative ABC transporter substrate binding lipoproteins. This is reinforced by the study performed by Miranda-CasoLuengo et al. (2005) that presented an *R. equi iupA* transposome mutant that failed to grow at low iron concentrations.

Another evident functional protein cluster is the collection of putative Mce family proteins, that are typically a large group of secreted or surface-exposed proteins that confer virulence under stress by facilitating entry into mammalian cells and survival within macrophages. This is entirely consistent with iron-dependent regulation of virulence gene expression in many bacterial systems (Meena and Rajni, 2010) The third notable protein cluster is a triad of sulfate/thiosulfate ABC transporter proteins that are involved with regulation of rhodococcal sulphur metabolism, including transport of extracellular sulphate across the cell envelope. The fourth protein cluster of potential interest involves a putative LytR family transcriptional regulator; signal peptidase and a penicillin-binding protein that are likely involved with cell wall biogenesis. The final protein cluster of interest arranges in a loose octahedral functional nexus linking two putative secreted cutinases with both lipoproteins and integral membrane proteins. The remaining nodes in Figure 7-18 were not predicted to be functionally related to other proteins identified in this study, however these may be of particular interest regarding cellular response to iron limitation within *R. equi* regarding transportation across the mycolate layer, therefore proteins that are predicted to be secreted or contain lipid anchors should

be explored further.

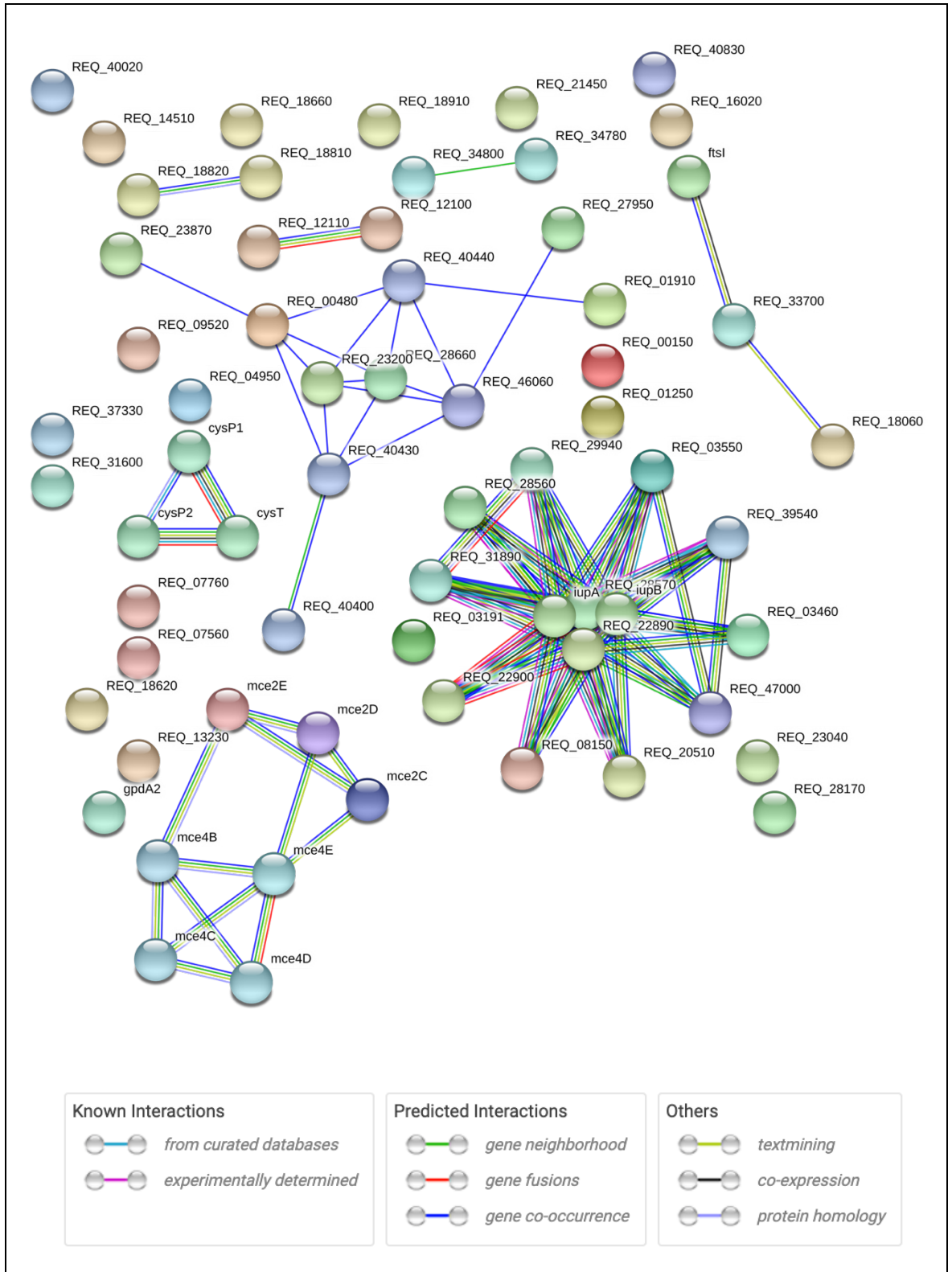


Figure 7-18: Evidence based STRING analysis for the multiple non-cytoplasmic proteins identified by the large-scale proteomic survey and *in silico* correlation methods. Protein identification is mapped to the gene identity listed in Table 7-6. Predicted interactions are listed in the legend.

7.6.5.4 Identification of potential iron transport systems

The non-cytoplasmic proteins in Table 7-6, upon network analysis via STRING (Figure 7-18), clustered four putative ABC transporter integral membrane subunits (Table 7-7) connected with ten putative ABC transporter substrate binding lipoproteins (

Table 7-8).

Table 7-7: Putative ABC transporter integral membrane subunits.

Gene ID	Accession	Description
iupA	CBH48450.1	putative ABC transporter integral membrane subunit IupA
iupB	CBH48451.1	putative ABC transporter integral membrane subunit IupB
REQ_28570	CBH48876.1	putative ABC transporter integral membrane subunit
REQ_22890	CBH48335.1	putative ABC transporter integral membrane subunit

Table 7-8 Putative ABC transporter substrate binding.

Gene ID	Accession	Description
REQ_03460	CBH46486.1	putative substrate binding lipoprotein
REQ_03550	CBH46494.1	putative substrate binding lipoprotein
REQ_08150	CBH46930.1	putative substrate binding lipoprotein
REQ_20510	CBH48109.1	putative ABC transporter substrate binding lipoprotein
REQ_22900	CBH48336.1	putative ABC transporter substrate binding lipoprotein
REQ_28560	CBH48875.1	putative ABC transporter substrate binding lipoprotein
REQ_29940	CBH49009.1	putative substrate binding lipoprotein
REQ_31890	CBH49190.1	putative substrate binding lipoprotein
REQ_39540	CBH49938.1	putative substrate binding lipoprotein
REQ_47000	CBH50642.1	putative ABC transporter substrate binding lipoprotein

The identification of the proteins IupA and IupB that are predicted to be ABC transporter components responsible for translocation of iron-bound complexes of the *R. equi* catecholate siderophore rhequibactin across the cytoplasmic membrane as predicted by Miranda-CasoLuengo et al. (2005) is an excellent indication that the correct experimental conditions were used. Interestingly, the authors of this paper describe an accumulation of catecholates within an *iupABC* transposon-mutant containing strain. Given that *R. equi* is predicted to synthesise two putative siderophores, it stands to reason that an alternative ABC transport system may be utilised for rhequichelin. The functional STRING analysis clustered two additional ABC transport membrane proteins each within a distinct operon, these may provide excellent targets for mutagenesis to characterise siderophore specific cytoplasmic translocation mechanisms. While the researchers above predict the mechanisms associated with siderophore-mediated cytoplasmic membrane translocation, they fail to discuss the implications of the rhodococcal cell-envelope region as a whole, and how the iron-siderophore

complexes translocate through the mycolate layer.

Unusually the putative ABC transporter integral membrane subunit protein encoded by REQ_22890 appears to contain a lipid -anchor yet is also contained within an operon that contains a putative ABC transporter substrate binding lipoprotein. This could provide a future avenue for exploration, relating to iron transport. The other proteins of interest include the ten putative substrate binding lipoproteins listed in

Table 7-8 and their function. The literature associated with Gram-positive bacteria denotes that lipid-modified substrate binding proteins anchored to the outer side of the cytoplasmic membrane facilitates translocation through the ABC transport system (Miethke and Marahiel, 2007), for example *B. subtilis* utilises five lipid-anchored substrate binding proteins as exclusive gateways for extracellular iron-siderophore-complex recognition prior to uptake into the cytoplasm (Ollinger et al., 2006).

For this study, there was no identification of the putative *R. equi* HupB protein that functions as a siderophore-complex transporter mechanism within the mycolate layer of *M. tuberculosis*, therefore, the remaining putative secreted and lipid-anchored proteins were compiled for further analysis (Table 7-9).

Table 7-9: A list of candidates identified to contain a secretory signal or lipid anchor signal

Gene	Accession	Description
REQ_00150	CBH46176.1	putative secreted lipase
REQ_00480	CBH46208.1	putative secreted cutinase
REQ_01250	CBH46276.1	secreted short chain dehydrogenase
REQ_04950	CBH46621.1	putative secreted protein
REQ_07760	CBH46891.1	putative secreted protein
REQ_09520	CBH47057.1	putative resuscitation-promoting factor rpfA
REQ_12110	CBH47308.1	putative secreted protein
REQ_13230	CBH47412.1	putative lipoprotein
REQ_14510	CBH47531.1	putative secreted protein
REQ_16020	CBH47679.1	putative ABC transporter substrate binding lipoprotein
REQ_18620	CBH47928.1	putative beta-lactamase
REQ_18660	CBH47932.1	putative ABC transporter substrate binding lipoprotein
REQ_18910	CBH47957.1	putative secreted protein
REQ_21450	CBH48198.1	putative oligopeptide/dipeptide ABC transporter substrate binding lipoprotein
REQ_22890	CBH48335.1	putative ABC transporter integral membrane subunit
REQ_23040	CBH48350.1	putative secreted chitinase
REQ_23200	CBH48366.1	putative lipoprotein
REQ_23870	CBH48429.1	putative secreted lipase
REQ_27950	CBH48813.1	putative membrane protein
REQ_28170	CBH48836.1	putative secreted peptidase
REQ_28660	CBH48885.1	putative lipoprotein
REQ_37330	CBH49720.1	putative secreted protein
REQ_40020	CBH49983.1	putative branched-chain amino acid ABC transporter substrate binding lipoprotein
REQ_40430	CBH50020.1	putative secreted peptidase
REQ_40440	CBH50021.1	putative integral membrane protein

These potential proteins of interest were analysed using STRING Co-occurrence analysis to the gene family pattern across a selection of relevant Corynebacteriales taxa (Figure 7-19). Co-occurrence was predicted by gene alignment; score graded by intensity. Interestingly, there is a significant difference between mycobacterial species and the *R. equi* gene selection, with many genes not present, this may perhaps refine the proteins of interest further if *R. equi* is to utilise an analogous mycolate transport system.

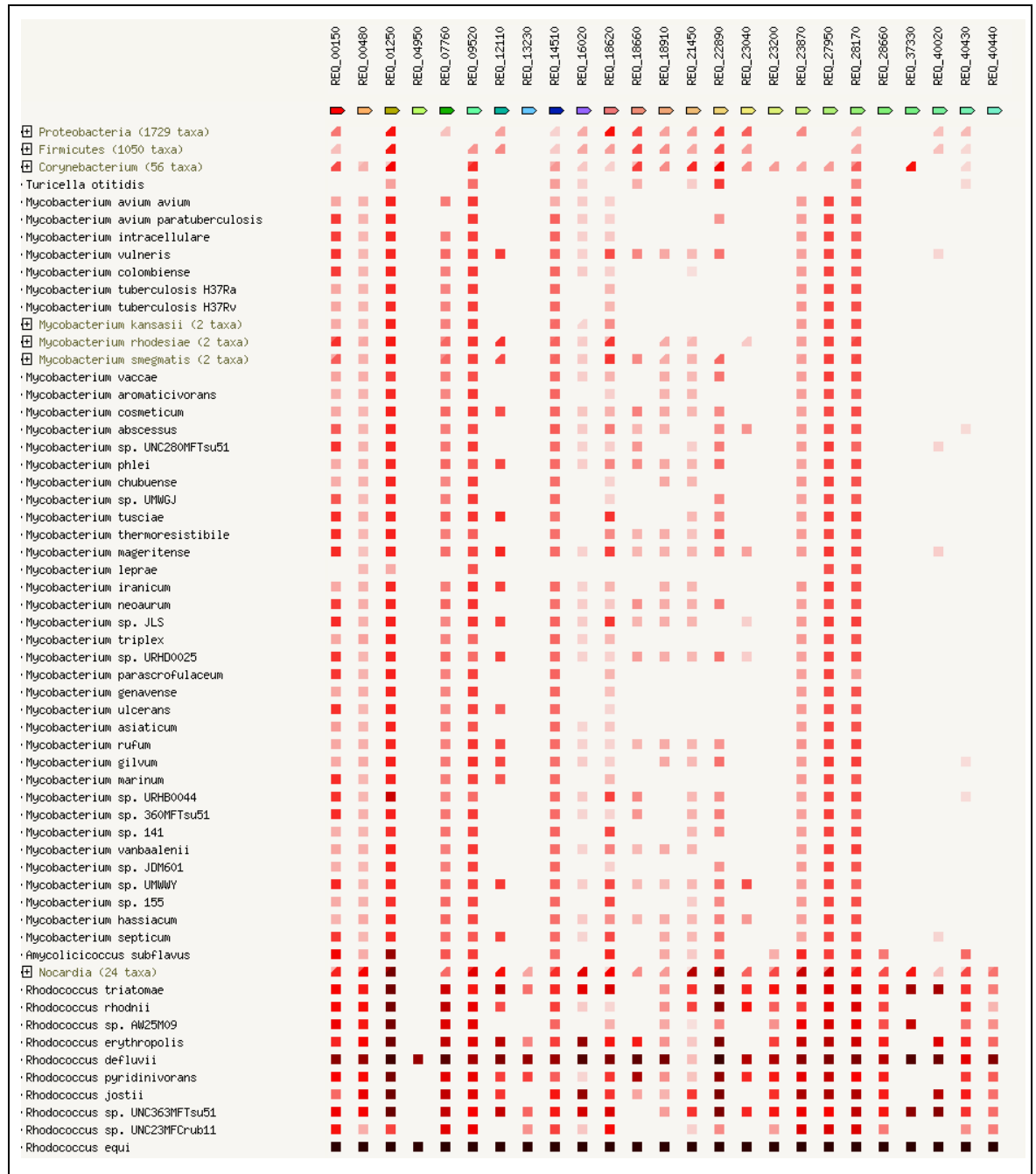


Figure 7-19: Gene co-occurrence between the candidate list generated in Table 7-9 from *R. equi* against other closely related bacterium from the order Corynebacteriales.

7.7 Discussion

The aims of the project were to explore the mechanisms utilised by *R. equi* to facilitate iron transport across the cell envelope, by firstly identifying proteins associated with isolated cell envelope and subsequent identification of any potential outer membrane receptors that may facilitate transport of iron across the mycolate.

The project was developed using the iron limiting conditions outlined in chapter 6; the subsequent proteomic analysis of the rhodococcal cellular material was performed to enhance the understanding of uptake mechanisms within acid-fast bacteria especially under nutritional stress.

7.7.1 Outcomes of the Proteomic study

A significant amount of the project was dedicated to identifying the most suitable fractionation conditions for isolation of the rhodococcal mycolate-containing cell wall region, comparing the unfractionated clarified lysate analysis with *in vitro* fractionation using the density gradient medium Percoll at 60 % (v/v) with the buoyant cell wall-containing phase solubilised using either SDS-buffer or Triton X-114 before downstream proteomic analysis. Comparison of these methodologies produced a range of positive and negative attributes that were considered for the experimental design for the final proteomic study. Utilisation of *in vitro* fractionation methods provided an enhanced view of retained hydrophobic proteins with a reduction of cytoplasmic contamination the fraction, however there was clear indication of false negative results resulting from transfer and recovery losses during sample processing, signifying that the more aggressive the isolation technique, the more limiting the analysis would be. Therefore, exploratory *in silico* localisation methods were employed to identify proteins relevant to cell envelope region through the use of a protein localisation pipeline to predict transmembrane topology and signal peptides using the bioinformatic tools TOPCONS (Tsirigos et al., 2015), PredLIPO (Bagos et al., 2008), SignalP (Almagro Armenteros et al., 2019) and Phobius (Kall et al., 2007).

The final large-scale study was performed using a whole-cell analysis with subsequent *in silico* fractionation to target proteins that may be involved with iron uptake mechanisms. The study identified presence of predicted cytoplasmic membrane-associated iron ABC transporter

components, although not relevant to the mycolate-associated mechanisms, their presence reinforces that correct experimental parameters regarding iron limitation were achieved, corroborating with previous literature (Miranda-CasoLuengo et al., 2005).

7.7.2 Project limitations

There are a number of technical limitations associated with this project, firstly arising from the GeLC-MS/MS approach; while this approach is typically synonymous with protein identification in both qualitative and quantitative methods, several associated limitations should be considered.

Firstly, artefactual peptide modifications may arise from in-gel proteomic methods, these can typically arise from incomplete gel polymerisation and may interact with cysteine residues present to reform disulphide bridges; to counteract this, both reduction and alkylation steps are performed when extracting peptides (Müller and Winter, 2017). Moreover, methionine residues are often susceptible to oxidation, likely due to excess persulphate within the gel. Both of these issues are partially mediated through indication of fixed and variable peptide modifications, in the form of carbamidomethylation and oxidation respectively when database searching for MS/MS ions.

However, the most prevalent issue relates to sample loss in GeLC-MS/MS experiments, ranging from incomplete protein solubilisation prior to electrophoretic separation and inadequate peptide following digestion (Paulo, 2016). Comparative to in-solution extraction methods, recovery of peptides can range from between 70 – 90 % (Havlis and Shevchenko, 2004), however the in-solution methodology was not yet established as a robust peptide extraction option at the institution. Other opportunities for peptide loss could occur from absorption to surfaces including pipette tips and microcentrifuge tubes; drying of samples in a vacuum concentrator, and during the electrospray ionisation. These occasions were hopefully mitigated by accurate and careful manipulation of the gel and peptide mixtures, and the utilisation of both biological and technical replications.

The utilisation of an unfractionated approach provided the opportunity to analyse the most significant proportion of peptides available, without potential loss through *in vitro* fractionation methods, however the subsequent *in silico* fractionation pipeline methodology excludes proteins based upon their predicted subcellular localisation. This methodology as a whole effectively circumvents

analysis of potential moonlighting proteins, that may be relevant to the cellular region of interest. If moonlighting proteins were to be explored in the future, a proteomic approach should be explored without risking cytoplasmic contamination throughout by cell lysis methods.

7.7.3 Development of an appropriate hypothesis for iron uptake mechanisms in *R. equi*

While no obvious candidates for iron transport across the mycolate region were identified in rhodococci, this chapter aims to raise the issue of the lack of discussion and subsequent investigation regarding the mycolate-related pathogenicity and survivability mechanisms associated with nutritional stress.

In contrast, mycobacteria have the ability to synthesise salicylate-containing siderophores known as mycobactins; carboxymycobactin, a water-soluble secreted siderophore and the hydrophobic mycobactin that is cell-surface associated (Snow and White, 1969), from which it is proposed that HupB, a 28-kDa iron-regulated cell wall-associated protein in *M. tuberculosis* arbitrates transfer from the iron-bound carboxymycobactin to the mycobactin (Yeruva et al., 2006).

As a result of no clear identification of an analogous HupB-like protein during this proteomic study, in combination with the siderophore predictions identified in chapter 3, it is highly unlikely that *R. equi* utilises a similar approach to iron uptake. Regardless of this, the acid-fast bacterium still requires a method of translocation of ferri-siderophores across the mycolate region. Given the dearth of evidence at this stage, it is proposed that *R. equi* utilises facilitated diffusion via a simple porin route for ingress of ferri-siderophore complexes into the periplasmic space (Figure 7-20), from where a substrate binding lipoprotein may facilitate transfer through an ABC transporter system such as the previously identified *iupABC* complex. Thus, the system found in *R. equi* might resemble that of a canonical Gram-positive active transporter of a ferri-siderophore, with all of the specific recognition events occurring in the relatively protected environment of the periplasm. The proposed system would be unusual in that the high-affinity receptor protein expected of such a transport system (Ratledge and Dover, 2000) would not be found at the outer surface of the bacterium. The periplasmic ferri-siderophore binding proteins of Gram-negative bacteria often have broader

specificities and poorer affinities for their ligands than their high affinity outer membrane receptor counterparts (Ratledge and Dover, 2000); biochemical analysis of ligand binding kinetics and comparison with their Gram-negative and Gram-positive functional counterparts might be informative in understanding whether these lipoproteins operate as primary high-affinity receptors or ancillary transport proteins.

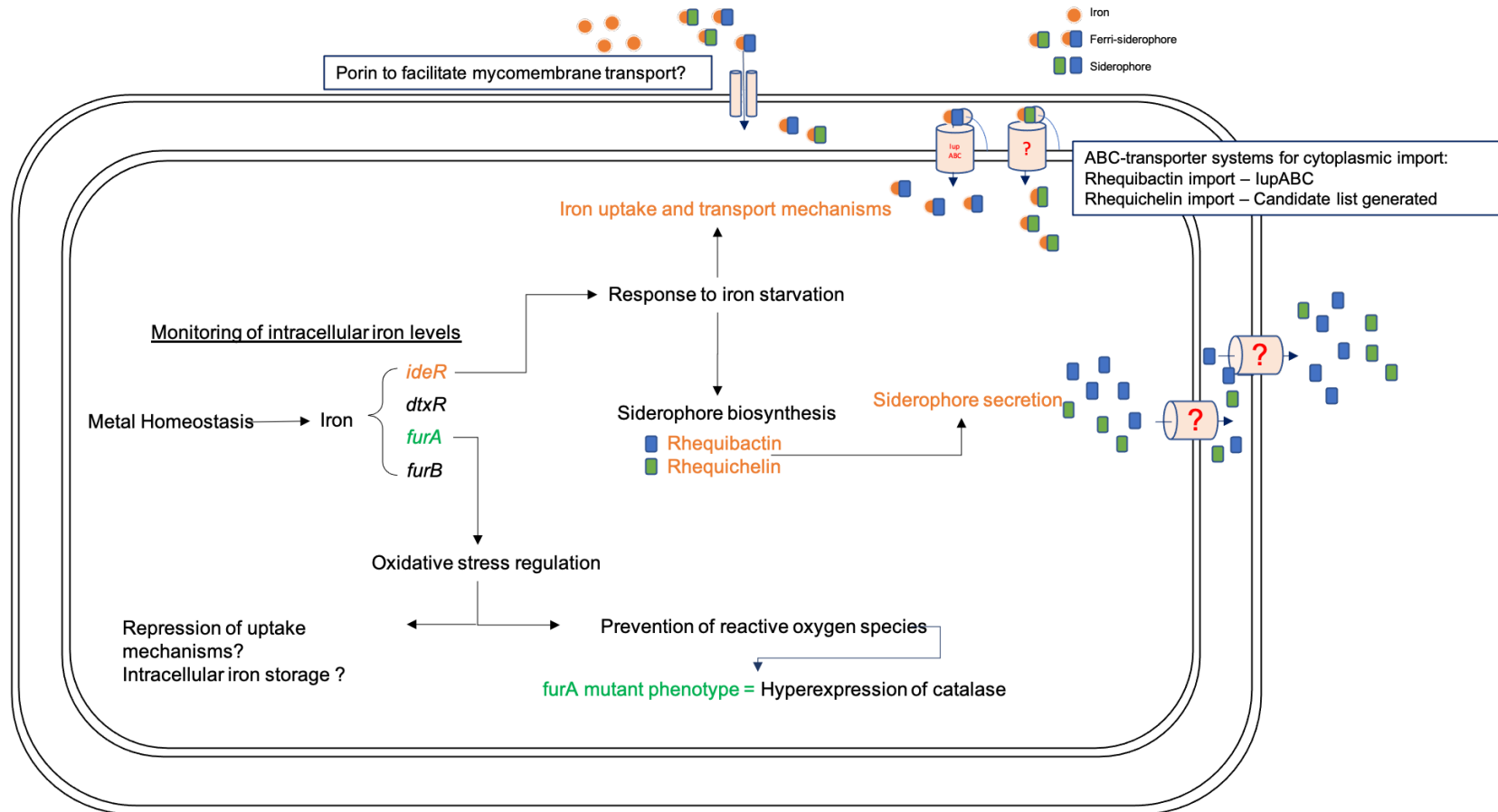


Figure 7-20: An updated *R. equi* iron homeostasis network map (after Chapter 7)

Data generated from the preliminary genomic survey, developed using in silico metal binding capacities, and the production of metal regulator mutants. Siderophore candidates were identified *via* untargeted and subsequent targeted metabolomic analyses. Transport mechanisms were targeted using in silico and in vitro fractionation coupled with proteomic analysis, shown in orange as further characterisation is needed. Specific transport components are displayed in red as no viable candidates were identified in this study.

8 Method optimisation for RNA isolation for use in
RNA-Seq analysis of *Rhodococcus equi* 103S
global gene expression under iron starvation

8.1 Introduction

Transcription, being the first stage of gene expression where RNA is produced from a DNA template by RNA polymerase is the fundamental intermediary linking the genome and proteome. Transcriptional studies have become a cornerstone of systems biology, facilitating quantification of gene expression. Since the completion of the Human Genome Project, transcriptomics has quickly become one of the most developed fields, with hopes of pulling back the veil on how genome expression and cellular function are intertwined (Dong and Chen, 2013). Transcriptomic analysis is performed by studying the transcriptome of an organism, in effect, the totality of its RNA transcripts under a certain physiological condition. (Dong and Chen, 2013; Lowe et al., 2017).

Initial gene expression studies were performed using low-throughput single transcript measurement methods such as Northern blotting and quantitative PCR, however transcript analysis has progressed to incorporate genome-wide quantification of gene expression (Kukurba and Montgomery, 2015), initially from expressed sequence tags and subsequent serial analysis of gene expression (SAGE) (Velculescu et al., 1995) through microarray analyses and more recently high-throughput RNA-sequencing (RNA-seq) (Nagalakshmi et al., 2010).

The development of microarrays facilitated the measurement of gene transcript abundance by hybridisation of fluorescently labelled transcripts to short immobilised oligonucleotide probes. The ability to multiplex array probes facilitated thousands of transcripts to be read simultaneously, thereby reducing associated time and labour costs. The microarray methodology encountered a range of limitations including a requirement of priori knowledge of the sequences in question, potential for cross-hybridisation events occurring between homologous sequences and difficulty in quantification of transcripts at both very low and very high expression levels (Kukurba and Montgomery, 2015).

Ultimately, transcriptomic profiling was performed using spotted oligonucleotide arrays (Shalon et al., 1996) and high density Affymetrix GeneChip arrays (Irizarry et al., 2003) until the late 2000's with the surfacing of the genomic era when RNA-seq methods was considered a viable option (Kukurba and Montgomery, 2015).

The rapid development of high-throughput sequencing technologies transformed the field of

transcriptomics by facilitating RNA analysis through sequencing of complementary DNA (cDNA), from which the calculated abundance was derived from the number of counts from each transcript. The first use of RNA-seq methodology in research was published by Bainbridge et al. (2006) with over 100,000 transcripts analysed using the ‘sequencing by synthesis’ methodology of Roche’s 454 pyro-sequencer to quantify relative transcript abundance. As previously alluded, the RNA-Seq era began in the late 2000’s with the competition between Solexa and Illumina facilitating the determination of over 1,000,000,000 transcripts (Kukurba and Montgomery, 2015; Oszolak and Milos, 2011).

Development of the RNA-sequencing technologies have improved the knowledge of both qualitative and quantitative aspects of transcriptomics, including a deeper understanding of transcript initiation sites, detection of alternative splicing and gene fusion transcripts as well as identification and indexing both sense and antisense transcripts (Kukurba and Montgomery, 2015).

8.1.1 Bacterial transcriptomics

As for eukaryotic organisms, analysis of the bacterial transcriptomes was instigated through development of the microarray technology, of which, high-density tiling arrays were at the forefront of bacterial transcript analysis to globally quantify gene expression without needing to account for genome annotations (Toledo-Arana and Solano, 2010). With the introduction of RNA-seq and the associated deep sequencing of cDNA, the shortcomings of microarray methodologies can be circumvented; with sequencing resolution down to single bases; and improved signal-noise ratios, a higher dynamic range can be achieved (Vivancos et al., 2010).

Stemming from the genomic era revolution and the race for completion of the human genome project, the developments within and access to sequencing technologies have resulted in the significant decrease in cost for next-generation sequencing far below the expected values portrayed by Moore’s Law (Muir et al., 2016). This in combination with the freely available, open source bioinformatic analysis tools presents RNA-seq technology as an appealing method of bacterial transcript analysis (Haas et al., 2012).

The utilisation of RNA-seq in bacterial transcriptomics has been paramount in generation of deep

and comprehensive transcriptome profiles, as the relative abundance of individual transcripts can vary by up to several orders of magnitude (Haas et al., 2012). Sufficiently deep profiling facilitates detection of quantification of low abundance transcripts, this is typically achieved through increasing the total read count per library generated, thereby producing sufficient reads to detect the biologically relevant transcripts that constitute a minority in the cDNA library. Secondly, the depletion of abundant transcripts by rRNA depletion prior to cDNA generation can be utilised to enrich the mRNA levels for quantification of rarer or lower expressed transcripts downstream (Giannoukos et al., 2012).

Ensuring that a transcriptome can be mapped both efficiently and effectively is essential for the majority of RNA-seq projects; and as such, the experiments must incorporate a balance of sequencing depth with the quantity of samples sequenced. Haas et al. (2012) discusses this approach in detail, indicating that use of RNA-seq for transcriptome mapping should utilise approaches focussing on depth that can detect rare transcripts for full coverage, with breadth of coverage (biological replicates) perhaps becoming more relevant than sequencing depth (reads per sample) for applications comparing gene expression levels under specific growth conditions or strain variants, with indication that a balance is between these parameters is essential for experimental design.

8.2 Aims of the chapter

The aim of this chapter is to reveal the genetic response to iron starvation; the elucidation of gene transcription under these conditions will provided a snapshot of total transcripts present in a cell. The resulting transcriptomic analysis will corroborate with other omics analysis methods in characterising the global iron regulatory mechanisms utilised in the veterinary pathogen *Rhodococcus equi* 103S.

8.3 Experimental Design

Given the funding and time limitations within this project, the experimental design was constructed to achieve a significant transcriptome coverage without exceedingly high numbers of transcript reads and the associated costs.

Initial consultation with in-house RNA-seq experts (Dr D Smith & Dr A Nelson, pers. comm),

resulted in a recommendation of ~10 million reads per sample for accuracy and sufficient depth for an appropriate level of transcript coverage (Haas et al., 2012).

The findings from a systematic analysis of RNA-seq profiling in bacterial transcriptomics performed by Haas et al. (2012) were consistent with this recommendation stating that their analysis identified that 5 – 10 million reads would typically be sufficient to detect the vast majority of expressed genes under a wide variety of experimental conditions apart from those genes that are expressed at extremely low levels. Furthermore, if RNA-seq analysis can be performed incorporating biological replicates, differential expression of genes can be detected with high statistical significance, with reads per sample reduced to as low as 2 – 3 million (Haas et al., 2012).

In order to define the recommended number of reads required for this experiment, a targeted search for RNA-seq transcript profiling was performed restricted to the taxon *Rhodococcus*. At the time of writing, *R. equi* transcription does not appear to have been analysed *via* RNA-seq in any capacity, although there are a number of studies that focus on other rhodococci. , While studying regulation of plasmid-encoded isoprene metabolism in *Rhodococcus* sp. ACPA1, ACPA4 and ACS1, (Crombie et al., 2015) , scrutinised isoprene-related gene expression using a time-course starvation experiment in an analogous manner to the iron-starvation experiments determined in chapter 6. Changes in gene expression were mapped by transcriptome sequencing using RNA-seq, wherein 475 million reads were performed in total; accounting for biological replicates resulted in a sequence read depth of between 3 and 5 million reads (and no less than 2 million reads) per sample, with the study citing the systematic analysis by Haas et al. (2012).

Given the stringent limitations associated with this project outlined previously, and the information available, the experiment was designed to map the transcriptome of wild-type *R. equi* 103S under iron limiting conditions and iron-replete conditions, with sufficient biological replicates. Furthermore, the scope also facilitated transcriptomic analysis of one of the four *R. equi* mutants (generated in Chapter 5) under iron replete conditions.

The mutant selected for transcriptomic analysis was selected based upon which would likely contribute the most to understanding the iron regulation within *R. equi*. Using the predictions from Chapter 4, the mutant *R. equi-ΔideR* was selected based upon the regulatory responses predicted

under iron starvation over the other predicted iron-regulatory mutant *R. equi-ΔfurA*, given the limited phenotypic and proteomic changes already observed for the latter. The other two mutants *R. equi-ΔdtxR* and *R. equi-ΔfurB* were not selected based upon their predicted binding with the other divalent cations, manganese and zinc respectively.

Using the relevant literature and calculation of the desired experimental conditions, the transcriptomic sequencing was to be performed on RNA harvested from: 1) Wildtype *R. equi* 103S, 2) Wildtype *R. equi* 103S under iron limitation, and 3) Mutant *R. equi-ΔideR*. With the experimental design incorporating three biological replicates and two technical replicates for each sample, totalling 18 runs which is the maximum viable runs based on the reagents available at 2.5 million reads per sample.

8.4 Results

8.4.1 Cell culture and media modifications

Typically, 10 mL of logarithmic phase cultures in a rich medium such as LB or BHI would achieve $0.5-1 \times 10^9$ cells, however the transcriptomic analysis performed was to identify differential expression resulting from iron limitation in the growth media or caused by gene knockouts under iron limitation.

Therefore, samples collected were expected to be a stressed stationary phase culture, harvested after ~5 days growth, identical to harvest conditions in the metabolomics and proteomics analyses. The cellular harvest was performed upon the visual cue of production of the red chromophore as a convenient biomarker for iron limitation.

8.4.2 Initial RNA isolation using RiboPure Bacteria Kit

Initial isolation and purification of *R. equi* total RNA was performed as described in the methods 2.15 using the RiboPure Bacteria RNA isolation kit (Ambion, Life Technologies), from which the RNA integrity was assessed using a Prokaryotic Bioanalyzer RNA 6000 Nano Assay yielding a score of 0 for all samples.

The lack of detectable RNA or even degraded material was a potential indicator that either the

rhodococcal cells were not lysing, likely as a result of the thick waxy mycolic acid-containing cell wall, or if lysing, they may have a particularly short RNA half-life. Extensive studies have been performed highlighting these issues within the Gram-positive mycolate-containing bacteria (Akhtar et al., 2011). Therefore, to improve isolation of intact high-quality RNA, the methodology was modified to incorporate an aggressive cell lysis procedure using a FastPrep FP120 Cell Disrupter. The isolation step was repeated and the RNA integrity subsequently analysed (Table 8-1).

Table 8-1: Prokaryote Total RNA Nano Series II Assay

to detect RNA concentration, rRNA ratio and RNA integrity in *R. equi* samples using an Agilent Bioanalyzer 2100 (Version B.02.08.SI648). For sample identification: WT – wildtype *R. equi* 103S; IDER – *R. equi*- Δ *ideR*; D – iron deplete conditions; R – iron replete conditions; numbers indicative of biological replicate. Ladder – A proprietary molecular marker

Sample	RNA Concentration:	rRNA Ratio [23s / 16s]:	RNA Integrity Number (RIN):
WT1-D	85.40598	0	5.2 (B.02.08)
WT2-D	38.49717	0.8934986	5 (B.02.08)
WT3-D	95.82452	0	5.4 (B.02.08)
WT1-R	203.358	0.9472562	6.9 (B.02.08)
WT2-R	245.9775	0.9842288	6.9 (B.02.08)
WT3-R	128.4064	0.9812167	6.5 (B.02.08)
IDER1-R	234.5266	1.160092	6.6 (B.02.08)
IDER2-R	269.706	1.233795	N/A (B.02.08)
IDER3-R	192.7845	1.339374	7.5 (B.02.08)
Ladder	150	N/A	N/A (B.02.08)

Although a marked improvement was achieved on previous attempts, the RNA integrity here was still lower than the minimum level generally considered sufficient ($RIN \geq 7$) (Jahn et al., 2008), with RIN values of between 9 and 10 considered ideal to accurately quantify gene expression; crucially, although RNA degradation may not result in complete loss of transcripts, it can significantly affect the associated estimated relative gene expression levels (Gallego Romero et al., 2014).

Analysis of the RNA electropherograms, using WT1-R (*R. equi* 103S, iron-replete) (RIN 6.5) as an example (Figure 8-1) suggests partial degradation of the RNA. The presence of multiple peaks in both the inter region and in the fast region is indicative of degradation of the 13S and 26S rRNA to an intermediate scale, correlating to examples provided by Agilent (Mueller et al., 2004), and a RIN value below the desired threshold. Given the apparent degradation at this stage of the rRNA, it is unlikely that the transcripts will be of sufficient quality to infer gene expression levels by RNA-seq.

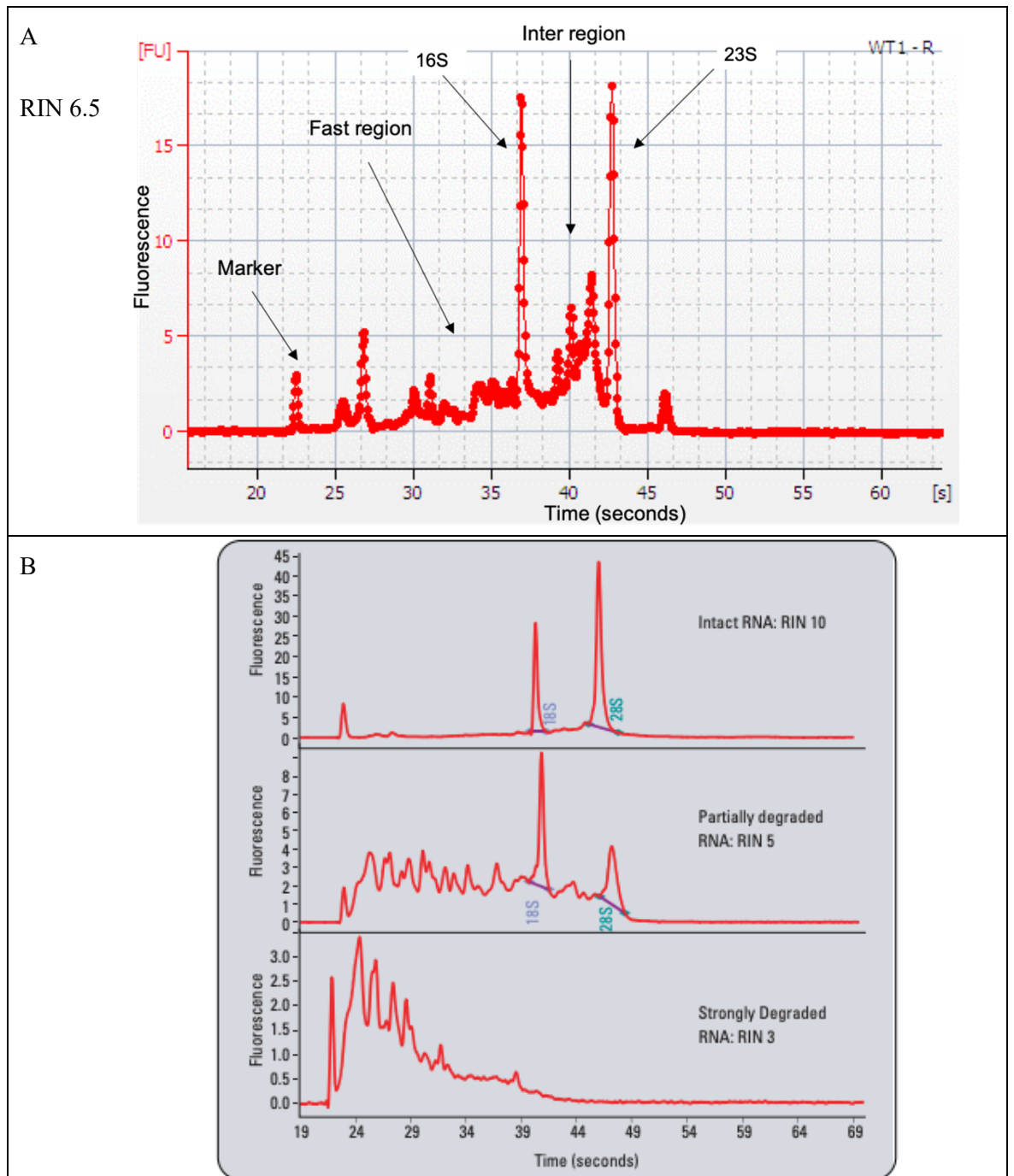


Figure 8-1: Electropherogram analysis of RNA integrity.

A: Wildtype *R. equi* 103S grown in iron replete minimal media conditions, harvested after 5 days growth. B: Examples of different RNA integrity levels (small subunit [16s/18s], large subunit [23s/28s]) (Mueller et al., 2004).

The method development for analysis of RNA integrity by the Agilent Bioanalyzer 2100 has facilitated a quality control procedure regulating an industry standard of best practice. With previously established acceptable RIN cut off values available, (which can vary depending on the type of downstream experiment) the incorporation of this quality control test can ensure that the downstream analysis is performed on RNA of validated acceptable quality above the threshold value,

and if the value falls below this threshold, the samples should be discarded and isolation repeated as demonstrated in the flowchart below (Figure 8-2; adapted from Mueller et al. (2004)).

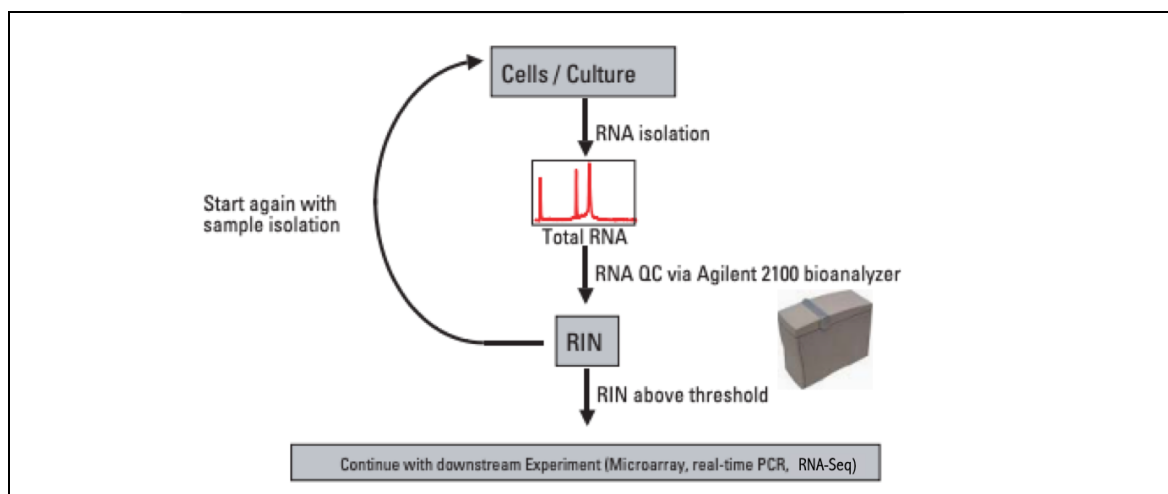


Figure 8-2: Quality control flowchart for RNA integrity analysis. Adapted from (Mueller et al., 2004)

8.5 Future methodology

Advancement of this project requires a significant improvement in quality yielded from the initial RNA isolation step given above. Incorporation of a RIN quality control parameter at this stage is essential in downstream analysis of RNA transcripts (Vermeulen et al., 2011).

8.5.1 Identification of methodology issues

Intriguingly, RNA extraction from mycolic acid-containing bacteria has been well documented over the last 25 years, and isolation of pure, intact RNA has consistently proved challenging (Jahn et al., 2008) through utilisation of both Trizol (Cheung et al., 1994; Payton and Pinter, 1999) and Hot Phenol (Barry et al., 1992; Monahan et al., 2001; Rustad et al., 2009) extraction methods.

Recently, (Venkataraman et al., 2013) performed a comparison of RNA isolation methods in the similarly mycolic acid-containing organism *Mycobacterium tuberculosis*, from which they discuss how the use of Trizol resulted in extensive smearing and RIN comparable with methods used in this chapter, with the conclusion of extensively degraded RNA. Hot Phenol extraction produced significantly higher RIN values (between 8.0 and 8.6), through the extensive use of toxic chemicals at high temperatures. This method was excluded for the *R. equi* extractions, given the inherent risks associated.

8.5.2 Method optimisation

An additional method was also performed for extraction of mycobacterial RNA, this 'Method 3' was an optimised technique using an RNeasy RNA isolation kit (QIAGEN) with a modified protocol, utilising an enhanced cellular lysis step, RNeasy spin column purification and two-step DNase digestion. The resulting QA analysis identified that this method could reproducibly isolate high yields of RNA, both pure and intact, with RIN values between 9.5 and 9.9; and was further confirmed to successfully isolate RNA from another mycolate-containing bacteria *Mycobacterium smegmatis* and the Gram-negative organism *Escherichia coli*.

8.5.2.1 Modified RNeasy methodology for isolation of *R. equi* RNA

In this optimised method described by Venkataraman et al. (2013), culture harvesting and centrifugation was performed as for previous methods; the cell pellet was then suspended in 100 µl of TE buffer (10 mM Tris pH 7.5, 1 mM EDTA pH 8.0) containing 20 mg/ml lysozyme (USB Corp., USA) and 10 µl Proteinase K (1 U/µl; Qiagen Inc., USA).

The cell lysis suspension was resuspended thoroughly and incubated at room temperature for 10 minutes with intermittent vortexing, after which 700 µl of RLT buffer (RNeasy kit, Qiagen Inc.) was added and further resuspended. The suspension was transferred to a RNase free 1.5 ml microcentrifuge tube containing 400 µl of acid-washed zirconia beads and lysed *via* bead-beating for 10 pulses of 20 seconds, each with intermittent chilling on ice.

The cell lysate was centrifuged at 12,000 x g for 20 seconds at room temperature to pellet cellular material and debris, and the supernatant was transferred to a fresh microcentrifuge tube, mixed with 590 µl of 80% ethanol and transferred to RNeasy spin column. The column was then centrifuged at 12,000 x g for 15 seconds and the flow-through discarded, then 350µl of RW1 buffer was added to the column filter and the column was centrifuged again for 15 seconds at 12,000 x g.

To streamline the process further, 10µl of RNase-free DNase diluted in 70µl of RDD buffer was added to the column filter and incubated at room temperature to facilitate on-column DNase digestion, removing any genomic DNA contamination. After 15 minutes incubation, 350µl of RW1 buffer was added to the column filter and centrifuged for 15 seconds at 12,000 x g, with two

subsequent column washes using RPE buffer. Finally, the RNA was eluted in 50 µl of RNase-free water.

To ensure complete removal of genomic DNA, the RNA samples were subjected to a second decontamination using an in-solution DNase treatment using DNA-Free (Ambion, Life Technologies, USA) as per manufacturer's instructions.

8.5.3 Importance of stressed growth conditions for RNA isolation methodology

Significantly, Venkataraman et al. (2013) also explored the isolation of mycobacterial RNA under a range of sub-optimal growth conditions, including pH stress, oxidative stress, medium starvation, antibiotic stress and latency. The utilisation of the modified RNeasy kit isolation method resulted in successful isolation of RNA under all conditions, with QA testing indicating isolation of highly pure and intact RNA using a Bioanalyzer 2100, akin to QA methods used in this chapter. RIN values reliably scored between 9.2 and 9.9, further highlighting that this modified methodology would likely be suitable for isolation of *R. equi* RNA under the desired media starvation conditions, without degradation

Furthermore, this research group also investigated downstream RNA methods including rRNA depletion for mRNA enrichment (Venkataraman et al., 2013), this is of particular importance, given the objective of expression profiling by RNA-seq. Their isolated rRNA from the modified RNeasy method, was depleted *via* subtractive hybridisation, with complete rRNA-depletion observed as well as acquisition of high-quality mRNA suitable for downstream RNA-seq transcript analysis for expression profiling.

8.6 Concluding remarks

Although this project did not advance as quickly as anticipated, it was determined that the utilisation of appropriate quality assurance methods would be paramount for impact when publishing the results, and to prevent the depletion of the valuable and limited resources. The methodology in this chapter for RNA isolation was the use of a RiboPure™-Bacteria Kit that used a Trizol based reagent (RNA-WIZ), although many attempts were made to improve cell lysis and prevention of degradation, the RNA integrity was comparable with other Trizol isolation methods described in the literature.

As the RNA integrity did not meet the assigned threshold, the project was not able to advance; with focus altered on identification of a refined method of isolation of intact RNA from mycolic acid containing bacteria. The identification of a refined methodology in the literature (Venkataraman et al., 2013) should prove useful for advancement of the project. The decision was made to not attempt the use of a hot phenol-based extraction given the associated risks, however going forward the modified RNeasy kit protocol should be explored for isolation of pure, intact *R. equi* rRNA, this in turn will facilitate procurement of mRNA for transcript analysis in gene expression profiling using RNA-seq.

Although not discussed at this stage, it will be important to maintain a high degree of quality assurance throughout the project stages including integrity checks after mRNA enrichment, and quality assessment of adapter-ligated DNA libraries as well as quality control during bioinformatic analysis, regarding read alignments and transcript assemblies, consideration of these variables, and maintenance or associated thresholds will ensure maximum impact can be achieved and will coordinate with other projects for the multi-omics approach to characterisation of the iron-regulated homeostatic mechanisms with the veterinary pathogen *R. equi*.

9 - General discussion

9.1 The importance of a Multi-Omics experimental approach

Multi-omics studies are becoming increasingly popular as researchers strive to achieve the *magnum opus* of Systems Biology: the ability to produce a comprehensive description of the complex interactions between biomolecules (Breitling, 2010). Datasets from genomic, transcriptomic, proteomic and metabolomic profiling experiments are integrated to generate a more holistic understanding of the relationships between cells, organisms, and communities. In the context of infection, these datasets specifically relate to pathogen growth, environmental adaptation, development and survival through utilisation of virulence factors, and ultimately, disease progression (Cho et al., 2006; Pinu et al., 2019; Yugi et al., 2016).

The competition of the post-genomic era has facilitated significant technological advancements in omics-related analytical techniques such as next-generation sequencing and high resolution tandem mass spectrometric analysis. Development of these technologies provided higher throughput data at significantly lower cost in a fraction of the time (Park and Kim, 2016). That, in turn, has facilitated larger-scale omics projects to become widely-accessible to the research community (Pinu et al., 2019), and naturally the instigation of multi-omics project are becoming an increasingly desirable practice to answer complex biological networking questions.

The challenges associated with omics integration is a pressing issue as, unsurprisingly, many of the techniques require an appropriate experimental design and specific analytical equipment; and as such are often not adequate to perform comparative studies as a result; successful implementation of more than two omics datasets is exceptionally rare (Biswapriya et al., 2019) due to the inherent differences associated with the datasets.

9.1.1 Designing Experiments Suitable for Multi-Omics Integration

It is evident at this stage that to facilitate integration of multi-omics data careful consideration must be taken for experimental design firstly for the individual -omics experiment, but also how the data integration will occur. To enable this integration step, it is imperative that the experimental design considers selection process as well as preparation, and storage of the biological samples suitable for the wide variety of analytical methods (Pinu et al., 2019). Ideally the results should be generated by

analysis of the same sample group to facilitate direct comparison under identical conditions, principally because omics analysis is considered a cellular snapshot of the physiological status of the cell (Peng et al., 2015). Moreover, sample cohort size is directly relevant to the expected hypothesis and should be incorporated into the experimental design; for example, investigation of a vague disease phenotype likely requires a large sample cohort given the small effect size, comparative to a strong disease phenotype that is likely to require a much smaller cohort.

It appears at this stage, the prevalent challenges faced when designing an experiment suitable for a multi-omics approach relate to the biological sample type. Interestingly, the review by Pinu et al. (2019), identified that metabolomics studies are the most versatile, offering the most compliance with other omics requirements, including genomic, transcriptomic and proteomic analyses. For example, metabolomic studies frequently require sample storage at $-80\text{ }^{\circ}\text{C}$, this is typically also beneficial for other studies given that storage under these conditions maintains both protein and RNA integrity for proteomics and transcriptomics respectively (Day and Stacey, 2008; Haas et al., 2017).

9.1.2 Methods of Multi-omics Data Integration

Despite the significant technological advancements in individual -omics fields and the particularly important growth seen in readily-accessible omics databases, the identification of novel biological pathways appears to be bottlenecked by the data integration steps of the multiple omics approaches (Broadhurst and Kell, 2006; Haas et al., 2017). This has been attributed to the input and data interpretation steps required by specialist researchers (Pinu et al., 2019), however providing the data integrity achieves the desired threshold, a selection of integration methods can be utilised.

Firstly, individual omics can be independently analysed, with key features networked to stipulate a post-analysis data integration methodology. This facilitates the production of a biomolecular pathway based upon interactions with the aforementioned key features. This method of integration has been successful for a range of investigations including a multi-omics approach to characterise the microbial biota of sewage sludge and their metabolism in anaerobic digesters under operational shocks (Beale et al., 2016), as well as a multi-omics approach to assess the surface water quality of an Australian river system (Beale et al., 2017).

Alternatively, omics data can be integrated prior to analysis and interpretation to identify statistical algorithm-derived connections between datasets using without reliance on human interpretation (Huang et al., 2017; Kuo et al., 2013; Pinu et al., 2019). This methodology is typically used in clinical studies where well-annotated databases are available (Rantalainen et al., 2006). Similarly, data integration can also be performed using a systems biology modelling approach, this methodology is most appropriate when investigating a well-defined biological network, with previous comprehensive omics data available and the aim of comparing novel experimental data to prediction models (Manninen et al., 2018; Wierling et al., 2007).

9.2 Initial Project Aims

As described in the introduction, the main research focus was to improve the understanding of the complex iron regulatory network within the veterinary pathogen *Rhodococcus equi*. The principal aims of the research was focused on iron homeostasis and uptake mechanisms.

The initial scope of the project was generated by utilisation of a genomic survey for metal regulation with *R. equi*. Any metalloregulatory genes of interest that may be responsible for transcription of essential virulence factors, were to be explored by recombinant gene expression studies and production of in-frame deletion mutations to address a hypothesis that complete deregulation might generate a phenotypic change and stimulate hyper-production of siderophores.

R. equi is predicted to synthesise two siderophores under iron starvation; a catecholate-hydroxamate mixed type-siderophore termed rhequibactin (Miranda-CasoLuengo et al., 2008) and a hydroxamate siderophore termed rhequichelin (Miranda-CasoLuengo et al., 2012), however they currently remain uncharacterised, likely due to low yielding synthesis complicating isolation of the siderophores. Furthermore, very little research has been presented for cellular transport mechanisms for essential nutrients such as iron in *R. equi*. Frequently, there is no consideration for the presence of the barrier to materials transfer represented by the mycomembrane, the structure that is often attributed to resistance to antimicrobials and pathogenicity of other better studied related organisms. Often the focus for transport studies are the mechanisms associated with translocation across the cytoplasmic membrane. As it currently stands, both regulation and uptake of these siderophores are uncharacterised.

To address these issues another project was instigated where *R. equi* was grown under iron-limiting conditions, to facilitate 1) identification of the *R. equi* siderophores, and ideally characterise them using a metabolomics approach. 2) identify the means associated with ferri-siderophores uptake across the mycomembrane using a proteomic approach. 3) Map global gene expression of *R. equi* under the given conditions, to infer a network of interactions associated with iron homeostasis.

9.3 Genomic survey of iron homeostasis with *Rhodococcus equi* 103S

The aim of this chapter was to audit the requirements for siderophore-mediated iron-acquisition by *R. equi* 103S, with this survey aiding transcriptome, metabolome and proteome data analysis.

Initially, functional candidates for regulation, biosynthesis and export, and ferri-siderophore uptake were assembled using a text-based genome mining approach. The bioinformatic investigation revealed that the *R. equi* genomic contains four putative genes predicted to encode putative iron regulator proteins: *ideR* (REQ_20130), *dtxR* (REQ_19260), *furA* (REQ_04740), *furB* (REQ_29130). While very little is known regarding *R. equi* iron regulation, previous literature highlights that iron regulatory proteins are responsible for controlling siderophore biosynthesis and transport, however it was interesting to understand the biological significance of four analogous regulatory components. In order to determine whether this apparent redundancy in iron-dependent transcriptional regulators was a conserved feature of genus *Rhodococcus* or an unusual feature of *R. equi* 103S, each candidate gene product sequence was queried using a BLASTP approach with outputs being restricted to *Rhodococcus*. In summary, 27 genomes were investigated, of which 25 appear to possess homologues of all four putative iron-dependent transcriptional regulator proteins, with none possessing less than three of them.

It is therefore unlikely that all of these proteins co-repress their regulons with ferrous iron, but rather exhibit interactions with other metals, and each protein is likely to possess a discrete independent physiological role. Furthermore, possession of two each of the Fur and IdeR/DtxR families, appears to be consistent with other related genera, of which the best characterised organism is *M. tuberculosis* H37Rv, which also has a complement of four such regulators (Rodriguez and Smith, 2003).

Each of these four regulators from *R. equi*, were targeted for molecular cloning and expression techniques, in order to facilitate regulator-binding site analysis, discussed in Chapter 4. Additionally,

each of the regulator genes were targeted to produce unmarked in-frame gene deletions to establish their influence on rhodococcal physiology independently, as discussed in Chapter 5.

9.3.1 Iron sequestration

As previously discussed, *R. equi* is predicted to synthesise two siderophores under iron starvation; rhequibactin (Miranda-CasoLuengo et al., 2008) and rhequichelin (Miranda-CasoLuengo et al., 2012), both of which are currently uncharacterised. An early research study that identified that *R. equi* does not produce sufficient siderophore under iron limitation to be detected in a Chromazurol S assay (Fiss and Brooks, 1991), yet subsequent bioinformatic analysis has undoubtedly displayed that *R. equi* has the capacity to produce multiple siderophoric compounds. However, production of the strain *R. equi* α 5 (Miranda-CasoLuengo et al., 2005), containing a transposome insertion in *iupABC*, induced a siderophore-like effect under iron sufficient conditions predicted to be due to the formation of an Fe³⁺-siderophore complex given the red pigmentation of the medium supernatant.

Therefore, bioinformatic analysis of putative biosynthetic gene clusters was performed with the aim of producing putative structures and associated masses that can be used in a targeted metabolomic approach for siderophore characterisation.

Firstly, to explore the strategies utilised by *R. equi* 103S to sequester iron from the extracellular environment the genome was investigated using a text-based genome mining approach similar to the regulator search. It is well established that siderophores can be synthesized by both non-ribosomal peptide synthetase (NRPS)-dependent and NRPS-independent pathways therefore functional candidates were identified by related genome annotation text.

In silico secondary metabolite structure predictions were produced based on nonribosomal peptide biosynthetic pathways retrieved from *R. equi* 103S DNA sequence data, which were compared with the output of the bioinformatic tool antiSMASH. The analysis identified NRPS clusters similar to Erythrochelin, and Heterobactin (57 % and 54 % similar respectively), providing evidence to support NRPS synthesis of both rhequichelin and rhequibactin. The remaining clusters appeared to be responsible for antibiotic synthesis.

Analysis of the rhequibactin biosynthetic gene cluster *via* the bioinformatic tool antiSMASH, in

combination with the previous literature, highlights that rhequibactin utilises a 2,3-hydroxybenzoic acid moiety for iron coordination, given the proximity of the upstream 2,3-DHB biosynthetic machinery, and previous experimental verification of catecholates moieties (Miranda-CasoLuengo et al., 2008). Utilising a stable version of antiSMASH v4.2.1 facilitated a consensus prediction method using the SANDPUMA ensemble (Blin et al., 2017), The assembly of the hypothetical structure for rhequibactin was performed with direction from Dr. D. Tetard, following the orthodox linear peptide assembly, producing a catecholate-hydroxamate mixed ligand model of the siderophore rhequibactin using the peptide monomers; 2,3-dihydroxybenzoic acid, lysine, ornithine (x2), and a cyclised δ -N-hydroxy-ornithine.

Similarly, the antiSMASH analysis for rhequichelin biosynthetic gene cluster corroborates the predicted domains identified in the literature (Miranda-CasoLuengo et al., 2012), hypothesising that the hydroxamate siderophore rhequichelin is comprised of the peptide monomers: N^5 -formyl- N^5 -hydroxyornithine -serine - N^5 -hydroxyornithine - N^5 -acyl- N^5 -hydroxyornithine. Due to the prediction of a N^5 -acyl- N^5 -hydroxyornithine peptide monomer, elucidation of an accurate mass clearly becomes inherently more challenging, therefore non-exclusive selection of potential acylation groups and their predicted masses were compiled.

Although molecular characterisation of the *R. equi* siderophores remains unsolved, the predictions made within this chapter assisted production of a refined candidate list for the identification of rhequichelin and rhequibactin.

9.4 Predicted Metal-binding capacities for the putative metal-dependent regulators of *Rhodococcus equi* 103S

At the outset, the aims relating to the Fur and DtxR superfamily members of *R. equi* 103S identified in Chapter 3 were to characterise these proteins by molecular cloning to support *in vitro* analysis of metal and DNA binding to establish their respective regulons.

Initially, the research project started by designing constructs for the four metal regulatory genes *ideR*, *dtxR*, *furA* and *furB* for heterologous gene expression in *E. coli* to facilitate physico-chemical binding analysis using ICP-OES.

Protein purification was attempted using heparin affinity chromatography to retain the predicted DNA-binding proteins, rather than the commonly used poly-histidine tag approach so as to not impair or influence the resulting metal-binding analysis. Unfortunately, this methodology was unsuccessful therefore the project methodology was amended to utilise a GST fusion tag to generate a readily purifiable fusion proteins. This was successfully achieved for the protein IdeR-GST; however, heterologous gene expression was unsuccessful for the other regulator fusions. While measurement of the IdeR protein metal-binding capacity would potentially infer a preference for a range of divalent cations, without the other three metalloregulatory proteins, the scope of the project unfortunately becomes rather limited, as ideally the metal-binding capacities would be performed as a comparative exercise.

Consequently, the project aims were compromised to infer metallo-preference for the binding sites of interest based upon homology studies with other closely related structures. This was performed by use of a *in silico* homology-based protein scaffold; significantly, all four putative metalloregulatory proteins in *R. equi* 103S were successfully modelled to the proteins from the closely related *M. tuberculosis* with a high degree of sequence similarity. Furthermore, each of the four proteins are likely responsible for discrete regulatory functions as indicated by their coordination preference for different transition metals.

Firstly, the *in-silico* protein model for *R. equi* IdeR produced a significantly high sequence template homology of 78 % to 1FX7-C(iron-dependent regulator (IdeR) from *M. tuberculosis*; with local sequence alignments identifying a clear conservation of the metal-coordinating residues, it is likely that IdeR preferentially coordinates Fe(II) ions based on the identified octahedral coordination geometry of IdeR from *M. tuberculosis*. This conclusion further corroborates the findings of the genomic survey and the iron-box sequences identified in the operator sequence of iron-responsive genes including siderophore biosynthesis, transport and iron storage genes, which are also prevalent in mycobacterial species.

The *in-silico* protein model for *R. equi* DtxR also produced a significantly high sequence template homology to the crystal structures 5ZR4-B and 5ZR6 (manganese regulator – MntR, *M. tuberculosis*, *apo* and *holo* forms); with one-to-one threading of 5ZR6 yielded a 73 % identity match. Upon

performing local sequence alignments *R. equi* DtxR clearly demonstrates a conservation of the metal-coordinating residues, of which there is a pentavalent bi-nuclear metal ion cluster that has been shown to be a specific coordination cluster associated with Manganese(II) and Magnesium in biological systems (Liu et al., 2007). The combination of this evidence, and the high overall sequence and structure conservation strongly infers that *R. equi* DtxR functions as a manganese regulatory protein homologous to MntR in *M. tuberculosis*.

The *in-silico* protein model for *R. equi* FurA was generated using the crystal structure 2FE3-B (peroxide operon regulator (PerR) from *Bacillus subtilis*); the percentage identity between the sequence and template was only 24 %, feasibly indicating a core structure. Therefore, a local sequence alignment was performed to compare *R. equi* FurA with both FurA from *M. tuberculosis* and PerR in *B. subtilis*. Interestingly, the metal-coordinating sites of FurA from *R. equi* also appear to be conserved these proteins; in all cases the proteins appear to function as repressors of catalase/peroxidase, it is therefore reasonable to hypothesise that *R. equi* FurA could represent a metal-dependent peroxide sensor, in addition to the supporting evidence of gene clustering between *furA* to the catalase gene *cat*.

Finally, the *in-silico* protein model for *R. equi* FurB has significant sequence homology to the crystal structure 2O03-A (zinc uptake regulator (Zur/FurB from *M. tuberculosis*); the percentage identity between the query sequence and template was 69 % suggesting high model accuracy. As for all of the other metalloregulatory proteins of interest, a local alignment of sequences between *R. equi* FurB and *M. tuberculosis* FurB identified a clear conservation of the metal-coordinating residues; the coordination ligands were arranged in a tetragonal geometric pattern characteristically favourable to the binding of zinc ions.

As previously discussed, the designation of four analogous regulatory genes would be biologically redundant and would likely contribute to the bacterium being outcompeted based upon unnecessary energy consumption. The original gene annotations listed were produced as an autonomous process, however the analysis performed here appears to show that the veterinary pathogen *R. equi* utilises a range of different divalent metal regulatory genes to maintain adequate transition metal homeostasis; therefore the combination of evidence available strongly suggests *R. equi* utilises *ideR* as an iron

regulator; *furB* as a zinc regulator, *dtxR* as a manganese regulator and *furA* as an oxidative stress response regulator.

9.5 Generation and *in vitro* characterisation of *Rhodococcus equi* 103S mutants for four putative metal regulatory genes: *ideR*, *dtxR*, *furA* and *furB*.

Building on the results of the genomic survey in chapter 3, and the predicted metal-binding capacities in chapter 4, the aim of this project was to further characterise the Fur and DtxR superfamily members of *R. equi* 103S by investigation of the regulatory influence *via* generation of unmarked in-frame gene deletions using the pSelAct suicide vector developed by van der Geize *et al.* (2008).

The mutant gene knockout constructs were successfully generated utilising a Gibson assembly style methodology and replicated in *E. coli* without issue. The recombinant plasmids were independently incorporated into *R. equi* chromosomal DNA, using apramycin resistance as positive selection screen for single-crossover homologous recombination event.

Efficient selection for a double-crossover homologous recombination event leading to excision of the plasmid from the chromosome proved particularly challenging despite following the previously optimisation described by van der Geize. Subsequently, a range of method optimisation attempts were made, and ultimately a four-fold increase of 5-fluorocytosine concentration facilitated successful counter selection.

The counter-selected second homologous recombination event can resolve in two ways; the reversion to a wild-type parental genotype or the generation of the unmarked in-frame deletion. Rigorous testing with a diagnostic colony PCR with the integrated pSelAct transformant produced by the first homologous recombination event demonstrated that assay design could be biased by extension time that could entirely obscure the signal from the parental gene. In the event of recovery of such a transformant in counter-selection screening, the assay must be capable of identifying both possible amplicons. These diagnostic assays were optimised to ensure that the larger amplicon was reliably detected. Significantly, secondary recombination events identifying the truncated gene mutations were successfully achieved for all four targets.

Subsequent phenotype characterisation identified a noteworthy change for *R. equi*- $\Delta furA$, which

exhibited an impaired growth rate and altered colony pigmentation. This gene shares a locus with another encoding a putative catalase; its gene product was markedly upregulated in the mutant strain, which was coincidentally hyper-resistant to hydrogen peroxide, which is consistent with its annotation as a catalase and its transcriptional repression by FurA. Surprisingly, the other mutants did not produce any identifiable phenotypic changes, and analysis of siderophore production using the Chromazurol S chelation assay indicated that no marked deregulation of siderophore synthesis occurred for any of the mutants.

Although it is surprising that functional de-repression of the IdeR regulon by gene deletion did not result in maximal production of siderophores akin to the exaggerated *in vitro* iron starvation responses characteristic of many microorganisms, this outcome suggests that siderophore biosynthesis in *R. equi* is heavily repressed by another regulatory mechanism. Three possibilities are evident: firstly, IdeR is the primary repressor of siderophore biosynthesis and export in *R. equi* but the mutagenesis process has generated another mutation(s) that has compromised siderophore biosynthesis; this will be resolved by whole genome sequencing of the mutant and also in recovery of another mutant from the integrated pSelAct transformant for comparison. Secondly, the *in-silico* analysis of repressor metal specificity described herein might be faulted and IdeR may not be an orthologue of *M. tuberculosis* IdeR. However, this seems unlikely due to the consistent pattern of metal ligand conservation across all of the four regulators when compared to their closest characterised homologues. Additionally, mutants in the other regulator genes did not result in de-repression of the siderophores so a case of mistaken identity seems unlikely. Arguments relating to additional, unintended mutations are equally as valid for this second scenario; whole genome sequencing will also address this possibility. Thirdly, another regulatory mechanism may influence the production of siderophores in *R. equi* and the other rhodococci that present a CAS-negative phenotype characterised by a muted response to iron limitation. The exaggerated responses exhibited by many bacteria may be counter-productive in nature. In an iron-limited environment subject to competition for iron within a community of microorganisms, the use of other nutrient resources on extensive siderophore biosynthesis may place the producing organism at a disadvantage if iron is not returned by the investment. The likelihood of success will be influenced by the absolute availability of the metal in that environment and the identity and number of the producing organisms'

competitors. Given the common ability of bacteria to engage in ferri-siderophore piracy, secreted siderophores ultimately become a community resource. The preservation of resources until a point at which return of iron is significant may provide an ecological advantage to *R. equi* and other CAS-negative rhodococci. This scenario is partly reminiscent of observations of the hyper-repressed production of (carboxy)mycobactins and the inability to culture new isolates of *Mycobacterium avium* subsp. paratuberculosis and *Mycobacterium avium*, which can be relieved by addition of ferri-mycobactin. Repeated culture in lab media led to the loss of this mycobactin-dependence and promoted siderophore biosynthesis suggesting a phenotypic / regulatory mechanism rather than a genetic lesion in siderophore biosynthesis was responsible (Barclay and Ratledge, 1983).

If a conserved regulatory mechanism were to override IdeR in some of the Corynebacteriaceae, it seems plausible that this might be addressed and identified by deploying random mutagenesis in appropriate mycobacteria and *R. equi* to relieve the inhibition and to identify affected sequences by mapping genetic changes by whole genome sequencing.

9.6 An analysis of the *Rhodococcus equi* 103S metabolome under iron limitation: An insight into siderophore mediated iron metabolism

The precursor for the multi-omics approach was to establish bacterial growth conditions that facilitate a biological response to iron starvation, that can be compared to iron-replete bacterial growth conditions. It was hypothesised that optimisation of these media conditions would induce a physiological response by secretion of siderophores, as predicted in the bioinformatic study performed earlier that hypothesised that *R. equi* has the capability to synthesis two distinct siderophores. Ultimately, this experimental design facilitated the multi-omics approach for analysis of secreted metabolites (metabolomics), a comparative analysis of cell-wall associated iron-regulated proteins (proteomics) and identification of global gene expression under iron starvation (transcriptomics).

The experimental conditions for iron limitation utilised lactate minimal media as previously described (Miranda-CasoLuengo et al., 2012; Miranda-CasoLuengo et al., 2005; Miranda-CasoLuengo et al., 2008), with further treatment conditions to sequester trace elements. The initial

experiment was performed using the modified conditions comparative to iron supplemented conditions, in an attempt to replicate the observed chromophore identified by Miranda-CasoLuengo et al. (2008). After a period of 120 hours a noticeable red pigment developed in the iron-limited culture, with the iron supplemented culture remaining a creamy-white pigment. The development of this pigment was used as an indication of an appropriate harvest time, given that *R. equi* does not appear to produce sufficient siderophore to be detected via the Chromazurol S assay.

Cellular material was separated by centrifugation, with the supernatant used for metabolomic analysis and cell pellets frozen for subsequent transcriptomic and proteomics analysis. In an attempt to detect ferri-siderophore complexes, aliquots of the supernatants were supplemented with ferric chloride, this however resulted in a precipitation event for the iron-limited supernatant and also transfer of the pigment from the supernatant into the precipitant. The experimental parameters were adjusted as a result, to incorporate analysis of the iron-supplemented supernatants to profile both the metabolome differences and the effect of iron supplementation.

The aforementioned samples were subject to an untargeted metabolite profiling strategy using both reversed phased methodology with a C18 column and a HILIC methodology using an ethylene bridge hybrid amide analytical column coupled with a heated electrospray ionisation - mass spectrometry, with observable differences between conditions. The positive ionisation mode reverse phased chromatography was further analysed with metabolomic profiles compared using chemometric analysis, confirming that 54.6 % of variance between samples was produced by biological differences occurring during growth in the different medium conditions, with 22 % of variance derived from the addition of ferric chloride to the bacterial supernatants and likely the changes caused by the precipitation. Siderophore candidates were selected based upon their comparative abundance in the conditions, firstly using a by-eye approach, that was reinforced by a ranking of statistical significance using a significance analysis of microarray method, with key candidates M0872 and M1067 having a corresponding mass of 435.13361 [M+H]⁺ and 463.13996 [M+H]⁺ respectively.

These identified candidates have potential to be used as biomarkers when utilising a targeted approach to identifying and characterising the *R. equi* siderophores. The metabolite M0872 is of particular interest as the loss of peak intensity when supplemented with ferric chloride appears to

correlate with the observed precipitation event, this also appears reminiscent of the heterobactins that were shown to have very low aqueous solubility (Carrano et al., 2001). Given that this metabolite is likely the chromophore, it should be possible to track the metabolite through various purification stages if adequately concentrated.

To facilitate this targeted approach, the iron-limited conditions identified earlier were replicated and scaled up accordingly, then harvested after the development of the red pigment. Cellular material was separated by centrifugation as before, with the bacterial supernatant concentrated via rotary evaporation, that tested positive using the Chromazurol S assay, further confirming the hypothesis that *R. equi* does in fact produce siderophores, but not at sufficient levels normally under iron limitation to be detected in a Chromazurol S assay.

The Chromoazurol S assay- positive concentrated bacterial supernatant was purified by passage through an Amberlite XAD-2 adsorption resin packed chromatography column; during the loading of the column the resin appeared to retain the pigment, that was transferred to the elution buffer correlating with a sharp observable UV peak. The fraction of interest suggested elution of target compounds including the apparent chromophore, which further tested positive using the Chromoazurol S assay.

The LC-MS analysis of the purified fraction of interest identifies one predominant peak that had a column retention time of ~2.8 minutes, and a corresponding mass of 435.1346 [M+H]⁺ therefore it is likely that this metabolite previously identified as the candidate M0872 is a siderophore . Further exploration of this candidate by MS² fragmentation pattern analysis, was unsuccessful but provides excellent scope for future characterisation studies.

In addition to this, a candidate list for alternative siderophoric metabolites was produced based upon a collection of compounds that were previously been identified as upregulated under iron limited conditions. While the attempts to remove the candidate M0872, that previously dominated the mass spectra were largely successful with the addition of an ferric chloride solution to precipitate the metabolite, it is apparent that not all of the compound was removed, therefore it would be difficult at this stage to assert that the observed siderophore activity was related to these other candidates without further chromatographic separation.

9.7 A comparative proteomic analysis of the composition of the rhodococcal cell wall under iron limitation

The aim of this project was to explore the mechanisms used to facilitate iron transport across the cell envelope in *R. equi*. This was to be achieved by firstly isolating cell envelope-associated proteins and subsequent identification of any potential outer membrane receptors that may facilitate transport of iron across the mycolic acid-containing cell wall. Similarly to the other omics research projects listed in this thesis, this project utilised the cellular material produced from the iron limiting conditions outlined in the metabolomics study; the subsequent proteomic analysis of the rhodococcal cell wall was performed to develop the literature associated with transport mechanisms within acid-fast bacteria especially under sub-optimal growth conditions such as nutritional stress.

As for the other omics projects, a large proportion of this project was dedicated to experimental design, in this instance identification of the most appropriate fractionation methodology for isolation of the rhodococcal mycolate-containing cell wall region to allow for further characterisation of predicted mycolate-layer transport mechanisms. Traditionally, isolation of the mycolate-containing cell envelope region was performed using *in vitro* density cellular fractionation methods and solubilisation of hydrophobic proteins using specific non-ionic detergents such as Triton X-114. However, development of bioinformatics prediction tools over the last decade has facilitated a process of performing a consensus prediction of membrane protein topology signal peptides without fractionation, thereby reducing the potential for sample loss through experimental procedures.

In this research a preliminary study was performed to identify a strategy that would provide the greatest scope for identification of potential iron transporters through a combined *in-vitro* and *in silico* comparative approach. The preliminary results indicated that while usage of *in vitro* fractionation methods may provide an enhanced view of retained hydrophobic proteins there is significant risk of false negative results due to transfer and recovery losses during sample processing. While this *in vitro* methodology may identify novel mycomembrane protein architecture, it is highly unlikely given none were identified during the preliminary stages.

The large-scale analysis was performed using an unfractionated protein isolation methodology on

the previously optimised growth conditions to promote iron limitation in combination with the *in-silico* bioinformatic pipeline to predict subcellular localisation based upon transmembrane topology and detection of signal peptides.

Proteomic analysis was performed using a GeLC-MS/MS approach, with fraction recombination prior to analysis to facilitate both technical and biological replicates. Proteome data analysis was implemented using MASCOT to generate un-interpreted MS/MS data; in tandem with the Progenesis LC-MS software to identify statistically significant differences in protein expression between iron-deplete and iron-replete cultures, of which 445 were identified to meet the desired threshold to be statistically significant and upregulated under iron limitation. After *in silico* processing the protein complexity was reduced by 86 %. Although cytoplasmic contamination was removed, there still remained membrane-associated components such as iron ABC transporters and while not relevant for identification of mycolate-cell wall associated mechanisms, their presence strengthens the judgement that iron limitation was achieved through the experimental parameters.

The lack of either a rhodococcal carboxymycobactin-like or mycobactin-like siderophore infers that the importation methods utilised by *R. equi* may differ substantially to *M. tuberculosis*. Interestingly, *M. tuberculosis* arbitrates transfer from the iron-bound carboxymycobactin to the mycobactin using an iron-regulated cell wall-associated protein HupB (Yeruva et al., 2006); while a BLASTP search identified an *R. equi* homologue, this was not found to be upregulated under the given conditions, further suggesting that *R. equi* utilises an alternative approach to ferri-siderophore uptake.

From the dataset no obvious candidates for mycolate-associated iron transport were identified, therefore it is likely that *R. equi* utilises a simple porin strategy for accumulation of ferri-siderophore complexes in to the pseudoperiplasmic space, from which a substrate binding lipoprotein is likely to transfer the siderophore complex through an ABC transporter system.

9.8 Method optimisation for RNA isolation for use in RNA-Seq analysis of *Rhodococcus equi* 103S global gene expression under iron starvation

The aim of this project was to reveal the genetic response to iron starvation, the elucidation of gene transcription under these conditions will corroborate with other omics analysis methods in characterising the global iron regulatory mechanisms utilised in the veterinary pathogen *Rhodococcus equi* 103S. To facilitate this integrated omics analysis, the experimental parameters were designed to utilise the same cultures as used in the metabolomics study.

Furthermore, due to funding and time limitations within this project, the experimental design was constructed to achieve a significant transcriptome coverage without exceedingly high amounts of transcript reads and the associated costs. The transcriptomic sequencing was to be performed on RNA harvested from: Wildtype *R. equi* 103S both under iron limitation and iron sufficiency (as performed in the other omics projects) and also *R.equi-ΔideR* under iron sufficiency to identify any potential dysregulation to iron-responsive genes including siderophore biosynthesis, transport and iron storage genes. Furthermore, to ensure adequate coverage, and statistical robustness the experimental design incorporated three biological replicates and two technical replicates with 2.5 million reads per sample.

The experimental isolation of *R. equi* total RNA performed as described in the manufacturer's protocol, however upon performing the quality assurance, the RNA integrity was significantly below the required threshold. Unfortunately, multiple attempts to improve cell lysis and prevention of RNA degradation were made, but the RNA integrity was still below the desired threshold.

RNA extraction from mycolic acid-containing bacteria has been well documented over the last 25 years, and isolation of pure, intact RNA has consistently proved challenging (Jahn et al., 2008) and interestingly the integrity results in this research are comparable with other Trizol isolation methods described in the literature (Akhtar et al., 2011; Payton and Pinter, 1999; Venkataraman et al., 2013).

Although time limitations impacted further progress, an ongoing literature survey identified that an optimised method for RNA isolation of mycolic acid-containing bacteria was established by Venkataraman et al. (2013) utilising a modified RNeasy kit protocol that was successful under a

range of sub-optimal growth conditions, including pH stress, oxidative stress, medium starvation and antibiotic stress. This methodology should be explored for isolation of pure, intact *R. equi* RNA, evidenced principally by the integrity of the rRNA, for transcript analysis in gene expression profiling using RNA-seq.

9.9 Data integration of the multi-omics study

Overall, the main objective for the research project was to start to map the complex iron regulatory network within *Rhodococcus equi*. Considering the availability of important analytical techniques such as next-generation sequencing and UPLC-MS/MS analysis, the use of an integrated multi-omics study appeared to be excellent choice to understand the complex biological network questions relating to how *R. equi* both regulates iron homeostasis and the methods employed to sequester iron in response to starvation.

Fundamentally, three main stages were considered essential to the success of the project: Firstly, the genomic survey of iron homeostasis within *R. equi* was critical as it provided precedence for understanding the transcriptional control of iron-responsive genes, as well as identification of the biosynthetic machinery responsible for the production of the putative *R. equi* siderophore with potential candidate masses generated. Secondly, with the genomic survey in-mind, the experimental parameters could be established, for which the foundation of the multi-omics approach was to establish medium conditions that promoted a biological response to iron starvation by secretion of low molecular weight chelatory peptides – siderophores. Finally, the decision was made to perform the post-analysis data integration methodology, this facilitated individual omics experiments to be designed, performed and analysed independently. With the assumed complexity associated with the iron regulatory network, the individual omics projects would each be responsible for analysis of different aspects of *R. equi* iron homeostatic mechanisms. These were a comparative analysis of secreted metabolites using metabolomics, a comparative proteomic investigation of cell-wall associated iron-regulatory mechanisms and identification of global gene expression under iron starvation using RNA-seq transcriptomics.

The outcomes of each individual omics study portrayed varying degrees of success; however, it is clear that the use of a post-analysis data integration methodology was effective, with clear predictions

made regarding the mechanisms *R. equi* utilises to regulate metal homeostasis. Figure 9-1- directly assigns putative identities for the transcriptional regulators identified during the genomic survey of iron homeostasis, collaborating with the *in-silico* homology-based models produced for each, from which the putatively iron transcriptional regulators explored further. Interestingly, the gene *furA* was predicted to be an essential component in redox sensing which was confirmed by the in-frame deletion mutant produced, and subsequent proteomic analysis and susceptibility testing with hydrogen peroxide. The other significant gene of interest *ideR*, was predicted to regulate iron-responsive genes including siderophore biosynthesis, transport and iron storage genes, based upon the presence of a specific iron-box sequence in the operator sequence of these genes. It was expected that under iron starvation, the iron responsive genes would facilitate de-repression the aforementioned genes. The metabolomics analysis does distinguish discrete differences between the biological conditions, and further facilitated putative masses for the *R. equi* siderophores. Furthermore, the proteomic analysis also indicated the utilisation of an iron starvation response, however identification of a putative mechanism for siderophore transport across the mycolate-containing cell wall was not successful, therefore one potential hypothesis for this may be the use of a channel forming porin to facilitate accumulation of siderophores at the surface of the cytoplasmic membrane.

Ideally, the transcriptomics project would have further informed the regulatory network of iron homeostasis within *R. equi* by mapping the global gene expression levels, however it remains essential that a high degree of quality assurance is maintained for this work to achieve maximum impact. On this basis, it was excluded from the multi-omics integration at this stage but will provide excellent grounds for future study.

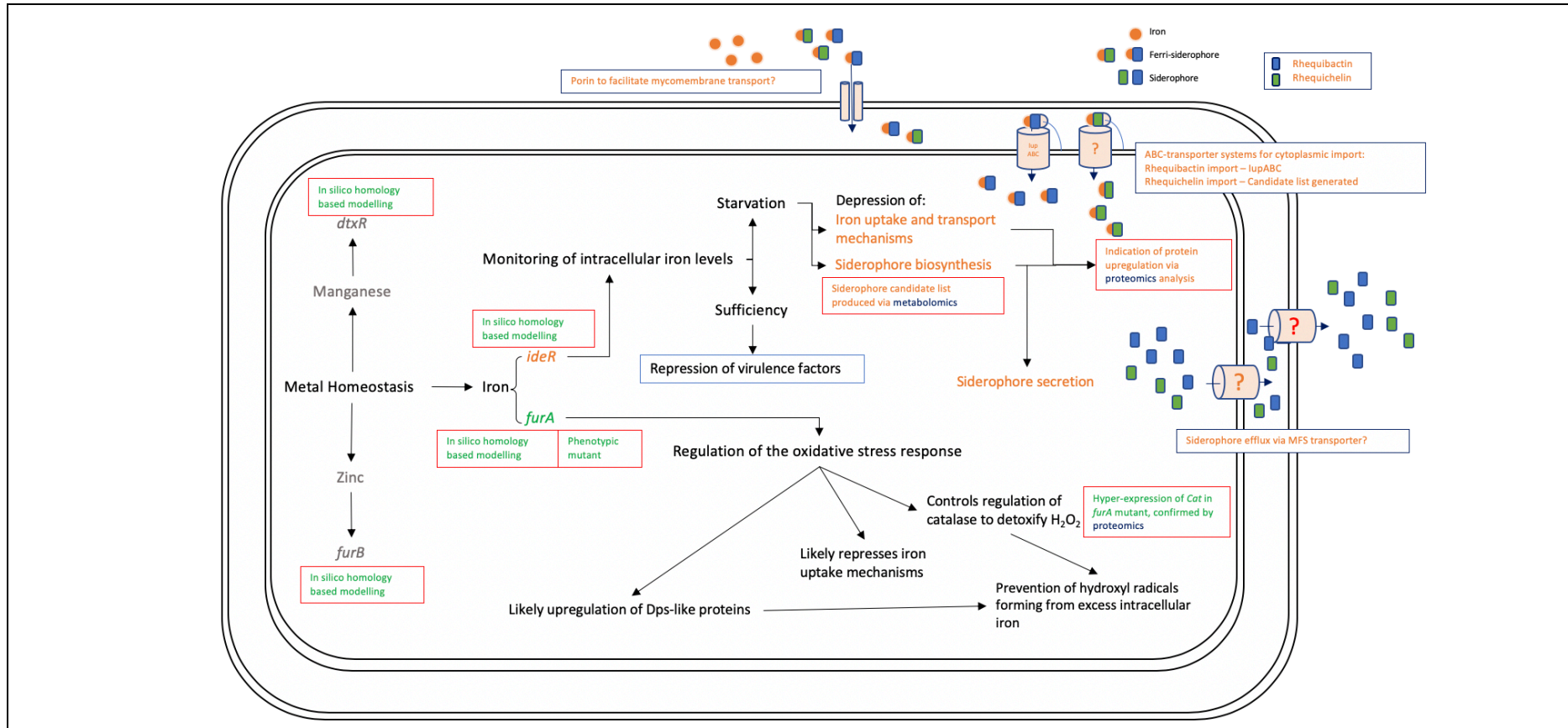


Figure 9-1:- An *R. equi* iron homeostasis network map generated from the results in this thesis.

Experimental procedures identified by red-bordered boxes; blue-bordered boxes indicate predictions. Green text indicates results with a successful experiments or characterised candidate. Orange text indicates further work is needed to fully characterise candidates. Red text indicates uncharacterised components or no candidates available.

9.10 Future work

Although the work presented here contributes to the fundamental knowledge surrounding iron homeostasis within *R. equi*, there is clear grounds for exploration to develop and explore the hypotheses described throughout.

9.10.1 Metal regulatory proteins

With the success of the *in-silico* homology-based modelling, and the prediction of metal binding capacities, this project has scope for further expansion to complete *in vitro* heterologous expression, purification and metal binding capacities of the regulatory proteins. The initial work produces a working hypothesis that can be further explored for accurate characterisation of these proteins. Further to this, when successful the production of these recombinant proteins would be suitable for subsequent macromolecular structure determination, such as X-ray crystallography analysis, again this would further develop the predictions made within this thesis.

9.10.2 Mutagenesis project and further candidates

As all four regulators were successfully produced, further mutant characterisation would be useful. Whole genome sequencing will confirm successful mutagenesis, while ensuring other off-site mutagenic events have not occurred to effect regulation. Priority should be given to *ideR* given the predicted iron-regulatory involvement, with *furA* being the least important, given the expected phenotype developed in collaboration with hyper expression of catalase.

In addition, an exploration of the concept of the super-repression of iron transport components by an orthologous system would be an interesting focus. Repeated culture of mutants (with and without UV-treatment to generate random mutations) and wild type may allow for mutation and relaxation of this system as seen in other lab conditioned strains of other bacteria. Candidate genes involved in this putative regulatory system would be revealed by polymorphisms on WGS. Omic analyses described hereafter would be enhanced by using this approach.

9.10.3 Metabolomic analysis of siderophore candidates

Further attempts at siderophore purification should be explored, perhaps *via* preparative HPLC using a range of alternative adsorption resins as described by other papers to identify the best resin to retain these siderophore compounds. The apparent production of a chromophore chelator should reinforce the results with a visual colorimetric change in the eluant.

Downstream of the collection, the use of an Orbitrap ID-X Tribrid mass spectrometer system would facilitate identification of unknown metabolites, through an automated process that can select precursor ions for subsequent fragmentation and structural elucidation.

9.10.4 Transcriptomics

For overall project impact transcriptomics is essential; detecting changes in gene expression levels will inform us of the essential components utilised in regulation of iron homeostasis within *R. equi*. Issues regarding collection of high-quality RNA has prevented this analysis from being performed within the timescale predicted. Therefore, identification of best practice and process optimisation was performed, by using the modified methodology with a RNeasy isolation kit (Qiagen) with subsequent analysis and confirmation of appropriate RNA integrity. If the RNA integrity at this point is above the required threshold then the downstream processing should be commenced facilitating transcriptome analysis under iron limitation.

10 References

- Akhtar, S., Sarkar, S., Mishra, A., Sarkar, D., 2011. A method to extract intact and pure RNA from mycobacteria. *Analytical biochemistry* 417, 286-288.
- Albrecht-Gary, A.M., Crumbliss, A.L., 1998. Coordination chemistry of siderophores: thermodynamics and kinetics of iron chelation and release. *Metal ions in biological systems* 35, 239-327.
- Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* 37, 420-423.
- Alvarez, H.M., 2010. *Biology of Rhodococcus*. Springer Berlin Heidelberg.
- An, Y.J., Ahn, B.E., Han, A.R., Kim, H.M., Chung, K.M., Shin, J.H., Cho, Y.B., Roe, J.H., Cha, S.S., 2009. Structural basis for the specialization of Nur, a nickel-specific Fur homolog, in metal sensing and DNA recognition. *Nucleic acids research* 37, 3442-3451.
- Anastasi, E., MacArthur, I., Scortti, M., Alvarez, S., Giguere, S., Vazquez-Boland, J.A., 2016. Pangenome and Phylogenomic Analysis of the Pathogenic Actinobacterium *Rhodococcus equi*. *Genome biology and evolution* 8, 3140-3148.
- Andrews, N.C., 1999. Disorders of iron metabolism. *The New England journal of medicine* 341, 1986-1995.
- Andrews, S.C., Robinson, A.K., Rodriguez-Quinones, F., 2003. Bacterial iron homeostasis. *FEMS Microbiol Rev* 27, 215-237.
- Arias, A.A., Lambert, S., Martinet, L., Adam, D., Tenconi, E., Hayette, M.-P., Ongena, M., Rigali, S., 2015. Growth of desferrioxamine-deficient *Streptomyces* mutants through xenosiderophore piracy of airborne fungal contaminations. *FEMS Microbiology Ecology* 91.
- Arora, N.K., Verma, M., 2017. Modified microplate method for rapid and efficient estimation of siderophore produced by bacteria. *3 Biotech* 7, 381-381.
- Auld, D.S., 2001. Zinc coordination sphere in biochemical zinc sites. *Biometals* 14, 271-313.
- Bachmann, B.O., Ravel, J., 2009. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods in enzymology* 458, 181-217.
- Bagg, A., Neilands, J.B., 1987. Ferric uptake regulation protein acts as a repressor, employing iron(II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* 26, 5471-5477.
- Bagos, P.G., Tsirigos, K.D., Liakopoulos, T.D., Hamodrakas, S.J., 2008. Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *Journal of proteome research* 7, 5082-5093.
- Baichoo, N., Wang, T., Ye, R., Helmann, J.D., 2002. Global analysis of the *Bacillus subtilis* Fur regulon and the iron starvation stimulon. *Molecular microbiology* 45, 1613-1629.
- Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., Mardis, E.R., Sadar, M.D., Siddiqui, A.S., Marra, M.A., Jones, S.J., 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7, 246.
- Barclay, R., Ratledge, C., 1983. Iron-binding compounds of *Mycobacterium avium*, *M. intracellulare*, *M. scrofulaceum*, and mycobactin-dependent *M. paratuberculosis* and *M. avium*. *Journal of bacteriology* 153, 1138-1146.
- Barry, T., Geary, S., Hannify, S., MacGearailt, C., Shalloo, M., Heery, D., Gannon, F., Powell, R., 1992. Rapid mini-preparations of total RNA from bacteria. *Nucleic acids research* 20, 4940-4940.
- Bayan, N., Houssin, C., Chami, M., Leblon, G., 2003. Mycomembrane and S-layer: two important structures of *Corynebacterium glutamicum* cell envelope with promising biotechnology applications. *Journal of Biotechnology* 104, 55-67.

- Beale, D.J., Karpe, A.V., Ahmed, W., Cook, S., Morrison, P.D., Staley, C., Sadowsky, M.J., Palombo, E.A., 2017. A Community Multi-Omics Approach towards the Assessment of Surface Water Quality in an Urban River System. *International journal of environmental research and public health* 14.
- Beale, D.J., Karpe, A.V., McLeod, J.D., Gondalia, S.V., Muster, T.H., Othman, M.Z., Palombo, E.A., Joshi, D., 2016. An 'omics' approach towards the characterisation of laboratory scale anaerobic digesters treating municipal sewage sludge. *Water research* 88, 346-357.
- Bell, K.S., Philp, J.C., Aw, D.W., Christofi, N., 1998. The genus *Rhodococcus*. *Journal of applied microbiology* 85, 195-210.
- Biswapriya, B.M., Carl, L., Michael, O., Laura, A.C., 2019. Integrated omics: tools, advances and future approaches. *Journal of Molecular Endocrinology* 62, R21-R45.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research* 47, W81-W87.
- Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T., Medema, M.H., 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic acids research* 45, W36-W41.
- Böcker, S., Letzel, M.C., Lipták, Z., Pervukhin, A., 2008. SIRIUS: decomposing isotope patterns for metabolite identification†. *Bioinformatics* 25, 218-224.
- Boland, C.A., Meijer, W.G., 2000. The iron dependent regulatory protein IdeR (DtxR) of *Rhodococcus equi*. *FEMS microbiology letters* 191, 1-5.
- Bordier, C., 1981. Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* 256, 1604-1607.
- Bosello, M., Robbel, L., Linne, U., Xie, X., Marahiel, M.A., 2011. Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in *Rhodococcus jostii* RHA1. *J Am Chem Soc* 133, 4587-4595.
- Bosello, M., Zeyadi, M., Kraas, F.I., Linne, U., Xie, X., Marahiel, M.A., 2013. Structural characterization of the heterobactin siderophores from *Rhodococcus erythropolis* PR4 and elucidation of their biosynthetic machinery. *Journal of natural products* 76, 2282-2290.
- Boukhalfa, H., Crumbliss, A.L.J.B., 2002. Chemical aspects of siderophore mediated iron transport. *15*, 325-339.
- Breitling, R., 2010. What is systems biology? *Frontiers in physiology* 1, 9-9.
- Bridges, S.M., Magee, G.B., Wang, N., Williams, W.P., Burgess, S.C., Nanduri, B., 2007. ProtQuant: a tool for the label-free quantification of MudPIT proteomics data. *BMC bioinformatics* 8 Suppl 7, S24-S24.
- Broadhurst, D.I., Kell, D.B., 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* 2, 171-196.
- Brune, I., Werner, H., Hüser, A.T., Kalinowski, J., Pühler, A., Tauch, A., 2006. The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics* 7, 21-21.
- Buchanan, R.E., 1915. Nomenclature of the Coccaceae. *The Journal of Infectious Diseases* 17, 528-541.
- Bullen, J.J., Griffiths, E., Edmiston, C.E., 1999. IRON AND INFECTION: MOLECULAR, PHYSIOLOGICAL AND CLINICAL ASPECTS, 2ND EDITION. *Shock* 12, 410.
- Burton, A.J., Giguère, S., Sturgill, T.L., Berghaus, L.J., Slovis, N.M., Whitman, J.L., Levering, C., Kuskie, K.R., Cohen, N.D., 2013. Macrolide- and Rifampin-Resistant *Rhodococcus equi* on a Horse Breeding Farm, Kentucky, USA. *Emerging Infectious Diseases* 19, 282-285.

- Byrne, B.A., Prescott, J.F., Palmer, G.H., Takai, S., Nicholson, V.M., Alperin, D.C., Hines, S.A., 2001. Virulence plasmid of *Rhodococcus equi* contains inducible gene family encoding secreted proteins. *Infection and immunity* 69, 650-656.
- Calderwood, S.B., Mekalanos, J.J., 1988. Confirmation of the Fur operator site by insertion of a synthetic oligonucleotide into an operon fusion plasmid. *J Bacteriol* 170, 1015-1017.
- Capdevila, D.A., Edmonds, K.A., Giedroc, D.P., 2017. Metallochaperones and metalloregulation in bacteria. *Essays in biochemistry* 61, 177-200.
- Carran, C.J., Jordan, M., Drechsel, H., Schmid, D.G., Winkelmann, G., 2001. Heterobactins: A new class of siderophores from *Rhodococcus erythropolis* IGTS8 containing both hydroxamate and catecholate donor groups. *Biometals* 14, 119-125.
- Carrano, C.J., Jordan, M., Drechsel, H., Schmid, D.G., Winkelmann, G., 2001. Heterobactins: A new class of siderophores from *Rhodococcus erythropolis* IGTS8 containing both hydroxamate and catecholate donor groups. *Biometals* 14, 119-125.
- Caza, M., Kronstad, J.W., 2013. Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans. *Frontiers in Cellular and Infection Microbiology* 3.
- Challis, G.L., 2008. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology (Reading, England)* 154, 1555-1569.
- Challis, G.L., Hopwood, D.A., 2003. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proceedings of the National Academy of Sciences* 100, 14555-14561.
- Challis, G.L., Ravel, J., Townsend, C.A., 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology* 7, 211-224.
- Chen, C., Zabad, S., Liu, H., Wang, W., Jeffery, C., 2018. MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic acids research* 46, D640-d644.
- Chen, S., Tong, X., Woodard, R.W., Du, G., Wu, J., Chen, J., 2008. Identification and characterization of bacterial cutinase. *The Journal of biological chemistry* 283, 25854-25862.
- Cheung, A.L., Eberhardt, K.J., Fischetti, V.A., 1994. A method to isolate RNA from gram-positive bacteria and mycobacteria. *Analytical biochemistry* 222, 511-514.
- Cho, C.R., Labow, M., Reinhardt, M., van Oostrum, J., Peitsch, M.C., 2006. The application of systems biology to drug discovery. *Current opinion in chemical biology* 10, 294-302.
- Chu, J., Vila-Farres, X., Inoyama, D., Ternei, M., Cohen, L.J., Gordon, E.A., Reddy, B.V.B., Charlop-Powers, Z., Zebroski, H.A., Gallardo-Macias, R., Jaskowski, M., Satish, S., Park, S., Perlin, D.S., Freundlich, J.S., Brady, S.F., 2016. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nature chemical biology* 12, 1004.
- Clarke, T.E., Tari, L.W., Vogel, H.J., 2001. Structural biology of bacterial iron uptake systems. *Current topics in medicinal chemistry* 1, 7-30.
- Claverys, J.P., 2001. A new family of high-affinity ABC manganese and zinc permeases. *Research in microbiology* 152, 231-243.
- Clish, C.B., 2015. Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud* 1, a000588-a000588.
- Cong, X., Yuan, Z., Wang, Z., Wei, B., Xu, S., Wang, J., 2018. Crystal structures of manganese-dependent transcriptional repressor MntR (Rv2788) from *Mycobacterium tuberculosis* in apo and manganese bound forms. *Biochemical and biophysical research communications* 501, 423-427.
- Creason, A.L., Davis, E.W., Putnam, M.L., Vandeputte, O.M., Chang, J.H., 2014. Use of whole genome sequences to develop a molecular phylogenetic framework for *Rhodococcus fascians* and the *Rhodococcus* genus. 5.

- Crombie, A.T., Khawand, M.E., Rhodius, V.A., Fengler, K.A., Miller, M.C., Whited, G.M., McGenity, T.J., Murrell, J.C., 2015. Regulation of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link in the global isoprene cycle. *Environ Microbiol* 17, 3314-3329.
- Crosa, J.H., Walsh, C.T., 2002. Genetics and Assembly Line Enzymology of Siderophore Biosynthesis in Bacteria. *Microbiology and Molecular Biology Reviews* 66, 223-249.
- Day, J.G., Stacey, G.N., 2008. Biobanking. *Molecular biotechnology* 40, 202-213.
- de Carvalho, C.C.C.R., Costa, S.S., Fernandes, P., Couto, I., Viveiros, M., 2014. Membrane transport systems and the biodegradation potential and pathogenicity of genus *Rhodococcus*. *Frontiers in Physiology* 5, 133.
- Deng, Z., Wang, Q., Liu, Z., Zhang, M., Machado, A.C.D., Chiu, T.-P., Feng, C., Zhang, Q., Yu, L., Qi, L., Zheng, J., Wang, X., Huo, X., Qi, X., Li, X., Wu, W., Rohs, R., Li, Y., Chen, Z., 2015. Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nature Communications* 6, 7642.
- Dhungana, S., Crumbliss, A.L., 2005. Coordination Chemistry and Redox Processes in Siderophore-Mediated Iron Transport. *Geomicrobiology Journal* 22, 87-98.
- Dhungana, S., Michalczyk, R., Boukhalfa, H., Lack, J.G., Koppisch, A.T., Fairlee, J.M., Johnson, M.T., Ruggiero, C.E., John, S.G., Cox, M.M., Browder, C.C., Forsythe, J.H., Vanderberg, L.A., Neu, M.P., Hersman, L.E., 2007. Purification and characterization of rhodobactin: a mixed ligand siderophore from *Rhodococcus rhodochrous* strain OFS. *Biometals* 20, 853-867.
- Dhungana, S., Miller, M.J., Dong, L., Ratledge, C., Crumbliss, A.L., 2003. Iron Chelation Properties of an Extracellular Siderophore Exochelin MN. *Journal of the American Chemical Society* 125, 7654-7663.
- Diaz-Mireles, E., Wexler, M., Todd, J.D., Bellini, D., Johnston, A.W., Sawers, R.G., 2005. The manganese-responsive repressor Mur of *Rhizobium leguminosarum* is a member of the Fur superfamily that recognizes an unusual operator sequence. *Microbiology (Reading, England)* 151, 4071-4078.
- Dong, Z., Chen, Y., 2013. Transcriptomics: advances and approaches. *Science China. Life sciences* 56, 960-967.
- Dubeau, M.P., Ghinet, M.G., Jacques, P.E., Clermont, N., Beaulieu, C., Brzezinski, R., 2009. Cytosine deaminase as a negative selection marker for gene disruption and replacement in the genus *Streptomyces* and other actinobacteria. *Applied and environmental microbiology* 75, 1211-1214.
- Dussurget, O., Rodriguez, M., Smith, I., 1996. An ideR mutant of *Mycobacterium smegmatis* has derepressed siderophore production and an altered oxidative-stress response. *Molecular microbiology* 22, 535-544.
- Dussurget, O., Smith, I., 1998. Interdependence of mycobacterial iron regulation, oxidative-stress response and isoniazid resistance. *Trends in Microbiology* 6, 354-358.
- Dussurget, O., Timm, J., Gomez, M., Gold, B., Yu, S., Sabol, S.Z., Holmes, R.K., Jacobs, W.R., Jr., Smith, I., 1999. Transcriptional control of the iron-responsive fxbA gene by the mycobacterial regulator IdeR. *Journal of bacteriology* 181, 3402-3408.
- Emery, T., 1982. Iron Metabolism in Humans and Plants: Understanding how microorganisms assimilate iron has important consequences for the health of both plants and humans. *American Scientist* 70, 626-632.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486-501.
- Epstein, C.J., 1967. Non-randomness of Ammo-acid Changes in the Evolution of Homologous Proteins. *Nature* 215, 355-359.
- Escobar, L., Perez-Martin, J., de Lorenzo, V., 1999. Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* 181, 6223-6229.

- Feese, M.D., Ingason, B.P., Goranson-Siekierke, J., Holmes, R.K., Hol, W.G., 2001. Crystal structure of the iron-dependent regulator from *Mycobacterium tuberculosis* at 2.0-Å resolution reveals the Src homology domain 3-like fold and metal binding function of the third domain. *J Biol Chem* 276, 5959-5966.
- Fiehn, O., 2002. Metabolomics--the link between genotypes and phenotypes. *Plant molecular biology* 48, 155-171.
- Fillat, M.F., 2014. The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators. *Archives of biochemistry and biophysics* 546, 41-52.
- Fiss, E., Brooks, G.F., 1991. Use of a siderophore detection medium, ethylene glycol degradation, and beta-galactosidase activity in the early presumptive differentiation of *Nocardia*, *Rhodococcus*, *Streptomyces*, and rapidly growing *Mycobacterium* species. 29, 1533-1535.
- Foster, A.W., Osman, D., Robinson, N.J., 2014. Metal Preferences and Metallation. *Journal of Biological Chemistry* 289, 28095-28103.
- Freestone, J.F., Hietala, S., Moulton, J., Vivrette, S., 1987. Acquired immunodeficiency in a seven-year-old horse. *Journal of the American Veterinary Medical Association* 190, 689-691.
- Fukushima, T., Allred, B.E., Raymond, K.N., 2014. Direct Evidence of Iron Uptake by the Gram-Positive Siderophore-Shuttle Mechanism without Iron Reduction. *ACS Chemical Biology* 9, 2092-2100.
- Fukushima, T., Allred, B.E., Sia, A.K., Nichiporuk, R., Andersen, U.N., Raymond, K.N., 2013. Gram-positive siderophore-shuttle with iron-exchange from Fe-siderophore to apo-siderophore by *Bacillus cereus* YxeB. *Proceedings of the National Academy of Sciences of the United States of America* 110, 13821-13826.
- Gallego Romero, I., Pai, A.A., Tung, J., Gilad, Y., 2014. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology* 12, 42.
- Garcia, D.E., Baidoo, E.E., Benke, P.I., Pingitore, F., Tang, Y.J., Villa, S., Keasling, J.D., 2008. Separation and mass spectrometry in microbial metabolomics. *Current opinion in microbiology* 11, 233-239.
- Garrity, G.M., 2014. Conservation of *Rhodococcus equi* (Magnusson 1923) Goodfellow and Alderson 1977 and rejection of *Corynebacterium hoagii* (Morse 1912) Ebersson 1918. *International journal of systematic and evolutionary microbiology* 64, 311-312.
- Gates, S.C., Sweeley, C.C., 1978. Quantitative metabolic profiling based on gas chromatography. *Clinical chemistry* 24, 1663-1673.
- Giannoukos, G., Ciulla, D.M., Huang, K., Haas, B.J., IZard, J., Levin, J.Z., Livny, J., Earl, A.M., Gevers, D., Ward, D.V., 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome biology* 13, r23.
- Giguère, S., Cohen, N.D., Keith Chaffin, M., Slovis, N.M., Hondalus, M.K., Hines, S.A., Prescott, J.F., 2011. Diagnosis, Treatment, Control, and Prevention of Infections Caused by *Rhodococcus equi* in Foals. *Journal of Veterinary Internal Medicine* 25, 1209-1220.
- Giguère, S., Lee, E.A., Guldbeck, K.M., Berghaus, L.J., 2012. In vitro synergy, pharmacodynamics, and postantibiotic effect of 11 antimicrobial agents against *Rhodococcus equi*. *Veterinary microbiology* 160, 207-213.
- Giguere, S., Prescott, J.F., 1997. Clinical manifestations, diagnosis, treatment, and prevention of *Rhodococcus equi* infections in foals. *Veterinary microbiology* 56, 313-334.
- Golovin, A., Dimitropoulos, D., Oldfield, T., Rachedi, A., Henrick, K., 2005. MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. *Proteins* 58, 190-199.
- Golub, B., Falk, G., Spink, W.W., 1967. Lung abscess due to *Corynebacterium equi*. Report of first human infection. *Annals of internal medicine* 66, 1174-1177.

- Goodfellow, M., Alderson, G., 1977. The actinomycete-genus *Rhodococcus*: a home for the "rhodochrous" complex. *Journal of general microbiology* 100, 99-122.
- Goodfellow, M., Thomas, E.G., Ward, A.C., James, A.L., 1990. Classification and Identification of *Rhodococci*. *Zentralblatt für Bakteriologie* 274, 299-315.
- Gorska, A., Sloderbach, A., Marszall, M.P., 2014. Siderophore-drug complexes: potential medicinal applications of the 'Trojan horse' strategy. *Trends in pharmacological sciences* 35, 442-449.
- Graves, P.R., Haystead, T.A.J., 2002. *Molecular Biologist's Guide to Proteomics*. *Microbiology and Molecular Biology Reviews* 66, 39-63.
- Gray, K.J., French, N., Lugada, E., Watera, C., Gilks, C.F., 2000. *Rhodococcus equi* and HIV-1 infection in Uganda. *The Journal of infection* 41, 227-231.
- Grigg, J.C., Cheung, J., Heinrichs, D.E., Murphy, M.E.P., 2010. Specificity of Staphyloferrin B Recognition by the SirA Receptor from *Staphylococcus aureus*. *The Journal of Biological Chemistry* 285, 34579-34588.
- Gubbens, J., Wu, C., Zhu, H., Filippov, D.V., Florea, B.I., Rigali, S., Overkleeft, H.S., van Wezel, G.P., 2017. Qinichelins, novel catecholate-hydroxamate siderophores synthesized via a multiplexed convergent biosynthesis pathway. *bioRxiv*, 163170.
- Gupta, M., Sajid, A., Sharma, K., Ghosh, S., Arora, G., Singh, R., Nagaraja, V., Tandon, V., Singh, Y., 2014. HupB, a nucleoid-associated protein of *Mycobacterium tuberculosis*, is modified by serine/threonine protein kinases in vivo. *Journal of bacteriology* 196, 2646-2657.
- Gurel, V., Lambert, K., Page, A.E., Loynachan, A.T., Huges, K., Timoney, J.F., Fettingner, M., Horohov, D.W., McMichael, J., 2013. Streptolysin-O/antibiotics adjunct therapy modulates site-specific expression of extracellular matrix and inflammatory genes in lungs of *Rhodococcus equi* infected foals. *Veterinary research communications* 37, 145-154.
- Gürtler, V., Seviour, R. 2010. Systematics of Members of the Genus *Rhodococcus* (Zopf 1891) Emend Goodfellow et al. 1998, In: Alvarez, H.M. (Ed.) *Biology of Rhodococcus*. Springer Berlin Heidelberg, 1-28.
- Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., Livny, J., 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* 13, 734.
- Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., Ralser, M., 2017. Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. *Current Opinion in Systems Biology* 6, 37-45.
- Hall, R.M., Ratledge, C., 1982. A simple method for the production of mycobactin, the lipid-soluble siderophore, from mycobacteria. *FEMS microbiology letters* 15, 133-136.
- Hall, R.M., Ratledge, C., 1986. Distribution and Application of Mycobactins for the Characterization of Species within the Genus *Rhodococcus*. 132, 853-856.
- Hantke, K., 2001. Iron and metal regulation in bacteria. *Current opinion in microbiology* 4, 172-177.
- Harrington, J.M., Crumbliss, A.L.J.B., 2009. The redox hypothesis in siderophore-mediated iron uptake. 22, 679-689.
- Havlis, J., Shevchenko, A., 2004. Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Analytical chemistry* 76, 3029-3036.
- He, L., Diedrich, J., Chu, Y.Y., Yates, J.R., 3rd, 2015. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Analytical chemistry* 87, 11361-11367.
- Helmann, J.D., 2014. Specificity of Metal Sensing: Iron and Manganese Homeostasis in *Bacillus subtilis*. *Journal of Biological Chemistry* 289, 28112-28120.
- Henderson, D.P., Payne, S.M., 1994. *Vibrio cholerae* iron transport systems: roles of heme and siderophore iron transport in virulence and identification of a gene associated with multiple iron transport systems. *Infection and immunity* 62, 5120-5125.

- Herbst, D.A., Boll, B., Zocher, G., Stehle, T., Heide, L., 2013. Structural Basis of the Interaction of MbtH-like Proteins, Putative Regulators of Nonribosomal Peptide Biosynthesis, with Adenylyating Enzymes. *Journal of Biological Chemistry* 288, 1991-2003.
- Hider, R.C., Kong, X., 2010. Chemistry and biology of siderophores. *Natural product reports* 27, 637-657.
- Higgins, K.A., Giedroc, D.P. 2013. Metal Specificity of Metallosensors, In: *Encyclopedia of Inorganic and Bioinorganic Chemistry*. 1-16.
- Holden, V.I., Bachman, M.A., 2015. Diverging roles of bacterial siderophores during infection. *Metallomics : integrated biometal science* 7, 986-995.
- Hollenstein, K., Frei, D.C., Locher, K.P., 2007. Structure of an ABC transporter in complex with its binding protein. *Nature* 446, 213.
- Holm, L., 2019. Benchmarking fold detection by DaliLite v.5. *Bioinformatics*.
- Holm, L., Laakso, L.M., 2016. Dali server update. *Nucleic acids research* 44, W351-W355.
- Huang, S., Chaudhary, K., Garmire, L., 2017. More is better: recent progress in multi-omics data integration methods. *Frontiers in Genetics* 8, 84.
- Huberts, D.H.E.W., van der Klei, I.J., 2010. Moonlighting proteins: An intriguing mode of multitasking. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803, 520-525.
- Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., Speed, T.P., 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research* 31, e15-e15.
- Irving, H., Williams, R.J.P., 1953. 637. The stability of transition-metal complexes. *Journal of the Chemical Society (Resumed)*, 3192-3210.
- Jahn, C.E., Charkowski, A.O., Willis, D.K., 2008. Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of microbiological methods* 75, 318-324.
- Ji, C.J., Yang, Y.M., Kim, J.H., Ryu, S.H., Youn, H., Lee, J.W., 2018. The roles of two O-donor ligands in the Fe(2+)-binding and H₂O₂-sensing by the Fe(2+)-dependent H₂O₂ sensor PerR. *Biochemical and biophysical research communications* 501, 458-464.
- Jones, A.L., Goodfellow, M., 2012. Genus *Rhodococcus* (Zopf 1891) emend. Goodfellow, Alderson and Chun 1998. In *Bergey's Manual of Systematic Bacteriology*, 437-496.
- Jones, A.L., Sutcliffe, I.C., Goodfellow, M., 2013a. *Prescottia equi* gen. nov., comb. nov.: a new home for an old pathogen. *Antonie van Leeuwenhoek* 103, 655-671.
- Jones, A.L., Sutcliffe, I.C., Goodfellow, M., 2013b. Proposal to replace the illegitimate genus name *Prescottia* Jones et al. 2013 with the genus name *Prescottella* gen. nov. and to replace the illegitimate combination *Prescottia equi* Jones et al. 2013 with *Prescottella equi* comb. nov. *Antonie van Leeuwenhoek* 103, 1405-1407.
- Jordan, M.C., Harrington, J.R., Cohen, N.D., Tsohis, R.M., Dangott, L.J., Weinberg, E.D., Martens, R.J., 2003. Effects of iron modulation on growth and viability of *Rhodococcus equi* and expression of virulence-associated protein A. *American journal of veterinary research* 64, 1337-1346.
- Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.
- Kall, L., Krogh, A., Sonnhammer, E.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic acids research* 35, W429-432.
- Kampfer, P., Dott, W., Martin, K., Glaeser, S.P., 2014. *Rhodococcus defluvii* sp. nov., isolated from wastewater of a bioreactor and formal proposal to reclassify [*Corynebacterium hoagii*] and *Rhodococcus equi* as *Rhodococcus hoagii* comb. nov. *International journal of systematic and evolutionary microbiology* 64, 755-761.
- Kedlaya, I., Ing, M.B., Wong, S.S., 2001. *Rhodococcus equi* Infections in Immunocompetent Hosts: Case Report and Review. *Clinical Infectious Diseases* 32, e39-e46.

- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015a. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* 10, 845-858.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015b. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845.
- Kelly, B.G., Wall, D.M., Boland, C.A., Meijer, W.G., 2002. Isocitrate lyase of the facultative intracellular pathogen *Rhodococcus equi*. *Microbiology (Reading, England)* 148, 793-798.
- Kem, M.P., Butler, A., 2015. Acyl peptidic siderophores: structures, biosyntheses and post-assembly modifications. *BioMetals* 28, 445-459.
- Kind, T., Fiehn, O., 2010. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* 2, 23-60.
- Klein, J.S., Lewinson, O., 2011. Bacterial ATP-driven transporters of transition metals: physiological roles, mechanisms of action, and roles in bacterial virulence. *Metallomics : integrated biometal science* 3, 1098-1108.
- Kobayashi, K., Nakagaki, M., Ishikawa, H., Iwai, K., O'Brian, M.R., Ishimori, K., 2016. Redox-Dependent Dynamics in Heme-Bound Bacterial Iron Response Regulator (Irr) Protein. *Biochemistry* 55, 4047-4054.
- Krewulak, K.D., Vogel, H.J., 2008. Structural biology of bacterial iron uptake. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778, 1781-1804.
- Krewulak, K.D., Vogel, H.J., 2011. TonB or not TonB: is that the question? *Biochemistry and cell biology = Biochimie et biologie cellulaire* 89, 87-97.
- Kukurba, K.R., Montgomery, S.B., 2015. RNA Sequencing and Analysis. *Cold Spring Harbor Protocols* 2015, pdb.top084970.
- Kunkle, C.A., Schmitt, M.P., 2003. Analysis of the *Corynebacterium diphtheriae* DtxR Regulon: Identification of a Putative Siderophore Synthesis and Transport System That Is Similar to the *Yersinia* High-Pathogenicity Island-Encoded Yersiniabactin Synthesis and Uptake System. *Journal of Bacteriology* 185, 6826-6840.
- Kuo, T.C., Tian, T.F., Tseng, Y.J., 2013. 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC systems biology* 7, 64.
- Kuyukina, M., Ivshina, I. 2010. Application of *Rhodococcus* in Bioremediation of Contaminated Environments, In: Alvarez, H.M. (Ed.) *Biology of Rhodococcus*. Springer Berlin Heidelberg, 231-262.
- Lamb, A.L., 2015. Breaking a pathogen's iron will: inhibiting siderophore production as an antimicrobial strategy. *Biochimica et biophysica acta* 1854, 1054-1070.
- Lautru, S., Oves-Costales, D., Pernodet, J.L., Challis, G.L., 2007. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology (Reading, England)* 153, 1405-1412.
- Lee, J.-W., Helmann, J.D., 2006. The PerR transcription factor senses H₂O₂ by metal-catalysed histidine oxidation. *Nature* 440, 363-367.
- Lee, J.W., Helmann, J.D., 2007. Functional specialization within the Fur family of metalloregulators. *Biomaterials* 20, 485-499.
- Letek, M., González, P., MacArthur, I., Rodríguez, H., Freeman, T.C., Valero-Rello, A., Blanco, M., Buckley, T., Cherevach, I., Fahey, R., Hapeshi, A., Holdstock, J., Leadon, D., Navas, J., Ocampo, A., Quail, M.A., Sanders, M., Scortti, M.M., Prescott, J.F., Fogarty, U., Meijer, W.G., Parkhill, J., Bentley, S.D., Vázquez-Boland, J.A., 2010. The Genome of a Pathogenic *Rhodococcus*: Cooptive Virulence Underpinned by Key Gene Acquisitions. *PLoS genetics* 6, e1001145.
- Letek, M., Ocampo-Sosa, A.A., Sanders, M., Fogarty, U., Buckley, T., Leadon, D.P., González, P., Scortti, M., Meijer, W.G., Parkhill, J., Bentley, S., Vázquez-Boland, J.A., 2008. Evolution of the

- Rhodococcus equi vap Pathogenicity Island Seen through Comparison of Host-Associated vapA and vapB Virulence Plasmids. *Journal of Bacteriology* 190, 5797-5805.
- Linne, U., Schwarzer, D., Schroeder, G.N., Marahiel, M.A., 2004. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *European journal of biochemistry* 271, 1536-1545.
- Litwin, C.M., Calderwood, S.B., 1993. Role of iron in regulation of virulence genes. *Clin Microbiol Rev* 6, 137-149.
- Liu, S., Perera, L., Pedersen, L., 2007. Binuclear manganese(II) complexes in biological systems. *Molecular Physics - MOL PHYS* 105, 2893-2898.
- Locher, K.P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J.P., Moras, D., 1998. Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* 95, 771-778.
- Louden, B.C., Haarmann, D., Lynne, A.M., 2011. Use of Blue Agar CAS Assay for Siderophore Detection. *J Microbiol Biol Educ* 12, 51-53.
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. *PLOS Computational Biology* 13, e1005457.
- Lu, G., Moriyama, E.N., 2004. Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in Bioinformatics* 5, 378-388.
- Lucarelli, D., Russo, S., Garman, E., Milano, A., Meyer-Klaucke, W., Pohl, E., 2007. Crystal Structure and Function of the Zinc Uptake Regulator FurB from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 282, 9914-9922.
- Luo, M., Fadeev, E.A., Groves, J.T., 2005. Mycobactin-mediated iron acquisition within macrophages. *Nature chemical biology* 1, 149-153.
- Ma, Z., Jacobsen, F.E., Giedroc, D.P., 2009. Coordination Chemistry of Bacterial Metal Transport and Sensing. *Chemical Reviews* 109, 4644-4681.
- Maciag, A., Dainese, E., Rodriguez, G.M., Milano, A., Provvedi, R., Pasca, M.R., Smith, I., Palu, G., Riccardi, G., Manganelli, R., 2007. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *Journal of bacteriology* 189, 730-740.
- Magnusson, A.H., 1923. New or interesting Swedish lichens. Lund.
- Maitra, A., Munshi, T., Healy, J., Martin, L.T., Vollmer, W., Keep, N.H., Bhakta, S., 2019. Cell wall peptidoglycan in *Mycobacterium tuberculosis*: An Achilles' heel for the TB-causing pathogen. *FEMS Microbiology Reviews* 43, 548-575.
- Makrai, L., Takai, S., Tamura, M., Tsukamoto, A., Sekimoto, R., Sasaki, Y., Kakuda, T., Tsubaki, S., Varga, J., Fodor, L., Solymosi, N., Major, A., 2002. Characterization of virulence plasmid types in *Rhodococcus equi* isolates from foals, pigs, humans and soil in Hungary. *Veterinary microbiology* 88, 377-384.
- Manninen, T., Acimovic, J., Havela, R., Teppola, H., Linne, M.L., 2018. Challenges in Reproducibility, Replicability, and Comparability of Computational Models and Tools for Neuronal and Glial Networks, Cells, and Subcellular Structures. *Frontiers in neuroinformatics* 12, 20.
- Mao, F., Dam, P., Chou, J., Oلمان, V., Xu, Y., 2009. DOOR: a database for prokaryotic operons. *Nucleic acids research* 37, D459-463.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research* 45, D200-d203.
- Masaki, T., Qu, J., Cholewa-Waclaw, J., Burr, K., Raaum, R., Rambukkana, A., 2013. Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection. *Cell* 152, 51-67.

- Meena, L.S., Rajni, 2010. Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *The FEBS Journal* 277, 2416-2427.
- Meijer, W.G., Prescott, J.F., 2004. *Rhodococcus equi*. *Veterinary research* 35, 383-396.
- Meyer, J.M., Abdallah, M.A., 1978. The Fluorescent Pigment of *Pseudomonas fluorescens*: Biosynthesis, Purification and Physicochemical Properties. *Microbiology (Reading, England)* 107, 319-328.
- Miethke, M., Marahiel, M.A., 2007. Siderophore-Based Iron Acquisition and Pathogen Control. *Microbiology and Molecular Biology Reviews* 71, 413-451.
- Miethke, M., Westers, H., Blom, E.-J., Kuipers, O.P., Marahiel, M.A., 2006. Iron starvation triggers the stringent response and induces amino acid biosynthesis for bacillibactin production in *Bacillus subtilis*. *Journal of bacteriology* 188, 8655-8657.
- Mietzner, T.A., Morse, S.A., 1994. The role of iron-binding proteins in the survival of pathogenic bacteria. *Annual review of nutrition* 14, 471-493.
- Milano, A., Forti, F., Sala, C., Riccardi, G., Ghisotti, D., 2001. Transcriptional Regulation of *furA* and *katG* upon Oxidative Stress in *Mycobacterium smegmatis*. *Journal of Bacteriology* 183, 6801.
- Miller, D.D., Berner, L.A., 1989. Is solubility in vitro a reliable predictor of iron bioavailability? *Biological Trace Element Research* 19, 11.
- Miranda-CasoLuengo, R., Coulson, G.B., Miranda-CasoLuengo, A., Vázquez-Boland, J.A., Hondalus, M.K., Meijer, W.G., 2012. The Hydroxamate Siderophore Rhequichelin Is Required for Virulence of the Pathogenic Actinomycete *Rhodococcus equi*. *Infection and immunity* 80, 4106-4114.
- Miranda-CasoLuengo, R., Duffy, P.S., O'Connell, E.P., Graham, B.J., Mangan, M.W., Prescott, J.F., Meijer, W.G., 2005. The Iron-Regulated *iupABC* Operon Is Required for Saprophytic Growth of the Intracellular Pathogen *Rhodococcus equi* at Low Iron Concentrations. *Journal of Bacteriology* 187, 3438-3444.
- Miranda-CasoLuengo, R., Prescott, J.F., Vazquez-Boland, J.A., Meijer, W.G., 2008. The intracellular pathogen *Rhodococcus equi* produces a catecholate siderophore required for saprophytic growth. *J Bacteriol* 190, 1631-1637.
- Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y., Finn, R.D., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research* 47, D351-d360.
- Mizuno, H., Ueda, K., Kobayashi, Y., Tsuyama, N., Todoroki, K., Min, J.Z., Toyo'oka, T., 2017. The great importance of normalization of LC-MS data for highly-accurate non-targeted metabolomics. *Biomedical Chromatography* 31, e3864.
- Moeck, G.S., Coulton, J.W., 1998. TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport. *Molecular microbiology* 28, 675-681.
- Monahan, I.M., Mangan, J.A., Butcher, P.D. 2001. Extraction of RNA from intracellular *Mycobacterium tuberculosis*, In: *Mycobacterium tuberculosis protocols*. Springer, 31-42.
- Montecchi-Palazzi, L., Beavis, R., Binz, P.A., Chalkley, R.J., Cottrell, J., Creasy, D., Shofstahl, J., Seymour, S.L., Garavelli, J.S., 2008. The PSI-MOD community standard for representation of protein modification data. *Nature biotechnology* 26, 864-866.
- Mootz, H.D., Finking, R., Marahiel, M.A., 2001. 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J Biol Chem* 276, 37289-37298.

- Morgan, J.W., Anders, E., 1980. Chemical composition of Earth, Venus, and Mercury. *Proceedings of the National Academy of Sciences of the United States of America* 77, 6973-6977.
- Mosser, D.M., Hondalus, M.K., 1996. *Rhodococcus equi*: an emerging opportunistic pathogen. *Trends Microbiol* 4, 29-33.
- Mueller, O., Lightfoot, S., Schroeder, A., 2004. RNA integrity number (RIN)–standardization of RNA quality control.
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J., Gerstein, M., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome biology* 17, 53-53.
- Müller, T., Winter, D., 2017. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Molecular & Cellular Proteomics*, mcp.M116.064048.
- Mundade, R., Ozer, H.G., Wei, H., Prabhu, L., Lu, T., 2014. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* 13, 2847-2852.
- Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology* 247, 536-540.
- Muscatello, G., Leadon, D.P., Klay, M., Ocampo-Sosa, A., Lewis, D.A., Fogarty, U., Buckley, T., Gilkerson, J.R., Meijer, W.G., Vazquez-Boland, J.A., 2007. *Rhodococcus equi* infection in foals: the science of ‘rattles’. *Equine Veterinary Journal* 39, 470-478.
- Mussap, M., Zaffanello, M., Fanos, V., 2018. Metabolomics: a challenge for detecting and monitoring inborn errors of metabolism. *Ann Transl Med* 6, 338-338.
- Nagalakshmi, U., Waern, K., Snyder, M., 2010. RNA-Seq: a method for comprehensive transcriptome analysis. *Current protocols in molecular biology* Chapter 4, Unit 4.11.11-13.
- Naka, H., Liu, M., Actis, L.A., Crosa, J.H., 2013. Plasmid- and chromosome-encoded siderophore anguibactin systems found in marine vibrios: biosynthesis, transport and evolution. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 26, 537-547.
- Nies, D.H., 2016. The biological chemistry of the transition metal "transportome" of *Cupriavidus metallidurans*. *Metallomics : integrated biometal science* 8, 481-507.
- Nikolaev, D.M., Shtyrov, A.A., Panov, M.S., Jamal, A., Chakchir, O.B., Kochemirovsky, V.A., Olivucci, M., Ryazantsev, M.N., 2018. A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction. *ACS Omega* 3, 7555-7566.
- Nikolaivits, E., Kanelli, M., Dimarogona, M., Topakas, E., 2018. A Middle-Aged Enzyme Still in Its Prime: Recent Advances in the Field of Cutinases. *Catalysts* 8, 612.
- Norrandar, J., Kempe, T., Messing, J., 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26, 101-106.
- O'Halloran, T.V., Culotta, V.C., 2000. Metallochaperones, an intracellular shuttle service for metal ions. *J Biol Chem* 275, 25057-25060.
- Ocampo-Sosa, A.A., Lewis, D.A., Navas, J., Quigley, F., Callejo, R., Scorti, M., Leadon, D.P., Fogarty, U., Vázquez-Boland, J.A., 2007. Molecular Epidemiology of *Rhodococcus equi* Based on *traA*, *vapA*, and *vapB* Virulence Plasmid Markers. *Journal of Infectious Diseases* 196, 763-769.
- Ollinger, J., Song, K.B., Antelmann, H., Hecker, M., Helmman, J.D., 2006. Role of the Fur regulon in iron transport in *Bacillus subtilis*. *J Bacteriol* 188, 3664-3673.
- Ooi, W.W., Srinivasan, J., 2004. Leprosy and the peripheral nervous system: basic and clinical aspects. *Muscle & nerve* 30, 393-409.
- Outten, F.W., Theil, E.C., 2009. Iron-based redox switches in biology. *Antioxid Redox Signal* 11, 1029-1046.

- Overbeck, A., 1891. Zur Kenntnis der Fettfarbstoff-Production bei Spaltpilzen Halle, [399]-416 p. pp.
- Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12, 87-98.
- Pagan-Ramos, E., Song, J., McFalone, M., Mudd, M.H., Deretic, V., 1998. Oxidative stress response and characterization of the *oxyR-ahpC* and *furA-katG* loci in *Mycobacterium marinum*. *J Bacteriol* 180, 4856-4864.
- Pandey, R., Rodriguez, G.M., 2014. IdeR is required for iron homeostasis and virulence in *Mycobacterium tuberculosis*. *Molecular microbiology* 91, 98-109.
- Pandey, S.D., Choudhury, M., Yousuf, S., Wheeler, P.R., Gordon, S.V., Ranjan, A., Sritharan, M., 2014. Iron-regulated protein HupB of *Mycobacterium tuberculosis* positively regulates siderophore biosynthesis and is essential for growth in macrophages. *Journal of bacteriology* 196, 1853-1865.
- Papp-Wallace, K.M., Maguire, M.E., 2006. Manganese transport and the role of manganese in virulence. *Annual review of microbiology* 60, 187-209.
- Park, S.T., Kim, J., 2016. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *Int Neurourol J* 20, S76-S83.
- Patterson, S.D., Aebersold, R.H., 2003. Proteomics: the first decade and beyond. *Nature genetics* 33 Suppl, 311-323.
- Paulo, J.A., 2016. Sample preparation for proteomic analysis using a GeLC-MS/MS strategy. 2016.
- Payton, M., Pinter, K., 1999. A rapid novel method for the extraction of RNA from wild-type and genetically modified kanamycin resistant mycobacteria. *FEMS microbiology letters* 180, 141-146.
- Pei, Y., Parreira, V., Nicholson, V.M., Prescott, J.F., 2007. Mutation and virulence assessment of chromosomal genes of *Rhodococcus equi* 103. *Canadian Journal of Veterinary Research* 71, 1-7.
- Peng, B., Li, H., Peng, X.-X., 2015. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* 6, 628-637.
- Peuckert, F., Ramos-Vega, A.L., Miethke, M., Schworer, C.J., Albrecht, A.G., Oberthur, M., Marahiel, M.A., 2011. The siderophore binding protein FeuA shows limited promiscuity toward exogenous triscatecholates. *Chemistry & biology* 18, 907-919.
- Pi, H., Helmann, J.D., 2017. Sequential induction of Fur-regulated genes in response to iron limitation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 114, 12785-12790.
- Pillai, S., Chellappan, S.P., 2015. ChIP on chip and ChIP-Seq assays: genome-wide analysis of transcription factor binding and histone modifications. *Methods in molecular biology (Clifton, N.J.)* 1288, 447-472.
- Pinu, F.R., Beale, D.J., Paten, A.M., Kouremenos, K., Swarup, S., Schirra, H.J., Wishart, D., 2019. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites* 9, 76.
- Pluháček, T.L., K; Ghosh, D; Milde, D; Novák, J; Havlíček, V.;, 2016. Characterization of microbial siderophores by mass spectrometry. *Mass Spectrom Rev.*
- Pohl, E., Holmes, R.K., Hol, W.G., 1998. Motion of the DNA-binding domain with respect to the core of the diphtheria toxin repressor (DtxR) revealed in the crystal structures of apo- and holo-DtxR. *J Biol Chem* 273, 22420-22427.
- Pohl, E., Holmes, R.K., Hol, W.G.J., 1999. Crystal Structure of the Iron-dependent Regulator (IdeR) from *Mycobacterium tuberculosis* Shows Both Metal Binding Sites Fully Occupied. *Journal of molecular biology* 285, 1145-1156.
- Prescott, J.F., 1991. *Rhodococcus equi*: an animal and human pathogen. *Clinical Microbiology Reviews* 4, 20-34.

- Putnam, M.L., Miller, M., 2006. Pathogenic Isolates of *Rhodococcus fascians* from New Hosts in the United States. *Plant Disease* 90, 526-526.
- Pym, A.S., Domenech, P., Honoré, N., Song, J., Deretic, V., Cole, S.T., 2001. Regulation of catalase-peroxidase (KatG) expression, isoniazid sensitivity and virulence by furA of *Mycobacterium tuberculosis*. *Molecular microbiology* 40, 879-889.
- Qiu, X., Verlinde, C.L.M.J., Zhang, S., Schmitt, M.P., Holmes, R.K., Hol, W.G.J., 1995. Three-dimensional structure of the diphtheria toxin repressor in complex with divalent cation co-repressors. *Structure* 3, 87-100.
- Que, Q., Helmann, J.D., 2000. Manganese homeostasis in *Bacillus subtilis* is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. *Molecular microbiology* 35, 1454-1468.
- Radwan, S.S., Sorkhoh, N.A., Fardoun, F., al-Hasan, R.H., 1995. Soil management enhancing hydrocarbon biodegradation in the polluted Kuwaiti desert. *Applied microbiology and biotechnology* 44, 265-270.
- Rane, M., Sayyed, R., Sb, C. 2005. Methods for Microbial Iron Chelator (Siderophore) Analysis, In, 475-492.
- Rantalainen, M., Cloarec, O., Beckonert, O., Wilson, I.D., Jackson, D., Tonge, R., Rowlinson, R., Rayner, S., Nickson, J., Wilkinson, R.W., Mills, J.D., Trygg, J., Nicholson, J.K., Holmes, E., 2006. Statistically integrated metabolomic-proteomic studies on a human prostate cancer xenograft model in mice. *Journal of proteome research* 5, 2642-2655.
- Ratledge, C., 2004. Iron, mycobacteria and tuberculosis. *Tuberculosis (Edinburgh, Scotland)* 84, 110-130.
- Ratledge, C., Dover, L.G., 2000. Iron metabolism in pathogenic bacteria. *Annual review of microbiology* 54, 881-941.
- Raymond, K.N., Allred, B.E., Sia, A.K., 2015. Coordination Chemistry of Microbial Iron Transport. *Accounts of Chemical Research* 48, 2496-2505.
- Ribbenstedt, A., Ziarrusta, H., Benskin, J.P., 2018. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLOS ONE* 13, e0207082.
- Riess, F.G., Elflein, M., Benk, M., Schiffler, B., Benz, R., Garton, N., Sutcliffe, I., 2003. The cell wall of the pathogenic bacterium *Rhodococcus equi* contains two channel-forming proteins with different properties. *Journal of bacteriology* 185, 2952-2960.
- Roberts, L.D., Souza, A.L., Gerszten, R.E., Clish, C.B., 2012. Targeted metabolomics. *Current protocols in molecular biology* Chapter 30, Unit30.32-30.32.24.
- Rochfort, S., 2005. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *Journal of natural products* 68, 1813-1820.
- Rodriguez, G.M., Smith, I., 2003. Mechanisms of iron regulation in mycobacteria: role in physiology and virulence. *Molecular microbiology* 47, 1485-1494.
- Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., 2002a. *ideR*, an Essential Gene in *Mycobacterium tuberculosis*: Role of IdeR in Iron-Dependent Gene Expression, Iron Metabolism, and Oxidative Stress Response. *Infection and immunity* 70, 3371-3381.
- Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., 2002b. *ideR*, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infection and immunity* 70, 3371-3381.
- Roosenberg, J.M., 2nd, Lin, Y.M., Lu, Y., Miller, M.J., 2000. Studies and syntheses of siderophores, microbial iron chelators, and analogs as potential drug delivery agents. *Current medicinal chemistry* 7, 159-197.

- Rose, A., Schraegle, S.J., Stahlberg, E.A., Meier, I., 2005. Coiled-coil protein composition of 22 proteomes--differences and common themes in subcellular infrastructure and traffic control. *BMC Evol Biol* 5, 66-66.
- Roxas, B.A.P., Li, Q., 2008. Significance analysis of microarray for relative quantitation of LC/MS data in proteomics. *BMC Bioinformatics* 9, 187.
- Rustad, T.R., Roberts, D.M., Liao, R.P., Sherman, D.R. 2009. Isolation of mycobacterial RNA, In: *Mycobacteria Protocols*. Springer, 13-22.
- Sanchez, R., Sali, A., 1997. Advances in comparative protein-structure modelling. *Current opinion in structural biology* 7, 206-214.
- Sangal, V., Goodfellow, M., Jones, A.L., Seviour, R.J., Sutcliffe, I.C. 2019. Refined Systematics of the Genus *Rhodococcus* Based on Whole Genome Analyses, In: Alvarez, H.M. (Ed.) *Biology of Rhodococcus*. Springer International Publishing, Cham, 1-21.
- Sangal, V., Jones, A.L., Goodfellow, M., Sutcliffe, I.C., Hoskisson, P.A., 2014. Comparative genomic analyses reveal a lack of a substantial signature of host adaptation in *Rhodococcus equi* ('*Prescottella equi*'). *Pathogens and Disease* 71, 352-356.
- Savini, V., Fazii, P., Favaro, M., Astolfi, D., Polilli, E., Pompilio, A., Vannucci, M., D'Amario, C., Di Bonaventura, G., Fontana, C., D'Antonio, D., 2012. Tuberculosis-like pneumonias by the aerobic actinomycetes *Rhodococcus*, *Tsukamurella* and *Gordonia*. *Microbes and infection / Institut Pasteur* 14, 401-410.
- Schaechter, M., 2009. *Encyclopedia of Microbiology*. Elsevier Science.
- Schmitt, M.P., Holmes, R.K., 1993. Analysis of diphtheria toxin repressor-operator interactions and characterization of a mutant repressor with decreased binding activity for divalent metals. *Molecular microbiology* 9, 173-181.
- Schmitt, M.P., Predich, M., Doukhan, L., Smith, I., Holmes, R.K., 1995. Characterization of an iron-dependent regulatory protein (IdeR) of *Mycobacterium tuberculosis* as a functional homolog of the diphtheria toxin repressor (DtxR) from *Corynebacterium diphtheriae*. *Infection and immunity* 63, 4284-4289.
- Shalon, D., Smith, S.J., Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research* 6, 639-645.
- Sibley, L.D., Franzblau, S.G., Krahenbuhl, J.L., 1987. Intracellular fate of *Mycobacterium leprae* in normal and activated mouse macrophages. *Infection and immunity* 55, 680-685.
- Silva, B., Faustino, P., 2015. An overview of molecular basis of iron metabolism regulation and the associated pathologies. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1852, 1347-1359.
- Singh, P.R., Bajaj, H., Benz, R., Winterhalter, M., Mahendran, K.R., 2015. Transport across the outer membrane porin of mycolic acid containing actinomycetales: *Nocardia farcinica*. *Biochim Biophys Acta* 1848, 654-661.
- Smith, A.T., Smith, K.P., Rosenzweig, A.C., 2014. Diversity of the metal-transporting P1B-type ATPases. *J Biol Inorg Chem* 19, 947-960.
- Smith, I., 2003. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews* 16, 463-496.
- Snow, G.A., White, A.J., 1969. Chemical and biological properties of mycobactins isolated from various mycobacteria. *Biochem J* 115, 1031-1050.
- Sritharan, M., 2016. Iron Homeostasis in *Mycobacterium tuberculosis*: Mechanistic Insights into Siderophore-Mediated Iron Uptake. *Journal of bacteriology* 198, 2399-2409.
- Staudenmaier, H., Van Hove, B., Yaraghi, Z., Braun, V., 1989. Nucleotide sequences of the fecBCDE genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. *J Bacteriol* 171, 2626-2633.

- Stojiljkovic, I., Baumler, A.J., Hantke, K., 1994. Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a fur titration assay. *Journal of molecular biology* 236, 531-545.
- Struyve, M., Moons, M., Tommassen, J., 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *Journal of molecular biology* 218, 141-148.
- Süssmuth, R.D.M., Andi, 2017. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition* 56, 3770-3821.
- Sutcliffe, I., Brown, A., Dover, L. 2010. The Rhodococcal Cell Envelope: Composition, Organisation and Biosynthesis, In: Alvarez, H.M. (Ed.) *Biology of Rhodococcus*. Springer Berlin Heidelberg, 29-71.
- Sweeney, C.R., Sweeney, R.W., Divers, T.J., 1987. *Rhodococcus equi* pneumonia in 48 foals: Response to antimicrobial therapy. *Veterinary microbiology* 14, 329-336.
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C.V., 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 47, D607-d613.
- Takai, S., Hines, S.A., Sekizaki, T., Nicholson, V.M., Alperin, D.A., Osaki, M., Takamatsu, D., Nakamura, M., Suzuki, K., Ogino, N., Kakuda, T., Dan, H., Prescott, J.F., 2000. DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. *Infection and immunity* 68, 6840-6847.
- Takai, S., Martens, R.J., Julian, A., Garcia Ribeiro, M., Rodrigues de Farias, M., Sasaki, Y., Inuzuka, K., Kakuda, T., Tsubaki, S., Prescott, J.F., 2003. Virulence of *Rhodococcus equi* Isolated from Cats and Dogs. *Journal of Clinical Microbiology* 41, 4468-4470.
- Tam, R., Saier, M.H., Jr., 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological reviews* 57, 320-346.
- Tan, C., Prescott, J.F., Patterson, M.C., Nicholson, V.M., 1995. Molecular characterization of a lipid-modified virulence-associated protein of *Rhodococcus equi* and its potential in protective immunity. *Can J Vet Res* 59, 51-59.
- Tang, J., 2011. Microbial metabolomics. *Curr Genomics* 12, 391-403.
- Tang, Q., Liu, Y.P., Yan, X.X., Liang, D.C., 2014. Structural and functional characterization of Cys4 zinc finger motif in the recombination mediator protein RecR. *DNA repair* 24, 10-14.
- Tao, X., Schiering, N., Zeng, H.Y., Ringe, D., Murphy, J.R., 1994. Iron, DtxR, and the regulation of diphtheria toxin expression. *Molecular microbiology* 14, 191-197.
- Taylor, K.M., Morgan, H.E., Johnson, A., Nicholson, R.I., 2005. Structure–function analysis of a novel member of the LIV-1 subfamily of zinc transporters, ZIP14. *FEBS Letters* 579, 427-432.
- Toledo-Arana, A., Solano, C., 2010. Deciphering the physiological blueprint of a bacterial cell: revelations of unanticipated complexity in transcriptome and proteome. *BioEssays : news and reviews in molecular, cellular and developmental biology* 32, 461-467.
- Tom-Yew, S.A.L., Cui, D.T., Bekker, E.G., Murphy, M.E.P., 2005. Anion-independent Iron Coordination by the *Campylobacter jejuni* Ferric Binding Protein. *Journal of Biological Chemistry* 280, 9283-9290.
- Topino, S., Galati, V., Grilli, E., Petrosillo, N., 2010. *Rhodococcus equi* infection in HIV-infected individuals: case reports and review of the literature. *AIDS patient care and STDs* 24, 211-222.
- Traoré, D.A.K., El Ghazouani, A., Ilango, S., Dupuy, J., Jacquamet, L., Ferrer, J.-L., Caux-Thang, C., Duarte, V., Latour, J.-M., 2006. Crystal structure of the apo-PerR-Zn protein from *Bacillus subtilis*. *Molecular microbiology* 61, 1211-1219.
- Tsirigos, K.D., Peters, C., Shu, N., Kall, L., Elofsson, A., 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic acids research* 43, W401-407.

- Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* 98, 5116-5121.
- Valero-Rello, A., Hapeshi, A., Anastasi, E., Alvarez, S., Scotti, M., Meijer, W.G., MacArthur, I., Vazquez-Boland, J.A., 2015. An invertron-like linear plasmid mediates intracellular survival and virulence in bovine isolates of *Rhodococcus equi*. *Infection and immunity*.
- van der Geize, R., de Jong, W., Hessels, G.I., Grommen, A.W.F., Jacobs, A.A.C., Dijkhuizen, L., 2008. A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. *Nucleic acids research* 36, e151-e151.
- Vazquez-Boland, J.A., Giguere, S., Hapeshi, A., MacArthur, I., Anastasi, E., Valero-Rello, A., 2013. *Rhodococcus equi*: the many facets of a pathogenic actinomycete. *Veterinary microbiology* 167, 9-33.
- Vázquez-Boland, J.A., Meijer, W.G., 2019. The pathogenic actinobacterium *Rhodococcus equi*: what's in a name? *Molecular microbiology* 112, 1-15.
- Vazquez-Boland, J.A., Prescott, J.F., Meijer, W.G., Leadon, P., Hines, S.A., 2009. *Rhodococcus equi* comes of age. *Equine Veterinary Journal* 41, 93-95.
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial Analysis of Gene Expression. *Science* 270, 484-487.
- Vellore, J.M., 2001. Iron Acquisition in *Rhodococcus erythropolis*: the Isolation of Mutant(s) that Do Not Produce a Siderophore. . East Tennessee State University, ELECTRONIC THESES AND DISSERTATIONS.
- Venkataraman, B., Gupta, N., Gupta, A., 2013. A robust and efficient method for the isolation of DNA-free, pure and intact RNA from *Mycobacterium tuberculosis*. *Journal of microbiological methods* 93, 198-202.
- Verma, D., Das, L., Gambhir, V., Dikshit, K.L., Varshney, G.C., 2015. Heterogeneity among Homologs of Cutinase-Like Protein Cut5 in *Mycobacteria*. *PLOS ONE* 10, e0133186.
- Vermeulen, J., De Preter, K., Lefever, S., Nuytens, J., De Vloed, F., Derveaux, S., Hellemans, J., Speleman, F., Vandesompele, J., 2011. Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic acids research* 39, e63.
- Vishniac, W., Santer, M., 1957. The thiobacilli. *Bacteriological reviews* 21, 195-213.
- Vissa, V.D., Brennan, P.J., 2001. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome biology* 2, Reviews1023.
- Vivancos, A.P., Güell, M., Dohm, J.C., Serrano, L., Himmelbauer, H., 2010. Strand-specific deep sequencing of the transcriptome. *Genome research* 20, 989-999.
- von Bargen, K., Polidori, M., Becken, U., Huth, G., Prescott, J.F., Haas, A., 2009. *Rhodococcus equi* Virulence-Associated Protein A Is Required for Diversion of Phagosome Biogenesis but Not for Cytotoxicity. *Infection and immunity* 77, 5676-5681.
- Waldron, K.J., Rutherford, J.C., Ford, D., Robinson, N.J., 2009. Metalloproteins and metal sensing. *Nature* 460, 823-830.
- Wandersman, C., Delepelaire, P., 2004. Bacterial iron sources: from siderophores to hemophores. *Annual review of microbiology* 58, 611-647.
- Wanger, A., Chavez, V., Huang, R.S.P., Wahed, A., Actor, J.K., Dasgupta, A. 2017. Chapter 5 - Biochemical Tests and Staining Techniques for Microbial Identification, In: Wanger, A., Chavez, V., Huang, R.S.P., Wahed, A., Actor, J.K., Dasgupta, A. (Eds.) *Microbiology and Molecular Diagnosis in Pathology*. Elsevier, 61-73.
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H., 2015. antiSMASH 3.0-a

- comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* 43, W237-W243.
- Weinstock, D.M., Brown, A.E., 2002. *Rhodococcus equi*: An Emerging Pathogen. *Clinical Infectious Diseases* 34, 1379-1385.
- Wierling, C., Herwig, R., Lehrach, H., 2007. Resources, standards and tools for systems biology. *Briefings in functional genomics & proteomics* 6, 240-251.
- Wilkens, S., 2015. Structure and mechanism of ABC transporters. *F1000Prime Reports* 7, 14.
- Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L., 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & genetic engineering reviews* 13, 19-50.
- Winkelstroter, L.K., Dolan, S.K., Fernanda Dos Reis, T., Bom, V.L., Alves de Castro, P., Hagiwara, D., Alowni, R., Jones, G.W., Doyle, S., Brown, N.A., Goldman, G.H., 2015. Systematic Global Analysis of Genes Encoding Protein Phosphatases in *Aspergillus fumigatus*. *G3 (Bethesda, Md.)* 5, 1525-1539.
- Xia, J., Psychogios, N., Young, N., Wishart, D.S., 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research* 37, W652-W660.
- Xia, J., Wishart, D.S., 2016. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Current protocols in bioinformatics* 55, 14.10.11-14.10.91.
- Yamada, H., Kobayashi, M., 1996. Nitrile hydratase and its application to industrial production of acrylamide. *Bioscience, biotechnology, and biochemistry* 60, 1391-1400.
- Yeruva, V.C., Duggirala, S., Lakshmi, V., Kolarich, D., Altmann, F., Sritharan, M., 2006. Identification and Characterization of a Major Cell Wall-Associated Iron-Regulated Envelope Protein (Irep-28) in *Mycobacterium tuberculosis*. *Clinical and Vaccine Immunology* 13, 1137-1142.
- Young, D., Robertson, B., 2001. Genomics: Leprosy — a degenerative disease of the genome. *Current Biology* 11, R381-R383.
- Yugi, K., Kubota, H., Hatano, A., Kuroda, S., 2016. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers. *Trends in biotechnology* 34, 276-290.
- Zahrt, T.C., Song, J., Siple, J., Deretic, V., 2001. Mycobacterial FurA is a negative regulator of catalase-peroxidase gene katG. *Molecular microbiology* 39, 1174-1185.
- Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., Yates, J.R., 3rd, 2013. Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* 113, 2343-2394.
- Zhao, X., Yu, H., Yu, S., Wang, F., Sacchettini, J.C., Magliozzo, R.S., 2006. Hydrogen Peroxide-Mediated Isoniazid Activation Catalyzed by *Mycobacterium tuberculosis* Catalase-Peroxidase (KatG) and Its S315T Mutant. *Biochemistry* 45, 4131-4140.
- Zhi, X.Y., Li, W.J., Stackebrandt, E., 2009. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *International journal of systematic and evolutionary microbiology* 59, 589-608.
- Zhu, W., Arceneaux, J.E., Beggs, M.L., Byers, B.R., Eisenach, K.D., Lundrigan, M.D., 1998. Exochelin genes in *Mycobacterium smegmatis*: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. *Molecular microbiology* 29, 629-639.
- Zondervan, N.A., Van Dam, J.C.J., Schaap, P.J., Martins dos Santos, V.A.P., Suarez-Diez, M., 2018. Regulation of Three Virulence Strategies of *Mycobacterium tuberculosis*: A Success Story. *International Journal of Molecular Sciences* 19, 347.
- Zopf, W., 1891. 4. W. Zopf: Ueber Ausscheidung von Fettfarbstoffen (Lipochromen) seitens gewisser Spaltpilze. *Berichte der Deutschen Botanischen Gesellschaft* 9, 22-28.

