

Identification and characterization of the
biological mechanisms of metaldehyde
degradation

Edward David Fuller

Doctor of Philosophy

University of York

Biology

March 2021

Abstract

Metaldehyde is a molluscicide used to protect agriculture and domestic crops. Following its application, metaldehyde has been demonstrated to wash-off into nearby water sources and as such has been detected above the EU statutory drinking water limit of 0.1 µg/L. Due to metaldehyde's chemical properties, current water treatment processes are economically unsustainable. Prior to work conducted within this thesis, several metaldehyde degrading bacteria had been characterised. However, the degradative enzymes responsible were unknown. The work provided here involved the discovery and verification of the first metaldehyde degrading enzymes. Random chemical mutagenesis of metaldehyde-degrading strain *Acinetobacter calcoaceticus* E1 using ethylmethanesulphonate led to the isolation of four mutants deficient in metaldehyde degradation. Comparative genomic analysis of the mutants resulted in the discovery of a gene cluster hypothesized to be responsible for metaldehyde degradation. This cluster contained a predicted Fe (II)/(alpha) ketoglutarate (αKG)-dependent dioxygenase (MahX), a lyase (MahY) and an aldehyde dehydrogenase (MahZ). Heterologous expression within *Escherichia coli* revealed MahX as the initial metaldehyde catalysing enzyme. Bioinformatic analysis based on MahX protein sequence allowed for the identification of a novel metaldehyde degrading protein MahS within the metaldehyde degrading strain *Sphingobium* sp CMET-H. Genomic analysis revealed *mahS* to be located within a 194 kbp conjugative plasmid. Protein characterization of MahX and MahS demonstrated catabolic activity of the enzymes *in vitro* and revealed the requirements for cofactors Fe²⁺, αKG and a reducing agent to achieve optimum degradation. This research expanded the knowledge regarding biological degradation of metaldehyde and provides the basis for targeted bioremediation and biomonitoring approaches.

Table of contents

Contents

Abstract.....	2
Table of contents	3
Table of figures	8
Table of tables	15
Acknowledgments.....	16
Author's Declaration	17
Chapter 1: General Introduction	18
1.1 Metaldehyde	18
1.2 Metaldehyde detection methods	24
1.3 Non-biological metaldehyde reduction and removal methods	26
1.4 Biotic degradation of metaldehyde	31
1.5 Protein evolution	36
1.6 Gene mobility within the environment	38
1.7 Bioremediation, Biosensing and Biomonitoring	40

1.71 Bioremediation	40
1.72 Detection through biological methods	43
1.73 Biomonitoring	45
1.8 Aims of the Project	47
Chapter 2: Materials and Methods	49
2.1 Microbiological Methods	49
2.11 Growth media	49
2.2 Molecular Methods	53
2.21 Agarose gel electrophoresis	53
2.22 Polymerase Chain Reaction	54
2.23 Primers	56
2.24 Restriction Digest	57
2.25 DNA Purification	57
2.26 Cloning methods	58
2.27 Preparation of chemically competent <i>E. coli</i> DH5 α and BL21 (DE3)	59
2.28 Heat shock transformation of competent <i>E. coli</i>	59

2.29 Whole genome sequencing	60
2.210 Sanger sequencing	60
2.3 Bioinformatic analysis	60
2.31 Basic Local Alignment Search Tool (BLAST) methods	60
2.32 Comparative genomic through BLAST	61
2.33 Phylogenetic analysis	61
2.34 Genomic analysis	62
2.4 Protein analysis	63
2.5 Metaldehyde degradation assays	68
Chapter 3: Identification and characterization of the gene(s) responsible for the initial	
step of metaldehyde degradation within <i>A. calcoaceticus</i> E1	74
3.1 Introduction	74
3.2 Results	81
3.21 Using chemical mutagenesis to obtain mutants lacking the ability to degrade	
metaldehyde	81
3.22 Bioinformatic analysis of <i>A. calcoaceticus</i> E1 mutants deficient in metaldehyde	
degradation	87

3.23 Identification of the metaldehyde degrading gene(s) through BLAST analysis	98
3.24 Characterization of the identified metaldehyde degrading gene	99
3.25 Examination of the genomic locations of the metaldehyde degrading genes	108
3.26 Identification of the metaldehyde degrading gene(s) through heterologous expression	112
3.27 Observing the effects of metaldehyde on mahX constructs	122
3.3 Discussion	124
Chapter 4: Characterisation of the metaldehyde degrading protein MahX	131
4.1 Introduction	131
4.2 Results	135
4.21 Bioinformatics characterization of MahX	135
4.22 Overexpression and characterization of MahX	144
4.23 <i>In vitro</i> characterization of MahX	161
4.3 Discussion	168
Chapter 5: Identification and characterization of novel metaldehyde degrading genes	173
5.1 Introduction	173
5.2 Results	176

5.21 <i>Sphingobium sp</i> CMET-H bioinformatics analysis	176
5.21 Protein modelling for MahS	178
5.22 Genomic location of <i>mahS</i> within <i>Sphingobium sp</i> CMET-H	186
5.23 <i>R. globerulus</i> HNO-A bioinformatics analysis.....	190
5.24 Protein modelling for MahR	192
5.25 Genomic location of <i>mahR</i> within <i>R. globerulus</i> HNO-A's	198
5.26 Phylogenetic analysis of MahS and MahR	200
5.27 Experimental verification of the bioinformatics analysis	203
5.3 Discussion	212
Chapter 6: Final Discussion	219
Reference list	224
Appendix	256

Table of figures

Figure 1-1. Structure of metaldehyde.....	19
Figure 1-2. Metaldehyde related water quality failures	20
Figure 1-3. Tons of metaldehyde sold within the EU 2019	22
Figure 1-4. Degradation and growth curve of isolated metaldehyde degrading microorganisms	35
Figure 2-1. Gas chromatography metaldehyde calibration curve.....	74
Figure 3-1 Growth assay of mutants deficient in metaldehyde degradation	88
Figure 3-2. Degradation assay of mutants deficient in metaldehyde degradation	89
Figure 3-3. Metaldehyde degradation assay with pregrown cells	91
Figure 3-4. BLAST alignment of MahX.....	96
Figure 3-5. InterPro analysis of MahX.....	97
Figure 3-6. Phylogenetic tree of MahX	100
Figure 3-7. BLAST alignment of MahY.....	99
Figure 3-8. Putative conserved domain of MahY	100

Figure 3-9. Phylogenetic tree of MahY	101
Figure 3-10. BLAST alignment of MahZ	102
Figure 3-11. Putative conserved domain of MahZ	102
Figure 3-12. Predicted domains of MahZ	103
Figure 3-13. Phylogenetic tree of MahZ	104
Figure 3-14. Schematic of catabolic contig	105
Figure 3-15. BLAST alignment of catabolic contig	107
Figure 3-16. BLAST alignment of catabolic contig from <i>Acinetobacter calcoaceticus</i> E1 WT against mutants.....	108
Figure 3-17. Schematic of mutation within catabolic contig	109
Figure 3 18. Schematic of vector pBR322	110
Figure 3 19. Schematic of vector pBR322 with <i>mahX</i>	110
Figure 3 20 Schematic of vector pBR322 with <i>mahY</i>	111
Figure 3-21. Schematic of vector pBR322 with <i>mahXY</i>	111
Figure 3-22. Schematic of vector pBR322 with <i>mahXYZ</i>	112
Figure 3-23. Agarose gel of <i>EcoR1</i> linearised pBR322	113

Figure 3-24 Agarose gel of PCR amplified <i>mahX</i> , <i>mahXY</i> and <i>mahXYZ</i>	114
Figure 3-25 Agarose gel of PCR amplified <i>promoter region</i> and <i>mahY</i>	115
Figure 3-26 Growth assay of <i>mahX</i> transformants	116
Figure 3-27 Final concentration metaldehyde degradation assay of <i>mah</i> transformants	117
Figure 3-28 Metaldehyde degradation over time with <i>mahX</i> transformants	118
Figure 3-29 Metaldehyde inhibition growth assay	120
Figure 3-30 Schematic of 1,4 dioxane and tetrahydrofuran degradation pathway	124
Figure 3-31 Schematic of metaldehyde degradation pathway	125
Figure 4-1. Schematic for the mechanism reaction for alpha ketoglutarate dependent dioxygenases	130
Figure 4-2. Computer ribbon model of MahX	133
Figure 4-3. Computer model of alpha ketoglutarate binding sites within MahX	135
Figure 4-5. Computer model of Fe ²⁺ binding sites within MahX	136
Figure 4-4. Multiple sequence alignment of MahX and highly similar sequences	137
Figure 4 6. Computer model of surface of MahX with ligand binding sies	138
Figure 4 6. Superposition model of MahX and 5DAP1.....	140

Figure 4-8. Schematic of pET21a.....	141
Figure 4 9. SDS-PAGE gel of MahX induction	142
Figure 4-10. SDS-PAGE of soluble and insoluble extract of <i>Escherichia coli</i> expressing MahX ...	143
Figure 4-11. SDS-PAGE of soluble extract at various induction time points in <i>Escherichia coli</i> expressing MahX	144
Figure 4-12. SDS-PAGE of soluble extracts following induction at various time points and temperatures	145
Figure 4-13. IMAC purified of His-tagged MahX	146
Figure -14. Bioinformatics solubility analysis	148
Figure 4-15. <i>In vivo</i> metaldehyde degradation assay of <i>E. coli</i> overexpressing MahX	149
Figure 4-16. Schematic of pETFPP_2 MBP-MahX	150
Figure 4-17. Agarose gel following <i>mahX</i> amplification	151
Figure 4-18. Agarose gel following colony PCR following <i>mahX</i> insertion into pETFPP_2	152
Figure 419. SDS-PAGE of MBP-MahX following induction	153
Figure 4-20. SDS-PAGE of MBP-MahX with various induction time points	154
Figure 4-21. SDS-PAGE of initial MBP-MahX solubility assay	155
Figure 4-22. SDS-PAGE of soluble MBP-MahX following various induction time points	156

Figure 4-23. <i>In vivo</i> metaldehyde degradation assay with <i>Escherichia coli</i> expressing MBP-MahX.....	157
Figure 4 24. Metaldehyde degradation assay with soluble crude extract	158
Figure 4 25. Metaldehyde degradation cofactor assay	159
Figure 4 26. Reducing agents effect on metaldehyde degradation assay	160
Figure 4-27. IMAC MBP-MahX purification	161
Figure 4-28. Purified MBP-MahX metaldehyde degradation assay	162
Figure 4-29. Amylose purification of MBP-MahX	163
Figure 4-30. Metaldehyde degradation assay of amylose purified MBP-MahX	164
Figure 5-1. BLAST alignment of MahX against <i>Sphingobium</i> sp. CMET-H	172
Figure 5-2. BLAST alignment MahS against <i>Acinetobacter calcoaceticus</i> E1	173
Figure 5-3. BLAST alignment of MahS	174
Figure 5-4. Putative conserved region of MahS	174
Figure 5-5. Computer generated model of MahS	176
Figure 5-6. Computer model of alpha ketoglutarate binding sites within MahS	179
Figure 5-7. Computer model of Fe ²⁺ binding sites within MahS	180
Figure 5-8. Surface fill model of MahS with ligand binding sites	181

Figure 5-9. Schematic of the genomic location of *mahS* 182

Figure 5-10. BLAST alignment of *mahS* catabolic contig 184

Figure 5-11. Graphic depiction of identified plasmid sequences within *Sphingobium* sp.CMET-H
with type 4 section system annotations 185

Figure 5-12. BLAST alignment of MahX against the genome of <i>Rhodococcus globerulus</i> HNO-A	186
Figure 5-13. BLAST alignment of MahR against the genome of <i>Acinetobacter calcoaceticus</i> E1	187
Figure 5-14. BLAST alignment of MahR against non-redundant protein database	188
Figure 5-15. Putative conserved domains of MahR	188
Figure 5-16. Computer model of MahR	189
Figure 5-17. Computer model of alpha ketoglutarate binding sites within MahR	191
Figure 5-18. Computer model of Fe ²⁺ binding sites within MahR	192
Figure 5-19. Surface fill model of MahR with ligand binding sites	193
Figure 5-20. Schematic of <i>mahR</i> genomic location	194
Figure 5-21. Phylogenetic tree of MahX, MahS and MahR	197
Figure 5-22. Multiple sequence alignment of MahX, MahR and MahS	197
Figure 5-23. Agarose gel following PCR amplification of <i>mahS</i>	201
Figure 5-24. Agarose gel following PCR amplification of <i>mahR</i>	202
Figure 5-25. SDS-PAGE following the overexpression of MBP-MahS	203
Figure 5-26. SDS-PAGE following the overexpression of MBP-MahR	204
Figure 5-27. <i>In vivo</i> metaldehyde degradation assay of <i>Escherichia coli</i> expressing MBP-MahS and MBP-MahR	205

Figure 5-28. SDS-PAGE gel of the soluble extract of <i>Escherichia coli</i> expressing MBP-MahS	207
Figure 5-29. Metaldehyde degradation assay with soluble crude extract of <i>Escherichia coli</i> expressing MBP-MahS	208
Figure 6-1: Predicted metaldehyde degradation pathway for MahS.....	217

List of tables

Table 1-1 Chemical and physical properties of metaldehyde	18
Table 2-1. Trace element solution concentration	51
Table 2-2. Thermal cycling conditions for GoTaq PCR.....	57
Table 2-3. Thermal cycling conditions for High Fidelity Phusion PCR.....	58
Table 2-4 Primers used throughout the investigation	59
Table 2-5 Plasmids used throughout the investigation	64
Table 2-6 Strains used throughout the investigation	65
Table 3-1. Whole genome sequencing quality statistics	92
Table 3-2. Comparative analysis of metaldehyde mutants.....	93
Table 3-3 <i>Acinetobacter calcoaceticus</i> E1 mutants similarity analysis	98
Table 3-4 ISfinder for 5' region of <i>mahX</i> catabolic contig	106
Table 3-5 ISfinder for 3' region of <i>mahX</i> catabolic contig	106
Table 3-6 Promoter identification analysis for <i>mahX</i>	108
Table 4-1. PDB similarity results based on computer generated model of MahX	134
Table 4-2- Gene ontology terms associated with MahX model	139

Table 5-1. PDB similarity results based on computer generated model of MahS.....	177
Table 5-2. PDB similarity results based on computer generated model of MahR	190
Table6-1. Affinity table of characterised Fe (II)/ (alpha) ketoglutarate-dependent oxygenases.....	219

Acknowledgments

I would like to thank my supervisors Prof James Moir, Dr Steve Johnson and Dr Chris Sinclair for all their help and advice with this PhD. I would also like to thank my wonderful friends and colleagues, whom without this would have not been possible. I am also deeply grateful for all the support I received from my family. Finally, I would like to thank NERC and Thames Water for the studentship that allowed me to conduct this thesis.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Edward David Fuller

30th March 2021

Chapter 3:

A modified version of 3.1 has been approved for publication as part of the review chapter :

Fuller, E., Castro-Gutiérrez, V., Cambronero-Heinrichs, J.C., Rodriguez-Rodríguez, C.E. 2020. Using Molecular Methods to Identify and Monitor Xenobiotic-degrading Genes for Bioremediation. In: Bidoia, E., Montagnolli, R. Biodegradation, Pollutants and Bioremediation Principles. Taylor and Francis Group LLC.

A modified version of 3.2 has been published in a peer reviewed journal paper:

Castro-Gutiérrez, V., Fuller, E., Thomas, J.C., Sinclair, C.J., Johnson, S., Helgason, T., Moir, J.W.B., 2020. Genomic basis for pesticide degradation revealed by selection, isolation and characterisation of a library of metaldehyde-degrading strains from soil. *Soil Biology and Biochemistry* 142.

doi:10.1016/j.soilbio.2019.107702.

Chapter 1: General Introduction

1.1 Metaldehyde

Metaldehyde is a molluscicide used to protect a wide range of agricultural, horticultural and domestic crops. It is a synthetic compound with a formula of $(\text{CH}_3\text{CHO})_4$ that is manufactured through the polymerization of four acetaldehyde molecules under cold and acidic conditions (Pauling and Carpenter, 1936). The synthesis of metaldehyde yields around ~ 7-15 % product with the remaining acetaldehyde converted into the liquid cyclic trimer paraldehyde (Eckert *et al.*, 2006). The chemical and physical information regarding this compound is located within table 1-1. Due to polar organic nature of metaldehyde, it is described as moderately soluble within water and highly soluble within methanol (Lewis *et al.*, 2016). The cyclic carbon structure as shown in Figure 1-1 provides metaldehyde with hydro and photolytic stability .

Table 1-1: The chemical and physical properties of metaldehyde (Lewis et al., 2016).

Molecular Mass	176.21 g/ mol
IUPAC Name	2,4,6,8-tetramethyl-1,3,5,7-tetraoxocane
Boiling Point	112 to 115 °C
Flash Point	36 to 40 °C
Water Solubility	0.188 g/L at 20 °C
Methanol Solubility	1.73 g/L at 20 °C
Log octanol / water partition coefficient (K_{ow})	0.12 at 20 °C
Physical Description	A white coloured crystalline solid

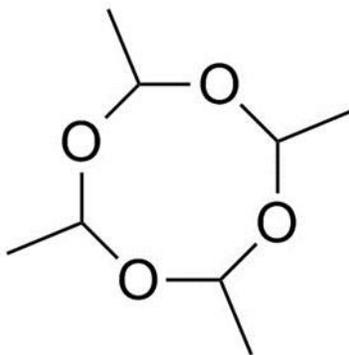


Figure 1-1: Sketetal structure of metaldehyde.

Originally developed in 1872, metaldehyde was first employed as a firelighter under the trade name of 'Meta-Fuel'. Due to its solid state and therefore ease of portability, the sublimation into acetaldehyde was utilized and led to the replacement of methylated spirits (Drury *et al.*, 1939). Despite its uses within the camping industry, metaldehyde is best known for its properties as a molluscicide.

Using spinning disc applicators, metaldehyde is applied to land in pellet form consisting of 1.5, 3.0 and 4.0 % weight (NRoSOS, 2012). The chemical and physical properties of metaldehyde allow for a highly mobile soil compound under moist conditions. As molluscs thrive in wet weather conditions, autumn/ winter is typically the season of greatest metaldehyde application. As such, short-term fluctuations of metaldehyde are reported in surface waters following rainfall runoff events (Castle *et al.*, 2018).

To ensure high quality water and long-term protection of water sources, the European Union Council Directive 98/83/EC set a blanket regulatory limit for individual and total pesticides to not exceed 0.1 µg/L and 0.5 µg/L respectively (Dolan *et al.*, 2013). Advancements in metaldehyde detection and monitoring technologies led to the first reported case of metaldehyde exceeding this limit in 2007 (Pendergrast, 2012). Following this incident, metaldehyde has since been detected above regulatory levels in reservoirs, watercourses and processed water (Mohamad Ibrahim *et al.*, 2019). As EU regulation states that water companies have a legal responsibility to both monitor and remove metaldehyde from drinking water, current water companies are under significant pressure to ensure they meet these standards. Despite the significant drop in both total contamination and metaldehyde breaches over the past 5 years, as shown in Figure 1-2, metaldehyde still accounts for the largest cause of pesticide breaches within England. As of 2019, the Drinking Water Inspectorate (DWI) identified metaldehyde as accounting for 50 % of pesticide contaminated water failures (Drinking Water Inspectorate, 2019).

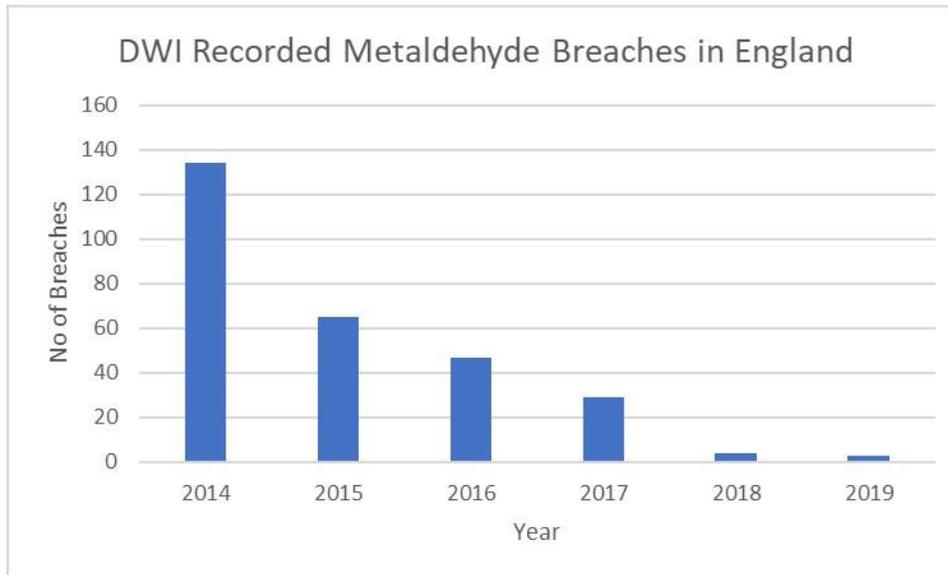


Figure 1-2: Annual metaldehyde breaches recorded by DWI within England from 2014 to 2019. Data obtained from <http://dwi.gov.uk/about/annual-report>.

Literature analysis revealed no readily available information regarding the presence of metaldehyde in drinking water within other EU countries. Although the concentrations of metaldehyde are not known, the usage can be inferred through both the amount of the molluscicide sold shown in Figure 2-2 and the alarming increase in metaldehyde based poisonings in Italy, Belgium and France (Berny *et al.*, 2010; De Roma *et al.*, 2017). Based on the above evidence and that arable crops are grown on 24 % of EU lands (European Commission, 2020), the metaldehyde contamination issue is likely to be found readily within numerous countries' water supplies. Evidence suggests that other countries, such as USA and Canada (Environment Agency, 2018; Government of Canada, 2020), readily utilize metaldehyde however as their regulatory limits are based on alternative criteria such as toxicity and environmental accumulation, they are unlikely to test or monitor the compound as extensively.

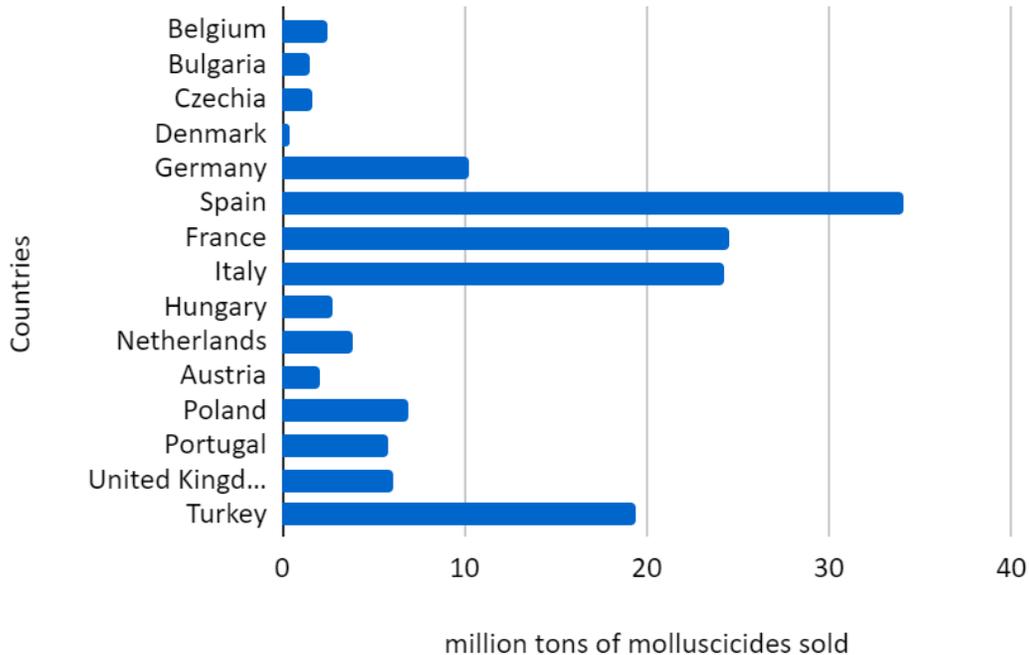


Figure 1-3 : Molluscicide sold within EU countries in 2019 (Eurostat,2021).

Within the acidic conditions of the mollusc's anterior digestive system, metaldehyde is depolymerised into acetaldehyde. The presence of acetaldehyde leads to damage and alterations to the mucocytes within the gut region leading to excess mucus secretion. This ultimately leads to the dehydration and eventual death of the pest (Triebkorn *et al.*, 1998). Within soil organisms with less acidic gut conditions (such as earthworms), metaldehyde appears to demonstrate no observable toxicity (Carritt and Kanwisher, 1959; Asfaw *et al.*, 2018; Wei *et al.*, 2019). To attract slugs and snails, metaldehyde is mixed or coated with bran or molasses. Although this approach creates a highly attractive and effective molluscicide, it also leads to consumption of the pesticide by nontargeted wildlife such as birds and mammals (Dolder and Volmer, 2003). Despite attempting to make the product unattractive to wildlife through colouring the product blue, numerous studies have observed toxicity of the compound within

both wild and domestic animals (Andreasen, 1993; Mineau *et al.*, 2001; Bieri, 2003). This is especially well recorded with companion animals, as metaldehyde is the second leading cause of canine poisoning within the UK (De Roma *et al.*, 2017; Mohamad Ibrahim *et al.*, 2019). Despite the lethal oral dose of metaldehyde to be 100 mg/kg in dogs, severe effects of ingestion can occur at significantly lower levels (Dolder and Volmer, 2003).

Numerous incidences of human toxicity brought on by metaldehyde consumption have also been reported. Depending on the amount of metaldehyde ingested, symptoms include drowsiness, convulsions and even death (Drury *et al.*, 1939; Cope *et al.*, 2006).

As other terrestrial invertebrates share the soil environments with snails and slugs, they are also exposed to the same environmental contaminants. As soil organisms, such as earthworms, cause no harm to crop yield and may convey beneficial properties to the soil, the effect on the greater ecosystem can be damaged through their increased mortality (Van Groenigen *et al.*, 2014). Therefore, various studies have been conducted to examine the effect of metaldehyde on numerous species of earthworm. Exposure of high concentrations of metaldehyde revealed negligible negative effects on earthworms tested. The results suggest that metaldehyde appears significantly less toxic than those of the alternative molluscicides iron phosphate and methiocarb (Langan and Shaw, 2006; Edwards *et al.*, 2009). As metaldehyde is frequently found within nearby water sources, through field run off, the effects on aquatic life are of great environmental importance. As such, numerous studies have been undertaken to determine what role the pesticide has on aquatic organisms such as non-targeted gastropods, pacific oysters, rainbow trout and water fleas (Environmental Protection Agency, 2006; Moreau *et al.*, 2015; Hallett *et al.*, 2016). Gastropod analysis revealed limited differences between metaldehyde

affected and control sites. Where laboratory studies were undertaken to determine tolerance to metaldehyde, concentrations significantly higher than those historically found within the environment demonstrated a reduction in survivorship (Hallett *et al.*, 2016). Such gastropod research has been verified through several studies and have all demonstrated no significant effect on gastropod mortality at environmental concentrations (Borlongan *et al.*, 1996; Coloso *et al.*, 1998). The leaching of metaldehyde into seawater and the effect this had on marine organisms was examined using Pacific oysters. Through exposing the oysters to the environmental concentration of metaldehyde (0.1 µg/L), significant reduction in the number of granulocytes and decreased phagocytic activity was observed. As such, the data suggests exposure to environmental levels of metaldehyde potentially increase Pacific oyster vulnerability to infectious agents (Moreau *et al.*, 2015). Studies conducted by the environmental protection agency demonstrated no significant toxicity upon acute exposure of metaldehyde to rainbow trout or water fleas (Environmental Protection Agency, 2006). As mortality rate based on acute exposure was the only criterion investigated, non-lethal harm caused by metaldehyde to these organisms would not have been identified.

Following advice from both the UK expert committee on pesticides and the Health and Safety Executive, the UK government will introduce a ban on metaldehyde application onto agricultural and domestic soils. This ban, commencing spring 2022, is due to the high environmental risk of metaldehyde application especially to birds and mammals (DEFRA, 2020). Currently, no ban is suggested within the EU where metaldehyde application and monitoring is still ongoing.

1.2 Metaldehyde detection methods

Recent technological advances have allowed for frequent and sensitive identification and monitoring of metaldehyde within surface and potable water. Since its first identification in surface waters in 2007, water companies have been under increasing pressure to remove the contaminant. Rapid and accurate detection of the molluscicide can be used to monitor concentrations over various locations and time points; it can also be utilized to prevent abstraction of contaminated water. As such, numerous studies have been undertaken to further optimize and improve these detection methods (Li *et al.*, 2010; Owlstone, 2015; Maher *et al.*, 2016; Schumacher *et al.*, 2016).

Although high levels of metaldehyde can be detected through numerous approaches, gas chromatography (GC) coupled with mass spectrometry (MS) or liquid chromatography (LC) with MS are typically employed for real world sample testing (Selim and Seiber, 1973; Alder *et al.*, 2006). Both GC MS and LC MS demonstrate appropriate sensitivity to be able to detect the regulatory limit (0.1 µg/L) and have been utilized at surface water testing sites and can allow for both online and offline detection of contamination (Iwata *et al.*, 1982; Paolo and Renzo, 1983).

GC analysis of metaldehyde can allow for cost effective, reliable and accurate monitoring to be performed. However, as aqueous solutions cannot be directly injected into the columns, significant sample preparation is required. Where lower limits of detection of metaldehyde need to be analyzed, concentration methods such as liquid-liquid extraction or solid phase extraction columns are often utilized (Castle *et al.*, 2019).

Unlike GC analysis, LC environmental samples can be directly injected within the instrument. As well as reduced sample preparation, when utilized as a triple quadrupole (LC/MS/MS) detection system metaldehyde detection can reach lower limits of low ng/L. Although significantly more expensive than alternative detection methods, LC-MS also allows for a wide range of analytes to be detected within the sample allowing numerous environmental contaminants to be co-detected (Alder *et al.*, 2006).

As the limitations of the above methods are their lack of portability, intensive sample preparation time and analysis time, an investigation undertaken by Maher *et al* (2016) led to the development of a metaldehyde detection method through reactive paper spray mass spectrometry. This approach utilized a lightweight and portable device, requiring little sample preparation and obtaining quantification results down to 0.05 ng/L. Despite the portability claim for the device, the weight and power requirements confine the testing to vehicles. Further work is required to demonstrate detection of metaldehyde within real world samples. Another recently developed metaldehyde detection instrument is the LONESTARTM (Owlstone, 2015). This device is highly portable and is capable of detecting metaldehyde below the regulatory limit. The Field Asymmetric Ion Mobility Spectrometry (FAIMS) device depolymerizes metaldehyde into four acetaldehyde molecules with the use of nitric acid. The acetaldehyde is then calculated based on

the area under the appropriate ion current peak. Despite requiring only 15 minutes, this approach requires significant sample preparation for portable field use and is best utilized as an online alarm system for metaldehyde contamination.

1.3 Non-biological metaldehyde reduction and removal methods

Catchment scale preventative approaches to metaldehyde contamination have been demonstrated to help avoid high energy and high cost engineering treatment solutions (Mohamad Ibrahim *et al.*, 2019). Current catchment initiatives, which seek to significantly reduce metaldehyde contamination within drinking water, include molluscicide substitution, education and preventing abstraction, amongst others. Through initiatives such as metaldehyde stewardship group's 'Get Pelletwise' and Anglian Water's 'Slug It Out', agricultural users were educated and advised in the correct storage, management and application of the pesticide (Anglian Water, 2019; Metaldehyde Stewardship Group, 2020). Both campaigns were considered to be highly successful in the reduction of metaldehyde use (Drinking Water Inspectorate, 2019; Mohamad Ibrahim *et al.*, 2019).

Metaldehyde is the most commonly used molluscicide in the UK, due to its relatively low cost and demonstrated efficacy as a molluscicide. Molluscicide importance to the agricultural industry is

such that its lack of application is predicted to cost the UK agricultural industry ~ £43.5 million (Nicholls, 2014). Therefore, suitable and sustainable alternative molluscicide compounds are required in order to protect arable crops in the event of prohibiting the use of the pesticide. The two most frequently used alternatives are methiocarb and ferric phosphate (Glen and Orsman, 1986; Edwards *et al.*, 2009). Methiocarb ($C_{11}H_{15}NO_2S$) is an insecticide, molluscicide and bird repellent. Its primary action mechanism is based on its ability to act as an acetylcholinesterase inhibitor. Despite its effectiveness as a molluscicide, mammalian toxicity is 10 times higher than metaldehyde and as such led EU banning of this pesticide in 2014 (Arena *et al.*, 2018). Therefore, the future utilization of a compound with greater toxicity to metaldehyde is unlikely. The alternative molluscicide ferric phosphate is the favored replacement for metaldehyde. This compound works through interfering with calcium metabolism within the mollusk's gut preventing further eating of the crops (Speiser and Kistler, 2002; Food and Authority, 2015). This approach has demonstrated significant lower toxicity to wildlife to that of metaldehyde however; several studies have demonstrated the high toxicity of the pesticide to terrestrial invertebrates such as earthworms (Edwards *et al.*, 2009). Despite the low toxicity of ferric phosphate to birds and mammals, the chelating compound ethylenediaminetetraacetic acid (EDTA) is often used to improve the effectiveness of the pesticide. EDTA accumulation within the environment is of growing concern and may prove to be a future complication with this molluscicide (Lanigan *et al.*, 2002; Wang *et al.*, 2017). As well as increased cost, another common complaint with the stakeholders regarding ferric phosphate is due to the relatively slow mollusk death following ingestion, slug carcasses which are notably visible with metaldehyde application are absent with ferric phosphate (Castle *et al.*, 2017). Replacement with this alternative therefore requires

further education with the stakeholders to provide greater product confidence. The success of combining education with replacement products is best demonstrated with the 'Slug It Out' campaign showing a 96 % reduction in metaldehyde levels through both stakeholder interaction and covering the additional costs of metaldehyde alternatives (Anglian Water, 2019).

A simple and effective approach to reduce xenobiotic contamination of the water supply is through building of swales. Grass swales, shallow vegetated open channels, provide buffer strips and following storm events prevent contamination of nearby water sources (Yousef *et al.*, 1987). These swales have previously been demonstrated to be effective in the preventing the runoff of several xenobiotic compounds such as pesticides and hydrocarbons (Gavrić *et al.*, 2019). Current ongoing trials by Thames Water aim to utilize these properties to greater reduce contamination of the drinking water (Pape, 2016).

Where the above preventative measures fail to reduce the metaldehyde contamination of the water sources, the temporary stopping of water abstraction can be undertaken. Following the spike, the contaminant can be diluted below the regulatory limits and normal abstraction can be resumed (Environment Agency, 2018). Due to the nature of this method, it is a temporary fix to the contamination problem and requires rapid and accurate detection methods.

As conventional drinking water treatment approaches are currently inadequate for cost effective and sustainable metaldehyde removal, the above preventative approaches are favored over removal itself (Cooke *et al.*, 2020). However, where water becomes contaminated water treatment processes are often required.

Granular activated carbon (GAC) based filtration has previously been shown to be effective in the removal of a wide range of xenobiotic compounds from drinking water such as tetracycline,

caffeine and DEET, amongst others (Fulazzaky and Omar, 2012; Sotelo *et al.*, 2014; Golovko *et al.*, 2020). However, due to the polarity of metaldehyde, dissolved organic carbon and xenobiotic compounds within the environment are capable of out competing metaldehyde for the GAC (Tao and Fletcher, 2013). Although studies have demonstrated reduction in metaldehyde where relatively high concentrations of the pesticide were tested, where typical environmental concentrations were analyzed little if any metaldehyde was removed (Tao and Fletcher, 2013; Li *et al.*, 2020). Metaldehyde removal through GAC methods also leads to a significant reduction in the bed lifetime and therefore the operational costs to regenerate and dispose of the carbon are both environmental and economically unsustainable (Busquets *et al.*, 2014).

As the small organic skeleton of metaldehyde dictates the weak interaction with GAC, Busquets *et al.* (2014) undertook the construction of 'designer' activated carbons. The study led to the synthesis of phenolic derived carbon compounds that revealed the ability to significantly reduce metaldehyde below the regulatory limit within the spiked water and was capable of performing significant reduction in the presence of high levels of background organic matter. However, the ability to remove metaldehyde below the regulatory limit in the presence of organic matter was not analyzed. Another limitation of this approach is the regenerative potential or costs were not discussed potentially suggesting the unsustainable nature of such a removal approach.

Through coupling absorption methods with the electrochemical destruction technology NyrexTM, metaldehyde was both absorbed and degraded down to CO₂ reducing the metaldehyde concentration to below the regulatory limit even in the presence of organic rich peat water (Nabeerasool *et al.*, 2015). Numerous other approaches utilized compounds such as nano-sized

zinc composites and resins in order to remove metaldehyde from water however due to limitations such as reduced removal ability and generation of toxic intermediate compounds they fail to provide adequate removal approaches suitable for industrial application (Doria *et al.*, 2013; Tao and Fletcher, 2014; Altarawneh *et al.*, 2020).

Advanced oxidation processes (AOP) represent a number of destructive reaction methods based up on the formation of hydroxyl radicals ($\cdot\text{OH}$). An AOP subgroup utilizes ultraviolet (UV) radiation to activate a number of different compounds, such as TiO_2 , H_2O and Fenton's reagent ($\text{H}_2\text{O}_2/\text{Fe}^{3+}$), in order to generate the highly reactive and quasi-unselective radicals (Matafonova and Batoev, 2018). These radicals in turn react with a wide range of organic compounds within the contaminated water thereby leading to the degradation and ultimate removal of the contaminants present. This approach has shown promise in the removal of numerous micropollutants present in the water supply including paracetamol, penicillin and metaldehyde. In the case of the latter, Autin *et al.* (2013) observed metaldehyde removal in efficiencies exceeding 90%. Despite the potential of AOPs as a highly effective micropollutant treatment, there are several limitations which prevent its further widespread utilization. Due to the quasi unselective nature of the free radicals, the vast majority of $\cdot\text{OH}$ are scavenged by natural organic matter and carbonate ions found in the water supply. Further studies also showed that in the case of UV/ TiO_2 high levels of background organic matter not only acted as a radical scavenger but also saturated the TiO_2 surface leading to a significant reduction in the efficacy of micropollutant removal (Autin *et al.*, 2012, 2013; Semitsoglou-Tsiapou *et al.*, 2016). The use of UV AOP also requires a high level of energy to overcome the dissolved organic carbon issues currently limiting its effectiveness. Despite these limitations, in 2015 Anglian Water built an AOP

water treatment system with the purpose of metaldehyde removal. Additional operation costs of £17 million and the inability to increase the consumer's utility bills by 21% means this approach is not, as of yet, an economically sustainable one to micropollutant remediation (Waste Water Treatment Online, 2016). Breakthroughs in light emitting diodes however, which may reduce energy consumption, or changes in regulation may make AOP remediation more economically viable in the future.

1.4 Biotic degradation of metaldehyde

In order to meet the appropriate regulatory demands, the metaldehyde manufacturing company Lonza conducted several investigations into the biotic degradation of metaldehyde. Although the degradative rates varied throughout the assays depending on the conditions investigated, biotic degradation of the molluscicide was observed (European Food Safety Authority, 2014). The metaldehyde removal ability of active sewage sludge was investigated numerous times. The metaldehyde reduction rates of the active sludge varied between 2.8 – 18.0 % over the 28 days of the assays (Wuthrich (1990); Lebertz (2008)). Despite demonstrating biotic degradative of metaldehyde within these assays, the regulatory set point of 20 % reduction was not met and therefore based on the outcomes of these assays metaldehyde is classified as a non-readily biodegradable compound.

Aerobic soil degradation undertaken by Juozenaite (2009), revealed following a lag phase of 5.8 – 19 days where no to very little metaldehyde was degraded, near complete degradation was observed in the days following. Radioactive assays failed to identify compounds in significant quantity upon analysis of the degradative product suggesting a highly reactive intermediate compound. (Möllerfield et al, 1993). Based on the evidence presented in the above studies, metaldehyde is suggested to be depolymerized into acetaldehyde molecules that are incorporated into central metabolism eventually generating CO₂.

Through use of ¹⁴C metaldehyde Balashova *et al.*, (2020) performed mineralization assays on soil samples obtained from three contrasting settings allotment, garden and agriculture. Following a 5-day incubation period, metaldehyde mineralization was observed in all soils examined with mineralization values ranging from 18 to 60 %. As all soils displayed significant levels of metaldehyde degradation, the abundance of metaldehyde degrading organisms within the environment and the catabolic competence of these soils was revealed. As well as demonstrating the ubiquitous nature of the degradative genes within the environment, examination of the mineralization conditions revealed soil texture has a significant role in metaldehyde mineralization whereby lighter soil textures demonstrating higher mineralization capacity. Further demonstrating the important role of soil microorganisms in the removal of metaldehyde from the environment, Simms *et al.*, (2006) revealed significant reduction in metaldehyde in soil inoculated with microorganisms relative to the sterilized control conditions. Based on the data collected, they theorized a metaldehyde pellet coated or containing antimicrobial properties with the economic advantage of preventing pesticide loss.

As previous investigations have demonstrated the biotic degradation of metaldehyde within various matrices and conditions, the industrial potential of this knowledge is applied in Rolph *et al.*, (2018). Operational sand filters promote both absorption and biosorption of contaminants and have demonstrated success in removing xenobiotic compounds such as bentazone, glyphosate, and p-nitrophenol (Taylor Eighmy *et al.*, 1992; Kowalczyk *et al.*, 2015; Randelovic *et al.*, 2016). Engineering and manipulation of the biofilm, through metaldehyde enrichment, conditions were optimized to maintain compliance degrading levels ($< 0.1 \mu\text{g/L}$) for >20 days and 80 % metaldehyde removal for > 40 days (Rolph *et al.*, 2018). This investigation revealed the potential industrial applications for obtaining further insight into metaldehyde degradative mechanisms and the genes responsible, through showing the potential of a cheap, chemical free biological metaldehyde treatment method.

Using solid and liquid culture enrichment methods, the first microorganisms capable of metaldehyde degradation were identified within domestic soils. Both strains, *Acinetobacter calcoaceticus* E1 and *Variovorax* strain E3, demonstrated the ability to utilise metaldehyde as a sole carbon and energy source (Thomas *et al.*, 2017). Degradation assays shown in Figure 1-4 reveal the industrial potential for such isolates as *A. calcoaceticus* E1 possess the ability to degrade metaldehyde below the regulatory limit ($0.1 \mu\text{g/L}$).

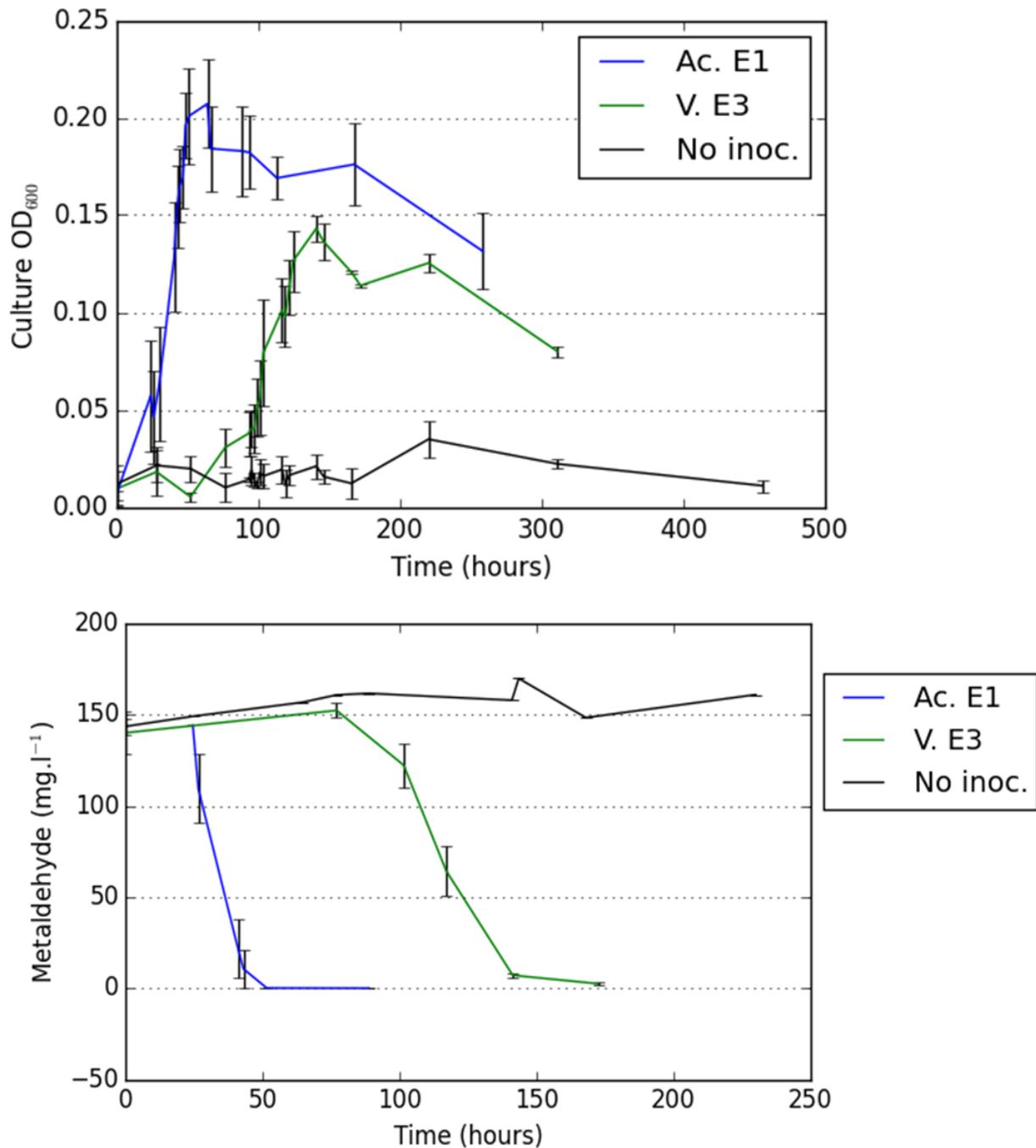


Figure 2-4- A. Mean OD₆₀₀ of *Acinetobacter* E1 (open circles), *Variovorax* E3 (filled circles) and not inoculated (filled triangles) in MSM culture supplemented with 0.85 mM metaldehyde. Error bars display the standard deviation of triplicate cultures. B. Mean metaldehyde concentration of culture media during the growth of *Acinetobacter* E1 (open circles), *Variovorax* E3 (filled circles) and not inoculated (filled triangles) in MSM culture supplemented with 0.85 mM .Error bars display standard deviation of triplicate cultures. Images reproduced from Thomas et al (2018).

The work conducted by Castro-Gutiérrez *et al.*, (2020) led to the isolation of two further diverse strains capable of metaldehyde degradation. The strains *Sphingobium sp.* CMET-H and *Rhodococcus globerulus* HMET-A, isolated from allotment and agricultural soils, demonstrated the ability to degrade metaldehyde. The isolated organism *Sphingobium sp.* CMET-H demonstrated a degradative rate greater than that of *A. calcoaceticus* E1 and as such shows greater industrial potential (Thomas *et al.*, 2017b; Castro-Gutiérrez *et al.*, 2020). Despite great advancements in knowledge surrounding the biotic degradation of metaldehyde and several microorganisms involved, prior to this work no degradative gene(s) or mechanism(s) were identified.

1.5 Protein evolution

Thousands of novel xenobiotic compounds have been designed and synthesised worldwide that display a wide variety of chemical properties. The addition of these novel compounds to an environment can cause significant evolutionary changes to the microbial community (Lawrence *et al.*, 2012). As bacteria rapidly produce new generations and typically form large populations, selective pressure and biological diversity enable evolution to work on existing mechanisms so that new metabolic resources can be exploited. Despite the relative speed through which

catabolic genes and pathways can be acquired, much degradation occurs through non-specific cometabolism and as such may not generate any cellular resource (Luo *et al.*, 2014). Numerous compounds, however, can be utilised by microorganisms for carbon, nitrogen, and energy sources. The speed at which compounds biodegrade is dependent on numerous factors such as environmental conditions (such as pH, temperature, aeration), chemical properties of the compound, and prior compound exposure (Spain and Van Veld, 1983; Providenti *et al.*, 1993). Xenobiotic compounds that are structural or chemically similar to well established or naturally occurring compounds typically show a higher rate of bioremediation than their novel and unique counterparts (Janssen *et al.*, 2005; Singh and Walker, 2006).

For an enzyme to catalyze a specific reaction, it must both bind to the substrate and position the substrate to the appropriate amino acid residues within the active site. As such, the overall structure and specific individual amino acid residues are essential to ensure effective catalysis (Dalby, 2003). The primary amino-acid sequence, which dictates the shape and activity of the enzyme, is ultimately encoded through nucleic acid base sequences. Mutation events within these nucleotide sequences can have significant consequences for the enzymes structure and therefore its specificity, sensitivity, and overall activity (Hynes *et al.*, 2009; Ishida, 2010; Morgenthaler *et al.*, 2019). Despite mutation events typically conveying negative or negligible effects, on occasion they can provide the organism with the capability to catabolize a substrate at improved efficiency or in some cases a new substrate altogether (Yuan *et al.*, 2017; Zhu *et al.*, 2019; Copley, 2020). As the initial function of the newly mutated protein must still be performed, gene duplication events can often be essential in providing evolutionary innovation whilst ensuring original functions are met (Innan and Kondrashov, 2010; Copley, 2020). Where the new

mutation encodes a novel advantageous function, such as degradation of a novel compound, positive selection encourages the preservation and distribution of the duplicated gene (Wagner, 2007). This new function is then often further improved through rapid evolution through additional mutations (Torgerson and Singh, 2004).

With regards to evolution towards pesticide degradation, a relevant example would be the pesticide atrazine. Within nature there exists no compounds that possess the same ring structure and as such it was originally considered to be resistant to biotic degradation (Monard *et al.*, 2010). However, with the pesticides widespread and high usage enzymes that possess the ability to initiate atrazine metabolism have been identified within numerous genera identified independently on 4 different continents (Udiković-Kolić *et al.*, 2012). The current evolutionary theory regarding these enzymes is based on the mechanism by which atrazine chlorohydrolase acts on the s-triazine ring of atrazine. This ring closely resembles that of pyrimidine ring compounds and as such selective pressure may have allowed for the evolution away from pyrimidine deamination and into atrazine metabolism. This is further supported by both enzyme belonging to the Amidohydrolase superfamily possessing several conserved features. Studies have also revealed the highly specific nature of atrazine chlorohydrolase (Udiković-Kolić *et al.*, 2012; Noor *et al.*, 2014). As the enzyme acts only on the chlorine or fluorine leaving group of an s-triazine ring this is therefore consistent with the idea of evolution by selective pressure towards s-triazine herbicide metabolism.

1.6 Gene mobility within the environment

The inheritance of genetic material from mother to daughter cell is dictated through the evolutionary process. Within the prokaryotic domain as well as vertical transfer of genetic traits, genetic interactions can occur horizontally. These horizontal gene transfer (HGT) events occur through three known mechanisms; transformation, conjugation and transduction (Wagner, 2007). These transfer mechanisms are mediated through mobile genetic elements (MGE) such as plasmids, transposons, and certain bacteriophages (Sobecky and Hazen, 2009).

Transposons are DNA sequences capable of replication and transposition within a genome. Between terminal inverted repeats, the DNA can be 'cut' from the genome and 'pasted' into to a new location. This process typically requires the transposase enzyme that is often encoded within the transposable element (Whittle *et al.*, 2002). Due to the nature of the replication mechanism, transposons are capable of duplicating genes found within the repeated sequences (Cerbin and Jiang, 2018). As discussed above, such duplication events allow for the development of novel enzymes and are therefore vital in microbial adaptation to novel compounds. As well as providing intracellular mobility, transposons are often located on transferable plasmids (Whittle *et al.*, 2002). As such, this allows for the replication and dissemination of the genetic information to the

wider microbial community. The importance of transposons for the microbial adaptation to novel metabolites can be inferred through the large numbers of catabolic gene clusters identified within insertion sequence sites (Tsuda *et al.*, 1999; Siddavattam *et al.*, 2003; Pandeeti *et al.*, 2012; Phale *et al.*, 2019).

Plasmids are extrachromosomal replicons found in both linear and circular forms. These are often transmissible and are capable of spreading genes to a broad and diverse range of microorganisms within the community (Meinhardt *et al.*, 1997; Yanagiya *et al.*, 2018). Positive selection ensures that transmission of genetic information that provides a selective advantage to the new host will further spread to the wider community. This can lead to bacteria acquiring numerous new traits such as antibiotic resistance, heavy metal resistance and the ability to utilize a new carbon source (Basta *et al.*, 2004; Bennett, 2008; Kothari *et al.*, 2019). In the case of the latter, degradative genes and clusters are often found on relatively large (90 to 1,100 kb) and highly mobile catabolic plasmids that confer the ability to degrade and utilize both naturally occurring and xenobiotic compounds. These plasmids have been associated with the degradation of numerous xenobiotic compounds such as the herbicide bromoxynil, the pesticide linuron and the insecticide carbofuran (Stalker and McBride, 1987; Ogram *et al.*, 2006; Werner *et al.*, 2020). Bacteriophages are viruses that possess the ability to infect and replicate within bacteria. In doing so, they can often package part of their host's DNA, such as genes encoding catabolic genes or antibiotic resistance, and propagate it among bacteria (Gunathilaka *et al.*, 2017). Their ubiquity and abundance ensure the effective dissemination of genetic material in a wide range of locations. Environments with a high abundance of phage and bacteria coupled with antibiotic

exposure have been revealed to encourage the spread of the resistance genes within the environment. This is demonstrated by the detection of phages carrying antibiotic resistance cassettes being found in environments such as sewage, soil and sludge (Colavecchio *et al.*, 2017; Gunathilaka *et al.*, 2017; Gómez-Gómez *et al.*, 2019). As such, bacteriophage aid in the transfer and acquisition of genes that can provide a positive selective advantage within the microbial community.

Within the well studied pesticide atrazine, genes involved with different mineralisation steps are found throughout numerous microorganisms. The microorganism *Pseudomonas* sp ADP however possesses the ability to completely mineralise the pesticide (De Souza *et al.*, 1998). This ability is achieved through possession of the plasmid PAD-1 which contains all 6 of the required *atz* genes. With regards to gene mobility, 3 of the 6 *atz* genes present upon the plasmid are flanked by insertion sequence sites and transposases (De Souza *et al.*, 1998; Meyer, 2010; Udiković-Kolić *et al.*, 2012). As such, this would strongly suggest multiple gene recruitment events that have allowed a single organism to completely degrade and metabolise a pesticide that had previously considered to be resistant to biotic degradation. Due to the natural selection advantage of possessing relevant catabolic genes, HGT mechanisms, with the help of MGEs, allow for the effective dissemination and propagation of catabolic genes within microbial communities.

1.7 Bioremediation, Biosensing and Biomonitoring

Technologies and methods can be developed to help combat metaldehyde pollution within the environment. Identification of the catabolic genes responsible for the initial catabolic step provides several opportunities for industrial application such as the ability quantify the pesticide, remove the pesticide and determine the environmental capability of degrading the pesticide.

1.71 Bioremediation

Bioremediation refers to the process of removal or transformation of contaminants within the environment typically through microbial activity. Using techniques such as biostimulation, bioaugmentation, and bioreactors, xenobiotic compounds can be effectively removed from the environment without the need for high-energy input or chemical application (Sarkar *et al.*, 2016; Kumar Singh *et al.*, 2020).

Where local microbial populations demonstrate sufficient degradative ability within the environment, biostimulation of existing degradative microorganisms can provide an effective, cheap and green approach to contaminant removal (Kanissery and Sims, 2011; Sarkar *et al.*, 2016). A well-studied example of this is following the Exxon Valdez Spill oil spill. Presence of hydrocarbon degrading microorganisms were already identified within the environment. As nitrogen was the growth limiting nutrient within the environment, addition of fertilizer lead to a

significant increase in the degradation rate of the contaminant (Kanissery and Sims, 2011). Concerning pesticide removal, biostimulation of soil with bird droppings has previously been shown to improve microbial degradation of the fungicide tebuconazole (Baćmaga *et al.*, 2019). As biostimulation utilizes the microbial community already present within the environment and numerous studies have shown the catabolic strains returning to normal concentration within their microbial community's post nutrient supplementation (Baćmaga *et al.*, 2019), this approach overcomes potential environmental and ethical considerations that may be considered in alternative methods. The Balashova *et al.* (2020) study revealed the ubiquitous nature of metaldehyde degraders within the environment, biostimulation provides a potentially effective approach when combined with alternative methods such as biobeds and sand filters, in the removal of metaldehyde.

Where a xenobiotic compound is typically not found within the environment, degradative genes may not be present to efficiently degrade the compound and prevent its accumulation. This lack of degradation may be overcome through bioaugmentation whereby preselected strains / mixed cultures are added to the contaminated environment. One example is the removal of lignin from wastewater generated by the pulp and paper industry (Baćmaga *et al.*, 2019). As microbial degraders of lignin are not commonly found within activated sludge, biodegradation within wastewater is typically slow thereby resulting in the accumulation of the compound. Through identifying microbial populations that demonstrate a high rate of lignin degradation and exposing this community to the contaminated water, removal of > 50 % of the lignin was observed (Priyadarshinee *et al.*, 2016; Baghel *et al.*, 2020). This approach has also been

demonstrated to effectively remove numerous compounds from wastewater such as cyanide, 3-Chloroaniline and Naphthalene (Park *et al.*, 2008; Bathe *et al.*, 2009; Xu *et al.*, 2015).

Where augmentation through genetic engineering methods are undertaken, microorganisms with qualities that show greater resistance of adverse conditions and demonstrate greater degradative efficiency can be utilized (Baćmaga *et al.*, 2019). Genetic modification and directed evolution of the degradative microorganisms to improve efficiency has been demonstrated to be effective in the removal of several xenobiotic compounds from the environment (Ang *et al.*, 2009a; Liu *et al.*, 2017). In the study conducted by Gong *et al.*, (2017), *Pseudomonas putida* KTUTGV underwent both genetic engineering and directed protein evolution to enhance its growth rate increase, increase thermal stability and improve trichloropropane (TCP) degradation efficiency. Continuous treatment for TCP through a biobed reactor proved high effective in removal of the contaminant and thereby demonstrated the industrial application of such an approach. Despite the above approaches showing the potential of genetic engineering to aid in xenobiotic removal from the environment limitations that prevent greater application include public opinion and current regulation surrounding GMOs (Marris, 2001; Health and Safety Executive, 2014; Cui and Shoemaker, 2018). To overcome these limitations, non-viable cell extracts can be utilized and have been shown to demonstrate promise. Expression of the degradative enzyme(s) has several advantages over live cell use such as; transport, storage, public opinion and ability to remove contaminants at high concentrations higher than most living organisms can tolerate (Azubuike *et al.*, 2016), it does however possess several limitations. As many catabolic enzymes require numerous cofactors, their environment must also contain these compounds. Also, the non-reproductive and non-regenerative nature of enzymes means this

approach requires continuous application within the environment (Azubiike *et al.*, 2016). The use of crude cell extracts and purified proteins have previously been utilized as bioremediation methods for numerous contaminants such as organophosphates, anthraquinonic dyes and TCP (Dvorak *et al.*, 2014; Yuan *et al.*, 2016; Thakur *et al.*, 2019). In the case of TCP, enzymes required for the conversion of the carcinogenic compound TCP into harmless glycerol were immobilized within a packed bed reactor. This approach demonstrated 97% removal of the initial contaminant and generated 78% of the final product during the period of operation (Dvorak *et al.*, 2014).

1.72 Detection through biological methods

Biosensors are analytical devices that utilize biological components to generate signals that are proportional to target analyte concentrations (Kohn and Plaxco, 2005). One of the most commonly known examples that demonstrates the great potential of biosensors is the glucose biosensor. In this example, immobilized glucose oxidase reacts with β -D-glucose. This reaction utilizes the redox cofactor Flavin Adenine Dinucleotide (FAD) that is subsequently reduced from FAD⁺ to FADH₂. Regeneration of this cofactor, through reacting with oxygen leads to the formation of hydrogen peroxide that subsequently oxidizes a platinum anode (Kohn and Plaxco, 2005). The number of electron transfers within this last step are thereby used as a proxy to determine the initial glucose concentration. As such, any reaction that utilizes a redox cofactor or can be coupled to one can be utilized as a biosensor. Based on this approach, the enzyme polyphenol oxidase was absorbed onto a glassy carbon electrode. This allowed for the monitoring of phenol within aqueous effluents down to detection limits of 10 nM (Kohn and Plaxco, 2005). Whole cell bioreporters work on a simple design principle whereby a promoter region, induced by the target analyte, is fused to a gene encoding a protein that can be easily detected (reporter

protein). As the analyte concentration determined the expression of the reporter protein, the signal generated by the reporter can be used as a proxy for the analyte concentration (Roggo and van der Meer, 2017). Construction of a wide variety of whole cell bioreporters have allowed contaminants such as heavy metals, alkanes and halogenated hydrocarbons to be detected in environments such as water and soil (D. Zhang *et al.*, 2012; Ul Haque *et al.*, 2013; Yoon *et al.*, 2016). These studies have demonstrated not only can whole cell bioreporters provide a cheap, sensitive, and rapid quantification method but also possess a high level of accuracy (within 3% when compared to traditional detection methods). Typically, sensitive bioreporters require two things: an inducible promoter region and a degradative gene. Previous work conducted by Thomas, (2016) identified exposure of *A. calcoaceticus* E1 to metaldehyde led to significant induction of an aldehyde dehydrogenase protein when compared to acetate induction. Therefore, the first criteria of an inducible region is met and as such requires only the second. Depending on the inducing compound and the degradative pathway, identification of the mechanism responsible for metaldehyde degradation is still required. Despite the great potential of whole cell bioreporters, several limitations of this approach do exist. The regulations surrounding GMO utilization prevent the full portability of this bioreporters being achieved (Health and Safety Executive, 2014). Another complication is inhibition and interference through contaminants within the environment reducing both the specificity and reliability of such devices.

1.73 Biomonitoring

Repeated application of xenobiotic compounds leads to a selective pressure on the local microbial communities. The presence and abundance of the compound degraders within these communities play a key role in determining the accumulation and biodegradation within the soil environment (Chikere *et al.*, 2011; Wu *et al.*, 2017). As mentioned previously, the HGT of catabolic genes allows for the propagation and distribution of these genes throughout the environment. Numerous studies have correlated the detectable degradative genes with the biodegradation potential of the environment (Holben *et al.*, 1992; Parekh *et al.*, 1995; Widada *et al.*, 2002). As such, robust and sensitive methods to determine the *in-situ* biodegradative ability and the resilience of the local environment to the xenobiotic is essential.

PCR based methods

Considered the gold standard of gene analysis, PCR based approaches have been demonstrated to be an effective method in the detection and monitoring of gene abundance within a given environment. Quantitative PCR (qPCR) provides real time analysis and allows for the detection and quantification of the gene abundance within the environment. Due to its fast,

sensitive and high throughput nature of qPCR, it has been utilized in numerous gene monitoring investigations (Kohno *et al.*, 2002; Monard *et al.*, 2013).

Within a given sample, qPCR only provides information regarding the genetic potential. As such, this method fails to consider the microbial conditions within the environment that may govern gene expression. In order to overcome these limitations, reverse transcription (RT) -qPCR can be undertaken to provide a clearer and more detail picture regarding the catabolic potential of the environment. Instead of DNA, RT-qPCR aims to detect and quantify RNA. As RNA is unstable and prone to degradation, the enzyme reverse transcriptase is used to convert the present RNA into the significantly more stable product cDNA (Kohno *et al.*, 2002). Following this, qPCR analysis is then performed on this product and analyzed accordingly. Several studies have demonstrated the good correlation between the catabolic gene expression and the degradation of the xenobiotic compound (Monard *et al.*, 2010, 2013).

Another gene monitoring method is through loop mediated isothermal amplification (LAMP). As this approach requires only a single constant temperature, as opposed to thermal cycling required for PCR based analysis, it allows for rapid, portable and cheap approach for determining the presence of degradative genes within the environment (Notomi *et al.*, 2000). To date, this method has shown to be effective in identifying the nitrite reductase encoding gene *nirS* within spiked saltwater samples (Zhang *et al.*, 2020). There is great industrial application and potential for LAMP to be utilized to determine the metaldehyde catabolic competence of a given environment or matrices such as sand filters and biobeds.

Through combination of the above applications, further optimization and engineering of bioremediation methods could be achieved. Effectively and readily possessing the ability to monitor the degradative genes and metaldehyde concentration within a given matrix allows for greater optimization of the bioremediation processes. As demonstrated throughout numerous studies there is degradative potential of metaldehyde within the environment (European Food Safety Authority, 2014; Thomas *et al.*, 2017b; Rolph *et al.*, 2018; Castro-Gutiérrez *et al.*, 2020) and through developing our understanding the mechanisms and understanding of these processes, our ability to optimise these processes through biostimulation or bioengineering can improve the rate at which this occurs.

1.8 Aims of the Project

- To identify the gene(s) responsible for metaldehyde degradation within the industrial relevant metaldehyde degrading strain *A. calcoaceticus* E1 (Chapter 3).
- To experimentally verifying the bioinformatics findings of the identified gene(s) through heterologous expression within a model organism (Chapter 3).
- To characterize the initial degradative metaldehyde degrading enzyme and demonstrate *in vitro* activity (Chapter 4).
- To identify key cofactors required for optimum degradative activity of the degradative enzyme (Chapter 4).
- To identify alternative metaldehyde degrading genes within metaldehyde degrading bacteria through bioinformatics analysis (Chapter 5).

- To perform bioinformatics analysis to determine HGT markers of any identified metaldehyde degrading gene (Chapter 5).
- To experimentally validate and characterize any potential metaldehyde degrading enzyme (Chapter 5).

Chapter 2: Materials and Methods

2.1 Microbiological Methods

2.11 Growth media

2.111 Minimal salts media

Unless stated otherwise, *Acinetobacter calcoaceticus* E1 was grown in minimal salts media (MSM) supplemented with 150 mg/ L (0.85 mM) of metaldehyde. MSM was prepared to achieve a final concentration of the following; 55 mM Na₂HPO₄, 11 mM KH₂PO₄, 6 mM NH₄Cl, 0.4 mM, and further supplemented with 2 mL of trace element solutions (Vishniac and Santer, 1957) found in table 2-1.

Table 2-1: The concentration of compounds found within the trace elements solution

	Concentration (mM)
Na ₂ EDTA	140
ZnSO ₄	7.6
CaCl ₂	37
MnCl ₂	25
FeSO ₄	18
(NH ₄)Mo ₇ O ₂₄	0.9
CuSO ₄	6.4
CoCl ₂	6.6

The salt solution was dissolved within 1 L of ultrapure water and autoclaved. Once cooled, metaldehyde (Sigma-Aldrich) was added and dissolved through continuous mixing overnight at room temperature. Where metaldehyde plates were created, 1.5 g of agarose was added to 200 mL of MSM liquid prior to autoclaving. Once cooled, 100 mg of metaldehyde was added to 200 mL of molten MSM media giving a final concentration of 4.3 mM. Where acetate media was created, sodium acetate was dissolved into ultrapure water to achieve a final concentration of 1M. This solution was then filtered sterilised using a 0.22 μm filter. This was added to the either cooled liquid or molten MSM agarose media to achieve a final concentration of 10 mM.

2.112 Lysogeny broth (LB) and LB plates

LB was prepared through dissolving 10 g of NaCl (Sigma-Aldrich), 10 g granulated tryptone (Melford) and 5 g of yeast extract (Oxoid) into 1 L of ultrapure water. Where plates were created, 3 g of granulated agar (bacteriological grade; Formedium) was added to the LB mixture and prepared to achieve a final volume of 200 mL. Media was autoclaved before use. Where antibiotics were supplemented, media was cooled to $\sim 50^\circ\text{C}$. Plates were poured within a Laminar flow to ensure appropriate sterility with ~ 20 mL poured into each petri dish. LB agar plates were then left within the hood for 30 mins to allow the agarose to solidify.

2.113 Glycerol stocks

Long term storage was required for bacterial strains and plasmids of importance. This was achieved through inoculating 10 mL of LB media with single colonies of the appropriate sample.

Where appropriate media was supplemented with antibiotics. Samples were then grown overnight at 30 °C at 180 rpm. Within a sterile Eppendorf tube, 500 µL of overnight culture was added to 500 µL of sterile 50 % (v/v) glycerol. Samples were then frozen at – 80 °C. To ensure adequate recovery, samples were tested 1 week after freezing. This was performed through streaking the samples on LB plates supplemented with the appropriate antibiotic.

2.114 Growth of *Acinetobacter calcoaceticus*

Throughout this work, the metaldehyde degrading strain *A. calcoaceticus* E1 was used. Streak plates were typically grown aerobically on metaldehyde at 30 °C for 48 hours. Once grown, samples were stored at 4 °C. Where liquid culture was required, *A. calcoaceticus* E1 was grown at 30 °C in 10 mL of LB within a 50 mL falcon or 50 mL of LB within a 250 mL conical flask. Growth of the *A. calcoaceticus* E1 mutants deficient in metaldehyde degradation were streaked on LB agar plates and grown at 30 °C.

2.115 Growth of *Escherichia coli*

E. coli strains DH5α and BL21 (DE3) were used frequently within this work. *E. coli* was grown at 37 °C LB, unless stated otherwise. Single colonies were obtained through streaking onto LB plates and 24 hour incubation at 37 °C. *E. coli* was grown aerobically within a 50 mL falcon, 250 mL conical flask and 2 L conical flask. Where cells expressed *mah* related genes, a growth temperature of 30 °C was used.

2.116 Random mutagenesis through ethyl methanesulfonate (EMS)

This method based on Geissdorfer *et al.*, (1995) was undertaken to generate and identify mutants unable to utilize metaldehyde. From a freshly streaked LB plate, *A. calcoaceticus* E1 was inoculated into 10 mL of LB media and incubated at 30 °C on an orbital shaker at 180 rpm. The overnight culture was then used to inoculate 20 mL of LB contained within a 50 mL Falcon tube, in order to generate a 1 in 100 dilution. The samples were then incubated (at 30 °C with shaking) until reaching mid log phase (OD_{600} of ~0.5). Once the appropriate OD was achieved, 180 μ L of EMS (Sigma-Aldrich) was added to the culture. The sample was then incubated at 30°C for 3 hours without shaking. Following this, 500 μ L of the EMS sample was then added to a 250 mL conical flask containing 40 mL of LB media. This mixture was incubated at 30 °C for 4 hours at 180 rpm. The culture was then serially diluted using LB. Various dilutions mixtures were then spread onto LB plates and incubated for 2 days at 30 °C. Where present, single colonies were replicate plated onto both minimal salt media (MSM) plates containing either metaldehyde or acetate. Plates were incubated at 30 °C for 3 days. Visual examination of the plates was undertaken to identify growth on acetate yet absence on metaldehyde plates. Where mutants were identified, they were subjected to whole genome sequencing at MicrobesNG. The no EMS control assay involved identical steps as above without the addition of the compound EMS.

2.118 A. calcoaceticus growth curve assay (Manual)

In triplicate, cells were grown onto LB plates overnight at 30 °C. Using a sterile sample swap, colonies were swabbed from each plate and emulsified into 10 mL of MSM. The OD_{600} of each of the samples was then calculated and used to inoculate 50 mL of MSM supplemented with

metaldehyde to final OD₆₀₀ of 0.001. Samples were then incubated at 30 °C at 180 rpm. To determine both OD₆₀₀ and the metaldehyde concentration through GC analysis, 1 mL samples were extracted at various time points of the assay. OD₆₀₀ was then calculated using a Jenway 6300 spectrophotometer.

2.119 Growth Curve Assay (Plate Reader)

Using an Epoch 2 Microplate spectrophotometer (BioTek), sample OD₆₀₀ over time was investigated. For all assays, samples were plated onto fresh LB plates and incubated at 30 °C overnight. Using a sterile swab, colonies were collected and emulsified within 10mL of MSM. The OD₆₀₀ of this media was determined using a Jenway 6305 spectrophotometer. Samples were then diluted within their appropriate growth culture to reach an OD₆₀₀ of 0.1. To each well of a 96 well microplate, 150 µL of diluted sample was added. Each sample was performed in biological triplicate. Samples were incubated for 24 hours at 30 °C with double orbital shaking.

2.2110 The effects of metaldehyde on growth on E. coli transformants

Inhibition growth analysis was performed using an Epoch 2 Microplate spectrophotometer (BioTek). *E. coli* samples were prepared as described previously in section 2.119. Samples were grown in a mixture containing 100 µL LB supplemented with tetracycline (10 µg/L) and 100 µL of MSM supplemented with/without metaldehyde (0.85 mM).

2.2 Molecular Methods

2.21 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments based on their size. The preparation of a 1.2 % agarose gel proceeded as follows; 0.6 g of agarose (Melford) was added to a 250 mL Erlenmeyer flask containing 50 mL of TBE buffer (90 mM Tris, 90 mM boric acid, 2 mM EDTA). The mixture was heated using a microwave to completely dissolve the agarose. Once cooled, 5 μ L of the DNA gel stain SYBR Safe (ThermoFisher) was added. The gel was then poured into a gel tray with the appropriately sized comb. DNA samples were mixed 5:1 with gel loading dye (NEB). The agarose gel was submerged under 0.5X TBE buffer. Gels were performed at room temperature for 60 mins at 100 V. Gel images were visualised on a GeneGenius Biol Imaging System (Syngene).

2.22 Polymerase Chain Reaction

PCR was used to amplify specific regions of DNA for both detection and cloning purposes. Where verification of the DNA region of interest was the aim, the low fidelity GoTaq (Promega) was used. The reagents added for a GoTaq reaction of 50 μ L were 10 μ L of 5X GoTaq Reaction buffer, 1 mM of forward and reverse primers, 0.25 μ L of GoTaq DNA polymerase (5U/ μ L) and 200 pg/ μ L of template DNA. Nuclease free water was then added to achieving the desired volume. The reaction conditions for this reaction are found in table 2.2

Table 2-2: Thermal Cycling Conditions for GoTaq PCR.

Step	Temperature	Time	No of Cycles
Initial			
Denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	30
Annealing	42-65 °C	1 min	30
Extension	72 °C	1 min/kb	30
Final Extension	72 °C	5 min	1
Hold	4 °C	Indefinite	1

Where cloning was the aim of the amplification, proofreading was required. As such, the polymerase Phusion High Fidelity (ThermoFisher) was used. For a 50 μ L reaction the following reagents were added 10 μ L of 5X Phusion HF Buffer, 1 μ L of 10 mM dNTPs, 10 mM of forward and reverse primers, 200 μ g/ μ L of template DNA, 0.5 μ L of Phusion High Fidelity DNA polymerase. Nuclease water was added to reach to the required reaction volume. The reaction conditions for the Phusion High Fidelity are located in table 2-3.

Table 2-3: Thermal Cycling Conditions for Phusion High Fidelity PCR.

Step	Temperature	Time	No of Cycles
Initial			
Denaturation	98 °C	30 sec	1
Denaturation	98 °C	10 sec	30
Annealing	42-65 °C	30 sec	30
Extension	72 °	30 sec/kb	30
Final Extension	72 °C	10 min	1
Hold	4 °C	Indefinite	1

For colony PCR, a 10 µL pipette tip was used to gently touch a single bacterial colony. The tip was then mixed into a 0.2 mL PCR tube containing 100 µL of nuclease free water. Following this, the sample was subjected to 95 °C for 10 minutes through the Px2 Thermal Cycler (Thermo Electron, MA, USA). From this sample, 2 µL was added to as template for the PCR. The remaining sample was stored at – 20 °C. Following PCR, all products were examined through gel electrophoresis

2.23 Primers

Oligonucleotide primers were purchased from Sigma and resuspended in nuclease free H₂O to a final concentration of 100 mM. Stock concentrations of 10 mM were prepared using nuclease free H₂O. Oligos used throughout the investigation are displayed in table 2-4.

Table 2-4: Primers used throughout the investigation

Primer Name	Sequence (5' to 3')	Purpose
contig X F	TTCGTCTTCAAGAATTCAAGCATCGTCA TTTTATTATGGTCAGC	Amplification of upstream region of mahX cluster
mahX F	TCAAACATGAGAATTTGCCTCCGCAATGG AACG	Amplification of mahX
mahX R	TGTCAAACATGAGAATTTTATCGTGTTGG GGAAGCGTC	Amplification of mahX
mahY F	GGAGAATGACATGAACAAATTACATCGA GTCGTTG	Amplification of mahY
mahY R	AGCTGTCAAACATGAGAATTTTATCGTGT TGGGGAAGCGTC	Amplification of mahY
Promoter mahY F	with CCCTTTCGTCTTCAAGAATTAAGCATCGT CATTTTATTATGGTCA	Amplification of promoter region
Promoter mahY R	with ATTTGTTTCATGTCATTCTCCTCTGTTAGTT CATTG	Amplification of promoter region
mah Z R	TGTCAAACATGAGAATTTCAAATTGACCC TCGCAAGCCA	Amplification of mahYZ
mbp-mahX F	TCCAGGGACCAGCAATGAAGCAAGAGCT TGAGTCCGC	Amplification of mbpmahX
mbp-mahX R	TGAGGAGAAGGCGCGTTAGCCGTTCAAC GTCGTTGCCTG	Amplification of mbpmahX
mbp-mahS F	TCCAGGGACCAGCAATGTCCGAGGTCGA CACCT	Amplification of mbpmahS
mbp-mahS R	TGAGGAGAAGGCGCGTCAGGCTTCCAGG CTCACC	Amplification of mbpmahS
mbp-mahS F	TCCAGGGACCAGCAATGACCGCGCGCT TGCTACA	Amplification of mbpmahR
mbp-mahR R	TGAGGAGAAGGCGCGTCACTTGCTGCTC TCGATGAAGTAGGAGA	Amplification of mbpmahR
pETFPP F	CGCGCCTTCTCCTCACATATGGCTAGC	Amplification of pETFPP
pETFPP R	TTGCTGGTCCCTGGAACAGAACTCC	Amplification of pETFPP

2.24 Endonuclease digest

Restriction digest was performed on pBR322 using the enzyme *EcoRI* (Promega). The reaction was performed using the protocol set out by the manufacturer and analysed through gel

electrophoresis. To prevent re-ligation, the DNA was subjected to alkaline phosphatase (NEB) following the manufacturer's protocol. Briefly, 1 pmol of DNA ends were combined with 2 μ L of CutSmart™ Buffer (x10) and 1 μ L of Quick CIP. This was then made up to a total of 20 μ L using nuclease free water. The mixture was incubated for 10 mins at 37 °C. Following this, the reaction was inactivated through heating the mixture to 80 °C for 2 mins.

2.25 DNA Purification

In order to undertake plasmid extraction, strains containing the relevant vector were grown in LB broth overnight supplemented with the appropriate antibiotic. Plasmids were harvested using a Qiagen MiniPrep kit following the manufacturer's instructions. Where DNA fragments were required for further downstream processes, post endonuclease digested products were purified through agarose gel electrophoresis. Using a clean scalpel, the band of interest was excised from the gel and transferred into a 2mL Eppendorf tube. The DNA was then extracted using a QIAquick Gel Extraction Kit (Qiagen) through following the manufacturer's protocol. Where PCR products were required for further downstream reactions, DNA was purified using QIAquick PCR purification kit. This was performed following the manufacturer's instructions. DNA was quantified through measuring the peak at 260 nm using a NanoDrop® ND-1000 Spectrophotometer (Thermo Scientific).

2.26 Cloning methods

For the purposes of identification of the metaldehyde degrading gene, constructs for heterologous expression were created. The NEBuilder[®] HiFi DNA Assembly Cloning kit was utilised to generate the required constructs. The genes *mahX*, *mahY* and *mahZ* were amplified using primers identified within table 2-4. Primers within this reaction were designed using Takara's primer design tool to contain regions of 20 -30 bp overlap to the plasmid fragment. PCR was performed using the High-Fidelity Phusion Polymerase (ThermoFisher). The second DNA fragment used was the *EcoRI* linearized and cleaned pBR322. Following the manufacturer's protocol, the reaction mixture was prepared on ice containing: 100 ng of vector, 50 ng of amplified DNA, 10 µL of NEBuilder HiFi DNA Assembly Master Mix. This mixture was then prepared up to a total volume of 20 µL using nuclease free H₂O. Both no insert and no master mix control were performed. The assembly reaction involved the incubation of the reaction mixtures for 60 mins at 50 °C using a Px2 Thermal Cycler (Thermo Electron, MA, USA). Samples were then stored at – 20°C for subsequent transformations.

Construction of a plasmid vector that provided increased solubility and demonstrated multiple methods of purification was undertaken through cloning of *mahX*, *mahR* and *mahS* into the *E. coli* expression vector pETFFP. This method also utilised the NEBuilder[®] HiFi DNA Assembly Cloning kit as discussed above. The genes *mahX*, *mahR*, *mahS* and the vector pETFFP, donated by L. Clark, were amplified using High Fidelity Phusion Polymerase and cleaned appropriately. All gene primers used were designed to contain regions of 20 -30 bp overlap

2.27 Preparation of chemically competent *E. coli* DH5 α and BL21 (DE3)

All initial cloning was undertaken within the *E. coli* strain DH5 α due to its high transformation efficiency. This method was used for both *E. coli* DH5 α and BL21 (DE3). Fresh colonies were streaked from glycerol stocks onto LB plates and incubated at 37 °C overnight. Colonies were picked and inoculated into 5 mL of LB and incubated overnight at 37 °C at 220 rpm. From the overnight culture, 500 μ L was transferred into a 250 mL conical flask containing 50 mL of LB. Cells were grown for ~ 3 hours until an OD₆₀₀ of 0.5-0.6 was achieved. The cells were then transferred to a 50 mL falcon tube and chilled on ice for 5 mins. Next, the cells were centrifuged using at 5,000 rpm for 10 mins at 4 °C. The supernatant was then discarded and the pellet resuspended into pre-chilled filtered sterilised 0.1 M CaCl₂. The cells were then further chilled for 20 mins on ice and then centrifuged again at 5,000 rpm for 10 mins at 4 °C. The supernatant was discarded and the pelleted cells resuspended into 2 mL of 70 mM CaCl₂ and 30 % (v/v) glycerol. Samples were aliquoted into sterile 1.5 mL Eppendorf tubes and frozen at – 80 °C (Angela *et al.*, 2017)

2.28 Heat shock transformation of competent *E. coli*

Previously prepared competent *E. coli* cells were removed from – 80°C and thawed on ice for 10 mins. Between 10-100 ng of DNA to be transformed was added to 50 μ L of the competent cells. The cells were then incubated on ice for 30 mins before being placed onto a 42 °C heat block for 60 sec. Cells were then rapidly cooled through placing them into ice for 2 mins. Following this, 500 μ L of LB was added to the Eppendorf tube. The cells were then incubated for 1 hour at 37 °C to allow the generation of antibiotic resistance proteins. Cells were then spread plated on LB plates supplemented with the appropriate antibiotic and incubated at 30 °C for 2-3 days.

2.29 Whole genome sequencing

Identified *A. calcoaceticus* E1 mutants were sent for WGS at MicrobeNG (Birmingham, UK). The sequencing was performed using an Illumina MiSeq platform using 2 x 250bp paired-end reads. BWA mem was utilised to ascertain the quality metric. De novo read assembly was performed through SPAdes (Bankevich *et al.*, 2012).

2.210 Sanger sequencing

Confirmation of cloning was performed using Sanger sequencing through GATC Biotech (Constance, Germany). Samples were submitted through the LightRun service using the sample submission protocol.

2.211 Strains and plasmids

Strains and plasmids that were used throughout this work are listed within table 2.5 and table 2.6.

Table 2.5: Plasmids used throughout this work

Plasmids	Description	Resistance	Source
pBR322	Vector used to determine gene required for initial metaldehyde degradation	ampicillin and tetracycline	ThermoFisherScientific
pBR322-mahX	pBR322 containing mahX with native Acinetobacter calcoaceticus E1 promoter	ampicillin and tetracycline	This work
pBR322-mahY	pBR322 containing mahY with native Acinetobacter calcoaceticus E1 promoter	ampicillin and tetracycline	This work
pBR322-mahXY	pBR322 containing mahX and mahY with native Acinetobacter calcoaceticus E1 promoter	ampicillin and tetracycline	This work
pBR322-mahXYZ	pBR322 containing mahX, mahY and mahZ with native Acinetobacter calcoaceticus E1 promoter	ampicillin and tetracycline	This work
pET21a	Vector used to express protein of choice with a C-terminally tagged his tag. IPTG inducible expression	ampicillin	Novagen
pET21a-mahXnhistag	pET21a containing maX with a C-terminally tagged his tag	ampicillin	GenScript (Designed within this work)
pETFPP_2	Vector used to express protein of choice with an N-terminally tagged his tag and mbp. IPTG inducible expression	kanamycin	(Fogg and Wilkinson, 2008)
pETFPP_2-mahX	pETFPP_2 containing mahX with an N terminal his tag and mbp	kanamycin	This work
pETFPP_2-mahS	pETFPP_2 containing mahS with an N terminal his tag and mbp	kanamycin	This work
pETFPP_2-mahR	pETFPP_2 containing mahR with an N terminal his tag and mbp	kanamycin	This work

Table 2.5: Strains used throughout this work

Strain	Genotype	Source
<i>E. coli</i> strains		Invitrogen
DH5 α	K-12 Φ 80dlacZ Δ M15 recA1 endA1 gyrA26 thi-1 supE44 relA1 deoR Δ (lacZYA-argF)U169	Novagen
BL21 (DE3)	F - ompT hsdSB (rB - , mB -) gal dcm (DE3)	Noavgen
<i>Acinetobacter</i> strains		
<i>Acinetobacter</i> E1		(Thomas <i>et al.</i> , 2017)
<i>Acinetobacter</i> RUH2202		Belgium Coordinated Collection of Microorganisms
Other strains		
<i>Sphingobium</i> CMET-H		(Castro-Gutiérrez <i>et al.</i> , 2020)
<i>Rhodococcus</i> globerulus HNO-A		(Castro-Gutiérrez <i>et al.</i> , 2020)

2.3 Bioinformatic analysis

2.3.1 Basic Local Alignment Search Tool (BLAST) methods

To find regions of similarity between both DNA and protein sequences, the program BLAST was used (Altschul *et al.*, 1990). Where alignment of two known sequences was undertaken, megablast were used. Where homology identification was the purpose, sequences were used to query either nucleotide collection database or the non-redundant protein database using BLASTn or BLASTP, respectively. Default settings were used unless stated otherwise.

2.32 Comparative genomic through BLAST

Comparative genomics through BLAST utilised a previously developed Python script (JC Thomas; available at <https://pypi.org/project/blast-score-ratio/>). This allowed for the predicted protein sequences from each genome to be queried against that of the *A. calcoaceticus* E1 wild type (WT). Output identified highest query matches for each predicted protein with an identity < 100%. Manual curation was undertaken to identify shared predicted protein sequences obtained through the analysis.

2.33 Phylogenetic analysis

To identify sequences of similar alignment, each protein sequence of interest was queried against the non-redundant protein sequences database using BLASTP. The sequences for the top 100 highest scoring alignments were then extracted. Through MEGA-X (Bankevich *et al.*, 2012), the multiple sequence alignment tool MUSCLE (Bankevich *et al.*, 2012) was used on the queried proteins and their most 100 most aligned sequences. Using MEGA-X, a maximum likelihood tree was generated with the image manual curated to reveal presence of relevant proteins. Default parameters were used throughout.

Phylogenetic analysis of MahX, MahR and MahS, involved the individual querying of the nonredundant protein database. For each query, the top 10 alignment sequences were extracted and duplications between the protein sequences were manually removed. Alignment and phylogenetic tree generation was performed as above.

2.34 Genomic analysis

Identification of insertion sequence sites was performed using the online tool ISfinder. BLASTn analysis was undertaken on 100 bp regions of the queried contig. The analysis utilised the BLASTn programme and the ISfinder database (Bankevich *et al.*, 2012). Default algorithm parameters were used. Where promoter prediction analysis was undertaken, the online programme SoftBerry BPROM was used (Li, 2011). DNA located 200 bp upstream from *mahX* was analysed using this tool. Plasmid analysis was undertaken using the plasmidSPAdes function in SPAdes genome assembler v3.11.1 (Antipov *et al.*, 2016). Identified contigs were curated through visual examination and BLAST analysis. Identified plasmid sequences were further analysed using the origin of transfer web-based tool OriTfinder (Li *et al.*, 2018). Data was input through FASTA format and default parameters were used.

2.36 Protein prediction

Functional analysis of protein was undertaken using InterPro (EMBL-EBI) (Hunter *et al.*, 2009). Queried protein sequences were input using FASTA format using the default settings. Further protein characterisation was undertaken using the protein structural prediction tool I-TASSER (Yang *et al.*, 2015). Protein sequences were submitted to the I-TASSER online server using FASTA format. The protein models generated through this approach were visualised and manually curated using the molecular visualization system PyMol. Where protein solubility was investigated, the online tool Protein-Sol was utilised (Hebditch *et al.*, 2017). Prediction protein sequences were submitted in single amino acid code FASTA format. Plasmid gene identification was performed using the online tool PATRIC. Samples were submitted as a single FASTA file using default parameters (Davis *et al.*, 2020).

2.37 Genomic visualisation

Examination of contig regions and vector design was undertaken using the online molecular software tool Benchling (Benchling, 2021). Whole genome visualisation and manual curation was performed using the genomic browser Sanger's Artemis (Hebditch *et al.*, 2017)

2.4 Protein analysis

2.41 Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

In order to observe protein size and presence, SDS-PAGE gels were used. The resolving gel was prepared through combining 4.1 mL dH₂O, 3.3 mL acrylamide (stock 30% acrylamide, 0.8% bisacrylamide (w/v); Protogel, National Diagnostics Atlanta, USA), 2.5 mL 1M Tris (pH 8.8), 0.1 mL 10 % (w/v) SDS, 32 µl 10% (w/v) ammonium persulphate (APS; Sigma-Aldrich) and 10 µl tetramethylethylenediamine (TEMED). The resolving gel mixture was transferred up to 2 cm from the top of the 1 mm gel plate. Isopropanol was used to overlay the resolving layer to ensure an even interface between both layers. The mixture was left at room temperature for 30 minutes, in order to allow the gel to set. The stacking gel, which consisted of the top 2 cm of the gel, was prepared through combining 6.1 mL dH₂O, 1.3 mL acrylamide, 2.5 mL 0.5M Tris (pH 6.8), 0.1 mL 10 % (w/v) SDS, 60 µl 10% (w/v) ammonium persulphate (APS; Sigma-Aldrich) and 100 µl TEMED. The SDS-PAGE tank running buffer consisted of 3 g of Tris base, 14.4 g of glycine and 1 g of SDS

dissolved within 1 L of H₂O. This solution was adjusted to achieve a final pH of 8.8, through addition of HCl.

The loading sample buffer (x2) was prepared through mixing the following; 100 mM Tris (pH 6.8), 4%(w/v) SDS, 0.2% (w/v) bromophenol blue (Sigma-Aldrich), 20% (v/v) glycerol and 200 mM β-mercaptoethanol (SigmaAldrich) (Cold Spring Harbour,2015). Cell lysate was prepared through the addition of loading buffer to sample pellet in a 1:1 mixture. This mixture was then heated to 98 °C for 5 minutes using a heat block. Samples were then centrifuged at 16,000g for 30 sec.

2.42 Coomassie protein staining

SDS-PAGE gels were stained using Coomassie Brilliant Blue (ThermoFisher). Following gel electrophoresis, the gels were placed into a tray containing the Coomassie stain. This consisted of 40 % methanol (Fisher Scientific), 10 % acetic acid (Fisher Scientific), 0.1% Coomassie Brilliant Blue R-250 (ThermoScientific), and made up with deionised H₂O. Gels were stained for 2 hours with gentle shaking at room temperature. To visualize the appropriate protein bands, gels were destained. This involved rinsing the gel with H₂O and resuspending the gel into the destaining solution containing 50 % H₂O, 40 % methanol and 10 % acetic acid. The de-staining procedure was performed until the protein bands were clearly visible.

2.43 Protein expression trials

Where protein production analysis was undertaken, *E. coli* BL21 containing the relevant plasmids were grown overnight at 37 °C in 5 mL of LB supplemented with the appropriate antibiotic within a 50 mL falcon tube. Following this, 50 µL of overnight cultured was transferred into a 250 mL

conical flask containing 50 mL of fresh LB supplemented with the appropriate antibiotic. Cells were grown for ~3 hours at 37 °C at 180 rpm until an OD₆₀₀ of 0.4-0.6 was achieved. Where induction was required, Isopropyl β- d-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich) was added to ensure a final concentration of 1 mM. Samples were then incubated at 30 °C whilst shaking at 180 rpm. Cells were collected at appropriate time points post induction and froze at -20 °C until further analysis. Where IPTG concentrations were investigated, reduced volume of IPTG was added to ensure appropriate final concentrations were achieved. Where temperature analysis was undertaken, samples were incubated post induction at 16 °C within an Excella E25 (New Brunswick).

2.44 Protein solubility assay

To determine protein solubility, the protein extraction reagent BugBuster (Merck Millipore) was utilised. Samples were grown and induced as previously described. From each appropriate condition, 1 mL of sample was extracted and transferred into a 2 mL Eppendorf tube. Samples were centrifuged for 5 mins at 17,000 g. The supernatant discarded and the pellet resuspended into 200 μL of BugBuster. Samples were then briefly vortexed and incubated for 15 mins at room temperature. From this mixture, 50 μL of sample was transferred and kept as the 'whole cell lysate'. The remaining sample was centrifuged for 5 mins at 17,000 g. The supernatant 'soluble extract' was carefully removed and used for further analysis.

Where the insoluble fraction was examined, the insoluble pellet was resuspended into 200 μL of a mixture containing 2 % (w/v) SDS and 100 mM DTT. The samples were then heated to 98 °C for

10 mins and then centrifuged at 17,000 g for 10 mins. The 'insoluble extract' supernatant was then carefully extracted and analysed through SDS-PAGE.

2.45 Obtaining soluble cell extract

All overexpressed proteins were performed in *E. coli* BL21 (DE3). *E. coli* expressing the proteins of interest were grown overnight in 25 mL of LB supplemented with the appropriate antibiotic. Following this, 5 mL of overnight culture was used to inoculate 2 x 500 mL of fresh LB containing the relevant antibiotic in 2 L flasks. Cells were grown to mid-log phase (0.4 -0.6) at 37 °C and then induced with 1 mM IPTG. Cells were then incubated at 30 °C overnight. Following IPTG induction, the cells were transferred to 1L plastic flasks and centrifuged for 4,500 rpm at 4 °C for 12 mins. The supernatant was discarded and the pellet resuspended into 35 mL of equilibrium buffer containing 20 mM Tris (pH 7.5) and 500 mM NaCl. Cell lysis was performed using a FisherBrand UltraSonic Sonicator with the samples kept on ice during the procedure. The conditions for which were 3 sec on 7 sec off with an initial output level of 7.0. Following the lysis of the cells, the crude extract was obtained through centrifugation of the samples at 15,000 rpm for 25 mins at 4 °C in a Sorvall LYNX 6000 Superspeed Centrifuge (ThermoFisher). The supernatant obtained was then carefully removed and filtered using a 0.45 µm filter. The solution at this point was used to perform soluble crude extract assays or for protein purification.

2.46 Immobilized metal affinity chromatography (IMAC)

Utilising the polyhistidine tag present on the overexpressed proteins, a 1 mL HisTrap excel column (GE Healthcare) was utilised to obtain pure protein. These columns are prepacked with Ni

Sepharose™ High Performance. Purification was performed using a hand-operated syringe. The column was equilibrated with 5 mL (5 column volumes (CV)) of wash buffer containing 20 mM Tris (pH 7.8), 500 mM NaCl and 20 mM imidazole. The sample was then loaded onto the column at an approximate rate of 2 mL/ min. Non bound protein was then removed through addition of 20 mL (20 CV) of wash buffer onto the column at a rate of ~5 mL/min. for ~4 mins To elute the bound protein, 5 mL (5CV) of elution buffer containing 20 mM Tris (pH 7.8), 500 mM NaCl and 500 mM imidazole was passed through the column at a speed of ~ 0.5 mL/ min. Multiple elution samples of ~ 0.75 mL were collected. Protein quantification was performed through measuring the absorbance values at A_{280} on a NanoDrop®ND-1000 Spectrophotometer (Thermo Scientific).

As imidazole possesses a high affinity for divalent metals and therefore may interfere with downstream assays, buffer exchange of the purified proteins was undertaken using a PD-10 column (Cytiva). For this buffer exchange method, the gravity protocol was utilised. To the column, 25 mL of equilibrium buffer was added and the flow through discarded. Following this, 2.5 mL of sample was added to the column. The flow through was then discarded and 3.5 equilibrium buffer was added to the column. The elution was collected within 2 x 2mL Eppendorf tubes.

2.47 Protein purification through amylose affinity

Purification through exploiting the fusion of the overexpressed protein to maltose binding protein was undertaken using a 5 mL MBPTrap™ - HP column. This column was prepacked with Dextrin Sepharose® High Performance to allow for protein purification. Purification was performed using a hand syringe. The column was equilibrated through addition of 30 mL of the

binding buffer which consisted of 20 mM Tris, 200 mM NaCl, 1 mM EDTA (pH 7.4). This was performed at approximately 5 mL/ min. The sample was then applied to the MBPTrap column at a rate of ~ 2 mL/ min. The column was then washed with 50 mL of binding buffer to remove all non-bound proteins. Purified protein was then eluted through addition of 10 mL of the maltose elution buffer. This buffer consisted of 10 mM Maltose, 20 mM Tris, 200 mM NaCl, 1 mM EDTA (pH 7.4). Sample quantity and quality was determined through the absorbance values at A_{280} on NanoDrop®ND-1000 Spectrophotometer (Thermo Scientific) and SDS-PAGE.

2.5 Metaldehyde degradation assays

2.51 Metaldehyde Detection through Gas Chromatography

To remove cell debris and other particulate matter, samples were centrifuged for 10 mins at 4000 g. From the supernatant, 400 μ L was extracted and mixed with 500 μ L of dichloromethane within a glass chromatography vial. Samples were then vortexed for 30 sec and allowed to settle for 30 mins. From the lower organic phase, 5 μ L was extracted and injected into the inlet of an Agilent 7820A gas chromatograph. The Agilent 7820A was fitted with a flame ionisation detector and HP-5 column. The parameters used were based on the work by Tao and Fletcher (2013). Briefly, the injection model was splitless. The injector and detector temperatures were set to 100 °C and 180 °C, respectively. The carrier gas utilised was helium (5.0 mL/ min) whilst hydrogen (30 mL/min) and compressed air (5 mL /min) were used as combustion gases for the FID. The oven temperature programme allowed for an initial temperature at 70 °C to be held for 1 min, followed by an increase to 100 °C at a rate of 7.5 °C / min. Following this, a target

temperature of 180 °C was set with an increase rate of 15 °C /min. This temperature was held for 6 mins. Through this method, a limit of quantification (LOQ) and limit of detection of 0.15 mg/L and 0.05 mg/L respectively was achieved. To quantify metaldehyde concentrations within samples, a calibration curve was created through generation of appropriate standards within minimal salt media (MSM) as shown in Figure 2-1. Where the area exceeded the LOQ, appropriate dilutions using MSM were undertaken.

GC Metaldehyde Calibration Curve

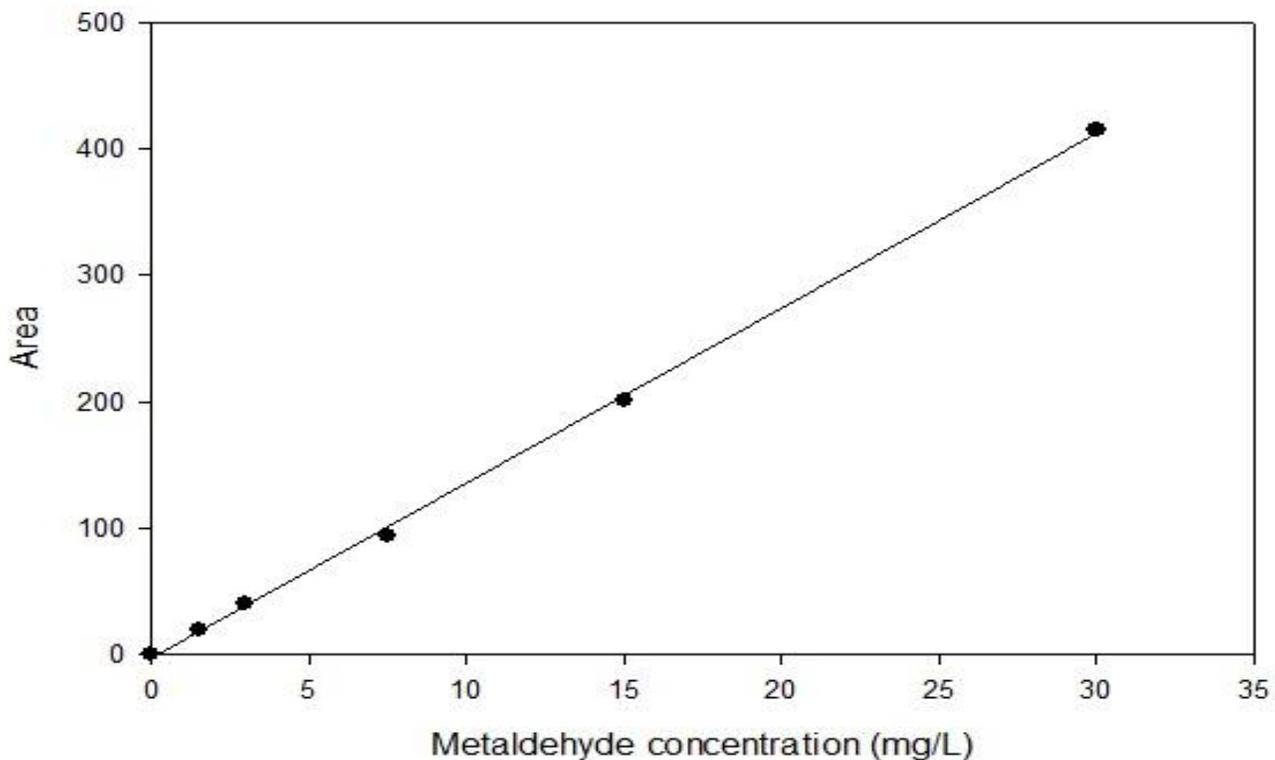


Figure 2-1: GC metaldehyde calibration curve generated using MSM spiked with metaldehyde and calculating the appropriate peak area. The R^2 for the calibration line was 0.99.

2.52 Degradative ability of *A. calcoaceticus* E1 mutants

In triplicate for each sample, each mutant and the *A. calcoaceticus* wild type were used to inoculate 10 mL of LB media. The samples were grown overnight at 30 °C at 180 rpm. Following incubation, sample OD₆₀₀ was calculated through spectrometer using a Jenway 6300. Each sample was normalized to an OD₆₀₀ of 1.0 for a total volume of 1 mL. Samples were centrifuged at 8,000 rpm for 5 minutes and the supernatant discarded. The pellet was then washed twice using 1 mL MSM using the same centrifuge conditions discussed above. Finally, the pellet was resuspended into 1 mL of MSM supplemented with 15 mg/L metaldehyde. Samples were incubated within 2 mL Eppendorf tubes for 3 hours at 30 °C. Following this, samples were subjected to GC analysis.

2.53 Determining metaldehyde degradation within *E. coli* pBR322 constructs

To determine the degradative ability of the heterologous expressed *mahX*, *mahY* and *mahXY* and *mahXYZ*, degradation assays were performed. *E. coli* DH5 α containing the relevant constructs were grown overnight in 10 mL of LB supplemented with tetracycline (10 μ g/mL) at 30°C. Following growth, cell OD₆₀₀ was calculated and the cells normalised to an OD₆₀₀ of 1.0 for a final volume of 10 mL. Samples were transferred to a sterile 50 mL Falcon tube and centrifuged for 10 mins at 6,000 g and the pellet washed in 10 mL of MSM. Following this, cells were centrifuged again for 10 mins at 6,000g and the pellet resuspended into 10 mL of MSM supplemented with metaldehyde (150 mg/L). Samples were then incubated at 30 °C at 180 rpm.

Throughout the assay, 1 mL was extracted and transferred into a 2 mL Eppendorf tube. Samples were then frozen at -20°C until GC analysis.

5.54 Determining metaldehyde degradation with cell free extracts

The degradative ability of an overexpressed protein was investigated through normalising cells by OD_{600} following induction. Within a 50 mL falcon tube, 10 mL of fresh LB was inoculated with *E. coli* BL21 (DE3) containing the relevant vector. Investigations utilising the pET21a vector involved supplementing the media with ampicillin ($100\ \mu\text{g}/\text{mL}$). Where pETFPP constructs were studied, kanamycin ($50\ \mu\text{g}/\text{mL}$) was supplemented. Samples were grown overnight at $37\ ^{\circ}\text{C}$ at 220 rpm. Following growth, 50 mL of fresh LB supplemented with the relevant antibiotic, was inoculated with 500 μL of overnight culture. Samples were then incubated for ~ 3 hours in order to reach mid-log phase (0.4- 0.6). When the appropriate sample OD_{600} was achieved, the induction compound Isopropyl β - d-1-thiogalactopyranoside (IPTG) was added to provide a final concentration of 1 mM. With the exception of constructs expressing MBP-MahS and MBP-MahR, all samples were incubated at $30\ ^{\circ}\text{C}$ overnight. Where constructs expressing MBP-MahS and MBP-MahR were analysed, an induction time of 3 hours was undertaken.

5.55 Metaldehyde disappearance assay with cell-free extracts

To determine the degradative activity of the soluble crude extract of MBP-MahX and MBP-MahS, degradation assays were undertaken. For MBP-MahX, the soluble crude extract was obtained as discussed in section 2.45. The extract for MBP-MahS was obtained through similar methods with the difference of a 3-hour induction. The initial soluble crude extract assay aimed to determine

the degradative activity of the enzyme when supplemented with $\text{Fe}_2\text{SO}_4 \cdot (\text{H}_2\text{O})_7$ (SLS), L- ascorbic acid (LAA)(Sigma) and alpha ketoglutarate (αKG)(Sigma). Total protein was calculated based on the A_{280} identified through NanoDrop analysis. In triplicate, to each 50 mL falcon tube 40 mg of soluble crude extract was added. The samples were then supplemented with the cofactors to ensure a final concentration of 1 mM was obtained. To phosphate buffer saline (PBS) (pH7.4) consisting of 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , KH_2PO_4 , 15 mg/L metaldehyde was added. This mixture was then added to the crude extract samples and incubated for 3.5 hours at 30 °C at 180 rpm. Samples were taken at the end point, transferred to 1.5 mL Eppendorf tubes and stored at -20 °C until further analysis.

Where cofactor analysis was undertaken, lack of enzyme or cofactor was substituted with PBS. Sample conditions were the same as discussed above with the variation in the time points collected. Within the cofactor crude extract assay, samples were extracted every 10 mins for metaldehyde detection.

2.56 Effect of reducing agent on metaldehyde degradation

Identification of the role of the reducing agents within metaldehyde degradation was undertaken through observing the metaldehyde concentration following 3 hours incubation. Crude extract was prepared using the same methods as discussed above. Within a 50 mL falcon tube, 40 mg of crude extract supplemented with 1 mM of $\text{Fe}_2\text{SO}_4 \cdot (\text{H}_2\text{O})_7$ and 1 mM αKG (Sigma-Aldrich) and 1 mM of either Dithiothreitol (DTT) (Promega), LAA (Sigma-Aldrich) or nicotinamide adenine dinucleotide (NADH) (Sigma-Aldrich). Reaction mixtures were prepared to a final volume of 10 mL containing PBS supplemented with 150 mg/L metaldehyde. Control samples were undertaken

containing PBS supplemented with metaldehyde, α KG and $\text{Fe}_2\text{SO}_4 \cdot (\text{H}_2\text{O})_7$. Samples were incubated for 3 hours at 30 °C whilst shaking at 180 rpm. Following incubation, samples were transferred into 2 mL Eppendorf tubes and frozen at -20 °C until further analysis.

2.57 Metaldehyde disappearance assay with purified protein samples

Metaldehyde degradation assays were conducted for protein purified through IMAC and dextrose purification methods. The IMAC purified protein assay was conducted within a 2 mL Eppendorf tube and consisted of mixing 0.176 mg of purified protein, 1 mM of Fe_2SO_4 , 1mM Lascorbic acid and 1 mM α KG with PBS supplemented with 150 mg/L metaldehyde to ensure a final volume of 1 mL. Where cofactors or enzyme was absent, PBS was added as a substitute. Samples were performed in triplicate at 30 °C whilst shaking at 180 rpm. Following 3.5 hours incubation, samples were frozen at -20 °C until GC analysis. The amylose purified protein assay was conducted with a 50 mL falcon and consisted of mixing 4 mg of purified protein with 1 mM of Fe_2SO_4 , 1mM L-ascorbic acid and 1 mM α KG with PBS supplemented with 150 mg/L metaldehyde to ensure a final volume of 10 mL. As described above, samples were stored at -20°C following 3.5 hour incubation at 30 °C.

Chapter 3: Identification and characterization of the gene(s) responsible for the initial step of metaldehyde degradation within *A. calcoaceticus* E1

3.1 Introduction

A previous study from our laboratory isolated and characterized the degradative ability of several strains capable of metaldehyde degradation and utilization (Thomas *et al.*, 2017). However, the mechanism by which this degradation occurs and the enzyme(s) responsible for the initial degradative step were not identified. The ability to identify and characterize novel catabolic enzymes allows for the development of new methods and technologies and as such is one of the most important steps in the development of bioremediation strategies (Widada *et al.*, 2002). The identification of novel catabolic genes can be achieved through numerous different approaches. These approaches include, but are not limited to; induction analysis, clone libraries, comparative genomics and mutagenesis. All of the above-mentioned approaches were attempted previously by Thomas (2016), in an attempt to identify the gene(s) responsible for metaldehyde degradation.

Regulation of gene expression prevents the synthesis of unnecessary enzymes, thereby saving energy and nutrients. Where expression is induced by the substrate, differences in either gene or protein expression levels may be used to identify the catabolic genes. By comparing either transcriptome or the proteome with and without substrate addition, inducible genes can be identified and information regarding the metabolic pathway can potentially be obtained (Pankaj *et al.*, 2016; Wang *et al.*, 2018). When a xenobiotic degrading microorganism is isolated and culturable, protein expression analysis provides a simple and effective method to identify inducible catabolic genes. Comparison of the proteomic induction profiles using traditional gel separation methods, such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or 2-dimensional electrophoresis (2D-GE), has proved to be an effective gene identification method (Seung *et al.*, 2007; Pankaj *et al.*, 2016). As such, this approach was used in the identification of a novel hydrolase capable of mineralizing the herbicide linuron. Extraction of the protein band/spot identified in the analysis, followed by mass spectrometry can reveal the sequence of the protein and ultimately the inducible gene (Bers *et al.*, 2011). Previous analysis, conducted by Thomas, (2016) performed an induction assay for *A. calcoaceticus* E1 using an SDS-PAGE analysis coupled with matrix-assisted laser desorption ionization time-of-flight analysis. Through comparing the protein profile of *A. calcoaceticus* E1 grown in MSM with metaldehyde against that of it grown in MSM with acetate, a visible band was observed in the former. MALDITOF analysis, coupled with InterPro examination, revealed the induced protein to be an aldehyde dehydrogenase (ALDH). This identified protein was later ruled out as being a metaldehyde degrading protein due to an identical sequence being present in the non-metaldehyde degrading type strain *Acinetobacter* RUH2202. No further differences in the protein

bands were observed. As the SDS-PAGE induction analysis identified no observable change in the protein profile in *A. calcoaceticus* E1, alternative identification methods would be required to identify the metaldehyde degradation gene.

Another potential approach in identifying the metaldehyde degrading gene is using genomic libraries. Genomic libraries are a traditional technique used in molecular biology, and when coupled with the ability to isolate and culture xenobiotic-degrading microorganisms, they can provide a powerful tool in the identification of catabolic genes (Widada *et al.*, 2002). Following the extraction and then either shearing or digesting of the genomic DNA of the xenobiotic degrading microorganism, the DNA fragments are cloned into a suitable vector and then transformed into an appropriate host organism (Widada *et al.*, 2002). Following this transformation, either a selection step, such as growing the colonies on the xenobiotic compound, or a screen approach, ranging from a colorimetric assay to observing zones of clearing surrounding the clones, is used to determine the presence of the catabolic gene(s) (Ruan *et al.*, 2013; Sadauskas *et al.*, 2017). Once the catabolic activity has been verified, the degradative colonies can be subjected to nucleotide sequencing and the relevant gene(s) identified. This approach has the advantage of requiring little or no information regarding the metabolic pathway or the degradative microorganism. This strategy has been shown to be very effective in the identification of numerous novel catabolic genes (Goyal and Zylstra, 1996; Jun Zhang *et al.*, 2012; Ruan *et al.*, 2013). One example is the identification an esterase capable of hydrolyzing a pyrethroid pesticide; by coupling the genomic library with a fast blue RR colorimetric screen, the degradative clone was successfully identified and verified (Ruan *et al.*, 2013). Despite the fact that numerous novel degradative genes have been identified using the library approach, several

limitations do exist. As the utilization of complex molecules can require multiple novel degradative steps, several novel genes may be required simultaneously to convey this ability across to another organism (Nordin *et al.*, 2005; Cámara *et al.*, 2007). Therefore, clones containing the degradative genes would not be able to grow on the xenobiotic compound and go undetected. A further limitation is based on the screening approach to gene identification. The screening reagent is typically a sophisticated substrate with chromogenic properties and as such may require information regarding either the metabolic pathway or the degradative mechanism in order to select the correct detection method (Cámara *et al.*, 2007). Due to the limited information regarding the degradative pathway and the ability to convey this degradative ability into a competent host strain, this approach was not used as a gene identification method for this investigation. The clone library method was previously attempted by Thomas (2016) within *E. coli*. The library generated no transformants capable of growing on metaldehyde plates.

Where multiple xenobiotic-degrading strains have been isolated, whole-genome sequencing and comparative genomics can be very powerful in the identification of novel genes. By identifying genes shared between three strains capable of mineralizing isoproturon, Yan *et al.*, (2016) were able to narrow down the number of candidates to 84 gene sequences. From this number, manual curation was possible and as such were able to identify and verify the isoproturon-mineralizing genes. This method has also been used to identify genes responsible for the degradation of other xenobiotic compounds such as chloroacetanilide (Cheng *et al.*, 2019). Since this strategy relies on a similar degradative mechanism and sequences to identify the novel gene(s), it is most powerful with a large dataset of degrading microorganisms or when combined

with complementary identification methods such as mutagenesis. One drawback is that as xenobiotic compounds may be degraded using different degrading pathways in different microorganisms, degrading genes may not necessarily be identified using this approach. This comparative genomic method was undertaken by Thomas (2016), using a BLAST Score-Ratio (BSR) analysis. Through comparing each predicted protein sequence against the reference genome and then against two other 'query' proteomes, rapid comparisons were made and the similarity between numerous genomes were evaluated. Using predicted protein sequences from *A. calcoaceticus* E1 and *Variovorax* E3 and non-metaldehyde degrading type strains *Acinetobacter* RUH2202, 34 predicted sequences were identified. The output results were for BSR scores of > 0.5 where predicted sequences of *A. calcoaceticus* E1 had a greater homology to *Variovorax boronicumulans* E3 than *A. calcoaceticus* RUH 2202. Of these 34 genes, none had properties consistent with a metaldehyde degrading gene, leading us to conclude that *A. calcoaceticus* E1 and *V. boronicumulans* E3 did not share a homologous metaldehyde degrading gene. In the study here, we used comparative genomics with a larger number of metaldehyde degrading genes which gave this approach a greater feasibility.

A final approach is through mutagenesis. Unlike the genomic library methods, which seek to insert the catabolic gene(s) into a non-native host, the mutagenesis approach aims to disrupt these genes in the degradative microorganism. As spontaneous mutations are a rare occurrence, active approaches to cause gene disruptions are undertaken (Widada *et al.*, 2002). For the identification of novel genes, these approaches can be simply divided into insertional mutagenesis and random mutagenesis, depending on the identification of the mutation sites.

Transposons are mobile elements capable of random insertion within a host's genome. These mobile elements can therefore be used as powerful molecular tools to disrupt genes throughout the genome, whilst simultaneously inserting selectable markers into the host genome to identify mutated genes (Kalindamar *et al.*, 2019). The advantage of this insertional mutagenesis approach, relative to random mutagenesis, is a reduced lethality rate and a potentially higher mutation rate (Seifert *et al.*, 1986). Transposon mutagenesis has been used successfully to identify numerous catabolic genes capable of degrading dichlorprop, long chain alkanes and rubber, amongst other compounds (Schleinitz *et al.*, 2004; Throne-Holst *et al.*, 2007; Kasai *et al.*, 2017; Qiu *et al.*, 2018). As transposons are typically transformed into the host cell through a suicide vector, the main limitation regarding this approach is that it is largely restricted to naturally competent microorganisms (Rabausch *et al.*, 2013). Thomas (2016) attempted to use transposon mutagenesis to generate a mutant library of *A. calcoaceticus* E1 strains unable to degrade metaldehyde. However, due to the non-model nature of *A. calcoaceticus* traditional molecular methods proved difficult. As transformation of *A. calcoaceticus* E1 strain, with the transposon vector led to stable maintenance of the supposed 'suicide plasmid' containing the transposon, screening for transposon mutants became impractical.

Random mutagenesis is the second mutagenic approach used to identify novel degradative genes. The aim of this strategy is to create either point mutations and/or deletions throughout the genome. This can be achieved typically through either radiation, such as ultraviolet (UV), or through the use of mutagenic compounds, such as ethyl methanesulfonate

(EMS), methyl methanesulfonate (MMS) and acridine, amongst others (Todd *et al.*, 1979; Ferguson and Denny, 1991; Geissdorfer *et al.*, 1995; Pan *et al.*, 2012; Shibai *et al.*, 2017). Chemical mutagens, which are typically favoured in random mutagenesis, exist in several classes with numerous modes of actions, ranging from nucleotide base analogs, whereby they resemble purine and pyridines, to chemical modifiers, where they are capable of modifying a single pair to create faulty base pairing (Goncharova and Kuzhir, 1989; Holroyd and Van Mourik, 2015). Using random chemical mutagenesis with the compound EMS, researchers obtained mutants deficient in the ability to degrade the alkane dodecane in *A. calcoaceticus*, ultimately leading to the identification of the initial dodecane degradative gene (Geißdörfer *et al.*, 1995). With the reduced costs to whole genome sequencing and the increased accessibility of bioinformatics packages, random mutagenesis is a gene identification method that can increasingly be applied to a wide variety of microorganisms (Ruan *et al.*, 2013). The main limitation of both mutagenesis methods are the non-targeted disruptions, whereby downstream metabolic genes can be mutated and therefore may give misleading results in the selection procedure (Ruan *et al.*, 2013). Although random mutagenesis provides an advantage over insertional mutagenesis by requiring no complex molecular techniques and therefore needing no information regarding the microorganism, it does suffer with the shortcoming of the handling of highly carcinogenic compounds.

As the EMS method has been demonstrated to be effective in gene identification with *A. calcoaceticus* previously, requires no further development of tools for genetic manipulation of

the degradative organisms, this approach was utilized in the identification of the initial metaldehyde degrading enzyme in *A. calcoaceticus* E1.

3.2 Results

3.21 Using chemical mutagenesis to obtain mutants lacking the ability to degrade metaldehyde

With the aim to generate and isolate *A. calcoaceticus* E1 mutants deficient in the ability to degrade and utilize metaldehyde, a random chemical mutagenesis screen was undertaken. The metaldehyde degrading strain *A. calcoaceticus* E1 was exposed to the mutagen ethyl methanesulfonate (EMS) and then plated at various dilutions onto LB agar. Following incubation, colonies were picked from the LB plates and plated onto both metaldehyde and acetate plates. Of the 1,000 colonies that were replicate plated following EMS treatment, four colonies were identified that showed a lack of growth on plates containing metaldehyde as the sole carbon source yet maintained the ability to grow on acetate as a control. From the acetate colonies, mutants unable to grow on metaldehyde were streaked onto fresh metaldehyde and acetate plates. This allowed for both the verification of the previous findings and for a morphological characterization of the single colonies to be undertaken. The observations following the restreaking verified the mutants' inability to utilize metaldehyde. To determine whether the mutation was due to EMS treatment or absence of selective pressure or through alternative

methods, a no EMS control assay was undertaken. Of the 1,000 colonies examined in the control assay, none were identified that had lost the ability to utilize metaldehyde.

Following the isolation of the four *A. calcoaceticus* E1 mutants unable to utilize metaldehyde, henceforth known as mutants 28, 73, 140 and 206, the ability to grow and degrade metaldehyde whilst in liquid MSM supplemented with metaldehyde was quantified. Samples were grown on LB plates and used to inoculate metaldehyde media to a starting OD₆₀₀ 0.001. The assay, as shown in Figure 3-1 and Figure 3-2, demonstrated growth of the *A. calcoaceticus* E1 wild type (WT) to a mean OD₆₀₀ of 0.203 within 26 hours of incubation. The four mutants showed no observable change in OD₆₀₀ throughout the assay. GC analysis revealed no significant change in metaldehyde concentration within the four mutants throughout the assay ($p < 0.05$). The *A. calcoaceticus* E1 WT showed degradation below the limit of detection within 36 hours of incubation.

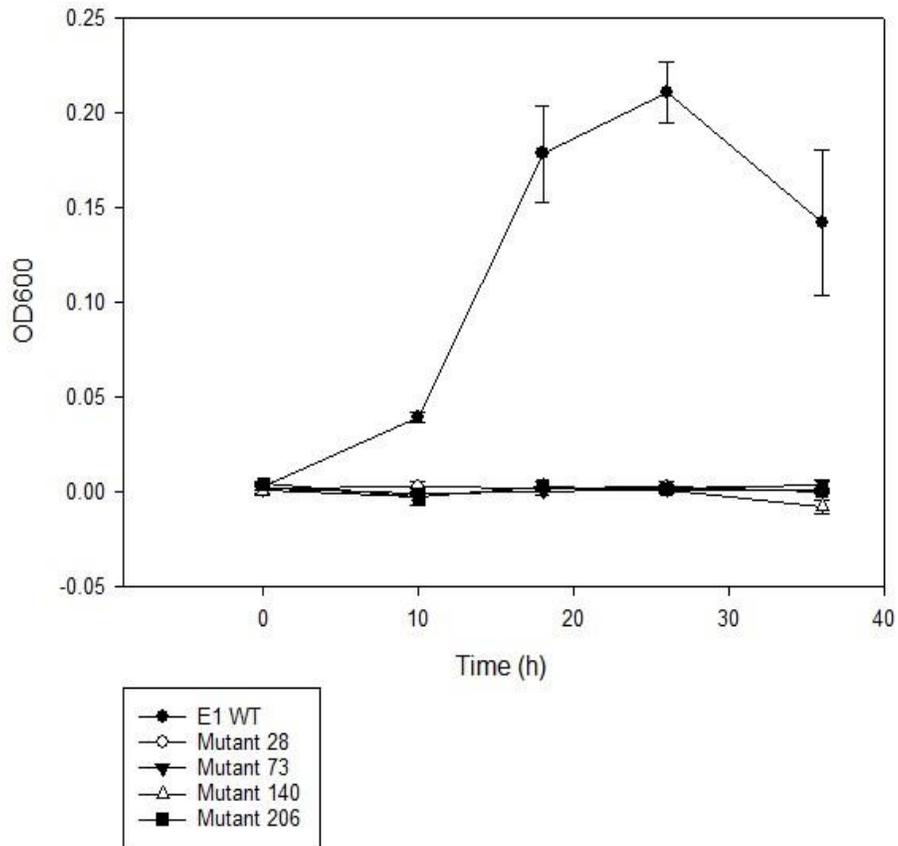


Figure 3-1: Mean cell OD₆₀₀ for A. calcaceticus E1 WT and 4 isolated mutants. Cells were grown in MSM supplemented with 150 mg/L metaldehyde. Error bars give the SD of biological replicates (n=3).

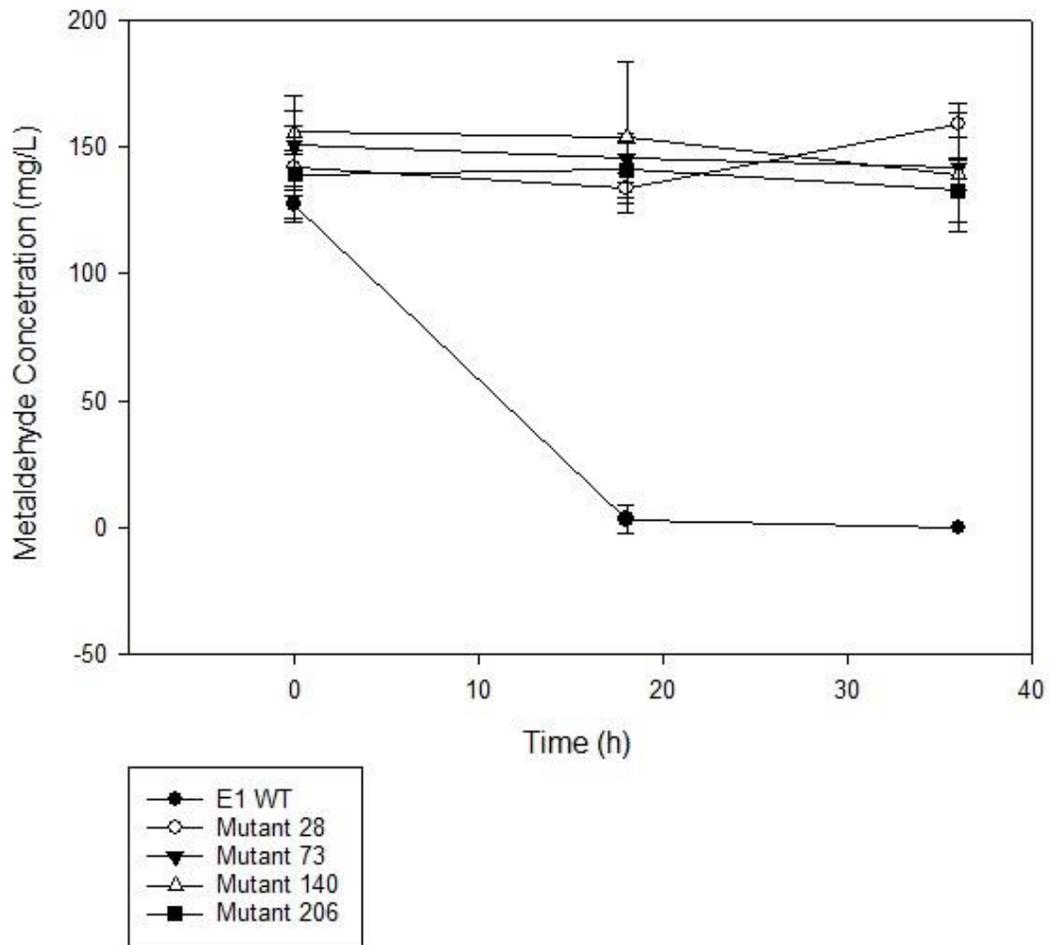


Figure 3-2: Metaldehyde concentration following inoculation with OD_{600} 0.001 of E1 WT or 4 isolated mutants into MSM supplemented with 150 mg/L metaldehyde. Error bars show the SD of biological replicates ($n=3$).

The lack of growth in metaldehyde media may be explained through the disruption of an intermediate metabolic enzyme, instead of the initial degradative enzyme. A degradation assay

where cells were pregrown in LB and then normalized by OD_{600} was undertaken. The results as shown in Figure 3-3 demonstrated no observable change in metaldehyde concentration in the four mutants. Metaldehyde was degraded below the limit of detection in the *A. calcoaceticus* E1 WT strain. The lack of growth in metaldehyde media and the inability to degrade metaldehyde demonstrate the isolation of mutants containing disruption of the crucial degradative gene. As the previous experiments demonstrated that the mutants isolated were deficient in the ability to degrade metaldehyde, the identification of the disrupted gene(s) was undertaken. Due to the high availability and accessibility of whole genome sequencing platforms, the four mutants and E1 WT were subjected to Illumina whole genome sequencing through MicrobesNG. The quality statistic for the five strains submitted are shown in table 3-1.

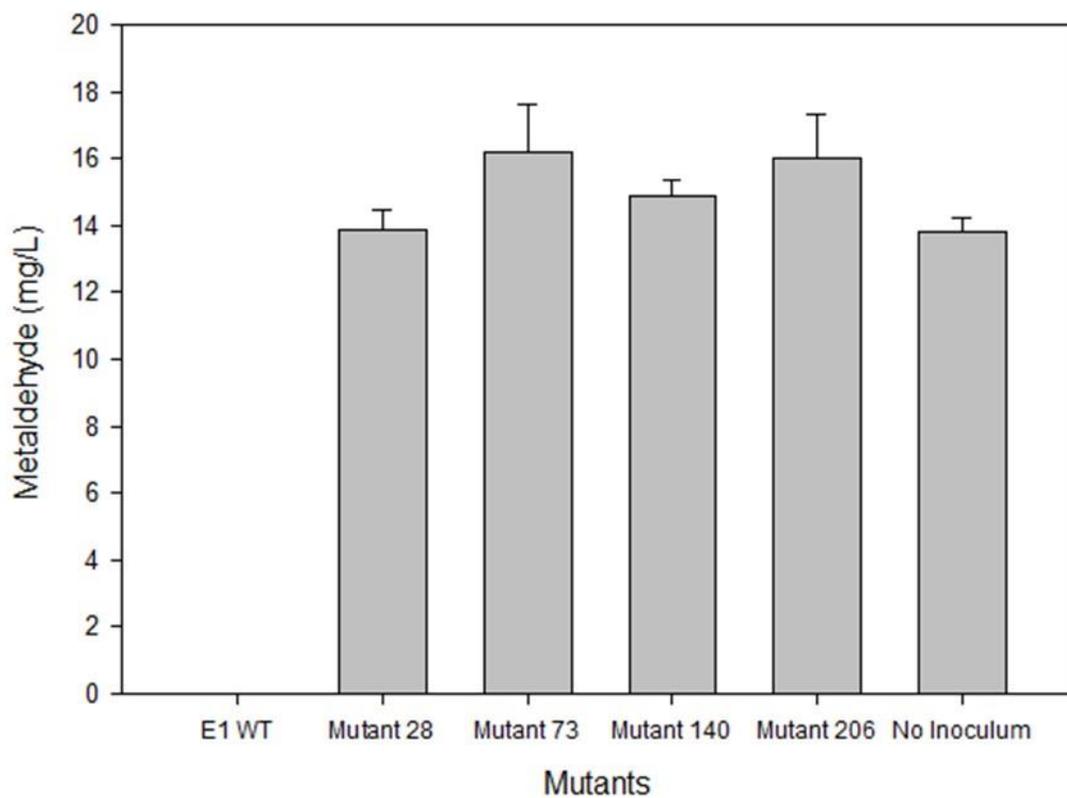


Figure 3-3: Metaldehyde degradation assay with samples pregrown in LB and normalised to $OD_{600} 1.0$. Cells were incubated with 15 mg/L metaldehyde for 3 hours. Samples were performed in biological triplicate with error bars showing the standard deviation.

Table 3-1: Quality statistics for whole genome sequencing of *A. calcoaceticus* E1 and the four mutants deficient in metaldehyde degradation. The N50 value is defined as the sequence length of the shortest contig at 50% of the total genome length.

	# contigs	Largest contig	Total length	N50	GC (%)	Coverage
E1 WT	98	377626	4383345	131564	38.7	30x
Mutant 28	106	377746	4385996	117495	38.7	30x
Mutant 73	103	377626	4380733	131794	38.7	30x
Mutant 140	97	377626	4382723	120850	38.7	30x
Mutant 206	102	377746	4383926	131794	38.7	30x

3.22 Bioinformatics analysis of *A. calcoaceticus* E1 mutants deficient in metaldehyde degradation

In order to determine which protein sequence(s) are responsible for the initial step of metaldehyde degradation, a BLAST analysis was undertaken. Using BLASTP, predicted protein sequences from *A. calcoaceticus* E1 WT was contrasted against that of the four mutants. This led to the identification of 25 sequences that possessed mutations between the *A. calcoaceticus* E1 WT and other strains. Through manual analysis of this list, based on which sequences showed <100 % identity between the wild type and all four mutants, a list of 14 predicted protein sequences remained. Through BLAST and InterPro analysis, the sequences were investigated to determine their potential as a metaldehyde degrading enzyme. Nomenclature surrounding the predicted proteins was based on their CDS position within *A. calcoaceticus* E1's genome.

Table 3.2- Bioinformatic analysis of the predicted protein sequences identified through comparative genomic analysis of E1 mutants unable to degrade metaldehyde. tBLASTn and InterPro analysis was undertaken to determine highest scoring BLAST match and domains to predict protein function.

Sequence ID	Highest scoring sequence	ID	Predicted Function
Sequence_0625	Uncharacterized protein [<i>Acinetobacter pittii</i>]	90%	Unable to predict
Sequence_2462	Putative pilus assembly protein FlE [<i>Acinetobacter baumannii</i>]	87%	Pilius assembly
Sequence_3430	AsmA family protein [<i>Biostraticola tofi</i>]	30%	Outer membrane protein assembly
Sequence_3463	2oxoglutarate dehydrogenase E1 component [<i>Acinetobacter pittii</i>],	85%	alpha-ketoglutarate decarboxylation
Sequence_3495	Hypothetical protein [<i>Acinetobacter colistiniresistens</i>]	99%	Protein binding and folding
Sequence_3935	Lytic transglycosylase domain-containing protein [<i>Acinetobacter calcoaceticus</i>]	100%	Murein degradation
Sequence_4226	Phytanoyl-CoA dioxygenase family protein [<i>Sphingobium japonicum</i>]	49%	Dioxygenase with potential for metaldehyde degradation
Sequence_4238	Hypothetical Protein AXA63_10640 [<i>Acinetobacter baumannii</i>]	50%	Unable to predict
Sequence_4275	Uncharacterized protein [<i>Salmonella sp.</i> NCTC 7297]	82%	Unable to predict
Sequence_4276	N/A	N/A	Unable to predict
Sequence_4277	Ubiquinone biosynthesis monooxygenase UbiB [<i>Escherichia coli</i> IS35]	100%	ubiquinone biosynthesis monooxygenase
Sequence_4278	MULTISPECIES: YhfH family protein [<i>Bacillus</i>]	100%	Unable to predict
Sequence_4279	PTS fructose-like transporter subunit IIB [<i>Escherichia coli</i>]	99%	phosphotransferase activity
Sequence_4280	N/A	N/A	Unable to predict

3.23 Identification of the metaldehyde degrading gene(s) through BLAST analysis

Further analysis into gene identification utilized the information obtained from V. Castro and J. Thomas of the Moir group (Thomas, 2016; Castro-Gutiérrez *et al.*, 2020). Blast-score-ratio (BSR) analysis was performed on two metaldehyde degrading strains (*A. calcoaceticus* E1 and *Acinetobacter bohemicus* Strain C), and their closely related type strains (RUH2202 and *A. bohemicus* ANC3994, respectively, neither of which can degrade metaldehyde). Protein sequences were identified that were highly similar within the different species yet absent within the non-degrading strains. This approach generated an output with 87 predicted protein sequences. The 87 predicted protein sequences were manually examined to remove sequences that had predicted functions related to non-catalytic activities and therefore highly unlikely to be responsible for the initial degradative step. The predicted functions removed included sequences responsible for mercury resistance, cation efflux pumps, DNA interaction motifs, such as transposases and integrases, and sequences < 65 AA residues in length. Following the manual examination of the dataset, 27 sequences remained. This curated list was used as a tBLASTN query to search against the whole genome sequences of the E1 WT and the four mutants. The aim of the BLAST analysis was to identify sequences from the BSR analysis that also contained mutations in strains that had lost the ability to degrade metaldehyde in the chemical mutagenesis study. Of the 27 sequences examined, a single predicted sequence, henceforth known as MahX, was identified as having <100 % identity in all the four mutants of the *A. calcoaceticus* E1 unable to degrade metaldehyde as shown in Table 3-3. This sequence was also identified in the previous

comparative genomic analysis under the sequence identity of 4226. Examination of the predicted sequence analysis revealed that within Mutant 73 two other predicted genes, contiguous with *mahX* were mutated. . These predicted protein sequences were labelled as MahY and MahZ . These sequences were further investigated along with MahX. BLAST examination of Mah X, Y and Z revealed that in mutant 73, the lower identity scores obtained were not due to in gene mutations within the sequences but contained whole gene deletions, and the alignments were against other significantly divergent sequences found within the *A. calacoaceticus* E1 genome. Analysis revealed that MahX was also identified in the previous bioinformatic analysis under the title of sequence_4226.

Table 3-3: Percent identity results with <100 % sequence similarity relative to the E1 WT. Predicted protein sequences of MahX, MahY and MahZ compared within each E1 mutant deficient in metaldehyde degradation.

	MahX	MahY	MahZ
E1 WT	100%	100%	100%
Mutant 28	96%	100%	100%
Mutant 73	31%	46%	44%
Mutant 140	96%	100%	100%
Mutant 206	96%	100%	100%

3.24 Characterization of the identified metaldehyde degrading gene

3.241 Examination of MahX

BLAST and InterPro analysis was conducted for MahX to further examine similar sequences and identify domains of interest. The results of the analysis, displayed in figure 3-4, show the high alignment similarity between MahX and a closely related phytanoyl-CoA dioxygenase family protein. InterPro analysis, shown in figure 3-5, also identified conserved domains related to the the 2-oxoglutarate (2OG) and Fe (II)-dependent oxygenase superfamily. This protein family are responsible for performing a diverse and complex set of reactions that range from hydroxylations to ring cleavages (Kuiper and Vissers, 2014). With regards to biodegradation, the 2OGFe (II) oxygenase superfamily has been identified as being responsible for the breakdown of numerous compounds including auxin, salicylic acid and jasmonic acid (León et al., 1995; Zhao et al., 2013; Smirnova et al., 2017).

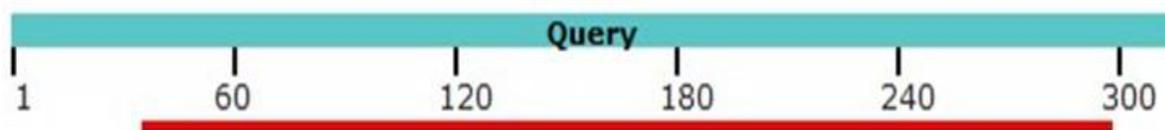


Figure 3-4: Alignment of MahX against the highest scored alignment result of phythanoyl-CoA dioxygenase family protein [Spingobium japonicum]]. Alignment and image was generated through BLASTP. The red alignment color denotes an alignment score of >200.

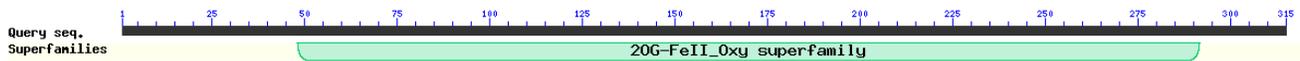


Figure 3-5: Putative conserved domain analysis of MahX generated through BLAST alignment analysis.

Phylogenetic analysis was conducted on MahX. The protein sequence was used to query the BLASTP non-redundant protein sequences database. The top 100 matches were then extracted and the multiple sequences aligned using a MUSCLE server hosted by MEGA. A phylogenetic tree was then generated, shown in Figure 3-6, by using these aligned sequences. The phylogenetic analysis demonstrates the relatively close relationship between MahX and a phytanoyl-CoA dioxygenase family protein [*Sphingobium japonicum*].

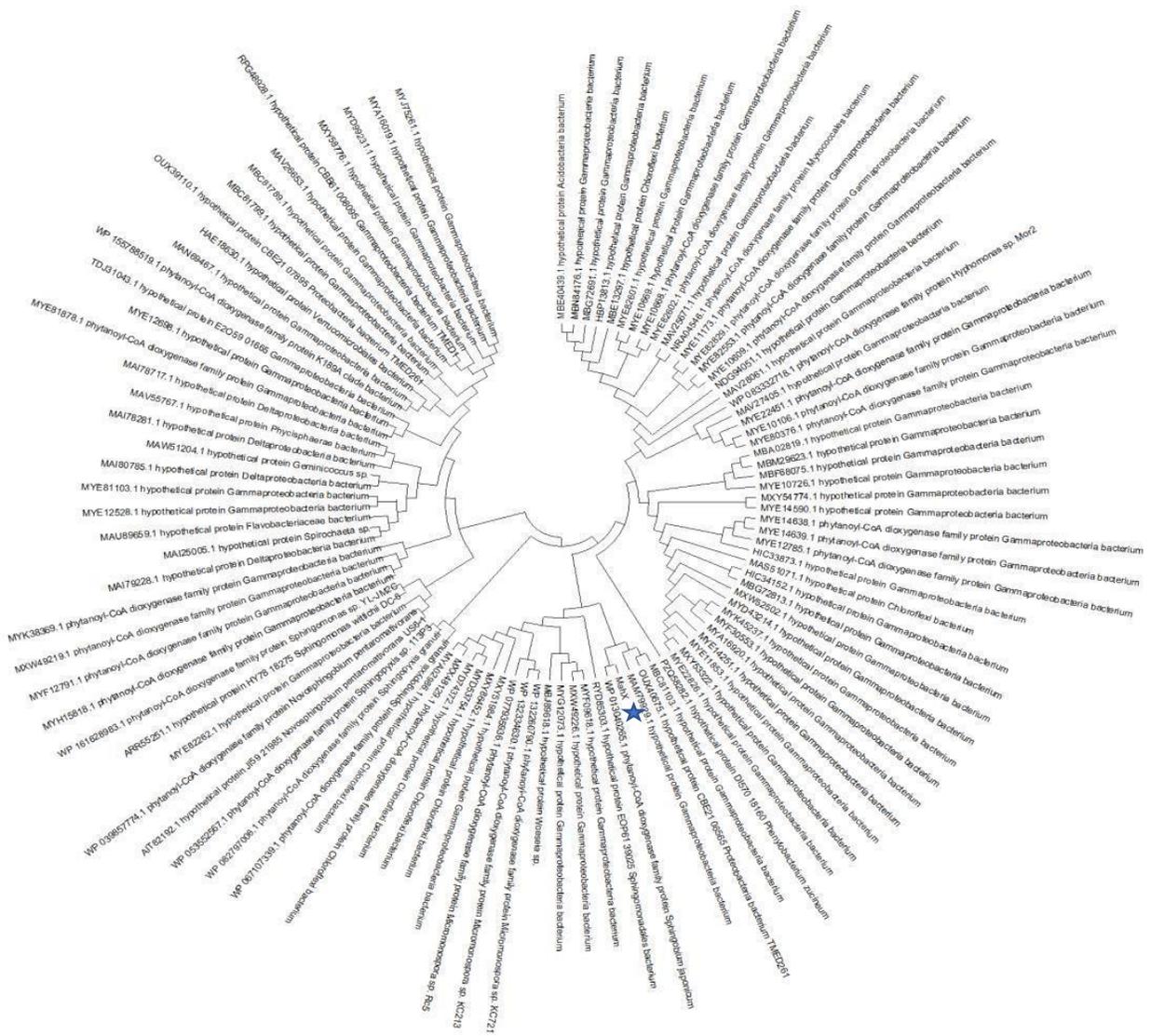


Figure 3-6: A maximum likelihood tree generated using alignments of MahX of *A. calcoaceticus* E1. A star has been used to highlight the MahX sequence.

3.242 Examination of MahY

MahY BLAST results revealed a hypothetical protein [*Chloroflexi* bacterium] as the top hit with a query coverage of 86%, an e-value of $4e-16$ and a 36% identity, as shown in Figure 3-7. The study of the conserved domains identified amino acid residues 4-99 within MahY as belonging to vicinal oxygen chelate (VOC) superfamily (Figure 3-8). This superfamily is found in a variety of structurally related metalloproteins, including type I extradiol dioxygenases and glyoxalase I (Liang *et al.*, 2017).

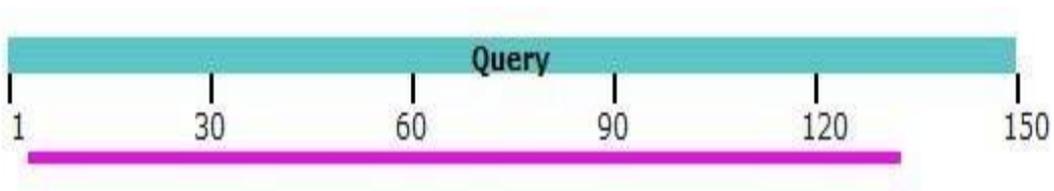


Figure 3-7: BLAST alignment graphic of MahY aligned against the top hit a hypothetical protein [*Chloroflexi* bacterium]. Pink line demonstrates an alignment score between 80-200.



Figure 3-8: Putative conserved domain analysis of MahY generated through BLAST alignment analysis. VOC superfamily domain alignment is displayed.

InterPro analysis revealed the 150 amino acid sequence for MahY was classified as belonging to the homologous superfamily of glyoxalase/bleomycin resistance protein/dihydroxybiphenyl dioxygenase (IPR029068). InterPro failed to identify any GO terms regarding this sequence. Phylogenetic analysis, shown in Figure 3-9, demonstrates the relationship between MahY and the closest related sequences which include hypothetical proteins from *Mycobacterium asiaticum* and a hypothetical Rhodococcus protein.

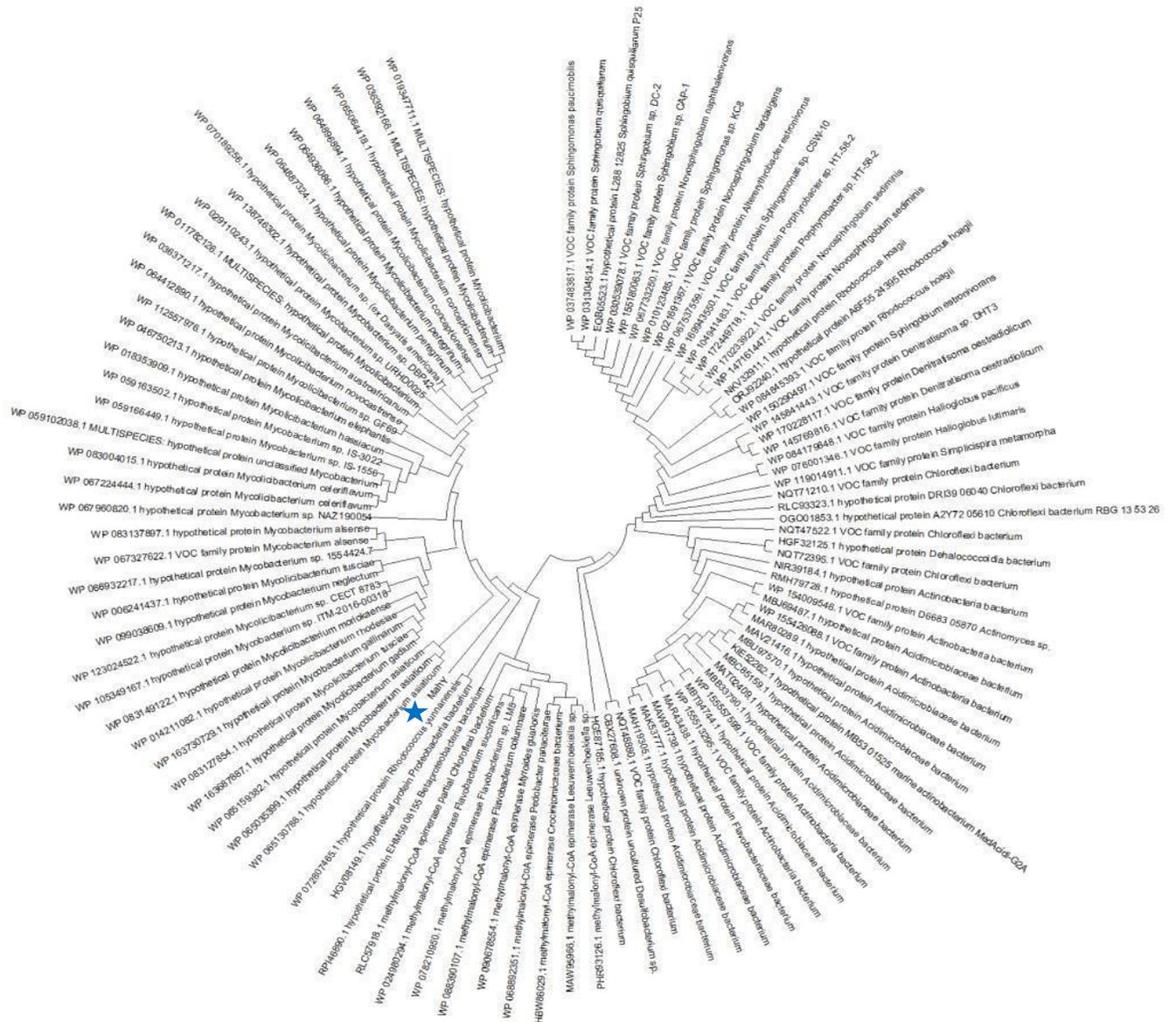


Figure 3-9: Maximum likelihood tree generated using alignments of MahY of *A. calcoaceticus* E1. A blue star further reveals the location of MahY.

3.243 Examination of MahZ

The highest scoring BLASTP score for MahZ was for aldehyde dehydrogenase family protein [*Panacagrimonas perspica*], as shown in Figure 3-10. This result demonstrated an 80% query coverage and 85% identity similarity.

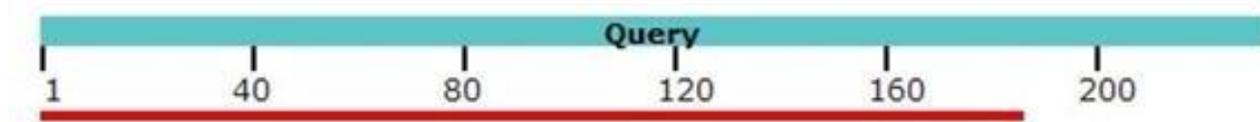


Figure 3-10: BLAST alignment graphic of MahZ aligned against the top hit of aldehyde dehydrogenase family protein [*Panacagrimonas perspica*]. Red line demonstrates an alignment score of ≥ 200 .

Investigation into the conserved domain revealed a hit between the amino acid residues of 25185 as belonging to the ALDH and PLN02466 superfamily, shown in Figure 3-11. InterPro analysis revealed the 232 amino acid sequences of MahZ identified as containing the domain of an aldehyde dehydrogenase between the residues of 30-186, shown in Figure 3-12.

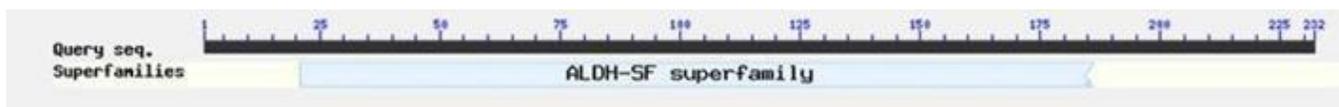


Figure 3-11 Putative conserved domain analysis of MahZ generated through BLAST alignment analysis.

GO terms for biological processes identified oxidation-reduction processes (GO:0055114) and for molecular function analysis as oxidoreductase activity (GO:0016491). The phylogenetic analysis, displayed in Figure 3-13, identified an aldehyde dehydrogenase protein family from *Panacagrimonas perspica* as the closest related sequence.

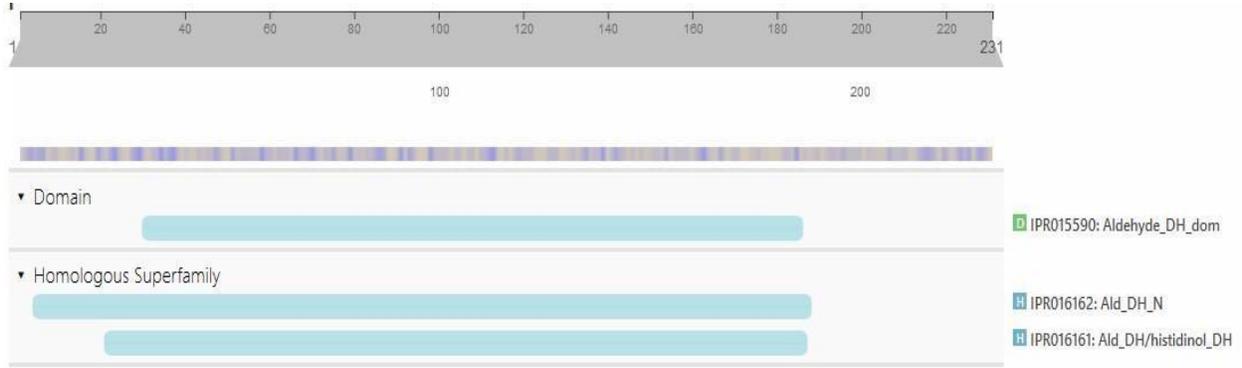


Figure 3-12: Predicted domains and important sites of MahZ generated through InterPro analysis.

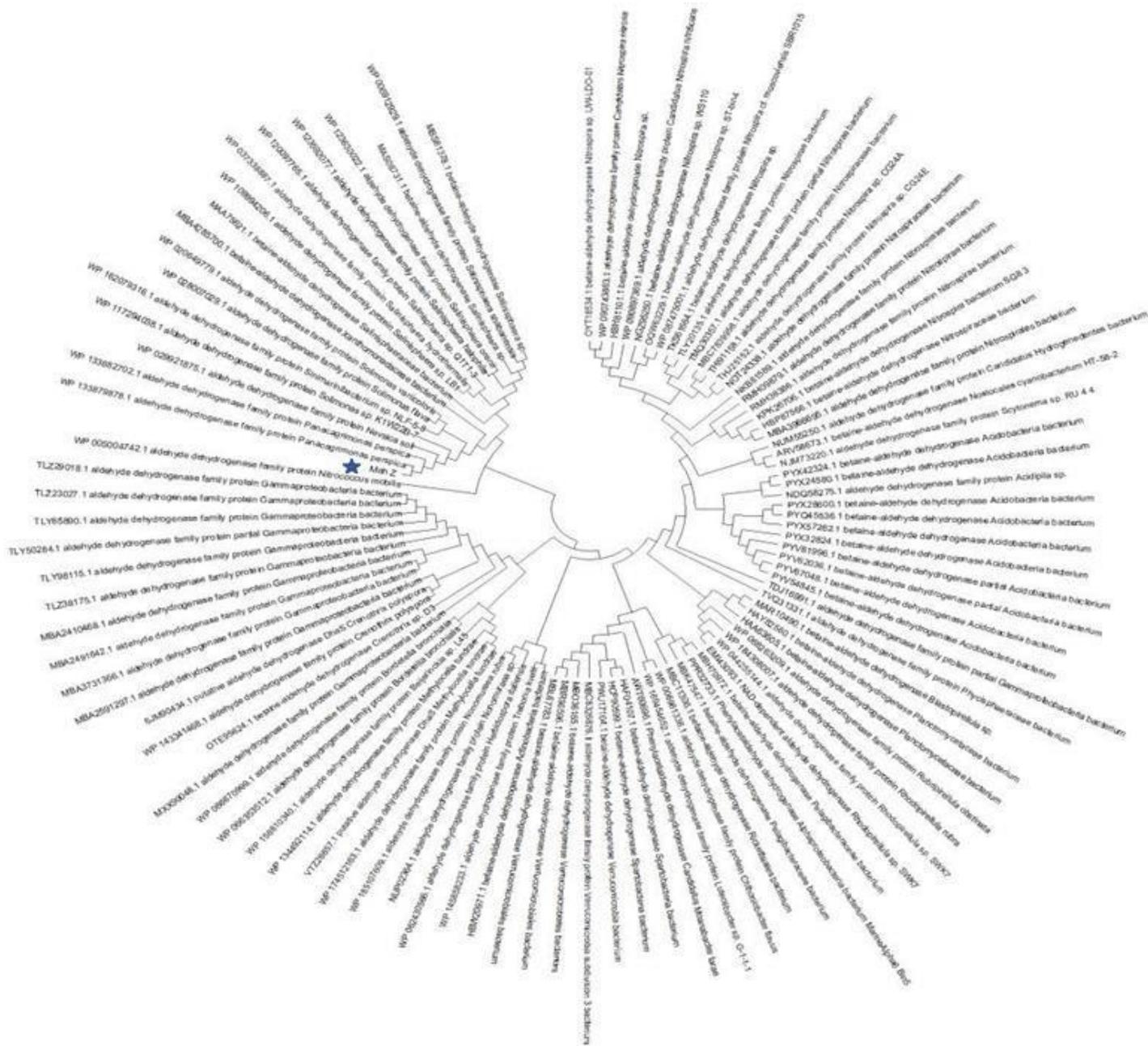


Figure 3-13 A maximum likelihood tree generated using alignments of MahZ of *A. calcoaceticus* E1. A star has been used to highlight the MahZ sequence .

3.25 Examination of the genomic locations of the metaldehyde degrading genes

Further examination of the identified sequences was performed using Sanger's genome browser Artemis. The location and orientation of the predicted sequences within *A. calcoaceticus* E1's genome was investigated. Visual analysis, shown in Figure 3-14, revealed that all three gene sequences reside on a 4229 bp contig along with a fourth gene of 1278 bp. This gene, found in the same orientation as three identified sequences, is located downstream from the gene cluster.

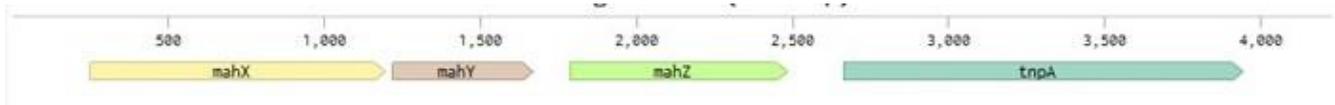


Figure 3-14: Visual schematic, generated using Benchling, of the 4 identified genes on the contig.

The translated sequence for the fourth gene was submitted for BLASTP analysis with a top result being a IS91 family transposase [*Pseudomonas putida*] with a query score and identity similarity of 99%. Based on the identification of a potential transposase on the contig, insertion sequence sites within the contig were investigated. IS finder, using the blastn tool, 100 bps from both ends of the contigs were examined (Siguier *et al.*, 2006). The results, displayed in Table 3-4 and Table 3-5, identified the regions as having relatively high similarity to the IS6 family on both the 5' and 3' end of the contig, with top e-values showing 3e-49 and 1e-51 respectively.

Table 3-4: ISFinder analysis results for the examination of the 5' end of the identified catabolic contig.

Sequences producing significant alignments	IS Family	Group	Origin	Score (bits)	E. value
ISAb52	IS6		Acinetobacter baumannii	190	3e-49
ISOur1	IS6		Oligella urethralis	190	3e-49
IS1006	IS6		Acinetobacter junii	182	6e-47
IS1008	IS6		Acinetobacter calcoaceticus	167	4e-42
IS1007	IS6		Acinetobacter sp.	36.2	0.009
ISAac3	IS200/IS605	IS200	Aggregatibacter actinomycetemcomitans	30.2	0.55
ISStin5	IS3	IS150	Streptococcus iniae	28.2	2.2
ISAcma25	IS630		Acaryochloris marina	28.2	2.2
ISFtu2	IS5	IS427	Francisella tularensis	28.2	2.2
ISLhe6	IS3	IS150	Lactobacillus helveticus	28.2	2.2
IS15DIV	IS6		Salmonella typhimurium	28.2	2.2
IS15DII	IS6		Salmonella panama	28.2	2.2
IS15DI	IS6		Salmonella panama	28.2	2.2
IS15	IS6		Salmonella panama	28.2	2.2
IS26	IS6		Proteus vulgaris	28.2	2.2
ISVal3	IS1595	ISPha2	Vibrio alginolyticus	26.3	8.6
ISCysp14	IS200/IS605	IS605	Cyanotheca sp.	26.3	8.6
ISMMac10	IS1634		Methanosarcina acetivorans	26.3	8.6
ISAac1	IS200/IS605	IS200	Actinobacillus actinomycetemcomitans	26.3	8.6
IS1216V	IS6		Enterococcus sp.	26.3	8.6
IS1216E	IS6		Enterococcus faecium	26.3	8.6
IS1216	IS6		Enterococcus hirae	26.3	8.6

Table 3-5 : ISFinder analysis results for the examination of the 3' end of the identified catabolic contig .

Sequences producing significant alignments	IS Family	Group	Origin	Score (bits)	E. value
ISAb52	IS6		Acinetobacter baumannii	198	1e-51
IS1007	IS6		Acinetobacter sp.	198	1e-51
IS1008	IS6		Acinetobacter calcoaceticus	190	3e-49
ISOur1	IS6		Oligella urethralis	190	3e-49
IS1006	IS6		Acinetobacter junii	75.8	1e-14
ISPPr9	IS6		Photobacterium profundum	32.2	0.14
ISAb20	IS3	IS407	Acinetobacter baumannii	30.2	0.55
ISAlw6	IS3	IS407	Acinetobacter lwoffii	28.2	2.2
ISCaje3	IS607		Campylobacter sp.	28.2	2.2
ISNml1	IS1380		Nakamurella multipartita	28.2	2.2
ISSpn6	IS200/IS605	IS200	Streptococcus pneumoniae	28.2	2.2
ISXoo7	IS5	IS5	Xanthomonas oryzae	28.2	2.2
ISPa16	IS5	IS5	Pseudomonas aeruginosa	28.2	2.2
ISPPu18	IS5	IS5	Pseudomonas putida	28.2	2.2
IS1646	IS5	IS5	Xanthomonas campestris	28.2	2.2
IS1479	IS5	IS5	Xanthomonas campestris	28.2	2.2
ISFop3	IS110		Francisella opportunistica	26.3	8.6
ISMmo2	IS1595	ISSod11	Magnetospirillum moscoviense	26.3	8.6
ISNma25	IS200/IS605	IS1341	Natrialba magadii	26.3	8.6
ISEnfa5	IS3	IS150	Enterococcus faecalis	26.3	8.6
ISPFi2	ISL3		Propionibacterium freudenreichii	26.3	8.6
ISMAe10	IS1634		Microcystis aeruginosa	26.3	8.6
ISBsp3	IS3	IS150	Bacillus sp.	26.3	8.6
ISBaov2	ISAs1		Bacteroides ovatus	26.3	8.6
ISEc23	IS66		Escherichia coli	26.3	8.6
ISCb4	IS3	IS150	Clostridium beijerinckii	26.3	8.6
ISGur8	IS21		Geobacter uraniumreducens	26.3	8.6

As the previous investigation demonstrated the insertion sites at both ends of the contig, further examination into the potential mobility of the contig was undertaken. BLASTn of the entire 4229 bp contig revealed a match with 43% query coverage and 99.8% identity similarity with a *Pseudomonas* sp. CT14 plasmid pCT14. Visual inspection of the alignment, shown in Figure 3-15, shows highest similarity regions occur from bp 2344 to 4102.

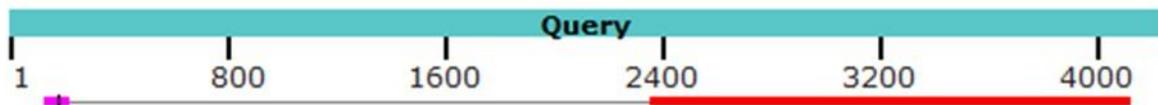


Figure 3-15: BLAST alignment graphic of the 4229 bp contig containing the identified metaldehyde degrading genes. Image shows similarity against the high scoring alignment of *Pseudomonas* sp. CT14 plasmid pCT14. Red line demonstrates an alignment score of ≥ 200 . Pink demonstrates an alignment score of 80 -200.

3.26 Promoter Analysis

As the three sequences, *mahX*, *mahY* and *mahZ*, appear clustered together on the contig, suggesting a potential catabolic operon, a promoter analysis was undertaken. The 245 bp region upstream from *mahX* was subjected to the Softberry program BPROM. The results, shown in Table 3.6, identified a promoter sequence 53 bp upstream from the predicted *mahX* translational start site. The confidence scores obtained for the -10 and -35 box were 78 and 37 respectively. Examination of upstream regions of the three other genes present on the contig identified no significant result.

Table 3-6 PBROM SOFTBERRY generated table based on the 250 bp sequence submitted

Length of sequence-	249		
Threshold for promoters -	0.20		
Number of predicted promoters -		1	
Promoter Pos:	196 LDF-	5.74	
-10 box at pos.	181 TGCTATACT	Score	78
-35 box at pos.	161 TTTCCT	Score	37

3.27 Identification of mutations within the degradative contig

As the initial BSR analysis only detected direct gene disruptions and therefore does not identify mutations outside the coding sequence, further examination of the contig was undertaken. Using the *A. calcoaceticus* E1 WT contig as the query sequence, the catabolic contig of the four mutants were examined. For mutants 28, 140 and 206, the BLAST alignment identified a 173 bp deletion within the contig, as shown in Figure 3-16. This deletion is found from 120 bp upstream from *mahX* to 53 bp into the coding region. Based on the promoter location, as identified above, the promoter region would be absent for the contigs in the four mutants, and the translational start for MahX would be absent. Based on all the previous bioinformatics analysis surrounding the metaldehyde degrading contig, an annotated schematic, shown in Figure 3-17, was designed.



Figure 3-16: BLAST schematic generated through query the identified contig containing the metaldehyde degrading genes against that of the *A. calcoaceticus* E1 mutants 28, 140, and 206. Result generated an identical output for each. Red lines demonstrate an alignment score of >200. Black line suggests an alignment score <40.

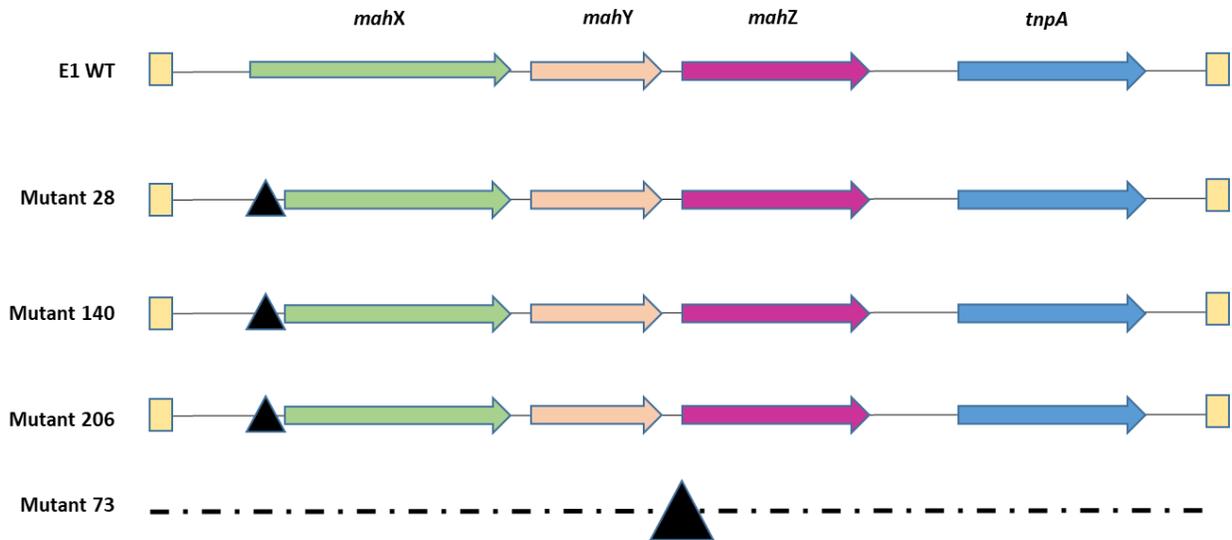


Figure 3 17: Schematic illustration of catabolic contig from *Acinetobacter calcoaceticus* E1 WT and derived mutants. In mutants 28, 140 and 206, BlastN analysis revealed a 173 bp deletion within the contig, which disrupted the start of the *mahX* gene. IS Finder identified insertion sites at both ends of the contig which are displayed by the yellow blocks. Δ indicate deleted regions.

3.28 Identification of the metaldehyde degrading gene(s) through heterologous expression

In order to verify the bioinformatics predictions of the metaldehyde degrading gene(s) and to determine the gene(s) responsible for the initial degradative step, the genes were inserted into the *E. coli* cloning vector pBR322. To identify the initial degradative gene, four vector constructs ; *mahX*, *mahY*, *mah XY*, *mah XYZ*, were created. All constructs, as shown in Figures 3-18, 19, 20, 21 and 22, contained the native *A. calcoaceticus* E1 promoter region, to provide the possibility for potential cloning within *A. calcoaceticus*, if required.

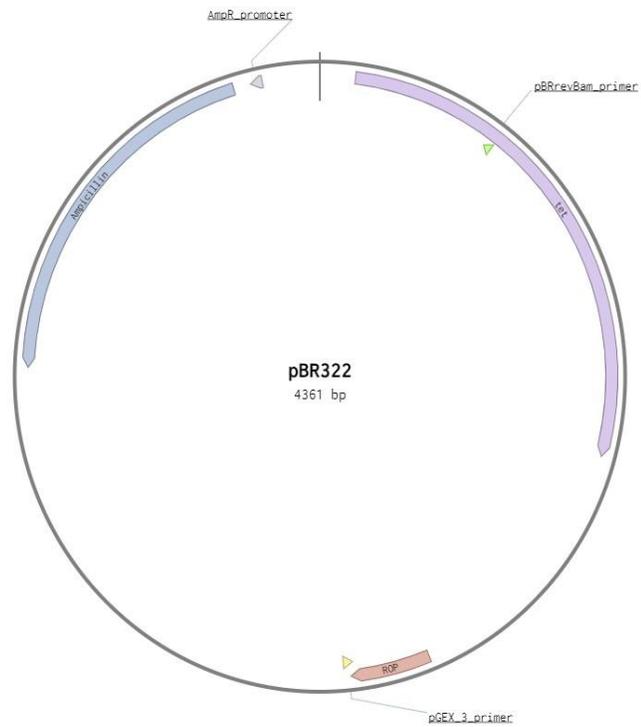


Figure 3-18: Schematic of pBR322 vector with key genes highlighted. Image obtained through benchling.

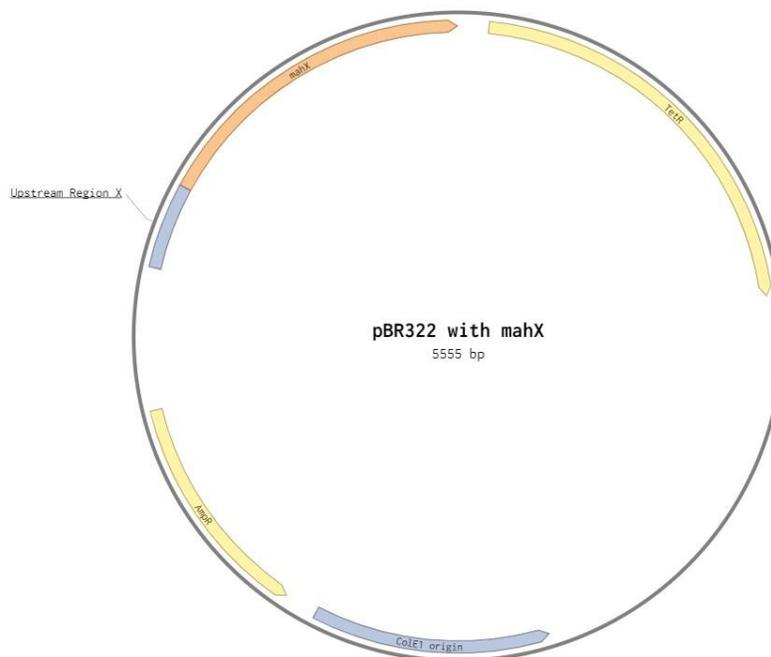


Figure 3-19: Schematic of pBR322 vector with mahX and upstream promoter region with other key genes highlighted. Image obtained through Benchling.

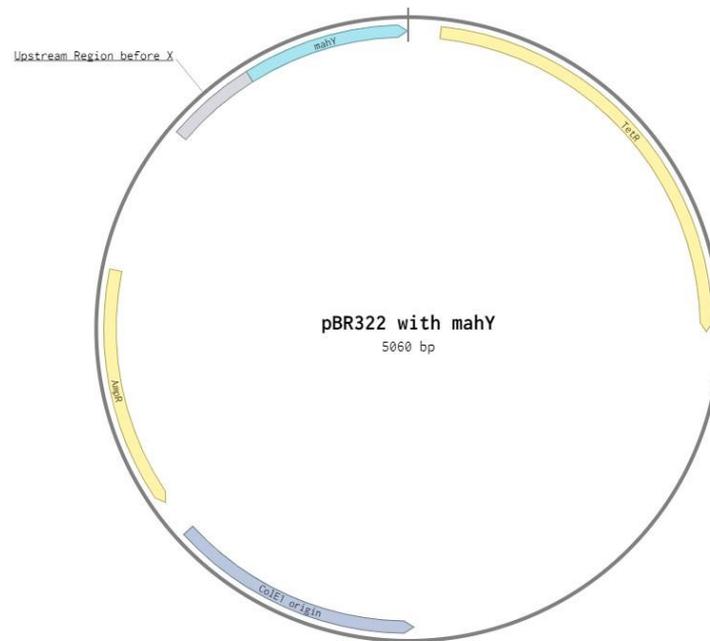


Figure 3-20: Schematic of pBR322 vector with mahY and upstream promoter region with other key genes highlighted. Image obtained through benchling.



Figure 3-21: Schematic of pBR322 vector with mahXY and upstream promoter region with other key genes highlighted. Image obtained through benchling.

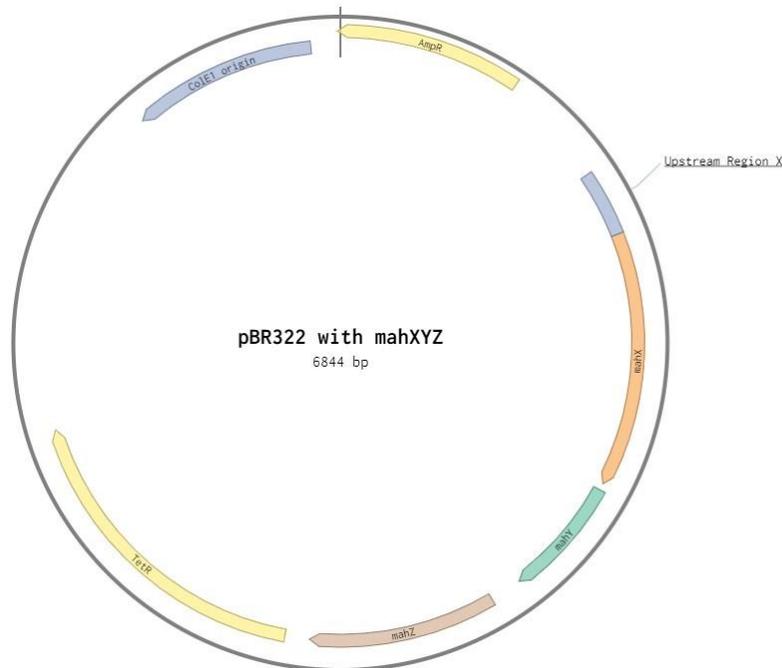


Figure 3 -22: Schematic of pBR322 vector with mahXYZ and upstream promoter region with other key genes highlighted. Image obtained through Benchling.

The *E. coli* cloning vector pBR322 was digested with *Eco*RI and gel extracted in preparation for insertion of the potential metaldehyde degrading genes, as revealed in Figure 3-33. Insertion regions were amplified using a high-fidelity polymerase and verified through gel analysis as shown in Figure 3-23 and Figure 3-24. Confirmation of successful ligation was performed using both colony PCR and Sanger sequencing. Examination of the all the transformed colonies identified morphological differences to the no insert control. Observational analysis identified the colonies containing the correct insert possessed an observable reduction in colony size, relative to the no insert control.

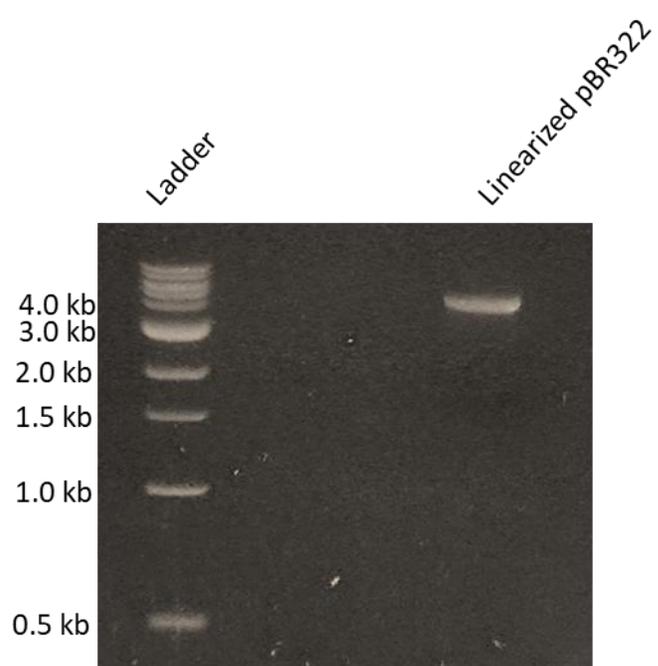


Figure 3-23: 1.2 % agarose gel of EcoRI linearised pBR322 post gel extraction. NEB 1kb ruler was used to estimate DNA size.

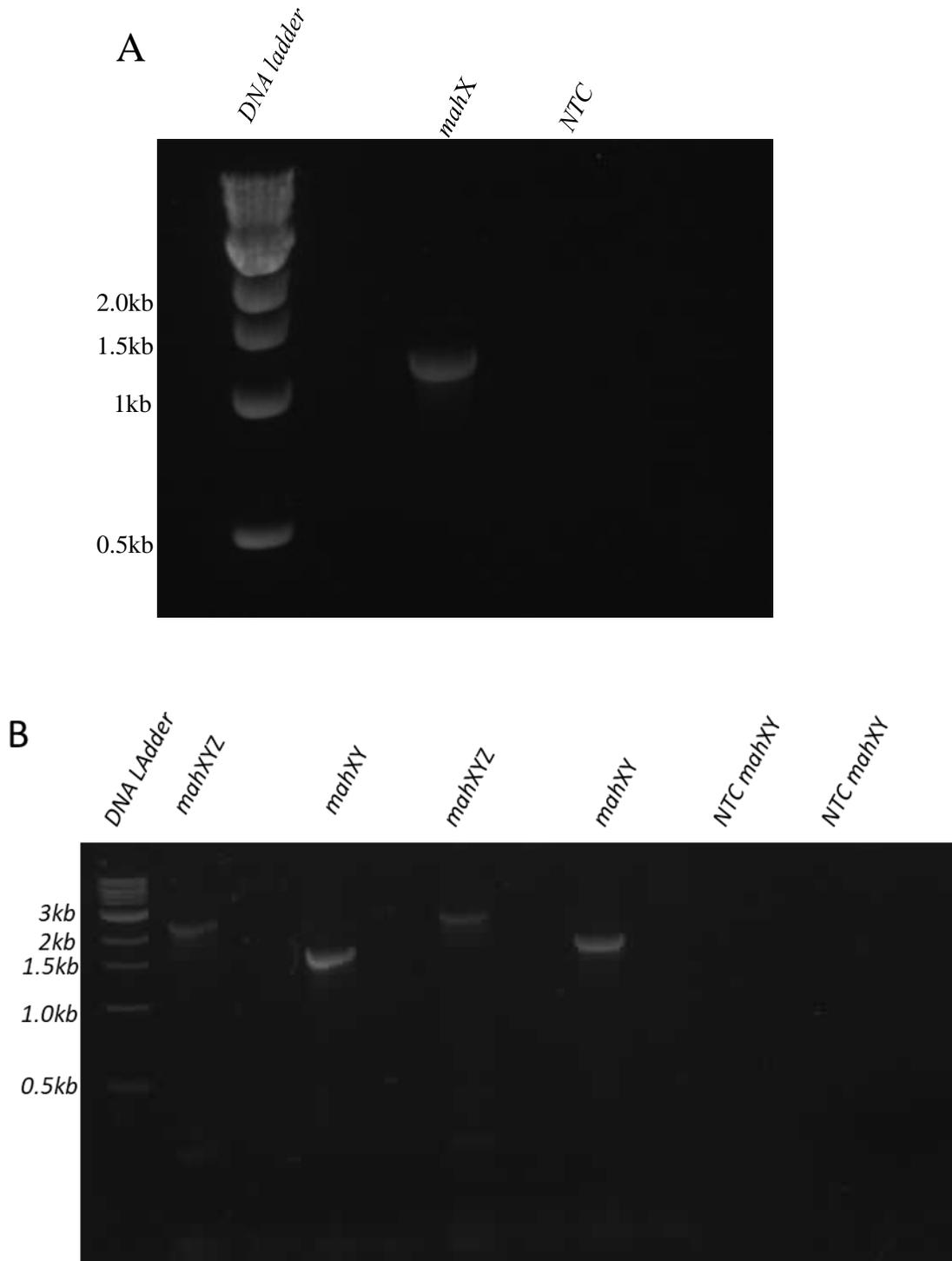


Figure 3 -24 Agarose gel (1.2 %) showing products post amplification from native host *A. calcoaceticus* E1. A 1kb quick load ladder was used for all gels A) Gene mahX and NTC B) Genes mahXY and mahXYZ with their respective NTCs.

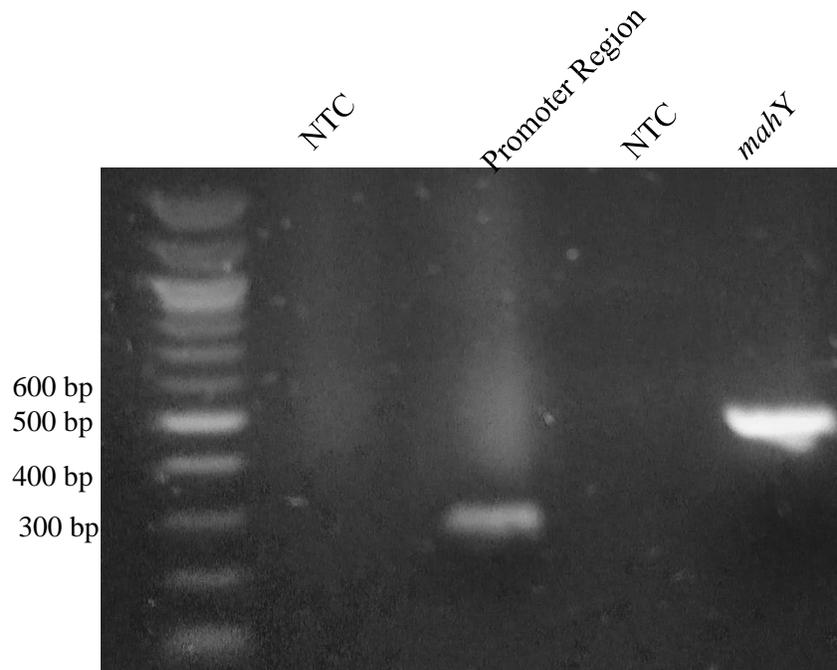


Figure 3-25: 1.2% agarose gel following PCR amplification of mahX promoter region and mahY. 100 bp ladder was used to estimate DNA size.

To determine whether transformation of the constructs conveyed the ability to utilize metaldehyde as a sole carbon source, transformed colonies were plated onto MSM metaldehyde plates. Visual examination identified no observable growth upon the plates. Growth curve analysis in liquid media, shown in Figure 3-26 allowed for the quantification of growth of the constructs within a metaldehyde media.

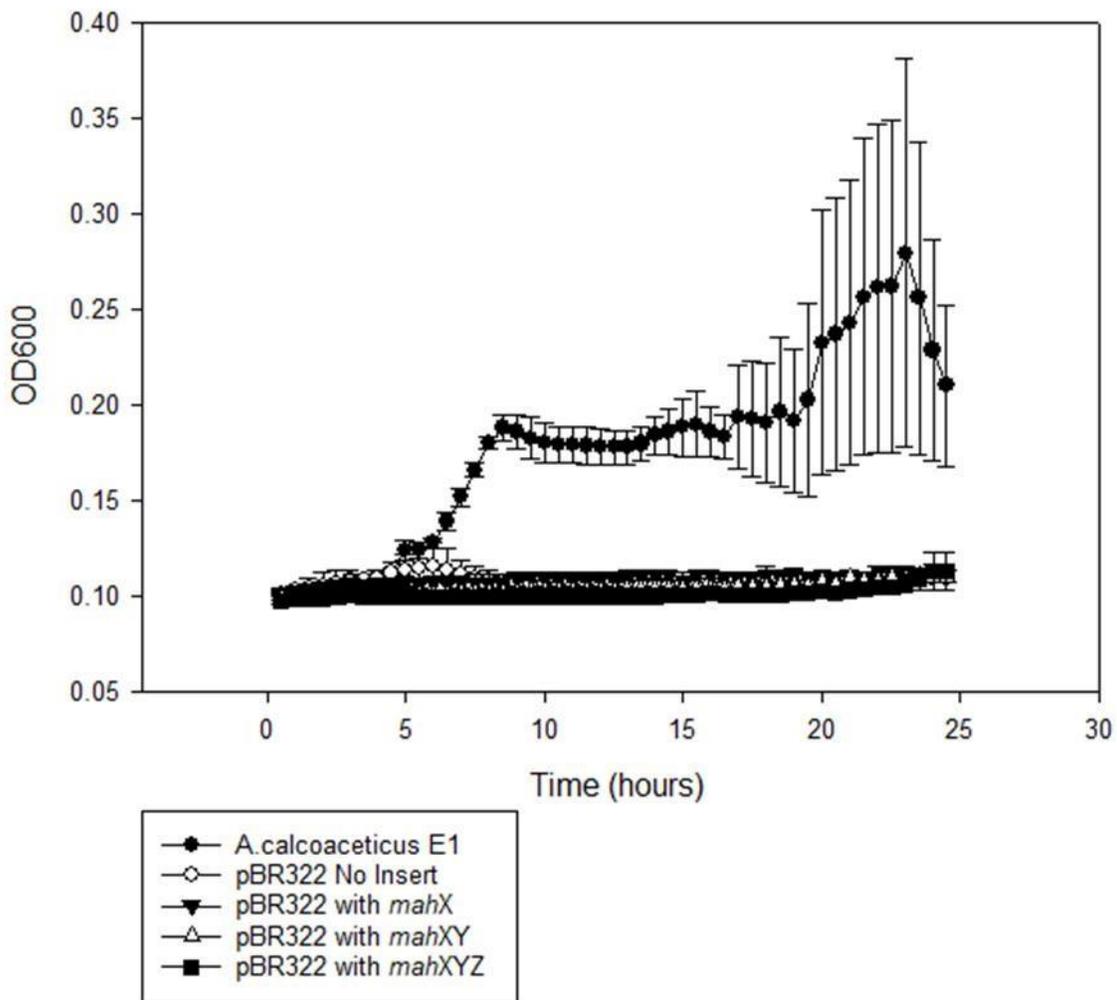


Figure 3-26: Growth assay with transformants containing *mahX*, *mahXY* and *mahXYZ* incubated within MSM supplemented with 150mg /L metaldehyde. Error bars display the standard deviation.

As the previous analysis identified no observable growth, an investigation was undertaken to determine whether if and which transformed *E. coli* strains could degrade metaldehyde. Constructs were grown overnight in LB with tetracycline and normalised to an OD₆₀₀ of 1.0. The four constructs and a no insert control were incubated with 15 mg/L of metaldehyde for 3.5 hours at 30 °C. Gas chromatography results, as shown in Figure 3-27, revealed degradation below the

limit of detection in constructs containing mahX, mahXY and mahXYZ. No measurable change in metaldehyde was observed in the mahY construct, relative to the no insert control.

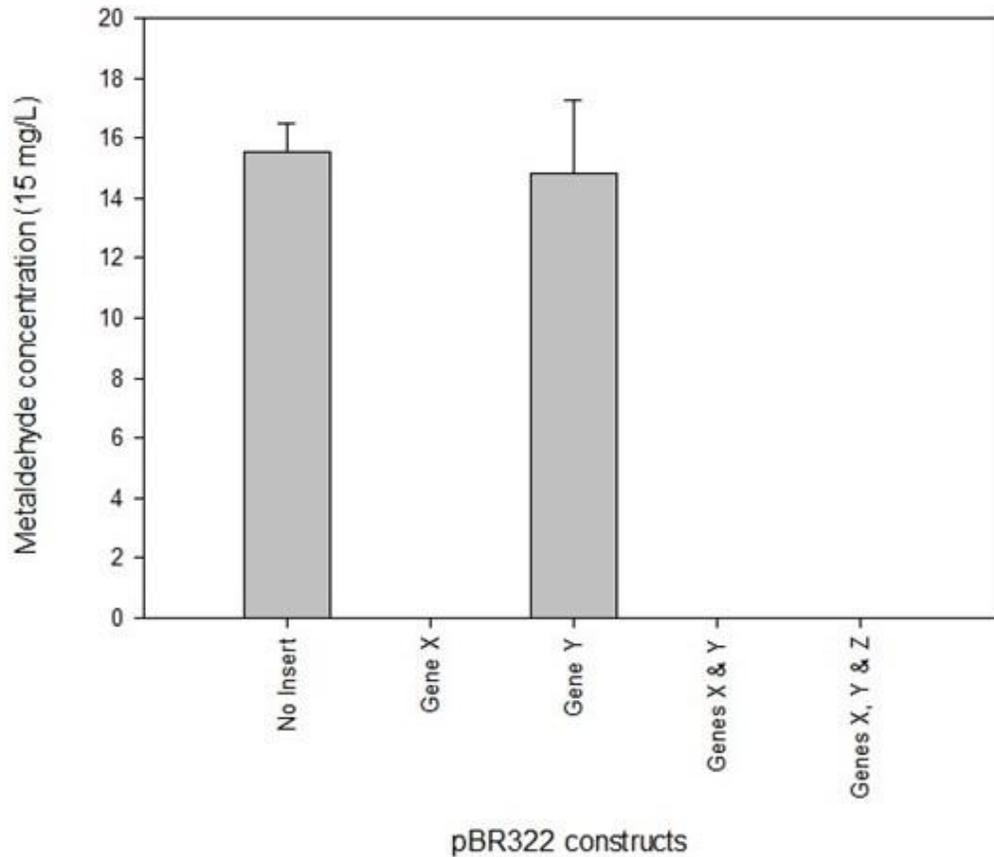


Figure 3.27- Figure 3-27 Degradation of metaldehyde by *E. coli* was measured following growth in LB liquid media. Degradation assay features pBR322 with various derivatives of this plasmid containing mahX, mahY, mahXY, and mahXYZ.

To determine the rate at which metaldehyde degrading within the constructs occurs, a degradation assay was undertaken. Like the previous assay, samples were grown overnight in LB

media with Tet and resuspended in MSM this time supplemented with 150 mg/L metaldehyde. Samples were extracted every 30 minutes to determine the metaldehyde concentration. All constructs containing *mahX* degraded metaldehyde below the LOD within 240 mins, as shown in Figure 3-28. The construct containing just *mahX* demonstrated degradation below the LOD within 180 mins.

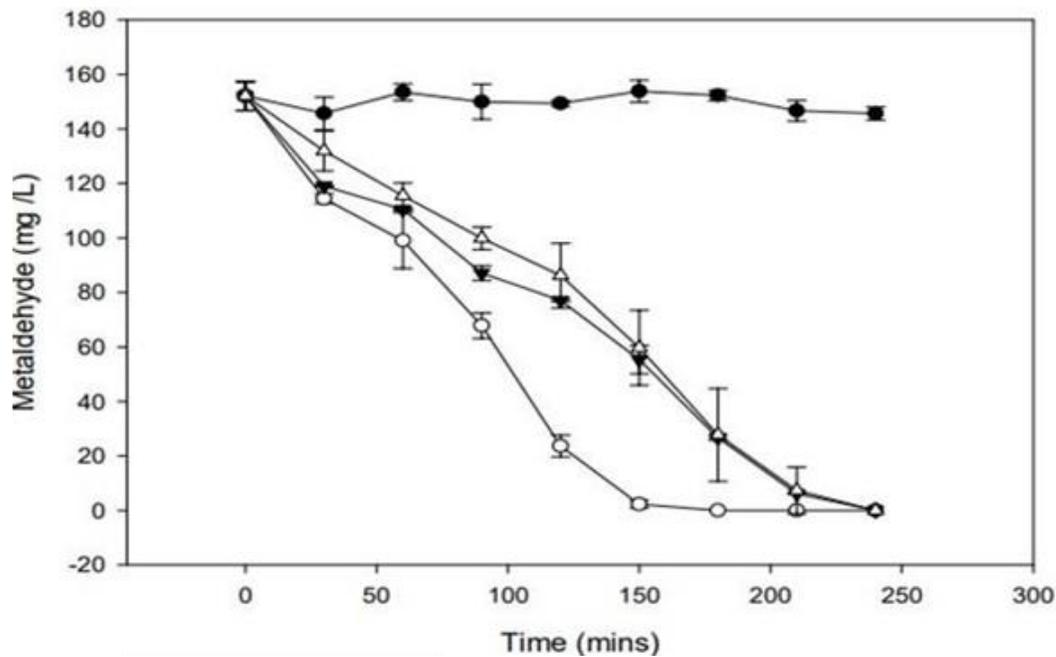


Figure 3-28- Degradation of metaldehyde by *E.coli* was measured following growth in LB liquid media. *E.coli* carrying plasmid pBR322 (solid circle), and derivatives of the plasmid containing *mahX* (open circle), *mahXY* (filled triangle) and *mahXYZ* (open triangle).

Based on an average reading over the 180 minutes, the rate of metaldehyde degradation in the *mahX* construct is 47.3 nmol/ min. The rate for both *mahXY* and *mahXYZ* is 35.5 nmol/min. The no insert control showed reduction in metaldehyde over the 240 mins of the assay.

3.29 Observing the effects of metaldehyde on mahX constructs

Although degradation of metaldehyde was observed, *E. coli* DH5 α was unable to utilize metaldehyde or its degradative products to support observable growth. Despite the lack of growth, the effect of the degradative product may provide greater insight into the causes for the lack of growth and properties of the metabolic pathway. Growth assays, shown in Figure 3-29, were conducted within *E. coli* containing mahX constructs incubated with a media containing both MSM and LB. Conditions were compared for conditions in which metaldehyde was supplemented against those without. Analysis of the end point values, following 24 hours incubation, revealed no significant change OD₆₀₀ was observed between with and without metaldehyde. (*p* values for no insert, mahX, mahXY and mahXYZ are .92, .19, .50 and .44, respectively).

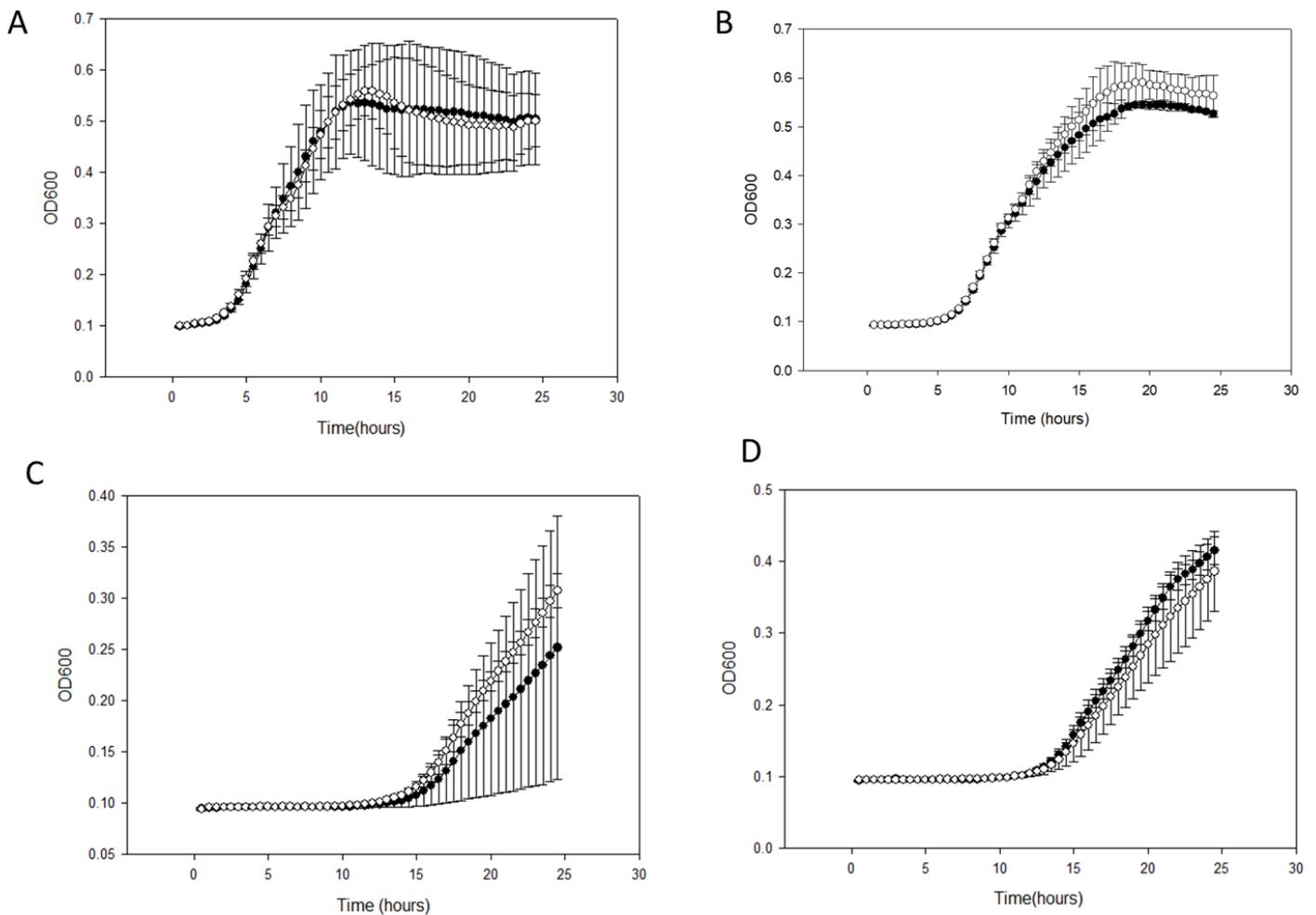


Figure 3-29: The OD₆₀₀ of *E. coli* DH5a transformants grown in LB and MSM supplemented with and without 0.43 mM metaldehyde. Open circles display no metaldehyde. Closed circles display with metaldehyde. A displays pBR322 without an insert. B. Displays pBR322 with *mahX*. C. Displays pBR322 with *mahXY*. D. Displays pBR322 with *mahXYZ*. All samples were performed in biological triplicate with error bars showing standard deviation.

3.3 Discussion

Previous studies demonstrated microbial degradation of metaldehyde as well as the isolation and characterization of metaldehyde degrading strains (Thomas, 2016; Castro-Gutiérrez *et al.*, 2020). However, prior to this investigation no knowledge of the degradative genes or a potential degradative pathway was known. This research set out to identify and characterize the first known metaldehyde degrading gene(s), using a mixture of both random mutagenesis and comparative genomics.

The EMS mutagenesis analysis led to the isolation and identified of four mutants, which demonstrated an inability to degrade and utilize metaldehyde, yet maintained their ability to grow on acetate. Typically, an ethylating agent such as EMS causes transition mutations throughout the genome of the organism (Sega, 1984). The genomic analysis of the mutants however suggests small deletions were present both 120 bp upstream and 53 bp into the 5' end of *mahX* in mutants 28, 140 and 206. This disruption of the promoter region and gene is what prevented the degradation of metaldehyde and therefore the ability to grow on the MSM metaldehyde plate. Examination of the genomic data for mutant 73 failed to identify the catabolic contig suggesting a possible deletion of the entire contig. Small deletions, following EMS exposure, although not as common as transition mutations, have been associated with EMS mutagenesis previously (Shukla and Auerbach, 1981). As the no EMS control identified no mutants deficient in metaldehyde degradation, the mutations are presumed to be due to the presence of the mutagen and not some other evolutionary process occurring during laboratory culture. Greater confidence was added by using comparative genomic analysis of the isolated mutants and combining this with the blast-score-ratio analysis using multiple different

metaldehyde degrading *Acinetobacter* isolates, as carried out by Castro and Thomas of the Moir laboratory (Thomas, 2016; CastroGutiérrez, 2020). This shows the great potential of random mutagenesis when coupled to whole genome sequencing as a method for identification of novel degradative genes in a non-model organism.

Numerous studies clearly demonstrate the role HGT plays in both the spread and the construction of novel degradative pathways (Stalker and McBride, 1987; Gunathilaka *et al.*, 2017; Werner *et al.*, 2020). This ability allows for the rapid adaptation of microbial communities to novel xenobiotic compounds. The construction of these pathways typically involves HGT whereby genes required for different degradative steps are recruited from both phylogenetically distant and similar microorganisms into a single host (Boto, 2010). The sequences for MahX, MahY and MahZ all show close relationships with various proteins outside the *Acinetobacter* genus. Analysis of the genomic location of *mahX*, led to the identification of *mahY* and *mahZ*. Based on the orientation and the reduced spacing between gene regions, it is very likely that the genes form a catabolic operon for metaldehyde degradation. This is further supported by the presence of a single identifiable promoter region upstream from *mahX*. The inducibility of this promoter however is yet to be studied and may shed light on not only the regulatory control of the catabolic genes but also provide insight into the evolutionary nature of the operon, as recently evolved catabolic processes are typically constitutively active (Cases and De Lorenzo, 2001). Further analysis, conducted by Castro *et al.* (2020), identified this operon in other numerous metaldehyde degrading strains. This therefore suggests both that the operon is likely to be required for the complete degradation of metaldehyde in certain strains and that it is likely to be transferred horizontally. The identification of both a transposase and insertion sequence sites on the

catabolic contig and the high BLAST similarity to *Pseudomonas sp.* CT14 plasmid, suggest it is very likely that the contig is a transposable element found on a plasmid. This would both explain the presence of the contig within other bacterial strains and the closest homologues of the 3 *mah* genes not being closely associated with *Acinetobacter*.

In order to generate a degradative pathway for metaldehyde in *A. calcoaceticus* E1, literature surrounding the degradative pathway of cyclic ethers were combined with the predicted functional information of the genes present on the catabolic operon. In the cyclic ether 1,4 dioxane, for example, the initial cleavage step is performed by a monooxygenase (Cases and De Lorenzo, 2001). In this example, the hydroxylation of an ether bond associated carbon leads to the cleavage of the stable ring structure creating the product of 2-hydroxyethoxyacetic acid. A similar degradative pathway, shown in Figure 3-30, is also characterized for the degradation of the cyclic ether tetrahydrofuran through another monooxygenase where the product formed is 4 hydroxybutyraldehyde (Sales *et al.*, 2013). Based on the above pathways and the information regarding the catabolic operon, it is believed that metaldehyde's cyclic structure is activated through the 2OG-Fe (II) oxygenase MahX. This cleavage leads to the production of a hemiacetal molecule 1, 3, 5, 7-tetramethyl-2, 4, 6-trioxa-1-hydroxy-7-octanone. The unstable nature of hemiacetals suggests chemical degradation of the compound will occur over minutes to hours (Chiang and Jerry Kresge, 1985). Based on the presence of *mahY* in the catabolic contig and on its predicted function, it seems likely that MahY acts a lyase, enzymes which are capable of catalyzing the cleavage of C-C and C-O bonds (Chiang and Jerry Kresge, 1985). This would allow the formation of a double bond leading to an iterative depolymerization of the compound. Products from this reaction include three acetaldehyde molecules and a single acetate as shown in Figure 3-31.

Following depolymerization, acetaldehyde molecules can be oxidized with MahZ into acetate, which is then incorporated the central metabolism.

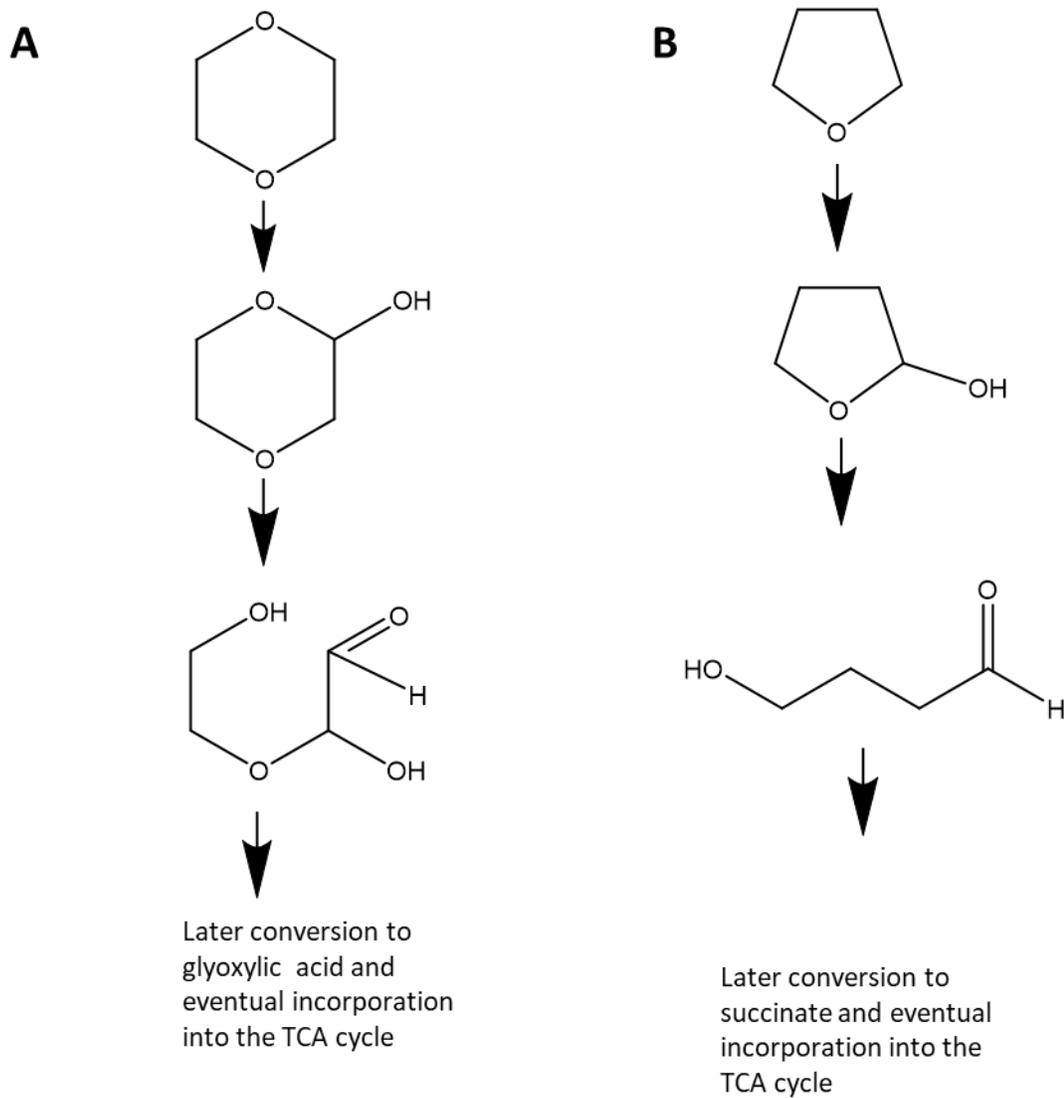


Figure 3-30: Schematic based on the work by C. Sales et al (2013). A displays the initial cleavage steps of 1, 4 dioxane through monooxygenase activity. B shows the initial cleavage step of tetrahydrofuran through monooxygenase activity.

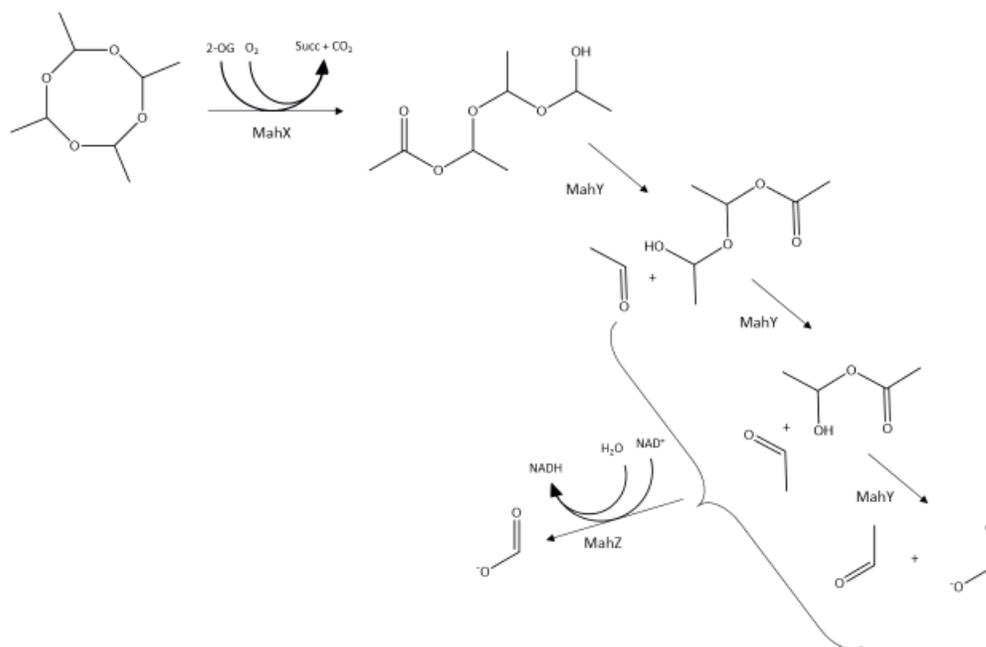


Figure 3.31- Predicted pathway for metaldehyde degradation. *MahX* is related to the 2 oxoglutarate (2-OG)-dependent dioxygenases which generate succinate and CO₂. *MahX* oxygenates metaldehyde to cleave the cyclic structure of metaldehyde producing a linear hemiacetal which is iteratively depolymerised into acetaldehyde- a shorter chain hemiacetal and eventually acetate. *MahZ* oxidises acetaldehyde into acetate in an NAD⁺ dependent reaction.

Heterologous Expression

Although the *E. coli* strains containing *mahX* demonstrated the ability to degrade metaldehyde below the detectable limit, the strain containing *mahXYZ* failed to utilize metaldehyde as a sole carbon source. There are several possibilities for the lack of growth when

utilising metaldehyde as a sole carbon source. The expression levels of the protein within *E.coli* expressing *mahXYZ* were not investigated and as such the metabolic burden of expressing 3 non-native proteins may have been too high for observable growth to occur. As well as this, the assay time points could have been extended to ~72 hours to allow for greater confidence in the results observed. Also, the proposed pathway requires the host microorganism to utilize acetate however, growth analysis of *E. coli* DH5 α demonstrated its inability to perform this function. This could potentially be overcome through using an acetate tolerant strain such as *E. coli* D138Y(Chong *et al.*, 2013). Alternatively, the origin region of the *Acinetobacter* plasmid pWH1266 could be inserted into the pBR322 constructs (Lucidi *et al.*, 2018). This could allow for the expression of the catabolic genes into a non-metaldehyde degrading *Acinetobacter* strain, such as RUH2202, allowing for analysis of the genes in a host closer to the native strain. This would further demonstrate the role and importance of all three genes in the breakdown and utilization of metaldehyde.

Conclusion

The recent reduction in both sequencing costs and the increased accessibility of the technology allows the identification of novel degradative genes to be performed without requiring pathway or genetic information. The identification of *mahX* through random mutagenesis combined with comparative genomics shows a simple pipeline for the identification of novel xenobiotic degradative genes when working with a non-model microorganism.

Chapter 4: Characterization of the metaldehyde degrading protein MahX

4.1 Introduction

Previous work conducted within Chapter 3 led to the discovery and experimental validation of the first identified metaldehyde-degrading enzyme, MahX. Bioinformatics analysis revealed this enzyme to belong to the family of Fe (II)/ (alpha) ketoglutarate-dependent dioxygenases. This enzyme superfamily is widely distributed throughout the kingdoms of life (Martinez and Hausinger, 2015). As such, the Fe (II)/(alpha)ketoglutarate-dependent dioxygenases have been demonstrated to catalyze a highly diverse set of reactions which include but are not limited to the following; DNA/RNA repair, antibiotic biosynthesis, lipid metabolism and organic compound biodegradation (Müller *et al.*, 2006; Martinez and Hausinger, 2015; Bräuer *et al.*, 2016; Wu *et al.*, 2016; Deepa *et al.*, 2018). Concerning biodegradation, dioxygenases have been shown to exhibit the ability to degrade substrates such as naphthalene and catechol as well as numerous other aromatic compounds (Martin and Mohn, 1999; Rodríguez-Salazar *et al.*, 2020)s.

As the Fe (II) / (alpha) ketoglutarate-dependent dioxygenases superfamily of enzymes is widely found and have been demonstrated to play essential roles within numerous biochemical reactions, their mechanisms of action and the cofactors required for optimal activity have been subjected to extensive research, shown in Figure 4-1. Dioxygenase reactions involve the incorporation of a single oxygen molecule, from molecular oxygen, into both the substrate and the cofactor/cosubstrate alpha ketoglutarate (α KG), with the oxidation of the latter leading to the generation of succinate and CO₂ (Martinez and Hausinger, 2015). As well as requiring α KG, several studies have demonstrated the need for both a divalent metal (such as iron or copper) (Steiner *et al.*, 2002; Grzyska *et al.*, 2010) and a reducing agent such as ascorbic acid (LAA)

(Pornsuwan *et al.*, 2017; Khan *et al.*, 2020). Within Fe (II)/ (alpha) ketoglutarate-dependent dioxygenases, the importance of the metal cofactor is best demonstrated by the near identical active site structural arrangements throughout the superfamily. Through the His-Asp/Glu-His triad, the metal ion is bound to the protein. The remaining three coordination sites are essential, as they are required for α KG and oxygen binding (Aik *et al.*, 2012; Martinez and Hausinger, 2015). The cosubstrate/cofactor α KG's importance within the α KG dioxygenase family is also demonstrated through structural conservation, as the double stranded β helix (DSBH a.k.the jelly-fold) is found within all crystallographic structures of α KG dioxygenases. The secondary structure elements of this 'jelly fold' appear crucial for binding of the molecule α KG binding and subsequently essential for binding of the primary substrate (Aik *et al.*, 2012; Srnec *et al.*, 2014; Martinez and Hausinger, 2015). The final cofactor identified for optimum dioxygenase activity within the literature is that of reducing agents. Due to the reactive nature of oxygen, several inhibitive processes can occur that may explain the important role of reducing agents within the catalytic reactions. The first complication is associated with the oxidation of the Fe $^{2+}$ into the inactive Fe $^{3+}$ -containing enzyme. This inhibitive process has been described previously in many dioxygenases within the literature (Kuiper and Vissers, 2014; Mahmood and Dunwell, 2020). Another complication, demonstrating the role of reducing agents, is through the decreasing of 'uncoupled turnover'. In this example, the enzyme bound α KG is oxidised thereby generating the succinate and CO₂ without the consumption of the target substrate. Presence of a reducing agent within the reaction has previously been demonstrated to reduce the occurrence of uncoupled turnover (Kuiper and Vissers, 2014; Mahmood and Dunwell, 2020). As such, the reducing agent has an indirect role within the catalytic reaction through the prevention of enzyme inactivation

and uncoupled turnover. Measurement of metaldehyde degradation through lysed cell suspension was previously attempted by Thomas, (2016) within the MahX native host *A. calcoaceticus* E1. As the previous cell lysis investigation was conducted without prior knowledge of the initial degradative enzyme, numerous cofactors (such as NAD⁺, NADH and coenzyme A) and pH conditions were attempted. Identification of the cofactors and conditions required for MahX degradation of metaldehyde is essential for verifying *in vitro* activity of MahX. Although, as mentioned previously, Fe²⁺/ α KG dependent dioxygenases require Fe²⁺, α KG and a reducing agent and are therefore typically utilized using whole cell catalysis, thereby alleviating the need for exogenous co-factor regeneration, crude and purified protein analysis is essential to further optimize this process and allow for further advances within metaldehyde bioremediation and biosensing methods. As such, the aim of the investigation was to characterize the enzyme using both bioinformatics analysis and experimental approaches to gain a greater understanding of the optimal conditions for the degradation of metaldehyde by MahX. Through gaining insight into the cofactors and conditions required for metaldehyde degradation within the model organism *E. coli* expressing MahX, greater information regarding the initial catalytic mechanism was obtained. The investigation therefore set out to overexpress an active MahX from a non-native host, to demonstrate activity within lysed cell suspensions and to obtain purified and active MahX protein. Through meeting these objectives, the potential industrial applications to the enzyme regarding both bioremediation and biosensing can be assessed.

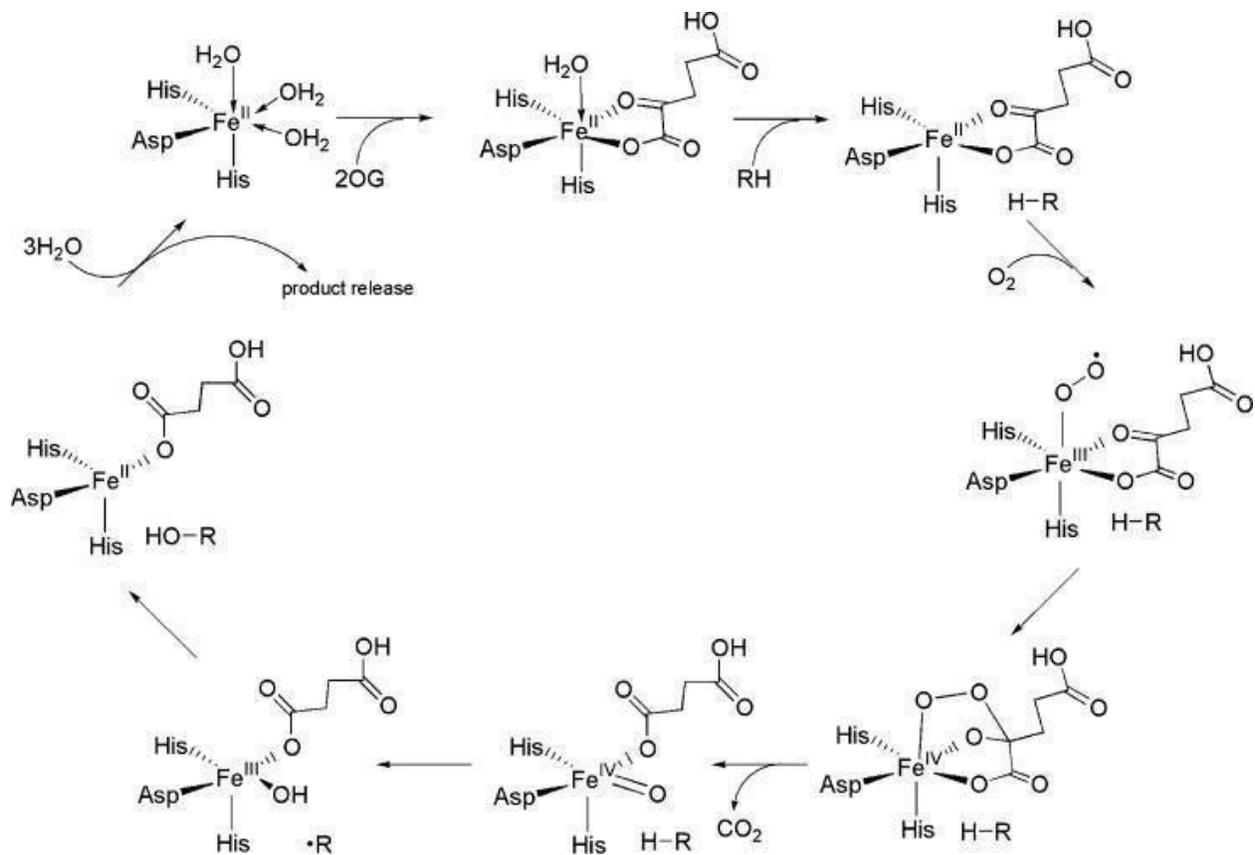


Figure 4-1 Schematic showing the postulated mechanism reaction for α KG dependent enzymes. ($S\text{-H}$ = prime substrate, $S\text{-OH}$ = prime product). The resting enzyme facial triad and 3 water molecules bind to the core iron molecule. Two water molecules are later displaced through α KG binding. Displacement of the final water molecule is through the binding of the substrate nearby, although not to the Fe^{2+} centre. This opens up a coordination site for the activation of O_2 . Dioxygen occupation of the iron ultimately leads to the presence of the reactive ferryl intermediate compound which is responsible for the oxidation of both the prime substrate and the α KG. With the release of the prime product, CO_2 and succinate, the enzyme is ready to start a new degradative cycle (Image obtained from Welford et al., (2005)).

4.2 Results

4.2.1 Bioinformatics characterization of MahX

Prediction of the protein structure of MahX allows for the further protein classification, and allows the potential function and cofactors to be identified. The model of MahX was generated using the protein structure prediction and structure-based annotation tool I-TASSER. This program uses a protein threading method, which utilizes key protein features such as their secondary structure, solvent accessibility, and sequence patterns to find potential structural templates for modelling. Analysis was undertaken to identify protein databank (PDB) files with predicted similar structures. The highest ranked threading template sequence was 5dapA. This result encodes for the protein Fe (II)/ (alpha) ketoglutarate-dependent dioxygenase AsqJ. This protein, isolated from *Aspergillus nidulans*, is responsible for catalyzing the synthesis of quinolone alkaloids. The coverage of this threading alignment, which is characterized by the number of aligned residues divided by the length of the query protein, was 0.87. Another quality value obtained is the normalized Z-score. This value provides the normalized threading alignment Z-score whereby values > 1.0 are considered 'good' alignments. The 5dapA result obtained a norm Z-score of 2.15. Further examination of the remaining 9 threading template sequences revealed all were also Fe (II)/ (alpha) ketoglutaratedependent dioxygenases. Using I-TASSER, coupled with the clustering algorithm SPICKER, five protein models for MahX were generated. A confidence score (C-score) of the generated models allowed the quality of the I-TASSER predicted models to be estimated based on both the threading alignment and convergence parameters of each structure assembly simulation. The C-score typically ranges between values of -5 and 2 with higher values signifying greater confidence in the model. The highest ranked MahX model, shown

in Figure 4-2 , obtained a C score of -1.17 . Another measure of predicted protein structure quality is the template modelling score (T-M score) and root-mean-square deviation (RMSD). Both the RMSD and T-M score allow for the structural comparisons of model alignments to be quantified through superimposing structural models and calculating the divergence of equivalent residues. The RMSD value is one of the most popular measures for structural similarity and provides effective comparative values. However, this calculates all residue distances with equal weight and is therefore prone to significant local structural deviation. The T-M score attempts to overcome this through counting all residues pairs using the Levitt–Gerstein weight and therefore can potentially provide a score that is more sensitive to the global topology. The RMSD for this model obtained a score of $8.9 \pm 4.6 \text{ \AA}$. For the T-M score, the values are calculated between 0 - 1 with 1 being an identical match- the model obtained a T-M score of 0.57 ± 0.15 was calculated.



Figure 4-2: Cartoon ribbon predicted model for MahX. Model generated through I-TASSER and imaged through PyMol.

Protein structure identity can provide functional information that sequence based alignment alone may overlook. By generating a MahX model, proteins recognised as structurally close were identified allowing for a greater understanding of dioxygenase requirements. Through querying the PDB using the T-M structural alignment program, 10 protein structures with a close structural similarity to the MahX model were identified and are shown in Table 4-1.

Table 4-1 I-TASSER generated similarity-based results based on querying PDB using T-M structural alignment program with the MahX model. T-M Score is a metric for assessing the similarity of protein structures. Identity is the % sequence identity in the structurally aligned region. Coverage denotes the alignment by T-M align and is equal to the number of structurally aligned residues divided by length of the query protein.

Rank	PDB ID	PDB Hit	T-M Score	Identity	Coverage
1	5dapA	Fe (II) /(alpha) ketoglutarate-dependent dioxygenase AsqJ	0.86	0.16	0.87
2	5m0tA	Alpha-ketoglutarate-dependent non-heme iron oxygenase EasH	0.82	0.17	0.86
3	4y5sA	Structure of FtmOx1 with a-Ketoglutarate as co-substrate	0.81	0.18	0.87
4	4naoA	Crystal structure of EasH	0.79	0.16	0.85
5	5zm2A	Fe (II) / (alpha) ketoglutarate-dependent dioxygenase AndA	0.79	0.18	0.82
6	5yblA	Fe (II) /(alpha) ketoglutarate-dependent dioxygenase AusE	0.77	0.17	0.83
7	6akzA	Crystal structure of GlcNAc Inducible Gene 2, GIG2 (DUF1479) from <i>Candida albicans</i>	0.69	0.09	0.84
8	6s0rA	The crystal structure of kanamycin B dioxygenase (KanJ) from <i>Streptomyces kanamyceticus</i> complex with nickel, sulphate and chloride	0.68	0.18	0.88
9	2csgA	Crystal Structure of the putative oxidoreductase from <i>Salmonella typhimurium</i> LT2	0.67	0.11	0.83
10	4xbzA	Crystal Structure of EvdO1 from <i>Micromonospora carbonacea</i> var. <i>aurantiaca</i>	0.66	0.15	0.80

As the PDB template 5dapA was identified as the highest ranked protein within the threading alignment analysis, its position as the highest ranked similar sequence, T-M score and coverage value was expected. Using the I-TASSER generated structural prediction model of MahX, the predicted functions and cofactors required were investigated using the programs COACH and COFACTOR. The analysis identified five ligand-binding sites within MahX. The highest ranked ligand-binding site was that for α KG. The confidence score, which ranges from 0-1 where a higher score means high reliability, was 0.43. The cluster size for this prediction, which is described as the total number of templates within the cluster, is 61. The 15 ligand binding site residues, shown in Figure 4-3, are located at the following positions; 105, 147, 149, 156, 159, 161, 177, 179, 192, 229, 235, 237, 246 and 250. The ligand-binding site displays the presence of the highly conserved double stranded β helix.

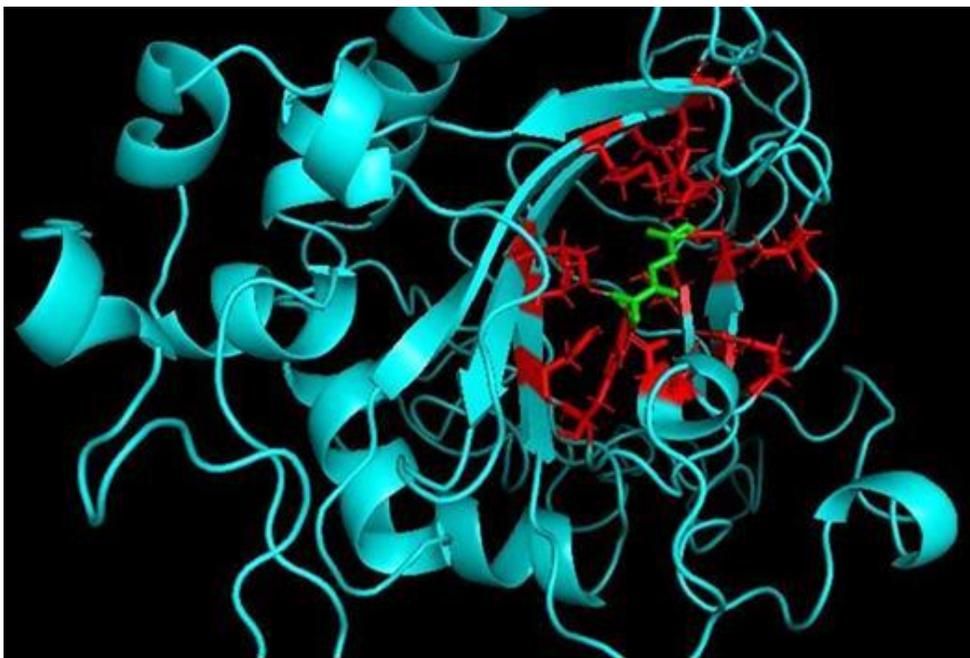


Figure 4-3: I-TASSER generated model of MahX. Red indicates α KG binding residues. The green structure displays the α KG molecule.

The second ranked ligand-binding site was that for Fe^{2+} . This prediction had a C-score of 0.22 and cluster size of 45. The analysis identified the His-Asp-His facial triad, shown in Figure 4-4, at residues 159,161 and 235. The importance of this ligand-binding site was further examined through multiple sequence alignment analysis of closely related protein sequences obtained through BLAST analysis. The His-Asp-His triad was demonstrated to be present in all sequences analysed showing the importance of these residues for enzyme activity, as shown in Figure-4-5.

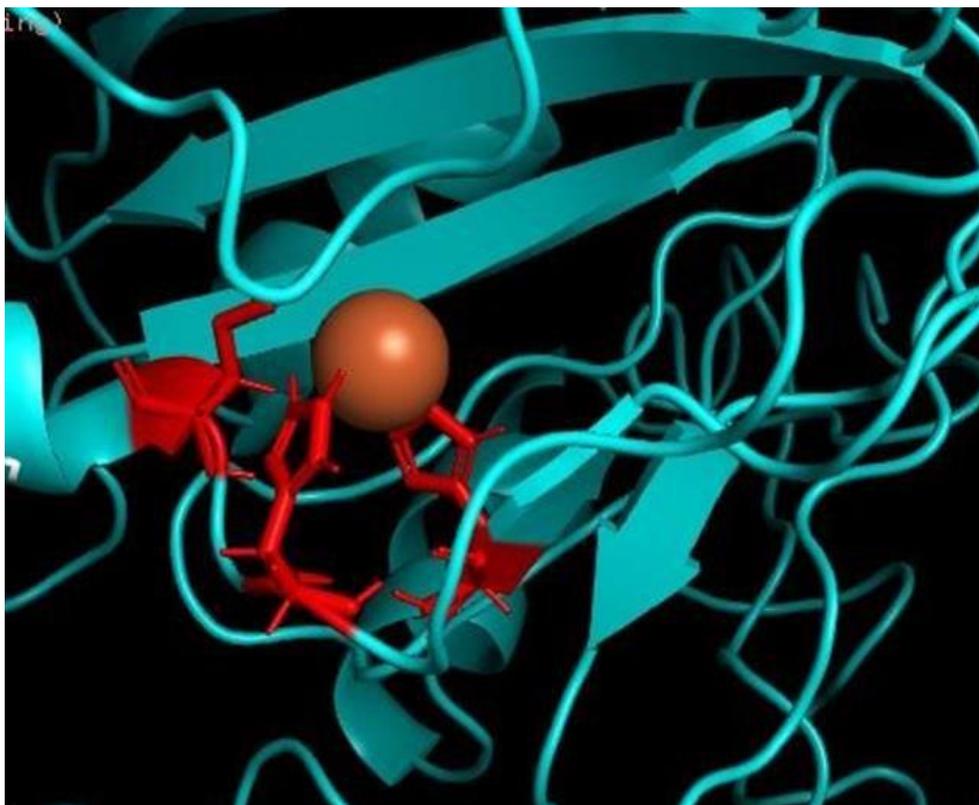


Figure 4-4: TASSER generated model of MahX. Red indicates Fe^{2+} binding His-Asp-His residues. The orange sphere represents the iron ion.

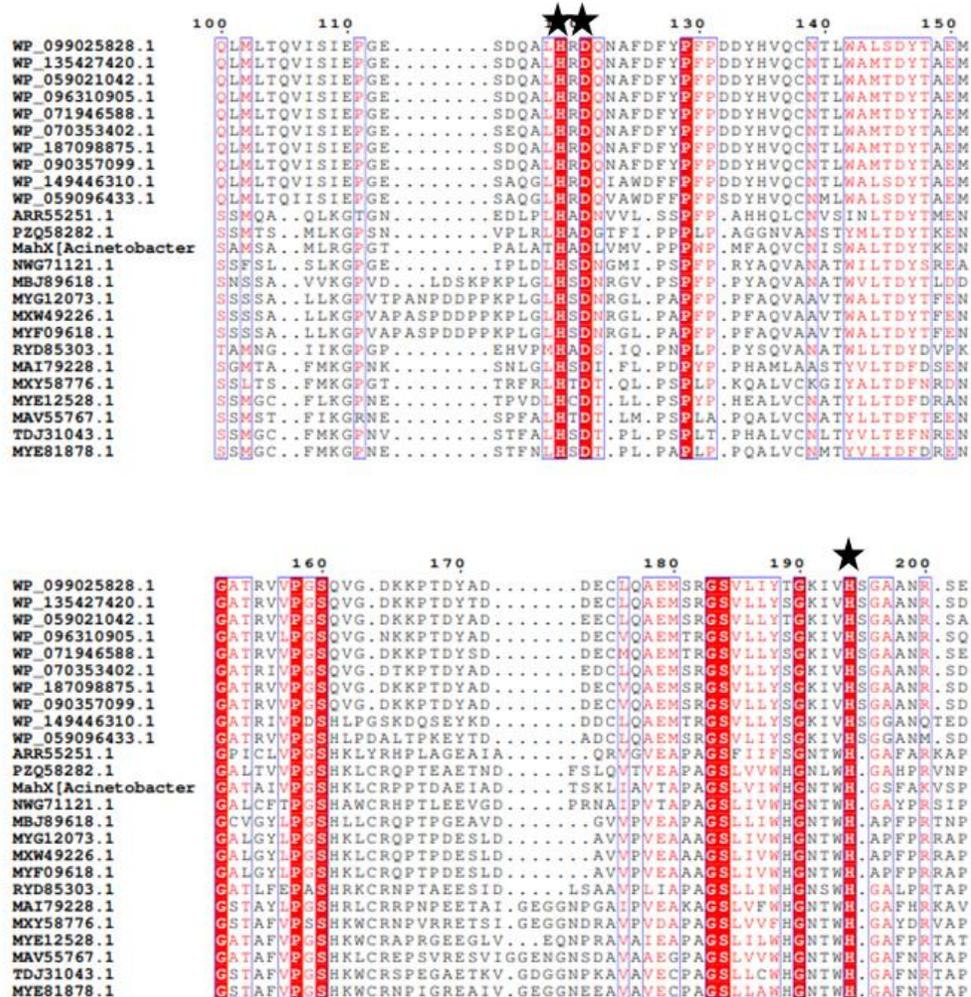


Figure 4-5: Multiple sequence alignment of sequences with high similarity to MahX identified through BLASTP analysis. Residues are coloured based on sequence similarity. With stars indicating conserved regions relating to Fe²⁺ binding. Image visualised on ESPrpt (2014).

The surface structure of the MahX model was simulated within molecular visualization system PyMol, to allow for a visual representation of the protein structure and ligand binding interactions within the model. Visual examination of the surface model, shown in Figure 4-6, revealed a ‘pocket’ within which the ligand binding sites reside. Residues were identified within this binding pocket through identifying residues within 5 Å of the ligands followed by manual curation. This analysis identified the presence of the following 12 residues: T96, S98, G100, E101, C107, H109, S145, M147, L156, V163, M164, I250.

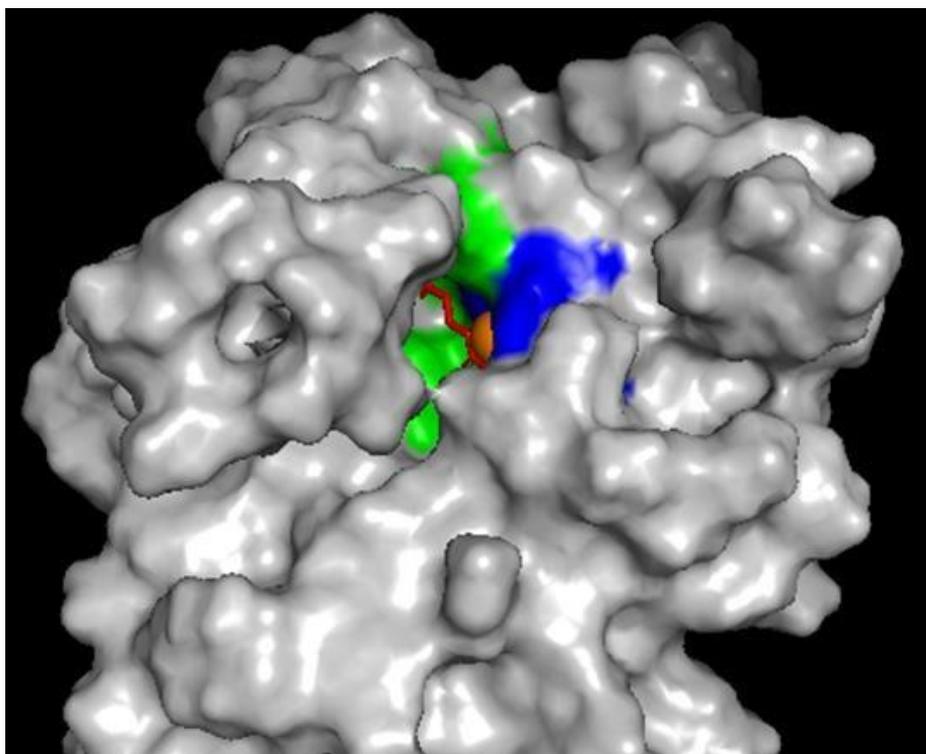


Figure 4-6: Surface visualisation of I-TASSER generated MahX model. Fe²⁺ and akG binding sites are highlighted blue and green, respectively. Image taken on PyMol

The consensus prediction of the molecular function of MahX was undertaken and describe potential activities associated with MahX. Each prediction has a GO-score, based upon the Cscore of the template and is shown in Table 4-2

Table 4- 2:Gene ontology terms associated with the predicted MahX protein model.

Molecular Function ID	Description	GO-Score
GO:0046872	Metal Binding Ion	0.59
GO:0016702	Oxidoreductase Activity	0.48
GO:0031406	Carboxylic Acid Binding	0.37
GO:0019842	Vitamin Binding	0.37
GO:0016706	2-oxoglutarate-dependent Dioxygenase Activity	0.37
GO:0005515	Protein Binding	0.34

Finally, further examination was conducted through superpositioning the MahX template with that of the 5DAPA. This allowed for conserved features to be identified within the modelling structure. The alignment, shown in figure 4-7, shows the conserved features relating to both the α KG and Fe^2 binding residues.



Figure 4-7-Superposition of MahX model (blue) with that of 5DAPA (orange). Key residues relating to α KG and Fe^{2+} binding are highlighted red.. Image generated using PyMol.

4.22 Expression and characterization of MahX

Initial experimental characterization of the metaldehyde degrading protein MahX required the construction of a vector containing *mahX* under an inducible promoter. As such, through using the gene synthesis company GenScript, *mahX* was cloned into the *E. coli* expression vector pET21a as shown in Figure 4-8. This construct was also designed to ensure a C terminal His tag

was fused to the MahX protein. This would allow for protein purification in downstream protein characterization assays.

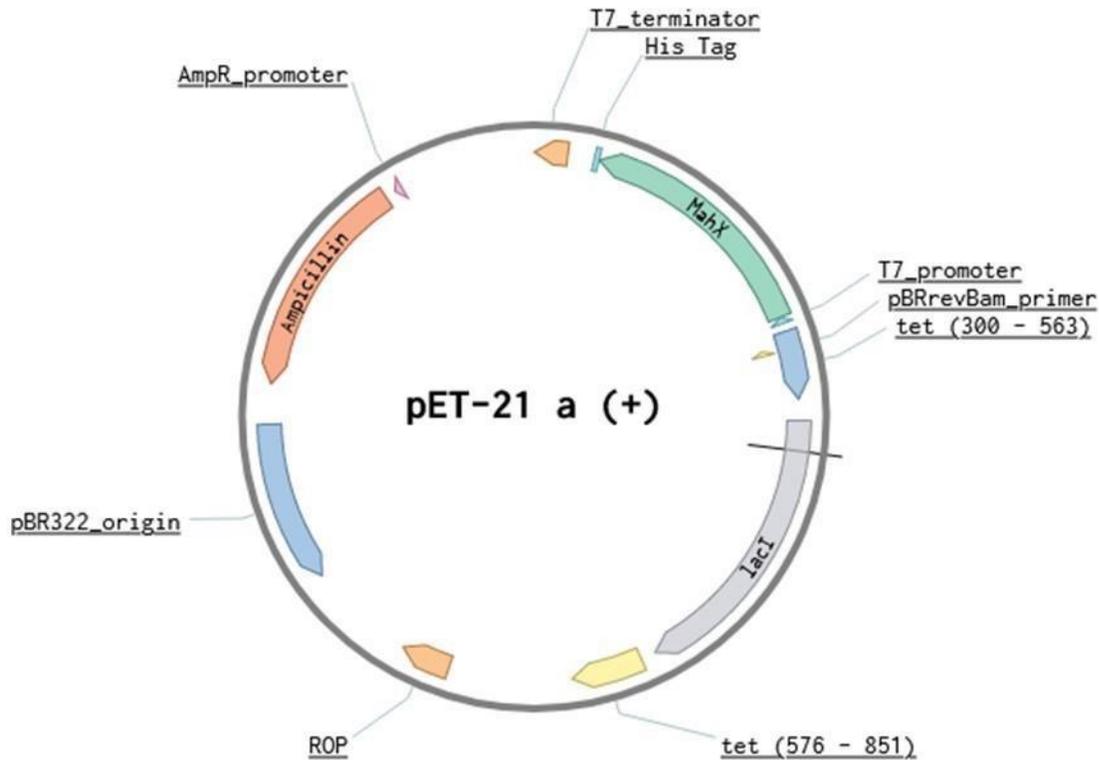


Figure 4-8: Schematic of the *E. coli* expression vector pET21a. Image generated through Benchling

Protein expression trials were undertaken with the aim to both demonstrate the inducible expression of MahX and to verify expression of a protein of the correct size. The expected size for the His tagged MahX is ~31 kDa. Visual inspection of the SDS-PAGE, shown in Figure 4-9, identified a band of significant density at ~ 32 kDa. No band at this position was observed in the no insert sample.

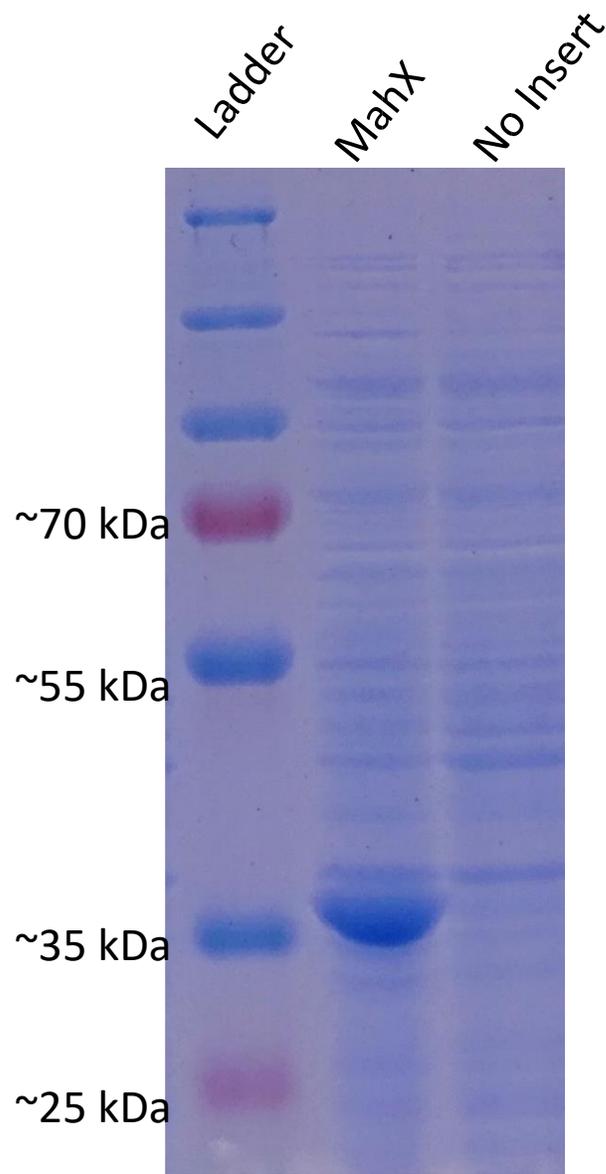


Figure 4-9: 10 % SDS-PAGE gel with *E. coli* BL21 pET21a no insert and MahX. 3 hour post IPTG induction .To each well, 15 μ L of protein sample was added. A pageruler ladder was used to estimate protein size.

As correct inducible expression was observed, optimization of protein expression was performed.

As purification of a protein typically requires the protein to be in a soluble state, the soluble extract of each of the induction time points was assessed. Initial solubility characterization was

performed following overnight induction. Following cell disruption with the detergent BugBuster, examination of the SDS-PAGE gel, shown in Figure 4-10, revealed observable presence of protein in the both soluble and insoluble extracts with greater density found in the latter.

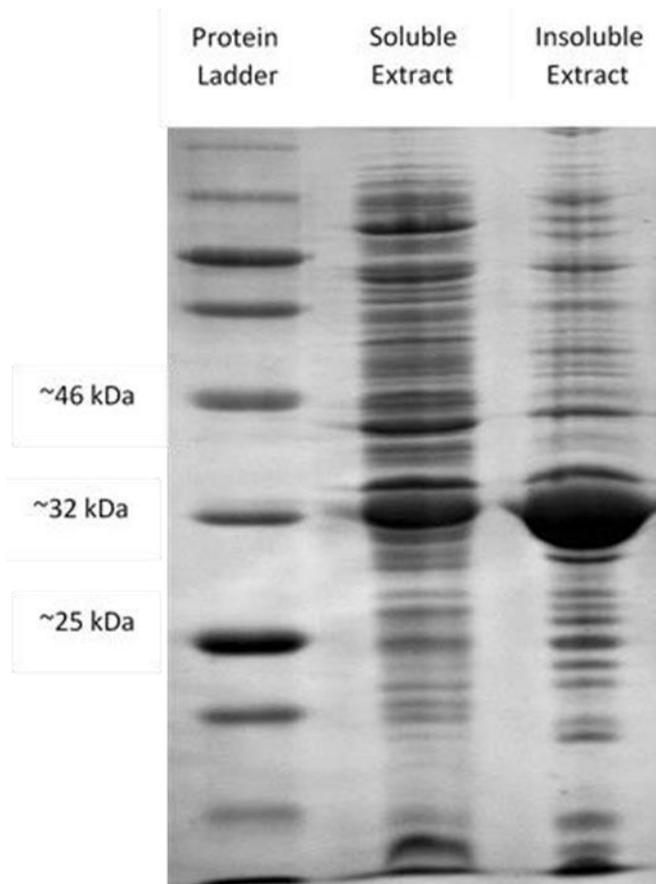


Figure 4-10: 10 % SDS PAGE of the soluble extract insoluble fraction of E. coli BL21 expressing MahX following overnight induction with 1 mM IPTG. PageRuler ladder was added to estimate protein size. 15 μ L of protein solution was loaded to each well.

As MahX was identified within the soluble extract, further analysis was undertaken to optimize the concentration of soluble protein obtained. Initial investigation, shown in Figure 4-11, examined the effect of induction time on MahX concentration within the soluble extract.

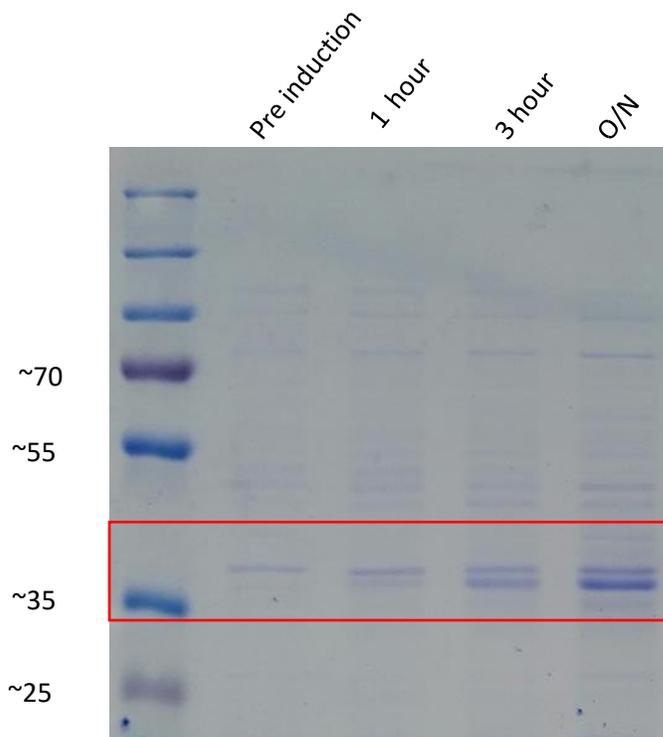


Figure 4-11: 10 % SDS-PAGE gel with soluble extract of *E. coli* BL21 expressing MahX following various induction time points. PageRuler was used to estimate protein size. 15 μ L of protein solution was loaded to each well.

Optimization of the protein solubility was undertaken through altering both the IPTG concentration and the temperature following IPTG induction. This solubility analysis, shown in Figure 4-12, revealed an IPTG concentration of 1 mM generated the greatest density of the His tagged MahX in the soluble extract. As a reduction in induction temperature to between 16-23 °C has been widely reported to increase protein solubility (Gutiérrez-González *et al.*, 2019), the solubility of MahX following incubation at 16 °C was examined. SDS-PAGE examination revealed

that the greatest expression of soluble protein occurred at 30 °C. Based on this analysis, the optimum conditions for soluble MahX was an IPTG concentration of 1mM, overnight induction at 30 °C.

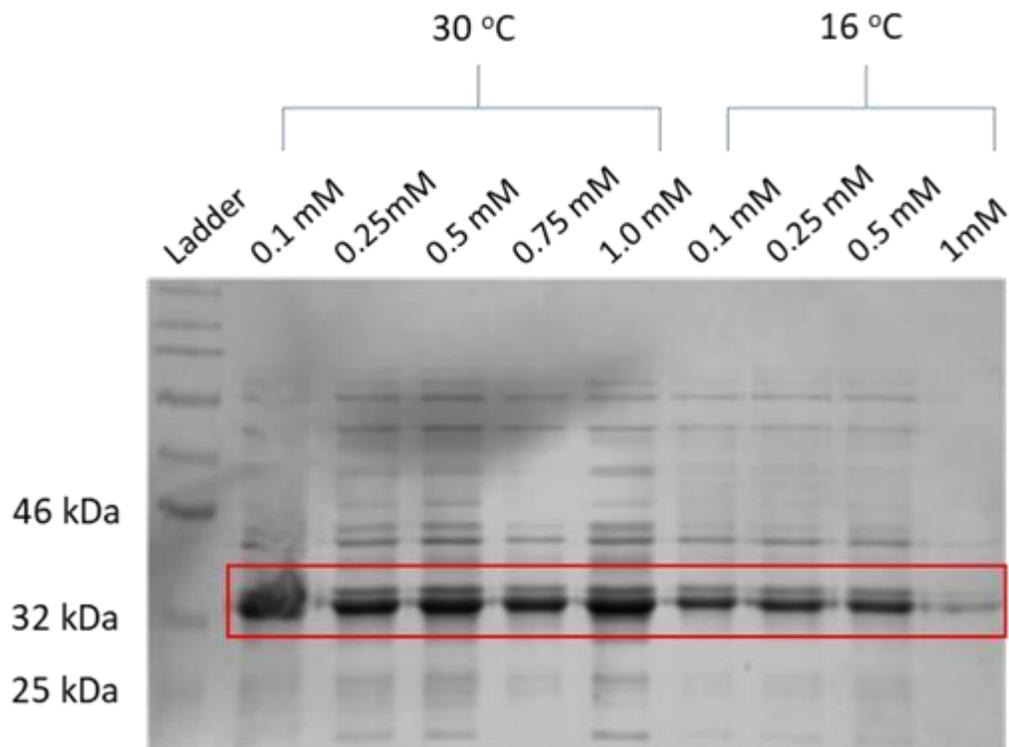


Figure 4-12: 10 % SDS-PAGE of the soluble protein extract following overnight incubation with various final concentrations of IPTG at either 16 °C or 30 °C. The red box highlights the expected MahX protein. Work was performed alongside J.Baines. Pagaruler was used to estimate protein size. 10 μ L of protein solution was loaded into each well.

Following the results from the solubility optimization assay, attempts were made to purify MahX.

This was performed using a 1 mL nickel based immobilized metal affinity chromatography (IMAC) HisTrap column. Examination of both the flow through and elution samples, shown in Figure 4-13, revealed no presence of the MahX protein. Further investigation identified the protein to be found within the insoluble pellet (image not shown).

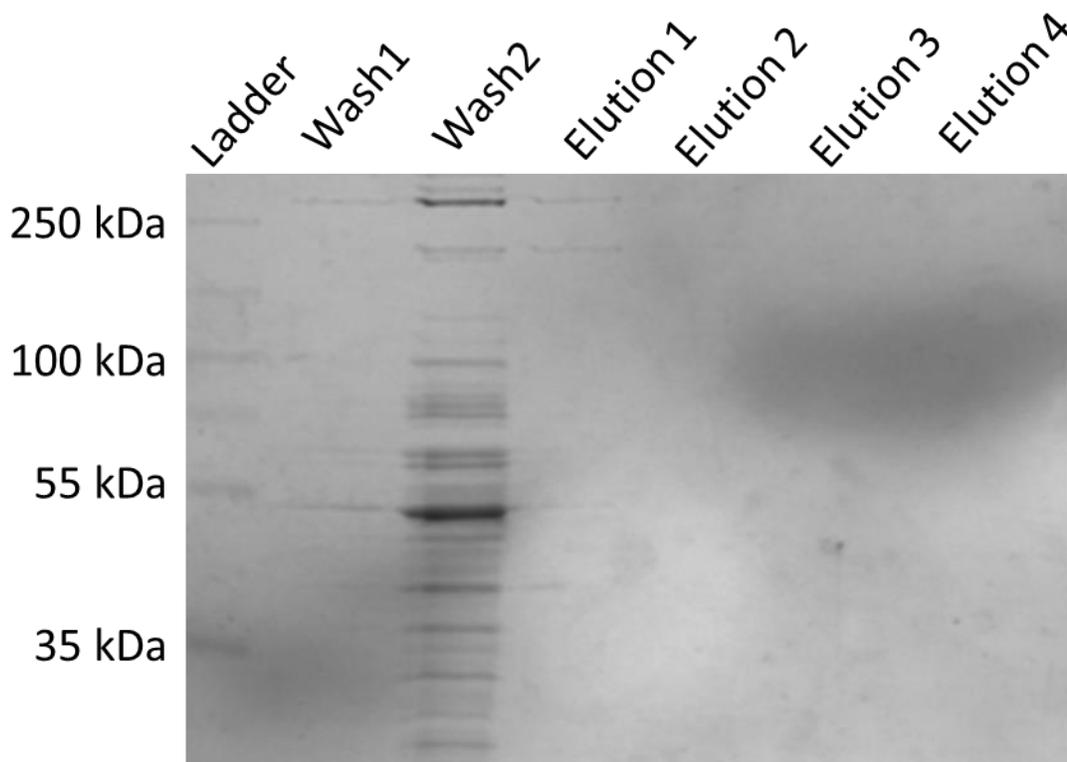


Figure 4-13 :12 % agarose gel of IMAC purification of His tagged-X. Wash and elution steps are shown. Pagenuler ladder was used to estimate protein size. Work was undertaken alongside J.Baines. 10 μ L of each sample was loaded into each well.

Numerous potential causes for the insoluble nature of the protein within the purification step were hypothesized. The solubility assays conducted previously involved cell disruption through detergent whereas cell lysis during purification was performed through sonication. As such, the potential cause of protein aggregation may have been due to the cell disruption method itself. Another possibility was that the proteins observed in the soluble extract were in fact nonfunctional inclusion bodies and the high concentration of proteins within the cell led to

protein aggregation. This theory is supported by the significantly larger bands present within the insoluble extracts. As both the BugBuster solubility assay and attempted purification method suggest the insolubility of the His tagged MahX, characterization through bioinformatics analysis was undertaken. The solubility tool Protein-Sol (Hebditch *et al.*, 2017), was used to determine the relative solubility of the His tagged MahX. The experimental dataset for Protein-Sol has a mean solubility of 0.45. Therefore, anything with a value lower is considered to be less soluble than the average *E. coli* protein within that dataset. The solubility value, as shown in Figure 4-14, for the tagged MahX was calculated to be 0.29 and therefore has a relatively low predicted solubility.

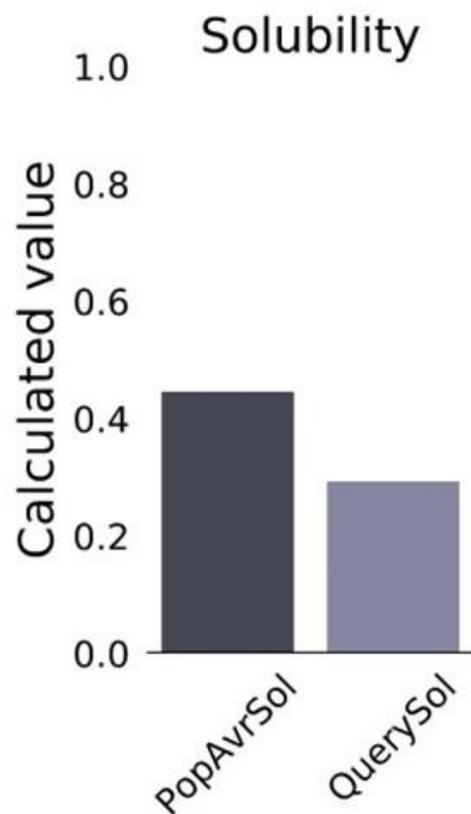


Figure 4-14: Relative solubility analysis compared against the mean solubility of the E. coli dataset. Calculations were undertaken using the Protein-Sol online tool

MahX-dependent metaldehyde disappearance was previously demonstrated in the heterologous expression assay within Chapter 2, when expressed from its native promoter within the *E. coli* strain DH5 α . As a crucial requirement of protein characterization requires active enzyme, the *in vivo* metaldehyde degradation of the His tagged MahX within the *E. coli* BL21 (DE3) strain was investigated. Through comparing the post-induced MahX samples with no vector BL21 control, the degradative ability of the overexpressed protein was determined. The results, shown in Figure 4-15, showed no significant change in metaldehyde over the time course of the assay, relative to the control (*p value .96*).

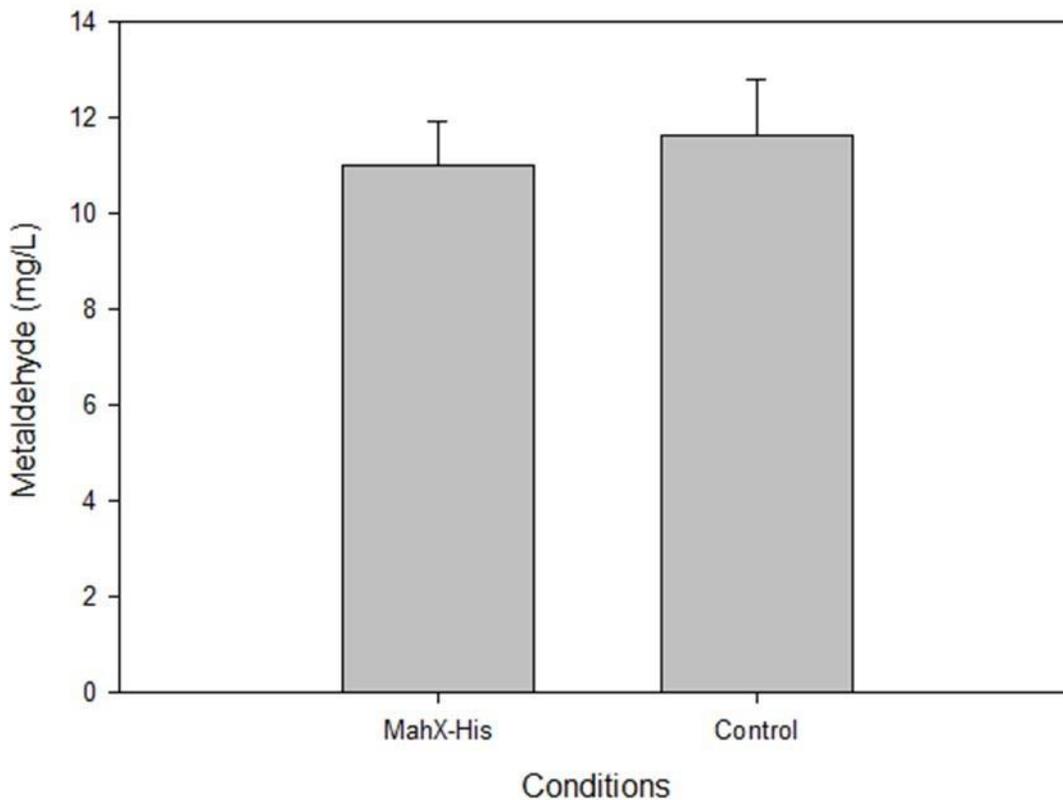


Figure 4-15: Metaldehyde concentration following 3.5-hour incubation with *E. coli* BL21 expressing MahX following overnight induction. A no vector BL21 sample was used as the control. Error bars display the standard deviation ($n=3$)

4.23 Expression and characterization of MBP-MahX

Several assays have demonstrated the insolubility of the His tagged MahX, potentially resulting from improper folding. A construct was created, shown in Figure 14-16, in order to overcome the solubility related issues. The protein maltose binding protein (MBP) has previously

demonstrated the ability to increase protein solubility (Kapust and Waugh, 1999; Sun *et al.*, 2011). As well as increasing protein solubility, the fusion of MahX with an MBP also provides an alternative purification method. The MahX protein was N-terminally fused to both a His tag and an MBP.

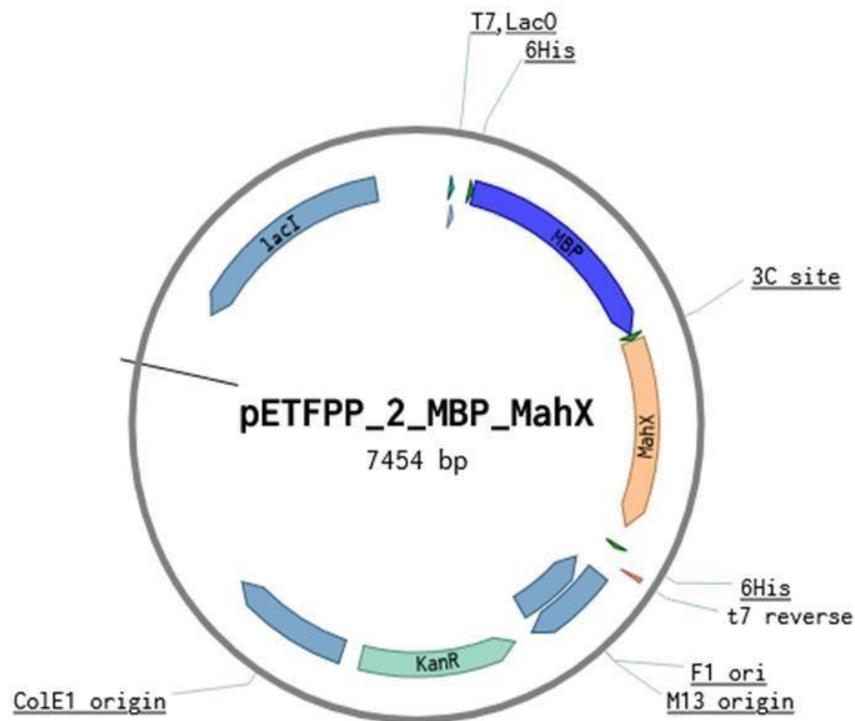


Figure 4-16: Schematic of the pETFPP_2 following the insertion of mahX. The vector contains an N-terminally fused his tag and mbp to mahX under the control of an IPTG regulated T7 promoter. Schematic generated using Benchling

The vector was constructed through PCR amplification of the *mahX* region, as shown in Figure 4-17, and ligated to the linearized vector donated by L.Clark (Potts Lab), using NEBuilder DNA assembly.

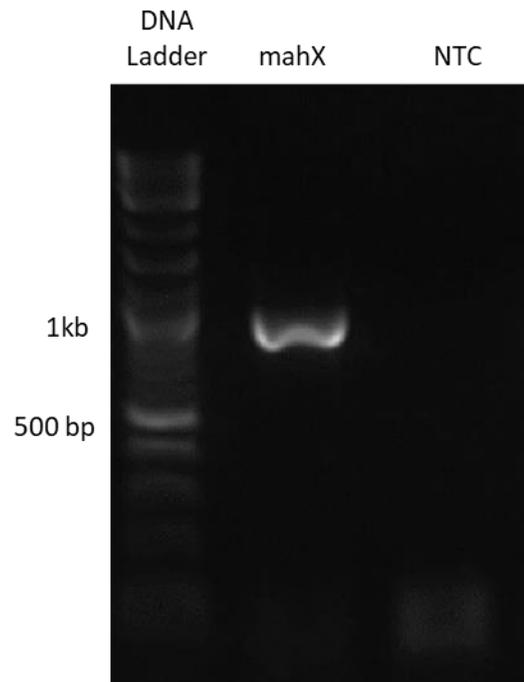


Figure 4-17 :A 1.2% agarose gel showing the *mahX* PCR product. A quick load log2 ladder was used to estimate DNA size.

Transformation resulted into 34 colonies. No colonies were identified on the no insert, no vector and no ligation mixture plates. Following a successful transformation step, the correct insert was verified through PCR and Sanger sequencing. As mentioned previously, the expected band size for the amplified *mahX* gene is 945 bp. The DNA gel, shown in Figure 4-18, verified the correct size of the banding by displaying a band ~ 1kb.

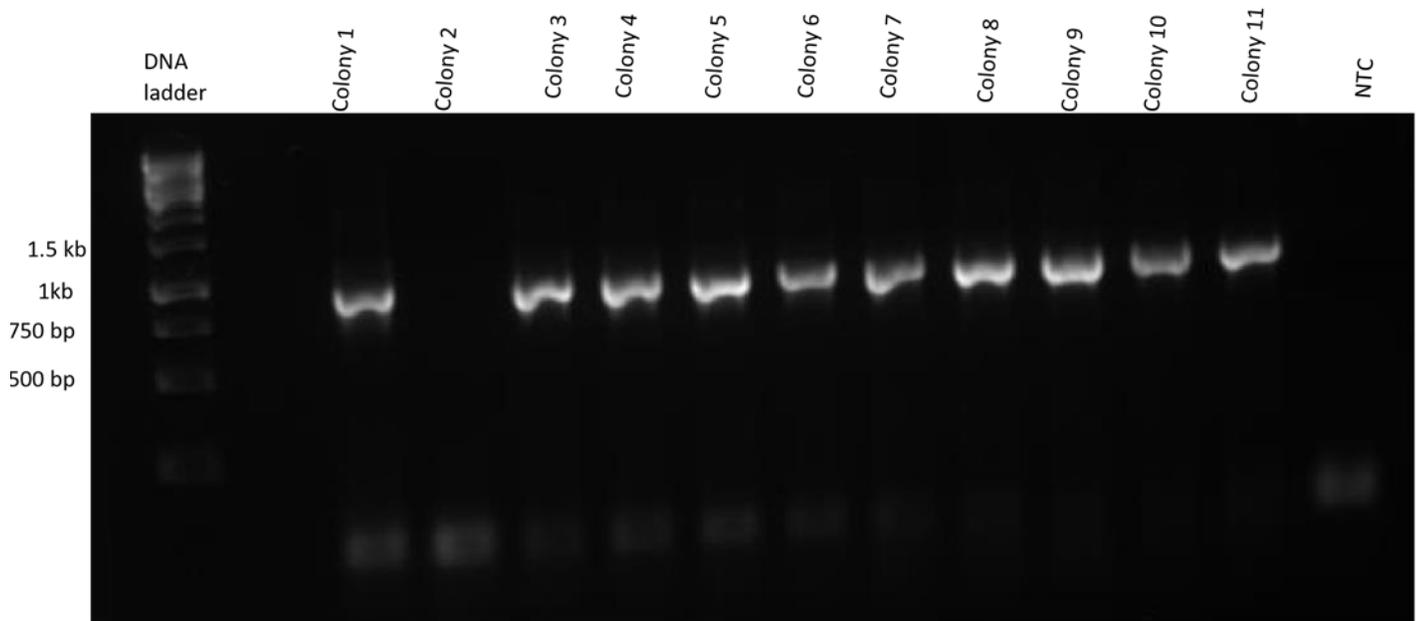


Figure 4-18: A 1.2% agarose gel showing 11 colony PCR products following ligation of mahX into pETFPP_2.A 1kb DNA ladder was used to estimate protein size.

The expected size for the MBP-MahX construct is 77 kDa and was confirmed through gel examination. The SDS-PAGE analysis, shown in Figure 4-19, revealed an induced band of ~70 kDa within the induced construct. No band was present in the no insert sample thereby confirming the correct protein was being induced and expressed.

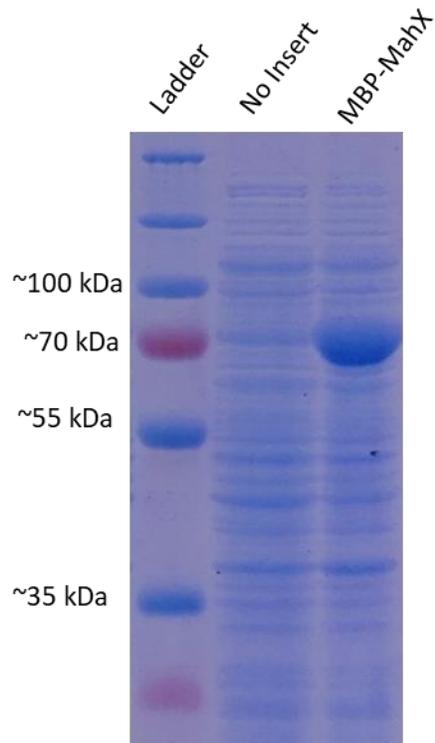


Figure 4-19: 10% SDS-PAGE gel with whole cell lysate of E. coli BL21 vector containing MBP-MahX following 3 hour induction. To each well, 15 μ L of protein estimate protein size.

The optimization of protein expression was conducted through observing the band intensity of various post induction time points.

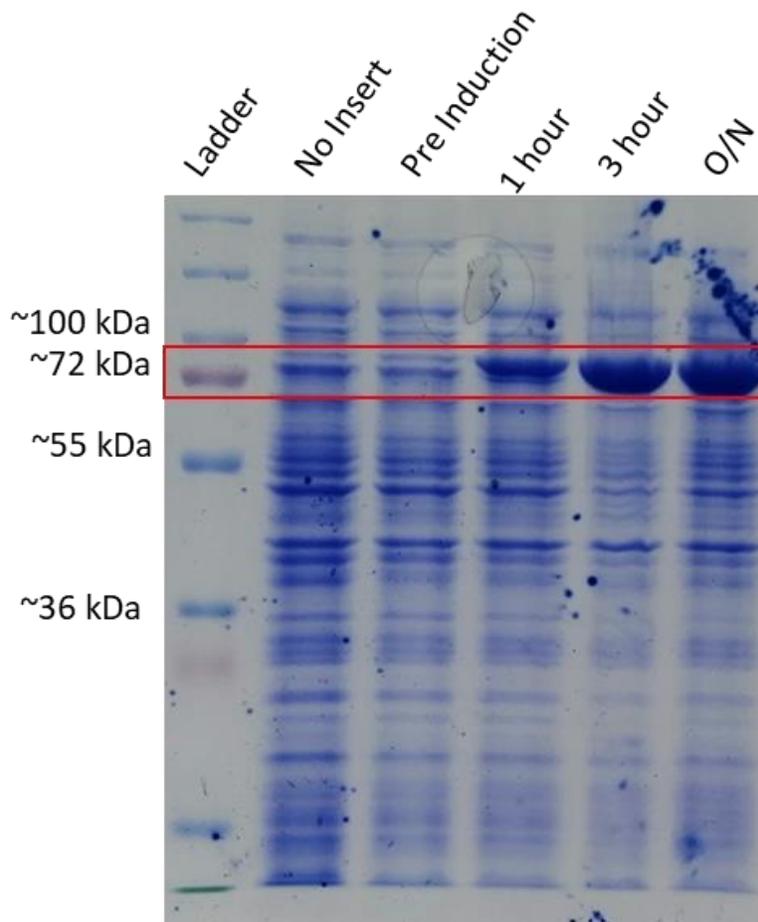


Figure 4-20: 10 % SDS-PAGE gel following E. coli BL21 expression of MBP-MahX. Whole cell lysate of pre induction, and 1 hour, 3 hour and overnight induction samples are shown. PAGERuler prestained ladder was used to estimate size. To each well, 15 μ L of protein sample was added.

Through visual examination of the SDS-PAGE gel in Figure 4-20, overnight induction generated the protein band of greatest density. As the protein solubility of the previous MahX construct was considered a reason for inclusion body presence and no observable degradative activity, the solubility of the MBP-MahX fusion protein was investigated. Through use of the detergent BugBuster, protein of overnight induced cells was compared against that of the soluble extract.

Visual examination, as shown in Figure 4-21, revealed significant presence of MBP-MahX within the soluble fraction.

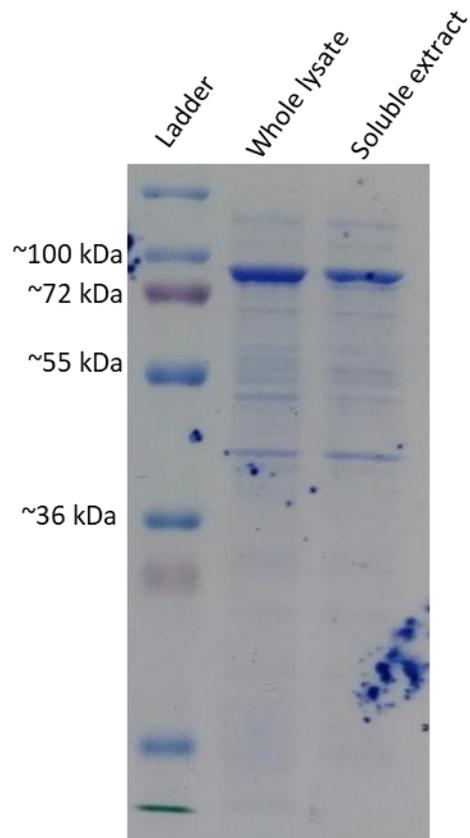


Figure 4-21:10 % SDS-PAGE of E. coli BL21 expressing MBP-MahX. Whole cell lysate and soluble extract are shown. Prestained PageRuler was used to estimate protein size. 15 μ L of sample was added to each well.

The effect of induction time on protein solubility was undertaken to identify the optimum time to harvest cells for downstream experiments and purification assays. Analysis, shown in Figure 4-22, revealed the greatest band density was present following overnight induction.

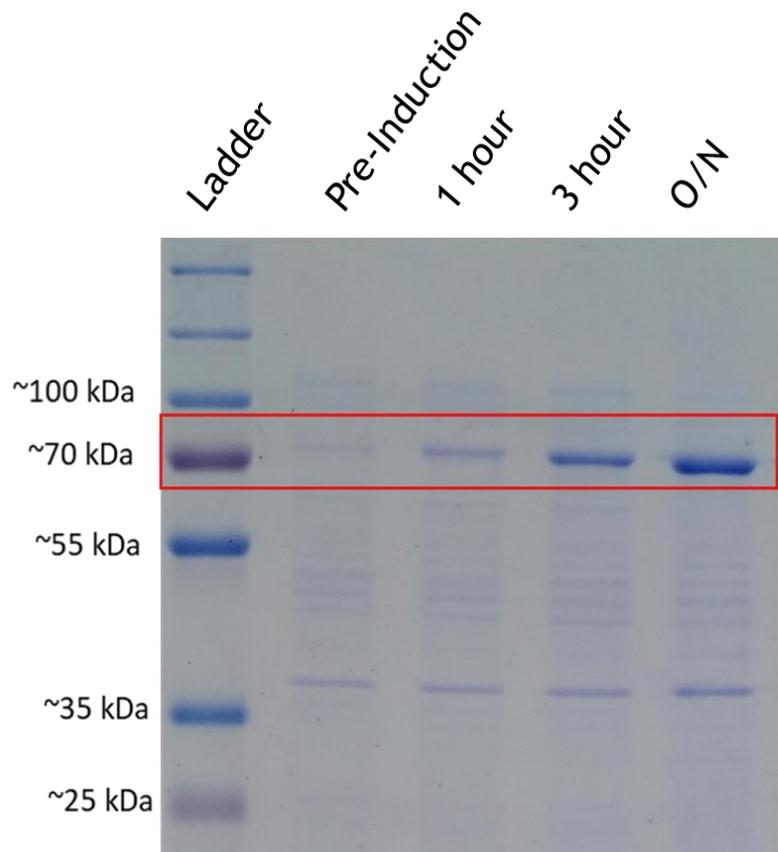


Figure 4-22: 10% SDS-PAGE of the soluble extract of MBP-MahX following BugBuster treatment. PageRuler was used to estimate the protein size. 15 μ L of protein sample was loaded to each well.

Due to the importance of demonstrating enzyme activity of the overexpressed protein, an *in vivo* degradation assay was conducted. Following 3.5 hours incubation with 15 mg/L of metaldehyde, the overnight induced MBP-MahX sample demonstrated removal of metaldehyde below the limit of detection shown in Figure 4-23. Demonstration of protein activity *in vivo* allowed for further characterization of MBP-MahX outside the cellular environment.

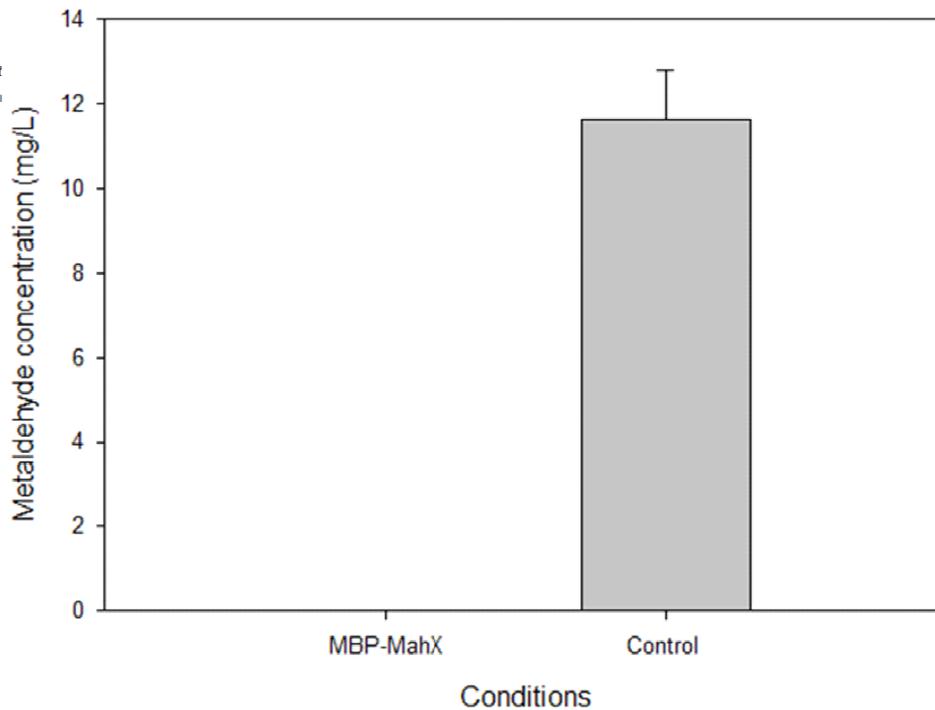


Figure 4-23: Metaldehyde concentrations following a 3.5-hour incubation of 15 mg/L metaldehyde. *In vivo* post induced cells normalised to OD600 of 1.0. Samples were performed in biological triplicate. Control sample was BL21 (DE3). Error bars show standard deviation.

4.24 In vitro characterization of MahX

As the MBP-MahX construct revealed the ability to degrade metaldehyde *in vivo*, further protein characterisation focused on analysing the soluble crude extract of MBP-MahX expressed within *E. coli* BL21. With the aim of the investigation to both determine soluble crude extract activity and to identify which cofactors are essential for optimum degradation, a metaldehyde degradation assay was undertaken. The initial assay involved comparing a no extract control supplemented with cofactors against that of the MBP-MahX crude extract supplemented with three cofactors identified through the literature and bioinformatic analysis. Following 3.5 hours incubation, metaldehyde was degraded within the soluble crude extract samples supplemented

with the cofactor α KG, LAA and Fe^{2+} , as shown in Figure 4-24. Comparisons against the no crude extract control, demonstrated that the difference observed was significant (*p* value. 0014). Of the 40 mg of crude extract protein used in this reaction, a metaldehyde removal rate of 0.7 nmol/min/mg of protein was calculated. As such, this assay exhibited the degradative potential of the crude extract and validated the findings identified previously.

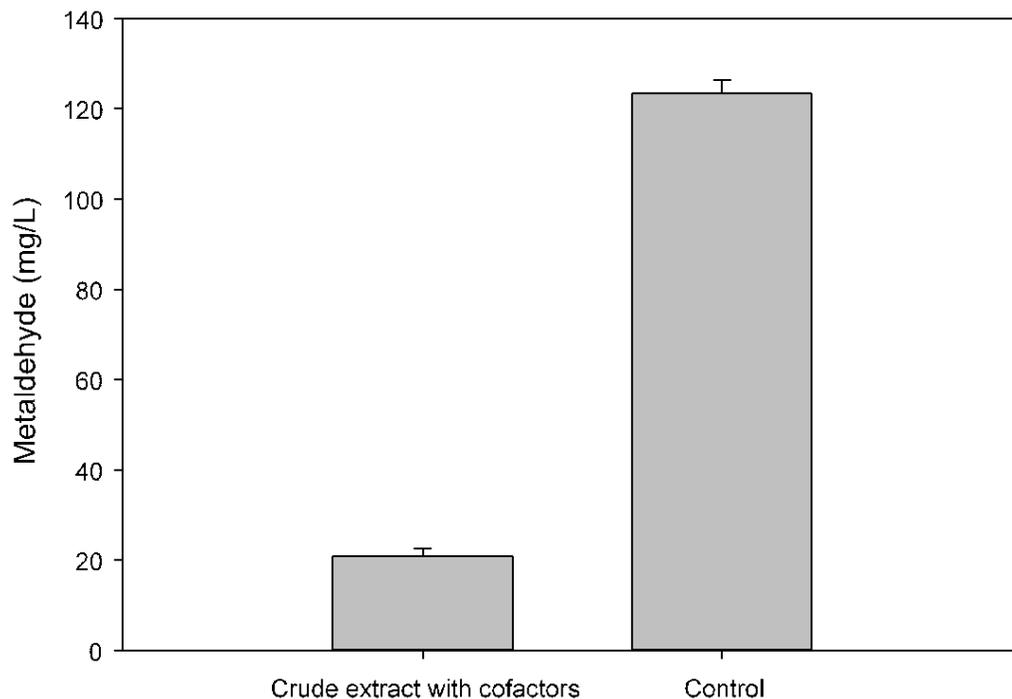


Figure 4-24: Metaldehyde degradation assay with 40 mg of overnight induced crude extract supplemented with 1mM of Fe^{2+} , LAA and α KG. A no extract control was used. Error bars show standard deviation (n=3).

As the presence of all three cofactors led to the significant removal of metaldehyde, a degradation assay to determine which combination of cofactors were essential was undertaken. Through observing the change in metaldehyde over 10-minute intervals, the rate of degradation at each time point was calculated. The greatest change in rate was observed in the samples

containing all three of the cofactors, as shown in Figure 4-25. All samples containing α KG showed the greatest degradative rate thereby verifying the importance of the cofactor. Following the initial 10 minutes degradation, a significant decrease in rate was observed in all sample conditions showing significant metaldehyde degradation.

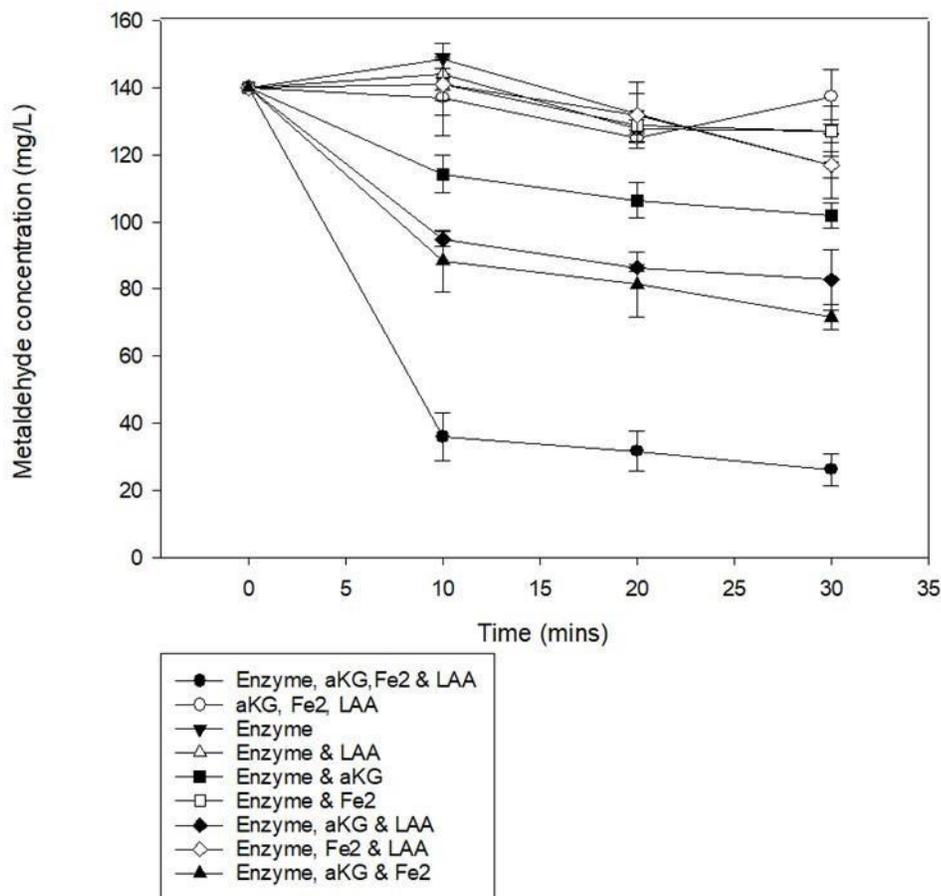


Figure 4-25: Metaldehyde degradation with 40 mg of overnight induced MBP-MahX BL21 soluble crude extract. Samples were supplemented with various combinations of Fe^{2+} , LAA and/or α KG. Error bars show the standard deviation ($n=3$).

As both the previous analysis demonstrated the importance of a reducing agent in metaldehyde degradation with MahX, the role of alternative reducing agents was undertaken. The two alternative reducing agents chosen for the assay were NADH and DTT. The former was chosen

due to its relative abundance within *E. coli* and therefore likely to assume the role of reducing agent during *in vivo* degradation (Gutiérrez-González *et al.*, 2019). All samples were supplemented with α KG and Fe^{2+} and a reducing agent demonstrated degradation of metaldehyde over the time course of the assay. Based on the metaldehyde concentration of the end time points, NADH displayed the greatest reduction of metaldehyde, as shown in Figure 4-26. This difference was significant relative to both the DTT and LAA samples (p value .015 and 0.001, respectively).

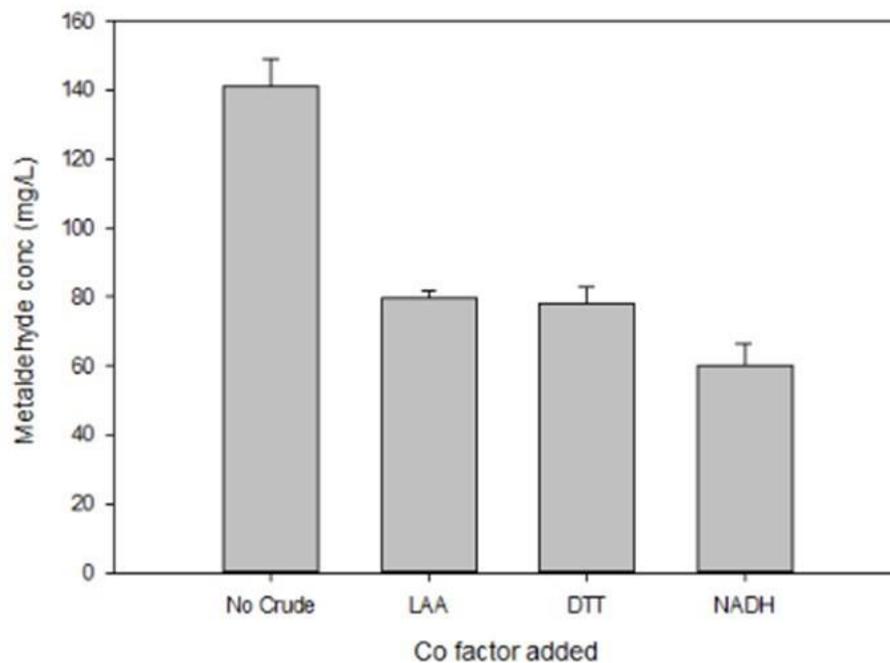


Figure 4-26: Metaldehyde concentration following incubation of 150mg/L with crude extract, Fe^{2+} , α KG and a reducing agent. Samples were incubated for 3 hours. Error bars show the standard deviation($n=3$).

Purification of active enzyme allows for further characterization of MahX as it would allow experimental structural analysis to be performed and would allow for accurate comparisons against other metaldehyde degrading proteins identified in future studies. Using the His tag of the MBP-MahX, a nickel IMAC purification method was undertaken. Purification was achieved with sufficient quantity of enzyme to perform degradative assays, as shown in Figure 4-27.

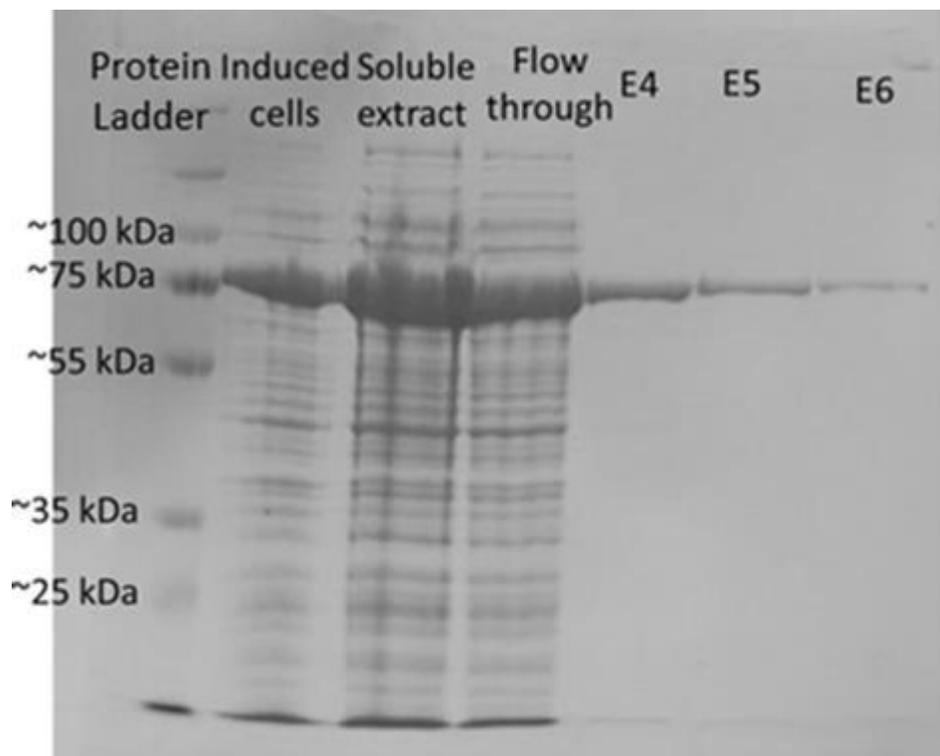


Figure 4-27: 12 % SDS-PAGE gel with stages stages of protein purification of MBP-MahX using a nickel based IMAC column. Elution samples shown are of the 3 highest protein concentrations as determined by A_{280} . 10 μ L of each sample was added to each well.

Through supplementation of the purified MBP-MahX protein with the cofactors identified in the previous analysis, the activity of the protein was assessed. Incubation of 0.179 mg of the enzyme supplemented with the cofactors failed to demonstrate a significant reduction in metaldehyde (p value .17), as shown in Figure 4-28.

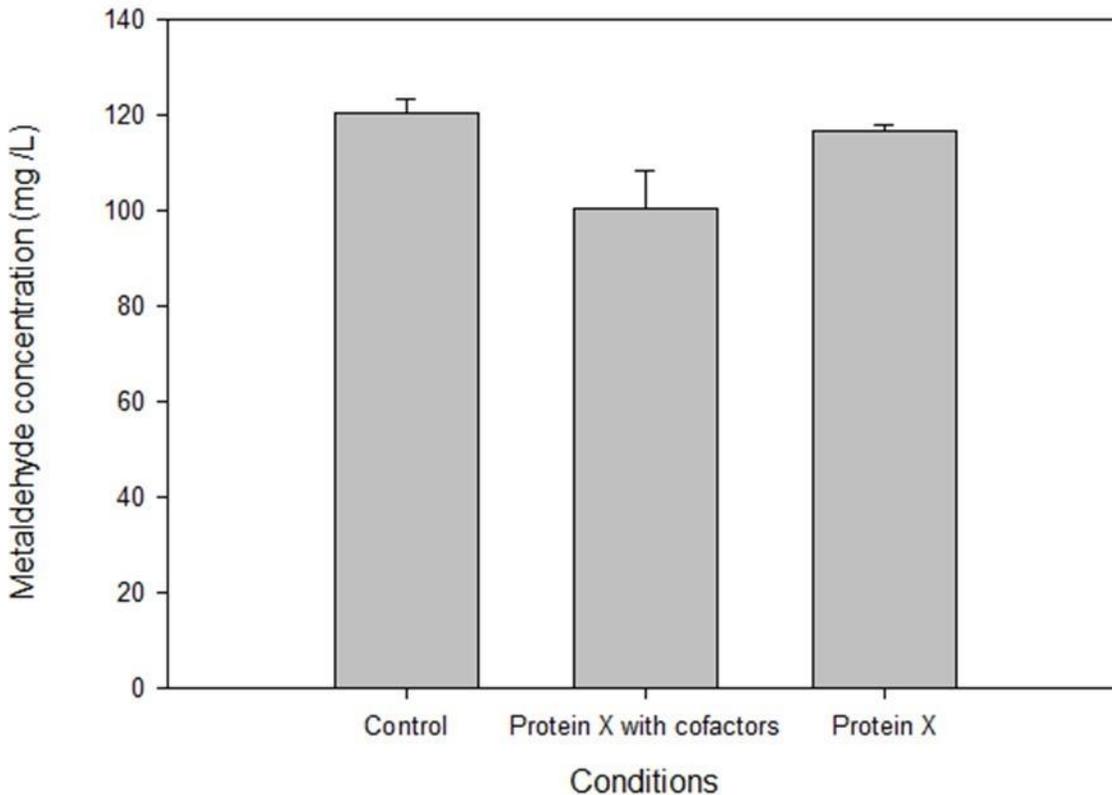


Figure 4-28: End point metaldehyde concentration following the incubation of the nickel IMAC purified MBP-MahX. 0.179 mg of purified protein was added and incubated with 1 mL of final concentration 120 mg/L of metaldehyde for 3.5 hours. Enzyme samples were performed both with and without cofactor. Control contained no enzyme or cofactor. Error bars display the standard deviation. ($n=3$).

The divalent metal nickel has previously demonstrated inhibitory effects towards Fe^{2+} / αKG dependent dioxygenases through the presence of nickel leading to the replacement of the ferrous iron in the catalytic centre (Chen *et al.*, 2010; Yin *et al.*, 2017). To determine whether nickel inhibition was leading to a reduction in MBP-MahX activity, the use of the MBP's affinity for amylose was utilized. The eluted fractions are shown in Figure 4-29.

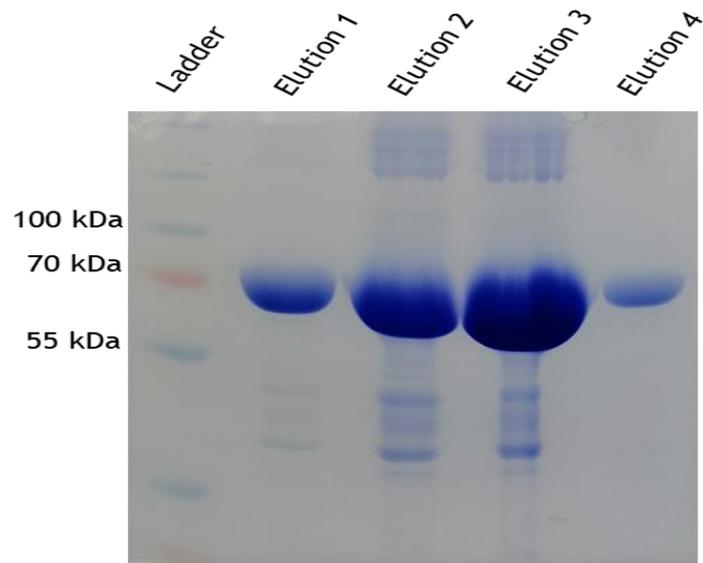


Figure 4-29: 10 % SDS-PAGE gel of elution samples following amylose purification of MBP-MahX. PageRuler ladder was used to approximate protein size. To each well, 15 μ L of protein sample was added.

Activity of the amylose purified MBP-MahX protein was examined through a degradation assays. Significant reduction of metaldehyde was observed in the purified protein relevant over the time course of the assay (p value .0002), as shown in Figure 4-30. Based on the loss of metaldehyde over the time course an enzyme activity of 8.1 nmol/min/mg of protein was calculated for the amylose purified protein.

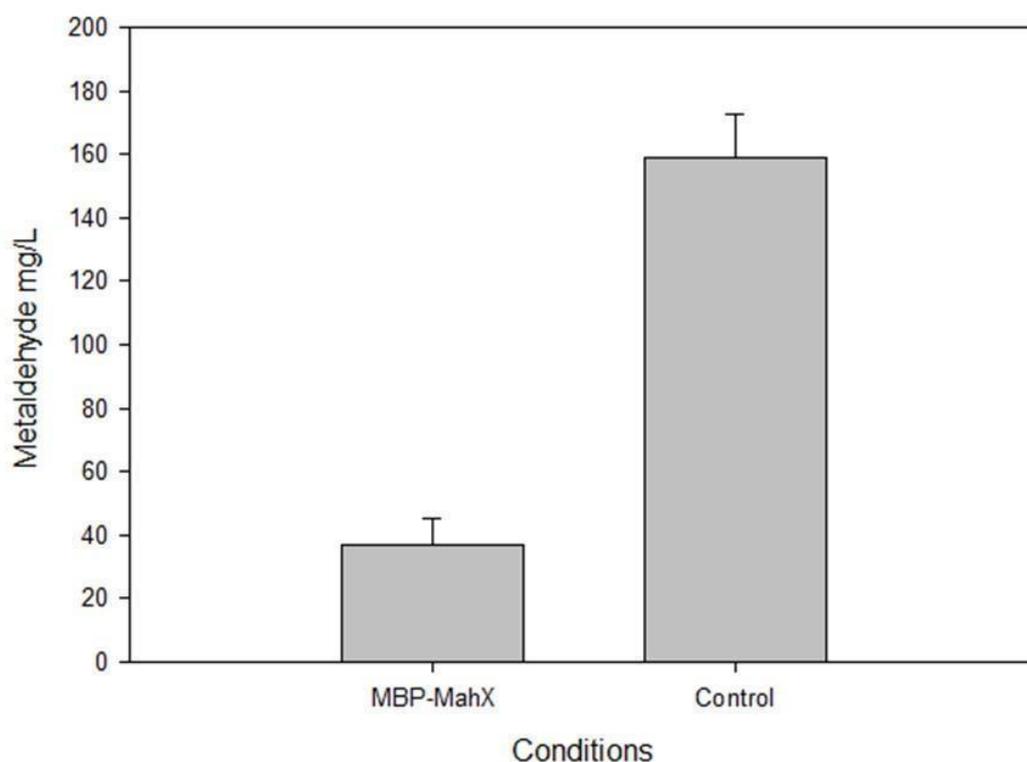


Figure 4-30: Metaldehyde degradation assay of the amylose purified MBP-MahX. 4mg of MBP-MahX was added and incubated with 10 mL of 150mg/L metaldehyde for 3.5 hours. Samples were performed in triplicate. A no extract control containing just buffer with cofactors was as a control.

4.3 Discussion

Work conducted within the previous chapter led to the discovery of the first identified metaldehyde degrading protein MahX. Further analysis revealed that the protein belongs to the family of Fe (II)/ (alpha) ketoglutarate-dependent dioxygenases and has demonstrated activity upon heterologous expression within *E. coli*. Although attempted previously, prior this work, metaldehyde degradation was yet to be observed *in vitro*. This research set out to further characterize the MahX protein using both experimental and bioinformatics approaches.

Protein modelling generated through I-TASSER allowed for a model of high alignment quality to be obtained. This therefore allowed for accurate identification of the required cofactors needed for the experimental assays as well as further verification of the role as a dioxygenase. Further examination of the protein structure itself could be achieved in future investigations through experimental approaches such as X-ray crystallography and would allow a more accurate protein model to be generated.

4.31 In vitro activity

Initial protein analysis involved the expression of a C terminal His tagged MahX. Experimental analysis revealed the induced expression of an enzyme of the correct size. However, further examination revealed no observable activity of the overexpressed protein. As both previous and subsequent work has demonstrated heterologous activity within the expression host *E. coli*, this suggests complications relating to the protein itself as opposed to alternative factors. As MahX lacking a His tag previously demonstrated significant degradative ability *in vivo*, this would suggest complications regarding either protein expression rate or the position of the tag. As the ferrous centre of alpha-ketoglutarate dioxygenases are essential for efficient activity, the chelating properties of the poly histidine tag may lead to either the sequestering of the iron within the solution or preventing effective iron coordination with the relevant ligands. Within this investigation, the lack of enzyme activity was overcome through fusing MahX with an N terminally bound MBP and His tag. As such, this would suggest the positioning of the His tag itself may have resulted in poor protein folding and the presence of inactive inclusion bodies. Alternatively, the

expression of a protein of relatively low solubility may have exacerbated the accumulation of inclusion bodies. As MBP has been demonstrated to improve both protein folding and solubility, both potential causes would have been overcome through use of MBP fusion (Kapust and Waugh, 1999; Sun *et al.*, 2011).

Prior to this work, no evidence of *in vitro* metaldehyde degradation had been observed. Identification of key cofactors, through model construction and literature analysis, identified three important cofactors needed for traditional Fe (II)/ (alpha) ketoglutarate-dependent dioxygenase activity. Through crude extract analysis coupled with examination of purified protein activity, the importance of Fe²⁺, α KG and a reducing agent was demonstrated. Further analysis to optimize *in vitro* degradation, revealed the role of NADH as the optimum reducing agent for MahX activity relative to DTT and LAA.

4.32 Purified Protein

As Fe (II)/ (alpha) ketoglutarate-dependent dioxygenase are dependent on ferrous iron for optimum activity, it was hypothesized that nickel within the purification column may be targeting these binding sites. Previous research has demonstrated the high sensitivity to nickel inhibition within Fe (II)/ (alpha) ketoglutarate-dependent dioxygenases and as such an alternative purification method was chosen (Davidson *et al.*, 2006; Chen *et al.*, 2010; Yin *et al.*, 2017). In order to ensure optimum protein activity and overcome any potential inhibition, the MBP tag was utilized. Binding to an amylose column thereby bypassing the need for divalent metals and allowing for further examination of MBP-MahX to be undertaken. This resulted in a significant

reduction in metaldehyde with an activity of 8.1 nmol/min/mg of protein. As the rates analyzed was based upon an end point, as opposed to frequently collected time point, the assumption is based on the rate being constant throughout the reaction. Based on the crude extract assay, following the first 10 minutes of the reaction, the rate drops at a significant rate. As such, further work should be conducted to provide more accurate and real time data values for each condition. This would provide not only greater insight into the potential inhibitory effects but would also allow for the key enzyme characteristics, such as V_{\max} and K_m to be calculated.

As MBP-MahX required numerous cofactors to ensure optimum metaldehyde degradation, the use of purified enzyme for industrial bioremediation appears unsustainable. Significant activity was observed *in vivo* and so the use of overexpressed whole cells would appear to be the most appropriate remediation approach. As cofactors and conditions are regenerated and maintained within the cellular environment, this ensures high degradative rates over greater durations. Such an approach has been utilised in numerous studies in the bioremediation and biotransformation of compounds such as heavy metals and organophosphates (Kostal *et al.*, 2004; Yang *et al.*, 2008; Ameen *et al.*, 2020).

MahX was demonstrated to require α KG for efficient metaldehyde degradation. Therefore, spectrophotometric bioassay may be constructed with the aim of quantifying metaldehyde. The decomposition of the cofactor α KG into succinate and CO_2 can be exploited to provide a simple and high throughput bioassay with the potential for detecting the concentration of the primary substrate. This can be achieved through coupling the formation of the succinate to the conversion

of NADH to NAD, with the use of the enzymes succinyl-coenzyme A synthetase, pyruvate kinase, and lactate dehydrogenase (Luo *et al.*, 2006). As well as providing a useful tool for determining the uncoupling turnover rate of MahX for future studies, this approach may be utilized to detect metaldehyde within a given matrix by using the NADH oxidation as a proxy for the metaldehyde concentration.

Cofactor analysis also identified the importance of a reducing agent within MahX activity. No binding site or experimental specificity was observed within the bioinformatic or experimental analysis as all reducing agents tested demonstrated activity. As such, the role of the reducing agent within metaldehyde degradation appears to be a protective one. This is in line with current literature surrounding reducing agent's role within dioxygenase activity (Kuiper and Vissers, 2014; Martinez and Hausinger, 2015). As relationship between metaldehyde degradation and NADH reduction is not considered stoichiometric, a biosensor based on direct NADH oxidation methods would not therefore be appropriate.

4.33 Conclusion

This work aimed to characterize the metaldehyde degrading enzyme MahX and determine the conditions for providing *in vitro* activity of the enzyme. This was achieved through both crude and purified conditions. Further characterization of the enzyme through determining the initial degradative rates within numerous conditions, such as temperature and pH, is essential to provide greater insight into the optimum degradative ability of MahX. With regards the bioremediation potential of MahX, further work should be conducted to determine the ability to degrade metaldehyde below the regulatory limit within spiked water samples. .

Chapter 5: Identification and characterization of novel metaldehyde degrading genes

5.1 Introduction

Work conducted within Chapter 3 and 4 led to the discovery and experimental verification of the first metaldehyde degrading protein MahX. This protein was identified as belonging to the protein superfamily of α KG-Fe (II) oxygenase superfamily and requiring alpha ketoglutarate, Fe^{2+} , and a reducing agent such as NADH to ensure optimum activity. The *mahX* gene has subsequently been identified within numerous metaldehyde degrading strains such as *Acinetobacter bohemicus* JMET-C, *Acinetobacter lwoffii* SMET-C and *Pseudomonas vancouverensis* SMET-B (Castro-Gutiérrez et al., 2020). Bioinformatic analysis conducted within chapter 3 revealed the predicted sequence with the highest similarity to MahX was found within *Sphingobium japonicum*. The *Sphingobium* genus are aerobic, Gram-negative bacteria and have been associated with the biodegradation of a wide range of xenobiotic compounds such as chlorinated phenols, linuron and aliphatic hydrocarbons, amongst others (Ji Zhang et al., 2012; Liu et al., 2012; Ahn et al., 2018; Jia et al., 2020). Recent work conducted by Castro Gutiérrez et al., (2020) led to the isolation and identification of two metaldehyde strains which have subsequently revealed to lack mahX. The first of which was *Sphingobium* sp. CMETH. This *Sphingobium* strain holds very important industrial significance for two reasons. The first is the degradative rate observed is greater than that of *A. calcoaceticus* E1. The second is the ability of *Sphingobium* sp. CMETH to degrade metaldehyde below the regulatory limit within an experimental slow sand filter (Castro-Gutiérrez, 2020). The second strain isolated was *Rhodococcus globerulus* HNO-A. The *Rhodococcus* genus are aerobic, Gram-positive bacteria commonly found within the environment, that are capable of degrading large

numbers of recalcitrant and toxic organic compounds (Kim et al., 2002; Yoshimoto et al., 2004; Hara et al., 2007; Zhang et al., 2013).

Identification of novel metaldehyde degrading genes, allows for the potential for not only a greater understanding of degradation mechanisms but also can also provide industrial importance. Through further developing the understanding of the metaldehyde degradative genes within the environment, the catabolic competence of a given matrix can be optimised to ensure greater efficiency and robustness. Furthermore, identification of conserved domains within metaldehyde degrading genes could lead to the development of universal primers. Further information regarding novel protein sequences and structures can also provide insight into protein optimisation to ensure greater protein sensitivity and efficiency.

In cases where the catabolic mechanism or a catabolic gene has been identified previously for a xenobiotic compound, a similarity analysis provides a simple and powerful tool in gene identification. As high sequence similarity suggests functional similarity, programs such as BLAST (Agarwala et al., 2018) can allow the xenobiotic-degrading microorganism's genome to be appropriately searched and analysed. By using this approach, numerous novel xenobiotic-degrading genes have been identified requiring only whole genome sequence information and a similarity program (Heiss et al., 2003; Kube et al., 2013; Nanthini et al., 2017). Despite the simplicity and effectiveness of the similarity analysis, this approach requires past information regarding the mechanism of action and/or the catabolic genes, and as such will not be effective in the identification of uncharacterized or completely novel xenobiotic compound degradation. As the work conducted in Chapter 3 led to the discovery of the first identified metaldehyde degrading protein MahX, BLAST analysis has the potential to provide greater insight into the degradative pathways in microorganisms lacking MahX.

Through identifying the initial metaldehyde degrading genes within *Sphingobium* CMET-H, further insight was examined to look at horizontal gene transfer mechanisms within these genes as well as providing insight into the evolution of metaldehyde degradation. Using bioinformatics analysis and sequence data obtained from Chapter 3, this work set out to identify and characterise potential metaldehyde degrading genes in *R. globerulus* HNO-A and *Sphingobium* sp. CMET-H. Once identified, experimental analysis was conducted to verify the bioinformatic findings and to characterise the degradative proteins.

5.2 Results

5.21 *Sphingobium* sp CMET-H bioinformatics analysis

Work conducted in previous chapters led to the discovery and experimental verification of the first identified metaldehyde degrading protein MahX. Although this protein has been identified within numerous metaldehyde degrading strains isolated, its absence was noted within the strains *Sphingobium* sp. CMET-H and *R. globerulus* HNO-A. To identify potential homologues

within these strains, the MahX amino acid sequence was used with tBLASTn to query their genomes.

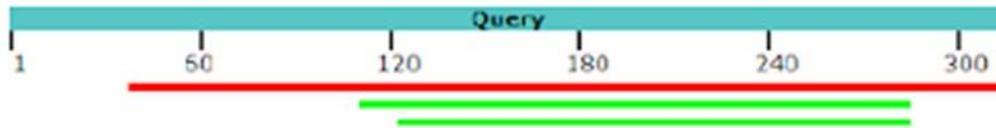


Figure 5-1: Alignment schematic generated by BLAST following the query of MahX against the *Sphingobium* sp. CMET-H genome using tBLASTN. The numbers show the amino acid length. The red colour denotes an alignment score of >200 whereas the green colour denotes an alignment score of 50 – 80.

As shown in Figure 5-1, BLAST analysis of *Sphingobium* sp. CMET-H identified three significant alignment results. The alignment sequence with the highest max score possessed a 57 % identity, a query coverage of 87 % and an e-value of $2e-108$ when queried with MahX. The identified protein, henceforth known as MahS, consists of 307 amino acid residues. Reciprocal BLAST was undertaken on the MahS to predict orthology. This involved querying the MahS protein against that of the native MahX genome *Acinetobacter calcoaceticus* E1. The results, shown in Figure 52, from this analysis identified a single hit with 57 % identity, 97 % query coverage and an E-value of $3-109$. Further analysis revealed this aligned sequence was that of MahX.

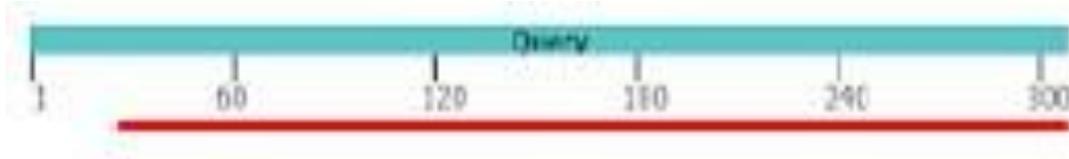


Figure 5-2: Alignment schematic generated by tBLASTn, following the querying of MahS against the genome of *A.calcoacticus* E1. The numbers show the amino acid length. The red colour denotes an alignment score of >200.

To provide further insight into the potential of MahS as a metaldehyde degrading protein, the genome of a closely related non-metaldehyde degrading strain was queried using tBLASTn. MahS alignment against the *Sphingobium chlorophenicum* NBRC 16172 genome revealed no significant matches.

As MahS demonstrated significant similarity to MahX but was not identified in the closely related non metaldehyde degrading strain *Sphingobium chlorophenicum* NBRC 16172 , further investigation into the predicted protein was undertaken. The 307 amino acid sequence of MahS was subjected to BLASTP analysis and was used to query the non-redundant protein sequence database. The highest max score alignment result obtained was that for with a phytanoyl-CoA dioxygenase family protein found within *Sphingobium japonicum*. This alignment possessed a 53% identity, 87% query coverage and an e-value of 1e-99, shown in Figure 5-3. BLAST conserved putative domain analysis also identified domains belonging to the superfamily of α KG -Fe (II) oxygenase superfamily. This result was between the residues 40 - 236 and possessed an e-value of 2.19e-04.



Figure 5-3: Alignment schematic generated by tBLASTn, following the querying of MahS against the non redundant protein database. The numbers show the amino acid length. The red colour denotes an alignment score of >200.

Further examination of the predicted protein was conducted using the protein family classification and domain predicting tool InterPro. Results of the analysis, shown in Figure 5-4, revealed that similar to MahX, MahS contains domains belonging the protein family of phytanoyl-CoA dioxygenase family (IPR008775).



Figure 5-4: Putative conserved domain analysis of MahS generated through BLAST alignment analysis.

5.21 Protein modelling for MahS

To determine the protein structure for MahS and further predict protein function, the protein model prediction software I-TASSER was used (Yang *et al.*, 2015). Initial threading alignment analysis, to identify protein templates with significant similarity was undertaken. Based on this threading alignment analysis, 5dapA was the highest ranked template. This output generated several values to demonstrate the alignment significance. The coverage value, which was calculated through dividing the aligned residues by the query length, for this alignment was 0.9. Another quality value generated was that for the Z-score which is calculated through comparing the difference between the raw and average scores in the units of standard deviation.

As such, alignments obtaining values greater than 1 are considered 'good'. MahS obtained a Zscore of 2.21 against 5dapA. The highest ranked alignment of 5dapA was revealed to encode Fe (II)/ (alpha) ketoglutarate-dependent dioxygenase AsqJ. It should be noted that this alignment was also highest ranked within the MahX analysis.

Through I-TASSER's pipeline, whereby I-TASSER generated models were analyzed through the clustering algorithm SPICKER program, five MahS protein models were generated. The protein models were ranked on their quality through numerous criteria such as C-scores, Template modelling score (T-M score) and RMSD (Root Mean Square Deviation). The protein model of highest quality, shown in Figure 5-5, obtained a C-Score of -0.41, this was calculated based on the significance of the threading template alignments and the convergence parameters of the model simulations. Other measurements of quality were also available such as RMSD and the T-M score. Both the RMSD and T-M score allow for the comparisons of model alignments to be quantified through superimposing structural models and calculating the divergence of equivalent residues. The RMSD value is one of the most popular measures for structural similarity and provides important comparative information. However, this calculates all residue distances with equal weight and is therefore prone to significant local structural deviation. The T-M score attempts to overcome this through counting all residues pairs using the Levitt-Gerstein weight and therefore can potentially provide a score that is more sensitive to the global topology. The RMSD for this model obtained a score of $7.1 \pm 4.2 \text{ \AA}$. For the T-M score, the values are calculated between 0-1 with 1 being an identical match- the model obtained a score of 0.66 ± 0.13

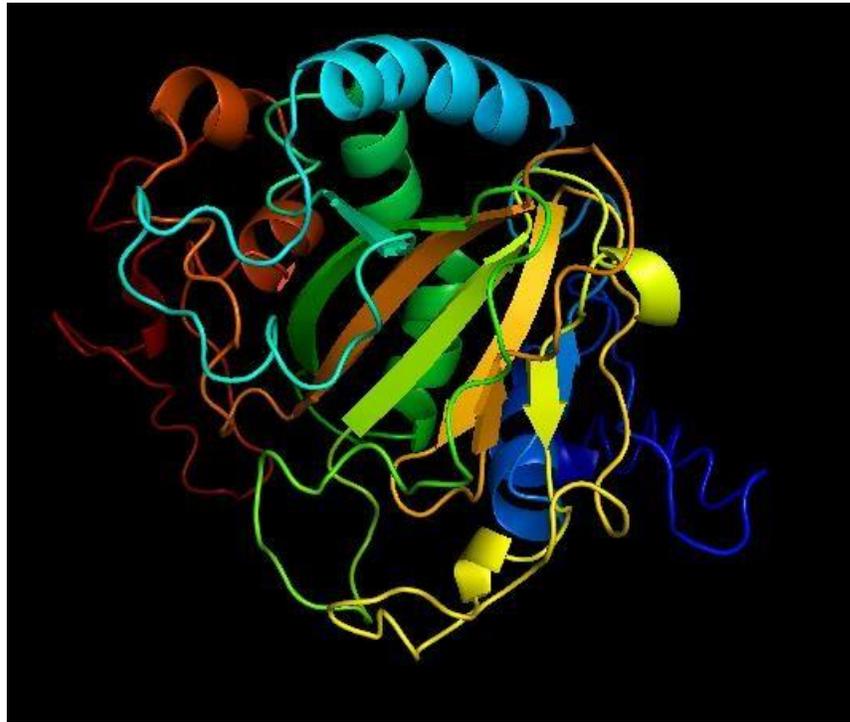


Figure 5-5: I-TASSER generated model for MahS. Image taken through PyMol

Following the creation of a suitable protein model for MahS, the PDB was queried using the T-M structural alignment program. This program generated a list of 10 structural analogs to the MahS model, shown in Table 5-1. Based on structural alignment similarities, the highest ranked alignment was that of 5dapA.

Table 5-1: I-TASSER generated similarity-based results based on querying PDB using T-M structural alignment program with the MahS model. T-M Score is a metric for assessing the similarity of protein structures. Identity is the % sequence identity in the structurally aligned region. Coverage denotes the alignment by T-M align and is equal to the number of structurally aligned residues divided by length of the query protein.

Rank	PDB ID	PDB Hit	T-M Score	Identity	Coverage
1	5dapA	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AsqJ	0.88	0.18	0.90
2	5m0tA	Alpha-ketoglutarate-dependent non-heme iron oxygenase EasH	0.84	0.17	0.90
3	4y5sA	Verruculogen synthase with α KG as co-substrate	0.84	0.18	0.90
4	4naoA	Crystal structure of EasH	0.82	0.13	0.87
5	5zm2A	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AndA	0.81	0.17	0.84
6	5yblA	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AusE	0.79	0.16	0.85
7	6akzA	Crystal structure of GlcNAc Inducible Gene 2, GIG2 (DUF1479) from <i>Candida albicans</i>	0.70	0.11	0.86
8	6s0rA	Kanamycin B dioxygenase (KanJ) from <i>Streptomyces kanamyceticus</i>	0.69	0.18	0.80
9	2csgA	Putative oxidoreductase from <i>Salmonella typhimurium</i> LT2	0.68	0.1	0.84
10	4xbzA	Crystal Structure of EvdO1 from <i>Micromonospora carbonacea</i> var. <i>aurantiaca</i>	0.67	0.15	0.81

Biological annotations of the model for MahS was undertaken using the COFACTOR and COACH programs. COFACTOR allows protein functions (ligand binding sites, EC and GO) through structural comparisons. Through this annotation program, two significant ligand-binding sites were identified. The first ligand site, based on the PDB hit with a non-heme alpha ketoglutarate dependent epimerase SnoN from nogalamycin biosynthesis (5ep9A), was for alpha ketoglutarate (α KG). The confidence score for this prediction, where the range is from 0-1 and the higher the score the more reliable the prediction was 0.45. COFACTOR works through TM align searching and therefore the cluster value reveals the number of templates within this cluster. For α KG, the MahS obtained a cluster score of 62. The ligand bind site residues identified, shown in Figure 5-6, were 94,136,138,145,148,150,166,168,181,218,224,226,235 and 239.



Figure 5-6: I-TASSER generated cartoon ribbon model of MahS. Identified α KG ligand binding sites are highlighted yellow. α KG is highlighted green. Image taken on PyMol.

The second ranked ligand-binding site was for Fe (II) with a C-score of 0.22 and cluster size of 46. The PDB hit for this ligand-binding site was for 2g19A , which encodes hypoxia-inducible factor prolyl hydroxylase (PHD2). Three ligand binding site residues predicted, shown in Figure 5-7, were H148, D150 and H 224.

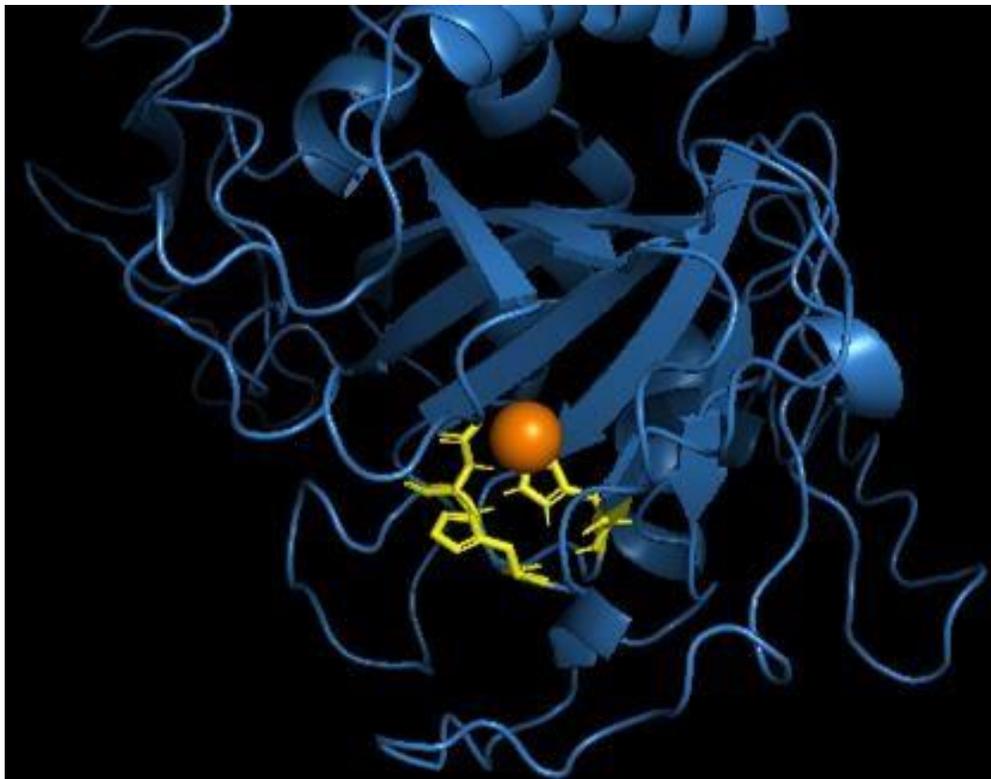


Figure 5-7: I-TASSER generated cartoon ribbon model for MahS. Fe²⁺ ligand binding sites are highlighted yellow. Fe²⁺ is represented as an orange sphere. Image taken on PyMol.

The surface structure of the model was simulated to allow for a visual representation of the protein structure and ligand binding interactions within the model. Visual examination of the surface model, shown in Figure 5-10, revealed a 'pocket' within which the ligand binding sites reside. Visual examination of the pocket using Pymol identified the following residues present: T85, H86, M88, E89, S13, L145, P146, H148, V152, M153, W218, Q250.

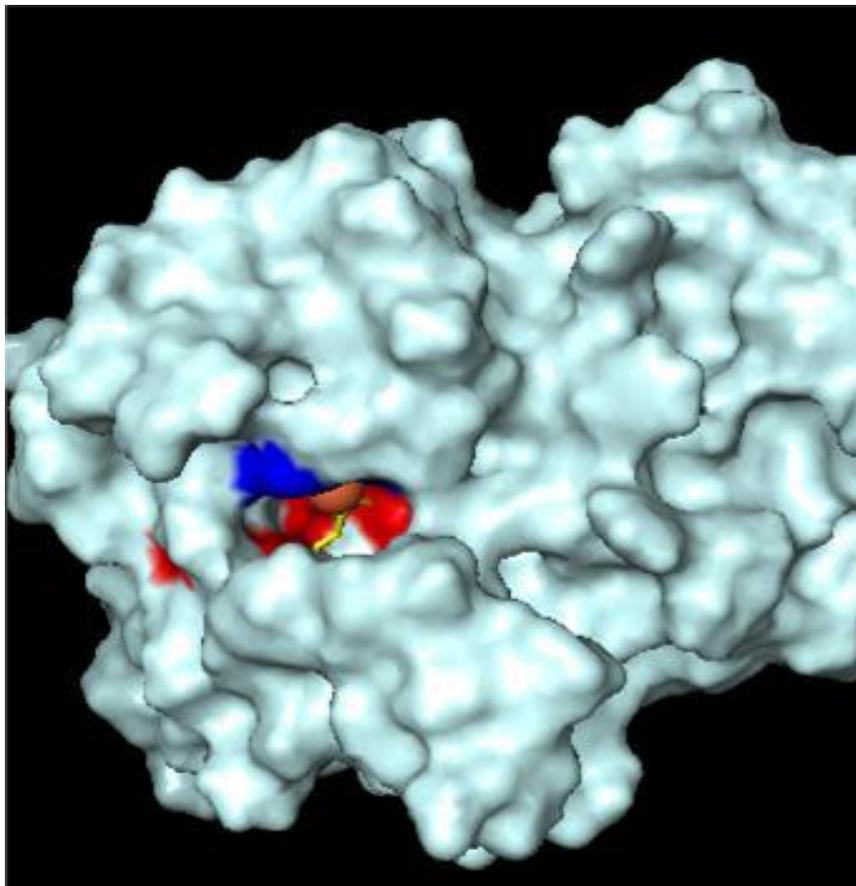


Figure 5-85: Surface area of I-TASSER generated model for MahS. Image was generated through PyMol. Orange sphere represents Fe²⁺, aKG structure is highlighted yellow. Ferrous binding residues are highlighted blue, AKG binding residues are highlighted red .

5.22 Genomic location of *mahS* within *Sphingobium sp* CMET-H

As the genomic location and nearby genes can provide information regarding both the potential metabolic pathway but also the mobility of the gene, the location of *mahS* inside *Sphingobium sp.* CMET-H's genome was examined. Using Sanger's Artemis genome for the analysis, *mahS* was shown to be located on a 5876 bp contig. Manual examination of the contig, shown in Figure 5-9, revealed that several genes were clustered in the same orientation near *mahS*.

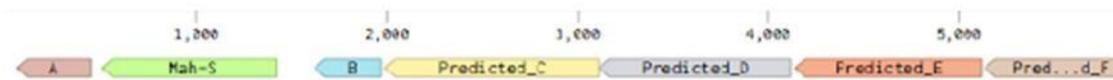


Figure 5-9: Schematic illustration of the 5876 bp contig containing *MahS* with *Spingobium* CMET-H. Predicted genes A,B,C,E and F and their orientation are depicted. Image generated using benchling.

InterPro analysis was undertaken on all the predicted proteins of the contig, to potentially provide insight into the metalehyde degradation pathway and provide greater insight into the mobility of the contig. The results revealed that Predicted_A has snoal-like domains present and a member of the NT2-like domain superfamily. This superfamily has several overlapping entries which include, but are not limited to, conjugal transfer proteins, bacterial virulence proteins and limonene-1,2-epoxide hydrolase (Sultana *et al.*, 2004). Predicted_B analysis revealed carboxymuconolactone decarboxylase-like domains present. Protein examination also identified Predicted_B belonging to the AhpD-like homologous superfamily. This family has been demonstrated to be involved in protocatechuate catabolism (Clarke *et al.*, 2011). Predicted_C demonstrates a domain relating to an alpha/beta hydrolase fold-1 on the C terminus. As such, this sequence was identified as belonging to the homologous superfamily of alpha/beta hydrolase fold. These folds are common in hydrolytic enzymes and have overlapping entries with

carboxylesterases and thioesterases (Holmquist, 2005). Examination of Predicted_D revealed transketolase-like, pyrimidine-binding domain and belonging to the superfamily of thiamin diphosphate-binding fold. Proteins in this family has several overlapping entries ranging from pyruvate flavodoxin/ferredoxin oxidoreductase to dehydrogenases (Chen *et al.*, 2018). Analysis revealed that Predicted_E belongs to the superfamily of thiamin diphosphate-binding fold domains. These are typically found in a number of dehydrogenases such as pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and 2-oxoisovalerate dehydrogenase. The predicted protein also possesses dehydrogenase E1 component domain (Chen *et al.*, 2018). Results for Predicted_F shows that it belongs to the family of AcoX-like. This family is predicted to be kinases with members being found in gene clusters responsible for acetoin catabolism in certain bacteria (Min Huang *et al.*, 1994).

Further characterization of the contig involved the querying of the 5876 bp contig against the nucleotide collection database using BLASTn. The results, shown in Figure 5-10, identified significant alignments related to *Paracoccus* plasmids, with the five highest scoring results relating to these sequences. The highest max score result returned was for *Paracoccus pantotrophus* strain DMS 2944 plasmid pPAN2. This result demonstrated a 77% identity, a query coverage of 54% and an E-value of 0.0. The sequences Predicted_A, Predicted_B and MahS are located upstream from this high scoring plasmid alignment sequence and therefore may form a catabolic cluster.

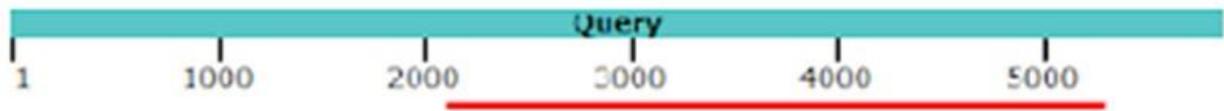


Figure 5-10: Schematic generated with BLAST through querying the 5876 bp contig through the non-redundant nucleotide sequences database. The number shows the bp positions and the red colour denotes an alignment score of >200.

As the BLAST result of the contig showed potential for *mahS* to be located within a plasmid, further examination of *Sphingobium* sp. CMET-H's genome was undertaken. plasmidSPAdes is a software tool used in the extraction and assembly of plasmid data from whole genome sequences (Antipov *et al.*, 2016). By using the read coverage of contigs to assist in distinguishing between chromosomal and plasmid DNA, *Sphingobium* sp. CMET-H's genomic data was examined. The results of the analysis revealed four predicted plasmids were identified. The plasmid of greatest interest to this investigation, which contains *mahS*, was identified over seven nodes and totalled 193,696 bp in size. Examination of these seven nodes was undertaken, to ascertain properties of the plasmid, using the online tool PATRIC (Davis *et al.*, 2020). Visual analysis revealed of the 200 proteins predicted, 17 were directly suggested to function in conjugative transfer pilus assembly. Further examination revealed that the predicted plasmid also contains a conjugative relaxosome accessory protein. Typically, conjugative plasmids consist of four required components; origin of transfer site (*oriT*), relaxase gene, a gene encoding a type 4 coupling protein (T4CP) and gene cluster for bacterial type IV secretion system (T4SS). The online tool *oriT* finder was undertaken to identify and verify the findings of the previous analysis

(Li *et al.*, 2018). Results of the query, shown in Figure 5-11 identified locations and presence of both T4SS and T4CP. OriT regions were not identified in this analysis.

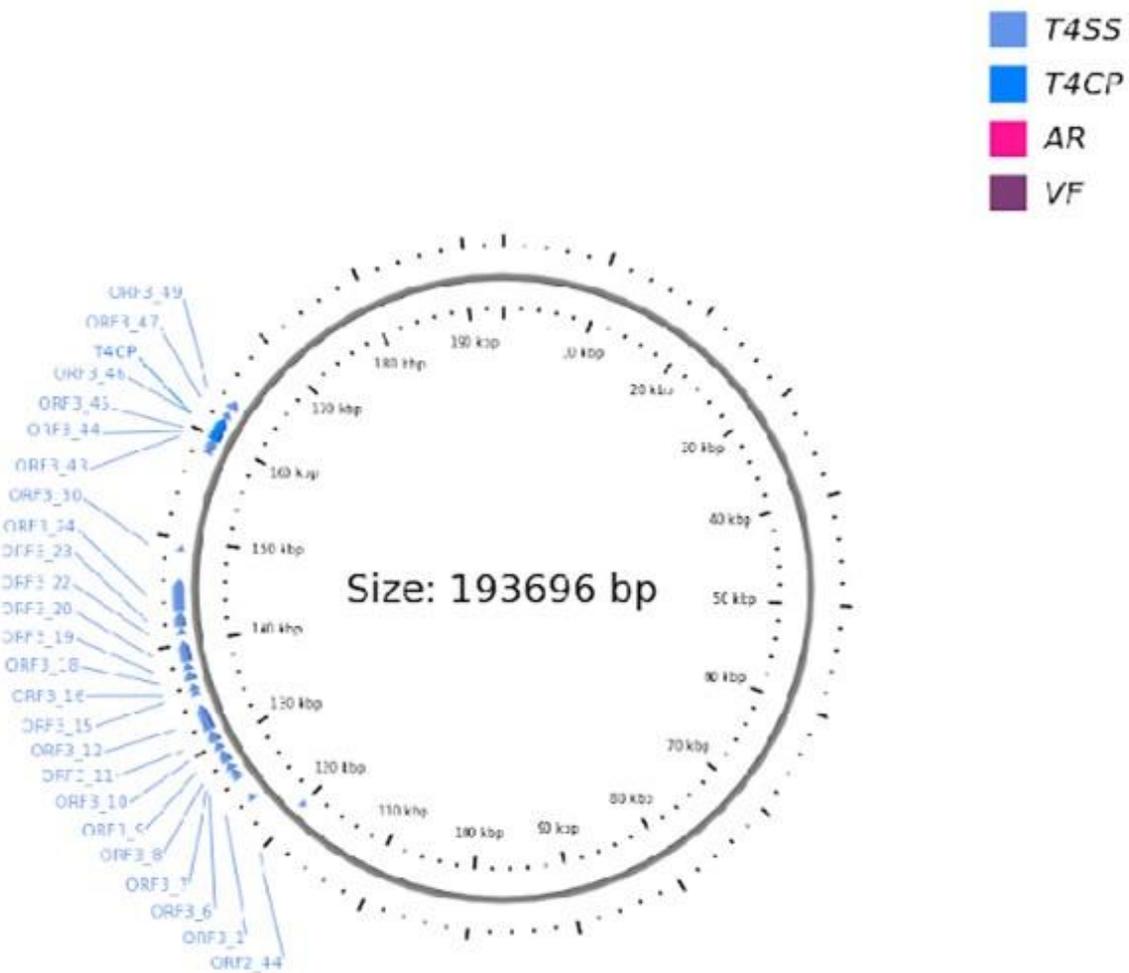


Figure 5-11: Graphic depiction of Type 4 secretion system and Type 4 coupling protein. Graphic constructed using oriT. Image displayed is a collection of identified contigs which are not assembled.

5.221 Examination for the presence of MahY homologue

As MahX was discovered in a potential operon upstream from the predicted protein of MahY, examination of the *Sphingobium* sp. CMET-H's genome was undertaken to identify potential similar sequences. BLAST analysis of MahY resulted in no significant alignment matches.

5.23 *R. globerulus* HNO-A bioinformatics analysis

To identify potential metaldehyde degrading protein sequences within *R. globerulus* HNO-A, MahX was used to query its genome. The alignment analysis identified a single result henceforth known as MahR. This predicted protein consists of 303 amino acid residues in length, as shown in Figure 5-12. Alignment of MahR to MahX reveals a 32% identity, a query coverage of 44% and an e value of $7e-15$.

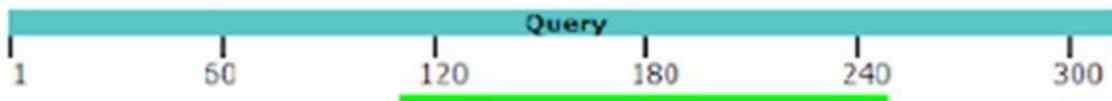


Figure 5-12: Graphic depicting the tBLASTn analysis results of MahX (the query) against the whole genome of *R. globerulus* HNO-A. The numbers show the amino acid length and the green colour denotes an alignment score of 50-80.

Reciprocal tBLASTn analysis of MahR against *A. calcoaceticus* E1 revealed a single significant alignment result. The result showed a 32% identity, a 46% query coverage and an e-

value of $9e-11$, as shown in Figure 5-13. Further analysis revealed that the aligned sequence was that of MahX.

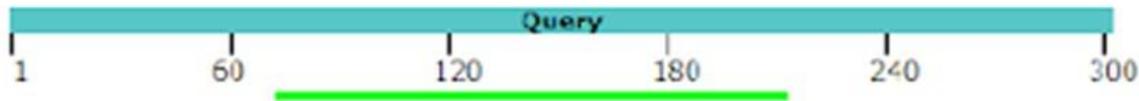


Figure 5-13: Graphic depicting the tBLASTn analysis results of MahR (the query) against the whole genome of *A. calcoaceticus* E1. The numbers show the amino acid length and the green colour denotes an alignment score of 50-80.

In order to determine whether the MahR is a likely metaldehyde degrading protein, analysis was undertaken to determine whether it was present in a closely related, nonmetaldehyde degrading type strain *Rhodococcus globerulus* NBRC 14531. Upon querying *R. globerulus* NBRC 14531's genome, the analysis revealed no significant similarities were found.

To identify similar sequences and predict potential function of MahR, the protein was queried against the non-redundant protein sequence database using BLASTP. The highest scoring result was for a phytanoyl-CoA dioxygenase family protein found in *Mycolicibacterium* sp. P9-22. This result, as shown in Figure 5-14, gave a 71% identity, 98% query coverage and an e-value of $1e-154$. Putative domains belonging to the 2-oxoglutarate (2OG) and Fe (II)-dependent oxygenase superfamily were also detected in the analysis, as shown in Figure 5-15.

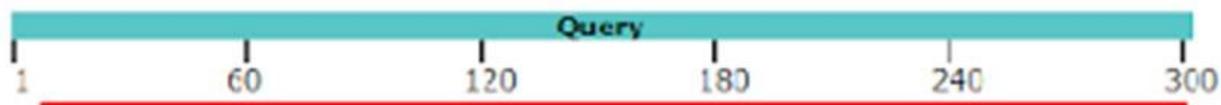


Figure 5-14: Graphic depicting the BLASTp analysis results of MahR (the query) against the non redundant protein sequences database. The numbers show the amino acid length and the red colour denotes an alignment score of >200.

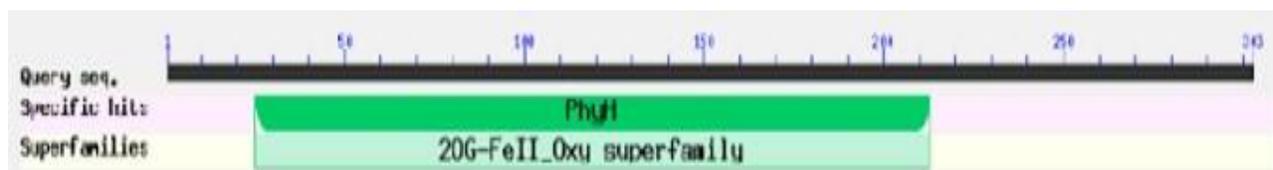


Figure 5-15 Putative conserved domain analysis of MahR generated through BLAST alignment analysis.

5.24 Protein modelling for MahR

To further investigate the potential function of MahR, a model was generated using the online tool I-TASSER. As discussed previously, I-TASSER use a threading-based approach to generate models using structural information obtained from templates within the PDB. Based on this analysis, 10 threading templates which were predicted to possess significant alignment to MahR were obtained. The PDB ranked highest was for 4y5sA. This database file encodes for verruculogen synthase with α KG as co-substrate. The threading alignment between MahR and 4y5sA generated a coverage score of 0.93 and a Z-score of 2.42.

Through I-TASSER, the SPICKER program was utilised to cluster and sort the I-TASSER generated MahX protein models. Following this, five models were generated along with

confidence scores for each model. The highest ranked model, shown in Figure 5-16, displayed a C-score of -0.07, a T-M score of 0.70 ± 0.12 , and an estimated RMSD of $6.4 \pm 3.9 \text{ \AA}$.

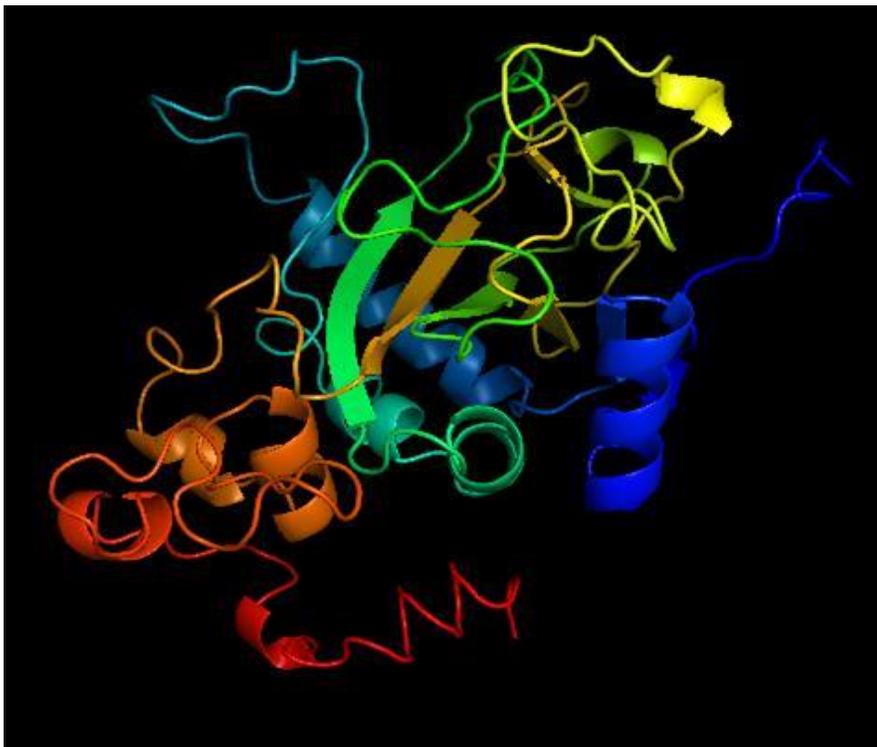


Figure 5-16: I-TASSER generated cartoon ribbon model of MahR. Image taken on PyMol.

The highest ranked model was then used to query the PDB to identify similar proteins. The highest ranked PDB hit was for 4y5sA, as shown in Table 5-2.

Table 5-2 I-TASSER generated similarity-based results based on querying PDB using T-M structural alignment program with the MahR model. T-M Score is a metric for assessing the similarity of protein structures. Identity is the % sequence identity in the structurally aligned region. Coverage denotes the alignment by T-M align and is equal to the number of structurally aligned residues divided by length of the query protein.

Rank	PDB ID	PDB Hit	T-M Score	Identity	Coverage
1	4y5sA	Verruculogen synthase with α KG as co-substrate	0.92	0.17	0.93
2	5dapA	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AsqJ	0.85	0.21	0.91
3	5m0tA	Alpha-ketoglutarate-dependent non-heme iron oxygenase EasH	0.83	0.21	0.90
4	5zm2A	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AndA	0.83	0.22	0.87
5	4naoA	Crystal structure of EasH	0.82	0.18	0.90
6	5yblA	Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AusE	0.79	0.25	0.86
7	6akzA	Crystal structure of GlcNAc Inducible Gene 2, GIG2 (DUF1479) from <i>Candida albicans</i>	0.72	0.12	0.88
8	4xbzA	Crystal structure of EvdO1 from <i>Micromonospora carbonacea</i> var. <i>aurantiaca</i>	0.69	0.14	0.82
9	4rgkA	Putative phytanoyl-CoA dioxygenase family protein YbiU from <i>Yersinia pestis</i>	0.68	0.12	0.83
10	2csgA	Crystal Structure of the putative oxidoreductase from <i>Salmonella typhimurium</i> LT2	0.68	0.11	0.83

Using the predicted structure both COFACTOR and COACH analysis was undertaken. The top two ligand-binding sites were for α KG and Fe^{3+} . The AKG hit possessed a C-score of 0.52 and cluster value of 64. The PDB hit for this binding site was for 5dapA, encoding Fe (II)/ (alpha)

ketoglutarate-dependent dioxygenase AsqJ. A model of the α KG binding sites was generated and is shown in Figure 5-17.

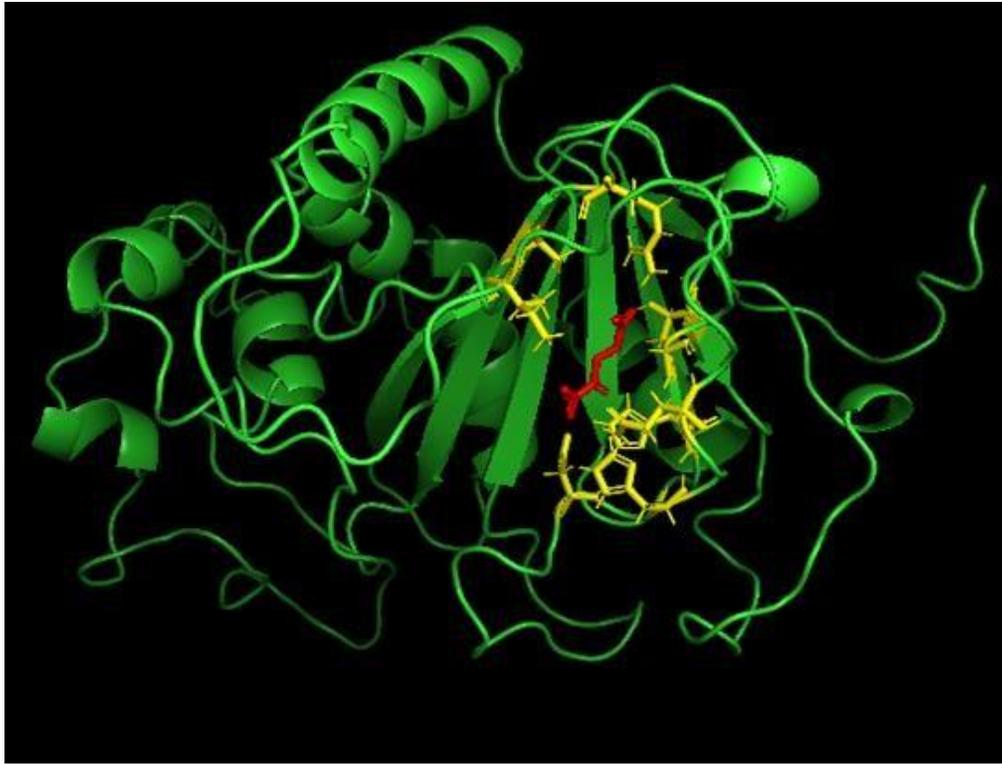


Figure5-17:I-TASSER generated model of MahR. α KG ligand binding sites are highlighted yellow. α KG is highlighted red.

The second binding site for Fe^{3+} and possessed a C-score of 0.26 and a cluster size of 40.

The PDB hit for this result was 3kT4A that encodes a component of the messenger ribonucleoprotein complex. Three ligand binding site residues were identified in this analysis H124, D126 and H200 and are shown in Figure 5-18.

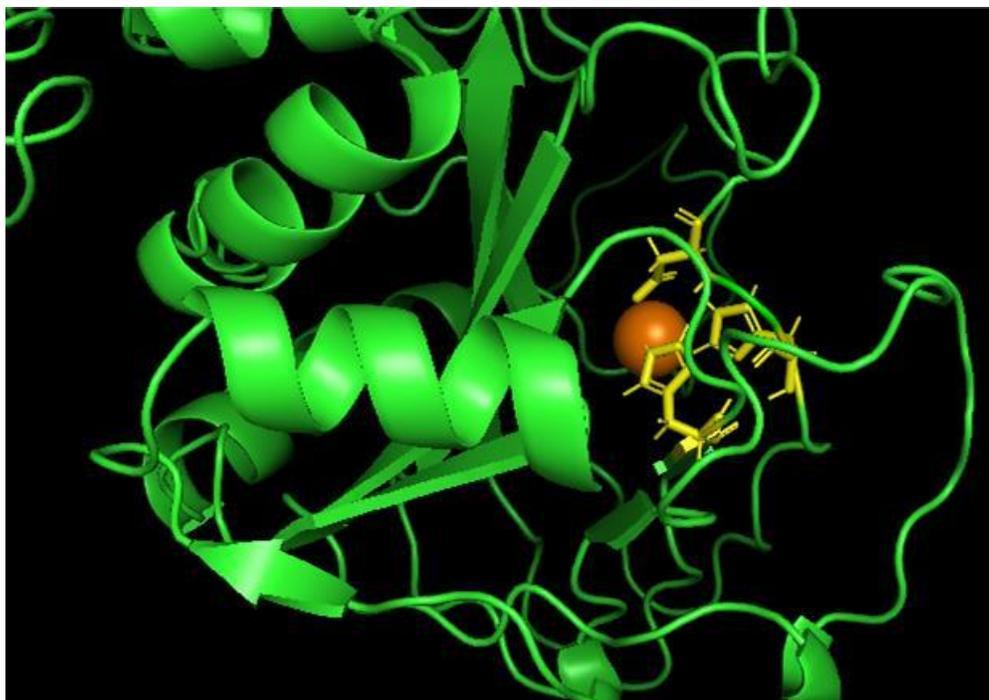


Figure 5-18:Cartoon ribbon model of MahR generated through I-TASSER. Fe³⁺ ligand binding sites are highlighted yellow. Fe³⁺ is depicted as an orange sphere.

The surface structure of the MahR model was simulated within molecular visualization system PyMol to allow for a visual representation of the protein structure and ligand binding interactions within the model. Visual examination of the surface model, shown in Figure 5-19, revealed a 'pocket' within which the ligand binding sites reside.

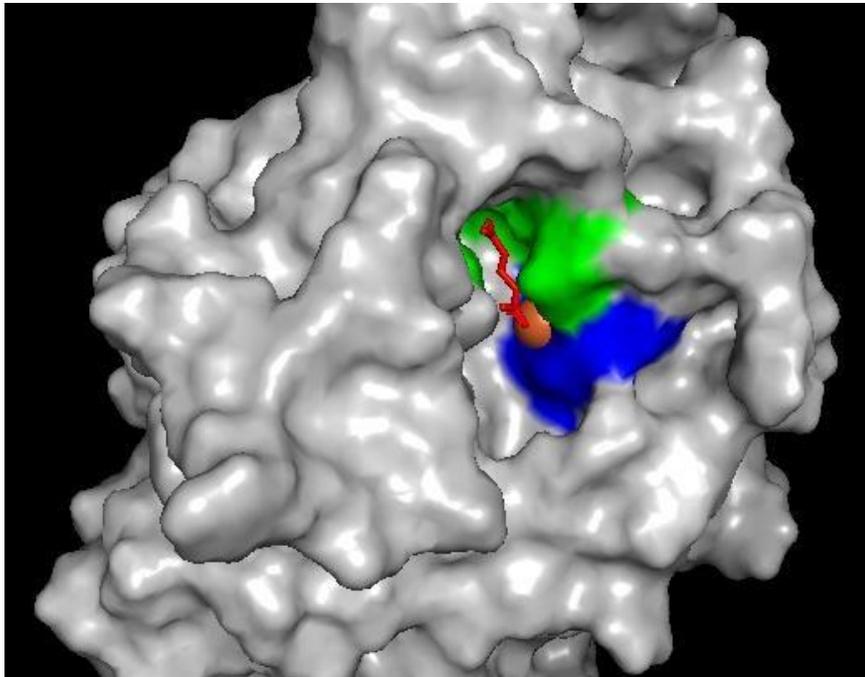


Figure 5-19: Surface fill model of MahS with ligand sites for aKG and Fe²⁺ highlighted in green and blue, respectively. Image taken through PyMol.

uper5.25 Genomic location of mahR within *R. globerulus* HNO-A's

Through utilisation of the genome browsing program Artemis, the genomic location of *mahR* was investigated. Analysis revealed that the 909 bp gene resided on a 221,240 bp contig. BLAST analysis was conducted on sequences 4500 bp up and downstream, to provide potential

insight into any catabolic gene clusters present. From this analysis, a schematic was designed to reveal nearby predicted genes. This is shown in Figure 5-20.

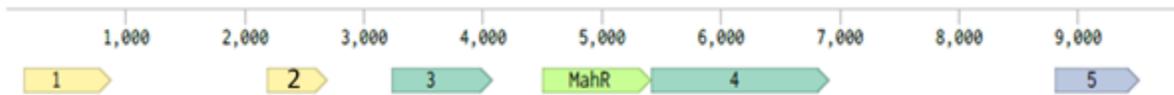


Figure 5-20 :Schematic of 9900bp region within genome of *R.globerulus* demonstrating 4500bp up and downstream from mahR. Five predicted genes were identified alongside mahR generated through Sanger's Artemis. Image generated through Benchling.

Predicted protein 1 demonstrated a significant alignment against alpha/beta hydrolase [*Rhodococcus* sp. MS16]. This generated an identity score of 100%, a query coverage of 99% and an e-value of $8e-173$. This family of protein have demonstrated a diverse catalytic functions that include epoxide hydrolase, hydroxynitrile lyase and dienelactone hydrolase, amongst others. Based on the broad range of predicted functions regarding this protein, its role within metaldehyde degradation could not be determined (Zheng *et al.*, 2016).

Predicted protein 2 was identified as a hypothetical protein EOP31_07365 [*Rhodococcus* sp.] The analysis revealed a 99% identity, 99% query and an e-value of $4e-111$. Based on the limited data available, the function of this protein is unknown.

Analysis of predicted protein 3 failed to identify an alignment of significance. As such, it is not possible to reach a judgement regarding the function of this predicted sequence.

Predicted protein 4 demonstrated significant alignment against APC family permease [*Microbacterium trichothecenolyticum*]. This match generated an identity score of 62%, query coverage of 92% and an E-value of 0.0. Proteins within this family are responsible for the transportation of amino acids, peptides and inorganic cations. Based on this analysis, this protein does not appear to have a role in metaldehyde degradation (Jack *et al.*, 2000).

Predicted protein 5 demonstrated significant alignment against type 1 glutamine amidotransferase domain-containing protein [*Rhodococcus* sp. WS3]. This score generated an identity score of 100%, a query coverage of 99% and an E-value of 1e-168. This domain is associated with the removal of ammonia groups from glutamine and transference onto a new substrate (Korolev *et al.*, 2002). As such, this protein does not appear to have a role within metaldehyde degradation.

BLAST analysis of the 9909 bp region failed to identify high alignment against plasmid regions. Analysis of *R. globerulus* HNO-A's entire genome through plasmidSPAdes also failed to identify the presence of a plasmid.

5.251 Investigation into MahY homologue

As predicted protein MahY was found to be present downstream from MahX in *A. calcoaceticus* E1, examination of the *R. globerulus* HNO-A genome was undertaken to identify potential similar sequences. BLAST analysis of MahY resulted in no significant alignment matches.

5.26 Phylogenetic analysis of MahS and MahR

Following the discovery of two potential metaldehyde degrading enzymes, a phylogenetic analysis was undertaken. BLASTP was conducted for MahS, MahR and MahX, through querying each sequence through the non-redundant protein sequence database. For each sample, 10 highest ranked sequences were aligned using the multiple sequence alignment tool MUSCLE. Phylogenetic tree construct was performed using MEGA with the maximum likelihood method and shown in Figure 5-21. Phylogenetic analysis demonstrates the close relationship between MahS and Mah-X, relative to Mah-R. The sequence Mah-R appears to have a closer relationship to the dioxygenases found in Mycobacterium.

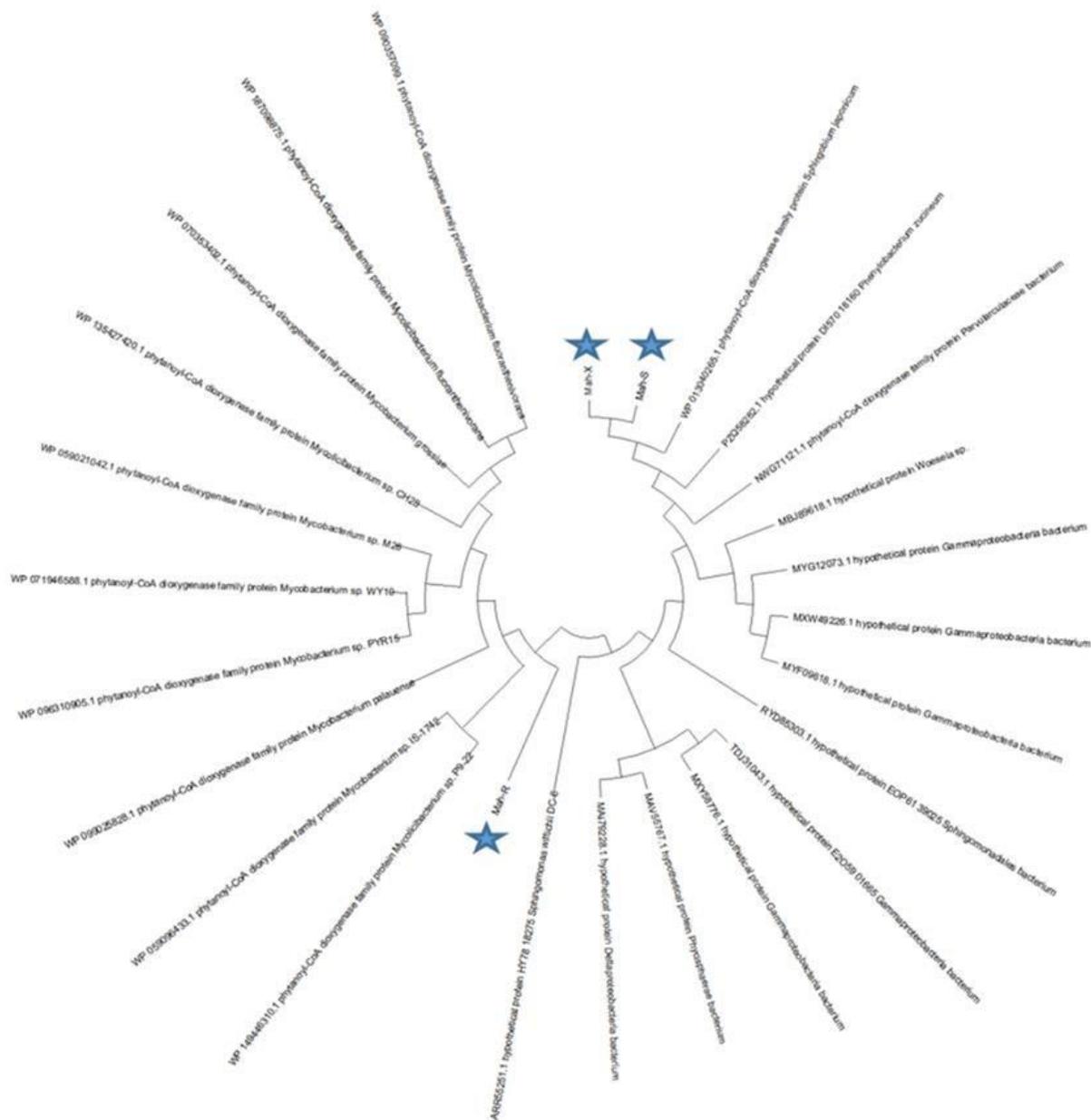


Figure5-21: Phylogenetic tree showing highest aligned sequences from BLAST analysis of the protein sequences against non-redundant database. The program MUSCLE was used to generate the alignment and the phylogenetic tree was performed using MEGA with the maximum likelihood method.

5.262 Multiple Sequence analysis of BLAST identified sequences

The evolutionary relationship between the identified protein sequences and the BLAST sequences identified previously were further analyzed through visual examination of ESPript generated multiple sequence alignment results (Robert and Gouet, 2014). As the amino acid residues within the ligand binding pockets for MahX and MahS were identified previously within section 4.21 and 5.21, the conservation of these regions was investigated. Of the 12 identified residues within the MahX binding pocket, three residues were identified as being conserved within both MahS and MahX. In relation to MahX, these amino acids were threonine 96, serine 145 and leucine 156. Relating these conservation residues to the previously identified MahR revealed the protein sequence possessed none of the three of the conserved regions. However, numerous other regions throughout the protein sequence were identified as being conserved. Further comparative analysis between MahX and MahS revealed numerous conserved regions identified between both predicted protein sequences, as shown in figure 5-22.

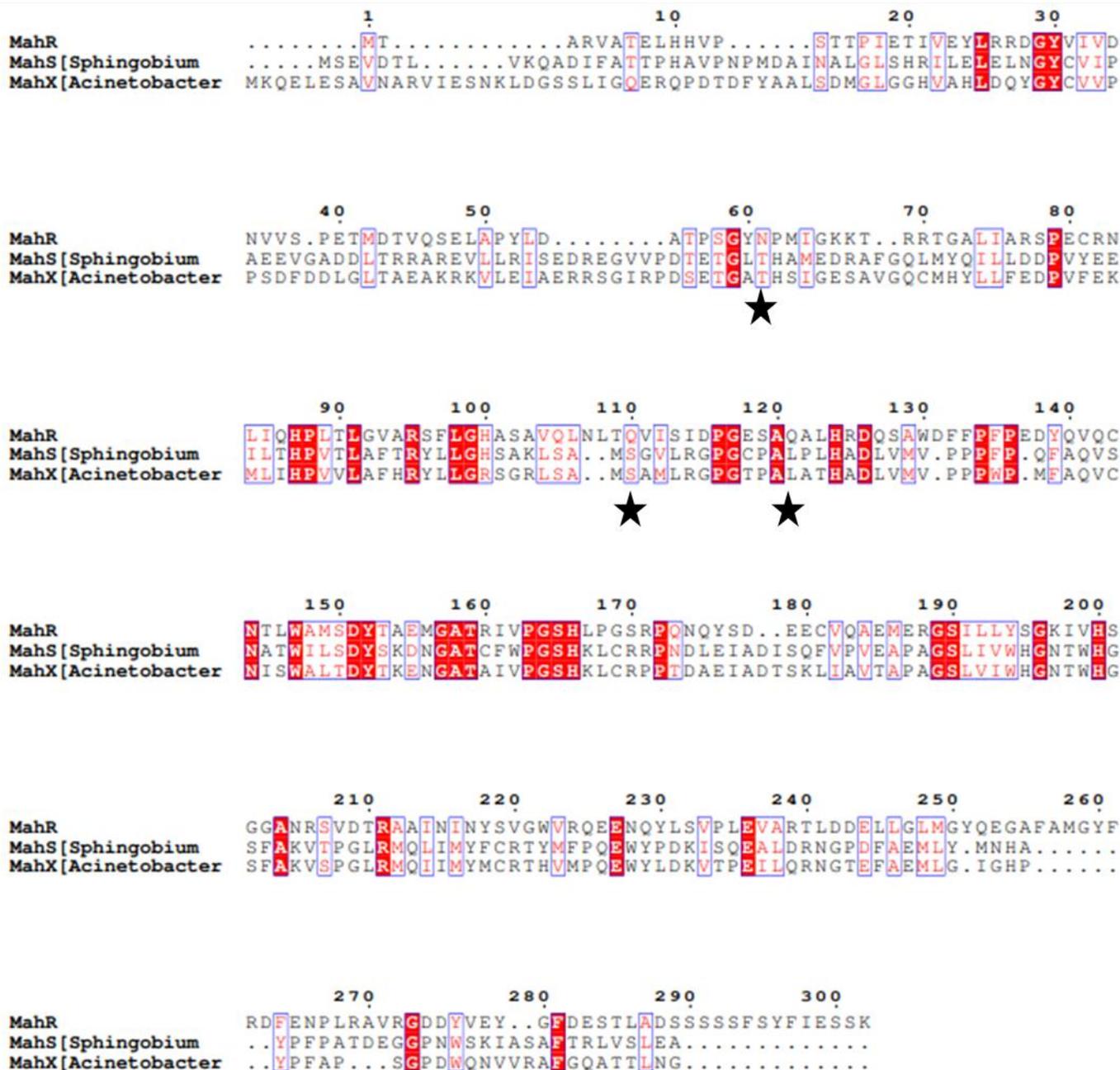


Figure 5-22: Multiple sequence alignment of MahX, MahS and MahR. Residues were coloured based on sequence identity. Stars demonstrate regions considered to be important for metaldehyde binding. Analysis performed on ClustalOmega with visualisation performed using ESPript (Robert and Gouet, 2014).

5.27 Experimental verification of the bioinformatics analysis

As the bioinformatics analysis showed potential for MahS and MahR to be responsible for metaldehyde degradation in their respective organisms, experimental verification of these findings were undertaken. Similar to the protein expression analysis conducted with MahX within Chapter 4, both *mahS* and *mahR* were cloned into the expression vector pETFPP_2. This vector allowed for both proteins to be overexpressed, under an IPTG inducible promoter, fused to both MBP and hexa HIS tag on the N-terminal. These fusions therefore provided not only increased solubility but also multiple potential protein purification steps for downstream analysis.

The genes *mahS* and *mahR* were colony PCR amplified, using a high-fidelity polymerase. Confirmation of gene amplification was performed using a 1.2 % agarose gel and shown in Figure 5-23 and Figure 5-24. Constructs were ligated using NEBuilder master mix and transformed into competent *E. coli* DH5 α . No ligation and no insert controls identified 0 colonies, when plated. Conditions with vector, ligase mixture and insert generated >150 colonies for both constructs.

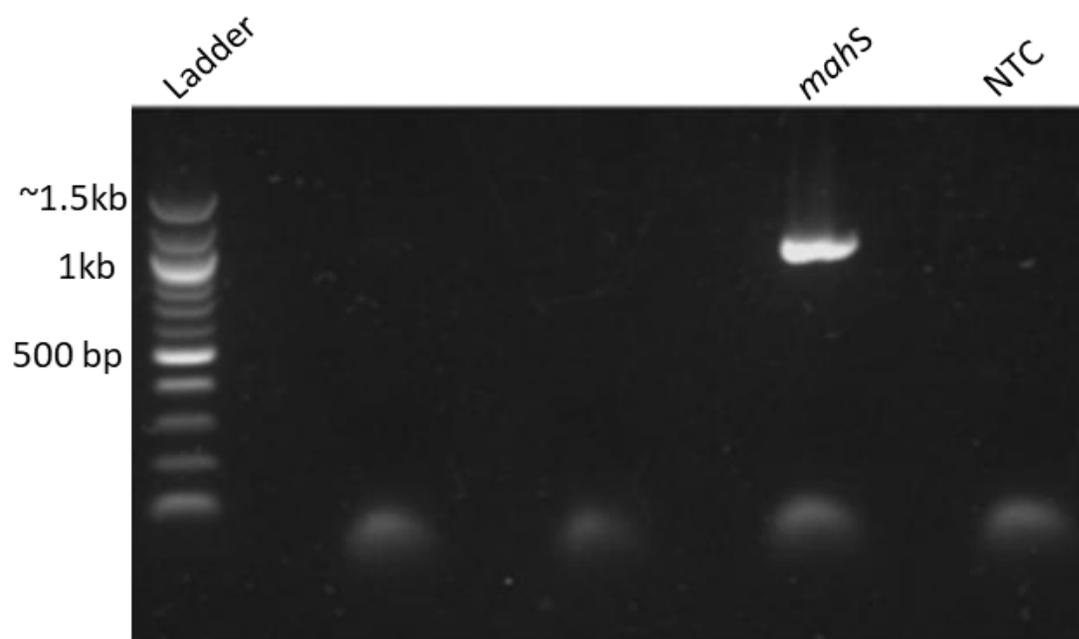


Figure 5-23: 1.2% agarose gel following mahS PCR amplification. A no template control was also performed. NEB 100 bp quickload DNA ladder used to estimate DNA size.

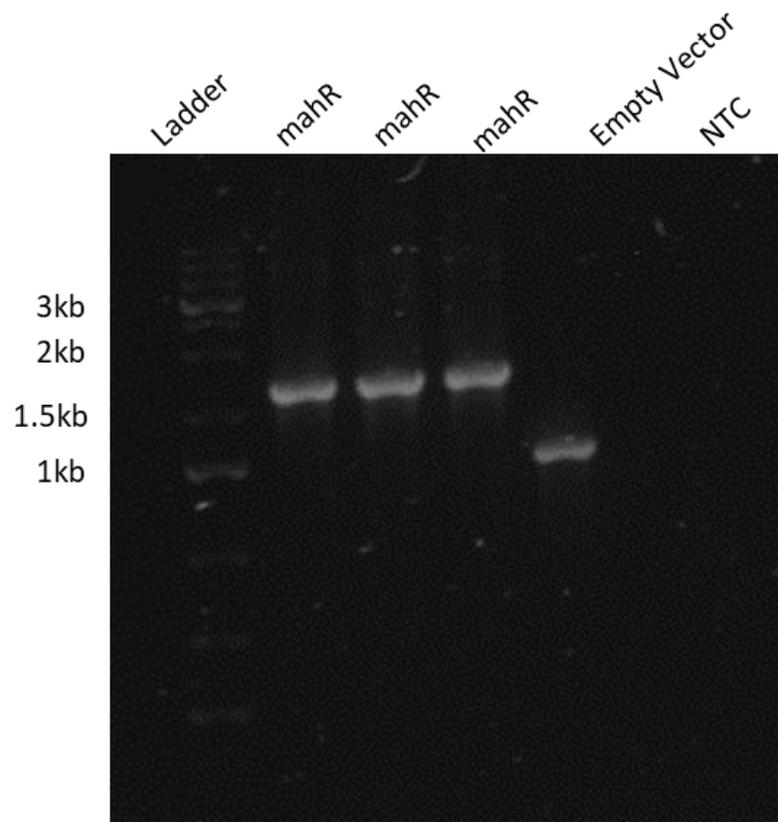


Figure5-24: 1.2% agarose gel following PCR amplification of mahR . An empty vector control and no template control was used to further demonstrate correct amplification. 1kb promega DNA was used to estimate DNA size.

To determine whether constructs were expressing the correct fusion protein of the expected size, protein expression trials were conducted. For MahS and MahR, the predicted protein fusion was expected to be ~78 kDa and ~76 kDa, respectively. Samples of different induction time points (1 hour and 3 hours relative to non-induced) were performed to both

demonstrate and optimize protein expression. Protein analysis, shown in Figure 5-25 and Figure 5-26, revealed the presence of noticeable bands in both the 1 hour and 3-hour post induction samples.

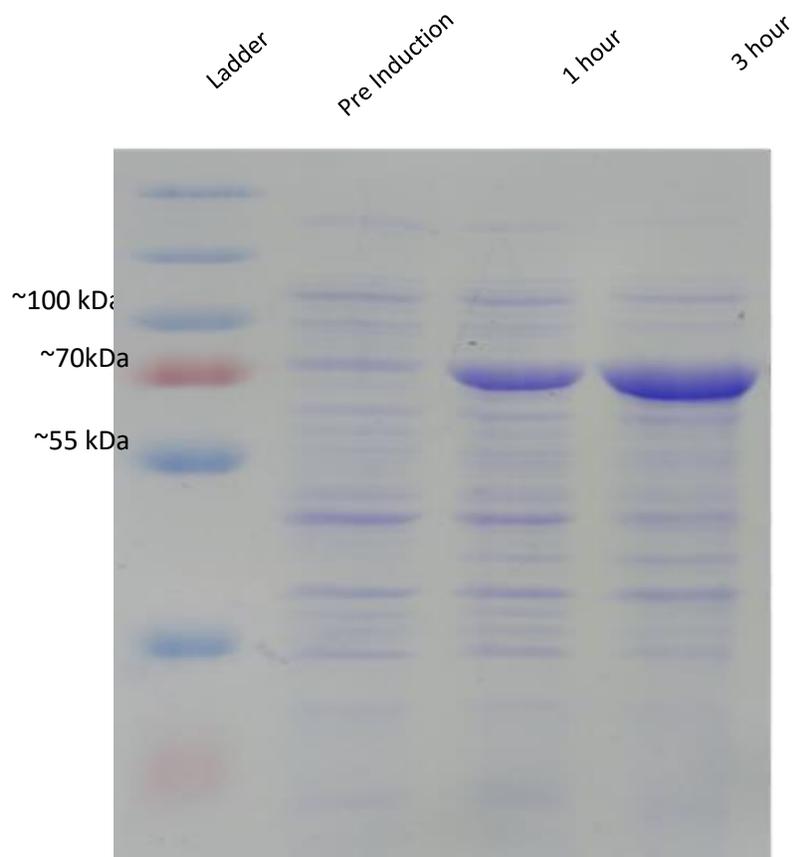


Figure 5-25:10% SDS PAGE gel with pre IPTG induced, 1 hour post induction and 3 hour post induction of IPTG induced MBP-MahS expressing E. coli BL21. PAGERuler was used to determine approximate size. 15 μ L of sample was loaded to each well.

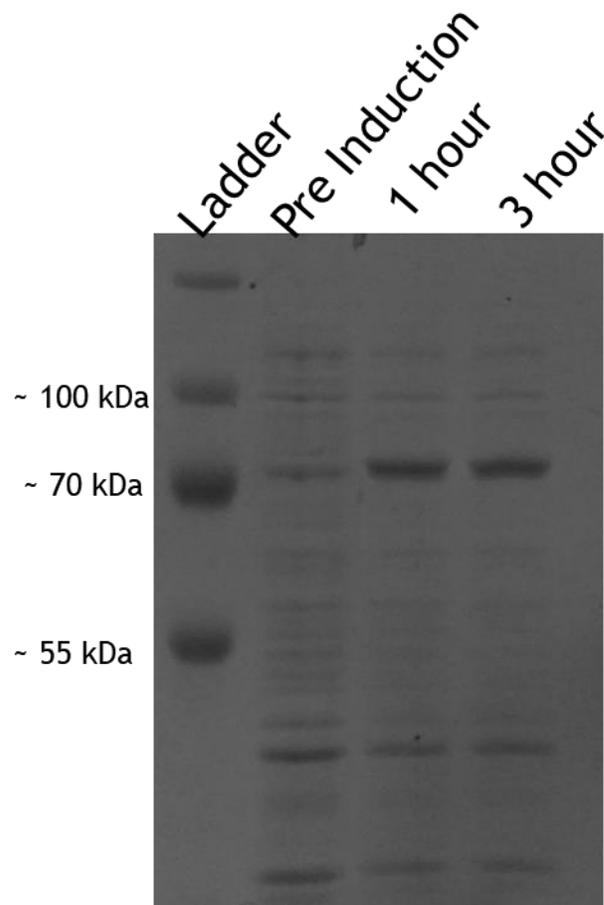


Figure 5-26:10% SDS PAGE gel with pre IPTG induced, 1 hour post induction and 3 hour post induction of IPTG induced MBP-MahR expressing *E. coli* BL21. PAGERuler was used to determine approximate size. To each well, 15 μ L of protein sample was added.

As both MBP-MahS and MBP-MahR displayed observable expression upon induction, an *in vivo* metaldehyde degradation assay was undertaken. Following 3 hour incubation post IPTG induction, induced *E. coli* BL21 expressing MBP-MahS, MBP-Mah-R or a no insert control was incubated with 10 mL of MSM supplemented with 15 mg/L metaldehyde. As shown in Figure 5-27, examination of the metaldehyde concentration post incubation revealed degradation below the limit of detection within the MBP-MahS samples. Inspection of the MBP-MahR revealed no significant change in metaldehyde degradation over the 3.5 hour of the assay relative to the no insert control (p value= .5).

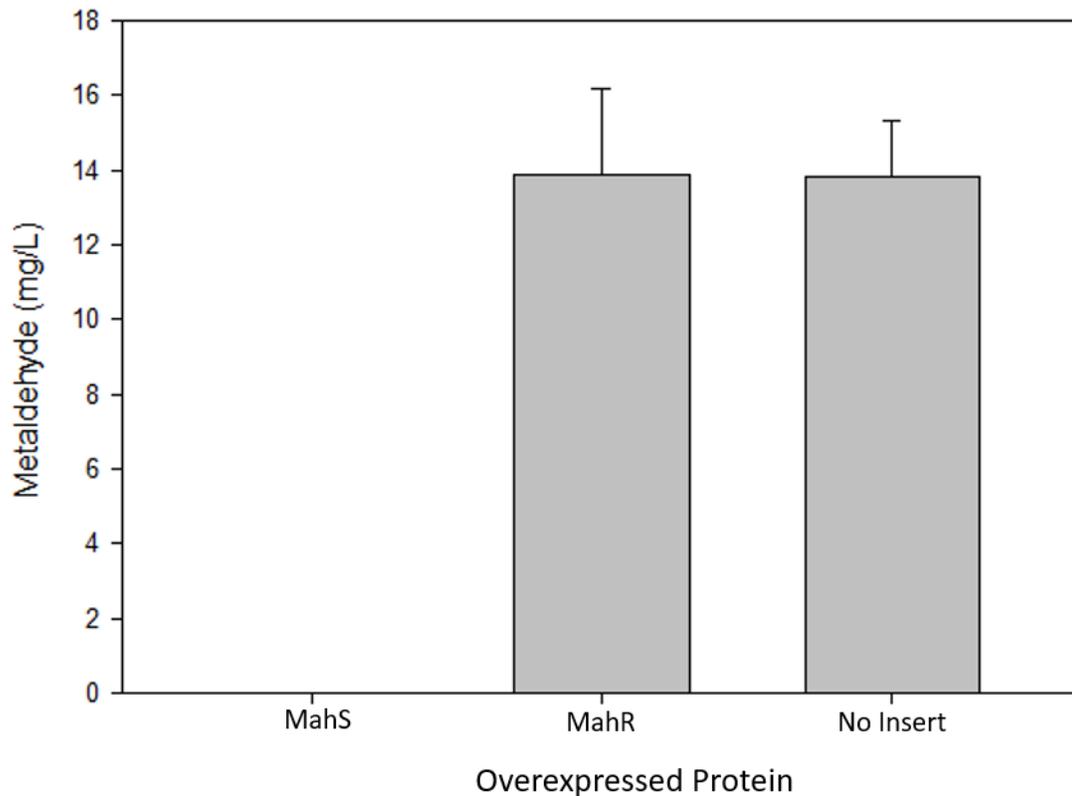


Figure 5-27: Metaldehyde concentration following 3.5 hours incubation with induced E. coli BL21 expressing MBP-MahS or MBP-MahR. A no insert control was used. All samples were performed in triplicate with error bars depicting standard deviation.

Following the verification of MahS as a metaldehyde degrading protein, further characterization of this enzyme was conducted. To determine whether the MahS fusion protein degrade metaldehyde outside the cell, the degradative ability of the soluble crude extract was investigated. This was first undertaken through demonstrating the presence of MBP-MahS within the soluble extract following 3-hour induction and cell lysis. As shown in Figure 5-28, visual

inspection of the SDS-PAGE gel revealed MBP-MahS in both the whole cell and the soluble extract post lysis.

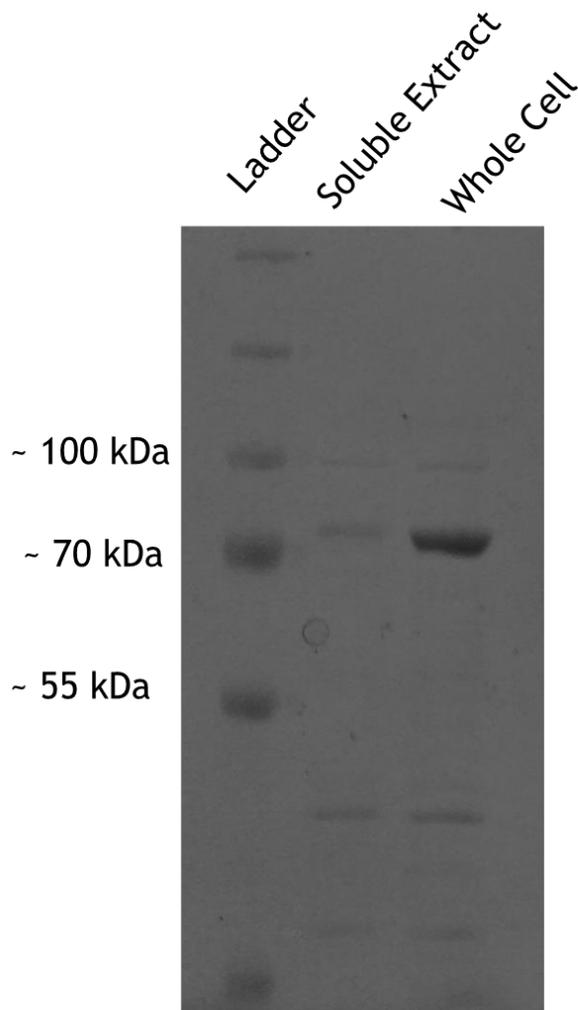


Figure 5-28: 10% SDS-PAGE gel of whole cell and soluble extract of MBP-MahS expressing cells following 3 hour induction. Pagemeruler was used to estimate protein size.

Based on the work conducted on MahX, coupled with the I-TASSER analysis, the cofactors supplemented for this assay were Fe^{2+} , αKG and L-AA. To not only determine crude extract ability but to identify if the same co-factors were required, the conditions of just enzyme, just co-factor and both were performed. As shown in Figure 5-29, significant reduction in metaldehyde was observed when both the co-factors and the enzyme were present, relative to a control containing

the cofactors but no enzyme. Similarly, in the absence of exogenous cofactors, the enzyme preparation itself yielded no significant metaldehyde removal *in vitro* (p value .2).

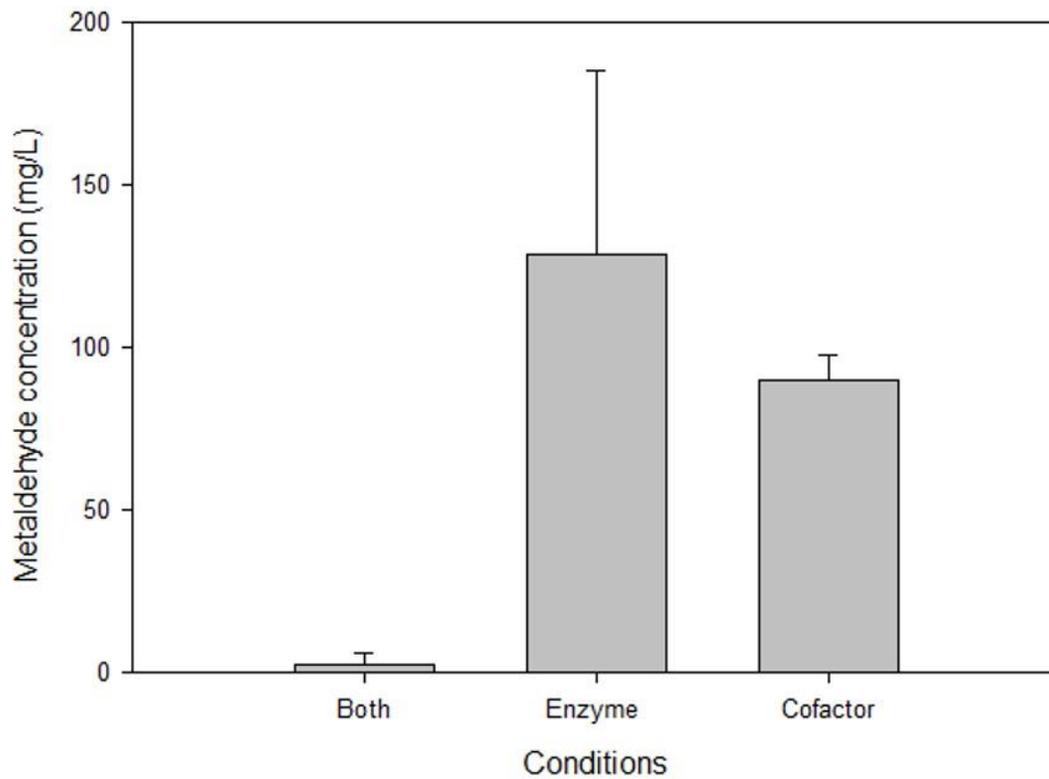


Figure 5-29: Metaldehyde concentration following 60 minute incubation of 112 mg/L with 40 mg/L crude extract of *E. coli* BL21 expressing *MahS*. Samples contained either crude extract and cofactors (Fe^{2+} , AKG and NADH), just crude extract (enzyme) or just cofactor. All samples were performed in triplicate with error bars depicting standard deviation.

5.3 Discussion

Prior to this investigation, MahX was the only identified metaldehyde degrading protein. Work conducted in this chapter aimed to identify other metaldehyde degrading enzymes found in metaldehyde degrading strains *Sphingobium sp.* CMET-H and *R. globerulus* HNO-A. Initial work was conducted through the use of a similarity-based analysis, whereby significant similarity can be used to infer functional homology. Where predicted metaldehyde proteins were identified, experimental verification of their degradative ability was undertaken.

5.3.1 *R. globerulus* HNO-A

Querying of *R. globerulus* HNO-A's genome using the MahX sequence identified a single significant result. The identified sequence (MahR) demonstrated a 32% identity to MahX, through BLASTP analysis. As the general rule for homology requires >30% identity sequence (Pearson, 2013), it is therefore considered significant enough to infer similar functional homology to MahX. The results from the reciprocal BLAST, where MahX was the most significant hit, adds confidence to the potential of an orthological relationship between MahR and MahX. The work conducted by (Castro-Gutiérrez *et al.*, 2020) demonstrated that the closely related *Rhodococcus globerulus* NBRC 14531 type strain was unable to degrade metaldehyde. Querying of this genome with MahR led to no significant result being obtained. Although absence of the protein in the type strain is not to be considered direct evidence of the function of MahR, its absence from NBRC 14531 does strengthen the argument for MahR as a metaldehyde degrading protein. Further examination of MahR was undertaken through looking at the predicted functions and

domains based on both the amino acid sequence and predicted structure. BLAST analysis revealed similarity to a phytanoyl-CoA dioxygenase family protein. Putative conserved domain results were also identified within MahR as being in a member of α KG, Fe²⁺ oxygenase superfamily. As MahX is a dioxygenase and this family has been demonstrated to be involved in numerous degradative processes this would be expected in a potential metaldehyde degrading protein.

Using the I-TASSER generated predicted protein model to query the PDB, verrucologen synthase was identified as the highest ranked and therefore possessed the greatest structural similarity to MahR. Further demonstration of similarities to a dioxygenase further supports the potential for MahR to be involved with the cleavage of metaldehyde's cyclic structure. Despite the strong bioinformatics prediction regarding MahR as a metaldehyde degrading enzyme, experimental evidence shows no significant removal of metaldehyde when incubated *in vivo*. This lack of degradative ability may be explained through *Rhodococcus* strains possessing a different intracellular milieu, relative to traditional model organisms such as *E. coli* and *B.substilis* (Nakashima and Tamura, 2004). As such, the overexpressed protein may be incorrectly processed or folded in *E. coli*. This limitation of the *E. coli* expression system may in future be overcome through expression of MahR in a strain closer to its native host. Furthermore, the construct design is important in achieving an active protein which was demonstrated with the expression of active MahX. As such, further modification of protein expression through removal/ altering of the fusion tags may demonstrate observable activity. Future investigations, relating to MahR, should attempt overexpression in the microorganism *Rhodococcus erythropolis*. This strain has been demonstrated to overcome several limitations of *E. coli* and allows the correct expression

of numerous proteins ranging from *Mycobacterium tuberculosis* recombinant glycoproteins to *Streptomyces* proteasome (Nakashima and Tamura, 2004; Vallecillo *et al.*, 2017). However, the possibility that *R. globerulus* HNO-A degrades metaldehyde through an enzymatic mechanism significantly different from *A. calcoaceticus* E1 cannot yet be dismissed. As the similarity-based analysis relies on a similar catalytic enzyme and/or mechanism, alternative pathways would be unlikely to be identified. As such, identification of the degradative gene(s) within *R. globerulus* HNO-A could be achieved through traditional gene identification methods such as mutagenesis, induction assays and clone libraries

5.32 *Sphingobium* sp. CMET-H

The BLAST analysis of *Sphingobium* sp. CMET-H revealed a 307 amino acid sequence with a 57% identity when queried with MahX. Bioinformatic analysis of MahS revealed both a predicted orthological relationship to MahX and an absence of the predicted protein within the closely related non-metaldehyde degrading type strain *Sphingobium chlorophenolicum* L-1. Further examination of MahS also demonstrated conserved regions belonging to the phytylCoA dioxygenase family protein. As discussed above, dioxygenases are involved in the initial catabolic reaction in numerous biodegradative pathways. Putative conserved domains were also identified in MahS, relating to the α KG -Fe (II) oxygenase superfamily. The I-TASSER predicted model identified Fe (II) / (alpha) ketoglutarate-dependent dioxygenase AsqJ as the highest ranked similar protein. This was also PDB result obtained for MahX during I-TASSER analysis. Similarity of predicted structure to MahX coupled with experimental validation of degradative activity would strong suggest a similar degradative mechanism.

Inspection of the genomic location of *mahS* identified a 5876 bp contig whereby seven predicted genes were in the same orientation and relatively close proximity. The maximum distance between the sequences observed was 195 bp. As this distance is below the 200 bp cut off point traditionally used in operon analysis, it further provides confidence to the operon hypothesis (Ermolaeva *et al.*, 2001; Pertea *et al.*, 2009). As the predicted genes for predicted_protein A, *mahS* and predicted_protein B showed no similarity to that of a plasmid, unlike the remaining predicted proteins, it may suggest these 3 genes belong to a catabolic cluster. Based on the degradative method described in Chapter 3 and observed in the degradation of other cyclic ethers, the initial degradative step would involve the oxygenation of the cyclic structure (Mahendra and Alvarez-Cohen, 2006; Sales *et al.*, 2013). This would cleave the stable cyclic structure and lead to the formation of the predicted hemiacetal product 1, 3, 5, 7-tetramethyl-2, 4, 6-trioxa-1-hydroxy-7-octanone. It may be possible that due to the mobile nature of the plasmid, nearby predicted sequences were part of a pre-existing degradative pathway and are not required for the utilization of metaldehyde. Due to the unstable nature of hemiacetal product and its breakdown occurring rapidly, a MahY-like protein may not necessarily be required for metaldehyde utilization in *Sphingobium sp.* CMET-H. Expression of MahS within the *Sphingobium chlorophenolicum* L-1 type strain could allow for further characterization of this pathway and determine which gene(s) are necessary for metaldehyde utilization. As both BLAST analysis of the contig and plasmidSPAdes both independently identified *mahS* as belonging to a plasmid, this seems to be highly likely. Visual examination of the predicted functions of the genes found on the mobile element, coupled with oriT analysis, suggest that the identified sequence is a conjugative plasmid. Typically, conjugative plasmids require; a relaxase gene, a gene encoding

a type 4 coupling protein (T4CP) and gene cluster for bacterial type IV secretion system and origin of transfer site (*oriT*) (Zechner *et al.*, 2012), of which the latter is absent. The absence of *oriT* may be due to the relatively poor characterization *Sphingobium* plasmids and therefore may not have been identified within the bioinformatics analysis. Further characterization of the plasmid through MinION sequencing could allow for greater confidence in this analysis and allow for a complete identification of the genes present. Further examination of the conjugative ability of the plasmid, through experimental means, could verify the ability of transferability of *mahS* through conveying the degradative ability to a non-degrading strain.

Experimental verification of the bioinformatics analysis revealed that MahS shows the ability to degrade metaldehyde both *in vivo* and *in vitro*. The crude extract assay demonstrated not only *in vitro* activity but also the cofactors required for metaldehyde degradation. Further investigation is required, however, to ascertain if all three cofactors are required for optimum degradation, similar to MahX. A recent study, within the Moir laboratory (Castro-Gutiérrez, 2020) demonstrated the industrial potential of the *Sphingobium sp.* CMET-H strain through its ability to degrade metaldehyde below the EU detection limit of 0.1 µg/L within a slow sand filter. As this was the only strain to achieve this, the identification of the metaldehyde degrading gene in this microorganism can allow, not only for the potential optimization of the MahS protein to increase efficiency, but allow for the detection and monitoring of the catabolic genes throughout the bioremediation process. Further characterization of the enzyme would require the examination of the enzyme's degradative ability when incubated with real world metaldehyde concentrations. As MahS and MahX demonstrate regions of conservation throughout their

sequence, the possibility of developing universal primers could further allow for the potential of monitoring of yet to be identified metaldehyde degrading genes. This would provide greater information regarding the degradative potential of a given matrix and allow for more accurate monitoring.

Manual examination of the structural models coupled with multiple sequence alignment analysis revealed three binding pocket residues (threonine, serine and leucine) which are conserved between MahX and MahS. All three residues were not found within any other sequences analysed suggesting a potential role in metaldehyde binding. Based on the chemical properties of metaldehyde and the presence of hydroxyl groups within two of the conserved residues, it is hypothesized that these residues may be essential in enzyme-substrate binding. Furthermore, as only 3 out of the 12 identified binding pocket residues found within MahX were also found within MahS, it would suggest that these regions have evolved towards metaldehyde binding independently.

5.33 Conclusion

This investigation set out to identify novel metaldehyde degrading genes within two metaldehyde degrading microorganisms. Through bioinformatics analysis, two predicted metaldehyde genes were identified and a single gene was experimental verified in the case of *Sphingobium* sp. CMET-H. Further work should look into developing universal primers to allow for greater and more accurate monitoring of metaldehyde degrading organisms within the environment. Further protein characterization of MahS is required to determine the K_m and V_{max} and allow more robust comparisons to be made against MahX.

Chapter 6: Final Discussion

Prior to the data generated from this research, it had been established that bacteria could utilise metaldehyde as a sole carbon source (Thomas *et al.*, 2017a; Balashova *et al.*, 2020). However, the enzymes responsible and the mechanism by which this degradation occurred were not known. From the data generated from this work, two metaldehyde degrading enzymes belonging to the superfamily of Fe (II)/ (alpha) ketoglutarate-dependent dioxygenase were identified and their ability to degrade metaldehyde *in vivo* and *in vitro* activity was demonstrated.

Random chemical mutagenesis identified a genetic locus hypothesized to contain the genes required for metaldehyde utilization within *A. calcoaceticus* E1. These genes, encoding a dioxygenase, a lyase and an aldehyde dehydrogenase, allowed a predicted degradation pathway to be theorized. Bioinformatics analysis of this gene cluster suggests that the genes are present within a transposable element located upon a plasmid. Potential future work for further characterization of the predicted degradative pathway could be performed using both chemical and molecular methods. In the case of the former, liquid chromatography mass spectrometry (LC-MS) could be performed on growth media supplemented with metaldehyde, purified dioxygenase (MahX) and the appropriate cofactors. Identification of the product would provide further confidence in the predicted pathway. Concerning molecular methods, complementing *mahX*, *mahXY* and *mahXYZ* into an *A. calcoaceticus* non-degrading strain such as RUH2202 would allow for the identification of the genes necessary for metaldehyde utilization. Through inserting the *ori* region of the *A. calcoaceticus* plasmid pWH1266 into the pBR322 vector, the role of each

of the identified genes could be further characterised. This could overcome the limitations current imposed with *E. coli* and ascertain whether utilisation of metaldehyde as a sole carbon source can occur within *A. calcoaceticus* when MahY and MahZ are absent. Through gaining a deeper understanding of the complete degradative mechanism, further optimisation could be achieved to ensure faster metaldehyde removal within *A. calcoaceticus* and thus improving the remediation properties of the organism. Interestingly, degradation of metaldehyde within *Sphingobium* sp HMET-C occurs without the presence of a MahY. This would suggest that MahY is not required and the spontaneous breakdown of the hemiacetal products can occur without the requirement of an enzyme, as shown in figure 6-1. Within *Sphingobium* sp HMET-C, acetaldehyde dehydrogenase enzymes are present therefore there is strong likelihood that this enzyme class plays a role within the conversion of the acetaldehyde molecules into acetate. As such, further analysis is required to ascertain the importance of MahY within *A. calcoaceticus* E1 for metaldehyde metabolism or whether its presence within the catabolic operon relates to an ancestral function.

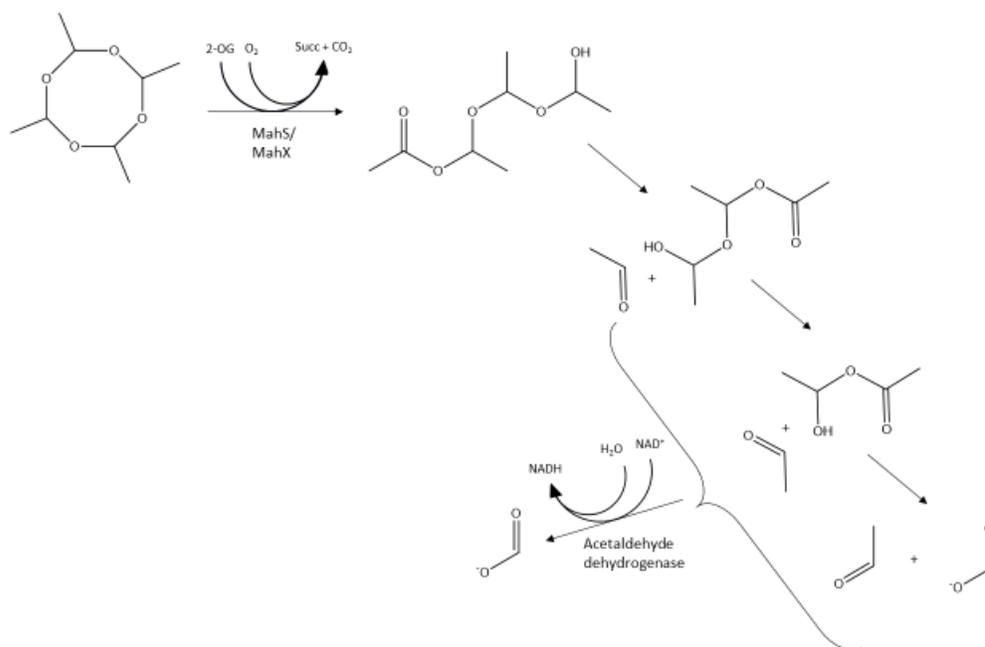


Figure 6-1: Predicted pathway for metaldehyde degradation. MahS or MahX is required for the cleavage of the cyclic structure through incorporation of a single molecule of oxygen into metaldehyde. The unstable nature of the hemiacetal molecules leads to the spontaneous break down of the intermediate products generating acetaldehyde through each iterative step. Acetaldehyde Dehydrogenase oxidases the acetaldehyde into acetate in a NAD^+ dependent manner.

Identification and characterisation of the MahX and MahS provided information regarding their potential in the bioremediation of metaldehyde from the environment. As both enzymes require the cofactors Fe^{2+} , α -Ketoglutarate (αKG) and a reducing agent to demonstrate observable metaldehyde degradation, bioremediation would be most effectively performed through whole cell methods, rather than utilising the isolated enzyme. Further augmentation of the isolated strains could allow for more efficient and degradation to occur. Through using directed evolution

approaches, genetically modified organism (GMO) limitations could be avoided and therefore allowing for an effective and easy to use bioremediation tool (Kumamaru *et al.*, 1998; Ang *et al.*, 2009b). Literature analysis, shown in table 6-1, revealed a range of potential affinity concentrations for another Fe (II)/ (alpha) ketoglutarate-dependent oxygenases. As the environment concentration of metaldehyde is considered a breach of water regulations upon exceeding 0.1 µg/L (0.6 pmol with regards to metaldehyde), these identified enzymes show potential limitations of utilising direct enzyme for bioremediative purposes. However, through utilising concentration methods significant reduction in metaldehyde may still be possible.

Table 6-1: Affinity concentrations of characterized Fe (II)/ (alpha) ketoglutarate-dependent oxygenases.

Enzyme	Substrate	Km	Source
Taurine dioxygenase	Taurine	55 μ M	(Kukor and Olsen, 1996)
Catechol 2,3-Dioxygenases	Catechol	5.5 μ M	(Eichhorn <i>et al.</i> , 1997)
SfnG	Dimethylsulfone	69 μ M	(Wicht, 2016)
Xanthine hydroxylase	Xanthine	45 μ M	(Montero-Morán <i>et al.</i> , 2007)
AtsK	Hexyl sulfate	40 pM	(Kahnert and Kertesz, 2000)
RdPa	(R) Enantiomers of mecoprop [2-(4-chloro-2-methylphenoxy)propanoic acid]	99 μ M	(Müller <i>et al.</i> , 2006)
HtXa	Hypophosphite	0.58 mM	(White and Metcalf, 2002)
DszB	Dibenzothiophene	8.2 μ m	(Nakayama <i>et al.</i> , 2002)

Next steps for the analysis of MahX and MahS would include determining whether the relatively much faster rate of degradation rate of metaldehyde by *Sphingobium* sp HMET-C over *A. calcoaceticus* E1 at low metaldehyde concentrations in pilot scale sand filters is related to enzymatic properties or through alternative factors. This could be achieved through further characterization of the purified proteins and through determining initial reaction velocity with various substrate concentrations. As such, important enzymatic values particularly the K_m could be calculated to allow effective comparisons to be made. Within the project, work was attempted to achieve this utilizing a Clark electrode with the aim of correlating oxygen consumption with metaldehyde turnover. However due to the enzymatic requirements of a reducing agent and ferrous ions, accurate measurements were unable to be calculated. To determine the industrial applications of heterologous expressed metaldehyde degrading proteins, future work should also determine whether removal below the regulatory limit of 0.1 μ g/L is achieved.

The utilisation of the cofactor α KG provides opportunity for a bioassay to be constructed with the aim of metaldehyde detection. Through coupling succinate formation with that of metaldehyde degradation, a spectrophotometric bioassay may be created (Luo *et al.*, 2006). Alternatively, as Fe (II)/(α)ketoglutarate-dependent dioxygenases require molecular oxygen as a co-substrate measuring of the change of oxygen through colormetric or amperometric methods could provide an alternative method for quantifying metaldehyde (Carritt and Kanwisher, 1959; Wei *et al.*, 2019). Bioassays based on these methods have the potential to provide high throughput, quick and effective metaldehyde detection methods and therefore would be incredibly useful within the water industry.

Identification and verification of two novel metaldehyde degrading genes allows for metaldehyde degrading genes within the environment to be actively monitored. This approach has practical applications that could be utilised within, for example, slow sand filters in water treatment works to monitor the presence and abundance of degradative genes within a given reactor. Sand filters have been demonstrated to be a clean and effective method for metaldehyde removal, but highly variable between slow sand filters, and therefore further optimisation of these processes would increase the viability of this approach for metaldehyde bioremediation (Taylor Eighmy *et al.*, 1992; Castle *et al.*, 2018). Through identification of conserved nucleotide regions within both *A. calcoaceticus* E1 and *Sphingobium sp* HMET-C, universal primers could be designed to allow for the monitoring of numerous metaldehyde genes potentially including those not yet identified.

Through combining protein modelling with multiple sequence alignment analysis, the Fe²⁺ αKG binding pocket was found to possess conserved residues between MahX and MahS. Of the 12 binding pocket residues identified for each protein, only three were conserved within both degrading enzymes. As no other sequence possessed all three residues, it may be that these are essential for metaldehyde binding. Further work regarding this could involve measuring the change of metaldehyde degradation following directed mutations of the identified residues. Furthermore, 9 out of 12 residues within MahX binding pocket were not identified as conserved within MahS. As such, this suggests the independent evolution toward metaldehyde binding. Through examining highly similar protein sequences that possess one or two of the conserved residues, further information regarding the evolution of dioxygenases toward metaldehyde degradation could be identified.

Despite the legislative change surrounding metaldehyde within the UK (DEFRA, 2020), the production and utilisation of metaldehyde throughout Europe and the United States (Berny *et al.*, 2010; Asfaw *et al.*, 2018; Environment Agency, 2018) demonstrates the importance of understanding and monitoring of metaldehyde degradation pathways. The data provided within this research provides the tools to further characterise these degradative enzymes and allows further work to identify and monitor these genes within the environment. As well as the direct implications of this research, the work conducted here has described a powerful and effective pipeline in the identification of novel genes in non-model organisms. Through combining random chemical mutagenesis with WGS and comparative genomics, novel enzymes can be discovered and their mechanisms further characterised.

Abbreviations

2D-GE	Two-dimensional gel electrophoresis
-------	-------------------------------------

AA	Amino Acid
α KG	Alpha-KetoGlutarate
AOP	Advanced Oxidation Processes
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
BSR	Blast Score Ratio
C-Score	Confidence Score
DEET	N,N-diethyl-meta-toluamide
DNA	Deoxyribonucleic acid
DSBH	Double Stranded Beta Helix
DTT	Dithiothreitol
DWI	Drinking Water Inspectorate
EDTA	Ethylenediaminetetraacetic acid
EMS	Ethyl methanesulfonate
FAD/FADH	Flavin adenine dinucleotide
FAIMS	Field asymmetric waveform ion mobility spectrometry
GAC	Granular Activated Carbon
GC	Gas Chromatography
GMO	Genetically Modified Organism
GO	Gene Ontology
HGT	Horizontal Gene Transfer
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropylthio- β -galactoside
IS	Insertion Sequence
LAA	Ascorbic Acid
LAMP	Loop-mediated isothermal amplification
LB	Lysogeny broth (LB)
LC	Liquid Chromatography
LOD	Limit of Detection
LOQ	Limit of Quantification
MBP	Maltose Binding Protein
MGE	Mobile Genetic Elements
MS	Mass Spectrometry
MSM	Minimal Salts Media
MUSCLE	MULTiple Sequence Comparison by Log-Expectation
NAD ⁺ /NADH	Nicotinamide adenine dinucleotide
OD	Optical Density

oriT	Origin of Transfer
PAC	Powder Activated Carbon
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDB	Protein Databank
qPCR	Quantitative Polymerase Chain Reaction
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
RT-qPCR	Reverse Transcription- Quantitative Polymerase Chain Reaction
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
T4CP	Type IV Coupling Protein
T4SS	Type IV Secretion System
TCA	Tricarboxylic acid cycle
TCP	1,2,3-trichloropropane
TEMED	Tetramethylethylenediamine
T-M Score	Template Modelling Score
UV	Ultraviolet
VOC	Vicinal oxygen chelate
WGS	Whole Genome Sequencing
WT	Wild Type

Reference list

- Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., et al. (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**: D8–D13.
- Ahn, J.H., Lee, S.A., Kim, S.J., You, J., Han, B.H., Weon, H.Y., and Lee, S.W. (2018) Biodegradation of organophosphorus insecticides with P–S bonds by two *Sphingobium* sp. strains. *Int. Biodeterior. Biodegrad.* **132**: 59–65.

Aik, W.S., McDonough, M.A., Thalhammer, A., Chowdhury, R., and Schofield, C.J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22**: 691–700.

Biol. **22**: 691–700.

Alder, L., Greulich, K., Kempe, G., and Vieth, B. (2006) Residue analysis of 500 high priority pesticides: Better by GC-MS or LC-MS/MS? *Mass Spectrom. Rev.* **25**: 838–865.

Altarawneh, M., Waters, D., Goh, B.M., Jiang, Z.T., El-Harbawi, M., and Yin, C.Y. (2020) Adsorptive interactions between metaldehyde and sulfonic functional group in ion exchange resin. *J. Mol. Liq.* **313**: 113555.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.

Ameen, F.A., Hamdan, A.M., and El-Naggar, M.Y. (2020) Assessment of the heavy metal bioremediation efficiency of the novel marine lactic acid bacterium, *Lactobacillus plantarum* MF042018. *Sci. Rep.* **10**: 1–11.

Andreasen, J.R. (1993) Metaldehyde toxicosis in ducklings. *J. Vet. Diagnostic Investig.* **5**: 500–501.

Ang, E.L., Obbard, J.P., and Zhao, H. (2009b) Directed evolution of aniline dioxygenase for enhanced bioremediation of aromatic amines. *Appl. Microbiol. Biotechnol.* **81**: 1063–1070.

Angela, Y., Wy, V., and Julius, A. (2017) Preparation of Calcium Competent *Escherichia coli* and Heat-Shock Transformation | jemi.microbiology.ubc.ca. **1**: 22–25.

Anglian Water (2019) Slug it Out. *Sliug It Out*.

- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P.A. (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* **32**: btw493.
- Araújo, W.L., Martins, A.O., Fernie, A.R., and Tohge, T. (2014) 2-Oxoglutarate: linking TCA cycle function with amino acid, glucosinolate, flavonoid, alkaloid, and gibberellin biosynthesis. *Front. Plant Sci.* **5**: 552.
- Arena, M., Auteri, D., Barmaz, S., Brancato, A., Brocca, D., Bura, L., et al. (2018) Peer review of the pesticide risk assessment of the active substance methiocarb. *EFSA J.* **16**: e05429.
- Asfaw, A., Maher, K., and Shucksmith, J.D. (2018) Modelling of metaldehyde concentrations in surface waters: A travel time based approach. *J. Hydrol.* **562**: 397–410.
- Autin, O., Hart, J., Jarvis, P., MacAdam, J., Parsons, S.A., and Jefferson, B. (2012) Comparison of UV/H₂O₂ and UV/TiO₂ for the degradation of metaldehyde: Kinetics and the impact of background organics. *Water Res.* **46**: 5655–5662.
- Autin, O., Hart, J., Jarvis, P., MacAdam, J., Parsons, S.A., and Jefferson, B. (2013) The impact of background organic matter and alkalinity on the degradation of the pesticide metaldehyde by two advanced oxidation processes: UV/H₂O₂ and UV/TiO₂. *Water Res.* **47**: 2041–2049.
- Azubuiké, C.C., Chikere, C.B., and Okpokwasili, G.C. (2016) Bioremediation techniques—classification based on site of application: principles, advantages, limitations and prospects. *World J. Microbiol. Biotechnol.* **32**:
- Baćmaga, M., Wyszowska, J., and Kucharski, J. (2019) Biostimulation as a process aiding tebuconazole degradation in soil. *J. Soils Sediments* **19**: 3728–3741.

- Baghel, S., Sahariah, B.P., and Anandkumar, J. (2020) Bioremediation of Lignin-Rich Pulp and Paper Industry Effluent. In, *Combined Application of Physico-Chemical & Microbiological Processes for Industrial Effluent Treatment Plant*. Springer Singapore, pp. 261–278.
- Balashova, N., Wilderspin, S., Cai, C., and Reid, B.J. (2020) Ubiquity of microbial capacity to degrade metaldehyde in dissimilar agricultural, allotment and garden soils. *Sci. Total Environ.* **704**:
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.
- Basta, T., Keck, A., Klein, J., and Stolz, A. (2004) Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains. *J. Bacteriol.* **186**: 3862–3872.
- Bathe, S., Schwarzenbeck, N., and Hausner, M. (2009) Bioaugmentation of activated sludge towards 3-chloroaniline removal with a mixed bacterial population carrying a degradative plasmid. *Bioresour. Technol.* **100**: 2902–2909.
- Bennett, P.M. (2008) Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. In, *British Journal of Pharmacology*. Wiley-Blackwell, p. S347.
- Berny, P., Caloni, F., Croubels, S., Sachana, M., Vandenbroucke, V., Davanzo, F., and Guitart, R. (2010) Animal poisoning in Europe. Part 2: Companion animals. *Vet. J.* **183**: 255–259.

- Bers, K., Leroy, B., Breugelmans, P., Albers, P., Lavigne, R., Sørensen, S.R., et al. (2011) A novel hydrolase identified by genomic-proteomic analysis of phenylurea herbicide mineralization by *Variovorax* sp. strain SRS16. *Appl. Environ. Microbiol.* **77**: 8754–8764.
- Bieri, M. (2003) The environmental profile of metaldehyde. *BCPC Symp. Proc.* 255–262.
- Borlongan, I.G., Coloso, R.M., and Blum, R.A. (1996) Use of metaldehyde as molluscicide in milkfish ponds. 205–212.
- Boto, L. (2010) Horizontal gene transfer in evolution: Facts and challenges. *Proc. R. Soc. B Biol. Sci.* **277**: 819–827.
- Bräuer, A., Beck, P., Hintermann, L., and Groll, M. (2016) Structure of the Dioxygenase AsqJ: Mechanistic Insights into a One-Pot Multistep Quinolone Antibiotic Biosynthesis. *Angew. Chemie - Int. Ed.* **55**: 422–426.
- Busquets, R., Kozynchenko, O.P., Whitby, R.L.D., Tennison, S.R., and Cundy, A.B. (2014) Phenolic carbon tailored for the removal of polar organic contaminants from water: A solution to the metaldehyde problem? *Water Res.* **61**: 46–56.
- Cámara, B., Bielecki, P., Kaminski, F., Dos Santos, V.M., Plumeier, I., Nikodem, P., and Pieper, D.H. (2007) A gene cluster involved in degradation of substituted salicylates via ortho cleavage in *Pseudomonas* sp. strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate. *J. Bacteriol.* **189**: 1664–1674.
- Carritt, D.E. and Kanwisher, J.W. (1959) An Electrode System for Measuring Dissolved Oxygen. *Anal. Chem.* **31**: 5–9.
- Cases, I. and De Lorenzo, V. (2001) The black cat/white cat principle of signal integration in bacterial promoters. *EMBO J.* **20**: 1–11.

Castle, G.D., Mills, G.A., Bakir, A., Gravell, A., Schumacher, M., Snow, K., and Fones, G.R. (2018) Measuring metaldehyde in surface waters in the UK using two monitoring approaches.

Environ. Sci. Process. Impacts **20**: 1180–1190.

Castle, G.D., Mills, G.A., Gravell, A., Jones, L., Townsend, I., Cameron, D.G., and Fones, G.R. (2017) Review of the molluscicide metaldehyde in the environment. *Environ. Sci. Water Res. Technol.* **3**: 415–428.

Castle, G.D., Mills, G.A., Gravell, A., Leggatt, A., Stubbs, J., Davis, R., and Fones, G.R. (2019) Comparison of different monitoring methods for the measurement of metaldehyde in surface waters. *Environ. Monit. Assess.* **191**:75 1-13

Castro-Gutiérrez, V., Fuller, E., Thomas, J.C., Sinclair, C.J., Johnson, S., Helgason, T., and Moir, J.W.B. (2020) Genomic basis for pesticide degradation revealed by selection, isolation and characterisation of a library of metaldehyde-degrading strains from soil. *Soil Biol. Biochem.* **142**: 107702.

Cerbin, S. and Jiang, N. (2018) Duplication of host genes by transposable elements. *Curr. Opin. Genet. Dev.* **49**: 63–69.

Chen, H., Giri, N.C., Zhang, R., Yamane, K., Zhang, Y., Maroney, M., and Costa, M. (2010) Nickel ions inhibit histone demethylase JMJD1A and DNA repair enzyme ABH2 by replacing the ferrous iron in the catalytic centers. *J. Biol. Chem.* **285**: 7374–7383.

Chen, P.Y.T., Aman, H., Can, M., Ragsdale, S.W., and Drennan, C.L. (2018) Binding site for coenzyme A revealed in the structure of pyruvate:ferredoxin oxidoreductase from *Moorella thermoacetica*. *Proc. Natl. Acad. Sci. U. S. A.* **115**: 3846–3851.

Chen, X., Wong, C.H., and Ma, C. (2019) Targeting the Bacterial Transglycosylase: Antibiotic

- Development from a Structural Perspective. *ACS Infect. Dis.* **5**: 1493–1504.
- Cheng, M., Yan, X., He, J., Qiu, J., and Chen, Q. (2019) Comparative genome analysis reveals the evolution of chloroacetanilide herbicide mineralization in *Sphingomonas wittichii* DC-6. *Arch. Microbiol.* **201**: 907–918.
- Chiang, Y. and Jerry Kresge, A. (1985) Kinetics of Hydrolysis of Acetaldehyde Ethyl Hemiacetal in Aqueous Solution. *J. Org. Chem.* **50**: 5038–5040.
- Chikere, C.B., Okpokwasili, G.C., and Chikere, B.O. (2011) Monitoring of microbial hydrocarbon remediation in the soil. *3 Biotech* **1**: 117–138.
- Chong, H., Yeow, J., Wang, I., Song, H., and Jiang, R. (2013) Improving Acetate Tolerance of *Escherichia coli* by Rewiring Its Global Regulator cAMP Receptor Protein (CRP). *PLoS One* **8**:
- Clarke, T.E., Romanov, V., Chirgadze, Y.N., Klomsiri, C., Kisselman, G., Wu-Brown, J., et al. (2011) Crystal structure of alkyl hydroperoxidase D like protein PA0269 from *Pseudomonas aeruginosa*: Homology of the AhpD-like structural family. *BMC Struct. Biol.* **11**: 27.
- Colavecchio, A., Cadieux, B., Lo, A., and Goodridge, L.D. (2017) Bacteriophages contribute to the spread of antibiotic resistance genes among foodborne pathogens of the Enterobacteriaceae family - A review. *Front. Microbiol.* **8**:
- Cold Spring Harbour (2015) SDS-PAGE Gel. *Cold Spring Harb. Protoc.* **2015**: pdb.rec087908.
- Coloso, R.M., Borlongan, I.G., and Blum, R.A. (1998) Use of metaldehyde as a molluscicide in semicommercial and commercial milkfish ponds. *Crop Prot.* **17**: 669–674.
- Cooke, A., Rettino, J., Flower, L., Filby, K., and Freer, A. (2020) Farming for water; catchment management initiatives for reducing pesticides. *Water Environ. J.* **34**: 679–691.

- Cope, R.B., White, K.S., More, E., Holmes, K., Nair, A., Chauvin, P., and Oncken, A. (2006) Exposure-to-treatment interval and clinical severity in canine poisoning: A retrospective analysis at a Portland Veterinary Emergency Center. *J. Vet. Pharmacol. Ther.* **29**: 233–236.
- Copley, S.D. (2020) Evolution of new enzymes by gene duplication and divergence. *FEBS J.* **287**: 1262–1283.
- Cui, K. and Shoemaker, S.P. (2018) Public perception of genetically-modified (GM) food: A Nationwide Chinese Consumer Study. *npj Sci. Food* **2**: 1–8.
- Dalby, P.A. (2003) Optimising enzyme function by directed evolution. *Curr. Opin. Struct. Biol.* **13**: 500–505.
- Davidson, T.L., Chen, H., Di Toro, D.M., D'Angelo, G., and Costa, M. (2006) Soluble nickel inhibits HIF-prolyl-hydroxylases creating persistent hypoxic signaling in A549 cells. *Mol. Carcinog.* **45**: 479–489.
- Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., et al. (2020) The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Res.* **48**: D606–D612.
- Deepa, A., Naveena, K., and Anindya, R. (2018) DNA repair activity of Fe(II)/2OG-dependent dioxygenases affected by low iron level in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **18**:
- DEFRA (2020) Outdoor use of metaldehyde to be banned to protect wildlife - GOV.UK. *Press Release*.
- Deutscher, J., Francke, C., and Postma, P.W. (2006) How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiol. Mol. Biol. Rev.* **70**: 939–1031.

- Dolan, T., Howsam, P., Parsons, D.J., and Whelan, M.J. (2013) Is the EU drinking water directive standard for pesticides in drinking water consistent with the precautionary principle? *Environ. Sci. Technol.* **47**: 4999–5006.
- Dolder, L.K. and Volmer, P.A. (2003) Metaldehyde toxicosis. *Vet. Med.* **98**: 213–215.
- Doria, F.C., Borges, A.C., Kim, J.K., Nathan, A., Joo, J.C., and Campos, L.C. (2013) Removal of metaldehyde through photocatalytic reactions using nano-sized zinc oxide composites. *Water. Air. Soil Pollut.* **224**: 1–9.
- Drinking Water Inspectorate (2019) Drinking water 2015: Summary of the Chief Inspector's report for drinking water in England. 1–97.
- Drury, J., LLwis, D.R., and Madel, G.A. (1939) Fatal Poisoning by "Meta Fuel" Tablets. *Br. Med. J.* **1**: 1283–1284.
- Dvorak, P., Bidmanova, S., Damborsky, J., and Prokop, Z. (2014) Immobilized synthetic pathway for biodegradation of toxic recalcitrant pollutant 1,2,3-trichloropropane. *Environ. Sci. Technol.* **48**: 6859–6866.
- Eckert, M., Fleischmann, G., Jira, R., Bolt, H.M., and Golka, K. (2006) Acetaldehyde. In, *Ullmann's Encyclopedia of Industrial Chemistry*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- Edwards, C.A., Arancon, N.Q., Vasko-Bennett, M., Little, B., and Askar, A. (2009) The relative toxicity of metaldehyde and iron phosphate-based molluscicides to earthworms. *Crop Prot.* **28**: 289–294.

Environment Agency (2018) The Environment Agency 's approach to groundwater protection. 1–56.

Environmental Protection Agency (2006) Registration Eligibility Decision for Metaldehyde.

Ermolaeva, M.D., White, O., and Salzberg, S.L. (2001) Prediction of operons in microbial genomes. *Nucleic Acids Res.* **29**: 1216–1221.

European Food Safety Authority (2014) CLH report proposal for harmonised classification and labelling. Substance Name: Medetomidine. **2008**: 1–59.

Ferguson, L.R. and Denny, W.A. (1991) The genetic toxicology of acridines. *Mutat. Res. Genet. Toxicol.* **258**: 123–160.

Food, E. and Authority, S. (2015) Conclusion on the peer review of the pesticide risk assessment of the active substance ferric phosphate. *EFSA J.* **13**: 1–50.

Fulazzaky, M.A. and Omar, R. (2012) Removal of oil and grease contamination from stream water using the granular activated carbon block filter. *Clean Technol. Environ. Policy* **14**: 965–971.

Gavrić, S., Leonhardt, G., Marsalek, J., and Viklander, M. (2019) Processes improving urban stormwater quality in grass swales and filter strips: A review of research findings. *Sci. Total Environ.* **669**: 431–447.

Geissdorfer, W., Frosch, S.C., Haspel, G., Ehrt, S., and Hillen, W. (1995) Two genes encoding proteins with similarities to rubredoxin and rubredoxin reductase are required for conversion of dodecane to lauric acid in *Acinetobacter calcoaceticus* ADP1. *Microbiology* **141**: 1425–1432.

- Glen, D.M. and Orsman, I.A. (1986) Comparison of molluscicides based on metaldehyde, methiocarb or aluminium sulphate. *Crop Prot.* **5**: 371–375.
- Golovko, O., De Brito, L., Cascone, C., Ahrens, L., Lavonen, E., and Kohler, S. (2020) Sorption Characteristics and Removal Efficiency of Organic Micropollutants in Drinking Water Using. *Mdpi* **12**: 1–14.
- Gómez-Gómez, C., Blanco-Picazo, P., Brown-Jaque, M., Quirós, P., Rodríguez-Rubio, L., CerdàCuellar, M., and Muniesa, M. (2019) Infectious phage particles packaging antibiotic resistance genes found in meat products and chicken feces. *Sci. Rep.* **9**: 1–11.
- Goncharova, R.I. and Kuzhir, T.D. (1989) A comparative study of the antimutagenic effects of antioxidants on chemical mutagenesis in *Drosophila melanogaster*. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **214**: 257–265.
- Gong, T., Xu, X., Che, Y., Liu, R., Gao, W., Zhao, F., et al. (2017) Combinatorial metabolic engineering of *Pseudomonas putida* KT2440 for efficient mineralization of 1,2,3trichloropropane. *Sci. Rep.* **7**: 1–12.
- Government of Canada (2020) Proposed Special Review Decision PSRD2020-02, Special Review for Metaldehyde and Its Associated End-use Products - Canada.ca. *Rev. Met. Its Assoc. Enduse.*
- Government of Canada, H.C.P.M.R.A. Pesticide Product Information - Health Canada.
- Goyal, A.K. and Zylstra, G.J. (1996) Molecular cloning of novel genes for polycyclic aromatic hydrocarbon degradation from *Comamonas testosteroni* GZ39. *Appl. Environ. Microbiol.* **62**: 230–236.

- Van Groenigen, J.W., Lubbers, I.M., Vos, H.M.J., Brown, G.G., De Deyn, G.B., and Van Groenigen, K.J. (2014) Earthworms increase plant production: a meta-analysis. *Sci. Rep.* **4**: 1–7.
- Grzyska, P.K., Appelman, E.H., Hausinger, R.P., and Proshlyakov, D.A. (2010) Insight into the mechanism of an iron dioxygenase by resolution of steps following the FeIV=O species. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 3982–3987.
- Gunathilaka, G.U., Tahlan, V., Mafiz, A.I., Polur, M., and Zhang, Y. (2017) Phages in urban wastewater have the potential to disseminate antibiotic resistance. *Int. J. Antimicrob. Agents* **50**: 678–683.
- Gutiérrez-González, M., Farías, C., Tello, S., Pérez-Etcheverry, D., Romero, A., Zúñiga, R., et al. (2019) Optimization of culture conditions for the expression of three different insoluble proteins in *Escherichia coli*. *Sci. Rep.* **9**: 1–11.
- Hallett, K.C., Atfield, A., Comber, S., and Hutchinson, T.H. (2016) Developmental toxicity of metaldehyde in the embryos of *Lymnaea stagnalis* (Gastropoda: Pulmonata) co-exposed to the synergist piperonyl butoxide. *Sci. Total Environ.* **543**: 37–43.
- Hara, H., Eltis, L.D., Davies, J.E., and Mohn, W.W. (2007) Transcriptomic analysis reveals a bifurcated terephthalate degradation pathway in *Rhodococcus* sp. strain RHA1. *J. Bacteriol.* **189**: 1641–1647.
- Health and Safety Executive (2014) The Genetically Modified Organisms (Contained Use) Regulations 2014. *Heal. Saf. Exec.* **2002**: 1–78.
- Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R., and Warwicker, J. (2017) ProteinSol: A web tool for predicting protein solubility from sequence. *Bioinformatics* **33**: 3098–3100.

- Heiss, G., Trachtmann, N., Abe, Y., Takeo, M., and Knackmuss, H.J. (2003) Homologous npdGl genes in 2,4-dinitrophenol- and 4-nitrophenol-degrading *Rhodococcus* spp. *Appl. Environ. Microbiol.* **69**: 2748–2754.
- Holben, W.E., Schroeter, B.M., Calabrese, V.G.M., Olsen, R.H., Kukor, J.K., Biederbeck, V.O., et al. (1992) Gene probe analysis of soil microbial populations selected by amendment with 2,4-dichlorophenoxyacetic acid. *Appl. Environ. Microbiol.* **58**: 3941–3948.
- Holmquist, M. (2005) Alpha Beta-Hydrolase Fold Enzymes Structures, Functions and Mechanisms. *Curr. Protein Pept. Sci.* **1**: 209–235.
- Holroyd, L.F. and Van Mourik, T. (2015) Stacking of the mutagenic base analogue 5-bromouracil: Energy landscapes of pyrimidine dimers in gas phase and water. *Phys. Chem. Chem. Phys.* **17**: 30364–30370.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al. (2009) InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37**: D211.
- Hynes, W., Johnson, C., and Stokes, M. (2009) A single nucleotide mutation results in loss of enzymatic activity in the hyaluronate lyase gene of *Streptococcus pyogenes*. *Microb. Pathog.* **47**: 308–313.
- Innan, H. and Kondrashov, F. (2010) The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **11**: 97–108.
- Ishida, T. (2010) Effects of point mutation on enzymatic activity: Correlation between protein electronic structure and motion in chorismate mutase reaction. *J. Am. Chem. Soc.* **132**: 7104–7118.

- Iwata, Y., Carman, G.E., Dinoff, T.M., and Gunther, F.A. (1982) Metaldehyde Residues on and in Citrus Fruits after a Soil Broadcast of a Granular Formulation and after a Spray Application to Citrus Trees. *J. Agric. Food Chem.* **30**: 606–608.
- Jack, D.L., Paulsen, I.T., and Saier, J. (2000) The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* **146**: 1797–1814.
- Janssen, D.B., Dinkla, I.J.T., Poelarends, G.J., and Terpstra, P. (2005) Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ. Microbiol.* **7**: 1868–1882.
- Jia, Y., Eltoukhy, A., Wang, J., Li, X., Hlaing, T.S., Aung, M.M., et al. (2020) Biodegradation of bisphenol a by sphingobium sp. YC-JY1 and the essential role of cytochrome P450 Monooxygenase. *Int. J. Mol. Sci.* **21**..
- Kalindamar, S., Lu, J., Abdelhamed, H., Tekedar, H.C., Lawrence, M.L., and Karsi, A. (2019) Transposon mutagenesis and identification of mutated genes in growth-delayed *Edwardsiella ictaluri*. *BMC Microbiol.* **19**: 55.
- Kanissery, R.G. and Sims, G.K. (2011) Biostimulation for the Enhanced Degradation of Herbicides in Soil. *Appl. Environ. Soil Sci.* **2011**: 1–10.
- Kapust, R.B. and Waugh, D.S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**: 1668–1674.
- Kasai, D., Imai, S., Asano, S., Tabata, M., Iijima, S., Kamimura, N., et al. (2017) Identification of natural rubber degradation gene in *Rhizobacter gummiphilus* NS21. *Biosci. Biotechnol.*

Biochem. **81**: 614–620.

Khan, A., Schofield, C.J., and Claridge, T.D.W. (2020) Reducing Agent-Mediated Nonenzymatic Conversion of 2-Oxoglutarate to Succinate: Implications for Oxygenase Assays.

ChemBioChem **21**: 2898–2902.

Kim, D., Kim, Y.S., Kim, S.K., Kim, S.W., Zylstra, G.J., Kim, Y.M., and Kim, E. (2002) Monocyclic aromatic hydrocarbon degradation by *Rhodococcus* sp. strain DK17. *Appl. Environ.*

Microbiol. **68**: 3270–3278.

Kohn, J.E. and Plaxco, K.W. (2005) Conversion of a maltose receptor into a zinc biosensor by computational design. *PNAS* **98**: 4955–4960.

Kohno, T., Sugimoto, Y., Sei, K., and Mori, K. (2002) Design of PCR Primers and Gene Probes for General Detection of Alkane-Degrading Bacteria. *Microbes Environ.* **17**: 114–121.

Korolev, S., Skarina, T., Evdokimova, E., Beasley, S., Edwards, A., Joachimiak, A., and Savchenko, A. (2002) Crystal structure of glutamine amidotransferase from *Thermotoga maritima*.

Proteins Struct. Funct. Genet. **49**: 420–422.

Kostal, J., Yang, R., Wu, C.H., Mulchandani, A., and Chen, W. (2004) Enhanced arsenic accumulation in engineered bacterial cells expressing ArsR. *Appl. Environ. Microbiol.* **70**:

4582–4587.

Kothari, A., Soneja, D., Tang, A., Carlson, H.K., Deutschbauer, A.M., and Mukhopadhyay, A.

(2019) Native Plasmid-Encoded Mercury Resistance Genes Are Functional and Demonstrate Natural Transformation in Environmental Bacterial Isolates. *mSystems* **4**:

Kowalczyk, A., Eyice, Ö., Schäfer, H., Price, O.R., Finnegan, C.J., van Egmond, R.A., et al. (2015)

Characterization of para-nitrophenol-degrading bacterial communities in river water by

- using functional markers and stable isotope probing. *Appl. Environ. Microbiol.* **81**: 6890–6900.
- Kube, M., Chernikova, T.N., Al-Ramahi, Y., Beloqui, A., Lopez-Cortez, N., Guazzaroni, M.E., et al. (2013) Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nat. Commun.* **4**:
- Kuiper, C. and Vissers, M.C.M. (2014) Ascorbate as a cofactor for Fe- and 2-oxoglutarate dependent dioxygenases: Physiological activity in tumour growth and progression. *Front. Oncol.* **4**:
- Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* **16**: 663–666.
- Kumar Singh, N., Pandey, S., Singh, R.P., Muzamil Gani, K., Yadav, M., Thanki, A., and Kumar, T. (2020) Bioreactor and bioprocess technology for bioremediation of domestic and municipal wastewater. In, *Bioremediation of Pollutants*. Elsevier, pp. 251–273.
- Langan, A.M. and Shaw, E.M. (2006) Responses of the earthworm *Lumbricus terrestris* (L.) to iron phosphate and metaldehyde slug pellet formulations. *Appl. Soil Ecol.* **34**: 184–189.
- Lanigan, R.S., Yamarik, T.A., and Andersen, F.A. (2002) Final report on the safety assessment of EDTA, calcium disodium EDTA, diammonium EDTA, dipotassium EDTA, disodium EDTA, TEA-EDTA, tetrasodium EDTA, tripotassium EDTA, trisodium EDTA, HEDTA, and trisodium HEDTA. *Int. J. Toxicol.* **21**: 95–142.
- Lawrence, D., Fiegna, F., Behrends, V., Bundy, J.G., Phillimore, A.B., Bell, T., and Barraclough, T.G. (2012) Species Interactions Alter Evolutionary Responses to a Novel Environment.

PLoS Biol. **10**: e1001330.

León, J., Shulaev, V., Yalpani, N., Lawton, M.A., and Raskin, I. (1995) Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 10413–10417.

Lewis, K.A., Tzilivakis, J., Warner, D.J., and Green, A. (2016) An international database for pesticide risk assessments and management. *Hum. Ecol. Risk Assess.* **22**: 1050–1064.

Li, C., Wu, Y.L., Yang, T., and Zhang, Y. (2010) Determination of metaldehyde in water by SPE and UPLC-MS-MS. *Chromatographia* **72**: 987–991.

Li, R.W. (2011) Metagenomics and its applications in agriculture, biomedicine, and environmental studies Nova Science Publisher's.

Li, X., Xie, Y., Liu, M., Tai, C., Sun, J., Deng, Z., and Ou, H.Y. (2018) OriTfinder: A web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.* **46**: W229–W234.

Li, Z., Li, J., Guo, Z., and Campos, L.C. (2020) Investigation of metaldehyde removal by powdered activated carbon from different water samples. *Environ. Sci. Water Res. Technol.* **6**: 1432–1444.

Liang, Y., Wan, N., Cheng, Z., Mo, Y., Liu, B., Liu, H., et al. (2017) Whole-genome identification and expression pattern of the vicinal oxygen chelate family in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **8**:

Liu, X., Gai, Z., Tao, F., Tang, H., and Xu, P. (2012) Carotenoids Play a Positive Role in the Degradation of Heterocycles by *Sphingobium yanoikuyae*. *PLoS One* **7**: e39522.

- Liu, X., Liang, M., Liu, Y., and Fan, X. (2017) Directed evolution and secretory expression of a pyrethroid-hydrolyzing esterase with enhanced catalytic activity and thermostability. *Microb. Cell Fact.* **16**: 81.
- Lucidi, M., Runci, F., Rampioni, G., Frangipani, E., Leoni, L., and Visca, P. (2018) New shuttle vectors for gene cloning and expression in multidrug-resistant *Acinetobacter* species. *Antimicrob. Agents Chemother.* **62**:
- Luo, L., Pappalardi, M.B., Tummino, P.J., Copeland, R.A., Fraser, M.E., Grzyska, P.K., and Hausinger, R.P. (2006) An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* **353**: 69–74.
- Luo, W., Zhu, X., Chen, W., Duan, Z., Wang, L., and Zhou, Y. (2014) Mechanisms and strategies of microbial cometabolism in the degradation of organic compounds - Chlorinated ethylenes as the model. *Water Sci. Technol.* **69**: 1971–1983.
- Mahendra, S. and Alvarez-Cohen, L. (2006) Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. *Environ. Sci. Technol.* **40**: 5435–5442.
- Maher, S., Jjunju, F.P.M., Damon, D.E., Gorton, H., Maher, Y.S., Syed, S.U., et al. (2016) Direct Analysis and Quantification of Metaldehyde in Water using Reactive Paper Spray Mass Spectrometry. *Sci. Rep.* **6**: 1–10.
- Mahmood, A.M. and Dunwell, J.M. (2020) 2-oxoglutarate-dependent dioxygenases: A renaissance in attention for ascorbic acid in plants. *PLoS One* **15**: e0242833.
- Marris, C. (2001) Public views on GMOs: Deconstructing the myths. Stakeholders in the GMO debate often describe public opinion as irrational. But do they really understand the public? *EMBO Rep.* **2**: 545–548.

- Martin, V.J.J. and Mohn, W.W. (1999) A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid- degrading bacterium *Pseudomonas abietaniphila* BKME-9. *J. Bacteriol.* **181**: 2675–2682.
- Martinez, S. and Hausinger, R.P. (2015) Catalytic mechanisms of Fe(II)- and 2-Oxoglutaratedependent oxygenases. *J. Biol. Chem.* **290**: 20702–20711.
- Matafonova, G. and Batoev, V. (2018) Recent advances in application of UV light-emitting diodes for degrading organic pollutants in water through advanced oxidation processes: A review. *Water Res.* **132**: 177–189.
- Meinhardt, F., Schaffrath, R., and Larsen, M. (1997) Microbial linear plasmids. *Appl. Microbiol. Biotechnol.* **47**: 329–336.
- Metaldehyde Stewardship Group (2020) Get Pelletwise! | PROMOTING RESPONSIBLE METALDEHYDE APPLICATIONS. *Promot. Responsible Met. Appl.*
- Min Huang, Oppermann, F.B., and Steinbüchel, A. (1994) Molecular characterization of the *Pseudomonas putida* 2,3-butanediol catabolic pathway. *FEMS Microbiol. Lett.* **124**: 141–150.
- Mineau, P., Baril, A., Collins, B.T., Duffe, J., Joerman, G., and Luttik, R. (2001) Pesticide acute toxicity reference values for birds. *Rev. Environ. Contam. Toxicol.* **170**: 13–74.
- Mohamad Ibrahim, I.H., Gilfoyle, L., Reynolds, R., and Voulvoulis, N. (2019) Integrated catchment management for reducing pesticide levels in water: Engaging with stakeholders in East Anglia to tackle metaldehyde. *Sci. Total Environ.* **656**: 1436–1447.
- Monard, C., Martin-Laurent, F., Devers-Lamrani, M., Lima, O., Vandenkoornhuyse, P., and Binet, F. (2010) atz gene expressions during atrazine degradation in the soil drilosphere. *Mol.*

Ecol. **19**: 749–59.

Monard, C., Martin-Laurent, F., Lima, O., Devers-Lamrani, M., and Binet, F. (2013) Estimating the biodegradation of pesticide in soils by monitoring pesticide-degrading gene expression.

Biodegradation **24**: 203–213.

Moreau, P., Burgeot, T., and Renault, T. (2015) In vivo effects of metaldehyde on Pacific oyster, *Crassostrea gigas*: comparing hemocyte parameters in two oyster families. *Environ. Sci.*

Pollut. Res. **22**: 8003–8009.

Morgenthaler, A.B., Kinney, W.R., Ebmeier, C.C., Walsh, C.M., Snyder, D.J., Cooper, V.S., et al. (2019) Mutations that improve efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. *Elife* **8**.

Müller, T.A., Fleischmann, T., Van Der Meer, J.R., and Kohler, H.P.E. (2006) Purification and characterization of two enantioselective α -ketoglutarate-dependent dioxygenases, RdpA and SdpA, from *Sphingomonas herbicidovorans* MH. *Appl. Environ. Microbiol.* **72**: 4853–4861.

Nabeerasool, M., Campen, A., Polya, D., Brown, N., and van Dongen, B. (2015) Removal of Metaldehyde from Water Using a Novel Coupled Adsorption and Electrochemical Destruction Technique. *Water* **7**: 3057–3071.

Nakashima, N. and Tamura, T. (2004) Isolation and characterization of a rolling-circle-type plasmid from *Rhodococcus erythropolis* and application of the plasmid to multiplerecombinant- protein expression. *Appl. Environ. Microbiol.* **70**: 5557–5568.

- Nanthini, J., Ong, S.Y., and Sudesh, K. (2017) Identification of three homologous latex-clearing protein (lcp) genes from the genome of *Streptomyces* sp. strain CFMR 7. *Gene* **628**: 146–155.
- Nicholls, C.J. (2014) Implications of not controlling slugs in oilseed rape and wheat in the UK. *Hgca* 4–9.
- Nordin, K., Unell, M., and Jansson, J.K. (2005) Novel 4-chlorophenol degradation gene cluster and degradation route via hydroxyquinol in *Arthrobacter chlorophenicus* A6. *Appl. Environ. Microbiol.* **71**: 6538–6544.
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., and Hase, T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **28**: e63.
- NRoSOS (2012) pellet stewardship. *Met. slug pellet Steward*.
- Ogram, A.V., Duan, Y.-P., Trabue, S.L., Feng, X., Castro, H., and Ou, L.-T. (2006) Carbofuran degradation mediated by three related plasmid systems. *FEMS Microbiol. Ecol.* **32**: 197–203.
- Owlstone (2015) Detection of Metaldehyde in Water Using a LONESTAR Portable Analyzer.
- Pan, X., Lei, B., Zhou, N., Feng, B., Yao, W., Zhao, X., et al. (2012) Identification of novel genes involved in DNA damage response by screening a genome-wide *Schizosaccharomyces pombe* deletion library. *BMC Genomics* **13**: 662.
- Pandeeti, E.V.P., Longkumer, T., Chakka, D., Muthyala, V.R., Parthasarathy, S., Madugundu, A.K., et al. (2012) Multiple mechanisms contribute to lateral transfer of an organophosphate degradation (opd) island in *sphingobium fuliginis* ATCC 27551. *G3 Genes, Genomes, Genet.* **2**: 1541–1554.

- Pankaj, Negi, G., Gangola, S., Khatri, P., Kumar, G., Srivastava, A., and Sharma, A. (2016) Differential expression and characterization of cypermethrin-degrading potential proteins in *Bacillus thuringiensis* strain, SG4. *3 Biotech* **6**.
- Paolo, M. and Renzo, B. (1983) Determination of metaldehyde in workroom air. *Bull. Environ. Contam. Toxicol.* **30**: 479–484.
- Pape, A. (2016) Smarter Water. *South Lake Style*.
- Parekh, N.R., Hartmann, A., Charnay, M.-P., and Fournier, J.-C. (1995) Diversity of carbofurandegrading soil bacteria and detection of plasmid-encoded sequences homologous to the *mcd* gene. *FEMS Microbiol. Ecol.* **17**: 149–160.
- Park, D., Lee, D.S., Kim, Y.M., and Park, J.M. (2008) Bioaugmentation of cyanide-degrading microorganisms in a full-scale cokes wastewater treatment facility. *Bioresour. Technol.* **99**: 2092–2096.
- Pauling, L. and Carpenter, D.C. (1936) The Crystal Structure of Metaldehyde. *J. Am. Chem. Soc.* **58**: 1274–1278.
- Pearson, W.R. (2013) An introduction to sequence similarity (“homology”) searching. *Curr. Protoc. Bioinforma.* **0 3**:
- Pendergrast, D. (2012) Briefing Briefing on Metaldehyde stewardship in autumn 2012. *NFU Brief.* 1–3.
- Pertea, M., Ayanbule, K., Smedinghoff, M., and Salzberg, S.L. (2009) OperonDB: A comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res.* **37**: D479–D482.

Phale, P.S., Shah, B.A., and Malhotra, H. (2019) Variability in assembly of degradation operons for naphthalene and its derivative, carbaryl, suggests mobilization through horizontal gene transfer.

Genes (Basel). **10**:

Pornsuwan, S., Maenpuen, S., Kamutira, P., Watthaisong, P., Thotsaporn, K., Tongsook, C., et al.

(2017) 3,4-Dihydroxyphenylacetate 2,3-dioxygenase from *Pseudomonas aeruginosa*: An Fe(II)-containing enzyme with fast turnover. *PLoS One* **12**: e0171135.

Prieto, A.I., Hernández, S.B., Cota, I., Pucciarelli, M.G., Orlov, Y., Ramos-Morales, F., et al. (2009)

Roles of the outer membrane protein *asmA* of *Salmonella enterica* in the control of *marRAB* expression and invasion of epithelial cells. *J. Bacteriol.* **191**: 3615–3622.

Priyadarshinee, R., Kumar, A., Mandal, T., and Dasguptamandal, D. (2016) Unleashing the

potential of ligninolytic bacterial contributions towards pulp and paper industry: key challenges and new insights. *Environ. Sci. Pollut. Res.* **23**: 23349–23368.

Providenti, M.A., Lee, H., and Trevors, J.T. (1993) Selected factors limiting the microbial

degradation of recalcitrant compounds. *J. Ind. Microbiol.* **12**: 379–395.

Qiu, J., Liu, B., Zhao, L., Zhang, Y., Cheng, D., Yan, X., et al. (2018) A novel degradation mechanism

for pyridine derivatives in *Alcaligenes faecalis* JQ135. *Appl. Environ. Microbiol.*

84:

Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., et al. (2013)

Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol.* **79**: 4551–4563.

Randelovic, A., Zhang, K., Jacimovic, N., McCarthy, D., and Deletic, A. (2016) Stormwater biofilter

treatment model (MPiRe) for selected micro-pollutants. *Water Res.* **89**: 180–191.

- Rodríguez-Salazar, J., Almeida-Juarez, A.G., Ornelas-Ocampo, K., Millán-López, S., Raga-Carbajal, E., Rodríguez-Mejía, J.L., et al. (2020) Characterization of a Novel Functional Trimeric Catechol 1,2-Dioxygenase From a *Pseudomonas stutzeri* Isolated From the Gulf of Mexico. *Front. Microbiol.* **11**: 1100.
- Roggo, C. and van der Meer, J.R. (2017) Miniaturized and integrated whole cell living bacterial sensors in field applicable autonomous devices. *Curr. Opin. Biotechnol.* **45**: 24–33.
- Rolph, C.A., Jefferson, B., Hassard, F., and Villa, R. (2018) Metaldehyde removal from drinking water by adsorption onto filtration media: Mechanisms and optimisation. *Environ. Sci. Water Res. Technol.* **4**: 1543–1552.
- De Roma, A., Miletti, G., D'Alessio, N., Rossini, C., Vangone, L., Galiero, G., and Esposito, M. (2017) Metaldehyde poisoning of companion animals: A three-year retrospective study. *J. Vet. Res.* **61**: 307–311.
- Ruan, Z., Zhai, Y., Song, J., Shi, Y., Li, K., Zhao, B., and Yan, Y. (2013) Molecular Cloning and Characterization of a Newly Isolated Pyrethroid-Degrading Esterase Gene from a Genomic Library of *Ochrobactrum anthropi* YZ-1. *PLoS One* **8**: e77329.
- Sadauskas, M., Vaitekunas, J., Gasparavičiute, R., and Meškys, R. (2017) Indole biodegradation in *Acinetobacter* sp. strain O153: Genetic and biochemical characterization. *Appl. Environ. Microbiol.* **83**.
- Sales, C.M., Grostern, A., Parales, J. V., Parales, R.E., and Alvarez-Cohen, L. (2013) Oxidation of the cyclic ethers 1,4-dioxane and tetrahydrofuran by a monooxygenase in two *Pseudonocardia* species. *Appl. Environ. Microbiol.* **79**: 7702–7708.

- Sarkar, J., Kazy, S.K., Gupta, A., Dutta, A., Mohapatra, B., Roy, A., et al. (2016) Biostimulation of indigenous microbial community for bioremediation of petroleum refinery sludge. *Front. Microbiol.* **7**: 1407.
- Schleinitz, K.M., Kleinsteuber, S., Vallaeyts, T., and Babel, W. (2004) Localization and characterization of two novel genes encoding stereospecific dioxygenases catalyzing 2(2,4dichlorophenoxy)propionate cleavage in *Delftia acidovorans* MC1. *Appl. Environ. Microbiol.* **70**: 5357–5365.
- Schumacher, M., Castle, G., Gravell, A., Mills, G.A., and Fones, G.R. (2016) An improved method for measuring metaldehyde in surface water using liquid chromatography tandem mass spectrometry. *MethodsX* **3**: 188–194.
- Sega, G.A. (1984) A review of the genetic effects of ethyl methanesulfonate. *Mutat. Res. Genet. Toxicol.* **134**: 113–142.
- Seifert, H.S., Chen, E.Y., So, M., and Heffron, F. (1986) Shuttle mutagenesis: A method of transposon mutagenesis for *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 735–739.
- Selim, S. and Seiber, J.N. (1973) Gas Chromatographic Method for the Analysis of Metaldehyde in Crop Tissue. *J. Agric. Food Chem.* **21**: 430–433.
- Semitsoglou-Tsiapou, S., Templeton, M.R., Graham, N.J.D., Hernández Leal, L., Martijn, B.J., Royce, A., and Kruithof, J.C. (2016) Low pressure UV/H₂O₂ treatment for the degradation of the pesticides metaldehyde, clopyralid and mecoprop - Kinetics and reaction product formation. *Water Res.* **91**: 285–294.

- Seung, I.K., Choi, J.S., and Kahng, H.Y. (2007) A proteomics strategy for the analysis of bacterial biodegradation pathways. *Omi. A J. Integr. Biol.* **11**: 280–294.
- Shibai, A., Takahashi, Y., Ishizawa, Y., Motooka, D., Nakamura, S., Ying, B.W., and Tsuru, S. (2017) Mutation accumulation under UV radiation in *Escherichia coli*. *Sci. Rep.* **7**: 1–12.
- Shukla, P.T. and Auerbach, C. (1981) Genetical tests for the frequency of small deletions among EMS-induced point mutations in *Drosophila*. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **83**: 81–89.
- Siddavattam, D., Khajamohiddin, S., Manavathi, B., Pakala, S.B., and Merrick, M. (2003) Transposon-like organization of the plasmid-borne organophosphate degradation (*opd*) gene cluster found in *Flavobacterium* sp. *Appl. Environ. Microbiol.* **69**: 2533–2539.
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**:
- Simms, L.C., Dawson, J.J.C., Paton, G.I., and Wilson, M.J. (2006) Identification of environmental factors limiting plant uptake of metaldehyde seed treatments under field conditions. *J. Agric. Food Chem.* **54**: 3646–3650.
- Singh, B.K. and Walker, A. (2006) Microbial degradation of organophosphorus compounds. *FEMS Microbiol. Rev.* **30**: 428–471.
- Smirnova, E., Marquis, V., Poirier, L., Aubert, Y., Zumsteg, J., Ménard, R., et al. (2017) Jasmonic Acid Oxidase 2 Hydroxylates Jasmonic Acid and Represses Basal Defense and Resistance Responses against *Botrytis cinerea* Infection. *Mol. Plant* **10**: 1159–1173.
- Sobecky, P.A. and Hazen, T.H. (2009) Horizontal gene transfer and mobile genetic elements in marine systems. *Methods Mol. Biol.* **532**: 435–453.

- Sotelo, J.L., Ovejero, G., Rodríguez, A., Álvarez, S., Galán, J., and García, J. (2014) Competitive adsorption studies of caffeine and diclofenac aqueous solutions by activated carbon. *Chem. Eng. J.* **240**: 443–453.
- Spain, J.C. and Van Veld, P.A. (1983) Adaptation of natural microbial communities to degradation of xenobiotic compounds: Effects of concentration, exposure time, inoculum, and chemical structure. *Appl. Environ. Microbiol.* **45**: 428–435.
- Speiser, B. and Kistler, C. (2002) Field tests with a molluscicide containing iron phosphate. *Crop Prot.* **21**: 389–394.
- Srnec, M., Wong, S.D., and Solomon, E.I. (2014) Excited state potential energy surfaces and their interactions in FeIVO active sites. *Dalt. Trans.* **43**: 17567–17577.
- Stalker, D.M. and McBride, K.E. (1987) Cloning and expression in *Escherichia coli* of a *Klebsiella ozaenae* plasmid-borne gene encoding a nitrilase specific for the herbicide bromoxynil. *J. Bacteriol.* **169**: 955–960.
- Steiner, R.A., Kalk, K.H., and Dijkstra, B.W. (2002) Anaerobic enzyme-substrate structures provide insight into the reaction mechanism of the copper-dependent quercetin 2,3-dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 16625–16630.
- Sultana, A., Kallio, P., Jansson, A., Wang, J.S., Niemi, J., Mäntsälä, P., and Schneider, G. (2004) Structure of the polyketide cyclase Snoal reveals a novel mechanism for enzymatic aldol condensation. *EMBO J.* **23**: 1911–1921.
- Sun, P., Tropea, J.E., and Waugh, D.S. (2011) Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner. *Methods Mol. Biol.* **705**: 259–274.

- Tao, B. and Fletcher, A.J. (2014) Catalytic degradation and adsorption of metaldehyde from drinking water by functionalized mesoporous silicas and ion-exchange resin. *Sep. Purif. Technol.* **124**: 195–200.
- Tao, B. and Fletcher, A.J. (2013) Metaldehyde removal from aqueous solution by adsorption and ion exchange mechanisms onto activated carbon and polymeric sorbents. *J. Hazard. Mater.* **244–245**: 240–250.
- Taylor Eighmy, T., Robin Collins, M., Spanos, S.K., and Fenstermacher, J. (1992) Microbial populations, activities and carbon metabolism in slow sand filters. *Water Res.* **26**: 1319–1328.
- Thakur, M., Medintz, I.L., and Walper, S.A. (2019) Enzymatic Bioremediation of Organophosphate Compounds—Progress and Remaining Challenges. *Front. Bioeng. Biotechnol.* **7**: 289.
- Thomas, J.C., Helgason, T., Sinclair, C.J., and Moir, J.W.B. (2017) Isolation and characterization of metaldehyde-degrading bacteria from domestic soils. *Microb. Biotechnol.* **10**: 1824–1829.
- Throne-Holst, M., Wentzel, A., Ellingsen, T.E., Kotlar, H.K., and Zotchev, S.B. (2007) Identification of novel genes involved in long-chain n-alkane degradation by *Acinetobacter* sp. strain DSM 17874. *Appl. Environ. Microbiol.* **73**: 3327–3332.
- Todd, P.A., Monti-Bragadin, C., and Glickman, B.W. (1979) MMS mutagenesis in strains of *Escherichia coli* carrying the R46 mutagenic enhancing plasmid: Phenotypic analysis of Arg⁺ revertants. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **62**: 227–237.

- Torgerson, D.G. and Singh, R.S. (2004) Rapid evolution through gene duplication and subfunctionalization of the testes-specific $\alpha 4$ proteasome subunits in drosophila. *Genetics* **168**: 1421–1432.
- Triebkorn, R., Christensen, K., and Heim, I. (1998) Effects of orally and dermally applied metaldehyde on mucus cells of slugs (*Deroceras reticulatum*) depending on temperature and duration of exposure. *J. Molluscan Stud.* **64**: 467–487.
- Tsuda, M., Tan, H.M., Nishi, A., and Furukawa, K. (1999) Mobile catabolic genes in bacteria. *J. Biosci. Bioeng.* **87**: 401–410.
- Ul Haque, M.F., Nadalig, T., Bringel, F., Schaller, H., and Vuilleumier, S. (2013) Fluorescencebased bacterial bioreporter for specific detection of methyl halide emissions in the environment. *Appl. Environ. Microbiol.* **79**: 6561–6567.
- UniProt (2009) yhfH - Uncharacterized protein YhfH - *Bacillus subtilis* (strain 168) - yhfH gene & protein. *UniProtKB*.
- Vallecillo, A.J., Parada, C., Morales, P., and Espitia, C. (2017) *Rhodococcus erythropolis* as a host for expression, secretion and glycosylation of *Mycobacterium tuberculosis* proteins. *Microb. Cell Fact.* **16**: 12.
- Vishniac, W. and Santer, M. (1957) the Thiobacilli, 12. *Bacteriol. Rev.* **21**: 195–213.
- Wagner, A. (2007) Rapid detection of positive selection in genes and genomes through variation clusters. *Genetics* **176**: 2451–2463.
- Wang, W., Wang, L., and Shao, Z. (2018) Polycyclic aromatic hydrocarbon (PAH) degradation pathways of the obligate marine PAH degrader *Cycloclasticus* sp. strain P1. *Appl. Environ. Microbiol.* **84**..

- Wang, Y., Zhang, B., Deng, T., and Li, F. (2017) Reclamation of EDTA by sodium tetraethylenepentamine-multi dithiocarbamate after soil washing process with EDTA. *Environ. Earth Sci.* **76**: 1–10.
- Waste Water Treatment Online (2016) Metaldehyde treatment would cost £600M, warns Anglian - WWT. *WWT*.
- Wei, Y., Jiao, Y., An, D., Li, D., Li, W., and Wei, Q. (2019) Review of dissolved oxygen detection technology: From laboratory analysis to online intelligent detection. *Sensors (Switzerland)* **19**.
- Welford, R.W.D., Kirkpatrick, J.M., McNeill, L.A., Puri, M., Oldham, N.J., and Schofield, C.J. (2005) Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett.* **579**: 5170–5174.
- Werner, J., Nour, E., Bunk, B., Spröer, C., Smalla, K., Springael, D., and Öztürk, B. (2020) PromA Plasmids Are Instrumental in the Dissemination of Linuron Catabolic Genes Between Different Genera. *Front. Microbiol.* **11**: 149.
- Whittle, G., Shoemaker, N.B., and Salyers, A.A. (2002) The role of Bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. *Cell. Mol. Life Sci.* **59**: 2044–2054.
- Widada, J., Nojiri, H., and Omori, T. (2002) Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation. *Appl. Microbiol. Biotechnol.* **60**: 45–59.

- Wu, L.F., Meng, S., and Tang, G.L. (2016) Ferrous iron and α -ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**: 453–470.
- Wu, M., Li, W., Dick, W.A., Ye, X., Chen, K., Kost, D., and Chen, L. (2017) Bioremediation of hydrocarbon degradation in a petroleum-contaminated soil and microbial population and activity determination. *Chemosphere* **169**: 124–130.
- Xu, P., Ma, W., Han, H., Jia, S., and Hou, B. (2015) Isolation of a naphthalene-degrading strain from activated sludge and bioaugmentation with it in a MBR treating coal gasification wastewater. *Bull. Environ. Contam. Toxicol.* **94**: 358–364.
- Yan, X., Gu, T., Yi, Z., Huang, J., Liu, X., Zhang, J., et al. (2016) Comparative genomic analysis of isoproturon-mineralizing sphingomonads reveals the isoproturon catabolic mechanism. *Environ. Microbiol.* **18**: 4888–4906.
- Yanagiya, K., Maejima, Y., Nakata, H., Tokuda, M., Moriuchi, R., Dohra, H., et al. (2018) Novel Self-Transmissible and Broad-Host-Range Plasmids Exogenously Captured From Anaerobic Granules or Cow Manure. *Front. Microbiol.* **9**: 2602.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**: 7–8.
- Yang, J., Yang, C., Jiang, H., and Qiao, C. (2008) Overexpression of methyl parathion hydrolase and its application in detoxification of organophosphates. *Biodegradation* **19**: 831–839.
- Yin, R., Mo, J., Dai, J., and Wang, H. (2017) Nickel(II) Inhibits Tet-Mediated 5-Methylcytosine Oxidation by High Affinity Displacement of the Cofactor Iron(II). *ACS Chem. Biol.* **12**: 1494–1498.

- Yoon, Y., Kim, S., Chae, Y., Kim, S.W., Kang, Y., An, G., et al. (2016) Simultaneous detection of bioavailable arsenic and cadmium in contaminated soils using dual-sensing bioreporters. *Appl. Microbiol. Biotechnol.* **100**: 3713–3722.
- Yoshimoto, T., Nagai, F., Fujimoto, J., Watanabe, K., Mizukoshi, H., Makino, T., et al. (2004) Degradation of estrogens by *Rhodococcus zopfii* and *Rhodococcus equi* isolates from activated sludge in wastewater treatment plants. *Appl. Environ. Microbiol.* **70**: 5283–5289.
- Yousef, Y.A., Hvitved-Jacobsen, T., Wanielista, M.P., and Harper, H.H. (1987) Removal of contaminants in highway runoff flowing through swales. *Sci. Total Environ.* **59**: 391–399.
- Yuan, H., Wu, J., Wang, X., Chen, J., Zhong, Y., Huang, Q., and Nan, P. (2017) Computational identification of amino-acid mutations that further improve the activity of a chalcone–flavonone isomerase from *Glycine max*. *Front. Plant Sci.* **8**.
- Yuan, X., Tian, G., Zhao, Y., Zhao, L., Wang, H., and Ng, T.B. (2016) Degradation of dyes using crude extract and a thermostable and pH-stable laccase isolated from *Pleurotus nebrodensis*. *Biosci. Rep.* **36**.
- Zechner, E.L., Lang, S., and Schildbach, J.F. (2012) Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. B Biol. Sci.* **367**: 1073–1087.
- Zhang, D., He, Y., Wang, Y., Wang, H., Wu, L., Aries, E., and Huang, W.E. (2012) Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. *Microb. Biotechnol.* **5**: 87–97.
- Zhang, Ji, Hong, Q., Li, Q., Li, C., Cao, L., Sun, J.Q., et al. (2012) Characterization of isoproturon biodegradation pathway in *Sphingobium* sp. YBL2. *Int. Biodeterior. Biodegrad.* **70**: 8–13.

- Zhang, Jun, Yin, J.G., Hang, B.J., Cai, S., He, J., Zhou, S.G., and Li, S.P. (2012) Cloning of a novel arylamidase gene from paracoccus sp. Strain FLN-7 that hydrolyzes amide pesticides. *Appl. Environ. Microbiol.* **78**: 4848–4855.
- Zhang, X., Huang, Y., Harvey, P.R., Li, H., Ren, Y., Li, J., et al. (2013) Isolation and Characterization of Carbendazim-degrading *Rhodococcus erythropolis* djl-11. *PLoS One* **8**: e74810.
- Zhang, X., Yang, Q., Zhang, Q., Jiang, X., Wang, X., Li, Y., et al. (2020) Rapid detection of cytochrome cd1-containing nitrite reductase encoding gene nirS of denitrifying bacteria with loop-mediated isothermal amplification assay. *Sci. Rep.* **10**: 1–10.
- Zhao, Z., Zhang, Y., Liu, X., Zhang, X., Liu, S., Yu, X., et al. (2013) A Role for a Dioxygenase in Auxin Metabolism and Reproductive Development in Rice. *Dev. Cell* **27**: 113–122.
- Zheng, Q., Wang, S., Duan, P., Liao, R., Chen, D., and Liu, W. (2016) An α/β -hydrolase fold protein in the biosynthesis of thiostrepton exhibits a dual activity for endopeptidyl hydrolysis and epoxide ring opening/macrocyclization. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 14318–14323.
- Zhu, W.L., Hu, S., Lv, C.J., Zhao, W.R., Wang, H.P., Mei, J.Q., et al. (2019) A single mutation increases the thermostability and activity of *aspergillus terreus* amine transaminase. *Molecules* **24**:

Appendix

MahX Protein Sequence

MKQELESAVNARVIESNKLDGSSSLIGQERQPDTDFYAALSDMGLGGHVAHLDQYGYCVPPSDFDDLGLTA
EAKRKVLEIAERRSGIRPDSETGATHSIGESAVGQCMHYLLFEDPVFEKMLIHPVVLAFHRYLLGRSGRLSAMS
AMLRGPGTPALATHADLVMVPPPWPMFAQVCNISWALTDYTKENGATAIVPGSHKLCRPPTDAEIADTSKL
IAVTAPAGSLVIWHGNTWHGSAKVPGLRMQIIMYMCRTHVMPQEWYLDKVTPEILQRNGTEFAEMLGI
GHPYFAPSGPDWQNVVRAFGQATTLNG*

MahS Protein Sequence

MSEVDTLVKQADIFATTPHAVPNPMDAINALGLSHRILELELNGYCVIPAEVVGADDLRRAREVLLRISEDRE
GVVPDTETGLTHAMEDRAFGQLMYQILLDDPVYEEILTHPVTLAFTRYLLGHSAKLSAMSGVLRGPGCPALPL
HADLVMVPPPPQFAQVSNATWILSDYSKDNATCFWPGSHKLCRRPNDLEIADISQFVPVEAPAGSLIVW
HGNTWHGSAKVPGLRMQLIMYFCRTYMFPQEWYDPKISQEALDRNGPDFAEMLYMNHAYFPATDEG
GPNWSKIASAFTRLVSLEA*

MahR Protein Sequence

MTARVATELHHVPSTTPIETIVEYLRRDGYVIVDNVSPETMDTVQSELAPYLDATPSGYNPMIGKKTRRTGA

LIARSPECRNLIQHPLTLGVARSLGHASAVQLNLTQVISIDPGESAQALHRDQSAWDFFPEDYQVQCNTL
WAMSDYTAEMGATRIVPGSHLPGSRPQNQYSDEECVQAEMERGSILLYSGKIVHSGGANRSVDTRAININ
YSVGWVRQEENQYLSVPLEVARTLDDPELLGLMGYQEGAFAMGYFRDFENPLRAVRGDDYVEYGFDESTLA
DSSSSFSYFISSK*

MahY Protein Sequence

MNKLHRVVAVKDLDEAAGRYERIFAVPFVRTGPYVASMGVKVAGAWGLGVELIQPMPGSDSQFAQDIQRHLNERG
EGLYGVVFQTRTMKSDIEHLEKNAFVAYGPTFSFSSSVLETEFGGAFSRFEETVFTPERLGYLVAAMDASPTR*

MahZ Protein Sequence

MDGDRGLTALLPSVKEFLGSPRKAFIDGKWVAAKSGKTFEVFNPATGTVIGHAAACEKADVDEAVKAARKA
FDTGPWTKMTPSERGRIIWKVGDLILKYTEELAQLESIDNGKPISIARAADVLAADIFHYMAGWATKIEGSSL
SLSVPYTPGVEYQAFTRKEPVGVVAQIIPWNFLLMAAWKLYLHSIHASHRQTQALQPPPPRTHSALPNRGR
ALRELAGAGLRGSI*

