MONASH University

Bacterial Methane Monooxygenase: from Molecular Evolution to Biotechnology

Craig David Osborne

B. Eng. (Chemical) (Hons) / B.Sci. (Chemistry, Biochemistry)

A thesis submitted for the degree of *Doctor of Philosophy*

Department of Chemical Engineering

Monash University, Clayton

Australia

December 2018

This page is intentionally blank

Copyright notice

Notice 1

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

This page is intentionally blank

Table of contents

List of figures

This page is intentionally blank

List of tables

This page is intentionally blank

Abbreviations

ATP	Adenosine triphosphate
AMO	Ammonia monooxygenase
AOB	Ammonia oxidising bacteria
BLASTp	Protein version of basic local alignment search tool
BMM	Bacterial multicomponent monooxygenase
bpd	Barrels per day
CuMMO	Copper membrane monooxygeanse
DME	Dimethyl ether
FAD	Flavin adenine dinucleotide
FCSR	Flavocytochrome c: sulphide dehydrogenase
GHG	Greenhouse gas
H_4F	Tetrahydrofolate
H_4MPT	Tetrahydromethanopterin
HAO	Hydroxylamine oxidase
HCP	Hybrid cluster protein
HGT	Horizontal gene transfer
ICM	Intracytoplasmic membrane
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilodaltons
MMO	Methane monooxygenase
MO	Monooxygenase
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
PBS	Phosphate buffered saline
pMMO	Particulate methane monooxygenase
pMMO2	Divergent particulate methane monooxygenase associated with low methane concentrations
ppm	Parts per million
PQQ	Pyrroloquinoline quinone
$PQQH_2$	Pyrroloquinoline quinol

pXMO	Uncharacterised membrane monooxygenase
RBS	Ribosomal binding site
RLP	RuBisCO like proteins
rpm	Revolutions per minute
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuMP	Ribulose monophosphate
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
sMMO	Soluble methane monooxygenase.
SQR	Sulphur quinone reductase
TIR	Translation initiation region

Abstract

Gas-to-liquid technologies involve the conversion of light hydrocarbon molecules such as methane to liquid alternatives, and has the potential to dramatically change the energy landscape, providing low or zero net-carbon fuels without requiring significant modifications to automotive and energy generation systems and infrastructure. A forerunner in this field involves methanol synthesis from methane, however, current cost and technology limitations restrict the viability of implementing this at small to medium scale which is required to utilise the majority of renewable and waste methane streams.

A potential solution to current production limitations is the use of alternative bioprocess technologies based on the soluble methane monooxygenase enzyme system from methanotrophs. Previous attempts at biotechnological application of this enzyme system faced major barriers either through native organism utilisation or heterologous expression; the former has been limited by difficulties with the use of the sMMO enzyme compromising the organism's primary metabolic pathway, and maintaining sufficient sMMO levels. The latter route has been hitherto unsuccessful in achieving biotechnologically relevant expression levels in highly tractable organisms.

In order to facilitate further development in this field, this thesis aimed to establish a deeper understanding of the soluble methane monooxygenase (sMMO) enzyme system, then use this knowledge to inform biotechnological applications. More specifically, it used bioinformatic approaches to probe the functional and evolutionary history of sMMO at the sequence, catalytic, (sub-)unit, substrate and organism level. This information was then used to further develop sMMO for biotechnological applications with either the native organism or purified enzyme complex.

Through combining molecular and taxonomic phylogenetic analysis, the first of three experimental chapters in this thesis demonstrated that the proteobacterial methanotrophs arose from the simultaneous horizontal gene transfer of three different methane oxidising systems, and that the ancestors to these enzyme systems were associated before they developed activity toward methane. Further, the evolutionary links between the methanotrophs to chemolithotrophic bacteria prompted a gene survey among methanotrophs for potential activity towards reduced gases, and revealed a previously unknown

capacity (at the genomic level). This has implications around methanotroph ecology, as well as biotechnological applications using non-pure methane feeds. The evolution of the sMMO enzyme system from the larger Bacterial Multicomponent Monooxygenase (BMM) family was investigated via bioinformatics in the second research chapter. This analysis revealed the primary evolutionary driver for methane activity in this family was an increase in oxidative power of the active centre with sMMO being the culmination of activity that developed stepwise from alkenes and longer chain alkanes. Such active site development had a destabilising effect on the catalytic centre, however, and methane activity evolved at the time of appearance of structural and assembly proteins which we postulate are present to counteract such destabilisation. The knowledge revealed by the bioinformatics analysis of sMMO was employed to design a targeted heterologous expression approach for the sMMO system. This resulted in high level expression of all required components of sMMO in *E. coli*, with tentative evidence of *in vivo* activity of the reconstituted system.

In summary, this PhD took a holistic approach to the development of sMMO for biotechnological applications. By gaining a deeper understanding of the evolutionary history of both sMMO and the methanotrophs themselves, it was able to provide significant insights into important prerequisites both of the organism and the protein complex that enabled methane activity. These insights then informed efforts towards heterologous expression attempts with sMMO, producing tentative positive results for reconstituted enzyme activity in *E. coli*.

List of publications

Published Journal Articles (included in this thesis)

1. Osborne CD, Haritos VS. (2018). Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in Proteobacteria. Molecular Phylogenetics and Evolution, Vol, page.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, expect where due reference is made in the text of the thesis.

This thesis includes *1* original papers published in peer reviewed journals. The core theme of the thesis is **"Bacterial Methane Monooxygenase: from Molecular Evolution to Biotechnology"**. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Chemical Engineering under the supervision of Associate Prof. Victoria Haritos and Associate Prof. Ashley Buckle.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of *chapter 3* my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status *(published, in press, accepted or returned for revision, submitted)*	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
3	Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria	Published	80% - Idea, computational work, analysis and write-up	1. Victoria Haritos 20% - Idea, reviewing and editing of paper	Monash staff

** If no co-authors, leave fields blank*

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: [redacted] **Date: 20/12/2018**

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: [redacted] **Date: 20/12/2018**

Acknowledgements:

First and foremost, I would like to thank my primary project supervisor Associate Professor Victoria Haritos for the endless hours of discussion, ideas and continued steady hand navigating me through this process in all respects. I think it goes without saying that I couldn't have done it without you. To all of the members of the Haritos group I am also deeply appreciative, but especially Melissa Mace and Huadong Peng, both of who started this journey at the same time as me and have shared in so many of its highs and lows. I would further like to thank my co-supervisor Associate Professor Ashley Buckle, as well as all members of his group for the many contributions and suggestions during group meetings: a particular thanks to Dr. David Hoke who contributed many hours in the lab supporting this project.

The assistance from Dr Chris Greening and the Greening group, especially Tent Jirapanjawat, has been invaluable, without which methanotroph growth during this project would not have been possible. I would also like to acknowledge Dr. David Steer from the Monash Biomedical Proteomics facility for aiding with protein mass spectrometry work. Thank you to my thesis committee examiners Dr. Lizhong He and Dr. Tom Peat, as well as the multiple chairpersons, for taking the time to provide both insights and suggestions which have been of significant benefit to the project. I also wish to acknowledge the financial support of Monash University and the Australian Government, who provided the Seed Funding and Australian Postgraduate Award respectively that made this project possible.

Thank you to the Monash Chemical Engineering Department and all the staff that constitute it, both the academics and administrative staff, of which have contributed in so many different ways, both large and small. Also thank you to the BEL Lab managers for the often thankless work they have done in keeping the lab running, as well as all other members of the BEL Lab for the countless hours of conversation and light hearted banter which made so many a long hour in the lab bearable.

To all Family and friends, who have stood by me through this long and sometimes difficult process. For all your love and support over all the years I am eternally grateful. Finally, I would like to thank God, for the countless blessings in my life that has made this all possible.

This page is intentionally blank

Dedication:

I dedicate this work to my Father.

Though you have said multiple times since your diagnosis that one of your biggest desires was to be able to see me graduate, for a while we did not even know if you would make it to this day: the day of my thesis submission. God willing, you will be able to see me graduate; but even if not, know you have been with me each step of the journey, and will also be there with me on that day as well.

For raising me and guiding me to the person I am today is a debt that I can never repay. Though I may not have always been the best at putting it into words, know that I love you always.

This page is intentionally blank

CHAPTER 1

INTRODUCTION

This page is intentionally blank

1.1. Background

Despite the majority of debate around climate change focusing on the reduction of carbon dioxide emissions, alternative terminology such as zero carbon emissions or climate neutrality are increasingly being used to acknowledge that mitigation of global warming will also require action around non-carbon dioxide greenhouse gasses (Rogelj et al. 2015). Of these, methane is the most significant, accounting for 17% of the current increase to global warming potential, with the major anthropogenic sources being landfills, manure management, livestock, and the extraction and processing of fossil fuels (Allen 2016).

Although methane is a major contributor to global warming, it also has the potential to play a major role in climate change mitigation. The high energy content of methane results in its utilisation as an energy commodity in the current fossil fuel economy. Unlike the other fossil fuels however, biologically-derived methane can also be readily produced with near zero net carbon dioxide emissions (Budzianowski and Postawa 2017). This is currently performed using anaerobic digesters which are capable of utilising diverse biomass materials: the ability to readily digest cellulose with minimal pre-treatment providing a marked advantage to cellulosic bioethanol (Golkowska and Greger 2013). In addition to this, technologies are already being commercialised for methane production from surplus power generation and atmospheric carbon dioxide, therefore facilitating the integration of fluctuating renewable electricity into present power grids (www.electrochaea.com).

Although methane is usable for many applications in its immediate form, its versatility is dramatically increased through chemical conversion to methanol. Examples include compatibility with present day flexi-fuel cars (Zhen and Wang 2015) and use as a diesel substitute through dehydration to dimethyl ether (DME), the latter already commercially viable with growing markets in China and India (Arcoumanis et al. 2008; Yang and Jackson 2012). The potential utility of methanol is so great that it has been proposed as the primary future universal energy commodity and industrial feedstock (Olah 2009). Current production technology that employs multi-stage reactors at high temperatures and pressures (Khoshtinat et al. 2010), however, have resulted in its economic viability only on large scales, with typical new gas-to-liquid processes targeting in excess of 50,000 bpd and priced at around $20 billion dollars (Haynes and Gonzalez 2014). This barrier means that a scalable and efficient method of

methanol production with low complexity as a 'Holy Grail' for the chemical industry (Van Beilen and Funhoff 2005; Horn and Schlogl 2014).

A possible solution to this problem exists in the form of the methanotrophs, a group of microorganisms that metabolise methane; or more specifically the systems that performs the first step in their primary metabolic pathway. This is performed by one of several enzymes, all of which are capable of converting methane to methanol with high specificity. Of these, soluble methane monooxygenase (sMMO) remains the most promising candidate for application in an industrial bioprocess, especially when using cell-free technology. Adaptation of this enzyme for use in an industrial process would therefore lead to a far more scalable system, enabling methanol to be produced not only on large scales, but also the critical small and medium scales (Park and Lee 2013).

Although significant effort has been invested in bioprocesses utilising native methanotrophic organisms in industrial and environmental applications, with examples of commercial scale production (calysta.com), progress remains limited and severely constrained by the nature of the native organisms (Strong et al. 2015). Alternative approaches such as heterologous expression in host cells or purified enzyme-based systems are potentially more amenable to many biotechnological applications, however, development in these areas has been stymied due to a lack of recombinant protein expression in tractable hosts. Both approaches would benefit significantly from resolution of numerous outstanding questions surrounding methanotrophs, methanotrophy and the methane monooxygenases and open the field of applications including bioremediation, fine chemical production and reduction in GHG emissions.

1.2. Scope of research

This research mainly involved an evolutionarily-informed approach to heterologous production of functional soluble methane monooxygenase; the thesis consists of major components of both bioinformatic analysis and recombinant protein expression design and experimentation. Research involving native methanotroph bacteria was limited to sourcing genes *via* open source databases and to facilitate development of an assay for detection of sMMO activity. Environmental sampling and growth experiments of methanotrophs were beyond the remit of this work.

The bioinformatic components of the thesis that examined the evolution of methanotrophy involved both the soluble and particulate monooxygenases, as well as the aerobic methanotrophic organisms themselves. Morphological and biochemical characteristics of these organisms such as intracytoplasmic membranes and other metabolic enzymes were also covered, however, only to an extent that they informed the evolution of methanotrophy in bacteria, and not as an independent analysis in themselves. Bioinformatic investigation of methanotroph genomes focussed on genes potentially involved in the metabolism of major gasses and did not cover all potential gasses, such as methylamine and carbon monoxide. The presence of both forms of the immediate metabolically downstream enzyme of the methane monooxygenases, methanol dehydrogenase, was also evaluated in these genomes, however the remainder of the genes involved in metabolic pathways for carbon assimilation and/or energy generation were beyond the scope of this thesis.

Heterologous expression experiments were targeted to the production of active sMMO in *E. coli*, with all genes selected based on bioinformatic analysis (as above) which provided insight into their proposed role in facilitating functional expression. Genes assessed for heterologous expression were sourced from *Methylococcus capsulatus (Bath)* solely. Numerous strategies were explored during the design, gene synthesis and expression stage in order to increase the likelihood of successful heterologous expression. This included trials with varying combinations of recombinant genes and expression conditions/methodology. During activity assays, only indirect methods of activity measurement were investigated, *vis a vis* the naphthalene assay, with direct assessment of methane to methanol conversion outside the present scope.

1.3. Research Objectives

The soluble methane monooxygenases present an excellent opportunity for biotechnology but have resisted prior attempts of functional expression in more tractable organisms. This project aimed to achieve heterologous expression and assembly of the multiple subunits of the enzyme in *Escherichia coli*, taking a distinctive approach that was informed by an evolutionary perspective on the development of this enzyme system. Specifically, the project aimed to:

- Understand the evolutionary origin of aerobic methane metabolising ability in bacteria through investigation of inheritance of the gene/protein systems
- Recapitulate the development of methane oxidation capability in the BMM family of monooxygenases via phylogenetic and sequence-structure-function analysis.
- Devise and employ a strategy for sMMO heterologous expression and assembly informed by both evolutionary and functional understanding.

1.4. Outline of thesis

Chapter 1: Introduction

It is intended to provide both the background and an overview of the drivers of the project. It clearly outlines the problem statement as well as project rationale and objectives. Further, the section considers biotechnological applications of methanotrophs and specifically sMMO in regards to greenhouse gas mitigation and production of bio-methanol.

Chapter 2: Review of Literature

The chapter provides a critical review of methanotrophy with a focus on soluble methane monooxygenase (sMMO). It begins by providing a broad introduction to the methanotrophs, including environmental and taxonomic distribution, metabolic pathways, and the important role they play in regulating atmospheric methane concentrations. It then progresses to a general overview of particulate methane monooxygenase and regulation of methane monooxygenase systems, before providing a detailed review on sMMO. The latter section makes up the majority of the report and it details the catalytic activity of sMMO from multiple perspectives including enzyme substrate promiscuity and the larger bacterial multicomponent monooxygenase family in which it resides. The review concludes with a summary of prior attempts at sMMO recombinant expression, highlighting the potential importance of operon associated non-catalytic components with particular reference to chaperonin analogues.

Chapter 3: The role of sMMO in methanotroph evolution and the implications for contemporary enzyme activities

In this chapter a bioinformatic investigation centred on molecular and taxonomic phylogenetic analysis showed for the first time that particulate and soluble monooxygenases entered ancestral species by multiple instances of simultaneous horizontal gene transfer. Common pre-existing physiological and metabolic attributes that likely supported conversion to methanotrophy are also identified. It provides evidence that prior to these enzyme systems developing methane oxidation capabilities, the membrane-bound and cytoplasmic monooxygenases were already both functionally and phylogenetically associated. The identification of an evolutionary association with species that were active towards hydrogen sulphide, short chain alkanes and ammonia also led to a genetic inventory being taken of reduced gas enzyme capability in methanotrophs and revealed the previously unknown extent of processing capability in some representatives towards all major reduced gasses.

Chapter 4: Linking enzyme function to evolution in protein sequence, cofactor presence and active site modifications in the BMM family

This chapter investigated the evolution of the BMM family, which includes sMMO, at three different bioinformatic levels: phylogenetic, component and sequence. The work reveals a progressive increase in oxidative potential in multiple branches of the evolutionary history of the BMM family, correlating with activity on progressively higher energy bonds. A strong correlation between C-H bond breaking reactivity and presence of assembly and stabilisation factors in the operon/genome was also found, with the development path towards methane coinciding with the presence of the highest number of these factors. It is hypothesised that these assembly and stabilisation factors are required to support an increasingly strained protein structure that is required to catalyse the most energetically demanding oxidations. Finally, a highly conserved sequence level change that occurred simultaneously with appearance of methane oxidation capability is identified, directly impacting the enzyme's active site and is postulated to be the primary evolutionary step through which the family has developed methane reactivity.

Chapter 5: Heterologous sMMO expression: combining components that evolved to develop methane activity

This chapter builds on the previous chapter's learnings to formulate a highly targeted approach for heterologous expression of the sMMO system in *E. coli*. This chapter presents the first known high level heterologous expression of all components required for the synthesis of an active methane monooxygenase system, as well as expression of the additional putative folding enzymes identified in chapter 4. It also achieved soluble expression of all units related to either folding or activity of the hydroxylase, as well as two out of the three subunits of the hydroxylase itself, all simultaneously. Naphthalene assay trials, though inconclusive due to repeatability issues, demonstrated results most parsimoniously explained through heterologous sMMO activity *in vivo*. Issues remain in the soluble expression of the final hydroxylase subunit, however, this work provides many promising directions for future research towards this goal.

Chapter 6: Conclusion and Future work

The chapter provides a concise summary of the work presented in the thesis, as well as providing a perspective into its significance in a wider research context. Further, the potential future direction of research aimed towards the achievement of the explicit goals of this project, as well as additional areas of high significance which this work has provided new insights to is elaborated.

1.5. References

Allen D. (2016) Attributing Atmospheric Methane to Anthropogenic Emission Sources. *Acc. Chem. Res.* 49, 1344–1350.

Arcoumanis C, Bae C, Crookes R, Kinoshita E. (2008) The potential of di-methyl ether (DME) as an alternative fuel for compression-ignition engines: A review. *Fuel* 87, 1014–1030.

Van Beilen JB, Funhoff EG. (2005) Expanding the alkane oxygenase toolbox: New enzymes and applications. *Curr. Opin. Biotechnol.* 16, 308–314.

Budzianowski WM, Postawa K. (2017) Renewable energy from biogas with reduced carbon dioxide footprint: Implications of applying different plant configurations and operating pressures. *Renew. Sustain. Energy Rev.* 68, 852–868.

Golkowska K, Greger M. (2013) Anaerobic digestion of maize and cellulose under thermophilic and mesophilic conditions - A comparative study. *Biomass and Bioenergy* 56, 545–554.

Haynes CA, Gonzalez R. (2014) Rethinking biological activation of methane and conversion to liquid fuels. *Nat. Chem. Biol.* 10, 331–339.

Horn R, Schlogl R. (2014) Methane Activation by Heterogeneous Catalysis. *Catal. Letters*, 23–39.

Khoshtinat M, Amin N a S, Noshadi I. (2010) A Review of Methanol Production from Methane Oxidation via Non-Thermal Plasma Reactor. *World Acad. Sci. Eng. Technol.* 4, 354–358.

Olah GA. (2009) Beyond oil and gas : the methanol economy. 2nd update. (Goeppert 1974- A, Prakash GKS, Service) WI (Online, editors.). Weinheim: Weinheim : Wiley-VCH

Park D, Lee J. (2013) Biological conversion of methane to methanol. *Korean J. Chem. Eng.* 30, 977–987.

Rogelj J, Schaeffer M, Meinshausen M, Knutti R, Alcamo J, Riahi K, Hare W. (2015) Zero emission targets as long-term global goals for climate protection. *Environ. Res. Lett.* 10.

Strong PJ, Xie S, Clarke WP. (2015) Methane as a resource: Can the methanotrophs add value? *Environ. Sci. Technol.* 49, 4001–4018.

Yang CJ, Jackson RB. (2012) China's growing methanol economy and its implications for energy and the environment. *Energy Policy* 41, 878–884.

Zhen X, Wang Y. (2015) An overview of methanol as an internal combustion engine fuel. *Renew. Sustain. Energy Rev.* 52, 477–493.

This page is intentionally blank

CHAPTER 2

LITERATURE REVIEW

This page is intentionally blank

2.1. The methanotrophic bacteria

The methanotrophs are a diverse grouping of microorganisms that have evolved to occupy a unique and highly specific position in the biosphere: utilising methane as both a carbon and energy source (Kalyuzhnaya et al. 2015). Members displaying this lifestyle appear in both the bacterial and archaeal kingdoms, with specific members capable of coupling methane oxidation not only to oxygen, but sulphate, iron, manganese, nitrate and nitrite reduction (Bhattarai et al. 2017). Methanotrophy has classically been differentiated based on its location in either oxic or anoxic environments, with the former and latter allocated to bacterial and archaeal methanotrophs respectively. This differentiation was supported by the requirement of diatomic oxygen by the methane oxidising systems of the bacterial methanotrophs (Ross and Rosenzweig 2017), and the highly oxygen sensitive nature of the methane activation system utilised by archaeal methanotrophs (Cedervall et al. 2010). The recent discovery of *Methylomirabilis* bacterial methanotrophs of the *NC10* phylum do not strictly obey this rule however (Wu et al. 2011), as they use nitrate reduction to generate oxygen 'intra-aerobically' for their methane monooxygenase system in anaerobic environments. Currently this appears to be only a single instance of an exception to this rule though, and thus does not yet necessitate a change in the fundamental paradigm.

The majority of bacterial methanotrophs are known to typically reside in terrestrial, aquatic and marine habitats on the border between oxic and anoxic environments where both oxygen and methane can be readily accessed (Knief 2015). It was previously believed that all were obligate methanotrophs, incapable of metabolising carbon sources containing C-C bonds (Smith and Murrell 2009). Recent discoveries however, in the form of several facultative bacterial methanotrophs capable of growing on varied substrates including acetate and ethanol (Dedysh et al. 2005; Theisen et al. 2005), have also provided exceptions to a former methanotrophy rule, though once again these are currently the only known exceptions.

Amongst the aerobic methanotrophs, thermophilic (growth over 40^{o}C), psychrophilic (growth below 15^{o}C), acidophilic (growth below pH 5), alkaliphilic (growth above pH 9) and halophilic (NaCl concentrations greater than 0.2M) traits are frequent (Smith and Murrell 2009; Semrau et al. 2010), with

extremes including growth at a pH of 0.5 (van Teeseling et al. 2014) as well as 15% NaCl (Heyer et al. 2005) identified. Such environmental flexibility enables aerobic methanotrophs to grow in virtually all biomes known to be methanogenic including swamps and fens, rice paddies, oceans, hydrothermal vents, permafrost and arctic saline lakes (Hanson and Hanson 1996; Bowman et al. 1997; Hirayama et al. 2013; Park and Lee 2013; Singleton et al. 2018). There is even demonstration of them forming endosymbiotic relationships with both plant and animal species (Raghoebarsing et al. 2005; Petersen and Dubilier 2009).

By consuming the methane generated in these environments, they have been noted to play a vital role in moderating atmospheric levels of methane (Semrau et al. 2011), with typical estimates indicating prevention of 80% or more of methane generated in many high emission biomes from reaching the atmosphere (Conrad and Rothfuss 1991; Hinrichs and Boetius 2003; W S Reeburgh 2007). This equates to the overall mitigation of aerobic and aerobic methanotrophs in the order of 0.7 billion tonnes per annum (William S Reeburgh 2007) of a greenhouse gas 25 times more potent than CO_2 (Dalton 2005). These values could even be a significant underestimate, with recent discoveries having identified methanogenesis occurring in oxic conditions, countering the former paradigm that methanogenesis only occurred under anoxia and also resolving the 'methane paradox' (Repeta et al. 2016). The primary mechanisms identified enabling methanogenesis in oxic environments is breakdown of methylphosphonate in marine environments (Karl et al. 2008) and acetoclastic activity in oxic soils (Angle et al. 2017). These have both been shown to significantly contribute to methane flux in these environments, the latter estimated to contribute up to 80% of the methane flux in wetland soils (Angle et al. 2017).

Further to this role mitigating initial methane release into the atmosphere, methanotrophic oxidation of atmospheric methane has also been identified, typically in upland soil locations (Bull et al. 2000). While the dominant methanotrophs in these biomes largely remain uncultured (Knief et al. 2003), their existence significantly expands the global methanotroph distribution into regions which are not net methane emitting. Estimates indicate the consumption of 5-10% of atmospheric methane through this process (Le Mer and Roger 2001), thus making methanotrophs not only indispensable in mitigating

methane emissions from terrestrial and marine biomes, but also a major player in its elimination once it has reached the atmosphere.

2.2. Methanotroph classification

From its earliest implementation almost half a century ago, the primary groupings for methanotroph classification system has remained largely unchanged despite the method for defining these categories undergoing successive iterations. This identifies two fundamental 'types', designated type I and type II: these initially differentiated based on intracytoplasmic membrane structure (Davies and Whittenbury 1970). The subsequent decades saw this classification system expanded to also incorporate fatty acid composition and C1 assimilation pathways, increasing the robustness of the original classification (though the latter did briefly result in the addition of a type X before its reassimilation into the type I (Hanson and Hanson 1996)). The advent and subsequent development of 16S taxonomy has now resulted in it becoming the dominant method for methanotroph classification, with type I and type II now clearly defined as the gammaproteobacterial and alphaproteobacterial methanotrophs respectively. Recent discoveries of non-proteobacterial methanotrophic species have necessitated the only major changes to this system to date, with the addition of a type III group to accommodate *Methylacidiphilum* species from the *Verrucomicrobia* phylum (Knief 2015). This may soon be further revised however to also accommodate 'intra-aerobic' members of the NC10 phylum (Ettwig et al. 2010) not yet incorporated into the current system. Secondary classification systems have also been recently proposed to further subdivide the type I and type II methanotrophs. No definitive conclusion has been reached however as to where and how such further subdivision should be made, with multiple proposals existing in the literature.

Biological methane metabolism for both biomass and energy can be divided into two components: the oxidation of methane to methanol; and the subsequent incorporation of methanol into the biological metabolic pathways. Whilst the former of these is unique to the methanotrophs and is the focus of much of this review, the latter not just occurs in methanotrophs but is also performed by many members of a much larger family of organisms known as the methylotrophs. Whilst the methylotroph family includes

methanotrophs under its strict definition of the ability to use reduced carbon substrates lacking the presence of carbon-carbon bonds as its sole source of carbon and energy (Chistoserdova et al. 2009), it also includes other members including those specialised towards growth on environmental concentrations of methanol whilst being unreactive to methane (Anthony 1982). As the majority of environmental methanol occurs as a by-product of plant and algal metabolism, predominantly the breakdown of pectin (Kist and Tate 2013), the majority of these specialised methanol oxidisers have developed close associations with these members (Sy et al. 2005). Some have been noted to have further developed this relationship with plants to the extent of true mutualistic symbiosis (Trotsenko et al. 2001), even to the extent of symbiotic incorporation into plant root tissue in a similar manner to root nodule symbioses by nitrogen fixing rhizobia (Sy et al. 2001).

The methanol incorporation pathway used by these methylotrophs as well as methanotrophs can be divided into two stages: oxidation of methanol to formaldehyde; then subsequent metabolic processes for attainment of energy from formaldehyde and its incorporation into biomass (figure 1). It has been noted previously that the process of methanol metabolism is 'modularised', with certain species containing different combinations of these modules to perform the same primary biological outcome (Chistoserdova 2011). For the methanol to formaldehyde conversion, the two known primary modules are the MxaF and XoxF enzymes, both of which are located periplasmically (Chu et al. 2016). Most species that oxidise methanol contain both of these modules, with MxaF previously believed to be the primary oxidant used by both methanotrophs and methylotrophs (Skovran et al. 2011). Relatively recently however it has been identified that XoxF contains a lanthanide catalytic centre (Pol et al. 2014), and supplementation of growth media with either lanthanum or cerium in some species results in XoxF becoming the primary methanol oxidant (Chu and Lidstrom 2016). Furthermore, there is also building evidence for XoxF being the primary environmental oxidant of methanol in key biomes (Vorobev et al. 2013; Chu and Lidstrom 2016). The implications of these findings is still being resolved, and it is uncertain how much this interchange effects the remainder of the C1 processing pathway that to date has almost solely been resolved with MxaF as the methanol oxidiser.

The remainder of the C1 metabolising pathway is much more diverse and complex, however can be divided into three main groups: those pathways that oxidise formaldehyde all the way down to carbon dioxide before reincorporation, those that oxidise down to formic acid, and those that directly incorporate formaldehyde (Chistoserdova et al. 2005). Members selecting the first option are the least common of the methanotrophs, with only non-proteobacterial members proposed to use this alternative (Khadem et al. 2011; Wu et al. 2011). Two different pathways have been implicated for the initial oxidation of formaldehyde using this alternative, both of which utilise C1 carriers (Chistoserdova et al. 2005). The first of these is the tetrahydrofolate (H_4F) pathway, components of which are widely distributed in biology and also crucial for many key biosynthetic pathways (Hitchings 1983). The alternative is the tetrahydromethanopterin (H4MPT) pathway, also being central in methanogenesis and archaeal methanotrophy (Ferry 1999; McGlynn 2017). Both carriers then release the C1 unit as formic acid, with final oxidation to carbon dioxide by one of multiple formate dehydrogenases (Hou et al. 2008; Versantvoort et al. 2018). Energy can then be generated through transfer of generated reducing

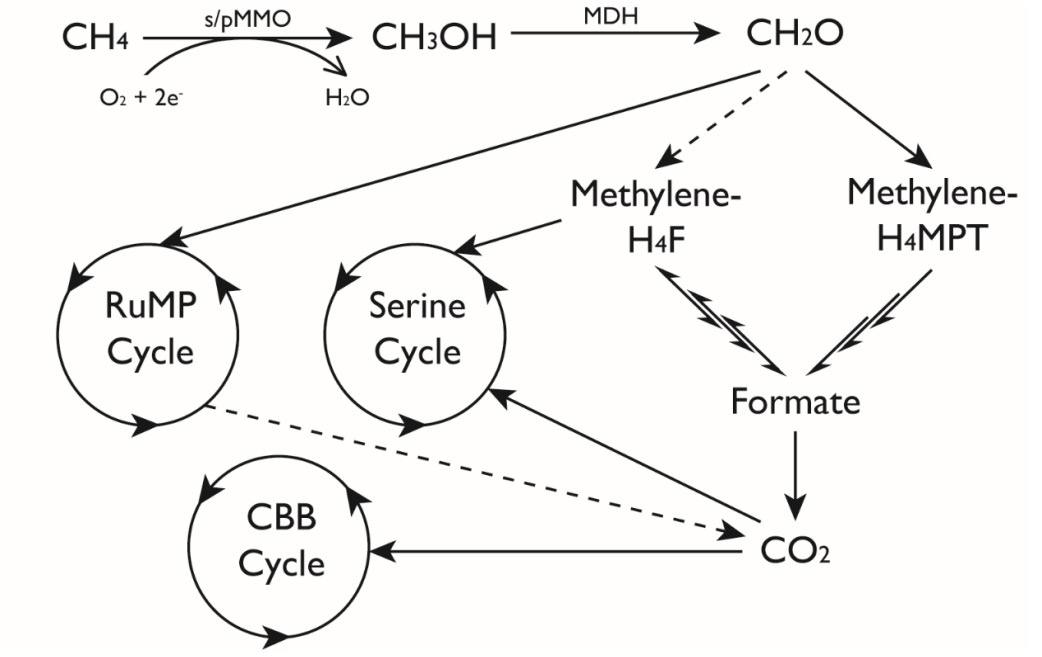


Figure 2.1: Metabolic pathway of methanotrophs. Identified methanotrophic pathways found in aerobic bacteria. Carbon assimilation occurs through one of the three cycles denoted in circles. Dashed lines indicate secondary or non-dominant pathways. H_4F, tetrahydrofolate; H_4MPT tetrahydromethanopterin; CBB cycle, Calvin-Benson-Bassham carbon fixation cycle; RuMP cycle, ribulose monophosphate formaldehyde assimilation cycle; MDH, methanol dehydrogenase; s/pMMO, either soluble or particulate methane monooxygenase.

equivalents to the electron transport chain, and carbon reincorporated through the CBB cycle (Khadem et al. 2011; Rasigraf et al. 2014).

Whilst this first option does not involve a unique method for the generation of C-C bonds using the formaldehyde derived C1 unit, such activity is central to the alternative two major pathways: the serine and RuMP cycle. For the serine cycle, this incorporation uses serine hydroxylmethyltransferase to synthesise its namesake from glycine and methylene-H_4F (Miyata et al. 1993). For each formaldehyde derived C1 unit incorporated in the cycle, a CO_2 unit is also incorporated using phosphoenolpyruvate carboxylase (Chistoserdova and Lidstrom 2013). Two carbon units are therefore incorporated during each turn of the cycle, these subsequently being spun off as acetyl-CoA (Korotkova et al. 2002). Though it was originally believed that methylene-H_4F was produced directly through interaction of formaldehyde and H_4F, it has recently been revealed that the C1 unit is actually bound to H_4F at the formic acid level, with subsequent reduction to generate the required methylene-H_4F (Crowther et al. 2008). The formic acid itself is generated through oxidation in the H_4MPT pathway, thus also placing this pathway of central importance to the serine cycle (Chistoserdova et al. 2009). This finding therefore means that the serine cycle is now placed in the category involving oxidation down to the formic acid level before subsequent incorporation.

The last C1 incorporation pathway, the RuMP cycle, is the only member to directly incorporate the C1 unit at the formaldehyde level. This is performed through its direct addition to ribolose-5-monophosphate (RuMP) using hexulosephosphate synthase, producing a C6 sugar (Kato et al. 1978). This is very closely analogous to the action of RuBisCO in the CBB pathway (Michelet et al. 2013), and indeed the two pathways share most of the regeneration enzymes in common. Furthermore, the regeneration enzymes are also shared with the pentose-phosphate pathway (Orita et al. 2006), with only the requirement of two additional enzymes, hexulosephosphate isomerase and the formerly mentioned hexulosephosphate synthase, to integrate the two. This enzymatic simplicity has resulted in the use of this pathway in, to the authors knowledge, all known attempts to create *de novo* methylotrophs (He et al. 2018; Meyer et al. 2018).

It has been noted that for the two bacterial classes in which both methanotrophs and non-methanotrophic methylotrophs appear, the primary C1 incorporation pathway is consistent within the class (Chistoserdova and Lidstrom 2013): the ribulose monophosphate (RuMP) pathway is dominant for both methane oxidising and non-methane oxidising gammaproteobacterial methylotrophs alike (exemplified by Methylophaga (Janvier et al. 1985; Villeneuve et al. 2013) and type I methanotrophs); whilst the same can be said for the serine pathway in the alphaproteobacterial methylotrophs (exemplified by Methylobacterium (Vuilleumier et al. 2009) and type II methanotrophs). Interestingly, this correlation of dominant C1 pathways with bacterial classes does not extend to the betaproteobacteria, which demonstrates both RuMP and serine pathway dominance in different members (Chistoserdova & Lidstrom 2013).

Despite these differences between carbon assimilation pathways, all bacterial methanotrophs utilise a common methane activation pathway involving initial oxidation to methanol *via* a member of the methane monooxygenase (MMO) family (Park and Lee 2013). Furthermore, the MMO can be split into two subcategories: membrane bound particulate methane monooxygenase (pMMO), and cytoplasmic soluble methane monooxygenase (sMMO) (Sazinsky and Lippard 2015).

2.3. Particulate methane monooxygenase

By far the more prevalent of the two different types of methane monooxygenase is pMMO, with numerous methanotrophic species capable of expressing only this form (Smith and Murrell 2009). For a long time it was believed that this enzyme was universal within methanotrophs (Holmes et al. 1995), and while this was dispelled upon the finding of several methanotrophic *Beijerinckiaceae* strains only containing sMMO (Theisen et al. 2005; Vorobev et al. 2011), these currently remains the only known exception to this rule. Whilst numerous strains contain both types of monooxygenase, in these there appears to be a strong preference towards pMMO being the dominant monooxygenase under the majority of conditions tested (Semrau et al. 2013)

The membrane bound nature of pMMO and its instability during purification has hampered attempts to characterise this enzyme, and it has only been relatively recently that the structure was resolved as a

membrane bound $\alpha_3\beta_3\gamma_3$ trimer (Lieberman and Rosenzweig 2005; Smith et al. 2011). Much uncertainty remains about the active site of this systems, including the nature of its catalytic core. Though this is known to be copper based, debate still exist as to its location and fundamental structure, with mono-, di- and tri-copper active centres proposed (Smith et al. 2011; Culpepper et al. 2014; Cao et al. 2018). The enzyme relies on molecular oxygen and an electron donor to facilitate methane oxidation, and though the biologically active reductant remains officially undetermined, strong support exists for it to be a quinol derivative (Smith and Dalton 2004).

The purification and crystallisation of a stable pMMO – methanol dehydrogenase complex has also raised the intriguing possibility that the reduced pyrrolopuinoline quinone ($PQQH_2$) generated during the oxidation of methanol to formaldehyde may be the *in vivo* electron donor (Myronova et al. 2006). This would therefore enable localised transfer of both methanol and reduction equivalents, resulting in an enzymatic complex converting methane directly to formaldehyde. This, in conjunction with the lower reduction potential of $PQQH_2$ that produces an energetic advantage in comparison to the NADH required by sMMO (Leak and Dalton 1986), may provide the biological basis for the preference. These advantages need to overcome pMMO's almost 10 fold lower reaction rate comparative to sMMO (Lee et al. 2006), which necessitates pMMO at levels up to 20% of a methanotrophs cellular protein (Sazinsky and Lippard 2015).

pMMO is just one member of a small family of enzymes known as the copper membrane monooxygenases (CuMMO) (Tavormina et al. 2011). The family also contains characterised members specialised towards ammonia (Norton et al. 2002) and members implicated in short chain hydrocarbon oxidation (Redmond et al. 2010; Rubin-Blum et al. 2017). Despite clear evidence of specialisation, the family is known to have poor selectivity for their primary substrate, with limited differentiation between ammonia and methane by members specialised to both the former and latter well documented (Stein et al. 2012; Zheng et al. 2014). The products of methane and especially ammonia oxidation are highly toxic to organisms (Stein and Klotz 2011), with evidence that many species containing CuMMO members have been required to incorporate systems to detoxify any co-oxidised products in addition to the metabolic process for their primary substrate (Jones and Morita 1983; Stein and Klotz 2011).

Interestingly, pMMO is not the only member from this CuMMO family that appears in methanotrophs. A largely uncharacterised member known as pXMO has been identified in select members of both the alphaproteobacterial and gammaproteobacterial methanotrophs (Tavormina et al. 2011). Though the primary substrate of this system has not yet been demonstrated due to isolation and purification difficulties, indicators point towards likely methane activity (Tavormina et al. 2011; Hainbuch 2015; Kits et al. 2015). What is known about the system is that expression is upregulated during times of extremely low oxygen concentrations (Hernandez et al. 2015; Kits et al. 2015), thus providing a tentative connection between the system and substrate availability.

2.4. Soluble methane monooxygenase

2.4.1. Overview

Originally identified as an independent monooxygenase through the analysis of methane oxidation in the soluble fraction of methanotroph cellular extract (Colby et al. 1977), sMMO has proved to be a far more tractable methane monooxygenase enzyme to study. Early analysis indicated that the principal components for active hydroxylase activity were the multimeric hydroxylase unit (MMOH) and associated reductase (MMOR) and regulatory (MMOB) components (Dalton and Leak 1985). Though low level oxidation activity can be induced in the hydroxylase unit alone through the use of hydrogen peroxide as a proxy for both O_2 and reducing equivalents (Andersson et al. 1991), the addition of regulatory and the reductase components facilitates a 150 fold increase in kinetics for methane oxidation (Liu, J. C. Nesheim, et al. 1995) and enables the use of the *in vivo* electron donor NAD(P)H (Dalton and Leak 1985).

MMOH itself consists of a 251kDa $\alpha_2\beta_2\gamma_2$ heterodimer (Rosenzweig et al. 1997), with a carboxylate bridged di-iron active site residing in an alpha helix bundle; a characteristic typical of the di-iron centre protein family (Nordlund and Eklund 1995). This active site is enveloped inside the α unit, and substrate ingress/egress is tightly controlled through a highly regulated series of channels and pores (Lee et al. 2013). Despite this extensive regulation of substrate access, oxygenase activity has been observed on an extensive range of organic substrates including alkanes, alkenes, ethers and ketones, as well as

alicyclic, heterocyclic and aromatic compounds (Colby et al. 1977; Jiang et al. 2010). There has also been identified activity towards halogenated organics (Fox et al. 1990) and the inorganic substrates ammonia and carbon monoxide (Colby et al. 1977; Dalton 1977). The capacity for sMMO to degrade many industrially and environmentally relevant chemicals has been frequently noted (Sullivan et al. 1998; Smith and Dalton 2004; Van Beilen and Funhoff 2005; Jiang et al. 2010; Torres Pazmiño et al. 2010), with specific investigations towards the capacity for trichloroethylene and N-nitrosodimethylamine degradation for bioremediation already conducted (Tsien et al. 1989; Sharp et al. 2005). These studies however used sMMO in the native organism, and encountered issues maintaining sMMO expression over the alternative pMMO.

The interchange between pMMO and sMMO in methanotrophs has been observed to occur due to environmental copper concentrations, with sMMO expression in *M. trichosporium* found to only occur at copper concentrations below 0.25μM (Murrell et al. 2000). Though the full mechanism for copper regulation in these systems has not yet been elucidated, extensive investigation has demonstrated a regulation at the transcription level by a σ^{54} promoter (Nielsen et al. 1997), in turn controlled by a

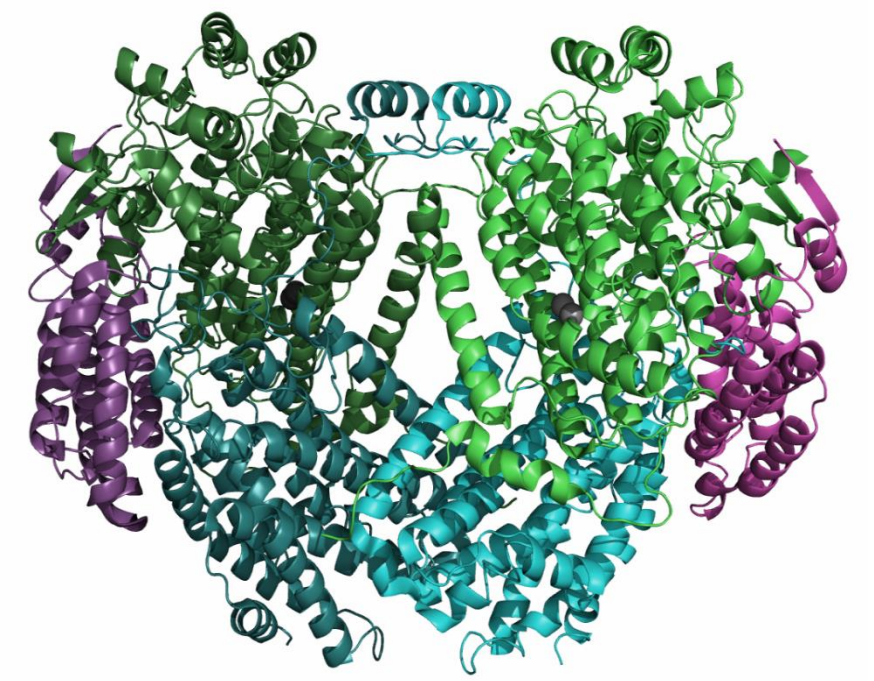


Figure 2.2: Ribbon diagram of MMOH demonstrating dyad symmetry. Structure coloured according to peptide chain: α-, β-, and γ- subunits depicted in *green, blue and purple* respectively, with Fe centres in grey. Symmetric unit rendered in lighter and darker shades respectively. Image produced on PyMol (PDB entry 4GAM).

regulatory σ^{54} factor transcribed by the *mmoR* gene located adjacent to the sMMO operon (Stafford et al. 2003). Furthermore, the *mmoD* gene inside the sMMO operon (Semrau et al. 2013), sMMO associated groEL homologue MMOG (Stafford et al. 2003; Scanlan et al. 2009), and copper sensitive chalkophore methanobactin (Semrau et al. 2013) have also been implicated in regulation at the transcription level.

M. silvestris is the only known exception to copper regulation of sMMO in methanotrophs, with active sMMO expression at copper concentrations as high as 1μM demonstrated (Theisen et al. 2005). This result is unsurprising however as *M. silvestris* is also the only currently known methanotroph lacking

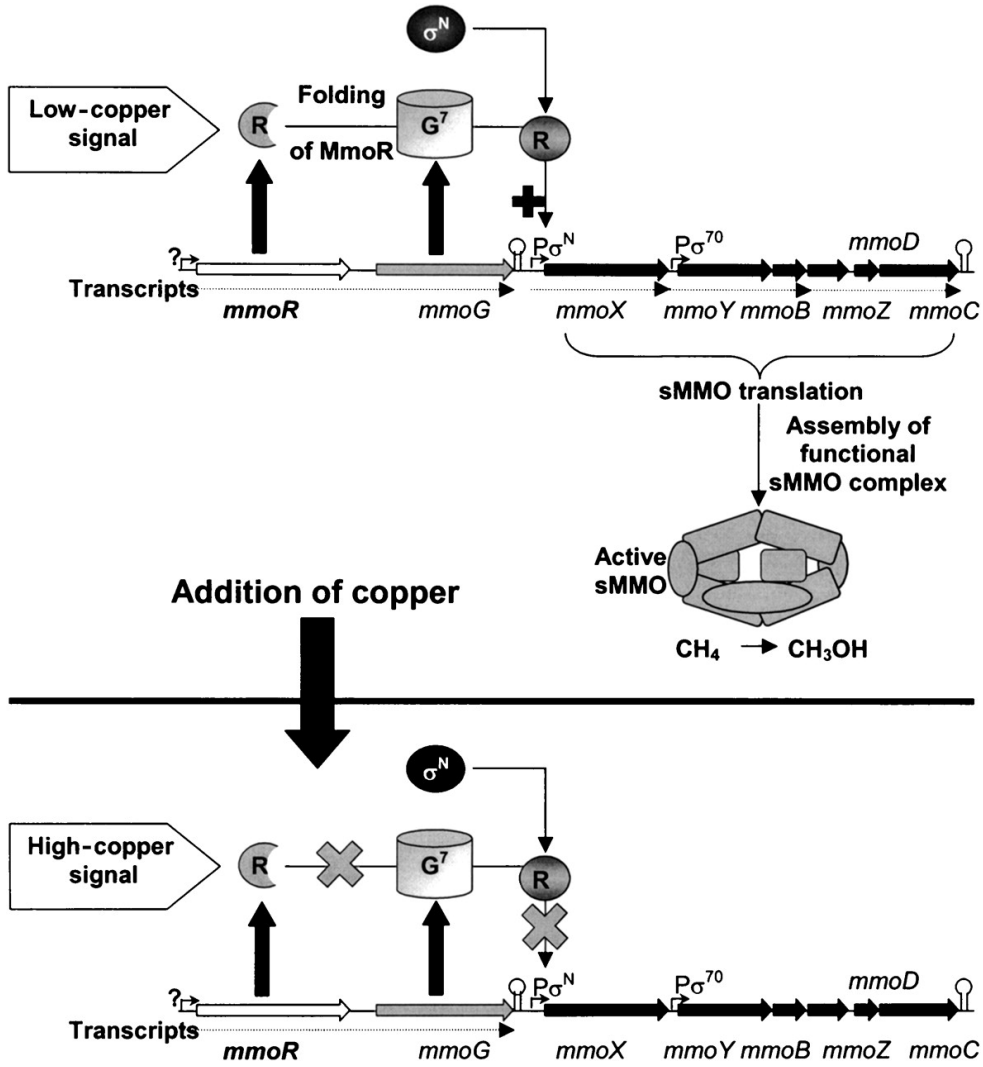


Figure 2.3: Proposed regulation pathway for sMMO expression. Model highlights differential expression under the absence (top) and presence (bottom) of copper. Model also depicts proposed role of MmoR and MmoG in facilitating sMMO operon (*mmoX* - *mmoC*) expression (reproduced with permission from Stafford et al. 2003).

the ability to produce pMMO (Theisen et al. 2005), therefore requiring constitutive expression of sMMO. It does however demonstrate that high copper concentration is not a limiting factor to sMMO activity *per se*.

sMMO exists in a much larger family of bacterial multicomponent monooxygenases (BMMs) including: ethane (Martin et al. 2014); butane (Dubbels et al. 2007); propane (Kotani et al. 2003); phenol (Sazinsky et al. 2006); toluene-2, toluene-3, toluene-4 and toluene/o-xylene (Whited and Gibson 1991; Olsen et al. 1994; Johnson and Olsen 1995; Cafaro et al. 2002); ethene and propene (Coleman and Spain 2003; Chan et al. 2005); as well as tetrahydrofuran (Thiemer et al. 2003) monooxygenases. Work into the evolutionary relationship of the BMM family has demonstrated that substrate specificity is strongly correlated with both its level of evolutionary divergence and the operonic gene order (Leahy et al. 2003; Notomista et al. 2003). Although neither the operonic gene rearrangement nor phylogenetic analysis of the genes themselves provide a clear evolutionary pathway of the BMM family, indications that both the α and β unit originally diverged from a common ancestor (Lundin et al. 2012) allow for putative modelling of evolutionary interrelation (Leahy et al. 2003). Such analysis places the phenol and toluene hydroxylases as closest members to the common ancestor, whilst confirming alternate lines of investigation that indicate sMMO to be evolutionarily the most diverged (Leahy et al. 2003).

Each BMM member utilises a heterodimeric hydroxylase with di-iron centred active site, and also requires multiple cofactors including a reductase and regulatory unit for physiological enzymatic function (Leahy et al. 2003; Notomista et al. 2003). Like sMMO, all BMM members utilise their corresponding reductase cofactor to transfer reducing equivalents from NAD(P)H to an oxidised di-iron centre of the hydroxylase; this transfer performed upon binding of the reduced reductase unit into a ‘binding canyon’ on the hydroxylase (Sazinsky and Lippard 2006; Wang et al. 2014). In most members this is believed to place the [2Fe-2S] cluster of the reductase in close enough proximity of the di-iron centre to directly transfer two electrons, resulting in the formation of the reduced Fe(II)-Fe(II) state at the active site of the hydroxylase (Wang et al. 2014). This process is less certain for the 4-component hydroxylases however as the Rieske-type fourth component was found to mediate electron transfer between the reductase and hydroxylase units (Pikus et al. 1996).

Once the hydroxylase active site has reached its reduced state, the binding canyon undergoes a change in specificity which promotes release of the reductase and binding of the regulatory unit (Liu et al. 1997; Zhang et al. 2006). Binding of the regulatory unit then modifies the orientation of residues in the vicinity of the active site (Lee et al. 2013; Acheson et al. 2014), facilitating subsequent monooxygenase activity of the hydroxylase unit (Liu, Jeremy C. Nesheim, et al. 1995), and also prevents further oxidation of the active site by additional binding of a reductase unit (Wang et al. 2014). The level of similarity of these systems in catalytic cycle and active site structure have been used to draw extensive parallels between these related enzymes in an attempt to elucidate broad structural and mechanistic functionality (Cafaro et al. 2004; Bochevarov et al. 2011; McCormick and Lippard 2011).

2.4.2. The active centre and reaction cycle

The 105kcal mol^{-1} C-H bond energy of methane means that its conversion to methanol is one of the most difficult reactions performed in nature (Sazinsky and Lippard 2015). sMMO and butane monooxygenase are unique in the BMM family as the only enzymes capable of oxidising methane (Sazinsky and Lippard 2006; Cooley et al. 2009), yet sMMO is also capable of (under non-physiological conditions) oxidising the majority of primary substrates of the other BMMs (Colby et al. 1977). It performs this reaction through one of the hallmarks of the BMM family, a di-iron catalytic centre (Leahy et al. 2003).

The di-iron centre resides in a ferritin-like four helix bundle (Lundin et al. 2012), with each helix contributing ligating residues to stabilise the active centre. Both irons are bound by ExxH motifs, these being derived from helix 2 and 4 of the helix bundle for Fe1 and Fe2 respectively. These, in conjunction with a further two glutamic acids from helix 1 and 3 comprise all amino acid components of the binding sphere (figure 4). In the resting ($MMOH_{ox}$) state of the enzyme, the remaining occupied co-ordinate sites of the oxidised $Fe^{III}Fe^{III}$ active centre are occupied by either bridging hydroxide ligands or, in the case of Fe1, a single water molecule (Sirajuddin and Rosenzweig 2015). These sites however undergo dramatic change as the active sites progresses through its reaction cycle.

MMOH is initially activated *via* electron donation by MMOR, reducing the active site to its $Fe^{II}Fe^{II}$ state (Sazinsky and Lippard 2015). Though extensive debate still remains regarding the subsequent oxygen binding and activation process, there is general consensus for a multi-stage activation pathway involving at least one peroxo intermediate (Tinberg and Lippard 2011). This pathway is typically resolved into an initial oxygen binding stage, designated O (Banerjee et al. 2015), before the transition through at least one of several proposed peroxo intermediates including cis-μ-1,2 peroxo, trans-μ-1,2 peroxo and $\mu\text{-}\eta^2\eta^2$ conformations (Rinaldo et al. 2007; Han and Noodleman 2008; Bochevarov et al. 2011). Spectroscopic and kinetic studies have indicated that the reaction cycle is likely to involve at least two of these peroxo intermediates, which have subsequently been designated P^* and P (Tinberg and Lippard 2009). Cleavage of the characteristic peroxide O-O bond then produces the enzymatically active di(μ-oxo) 'diamond core' known as Q (Gherman et al. 2001). This state has been noted to be the most powerful oxidant in biology (Rosenzweig 2015), and the only intermediate in the reaction cycle capable of oxidising methane (Ambundo et al. 2002; Beauvais and Lippard 2005a).

It has been noted however that though the Q state is the only reaction cycle intermediate capable of oxidising methane, at least one other reaction intermediate is capable of C-H bond cleavage in other substrates, and in some instances is more reactive towards select substrates than the Q state (Beauvais and Lippard 2005b; Tinberg and Lippard 2010). Active site modelling (Bochevarov et al. 2011), in

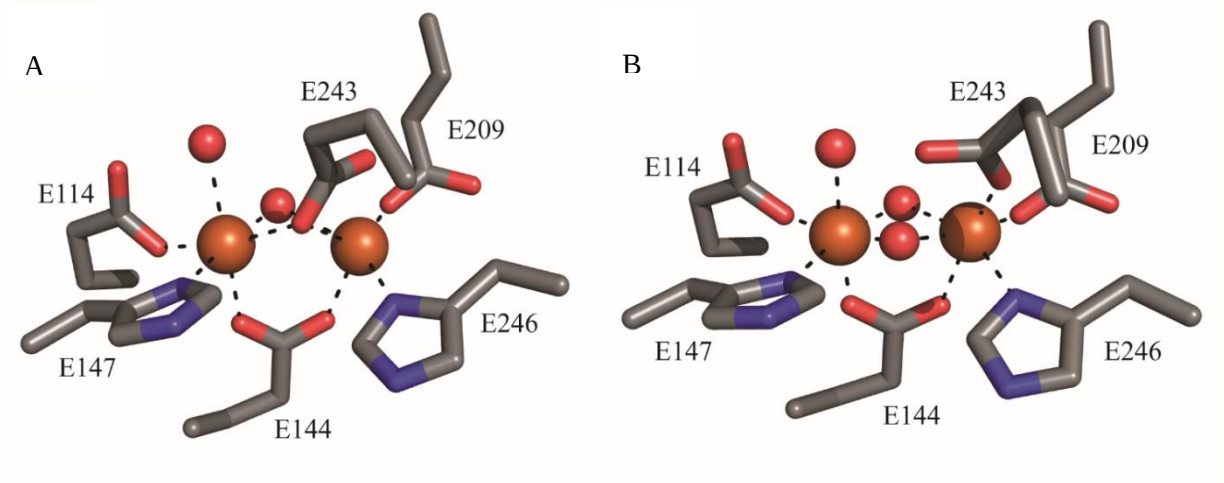


Figure 2.4: sMMO hydroxylase active site structures. (A) Oxidised di-iron active site ($MMOH_{Ox}$) of *M. capsulatus* (PDB accession number 1MTY) and (B) Reduced di-iron active site ($MMOH_{Red}$) of *M. capsulatus* (PDB accession number 1FYZ). Images produced on PyMol.

conjunction with the lack of observed high valence intermediates (Tinberg et al. 2011), has indicated most BMMs inability to cleave the peroxo O-O bond, strongly indicating the locus of functional divergence at a catalytic level characterising sMMO and its respective activity towards methane.

2.4.3. MMOB and its role in hydroxylase regulation

Of primary importance to this reaction cycle is the presence of MMOB, which has been demonstrated to moderate the kinetics of individual reaction stages: the most noteworthy being an increase in oxygen association and P state formation of ~1000 fold (Liu, J. C. Nesheim, et al. 1995). The 15.9kDa MMOB unit binds in a 2:1 ratio to the 'canyon' regions of MMOH produced at the interface between the two protomers, with the binding pocket lying in the region of closest approach to the active site (Figure 2.6) (Lee et al. 2013). Though this binding site is in agreement with the findings of regulatory units for other BMMs, a unique and indispensable N-terminal sequence has been identified in MMOB, that when bound to MMOH forms a highly stable loop conformation, suggesting additional functionality in sMMO (Lee et al. 2013).

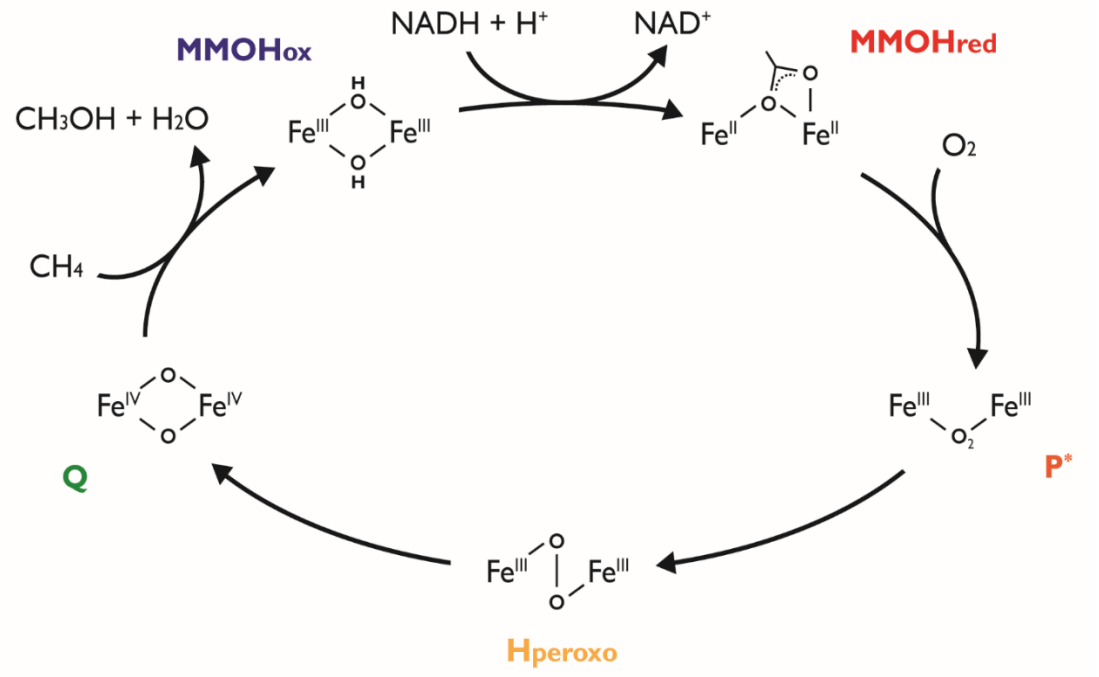


Figure 2.5: Soluble methane monooxygenase reaction cycle including proposed structure of active site intermediates. Names for proposed intermediates are in coloured text.

Site directed mutagenesis and double electron-electron resonance studies have implicated this N-terminal region as a potential tether or hinge (Lee et al. 2013; Wang and Lippard 2014), which would maintain MMOB's proximity during the low affinity oxidised stage of the MMOH catalytic cycle (Kazlauskaite et al. 1996). There is strong evidence for the binding of MMOR to the same region (Figure 2.7) (Wang et al. 2014), which would necessitate the transient nature of this binding and also indicate a possible mechanistic role in preventing further electron donation from MMOR during the reduced catalytic cycle stage through competitive binding (Lee et al. 2013). Such donation would result in the further reduction of the active centre peroxo intermediate and the formation of a futile cycle (Wang et al. 2014). The proposed mechanism would also resolve earlier findings that indicated non-competitive MMOB and MMOR binding, from which separate binding sites was initially inferred (Gassner and Lippard 1999).

The lack of gross structural changes in MMOH upon MMOB binding belies the crucial conformational changes that it induces in the hydroxylase unit. Though alterations to the redox potential (Fox et al. 1991) and structural stability (Pulver et al. 1994; DeWitt et al. 1995; Davydov et al. 1999) in the di-iron centre upon regulatory unit binding have been repeatedly demonstrated, the precise mechanism for such modifications remains elusive. There is greater clarity around the mechanism through which MMOB alters the availability of a 2.0Å wide hydrophilic 'pore' that provides the most direct access pathway to

sMMO subunits involved in catalytic cycle		
Name	Designation	Depicted colour
MMOH - α	Hydroxylase subunit	Green
- β	Hydroxylase subunit	Cyan
- γ	Hydroxylase subunit	Magenta
MMOB	Regulatory unit	Yellow
MMOR	Reductase unit	Not shown

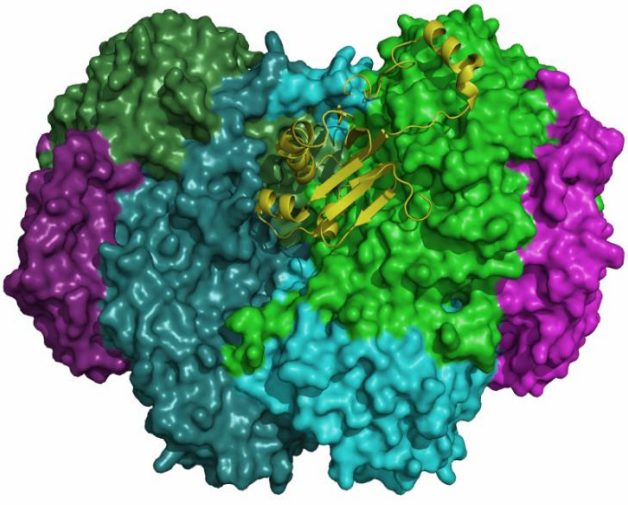


Figure 2.6: Identification of all subunits involved in sMMO catalytic cycle, with space filling model of active intermediate including regulatory unit. Unit designation in space filling model described in table. Note that MMOB unit represented as cartoon structure.

the active site (Sazinsky and Lippard 2005). This channel is produced by the movement of a π loop in helix E of the active site, and has analogous motifs in numerous members of the di-iron centre family (Sazinsky et al. 2006; Sazinsky and Lippard 2006; Murray and Lippard 2007) despite the closest members in the BMM family typically having much larger pore diameters (Sazinsky et al. 2006).

Crystal structures of oxidised MMOH ($MMOH_{ox}$) indicate that this state favours an open conformation for this pore region, facilitating access to the active site (Lee et al. 2013). Comparatively, the binding of MMOB to the hydroxylase induces conformational change, involving the hydrophilic interaction of Thr111 in MMOB and Asn214 in MMOH, drawing a π helix loop adjacent to the active centre outwards and thus occluding this access pathway (Brazeau and Lipscomb 2003; Lee et al. 2013). Furthermore, the binding position of MMOB also physically encapsulates the pore, which also is believed to contribute to the restriction in molecular access (Whittington and Lippard 2001; Lee et al. 2013).

MMOB has also been implicated in the regulation of a large internal cavity network, hypothesised to be the access pathway for methane and oxygen (Rosenzweig et al. 1997; Song et al. 2011). The first observation of such a pathway in the BMM family was a 40Å long tunnel leading from the active site to the enzyme surface in phenol hydroxylase (Figure 2.8) (Sazinsky and Lippard 2005). Similar motifs have since also been elucidated in the additional BMM members toluene/o-xylene monooxygenase

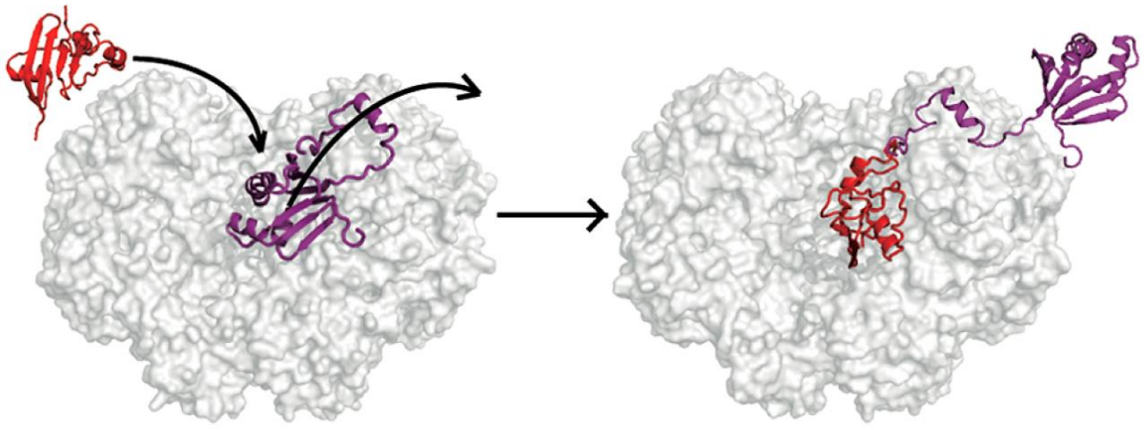


Figure 2.7: Proposed role of 'hinge region' in regulatory unit function, facilitating moderation of reductase access to binding cavity. MMOB (purple) and MMOR (red) depicted interacting with hydroxylase unit (grey), demonstrating both binding (left) and tethering (right) abilities of regulatory unit (reproduced with permission from (Wang et al. 2014)).

(ToMO) and sMMO (McCormick and Lippard 2011), hence exhibiting conserved structural morphology in all crystallised BMMs to date.

The BMM regulatory unit, including the methane monooxygenase associated MMOB, has been demonstrated to play a vital role in the regulation of these cavities through gating effects at cavity interfaces. Control for this regulatory system has been predominantly observed in the region adjacent to the active site, and has been referred to as the 'leucine gate' in MMO due to the predominant role of Leu110 in regulation through steric hindrance (Borodina et al. 2007). Comparative crystal structures of the oxidised and MMOB bound hydroxylase unit has demonstrated a repositioning of this residue, along with Phe188, upon MMOB binding to facilitate access through the hydrophobic network to the active site. This gateway has demonstrated inverse concomitant regulation of the hydrophilic pore, with accessibility being induced by MMOB binding (Lee et al. 2013).

A surprising observation arising from *in vitro* studies was that the presence of MMOB eliminated hydrogen peroxide induced activity in the hydroxylase unit: this is despite catalytic activity being

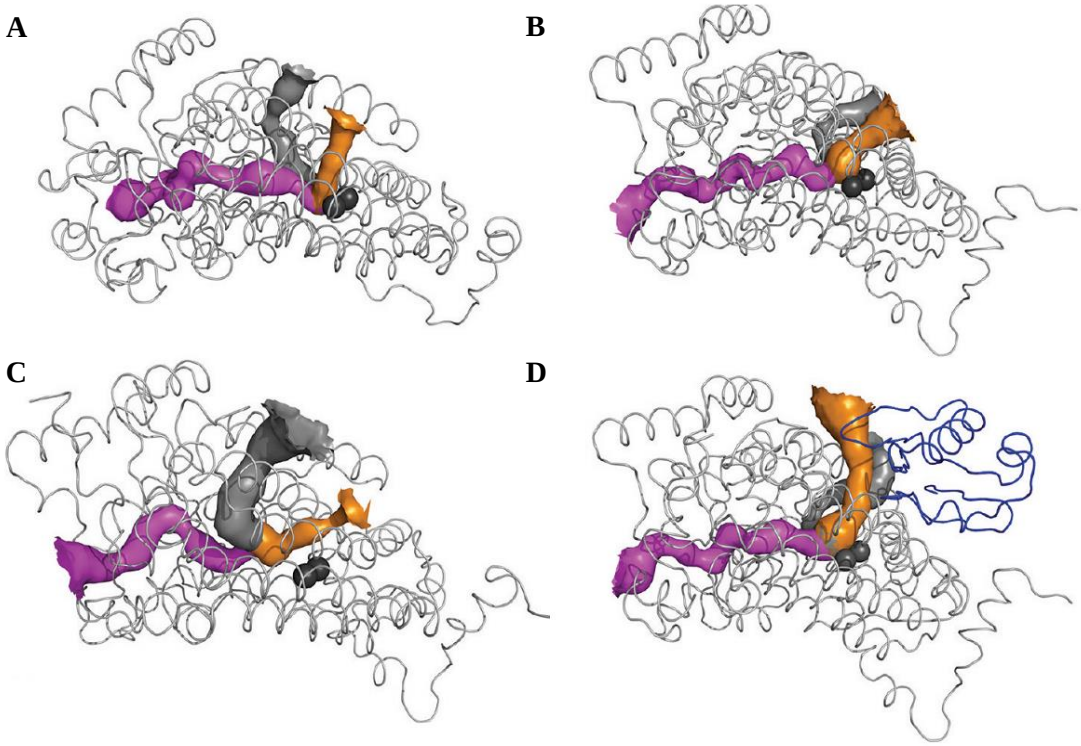


Figure 2.8: Internal cavity network represented in multiple members of the bacterial multicomponent monooxygenase family. Stylised primary cavity network (purple), pore region (orange), and alternative access cavity (charcoal) depicted for MMOH **(a)**, ToMOH **(b)**, PHH **(c)** and regulator unit bound PHH **(d)** (reproduced with permission from McCormick & Lippard 2011).

observed on methane in isolated assays (Jiang et al. 1993). Hydrogen peroxide has been proposed as an expedient method for the activation of the di-iron centre to facilitate methane conversion, especially on an industrial scale (Park and Lee 2013), yet the catalytic rate of the native enzyme under such conditions is insufficient to facilitate an economically viable industrial process. Despite numerous proposals for alternative electron donors and efficient NADH regeneration methods, none have yet achieved commercial viability (Torres Pazmiño et al. 2010). The future use of sMMO in methanol production is therefore envisioned to either require maturation of one of these alternative reduction methods, extensive metabolic engineering of the wild type organism, or the development of sMMO variants with high hydrogen peroxide activity (Park and Lee 2013).

2.5. Expression of sMMO and the role of MMOD

Progress in the detailed characterisation and rational design of high activity sMMO mutants has to date been hindered by the lack of expedient systems for enzyme production. Among the few reported expressions of active hydroxylase in heterologous host organisms (Jahng and Wood 1994; Jahng et al. 1996; Lloyd et al. 1999), none have been achieved in highly tractable organisms (West et al. 1992; Murrell et al. 2000) and expression at levels greater than in wild type has only been observed when using genetically modified methanotrophic organisms as hosts (Lloyd et al. 1999). Furthermore, to date *M. trichosporium* has been the only host in which mutant sMMO has been successfully produced (Smith et al. 2002; Borodina et al. 2007), and never in sufficient quantities for kinetic characterisation (Song et al. 2011). These issues have been due to the inability to produce sufficient levels of active hydroxylase unit, as high level expression of the reductase and regulatory unit in *E. coli* was first achieved over 25 years ago (West et al. 1992). Though new techniques for methanotroph culturing to high densities have recently been developed (Yu et al. 2009), a method for expedient production of both wild type and mutant sMMO still remains elusive, and remains of vital importance for the further development of the field.

In contrast to the issues encountered during attempted sMMO expression in *E. coli*, numerous other members of the BMM family have been successfully expressed using comparable systems. Early

success expressing toluene/o-xylene monooxygenase (Cafaro et al. 2002) was quickly followed by phenol monooxygenase (Cafaro et al. 2004) and alkene monooxygenase (Champreda et al. 2004). More recently, the successful expression of propane monooxygenase (Furuya, Hayashi, and Kino 2013) has resulted in sMMO being the only major phylogenetic radiation in the BMM family to not have a member successfully expressed in *E. coli*. The expression of both toluene/o-xylene and alkene monooxygenase was found to only require the genes explicitly involved in catalysis for active heterologous production (Cafaro et al. 2002; Champreda et al. 2004), demonstrating that the BMM family as a whole has not developed, nor does it require, additional factors for folding or active site assembly. Furthermore, the additional factors required for heterologous expression of phenol hydroxylase and propane monooxygenase were both present either in the BMM operon itself or immediately adjacent to it. These were found to be a small (~10kDa) cofactor-less accessory protein and a GroEL analogue for phenol hydroxylase and propane monooxygenase respectively (Powlowski et al. 1997; Furuya, Hayashi, Semba, et al. 2013).

The sMMO operon consists of six open reading frames. Half of these (*mmoX*, *mmoY* and *mmoZ*) produce the constituent α, β and γ units of MMOH, whilst *mmoB* and *mmoC* encode the regulatory

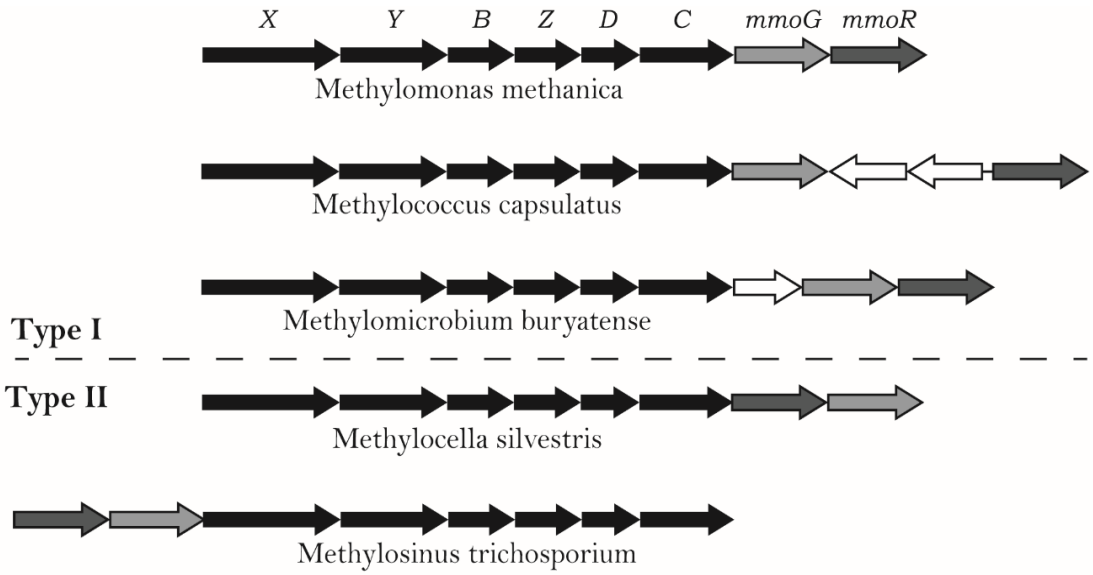


Figure 2.9: Gene organization of sMMO gene cluster in methanotrophs. sMMO operon represented in black, with associated mmoG and mmoR genes in light and dark grey respectively. Additional genes not conserved in vicinity of sMMO operon in white. Selected representative methanotrophs containing respective gene order specified below each sequence, with demarcation between clusters found in the type I and type II methanotrophs shown.

MMOB and reductase MMOR units, respectively (Figure 2.9) (Merkx et al. 2001). The additional gene in the operon, designated *mmoD,* is positioned between *mmoZ* and *mmoC* and has been demonstrated to contain an actively expressed protein, designated as both MMOD (sometimes MmoD) and OrfY in the literature. Despite only a low level of protein sequence conservation in MMOD being observed across methanotrophic species, a central region exhibiting around 45% conservation has been identified (Merkx and Lippard 2002).

Investigations of MMOD have revealed a 12kDa protein (Merkx and Lippard 2002) expressed to a level of around 1-2% of that for MMOH (Merkx et al. 2001). Though not directly required for sMMO activity, and indeed demonstrating inhibitory affects to MMOH activity (Merkx and Lippard 2002), gene knockouts have indicated that this protein is required for production of the active enzyme system (Semrau et al. 2013). The protein has been shown to bind to MMOH, with a higher affinity for the apo-protein (Sazinsky et al. 2004). Demonstrations of competitive binding with MMOB has also indicated the possibility of an overlapping binding region for both proteins (Merkx and Lippard 2002).

Despite these findings, a tentative assignment of the role of MMOD has been made as a transcriptional regulator (Semrau et al. 2013). This has been due to findings in marker exchange mutagenesis experiments, in which dramatically lower sMMO operon expression levels were observed after the substitution of all operonic genes with an expression marker (genes X – C in Figure 2.9) (Semrau et al. 2013). This, in conjunction with the lack of assigned physiological function of MMOD as compared to all other genes contained in the operon, has resulted in the subsequent transcriptional activator designation, though such assignment has often not been considered conclusive (Sazinsky and Lippard 2015).

Despite low sequence similarities, the question still remains of whether the DmpK accessory protein found to be required for heterologous phenol hydroxylase expression is an analogue to MMOD (Powlowski et al. 1997). These have demonstrated morphological similarities such as size and absence of cofactors, as well as parallel physiological characteristics: that is, binding to the hydroxylase unit in a competitive fashion relative to the regulatory component, relative low level expression comparative

to the hydroxylase unit, decrease in hydroxylase activity when present in high levels, and an indispensable role in the production of active enzyme (Powlowski et al. 1997; Izzo et al. 2011). In contrast to MMOD however, DmpK has been demonstrated not to be required for transcription of the hydroxylase unit (Semrau et al. 2013), and there is evidence of its direct requirement in the formation of the active hydroxylase unit (Izzo et al. 2011); the latter having been difficult to assess in MMOD to date.

Of particular relevance to sMMO formation was the finding that DmpK bound to the phenol hydroxylase increased the stability of the monomeric α/β/γ unit, as opposed to the catalytically active dimeric form (Izzo et al. 2011). In conjunction with the increased binding ability for the iron co-factor in this form, the primary role of DmpK was therefore hypothesised to be involved in both the formation and reformation of active di-iron centres. Though comparative experiments using MMOD bound methane monooxygenase indicated decreased iron retention in the active site (Sazinsky et al. 2004), the precise nature of MMOD's role regarding both transcription and protein assembly still remains unresolved.

2.6. GroEL and its potential role in sMMO expression

An interesting observation has been noted for several members of the BMM family: a GroEL analogue divergent from classical members of the family has been observed to be retained in close proximity to the respective monooxygenase operon. The characterised members that exhibit this phenomenon are the propane (Kotani et al. 2003; Sharp et al. 2007), butane (Kurth et al. 2008) and methane monooxygenases (Csáki et al. 2003; Stafford et al. 2003; Theisen et al. 2005); for each group this GroEL analogue has been demonstrated to be required for the successful expression of the active hydroxylase in the native organism (Stafford et al. 2003; Kurth et al. 2008; Furuya, Hayashi, Semba, et al. 2013). Recent success expressing propane monooxygenase in heterologous organisms including *E. coli* has also demonstrated that this GroEL analogue was required for the successful expression of the active hydroxylase at the translational level (Furuya, Hayashi, Semba, et al. 2013; Furuya, Hayashi, and Kino 2013).

The archetype GroEL, otherwise known as chaperonin, belongs to a wider family of chaperone proteins that facilitate proteostasis through assisting partially folded and mis-folded proteins in attaining their native conformation (Ellis 2006). The potential specialisation of the BMM associated GroEL analogues, and inability to replace with conventional GroEL analogues, is surprising due to the conventionally promiscuous nature of these protein folding enzymes. The single copy found in *E. coli* has been found to associate with up to 50% of mis-folded *E. coli* proteins (Viitanen et al. 1992), ~250-300 of which have been identified as binding in a stable and highly reproducible fashion (Houry et al. 1999; Kerner et al. 2005). Though debate continues regarding the detailed mechanism for GroEL chaperone activity (Lin and Rye 2006), general consensus exists regarding its role in sequestering unfolded and mis-folded proteins and providing a conducive environment for correct folding, termed the 'Anfinsen cage' (Ellis 2003). This facilitates attainment of native structure and destabilises erroneous conformations resultant from kinetic folding traps in proteins (Dahiya and Chaudhuri 2014).

Although for the majority of substrates folding will occur intrinsically in typical *in vitro* conditions, between 20-30% of the cellular cytoplasm is typically constituted by macromolecules, and the subsequent high level of steric crowding can result in protein agglomeration before the native state is

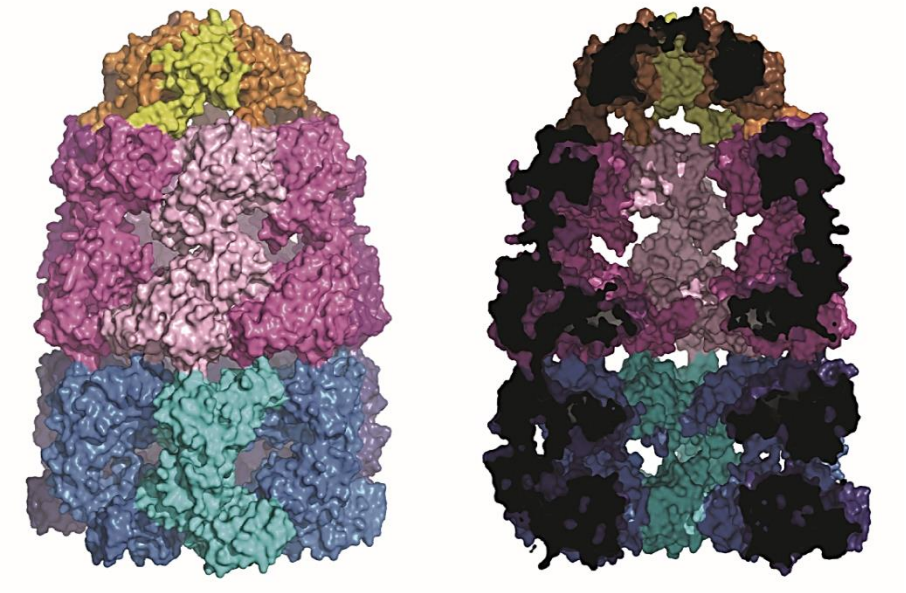


Figure 2.10: Space filling model of GroES bound GroEL. Surface structural (right) and internal cut-away (left) representations. Unbound 'trans' cavity (blue), bound 'cis' cavity (pink) and GroES cap (orange) depicted, with individual monomers for each region shown in cyan, lt. pink and yellow respectively.

attained (Ellis 2001). Furthermore, a small subset of proteins have an obligatory dependency on GroEL for attainment of native conformation, providing direct necessitation of such chaperones (Kerner et al. 2005).

GroEL's tertiary and quaternary structure is crucial for its function. Each 60kDa GroEL unit consists of three domains: an equatorial base that contains an ATP binding region, intermediate hinge domain, and an apical domain including the substrate and GroES binding components (Figure 2.10) (Wang et al. 2002). In GroEL's functional form, these units are oligomerised to the native tetradecameric structure, with dual heptameric rings producing two adjoining hollow cavities located in the protein mesostructure, both facing the external environment. The heptameric GroES cofactor is designed to cap these cavities, producing an encapsulated space of slightly under 200,000$Å^3$ (Figure 2.10) (Clare et al. 2006), the so called 'Anfinsen cage', capable of accommodating proteins of up to 60kDa, dependant on morphological characteristics (Saibil 2013). Though no specific binding sequences have been identified for any member of the chaperonin family (Houry et al. 1999), a propensity to bind exposed hydrophobic regions has been observed and is believed to be their primary mechanism for identifying misfolded proteins (Saibil 2013).

Whilst the presence of a single copy of the GroEL gene is most prevalent in bacterial species, nearly 30% exhibit at least two copies (Lund 2009), with high levels of proliferation being observed in numerous species including *Bradyrhizobium japonicum*, with the highest recorded number of seven variants (Kumar et al. 2015). Whilst the importance of GroEL gene multiplication in facilitating tight regulation expression has been clearly demonstrated (Kondrashov and Kondrashov 2006; Lund 2009), a significant body of evidence is also pointing towards the existence of divergent GroEL members facilitating folding of a wider array of substrates, with clear examples in *Mycobacterium smegmatis* (Kim et al. 2003; Ojha et al. 2005), *Mycobacterium tuberculosis* (Hu et al. 2008), *Corynebacterium glutamicum* (Barreiro et al. 2005) and *Streptomyces albus* (Servant et al. 1994) all demonstrating non-synonymous function across GroEL variants in the organism.

2.7. References

Acheson JF, Bailey LJ, Elsen NL, Fox BG. (2014) Structural basis for biomolecular recognition in overlapping binding sites in a diiron enzyme system. *Nat. Commun.* 5, 5009.

Ambundo EA, Lippard SJ, Friesner RA. (2002) Reactions of methane monooxygenase intermediate Q with derivatized methanes. *J. Am. Chem. Soc.* 124, 8770–8771.

Andersson KK, Froland W a., Lee S-K, Lipscomb JD. (1991) Peroxide coupled turnover by methane monooxygenase hydroxylase component. *J. Inorg. Biochem.* 43, 557.

Angle JC, Morin TH, Solden LM, Narrowe AB, Smith GJ, Borton MA, Rey-Sanchez C, Daly RA, Mirfenderesgi G, Hoyt DW, et al. (2017) Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat. Commun.* 8, 1–9.

Anthony C. (1982) The biochemistry of methylotrophs. Academic Press

Banerjee R, Proshlyakov Y, Lipscomb JD, Proshlyakov D a. (2015) Structure of the key species in the enzymatic oxidation of methane to methanol. *Nature* 518.

Barreiro C, González-Lavado E, Brand S, Tauch A, Martín JF. (2005) Heat shock proteome analysis of wild-type Corynebacterium glutamicum ATCC 13032 and a spontaneous mutant lacking GroEL1, a dispensable chaperone. *J. Bacteriol.* 187, 884–889.

Beauvais LG, Lippard SJ. (2005a) Reactions of the diiron(IV) intermediate Q in soluble methane monooxygenase with fluoromethanes. *Biochem. Biophys. Res. Commun.* 338, 262–266.

Beauvais LG, Lippard SJ. (2005b) Reactions of the Peroxo Intermediate of Soluble Methane Monooxygenase Hydroxylase with Ethers. *J. Am. Chem. Soc.* 127, 7370–7378.

Van Beilen JB, Funhoff EG. (2005) Expanding the alkane oxygenase toolbox: New enzymes and applications. *Curr. Opin. Biotechnol.* 16, 308–314.

Bhattarai S, Cassarini C, Gonzalez-Gil G, Egger M, Slomp CP, Zhang Y, Esposito G, Lens PNL. (2017) Anaerobic Methane-Oxidizing Microbial Community in a Coastal Marine Sediment: Anaerobic Methanotrophy Dominated by ANME-3. *Microb. Ecol.* 74, 608–622.

Bochevarov AD, Li J, Song WJ, Friesner R a., Lippard SJ. (2011) Insights into the different dioxygen activation pathways of methane and toluene monooxygenase hydroxylases. *J. Am. Chem. Soc.* 133, 7384–7397.

Borodina E, Nichol T, Dumont MG, Smith TJ, Murreil JC. (2007) Mutagenesis of the "leucine gate" to explore the basis of catalytic versatility in soluble methane monooxygenase. *Appl. Environ. Microbiol.* 73, 6460–6467.

Bowman JP, McCammon S a., Skerratt JH. (1997) Methylosphaera hansonii gen. nov., sp. nov., a psychrophilic, group I methanotroph from Antarctic marine-salinity, meromictic lakes. *Microbiology* 143, 1451–1459.

Brazeau BJ, Lipscomb JD. (2003) Key amino acid residues in the regulation of soluble methane monooxygenase catalysis by component B. *Biochemistry* 42, 5618–5631.

Bull ID, Parekh NR, Hall GH, Ineson P, Evershed RR. (2000) Detection and classification of atmospheric methane oxidizing bacteria in soil. *Nature* 405, 175–178.

Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Di Donato A. (2004) Phenol hydroxylase and toluene/o-xylene monooxygenase from Pseudomonas stutzeri OX1: Interplay between two enzymes. *Appl. Environ. Microbiol.* 70, 2211–2219.

Cafaro V, Scognamiglio R, Viggiani A, Izzo V, Passaro I, Notomista E, Dal Piaz F, Amoresano A, Casbarra A, Pucci P, et al. (2002) Expression and purification of the recombinant subunits of toluene/o-xylene monooxygenase and reconstitution of the active complex. *Eur. J. Biochem.* 269, 5689–5699.

Cao L, Caldararu O, Rosenzweig AC, Ryde U. (2018) Quantum Refinement Does Not Support Dinuclear Copper Sites in Crystal Structures of Particulate Methane Monooxygenase. *Angew. Chemie - Int. Ed.* 57, 162–166.

Cedervall PE, Dey M, Pearson AR, Ragsdale SW, Milmot CM. (2010) Structural insight into methyl-Coenzyme M reductase chemistry using Coenzyme B analogues. *Biochemistry* 49, 7683–7693.

Champreda V, Zhou N-Y, Leak DJ. (2004) Heterologous expression of alkene monooxygenase components from Xanthobacter autotrophicus Py2 and reconstitution of the active complex. *FEMS Microbiol. Lett.* 239, 309–318.

Chan CK, Chion K, Askew SE, Leak DJ, Icrobiol a PPLENM. (2005) Cloning , Expression , and Site-Directed Mutagenesis of the Propene Monooxygenase Genes from Mycobacterium sp . Strain M156. *Appl. Environ. Microbiol.* 71, 1909–1914.

Chistoserdova L. (2011) Modularity of methylotrophy, revisited. *Environ. Microbiol.* 13, 2603–2622.

Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. (2009) The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* 63, 477–499.

Chistoserdova L, Lidstrom ME. (2013) Aerobic methylotrophic prokaryotes. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes: Prokaryotic Physiology and Biochemistry. 4th ed. Springer Berlin Heidelberg. p. 267–285.

Chistoserdova L, Vorholt JA, Lidstrom ME. (2005) A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. *Genome Biol.* 6, 208.

Chu F, Beck DAC, Lidstrom ME. (2016) MxaY regulates the lanthanide-mediated methanol dehydrogenase switch in *Methylomicrobium buryatense*. *PeerJ* 4, e2435.

Chu F, Lidstrom ME. (2016) XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph Methylomicrobium buryatense. *J. Bacteriol.* 198, 1317–1325.

Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR. (2006) An Expanded Protein Folding Cage in the GroEL–gp31 Complex. *J. Mol. Biol.* 358, 905–911.

Colby J, Stirling DI, Dalton H. (1977) The soluble methane mono-oxygenase of Methylococcus capsulatus (Bath). Its ability to oxygenate n-alkanes, n-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem. J.* 165, 395–402.

Coleman N V, Spain JC. (2003) Epoxyalkane: Coenzyme M Transferase in the Ethene and Vinyl Chloride Biodegradation Pathways of Mycobacterium Strain JS60. *J. Bacteriol.* 185, 5536–5545.

Conrad R, Rothfuss F. (1991) Methane oxidation in the soil surface layer of a flooded rice field and the effect of ammonium. *Biol. Fertil. Soils* 12, 28–32.

Cooley RB, Dubbels BL, Sayavedra-Soto L a., Bottomley PJ, Arp DJ. (2009) Kinetic characterization of the soluble butane monooxygenase from Thauera butanivorans, formerly "Pseudomonas butanovora." *Microbiology* 155, 2086–2096.

Crowther GJ, Kos??ly G, Lidstrom ME. (2008) Formate as the main branch point for methylotrophic metabolism in Methylobacterium extorquens AM1. *J. Bacteriol.* 190, 5057–5062.

Csáki R, Bodrossy L, Klem J, Kovács KL, Murrell JC. (2003) Genes involved in the copper-dependent regulation of soluble methane monooxygenase of Methylococcus capsulatus (Bath): Cloning, sequencing and mutational analysis. *Microbiology* 149, 1785–1795.

Culpepper M a, Cutsail Iii GE, Gunderson W a, Hoffman BM, Rosenzweig AC. (2014) Identification of the Valence and Coordination Environment of the Particulate Methane Monooxygenase Copper Centers by Advanced EPR Characterization. *J. Am. Chem. Soc.* 136, 11767–11775.

Dahiya V, Chaudhuri TK. (2014) Chaperones GroEL/GroES accelerate the refolding of a multidomain protein through modulating on-pathway intermediates. *J. Biol. Chem.* 289, 286–298.

Dalton H. (1977) Ammonia oxidation by the methane oxidising bacterium Methylococcus capsulatus strain bath. *Arch. Microbiol.* 114, 273–279.

Dalton H. (2005) The Leeuwenhoek Lecture 2000 the natural and unnatural history of methane-oxidizing bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1207–1222.

Dalton H, Leak DJ. (1985) Mechanistic Studies on the Mode of Action of Methane Monooxygenase. In: Degn H, Cox RP, Toftlund H, editors. Gas Enzymology. Springer Netherlands. p. 169–186.

Davies SL, Whittenbury R. (1970) Fine structure of methane and other hydrocarbon-utilizing bacteria. *J. Gen. Microbiol.* 61, 227–232.

Davydov R, Valentine AM, Komar-Panicucci S, Hoffman BM, Lippard SJ. (1999) An EPR study of the dinuclear iron site in the soluble methane monooxygenase from Methylococcus capsulatus (Bath) reduced by one electron at 77 K: The effects of component interactions and the binding of small molecules to the diiron(III) center. *Biochemistry* 38, 4188–4197.

Dedysh SN, Knief C, Dunfield PF. (2005) Methylocella species are facultatively methanotrophic. *J. Bacteriol.* 187, 4665–4670.

DeWitt JG, Rosenzweig AC, Salifoglou A, Hedman B, Lippard SJ, Hodgson KO. (1995) X-ray Absorption Spectroscopic Studies of the Diiron Center in Methane Monooxygenase in the Presence of Substrate and the Coupling Protein of the Enzyme System. *Inorg. Chem.* 34, 2505–2515.

Dubbels BL, Sayavedra-Soto LA, Arp DJ. (2007) Butane monooxygenase of "Pseudomonas butanovora": Purification and biochemical characterization of a terminal-alkane hydroxylating diiron monooxygenase. *Microbiology* 153, 1808–1816.

Ellis RJ. (2001) Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* 11, 114–119.

Ellis RJ. (2003) Protein Folding: Importance of the Anfinsen Cage. *Curr. Biol.* 13, 881–883.

Ellis RJ. (2006) Molecular chaperones: assisting assembly in addition to folding. *Trends Biochem. Sci.* 31, 395–401.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MMM, Schreiber F, Dutilh BE, Zedelius J, de Beer D, et al. (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 464, 543–548.

Ferry JG. (1999) Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol. Rev.* 23, 13–38.

Fox BG, Borneman JG, Wackett LP, Lipscomb JD. (1990) Haloalkene oxidation by the soluble methane monooxygenase from Methylosinus trichosporium OB3b: mechanistic and environmental implications. *Biochemistry* 29, 6419–6427.

Fox BG, Liu Y, Dege JE, Lipscomb JD. (1991) Complex Formation between the Protein Components of Methane Monooxygenase from Methylosinus trichosporium OB3b. *J. Biol. Chem.* 266, 540–550.

Furuya T, Hayashi M, Kino K. (2013) Reconstitution of active mycobacterial binuclear iron monooxygenase complex in escherichia coli. *Appl. Environ. Microbiol.* 79, 6033–6039.

Furuya T, Hayashi M, Semba H, Kino K. (2013) The mycobacterial binuclear iron monooxygenases require a specific chaperonin-like protein for functional expression in a heterologous host. *FEBS J.* 280, 817–826.

Gassner GT, Lippard SJ. (1999) Component interactions in the soluble methane monooxygenase system from Methylococcus capsulatus (Bath). *Biochemistry* 38, 12768.

Gherman BF, Dunietz BD, Whittington D a., Lippard SJ, Friesner R a. (2001) Activation of the C-H bond of methane by intermediate Q of methane monooxygenase: A theoretical study [10]. *J. Am. Chem. Soc.* 123, 3836–3837.

Hainbuch S. (2015) Novel deep branching Cu-containing membrane-bound monooxygenases : distribution and function. *Thesis*, 145.

Han WG, Noodleman L. (2008) Structural model studies for the peroxo intermediate P and the reaction pathway from P to Q of methane monooxygenase using broken-symmetry density functional calculations. *Inorg. Chem.* 47, 2975–2986.

Hanson RS, Hanson TE. (1996) Methanotrophic bacteria. *Microbiol. Rev.* 60, 439–471.

He H, Edlich-Muth C, Lindner SN, Bar-Even A. (2018) Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of E. coli. *ACS Synth. Biol.* 7, 1601–1611.

Hernandez ME, Beck DAC, Lidstrom ME, Chistoserdova L. (2015) Oxygen availability is a major factor in determining the composition of microbial communities involved in methane oxidation. *PeerJ* 3, e801.

Heyer J, Berger U, Hardt M, Dunfield PF. (2005) Methylohalobius crimeensis gen. nov., sp. nov., a moderately halophilic, methanotrophic bacterium isolated from hypersaline lakes of Crimea. *Int. J. Syst. Evol. Microbiol.* 55, 1817–1826.

Hinrichs K-U, Boetius A. (2003) The Anaerobic Oxidation of Methane: New Insights in Microbial Ecology and Biogeochemistry. *Ocean Margin Syst. SE - 28*, 457–477.

Hirayama H, Fuse H, Abe M, Miyazaki M, Nakamura T, Nunoura T, Furushima Y, Yamamoto H, Takai K. (2013) Methylomarinum vadi gen. nov., sp. nov., a methanotroph isolated from two distinct marine environments. *Int. J. Syst. Evol. Microbiol.* 63, 1073–1082.

Hitchings GH. (1983) Functions of tetrahydrofolate and the role of dihydrofolate reductase in cellular metabolism. In: Hitchings GH, editor. Inhibition of folate metabolism in chemotherapy. p. 11–23.

Holmes a J, Costello a, Lidstrom ME, Murrell JC. (1995) Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* 132, 203–208.

Hou S, Makarova KS, Saw JHW, Senin P, Ly B V., Zhou Z, Ren Y, Wang J, Galperin MY, Omelchenko M V., et al. (2008) Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. *Biol. Direct* 3, 1–25.

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* 402, 147.

Hu Y, Henderson B, Lund P a., Tormay P, Ahmed MT, Gurcha SS, Besra GS, Coates ARM. (2008) A Mycobacterium tuberculosis mutant lacking the groEL homologue cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. *Infect. Immun.* 76, 1535–1546.

Izzo V, Leo G, Scognamiglio R, Troncone L, Birolo L, Di Donato A. (2011) PHK from phenol hydroxylase of Pseudomonas sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases. *Arch. Biochem. Biophys.* 505, 48–59.

Jahng D, Kim CS, Hanson RS, Wood TK. (1996) Optimization of trichloroethylene degradation using soluble methane monooxygenase of Methylosinus trichosporium OB3b expressed in recombinant bacteria. *Biotechnol. Bioeng.* 51, 349–359.

Jahng D, Wood TK. (1994) Trichloroethylene and chloroform degradation by a recombinant pseudomonad expressing soluble methane monooxygenase from Methylosinus trichosporium OB3b. *Appl. Environ. Microbiol.* 60, 2473–2482.

Janvier M, Frehel C, Grimont F, Gasser F. (1985) Methylophaga marina gen. nov., sp. nov. and methylophaga thalassica sp. nov., marine methylotrophs. *Int. J. Syst. Bacteriol.* 35, 131–139.

Jiang H, Chen Y, Jiang P, Zhang C, Smith TJ, Murrell JC, Xing XH. (2010) Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. *Biochem. Eng. J.* 49, 277–288.

Jiang Y, Wilkins PC, Dalton H. (1993) Activation of the hydroxylase of sMMO from Methylococcus capsulatus (Bath) by hydrogen peroxide. *Biochim. Biophys. Acta* 1163, 105–112.

Johnson GR, Olsen RH. (1995) Nucleotide sequence analysis of genes encoding a toluene/benzene-2-monooxygenase from Pseudomonas sp. strain JS150. *Appl. Environ. Microbiol.* 61, 3336–3346.

Jones RD, Morita RY. (1983) Methane oxidation by Nitrosococcus oceanus and Nitrosomonas europaea. *Appl. Environ. Microbiol.* 45, 401–410.

Kalyuzhnaya MG, Puri AW, Lidstrom ME. (2015) Metabolic engineering in methanotrophic bacteria. *Metab. Eng.* 29, 142–152.

Karl DM, Beversdorf L, Björkman KM, Church MJ, Martinez A, Delong EF. (2008) Aerobic production of methane in the sea. *Nat. Geosci.* 1, 473–478.

Kato N, Ohashi H, Tani Y, Ogata K. (1978) 3-Hexulosephosphate synthase from Methylomonas aminofaciens 77a Purification, properties and kinetics. *BBA - Enzymol.* 523, 236–244.

Kazlauskaite J, Hill H a O, Wilkins PC, Dalton H. (1996) Direct electrochemistry of the hydroxylase of soluble methane monooxygenase from Methylococcus capsulatus (Bath). *Eur. J. Biochem.* 241, 552–556.

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* 122, 209–220.

Khadem AF, Pol A, Wieczorek A, Mohammadi SS, Francoijs KJ, Stunnenberg HG, Jetten MSM, Op den Camp HJM. (2011) Autotrophic methanotrophy in verrucomicrobia: Methylacidiphilum fumariolicum solv uses the calvin-benson-bassham cycle for Carbon Dioxide Fixation. *J. Bacteriol.* 193, 4438–4446.

Kim AI, Ghosh P, Aaron M a., Bibb L a., Jain S, Hatfull GF. (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol. Microbiol.* 50, 463–473.

Kist J, Tate RL. (2013) Phylogeny of bacterial methylotrophy genes reveals robustness in Methylobacterium mxaF sequences and mxa operon construction. *Soil Biol. Biochem.* 59, 49–57.

Kits KD, Klotz MG, Stein LY. (2015) Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp. nov. type strain FJG1. *Environ. Microbiol.* 17, 3219–3232.

Knief C. (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Front. Microbiol.* 6.

Knief C, Lipski A, Dunfield PF. (2003) Diversity and activity of methanotrophic bacteria in different upland soils. *Appl. Environ. Microbiol.* 69, 6703–6714.

Kondrashov F a., Kondrashov AS. (2006) Role of selection in fixation of gene duplications. *J. Theor. Biol.* 239, 141–151.

Korotkova N, Chistoserdova L, Kuksa V, Lidstrom ME. (2002) Glyoxylate regeneration pathway in the methylotroph Methylobacterium extorquens AM1. *J. Bacteriol.* 184, 1750–1758.

Kotani T, Yamamoto T, Yurimoto H, Sakai Y, Kato N. (2003) Propane Monooxygenase and NAD + -Dependent Secondary Alcohol Dehydrogenase in Propane Metabolism by Gordonia sp. Strain TY-5. *J. Bacteriol.* 185, 7120–7128.

Kumar CMS, Mande SC, Mahajan G. (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. *Cell Stress Chaperones* 20, 555–574.

Kurth EG, Doughty DM, Bottomley PJ, Arp DJ, Sayavedra-Soto L a. (2008) Involvement of BmoR and BmoG in n-alkane metabolism in "Pseudomonas butanovora." *Microbiology* 154, 139–147.

Leahy JG, Batchelor PJ, Morcomb SM. (2003) Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* 27, 449–479.

Leak DJ, Dalton H. (1986) Growth yields of methanotrophs 2. A theoretical analysis. *Appl. Microbiol. Biotechnol.* 23, 477–481.

Lee SJ, McCormick MS, Lippard SJ, Cho U-S. (2013) Control of substrate access to the active site in methane monooxygenase. *Nature* 494, 380–384.

Lee SW, Keeney DR, Lim DH, Dispirito A a., Semrau JD. (2006) Mixed pollutant degradation by

Methylosinus trichosporium OB3b expressing either soluble or particulate methane monooxygenase: Can the tortoise beat the hare. *Appl. Environ. Microbiol.* 72, 7503–7509.

Lieberman RL, Rosenzweig AC. (2005) Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature* 434, 177–182.

Lin Z, Rye H. (2006) GroEL-mediated protein folding: Making the impossible, possible. *Crit. Rev. Biochem. Mol. Biol.* 41, 211–239.

Liu Y, Nesheim JC, Lee SK, Lipscomb JD. (1995) Gating effects of component B on oxygen activation by the methane monooxygenase hydroxylase component. *J. Biol. Chem.* 270, 24662–24665.

Liu Y, Nesheim JC, Paulsen KE, Stankovich MT, Lipscomb JD. (1997) Roles of the methane monooxygenase reductase component in the regulation of catalysis. *Biochemistry* 36, 5223–5233.

Liu Y, Nesheim JC, Paulsen KE, Wallar BJ, Stankovich MT, Lipscomb JD. (1995) Roles of component B and reductase in methane monooxygenase (MMO) turnover. *J. Inorg. Biochem.* 59, 368.

Lloyd JS, De Marco P, Dalton H, Murrell JC. (1999) Heterologous expression of soluble methane monooxygenase genes in methanotrophs containing only particulate methane monooxygenase. *Arch. Microbiol.* 171, 364–370.

Lund PA. (2009) Multiple chaperonins in bacteria – why so many? *FEMS Microbiol. Rev.* 33, 785–800.

Lundin D, Poole AM, Sjoberg B-M, Hogbom M. (2012) Use of Structural Phylogenetic Networks for Classification of the Ferritin-like Superfamily. *J. Biol. Chem.* 287, 20565–20575.

Martin KE, Ozsvar J, Coleman N V. (2014) SmoXYB1C1Z of Mycobacterium sp. Strain NBB4: a Soluble Methane Monooxygenase (sMMO)-Like Enzyme, Active on C2 to C4 Alkanes and Alkenes. *Appl. Environ. Microbiol.* 80, 5801–5806.

McCormick MS, Lippard SJ. (2011) Analysis of Substrate Access to Active Sites in Bacterial Multicomponent Monooxygenase Hydroxylases: X-ray Crystal Structure of Xenon-Pressurized Phenol Hydroxylase from Pseudomonas sp. OX1. *Biochemistry* 50, 11058–11069.

McGlynn SE. (2017) Energy Metabolism during Anaerobic Methane Oxidation in ANME Archaea. *Microbes Environ.* 32, 5–13.

Le Mer J, Roger P. (2001) Production, oxidation, emission and consumption of methane by soils: A review. *Eur. J. Soil Biol.* 37, 25–50.

Merkx M, Kopp D a., Sazinsky MH, Blazyk JL, Muandller J, Lippard SJ. (2001) Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: A tale of two irons and three proteins. *Angew. Chemie - Int. Ed.* 40, 2782–2807.

Merkx M, Lippard SJ. (2002) Why OrfY? Characterization of MMOD, a long overlooked component of the soluble methane monooxygenase from Methylococcus capsulatus (Bath). *J. Biol. Chem.* 277, 5858.

Meyer F, Keller P, Hartl J, Gröninger OG, Kiefer P, Vorholt JA. (2018) Methanol-essential growth of Escherichia coli. *Nat. Commun.* 9.

Michelet L, Zaffagnini M, Morisse S, Sparla F, Pérez-Pérez ME, Francia F, Danon A, Marchand CH, Fermani S, Trost P, et al. (2013) Redox regulation of the Calvin–Benson cycle: something old, something new. *Front. Plant Sci.* 4, 1–21.

Miyata A, Yoshida T, Yamaguchi K, Yokoyama C, Tanabe T, Toh H, Mitsunaga T, Izumi Y. (1993) Cloning and expression of the gene for serine hydroxylmethyltransferase from an obligate methylotroph Hyphomicrobium methylovorum GM2. *Eur. J. Biochem.* 212, 745–750.

Murray LJ, Lippard SJ. (2007) Substrate trafficking and dioxygen activation in bacterial multicomponent monooxygenases. *Acc. Chem. Res.* 40, 466–474.

Murrell JC, Gilbert B, McDonald IR. (2000) Molecular biology and regulation of methane monooxygenase. *Arch. Microbiol.* 173, 325–332.

Myronova N, Miyaji A, Dalton H, Kitmitto A, Collins RF. (2006) Three-dimensional structure determination of a protein supercomplex that oxidizes methane to formaldehyde in Methylococcus capsulatus (Bath). *Biochemistry* 45, 11905–11914.

Nielsen AK, Gerdes K, Murrell JC. (1997) Copper dependent reciprocal transcriptional regulation of methane monooxygenase genes in Methylococcus capsulatus and Methylosinus trichosporium. *Mol. Microbiol.* 25, 399–409.

Nordlund P, Eklund H. (1995) Di-iron—carboxylate proteins. *Curr. Opin. Struct. Biol.* 5, 758–766.

Norton JM, Alzerreca JJ, Suwa Y, Klotz MG. (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch. Microbiol.* 177, 139–149.

Notomista E, Lahm A, Di Donato A, Tramontano A. (2003) Evolution of Bacterial and Archaeal Multicomponent Monooxygenases. *J. Mol. Evol.* 56, 435–445.

Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF. (2005) GroEL1: A dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* 123, 861–873.

Olsen RH, Kukor JJ, Kaphammer B. (1994) A novel toluene-3-monooxygenase pathway cloned from Pseudomonas pickettii PKO1. *J. Bacteriol.* 176, 3749–3756.

Orita I, Sato T, Yurimoto H, Kato N, Atomi H, Imanaka T, Sakai Y. (2006) The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon Thermococcus kodakaraensis. *J. Bacteriol.* 188, 4698–4704.

Park D, Lee J. (2013) Biological conversion of methane to methanol. *Korean J. Chem. Eng.* 30, 977–987.

Petersen JM, Dubilier N. (2009) Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* 1, 319–335.

Pikus JD, Studts JM, Achim C, Kauffmann KE, Münck E, Steffan RJ, McClay K, Fox BG. (1996) Recombinant toluene-4-monooxygenase: Catalytic and Mossbauer studies of the purified diiron and Rieske components of a four-protein complex. *Biochemistry* 35, 9106–9119.

Pol A, Barends TRM, Dietl A, Khadem AF, Eygensteyn J, Jetten MSM, Op den Camp HJM. (2014) Rare earth metals are essential for methanotrophic life in volcanic mudpots. *Environ. Microbiol.* 16, 255–264.

Powlowski J, Sealy J, Shingler V, Cadieux E. (1997) On the role of DmpK, an auxiliary protein associated with multicomponent phenol hydroxylase from Pseudomonas sp. strain CF600. *J. Biol. Chem.* 272, 945–951.

Pulver S, Froland W a, Fox BG, Lipscomb JD, Solomon EI. (1994) Spectroscopic studies of the coupled binuclear non-heme iron active site in the fully reduced hydroxylase component of methane monooxygenase: comparison to deoxy and deoxy-azide hemerythrin. *J. Am. Chem. Soc.* 116, 4529.

Raghoebarsing A a, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts M, Derksen J, Jetten MSM, Schouten S, Sinninghe Damsté JS, Lamers LPM, et al. (2005) Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* 436, 1153–1156.

Rasigraf O, Kool DM, Jetten MSM, Sinninghe Damsté JS, Ettwig KF. (2014) Autotrophic carbon dioxide fixation via the Calvin-Benson-Bassham cycle by the denitrifying methanotroph "Candidatus Methylomirabilis oxyfera." *Appl. Environ. Microbiol.* 80, 2451–2460.

Redmond MC, Valentine DL, Sessions AL. (2010) Identification of novel methane-, ethane-, and propane-oxidizing bacteria at marine hydrocarbon seeps by stable isotope probing. *Appl. Environ. Microbiol.* 76, 6412–6422.

Reeburgh WS. (2007) Oceanic methane biogeochemistry. *Am. Chem. Soc.* 107, 486–513.

Reeburgh WS. (2007) Oceanic Methane Biogeochemistry. *Chem. Rev.* 107, 486–513.

Repeta DJ, Ferrón S, Sosa OA, Johnson CG, Repeta LD, Acker M, Delong EF, Karl DM. (2016) Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nat. Geosci.* 9, 884–887.

Rinaldo D, Friesner RA, Philipp DM, Lippard SJ. (2007) Intermediates in dioxygen activation by methane monooxygenase: A QM/MM study. *J. Am. Chem. Soc.* 129, 3135–3147.

Rosenzweig AC. (2015) Breaking methane. *Nature* 518.

Rosenzweig AC, Brandstetter H, Whittington D a., Nordlund P, Lippard SJ, Frederick C a. (1997) Crystal structures of the methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath): Implications for substrate gating and component interactions. *Proteins Struct. Funct. Genet.* 29, 141–152.

Ross MO, Rosenzweig AC. (2017) A tale of two methane monooxygenases. *J. Biol. Inorg. Chem.* 22, 1173–1178.

Rubin-Blum M, Antony CP, Borowski C, Sayavedra L, Pape T, Sahling H, Bohrmann G, Kleiner M, Redmond MC, Valentine DL, et al. (2017) Short-chain alkanes fuel mussel and sponge Cycloclasticus symbionts from deep-sea gas and oil seeps. *Nat. Microbiol.* 2.

Saibil H. (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* 14, 630–642.

Sazinsky MH, Dunten PW, McCormick MS, DiDonato A, Lippard SJ. (2006) X-ray structure of a hydroxylase-regulatory protein complex from a hydrocarbon-oxidizing multicomponent monooxygenase, Pseudomonas sp. OX1 phenol hydroxylase. *Biochemistry* 45, 15392–15404.

Sazinsky MH, Lippard S. (2005) Product bound structures of the soluble methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath): Protein motion in the alpha-subunit. *J. Am. Chem. Soc.* 127, 5814–5825.

Sazinsky MH, Lippard SJ. (2006) Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc. Chem. Res.* 39, 558–566.

Sazinsky MH, Lippard SJ. (2015) Methane monooxygenase: Functionalizing methane at iron and copper. In: Metal Ions in Life Sciences. Vol. 15. p. 205–256.

Sazinsky MH, Merkx M, Cadieux E, Tang S, Lippard SJ. (2004) Preparation and X-ray structures of metal-free, dicobalt and dimanganese forms of soluble methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath). *Biochemistry* 43, 16263–16276.

Scanlan J, Dumont MG, Murrell JC. (2009) Involvement of MmoR and MmoG in the transcriptional activation of soluble methane monooxygenase genes in Methylosinus trichosporium OB3b. *FEMS Microbiol. Lett.* 301, 181–187.

Semrau JD, DiSpirito AA, Vuilleumier S. (2011) Facultative methanotrophy: false leads, true results, and suggestions for future research. *FEMS Microbiol. Lett.* 323, 1–12.

Semrau JD, DiSpirito AA, Yoon S. (2010) Methanotrophs and copper. *FEMS Microbiol. Rev.* 34, 496–531.

Semrau JD, Jagadevan S, DiSpirito AA, Khalifa A, Scanlan J, Bergman BH, Freemeier BC, Baral BS, Bandow NL, Vorobev A, et al. (2013) Methanobactin and MmoD work in concert to act as the 'copper-switch' in methanotrophs. *Environ. Microbiol.* 15, 3077–3086.

Servant P, Thompson CJ, Mazodier P. (1994) Post-transcriptional regulation of the groEL1 gene of Streptomyces albus. *Mol. Microbiol.* 12, 423–432.

Sharp JO, Sales CM, LeBlanc JC, Liu J, Wood TK, Eltis LD, Mohn WW, Alvarez-Cohen L. (2007) An inducible propane monooxygenase is responsible for N- nitrosodimethylamine degradation by Rhodococcus sp. strain RHA1. *Appl. Environ. Microbiol.* 73, 6930–6938.

Sharp JO, Wood TK, Alvarez-Cohen L. (2005) Aerobic biodegradation of N-nitrosodimethylamine (NDMA) by axenic bacterial strains. *Biotechnol. Bioeng.* 89, 608–618.

Singleton CM, McCalley CK, Woodcroft BJ, Boyd JA, Evans PN, Hodgkins SB, Chanton JP, Frolking S, Crill PM, Saleska SR, et al. (2018) Methanotrophy across a natural permafrost thaw environment. *ISME J.* 1, 1–15.

Sirajuddin S, Rosenzweig AC. (2015) Enzymatic Oxidation of Methane. *Biochemistry* 54, 2283–2294.

Skovran E, Palmer AD, Rountree AM, Good NM, Lidstrom ME. (2011) XoxF is required for expression of methanol dehydrogenase in Methylobacterium extorquens AM1. *J. Bacteriol.* 193, 6032–6038.

Smith SM, Rawat S, Telser J, Hoffman BM, Stemmler TL, Rosenzweig AC. (2011) Crystal structure and characterization of particulate methane monooxygenase from Methylocystis species strain M. *Biochemistry* 50, 10231–10240.

Smith TJ, Dalton H. (2004) Biocatalysis by methane monooxygenase and its implications for the petroleum industry. *Stud. Surf. Sci. Catal.* 151, 177–192.

Smith TJ, Murrell JC. (2009) Methanotrophs. In: Encyclopedia of Industrial Biotechnology. John Wiley & Sons, Inc.

Smith TJ, Slade SE, Burton NP, Murrell JC, Dalton H. (2002) Improved system for protein engineering of the hydroxylase component of soluble methane monooxygenase. *Appl. Environ. Microbiol.* 68, 5265–5273.

Song WJ, Gucinski G, Sazinsky MH, Lippard SJ. (2011) Tracking a defined route for O2 migration in

a dioxygen-activating diiron enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14795–14800.

Stafford GP, Scanlan J, McDonald IR, Murell JC. (2003) rpoN, mmoR and mmoG, genes involved in regulating the expression of soluble methane monooxygenase in Methylosinus trichosporium OB3b. *Microbiology* 149, 1771–1784.

Stein LY, Klotz MG. (2011) Nitrifying and denitrifying pathways of methanotrophic bacteria. *Biochem. Soc. Trans.* 39, 1826–1831.

Stein LY, Roy R, Dunfield PF. (2012) Aerobic Methanotrophy and Nitrification: Processes and Connections. *eLs*, DOI: 10.1002/9780470015902.a0022213.

Sullivan JP, Dickinson D, Chase HA. (1998) Methanotrophs, Methylosinus trichosporium OB3b, sMMO, and their application to bioremediation. *Crit. Rev. Microbiol.* 24, 335–373.

Sy A, Giraud E, Jourand P, Garcia N, Willems A, De Lajudie P, Prin Y, Neyra M, Gillis M, Boivin-Masson C, et al. (2001) Methylotrophic Methylobacterium bacteria nodulate and fix nitrogen in symbiosis with legumes. *J. Bacteriol.* 183, 214–220.

Sy A, Timmers ACJ, Knief C, Vorholt J a. (2005) Methylotrophic metabolism is advantageous for *Methylobacterium extorquens* during colonization of *Medicago truncatula* under competitive conditions. *Appl. Environ. Microbiol.* 71, 7245–7252.

Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG. (2011) A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ. Microbiol. Rep.* 3, 91–100.

van Teeseling MCF, Pol A, Harhangi HR, van der Zwart S, Jetten MSM, Op den Camp HJM, van Niftrik L. (2014) Expanding the verrucomicrobial methanotrophic world: Description of three novel species of Methylacidimicrobium gen. nov. *Appl. Environ. Microbiol.* 80, 6782–6791.

Theisen AR, Ali MH, Radajewski S, Dumont MG, McDonald IR, Murrell JC, Dunfield PF, Dedysh SN, Miguez CB. (2005) Regulation of methane oxidation in the facultative methanotroph Methylocella silvestris BL2. *Mol. Microbiol.* 58, 682–692.

Thiemer B, Andreesen JR, Schräder T. (2003) Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in Pseudonocardia sp. strain K1. *Arch. Microbiol.* 179, 266–277.

Tinberg CE, Lippard SJ. (2009) Revisiting the mechanism of dioxygen activation in soluble methane monooxygenase from M. capsulatus (Bath): Evidence for a multi-step, proton-dependent reaction pathway. *Biochemistry* 48, 12145–12158.

Tinberg CE, Lippard SJ. (2010) Oxidation reactions performed by soluble methane monooxygenase hydroxylase intermediates H peroxo and Q proceed by distinct mechanisms. *Biochemistry* 49, 7902–7912.

Tinberg CE, Lippard SJ. (2011) Dioxygen activation in soluble methane monooxygenase. *Acc. Chem. Res.* 44, 280–288.

Tinberg CE, Song WJ, Izzo V, Lippard SJ. (2011) Multiple roles of component proteins in bacterial multicomponent monooxygenases: Phenol hydroxylase and toluene/ o-xylene monooxygenase from Pseudomonas sp. OX1. *Biochemistry* 50, 1788–1798.

Torres Pazmiño DE, Winkler M, Glieder a., Fraaije MW. (2010) Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.* 146, 9–24.

Trotsenko YA, Ivanova EG, Doronina N V. (2001) Aerobic methylotrophic bacteria as phytosymbionts. *Microbiology* 70, 623–632.

Tsien HC, Brusseau G a., Hanson RS, Wackett LP. (1989) Biodegradation of trichloroethylene by Methylosinus trichosporium OB3b. *Appl. Environ. Microbiol.* 55, 3155–3161.

Versantvoort W, Guerrero-Cruz S, Speth DR, Frank J, Gambelli L, Cremers G, van Alen T, Jetten MSM, Kartal B, Op den Camp HJM, et al. (2018) Comparative genomics of Candidatus Methylomirabilis species and description of Ca. Methylomirabilis lanthanidiphila. *Front. Microbiol.* 9, 1–10.

Viitanen P V, Gatenby a a, Lorimer GH. (1992) Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. *Protein Sci.* 1, 363–369.

Villeneuve C, Martineau C, Mauffrey F, Villemur R. (2013) Methylophaga nitratireducenticrescens sp. nov. and Methylophaga frappieri sp. nov., isolated from the biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome. *Int. J. Syst. Evol. Microbiol.* 63, 2216–2222.

Vorobev A, Beck DAC, Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L. (2013) Comparative transcriptomics in three *Methylophilaceae* species uncover different strategies for environmental adaptation. *PeerJ* 1, e115.

Vorobev A V., Baani M, Doronina N V., Brady AL, Liesack W, Dunfield PF, Dedysh SN. (2011) Methyloferula stellata gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. *Int. J. Syst. Evol. Microbiol.* 61, 2456–2463.

Vuilleumier S, Chistoserdova L, Lee MC, Bringel F, Lajus A, Yang Z, Gourion B, Barbe V, Chang J, Cruveiller S, et al. (2009) Methylobacterium genome sequences: A reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. *PLoS One* 4.

Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS. (2002) Directed Evolution of Substrate-Optimized GroEL/S Chaperonins. *Cell* 111, 1027–1039.

Wang W, Iacob RE, Luoh RP, Engen JR, Lippard SJ. (2014) Electron transfer control in soluble methane monooxygenase. *J. Am. Chem. Soc.* 136, 9754–9762.

Wang W, Lippard SJ. (2014) Diiron oxidation state control of substrate access to the active site of soluble methane monooxygenase mediated by the regulatory component. *J. Am. Chem. Soc.* 136, 2244–2247.

West CA, Salmond GPC, Dalton H, Murrell JC. (1992) Functional expression in Escherichia coli of proteins B and C from soluble methane monooxygenase of Methylococcus capsulatus (Bath). *J. Gen. Microbiol.* 138, 1301–1307.

Whited GM, Gibson DT. (1991) Toluene-4-monooxygenase, a three-component enzyme system that catalyzes the oxidation of toluene to p-cresol in Pseudomonas mendocina KR1. *J. Bacteriol.* 173, 3010–3016.

Whittington D a., Lippard SJ. (2001) Crystal structures of the soluble methane monooxygenase hydroxylase from methylococcus capsulatus (Bath) demonstrating geometrical variability at the dinuclear iron active site. *J. Am. Chem. Soc.* 123, 827–838.

Wu ML, Ettwig KF, Jetten MSM, Strous M, Keltjens JT, Niftrik L van. (2011) A new intra-aerobic

metabolism in the nitrite-dependent anaerobic methane-oxidizing bacterium *Candidatus* ‘Methylomirabilis oxyfera.’ *Biochem. Soc. Trans.* 39, 243–248.

Yu Y, Ramsay JA, Ramsay BA. (2009) Production of soluble methane monooxygenase during growth of Methylosinus trichosporium on methanol. *J. Biotechnol.* 139, 78–83.

Zhang J, Wallar BJ, Popescu C V, Renner DB, Thomas DD, Lipscomb JD. (2006) Methane monooxygenase hydroxylase and B component interactions. *Biochemistry* 45, 2913–2926.

Zheng Y, Huang R, Wang BZ, Bodelier PLE, Jia ZJ. (2014) Competitive interactions between methane- and ammonia-oxidizing bacteria modulate carbon and nitrogen cycling in paddy soil. *Biogeosciences* 11, 3353–3368.

This page is intentionally blank

CHAPTER 3

THE ROLE OF SMMO IN METHANOTROPH EVOLUTION AND THE IMPLICATIONS FOR CONTEMPORARY ENZYME ACTIVITIES

This page is intentionally blank

Molecular Phylogenetics and Evolution 129 (2018) 171–181

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria

Craig D. Osborne, Victoria S. Haritos*

Department of Chemical Engineering, Monash University, Wellington Road, Clayton 3800, Australia

ARTICLE INFO

Keywords:
pMMO
sMMO
CuMMO
Co-evolution
Chemolithoautotroph
Greenhouse gas

ABSTRACT

The critical role that bacterial methanotrophs have in regulating the environmental concentrations of the potent greenhouse gas, methane, under aerobic conditions is dependent on monooxygenase enzymes which oxidise the substrate as both a carbon and energy source. Despite the importance of these organisms, the evolutionary origins of aerobic methane oxidation capability and its relationship to proteobacterial evolution is not well understood. Here we investigated the phylogenetic relationship of proteobacterial methanotrophs with related, non-methanotrophic bacteria using 16S rRNA and the evolution of two forms of methane monooxygenase: membrane bound (pMMO and pXMO) and cytoplasmic (sMMO). Through analysis we have concluded that extant proteobacterial methanotrophs evolved from up to five ancestral species, and that all three methane monooxygenase systems, pMMO, pXMO and sMMO, were likely present in the ancestral species (although pXMO and sMMO are not present in most of the present day methanotrophs). Here we propose that the three monooxygenase systems entered the ancestral species by horizontal gene transfer, with these likely to have pre-existing physiological and metabolic attributes that supported conversion to methanotrophy. Further, we suggest that prior to these enzyme systems developing methane oxidation capabilities, the membrane-bound and cytoplasmic monooxygenases were already both functionally and phylogenetically associated. These results not only suggest that sMMO and pXMO have a far greater role in methanotrophic evolution than previously understood but also implies that the co-inheritance of membrane bound and cytoplasmic monooxygenases have roles additional to that of supporting methanotrophy.

1. Introduction

Since 1750, atmospheric methane levels have increased by 2.5 fold, now making it the second most significant greenhouse gas after CO_2, accounting for 17% of the current increase to global warming potential (Allen, 2016). The methanotrophs, organisms capable of consuming methane as their carbon and energy source, have been identified as one of the key natural mechanisms for regulating these atmospheric methane concentrations (Semrau et al., 2011), with estimates that these organisms consume around 30 million tonnes per annum of atmospheric methane (Kolb, 2009; Smith et al., 2000). More importantly is the role of methanotrophs in preventing the initial release of methane to the atmosphere, in which it is believed that they can attenuate methane emissions by over 80% in some high generation biomes (Conrad and Rothfuss, 1991; Frenzel et al., 1992). Their ability to utilize methane as a sole energy source correlates with their identification in environments as diverse as swamps, rice paddies, oceans, deciduous woods and arctic saline lakes (Bowman et al., 1997; Hanson and Hanson, 1996; Park and Lee, 2013).

The methanotrophs form an important subsection of the much larger methylotrophic group, which grow on reduced carbon substrates lacking carbon-carbon bonds (Chistoserdova, 2011). Specific members of these methanotrophs can couple methane oxidation to the reduction of sulfate, iron, manganese, nitrate and nitrite, as well as the more typical oxygen (Bhattarai et al., 2017). Those that use anoxic methods for methane oxidation remain poorly characterised, with strains specific to the majority of these oxidants yet to be isolated in pure culture (Chistoserdova et al., 2005). This is contrasted by the enormous body of literature detailing the biochemistry, metabolic pathways, physiology and ecology of the aerobic methanotrophs.

Abbreviations: AMO, Ammonia monooxygenase; AOB, Ammonia oxidising bacteria; BMM, Bacterial multicomponent monooxygenase; CuMMO, Copper membrane monooxygenase; HGT, Horizontal gene transfer; ICM, Intracytoplasmic membrane; pMMO, Particulate methane monooxygenase; pXMO, Uncharacterised membrane monooxygenase; sMMO, Soluble methane monooxygenase

* Corresponding author.
E-mail address: victoria.haritos@monash.edu (V.S. Haritos).

https://doi.org/10.1016/j.ympev.2018.08.010
Received 8 May 2018; Received in revised form 1 August 2018; Accepted 19 August 2018
Available online 24 August 2018

Ever since detailed characterisation of methanotrophs began over 60 years ago (Dworkin and Foster, 1956), the classification scheme constantly evolved to accommodate new findings in the field. Early classification primarily used intracytoplasmic membrane morphology, fatty acid composition and methane assimilation pathway to differentiate three different methanotrophic 'types' (Hanson and Hanson, 1996): I, II, and X. Subsequent revisions to this system have typically incorporated the type X into type I (usually as type Ib) (Danilova et al., 2016). The classification into either type I or type II delineates methanotrophs at the taxonomic class level, that is, derived from either the Gammaproteobacteria or Alphaproteobacteria (Stein et al., 2012). More recent findings have also revealed methanotrophs outside the Proteobacteria including Verrucomicrobia, for which an additional type III has been ascribed, and members of the only known bacterial intra-aerobic methanotrophs of the NC10 phylum (typified by the Methylomirabilis genus), which have yet to be incorporated into the classification system (Knief, 2015).

All bacterial methanotrophs rely on one of two forms of the enzymatic methane monooxygenase system, in conjunction with molecular oxygen and a reducing equivalent, to perform the initial oxidation of methane to methanol (Sirajuddin and Rosenzweig, 2015). The membrane bound form of this enzyme, particulate methane monooxygenase (pMMO), was long thought to be universal within the bacterial methanotrophs (Holmes et al., 1995). Recent discoveries have demonstrated this not strictly correct, with the alternative cytoplasmic form of the methane monooxygenase, soluble methane monooxygenase (sMMO) being the sole methane oxidising unit in several species (Dunfield et al., 2003; Vorobev et al., 2011). Despite this, those containing only pMMO are the dominant form of methanotrophs identified to date, with those that exhibit both enzyme systems typically demonstrating a strong functional preference towards pMMO (Semrau et al., 2013).

It has long been acknowledged that pMMO and the ammonia monooxygenase (AMO) critical to the metabolic lifestyle of ammonia oxidising bacteria (AOB) are closely related (Holmes et al., 1995). Both belong to a larger family of copper membrane monooxygenases (CuMMO), also containing less characterised members including pXMO, also known to be present in some methanotrophic species (Tavormina et al., 2011). sMMO by contrast, exists in a much larger and more elucidated family of bacterial multicomponent monooxygenases (BMM) known for its high diversity of substrates (Leahy et al., 2003; Notomista et al., 2003).

The occurrence of methanotrophy is restricted to relatively few genera in the Proteobacteria due to the metabolic difficulty for organisms to adopt or diversify from the 'specialist' lifestyle and these occur in mostly methanotroph-specific clades (Tamas et al., 2014). Despite the extensive effort and revisions of methanotroph classification (Knief, 2015; Semrau et al., 2010; Stein et al., 2012; Tamas et al., 2014; Trotsenko and Murrell, 2008) key questions of the evolution of methanotrophy and its distribution amongst Proteobacteria remain, such as, how methanotrophy arose in Proteobacteria, how far did it spread and what organisms were the likely proto-methanotrophs?

Here we investigated the phylogenetic relationship of proteobacterial methanotrophs with related, non-methanotrophic bacteria using 16S rRNA and the evolution of two forms of methane monooxygenase: membrane bound (pMMO and pXMO) and cytoplasmic (sMMO). Through such analysis we have determined the most parsimonious explanation for the current taxonomic arrangement of present day proteobacterial methanotrophs was five separate speciation events, with each event involving the horizontal transfer of three evolutionary distinct methane oxidation systems to ancestral species with select traits. These results place in a new light the physiological importance of each of the three systems, and point towards a more complex interplay between them that has both enabled and sustained methanotrophy.

2. Materials and methods

2.1. 16S phylogenetic analysis of Proteobacteria related to methanotrophs

Proteobacteria *Methylomonas methanica* (AF304196), *Methylococcus capsulatus* (AJ563935), *Methylohalobius crimeensis* (AJ581837), *Methylosinus trichosporium* (Y18947) and *Methylocella silvestris* (AJ491847) were selected as representatives of methanotrophic subgroups Ia-c and IIa-b, respectively. The subgroup archetypical behaviour used as the selection criteria for the first four representatives does not exist for the IIb subgroup, and therefore this species was selected as one of the subgroups best characterised members. 16S sequences for each species were obtained from the SILVA LTP 123 database (Sept 2015 release) (Yarza et al., 2010) using a localised BLASTn search of the V4 region, guided by positions 488–746 of *E. coli* as a reference frame (Yang et al., 2016). The closest 100 representatives for each reference sequence were selected, with the three Gammaproteobacteria and two Alphaproteobacteria searches then merged to produce the combined datasets containing 227 and 122 species, respectively. The two datasets contained all currently identified methanotrophs present in the SILVA LTP 123 dataset. All 16S sequences from both datasets were extracted from the SILVA LTP 123 database then aligned with MAFFT v7 (Katoh and Standley, 2013) using the Q-INS-i method (Katoh and Toh, 2008).

Phylogenetic analysis of the aligned datasets was initially conducted using a Bayesian Markov chain Monte Carlo method implemented in BEAST v1.8 (Drummond et al., 2012; Drummond and Rambaut, 2007) using a GTR + G + I model and an uncorrelated lognormal relaxed clock. Datasets were run for 60 million and 20 million generations for the Gammaproteobacteria and Alphaproteobacteria dataset, respectively, with sampling every 1000th generation. The maximum clade credibility tree and posterior probabilities was then evaluated using TreeAnnotator v1.8.2 (Drummond et al., 2012). Phylogenetic confidence for the Bayesian MCMC constructed phylogenies were also assessed using the Maximum Likelihood methodology implemented in RAxML (Stamatakis, 2014) using 1000 bootstrap replicates and the GTR + G + I model. Due to low bootstrap stability observed within both the Ia and Ib group, bootstrap replicates were manually inspected and the positioning of three species, *Methylosphaera hansonii, Methylogaea oryzae* and *Cycloclasticus pugetii* showed lability within and outside their associated methanotrophic cluster. After excising the three species from both the Bayesian reference tree and sequence dataset, reanalysis of bootstrap support resulted in significant improvement in phylogenetic stability. All phylogenetic results were visualised using the graphical phylogenetic viewer FigTree (tree.bio.ed.ac.uk/software/figtree/).

2.2. Phylogenetic analysis of MMO proteins

pmoB from *M. capsulatus* Bath (G1UBD1) was used in BLASTp analysis of the Joint Genome Institute IMG/ER database (https://img.jgi.doe.gov/) to identify CuMMO sequences. The use of complete genomes in this database enabled confirmation of the presence of all genes coding for subunits required for enzymatic activity and gene arrangement in the operon, as well as the determination of all monooxygenase genes of interest in the species.

The sequences were aligned using MAFFT v7 and the L-INS-i method, after which the dataset was manually curated to remove sequences with large indels. To avoid the domination of the dataset by frequently sequenced species, representative sequences from one strain per species were randomly selected for analysis. Where sequences were present from strains identified to genus only and species level, the former sequence was removed. Finally, any identical copies of the CuMMO gene contained within a specific genome were removed, with the resultant dataset reduced to 88 members. Phylogenetic analysis and post-processing was performed as outlined under 16S unit methodology

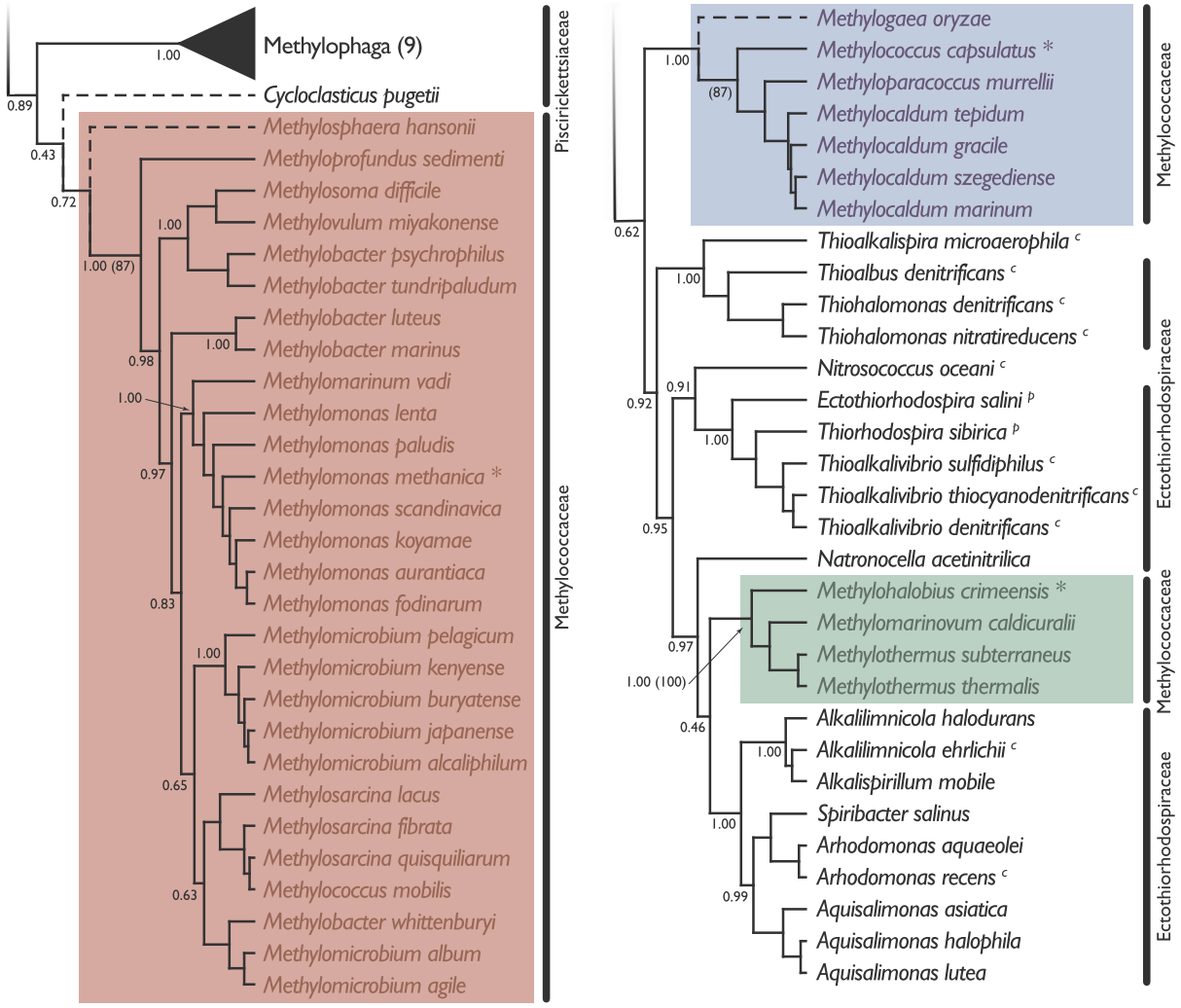


Fig. 1. Truncated gammaproteobacterial phylogeny featuring all methanotrophic species in the class and closely related taxa. The extended tree including accession numbers is given in Suppl Fig. S1. The phylogeny was based on Bayesian Markov chain Monte Carlo 16S phylogenetic analysis of 227 sequences most closely related to gammaproteobacterial methanotrophs. All methanotrophic species in the dataset were captured within the two phylogenetic subregions displayed. Methanotrophic species from Types Ia, b and c are coloured red, blue and green respectively, with the common phylogenetic region to methanotrophs of each type shaded in the respective colour. Posterior probabilities are displayed at all major nodes, with maximum likelihood values for key nodes in brackets. Dashed lines denote species removed for the maximum likelihood bootstrap analysis, performed on the Bayesian derived phylogenetic topology. The exclusive Methylophaga branch has been collapsed for clarity, with the number of species shown in parenthesis and main proteobacterial families are indicated adjacent to the phylogenies. * indicates the species selected as the exemplar of the methanotroph 'type' species for BLAST analysis. P and C superscript denote species capable of photoautotrophy and chemolithoautotrophy respectively.

with the exception of the Bayesian analysis which was conducted to 10 million generations, and the substitution model for both Bayesian and Maximum likelihood, for which the WAG + G + I model was used. The genome of randomly selected members of each phylogenetic cluster was examined, and the presence of all three CuMMO genes required for catalytic activity confirmed.

Sequence diversity of the catalytic subunit of BMM was obtained through BLASTp analysis of the IMG/ER database using a translation of *mmoX* (active subunit of sMMO) from *M. capsulatus* Bath (P22869) at 1e-10 cut-off limit. Preliminary alignment and phylogenetic analysis of the 500 sequences resulting from the search indicated that much of the larger BMM family was also captured in the search, with characterised representatives from the phenol hydroxylase, toluene monooxygenase and alkene monooxygenase-containing strains (Leahy et al., 2003) identified in the dataset. Examination of the phylogenetic cluster closest to sMMO (assumed by the inclusion of proteins from characterized methanotrophs) were two small groups of related BMM having characterized ethane and butane monooxygenase activities; both groups were highly stable in their location, which was adjacent to the main BMM phylogeny on the same branch as the sMMO cluster.

The sequences contained within the sMMO radiation and the two additional clusters as described above were excised from the original BLAST results. To this dataset, the *amoC* unit of *Gordonia rubripertincta* (formerly *Rhodococcus corallinus*) (Q53027) was added as the closest member to the sMMO cluster observed in previous BMM family analysis (Leahy et al., 2003; Notomista et al., 2003) and *bmoX* from *Thauera butanivorans* as the only fully characterised butane monooxygenase (Dubbels et al., 2007). For each member, the presence of regulatory and reductase components, as well as additional hydroxylase subunits were confirmed in the respective operons, thus confirming all units required for catalytic activity. The sequences were aligned as previously described using MAFFT v7 and the L-INS-i method, with subsequent removal of sequences with large indels and a *mmoX* gene fragment from *Plasmodium yoelii yoelii*. Phylogenetic analysis and post-processing was performed as outlined under 16S unit methodology with the exception of the Bayesian analysis which was conducted to 10 million generations, and the substitution model for both Bayesian and Maximum likelihood, for which the WAG + G + I model was used.

3. Results

3.1. Proteobacterial 16S analysis reveals five distinct methanotroph radiations amidst photo- and chemolitho-autotrophs and methylotrophs

For the examination of methanotrophic groupings and identification of related non-methanotrophic taxa, we used the SILVA LTP database as a high resolution neutrally curated 16S dataset (Yarza et al., 2010). This database had the notable absence of representatives of either the *Methylomirabilis* or *Methylacidiphilum* methanotrophic genera. These genera were not inserted into the dataset as it was desired to avoid manual editing of the database to prevent potential data biasing, and they were deemed unlikely to provide significant taxonomic clarification, nor identify any closely related non-methanotrophic species. This was based on the fact that they resided in the NC10 phylum and *Methyloacidiphilales* order respectively, both of which are candidate classifications and remain to be accurately taxonomically classified (Ettwig et al., 2010; Op den Camp et al., 2009). No additional absence of bacterial methanotrophic genera beyond these two instances were identified.

In our phylogenetic analysis of the datasets generated from collation of BLAST results around known proteobacterial methanotrophic species, methanotrophic Proteobacteria were observed to exist in two narrow phylogenetic lineages consistent with those previously described (Bowman, 2005; Tamas et al., 2014). Bayesian analysis conducted on the gammaproteobacterial type I dataset showed complete support for Ib and Ic sub-groups as monophyletic and exclusively methanotrophic (Fig. 1). Similarly, there was strong support for monophyletic methanotrophy for the Ia sub-group (posterior probability - 0.72), increasing to complete support when the unstable *Methylosphaera hansonii* was excluded (Fig. 1). ML bootstrap analysis also supported this conclusion following removal of several unstable species (Fig. 1). Although overall tree instability prevented clear elucidation of relationships between all type I gamma methanotrophic subgroups (Fig. S1), Bayesian analysis resolved the region immediately surrounding each subgroup. Whilst the phylogenetic relatives of the Ia group were dominated by the methylotrophic species from *Methylophaga*, the analysis supported their closest relative being the hydrocarbon-degrading *Cycloclasticus pugetii* (Fig. 1). The taxa surrounding the Ic group in the phylogenetic tree were almost exclusively species of the Ectothiorhodospiraceae, with strong representation of chemolithoautotrophic bacteria (Fig. 1) with two exceptions; *Thioalkalispira microaerophila*, an alkaliphilic sulphur oxidiser and *Nitrosococcus oceani,* the closest extant species to the original Ammonia Oxidising Bacteria (AOB) type strain (Klotz et al., 2006). There was also moderate support in the Bayesian analysis (posterior probability > 0.60) for the association of the type Ib methanotrophic clade with the Ectothiorhodospiraceae + Methylococcaceae (Ic) cluster, thus placing them in close phylogenetic proximity to the Ic methanotrophs (Fig. 1) but did not support the possibility of a single ancestor to both the Ib and Ic groups.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.08.010.

For the methanotrophic Alphaproteobacteria, phylogenetic analysis showed complete Bayesian support for the type IIa but not type IIb group as monophyletic and exclusively methanotrophic radiations (Fig. 2). Two non-methanotrophic species, *Methylovirgular ligni* and *Methylorosula polaris*, were present within the IIb radiation (Fig. 2) with high level support (posterior probability > 0.90), ruling out the possibility of the methanotrophic species forming an exclusive group within this cluster. Furthermore, there was complete Bayesian support for the presence of non-methanotrophic purple non-sulphur *Rhodoblastus* spp. between types IIa and IIb (Fig. 2) extending the lack of exclusive methanotrophic behaviour to the region between type IIa and IIb methanotrophs. According to the phylogeny, the closest relatives of type II methanotrophs were photo- and chemolitho-autotrophs: purple non-sulphur *Rhodoblastus* spp. and *Rhodoplanes* spp., sulphur oxidising

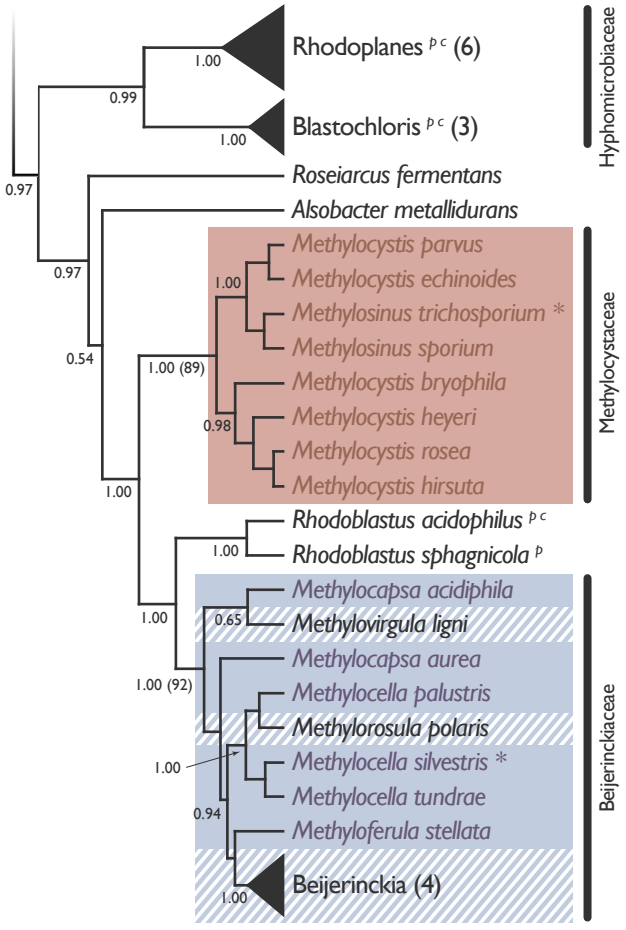


Fig. 2. Truncated alphaproteobacterial phylogeny based on 16S analysis featuring all methanotrophic species in the class and closely related taxa. The extended tree including accession numbers is given in Suppl Fig. S2. The phylogeny was based on Bayesian Markov chain Monte Carlo 16S phylogenetic analysis of 122 sequences most closely related to alphaproteobacterial methanotrophs. Methanotrophic species from Type IIa and b are coloured red and blue respectively, with the common phylogenetic region to methanotrophs of each type shaded in the respective colour. Hatched shading denotes regions within the common phylogenetic region with non-methanotrophic species present. Posterior probabilities are displayed at all major nodes. Dashed lines denote phylogenetic regions removed for ML bootstrap analysis of Bayesian derived phylogenetic topology, with results for key nodes included in parenthesis. Several non-methanotrophic branches exclusive at the genus level have been collapsed for clarity with the number of species in these branches denoted in parenthesis. The main proteobacterial families are indicated adjacent to the phylogenies. * indicates the species selected as the exemplar of the methanotroph 'type' species for BLAST analysis. P and C superscript denote species capable of photoautotrophy and chemolithoautotrophy respectively.

Blastochloris spp., purple non-sulfur *Roseiarcus fermentans* (Kulichevskaya et al., 2014) and the chemo-organotrophic *Alsobacter metallidurans* (Bao et al., 2014) (Fig. 2).

3.2. Two major radiations of methanotrophic CuMMOs

The IMG/ER database was selected for both CuMMO and BMM analysis due to the diversity of annotated full genome sequences which was crucial for the identification of additional members of the CuMMO or BMM families in the respective genomes. A BLAST analysis of this database for CuMMO members centred on *pmoB* from *M. capsulatus* Bath identified 312 genes in the database using an E value cut-off limit of 1e-10; increasing the threshold to 1e-5 resulted in just two additional sequences, both originating from the Archaeal genus *Methanosarcina*. As

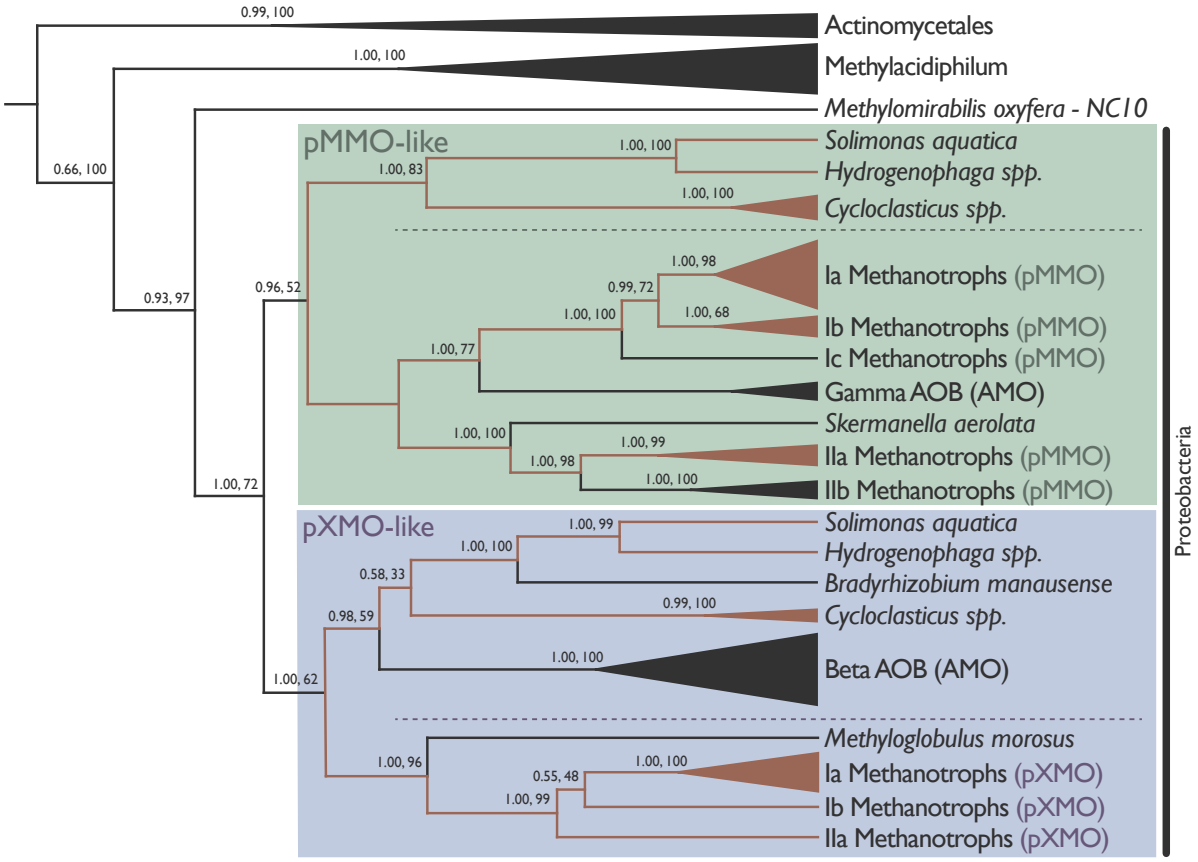


Fig. 3. Bayesian phylogenetic topology of translated CuMMO *pmoB* sequences from bacteria. The protein translation of 88 *pmoB* or homologous genes spanning the diversity of CuMMO in bacteria were obtained, aligned using MAFFT, and phylogeny constructed using BEAST v1.8. Several nodes were collapsed based on either congruence at the genus level or similar metabolic function. The first number at each node corresponds to the posterior probabilities determined from the Bayesian analysis, the second designates the ML bootstrap scoring for the same Bayesian maximum credibility tree using RaXML (WAG + G + I model). The regions designated as pMMO-like and pXMO-like are shown in green and blue respectively, with methanotrophs containing the namesake pMMO and pXMO genes identified in parenthesis. Dashed lines indicate the division between methanotroph-dominant and non-methanotroph functions. Branches where there is direct congruence between the pMMO-like and pXMO-like are indicated as red lines and the species representing Proteobacteria are shown to the right of the tree.

sequences from the Archaea are the most divergent members of the CuMMO family (Stein et al., 2012; Tavormina et al., 2011) this indicates the full scope of previously identified bacterial CuMMO species was captured within the search. However, due to the high level of divergence of the *Methanosarcina* CuMMO sequences, they were excluded from analysis to avoid potential long-branch attraction artefacts (Bergsten, 2005). To prevent any potential biasing of the dataset, all subsequent curation was performed using objective parameters applied to the whole dataset.

Phylogenetic analysis of the dataset revealed that the bacterial CuMMO phylogeny was divided into two branches at an ancestral node (Fig. 3, Supplementary Fig. 3): the first to diverge was a cluster of Actinomycetales CuMMO sequences, with the only previously identified activity being in *Mycobacterium chubuense* NBB4, with highest activity towards C_2 – C_4 alkanes (Coleman et al., 2012). The second CuMMO branch diverged further into 3: the oldest of these were proteins from the non-bacterial methanotrophic Verrucomicrobia (*Methylacidiphilum* spp.), next was a single representative of the NC10 lineage *Methylomirabilis oxyfera,* with the most recent branch containing all proteobacterial CuMMO proteins (Fig. 3). All organisms contained the three conserved CuMMO genes with all but the explicit pXMO cluster in methanotrophs retaining the canonical 'C-A-B' gene order (Baani and Liesack, 2008).

The proteobacterial CuMMO radiation divided into pMMO- and pXMO-like sequences as differentiated in Fig. 3. There was a striking symmetry in tree topology within the pMMO- and pXMO groups; each was split into two similar subtrees. In each subtree, one contained CuMMO proteins from predominantly methanotrophs and the other, from non-methanotrophs (divided by dotted lines, Fig. 3). The pMMO-like proteins included the well-characterised type I and II methanotrophic pMMO and the gammaproteobacterial AMO which clustered with the type I methanotrophic proteins. Previously unreported is the CuMMO of *Skermanella aerolata*, an alphaproteobacterium with no methanotrophic characteristics yet identified, which also clustered with the type II pMMO (Fig. 3). Interestingly, *Sk. aerolata* is not related to the type II methanotrophs based on 16S and this organism was positioned outside of our methanotrophic Alphaproteobacteria phylogenetic analysis (Suppl Fig. 2).

In addition to the many type Ia methanotrophs containing copies of both pMMO and pXMO, this was also observed in the sole representatives *Methylocaldum* sp. *175* and *Methylocystis rosea* from type Ib and type IIa methanotrophs, respectively. These findings, therefore, showed that the majority of methanotrophic proteobacterial types (3 out of 5) contain representatives that possess both pMMO and pXMO systems. Furthermore, while the overall tree topology of the pXMO sequences closely resembled that of the pMMO, it differed in *Methyloglobulus morosus*, which has a *pmo* gene and two non-synonymous *pxm* gene copies from separate origins. One pXMO protein resides within the type Ia (within the collapsed Ia cluster in Fig. 3) whilst the second protein lies in a single branch adjacent to the main methanotroph pXMO. Retaining dual CuMMO (pMMO- and pXMO-like) genes was not limited to methanotrophic species. CuMMO proteins from three non-methanotrophs: *Hydrogenophaga* spp. (Betaproteobacteria), *Solimonas aquatica* (Alphaproteobacteria), and an aromatic hydrocarbon-

degrading *Cycloclasticus spp* (Gammaproteobacteria) clustered together in both the pMMO and pXMO groups despite the host organisms being taxonomically distinct. The former was poorly taxonomically characterised in the JGI database, however contained a > 98% similarity of multiple universally conserved protein genes to *Hydrogenophaga taeniospiralis*. This, in conjunction with the 100% sequence identity of the *pmo*-like gene to that found in *Hydrogenophaga* sp. T4 resulted in to our current designation. Also found to cluster with the pXMO from non-methanotrophs was the CuMMO from the alphaproteobacterium *Bradyrhizobium manausense* and the betaproteobacterial AOB (Fig. 3).

3.3. Phylogenetic relationship of sMMO and closely related BMM

The evolutionary history of the alternative methane oxidising module to CuMMO in bacteria, the soluble methane monooxygenase (sMMO), was also investigated in this study. In examining the phylogeny of the larger BMM family in which sMMO resides, we limited the scope to the region immediately surrounding sMMO. To achieve this, the ethane monooxygenases from the Actinomycetales: *My. rhodesia NBB3* and *chubuense NBB4,* were used to delimit the BMM members from the original BLAST search results of the JGI database that are closely related but functionally distinct to sMMO. Similarly, the soluble butane monooxygenase from *Thauera butanivorans* was added to the analysis as a well-characterised, short chain hydrocarbon-active BMM that is closely related to sMMO. The resulting BMM alpha unit protein sequence phylogenetic tree included sequences from all identified methanotrophs, identified ethane and butane monooxygenases, and uncharacterised members from *So. aquatica* and *Sk. aerolata* (Suppl Fig. 4). Bayesian molecular clock analysis showed the ethane oxidising Actinomycetales were first to diverge from a putative ancestral species, followed by a division into two subsequent branches: one represented by *So. aquatica/Th. butanivorans* BMMs and the other leading to the methanotroph radiation (Suppl Fig. 4). The BMM protein from *Sk. aerolata* resided inside the larger methanotrophic cluster with low support for specific location (Suppl Fig. 4). However, there was certainty for its placement either directly adjacent to or within the alphaproteobacterial type IIb methanotrophic group in the cluster.

3.4. Alignment of phylogenetic trees for methane-oxidizing proteins

The protein sequence phylogenetic trees for CuMMO and BMM were simplified and aligned as shown in Fig. 4. It was clear that the trees based on protein sequence relationships did not agree with phylogeny based on 16S. Here, the methanotrophs grouped in accordance with their respective Class (Alphaproteobacteria and Gammaproteobacteria) but the remainder of the organisms were randomly ordered and did not follow expected evolutionary order (as compared with the inset diagram Fig. 4). This was most apparent with the non-methanotrophic region of the pXMO radiation, with the occurrence of sequences from both Betaproteobacteria and Gammaproteobacteria separated by representatives of a different class.

Aligning the phylogenetic trees for the three monooxygenase sequence families results in a complete agreement of topologies (Fig. 4), that is, where an organism shares either two or three different monooxygenases (pMMO-like, pXMO-like or BMM) the sequences have the same evolutionary history suggesting they were gained concurrently and have evolved together. Furthermore, comparison of the BMM and CuMMO protein sequence trees revealed a striking correlation where nearly all organisms containing BMM also had CuMMO-coding genes (Fig. 4). The exceptions were the sMMO-only methanotrophs from type IIb and *Th. butanivorans* which relies on its BMM for the short chain alkane oxidation essential to its metabolism (Dubbels et al., 2009; Sluis et al., 2002). Although sMMO was present in a minority of methanotrophs that also possessed CuMMO, four out of the five methanotrophic types (Ia, Ib, IIa, IIb) were represented by at least one species where both sMMO and CuMMO are present. Furthermore, organisms from methanotroph types Ia, Ib, IIa also contained pMMO and pXMO (Fig. 4) but only *Methylocaldum sp. 175* contained all three: sMMO, pMMO and pXMO, in the single organism (Suppl Fig. 3 and 4). Among the non-methanotrophs, *So. aquatica* also possessed genes coding for the 3 types of monooxygenase. The Actinomycetales, *My. rhodesia NBB3* and *My. chubuense NBB4,* have a BMM and a single CuMMO as the latter had diverged before the proposed duplication of the CuMMO gene in Proteobacteria (Fig. 3) and will be described further below.

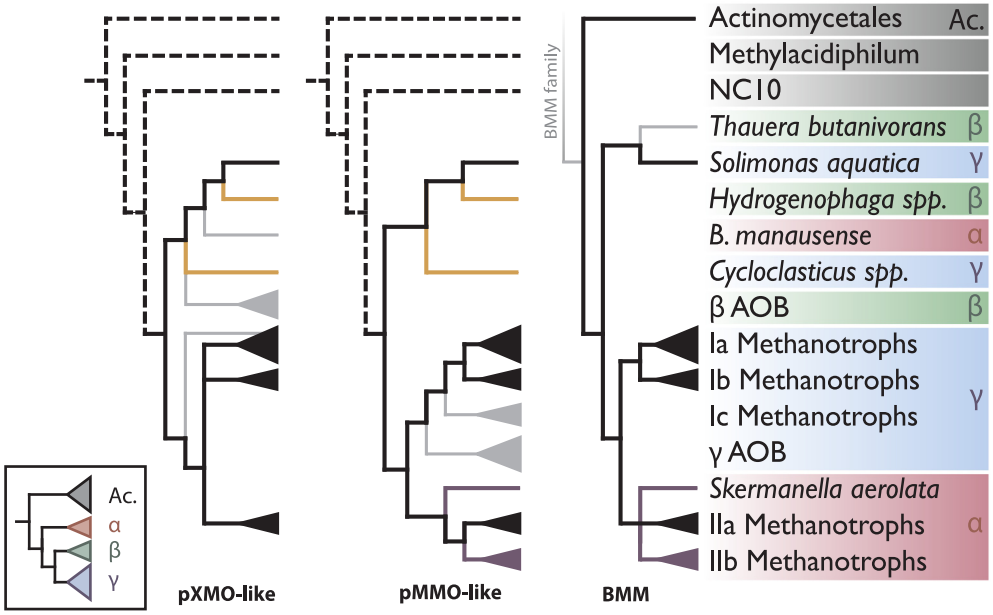


Fig. 4. Phylogenetic comparison of the pmo-like and pxm-like CuMMO and the sMMO related BMM protein sequence tree. pXMO-like and pMMO-like region were extracted from larger CuMMO analysis based on *pmoB* gene and equivalent homologues (Fig. 3 and Supp Fig. S3). The BMM phylogeny was generated in BEAST from aligned, translated gene sequences of putative di-iron centred subunits of BMM complexes closely related to sMMO. Branches with a common evolution in all three phylogenies are shown in black, whereas those common to pMMO-like and either pXMO-like or BMM are shown as orange and purple lines, respectively. Branches with no correlate in the comparative phylogenies are shown as grey lines. The outgroup used in the BMM tree, *amoC* from *Gordonia rubripertincta*, is a more distant relative in the BMM family; it therefore also indicates the location of the common branch to the larger BMM family (see Fig. S4). Dashed lines indicate the area of the CuMMO phylogeny prior to the pXMO-like and pMMO-like divergence and are therefore common to both trees. Representatives of methanotroph types and both classes of AOB have been collapsed based on the functional association of the CuMMO/BMM member in these groups. Taxonomic names and groups have been shaded based on bacterial class or order from which they belong, with the corresponding designation located to the far right of the figure. α – Alphaproteobacteria, β – Betaproteobacteria, γ – Gammaproteobacteria, Ac. – Actinomycetales. The inset illustrates the current consensus for evolutionary relationship of the Actinomycetales and proteobacterial classes and orders identified in the main figure (the location of NC10 and Methylacidiphilum have been excluded due to current taxonomic uncertainty).

4. Discussion

4.1. Methanotrophy in the Proteobacteria resides in five separate clades

The recent proliferation of genome information for Proteobacteria *via* mass sequencing highlights the limitations of a single level of differentiation within these methanotrophic groups, and thus numerous proposals for subdivisions within the types now exist (reviewed in Knief (2015)). We sought to resolve the current categorisation ambiguities by creating a proteobacterial methanotroph 16S phylogenetic distribution that also contained sequences from neighbouring non-methanotrophic species.

In our analysis, there was strong statistical support for five clades of proteobacterial methanotrophs in accordance with some of the most comprehensive classifications systems in the literature (Dumont et al., 2014). We have therefore maintained the naming convention (types Ia-c and IIa, b) used by those authors. 4 out of the 5 of these clades were found to be exclusively methanotrophic. Additionally, there is strong support for non-methanotrophs positioned between methanotrophic clades in the 16S phylogeny. In investigating the Beijerinkiacecae which include members of the type IIb methanotrophs, Tamas et al. (2014) also noted this family had both methanotrophic and non-methanotrophic members and proposed that the latter arose through rare and difficult reversion from that lifestyle. Our analysis provides the first clear demonstration of phylogenetic results showing the monophyletic nature of all 5 radiations, as well as the polyphyletic nature of these in relation to each other. The polyphyletic origin of methanotrophs has important implications for environmental biology and biotechnology where each radiation should be considered separately in functional analysis and in their potential for divergent behaviour.

4.2. The likely proto-methanotrophs

A characteristic almost universal in both the proteobacterial methanotrophs and AOB is the existence of ICM structures (Brantner et al., 2002; Fiencke and Bock, 2006) which provide sufficient (intra)cytoplasmic membrane capacity for the levels of the CuMMO membrane complex typically observed in these organisms (Sazinsky and Lippard, 2015). ICM with vesicular membrane disk morphology are present in almost all gammaproteobacterial methanotrophs, gamma AOB (e.g. *Ni. oceanus*) (Murray and Watson, 1965) and members of the purple sulfur bacteria from Ectothiorhodospira (Ramana et al., 2010; Remsen et al., 1968). These organisms were closely related to methanotrophs in our 16S phylogenetic analysis (Fig. 1), yet the role of the ICM in the metabolic processes for each of these groupings are markedly different. Similarly, ICM structures in the form of proliferated peripheral membrane layers characteristic of the type II methanotrophs are also observed in related organisms of the non-methanotrophic *Rhodoplanes* and *Rhodoblastus* genera (Hiraishi and Ueda, 1994; Kulichevskaya et al., 2006). Thus, we propose that methanotrophy arose in organisms that already contained suitable membrane structures to support the metabolic process such as the organisms that remain close in phylogenetic relationship to the current methanotrophs. One exception is *Cycloclasticus* spp., a genus closely related to the 1a methanotrophs. It contains CuMMO, however the presence of ICM structures has yet to be demonstrated. There are, perhaps, additional characteristics in related organisms that support CuMMO-based metabolic activity beyond membrane structures.

4.3. Co-inheritance of CuMMO and BMM

The CuMMO family has previously been shown to be dominated by members of the methane oxidising pMMO group, and ammonia oxidising AMO group. The family also contains numerous less characterised members, these include: pXMO, which is also present in methanotrophs and has been implicated, though yet to be directly shown, to oxidise methane (Hainbuch, 2015; Kits et al., 2015; Tavormina et al., 2011); highly divergent ammonia oxidising members present in archaeal species (Stein et al., 2012; Tavormina et al., 2011); and members optimised for short chain alkane activity (Coleman et al., 2012). Whilst previous analysis has demonstrated the overall phylogenetic topology of these groups (Tavormina et al., 2011), by also including uncharacterised members of this family our analysis revealed previously unidentified features of this family.

In our analysis, the CuMMO proteins from Proteobacteria were grouped in a single radiation, the most recent to diverge from the common ancestor, with those of the non-proteobacterial methanotrophs and ethane oxidising Actinomycetales diverging prior as previously reported (Stein et al., 2012; Tavormina et al., 2011). The gammaproteobacterial AMO were located inside the pMMO cluster of methanotrophic Proteobacteria whereas pXMO members formed a separate clade consistent with Tavormina et al. (2011). Newly identified in this study was the CuMMO of *Sk. aerolata* located inside the pMMO cluster. Also identified here, the presence of both pMMO- and pXMO-like proteins in the non-methanotrophs *So. aquatica, Hydrogenophaga* and *Cycloclasticus* shows that having dual CuMMO proteins is not exclusive to methanotrophs but suggests a fundamental characteristic of the CuMMO system for which having dual divergent copies may impart a significant physiological advantage.

The formation of two branches of diverged CuMMO complexes (pMMO and pXMO-like) in Proteobacteria, presumably arising from gene duplication, would have had to occur around the time of the transferral of the CuMMO operon into the first proteobacterial ancestor. Subsequent to duplication, both complexes underwent a common radiation pattern to generate similar pMMO- and pXMO-like subtree topologies. The most likely explanation for this result is that these two complexes were co-inherited, with the dual systems providing a greater evolutionary advantage than one of either complex.

The alternative system for methane oxidation in bacteria involves the cytoplasmic sMMO. The interchange between these two methane oxidising systems, pMMO and sMMO, has been observed to occur due to environmental copper concentrations, with sMMO expression induced at very low copper concentrations (Murrell et al., 2000; Prior and Dalton, 1985; Stanley et al., 1983). This equates with the necessity of copper to form the pMMO active site, compared to that of sMMO which utilises an iron centre (Wang et al., 2017).

The sMMO enzyme system exists in a much larger and more elucidated family in comparison to CuMMO, known as the bacterial multicomponent monooxygenases (BMM) (Leahy et al., 2003; Notomista et al., 2003). This family contains a high number of characterised members, with a demonstrated diverse substrate range including both saturated and unsaturated short chain hydrocarbons and aromatic compounds (Dubbels et al., 2007; Kotani et al., 2003; Martin et al., 2014; Sazinsky and Lippard, 2006). As previous analysis had demonstrated both the evolutionary origin of this family (Leahy et al., 2003; Lundin et al., 2012), as well as the phylogenetic interrelationship of members containing specific enzymatic capabilities (Leahy et al., 2003; Notomista et al., 2003) here we focussed on BMMs from organisms included in and adjacent to the sMMO radiation.

Alignment of the BMM phylogenetic tree with those of the two CuMMO radiations not only revealed a correlated phylogenetic history for the three monooxygenase systems within the Proteobacteria but also demonstrated that the ethane-degrading Actinomycetales contained CuMMO and BMM family proteins in a similar evolutionary relationship. This is significant as it suggests that in the period before the proposed duplication event giving rise to the pMMO- and pXMO-like groupings, the ancestral gene was associated with BMM. Also, that the CuMMO and BMM ancestral operons were co-inherited many times, with their number expanding to three independent monooxygenase systems after the CuMMO gene duplication event.

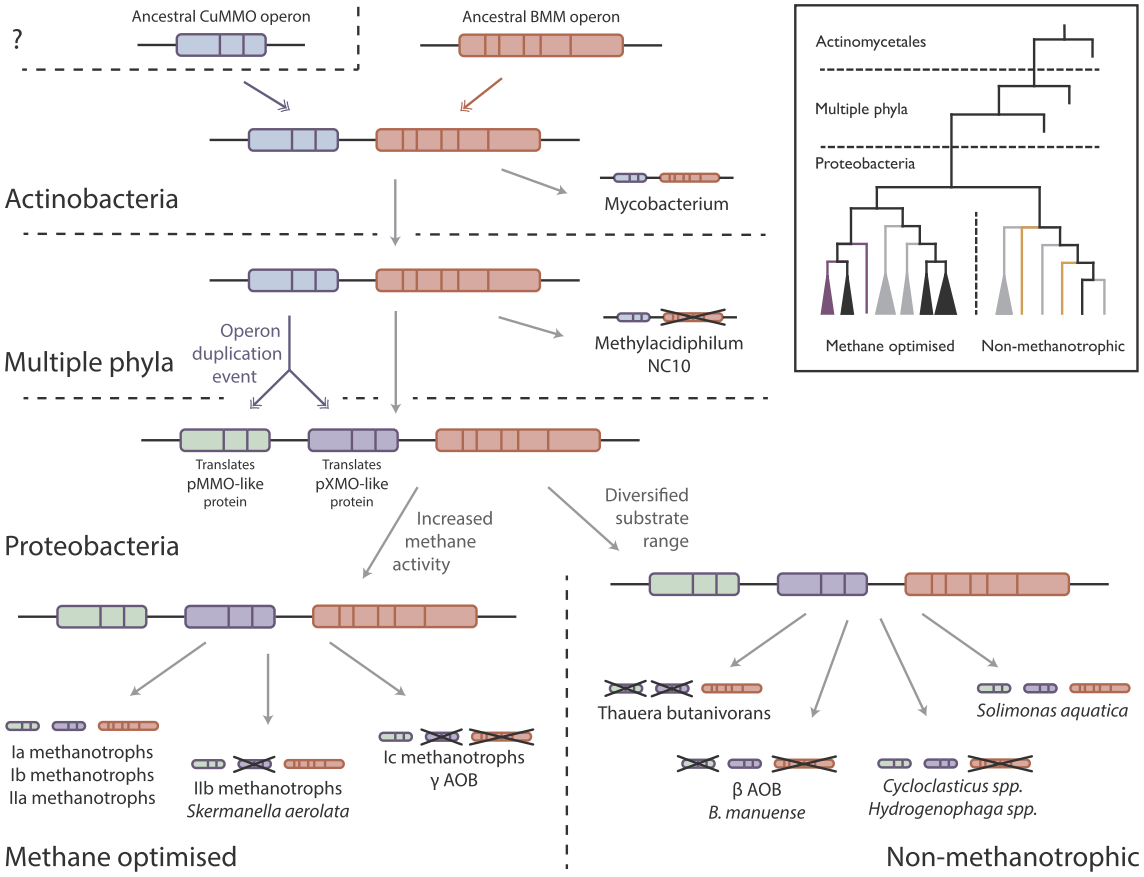


Fig. 5. Proposed evolutionary pathways and inheritance of CuMMO and BMM genes that explain the range of gene combinations found in extant methanotrophs and related species. The extant combinations can be explained by gain of the ancestral CuMMO/BMM operon through LGT, subsequent duplication of the CuMMO and gaining of methane activity, followed by a range of outcomes including: gene loss (pXMO, pMMO, or BMM) or further duplication and retention (pXMO). Coloured double headed arrows denote movement of respective coloured operon. Insert denotes combined phylogenetic history of CuMMO and BMM genes as outlined in Fig. 4. Gene cluster adjacency as well as operon order and divisions are for illustrative purposes only, with the later intended to infer the large multi-gene nature of these components, and not the explicit gene order occurring therein.

4.4. HGT accounts for spread of methanotrophy, not vertical descent

Previous investigations of CuMMO phylogeny that focussed on subsets of pMMO, pXMO and AMO suggested the protein phylogenies and bacterial evolution were correlated (Knief, 2015; Kolb et al., 2003; Op den Camp et al., 2009; Stein et al., 2012). This relationship was used to support the origin of aerobic methanotrophy in a common ancestor to the Alphaproteobacteria and Gammaproteobacteria (Battistuzzi et al., 2004). Even assuming an early instance of gene duplication to produce the pMMO- and pXMO-like subtrees, our analysis, in drawing from a wider CuMMO cohort, indicates numerous inconsistencies between the CuMMO protein sequence tree and proteobacterial species evolution (Fig. 4). These inconsistencies suggest a significant role of HGT in the spread of CuMMO and casts doubt on the notion of methanotrophy inherited in the Proteobacteria by vertical descent. HGT also provides a logical explanation for the presence of dual pXMO members in *Me. morosus* from different origins, which is difficult to explain by vertical descent.

Strong support for the origins of methanotrophy through HGT also comes from the identification of methanotrophic-related genes in *Sk. aerolata*. The genomic co-location of the particulate and soluble monooxygenase operons, absence of these genes in closely related species, and evolutionary relatedness of genes/proteins to comparable systems in the phylogenetically distant methanotrophs, each independently provides compelling evidence of a simultaneous HGT event of monooxygenases into this species. The location of the branch point of both particulate and soluble monooxygenase genes (Fig. 3 and Supp. Fig. 4) also indicates a divergence time in the vicinity of that for the proteobacterial methanotrophic speciation events. Collectively, this provides major support for the existence of a laterally transferrable element which included both sMMO and pMMO-coding operons that existed around the origin of proteobacterial methanotrophy. This is true regardless of whether the *pmo* and *bmm* genes present in *Sk. aerolata* are demonstrated to be functional and active on methane, the positive demonstration of which would have significant ramifications for our understanding of methanotrophy.

The 5 methanotroph types were identified as separate clades in our 16S phylogenetic tree with many cases of non-methanotrophs firmly located between these. Combining our phylogenetic analysis with the Tamas et al. (2014) proposition of an expected difficulty for organisms to revert from a methanotrophic lifestyle, the most logical conclusion is that methanotrophy in Proteobacteria arose from at least 5 independent HGT events. This contrasts with previous studies that have either suggested HGT events occurred in limited regions (Notomista et al., 2003; Tamas et al., 2014), or lacked incongruent evolutionary signals (Knief, 2015; Kolb et al., 2003; Stein et al., 2012). The likelihood of these methanotrophic clades arising from HGT into different ancestral species has implications including the importance of examining their functionality and behaviour as separate and raises questions about some current taxonomic classifications.

4.5. Functional role(s) for CuMMO systems?

Despite the clear differentiation of the pMMO- and pXMO subtrees

in the CuMMO phylogeny, and indication of their co-inheritance, little is known of the functional benefit of possessing two systems. While all AOB contain a single CuMMO complex with high specificity for ammonia, this can be either the pMMO- or pXMO-like groups depending on whether it is a gamma- or alphaproteobacterial AOB, respectively. Thus, gaining new substrate specificity e.g. for ammonia, is not unique to any one CuMMO type. Similarly, both pMMO and pXMO are found in obligate methanotrophic bacteria, and although pXMO has yet to be directly demonstrated to be active towards methane, there are indications that it has a role in methane oxidation under low oxygen tensions (Hainbuch, 2015; Hernandez et al., 2015; Kits et al., 2015). While pMMO and pXMO have overlapping substrate ranges and the initial duplication and subsequent divergence of the gene was unlikely to have been driven by differing substrate specificities, it does raise the possibility that the two systems have undergone a degree of specialisation to adapt to specific environmental conditions. Possessing versions of both systems may have supported survival under a more diverse and variable set of environmental conditions than that of a single CuMMO. The subsequent lack of monooxygenase diversity in most extant methanotrophs may therefore be the result of these species inhabiting more specific or stable ecological niches in the current day.

By contrast, much stronger substrate differentiation is apparent within the pMMO- and pXMO-like proteins, where one branch contains all proteobacterial methanotrophs and the other is devoid of them. Substrates for the non-methane utilizing enzymes include ammonia (beta AOB), aromatic compounds (*Cycloclasticus* spp. (Dyksterhouse et al., 1995)) and possibly hydrocarbons (*So. aquatica*). Importantly, the functional split towards methane or non-methane activities developed soon after the duplication event that gave rise to pMMO- and pXMO-like branches (Fig. 5).

4.6. Proposed evolutionary path of methanotrophy in Proteobacteria

The butane monooxygenase of *T. butanivorans* appears in our analysis as the only characterised member of the closest phylogenetic radiation to the sMMO cluster (Fig. 4); it is also the only BMM outside of the sMMO grouping known thus far to have activity towards methane (Cooley et al., 2009). The close proximity of the only other methane active BMM to sMMO supports the proposal of Leahy et al. (2003) that the ability to oxidise methane in the BMM is a relatively recent evolutionary event, occurring late in the branch leading to the sMMO radiation. All BMM sequences with methane oxidation activity reside within the Proteobacteria and the non-methane forms are mainly from Actinomycetales (Kotani et al., 2003; Leahy et al., 2003; Notomista et al., 2003). The closest phylogenetic radiation to methanotrophs containing Actinomycetales in both CuMMO and BMM evolutionary trees includes several species not only demonstrating the presence of genes for both systems but these are located tandemly in the genome. Furthermore, the CuMMO from the Actinomycetales species mimic the BMM in the organisms in being inactive towards methane (Coleman et al., 2012). It is therefore plausible that methane-inactive CuMMO and BMM genes became associated in an ancestral Actinomycetales with subsequent development of methane oxidation capability in both the particulate and soluble enzyme systems once the genes were transferred out of the Order.

The most parsimonious interpretations of our results is that once outside of the Actinomycetales, the CuMMO/BMM units evolved methane oxidation capability and this was transferred into the NC10 phylum and *Methylacidiphilum* sp., conferring methanotrophy (as proposed in Fig. 5). Following this, there was a duplication of the CuMMO operon resulting in the pMMO- and pXMO-like branches at a timepoint close to the appearance of these genes in Proteobacteria (Fig. 5). Divergence within each of the three monooxygenase systems resulted in one branch becoming more specialised to methane, with insertion of all three monooxygenase systems (pMMO, pXMO and sMMO) into at least five pre-methanotrophic ancestors resulting in methanotrophy in the Proteobacteria from which all currently identified forms descended from. Each ancestor must have either already contained, or inherited simultaneously, the metabolic systems to process the methanol product to avoid cytotoxicity issues and provide an evolutionary benefit. These systems could easily have been co-opted from pre-existing methylotrophs, with some instances of HGT already being demonstrated (Kalyuzhnaya et al., 2005). The alternative branch, whose proteobacterial members adopted or maintained more generalised monooxygenase activity, arose from multiple instances of CuMMO/BMM transfer resulting in the current diversity seen in the non-methanotrophs today (Fig. 5).

Key to this proposed evolutionary path is the concurrent inheritance of CuMMO and BMM-coding operons, with the later CuMMO diversification towards pMMO- and pXMO-like systems extending this co-inheritance to three functional monooxygenase systems, as described in Fig. 5. Although previous studies have investigated either the phylogeny of the BMM or subsections of the CuMMO family independently, it is by our undertaking a comprehensive study of these two systems that the previously undetected evolutionary connection between the two systems was revealed.

The combined size of the three complete operons is in excess of 10 kb, compared to the typical HGT event composed of short DNA segments of one to several genes (Wiedenbeck and Cohan, 2011), suggesting an exceptionally strong evolutionary driving force for the coinheritance of these systems. Furthermore, such event may not have been limited to genes for methane oxidation but may include pathways for processing the methanol produced. For example, type Ib gammaproteobacterial methanotrophs contain the required elements for the RuMP C1 processing pathway and the serine pathway, where the latter is more commonly associated with the Alphaproteobacteria (Hanson and Hanson, 1996). Large scale HGT events, although only constituting a small portion of overall HGT events, have still been noted to occur at appreciable rates in bacteria (Dougherty et al., 2014).

As the reconstructed evolutionary pathway in Fig. 5 shows, HGT of methane monooxygenase genes into non-methanotrophic but metabolically suitable pre-methanotrophs can explain the existence of all known types of methanotrophic Proteobacteria. It appears that some, if not all, of these methanotrophic HGT events involved the transfer of genes for three methane monooxygenase systems. However, there remains unsolved questions. Why were all three monooxygenase operons present in so many of the speciation events and what triggered both enzyme system's specialisation into methane oxidation? This is especially significant considering the low incidence of pXMO and sMMO operons and the dominance of pMMO in extant methanotrophs. Here we raise questions about the relationship between the membrane and soluble methane monooxygenases and about the main evolutionary drivers of the three methane monooxygenase systems, yet it does reconcile the apparent importance of all three in early methanotroph speciation in the Proteobacteria.

5. Conclusion

Based on several lines of analysis we have concluded that all currently known methanotrophs reside in one of five distinct radiations located in a narrow phylogenetic region of either the Alphaproteobacteria or Gammaproteobacteria, each of which we propose to have arisen from independent horizontal gene transfer events. Prior to these HGT events, the likely pre-methanotrophic ancestors were photo- and chemolithoautotrophic bacteria, for which the current day descendants remain closely phylogenetically related. Traits such as the morphology of ICMs appear to correlate more closely with these phytogenic relatives than with metabolic function.

The three different methane monooxygenases responsible for methane oxidation: membrane bound pMMO and pXMO and the cytoplasmic sMMO were likely present in most, if not all, of the ancestral species. This is surprising considering the low level of retention of both

the pXMO and sMMO systems in extant methanotrophs. Significantly, prior to the development of methane oxidation activity in these monooxygenases, the membrane bound and cytoplasmic monooxygenase systems were already both functionally and phylogenetically associated, suggesting their functional association plays a far more fundamental role beyond methanotrophy.

Acknowledgements

The authors wish to thank Dr Chris Greening for helpful advice on the manuscript and Assoc. Prof Ashley Buckle for general guidance.

Funding

This work was supported by a postgraduate research scholarship awarded to CDO by Monash University but the research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author's declaration of interests

None.

References

Allen, D., 2016. Attributing atmospheric methane to anthropogenic emission sources. Acc. Chem. Res. 49, 1344–1350. https://doi.org/10.1021/acs.accounts.6b00081.

Baani, M., Liesack, W., 2008. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. Proc. Natl. Acad. Sci. U. S. A. 105, 10203–10208. https://doi.org/10.1073/pnas.0702643105.

Bao, Z., Sato, Y., Fujimura, R., Ohta, H., 2014. Alsobacter metallidurans gen. nov., sp. nov., a thallium-tolerant soil bacterium in the order Rhizobiales. Int. J. Syst. Evol. Microbiol. 64, 775–780. https://doi.org/10.1099/ijs.0.054783-0.

Battistuzzi, F.U., Feijao, A., Hedges, S.B., 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol. Biol. 4, 44. https://doi.org/10.1186/1471-2148-4-44.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.

Bhattarai, S., Cassarini, C., Gonzalez-Gil, G., Egger, M., Slomp, C.P., Zhang, Y., Esposito, G., Lens, P.N.L., 2017. Anaerobic methane-oxidizing microbial community in a coastal marine sediment: anaerobic methanotrophy dominated by ANME-3. Microb. Ecol. 74, 608–622. https://doi.org/10.1007/s00248-017-0978-y.

Bowman, J.P., 2005. Methylococcales. In: Garrity, G., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey's Manual of Systematic Bacteriology. Springer, New York, NY, pp. 248–270. https://doi.org/10.1007/0-387-28022-7_7.

Bowman, J.P., McCammon, S.a., Skerratt, J.H., 1997. Methylosphaera hansonii gen. nov., sp. nov., a psychrophilic, group I methanotroph from Antarctic marine-salinity, meromictic lakes. Microbiology 143, 1451–1459. https://doi.org/10.1099/00221287-143-4-1451.

Brantner, C.A., Remsen, C.C., Owen, H.A., Buchholz, L.A., Collins, M., 2002. Intracellular localization of the particulate methane monooxygenase and methanol dehydrogenase in Methylomicrobium album BG8. Arch. Microbiol. 178, 59–64. https://doi.org/10.1007/s00203-002-0426-2.

Chistoserdova, L., 2011. Modularity of methylotrophy, revisited. Environ. Microbiol. 13, 2603–2622. https://doi.org/10.1111/j.1462-2920.2011.02464.x.

Chistoserdova, L., Vorholt, J.A., Lidstrom, M.E., 2005. A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. Genome Biol. 6, 208. https://doi.org/10.1186/gb-2005-6-2-208.

Coleman, N.V., Le, N.B., Ly, M.a., Ogawa, H.E., McCarl, V., Wilson, N.L., Holmes, A.J., 2012. Hydrocarbon monooxygenase in Mycobacterium: recombinant expression of a member of the ammonia monooxygenase superfamily. ISME J. 6, 171–182. https://doi.org/10.1038/ismej.2011.98.

Conrad, R., Rothfuss, F., 1991. Methane oxidation in the soil surface layer of a flooded rice field and the effect of ammonium. Biol. Fertil. Soils 12, 28–32. https://doi.org/10.1007/BF00369384.

Cooley, R.B., Dubbels, B.L., Sayavedra-Soto, L.a., Bottomley, P.J., Arp, D.J., 2009. Kinetic characterization of the soluble butane monooxygenase from Thauera butanivorans, formerly "Pseudomonas butanovora". Microbiology 155, 2086–2096. https://doi.org/10.1099/mic.0.028175-0.

Danilova, O.V., Suzina, N.E., Van De Kamp, J., Svenning, M.M., Bodrossy, L., Dedysh, S.N., 2016. A new cell morphotype among methane oxidizers: a spiral-shaped obligately microaerophilic methanotroph from northern low-oxygen environments. ISME J. 10, 2734–2743. https://doi.org/10.1038/ismej.2016.48.

Dougherty, K., Smith, B.A., Moore, A.F., Maitland, S., Fanger, C., Murillo, R., Baltrus, D.A., 2014. Multiple phenotypic changes associated with large-scale horizontal gene transfer. PLoS One 9. https://doi.org/10.1371/journal.pone.0102170.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. https://doi.org/10.1186/1471-2148-7-214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. https://doi.org/10.1093/molbev/mss075.

Dubbels, B.L., Sayavedra-Soto, L.A., Arp, D.J., 2007. Butane monooxygenase of "Pseudomonas butanovora": Purification and biochemical characterization of a terminal-alkane hydroxylating diiron monooxygenase. Microbiology 153, 1808–1816. https://doi.org/10.1099/mic.0.2006/004960-0.

Dubbels, B.L., Sayavedra-soto, L.A., Bottomley, P.J., Arp, D.J., 2009. Thauera butanivorans sp. nov., a C2–C9 alkane-oxidizing bacterium previously referred to as 'Pseudomonas butanovora'. Int. J. Syst. Evol. Microbiol. 1576–1578. https://doi.org/10.1099/ijs.0.000638-0.

Dumont, M.G., Lüke, C., Deng, Y., Frenzel, P., 2014. Classification of pmoA amplicon pyrosequences using BLAST and the lowest common ancestor method in MEGAN. Front. Microbiol. 5, 1–11. https://doi.org/10.3389/fmicb.2014.00034.

Dunfield, P.F., Khmelenina, V.N., Suzina, N.E., Trotsenko, Y.A., Dedysh, S.N., 2003. Methylocella silvestris sp. nov., a novel methonotroph isolated from an acidic forest cambisol. Int. J. Syst. Evol. Microbiol. 53, 1231–1239. https://doi.org/10.1099/ijs.0.02481-0.

Dworkin, M., Foster, J.W., 1956. Studies on Pseudomonas methanica (SÖHNGEN) nov. comb. J. Bacteriol. 72, 646–659.

Dyksterhouse, S.E., Gray, J.P., Herwig, R.P., Cano Lara, J., Staley, J.T., 1995. Cycloclasticus pugetii gen. nov., sp. now., an aromatic hydrocarbon-degrading bacterium from marine sediments. Int. J. Syst. Bacteriol. 45, 116–123. https://doi.org/10.1099/00207713-45-1-116.

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., Gloerich, J., Wessels, H.J.C.T., van Alen, T., Luesken, F., Wu, M.L., van de Pas-Schoonen, K.T., Op den Camp, H.J.M., Janssen-Megens, E.M., Francoijs, K.-J., Stunnenberg, H., Weissenbach, J., Jetten, M.S.M., Strous, M., 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature 464, 543–548. https://doi.org/10.1038/nature08883.

Fiencke, C., Bock, E., 2006. Immunocytochemical localization of membrane-bound ammonia monooxygenase in cells of ammonia oxidizing bacteria. Arch. Microbiol. 185, 99–106. https://doi.org/10.1007/s00203-005-0074-4.

Frenzel, P., Rothfuss, F., Conrad, R., 1992. Oxygen profiles and methane turnover in a flooded rice microcosm. Biol. Fertil. Soils 14, 84–89. https://doi.org/10.1007/BF00336255.

Hainbuch, S., 2015. Novel deep branching Cu-containing membrane-bound monooxygenases : distribution and function. Thesis 145.

Hanson, R.S., Hanson, T.E., 1996. Methanotrophic bacteria. Microbiol. Rev. 60, 439–471.

Hernandez, M.E., Beck, D.A.C., Lidstrom, M.E., Chistoserdova, L., 2015. Oxygen availability is a major factor in determining the composition of microbial communities involved in methane oxidation. e801. PeerJ 3. https://doi.org/10.7717/peerj.801.

Hiraishi, a., Ueda, Y., 1994. Rhodoplanes gen. nov., a New Genus of Phototrophic Bacteria Including Rhodopseudomonas rosea as Rhodoplanes roseus comb. nov. and Rhodoplanes elegans sp. nov. Int. J. Syst. Bacteriol. 44, 665–673. https://doi.org/10.1099/00207713-44-4-665.

Holmes, a.J., Costello, a., Lidstrom, M.E., Murrell, J.C., 1995. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. FEMS Microbiol. Lett. 132, 203–208. https://doi.org/10.1111/j.1574-6968.1995.tb07834.x.

Kalyuzhnaya, M.G., Korotkova, N., Marx, C.J., Lidstrom, M.E., Chistoserdova, L., Crowther, G., 2005. Analysis of gene islands involved in methanopterin-linked C 1 transfer reactions reveals new functions and provides evolutionary insights. J. Bacteriol. 187, 4607–4614. https://doi.org/10.1128/JB.187.13.4607.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298. https://doi.org/10.1093/bib/bbn013.

Kits, K.D., Klotz, M.G., Stein, L.Y., 2015. Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp. nov. type strain FJG1. Environ. Microbiol. 17, 3219–3232. https://doi.org/10.1111/1462-2920.12772.

Klotz, M.G., Arp, D.J., Chain, P.S.G., El-Sheikh, A.F., Hauser, L.J., Hommes, N.G., Larimer, F.W., Malfatti, S.A., Norton, J.M., Poret-Peterson, A.T., Vergez, L.M., Ward, B.B., 2006. Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium Nitrosococcus oceani ATCC 19707. Appl. Environ. Microbiol. 72, 6299–6315. https://doi.org/10.1128/AEM.00463-06.

Knief, C., 2015. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.01346.

Kolb, S., 2009. The quest for atmospheric methane oxidizers in forest soils. Environ. Microbiol. Rep. 1, 336–346. https://doi.org/10.1111/j.1758-2229.2009.00047.x.

Kolb, S., Knief, C., Stubner, S., Conrad, R., 2003. Quantitative detection of methanotrophs in soil by novel pmoA- targeted real-time PCR assays. Appl. Environ. Microbiol. 69, 2423–2429. https://doi.org/10.1128/AEM.69.5.2423.

Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y., Kato, N., 2003. Propane monooxygenase and NAD + -dependent secondary alcohol dehydrogenase in propane metabolism by Gordonia sp. strain TY-5. J. Bacteriol. 185, 7120–7128. https://doi.org/10.1128/JB.185.24.7120.

Kulichevskaya, I.S., Danilova, O.V., Tereshina, V.M., Kevbrin, V.V., Dedysh, S.N., 2014. Descriptions of Roseiarcus fermentans gen. nov., sp. nov., a bacteriochlorophyll a-containing fermentative bacterium related phylogenetically to alphaproteobacterial methanotrophs, and of the family Roseiarcaceae fam. nov. Int. J. Syst. Evol. Microbiol. 64, 2558–2565. https://doi.org/10.1099/ijs.0.064576-0.

Kulichevskaya, I.S., Guzev, V.S., Gorlenko, V.M., Liesack, W., Dedysh, S.N., 2006.

Rhodoblastus sphagnicola sp. nov., a novel acidophilic purple non-sulfur bacterium form Sphagnum peat bog. Int. J. Syst. Evol. Microbiol. 56, 1397–1402. https://doi.org/10.1099/ijs.0.63962-0.
Leahy, J.G., Batchelor, P.J., Morcomb, S.M., 2003. Evolution of the soluble diiron monooxygenases. FEMS Microbiol. Rev. 27, 449–479. https://doi.org/10.1016/S0168-6445(03)00023-8.
Lundin, D., Poole, A.M., Sjoberg, B.-M., Hogbom, M., 2012. Use of Structural Phylogenetic Networks for Classification of the Ferritin-like Superfamily. J. Biol. Chem. 287, 20565–20575. https://doi.org/10.1074/jbc.M112.367458.
Martin, K.E., Ozsvar, J., Coleman, N.V., 2014. SmoXYB1C1Z of Mycobacterium sp. Strain NBB4: a soluble methane monooxygenase (sMMO)-like enzyme, active on C2 to C4 alkanes and alkenes. Appl. Environ. Microbiol. 80, 5801–5806. https://doi.org/10.1128/AEM.01338-14.
Murray, R.G., Watson, S.W., 1965. Structure of nitrosocystis oceanus and comparison with nitrosomonas and nitrobacter. J. Bacteriol. 89, 1594–1609.
Murrell, J.C., McDonald, I.R., Gilbert, B., 2000. Regulation of expression of methane monooxygenases by copper ions. Trends Microbiol. 8, 221–225. https://doi.org/10.1016/S0966-842X(00)01739-X.
Notomista, E., Lahm, A., Di Donato, A., Tramontano, A., 2003. Evolution of bacterial and archaeal multicomponent monooxygenases. J. Mol. Evol. 56, 435–445. https://doi.org/10.1007/s00239-002-2414-1.
Op den Camp, H.J.M., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S., Jetten, M.S.M., Birkeland, N.K., Pol, A., Dunfield, P.F., 2009. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. Environ. Microbiol. Rep. 1, 293–306. https://doi.org/10.1111/j.1758-2229.2009.00022.x.
Park, D., Lee, J., 2013. Biological conversion of methane to methanol. Korean J. Chem. Eng. 30, 977–987. https://doi.org/10.1007/s11814-013-0060-5.
Prior, S.D., Dalton, H., 1985. The effect of copper ions on membrane content and methane monooxygenase activity in methanol-grown cells of Methylococcus capsulatus (Bath). Microbiology 131, 155–163. https://doi.org/10.1099/00221287-131-1-155.
Ramana, V.V., Sasikala, C., Ramaprasad, V.E.V., Ramana, C.V., 2010. Description of Ectothiorhodospira salini sp. nov. J. Gen. Appl. Microbiol. 56, 313–319. https://doi.org/https://doi.org/10.2323/jgam.56.313.
Remsen, C.C., Watson, S.W., Waterbury, J.B., Trüper, H.G., 1968. Fine structure of Ectothiorhodospira mobilis Pelsh. J. Bacteriol. 95, 2374–2392.
Sazinsky, M.H., Lippard, S.J., 2015. Methane monooxygenase: Functionalizing methane at iron and copper. Metal Ions in Life Sci. 205–256. https://doi.org/10.1007/978-3-319-12415-5-6.
Sazinsky, M.H., Lippard, S.J., 2006. Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. Acc. Chem. Res. 39, 558–566. https://doi.org/10.1021/ar030204v.
Semrau, J.D., DiSpirito, A.A., Vuilleumier, S., 2011. Facultative methanotrophy: false leads, true results, and suggestions for future research. FEMS Microbiol. Lett. 323, 1–12. https://doi.org/10.1111/j.1574-6968.2011.02315.x.
Semrau, J.D., DiSpirito, A.A., Yoon, S., 2010. Methanotrophs and copper. FEMS Microbiol. Rev. 34, 496–531. https://doi.org/10.1111/j.1574-6976.2010.00212.x.
Semrau, J.D., Jagadevan, S., DiSpirito, A.A., Khalifa, A., Scanlan, J., Bergman, B.H., Freemeier, B.C., Baral, B.S., Bandow, N.L., Vorobev, A., Haft, D.H., Vuilleumier, S., Murrell, J.C., 2013. Methanobactin and MmoD work in concert to act as the 'copper-switch' in methanotrophs. Environ. Microbiol. 15, 3077–3086. https://doi.org/10.1111/1462-2920.12150.
Sirajuddin, S., Rosenzweig, A.C., 2015. Enzymatic oxidation of methane. Biochemistry 54, 2283–2294. https://doi.org/10.1021/acs.biochem.5b00198.
Sluis, M.K., Sayavedra-Soto, L.a., Arp, D.J., 2002. Molecular analysis of the soluble butane monooxygenase from "Pseudomanas butanovora". Microbiology 148, 3617–3629.
Smith, K.A., Dobbie, K.E., Ball, B.C., Bakken, L.R., Sitaula, B.K., Hansen, S., Brumme, R., Borken, W., Christensen, S., Priemé, A., Fowler, D., Macdonald, J.A., Skiba, U., Klemedtsson, L., Kasimir-Klemedtsson, A., Degórska, A., Orlanski, P., 2000. Oxidation of atmospheric methane in Northern European soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. Glob. Chang. Biol. 6, 791–803. https://doi.org/10.1046/j.1365-2486.2000.00356.x.
Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stanley, S.H., Prior, S.D., Leak, D.J., Dalton, H., 1983. Copper stress underlies the fundamental change in intracellular location of methane mono-oxygenase in methane-oxidizing organisms: Studies in batch and continuous cultures. Biotechnol. Lett. 5, 487–492. https://doi.org/10.1007/BF00132233.
Stein, L.Y., Roy, R., Dunfield, P.F., 2012. Aerobic methanotrophy and nitrification: processes and connections. eLs DOI: 10.1002/9780470015902.a0022213. https://doi.org/10.1002/9780470015902.a0022213.
Tamas, I., Smirnova, A.V., He, Z., Dunfield, P.F., 2014. The (d)evolution of methanotrophy in the Beijerinckiaceae–a comparative genomics analysis. ISME J. 8, 369–382. https://doi.org/10.1038/ismej.2013.145.
Tavormina, P.L., Orphan, V.J., Kalyuzhnaya, M.G., Jetten, M.S.M., Klotz, M.G., 2011. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. Environ. Microbiol. Rep. 3, 91–100. https://doi.org/10.1111/j.1758-2229.2010.00192.x.
Trotsenko, Y.A., Murrell, J.C., 2008. Metabolic aspects of aerobic obligate methanotrophy. Adv. Appl. Microbiol. 63, 183–229. https://doi.org/10.1016/S0065-2164(07)00005-6.
Vorobev, A.V., Baani, M., Doronina, N.V., Brady, A.L., Liesack, W., Dunfield, P.F., Dedysh, S.N., 2011. Methyloferula stellata gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. Int. J. Syst. Evol. Microbiol. 61, 2456–2463. https://doi.org/10.1099/ijs.0.028118-0.
Wang, V.C., Maji, S., Chen, P.P., Lee, H.K., Yu, S.S., Chan, S.I., 2017. Alkane oxidation: methane monooxygenases, related enzymes, and their biomimetics. Chem. Rev. 117, 8574–8621. https://doi.org/10.1021/acs.chemrev.6b00624.
Wiedenbeck, J., Cohan, F.M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35, 957–976. https://doi.org/10.1111/j.1574-6976.2011.00292.x.
Yang, B., Wang, Y., Qian, P.-Y., 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinf. 17, 135. https://doi.org/10.1186/s12859-016-0992-y.
Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.H., Glöckner, F.O., Rosselló-Móra, R., 2010. Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. 33, 291–299. https://doi.org/10.1016/j.syapm.2010.08.001.

3.2. Introduction

3.2.1. The mixed gasses of methanotrophic biomes

The natural biomes in which methane is generated, either by biological or thermogenic means, also generates a range of additional reduced gasses including H_2, H_2S, NH_3 and short chain alkanes in varying amounts (Khalil 2000). Although these are usually present in trace gas concentrations, higher levels are not uncommon, with H_2 and H_2S in some locations exceeding 4.5% (45000ppm) and 8.0% (80000ppm) respectively (Sigvaldason and Elísson 1968; Dando et al. 1995). Anthropogenically produced biogas, although exhibiting concentrations lower than the extremes found in nature, have been demonstrated to exceed 5000ppm both in hydrogen sulphide and ammonia concentration in some systems (Pagans et al. 2006; Chen et al. 2015).

Such concentrations can have major impact on methanotroph growth and viability: a reduction of methane oxidation rate of over 65% has been shown for both hydrogen sulphide and ammonia concentrations as low as 1000ppm in some strains (Cáceres et al. 2014), with complete inhibition by hydrogen sulphide at 3000ppm (Zhang et al. 2016). Although strains exist with significantly higher tolerance, these have typically been identified through random screening (Nyerges and Stein 2009) or isolation from selected locations (Zhang et al. 2016). As the key gene/enzyme systems facilitating these metabolic pathways are known, an alternative approach to identifying organisms able to withstand the conditions using genetic analysis is also possible.

3.2.2. Genes as a proxy for function

Bacteria are known to exhibit a strong genomic deletional bias (Mira et al. 2001), creating a driver towards small, highly functional genomes whilst minimising the presence of pseudogenes and non-coding regions (Andersson and Andersson 2001). In light of this, whilst the presence of a particular gene does not explicitly demonstrate its activity in the host organism, it does indicate that the most likely reason for the retention of particular genetic material is that it provides an evolutionary advantage to the host, most likely through the functional activity of a protein it transcribes. Identification of candidate genes encoding highly similar sequences to characterised proteins creates a strong likelihood

that the organism is capable of the same function as that characterised in the protein, a fact regularly used in metabolic pathways identification from genomes (Kotera and Goto 2016). The likelihood of misidentification due to pseudogenes or highly functionally divergent genes can further be reduced through correlation of highly conserved residues, principally those involved in active site formation and metal binding, between characterised members and the candidate gene. *In vivo* demonstration of activity is ultimately required for confirmation of any particular functionality, however is often difficult due to uncertainty around regulatory mechanisms, and thus expression profiles (Yuan et al. 2007), making genetic analysis an expedient proxy for identification of enzymatic capabilities.

The proliferation of publicly available methanotroph genomes, especially in the last half decade, has enabled the identification of genetic capabilities and diversity within the methanotrophs to an extent not previously possible. This has resulted in the discovery of unexpected metabolic capabilities in some members of the methanotrophs, including growth of specific members on propane and hydrogen (Crombie and Murrell 2014; Carere et al. 2017). Recent work by Osborne and Haritos (2018) has implicated both ancestral and lateral gene transfer links between the methanotrophs and organisms that metabolise reduced gas substrates, especially chemolithotrophs. Therefore, potential metabolic capability of the methanotrophs towards these gasses was investigated through genomic analysis.

3.3. Method

Candidate methanotroph genomes were identified in the Joint Genome Institute (JGI) IMG/ER database (https://img.jgi.doe.gov/) based on the presence of either putative PmoB or MmoX sequences. These were then filtered to remove both the organisms not characterised down to the genus level, as well as limit the number of genomes from each identified species to one. In addition to this, *Skermanella aerolata KACC 11604* was removed as the only member appearing in the dataset to originate from a genus not previously identified to be associated with methanotrophy.

Production of the genomic table was, unless otherwise specified, performed using the same generalised procedure: Firstly, the literature was reviewed to identify characterised enzymes/proteins containing the function of interest. Where possible, the sequences of interest were derived from literature sources that

provided multiple phylogenetically diverse characterised enzymes, thus encompassing more of the evolutionary diversity for the set reactivity than a single representative. Consecutive BLASTp searches (E value 1e-5) using each of the identified reference sequences were then conducted using the previously identified methanotroph genomes in the JGI IMG/ER database. The BLASTp results and reference sequences were aligned in MAFFT v7 using the L-INS-i method (Katoh and Standley 2013). All fragment sequences were removed unless they came from a draft genome in which a single base indel resulted in a frameshift of what would otherwise result in a full length transcribed protein. In these cases the translated sequence was read from the genome after removal of the identified indel. Preliminary phylogenies were constructed in MEGA6 (Tamura et al. 2013), with results that clustered closer to sequences with an alternative functional activity, or lacking the required residues to form the catalytic active site also removed. The reduced dataset was then realigned in MAFFT v7 and the final phylogeny constructed in RaXML using the WAG+G+I methodology (Stamatakis 2014). The genomic table was then populated based on phylogenetic clustering relative to the reference sequences derived from literature.

3.4. Results

Examination of the JGI IMG/ER database revealed 72 genomes containing either a sMMO or pMMO operon. This included 37 unique methanotrophic species, as well as an additional 12 genomes with taxonomic classification at genus level. The remaining genomes were either not characterised to the genus level or were additional strains from an already represented species, and were removed from the dataset. The remaining genomes were sorted based on the most current methanotrophic classification system (Knief 2015; Osborne and Haritos 2018) into their respective groups and subgroups. Further demarcation was then made into clusters within these subgroups based on GroEL phylogeny (see table 1 and Appendix figure A7.5), most of which was shown to be in close agreement with taxonomic segregation.

For these genomes, analysis of methane oxidation capacity was performed for both the soluble (sMMO) and particulate (CuMMO) enzyme systems. For those genomes containing sMMO, none contained

more than a single copy of the operon. Divergent members were also almost non-existent, the only exception being in *M. silvestris* for which the divergent member could be assigned based on analysis conducted in chapter 4 as a propane monooxygenase (appearing in Table 3.1 under the alkane substrate heading). For the CuMMO analysis, it was noted that whilst *pmoA* and *pmoB* genes only ever occurred in instances of the full pMMO operon (*pmoC-A-B*), lone *pmoC* genes appear frequently, with four or more instances observed in some type Ib and IIa genomes (results not shown). Therefore, to limit the scope of this study to only the full CuMMO operon in instances of pMMO, PmoB was chosen for the reference sequence used during BLASTp searches.

CuMMO search results identified that all genomes apart from *M. silvestris* and *M. stellata* contained at least one main group pMMO, with dual copies of nearly identical genes frequently observed, especially in the type Ib and IIa methanotrophs. Two clusters not constituting main group pMMO sequences that have been previously identified in the literature also appeared in this analysis: pXMO (Tavormina et al. 2011) and pMMO2 (Baani and Liesack 2008). The former was clearly identified by its high level of divergence and distinct independent clustering in CuMMO phylogenetic analysis (Tavormina et al. 2011; Osborne and Haritos 2018). In contrast, differentiating pMMO2 sequences required greater judgement due to a much lower level of differentiation from conventional pMMO sequences. All pMMO2 members occurred in the type IIa methanotrophs, and their phylogenetic cluster appears

Table 3.1: Genetic analysis of aerobic methanotrophs for reduced gas handling systems. Each methanotrophic species is accompanied by representative strain designation used for genetic analysis. Species are differentiated into methanotrophic subgroups by heavy boarders, with light line dividers denoting phylogenetic clusters according to GroEL analysis (figure A1). Genes are categorised based on a functional hierarchical system, with substrate as the primary classification, followed by functional group, and lastly phylogenetic clusters as designated in the literature. Crosses denote gene presence, with multiple crosses in the same column representing multiple evolutionarily similar copies. Forward slashes represent divergent members closely associated with cluster. pMMO, particulate methane monooxygenase; pMMO2, divergent type II particulate methane monooxygenase; pXMO, uncharacterised membrane monooxygenase; sMMO, soluble methane monooxygenase; mxaF, calcium dependant methanol dehydrogenase; xoxF, lanthanide dependant methanol dehydrogenase; HAO, hydroxylamine oxidase; cytL, cytochrome P460; HCP, hybrid cluster protein; alkB, membrane bound alkane monooxygenase; BMM, bacterial multicomponent monooxygenase; PMO, propane monooxygenase; FCSD, favocytochrome C – sulphide dehydrogenase; SQR, sulphide quinone reductase; RuBisCO, Ribulose-1,5-bisphosphate carboxylase/oxygenase.

		Substrate	Methane				Methanol						Hydroxylamine			Alkane	
		Group	CuMMO			BMM	MxaF	XoxF					Oxidase		Red.	alkB	BMM
		Cluster	pMMO	pMMO2	pXMO	sMMO		F5	F5'	F3	F2	F1	HAO	cytL	HCP		PMO
Ia	Methylomonas methanica	MC09	X			X	X	X						X/			
	Methylomonas lenta	R-45370	X			X	X	X						X			
	Methylomonas koyamae	JCM 16701	X		X		X	X						X			
	Methylomonas sp.	LW13	X		X	X	X	X						X			
	Methylomonas sp.	MK1	X		XX	X	X	X						X			
	Methylomonas denitrificans	FGJ1	X					X						X			
	Methylomonas sp.	11b	X		X	X	X	X						X			
	Methyloglobulus morosus	KoM1	X		XXX		X	X					X	X		X	
	Methylosarcina fibrata	AML-C10	X		X		X	X					X			X	
	Methylosarcina lacus	LW14	X				X	X					X	X	X		
	Methylomicrobium agile	ATCC 35068	X		X		X	X					X	X		X	
	Methylomicrobium album	BG8	X		X		X	X					X	X		X	
	Methylobacter tundripaludum	SV96	X		X		X	X							X		
	Methylovulum miyakonense	HT12	X			X	X	X									
	Methylovulum psychrotolerans	HV10 M2	X				X	X									
	Methylobacter sp.	BBA5.1	X		X		X	X		X			X	X	X		
	Methylobacter luteus	IMV-B-3098T	X		X		X	X		X			X	X	X		
	Methylobacter marinus	A45	X		X		X	X		X			X	X	X		
	Methylobacter whittenburyi	ACM 3310	X		X		X	X		X			X	X	X		
	Methylomicrobium alcaliphilum	20Z	X				X	X					X	X			
	Methylomicrobium kenyense	AMO1	X				X	X					X	X			
	Methylomicrobium buryatense	5G	X			X	X	X					X	X	X		
	Methylomarinum vadi	IT-4/GN	X				X	X						XX		X	
Ib	Methylogaea oryzae	JCM 16910	X				X	X					X	X	X		
	Methyloterricola oryzae	73a	XX		X		X	X					X	X/		X	
	Methylomagnum ishizawai	175	XX		X	X	X	X						X		X	
	Methylococcus capsulatus	Bath	XX			X	X	X					X	X		X	
	Methylocaldum szegediense	O-12	X				X	X		X			X	X	X	X	
Ic	Methylohalobius crimeensis	10Ki	XX				X	X					X	X		X	
IIa	Methylosinus sp.	LW3	XX	X		X	X	X	X								
	Methylosinus sp.	PW1	XX	X		X	X	X	X								
	Methylosinus sp.	LW4	XX	X		X	X	X	X					/			
	Methylosinus sp.	R-45379	X	X	X	X	X	X	X								
	Methylocystis sp.	LW5	XX	X		X	X	X	X								
	Methylosinus trichosporium	OB3b	XX			X	X		X						X		
	Methylocystis parvus	OBBP	X	X			X	XX					X		X		
	Methylocystis sp.	Rockwell	XX				X	XX					X				
	Methylocystis sp.	SC2	XX	X			X	X		X			X		X		
	Methylocystis sp.	SB2	XX		X		X	X					X		X		
	Methylocystis rosea	SV97T	XX		X		X	X		X			X		X		
IIb	Methylocella silvestris	BL2				X	X	X		X		X		X			X
	Methyloferula stellata	AR4				X	X	XXX		X		X		X			
	Methylocapsa aurea	KYG T	X				X	X							X		
	Methylocapsa palsarum	NE2	X				X	X									
	Methylocapsa acidiphila	B2	X				X	X									
III	Methylacidiphilum fumarolicum	SolV	XXX								X		X				
	Methylacidiphilum kamchatkense	Kam1	XXX								X		X				
	Methylacidiphilum infernorum	V4	XXX								X		X				
	Methylomirabilis oxyfera sp.	Australia	X				X				X	X	/				

			Substrate	Hydrogen							Hydrogen sulphide				CO_2			
			Group	NiFe Hydrogenase							FCSD	SQR			RuBisCO			
			Cluster	1D	1E	1H	2A	2B	3B	3D		Type I	Type II	Type III	Green (IA)	Red (IC)	IE	II
Ia	Methylomonas methanica	MC09		X			X			X	X	X		X				
	Methylomonas lenta	R-45370								X	X	X		X				
	Methylomonas koyamae	JCM 16701					X			X	X	X						
	Methylomonas sp.	LW13		X						X	X	XX						
	Methylomonas sp.	MK1		X						X	X	XX						
	Methylomonas denitrificans	FGJ1		X						X	X	XX						
	Methylomonas sp.	11b		X	X					X	X	XX						
	Methyloglobulus morosus	KoM1											X					
	Methylosarcina fibrata	AML-C10								X			X					
	Methylosarcina lacus	LW14								X			X					
	Methylomicrobium agile	ATCC 35068					X			X			X					
	Methylomicrobium album	BG8					X			X			X					
	Methylobacter tundripaludum	SV96		X			X			X		X	X					
	Methylovulum miyakonense	HT12		X			X			X		X	X					
	Methylovulum psychrotolerans	HV10 M2		X			X			X		X	X					
	Methylobacter sp.	BBA5.1		X						X		X						
	Methylobacter luteus	IMV-B-3098T		X						X		X		X				
	Methylobacter marinus	A45		X						X		X						
	Methylobacter whittenburyi	ACM 3310		X						X		X						
	Methylomicrobium alcaliphilum	20Z					X			X		X	X	X				
	Methylomicrobium kenyense	AMO1								X		X						
	Methylomicrobium buryatense	5G					X			X		X		X				
	Methylomarinum vadi	IT-4/GN								X	X	X						
Ib	Methylogaea oryzae	JCM 16910		X			X			X		X	X					X
	Methyloterricola oryzae	73a		X	/					X	X	X	X					X
	Methylomagnum ishizawai	175		XX						X	X	X	X		X			
	Methylococcus capsulatus	Bath		X						X	X	X			X			
	Methylocaldum szegediense	O-12		X					XX		X		X		X			
Ic	Methylohalobius crimeensis	10Ki								X			XX					
IIa	Methylosinus sp.	LW3		X				X	X				X	X				
	Methylosinus sp.	PW1		X				X	X									
	Methylosinus sp.	LW4		X				X	X				X	X				
	Methylosinus sp.	R-45379		X				X	X				X	X				
	Methylocystis sp.	LW5		X				X	X				X	X				
	Methylosinus trichosporium	OB3b		X		X		X	X				X	X				
	Methylocystis parvus	OBBP		X		X			X			X						
	Methylocystis sp.	Rockwell		X				X	X			X						
	Methylocystis sp.	SC2		X		X		X	X			X						
	Methylocystis sp.	SB2				X			X			X						
	Methylocystis rosea	SV97T				X			X			X						
IIb	Methylocella silvestris	BL2					X			X						X		
	Methyloferula stellata	AR4					X			X		X				X		
	Methylocapsa aurea	KYG T				X	X					X				X		
	Methylocapsa palsarum	NE2										X				X		
	Methylocapsa acidiphila	B2					X						X			X		
III	Methylacidiphilum fumarolicum	SolV		X		X			X								X	
	Methylacidiphilum kamchatkense	Kam1		X		X			X					X			X	
	Methylacidiphilum infernorum	V4		X					X					X			X	
	Methylomirabilis oxyfera sp.	Australia														X		

immediately adjacent to that of conventional pMMO from IIa members (Appendix Figure A7.6). The two clusters can be differentiated at a phylogenetic level as the pMMO cluster contains representatives from all IIa methanotrophs, whilst the pMMO2 group contained only a smaller subset of IIa. Despite evidence that there are distinct expression profiles for each of the three copies of pMMO in *Verrucomicrobia*, it is not known whether these have true divergent functions, and therefore have all been tentatively placed as main group pMMO sequences due to a current lack of evidence to the contrary.

As pyrroloquinoline quinone (PQQ) dehydrogenases are known to be the principle methanol oxidisers in methanotrophs (Keltjens et al. 2014), the examination into methanol dehydrogenases was limited to this evolutionarily related family. The classic differentiation between the MxaFJ and XoxF systems in this larger family has recently been extended to show much more diversity in the XoxF members than previously identified (Chistoserdova 2011). The updated method for differentiation of this larger family, including differentiation of XoxF members in clusters 1-5 (Keltjens et al. 2014; Taubert et al. 2015), has been recapitulated in this work. The only addendum made to the current classification system was the inclusion of XoxF5' as a sub-group within in the overall XoxF5 cluster; it contained sequences from type IIa genomes and was located independently but immediately adjacent to the main group XoxF5 sequences from type IIa methanotrophs. XoxF5' could be readily differentiated due to it containing only a smaller subset of the IIa members, the manner and location strikingly similar to that found for pMMO2 sequences.

In addition to the propane monooxygenase in *M. silvestris* identified during sMMO analysis, the di-iron centred and membrane bound AlkB was also investigated as a major short chain alkane oxidising alternative (Austin et al. 2000). Using AlkB from *Pseudomonas oleovorans* as a reference sequence, a small, related cluster was identified in the methanotrophs. These contained at least 25% identity to the reference sequence over the central region of ca. 200 residues, as well as conservation of all 9 histidine residues identified to be indispensable for catalytic function in *P. oleovorans* (Shanklin and Whittle 2003).

Methanotrophs oxidise ammonia through the lack of specificity of their CuMMO (and potentially sMMO) enzymes and they have been shown repeatedly to produce hydroxylamine *in vivo* (Dalton 1977; Nyerges and Stein 2009; Stein et al. 2012; Mohammadi et al. 2017). The hydroxylamine formed can either be reduced back

to ammonia or further oxidised to either nitrite or nitric oxide depending on the system used (Campbell et al. 2011). Catalysing the former are the hydroxylamine reductases (also known as hybrid cluster protein), and these were identified in methanotrophs using reference sequences for each of the three different HCP types previously identified (Van Den Berg et al. 2000). Sequences which lacked the required binding residues for the hybrid cluster were removed.

Hydroxylamine oxidation in ammonia oxidising bacteria is primarily conducted by the HAO system, the HaoA of which contains seven c-type haem groups and one P460 centre (Lipscomb et al. 1982). An alternative and structurally unrelated oxidase system known as CytL (also cyr, Cytochrome P460) also uses a P460 centre, however lacks any additional c-type haem (Elmore et al. 2007). All identified HaoA sequences contained eight CxxCH binding motifs required for haem binding, with close sequence agreement to the reference *Nitrosomonas europaea* sequence (Q50925) (Lipscomb et al. 1982) in all but *Methylomirabilis oxyfera*. CytL identification was complicated by its similarity to nitric oxide reducing CytS, which does not contain a P460 centre, nor is it known to oxidise hydroxylamine (Elmore et al. 2007). Reference sequences of both CytL and CytS were selected from *M. capsulatus* (AAU93287 and AAU91546 identified in (Stein and Klotz 2011)), the former of which has been demonstrated to exhibit hydroxylamine oxidase activity (Zahn et al. 1994; Bergmann et al. 1998). An additional characterised CytL reference sequence from *Nitrosomonas europeae* (Q50927) (Pearson et al. 2007) was also included.

Phylogenetic analysis of extracted sequences showed a clearly identifiable cluster of genomes containing putative CytS sequences, all of which were also found to also contain CytL. Most of the genes identified as CytL were closely related to the characterised CytL in *M. capsulatus*, although it was noted that two subgroups did occur, with only *M. vadi* containing a member in both. Of the four sequences outside these groups, *M. methanica* demonstrated high similarity to CytL from *N. europaea*, whilst the underrepresented type II methanotrophs also contained a single non-clustered sequence.

Of the three different types of hydrogenases: [NiFe], [FeFe] and [Fe], only members from the [NiFe] family could be identified in the methanotroph genomes. Preliminary search results based on reference [NiFe] sequences derived from Greening et al. (2016) yielded a high frequency of hydrogenase candidate genes, as

well as sequences showing closer relationships to alternative families, especially NADH dehydrogenases. By removing all sequences not displaying the universally conserved CxxC sequence for [NiFe]-hydrogenases at both the L1 and L2 motif (Greening et al. 2016), the dataset was reduced to those members closely clustered with the reference sequences of one of the [NiFE]-hydrogenase groups (Appendix Figure A7.8). Of the resulting seven hydrogenase groups existing in methanotroph genomes, nearly all were across at least two taxonomic classes, with the exceptions being the 1E and 2B cluster. Whilst the latter was well represented in the type IIa methanotrophs, the 1E members were only represented by two strains: of these two, the sequence that appears in *M. sp. 11b* is the closest to the reference 1E sequence and shares a higher level of sequence identity with it than any other reference sequence. It does not however contain the gene order that is characteristic of a 1E member (results not shown). The correlation of these two members with the assigned 1E hydrogenases should therefore be considered more tentative than the other hydrogenase assignments.

Sulphur quinone reductase (SQR) reference sequences were derived from Marcia et al. (2010) and Gregersen et al. (2011), which differentiated five different groups of SQR, along with the related ferricytochrome-C sulphide reductases (FCSR). Identified sequences correlated well as a type I, II, or III SQR or FCSD, with the exception of several highly divergent members that were implicated in nitrite reduction and removed prior to final phylogenetic analysis. Whilst FCSD was only ever found in the type I methanotrophs, all SQR members occurred in at least some of the type I and type II members, though none were found in all genomes of any methanotrophic group or subgroup.

Multiple literature sources (Tabita et al. 2007; Badger and Bek 2008; Tabita et al. 2008; Khadem et al. 2011; Hauser et al. 2015) were drawn on for RuBisCO gene analysis in order to obtain sufficient reference genes to confidently assign all previously reported groups and sub-groups. Despite the low prevalence of RuBisCO genes found in the methanotroph genomes, four different RuBisCO forms were identified, along with RuBisCO like proteins (RBP) (Appendix Figure A7.10). These were mostly of the hexadecameric form I type, with both green (IA) and red (IC) identified, along with divergent verrucomicrobial sequences designated by Khadem et al. (2011) as type IE. The dimeric type II form was also identified in some of the IB methanotrophs.

3.5. Discussion

3.5.1. Genetics from methane to formaldehyde

While much has already been written on the diversity and distribution of methane oxidising systems in the methanotrophs (Tavormina et al. 2011; Liebner and Svenning 2013; Knief 2015; Osborne and Haritos 2018), the level of resolution of this gene analysis does provide an important new observation: the presence of any of the methane handling systems is not mutually exclusive of any other. This is exemplified, quite astonishingly, in Methylosinus sp. R-45379 which contains all four of the divergent methane handling systems. This is significant as it negates the possibility of mutual exclusivity as a driver for present day methane monooxygenase system distribution. Furthermore, it indicates that the capacity for regulation of more than two independent membrane bound methane monooxygenases exists, and provides a pathway for characterisation of pXMO not previously possible.

The pMMO2 members in methanotrophs were also noted to share a highly similar distribution to that of the XoxF5'. Furthermore, the phylogenetic location of each divergent group relative to the dominant radiation (pMMO and XoxF5 respectively) was almost identical for both groups. This suggests that a previously unconnected functional correlation exists between these two groups, which would have implications for atmospheric methane oxidation (Baani and Liesack 2008). The sporadic nature of XoxF1 and XoxF3 gene distribution also flags these as of interest, especially XoxF1 due to its exclusive appearance in the evolutionarily distant and highly unorthodox sMMO-only and 'intra-aerobic' methanotrophs.

3.5.2. Methanotroph metabolic capability towards reduced gasses extrapolated from genomics

Aside from the primary substrate of methane, this study suggests that metabolic capability to oxidise other reduced gasses exists in the majority of methanotrophs. Further, these genes do not occur as single systems distributed amongst the methanotrophs for each respective substrate, but in every case involves multiple divergent gene/enzyme systems, usually with clearly differentiated capabilities. These systems tend to be distributed sporadically throughout the extant members, showing poor correlation to species taxonomy.

Some level of metabolic capability towards hydrogen appears in genomes of almost all methanotrophs, with the majority of species containing multiple systems drawn from one of seven hydrogenase sub-groups for handling this reduced gas. These systems can be differentiated into the membrane bound H_2-uptake (group 1), cytosolic H_2-uptake (group 2) and cytosolic bidirectional (group 3) [NiFe]-hydrogenases (Greening et al. 2016), the only [NiFe]-hydrogenase group not represented in the methanotrophs being specific for H_2 generation. Of the seven sub-groups identified, three have been linked with aerobic respiration (ID, IH and 2A), whilst all but the 2B sub-group have been shown to be oxygen tolerant. The prevalence of these genes means that over 80% of the methanotrophs examined contain hydrogenase genes associated with aerobic respiration of H_2.

The metabolic capability for hydrogen sulphide oxidation is also highly prevalent in the methanotroph genomes. Whilst functionally the hydrogen sulphide oxidising systems can be divided based on the electron acceptor as either ferricytochrome C sulphide dehydrogenases (FCSD) or sulphur: quinone reductases (SQR), both are evolutionarily related group I flavoprotein disulphide reductases with only the C terminal electron transfer region being non-homologous between the two types (Marcia et al. 2010). The SQR can then be further divided into six types, three of which appear in the methanotrophs. The physiological role of only one of these three groups (type I) has been characterised, which was demonstrated to involve both detoxification and sulphide-dependant respiration in certain species (Marcia et al. 2010). Such classification is further complicated by the demonstration of type I sequences in the methanotroph *M. capsulatus* oxidising formaldehyde (Zahn et al. 2001). This formaldehyde dehydrogenase had a 64% sequence identity to the characterised SQR of *Rhodobacter capsulatus*, which is far higher than the relationship of *R. capsulatus* to some other characterised type I SQR members which can be as low as 40% (results not shown). Even excluding these type I SQR members, over 65% of the methanotrophs investigated contained genes for sulphide oxidising systems. The presence of the type III SQR was of particular note as it appears in proteobacterial methanotrophs from both the alpha- and gamma-class as well as the verrucomicrobial strains, yet the type III SQRs ‘belong’ to the green sulphur bacteria (Chlorobiaceae) and Archaea (Sulfulobales) (Marcia et al. 2010).

Short chain alkane (C_2-C_4) oxidation is mediated in bacteria through members of either the BMM or CuMMO family (Van Beilen and Funhoff 2007). Although each of the methane monooxygenase systems belong to one of these two respective families, and poor substrate selectivity does result in short chain oxidation *in vivo*

(Leadbetter and Foster 1960), no growth on these substrates has ever been identified with methane monooxygenase as the primary oxidant: the only methanotroph to date shown to grow on a short chain alkanes was found to contain an additional BMM member specialised towards this substrate (Crombie and Murrell 2014). With low activity recorded towards short chain alkanes, it was therefore surprising to find the presence of AlkB genes, which are active towards medium and long chain hydrocarbons (Austin et al. 2000), in some methanotrophs. While medium and long chain hydrocarbons are not present in most of the methanotrophic biomes which contain biologically-generated methane, methanotrophs are known to inhabit biomes such as hydrocarbon seeps which are ready sources of such substrates (Yan et al. 2006). Specific cytochrome P450 systems have also been shown to be active towards alkanes (Van Beilen et al. 2006) and potentially provide another mechanism for alkane oxidation in methanotrophs. Numerous putative P450 genes were identified in methanotroph genomes (results not shown), but at this stage there is no sequence-function relationship known for differentiating alkane oxidising P450 genes from those P450 members oxidising alternative substrates.

The presence of hydroxylamine oxidase genes in methanotrophs is unsurprising, as both methane monooxygenase systems are known to co-oxidise ammonia due to their low level of specificity towards the primary substrate (Stein et al. 2012), and the subsequently formed hydroxylamine is highly toxic (Vajrala et al. 2013). The interaction between HAO and CytL *in vivo* for hydroxylamine oxidation is yet to be fully resolved, as even though both have been shown to oxidise hydroxylamine to nitrite (Stein and Klotz 2011), proposals exist for the primary role of CytL being the detoxification of NO produced as a by-product of HAO oxidation (Elmore et al. 2007). As the number of methanotrophs containing both CytL and HAO concurrently is less than the number containing only one of these two genes, this proposal is difficult to reconcile with extant distribution. Alternative roles for CytL in hydroxylamine respiration as opposed to NO detoxification has also been proposed (Bergmann et al. 2000), and would appear to better reconcile the present distribution of these genes in extant methanotroph genomes. The appearance of hydroxylamine reductase genes (HCP) in the methanotrophs is more surprising, as when combined with CuMMO would theoretically form a futile and energy consuming cycle. The existence of this cycle would result in energy expenditure to return hydroxylamine to the available ammonia pool and may be of particular significance in ammonia deficient environments, including those in which methanotrophic nitrogen fixation occurs.

3.5.3. Implications for methanotroph metabolism

The majority of methanotrophs live at the interface between oxic and anoxic zones (Reim et al. 2012): locations with complex gas mixtures typically involving air, multiple reduced gasses, and significant carbon dioxide levels. As both ammonia and hydrogen sulphide have been shown to be competitive inhibitors of methane oxidation in some methanotrophs (Lee et al. 2015), they have important implications for these organisms in locations where these rise above trace gas levels (Dando et al. 1995). This competitive inhibition has been shown in the case of ammonia to involve redox cycling of methane monooxygenase systems, therefore consuming reducing equivalents and producing hydroxylamine with implications for methanotrophic growth (Nyerges and Stein 2009). Especially if reducing equivalents are also consumed during methane monooxygenase inhibition by other reduced gasses, the most significant of these being hydrogen sulphide and hydrogen, then it would place gas composition as an important factor driving methanotroph community distribution and diversity.

One mechanism to reduce the impact of reduced gasses on methane metabolism would be to improve the selectivity of the methane monooxygenase system, and indeed this may well prove to be the basis for some of the diversity that occurs within this system in bacteria (Osborne and Haritos 2018). Considering that increased enzyme selectivity, where evolved, often comes with significant trade-offs such as lower enzyme velocity (Erb and Zarzycki 2018), balancing mechanisms may also be required for survival. Incorporation of gene/enzyme systems that consume these competing gasses is an alternative means to reduce concentrations reaching the monooxygenase systems, and it is striking the proliferation of these systems that exists in the methanotrophs. Whilst some of the reduced gas-metabolising systems have a primary role in detoxification, there is the presence of at least one system known to be involved in respiration for each of these gasses in at least some of the methanotrophs.

The presence of these genes provides the intriguing possibility that methanotrophs not only reduce the impact of these gasses on the methane monooxygenase system, but have also incorporated the capability of extracting energy from these substrates. The extreme of such behaviour would be the ability to transition to a chemolithotrophic lifestyle, and indeed this has been observed in the verrucomicrobial methanotrophs (Carere et al. 2017). For the proteobacterial methanotrophs though, facultative methanotrophy has only ever been observed towards organic compounds (Semrau et al. 2011). A major limiting factor to this would be the method

for carbon assimilation, which in proteobacterial methanotrophs has always been shown to involve partially reduced C1 intermediates (Hanson and Hanson 1996). It is highly surprising therefore that RuBisCO genes were found in all Ib and IIb methanotrophs, which is not required in the dominant methanotrophic carbon assimilation pathways (Hanson and Hanson 1996). Even more surprising is that the RuBisCO genes in the type Ib are not monophyletic, but encode two highly divergent RuBisCO systems. The existence of two separate systems, yet the conservation of at least one of these systems in all known type Ib methanotrophs, implies a strong evolutionary driving factor to retain such a capability. This suggests a link between these organisms and autotrophic growth, and raises the question that if autotrophic growth occurs whether it is indeed chemolithoautotrophic, which from this gene analysis, multiple methanotrophs appear to have the capability for. Even in those that don't contain genes that would support autotrophy, the question arises as to whether some methanotrophs extract energy from reduced gasses and obtain a competitive advantage in the typical environmental conditions that they occur in. Indeed, indications of such behaviour may have already been observed (King and Schnell 1994; Cáceres et al. 2014), the behaviour of which can readily be explained from the gene inventory undertaken here.

3.6. Conclusion

Methane is not generated in isolation in natural environments, with biological and/or abiotic reactions typically resulting in the release of gas mixtures containing at least trace concentrations of additional reducing gasses. This analysis of methanotroph genomes indicates they have the gene/enzyme capability to not only limit the impact of reduced gasses on methane monooxygenase activity, but actively extract energy from them though aerobic pathways. Some even contain RuBisCO genes which points toward chemolithotrophic growth in these members. The diversity and sporadic distribution of genes metabolising reduced gasses complicates any correlations being made between genus and capability, and thus methanotrophic strains must be examined on a case by case basis. *In vivo* demonstration of mixed gas metabolism and growth would have major implications in multiple fields including biotechnology, environmental microbiology and agricultural science.

3.7. References

Andersson JO, Andersson SGE. (2001) Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. *Mol. Biol. Evol.* 18, 829–839.

Austin RN, Chang HK, Zylstra GJ, Groves JT. (2000) The non-heme diiron alkane monooxygenase of Pseudomonas oleovorans (AlkB) hydroxylates via a substrate radical intermediate. *J. Am. Chem. Soc.* 122, 11747–11748.

Baani M, Liesack W. (2008) Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10203–10208.

Badger MR, Bek EJ. (2008) Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle. *J. Exp. Bot.* 59, 1525–1541.

Van Beilen JB, Funhoff EG. (2007) Alkane hydroxylases involved in microbial alkane degradation. *Appl. Microbiol. Biotechnol.* 74, 13–21.

Van Beilen JB, Funhoff EG, van Loon A, Just A, Kaysser L, Bouza M, Holtackers R, Rothlisberger M, Li Z, Witholt B. (2006) Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. *Appl. Environ. Microbiol.* 72, 59–65.

Van Den Berg WAM, Hagen WR, Van Dongen WMAM. (2000) The hybrid-cluster protein ('prismane protein') from Escherichia coli. *Eur. J. Biochem* 267, 666–676.

Bergmann DJ, Zahn JA, DiSpirito AA. (2000) Primary structure of cytochrome c' of Methylococcus capsulatus Bath: Evidence of a phylogenetic link between P460 and c'-type cytochromes. *Arch. Microbiol.* 173, 29–34.

Bergmann DJ, Zahn JA, Hooper AB, DiSpirito AA. (1998) Cytochrome P460 genes from the methanotroph Methylococcus capsulatus bath. *J. Bacteriol.* 180, 6440–6445.

Cáceres M, Gentina JC, Aroca G. (2014) Oxidation of methane by Methylomicrobium album and Methylocystis sp. in the presence of H2S and NH3. *Biotechnol. Lett.* 36, 69–74.

Campbell MA, Nyerges G, Kozlowski JA, Poret-Peterson AT, Stein LY, Klotz MG. (2011) Model of the molecular basis for hydroxylamine oxidation and nitrous oxide production in methanotrophic bacteria. *FEMS Microbiol. Lett.* 322, 82–89.

Carere CR, Hards K, Houghton KM, Power JF, McDonald B, Collet C, Gapes DJ, Sparling R, Boyd ES, Cook GM, et al. (2017) Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. *ISME J.* 11, 2599–2610.

Chen XY, Vinh-Thang H, Ramirez AA, Rodrigue D, Kaliaguine S. (2015) Membrane gas separation technologies for biogas upgrading. *RSC Adv.* 5, 24399–24448.

Chistoserdova L. (2011) Modularity of methylotrophy, revisited. *Environ. Microbiol.* 13, 2603–2622.

Crombie AT, Murrell JC. (2014) Trace-gas metabolic versatility of the facultative methanotroph Methylocella silvestris. *Nature* 510, 148–151.

Dalton H. (1977) Ammonia oxidation by the methane oxidising bacterium Methylococcus capsulatus strain bath. *Arch. Microbiol.* 114, 273–279.

Dando PR, Hughes JA, Leahy Y, Niven SJ, Taylor LJ, Smith C. (1995) Gas venting rates from submarine hydrothermal areas around the island of Milos, Hellenic Volcanic Arc. *Cont. Shelf Res.* 15, 913–929.

Elmore BO, Bergmann DJ, Klotz MG, Hooper AB. (2007) Cytochromes P460 and c′-beta; A new

family of high-spin cytochromes c. *FEBS Lett.* 581, 911–916.

Erb TJ, Zarzycki J. (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2fixing enzyme. *Curr. Opin. Biotechnol.* 49, 100–107.

Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H 2 is a widely utilised energy source for microbial growth and survival. *ISME J.* 10, 761–777.

Gregersen LH, Bryant DA, Frigaard N-U. (2011) Mechanisms and Evolution of Oxidative Sulfur Metabolism in Green Sulfur Bacteria. *Front. Microbiol.* 2, 1–14.

Hanson RS, Hanson TE. (1996) Methanotrophic bacteria. *Microbiol. Rev.* 60, 439–471.

Hauser T, Popilka L, Hartl FU, Hayer-hartl M. (2015) Role of auxiliary proteins in Rubisco biogenesis and function. *Nat. plants* 1, 1–11.

Katoh K, Standley DM. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

Keltjens JT, Pol A, Reimann J, Op Den Camp HJM. (2014) PQQ-dependent methanol dehydrogenases: Rare-earth elements make a difference. *Appl. Microbiol. Biotechnol.* 98, 6163–6183.

Khadem AF, Pol A, Wieczorek A, Mohammadi SS, Francoijs KJ, Stunnenberg HG, Jetten MSM, Op den Camp HJM. (2011) Autotrophic methanotrophy in verrucomicrobia: Methylacidiphilum fumariolicum solv uses the calvin-benson-bassham cycle for Carbon Dioxide Fixation. *J. Bacteriol.* 193, 4438–4446.

Khalil MAK. (2000) Atmospheric methane : its role in the global environment.

King GM, Schnell S. (1994) Ammonium and nitrite inhibition of methane oxidation by Methylobacter albus BG8 and Methylosinus trichosporium OB3b at low methane concentrations. *Appl. Environ. Microbiol.* 60, 3508–3513.

Knief C. (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Front. Microbiol.* 6.

Kotera M, Goto S. (2016) Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis. *Biophys. Physicobiol.* 13, 117–126.

Leadbetter ER, Foster JW. (1960) Bacterial oxidation of gaseous alkanes. *Arch. Mikrobiol.* 35, 92–104.

Lee E-H, Moon K-E, Kim TG, Lee S-D, Cho K-S. (2015) Inhibitory effects of sulfur compounds on methane oxidation by a methane-oxidizing consortium. *J. Biosci. Bioeng.* 120, 670–676.

Liebner S, Svenning MM. (2013) Environmental transcription of mmoX by methane-oxidizing Proteobacteria in a subarctic palsa peatland. *Appl. Environ. Microbiol.* 79, 701–706.

Lipscomb JD, Andersson KK, Muenck E, Kent TA, Hooper AB. (1982) Resolution of multiple heme centers of hydroxylamine oxidoreductase from Nitrosomonas. 2. Moessbauer spectroscopy. *Biochemistry* 21, 3973–3976.

Marcia M, Ermler U, Peng G, Michel H. (2010) A new structure-based classification of sulfide: Quinone oxidoreductases. *Proteins Struct. Funct. Bioinforma.* 78, 1073–1083.

Mira A, Ochman H, Moran NA. (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596.

Mohammadi SS, Pol A, van Alen T, Jetten MSM, Op den Camp HJM. (2017) Ammonia Oxidation and Nitrite Reduction in the Verrucomicrobial Methanotroph Methylacidiphilum fumariolicum SolV. *Front. Microbiol.* 8, 1–11.

Nyerges G, Stein LY. (2009) Ammonia cometabolism and product inhibition vary considerably among species of methanotrophic bacteria. *FEMS Microbiol. Lett.* 297, 131–136.

Osborne CD, Haritos VS. (2018) Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria. *Mol. Phylogenet. Evol.*, 0–1.

Pagans E, Barrena R, Font X, Sánchez A. (2006) Ammonia emissions from the composting of different organic wastes. Dependency on process temperature. *Chemosphere* 62, 1534–1542.

Pearson AR, Elmore BO, Yang C, Ferrara JD, Hooper AB, Wilmot CM. (2007) The crystal structure of cytochrome P460 of Nitrosomonas europaea reveals a novel cytochrome fold and heme - Protein cross-link. *Biochemistry* 46, 8340–8349.

Reim A, Luke C, Krause S, Pratscher J, Frenzel P. (2012) One millimetre makes the difference: high-resolution analysis of methane-oxidizing bacteria and their specific activity at the oxic-anoxic interface in a flooded paddy soil. *ISME J.* 6, 2128–2139.

Semrau JD, DiSpirito AA, Vuilleumier S. (2011) Facultative methanotrophy: false leads, true results, and suggestions for future research. *FEMS Microbiol. Lett.* 323, 1–12.

Shanklin J, Whittle E. (2003) Evidence linking the Pseudomonas oleovorans alkane ω-hydroxylase, an integral membrane diiron enzyme, and the fatty acid desaturase family. *FEBS Lett.* 545, 188–192.

Sigvaldason GE, Elísson G. (1968) Collection and analysis of volcanic gases at Surtsey Iceland. *Geochim. Cosmochim. Acta* 32, 797–805.

Stamatakis A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.

Stein LY, Klotz MG. (2011) Nitrifying and denitrifying pathways of methanotrophic bacteria. *Biochem. Soc. Trans.* 39, 1826–1831.

Stein LY, Roy R, Dunfield PF. (2012) Aerobic Methanotrophy and Nitrification: Processes and Connections. *eLs*, DOI: 10.1002/9780470015902.a0022213.

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. (2007) Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiol. Mol. Biol. Rev.* 71, 576–599.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. (2008) Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* 59, 1515–1524.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.

Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, Murrell JC. (2015) XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. *Environ. Microbiol.* 17, 3937–3948.

Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG. (2011) A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ. Microbiol. Rep.* 3, 91–100.

Vajrala N, Martens-Habbena W, Sayavedra-Soto LA, Schauer A, Bottomley PJ, Stahl DA, Arp DJ. (2013) Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc. Natl. Acad. Sci.* 110, 1006–1011.

Yan T, Ye Q, Zhou J, Zhang CL. (2006) Diversity of functional genes for methanotrophs in sediments associated with gas hydrates and hydrocarbon seeps in the Gulf of Mexico. *FEMS Microbiol. Ecol.*

57, 251–259.

Yuan Y, Guo L, Shen L, Liu JS. (2007) Predicting gene expression from sequence: A reexamination. *PLoS Comput. Biol.* 3, 2391–2397.

Zahn JA, Bergmann DJ, Boyd JM, Kunz RC, DiSpirito AA. (2001) Membrane-associated quinoprotein formaldehyde dehydrogenase from Methylococcus capsulatus Bath. *J. Bacteriol.* 183, 6832–6840.

Zahn JA, Duncan C, DiSpirito AA. (1994) Oxidation of hydroxylamine by cytochrome P-460 of the obligate methylotroph Methylococcus capsulatus Bath. *J. Bacteriol.* 176, 5879–5887.

Zhang W, Ge X, Li YF, Yu Z, Li Y. (2016) Isolation of a methanotroph from a hydrogen sulfide-rich anaerobic digester for methanol production from biogas. *Process Biochem.* 51, 838–844.

This page is intentionally blank

CHAPTER 4

LINKING ENZYME FUNCTION TO EVOLUTION IN PROTEIN SEQUENCE, COFACTOR PRESENCE AND ACTIVE SITE MODIFICATIONS IN THE BMM FAMILY

This page is intentionally blank

4.1. Introduction

The bacterial multicomponent monooxygenase (BMM) enzyme family are capable of the selective oxidation of a vast array of industrially and environmentally important alkanes, alkenes, arenes, ethers and ketones, as well as halogenated versions of these substrates (Nordlund et al. 1990; Whited and Gibson 1991a; Small and Ensign 1997; Kotani et al. 2003; Jiang et al. 2010; Coleman et al. 2011). Although the substrate range of growth-linked metabolites in each specific enzyme system is quite narrow, the number of cometabolites is typically quite expansive: a property exemplified by soluble methane monooxygenase (sMMO), for which over 50 active substrates, including those identified as primary substrates for all other BMM members, have been identified (Colby et al. 1977; Jiang et al. 2010). Individual enzyme members also display high regiospecificity in oxidation such as towards primary or secondary C-H bonds (Arp 1999; Furuya et al. 2011), and in the ortho, meta and para positions of substituted aromatic compounds (Whited and Gibson 1991a; Olsen et al. 1994; Newman and Wackett 1995). Such enzyme specificities have drawn biotechnological interest from the energy, food, bioremediation and speciality chemicals sectors (Jahng et al. 1996; Xin et al. 2004; Torres Pazmiño et al. 2010; Strong et al. 2015; Matassa et al. 2016). Amongst its members, the BMM enzyme that has, to date, generated the most research and application interest is sMMO, which generates the most powerful oxidant in nature in its methane oxidising active site (Rosenzweig 2015).

The BMM family, as well as closely related archaeal multicomponent monooxygenases, are di-iron enzymes constituting a minimum of hydroxylase, reductase and regulatory units (Tinberg et al. 2011). The hydroxylase has a dimeric structure with each monomer consisting of active site containing α sub-units, as well as a β and in some instances γ subunits (Sazinsky and Lippard 2006; Furuya, Hayashi, and Kino 2013). Both the α and β units contain a ferritin-like structure (Lundin et al. 2012) with a common evolutionary origin (Leahy et al. 2003). Whilst some members of the BMM enzyme family remain either poorly or, as yet, biochemically uncharacterised (Coleman et al. 2011), the opposite is true for sMMO (Sazinsky and Lippard 2015) which has been heavily investigated due to its ability to break the C-H bond of methane (Dalton 2005).

To achieve C-H bond breakage in methane, a Q state constituting a 'diamond core' active centre configuration is known to be required (Whittington and Lippard 2001; Beauvais and Lippard 2005a; Tinberg and Lippard 2009), and proton tunnelling has also been implicated (Beauvais and Lippard 2005a; Zheng and Lipscomb 2006). This Q state is generated from a H_{peroxo} intermediate through cleavage of the O-O bond, with the conversion indicated to be proton transfer dependant (Tinberg and Lippard 2009). The repositioning of the E240 residue in the hydroxylase alpha-subunit during the catalytic cycle from externally facing to the secondary catalytic sphere has resulted in it being implicated in this proton transfer process (Lee et al. 2013), however its involvement in this proton transfer has never been directly shown. In the other two well characterised BMM members, toluene monooxygenase and phenol hydroxylase, this glutamate is substituted by a glutamine (McCormick and Lippard 2011), and neither enzyme requires formation of a Q state for catalytic activity. Catalysis in these members occurs while the enzyme is in the H_{peroxo} state (Murray et al. 2007), an analogue to the intermediate formed in sMMO which is also catalytically active and proceeds *via* a distinct mechanism to that for the Q state (Figure 4.1) (Tinberg and Lippard 2010).

In addition to the genes coding for the four canonical (sub-)units, BMMs harbour additional genes in their operons (Merkx and Lippard 2002) including those that produce an additional γ sub-unit for the hydroxylase (Sluis et al. 2002), those involved in electron transfer (Mitchell et al. 2002), and the remainder which are loosely defined as accessory proteins (Izzo et al. 2011). Only those possessing the

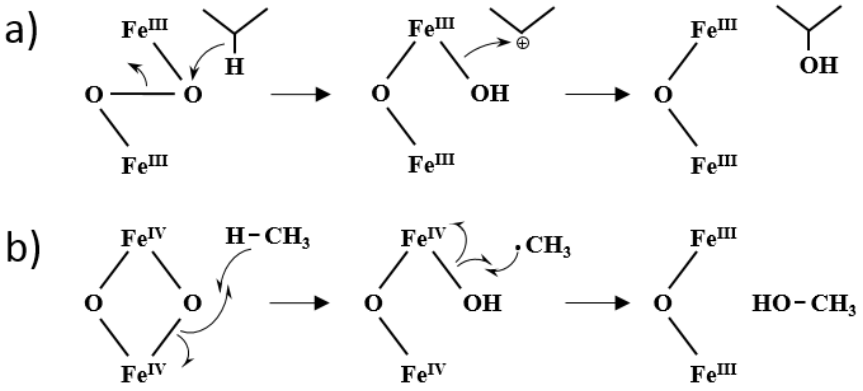


Figure 4.1: Catalytic mechanisms for both H_{peroxo} and Q state in sMMO as proposed by Tinberg and Lippard (2010). H_{peroxo} mechanism is shown in **a)**, and Q state mechanism in **b)**

electron transfer group are widely used for BMM differentiation, with the members containing this extra group defined as '4-component' monooxygenases (Pikus et al. 1996; Tinberg et al. 2011). In addition to the genes described above, others are commonly found in regions adjacent to the BMM operon such as analogues of GroEL, which have been identified in association with sMMO and propane MO (Stafford et al. 2003; Furuya, Hayashi, Semba, et al. 2013): GroEL being a component of a universally identified set of protein folding machinery (Bukau and Horwich 1998; Saibil 2013).

Here, we have applied bioinformatics approaches to investigate the evolutionary development of the BMM family enzymes in relation to known catalytic activities and noted presence or absence of associated gene/proteins. By combining three levels of analysis: phylogenetic, operon and sequence-function based across the entirety of the BMM phylogeny, we have been able to link much of the current understanding of monooxygenase activities to two changes in active site intermediate formation, each relieving constraints on the maximal C-H bond dissociation energy for oxidation. Further, we propose that enzyme assembly and stabilisation factors have independently-evolved multiple times in the BMM family to support enzymes that oxidise increasingly difficult substrates. Finally, we find that evolution of activity toward methane in the BMM occurs simultaneously with changes in active site residues of the hydroxylase units, implicated in enabling Q state formation.

4.2. Materials and Methods

4.2.1. Construction of the BMM protein sequence phylogeny

The BMM phylogeny was constructed using the translated nucleic acid sequences of the α-hydroxylase subunits of sMMO obtained from the Joint Genome Institute IMG/ER database (https://img.jgi.doe.gov/). The sequences were initially identified through sequential iterative searches using BLASTp, based on the *Methanococcus capsulatus (Bath)* α-hydroxylase unit (P22869) query sequence. The BLAST searches were conducted with an E value cut-off of 1e-3 and were ceased when multiple non-BMM proteins appeared in the search. BMM were identified by inspection of the operon associated with the α-hydroxylase unit and whether this contained the prerequisite components of hydroxylase α and β-subunits, reductase and regulatory units (Notomista et al. 2003).

In total, 1728 sequences were identified using this method, and were subsequently aligned using MAFFT v7 with the L-INS-i method (Katoh and Standley 2013). Sequence fragments obtained from consecutive reading frames which when combined gave a full length gene sequences, were recombined *in silico*. The sequence dataset was then manually curated and any incomplete sequences or those not containing the dual DExRH iron binding motifs were removed. A preliminary phylogeny was constructed using MEGA (version 6) (Tamura et al. 2013) and used to guide the reduction of redundant sequences in the dataset. Phylogenetic clusters consisting of multiple representatives from the same genus were reduced to a single representative. The final dataset of 342 sequences was realigned using the L-INS-i method in MAFFT v7. Phylogenetic analysis was performed in RAxML (Stamatakis 2014) using the WAG+G+I model, with 1000 bootstrap replicates obtained using the fast bootstrap algorithm (Stamatakis et al. 2008). The resultant phylogenies were then visualised in FigTree (tree.bio.ed.ac.uk/software/figtree/).

4.2.2. Identification of BMM operon-associated genes

Operon structure and presence of BMM components were visually inspected for monooxygenase representatives in each phylogenetic radiation (e.g. toluene-, aromatic-, alkene-) in the α-hydroxylase unit phylogenetic analysis (described above). Operons were viewed using the Joint Genome Institute IMG/ER viewer (https://img.jgi.doe.gov/), with gene function inferred based on pfam and/or KEGG orthology assignment: α and β subunits being associated with ferritin-like superfamily (α always the larger of the two); reductase units containing 2Fe-2S and FAD domains; and regulatory units belonging to the monooxygenase component MmoB/DmpM superfamily (SSF56029). For the remaining units, if not explicitly associated with a BMM component in Pfam or KEGG orthology, nor containing any alternative associations, BLASTp searches were used to determine the frequency of gene occurrence in BMM operons, and identify similar sequences with clearer characterisation.

4.2.3. Retrieval, alignment and phylogenetic analysis of bacterial GroEL proteins

BLASTp analysis was conducted with the curated UniProtKB/SwissProt database (http://www.uniprot.org/) based on the archetypical GroEL sequence from *E. coli* (P0A6F5) and an E cutoff value of 1e0. The initial 1198 sequences returned were reduced to 989 by elimination of fragmented sequences, non-bacterial chaperonin units, and members not associated with the Cpn60_TCP1 Pfam. To these, 14 additional sequences identified during BMM operon analysis as GroEL analogues and selected to represent the diversity identified therein were added: 6 associated with propane MO; 4 associated with methane MO and 2 from each of butane MO and propane/dioxane MO BMM group. GroEL sequences were aligned using the L-INS-i method in MAFFT v7. Phylogenetic analysis was conducted using the WAG+G model in RAxML, with 100 bootstrap replicates generated using the fast bootstrap algorithm. Phylogenetic relationships were visualised and BMM-associated GroEL proteins located using FigTree. Subsequently, methane and propane associated-GroEL were identified among proteins obtained in the original UniProtKB/SwissProt database search.

4.2.4. Other

Protein crystal structures obtained from the Protein Databank (www.rscb.org) were visualised using PyMol, with molecular distances determined using the built-in measurement tool. Conservation of residues at specific locations in protein sequences was determined by manual inspection of aligned sequences using the sequence viewer in MEGA v7, with subsequent representation undertaken using WebLogo (https://weblogo.berkeley.edu/).

4.3. Results

4.3.1. Updated phylogeny of the BMM family

Phylogenetic analysis of the BMM based on the α-hydroxylase translated protein sequences was dominated by three distinct clusters (Figure 4.2). The *in vivo* activity of multiple members from each cluster have been characterised (Bertoni et al. 1996; Arai et al. 1998; Kotani et al. 2003), which accords with their designation as the toluene monooxygenase, phenol hydroxylase and propane monooxygenase

cluster respectively (Leahy et al. 2003; Sazinsky and Lippard 2006). Whilst this classical naming system will be used throughout this paper, it is important to note that it contains a degree of oversimplification, with members of each of these clusters displaying at least some degree of substrate promiscuity, and at least one additional substrate per cluster supports growth in the host organisms. Beyond the three dominant clusters, there are BMM clusters with activities toward substrates including methane, ethane, butane, ethene, propene, and tetrahydrofuran as per reports (Colby and Dalton 1976; Coleman and Spain 2003; Thiemer et al. 2003; Chan et al. 2005; Cooley et al. 2009; Martin et al. 2014). There is also a growing body of evidence that the previously uncharacterised 'group 6' BMM members have growth

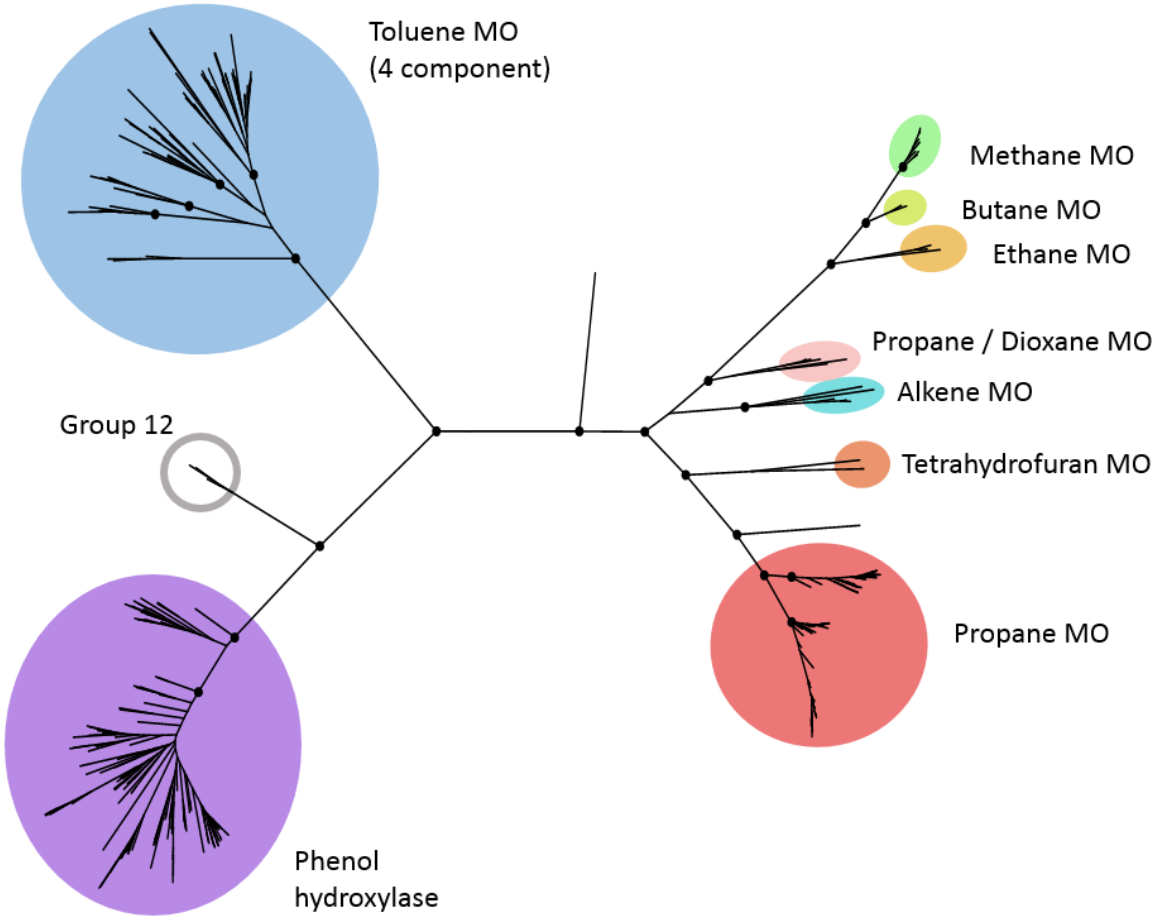


Figure 4.2: Phylogeny of the BMM family based on hydroxylase α units including putative assignments of functional regions based on characterised members. The full scope of BMM divergence in the JGI database was captured in 342 α unit sequences of BMM members. Subsequent alignment was conducted using MAFFT, with phylogenetic analysis including 1000 bootstrap replicates performed in RAxML. Nodes with >90% support have been indicated in black circles. Major clusters have been assigned based on functionally characterised members therein, with the sole cluster lacking any characterised member designated as group 12 in accordance with the standard naming conventions used in literature for formerly uncharacterised BMM clusters.

linked activity towards propane, dioxane and tetrahydrofuran (Kawashima et al. 2006; He et al. 2017; Deng et al. 2018). BMMs represented by one small cluster in the tree lack any characterisation to date; as this cluster lacks any previous designation in the literature, it is herein referred to as 'group 12' (Figure 4.2). Of the members not appearing as part of larger radiations, a representative from the gammaproteobacterial species *Solimonas soli* is of note as it appears in the central evolutionary region located between the two major phylogenetic radiations: those dominated the aromatic and saturated substrates. As there is only a single representative in this location, and there is to date no evidence of the system being active, further investigation of the BMM member in this species was not conducted as part of this study. It does however provide an interesting opportunity for future study and characterisation.

The larger toluene monooxygenase branch, containing the toluene cluster, also contains an outlier group which is active on both aromatics and short chain alkenes; the microorganisms represented in this outlier are capable of further metabolism of both substrate groups (Small and Ensign 1997; Zhou et al. 1999) (Figure 4.2). Interestingly, short chain alkene-acting monooxygenases appear in two distinct branches of the phylogenetic tree (Figure 4.2). This is underscored by the cluster in the alkane/alkene branch being a conventional three component monooxygenase (Chan et al. 2005), whilst the alternative from within the toluene monooxygenase branch contains an additional rieske-like protein characteristic of the 4-component monooxygenase (Small and Ensign 1997), thus also differentiating them at a physiological level.

4.3.2. Evolution of the BMM and hydrocarbon bond cleavage activity

From the ancestral node of the phylogenetic tree, three distinct evolutionary trajectories are observed (Figure 4.3): one that resulted in the phenol hydroxylase branch, another with activity towards toluene and short chain alkenes (4-component), and the final direction of the alkane/alkene monooxygenase branch for which the short chain alkene monooxygenases (3-component) are the first contemporary representatives. In all of the three evolutionary directions from the trifurcation point observed in Figure 4.2, the primary activity was initially toward unsaturated substrates. In one of the three directions

however, a major transition occurred enabling hydroxylation of saturated substrates for the first time (Figure 4.2 & 4.3).

Activity towards saturated compounds appeared around a similar time-point to when the alkene/alkane branch diverged in two distinct directions (Figure 4.2). One sub-branch contains members able to utilise tetrahydrofuran and propane (the propane sub-branch), whilst the second branch contains members that evolved to have activity toward a range of short chain alkanes and cyclic ethers, with the most distant members of the branch having methane activity (the methane sub-branch) (Figure 4.2 & 4.3).

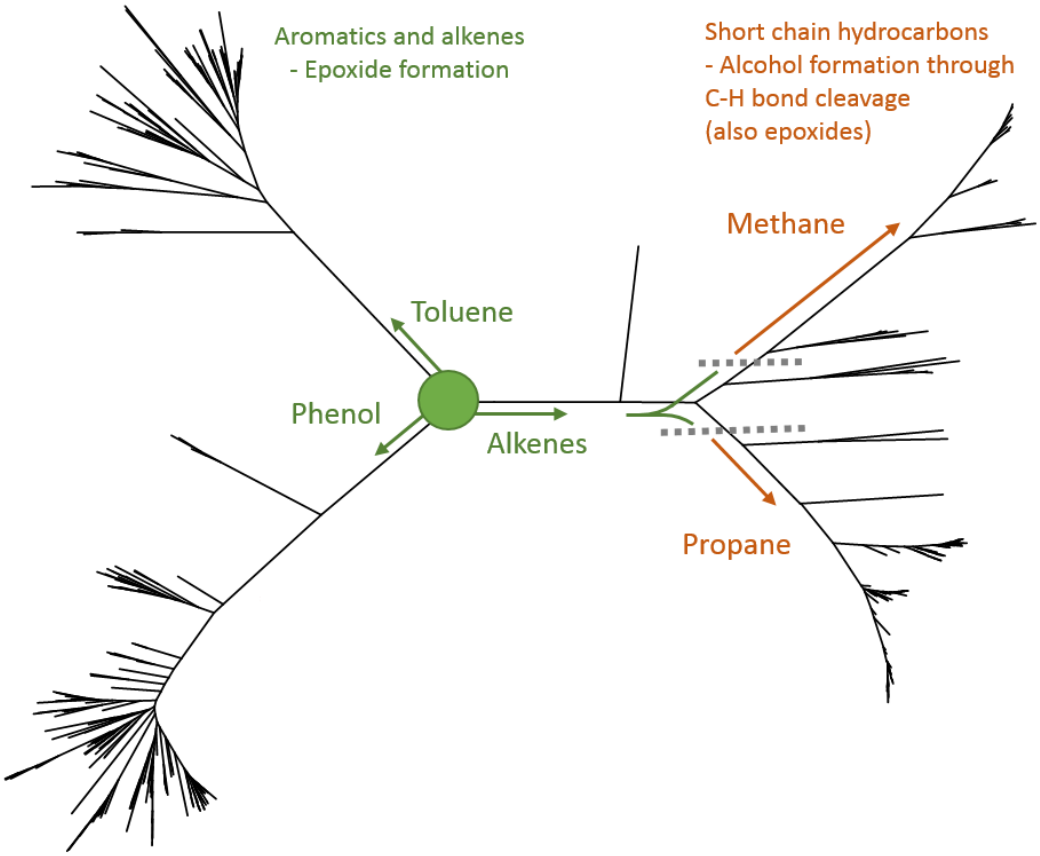


Figure 4.3: Evolutionary pathway of BMM family including transition regions from unsaturated to saturated substrates. The proposed evolutionary trajectories of the BMM family has been overlayed on the phylogeny contained in figure 1. The green circle denotes the approximate location of the common ancestor to the BMM family, with each of the three subsequent evolutionary directions labelled according to the major unsaturated substrate of this branch. Further elucidation is also given for the subsequent evolutionary directions of the alkenes branch, with the major bifurcation point highlighted and transition from unsaturated to saturated substrate capabilities indicated through the colour transition from green to orange.

At a superficial level the two sub-branches in the alkene/alkane branch appear to oxidise similar saturated hydrocarbon substrates, however, an examination of bond dissociation energies required for the oxidation steps demonstrates the difference in reactivities between members of the two branches. In Figure 4.4, bond energy values have been overlaid on the sub-branches of the phylogenetic tree that represent different intervening substrate activities. Here, propane and methane monooxygenase activities represent two extremes of this region of the phylogenetic tree and the C-H bond energies differ

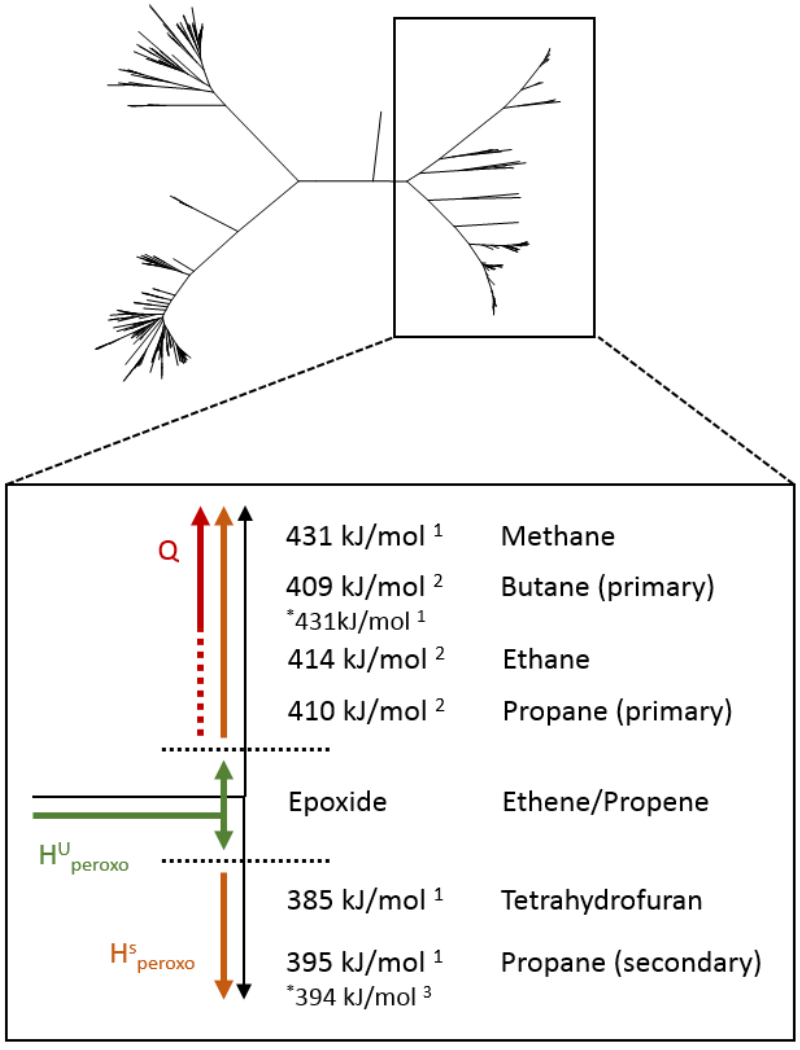


Figure 4.4: Alkane active region of BMM phylogeny indicating primary substrates, substrate bond energies, and proposed active intermediate involved in bond cleavage. Bond dissociation energy of the primary substrate (hydroxylation location in parenthesis if required) indicated, with additional BDE values denoted with * being alternative substrates for preceding BMM member with equivalent or higher BDE than primary substrate (methane and acetone for butane and propane MO respectively). Coloured arrows indicate proposed catalytic site configurations active in each group, specifying major transitions that enabled activity towards additional substrates. Colours correspond to those used in active site intermediate diagram (Figure 4.9). [1] Lange 1999, [2] Gribov et al. 2003, [3] Bordwell et al. 1988

by 36 kJ mol^{-1}. Importantly, in the propane branch the oxidation of propane has been shown to occur at the secondary carbon (Kotani et al. 2003). While tentative evidence for certain propane monooxygenase also oxidising propane at the primary position has been reported, to date this has only been demonstrated at the organism level (Kawashima et al. 2006). Activity towards ethane was also demonstrated in these organisms (Kawashima et al. 2006), and this alkane is not considered a propane monooxygenase substrate (Kotani et al. 2003). As it is possible that a separate alkane monooxygenase is present in the microorganism, this could also be responsible for the primary propane monooxygenase activity. Assuming that earlier findings of propane monooxygenase being only active towards the secondary propane positions is correct for all propane monooxygenase members (Kotani et al. 2003), this would suggest that enzymes of this branch reach a maximum bond energy cleavage under 400 kJ/mol. By comparison, all characterised members of the methane sub-branch are able to oxidise substrates of at least 410 kJ/mol, with the requirement for methane oxidation climbing to 431 kJ/mol. The difference between C-H of ethane and methane bond energy alone has previously been noted to equate to a predicted decrease in kinetic reaction rate of three orders of magnitude (Zheng and Lipscomb 2006).

4.3.3. Two distinct gamma subunits exist within the BMMs

The γ subunit was observed as a third component of the hydroxylase unit in two distinct regions of the BMM family (Figure 4.5a). The first region includes all toluene monooxygenases, phenol hydroxylases, and representatives in between these two groups; this region designated as 'aromatic γ units'. The second region, described as 'alkane γ units' encompasses the methane, ethane and butane monooxygenase clusters.

There is very low to non-existent identity (≤ 15%) between protein sequences of the alkane and aromatic γ subunits (data not shown). Furthermore, the alkane γ unit has ca. 1.5-fold longer sequence than the aromatic subunit and a comparison of their crystal structures also demonstrates no correspondence at any of the higher protein structure levels (Figure 4.6). In terms of secondary structure, the two types have different protein fold classes, with the aromatic γ units composed of an α+β fold, whilst the alkane γ unit exhibits an all-α structure. At the quaternary level, both phenol and toluene monooxygenase

crystal structures contain the aromatic γ unit mounted in almost identical locations in a pseudo-axial position on the α unit, whilst the corresponding alkane γ unit from methane monooxygenase resides at

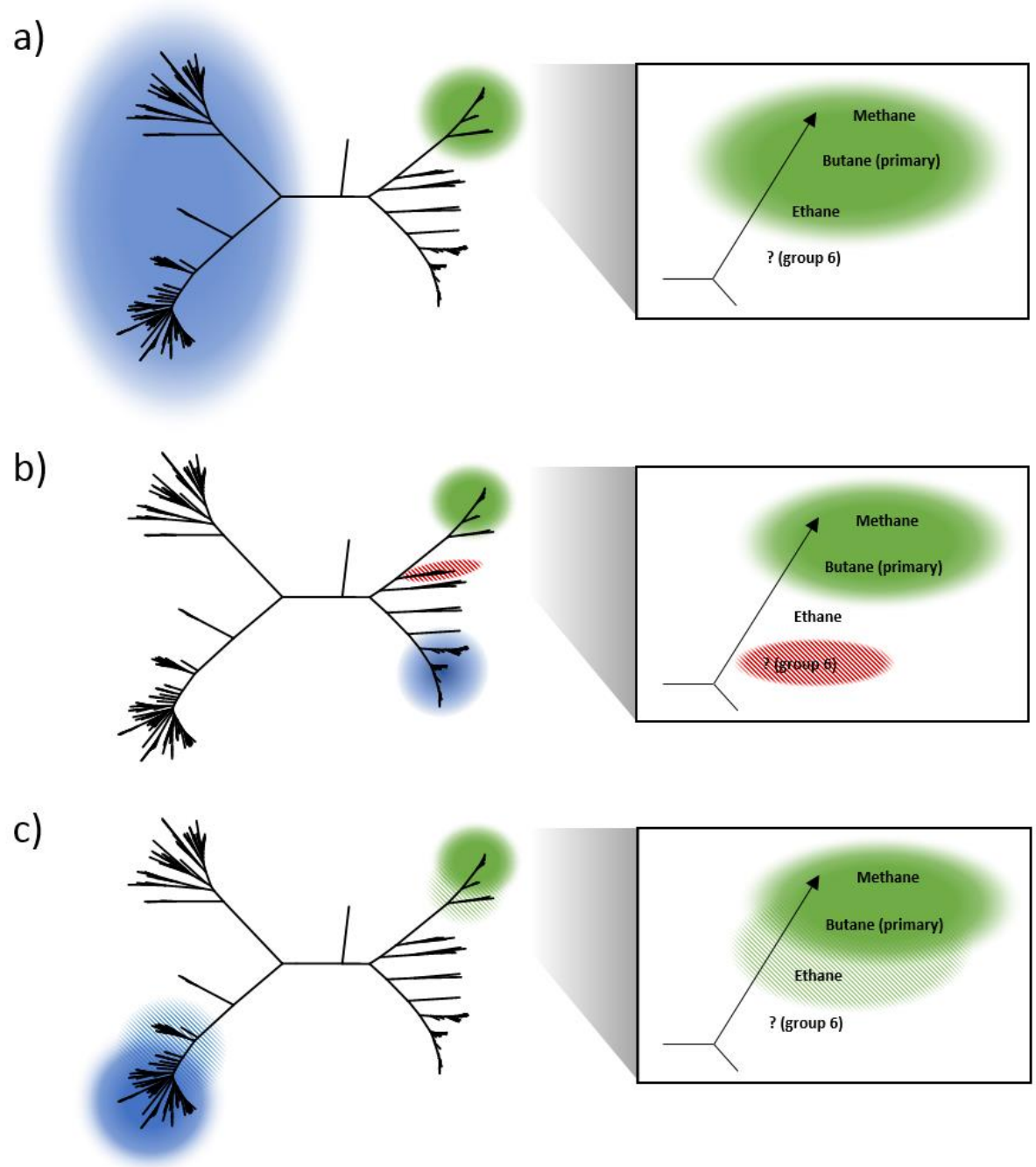


Figure 4.5: Distribution of additional putative assembly and folding factors in the BMM phylogeny. a) Locations in the BMM phylogeny in which additional gamma unit appears in BMM operon. Locations in blue share sequence homology with gamma units appearing in phenol hydroxylase and toluene monooxygenase, whilst locations in green contain greater likeness to sMMO gamma unit. **b)** Appearance of GroEL homologue appearing in association with BMM operon. Distinct clusters have been emphasised using different colours, with hatched region denoting only partial presence in those regions. **c)** Presence of additional unit in the BMM operon displaying inhibitory binding to hydroxylase. Distribution separated into locations displaying conformational likeness to PhK (blue) and MMOD (green), with hatched region indicating partial presence of additional unit in these members.

the interface between α and β subunits (Figure 4.6). A comparison of the binding sites of the core of the respective γ units to the larger α/β component shows no overlap, with only the tail regions exhibiting similarity in binding locations.

4.3.4. Additional proteins associated with BMM

4.3.4.1. Divergent GroEL proteins

Although not located within the operon, several members of the BMM phylogeny contain a conserved gene encoding a GroEL analogue in close association with the operon. The appearance of GroEL analogues occurs in three distinct locations within the BMM phylogeny (Figure 4.5b): the first group

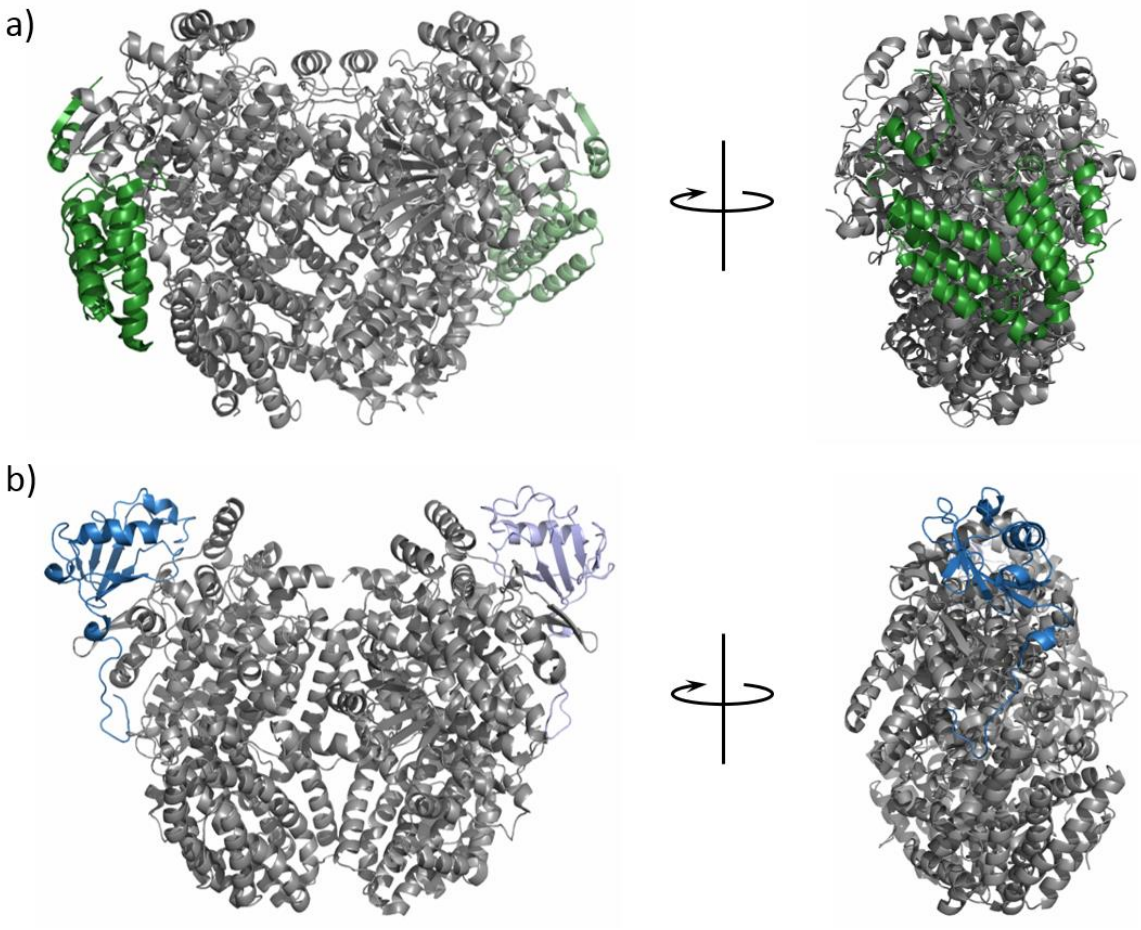


Figure 4.6: Crystal structures of sMMO and phenol hydroxylase highlighting position of gamma unit. Front (left) and side (right) views using ribbon diagram representations of **a)** sMMO hydroxylase unit (PDB accession 4GAM) depicting the gamma unit in green and remaining structure in grey. **b)** Phenol hydroxylase (PDB accession 2INP) depicting the gamma unit in blue with remaining structure in grey.

of GroEL analogues is specifically found in association with the propane monooxygenase cluster, the second for a small subsection of propane/dioxane monooxygenases, and the last is associated with all members of the methane and butane monooxygenases. All bacterial species identified to contain BMM with associated GroEL analogues were also found to also contain conventional GroEL genes in their genome (Figure 4.7).

The BMM family-associated GroEL analogues were further investigated to determine their relationship to conventional GroEL sequences *via* a phylogenetic examination of an extensive dataset (ca. 1000 sequences) encompassing the diversity of bacterial GroEL. Divergent GroEL proteins were located in four distinct phylogenetic clusters: the methane, ethane and butane MO-associated members; the

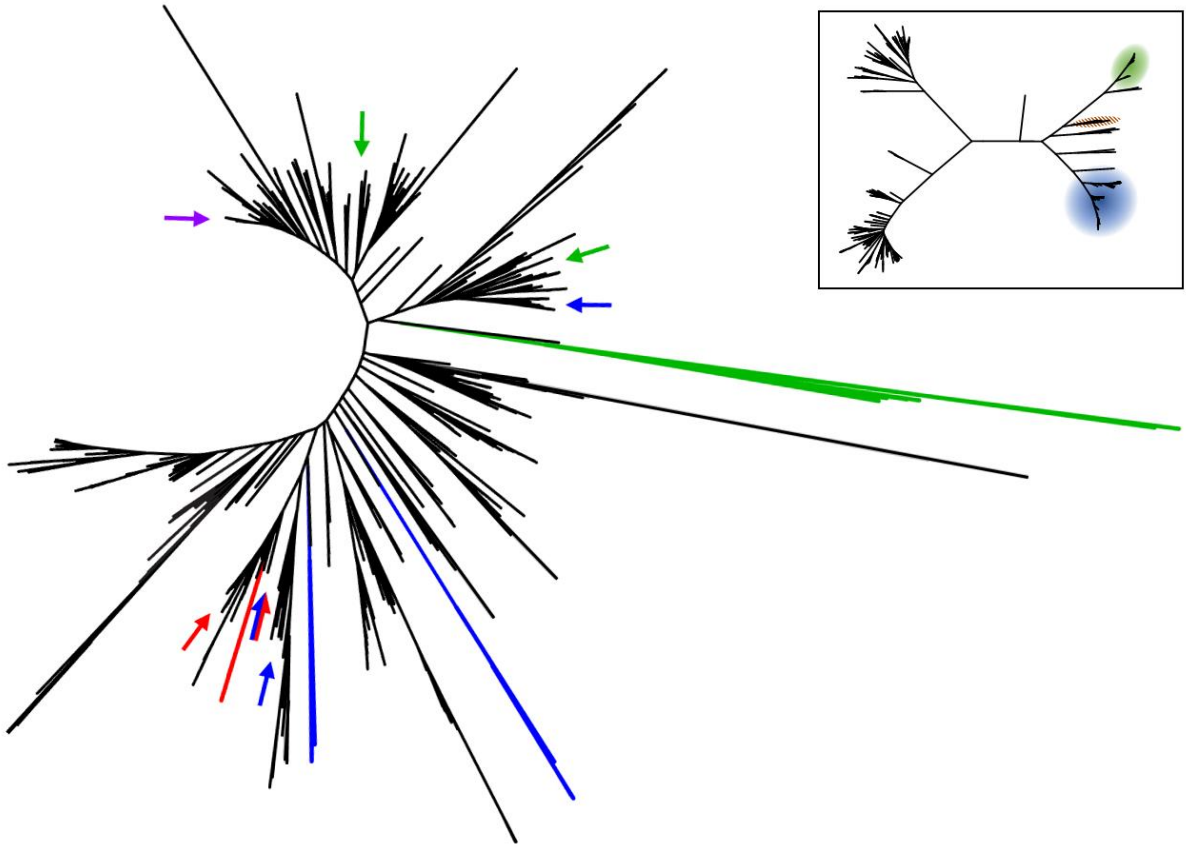


Figure 4.7: Presence of BMM associated members in phylogenetic diversity of GroEL homologues in bacteria. Unrooted phylogeny based on full diversity of the Cpn60_TCP1 pfam in bacteria. ~1000 sequences were aligned using MAFFT v7 with subsequent phylogenies constructed in RAxML. Green, blue and red designate members found to be in association with BMM of the methane/butane, propane and propane/dioxane MO clusters respectively. Purple arrow indicates location of *E. coli* GroEL, with remaining arrows indicate conventional GroELs present in representatives of BMM-associated GroEL (colours correspond with respective BMM group). Insert shows location in BMM phylogeny of respective members having associated GroEL homologue.

propane/dioxane MO-associated members; and two clusters containing the propane MO-associated members (Figure 4.7), the latter differentiated in accordance with the phylum of the host organism (results not shown). Between each of these clusters was a substantial region of conventional GroEL sequences, suggesting that GroEL specialisation towards the BMM family arose multiple times and these were likely derived from conventional GroEL sequences. The derived nature is also reflected in the non-contiguous nature of GroEL presence in BMM phylogeny, with the discrete clustering seen in the GroEL phylogeny reflecting that observed in the BMM phylogeny (Figure 4.7 and insert), the only exception being the contiguous nature of the two propane monooxygenase-associated GroEL groups.

In the methane, ethane and butane monooxygenase group, the BMM-associated GroEL were found to have ≤40% protein sequence identity to the conventional orthologue in their respective species, with this value being around 50% for propane/dioxane and propane monooxygenase members. By comparison, the relationship of conventional GroEL sequences across alpha and gammaproteobacterial methanotroph representatives was >65% identity, and between the actinomycetales and proteobacterial representatives were ≥50% (Appendix Table A7.1).

4.3.4.2. Multiple, unrelated subunit assembly proteins

The proteins PhK, DmpK, P0 and Fp (Nordlund et al. 1990; Notomista et al. 2003; Izzo et al. 2011) refer to closely related orthologues that are associated with phenol hydroxylases in a narrow phylogenetic region of the BMM (Figure 4.5c). This potential assembly protein is absent in microorganisms whose BMM diverged earliest from the common ancestor, as well as the entire toluene monooxygenase branch, but appears in some members of the phenol hydroxylase branch, with increased occurrence in those members diverging later from the common ancestor (Figure 4.5c). There is another small uncharacterised protein comparable to the PhK-type that appears in the operon of the methane, butane, and some ethane monooxygenases: all of these sequences occur in BMM operons forming a single cluster located on the methane sub-branch of the BMM phylogeny (green shading in Figure 4.5c). While these encoded proteins are analogous in terms of being low molecular weight and found within operons of the BMM, there is no correlation (<10% identity) between the two proteins at the sequence

level. This, in conjunction with their occurrence with BMM from separate regions of the phylogeny (Figure 4.5c), suggests these putative assembly protein groups have separate evolutionary origins.

4.3.5. An additional negative charge occurs in the active site secondary sphere of methane active MO

Examination of aligned hydroxylase sequences from the short chain alkane-metabolising BMM members was used to reveal key residues/regions responsible for different activities of BMM members. By focussing on the evolutionary region where methane activity appeared, and restricting the analysis to residues that are both conserved in all methane-active BMM members, and conserved as a different residue for all members with other activities, a single site within close proximity to the active site was

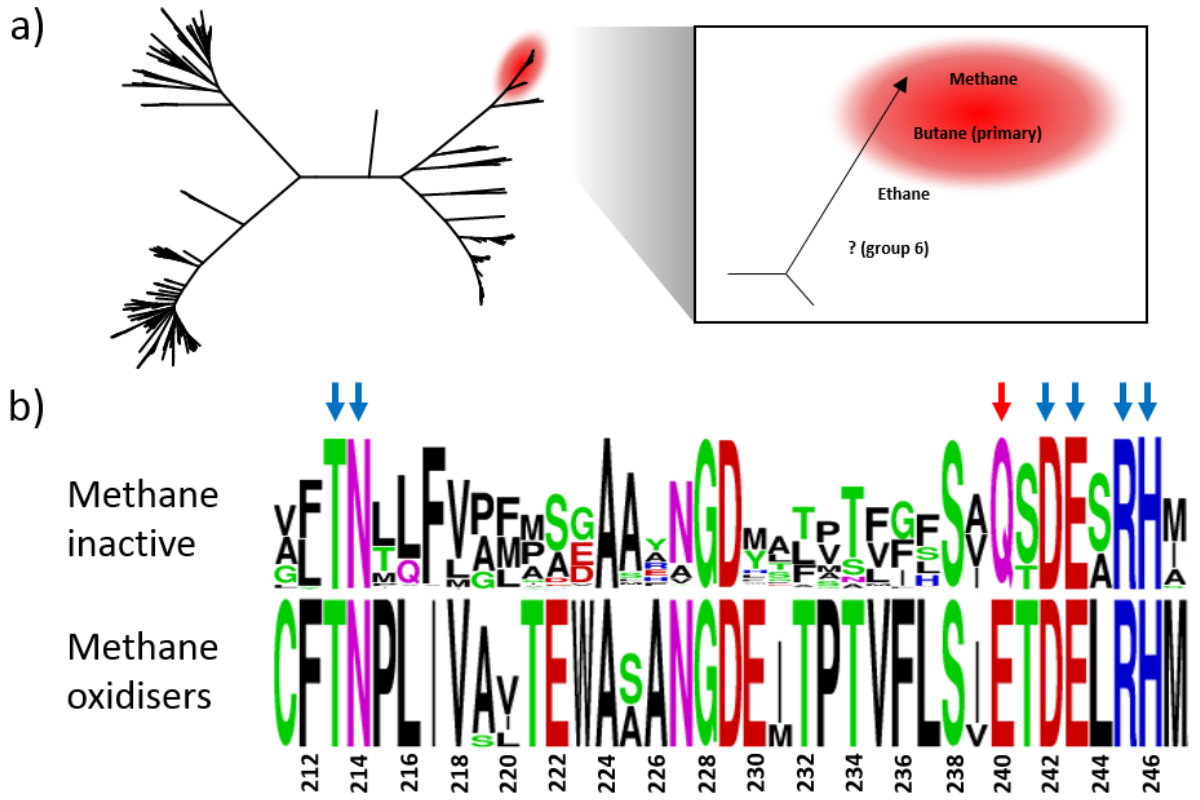


Figure 4.8: Phylogenetic and sequence based location of conserved glutamine change to conserved glutamate residue in the BMM. a) Location in the BMM phylogeny of transition to strictly conserved glutamate residue in MMOH α-subunit, designated E240 in *M. capsulatus*, from otherwise fully conserved glutamine in all other BMM members. Red shading denotes location of glutamate residue. **b)** Sequence logos of BMM hydroxylase α-subunits differentiated into members demonstrating methane reactivity to those inactive towards this substrate. Selected display region includes location of conserved glutamine / glutamate residue and several universally conserved residues involved with catalytic function: these are emphasised using red and blue arrows respectively. Numbering of residues based on that for *M. capsulatus* MMOX.

identified. This location, corresponding to MMOX residue E240 in *M. capsulatus*, demonstrated complete conservation as glutamate in methane oxidising members and glutamine in all other BMM members (Figure 4.8 and 4.11). Through comparison of sequence alignments together with sMMO hydroxylase crystal structures containing both outwards and catalytic core facing E240 geometries, we could rule out the complementary introduction of a positively charged residue in either the catalytic core, or in the vicinity of E240 occurring at the same evolutionary point as the appearance of the glutamate residue.

4.4. Discussion

4.4.1. Evolution of BMM reactivity towards saturated hydrocarbons and ultimately, methane

A significant number of the BMM enzymes have primary activity towards unsaturated substrates in the form of either aromatics or alkenes; these initially react to form epoxides (Saeki and Furuhashi 1994), although for aromatics there is typically a rearrangement process from the epoxide (arene oxide) to form hydroxylated products (Dalton and Leak 1985; Whited and Gibson 1991b). BMM activity towards unsaturated substrates involves reaction initiation *via* electron withdrawal from a π orbital (Beauvais and Lippard 2005b) but reactivity towards saturated compounds necessitates a direct attack on more recalcitrant C-H bonds (Tinberg and Lippard 2010). Following the evolutionary pathway shown in Fig 4.2 it is clear that significant changes occurred in the evolution of BMM genes/enzymes to enable activity firstly towards saturated hydrocarbons, and then further modifications allowed activity towards the most difficult bond oxidised in biology, the C-H bond of methane (Rosenzweig 2015).

BMM activity towards saturated substrates appears in both sub-branches of the larger alkane/alkene branch, evolving from an ancestral alkene-active systems (Figure 4.3). The chain length of the saturated hydrocarbon substrates differ between the two sub-branches, whereby one is limited to propane and the other can act on the smallest alkane substrate (methane) as well as longer chain hydrocarbons. Other than a small ethane monooxygenase group, the butane monooxygenases and sMMO, no other BMM has been clearly shown to oxidise ethane (Pilkington and Dalton 1990; Cooley et al. 2009; Martin et al. 2014) whereas a broad range of BMM can metabolise ethene (Pilkington and Dalton 1990; Small and

Ensign 1997; Dubbels et al. 2007; Coleman et al. 2011). As ethene and ethane have almost identical size and are both non-polar, it is unlikely that the activity differentiation is due to substrate access to active sites, but indicates significant changes to the active site catalysis over the evolutionary history of the BMM family. As BMM that are capable of oxidising saturated substrates are also reactive towards unsaturated compounds (Furuya et al. 2011) we can conclude that it is an increase in reactivity during evolution and not a change in substrate accessibility *per se* that enabled BMM reactivity towards saturated substrates.

To date, only the methane and butane monooxygenases have been shown capable of oxidising the C-H bond of methane (Cooley et al. 2009) which requires formation of the Q state in the di-iron core (Kopp and Lippard 2002) whereas an alternative transition state, H_{peroxo}, has been demonstrated to be active towards numerous substrates (Beauvais and Lippard 2005b). Two different H_{peroxo} states are differentiated, neither of which are an intermediate towards the formation of the other (Song et al. 2009) (Figure 4.9): One which is capable of oxidising unsaturated substrates via an epoxide intermediate (designated here H^{U}_{peroxo}); the other able to directly oxidise C-H bonds (designated here H^{S}_{peroxo}), and is a transition intermediate to the methane active Q state (Han and Noodleman 2008; Song et al. 2009). The two enzyme states capable of directly oxidising C-H bonds, H^{s}_{peroxo} and Q, likely proceed by distinct reaction mechanisms, with H^{S}_{peroxo} using hydride abstraction compared to radical formation by the Q state (Tinberg and Lippard 2010) (Figure 4.9).

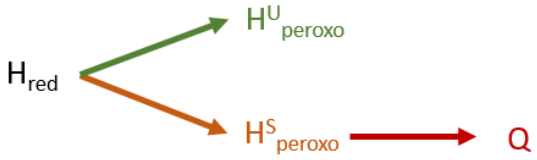


Figure 4.9: Interactions of different di-iron core reaction intermediates in BMM family. H_{red} denotes the activated reduced but not oxygen-bound state of the di-iron core. The mutually exclusive nature of the peroxide states H^{U}_{peroxo} and H^{S}_{peroxo} after oxygen addition are shown, with only the latter capable of oxidising saturated compounds. The Q state capable of methane oxidation is also illustrated, emphasising the prerequisite transition from the saturated substrate active peroxo state.

An apparent anomaly in the evolutionary trajectory of the methane branch is the butane-monooxygenases, which are also capable of oxidising methane and therefore must form the active site Q state. Based on bond energies for oxidation of the primary carbon of butane, these monooxygenases have a lower requirement compared with ethane and propane oxidation, yet the monooxygenases for these substrates evolved earlier (Figure 4.3), and without requiring comparative methane reactivity (Martin et al. 2014). Perhaps the answer to what seems like an anomaly lies in the limited distribution of BMM genes amongst the proteobacteria. Regarding alkane oxidation, only the methane, butane and propane monooxygenases appear in proteobacteria, with the remainder strictly appearing in gram positive members (figure A7.12). Given the infrequency of transfer of these genes across the gram positive / gram negative barrier, it may have been most evolutionarily expedient to obtain butane oxidation capabilities through the repurposing of an ancestral methane monooxygenase member. Although propane monooxygenases present in the proteobacteria could also fulfil this role, its specificity towards the secondary position may have been unfavourable due to the alternate metabolic pathways required for complete metabolism (Arp 1999).

The current proposal suggests that no member of the BMM family aside from the methane and butane monooxygenases to be capable of forming the Q state, and furthermore, that the H^{S}_{peroxo} state is capable of oxidising C-H bonds up to the bond energy level occurring in ethane: this point has yet to be demonstrated (Tinberg and Lippard 2010). However, it has been shown that the H^{S}_{peroxo} state is active towards ethers and is also capable of oxidising methanol (Beauvais and Lippard 2005b; Tinberg and Lippard 2010), the latter bond possessing a higher dissociation energy than propane (Ruscic 2015). This demonstrates that based on bond energies, the propane evolutionary sub-branch of the BMM family is capable of oxidising all substrates occurring in this sub-branch with the H^{S}_{peroxo} state. We therefore propose the Q state developed only on the methane evolutionary sub-branch, with the location and nature of the butane monooxygenase supporting this occurring after the separation of the ethane monooxygenases from the common ancestor (Figure 4.4). Furthermore, we propose that the distinct transition in the BMM family from reactivity toward unsaturated to saturated substrates resulted from

the development of the H^{S}_{peroxo} active site development from the earlier H^{U}_{peroxo} form, and that this transition was crucial in enabling the eventual development of the Q state.

4.4.2. Gamma subunit in the hydroxylase complex has evolved twice

The presence of γ subunit genes in operons of BMM members was overlaid on the hydroxylase subunit phylogenetic tree and these were found in two locations – an aromatic substrate-associated unit and the other restricted to operons of the ethane/butane/methane monooxygenases. The two types of γ units have no significant sequence identities, a point previously noted by Notomista et al. (2003), who furthermore proposed that the aromatic γ unit was inherited together with the common α and β hydroxylase ancestor to all BMM. This suggestion, in conjunction with our findings, indicates that the appearance of the aromatic γ unit in extant BMM members is due to common decent, with its absence in the alkane/alkene branch being due to an early loss of the gene. The genetic origin of the alkane γ unit is more obscure, it was unlikely to have been evolved from the aromatic γ unit, yet no other sources were identified in our analysis.

4.4.3. Divergent GroEL proteins have evolved for folding BMM subunits

The level of divergence in GroEL protein sequences associated with monooxygenases of the alkane/alkene branch (Figure 4.5c and 4.6), and especially with the methane, ethane and butane monooxygenases, is amongst the highest observed in bacteria (results not shown). Divergent members of the GroEL family often have different substrate specificities compared with conventional GroEL, can be regulated with the expression of specific proteins (Fischer et al. 1993) and develop specialisation at the loss of normal substrate activity (Wang et al. 2002). The substrate range of conserved GroEL almost solely consists of proteins with an alpha-beta structural motif (Houry et al. 1999), a fold absent from the all-alpha helix α and β subunits of all BMM hydroxylases. Rare examples of non-BMM all-alpha helical structures do exist as substrates for GroEL however, principally consisting of members of the ferritin-like superfamily (Houry et al. 1999) to which the α and β hydroxylase subunits belong. This indicates that GroEL has been important in the broader evolutionary history of this fold: an occurrence

highly unorthodox for all-alpha helical folds. In the case of propane monooxygenase at least one of the hydroxylase subunits have been shown to be a substrate for a divergent GroEL, further supporting the argument that the multiple instances of divergent GroEL evolved in BMM-containing microorganisms to assist folding of these proteins.

Prior attempts to demonstrate that a divergent GroEL associated with the sMMO operon was required for correct folding of the α and β subunits were unsuccessful due to the GroEL-deletion mutant organisms not transcribing the hydroxylase genes (Stafford et al. 2003). From these findings, the divergent GroEL gene in sMMO was assumed to be primarily a transcription regulator (Scanlan et al. 2009). However, we maintain that the primary role of the divergent GroEL in sMMO is indeed folding of either the α and/or β subunits, and this assertion has support from other studies. The closest evolutionary relative of sMMO in both the hydroxylase units and divergent GroEL proteins, butane monooxygenase, has not shown a similar transcription regulation role for its GroEL gene (Kurth et al. 2008), and the GroEL from propane monooxygenase has been shown to be required at the translation/folding stage of its BMM production (Furuya, Hayashi, Semba, et al. 2013). Furthermore, having a regulatory role in transcription does not preclude a concomitant protein folding function, and has precedents in which regulatory factors require GroEL-mediated folding (Govezensky et al. 1991), thus ensuring upregulation only occurs in the presence of sufficient GroEL to conduct its primary folding role (Kumar et al. 2015).

We propose that initial GroEL duplication and association with BMM was to facilitate contemporaneous regulation of GroEL with its respective BMM member, with precedence for such an event occurring in other GroEL substrates (Lund 2009). This is likely to have occurred at least three times, resulting in non-contiguous clusters for the propane/dioxane monooxygenases, propane monooxygenase and the methane/butane/ethane monooxygenase for both the BMM and GroEL phylogenies. A potential fourth instance of GroEL gene duplication resulting in two separate propane monooxygenases groups is also possible, though contains weaker support. The co-expression of the GroEL analogue with the BMM member then allowed specialisation towards the α and/or β units of the hydroxylase, occurring to a greater extent in BMM clusters that are clearly defined to be alkane active

rather than for the propane/dioxane monooxygenase lineage. In the alkane-active specialised GroEL members, this occurred to such an extent that this protein can no longer be substituted with the conventional housekeeping GroEL (Furuya, Hayashi, Semba, et al. 2013). Taken together, this analysis indicates the development of specialist GroEL enabled BMM activity on alkanes; the majority of alkane-active BMM members, and all of those most recently evolved, requiring a divergent GroEL for function. This proposal is not definitive though, due to the lack of divergent GroEL associated with the ethane monooxygenases, but in this group other specialised folding/stability proteins also appeared and were associated with this branch (Figure 4.5). In summary, we propose that the need for folding assistance or stabilising proteins developed together with the ability of the hydroxylases to act on alkanes and these were included in a cumulative fashion, with Q-state BMM members (butane and methane monooxygenase) requiring all three stabilising components, whilst those active on alkanes but not forming the Q state (ethane and propane) only required one.

4.4.4. Convergent evolution of putative folding proteins

Two disparate types of assembly proteins were identified in our analysis as associated with hydroxylase subunits from different sections of the phylogenetic tree. While PhK is not explicitly required for expression of the active hydroxylase, its presence has been shown to significantly increase the production of active hydroxylase and facilitate iron insertion into the hydroxylase active site (Powlowski et al. 1997; Izzo et al. 2011). Characterisation studies have revealed MMOD interacts with the sMMO hydroxylase unit, demonstrating binding in a 2:1 ratio, inhibition of monooxygenase activity, competition with the regulatory protein and preferential binding to the apo-hydroxylase (Merkx and Lippard 2002). Such behaviour mirrors that previously observed for PhK, along with more general similarities such as unit size and comparative expression level to the hydroxylase (Powlowski et al. 1997; Izzo et al. 2011).

Due to changes in gene regulation following deletion of the sMMO operon (Semrau et al. 2013), and being the only gene in the sMMO operon without a clearly defined role (Sazinsky and Lippard 2015), MMOD has been tentatively assigned in the literature as a regulatory factor (DiSpirito et al. 2016).

Conversely, the similarity between PhK and MMOD of the sMMO operon were previously noted in interactions with their respective hydroxylase units (Merkx and Lippard 2002; Izzo et al. 2011), providing support for evolutionary functional convergence. Hence, although the MMOD unit may have been co-opted for a regulatory role, due to the majority of its functional characteristics to date mirroring that of PhK, as well as the evolutionary implications of this work, in-principle support remains for a primary role associated with the sMMO hydroxylase unit.

4.4.5. Operon structure and Last Common Ancestor (LCA) of the BMM family

In identifying the last common ancestor (LCA) of the BMM, our interpretation of the analysis differs from that of Notomista et al. (2003) whose conclusions were based on the nucleic acid-level similarities between the phenol-cresol hydroxylase operon from *Pseudomonas. pickettii* (PKO1) and that of toluene benzene 2-monooxygenase (phenol hydroxylase) in *Pseudomonas. putida* (JS150). They noted that there was a minimum 72% nucleic acid sequence agreement between the phenol-cresol hydroxylase operon and the toluene benzene 2-monooxygenase, which rose to 98.8% for a region not translated in *P. pickettii*. The region that was translated produced a single phenol-cresol hydroxylase unit in *P. pickettii* and corresponds to the tail end of the α, γ and reductase units of the BMM in *P. putida*. Despite the high level of nucleic acid similarity between operons, at the protein level this correlation was negligible, reaching a maximum of 28% for a short stretch corresponding to the γ unit in the toluene benzene 2-monooxygenase.

Based on their findings Notomista et al. (2003) proposed that the ancestral BMM formed from the recombination of a pre-BMM sequence with an ancestral phenol-cresol hydroxylase. The former contributed an assembly, β, regulatory and partial α component, whilst the latter contributed the remaining α tail, γ, and reductase units. The inclusion of the assembly unit in their proposal is significant, as our research demonstrates that the only present-day members containing this unit lie within the phenol hydroxylase cluster. This would imply a common ancestor in the phenol hydroxylase cluster (Figure 4.10a), and thus contradict the findings by Leahy et al. (2003) derived from phylogenetic analysis (Figure 4.10b).

Our review of the data provided by Notomista et al. (2003) notes that the proposal does not strictly require the presence of the assembly unit in the common ancestor, allowing it to appear later in evolution whilst retaining the overall intent. Removal of this strict criteria opens up greater possibilities for the location of the common ancestor. Based on the common ancestor containing α, γ and reductase genes in that order (contributed from the phenol-cresol hydroxylase ancestor), all but the toluene and phenol hydroxylase region of Figure 4.2 can be eliminated due to a lack of phenol hydroxylase-like γ unit, and the toluene monooxygenase cluster can further be removed due to their different gene order. This revised proposal for the last common ancestor (Figure 4.10c) now includes an overlap of the region also proposed by Leahy et al. (2003) and enables the reconciliation of the two findings: the most likely location of the common ancestor is in close vicinity of the trifurcation point highlighted in Figure 4.34.10d. This location was further supported by molecular clock analysis of the BMM phylogeny

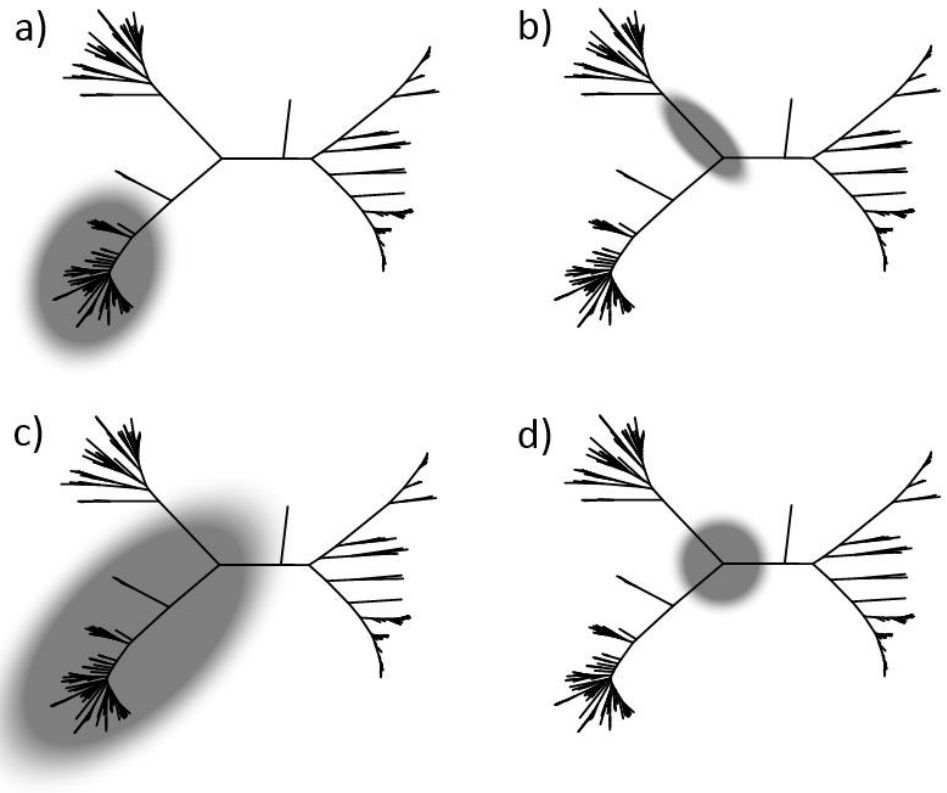


Figure 4.10: Various proposals for location of common ancestor for BMM phylogeny. Grey shading denotes identified region for ancestral node in each proposal. **a)** Possible region based on strict proposal by Notomista et al. (2003) requiring a common ancestor that possesses an aromatic γ hydroxylase sub-unit. **b)** Proposal by Leahy et al. (2003) based on phylogenetic analysis of both α and γ hydroxylase subunits. **c)** revision of Notomista et al. (2003) removing strict requirement of γ hydroxylase sub-unit. **d)** Current proposal for last common ancestor which is supported by the analysis and harmonises Leahy et al. and revised Notomista et al. findings.

(Appendix figure A7.11), which again placed the ancestral node adjacent to this trifurcation point. Based on this analysis, we deduce that the last common ancestor of the BMM was likely active on short chain alkenes or aromatics, with a high likelihood that it had a degree of promiscuity towards both.

4.4.6. Active site residue changes correlate with development of methane activity in sMMO

A change in amino acid from glutamine to glutamate in a protein results in increased Lewis Acidity of otherwise largely comparable amino acids. While glutamate residues are often involved in protein stabilisation through the formation of salt bridges with complementary positively charged amino acids in their local vicinity (Bosshard et al. 2004), our investigation found no such complementary residues in the case of E240 of the α-hydroxylase subunit of sMMO. The glutamine to glutamate conversion would be expected to have negligible impact during much of the catalytic cycle when the side chain of this residue has been shown to be externally facing (Bochevarov et al. 2011), however its entry into the active site secondary sphere upon regulatory unit binding places it in a position to directly affect the environment of the di-iron core (Figure 4.11). Regulatory unit binding has been previously noted to also coincide with a significant change the redox potential of the active site, reinforcing the idea that a significant change in environment occurs at this point (Fox et al. 1991).

Depending on the protonated state of the E240 sidechain when it rearranges to its inward-facing configuration, this residue would either introduce a negative charge or a highly acidic proton into the active site secondary sphere. The knowledge that the transition to the Q state is proton dependant (Tinberg and Lippard 2009) strongly implicates the latter, contingent on the appearance of the Q state at this point in evolution of this BMM sub-branch. The proton transfer required for Q state formation has been proposed to occur on the OH^-/H_2O ligand bound to Fe2 of the di-iron core (Tinberg and Lippard 2009). Although crystal structures of sMMO indicate that E240 is too far away from this ligand for direct proton transfer, an additional water molecule providing a direct hydrogen bond network from the comparative glutamine residue (Q228) to this ligand has been identified in toluene monooxygenase (Figure 4.11) (Bailey et al. 2008). The additional water molecule was identified when the toluene

monooxygenase was crystallised under the physiologically-active reduced and regulatory-bound state, a state that sMMO hydroxylase is yet to be crystallised in. The presence of an additional water molecule in the equivalent position in sMMO could then indirectly transfer the required proton to the OH^-/H_2O ligand through a Grotthuss (proton hopping) mechanism (Rich and Maréchal 2013).

The transition of E240 from external- to core-facing in a deprotonated state cannot be fully ruled out but this is deemed less likely due to the destabilising effects of the introduction of an unbalanced negative charge into the catalytic core. However, many observations around the Q state and BMM evolution could be explained in light of this mechanism, such as altered redox values including the dissimilarity between iron cores upon Q state formation, and the requirement of multiple stabilising factors in BMM members that produce this intermediate.

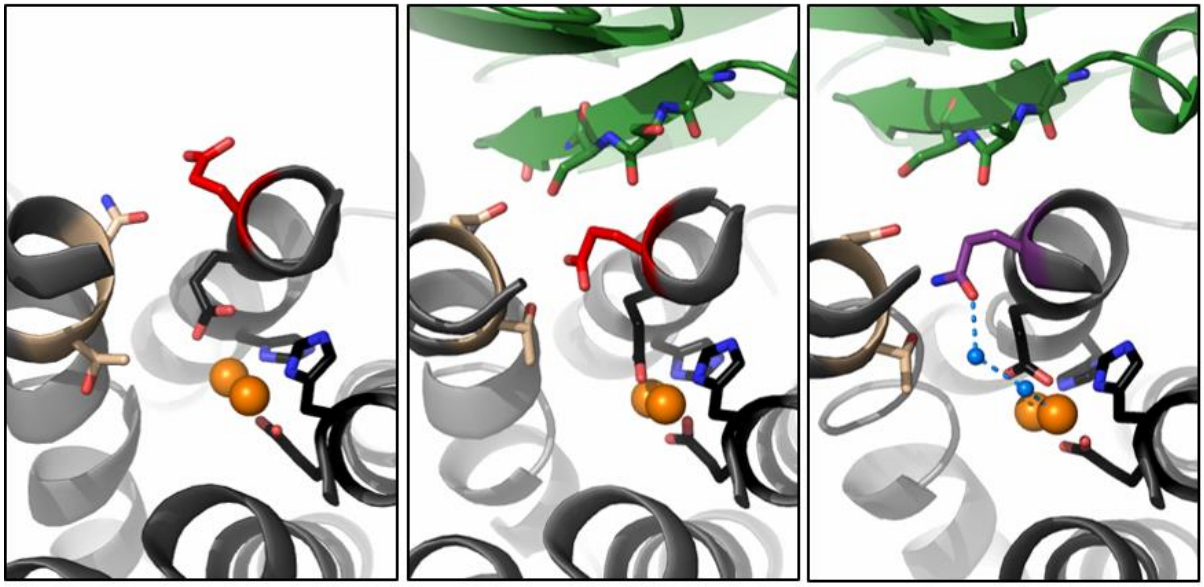


Figure 4.11: Ribbon diagram depictions of di-iron core of BMM members in different stages of the catalytic cycle. Grey and green ribbon diagram components represent the hydroxylase α sub-unit and regulatory component respectively. Iron atoms are rendered orange, with key binding residues in black, and universally conserved threonine and asparagine in beige stick figure representation. **Left** - Structure of oxidised and unbound sMMO active site (1FZ1). E240 residue depicted in red emphasising its side chain being externally facing in this configuration. **Centre** – Structure of oxidised and bound sMMO active site (4GAM). Transition of E240 residue in red to its catalytic core facing configuration is shown. **Right** – Structure of reduced and bound toluene monooxygenase active site (3DHI). The Q228 residue is depicted in magenta, the analogue of E240 in sMMO. Water molecules are shown as blue spheres, with proposed hydrogen bonding network depicted as blue dashed lines.

4.4.7. Piecing together the whole picture – a proposal for the appearance of methane activity in the BMM family.

In this analysis we have shown that initial evolution of the BMM family proceeded in three directions from an alkene/aromatic active ancestor, a proposal which also reconciles two former but contradictory suggestions for the ancestral root of the BMM family. During a relatively short evolutionary period, we propose one of these branches underwent both a split and development of a more reactive catalytic intermediate (H^{S}_{peroxo}), resulting in the creation of the subsequent methane and propane sub-branches. Further evolution, including a critical mutation of a glutamine to glutamate residue in the methane sub-branch then enabled formation of the Q state; this state being active towards methane and appearing at the same time that methane activity appeared in this sub-branch.

The increase in reactivity of the active site had a destabilising effect on the overall hydroxylase structure, requiring first the capture and then specialisation of stability factors involved in structural, folding and assembly functions to offset this in the hydroxylase unit. These additional elements did not facilitate specific modifications that occurred in the hydroxylase unit, as is evidenced by the multiple instances of convergent evolution of factors in disparate regions of the BMM tree, but were required for a stabilising role. Only in the late evolutionary stages of the methane sub-branch did all three of these elements coalesce in one location of the tree, providing sufficient stability to offset the destabilising effect of the generation of the Q state, and thus enabling the appearance of methane activity in the BMM family.

4.5. References

Arai H, Akahira S, Ohishi T, Maeda M, Kudo T. (1998) Adaptation of Comamonas testosteron TA441 to utilize phenol: Organization and regulation of the genes involved in phenol degradation. *Microbiology* 144, 2895–2903.

Arp DJ. (1999) Butane metabolism by butane-grown "Pseudomonas butanovora." *Microbiology* 145, 1173–1180.

Bailey LJ, McCoy JG, Phillips GN, Fox BG. (2008) Structural Consequences of Effector Protein Complex Formation in a Diiron Hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19194–19198.

Beauvais LG, Lippard SJ. (2005a) Reactions of the diiron(IV) intermediate Q in soluble methane monooxygenase with fluoromethanes. *Biochem. Biophys. Res. Commun.* 338, 262–266.

Beauvais LG, Lippard SJ. (2005b) Reactions of the Peroxo Intermediate of Soluble Methane Monooxygenase Hydroxylase with Ethers. *J. Am. Chem. Soc.* 127, 7370–7378.

Bertoni G, Bolognese F, Galli E, Barbieri P. (1996) Cloning of the genes for and characterization of the early stages of toluene and o-xylene catabolism in Pseudomonas stutzeri OX1. *Appl. Environ. Microbiol.* 62, 3704–3711.

Bochevarov AD, Li J, Song WJ, Friesner R a., Lippard SJ. (2011) Insights into the different dioxygen activation pathways of methane and toluene monooxygenase hydroxylases. *J. Am. Chem. Soc.* 133, 7384–7397.

Bosshard HR, Marti DN, Jelesarov I. (2004) Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* 17, 1–16.

Bukau B, Horwich AL. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92, 351–366.

Chan CK, Chion K, Askew SE, Leak DJ, Icrobiol a PPLENM. (2005) Cloning , Expression , and Site-Directed Mutagenesis of the Propene Monooxygenase Genes from Mycobacterium sp . Strain M156. *Appl. Environ. Microbiol.* 71, 1909–1914.

Colby J, Dalton H. (1976) Some properties of a soluble methane mono-oxygenase from Methylococcus capsulatus strain Bath. *Biochem. J.* 157, 495–497.

Colby J, Stirling DI, Dalton H. (1977) The soluble methane mono-oxygenase of Methylococcus capsulatus (Bath). Its ability to oxygenate n-alkanes, n-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem. J.* 165, 395–402.

Coleman N V., Yau S, Wilson NL, Nolan LM, Migocki MD, Ly MA, Crossett B, Holmes AJ. (2011) Untangling the multiple monooxygenases of Mycobacterium chubuense strain NBB4, a versatile hydrocarbon degrader. *Environ. Microbiol. Rep.* 3, 297–307.

Coleman N V, Spain JC. (2003) Epoxyalkane: Coenzyme M Transferase in the Ethene and Vinyl Chloride Biodegradation Pathways of Mycobacterium Strain JS60. *J. Bacteriol.* 185, 5536–5545.

Cooley RB, Dubbels BL, Sayavedra-Soto L a., Bottomley PJ, Arp DJ. (2009) Kinetic characterization of the soluble butane monooxygenase from Thauera butanivorans, formerly "Pseudomonas butanovora." *Microbiology* 155, 2086–2096.

Dalton H. (2005) The Leeuwenhoek Lecture 2000 the natural and unnatural history of methane-

oxidizing bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1207–1222.

Dalton H, Leak DJ. (1985) Mechanistic Studies on the Mode of Action of Methane Monooxygenase. In: Degn H, Cox RP, Toftlund H, editors. Gas Enzymology. Springer Netherlands. p. 169–186.

Deng D, Li F, Li M. (2018) A Novel Propane Monooxygenase Initiating Degradation of 1,4-Dioxane by Mycobacterium dioxanotrophicus PH-06. *Environ. Sci. Technol. Lett.* 5, 86–91.

DiSpirito AA, Semrau JD, Murrell JC, Gallagher WH, Dennison C, Vuilleumier S. (2016) Methanobactin and the Link between Copper and Bacterial Methane Oxidation. *Microbiol. Mol. Biol. Rev.* 80, 387–409.

Dubbels BL, Sayavedra-Soto LA, Arp DJ. (2007) Butane monooxygenase of "Pseudomonas butanovora": Purification and biochemical characterization of a terminal-alkane hydroxylating diiron monooxygenase. *Microbiology* 153, 1808–1816.

Fischer HM, Babst M, Kaspar T, Acuña G, Arigoni F, Hennecke H. (1993) One member of a gro-ESL-like chaperonin multigene family in Bradyrhizobium japonicum is co-regulated with symbiotic nitrogen fixation genes. *EMBO J.* 12, 2901–2912.

Fox BG, Liu Y, Dege JE, Lipscomb JD. (1991) Complex Formation between the Protein Components of Methane Monooxygenase from Methylosinus trichosporium OB3b. *J. Biol. Chem.* 266, 540–550.

Furuya T, Hayashi M, Kino K. (2013) Reconstitution of active mycobacterial binuclear iron monooxygenase complex in escherichia coli. *Appl. Environ. Microbiol.* 79, 6033–6039.

Furuya T, Hayashi M, Semba H, Kino K. (2013) The mycobacterial binuclear iron monooxygenases require a specific chaperonin-like protein for functional expression in a heterologous host. *FEBS J.* 280, 817–826.

Furuya T, Hirose S, Osanai H, Semba H, Kino K. (2011) Identification of the monooxygenase gene clusters responsible for the regioselective oxidation of phenol to hydroquinone in mycobacteria. *Appl. Environ. Microbiol.* 77, 1214–1220.

Govezensky D, Greener T, Segal G, Zamir a. (1991) Inolvement of GroEL in nif gene regulation and assembly. *J. Bacteriol.* 173, 6339–6346.

Han WG, Noodleman L. (2008) Structural model studies for the peroxo intermediate P and the reaction pathway from P to Q of methane monooxygenase using broken-symmetry density functional calculations. *Inorg. Chem.* 47, 2975–2986.

He Y, Mathieu J, Yang Y, Yu P, Da Silva MLB, Alvarez PJJ. (2017) 1,4-Dioxane Biodegradation by Mycobacterium dioxanotrophicus PH-06 Is Associated with a Group-6 Soluble Di-Iron Monooxygenase. *Environ. Sci. Technol. Lett.* 4, 494–499.

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* 402, 147.

Izzo V, Leo G, Scognamiglio R, Troncone L, Birolo L, Di Donato A. (2011) PHK from phenol hydroxylase of Pseudomonas sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases. *Arch. Biochem. Biophys.* 505, 48–59.

Jahng D, Kim CS, Hanson RS, Wood TK. (1996) Optimization of trichloroethylene degradation using soluble methane monooxygenase of Methylosinus trichosporium OB3b expressed in recombinant

bacteria. *Biotechnol. Bioeng.* 51, 349–359.

Jiang H, Chen Y, Jiang P, Zhang C, Smith TJ, Murrell JC, Xing XH. (2010) Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. *Biochem. Eng. J.* 49, 277–288.

Katoh K, Standley DM. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

Kawashima Y, Kato N, Sakai Y, Kotani T, Yurimoto H. (2006) Gene structure and regulation of alkane monooxygenases in propane-utilizing Mycobacterium sp. TY-6 and Pseudonocardia sp. TY-7. *J. Biosci. Bioeng.* 102, 184–192.

Kopp D a., Lippard SJ. (2002) Soluble methane monooxygenase: Activation of dioxygen and methane. *Curr. Opin. Chem. Biol.* 6, 568–576.

Kotani T, Yamamoto T, Yurimoto H, Sakai Y, Kato N. (2003) Propane Monooxygenase and NAD + -Dependent Secondary Alcohol Dehydrogenase in Propane Metabolism by Gordonia sp. Strain TY-5. *J. Bacteriol.* 185, 7120–7128.

Kumar CMS, Mande SC, Mahajan G. (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. *Cell Stress Chaperones* 20, 555–574.

Kurth EG, Doughty DM, Bottomley PJ, Arp DJ, Sayavedra-Soto L a. (2008) Involvement of BmoR and BmoG in n-alkane metabolism in "Pseudomonas butanovora." *Microbiology* 154, 139–147.

Leahy JG, Batchelor PJ, Morcomb SM. (2003) Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* 27, 449–479.

Lee SJ, McCormick MS, Lippard SJ, Cho U-S. (2013) Control of substrate access to the active site in methane monooxygenase. *Nature* 494, 380–384.

Lund PA. (2009) Multiple chaperonins in bacteria – why so many? *FEMS Microbiol. Rev.* 33, 785–800.

Lundin D, Poole AM, Sjoberg B-M, Hogbom M. (2012) Use of Structural Phylogenetic Networks for Classification of the Ferritin-like Superfamily. *J. Biol. Chem.* 287, 20565–20575.

Martin KE, Ozsvar J, Coleman N V. (2014) SmoXYB1C1Z of Mycobacterium sp. Strain NBB4: a Soluble Methane Monooxygenase (sMMO)-Like Enzyme, Active on C2 to C4 Alkanes and Alkenes. *Appl. Environ. Microbiol.* 80, 5801–5806.

Matassa S, Boon N, Pikaar I, Verstraete W. (2016) Microbial protein: future sustainable food supply route with low environmental footprint. *Microb. Biotechnol.* 9, 568–575.

McCormick MS, Lippard SJ. (2011) Analysis of Substrate Access to Active Sites in Bacterial Multicomponent Monooxygenase Hydroxylases: X-ray Crystal Structure of Xenon-Pressurized Phenol Hydroxylase from Pseudomonas sp. OX1. *Biochemistry* 50, 11058–11069.

Merkx M, Lippard SJ. (2002) Why OrfY? Characterization of MMOD, a long overlooked component of the soluble methane monooxygenase from Methylococcus capsulatus (Bath). *J. Biol. Chem.* 277, 5858.

Mitchell KH, Studts JM, Fox BG. (2002) Combined participation of hydroxylase active site residues and effector protein binding in a para to ortho modulation of toluene 4-monooxygenase regiospecificity. *Biochemistry* 41, 3176.

Murray LJ, Naik SG, Ortillo DO, García-Serres R, Lee JK, Huynh BH, Lippard SJ. (2007) Characterization of the Arene-Oxidizing Intermediate in ToMOH as a Diiron(III) Species. *J. Am. Chem. Soc.* 129, 14500–14510.

Newman LM, Wackett LP. (1995) Purification and characterization of toluene 2-monooxygenase from Burkholderia cepacia G4. *Biochemistry* 34, 14066–14076.

Nordlund I, Powlowski J, Shingler V. (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from Pseudomonas sp. strain CF600. *J. Bacteriol.* 172, 6826–6833.

Notomista E, Lahm A, Di Donato A, Tramontano A. (2003) Evolution of Bacterial and Archaeal Multicomponent Monooxygenases. *J. Mol. Evol.* 56, 435–445.

Olsen RH, Kukor JJ, Kaphammer B. (1994) A novel toluene-3-monooxygenase pathway cloned from Pseudomonas pickettii PKO1. *J. Bacteriol.* 176, 3749–3756.

Pikus JD, Studts JM, Achim C, Kauffmann KE, Münck E, Steffan RJ, McClay K, Fox BG. (1996) Recombinant toluene-4-monooxygenase: Catalytic and Mossbauer studies of the purified diiron and Rieske components of a four-protein complex. *Biochemistry* 35, 9106–9119.

Pilkington SJ, Dalton H. (1990) Soluble methane monooxygenase from Methylococcus capsulatus bath. In: Mary EL, editor. Methods in Enzymology. Vol. Volume 188. Academic Press. p. 181–190.

Powlowski J, Sealy J, Shingler V, Cadieux E. (1997) On the role of DmpK, an auxiliary protein associated with multicomponent phenol hydroxylase from Pseudomonas sp. strain CF600. *J. Biol. Chem.* 272, 945–951.

Rich PR, Maréchal A. (2013) Functions of the hydrophilic channels in protonmotive cytochrome c oxidase. *J. R. Soc. Interface* 10, 20130183.

Rosenzweig AC. (2015) Breaking methane. *Nature* 518.

Ruscic B. (2015) Active Thermochemical Tables: Sequential Bond Dissociation Enthalpies of Methane, Ethane, and Methanol and the Related Thermochemistry. *J. Phys. Chem. A* 119, 7810–7837.

Saeki H, Furuhashi K. (1994) Cloning and characterization of a Nocardia corallina B-276 gene cluster encoding alkene monooxygenase. *J. Ferment. Bioeng.* 78, 399–406.

Saibil H. (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* 14, 630–642.

Sazinsky MH, Lippard SJ. (2006) Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc. Chem. Res.* 39, 558–566.

Sazinsky MH, Lippard SJ. (2015) Methane monooxygenase: Functionalizing methane at iron and copper. In: Metal Ions in Life Sciences. Vol. 15. p. 205–256.

Scanlan J, Dumont MG, Murrell JC. (2009) Involvement of MmoR and MmoG in the transcriptional activation of soluble methane monooxygenase genes in Methylosinus trichosporium OB3b. *FEMS Microbiol. Lett.* 301, 181–187.

Semrau JD, Jagadevan S, DiSpirito AA, Khalifa A, Scanlan J, Bergman BH, Freemeier BC, Baral BS, Bandow NL, Vorobev A, et al. (2013) Methanobactin and MmoD work in concert to act as the ‘copper-switch’ in methanotrophs. *Environ. Microbiol.* 15, 3077–3086.

Sluis MK, Sayavedra-Soto L a., Arp DJ. (2002) Molecular analysis of the soluble butane monooxygenase from "Pseudomanas butanovora." *Microbiology* 148, 3617–3629.

Small F, Ensign S. (1997) Alkene monooxygenase from Xanthobacter strain Py2. *J. Biol. Chem.* 272, 24913–24920.

Song WJ, Behan RK, Naik SG, Huynh BH, Lippard SJ. (2009) Characterization of a Peroxodiiron(III) Intermediate in the T201S Variant of Toluene/o-Xylene Monooxygenase Hydroxylase from Pseudomonas sp. OX1. *J. Am. Chem. Soc.* 131, 6074–6075.

Stafford GP, Scanlan J, McDonald IR, Murell JC. (2003) rpoN, mmoR and mmoG, genes involved in regulating the expression of soluble methane monooxygenase in Methylosinus trichosporium OB3b. *Microbiology* 149, 1771–1784.

Stamatakis A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.

Stamatakis A, Hoover P, Rougemont J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771.

Strong PJ, Xie S, Clarke WP. (2015) Methane as a resource: Can the methanotrophs add value? *Environ. Sci. Technol.* 49, 4001–4018.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.

Thiemer B, Andreesen JR, Schräder T. (2003) Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in Pseudonocardia sp. strain K1. *Arch. Microbiol.* 179, 266–277.

Tinberg CE, Lippard SJ. (2009) Revisiting the mechanism of dioxygen activation in soluble methane monooxygenase from M. capsulatus (Bath): Evidence for a multi-step, proton-dependent reaction pathway. *Biochemistry* 48, 12145–12158.

Tinberg CE, Lippard SJ. (2010) Oxidation reactions performed by soluble methane monooxygenase hydroxylase intermediates H peroxo and Q proceed by distinct mechanisms. *Biochemistry* 49, 7902–7912.

Tinberg CE, Song WJ, Izzo V, Lippard SJ. (2011) Multiple roles of component proteins in bacterial multicomponent monooxygenases: Phenol hydroxylase and toluene/ o-xylene monooxygenase from Pseudomonas sp. OX1. *Biochemistry* 50, 1788–1798.

Torres Pazmiño DE, Winkler M, Glieder a., Fraaije MW. (2010) Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.* 146, 9–24.

Vangnai AS, Arp DJ, Sayavedra-soto LA. (2002) Two Distinct Alcohol Dehydrogenases Participate in Butane Metabolism by Pseudomonas butanovora. *J. Bacteriol.* 184, 1916–1924.

Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS. (2002) Directed Evolution of Substrate-Optimized GroEL/S Chaperonins. *Cell* 111, 1027–1039.

Whited GM, Gibson DT. (1991a) Toluene-4-monooxygenase, a three-component enzyme system that catalyzes the oxidation of toluene to p-cresol in Pseudomonas mendocina KR1. *J. Bacteriol.* 173, 3010–3016.

Whited GM, Gibson DT. (1991b) Toluene-4-monooxygenase, a three-component enzyme system that

catalyzes the oxidation of toluene to p-cresol in Pseudomonas mendocina KR1. *J. Bacteriol.* 173, 3010–3016.

Whittington D a., Lippard SJ. (2001) Crystal structures of the soluble methane monooxygenase hydroxylase from methylococcus capsulatus (Bath) demonstrating geometrical variability at the dinuclear iron active site. *J. Am. Chem. Soc.* 123, 827–838.

Xin J, Cui J, Niu J, Hua S, Xia C, Li S, Zhu L. (2004) Production of methanol from methane by methanotrophic bacteria. *Biocatal. Biotransformation* 22, 225–229.

Zheng H, Lipscomb JD. (2006) Regulation of methane monooxygenase catalysis based on size exclusion and quantum tunneling. *Biochemistry* 45, 1685–1692.

Zhou NY, Jenkins A, Chan Kwo Chion CK, Leak DJ. (1999) The alkene monooxygenase from Xanthobacter strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol. *Appl. Environ. Microbiol.* 65, 1589–1595.

CHAPTER 5

HETEROLOGOUS SMMO EXPRESSION: COMBINING COMPONENTS THAT EVOLVED TO DEVELOP METHANE ACTIVITY

This page is intentionally blank

5.1. Introduction

While heterologous protein expression has become commonplace for numerous applications in biotechnology, methods for consistent and high level expression of many proteins and protein families remains elusive (Rosano and Ceccarelli 2014). A highly desirable target for heterologous protein expression is soluble methane monooxygenase (sMMO) due to its potential utility in biotechnological applications and drawbacks of working in the native organisms. Yet despite over two decades of attempts, this remains an example of a system for which high level expression in a tractable organism is yet to be achieved (Jahng and Wood 1994; Jahng et al. 1996; Smith et al. 2002; Borodina et al. 2007).

Functional sMMO heterologous expression has been achieved in three non-methanotrophic species to date: *Pseudomonas putida*, *Agrobacterium tumefaciens* and *Rhizobium meliloti*, with low level expression and difficulty isolating active hydroxylase reported in all instances (Jahng and Wood 1994; Jahng et al. 1996). Higher levels of expression have been achieved in methanotrophic species using a large genetic insert including the entire sMMO operon, *mmoG* and the σ^{54} regulatory factor (Lloyd et al. 1999); as the methanotrophic host only originally contained methane monooxygenase of the pMMO type this does constitute heterologous expression, however they did not succeed in improving tractability of the host organism.

All sMMO heterologous expression attempts contained in the literature to date have also relied on polycistronic systems for insertion and expression of the required genes. Subsequent developments in the field of recombinant technology now provide alternative highly expedient systems in which multiple genes can all be simultaneously expressed on monocistronic constructs, an example of which is the Novagen Duet vector system that can facilitate simultaneous expression of up to eight candidate genes in an *E. coli* host (Rosano and Ceccarelli 2014). This system has the benefits over most traditional polycistronic systems in that the use of multiple vectors facilitate the ability to readily add and remove selected genes to produce desired expression combinations without requiring extensive subsequent cloning procedures, and eliminating the varying expression levels that occur in long operons due to relative position from start of mRNA constructs (Lim et al. 2011).

To the authors knowledge, all heterologous sMMO expression attempts to date have used genetic material derived from either *Methylosinus trichosporium OB3b* or *Methylococcus capsulatus (Bath)* (West et al. 1992; Jahng and Wood 1994; Jahng et al. 1996; Lloyd et al. 1999; Dumont 2004), these being alphaproteobacterial and gammaproteobacterial methanotrophic type strains respectively (Whittenbury et al. 1970). For both, the sMMO has been extensively characterised, including resolved hydroxylase crystal structures (Elango et al. 1997; Rosenzweig et al. 1997). Key functional differentia between these two strains include the methane assimilation pathway, intracytoplasmic membrane morphology and optimal growth temperatures: the latter being mesophilic for *Methylosinus* and thermophilic for *Methylococcus* (Jiang et al. 2010).

The most common method for the detection of sMMO activity in native methanotrophs is the naphthalene assay (Brusseau et al. 1990; Graham et al. 1992; Pacheco-Oliver et al. 2002; Kalidass et al. 2015), the method also applied to heterologous expression (Lloyd et al. 1999). This method relies on the co-oxidation of naphthalene by the sMMO system to naphthol, then the subsequent detection of this product through its reaction with a diazonium species to form an azo dye (Brusseau et al. 1990). While this method has been used, *via* spectroscopic detection of the azo dye, as a quantitative method (Yu et al. 2009; Kalidass et al. 2015), its predominant application in the sMMO literature is as a qualitative measure of activity (Dedysh et al. 2002; Pacheco-Oliver et al. 2002; Dunfield et al. 2003; Larsen and Karlsen 2016).

Proteins that have had difficulty in heterologous expression often exhibit characteristics such as specific active centre types or requirement for membrane solubility (Wagner et al. 2008) but these are not the issue for sMMO. Indeed, the majority of sMMO's larger protein family, the Bacterial Multicomponent Monooxygenases (BMM), have been successfully expressed using standard heterologous expression methods in *E. coli* (Cafaro et al. 2002; Cafaro et al. 2004; Champreda et al. 2004; Furuya, Hayashi, and Kino 2013). Of particular note in expression of some BMM members-was the requirement for specialist chaperone factor genes to achieve functional protein expression (Izzo et al. 2011; Furuya, Hayashi, Semba, et al. 2013). The genes for chaperone proteins were identified as residing either within the BMM operon or directly adjacent to it. Furthermore, in the case of propane monooxygenase, though one of

these genes clearly exhibited orthology to the chaperonin GroEL, native GroEL from *E. coli* could not supplement its function in a heterologous system, indicating divergent function (Furuya, Hayashi, and Kino 2013).

Two putative chaperone proteins, MMOG and MMOD, previously identified as associated with the sMMO operon, exhibit similar characteristics to their analogues found in other members of the BMM family (Merkx and Lippard 2002; Stafford et al. 2003). Both have been determined not to play a functional role in the catalytic cycle of sMMO, and *in vivo* experiments have indicated that MMOD, when bound to the hydroxylase, inhibits activity (Merkx and Lippard 2002) and is present in a ratio of approximately 1:50 to that of the MMOX subunit of sMMO hydroxylase (Merkx et al. 2001). Control of such unbalanced expression ratios can now be achieved in heterologous expression due to recent developments in understanding of ribosomal binding site (RBS) affinities (Espah Borujeni et al. 2014), with the resulting creation of the RBS calculator (Salis 2011). This tool provides general approximation as to the relative strength of a RBS in the form of a translation initiation rate (TIR), and systems for producing RBS activity to within a desired target range (https://salislab.net/software).

The heterologous expression strategies described here draw on an extensive examination of the evolution of the sMMO system (Chapter 4) and review of the heterologous expression literature for the BMM family, and extends these learnings to heterologous expression of sMMO in *E. coli*. Detailed expression system design enabled the testing of multiple hypotheses relating to requirements for successful sMMO expression, principle amongst these that a full contingent of eight genes, including two assembly factors, are required for active sMMO expression in a heterologous organism. This work also improves on the classical sMMO naphthalene assay, increasing sensitivity to provide detection below levels previously possible.

5.2. Experimental

5.2.1. sMMO gene and protein sequences

All genes used were initially obtained from the *Methylococcus capsulatus (Bath)* genome contained in the JGI Integrated Microbial Genome database (img.jgi.doe.gov). Potential sequencing errors were

checked through cross-referencing of two additional *M. capsulatus* genomes present in the JGI database (Texas and ATCC 19069), to ensure the fidelity of the translated proteins. Untagged protein constructs were selected for initial expression trials, however due to the inherent benefits of protein tagging, in-frame His tags were included on MMOX and MMOY constructs subsequent to the C-terminal stop codon. Use of this method allows a single point mutation in the stop codons to result in the inclusion of the tag. Synthesis of all nine constructs (Table 5.1) were performed by GenScript, with application of their proprietary codon optimisation algorithm for each gene.

MMOX and MMOD constructs were both inserted into the pETDuet-1 plasmid to facilitate fixed ratio expression between the pair. Due to the functional connection between the MMOY-MMOZ pair and MMOG-GroES pair, they were also linked on their own respective plasmids: pCDFDuet-1 and pCOLADuet-1 (Figure 5.2). Though the initial design enabled the co-factors MMOR and MMOB to be placed on the separate pACYCDuet-1 plasmid, the later use of the Lemo21 cell line prevented this additional Duet plasmids from being used. The system was therefore modified with partnering of genes

Table 5.1: Summary of genetic constructs as synthesised by GenScript. Table includes the nature of the ribosomal binding site (RBS) if present, all restriction sites explicitly specified for the project and the presence of any purification tags.

Construct name	Preceding restriction sites	RBS	Gene	Tag	Succeeding restriction sites	Initial plasmid
MMOG	BamHI/NheI/Nco1	Yes	mmoG	*Flag	HindIII	pUC57
GroES	HindIII	Yes	groES		NotI	pRSFDuet-1
GroES-His	Nde1	No	groES	His	Nhe1/BamHI	pET21a
MMOX	NdeI	Designed	mmoX	-	AvrII	pETDuet-1
MMOY	NdeI	Designed	mmoY	-	AvrII	pCDFDuet-1
MMOZ	BamHI/NcoI	No	mmoZ	-	HindIII	pUC57
MMOD	XbaI/NcoI	Designed	mmoD	-	HindIII	pUC57
MMOB	HindIII	Yes	mmoB	-	AvrII/XhoI	pUC57
MMOR	BamHI/NdeI	No	mmoC	-	HindIII/AvrII	pET21a

* Asterix denotes stop codon before tag

that had a close functional association in bicistronic constructs: MMOG with GroES and MMOR with MMOB.

To further facilitate achieving a fixed ratio in expression, ribosomal Binding Site (RBS) design was also used. The primary consideration was to ensure a significant excess of MMOX relative to MMOD, however an idealised target used for the design was a 1:50:50:50 ratio of MMOD:MMOX:MMOY:MMOZ: the level of the latter three determined from stoichiometric assembly ratios. Though, practical constraints limited the attainment of the ideal values, the final *in silico* design ratio achieved, based on TIR values, was 2:50:25:25 for the MMOD:MMOX:MMOY:MMOZ genes respectively. Further refinement was not pursued, as the current limit of the RBS predictive technology as well as other unquantified system variables indicated negligible benefits from such efforts.

5.2.2. Cloning and sub-cloning protocol

All GenScript synthesised plasmids and blank vectors were transformed in New England Biolabs Inc (NEB) Turbo Competent *E. coli* (High Efficiency) using their recommended High Efficiency Transformation Protocol (www.neb.com). Cells were then grown using SOC media with corresponding antibiotic overnight in an orbital shaker at 37°C, 250rpm. Plasmids were harvested using Invitrogen ChargeSwitch-Pro Plasmid Miniprep Kits. Plasmid extract concentration was quantified using a Thermo Scientific NanoDrop Lite Spectrophotometer, however these measurements were often found

Table 5.2: Ribosomal binding site predictions for both initial constructs and final designs. TIR values for both the gene inserts prior to design process, and final predicted values. Ratios of both the theoretical 'optimum ratio' as well as final values are also included.

Plasmid	Gene	Initial expression prediction (TIR)	Optimum ratio	Final design Expression prediction (TIR)	Final design ratio (approx.)
pETDuet-1	mmoD	12493	1	2065	2
	mmoX	4948	50	46327	50
pCDFDuet-1	mmoY	6689	50	26995	25
	mmoZ	27709	50	27709	25

to overestimate actual plasmid concentrations. When alternative analysis of plasmid concentration such as visual interpretation of stained agarose gel bands against reference samples was also available, these alternative values were used instead.

For those plasmids that required subsequent sub-cloning to produce the functional vectors, a restriction based ligation procedure using dual unique restriction sites was used to ensure the correct orientation of the subsequent insert. The initial backbone and insert plasmids were double digested using NEB high fidelity restriction enzymes and recommended protocol according to Table 5.3. Digests were then separated using a pre-stained (SYBR Safe DNA gel stain) agarose gel (typically 1%) run at 70V using TAE buffer. Identified bands were then excised and DNA extracted using a Zymoclean Gel DNA Recovery Kit. Ligation was conducted using NEB T4 DNA ligase and recommended protocol, with insert to backbone molar ratio of either 2:1 or 3:1.

The ligated products were then transformed into NEB Turbo Competent *E. coli* (High Efficiency) as per transformation procedure above. To confirm correct ligation, colonies were then selected and grown using SOC media with corresponding antibiotic overnight in an orbital shaker at 37°C 250rpm, with subsequent plasmid harvesting using an Invitrogen ChargeSwitch-Pro Plasmid Miniprep Kit. Single and double digests were conducted according to the restriction enzymes specified in Table 5.3, then run on an agarose gel, all as per above protocol. Only those colonies demonstrating correct plasmid size on the single digest and correct corresponding dual bands on the double digests were selected for subsequent use.

Table 5.3: Sub-cloning conducted to produce final plasmid constructs. Restriction sites used in dual restriction based ligation are specified, along with insert and backbone plasmids used.

Insert	Backbone	Restriction sites / digests	Ligated product
MMOG	pRSF-1b	NcoI / HindIII	pRSF-mmoG
MMOZ	pCDFD-mmoY	NcoI / HindIII	pCDFD-Y/Z
MMOD	pETD-mmoX	NcoI / HindIII	pETD-X/D
MMOB/MMOR	pRSFD-G/S	NdeI / AvrII	pRSFD-B/R//G/S

5.2.3. Preparation of chemically competent cells

Cell line to be made chemically competent was grown in SOB media with antibiotics, as per requirements of plasmids already present in cell line, until reaching an OD_{600} ~0.40 as measured on a Hach DR 5000 UV-Vis Spectrophotometer. Cells were then cooled on iced for 30 min and centrifuged at 3000*xg* and 4°C for 15 min. The supernatant was decanted, and the cell pellet resuspended in ice cold 100 mM $MgCl_2$. The resulting suspension was then centrifuged at 3000*xg* and 4°C for 15 min, decanted, and the pellet resuspended in ice cold 100 mM $CaCl_2$. After holding on ice for 30 min, centrifugation was conducted at a reduced speed of 2000*xg* (at 4°C and 15 min), with subsequent decanting and resuspension in an ice cold 85 mM $CaCl_2$, 15% glycerol solution. A final centrifugation was performed at 1000*xg*, 4°C for 15 min, with decanting and resuspension in further $CaCl_2$/Glycerol solution. The final product was then aliquoted into cryo-tubes and snap frozen in a dry ice / acetone bath for storage at -80°C.

5.2.4. Protein expression

To produce protein expression cell strains, either NEB BL21(DE3) or Lemo21(DE3) competent *E. coli* were transformed using their recommended High Efficiency Transformation Protocols (www.neb.com). This procedure was conducted directly when either one or two plasmids were required in the host strain. When triple transformants were required, dual transformants were returned to a competent state according to the procedure outlined in section 5.15.2.3, then the additional plasmid was inserted according to the High Efficiency Transformation Protocol used for the original cell lines.

A single colony from transformant agar plates was then selected and transferred to SOC outgrowth media, with subsequent growth at 37°C 250 rpm. Optical density was measured during growth until mid-log phase (OD_{600} ~0.5) was reached, as measured on a Hach DR 5000 UV-Vis Spectrophotometer. This was then used for starter culture in subsequent cell expression media, with transfer size varied to standardise initial culture amount. For auto-induction, the cell lines were transferred directly to auto-induction media, and grown at the respective trial temperature (using orbital shaker incubator at 250 rpm) for the allocated time, typically overnight. For IPTG induction, starter culture was transferred to

TB media, and grown at 37°C 250 rpm until an OD_{600} 0.4 - 0.6 was reached. IPTG, as well as rhamnose if Lemo21 turn-down was being used, was then added to specified concentration. The resulting media was then placed at the respective trial temperature (using orbital shaker incubator at 250 rpm) for the required time.

5.2.5. SDS-PAGE analysis

Cell samples were centrifuged at 16000*xg* for 5 min and the supernatant decanted. The cell pellets were then resuspended in phosphate buffer solution to an OD_{600} of 10, from which 500 μL samples were extracted and transferred to an ice bath. Samples were then sonicated (125 W Qsonica) whilst on ice at 90% for four cycles of 15 s with 45 s rest between. Samples were then spun again at 16000*xg* for 10 min and the supernatant decanted: the first 50 μL of this supernatant being collected as the soluble fraction. The pellet was resuspended in 500 μL phosphate buffer, with 50 μL extracted as the insoluble fraction. To all 50 μL samples, 17.5 μL of x4 SDS loading dye was added, then each was heated to 95°C for 5 min. SDS-PAGE was conducted on pre-cast Bolt 4-12% Bis-Tris Plus Gels, run in MES buffer, initially at 90 V then increased to 120 V after 10 min. Gels were stained using SimplyBlue SafeStain according to recommended microwave procedure (www.thermofisher.com).

Alternative lysis methods involved replacement of phosphate buffer with either Triton-X 100 lysis buffer or BugBuster Protein Extraction Reagent, the latter also replacing sonication steps with incubation at 37°C and 250 rpm.

5.2.6. Growth of native methanotroph cultures

To grow *Methylosinus trichosporium (OB3b)* cultures, autoclaved 125 ml serum bottles were filled with 10 ml NMS growth media either lacking copper (Cu-) or with 1 μM copper sulphate added (Cu+). 500 μL of previous *M. trichosporium (OB3b)* culture was then added and the serum bottles sealed with a crimp top airtight septum. Methane was then injected through the septa until headspace concentration reached 20%. Cultures were then grown at 30°C 120 rpm until an OD_{600} of 0.1-0.3 was reached, with sampling conducted *via* syringe through the septa.

5.2.7. Naphthalene assay for the detection of sMMO activity

Naphthalene assays were conducted using a modified method based on that outlined by Yu et al. (2009). Cell samples were centrifuged at 16000*xg* for 5 min and the supernatant decanted. The cell pellet was then redissolved in nitrate mineral salts (NMS) methanotroph growth media to an OD_{600} of 0.5 unless otherwise specified. 1 ml was then transferred to a fresh 1.7 ml microtube, and ~10 mg of ground naphthalene crystals was added at a medium-fine sand consistency. Samples were then incubated at 30°C and 250 rpm for between 0.5 – 2 h for *M. trichosporium* cultures, or 4 – 16 h for *E. coli* cultures. After incubation, the samples were vigorously shaken for several seconds; then after standing for 30 s, 800 µL was extracted and transferred to a 1 ml cuvette, avoiding any settled material. 200 µL of freshly prepared Fast Blue B solution (0.2% Fast Blue B dye in deionised water) was added to each 800 µL sample and the preparation mixed. Absorbance of the samples was measured immediately before and after addition of Fast Blue B solution at 525 nm on a Hach DR 5000 UV-Vis Spectrophotometer, with the difference in absorbance values determined.

Table 5.4: Table of genes used in this study: List of all putative sMMO associated genes, as well as the implicitly linked groES. Translated protein size has also been included with theoretical values where no literature values exist.

Gene	Function	Protein name	Translated protein size (kDa)
mmoX	Hydroxylase unit (α)	MMOX	60.6 [1]
mmoY	Hydroxylase unit (β)	MMOY	45.0 [1]
mmoZ	Hydroxylase unit (γ)	MMOZ	19.8 [1]
mmoB	Regulatory unit	MMOB	15.9 [2]
mmoC	Reductase unit	MMOR	38.5 [3]
mmoD	Putative assembly unit	MMOD	12 [4]
mmoG	Putative specialised chaperonin	MMOG	59.5*
groES	GroEL accessory protein	GroES	10.5*

Corresponding gene function and protein weight for components used in study from *M. capsulatus*: [1] - (Merkx et al. 2001), [2] - (Walters et al. 1999), [3] - (Gassner and Lippard 1999), [4] - (Merkx and Lippard 2002), * - Theoretical

5.3. Results

5.3.1. Expression of components in individual and dual constructs

As an initial step towards the expression of the entire sMMO construct, the heterologous expression of putative specialised chaperonin MMOG using pRSF-MMOG transformants in T7 promoter containing BL21(DE3) was undertaken. IPTG and autoinduction based expression were trialled, with both methods successfully expressing MMOG at significant levels, as indicated by SDS-PAGE analysis (Figure 5.1). Autoinduction at 25°C demonstrated the highest soluble expression amongst the conditions tested, with the lower levels during IPTG induction found to largely independent of IPTG concentration used (between the range 1 mM to 0.1 mM) (Figure 5.1). The demonstration of soluble MMOG in SDS-PAGE fractions was found to be highly dependent on cellular lysing protocol, with only sonication in phosphate buffer solution providing soluble MMOG in significant quantities. Alternative methods

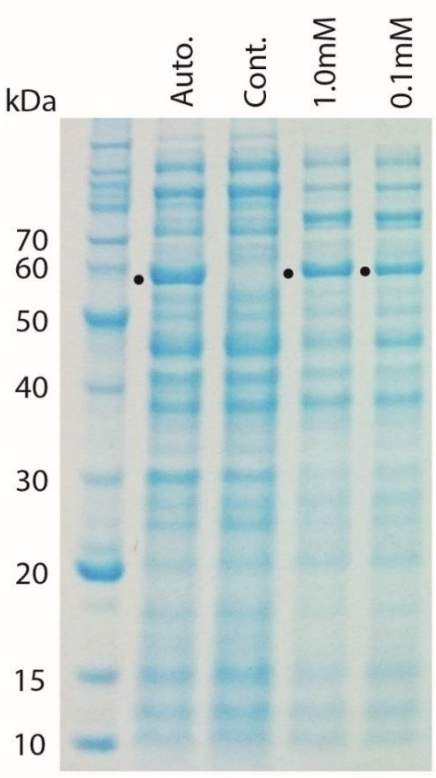


Figure 5.1: SDS-PAGE demonstrating comparison of MMOG expression methods. Soluble fractions of BL21(DE3) strains with pRSF-MMOG plasmid shown using either overnight autoinduction (Auto.) or IPTG induction (concentrations referenced) at 25°C. MMOG protein expression band emphasised with dots. Blank pRSF plasmid in BL21(DE3) used as control (Cont.). Protein ladder with adjacent molecular weights also included in right hand lane.

including sonication using Triton-based lysis buffer and chemical lysis methods proved to trigger protein aggregation of the formerly soluble MMOG component (results not shown).

GroES expression was also tested using a BL21(DE3) pET21-GroES transformant under corresponding conditions and found to express high levels of both soluble and insoluble protein. Further trials were also conducted to examine simultaneous MMOG/GroES expression. Dual transformant strains of BL21(DE3) were produced, either using pCOLA – MMOG/GroES plasmid, or pRSF-MMOG and pET21-GroES (or pET21-GroES(His)) as a dual transformant. Overnight autoinduction trials at 25°C showed reduced but still significant levels of soluble expression of both the MMOG and GroES components in all systems as compared with single component expression systems (Figure 5.2 & Figure 5.).

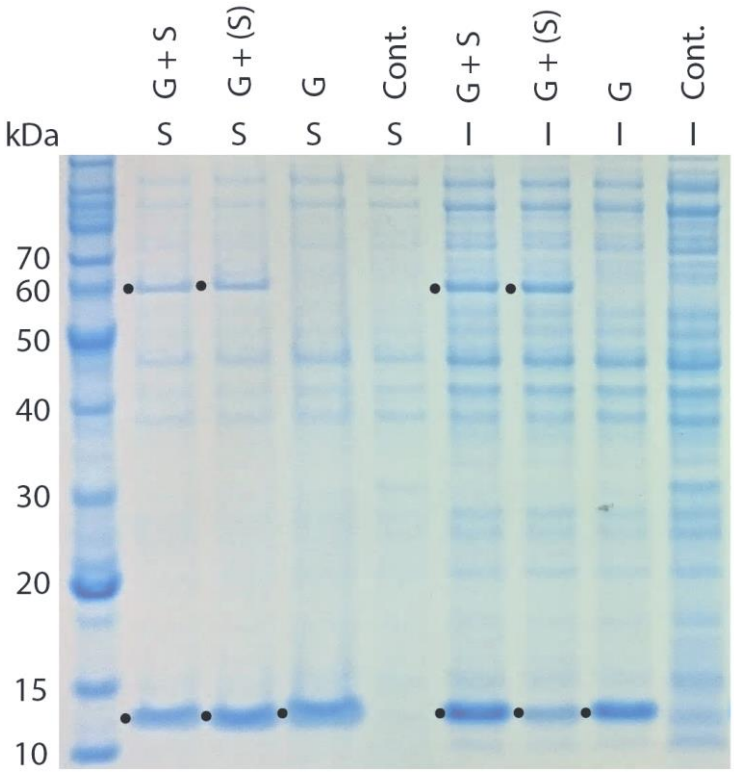


Figure 5.2: SDS-PAGE demonstrating soluble and insoluble expression of both MMOG and GroES. Soluble (S) and insoluble (I) fractions shown for BL21(DE3) strains with pRSF-mmoG/pET-groES (G+S), pRSF-mmoG/pET-groES(His) (G+(S)) or pRSF-mmoG (G). Overnight autoinduction at 25°C were used for all samples. Dots indicate molecular weight of proteins corresponding to inserted genes in respective strain. Blank pRSF plasmid in BL21(DE3) used as control (Cont.). Protein ladder with adjacent molecular weights also included in right hand lane.

Heterologous protein expression was demonstrated for all three MMOH subunits (α, β and γ) and MMOD using the pETD-X/D and pCDFD-Y/Z plasmids transformed into BL21(DE3) strains. Autoinduction at 37°C demonstrated high levels of gene expression with SDS-PAGE analysis indicating the four respective proteins dominating their expression profiles (Figure 5.). Comparable results was also observed during IPTG induction (0.2 mM at 37°C). Reducing autoinduction temperature to 25°C resulted in the MMOY MMOZ and MMOD component appearing predominantly in the soluble fraction, however MMOX did not undergo a comparable shift. The identity of SDS-PAGE expression bands for all proteins of interest were confirmed by mass spectrometry at least once for each gene construct, with the protein coverage indicating the bands contained undigested, full sequence sMMO components (See Appendix A7.3.4). The expression bands in subsequent trials were confirmed through comparison to these known reference bands.

5.3.2. Expression trials using triple transformant strains

Multiple attempts at transformation of BL21(DE3) strains with both pETD-X/D and pCDFD-Y/Z simultaneously by conventional heat shock methodology proved unsuccessful. Both plasmids were successfully transformed in this system as individuals, as well as the simultaneous transformant of pETD-X/D and pCDF-1b (blank). Furthermore, transformation efficiency was comparable between pCDF-1b (blank) and pCDFD-Y/Z at the plasmid concentrations used, indicating an expected correlation also between the transformation efficiency of the double transformants pETD-X/D + pCDFD-Y/Z and pETD-X/D + pCDF-1b (blank). Despite this, whilst 1 ml of transformation culture yielded 19 colonies using the pETD-X/D + pCDF-1b (blank) combination, no colonies were observed even after multiple independent attempts for the equivalent pETD-X/D + pCDFD-Y/Z transformation. Alternative transformations using the Lemo21 strain with pETD-X/D + pCDFD-Y/Z did prove successful, yielding ten transformant colonies per ml of transformation culture. To further confirm results in BL21(DE3) strains, a comparative trial using BL21(DE3) transformants under identical conditions to those found successful for Lemo21 transformation was performed and again found unsuccessful.

Simultaneous pETD-X/D + pCDFD-Y/Z + pCOLAD-ES/G triple transformants were attempted but proved unsuccessful, likely due to low transformation efficiency. Double transformant pETD-X/D + pCDFD-Y/Z cultures were therefore returned to a competent state, and transformed with either pCOLAD-ES/G or pRSF-MMOG: this achieving stable triple transformants containing the MMOX/MMOY/MMOZ/MMOD/MMOG ± GroES genetic inserts.

Initial expression attempts with X/D – Y/Z, X/D – Y/Z – G and X/D – Y/Z – G/S strains were conducted using overnight autoinduction expression to 37°C (Figure 5.). Translated proteins from all inserted

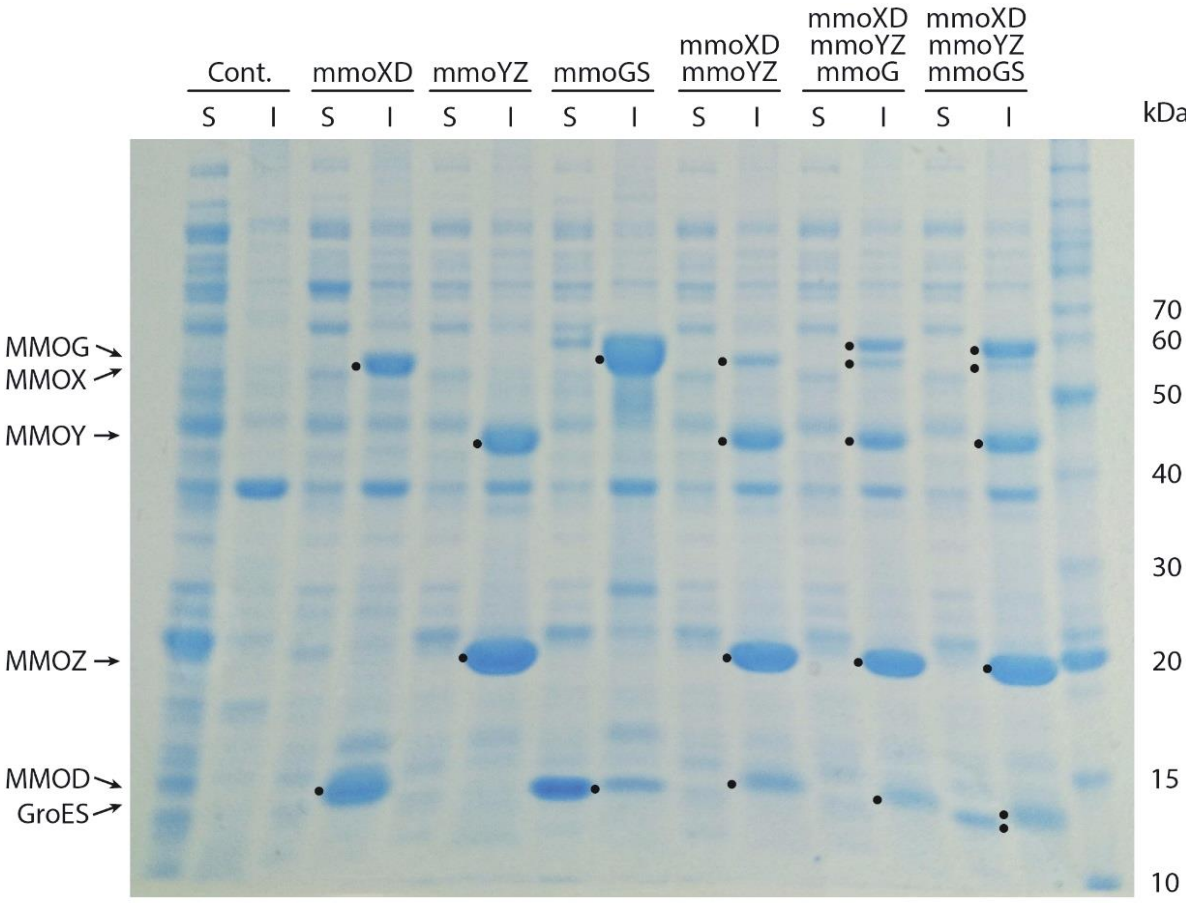


Figure 5.3: SDS-PAGE showing expression profiles of combinations of different constructs containing heterologous sMMO genes required for hydroxylase formation and assembly in *E. coli*. Soluble (S) and insoluble (I) fractions shown for various strains of Lemo21 *E. coli*. mmoXD denotes the pETD-X/D plasmid, mmoYZ denotes the pCDFD-Y/Z plasmid, mmoG denotes the pRSF-MMOG plasmid and mmoGS denotes the pCOLAD-S/G plasmid. Dots indicate molecular weight of proteins corresponding to inserted genes in respective strain, with corresponding identity of each dot at far left. Overnight autoinduction at 37°C were used for all samples without rhamnose supplimentation. Blank pCOLADuet-1 plasmid under identical expression conditions was used as control (Cont.). Protein ladder with adjacent molecular weights also included in left hand lane.

genes could be visualised on SDS-PAGE in at least one of the soluble or insoluble fractions, demonstrating both the stability of the plasmids in the transformants, and the ability to express from all plasmids simultaneously. In all strains however, MMOZ dominated the expression profile, with very low levels of MMOX expression. Reducing autoinduction temperatures to 25°C or 20°C resulted in much lower overall heterologous protein expression levels, and an increase in protein solubility as was seen to the individual plasmid trials, however this action also reduced MMOX and MMOD expression to levels below that which could be clearly visualised on SDS-PAGE gels (Figure 5.4). IPTG induction was also trialled, using varying concentrations of IPTG (1 mM, 0.5 mM, 0.2 mM and 0.05 mM) at both

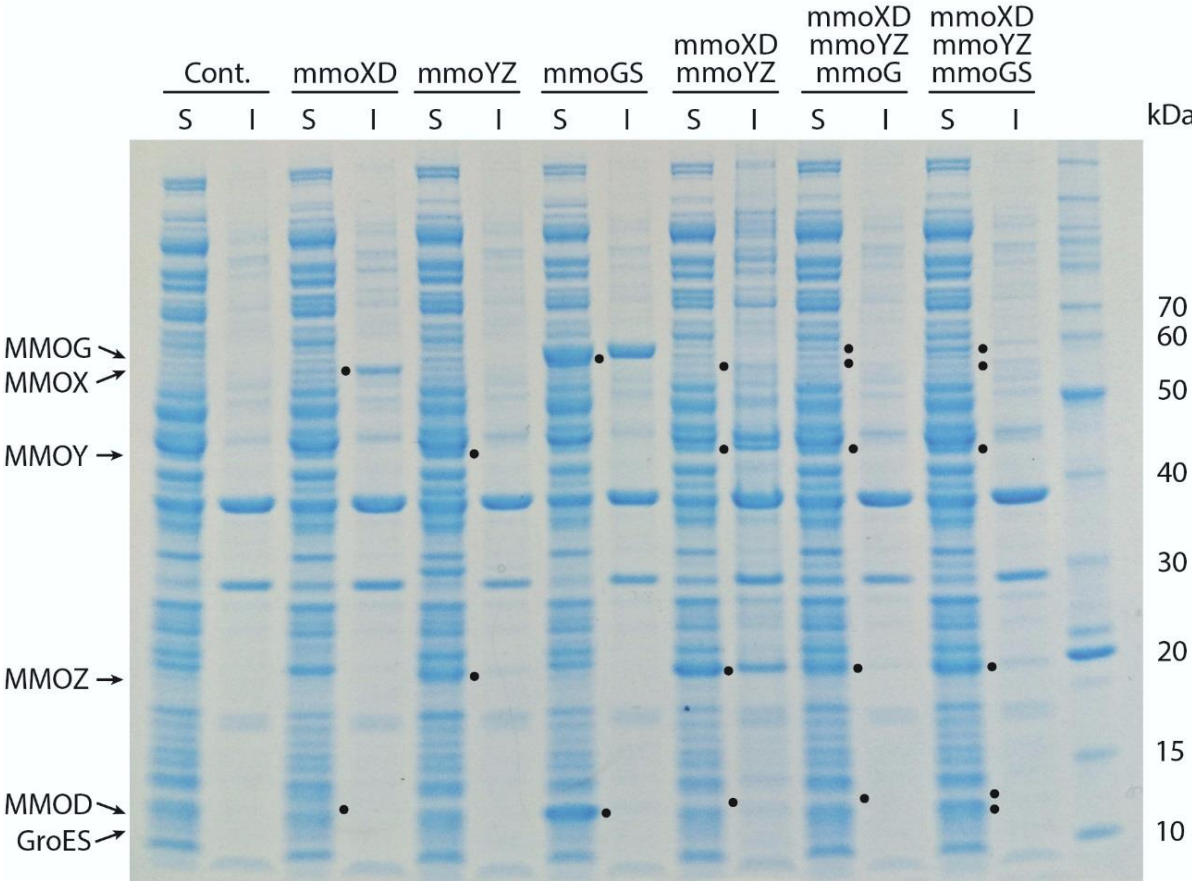


Figure 5.4: SDS-PAGE showing reduced temperature expression profiles of combinations of different constructs containing heterologous sMMO genes in *E. coli*. Soluble (S) and insoluble (I) fractions shown for various strains of Lemo21 *E. coli*. mmoXD denotes the pETD-X/D plasmid, mmoYZ denotes the pCDFD-Y/Z plasmid, mmoG denotes the pRSF-MMOG plasmid and mmoGS denotes the pCOLAD-S/G plasmid. Dots indicate molecular weight of proteins corresponding to inserted genes in respective strain, with corresponding identity of each dot at far right. Overnight autoinduction at 25°C were used for all samples without rhamnose supplimentation. Blank pCOLADuet-1 plasmid under identical expression conditions was used as control (Cont.). Protein ladder with adjacent molecular weights also included in left hand lane.

30°C and 37°C. Samples were extracted for SDS-PAGE analysis at 3 h, 5 h and 19 h after induction with IPTG, however did not show any significant improvement in protein expression compared to autoinduction trials.

Further to these efforts, several more advanced strategies were trialled to achieve solubilisation of all hydroxylase components: growth media supplemented with either iron sulphate (Fe^{2+}) or ferric citrate (Fe^{3+}) (both at 200 μM) was used to facilitate formation of the di-iron active centre of the hydroxylase; turn-down compatibility of the Lemo21 strains trialled (at both 25°C and 37°C) to decrease strain on the host organism; and heat shock at 42°C at various points either prior or during expression was also conducted to induce increased levels of native chaperone systems in the *E. coli* host. None of these methods resulted in clear soluble expression of all hydroxylase components (results not shown).

5.3.3. Naphthalene assay optimisation

Using *Methylosinus trichosporium OB3B* cultures grown in low copper NMS, the standard naphthalene assay was reproduced, with confirmation of the presence of active sMMO via red-pink colour formation (Figure 5.A). This result was only able to be reproduced in whole cells, with lysed cells from the same system unresponsive to the assay. Spectroscopic analysis of the solution gave an absorption peak for the conjugated dye at ca. 530 nm, in accordance with prior reports (Brusseau et al. 1990). Over the period of around 30 min, it was noted that the dye transitioned from the former red-pink hue towards an orange tint (Figure 5.a & 5.5b); this occurred concurrently with the control samples (methanotrophic cultures of *Methylocystis SC2* lacking the sMMO operon) adopting a yellow tint from its former clear/white solution typical of moderate cell densities. Spectroscopically this transition could be identified in both experimental and control samples as a gradual increase of absorption at 540 nm. This transition could be halted via the acidification of the solution with glacial acetic acid, producing a stable, vivid pink product (Figure 5.c).

To examine the spectroscopic sensitivity of the standard assay methodology, serial dilutions of the *M. trichosporium* culture were performed to produce OD_{600} concentrations between 1 – 0.001, with subsequent standard assay methodology performed. Clear spectroscopic differentiation between sMMO

containing and control samples was only possible down to an OD_{600} of around 0.1, with differentiation below this primarily prevented by interference from cell induced light scattering. This issue was found to be persistent despite attempts to control for such variations.

Visual identification of the presence of the azo dye proved more effective than spectroscopic methods at identifying weak signals, with sensitivity of visual identification further increased by centrifugation of the sample and characterisation based on cell pellet colour. Cellular lysis methods were performed on samples immediately prior to staining, but did not result in the partitioning the azo dye into the supernatant fraction. Solvent extraction of the cell pellet was found to be effective, with the stain preferentially partitioning into hydrophobic solvents. Of the solvents tested, ethyl acetate was found to most efficient to extract the stain, and provides a potential pathway for improvement of spectroscopic detection of dye formation. Due to identification of false negative results, cell free experiments were

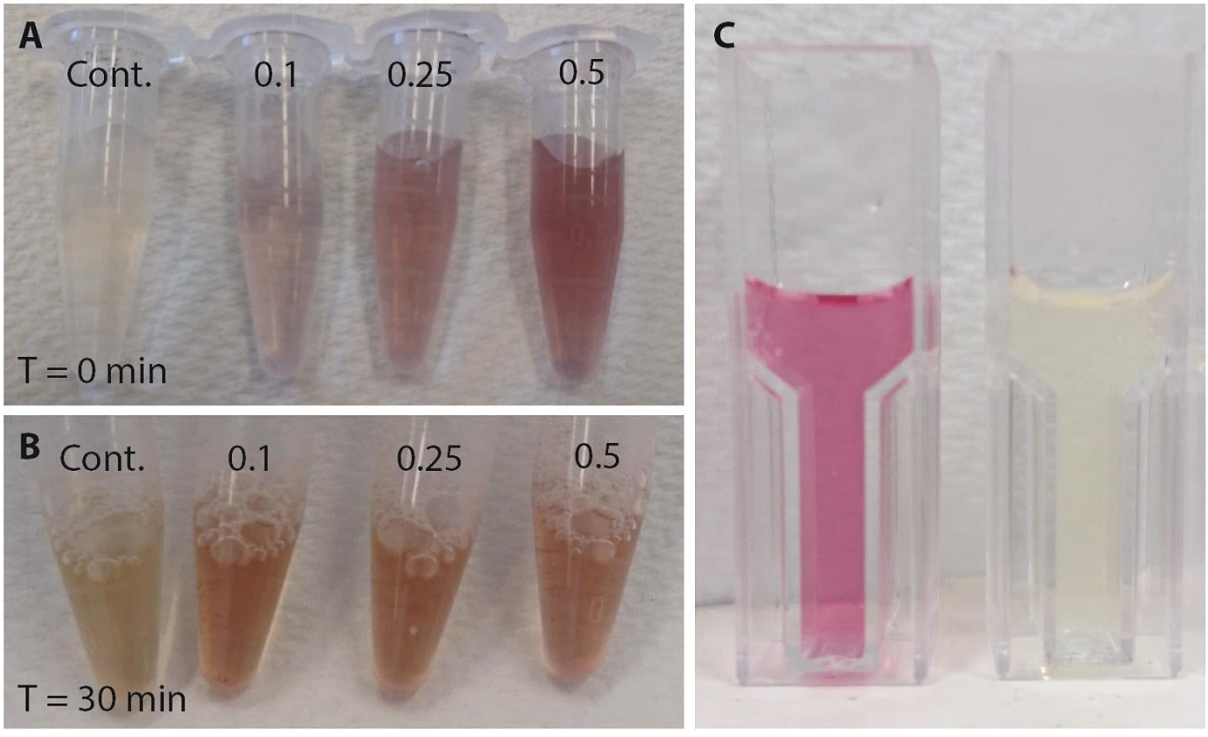


Figure 5.5: *M. trichosporium* after naphthalene assay at different cell concentrations and time intervals. Preparations of *M. trichosporium OB3b* grown in copper deficient NMS media after 2 hour naphthalene assay. Value indicate OD_{600} concentrations for each tube both **A)** immediately after addition of Fast blue B dye and **B)** 30 minutes after addition of dye. Control samples are *M. trichosporium* at 0.5 OD_{600} grown in copper rich media. **C)** Dye was also stabilised through addition of glacial acetic acid (right) showing transition to pink hue. Sample (left) is compared to control (right)

conducted under various conditions to identify the limitations of the assay methodology. The strongest response occurred under basification of the solutions (pH of 9.5), producing an orange hue similar to that formed in native samples over time (Figure 5.b), however additional parameters were also identified that affected the level of dye formation (results not shown).

5.3.4. Assessing heterologous sMMO expression *via* naphthalene assay

As the naphthalene assay for sMMO activity was restricted to whole cells, its application to heterologous expression required the simultaneous expression of all components involved in active turnover within the same cell. Towards this aim, MMOB and MMOR constructs were produced *de novo* from codon optimised genes whose sequence was obtained from *M. capsulatus*. Individual expression of each gene was successful with both showing almost complete soluble expression after autoinduction at 25°C (Appendix figure A7.14).

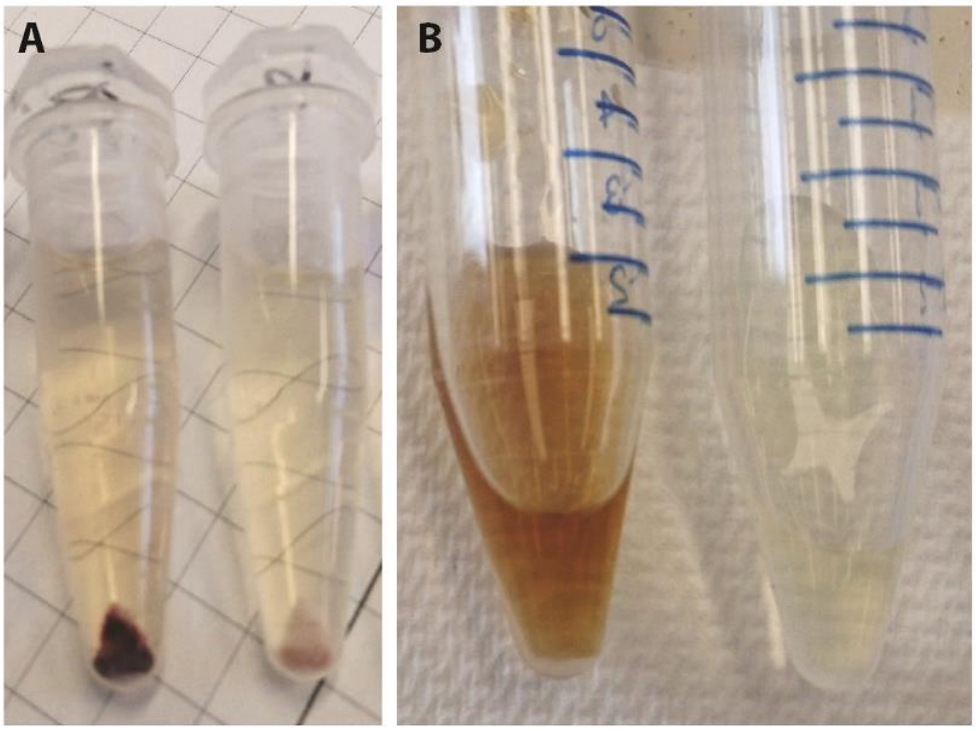


Figure 5.6: Cell pellets from M. trichosporium after naphthalene assay and cell free trials of naphthalene assay at varying pH. A) M. trichosporium samples grown in either Cu^- (left) or Cu^+ (right) media after naphthalene assay. In both cases OD_{600} of 1 was used during the assay. **B)** Cell free NMS media otherwise prepared as per standard cell preparations for naphthalene assay. High pH (9.5) sample due to potassium hydroxide addition (left) compared to control (right).

Due to the use of Lemo21 cell lines, only three of the four mutually compatible Duet plasmids could be simultaneously added due to antibiotic and replicon overlap with the pLemo plasmid (utilising pACYC backbone). To enable insertion of all eight genes into a single Lemo21 cell line, bicistronic constructs were produced of both MMOB/MMOR and MMOG/GroES, with insertion into the pRSFDuet-1 vector creating the pRSFD-B/R//G/S plasmid. Autoinduction using this plasmid in Lemo21 cell lines

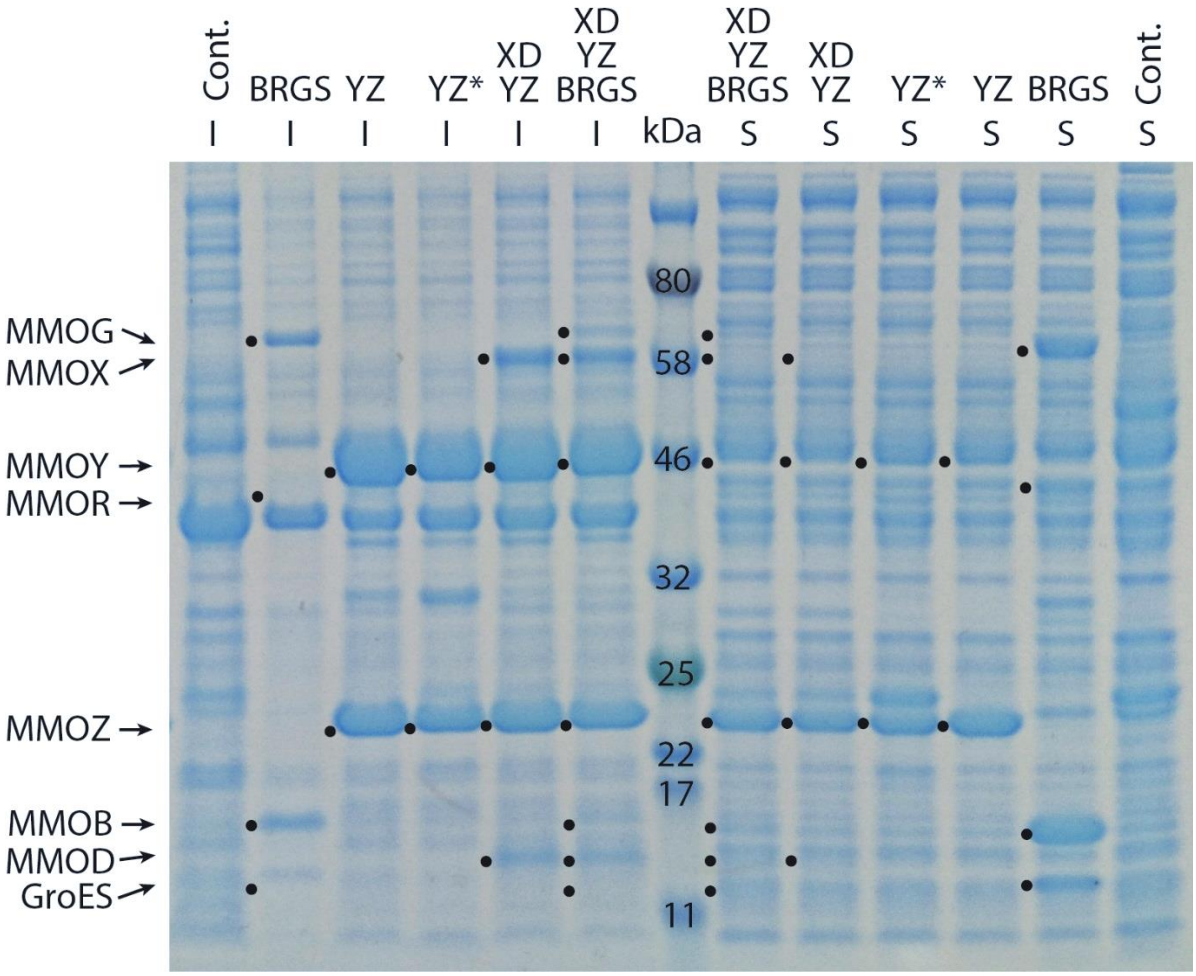


Figure 5.7: SDS-PAGE showing expression profiles of combinations of three different constructs containing heterologous sMMO genes including those for the reductase and regulatory units. Soluble (S) and insoluble (I) fractions shown for various strains in Lemo21 *E. coli* strains, or BL21(DE3) for strains with asterisks. Soluble and insoluble components are from same isolation and are matched around a vertical symmetry. YZ denotes the pCDFD-Y/Z plasmid, XD denotes the pETD-X/D plasmid and BRGS denotes the pRSFD-B/R//G/S plasmid. Dots indicate molecular weight of proteins corresponding to inserted genes in respective strain, with corresponding identity of each dot at far right. Overnight autoinduction at 30°C were used for all samples. Blank pCOLADuet-1 plasmid under identical expression conditions was used as control (Cont.). Protein ladder with superimposed molecular weights also included in centre lane.

demonstrated high level expression of all four genes, with trials at 30°C demonstrating all four components expressed in the soluble fraction (Figure 5.).

For naphthalene trials, the triple transformant capable of expressing all eight genes simultaneously was created using the pETD-X/D, pCDFD-Y/Z and pRSFD-B/R//G/S plasmids. Autoinduction trials showed lower expression levels of all components compared with the cell lines containing individual plasmids, with the highest reduction observed for the pRDFD-B/R//G/S plasmid. For those components that could still clearly be identified on SDS-PAGE, a reduction in the relative level of expressed protein in the soluble fraction was also observed (Figure 5.).

Results in line with successful active expression were obtained when these triple transformants underwent the naphthalene assay, with red cell pellet staining indicating active sMMO had been assembled. This result was strongest under 1.0 mM IPTG expression at 30°C in a 7 h incubation (Figure 5.), but also identified during IPTG induction at 25°C and autoinduction at both temperatures. The result was found to be IPTG concentration dependant, with strongest colour generation in the naphthalene assay occurring at higher IPTG concentrations (Figure 5.). While this result was observed in multiple instances, a positive assay result indicated by red pellet staining was not observed consistently, with some trials conducted under seemingly similar conditions lacking positive results. The inconsistency

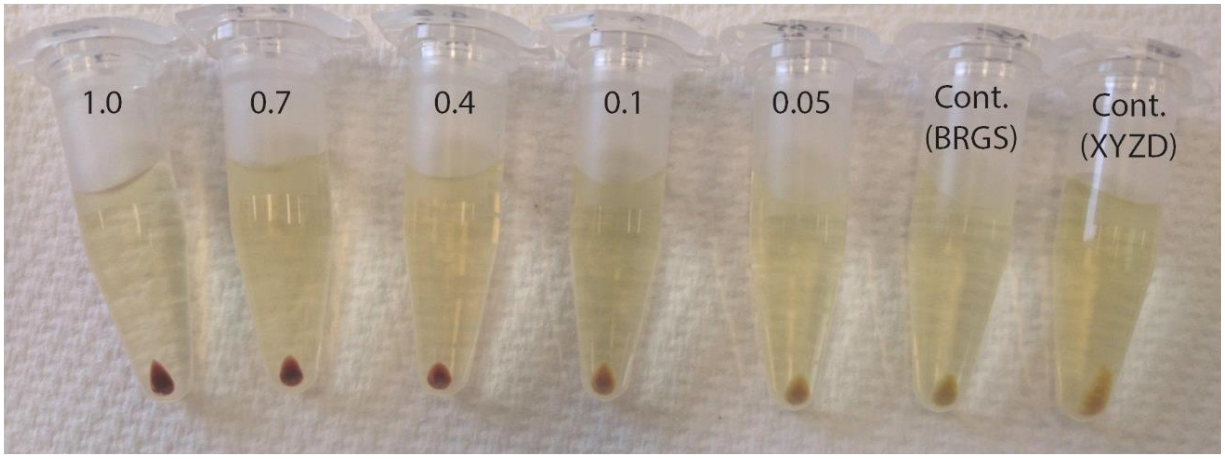


Figure 5.8: Naphthalene assays on E. coli transformants containing all genes required for active sMMO expression. Lemo21 strains containing the pETD-X/D, pCDFD-Y/Z and pRSFD-B/R//G/S plasmids after IPTG induction at 30°C and overnight naphthalene assaying. Positive results identified by red staining in cell pellet. Values indicate IPTG concentration in mM. Controls are Lemo21 strains either containing the pRSFD-B/R//G/S (BRGS) or both pETD-X/D and pCDFD-Y/Z (XYZD) plasmids.

was further complicated by a pink/red colour change observed in some trials from control samples lacking naphthalene, however standard control samples containing naphthalene as per naphthalene assay protocol did not exhibit comparable false positives. Positive results in naphthalene deficient samples were even seen to occur in trials where all naphthalene added samples, including those indicated to contain sMMO activity through prior positive assay results, yielded negative results (Appendix figure A7.15).

5.4. Discussion

This work provides a major step forward towards heterologous expression of sMMO in a highly tractable organism, not only demonstrating high level expression of all eight proteins that we consider necessary to achieve production of active sMMO, with seven of these produced in soluble form. Furthermore, though naphthalene assay results remain inconclusive, findings including an IPTG concentration dependant response can be most parsimoniously explained through *in vivo* sMMO activity in the *E. coli* host.

5.4.1. Simultaneous heterologous expression of sMMO components

Although this work is not the first to attempt simultaneous expression of all eight required genes associated with sMMO in *E. coli* (Dumont 2004), it is the first to achieve sufficient levels of all heterologous proteins for clear identification. This advance was most likely due to the use of *de novo* synthesised genes including codon optimisation, compared with previous attempts that utilised the operon directly from the native methanotroph host (Dumont 2004). The use of RBS to achieve required expression ratios of set proteins was partially successful, with reduced temperature of incubation and autoinduction mode consistently producing higher ratio of MMOX relative to MMOD, a critical issue for proper assembly and activity. The preferred ratio of expression was not observed for higher temperature induction experiments. However, it should be stated that protein subunit concentrations were not quantitatively analysed to definitively confirm that MMOX was in excess on a molar basis.

Despite extensive effort to design the expression system to balance gene expression between plasmids, especially for correlating production ratios between hydroxylase units, genes on pCDFD-Y/Z always expressed at significantly higher levels than pETD-X/D. While the Duet plasmid system is meant to provide that balance, in practice plasmid balancing was a major limitation for our genes during these experiments. This issue may be circumvented in future work through combining MMOX, MMOY and MMOZ into a polycistronic construct, linked on a duet plasmid to MMOD, MMOG or both. This arrangement should retain the majority of the Duet system interchangeability benefits whilst limiting imbalance issues, and would provide spatial proximity of the hydroxylase subunits both to each other and to their putative folding factors.

Attainment of soluble expression of all but the MMOX component (as demonstrated by SDS-PAGE analysis of the soluble fraction from *E. coli*) pointed to the most likely limitation to achieving high level heterologous activity of sMMO. MMOX, the α subunit of the hydroxylase, was detected in the insoluble fraction, most probably as a result of aggregation of the mis-folded nascent protein, or lack of incorporation of a correctly folded intermediate, into the supramolecular assembly; these factors not being mutually exclusive.

One of the proposed assembly proteins, MMOG, is an analogues of chaperonin (GroEL), a universal protein required for the correct folding of many cellular proteins. Although previous research into the MMOG component implicated the gene/protein in a transcriptional role during sMMO expression in the native organism, this does not preclude it from also having a role during sMMO translation. Indeed, its analogue in the closest evolutionary relative to sMMO in the BMM family has been explicitly shown to only be required during the translational stage of hydroxylase expression. This, in conjunction with the soluble expression of all other components in the sMMO operon, is strongly indicative of the primary role of MMOG being in folding of the MMOX component. Though a role for MMOG in hydroxylase assembly cannot be ruled out, the internal cavity for conventional GroEL in which folding occurs is limited to ca. 60 kDa, matching the size of the MMOX component alone. Therefore, if the mechanism of folding assistance is conserved between GroEL and the divergent MMOG homologue, then the larger supramolecular sMMO complex, even if only partially assembled, would be too large

for internalisation in the protein. Clearly, a deeper understanding of the nature of any interaction of MMOX with MMOG is needed and could improve outcomes in future heterologous expression and assembly of sMMO.

5.4.2. Indicators of activity and critical limitations of the naphthalene assay

A clear confirmation of successful assembly of active recombinant sMMO was prevented by inconsistent results of the naphthalene assay; improvement of the assay and resolution of the reasons for inconsistency is central to further progress of this research. A major limitation of the naphthalene assay was the lack of robustness of the method, in particular, its tendency to show high sensitivity to assay media components. Control samples using native *M. trichosporium OB3b* in either NMS or 1181 media produced expected positive results, however, the use of PBS or nitrogen-free mineral salts media with these microorganisms appeared to prevent dye formation. Not only false negatives but also false positives were observed using this methodology; several instances of the appearance of pink/red staining were identified in control samples, however were only seen to occur when the naphthalene substrate was omitted. While this colouration was more distinctly an orange shade, corresponding to the colour also identified to occur as positive samples were left over time, this colouration was difficult to differentiate from the pink-red staining of true positives especially when the response was weak.

5.4.3. Combination of hydroxylase units resulting in cytotoxicity

An unexpected result of this work was the identification of likely cytotoxicity issues upon combination of all three hydroxylase units and MMOD, even without the additional units required for sMMO activity. Individual transformation of the gene pairs in their respective plasmids were consistently successful, however, simultaneous transformation of the plasmids for the three hydroxylase units in standard BL21 (DE3) strains never produced viable colonies. Control trials were able to rule out common transformation issues such as plasmid stability, dual plasmid incompatibility and individual plasmid transformation efficiency.

Excluding the aforementioned possibilities, protein cytotoxicity is one of the most common factors preventing successful transformation, with low levels of constitutive expression of the cytotoxic protein resulting from a standard lactose-induced promotor causing cell death (Dumon-Seignovert et al. 2004). Successful transformation of genes for 'toxic' proteins can be achieved by eliminating 'leaky' constitutive expression, typically through the use of *E. coli* cell lines containing the T7 lysozyme capable of inhibiting the T7 polymerase under non-promoted conditions, such as Lemo21 (Studier 1991; Wagner et al. 2008). The trials using such a cell line described here were successful, even under identical conditions to those in which BL21 (DE3) expression were not, indicating that a protein-protein interaction is occurring on combining the pETD-X/D and pCDFD-Y/Z plasmids, resulting in cytotoxic effects.

5.4.4. Impact of initial design parameters

A key decision during the design stage was the selection of untagged proteins for expression. Although the addition of protein tags typically has negligible effects in protein expression, and can often provide additional benefits such as increased protein expression and solubility (Walls and Loughran 2011), the additional sequence may also destabilise the protein native state or interfere with the correct protein-protein interactions (Kapil et al. 2016). Both of these are significant possibilities for the sMMO system.

The lack of affinity tags for recombinant proteins limited the means of their identification to standard SDS-PAGE, and this was an issue for determining presence of these proteins in soluble protein fractions, especially under lower temperature incubation conditions which reduced recombinant protein expression. However, as formation of active sMMO was probably limited by either incorrect folding of the α sub-unit or hydroxylase assembly, selecting untagged proteins for expression was the correct decision. Furthermore, we hypothesised that the MMOX – MMOG interactions are particularly significant for this system and therefore inclusion of tags had the potential to impact the interaction between MMOX and MMOG, as well as increase the size of MMOX beyond that accessible to the MMOG internal folding cavity.

Another important consideration during the design phase was the selection of source organism for the sMMO genetic material. While *M. trichosporium OB3b* was a strong candidate for sMMO genes due to its prevalent use in the literature and high level of characterisation, *M. capsulatus (Bath)* was ultimately selected on the basis that *M. trichosporium* is not viable at 37°C, the standard expression conditions for *E. coli*: this suggested that proteins derived from *M. trichosporium* may not be stable at such temperatures. The highest levels of soluble heterologous protein expression however occurred at 25°C, the optimum growth temperature for *M. trichosporium*, thus indicating this species as a possible alternative for future trials.

5.5. Conclusion

This chapter presents the first known high level heterologous expression of all components required for the synthesis of an active methane monooxygenase system, as well as expression of the additional putative folding enzymes identified in Chapter 4. The investigation verified soluble expression of all units related to either folding or activity of the hydroxylase, as well as two out of the three subunits of the hydroxylase itself, all simultaneously. Subsequent naphthalene assay trials, though inconclusive due to repeatability issues, demonstrated results most parsimoniously explained through heterologous sMMO activity *in vivo*. This work provides significant progress towards the achievement of heterologous expression of active sMMO in a highly tractable organism, and will help to direct future research towards this goal.

5.7. References

Borodina E, Nichol T, Dumont MG, Smith TJ, Murreil JC. (2007) Mutagenesis of the "leucine gate" to explore the basis of catalytic versatility in soluble methane monooxygenase. *Appl. Environ. Microbiol.* 73, 6460–6467.

Brusseau G a., Tsien HC, Hanson RS, Wackett LP. (1990) Optimization of trichloroethylene oxidation by methanotrophs and the use of a colorimetric assay to detect soluble methane monooxygenase activity. *Biodegradation* 1, 19–29.

Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Di Donato A. (2004) Phenol hydroxylase and toluene/o-xylene monooxygenase from Pseudomonas stutzeri OX1: Interplay between two enzymes. *Appl. Environ. Microbiol.* 70, 2211–2219.

Cafaro V, Scognamiglio R, Viggiani A, Izzo V, Passaro I, Notomista E, Dal Piaz F, Amoresano A, Casbarra A, Pucci P, et al. (2002) Expression and purification of the recombinant subunits of toluene/o-xylene monooxygenase and reconstitution of the active complex. *Eur. J. Biochem.* 269, 5689–5699.

Champreda V, Zhou N-Y, Leak DJ. (2004) Heterologous expression of alkene monooxygenase components from Xanthobacter autotrophicus Py2 and reconstitution of the active complex. *FEMS Microbiol. Lett.* 239, 309–318.

Dedysh SN, Khmelenina VN, Suzina NE, Trotsenko YA, Semrau JD, Liesack W, Tiedje JM. (2002) Methylocapsa acidiphila gen. nov., sp. nov., a novel methane-oxidizing and dinitrogen-fixing acidophilic bacterium from Sphagnum bog. *Int. J. Syst. Evol. Microbiol.* 52, 251–261.

Dumon-Seignovert L, Cariot G, Vuillard L. (2004) The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* 37, 203–206.

Dumont MG. (2004) Diversity , Mutagenesis and Recombinant Expression of the Soluble Methane Monooxygenase.

Dunfield PF, Khmelenina VN, Suzina NE, Trotsenko YA, Dedysh SN. (2003) Methylocella silvestris sp. nov., a novel methonotroph isolated from an acidic forest cambisol. *Int. J. Syst. Evol. Microbiol.* 53, 1231–1239.

Elango N, Radhakrishnan R, Froland W a, Wallar BJ, Earhart C a, Lipscomb JD, Ohlendorf DH. (1997) Crystal structure of the hydroxylase component of methane monooxygenase from Methylosinus trichosporium OB3b. *Protein Sci.* 6, 556–568.

Espah Borujeni A, Channarasappa AS, Salis HM. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42, 2646–2659.

Furuya T, Hayashi M, Kino K. (2013) Reconstitution of active mycobacterial binuclear iron monooxygenase complex in escherichia coli. *Appl. Environ. Microbiol.* 79, 6033–6039.

Furuya T, Hayashi M, Semba H, Kino K. (2013) The mycobacterial binuclear iron monooxygenases require a specific chaperonin-like protein for functional expression in a heterologous host. *FEBS J.* 280, 817–826.

Gassner GT, Lippard SJ. (1999) Component interactions in the soluble methane monooxygenase system from Methylococcus capsulatus (Bath). *Biochemistry* 38, 12768.

Graham DW, Korich DG, LeBlanc RP, Sinclair NA, Arnold RG. (1992) Applications of a colorimetric plate assay for soluble methane monooxygenase activity. *Appl. Environ. Microbiol.* 58, 2231–2236.

Izzo V, Leo G, Scognamiglio R, Troncone L, Birolo L, Di Donato A. (2011) PHK from phenol hydroxylase of Pseudomonas sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases. *Arch. Biochem. Biophys.* 505, 48–59.

Jahng D, Kim CS, Hanson RS, Wood TK. (1996) Optimization of trichloroethylene degradation using soluble methane monooxygenase of Methylosinus trichosporium OB3b expressed in recombinant bacteria. *Biotechnol. Bioeng.* 51, 349–359.

Jahng D, Wood TK. (1994) Trichloroethylene and chloroform degradation by a recombinant pseudomonad expressing soluble methane monooxygenase from Methylosinus trichosporium OB3b. *Appl. Environ. Microbiol.* 60, 2473–2482.

Jiang H, Chen Y, Jiang P, Zhang C, Smith TJ, Murrell JC, Xing XH. (2010) Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. *Biochem. Eng. J.* 49, 277–288.

Kalidass B, Ul-Haque MF, Baral BS, DiSpirito A a, Semrau JD. (2015) Competition between metals for binding to methanobactin enables expression of soluble methane monooxygenase in the presence of copper. *Appl. Environ. Microbiol.* 81, 1024–1031.

Kapil D, Gelman H, Thu CTH, Guin D, Gruebele M. (2016) The Effect of Fluorescent Protein Tags on Phosphoglycerate Kinase Stability Is Nonadditive. *J. Phys. Chem. B* 120, 2878–2885.

Larsen Ø, Karlsen OA. (2016) Transcriptomic profiling of Methylococcus capsulatus (Bath)during growth with two different methane monooxygenases. *Microbiologyopen* 5, 254–267.

Lim HN, Lee Y, Hussein R. (2011) Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci.* 108, 10626–10631.

Lloyd JS, De Marco P, Dalton H, Murrell JC. (1999) Heterologous expression of soluble methane monooxygenase genes in methanotrophs containing only particulate methane monooxygenase. *Arch. Microbiol.* 171, 364–370.

Merkx M, Kopp D a., Sazinsky MH, Blazyk JL, Muanẗller J, Lippard SJ. (2001) Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: A tale of two irons and three proteins. *Angew. Chemie - Int. Ed.* 40, 2782–2807.

Merkx M, Lippard SJ. (2002) Why OrfY? Characterization of MMOD, a long overlooked component of the soluble methane monooxygenase from Methylococcus capsulatus (Bath). *J. Biol. Chem.* 277, 5858.

Pacheco-Oliver M, McDonald IR, Groleau D, Murrell JC, Miguez CB. (2002) Detection of methanotrophs with highly divergent pmoA genes from Arctic soils. *FEMS Microbiol. Lett.* 209, 313–319.

Rosano GL, Ceccarelli EA. (2014) Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* 5, 1–17.

Rosenzweig AC, Brandstetter H, Whittington D a., Nordlund P, Lippard SJ, Frederick C a. (1997) Crystal structures of the methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath): Implications for substrate gating and component interactions. *Proteins Struct. Funct.*

Genet. 29, 141–152.

Salis HM. (2011) The Ribosome Binding Site Calculator.

Smith TJ, Slade SE, Burton NP, Murrell JC, Dalton H. (2002) Improved system for protein engineering of the hydroxylase component of soluble methane monooxygenase. *Appl. Environ. Microbiol.* 68, 5265–5273.

Stafford GP, Scanlan J, McDonald IR, Murell JC. (2003) rpoN, mmoR and mmoG, genes involved in regulating the expression of soluble methane monooxygenase in Methylosinus trichosporium OB3b. *Microbiology* 149, 1771–1784.

Studier FW. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37–44.

Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO, et al. (2008) Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14371.

Walls D, Loughran ST. (2011) Tagging recombinant proteins to enhance solubility and aid purification. In: Walls D, Loughran ST, editors. Methods in molecular biology. p. 151–175.

Walters KJ, Gassner GT, Lippard SJ, Wagner G. (1999) Structure of the soluble methane monooxygenase regulatory protein B. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7877–7882.

West CA, Salmond GPC, Dalton H, Murrell JC. (1992) Functional expression in Escherichia coli of proteins B and C from soluble methane monooxygenase of Methylococcus capsulatus (Bath). *J. Gen. Microbiol.* 138, 1301–1307.

Whittenbury R, Phillips KC, Wilkinson JF. (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. *J. Gen. Microbiol.* 61, 205–218.

Yu Y, Ramsay JA, Ramsay BA. (2009) Production of soluble methane monooxygenase during growth of Methylosinus trichosporium on methanol. *J. Biotechnol.* 139, 78–83.

This page is intentionally blank

CHAPTER 6

CONCLUSION AND FUTURE WORK

This page is intentionally blank

6.1. Conclusion

The research presented in this thesis marks a significant development in our understanding of the nature and evolutionary history of the bacterial methanotrophs, drawing together previously unknown aspects of their evolutionary history that inform our understanding of the critical enzyme systems that facilitate methanotrophy: the methane monooxygenases. Then, by leveraging this knowledge towards sMMO heterologous expression, major progress was also achieved towards attainment of an active recombinant sMMO system in a highly tractable organism.

Critical to the elucidation of many of the conclusions around the evolutionary history of both methanotrophy and the methane monooxygenases has been the amalgamation of results obtained using methodology developed across different disciplines, each independently insufficient to reveal the underlying factor/behaviour. This has included the use of phylogenetic taxonomy, catalytic mechanisms, molecular evolution, protein-sequence-structure analysis, operon based analysis, protein-protein interactions and molecular biology. By drawing together results from these different areas of investigation, this work has brought new knowledge to several areas of our current understanding of methanotrophy.

Whilst the pMMO enzyme system has classically been the focal point for understanding bacterial methanotrophy, in this study the finding that methanotrophy arose in the Proteobacteria from horizontal gene transfer of three different methane monooxygenase systems places a renewed emphasis on the roles played by the two other methane monooxygenases: sMMO and pXMO. Furthermore, by revealing that the ancestral families of these methane monooxygenase enzyme systems were already linked prior to each developing activity on methane, and have remained linked in evolutionary branches active towards other substrates, suggests mechanisms for association that are more fundamental than explicit methane reactivity. Also in this analysis it was shown that horizontal gene transfer giving rise to the bacterial methanotrophs was an infrequent event. This was proposed to be due to characteristics of pre-methanotrophic ancestors such as extensive intracytoplasmic membranes, some of which are still

identifiable both in the present day methanotrophic and non-methanotrophic descendants from this ancestral event.

sMMO itself was shown in this research to be at the end of an evolutionary branch which has progressively become more reactive towards higher energy C-H bonds. The progressive increase in reactivity of the active centre appears to have developed at the cost of reduced stability of the hydroxylase unit, and required the simultaneous development of additional specialised auxiliary proteins to support the catalytic unit. These factors facilitate folding, assembly and stability of the hydroxylase unit, and demonstrate convergent evolution with accessory proteins appearing elsewhere in the BMM family. Through protein sequence- structure analysis of active site residue changes that are concurrent with development of methane reactivity, this research developed an evolutionarily-informed hypothesis of active site intermediate formation that enabled catalytic activity towards methane.

The updated analysis of the evolutionary development of sMMO undertaken in this thesis which highlighted the importance of specialised auxiliary proteins, was then employed in the design strategy for heterologous expression of this enzyme system. The strategy not only included the co-expression of these factors but also addressed the balance of expression ratios of specific auxiliary factor to catalytic unit to prevent inhibition, and implementation of heat shock during expression to increase native production of elements of the chaperone pathway. This advanced heterologous expression strategy for sMMO in *E. coli* delivered all the required components and subunits at high level and gave soluble expression of all but one of these units, also producing tentative evidence of *in vivo* hydroxylase activity *via* the naphthalene assay.

6.2. Future work

The most immediate unresolved question arising from this work is to whether successful heterologous expression of the sMMO system was achieved in *E. coli*. Though high level expression of all components, and tentative positive activity assay results were achieved, confirmation of these results is of central importance. As the use of an azo dye in the reporting system was determined to be the most

likely cause of the ambiguities during activity trials, use of an alternative method for evaluating naphthol production would likely resolve the current method limitations. Contingent on the confirmation of earlier positive results, direct demonstration of methane oxidation would then be required to conclusively demonstrate sMMO activity. This system could also be used to test whether both MMOD and MMOG are required for the production of active sMMO hydroxylase, one of the central hypothesis in the thesis.

Neither the biomes in which methanotrophs inhabit, nor potential feedstocks for future biotechnological applications, consist of pure methane: the gasses that methanotrophs are exposed to are complex mixtures with at least trace levels of additional reduced gasses. This work revealed a link between methanotrophs and these reduced gasses at both at an evolutionary and genetic inventory level. Though the literature contains studies examining the growth of specific methanotroph strains in the presence of either ammonia, hydrogen sulphide or hydrogen, there has been no continuity of strains used, nor comparison of these results to the comparative genetic inventory of the specific strain. A guided investigation of co-metabolism of reduced gasses in the methanotrophs is therefore important to determine the true capacity in this area, with specific selection of trial strains based on genetic inventories such as is found in this work. This will facilitate a better understanding of the interaction of methanotrophs with complex gasses that are more environmentally and technologically relevant.

Testable hypothesis were also generated during BMM research towards substrate ranges of specific active site intermediates, and active site modifications that enabled activity towards methane. Comparison of BMM substrate ranges with spectroscopic determination of active site intermediates, in particular on the ethane monooxygenases and three component alkene monooxygenases, would be able to directly test the former proposal. Evaluation of the latter could be tested through reversion of either the sMMO or BMO hydroxylase to the active site residues contained in the remainder of the family. If methane activity was lost without complete loss of activity of the active site, this would confirm this hypothesis.

In addition to this work opening up multiple immediate directions for future research, through challenging our current understanding of methanotrophy and the methane monooxygenases it also generates more fundamental questions, the answer of which will be of major importance to multiple current research fields. The evidence presented in this work demonstrating horizontal gene transfer of methane monooxygenase genes as the origin of methanotrophy in bacteria contrasts previous proposals of vertical genetic descent and necessitates specific methanotrophic transition events for each current radiation. Considering the significance of the methanotrophs in the biogeosphere, the question of when each specific radiation arose, and the fate of the methane that these radiations subsequently consumed prior to their appearance is currently unresolved. The limited number of these HGT events also implicates a specific set of characteristics in these pre-methanotrophic ancestors that enabled this transition. Understanding the nature of these pre-methanotrophic organisms, and why methanotrophy specifically arose in these species, could not only shed light on current methanotroph characteristics including the nature of obligate methanotrophy, but also facilitate current biotechnological efforts to generate *de novo* methanotrophic species.

The lateral movement of a conserved group of three distinct monooxygenase systems through the proteobacteria, giving rise to not only methanotrophic but ammonia and short chain alkane oxidising bacteria, also points towards a currently uncharacterised degree of mutualism between these monooxygenase systems. A deeper understanding of these systems will not only require the identification of the functional characteristics causing three enzyme systems with primary activity on the same substrate displaying mutualistic behaviour, but also why such interaction is no longer necessary in the majority of extant methanotrophic species. As methane monooxygenase systems divergent from the (almost) universal pMMO system in methanotrophs have been implicated in several important areas of methanotrophy under current investigation including atmospheric and intra-aerobic methane oxidation, understanding the interaction between pMMO and these alternative systems has the potential to facilitate major progress in these areas.

CHAPTER A7

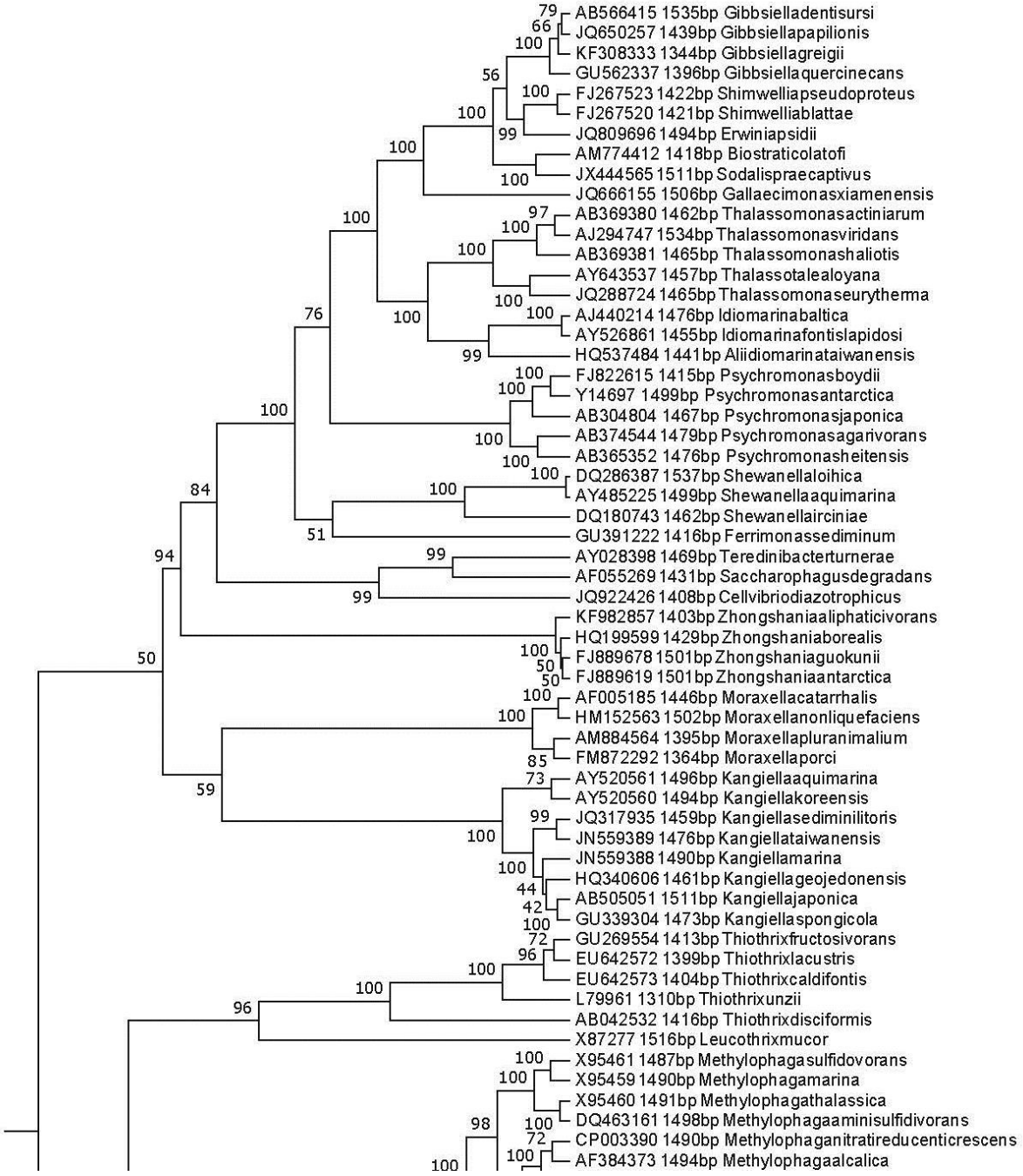
APPENDIX

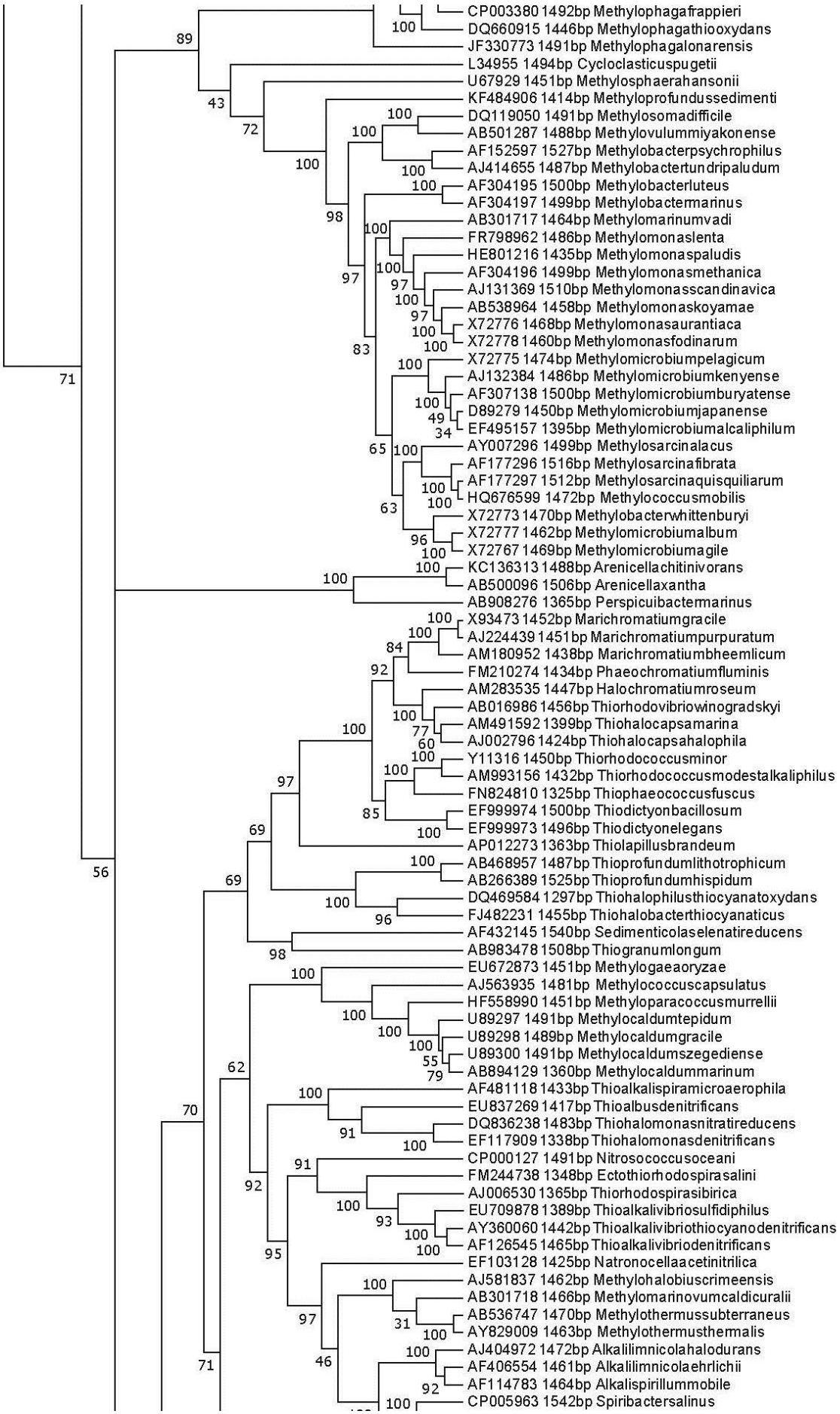
This page is intentionally not blank

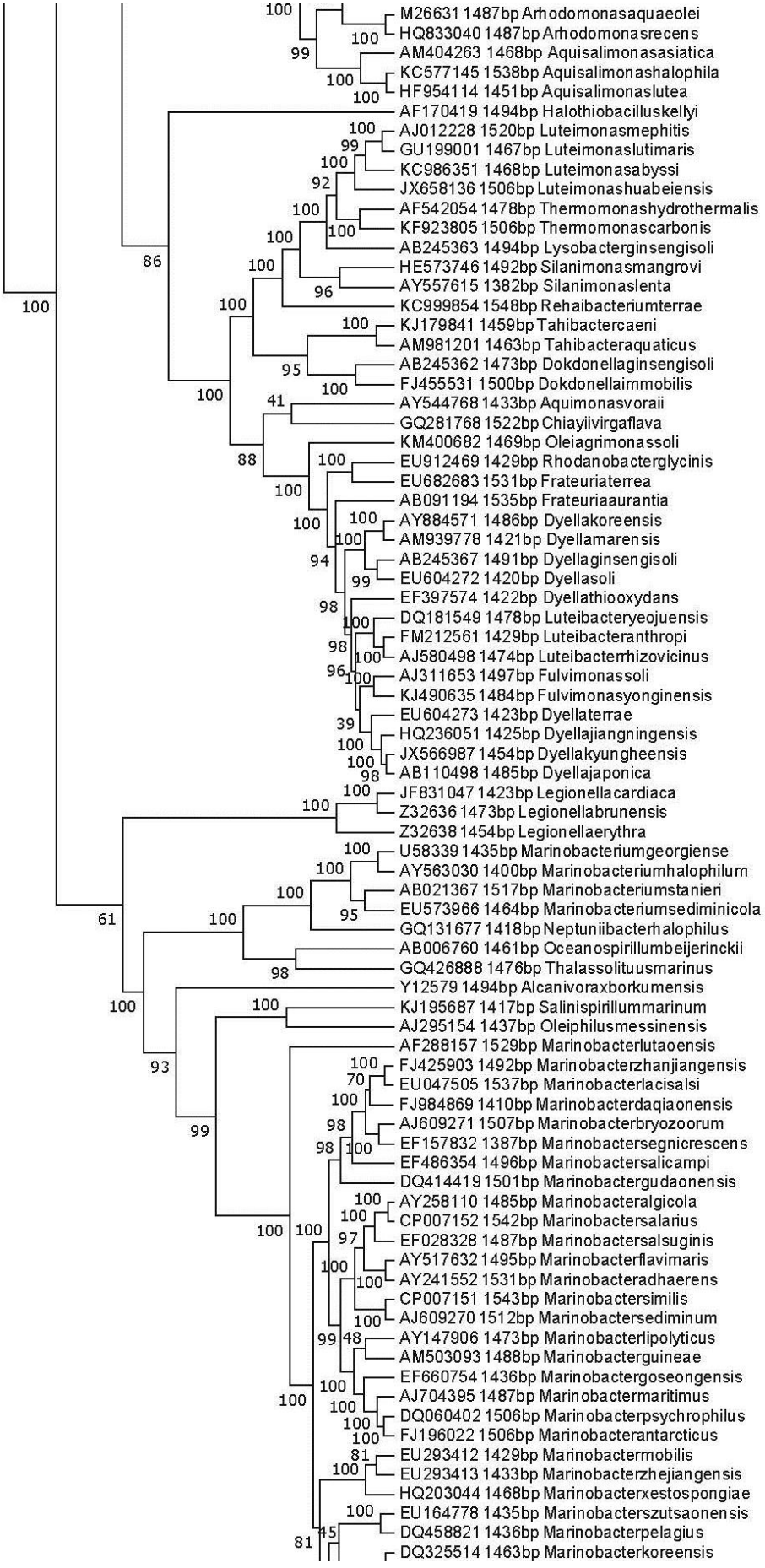
A7.1. Supplementary figures

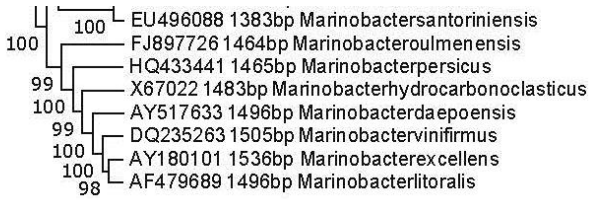
A7.1.1. Supplementary figures for section 3.1.

Supplementary Figure (A7.)1: 16S based phylogeny of region surrounding gammaproteobacterial methanotrophs. Taxa were obtained using sequential BLAST analyses of the SILVA LTP database centred on 16S ribosome units of three selected methanotrophs characterising each of the major gammaproteobacterial subgroups. The 100 closest results from each BLAST analysis were then merged to a single dataset and aligned using the Q-INS-i methodology in MAFFT. Phylogeny of the resulting 227 sequences was inferred using Bayesian Markov chain Monte Carlo analysis using the GTR+G+I model and an uncorrelated relaxed molecular clock in BEAST. Posterior probabilities have been included for all nodes (as percentage values), as well as accession numbers prior to species name.

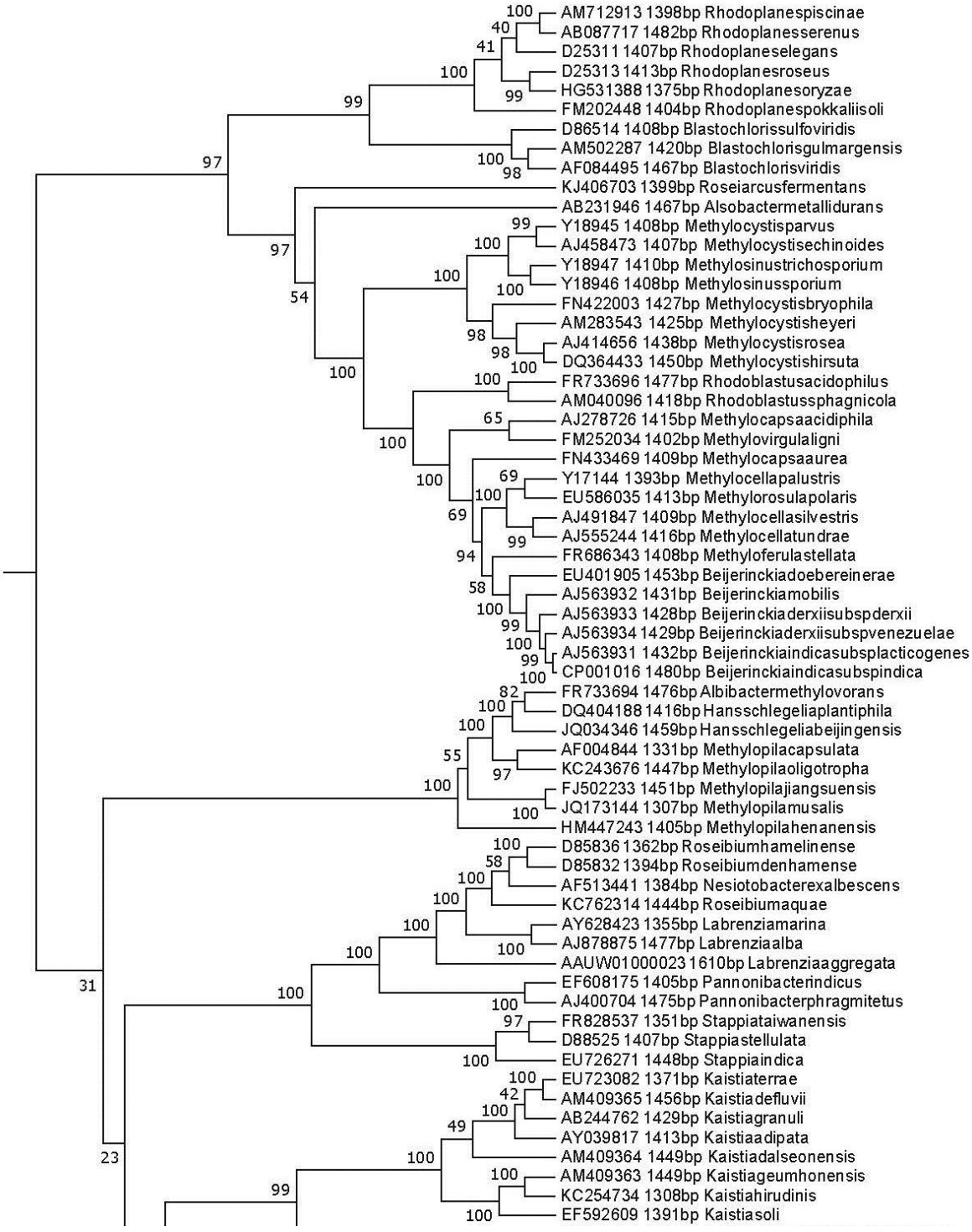


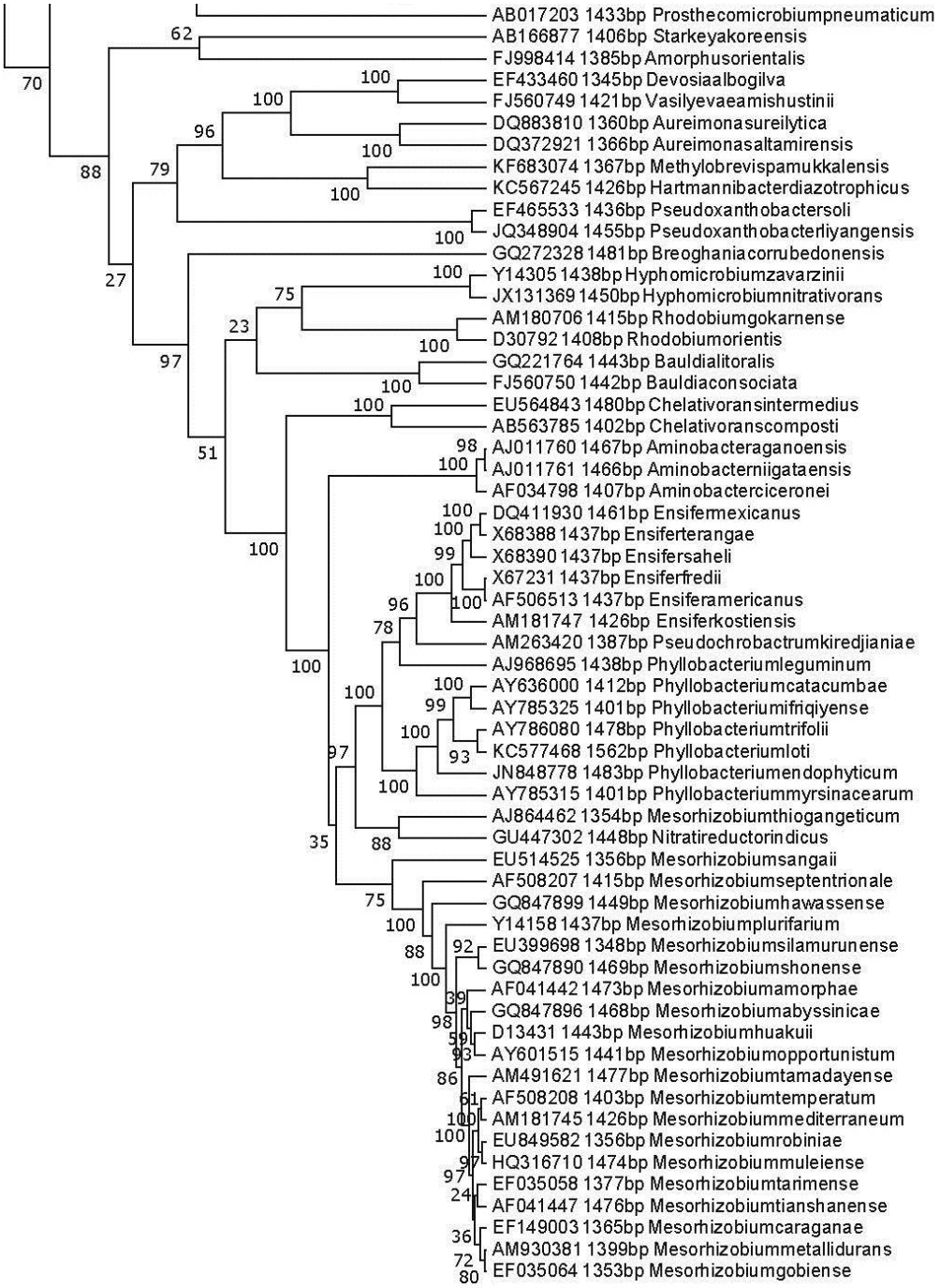

CP003380 1492bp Methylophagafrappieri
DQ660915 1446bp Methylophagathiooxydans
JF330773 1491bp Methylophagalonarensis
L34955 1494bp Cycloclasticuspugetii
U67929 1451bp Methylosphaerahansonii
KF484906 1414bp Methyloprofundussedimenti
DQ119050 1491bp Methylosomadifficile
AB501287 1488bp Methylovulummiyakonense
AF152597 1527bp Methylobacterpsychrophilus
AJ414655 1487bp Methylobactertundripaludum
AF304195 1500bp Methylobacterluteus
AF304197 1499bp Methylobactermarinus
AB301717 1464bp Methylomarinumvadi
FR798962 1486bp Methylomonaslenta
HE801216 1435bp Methylomonaspaludis
AF304196 1499bp Methylomonasmethanica
AJ131369 1510bp Methylomonasscandinavica
AB538964 1458bp Methylomonaskoyamae
X72776 1468bp Methylomonasaurantiaca
X72778 1460bp Methylomonasfodinarum
X72775 1474bp Methylomicrobiumpelagicum
AJ132384 1486bp Methylomicrobiumkenyense
AF307138 1500bp Methylomicrobiumburyatense
D89279 1450bp Methylomicrobiumjapanense
EF495157 1395bp Methylomicrobiumalcaliphilum
AY007296 1499bp Methylosarcinalacus
AF177296 1516bp Methylosarcinafibrata
AF177297 1512bp Methylosarcinaquisquiliarum
HQ676599 1472bp Methylococcusmobilis
X72773 1470bp Methylobacterwhittenburyi
X72777 1462bp Methylomicrobiumalbum
X72767 1469bp Methylomicrobiumagile
KC136313 1488bp Arenicellachitinivorans
AB500096 1506bp Arenicellaxantha
AB908276 1365bp Perspicuibactermarinus
X93473 1452bp Marichromatiumgracile
AJ224439 1451bp Marichromatiumpurpuratum
AM180952 1438bp Marichromatiumbheemlicum
FM210274 1434bp Phaeochromatiumfluminis
AM283535 1447bp Halochromatiumroseum
AB016986 1456bp Thiorhodovibriowinogradskyi
AM491592 1399bp Thiohalocapsamarina
AJ002796 1424bp Thiohalocapsahalophila
Y11316 1450bp Thiorhodococcusminor
AM993156 1432bp Thiorhodococcusmodestalkaliphilus
FN824810 1325bp Thiophaeococcusfuscus
EF999974 1500bp Thiodictyonbacillosum
EF999973 1496bp Thiodictyonelegans
AP012273 1363bp Thiolapillusbrandeum
AB468957 1487bp Thioprofundumlithotrophicum
AB266389 1525bp Thioprofundumhispidum
DQ469584 1297bp Thiohalophilusthiocyanatoxydans
FJ482231 1455bp Thiohalobacterthiocyanaticus
AF432145 1540bp Sedimenticolaselenatireducens
AB983478 1508bp Thiogranumlongum
EU672873 1451bp Methylogaeaoryzae
AJ563935 1481bp Methylococcuscapsulatus
HF558990 1451bp Methyloparacoccusmurrellii
U89297 1491bp Methylocaldumtepidum
U89298 1489bp Methylocaldumgracile
U89300 1491bp Methylocaldumszegediense
AB894129 1360bp Methylocaldummarinum
AF481118 1433bp Thioalkalispiramicroaerophila
EU837269 1417bp Thioalbusdenitrificans
DQ836238 1483bp Thiohalomonasnitratireducens
EF117909 1338bp Thiohalomonasdenitrificans
CP000127 1491bp Nitrosococcusoceani
FM244738 1348bp Ectothiorhodospirasalini
AJ006530 1365bp Thiorhodospirasibirica
EU709878 1389bp Thioalkalivibriosulfidiphilus
AY360060 1442bp Thioalkalivibriothiocyanodenitrificans
AF126545 1465bp Thioalkalivibriodenitrificans
EF103128 1425bp Natronocellaacetinitrilica
AJ581837 1462bp Methylohalobiuscrimeensis
AB301718 1466bp Methylomarinovumcaldicuralii
AB536747 1470bp Methylothermussubterraneus
AY829009 1463bp Methylothermusthermalis
AJ404972 1472bp Alkalilimnicolahalodurans
AF406554 1461bp Alkalilimnicolaehrlichii
AF114783 1464bp Alkalispirillummobile
CP005963 1542bp Spiribactersalinus


M26631 1487bp Arhodomonasaquaeolei
HQ833040 1487bp Arhodomonasrecens
AM404263 1468bp Aquisalimonasasiatica
KC577145 1538bp Aquisalimonashalophila
HF954114 1451bp Aquisalimonaslutea
AF170419 1494bp Halothiobacilluskellyi
AJ012228 1520bp Luteimonasmephitis
GU199001 1467bp Luteimonaslutimaris
KC986351 1468bp Luteimonasabyssi
JX658136 1506bp Luteimonashuabeiensis
AF542054 1478bp Thermomonashydrothermalis
KF923805 1506bp Thermomonascarbonis
AB245363 1494bp Lysobacterginsengisoli
HE573746 1492bp Silanimonasmangrovi
AY557615 1382bp Silanimonaslenta
KC999854 1548bp Rehaibacteriumterrae
KJ179841 1459bp Tahibactercaeni
AM981201 1463bp Tahibacteraquaticus
AB245362 1473bp Dokdonellaginsengisoli
FJ455531 1500bp Dokdonellaimmobilis
AY544768 1433bp Aquimonasvoraii
GQ281768 1522bp Chiayiivirgaflava
KM400682 1469bp Oleiagrimonassoli
EU912469 1429bp Rhodanobacterglycinis
EU682683 1531bp Frateuriaterrea
AB091194 1535bp Frateuriaaurantia
AY884571 1486bp Dyellakoreensis
AM939778 1421bp Dyellamarensis
AB245367 1491bp Dyellaginsengisoli
EU604272 1420bp Dyellasoli
EF397574 1422bp Dyellathiooxydans
DQ181549 1478bp Luteibacteryeojuensis
FM212561 1429bp Luteibacteranthropi
AJ580498 1474bp Luteibacterrhizovicinus
AJ311653 1497bp Fulvimonassoli
KJ490635 1484bp Fulvimonasyonginensis
EU604273 1423bp Dyellaterrae
HQ236051 1425bp Dyellajiangningensis
JX566987 1454bp Dyellakyungheensis
AB110498 1485bp Dyellajaponica
JF831047 1423bp Legionellacardiaca
Z32636 1473bp Legionellabrunensis
Z32638 1454bp Legionellaerythra
U58339 1435bp Marinobacteriumgeorgiense
AY563030 1400bp Marinobacteriumhalophilum
AB021367 1517bp Marinobacteriumstanieri
EU573966 1464bp Marinobacteriumsediminicola
GQ131677 1418bp Neptuniibacterhalophilus
AB006760 1461bp Oceanospirillumbeijerinckii
GQ426888 1476bp Thalassolituusmarinus
Y12579 1494bp Alcanivoraxborkumensis
KJ195687 1417bp Salinispirillummarinum
AJ295154 1437bp Oleiphilusmessinensis
AF288157 1529bp Marinobacterlutaoensis
FJ425903 1492bp Marinobacterzhanjiangensis
EU047505 1537bp Marinobacterlacisalsi
FJ984869 1410bp Marinobacterdaqiaonensis
AJ609271 1507bp Marinobacterbryozoorum
EF157832 1387bp Marinobactersegnicrescens
EF486354 1496bp Marinobactersalicampi
DQ414419 1501bp Marinobactergudaonensis
AY258110 1485bp Marinobacteralgicola
CP007152 1542bp Marinobactersalarius
EF028328 1487bp Marinobactersalsuginis
AY517632 1495bp Marinobacterflavimaris
AY241552 1531bp Marinobacteradhaerens
CP007151 1543bp Marinobactersimilis
AJ609270 1512bp Marinobactersediminum
AY147906 1473bp Marinobacterlipolyticus
AM503093 1488bp Marinobacterguineae
EF660754 1436bp Marinobactergoseongensis
AJ704395 1487bp Marinobactermaritimus
DQ060402 1506bp Marinobacterpsychrophilus
FJ196022 1506bp Marinobacterantarcticus
EU293412 1429bp Marinobactermobilis
EU293413 1433bp Marinobacterzhejiangensis
HQ203044 1468bp Marinobacterxestospongiae
EU164778 1435bp Marinobacterszutsaonensis
DQ458821 1436bp Marinobacterpelagius
DQ325514 1463bp Marinobacterkoreensis

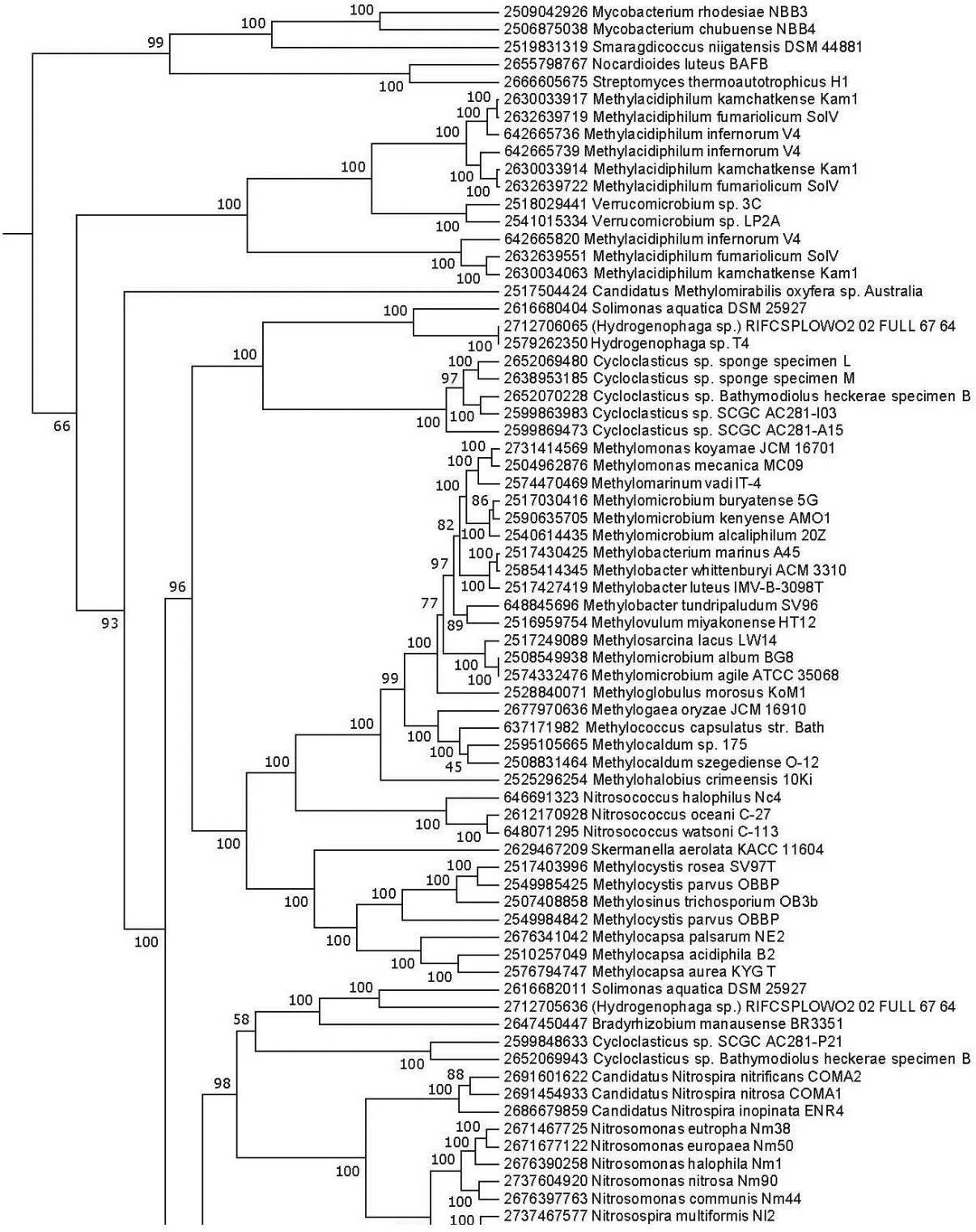

100
EU496088 1383bp Marinobactersantoriniensis
100
FJ897726 1464bp Marinobacteroulmenensis
HQ433441 1465bp Marinobacterpersicus
99
X67022 1483bp Marinobacterhydrocarbonoclasticus
100
AY517633 1496bp Marinobacterdaepoensis
99
DQ235263 1505bp Marinobactervinifirmus
100
AY180101 1536bp Marinobacterexcellens
100
98
AF479689 1496bp Marinobacterlitoralis

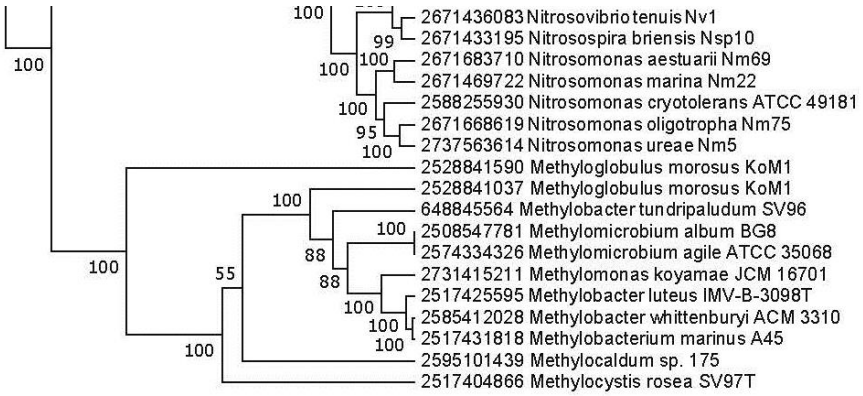
Supplementary Figure (A7.)2: 16S based phylogeny of region surrounding alphaproteobacterial methanotrophs. Taxa were obtained using sequential BLAST analyses of the SILVA LTP database centred on 16S ribosome units of two selected methanotrophs characterising each of the major alphaproteobacterial subgroups. The 100 closest results from each BLAST analysis were then merged to a single dataset and aligned using the Q-INS-i methodology in MAFFT. Phylogeny of the resulting 122 sequences was inferred using Bayesian Markov chain Monte Carlo analysis using the GTR+G+I model and an uncorrelated relaxed molecular clock in BEAST. Posterior probabilities have been included for all nodes (as percentage values), as well as accession numbers prior to species name.



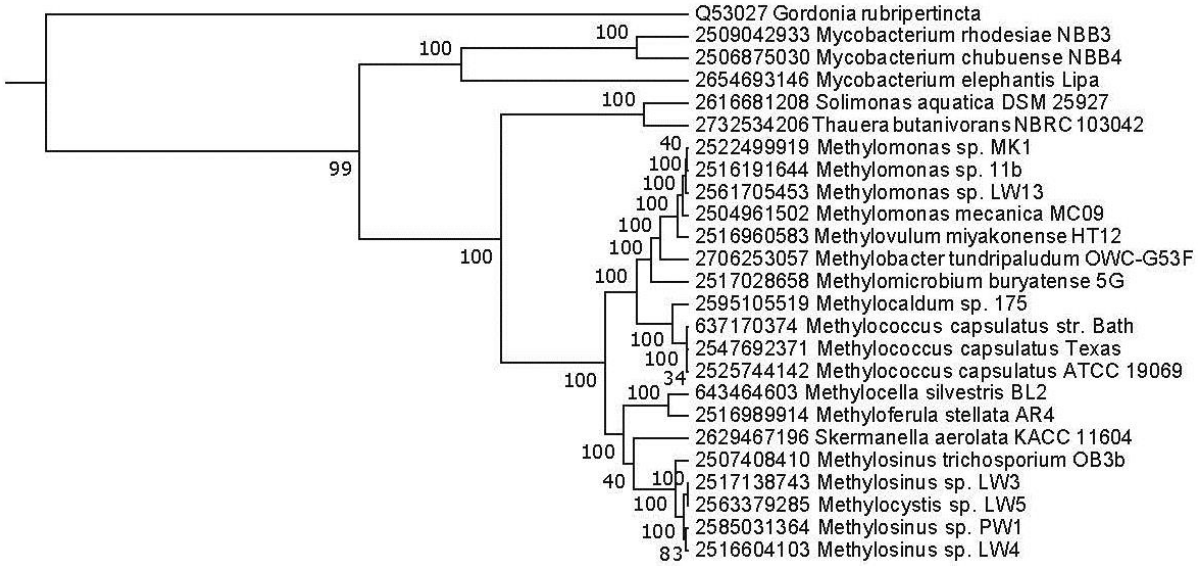

AB017203 1433bp Prosthecomicrobiumpneumaticum
AB166877 1406bp Starkeyakoreensis
FJ998414 1385bp Amorphusorientalis
EF433460 1345bp Devosiaalbogilva
FJ560749 1421bp Vasilyevaeamishustinii
DQ883810 1360bp Aureimonasureilytica
DQ372921 1366bp Aureimonasaltamirensis
KF683074 1367bp Methylobrevispamukkalensis
KC567245 1426bp Hartmannibacterdiazotrophicus
EF465533 1436bp Pseudoxanthobactersoli
JQ348904 1455bp Pseudoxanthobacterliyangensis
GQ272328 1481bp Breoghaniacorrubedonensis
Y14305 1438bp Hyphomicrobiumzavarzinii
JX131369 1450bp Hyphomicrobiumnitrativorans
AM180706 1415bp Rhodobiumgokarnense
D30792 1408bp Rhodobiumorientis
GQ221764 1443bp Bauldialitoralis
FJ560750 1442bp Bauldiaconsociata
EU564843 1480bp Chelativoransintermedius
AB563785 1402bp Chelativoranscomposti
AJ011760 1467bp Aminobacteraganoensis
AJ011761 1466bp Aminobacterniigataensis
AF034798 1407bp Aminobacterciceronei
DQ411930 1461bp Ensifermexicanus
X68388 1437bp Ensiferterangae
X68390 1437bp Ensifersaheli
X67231 1437bp Ensiferfredii
AF506513 1437bp Ensiferamericanus
AM181747 1426bp Ensiferkostiensis
AM263420 1387bp Pseudochrobactrumkiredjianiae
AJ968695 1438bp Phyllobacteriumleguminum
AY636000 1412bp Phyllobacteriumcatacumbae
AY785325 1401bp Phyllobacteriumifriqiyense
AY786080 1478bp Phyllobacteriumtrifolii
KC577468 1562bp Phyllobacteriumloti
JN848778 1483bp Phyllobacteriumendophyticum
AY785315 1401bp Phyllobacteriummyrsinacearum
AJ864462 1354bp Mesorhizobiumthiogangeticum
GU447302 1448bp Nitratireductorindicus
EU514525 1356bp Mesorhizobiumsangaii
AF508207 1415bp Mesorhizobiumseptentrionale
GQ847899 1449bp Mesorhizobiumhawassense
Y14158 1437bp Mesorhizobiumplurifarium
EU399698 1348bp Mesorhizobiumsilamurunense
GQ847890 1469bp Mesorhizobiumshonense
AF041442 1473bp Mesorhizobiumamorphae
GQ847896 1468bp Mesorhizobiumabyssinicae
D13431 1443bp Mesorhizobiumhuakuii
AY601515 1441bp Mesorhizobiumopportunistum
AM491621 1477bp Mesorhizobiumtamadayense
AF508208 1403bp Mesorhizobiumtemperatum
AM181745 1426bp Mesorhizobiummediterraneum
EU849582 1356bp Mesorhizobiumrobiniae
HQ316710 1474bp Mesorhizobiummuleiense
EF035058 1377bp Mesorhizobiumtarimense
AF041447 1476bp Mesorhizobiumtianshanense
EF149003 1365bp Mesorhizobiumcaraganae
AM930381 1399bp Mesorhizobiummetallidurans
EF035064 1353bp Mesorhizobiumgobiense

Supplementary Figure (A7.)3: Phylogenetic relationship of the CuMMO family based on pmoB and homologous sequences. A curated dataset of 88 translated gene sequences were selected to represent the full range of diversity of bacterial pmoB genes and their homologues identified in the Joint Genome Institute IMG/ER database. Alignment was conducted in MAFFT using the L-INS-i method, with subsequent Bayesian Markov chain Monte Carlo analysis using the WAG+G+I model and uncorrelated relaxed molecular clock in BEAST. The consensus tree with nodal posterior probabilities from this analysis is shown, as well as accession numbers from the JGI database for each gene.




2671436083 Nitrosovibrio tenuis Nv1
2671433195 Nitrosospira briensis Nsp10
2671683710 Nitrosomonas aestuarii Nm69
2671469722 Nitrosomonas marina Nm22
2588255930 Nitrosomonas cryotolerans ATCC 49181
2671668619 Nitrosomonas oligotropha Nm75
2737563614 Nitrosomonas ureae Nm5
2528841590 Methyloglobulus morosus KoM1
2528841037 Methyloglobulus morosus KoM1
648845564 Methylobacter tundripaludum SV96
2508547781 Methylomicrobium album BG8
2574334326 Methylomicrobium agile ATCC 35068
2731415211 Methylomonas koyamae JCM 16701
2517425595 Methylobacter luteus IMV-B-3098T
2585412028 Methylobacter whittenburyi ACM 3310
2517431818 Methylobacterium marinus A45
2595101439 Methylocaldum sp. 175
2517404866 Methylocystis rosea SV97T

Supplementary Figure (A7.)4: Subsection of BMM phylogeny displaying relationship of sMMO sequences and their closest relatives. Analysis was conducted using sequences identified through pBLAST of the Joint Genome Institute IMG/ER database, centred on the protein translation of mmoX. Sequences were aligned with MAFFT using the L-INS-I method, with initial larger phylogenetic analysis pruned to the region including and immediately adjacent to sMMO. Detailed phylogeny was constructed using Bayesian Markov chain Monte Carlo methodology, WAG+G+I model and an uncorrelated relaxed molecular clock implemented in BEAST. The consensus tree with nodal posterior probabilities from this analysis is shown, as well as accession numbers from the JGI database for each gene.



A7.1.2. Supplementary phylogenetic trees

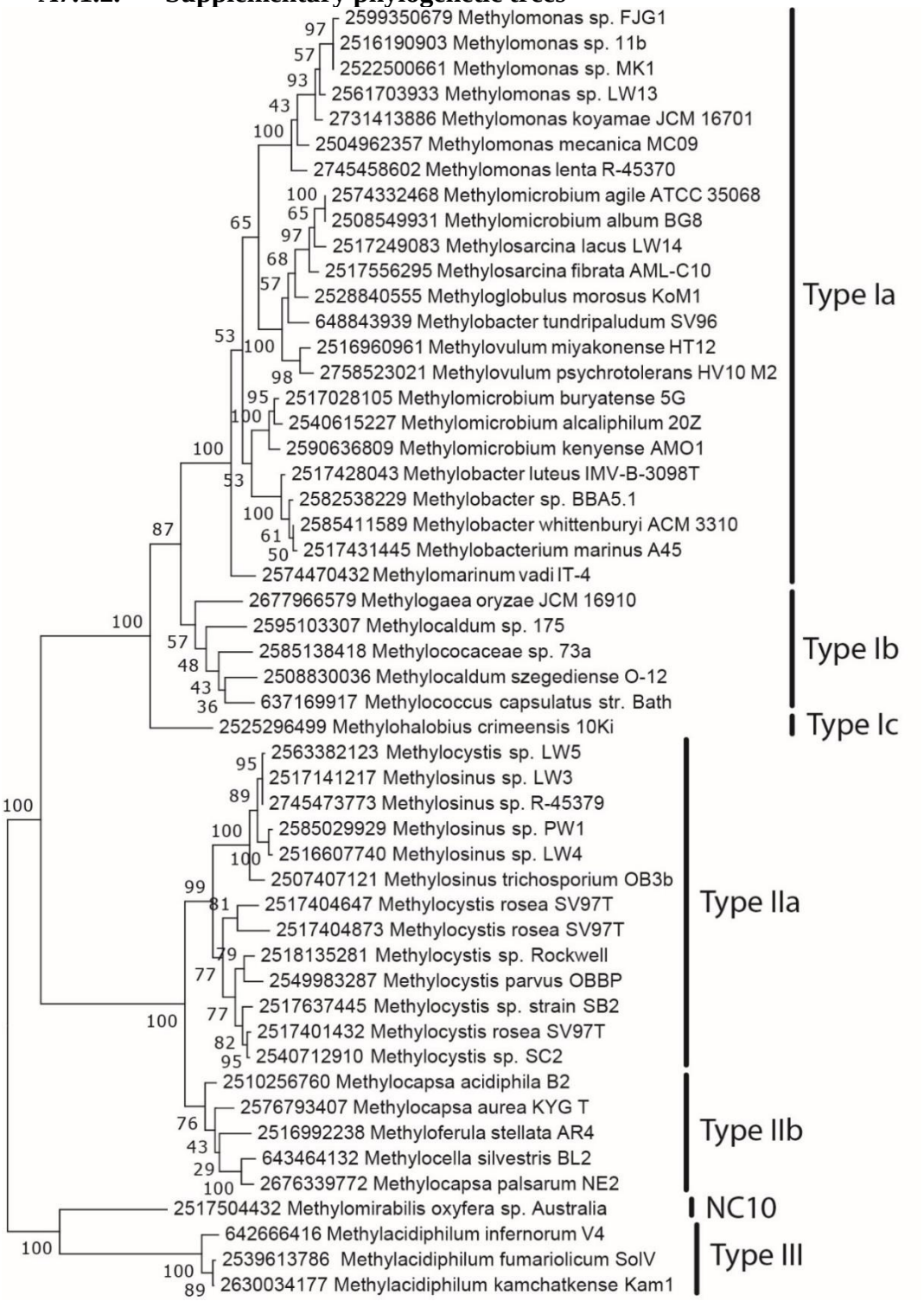
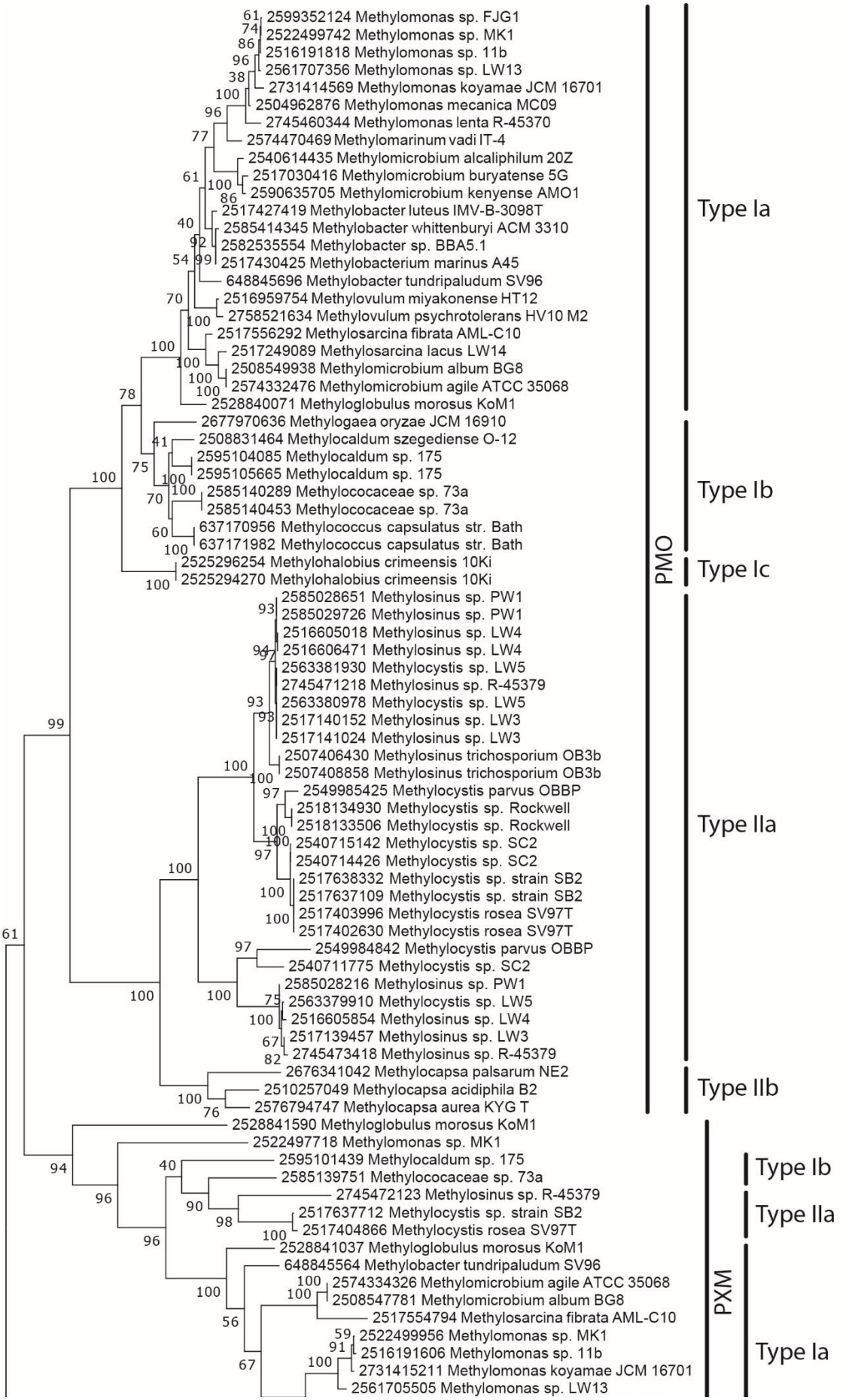


Figure A7.5: Phylogenetic analysis of methanotrophic species based on conventional GroEL sequences. Sequences of each methanotrophic strain used in genome analysis are included, with JGI unique gene identifier for GroEL gene preceding species name. Methanotrophic subgroups are designated on the right. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.


2599352124 Methylomonas sp. FJG1
2522499742 Methylomonas sp. MK1
2516191818 Methylomonas sp. 11b
2561707356 Methylomonas sp. LW13
2731414569 Methylomonas koyamae JCM 16701
2504962876 Methylomonas mecanica MC09
2745460344 Methylomonas lenta R-45370
2574470469 Methylomarinum vadi IT-4
2540614435 Methylomicrobium alcaliphilum 20Z
2517030416 Methylomicrobium buryatense 5G
2590635705 Methylomicrobium kenyense AMO1
2517427419 Methylobacter luteus IMV-B-3098T
2585414345 Methylobacter whittenburyi ACM 3310
2582535554 Methylobacter sp. BBA5.1
2517430425 Methylobacterium marinus A45
648845696 Methylobacter tundripaludum SV96
2516959754 Methylovulum miyakonense HT12
2758521634 Methylovulum psychrotolerans HV10 M2
2517556292 Methylosarcina fibrata AML-C10
2517249089 Methylosarcina lacus LW14
2508549938 Methylomicrobium album BG8
2574332476 Methylomicrobium agile ATCC 35068
2528840071 Methyloglobulus morosus KoM1
2677970636 Methylogaea oryzae JCM 16910
2508831464 Methylocaldum szegediense O-12
2595104085 Methylocaldum sp. 175
2595105665 Methylocaldum sp. 175
2585140289 Methylococcaceae sp. 73a
2585140453 Methylococcaceae sp. 73a
637170956 Methylococcus capsulatus str. Bath
637171982 Methylococcus capsulatus str. Bath
2525296254 Methylohalobius crimeensis 10Ki
2525294270 Methylohalobius crimeensis 10Ki
2585028651 Methylosinus sp. PW1
2585029726 Methylosinus sp. PW1
2516605018 Methylosinus sp. LW4
2516606471 Methylosinus sp. LW4
2563381930 Methylocystis sp. LW5
2745471218 Methylosinus sp. R-45379
2563380978 Methylocystis sp. LW5
2517140152 Methylosinus sp. LW3
2517141024 Methylosinus sp. LW3
2507406430 Methylosinus trichosporium OB3b
2507408858 Methylosinus trichosporium OB3b
2549985425 Methylocystis parvus OBBP
2518134930 Methylocystis sp. Rockwell
2518133506 Methylocystis sp. Rockwell
2540715142 Methylocystis sp. SC2
2540714426 Methylocystis sp. SC2
2517638332 Methylocystis sp. strain SB2
2517637109 Methylocystis sp. strain SB2
2517403996 Methylocystis rosea SV97T
2517402630 Methylocystis rosea SV97T
2549984842 Methylocystis parvus OBBP
2540711775 Methylocystis sp. SC2
2585028216 Methylosinus sp. PW1
2563379910 Methylocystis sp. LW5
2516605854 Methylosinus sp. LW4
2517139457 Methylosinus sp. LW3
2745473418 Methylosinus sp. R-45379
2676341042 Methylocapsa palsarum NE2
2510257049 Methylocapsa acidiphila B2
2576794747 Methylocapsa aurea KYG T
2528841590 Methyloglobulus morosus KoM1
2522497718 Methylomonas sp. MK1
2595101439 Methylocaldum sp. 175
2585139751 Methylococcaceae sp. 73a
2745472123 Methylosinus sp. R-45379
2517637712 Methylocystis sp. strain SB2
2517404866 Methylocystis rosea SV97T
2528841037 Methyloglobulus morosus KoM1
648845564 Methylobacter tundripaludum SV96
2574334326 Methylomicrobium agile ATCC 35068
2508547781 Methylomicrobium album BG8
2517554794 Methylosarcina fibrata AML-C10
2522499956 Methylomonas sp. MK1
2516191606 Methylomonas sp. 11b
2731415211 Methylomonas koyamae JCM 16701
2561705505 Methylomonas sp. LW13
Type Ia
Type Ib
Type Ic
Type IIa
Type IIb
PMO
Type Ib
Type IIa
Type Ia
PXM

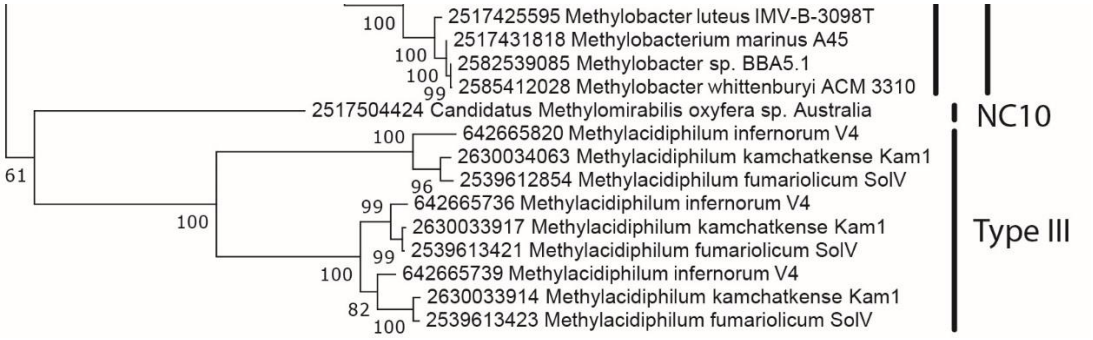


Figure A7.6: Detailed phylogenetic analysis of inferred CuMMO sequences in methanotrophs. Analysis based on the PmoB component of CuMMO and its homologues. In all cases, JGI unique gene identifier is listed prior to the species from which it was obtained. Methanotrophic subgroups to which the species containing the respective gene belongs is indicated on far right, with the designation of the respective gene as either PMO (particulate methane monooxygenase) or PXM (uncharacterised copper membrane monooxygenase) preceding where relevant. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.


2517430322 Methylobacterium marinus A45
2585414428 Methylobacter whittenburyi ACM 3310
2582538519 Methylobacter sp. BBA5.1
2517426653 Methylobacter luteus IMV-B-3098T
2517030652 Methylomicrobium buryatense 5G
2540617384 Methylomicrobium alcaliphilum 20Z
2590634458 Methylomicrobium kenyense AMO1
2574469073 Methylomarinum vadi IT-4
2528838472 Methyloglobulus morosus KoM1
2758520611 Methylovulum psychrotolerans HV10 M2
2516960807 Methylovulum miyakonense HT12
2745457400 Methylomonas lenta R-45370
2504963471 Methylomonas mecanica MC09
2731414952 Methylomonas koyamae JCM 16701
2516189055 Methylomonas sp. 11b
2522497092 Methylomonas sp. MK1
2561704090 Methylomonas sp. LW13
2517246851 Methylosarcina lacus LW14
2508548816 Methylomicrobium album BG8
2574335385 Methylomicrobium agile ATCC 35068
648843491 Methylobacter tundripaludum SV96
2517556682 Methylosarcina fibrata AML-C10
2508832546 Methylocaldum szegediense O-12
2595105005 Methylocaldum sp. 175
2585139491 Methylococaceae sp. 73a
637169987 Methylococcus capsulatus str. Bath
2576792736 Methylocapsa aurea KYG T
643463801 Methylocella silvestris BL2
2516607567 Methylosinus sp. LW4
2585030104 Methylosinus sp. PW1
2517141394 Methylosinus sp. LW3
2745472053 Methylosinus sp. R-45379
2563382300 Methylocystis sp. LW5
2549984483 Methylocystis parvus OBBP
2518136069 Methylocystis sp. Rockwell
2507408026 Methylosinus trichosporium OB3b
2676341710 Methylocapsa palsarum NE2
2540714167 Methylocystis sp. SC2
2517402041 Methylocystis rosea SV97T
2517635397 Methylocystis sp. strain SB2
2510253721 Methylocapsa acidiphila B2
2516991551 Methyloferula stellata AR4
2525294493 Methylohalobius crimeensis 10Ki
2517502116 Candidatus Methylomirabilis oxyfera sp. Australia
2522499177 Methylomonas sp. MK1
2599350338 Methylomonas sp. FJG1
2516192404 Methylomonas sp. 11b
2561706408 Methylomonas sp. LW13
2731415123 Methylomonas koyamae JCM 16701
2504961632 Methylomonas mecanica MC09
2745457963 Methylomonas lenta R-45370
2528841358 Methyloglobulus morosus KoM1
2758521763 Methylovulum psychrotolerans HV10 M2
2516961404 Methylovulum miyakonense HT12
648841408 Methylobacter tundripaludum SV96
2517247151 Methylosarcina lacus LW14
2517555720 Methylosarcina fibrata AML-C10
2574334287 Methylomicrobium agile ATCC 35068
2508547743 Methylomicrobium album BG8
2517424835 Methylobacter luteus IMV-B-3098T
2582536128 Methylobacter sp. BBA5.1
2585414654 Methylobacter whittenburyi ACM 3310
2517430095 Methylobacterium marinus A45
2574470365 Methylomarinum vadi IT-4
2590634523 Methylomicrobium kenyense AMO1
2540617435 Methylomicrobium alcaliphilum 20Z
2517030556 Methylomicrobium buryatense 5G
2525295022 Methylohalobius crimeensis 10Ki
2508833647 Methylocaldum szegediense O-12
2677968181 Methylogaea oryzae JCM 16910
637169521 Methylococcus capsulatus str. Bath
mxaF

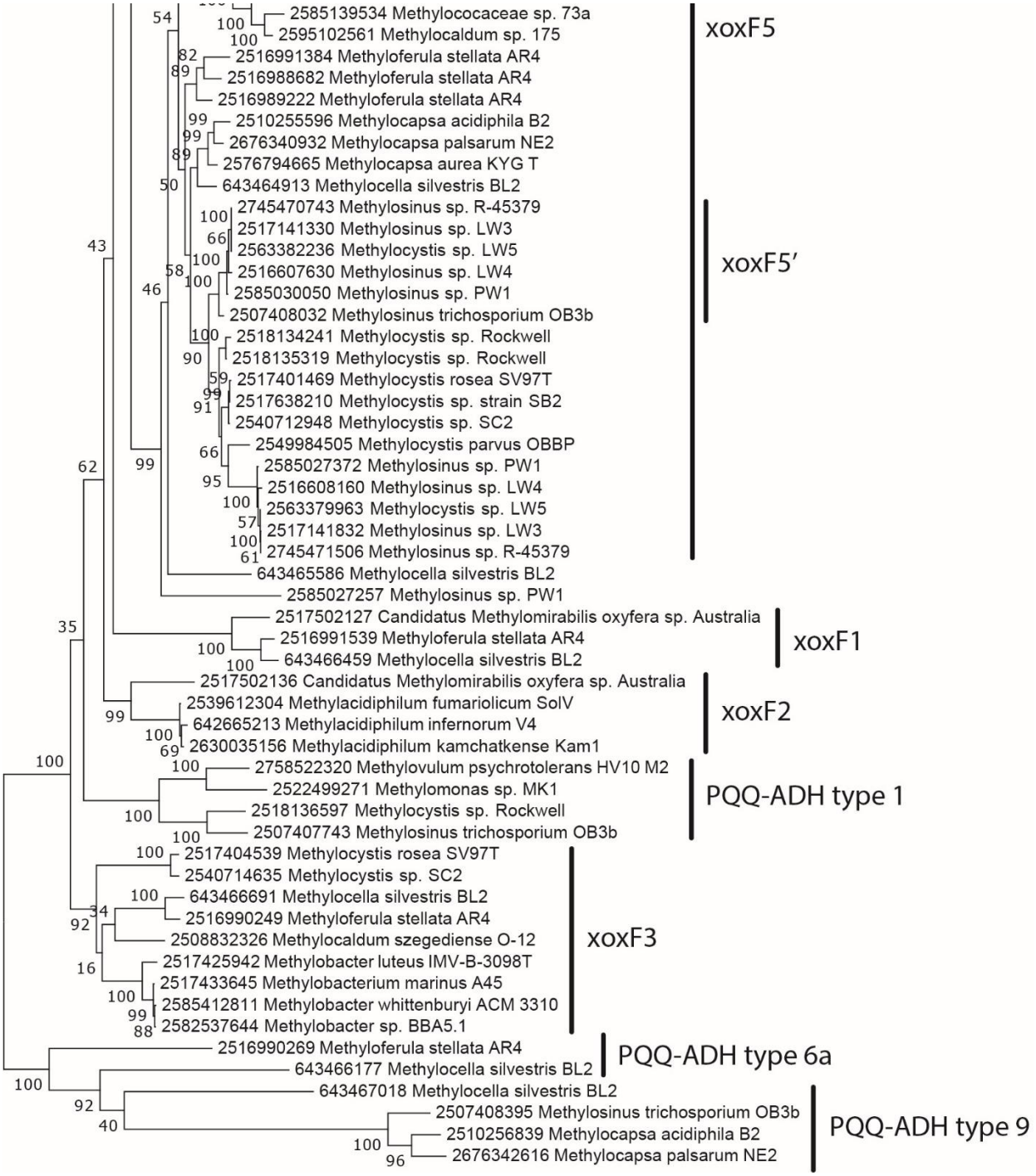
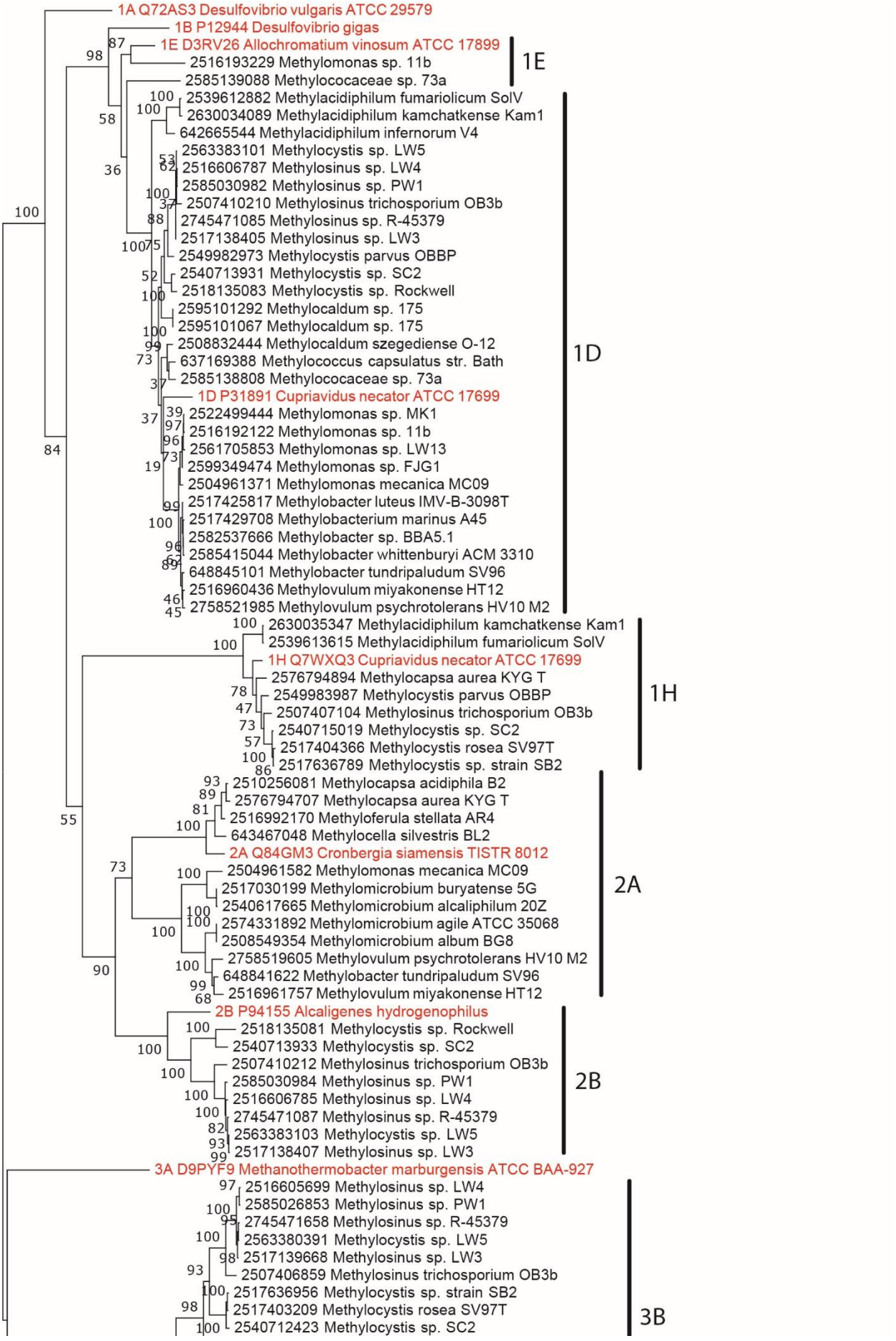


Figure A7.7: Detailed phylogenetic analysis of inferred PQQ containing alcohol dehydrogenase sequences in methanotrophs. In all cases, JGI unique gene identifier is listed prior to the species from which it was obtained. PQQ grouping as designated by Keltjens et al. (2014) appear on far right. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.


1A Q72AS3 Desulfovibrio vulgaris ATCC 29579
1B P12944 Desulfovibrio gigas
1E D3RV26 Allochromatium vinosum ATCC 17899
2516193229 Methylomonas sp. 11b
2585139088 Methylococcaceae sp. 73a
1E
2539612882 Methylacidiphilum fumariolicum SolV
2630034089 Methylacidiphilum kamchatkense Kam1
642665544 Methylacidiphilum infernorum V4
2563383101 Methylocystis sp. LW5
2516606787 Methylosinus sp. LW4
2585030982 Methylosinus sp. PW1
2507410210 Methylosinus trichosporium OB3b
2745471085 Methylosinus sp. R-45379
2517138405 Methylosinus sp. LW3
2549982973 Methylocystis parvus OBBP
2540713931 Methylocystis sp. SC2
2518135083 Methylocystis sp. Rockwell
2595101292 Methylocaldum sp. 175
2595101067 Methylocaldum sp. 175
2508832444 Methylocaldum szegediense O-12
637169388 Methylococcus capsulatus str. Bath
2585138808 Methylococcaceae sp. 73a
1D P31891 Cupriavidus necator ATCC 17699
2522499444 Methylomonas sp. MK1
2516192122 Methylomonas sp. 11b
2561705853 Methylomonas sp. LW13
2599349474 Methylomonas sp. FJG1
2504961371 Methylomonas mecanica MC09
2517425817 Methylobacter luteus IMV-B-3098T
2517429708 Methylobacterium marinus A45
2582537666 Methylobacter sp. BBA5.1
2585415044 Methylobacter whittenburyi ACM 3310
648845101 Methylobacter tundripaludum SV96
2516960436 Methylovulum miyakonense HT12
2758521985 Methylovulum psychrotolerans HV10 M2
1D
2630035347 Methylacidiphilum kamchatkense Kam1
2539613615 Methylacidiphilum fumariolicum SolV
1H Q7WXQ3 Cupriavidus necator ATCC 17699
2576794894 Methylocapsa aurea KYG T
2549983987 Methylocystis parvus OBBP
2507407104 Methylosinus trichosporium OB3b
2540715019 Methylocystis sp. SC2
2517404366 Methylocystis rosea SV97T
2517636789 Methylocystis sp. strain SB2
1H
2510256081 Methylocapsa acidiphila B2
2576794707 Methylocapsa aurea KYG T
2516992170 Methyloferula stellata AR4
643467048 Methylocella silvestris BL2
2A Q84GM3 Cronbergia siamensis TISTR 8012
2504961582 Methylomonas mecanica MC09
2517030199 Methylomicrobium buryatense 5G
2540617665 Methylomicrobium alcaliphilum 20Z
2574331892 Methylomicrobium agile ATCC 35068
2508549354 Methylomicrobium album BG8
2758519605 Methylovulum psychrotolerans HV10 M2
648841622 Methylobacter tundripaludum SV96
2516961757 Methylovulum miyakonense HT12
2A
2B P94155 Alcaligenes hydrogenophilus
2518135081 Methylocystis sp. Rockwell
2540713933 Methylocystis sp. SC2
2507410212 Methylosinus trichosporium OB3b
2585030984 Methylosinus sp. PW1
2516606785 Methylosinus sp. LW4
2745471087 Methylosinus sp. R-45379
2563383103 Methylocystis sp. LW5
2517138407 Methylosinus sp. LW3
2B
3A D9PYF9 Methanothermobacter marburgensis ATCC BAA-927
2516605699 Methylosinus sp. LW4
2585026853 Methylosinus sp. PW1
2745471658 Methylosinus sp. R-45379
2563380391 Methylocystis sp. LW5
2517139668 Methylosinus sp. LW3
2507406859 Methylosinus trichosporium OB3b
2517636956 Methylocystis sp. strain SB2
2517403209 Methylocystis rosea SV97T
2540712423 Methylocystis sp. SC2
3B

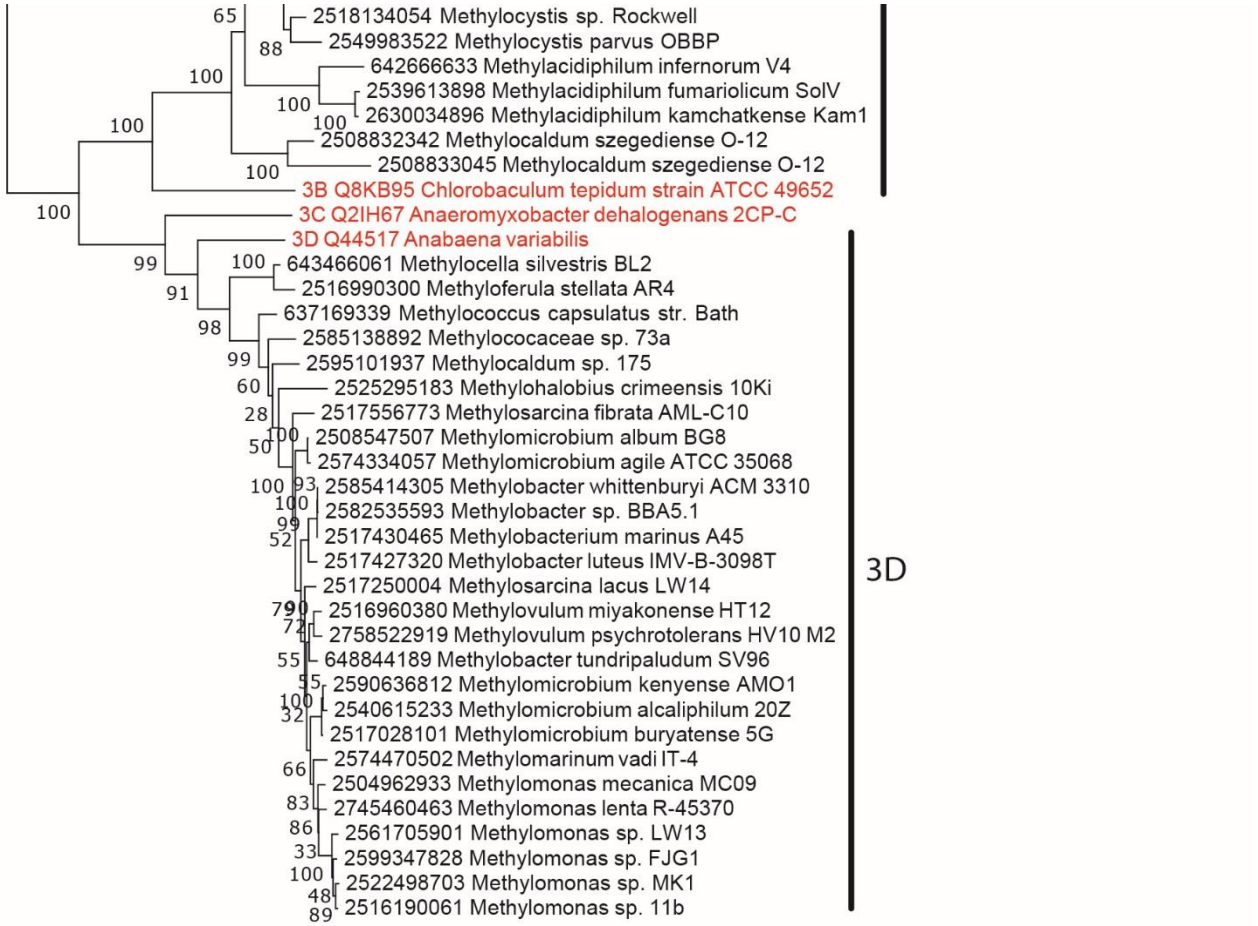
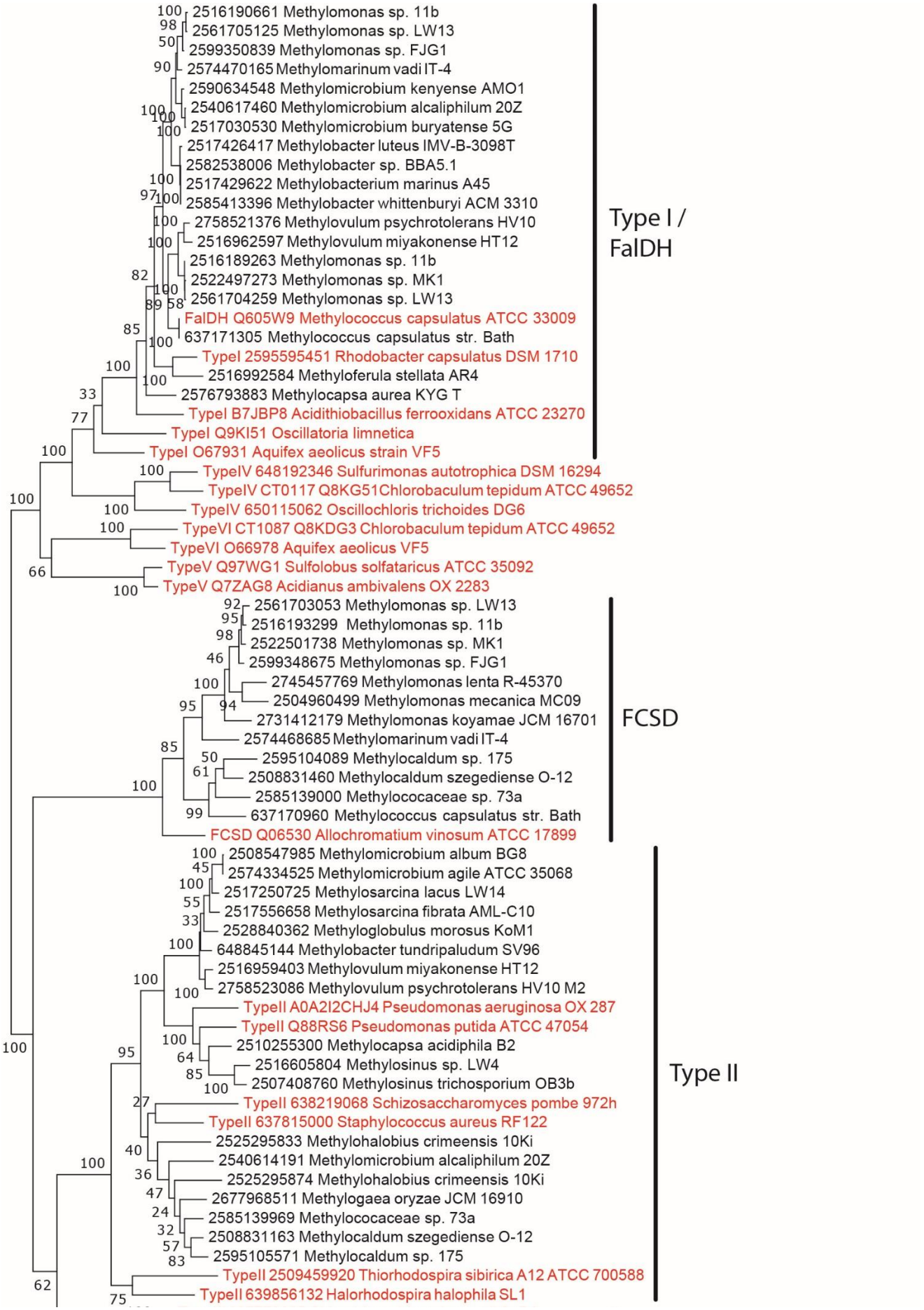


Figure A7.8: Detailed phylogenetic analysis of inferred NiFe hydrogenase sequences in methanotrophs. For all methanotrophic entries, JGI unique gene identifier is listed prior to the species from which it was obtained. For non-methanotrophic references sequences highlighted in red, hydrogenase group designation and accession numbers precede each entry. Hydrogenase group designation of methanotrophic genes appears on far right where appropriate. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.


2516190661 Methylomonas sp. 11b
2561705125 Methylomonas sp. LW13
2599350839 Methylomonas sp. FJG1
2574470165 Methylomarinum vadi IT-4
2590634548 Methylomicrobium kenyense AMO1
2540617460 Methylomicrobium alcaliphilum 20Z
2517030530 Methylomicrobium buryatense 5G
2517426417 Methylobacter luteus IMV-B-3098T
2582538006 Methylobacter sp. BBA5.1
2517429622 Methylobacterium marinus A45
2585413396 Methylobacter whittenburyi ACM 3310
2758521376 Methylovulum psychrotolerans HV10
2516962597 Methylovulum miyakonense HT12
2516189263 Methylomonas sp. 11b
2522497273 Methylomonas sp. MK1
2561704259 Methylomonas sp. LW13
FalDH Q605W9 Methylococcus capsulatus ATCC 33009
637171305 Methylococcus capsulatus str. Bath
TypeI 2595595451 Rhodobacter capsulatus DSM 1710
2516992584 Methyloferula stellata AR4
2576793883 Methylocapsa aurea KYG T
TypeI B7JBP8 Acidithiobacillus ferrooxidans ATCC 23270
TypeI Q9KI51 Oscillatoria limnetica
TypeI O67931 Aquifex aeolicus strain VF5
TypeIV 648192346 Sulfurimonas autotrophica DSM 16294
TypeIV CT0117 Q8KG51Chlorobaculum tepidum ATCC 49652
TypeIV 650115062 Oscillochloris trichoides DG6
TypeVI CT1087 Q8KDG3 Chlorobaculum tepidum ATCC 49652
TypeVI O66978 Aquifex aeolicus VF5
TypeV Q97WG1 Sulfolobus solfataricus ATCC 35092
TypeV Q7ZAG8 Acidianus ambivalens OX 2283
2561703053 Methylomonas sp. LW13
2516193299 Methylomonas sp. 11b
2522501738 Methylomonas sp. MK1
2599348675 Methylomonas sp. FJG1
2745457769 Methylomonas lenta R-45370
2504960499 Methylomonas mecanica MC09
2731412179 Methylomonas koyamae JCM 16701
2574468685 Methylomarinum vadi IT-4
2595104089 Methylocaldum sp. 175
2508831460 Methylocaldum szegediense O-12
2585139000 Methylococcaceae sp. 73a
637170960 Methylococcus capsulatus str. Bath
FCSD Q06530 Allochromatium vinosum ATCC 17899
2508547985 Methylomicrobium album BG8
2574334525 Methylomicrobium agile ATCC 35068
2517250725 Methylosarcina lacus LW14
2517556658 Methylosarcina fibrata AML-C10
2528840362 Methyloglobulus morosus KoM1
648845144 Methylobacter tundripaludum SV96
2516959403 Methylovulum miyakonense HT12
2758523086 Methylovulum psychrotolerans HV10 M2
TypeII A0A2I2CHJ4 Pseudomonas aeruginosa OX 287
TypeII Q88RS6 Pseudomonas putida ATCC 47054
2510255300 Methylocapsa acidiphila B2
2516605804 Methylosinus sp. LW4
2507408760 Methylosinus trichosporium OB3b
TypeII 638219068 Schizosaccharomyces pombe 972h
TypeII 637815000 Staphylococcus aureus RF122
2525295833 Methylohalobius crimeensis 10Ki
2540614191 Methylomicrobium alcaliphilum 20Z
2525295874 Methylohalobius crimeensis 10Ki
2677968511 Methylogaea oryzae JCM 16910
2585139969 Methylococcaceae sp. 73a
2508831163 Methylocaldum szegediense O-12
2595105571 Methylocaldum sp. 175
TypeII 2509459920 Thiorhodospira sibirica A12 ATCC 700588
TypeII 639856132 Halorhodospira halophila SL1
Type I / FalDH
FCSD
Type II

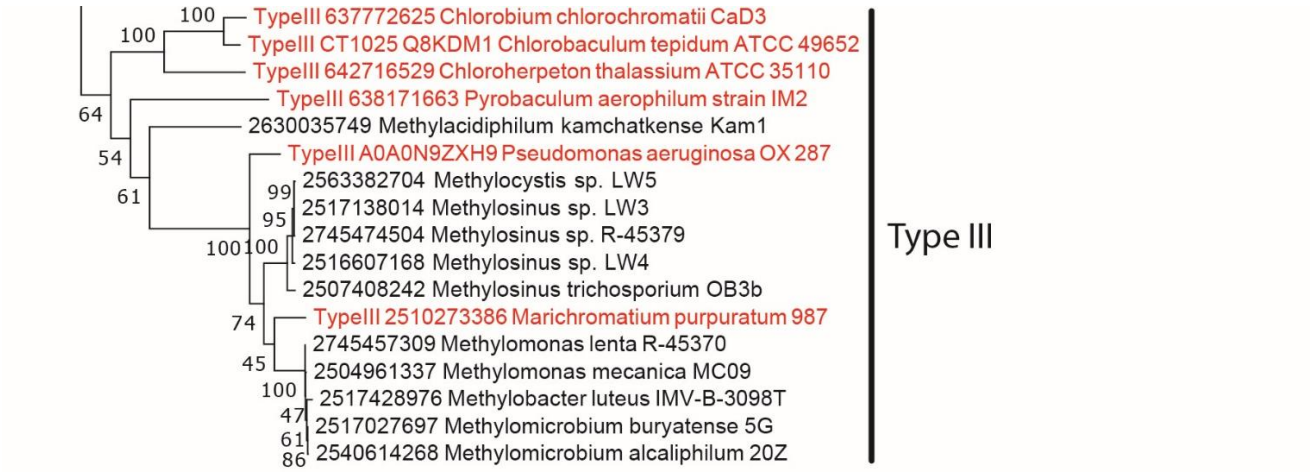


Figure A7.9: Detailed phylogenetic analysis of inferred sulphide reductase sequences in methanotrophs. For all methanotrophic entries, JGI unique gene identifier is listed prior to the species from which it was obtained. For non-methanotrophic references sequences highlighted in red, the first entry in each tag denotes the sulphur quinone reductases functional group from which the member belongs to, or designates the member as a flavocytochrome C – quinone dehydrogenase (FCSD). The second entry is either the protein accession number for alphanumeric entries, or JGI unique gene identifier for numeric only entries. Sulphide reductase group designation of methanotrophic genes appears on far right where appropriate. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.

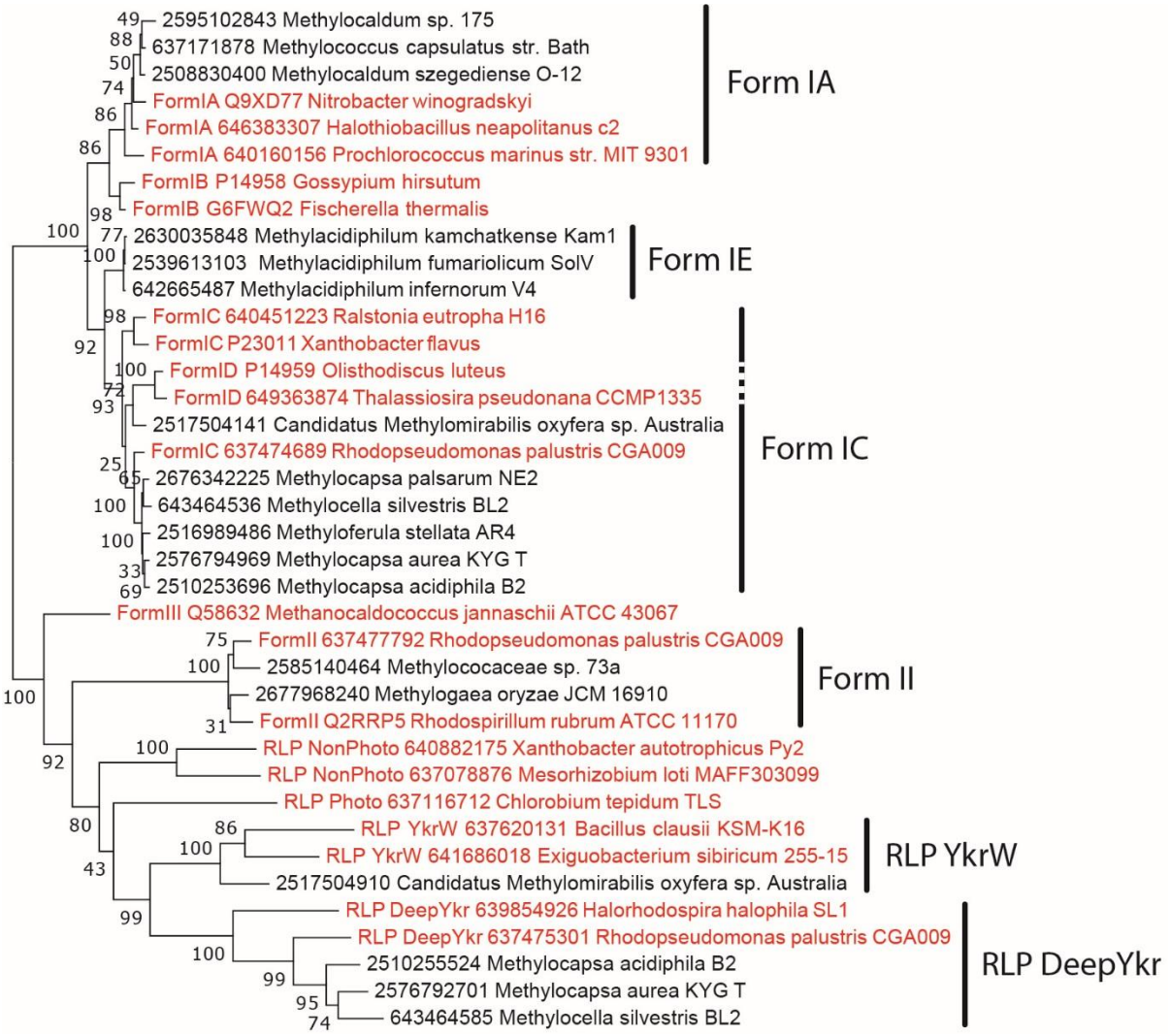


Figure A7.10: Detailed phylogenetic analysis of inferred RuBisCO and RuBisCO-like sequences in methanotrophs. For all methanotrophic entries, JGI unique gene identifier is listed prior to the species from which it was obtained. For non-methanotrophic references sequences highlighted in red, RuBisCO group designation and accession numbers precede each entry. RuBisCO group designation of methanotrophic genes appears on far right where appropriate, with broken line representing the appearance of Form ID RuBisCO genes inside the IC cluster. Sequences were aligned in MAFFT v7, with phylogeny constructed in RAxML using WAG+G+I methodology. Bootstrap values are present at each node.

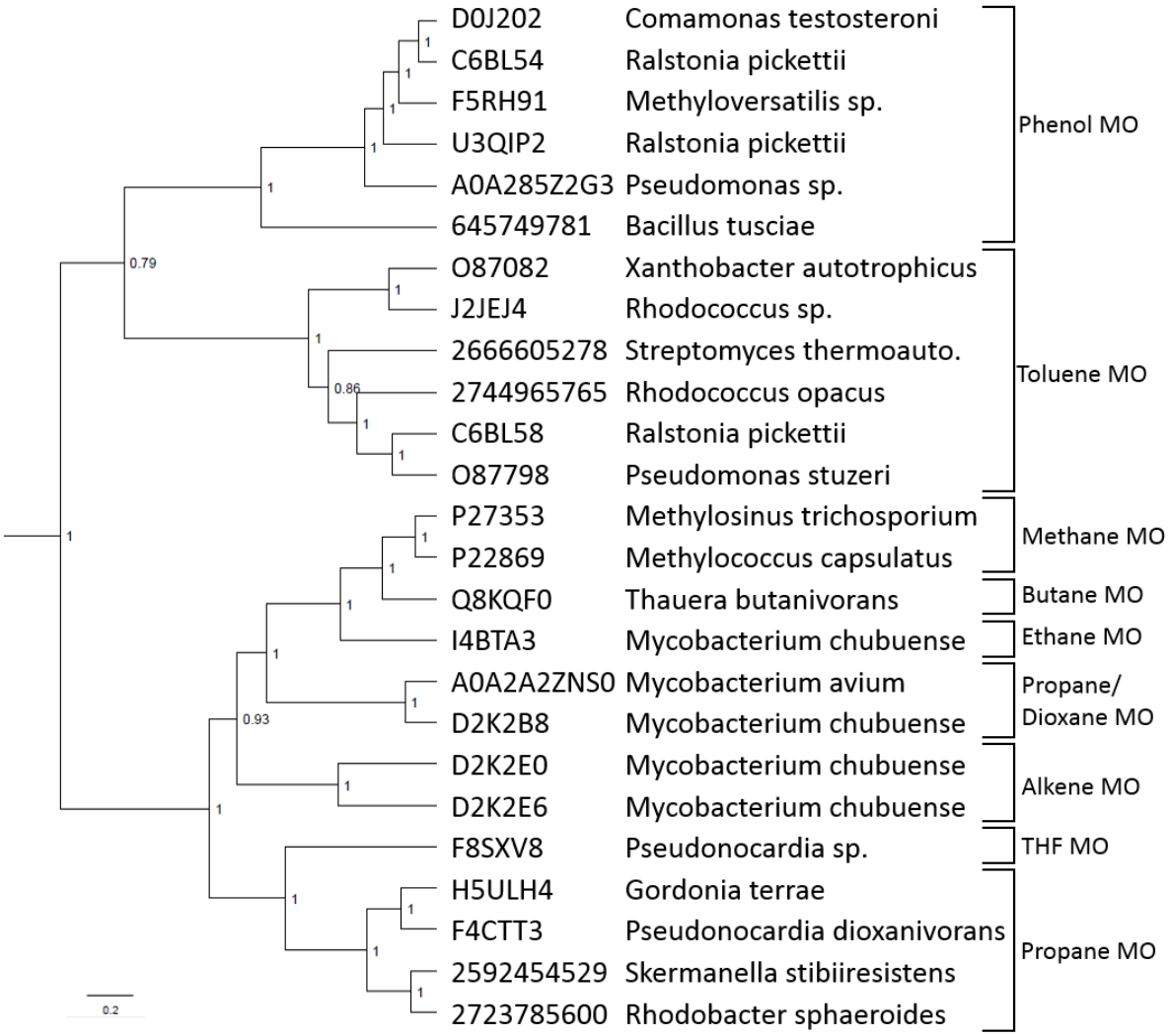


Figure A7.11: Bayesian analysis of BMM phylogeny using reduced dataset and including ancestral root. Reduced dataset was constructed from alpha hydroxylase subunit representatives of each functional radiation identified in Figure 4.2, with preference towards characterised members. Protein sequences were aligned in MAFFT v7, and phylogenetic analysis performed in BEAST v1.8.2 using relaxed clock methodology for 10 million states. Accession identifiers are included prior to the corresponding names of host species for each BMM member, with alpha-numeric and numeric accession values corresponding to the UniProt and JGI database respectively. Posterior probabilities are represented at each node, and functional group of BMM members as per Figure 4.2 is included at far right.

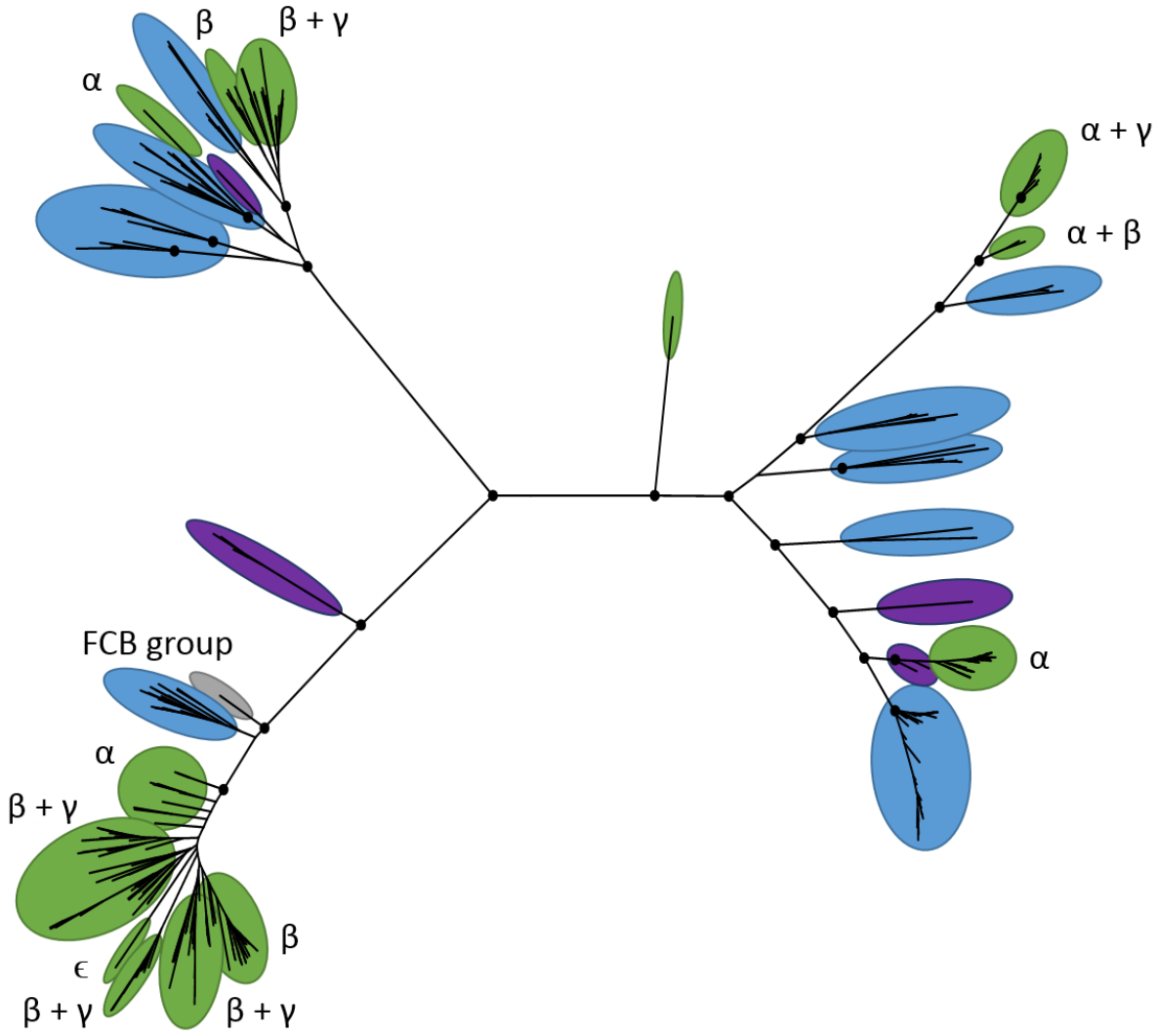


Figure A7.12: Phylum and class level differentiation of BMM host organisms. Differentiation has been overlayed on alpha unit BMM phylogeny as represented in figure 4.2. Blue magenta and green denote Actinobacteria, Firmicute and Proteobacteria clusters respectively. Proteobactera clusters are further differentiated based on class level representatives contained therein. A single representative of the Fibrobacteres, Chlorobi and bacteroidetes (FCB group) was also identified.

A7.1.3. Other supplementary figures

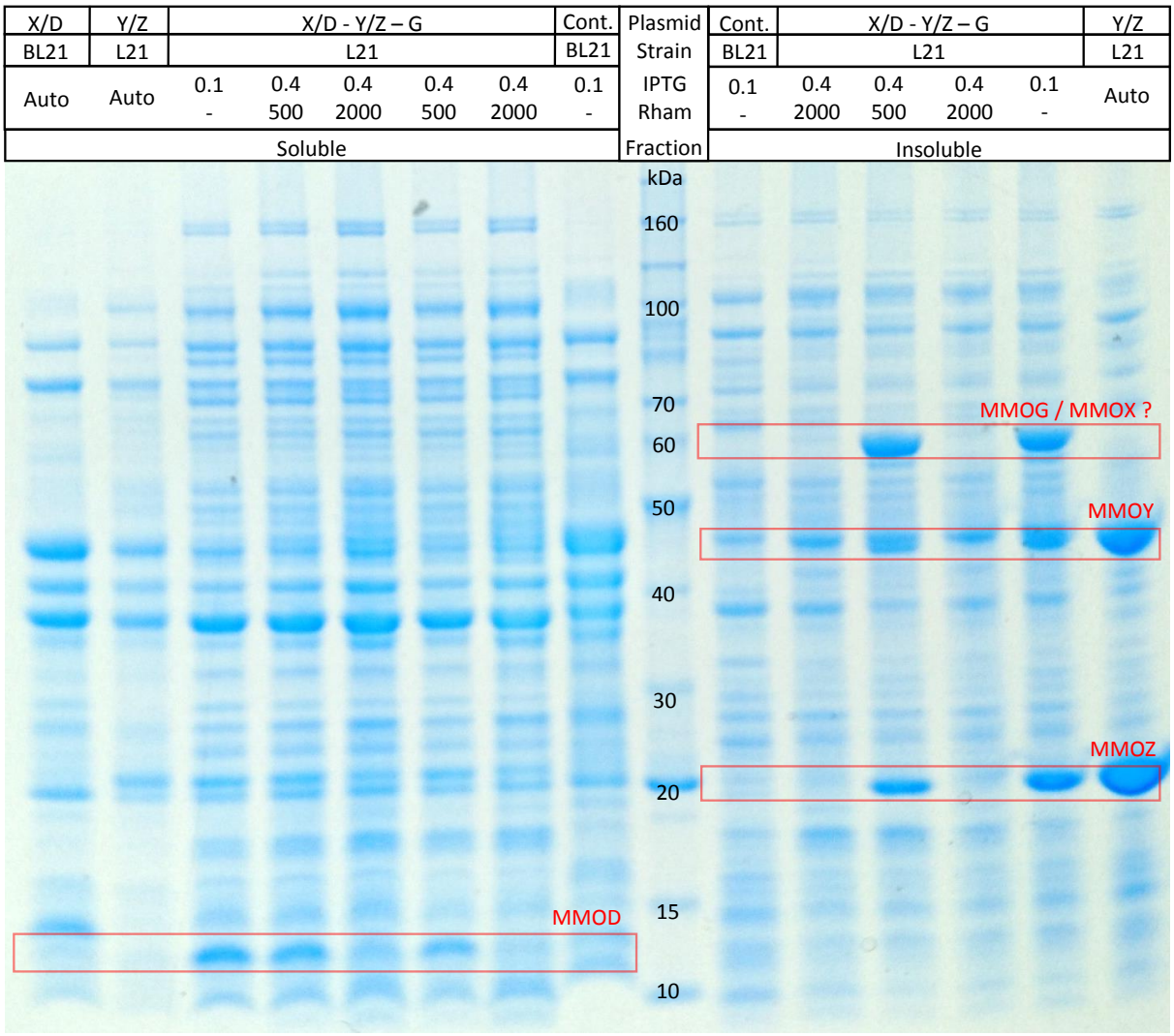


Figure A7.13: SDS-PAGE profile of X/Y/Z/D/G Lemo21 transformants using both moderate and high level rhamnose turn-down. The presence of the pCOLAD-MMOG (G), pETD-X/D (X/D) and pCDFD-Y/Z (Y/Z) plasmid in each strain is specified, along with the use of either BL21(DE3) (BL21) or Lemo21 (L21) cell lines. Expression conditions were either overnight autoinduction at 37°C (Auto) or IPTG induction with or without rhamnose (IPTG concentration in mM, rhamnose in μM) at 25°C with sampling after 24h. Multiple protein expression bands have been highlighted for emphasis, protein ladder is also included in central lane.

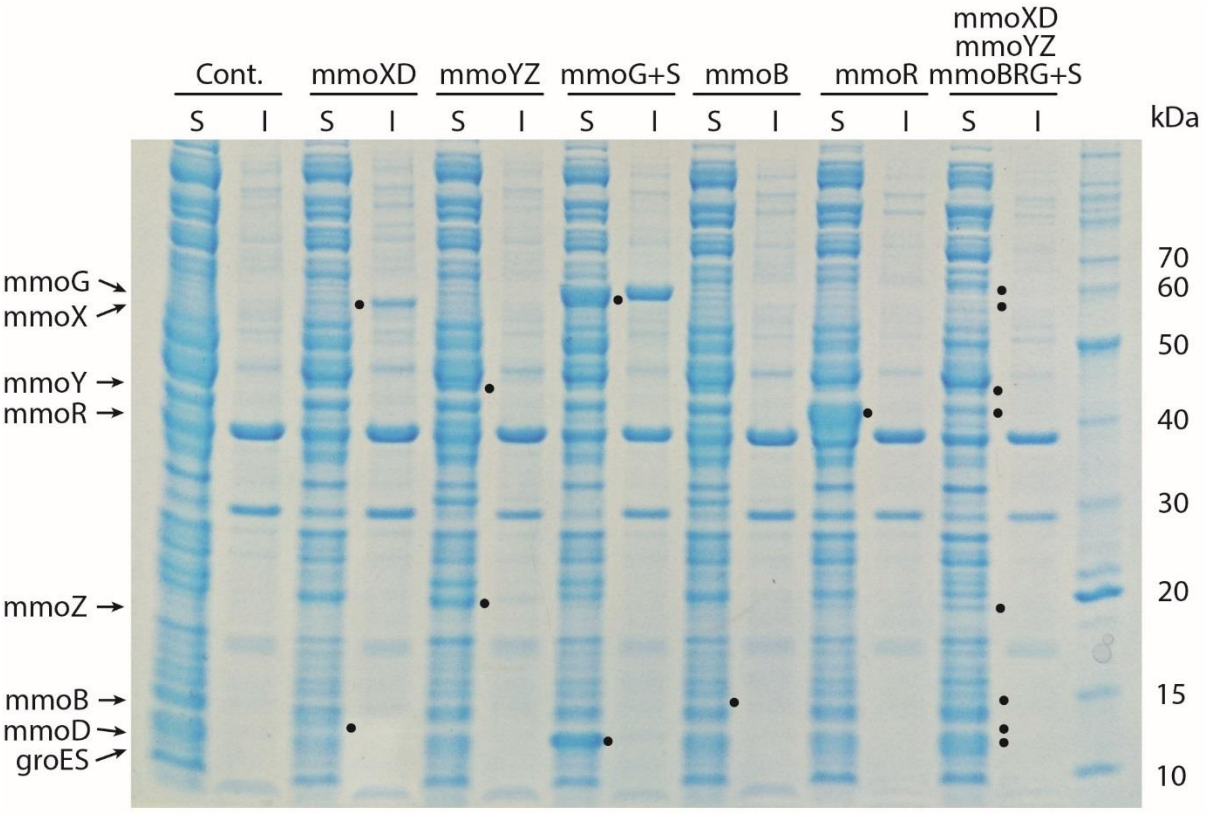


Figure A7.14: SDS-PAGE showing low temperature expression profiles of combinations of three different constructs containing heterologous sMMO genes including those for the reductase and regulatory units. Soluble (S) and insoluble (I) fractions shown for various strains of Lemo21 E. coli. mmoXD denotes the pETD-X/D plasmid, mmoYZ denotes the pCDFD-Y/Z plasmid, mmoG+S denotes the pCOLAD-S/G plasmid and mmoBRG+S denotes the pRSFD-B/R//G/S plasmid. Dots indicate molecular weight of proteins corresponding to inserted genes in respective strain, with corresponding identity of each dot at far right. Overnight autoinduction at 25°C were used for all samples. Blank pCOLADuet-1 plasmid under identical expression conditions was used as control (Cont.). Protein ladder with adjacent molecular weights also included in right hand lane.

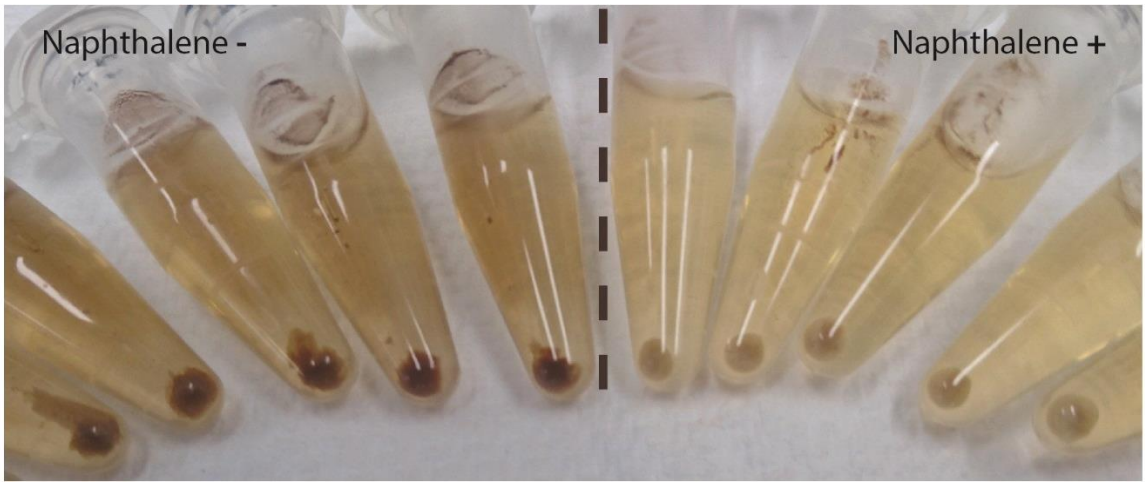


Figure A7.15: Demonstration of false positive in the naphthalene assay occurring in control samples lacking added naphthalene, with comparison to those under standard trial conditions. Samples were used from IPTG expression trials at 30°C using *E.coli* containing all eight heterologous sMMO genes. Samples from five trial runs induced at different IPTG concentrations were used, each trial run being used under both assay (naphthalene +) and control (naphthalene -) conditions. No assay conditions apart from the presence / absence of naphthalene were changed between samples.

A7.2. Supplementary tables

Table A7.1: Sequence identity of divergent GroEL associated with BMM groups and conventional GroEL sequences contained in respective species. All identity values are quotes as percentages. Sequences are either denoted as conventional (Conv.), or with the respective substrate of the BMM group. All sequences were obtained from the JGI IMG/ER database.

		M.cap Conv.	***M.tri*** Conv.	***S.aqu*** Conv.	***M.sme*** Conv.	***M.chu*** Conv.	***M.cap*** Meth.	***M.tri*** Meth.	***S.aqu*** But.	***M.sme*** Pro.	***M.chu*** Grp 6
M. capsulatus (Bath)	Conv.	x	67	79	54	54	39	42	31	49	51
M. trichosporium (OB3b)	Conv.		x	65	51	50	39	41	31	46	47
S. aquatica	Conv.			x	52	53	39	41	31	48	50
M. smegmatis (MC2 155)	Conv.				x	89	39	38	30	48	52
M. chubuense (NBB4)	Conv.					x	38	37	30	49	51
M. capsulatus (Bath)	Methane						x	43	32	35	37
M. trichosporium (OB3b)	Methane							x	32	35	36
S. aquatica	Butane								x	28	29
M. smegmatis (MC2 155)	Propane									x	47
M. chubuense (NBB4)	Grp 6										x

Table A7.2: Bacterial strains and plasmids used in heterologous sMMO trials

Strain or plasmid	Description	Source
Strain		
E. coli Turbo	Cloning host	NEB
E. coli BL21 (DE3)	Expression host	NEB
E. coli Lemo 21	Expression host with Lemo 21 plasmid	NEB
Plasmid		
pCDF-1b	Vector used for cloning	Merck
pRSF-1b	Vector used for cloning	Merck
pCOLADuet-1	Vector used for cloning	Merck
pUC57-MMOG	pUC57 containing synthesised MMOG construct for cloning	GenScript
pUC57-MMOR	pUC57 containing synthesised MMOR construct for cloning	GenScript
pUC57-MMOB	pUC57 containing synthesised MMOB construct for cloning	GenScript
pUC57-MMOZ	pUC57 containing synthesised MMOZ construct for cloning	GenScript
pUC57-MMOD	pUC57 containing synthesised MMOD construct for cloning	GenScript
pET21-GroES	pET21-1a containing synthesised GroES construct	GenScript
pET21-GroES(His)	pET21-1a containing synthesised His tagged GroES construct	GenScript
pET21-MMOR	pET21a containing MMOR construct subcloned from pUC57-MMOR	Genscript
pCDFD-MMOY	pCDFDuet-1 containing synthesised MMOY construct in MCS-1	Genscript
pETD-MMOX	pETDuet-1 containing synthesised MMOX construct in MCS-1	Genscript
pRSFD-MMOG	pRSFDuet-1 containing synthesised MMOG construct in MCS-2	Genscript
pCOLAD-GroES	pCOLADuet-1 containing synthesised GroES construct in MCS-1	Genscript
pCOLAD-S/G	pCOLADuet-1 containing synthesised GroES and mmoG construct in MCS-1 and MCS-2 respectively	Genscript
pRSFD-G/S	pRSFD-mmoG containing GroES construct subcloned from pUC57-GroES including preceding RBS sequence	Genscript
pET21-B/R	pET21-MMOR containing MMOB construct subcloned from pUC57-MMOB including preceding RBS sequence	Genscript
pRSF-MMOG	pRSF-1b containing MMOG construct subcloned from pUC57-MMOG	This study
pRSFD-B/R//G/S	pRSFD-G/S containing MMOB and MMOR including RBS situated between from pET21-B/R in MCS-1	This study
pETD-X/D	pETD-MMOX containing MMOD construct subcloned from pUC57-MMOD in MCS-2	This study
pCDFD-Y/Z	pCDFD-MMOY containing MMOZ construct subcloned from pUC57-MMOZ in MCS-2	This study

Table A7.3: Standard antibiotic concentrations used during growth of *E. coli* strains

Antibiotic	Concentration*	Corresponding plasmids
Ampicillin	100μg/ml	pET21-1b, pETDuet-1, pUC57
Chloramphenicol	33μg/ml	(Lemo21)
Kanamycin	30μg/ml	pRSF-1b, pCOLADuet-1
Streptomycin	25μg/ml	pCDF-1b, pCDFDuet-1

* As per recommendation for *E. coli* Turbo, BL21(DE3) and Lemo21 cell lines (www.neb.com)

A7.3. Other supplementary data

A7.3.1. Synthesised gene sequences

Gene name: MMOX

Key features: RBS(designed)/NdeI/*mmoX*/AvrII

Sequence:

AAGCTTGCGGCCGCATAATGCTTAAGTCGAACAGAAAGTAATCGTATTGTACACGGCCG
CATAATCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCAT
CTCGCAGAACGAGGGAATAGGGAGGACTACATATGGCGCTGAGCACCGCGACCAAAGC
GGCGACCGATGCGCTGGCGGCGAACCGTGCGCCGACCAGCGTGAACGCGCAGGAAGTTC
ACCGTTGGTTGCAGAGCTTCAACTGGGACTTTAAAAACAACCGTACCAAGTACGCGACC
AAGTATAAAATGGCGAACGAGACCAAGGAACAGTTCAAACTGATCGCGAAGGAGTACG
CGCGTATGGAAGCGGTGAAGGACGAGCGTCAGTTTGGCAGCCTGCAAGATGCGCTGACC
CGTCTGAACGCGGGTGTGCGTGTTCACCCGAAATGGAACGAGACCATGAAGGTGGTTAG
CAACTTCCTGGAAGTGGGCGAATACAACGCGATCGCGGCGACCGGTATGCTGTGGGATA
GCGCGCAGGCGGCGGAACAGAAGAACGGTTACCTGGCGCAGGTGCTGGACGAGATTCGT
CACACCCACCAATGCGCGTATGTTAACTACTATTTTGCGAAAAACGGTCAGGACCCGGCG
GGTCACAACGATGCGCGTCGTACCCGTACCATTGGTCCGCTGTGGAAGGGTATGAAACG
TGTGTTCAGCGACGGCTTTATTAGCGGTGATGCGGTGGAATGCAGCCTGAACCTGCAACT
GGTTGGCGAGGCGTGCTTCACCAACCCGCTGATCGTGGCGGTTACCGAATGGGCTGCGG
CGAACGGTGACGAGATCACCCCGACCGTTTTTCTGAGCATTGAGACCGATGAACTGCGTC
ACATGGCGAACGGTTACCAGACCGTGGTTAGCATTGCGAACGACCCGGCGAGCGCGAAA
TATCTGAACACCGATCTGAACAACGCGTTCTGGACCCAGCAAAAGTACTTTACCCCGGTG
CTGGGCATGCTGTTCGAATATGGTAGCAAGTTTAAAGTGGAGCCGTGGGTTAAAACCTG
GAACCGTTGGGTGTACGAAGATTGGGGTGGCATCTGGATTGGTCGTCTGGGCAAGTATG
GTGTTGAGAGCCCGCGTAGCCTGAAGGACGCGAAACAGGATGCGTACTGGGCGCACCAC
GACCTGTACCTGCTGGCGTATGCGCTGTGGCCGACCGGCTTCTTTCGTCTGGCGCTGCCG
GACCAAGAGGAAATGGAGTGGTTCGAAGCGAACTATCCGGGTTGGTACGATCACTATGG
TAAAATCTATGAGGAATGGCGTGCGCGTGGCTGCGAAGACCCGAGCAGCGGTTTCATCC
CGCTGATGTGGTTTATTGAGAACAACCACCCGATCTACATTGATCGTGTGAGCCAGGTTC
CGTTTTGCCCGAGCCTGGCGAAGGGTGCGAGCACCCTGCGTGTTCACGAATATAACGGCC
AGATGCACACCTTCAGCGACCAATGGGGTGAACGTATGTGGCTGGCGGAGCCGGAACGT
TACGAGTGCCAGAACATCTTTGAGCAATATGAAGGCCGTGAGCTGAGCGAAGTGATTGC
GGAGCTGCACGGCCTGCGTAGCGACGGTAAAACCCTGATCGCGCAACCGCACGTTCGTG
GTGATAAACTGTGGACCCTGGACGATATTAAGCGTCTGAACTGCGTGTTCAAAAACCCG
GTTAAGGCGTTTAACTAACCTAGG

Translated protein sequence:

MALSTATKAATDALAANRAPTSVNAQEVHRWLQSFNWDFKNNRTKYATKYKMANETKEQ
FKLIAKEYARMEAZKDERQFGSLQDALTRLNAGVRVHPKWNETMKVVSNFLEVGEYNAIA
ATGMLWDSAQAAEQKNGYLAQVLDEIRHTHQCAYVNYYFAKNGQDPAGHNDARRTRTIGP
LWKGMKRVFSDGFISGDAVECSLNLQLVGEACFTNPLIVAVTEWAAANGDEITPTVFLSIETD
ELRHMANGYQTVVSIANDPASAKYLNTDLNNAFWTQQKYFTPVLGMLFEYGSKFKVEPWV
KTWNRWVYEDWGGIWIGRLGKYGVESPRSLKDAKQDAYWAHHDLYLLAYALWPTGFFRL
ALPDQEEMEWFEANYPGWYDHYGKIYEEWRARGCEDPSSGFIPLMWFIENNHPIYIDRVSQV
PFCPSLAKGASTLRVHEYNGQMHTFSDQWGERMWLAEPERYECQNIFEQYEGRELSEVIAEL
HGLRSDGKTLIAQPHVRGDKLWTLDDIKRLNCVFKNPVKAFN

Gene name: MMOY

Key features: RBS(designed)/NdeI/*mmoY*/AvrII

Sequence:

AAGCTTGCGGCCGCATAATGCTTAAGTCGAACAGAAAGTAATCGTATTGTACACGGCCG
CATAATCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCAT
CTGTACAAGTAGACTAGGAAGGAGGTATACATATGAGCATGCTGGGCGAGCGTCGTCGT
GGTCTGACCGATCCGGAAATGGCGGCGGTGATCCTGAAGGCGCTGCCGGAAGCGCCGCT
GGACGGTAACAACAAGATGGGCTACTTCGTGACCCCGCGTTGGAAACGTCTGACCGAGT
ACGAAGCGCTGACCGTTTATGCGCAGCCGAACGCGGACTGGATTGCGGGTGGCCTGGAC
TGGGGCGATTGGACCCAAAAGTTCCACGGTGGCCGTCCGAGCTGGGGTAACGAGACCAC
CGAACTGCGTACCGTGGACTGGTTTAAACATCGTGATCCGCTGCGTCGTTGGCATGCGCC
GTACGTTAAGGACAAAGCGGAGGAATGGCGTTACACCGATCGTTTCCTGCAAGGTTATA
GCGCGGACGGCCAAATCCGTGCGATGAACCCGACCTGGCGTGATGAGTTCATTAACCGT
TACTGGGGTGCGTTCCTGTTTAACGAATATGGCCTGTTTAACGCGCACAGCCAAGGTGCG
CGTGAGGCGCTGAGCGATGTGACCCGTGTTAGCCTGGCGTTCTGGGGTTTTGACAAGATC
GATATTGCGCAGATGATCCAACTGGAACGTGGCTTCCTGGCGAAAATTGTGCCGGGTTTT
GATGAGAGCACCGCGGTTCCGAAGGCGGAGTGGACCAACGGCGAAGTGTACAAAAGCG
CGCGTCTGGCGGTGGAGGGTCTGTGGCAGGAAGTTTTCGATTGGAACGAAAGCGCGTTT
AGCGTGCACGCGGTTTATGACGCGCTGTTCGGCCAGTTTGTTCGTCGTGAGTTCTTTCAAC
GTCTGGCGCCGCGTTTCGGTGACAACCTGACCCCGTTCTTTATCAACCAGGCGCAAACCT
ACTTTCAAATTGCGAAGCAGGGCGTGCAGGACCTGTACTATAACTGCCTGGGTGACGATC
CGGAGTTCAGCGACTATAACCGTACCGTTATGCGTAACTGGACCGGCAAATGGCTGGAG
CCGACCATTGCGGCGCTGCGTGATTTCATGGGTCTGTTTGCGAAGCTGCCGGCGGGCACC
ACCGACAAAGAGGAAATTACCGCGAGCCTGTACCGTGTGGTTGACGATTGGATCGAGGA
CTATGCGAGCCGTATTGATTTTAAAGCGGACCGTGATCAGATCGTGAAGGCGGTTCTGGC
GGGTCTGAAATAACCTAGG

Translated protein sequence:

MSMLGERRRGLTDPEMAAVILKALPEAPLDGNNKMGYFVTPRWKRLTEYEALTVYAQPNA
DWIAGGLDWGDWTQKFHGGRPSWGNETTELRTVDWFKHRDPLRRWHAPYVKDKAEEWR
YTDRFLQGYSADGQIRAMNPTWRDEFINRYWGAFLFNEYGLFNAHSQGAREALSDVTRVSL
AFWGFDKIDIAQMIQLERGFLAKIVPGFDESTAVPKAEWTNGEVYKSARLAVEGLWQEVFD
WNESAFSVHAVYDALFGQFVRREFFQRLAPRFGDNLTPFFINQAQTYFQIAKQGVQDLYYNC
LGDDPEFSDYNRTVMRNWTGKWLEPTIAALRDFMGLFAKLPAGTTDKEEITASLYRVVDDW
IEDYASRIDFKADRDQIVKAVLAGLK

Gene name: MMOZ

Key features: BamHI/NcoI/*mmoZ*/HindIII

Sequence:

GGATCCCATGGCGAAGCTGGGTATCCACAGCAACGACACCCGTGATGCGTGGGTGAACA
AAATTGCGCAGCTGAACACCCTGGAGAAGGCGGCGGAAATGCTGAAACAATTCCGTATG
GACCACACCACCCCGTTTCGTAACAGCTACGAGCTGGACAACGATTATCTGTGGATCGAA
GCGAAGCTGGAGGAAAAAGTGGCGGTTCTGAAGGCGCGTGCGTTCAACGAGGTTGACTT
TCGTCACAAGACCGCGTTCGGTGAAGACGCGAAAAGCGTGCTGGATGGCACCGTTGCGA
AAATGAACGCGGCGAAGGATAAATGGGAGGCGGAAAAGATCCACATTGGTTTCCGTCAG
GCGTACAAACCGCCGATTATGCCGGTGAACTATTTTCTGGACGGCGAGCGTCAACTGGGC
ACCCGTCTGATGGAACTGCGTAACCTGAACTACTATGATACCCCGCTGGAGGAACTGCGT
AAGCAGCGTGGCGTTCGTGTGGTTCACCTGCAAAGCCCGCACTAAAAGCTT

Translated protein sequence:

MAKLGIHSNDTRDAWVNKIAQLNTLEKAAEMLKQFRMDHTTPFRNSYELDNDYLWIEAKL
EEKVAVLKARAFNEVDFRHKTAFGEDAKSVLDGTVAKMNAAKDKWEAEKIHIGFRQAYKP
PIMPVNYFLDGERQLGTRLMELRNLNYYDTPLEELRKQRGVRVVHLQSPH

Gene name: MMOD

Key features: XbaI/RBS(designed)/NcoI/*mmoD*/HindIII

Sequence:

TCTAGAGCAACCCCTAGAATAATAACAGGGCAAACCATGGTGGAGAGCGCGTTCCAGCC
GTTTAGCGGTGACGCGGATGAATGGTTCGAGGAACCGCGTCCGCAAGCGGGCTTCTTTCC
GAGCGCGGACTGGCACCTGCTGAAGCGTGATGAGACCTACGCGGCGTATGCGAAAGACC
TGGATTTTATGTGGCGTTGGGTGATCGTTCGTGAGGAACGTATTGTGCAGGAAGGTTGCA
GCATCAGCCTGGAGAGCAGCATTCGTGCGGTGACCCACGTTCTGAACTACTTTGGTATGA
CCGAGCAGCGTGCGCCGGCGGAAGACCGTACCGGTGGCGTTCAACACTAAAAGCTT

Translated protein sequence:

MVESAFQPFSGDADEWFEEPRPQAGFFPSADWHLLKRDETYAAYAKDLDFMWRWVIVREE
RIVQEGCSISLESSIRAVTHVLNYFGMTEQRAPAEDRTGGVQH

Construct name: MMOB

Key features: HindIII/RBS/*mmoB*/AvrII/XhoI

Sequence:

AAGCTTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAAATGAGCGTGAAC
AGCAACGCGTACGATGCGGGTATCATGGGCCTGAAGGGTAAAGACTTCGCGGATCAGTT
CTTTGCGGACGAGAACCAAGTGGTTCACGAAAGCGATACCGTGGTTCTGGTGCTGAAGA
AAAGCGACGAGATTAACACCTTTATCGAGGAAATTCTGCTGACCGATTACAAGAAAAAC
GTTAACCCGACCGTGAACGTTGAGGACCGTGCGGGCTATTGGTGGATCAAGGCGAACGG
TAAAATTGAAGTGGACTGCGATGAGATCAGCGAACTGCTGGGCCGTCAGTTCAACGTTT
ACGACTTTCTGGTGGATGTTAGCAGCACCATCGGCCGTGCGTATACCCTGGGTAACAAGT
TCACCATTACCAGCGAGCTGATGGGTCTGGACCGTAAACTGGAAGATTATCACGCGTAA
CCTAGGCTCGAG

Translated protein sequence:

MSVNSNAYDAGIMGLKGKDFADQFFADENQVVHESDTVVLVLKKSDEINTFIEEILLTDYKK
NVNPTVNVEDRAGYWWIKANGKIEVDCDEISELLGRQFNVYDFLVDVSSTIGRAYTLGNKFT
ITSELMGLDRKLEDYHA

Construct name: MMOR

Key features: BamHI/NdeI/*mmoC*/HindIII/AvrII

Sequence:

GGATCCCATATGCAGCGTGTTCACACCATCACCGCGGTGACCGAAGACGGCGAGAGCCT
GCGTTTCGAATGCCGTAGCGACGAGGATGTTATCACCGCGGCGCTGCGTCAAAACATTTT
TCTGATGAGCAGCTGCCGTGAAGGTGGCTGCGCGACCTGCAAGGCGCTGTGCAGCGAGG
GCGACTACGATCTGAAAGGTTGCAGCGTTCAGGCGCTGCCGCCGGAGGAAGAGGAAGAG
GGCCTGGTGCTGCTGTGCCGTACCTACCCGAAGACCGACCTGGAAATCGAGCTGCCGTAT
ACCCACTGCCGTATTAGCTTCGGTGAAGTGGGCAGCTTTGAAGCGGAAGTGGTTGGTCTG
AACTGGGTTAGCAGCAACACCGTGCAGTTCCTGCTGCAAAAGCGTCCGGATGAATGCGG
TAACCGTGGCGTTAAATTCGAGCCGGGTCAATTTATGGACCTGACCATTCCGGGCACCGA
TGTGAGCCGTAGCTATAGCCCGGCGAACCTGCCGAACCCGGAGGGCCGTCTGGAGTTCC
TGATTCGTGTTCTGCCGGAAGGTCGTTTTAGCGACTATCTGCGTAACGATGCGCGTGTGG
GCCAAGTTCTGAGCGTGAAGGGTCCGCTGGGCGTGTTCGGTCTGAAAGAGCGTGGCATG
GCGCCGCGTTACTTTGTGGCGGGTGGCACCGGTCTGGCGCCGGTGGTTAGCATGGTGCGT

CAGATGCAAGAATGGACCGCGCCGAACGAGACCCGTATCTACTTCGGCGTTAACACCGA
ACCGGAGCTGTTTTATATTGACGAACTGAAGAGCCTGGAGCGTAGCATGCGTAACCTGA
CCGTTAAAGCGTGCGTGTGGCACCCGAGCGGTGATTGGGAAGGTGAACAGGGTAGCCCG
ATTGATGCGCTGCGTGAGGACCTGGAGAGCAGCGACGCGAACCCGGACATCTATCTGTG
CGGCCCGCCGGGTATGATTGATGCGGCGTGCGAGCTGGTTCGTAGCCGTGGCATTCCGGG
TGAACAAGTGTTCTTTGAGAAATTTCTGCCGAGCGGTGCGGCGTAAAAGCTTCCTAGG

Translated protein sequence:

MQRVHTITAVTEDGESLRFECRSDEDVITAALRQNIFLMSSCREGGCATCKALCSEGDYDLK
GCSVQALPPEEEEEGLVLLCRTYPKTDLEIELPYTHCRISFGEVGSFEAEVVGLNWVSSNTVQ
FLLQKRPDECGNRGVKFEPGQFMDLTIPGTDVSRSYSPANLPNPEGRLEFLIRVLPEGRFSDYL
RNDARVGQVLSVKGPLGVFGLKERGMAPRYFVAGGTGLAPVVSMVRQMQEWTAPNETRIY
FGVNTEPELFYIDELKSLERSMRNLTVKACVWHPSGDWEGEQGSPIDALREDLESSDANPDIY
LCGPPGMIDAACELVRSRGIPGEQVFFEKFLPSGAA

Construct name: MMOG

Key features: BamHI/NheI/RBS/Nco1/*mmoG*/(Flag-tag)/AvrII/HindIII

Sequence:

GGATCCGCTAGCGAAGGAGATATACCATGGCGAAAGAGGTGGTTTACCGTGGCAGCGCG
CGTCAGCGTATGATGCAGGGCATTGAGATTCTGGCGCGTGCGGCGATTCCGACCCTGGGC
GCGACCGGTCCGAGCGTGATGATCCAGCACCGTGCGGACGGTCTGCCGCCGATTAGCAC
CCGTGATGGTGTGACCGTTGCGAACAGCATCGTGCTGAAGGACCGTGTTGCGAACCTGG
GCGCGCGTCTGCTGCGTGACGTTGCGGGCACCATGAGCCGTGAGGCGGGTGATGGCACC
ACCACCGCGATCGTGCTGGCGCGTCACATTGCGCGTGAAATGTTCAAGAGCCTGGCGGTT
GGTGCGGACCCGATCGCGCTGAAACGTGGTATTGATCGTGCGGTGGCGCGTGTTAGCGA
AGACATCGGCGCGCGTGCGTGGCGTGGTGATAAGGAGAGCGTGATTCTGGGCGTGGCGG
CGGTTGCGACCAAAGGTGAACCGGGCGTTGGTCGTCTGCTGCTGGAGGCGCTGGACGCG
GTGGGCGTTCACGGTGCTGTGAGCATCGAGCTGGGTCAGCGTCGTGAAGACCTGCTGGA
TGTGGTTGACGGCTACCGTTGGGAAAAGGGTTACCTGAGCCCGTATTTCGTGACCGATCG
TGCGCGTGAGCTGGCGGAGCTGGAAGATGTTTATCTGCTGATGACCGACCGTGAAGTGG
TTGATTTTATCGACCTGGTGCCGCTGCTGGAGGCGGTTACCGAAGCGGGTGGCAGCCTGC
TGATTGCGGCGGACCGTGTGCACGAGAAAGCGCTGGCGGGTCTGCTGCTGAACCACGTG
CGTGGTGTTTTCAAGGCGGTGGCGGTTACCGCGCCGGGTTTTGGTGATAAACGTCCGAAC
CGTCTGCTGGACCTGGCGGCGCTGACCGGTGGCCGTGCGGTTCTGGAAGCGCAAGGCGA
TCGTCTGGACCGTGTGACCCTGGCGGATCTGGGTCGTGTTCGTCGTGCGGTGGTTAGCGC
GGACGATACCGCGCTGCTGGGTATTCCGGGCACCGAGGCGAGCCGTGCGCGTCTGGAAG
GTCTGCGTCTGGAGGCGGAACAGTACCGTGCGCTGAAGCCGGGTCAAGGTAGCGCGACC
GGTCGTCTGCACGAGCTGGAGGAAATCGAAGCGCGTATTGTGGGCCTGAGCGGCAAGAG
CGCGGTGTACCGTGTTGGTGGCGTGACCGACGTTGAGATGAAAGAACGTATGGTTCGTAT
TGAGAACGCGTATCGTAGCGTGGTTAGCGCGCTGGAGGAAGGTGTGCTGCCGGGTGGCG
GTGTTGGCTTTCTGGGTAGCATGCCGGTGCTGGCGGAGCTGGAAGCGCGTGACGCGGAT
GAAGCGCGTGGCATCGGTATTGTTCGTAGCGCGCTGACCGAGCCGCTGCGTATCATTGGT
GAAAACAGCGGTCTGAGCGGCGAGGCGGTGGTTGCGAAAGTTATGGATCACGCGAATCC
GGGTTGGGGTTATGATCAAGAGAGCGGCAGCTTCTGCGACCTGCACGCGCGTGGTATCT
GGGATGCGGCGAAGGTGCTGCGTCTGGCGCTGGAAAAAGCGGCGAGCGTTGCGGGCACC
TTCCTGACCACCGAGGCGGTGGTTCTGGAAATTCCGGACACCGATGCGTTTGCGGGCTTT
AGCGCGGAGTGGGCGGCGGCGACCCGTGAAGACCCGCGTGTTTAAGGATCCGACTACAA
GGACGACGATGACAAGTAACCTAGCAAGCTT

Translated protein sequence:

MAKEVVYRGSARQRMMQGIEILARAAIPTLGATGPSVMIQHRADGLPPISTRDGVTVANSIV
LKDRVANLGARLLRDVAGTMSREAGDGTTTAIVLARHIAREMFKSLAVGADPIALKRGIDRA
VARVSEDIGARAWRGDKESVILGVAAVATKGEPGVGRLLLEALDAVGVHGAVSIELGQRRE
DLLDVVDGYRWEKGYLSPYFVTDRARELAELEDVYLLMTDREVVDFIDLVPLLEAVTEAGG
SLLIAADRVHEKALAGLLLNHVRGVFKAVAVTAPGFGDKRPNRLLDLAALTGGRAVLEAQG
DRLDRVTLADLGRVRRAVVSADDTALLGIPGTEASRARLEGLRLEAEQYRALKPGQGSATG
RLHELEEIEARIVGLSGKSAVYRVGGVTDVEMKERMVRIENAYRSVVSALEEGVLPGGGVGF
LGSMPVLAELEARDADEARGIGIVRSALTEPLRIIGENSGLSGEAVVAKVMDHANPGWGYDQ
ESGSFCDLHARGIWDAAKVLRLALEKAASVAGTFLTTEAVVLEIPDTDAFAGFSAEWAAATR
EDPRV

Construct name: GroES

Key features: HindIII/RBS/*groES*/NotI

Sequence:

AAGCTTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAAATGAAGATTCGC
CCGCTGCATGACCGTGTTGTTGTGATTCGTCGTGAAGAGGAGAAAACCAGCCCGGGTGG
CATTGTGATTCCGGACACCGCGAAGGAGAAACCGATCAAGGGTGAAATTGTGGCGGTTG
GCACCGGCAAAGTGCTGGACAACGGTCAAGTGCGTCCGCTGGCGGTTAAGGCGGGTGAT
ACCGTTCTGTTCGGCAAGTACAGCGGCACCGAGATCAAAATTGACGGCACCGAGTATCT
GATGCTGCGCGAAGACGACATTATGGGCGTTATTGAAAGCTAAGCGGCCGC

Translated protein sequence:

MKIRPLHDRVVVIRREEEKTSPGGIVIPDTAKEKPIKGEIVAVGTGKVLDNGQVRPLAVKAGD
TVLFGKYSGTEIKIDGTEYLMLREDDIMGVIES

Construct name: GroES-His

Key features: Nde1/*groES*/His-tag/Nhe1/BamHI

Sequence:

CATATGAAGATTCGCCCGCTGCATGACCGTGTTGTTGTGATTCGTCGTGAAGAGGAGAAA
ACCAGCCCGGGTGGCATTGTGATTCCGGACACCGCGAAGGAGAAACCGATCAAGGGTGA
AATTGTGGCGGTTGGCACCGGCAAAGTGCTGGACAACGGTCAAGTGCGTCCGCTGGCGG
TTAAGGCGGGTGATACCGTTCTGTTCGGCAAGTACAGCGGCACCGAGATCAAAATTGAC
GGCACCGAGTATCTGATGCTGCGCGAAGACGACATTATGGGCGTTATTGAAAGCAGCCA
CCATCACCATCACCATTAAGCTAGCGGATCC

Translated protein sequence:

MKIRPLHDRVVVIRREEEKTSPGGIVIPDTAKEKPIKGEIVAVGTGKVLDNGQVRPLAVKAGD
TVLFGKYSGTEIKIDGTEYLMLREDDIMGVIESSHHHHHH

A7.3.2. Standard media and buffer recipes

SOB Media

Tryptone	2% (w/v)
Yeast extract	0.5% (w/v)
NaCl	10mM
KCl	2.5mM
$MgCl_2$	10mM
$MgSO_4$	10mM

Phosphate buffer (7.5 pH)

KH_2PO_4	8mM
Na_2HPO_4	42mM
Glycerol	10%(v/v)

- pH balance to 7.5

Triton-X 100 lysis buffer

Tris-HCl	50mM
NaCl	150mM
Triton-X 100	1%
EDTA	5mM

- pH balance to 7.4

SOC Media

Tryptone	2% (w/v)
Yeast extract	0.5% (w/v)
NaCl	10mM
KCl	2.5mM
$MgCl_2$	10mM
$MgSO_4$	10mM
Glucose	20mM

TB Media

Tryptone	1.2% (w/v)
Yeast extract	2.4% (w/v)
Glycerol	0.4% (v/v)
KH_2PO_4	17mM
Na_2HPO_4	72mM

Autoinduction media

Tryptone	2% (w/v)
Yeast extract	0.5% (w/v)
NaCl	170mM
Glycerol	0.6% (v/v)
Lactose	0.2% (w/v)
Glucose	10mM
KH_2PO_4	42mM
Na_2HPO_4	22mM

TAE running buffer

Tris-base	40mM
Acetic acid	20mM
EDTA	1mM

LB agar

Tryptone	1% (w/v)
Yeast extract	0.5% (w/v)
NaCl	170mM
Agar	1.5% (w/v)

TBS

Tris	50mM
NaCl	150mM

- pH balanced to 7.6

MES running buffer (x20)

MES	50mM
Tris	50mM
EDTA	1mM
SDS	0.1% (w/v)

- pH balanced to 7.3

SDS loading dye (x4)

Tris-buffer	250mM (pH – 6.8)
SDS	10% (w/v)
Glycerol	40% (v/v)
B-mecaptoethanol	5% (v/v)
Bromophenol blue	0.01% (w/v)

A7.3.3. Methanotroph growth media

Component	*Mineral salts*	*1181*
Major (mM)		
MgSO4	4.0	0.8
KH_2PO_4	2.0	2.0
Na_2HPO_4	2.0	2.0
NaH_2PO_4		0.9
NH_4Cl	10 (for AMS)	
KNO_3	10 (for NMS)	10
Intermediate (µM)		
$FeNH_4$-*EDTA*		11
$CaCl_2$	900	17
Trace (nM)		
$FeSO_4$	720	
$FeNH_4$-*EDTA*	620	
$ZnSO_4$	17	17
$MnCl_2$	7.6	7.6
H_3BO_3	240	240
Na_2MoO_4	6.2	6.2
$CoCl_2$	42	
$NiCl_2$	4.2	
$CuCl_2$	1000 (for +Cu)	3.0

A7.3.4. Protein mass spectrometry results

All proteomic analysis was conducted at the Monash Biomedical Proteomics Facility.

Instrumentation:
Dionex Ultimate 3000 RSLCnano – Thermo Scientific
QExtractive Plus 1 – Thermo Scientific
Acclaim PepMap RSLS – 75µm x 50cm, nanoViper, C18, 2µm, 100Å – Thermo Scientific
Acclaim PepMap 100 – 100µm x 2cm, nanoViper, C18, 5µm, 100Å – Thermo Scientific

Fragment analysis:

Search engine:	Mascot V2.4
Max missed cleavages:	2
Database:	Uniprot *E. coli* reference strain K12
	Uniprot *Methylococcus capsulatus* (strain ATCC 33009 / NCIMB 11132)

Analysis parameters:

Digest:	Trypsin
Peptide tolerance:	15 ppm
Fragment tolerance:	20 mmu

Individual results for analysis of putative MMOX gel band

Top 10 hits:

P22869	Methane monooxygenase component A alpha chain OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233
P0ABB0	ATP synthase subunit alpha OS=Escherichia coli (strain K12) GN=atpA PE=1 SV=1
P13035	Aerobic glycerol-3-phosphate dehydrogenase OS=Escherichia coli (strain K12) GN=glpD PE=1 SV=3
P18798	Methane monooxygenase component A beta chain OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233
P25714	Membrane protein insertase YidC OS=Escherichia coli (strain K12) GN=yidC PE=1 SV=2
P0A9P0	Dihydrolipoyl dehydrogenase OS=Escherichia coli (strain K12) GN=lpdA PE=1 SV=2
P0CB39	Phosphoethanolamine transferase EptC OS=Escherichia coli (strain K12) GN=eptC PE=1 SV=1
P0A6Y8	Chaperone protein DnaK OS=Escherichia coli (strain K12) GN=dnaK PE=1 SV=2
P0A6F3	Glycerol kinase OS=Escherichia coli (strain K12) GN=glpK PE=1 SV=2
P0ADG7	Inosine-5'-monophosphate dehydrogenase OS=Escherichia coli (strain K12) GN=guaB PE=1 SV=1

Sequence results for P22869:

Mass: 61007 Score: 52325 Matches: 1817(1187) Sequences: 223(146) emPAI: 10179979.45

Sequence coverage : 96% (506/527)

MALSTATKAATDALAANRAPTSVNAQEVHRWLQSFNWDFKNNRTKYATKYKMANETKEQ
FKLIAKEYARMEAVKDERQFGSLQDALTRLNAGVRVHPKWNETMKVVSNFLEVGEYNAIA
ATGMLWDSAQAAEQKNGYLAQVLDEIRHTHQCAYVNYYFAKNGQDPAGHNDARRTRTIGP
LWKGMKRVFSDGFISGDAVECSLNLQLVGEACFTNPLIVAVTEWAAANGDEITPTVFLSIETD
ELRHMANGYQTVVSIANDPASAKYLNTDLNNAFWTQQKYFTPVLGMLFEYGSKFKVEPWV
KTWNRWVYEDWGGIWIGRLGKYGVESPRSLKDAKQDAYWAHHDLYLLAYALWPTGFFRL
ALPDQEEMEWFEANYPGWYDHYGKIYEEWRARGCEDPSSGFIPLMWFIENNHPIYIDRVSQV
PFCPSLAKGASTLRVHEYNGQMHTFSDQWGERMWLAEPERYECQNIFEQYEGRELSEVIAEL
HGLRSDGKTLIAQPHVRGDKLWTLDDIKRLNCVFKNPVKAFN

Individual results for analysis of putative MMOY gel band

Top 10 hits:

P18798	Methane monooxygenase component A beta chain OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233
P0CE47	Elongation factor Tu 1 OS=Escherichia coli (strain K12) GN=tufA PE=1 SV=1
P0C8J8	D-tagatose-1,6-bisphosphate aldolase subunit GatZ OS=Escherichia coli (strain K12) GN=gatZ PE=1 SV=1
P0A6B7	Cysteine desulfurase IscS OS=Escherichia coli (strain K12) GN=iscS PE=1 SV=1
P0A799	Phosphoglycerate kinase OS=Escherichia coli (strain K12) GN=pgk PE=1 SV=2
P0ABH7	Citrate synthase OS=Escherichia coli (strain K12) GN=gltA PE=1 SV=1
P0AE06	Multidrug efflux pump subunit AcrA OS=Escherichia coli (strain K12) GN=acrA PE=1 SV=1
P0ABU2	Ribosome-binding ATPase YchF OS=Escherichia coli (strain K12) GN=ychF PE=1 SV=2
P0A953	3-oxoacyl-[acyl-carrier-protein] synthase 1 OS=Escherichia coli (strain K12) GN=fabB PE=1 SV=1
P0AAI5	3-oxoacyl-[acyl-carrier-protein] synthase 2 OS=Escherichia coli (strain K12) GN=fabF PE=1 SV=2

Sequence results for P18798:

Mass: 45161 Score: 45414 Matches: 1702(1067) Sequences: 212(139) emPAI: 44986774.42

Sequence coverage: 94% (366/389)

MSMLGERRRGLTDPEMAAVILKALPEAPLDGNNKMGYFVTPRWKRLTEYEALTVYAQPNA
DWIAGGLDWGDWTQKFHGGRPSWGNETTELRTVDWFKHRDPLRRWHAPYVKDKAEEWR
YTDRFLQGYSADGQIRAMNPTWRDEFINRYWGAFLFNEYGLFNAHSQGAREALSDVTRVSL
AFWGFDKIDIAQMIQLERGFLAKIVPGFDESTAVPKAEWTNGEVYKSARLAVEGLWQEVFD
WNESAFSVHAVYDALFGQFVRREFFQRLAPRFGDNLTPFFINQAQTYFQIAKQGVQDLYYNC
LGDDPEFSDYNRTVMRNWTGKWLEPTIAALRDFMGLFAKLPAGTTDKEEITASLYRVVDDW
IEDYASRIDFKADRDQIVKAVLAGLK

Individual results for analysis of putative MMOZ gel band

Top 10 hits:

P11987	Methane monooxygenase component A gamma chain OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233
P0AGD3	Superoxide dismutase [Fe] OS=Escherichia coli (strain K12) GN=sodB PE=1 SV=2
P62399	50S ribosomal protein L5 OS=Escherichia coli (strain K12) GN=rplE PE=1 SV=2
P0AE08	Alkyl hydroperoxide reductase subunit C OS=Escherichia coli (strain K12) GN=ahpC PE=1 SV=2
P0A6G7	ATP-dependent Clp protease proteolytic subunit OS=Escherichia coli (strain K12) GN=clpP PE=1 SV=1
P63224	Phosphoheptose isomerase OS=Escherichia coli (strain K12) GN=gmhA PE=1 SV=1
P0AG55	50S ribosomal protein L6 OS=Escherichia coli (strain K12) GN=rplF PE=1 SV=2
P0A7A9	Inorganic pyrophosphatase OS=Escherichia coli (strain K12) GN=ppa PE=1 SV=2
P0A7B8	ATP-dependent protease subunit HslV OS=Escherichia coli (strain K12) GN=hslV PE=1 SV=2
P69783	Glucose-specific phosphotransferase enzyme IIA component OS=Escherichia coli (strain K12) GN=crr PE=1 SV=2

Sequence results for P11987:

Mass: 19834 Score: 17213 Matches: 695(471) Sequences: 57(41) emPAI: 74067.35

Sequence coverage: 89% (151 / 170)

MAKLGIHSNDTRDAWVNKIAQLNTLEKAAEMLKQFRMDHTTPFRNSYELDNDYLWIEAKL
EEKVAVLKARAFNEVDFRHKTAFGEDAKSVLDGTVAKMNAAKDKWEAEKIHIGFRQAYKP
PIMPVNYFLDGERQLGTRLMELRNLNYYDTPLEELRKQRGVRVVHLQSPH

Individual results for analysis of putative MMOD gel band

Top 10 hits:

P22867	Methane monooxygenase component D OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233 GN=mmoD PE=
P22869	Methane monooxygenase component A alpha chain OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) OX=243233
P0A7K2	50S ribosomal protein L7/L12 OS=Escherichia coli (strain K12) GN=rplL PE=1 SV=2
P0A6X3	RNA-binding protein Hfq OS=Escherichia coli (strain K12) GN=hfq PE=1 SV=2
P0C018	50S ribosomal protein L18 OS=Escherichia coli (strain K12) GN=rplR PE=1 SV=1
P0ADZ7	UPF0092 membrane protein YajC OS=Escherichia coli (strain K12) GN=yajC PE=1 SV=1
P0A7R5	30S ribosomal protein S10 OS=Escherichia coli (strain K12) GN=rpsJ PE=1 SV=1
P0AG48	50S ribosomal protein L21 OS=Escherichia coli (strain K12) GN=rplU PE=1 SV=1
P0A7K6	50S ribosomal protein L19 OS=Escherichia coli (strain K12) GN=rplS PE=1 SV=2
P69776	Major outer membrane lipoprotein Lpp OS=Escherichia coli (strain K12) GN=lpp PE=1 SV=1

Sequence results for P22867:

Mass:11992 Score:19874 Matches:911(444) Sequences:78(49) emPAI:11932857087216.55

Sequence coverage : 97% (100 / 103)

MVESAFQPFSGDADEWFEEPRPQAGFFPSADWHLLKRDETYAAYAKDLDFMWRWVIVREE
RIVQEGCSISLESSIRAVTHVLNYFGMTEQRAPAEDRTGGVQH

Individual results for analysis of putative MMOG gel band

Note: Results were obtained on Ultimate 3000 nanao HPLC / MicroTOFq quadrupole TOF

Sequence results:

Sequence coverage: 59% (330 / 559)

MAKEVVYRGSARQRMMQGIEILARAAIPTLGATGPSVMIQHRADGLPPISTRDGVTVANSIV
LKDRVANLGARLLRDVAGTMSREAGDGTTTAIVLARHIAREMFKSLAVGADPIALKRGIDRA
VARVSEDIGARAWRGDKESVILGVAAVATKGEPGVGRLLLEALDAVGVHGAVSIELGQRRE
DLLDVVDGYRWEKGYLSPYFVTDRARELAELEDVYLLMTDREVVDFIDLVPLLEAVTEAGG
SLLIAADRVHEKALAGLLLNHVRGVFKAVAVTAPGFGDKRPNRLLDLAALTGGRAVLEAQG
DRLDRVTLADLGRVRRAVVSADDTALLGIPGTEASRARLEGLRLEAEQYRALKPGQGSATG
RLHELEEIEARIVGLSGKSAVYRVGGVTDVEMKERMVRIENAYRSVVSALEEGVLPGGGVGF
LGSMPVLAELEARDADEARGIGIVRSALTEPLRIIGENSGLSGEAVVAKVMDHANPGWGYDQ
ESGSFCDLHARGIWDAAKVLRLALEKAASVAGTFLTTEAVVLEIPDTDAFAGFSAEWAAATR
EDPRV