Engineering Membranes in *Escherichia coli*: The Magnetosome, LemA Protein Family and Outer Membrane Vesicles

Rokas Juodeikis

2016

A thesis submitted to the School of Biosciences, University of Kent for the title of PhD in Biochemistry.

Abstract

Magnetosomes are membranous organelles found in magnetotactic bacteria (MTB). The organelle consist of ferromagnetic crystals housed within a lipid bilayer chained together by an actin-like filament and allows MTB to orient within magnetic fields. The genetic information required to produce these organelles has been linked to four different operons, encoding for 30 genes. These membranous organelles and the magnetic minerals housed within have various biotechnological applications, therefore enhanced recombinant production of such structures in a model organism holds significant potential. The research described in this thesis is focuses on the production of recombinant magnetosomes in the model organism *Escherichia coli.*

Cloning the genes involved in the generation of the organelle individually or in various combinations resulted in the construction of over 100 different plasmids, compatible with the model organism. SDS-PAGE and electron microscopy analysis was used to characterise *E. coli* cells harbouring these constructs. The observation of electron dense particles, arranged in a chain structure, show that magnetosome generation in the model organism is possible, but is highly dependent on the growth conditions used. The need for specific growth conditions is later backed up by the analysis of the maturation of the cytochrome c proteins involved in magnetosome biomineralisation, which can only be correctly processed under certain conditions.

Individual production of two different magnetosome proteins, MamQ or MamY, allowed the generation of various membranous structures in *E. coli* observed in 48.9% and 56.2% of the whole population of cells respectively. Combinations of these with MamI, MamL or MamB in a variety of combinations led to a variation in the phenotype observed. Bioinformatics analysis of MamQ led to the discovery of a novel membrane restructuring protein family, the LemA protein family, present in a broad range of bacteria. Four different LemA proteins from *Bacillus megaterium, Clostridium kluyveri, Brucella melitensis* or *Pseudomonas aeruginosa* were then produced in *E. coli* and the analysis of the resulting strains revealed the presence of novel intracellular membranous structures which vary in size, form and localisation. Furthermore, when attempts were made to target these proteins for the modification of the outer membrane, a mechanism for increased outer membrane vesicle generation was serendipitously discovered and different effects of these proteins were once again observed.

Together, the results described shows good evidence for recombinant magnetosome production in *E. coli* and opens a new avenue of membrane engineering in this commonly used organism. Such membranous structures have various biotechnological applications, such as enhanced metabolic engineering potential or specialised lipid vesicle production.

Table of Contents

Figures

Tables

Abbreviations

CAI	Codon adaptation index
CDF	Cation diffusion facilitator
LB	Lysogeny Broth
MAI	Magnetosome island
MFS	Major facilitator superfamily
MTB	Magnetotactic bacteria
OMV	Outer membrane vesicle
RFP	Red Fluorescent protein
TMH	Transmembrane helix
TRP	Tetricopeptide

Acknowledgments

First of all, I would like to acknowledge my mother, Milda, who did everything in her power to enable me to pursuit my dreams in higher education.

I am extremely grateful to Professor Martin Warren and Professor Mark Smales not only for giving me the opportunity to work on this project, but also providing great guidance, encouragement and advice throughout the years.

During the course of my PhD at The University of Kent, I encountered a large number of incredible individuals all of whom contributed in some way to my experience here. Firstly, I would like to thank Evelyne Deery, who provided great support for all of the molecular biology work and Ian Brown who gave me excellent training and guidance for everything related to electron microscopy. Also, to my good friend Matthew Lee (aka. Matty Boo), for a great amount of discussion and support.

I would also like to thank all the members of the Warren Lab for creating an enjoyable work environment and providing advice. In particular, I would like to mention Joseph Baker, for being a 'bro', Emi Nemoto-Smith, for her infective positive attitude, Matthias Mayer, for encouraging me to 'do it now', David Palmer, for talking 'nerdy' with me, Maria Stanley, for only pretending to be a grown up, and the post-docs, Steffi Frank, Susanne Schroeder and Andrew Lawrence, for putting up with my silliness.

Furthermore, I thank all the wonderful people I met in the university during these four years, especially Linas Tamosaitis, Chris Miller and Ben Blakeman for being great friends.

Finally, I would like to thank all my friends and family in Lithuania for the warm welcome every time I find the time to go back home.

Chapter 1

Introduction

1. Introduction

For years, one of the defining features of prokaryotic organisms was the lack of any membranous organelles. Recently, this view has been changing with the characterisation of bacterial organelles such as chromatophores, the photosynthetic organelles found in photosynthetic bacteria, nuclear envelopes, present in the *Planctomycetes* species, outer membrane vesicles (OMVs), produced by a range of different gram-negative bacteria and magnetosomes, characteristic of a variety of magnetotactic bacteria (Saier and Bogdanov, 2013).

Magnetosomes (**Figure 1.1**) are bacterial organelles that consist of a lipid membrane and a crystalline magnetic mineral, which are arranged in a chain within the cytoplasm (Gorby *et al.,* 1988). In recent years, these organelles have been extensively studied as they provide an excellent model for bacterial compartmentalisation and biomineralisation. Furthermore, the magnetic nano-particles and the membranous organelles produced have potential use in a variety of biotechnological and medical applications, some of which are discussed later (Yan *et al.,* 2012).

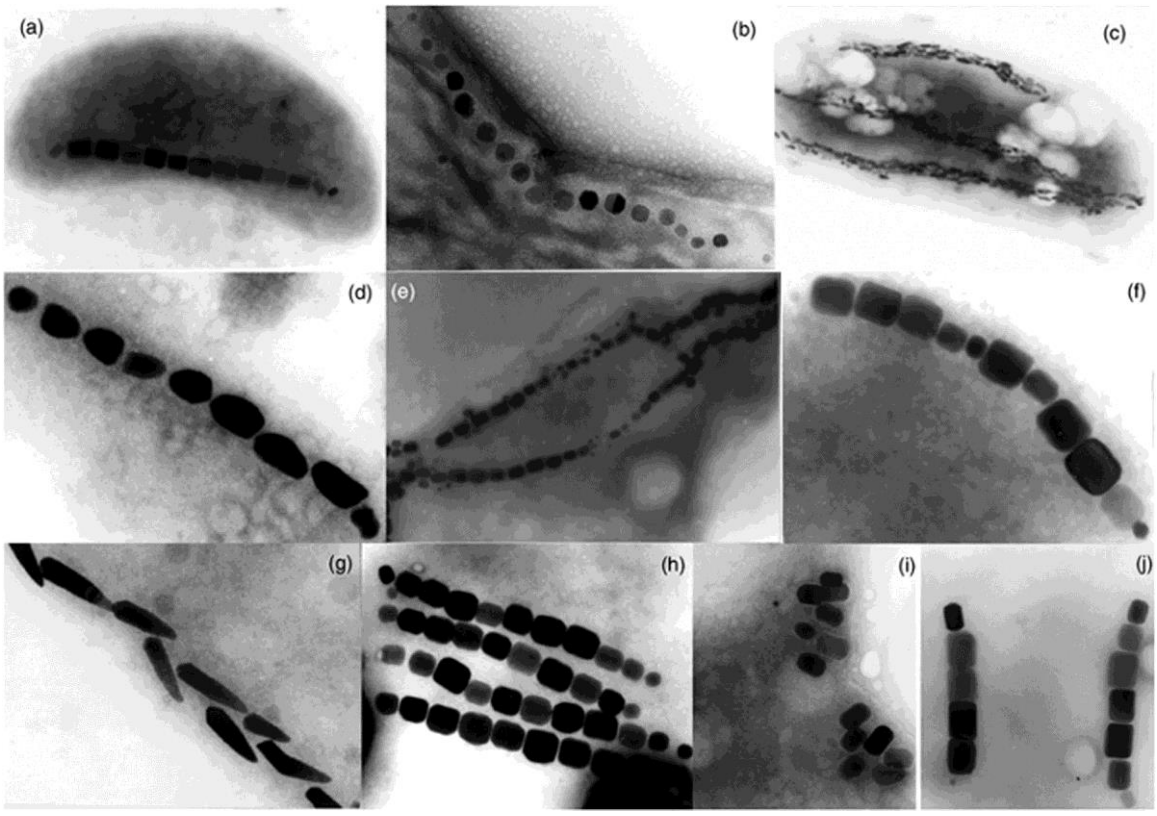


Figure 1.1. Transmission elecron micrographs of magnetosomes.
Transmission elecron micrographs of various arrangements of magnetosomes in different bacterial strains (Schuler, 2008).

1.1. Discovery of Magnetotactic Bacteria

The first observation of magnetotactic bacteria was made by Salvatore Bellini and is described in work carried out in 1963, which was distributed in small numbers through a variety of Italian universities. While observing drainage water and sediment from a pothole using light microscopy methods, Bellini noticed that the majority of microorganisms migrated to the same geographical direction. By varying the sample and environmental conditions, it was concluded that the organisms migrated towards magnetic north. Additional experimentation revealed that the introduction of an artificial magnetic field could alter this behaviour and was dependent on the availability of iron. In later work, Bellini also discovered that these organisms prefer micro-anaerobic conditions. Due to the growing interest in the field, the research carried out was later translated into English (Bellini, 2009a; b) thanks to the initiative taken by Richard Frankel (Frankel, 2009).

As Bellini's work was only published recently, the discovery of magnetotactic bacteria is usually attributed to Richard Blakemore (Blakemore, 1975). While analysing sediments from near Woods Hole, Massachusetts, Blakemore also identified a group of motile bacteria that migrated towards magnetic north. This directional movement was not altered by the positioning of the microscope or varying environmental conditions, such as the direction of light. As with Bellini, Blakemore noticed that the introduction of a magnetic field near the sample altered the migration of the cells. Furthermore, using electron microscopy and electron excitation analysis, Blakemore was able to show that these cells contained electron dense particles, consisting of iron, localised within membranous compartments and hypothesised that these crystal-like particles were a permanently magnetic mineral, such as magnetite.

1.2. Magnetotactic Bacteria

Magnetotactic bacteria are a group of diverse motile gram-negative bacteria that contain magnetosomes and grow in the oxic-anoxic transition zone, anaerobic zone or both in diverse aqueous environments (Bazylinski *et al.*, 1995; Simmons *et al.,* 2004; Lefevre and Bazylinski*,* 2013).

A number of magnetotactic bacteria have been cultured in axenic culture. Most of these are from the *Magnetospirillum* genus and include *M. magnetotacticum*, the first magnetotactic bacteria to be grown in a pure culture (Blakemore *et al.,* 1979; Maratea and Blakemore, 1981; Schleifer *et al.,* 1991), *M. gryphiswaldense* (Schleifer *et al.,* 1991) and *M. magneticum* (Matsanuga *et al*, 1991). A variety of other magnetotactic bacteria, which to date have not been fully characterised, have also been cultured in an axenic culture. These organisms include *Magnetovibrio blakemorei* (Bazylinski *et al.,* 1988; Bazylinski *et al.,* 2013), marine coccus MC-1 (Meldrum *et al.*, 1993; DeLong *et al.,* 1993), a marine spirillum MMS-1 (Meldrum *et al.*, 1993), *Desulfovibrio magneticus* (Sakaguchi *et al*, 1993, 2002), *Candidatus* Desulfamplus magnetomortis BW-1 (Lefevre *et al*., 2011a), *Magnetofaba australis* (Morillo *et al*., 2014) and a variety of obligately alkaliphilic strains (ML-1, ZZ-1 and AV-1) (Lefevre *et al.*, 2011b).

A number of magnetotactic bacteria have also been characterised without obtaining an axenic culture. These include a variety of multi-cellular prokaryotic organisms (Rodgers *et al.,* 1990; Keim *et al.*, 2004; Abreu *et al.,* 2006; 2007), incredibly large bacteria (up to 10 µm in length and up to 2 µm in diameter) (Jogler *et al.,* 2010) and thermophilic bacteria (Lefevre *et al.,* 2010).

Complete genome sequencing has been performed on some of these organisms including *M. magneticum* AMB-1 (Matsunaga *et al*., 2005), *D. magneticus* RS-1 (Nakazawa *et al.,* 2009), marine coccus MC-1 (Schubbe *et al.,* 2009), *M. gryphiswaldense* (Wang *et al.,* 2014) and, most recently, *M. magnetotacticum* MS-1 (Smalley *et al.,* 2015).

1.3. Magnetosomes

One of the defining features of magnetotactic bacteria is the presence of magnetosomes, intracellular membranous organelles, which contain magnetic iron crystals within their lumen (Gorby *et al.,* 1988). The term magnetosome was used due to the magnetic properties of the organelles (Balkwill *et al.,* 1980). These membranous organelles are further organised by alignment on actin-like filaments (Komeili *et al.*, 2006) via an anchoring protein (Scheffel *et al.,* 2006).

The magnetosome membrane has been extensively studied in the *Magnetospirillum* species and has been observed using regular electron microscopy methods (Gorby *et al.,* 1988; Komeili *et al.,* 2004) and, more recently, re-constructed using cryo-electron tomography methods (**Figure 1.2**) (Komeili *et al.,* 2006; Scheffel *et al.,* 2006). Due to the magnetic crystal present in these compartments, the magnetosome membranes can be easily isolated from other bacterial membranous content by density ultracentrifugation and magnetic separation (Grunberg *et al.,* 2001; 2004), which allows the analysis of the biochemical composition of this membrane. The lipid composition of the magnetosome membrane resembles that of the cytoplasmic membrane, although the relative amounts of these lipids differ with specific enrichement of C16:1 and C18:1 unsaturated fatty acids (Gorby *et al.,* 1988; Grunberg *et al.,* 2004; Tanaka *et al.,* 2006). The membrane-protein composition is however very specific for the magnetosome (Grunberg *et al.,* 2001; Tanaka *et al.,* 2006). Indeed, the identification of the proteins localised to the magnetosome membranes helped inform the identification of the operons involved in organelle formation and is discussed in the next section.

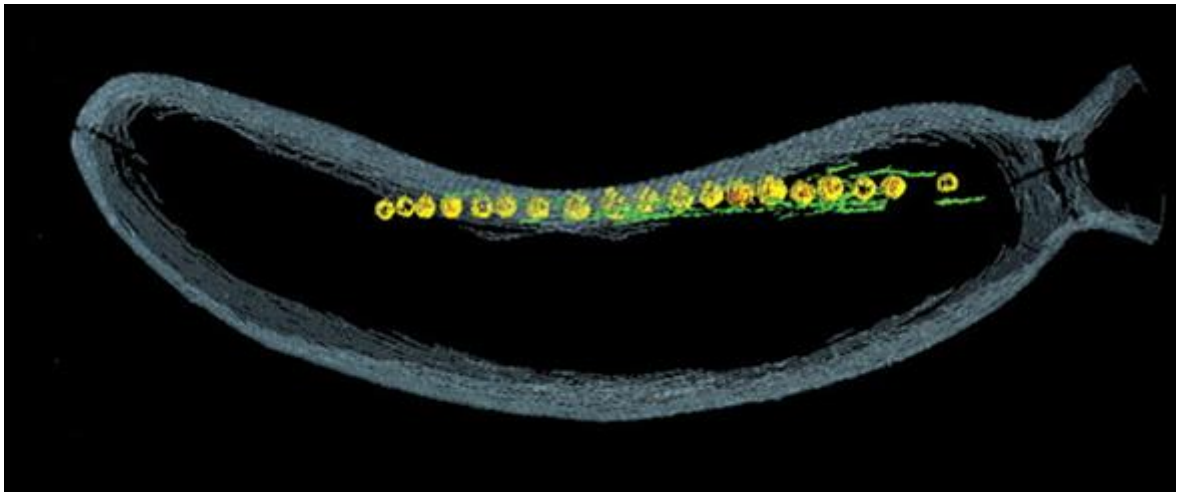


Figure 1.2. Three-dimensional reconstruction of *M. magneticum.*
Three-dimensional reconstruction of wild-type *M. magneticum* cell (Komeili *et al.*, 2006). Magnetosomes are shown in yellow and actin-like filaments are shown in green.

Depending on the organism producing these organelles, the ferromagnetic crystal housed within magnetosomes can be either magnetite (Fe_3O_4) or iron sulphide greigite (Fe_3S_4) (Bazylinski *et al.,* 1993). The size of the crystal is usually between 30 and 120 nm (Schuler, 2008), although much larger 250 nm crystals have been previously isolated (Lins *et al.,* 2006). Such crystals show single-domain permanent magnetism both when composed out of magnetite (Butler and Banerjee, 1975) or greigite (Diaz-Ricci and Kirschvink, 1992). At ambient temperatures smaller crystals would not be permanently magnetic, while larger crystals would have multiple magnetic domains with opposing magnetic orientations, reducing the total magnetism of the crystal (Schuler, 2008). This suggests the cells producing the organelles have evolved methods to control the size of the crystals within the narrow optimum range. Variation in the crystal morphology found within magnetosomes is also observed between different strains (**Figure 1.1**). Most of these morphologies are different to those obtained by chemical synthesis, suggesting a controlled mechanism for crystal growth within the organelles (Schuler, 2008).

1.4. Genes Involved in the Formation of the Organelle

The majority of genes involved in magnetosome formation were identified using reverse genetics. The initial experiment carried out identified the *mam22* gene (equivalent to *mamA* in *M. gryphiswaldense*) in *M. magneticum,* and showed that the gene product is involved in the formation of the organelle (Okuda *et al.,* 1996). Similar studies have been carried out in *M. gryphiswaldense* resulted in the identification of 22 further genes involved

in the process of magnetosome formation (Grunberg *et al.*, 2001; 2004). These genes were shown to localise to a genomic cluster referred to as the 'magnetosome island' (MAI) (Schubbe *et al*., 2003; Ullrich *et al*., 2005), as it shares features common to other genomic fragments described as 'pathogenicity' or 'environmental' islands (Dobrindt *et al.,* 2004). Such islands are known to be transferred between organisms via horizontal gene transfer. Similar genomic islands were also found in *M. magneticum* (Fukuda *et al*., 2006) and, to a lesser degree of conservation, in other magnetotactic organisms (Jogler and Schuler, 2009).

The MAI in *M. gryphiswaldense* is 130 kb (**Figure 1.3**) and has been subject to extensive genetic studies. One of the first of these studies involved the characterisation of a non-magnetic mutant strain, which had lost a 40 kb fragment from the MAI (Ullrich and Schuler, 2010). This fragment included three operons: *mamGFDC*, *mms6* and *mamAB*, which contained multiple genes identified by the reverse genetics approach. A previous study had shown that the *mamGFDC* operon was not necessary for magnetosome formation, but played a role in the size control of the magnetic crystal (Scheffel *et al.,* 2008). A study carried out in a related organism, *M. magneticum,* showed similar mineralisation defects when parts of *mamGFDC* and *mms6* operons were deleted (Lohbe *et al*., 2011). Furthermore, the research demonstrated that the deletion of the *mamAB* operon resulted in cells with no magnetosomes. Analysis of single deletions from the operon was also carried out, identifying four genes crucial for magnetosome membrane formation (*mamI*, *mamL*, *mamQ* and *mamB*) and four genes crucial for biomineralisation (*mamE*, *mamM*, *mamN* and *mamO*). Similar studies have also been carried out in *M. gryphiswaldense* (Lohbe *et al.,* 2011; 2014). In one of these studies, another operon from the MAI, *mamXY*, was also shown to be involved, but was not essential in the formation of the organelle. More recently, it has been shown that the recombinant expression of these operons (*mamGFDC, mms6, mamAB* and *mamYX*) in *Rhodospirillum rubrum* yields the functional production of magnetosomes (Kolinko *et al.,* 2014) and the overproduction of these operons in *M. gryphiswaldense* leads to a greatly increased number of the organelles within the cell (Lohbe *et al.,* 2016).

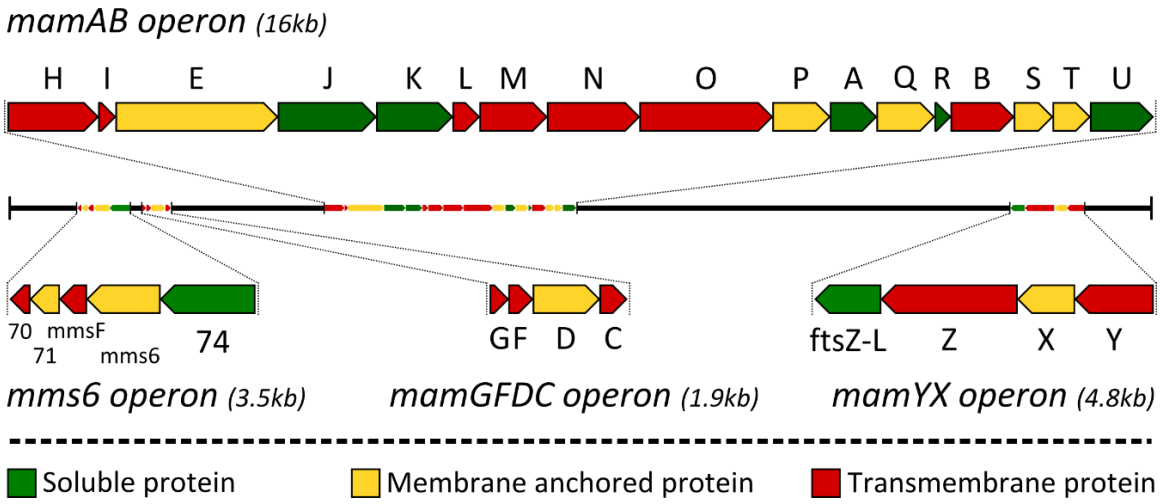


Figure 1.3. The magnetosome island.
The four operons depicted have been shown to be involved in magnetosome generation.

1.5. Mechanism of Magnetosome Formation

The magnetosome formation mechanism still remains unclear, although it has been well established that the proteins involved differ from those involved in eukaryotic organelle formation. There is good evidence to suggest a step-wise mechanism (**Figure 1.4**) (Murat *et al.*, 2010a; Lohbe *et al.,* 2011). These steps include membrane vesicle formation, crystal nucleation, organelle alignment and crystal maturation.

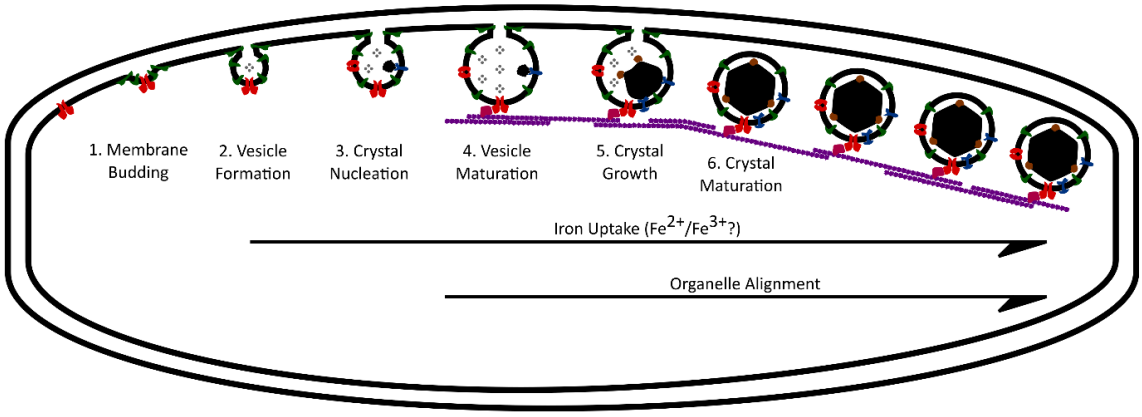


Figure 1.4. A schematic representation of magnetosome formation.
In red – initiatory protein; green – membrane restructuring proteins; blue – nucleation protein; purple – actin-like protein; gray – soluble iron; black – magnetite crystal; brown – crystal maturation proteins;

1.5.1. Vesicle Formation

Vesicle formation can further be broken down into three steps: membrane budding, vesicle formation and vesicle maturation. Recently, it has been shown that in *M. gryphiswaldense* only one protein, MamB, is necessary for magnetosome membrane formation (Raschdorf *et al.,* 2016). In a closely related organism, *M. magneticum*, it was shown that the deletion of one of four genes, including *mamB* (also *mamI*, *mamL* and *mamQ*), completely abolished magnetosome membrane formation (Murat *et al.,* 2010a). Interestingly, the deletion of *mamI* in *M. gryphiswaldense*, had little effect on magnetosome membrane formation, while the deletion of *mamL* or *mamQ* showed a decrease in organelle size, but did not completely abolish it (Raschdorf *et al.,* 2016). Significantly, the same study showed that co-expression of *mamB*, *mamQ*, *mamL* and *mamM* is not sufficient to form membranous vesicles. From these experiments it is clear that MamB plays a key role in membrane budding, but there are other factors involved, which may have homologous functions, therefore the deletion of one of these does not necessarily result in a null phenotype. Proteins involved in vesicle formation have been defined further, as recombinant expression of *mamLQBIEMO* has been shown to restore magnetosome membrane formation, but not mineralisation, in a *M. gryphiswaldense* strain lacking the *mamAB* operon (Raschdorf *et al*., 2016). It is worth noting that the expression of *mamLQRBM* did not rescue this phenotype and further constructs, containing different genes, were not examined. Recently, it has been shown that upon crystal nucleation the magnetosome membrane undergoes a further increase in size, which can be described as vesicle maturation (Cornejo *et al.*, 2016). The precise mechanism of this process and the proteins involved have still to be identified. Furthermore, all of these experiments were carried out in a strain that had functional *mamGFDC*, *mms6* and *mamXY* operons, so therefore their contribution to the process can not be omitted. One such example of a possible contributors is MamY, produced from the *mamXY* operon. Δ*mamY* strains produce larger magnetosome vesicles and the protein has been shown to induce liposome tubulation *in vitro* (Tanaka *et al*., 2010).

1.5.2. Crystal Nucleation

Crystal nucleation is the moment when the soluble iron becomes an insoluble mineral. From knock-out studies in *M. magneticum*, at least 4 proteins are likely to be involved in iron nucleation (MamE, MamM, MamN and MamO) (Murat *et al.,* 2010a), while in *M.*

gryphiswaldense it has been shown that MamE, MamM, MamO, MamB, MamQ and possibly MamL are essential for membrane formation and nucleation to take place (Lohbe *et al.,* 2014). As the majority of these are involved in the formation of the magnetosome membrane (see above), iron nucleation may be dependent on MamE, MamM and MamO, which are conserved between the two *Magnetospirillum* species.

MamE is a multi-domain protein, containing a protease domain, three magnetochrome domains and two structural PDZ domains, which are found in signalling proteins. Magnetochrome domains are cytochrome c domains with a unique arrangement, which is only found in magnetotactic bacteria (Siponen *et al.,* 2012). The protein has been suggested to be involved in protein localisation to the organelle (Murat *et al.,* 2010a). Interestingly, cells producing mutated variants of MamE, without a functional protease domain or the magnetochrome domains, are still able to mediate crystal nucleation, but the crystals observed do not exceed 20 nm (Quinlan *et al.,* 2011). This suggests that the protease activity and the haem-binding motifs are necessary for crystal growth, but are not involved in nucleation.

MamM and MamB belong to the cation diffusion facilitator (CDF) family. The members of this family are involved in metal ion homeostasis and generally transport divalent transition metal cations, including Fe^{2+} (Paulsen and Saier, 1997). Both of these proteins are conserved in various magnetotactic bacteria and are predicted to have multiple roles, including crystal nucleation (Uebe *et al.,* 2011). MamM has been shown to be necessary for biomineralisation and various mutations in the predicted active site of the transporter protein can limit this process. It also stabilises MamB, a protein involved in protein targeting to the organelle and iron import. Significantly, any attempted mutation in MamB abolished magnetosome formation completely. This further highlights the importance of this protein as a protein involved in all stages of organelle formation (Uebe *et al.,* 2011).

MamO has been recently shown to have a degenerative protease domain and to promote magnetosome formation by two non-catalytic activities (Hershey *et al.,* 2016). The first activity is the activation of MamE protease, which is not involved in crystal nucleation (discussed previously). The second activity involves transition metal binding, although this was only shown with Ni^{2+} due to technical limitations. The active site of the protein is large enough to easily accommodate Fe^{2+}, which is the suggested substrate. *M. magneticum* cells, producing MamO with a mutation in the metal binding site, shows no crystal

nucleation. This suggests that the metal binding site is necessary for crystal nucleation to take place.

Overall, it is clear that at least four proteins are involved in the nucleation step: MamE and MamB, playing a role in protein targeting, MamM, stabilising MamB, and MamO, via its metal-binding activity. The metal transport activities of MamB and MamM may also be involved, but have not been shown to be necessary for crystal nucleation and possibly only influence crystal growth.

1.5.3. Organelle Alignment

Magnetosome alignment is, arguably, the best understood stage of the organelle formation. This is due to relatively few proteins involved as miss-alignment of the organelles, to date, has only been observed in two mutant strains: Δ*mamK* and Δ*mamJ* (Komeili *et al*., 2006; Scheffel *et al*., 2006). MamK, a homolog of the bacterial actin-like protein MreB, forms filamentous structures *in vivo* (**Figure 1.2**), which bind magnetosomes via an acidic protein, MamJ, thereby aligning the organelles throughout the cell. Furthermore, MamJ has an additional regulatory effect on MamK (Draper *et al.,* 2011). The exact molecular mechanisms behind the interaction of MamJ with the organelle or MamK are still unclear.

1.5.4. Crystal Maturation

Crystal maturation, which includes crystal growth, is a complex biochemical process and involves the largest proportion of the magnetosome proteins. A knockout of almost any gene from the four operons, suggested to be involved in magnetosome formation, has an effect on this process (Scheffel *et al.,* 2008; Murat *et al.,* 2010a; 2012; Lohbe *et al.,* 2011; 2014).

One of these operons, *mms6*, contains five genes, of which four have been shown to be produced during magnetosome synthesis (Grunberg *et al.*, 2001; 2004). Two of these, *mms6* and *mmsF*, seem to promote crystal growth, while the other two, *mg4070* and *mg4071*, seem to inhibit it (Lohbe *et al.,* 2016). Mms6 is the best studied protein from the operon and is suggested to function by displaying a charged surface for specific iron binding (Staniland and Rawlings, 2016). This protein has been used for controlled chemical synthesis of magnetite, allowing the production cubo-octahedral magnetite crystals with a narrow size distribution (Amemiya, 2007). MmsF, a protein with a predicted homologous

function *in vivo*, has been shown to form nanostructures termed proteinosomes and can also be used to improve the chemical synthesis of magnetite particles (Rawlings *et al*., 2014). Interestingly, a *M. magneticum* mutant, lacking *mms6* and *mamGFDC* operons, has a similar mineralisation defect when compared to a Δ*mmsF* strain (Murat *et al.,* 2012a). Both of these phenotypes can be rescued, partially for the double operon mutant and fully for the single mutant, by recombinant expression of *mmsF.*

Deletion of the whole *mamGFDC* operon does not result in a non-magnetic phenotype, which is striking as combined the proteins expressed from the operon have been shown to make up around 35% of the whole magnetosome proteome (Grunberg *et al*., 2004; Scheffel *et al.,* 2008). The observed phenotype is a decrease in size (about 25%) and number of crystals produced. This phenotype could be recovered partially by recombinantly expressing any single gene from the operon and further by expressing two genes (Scheffel *et al.,* 2008). Practically full recovery could be observed when any three of the genes in any combination or the full operon was expressed recombinantly, suggesting interchangeable functionality. One of the proteins in the operon, MamF, has high similarity to MmsF (62.6% amino acid identity and 88.8% similarity in *M. gryphiswaldense*) and is also able to form proteainosome structures but do not improve chemical magnetite synthesis (Rawlings *et al*., 2014). Interestingly, even with such high similarity, it does not rescue the Δ*mmsF* phenotype in *M. magneticum* (Murat *et al.,* 2012a).

Multiple genes from the *mamAB* operon have been shown to be necessary for crystal maturation in *M. magneticum* and *M. gryphiswaldense* (Murat *et al*., 2010a; Lohbe *et al*., 2014). Slight differences are observed between the proteins involved in the different strains and, due to the relevance to this work, the discussion will be focused on the *M. gryphiswaldense* model organism. The proteins from the *mamAB* operon, suggested to be involved only in the maturation process of the organelle, are MamH, MamI, MamN, MamP, MamR, MamS and MamT (Raschdorf *et al.,* 2013; Lohbe *et al*., 2014).

MamH is a member of the major facilitator superfamily (MFS). This family of proteins are known to transport small solutes in response to chemiosmotic ion gradients (Pao *et al.,* 1998). Deletion of *mamH* results in a decrease of magnetite particle number and size (Raschdorf *et al.,* 2013). Although the function of any closely-related proteins to MamH is not known, there are examples of MFS members transporting chelated iron, which is the predicted function for MamH to date (Lesuisse *et al.*, 1998; Chatfield *et al.*, 2012).

In contrast to *M. magneticum,* where MamI plays a crucial role in magnetosome membrane generation (Murat *et al*., 2010a), it has been shown that in *M. gryphiswaldense* this protein only plays a role in magnetosome maturation (Lohbe *et al.,* 2014). Δ*mamI* cells were found to produce smaller (around 40% the size of the wild-type) electron-dense particles with highly irregular elongated morphologies and at a reduced frequency. Furthermore, the cells producing these particles showed no magnetic response and no magnetite was present in these cells, suggesting that MamI plays a crucial role in the generation of the magnetic mineral.

MamN has been suggested to function as an ATP-driven proton pump, regulating the pH within the organelle (Schuler, 2008). A knockout strain of *mamN* produces fewer (about a third) and smaller (around half the size of the wild-type) magnetite particles, which results in a reduced magnetic response (Lohbe *et al.,* 2014). As precipitation of magnetite produces a large amount of protons, leading to a change of pH in the organelle, their export may play an important, but not essential, role in the synthesis of the particles as the maintanence of the periplasmic pH is part of general cellular metabolism.

MamP is one of the magnetochrome proteins involved in magnetosome synthesis and the only one to have its structure solved (Siponen *et al.,* 2013). The reduced state of the protein is able to oxidise Fe^{2+} to magnetite *in vitro*. Δ*mamP M. gryphiswaldense* cells still produce magnetite crystals and retain a magnetic response (Lohbe *et al.,* 2014). These crystals are larger, but are often flanked by smaller, poorly crystalline electron dense particles.

MamR is a protein of unknown function, with a helix-turn-helix DNA binding motif. Strains, with the genetic knockout of the corresponding gene, have a very similar magnetic response when compared to the wild-type cells (Lohbe *et al.,* 2014). The resulting phenotype is a small, compared to other gene knockouts, (about 20%) reduction in the average magnetosome size and impairment of the chain assembly in some, but not all, of the cells observed by electron microscopy.

Deletion of *mamS* results in a reduction of magnetosome particle size, but the magnetic response or the frequency at which these particles are produced, is not impaired (Lohbe *et al.,* 2014). MamS contains no known putative domains and has no known significant homology to any protein outside of magnetotactic bacteria. It has amino acid sequence similarities with the magnetosome proteins MamE and MamX, but has no magnetochrome domains, a characteristic shared by both of these proteins.

As with the deletion of *mamS,* a Δ*mamT* strain results in an effect on the morphology of the magnetite crystal produced, but not the frequency of the crystals or the magnetic response (Lohbe *et al.,* 2014). Some of the crystals were similar to wild-type size and formed "pseudochains", but a larger proportion were much smaller, bringing the average size down to about 80% of the wild-type. MamT is one of four magnetochrome proteins suggested to play a role in magnetosome generation and is thought to play a similar role to MamP in redox balance of Fe^{2+} and Fe^{3+}, which is key for magnetite synthesis (Siponen *et al.,* 2012).

Some of the proteins produced from the *mamAB* operon, involved in the earlier stages of magnetosome synthesis, are likely to be involved in magnetosome maturation. A good example of this is MamE. As discussed previously, a mutation in the putative protease domain or the magnetochrome motifs of MamE does not impair crystal nucleation, but stops crystal growth at around 20 nm (Quinlan *et al.,* 2011). This suggests a requirement of a proteolytic activation for some of the maturation factors and a contribution of the MamE magnetochrome domains to crystal growth. Contributions made by other proteins involved in previous stages are difficult to study as genetic knockouts of these impair the organelle formation at earlier stages.

The final operon involved in magnetosome maturation is the *mamXY* operon which encodes four proteins. The deletion of the whole operon results in a decrease in the cellular magnetic response and a decrease in average magnetosome size (to about 50% of the wild-type) (Lohbe *et al*., 2011). Furthermore, electron microscopy analysis revealed a variation between the magnetosomes observed with some lacking any electron dense particles. Single deletion strains for all four proteins have been generated in the *Magnetospirillum* species. MamX contains two PDZ domains and two magnetochrome domains. Δ*mamX* strains*,* or strains producing MamX without functional magnetochrome domains, produce small, irregularly shaped particles together with wild-type like particles (Raschdorf *et al.,* 2013; Yang *et al*., 2013). These cells show a slight reduction in the magnetic response. Deletion of a non-related gene from the operon, *mamZ*, shows a similar phenotype. MamZ has two domains: a MFS domain, which is similar to that of MamH, and a ferric reductase domain (Raschdorf *et al.,* 2013). Interestingly, a double deletion of *mamZ* and *mamH* shows an additive effect on biomineralisation, suggesting an overlapping role for the two proteins. A further protein from the operon suggested to be involved in the process is FtsZ-like. It is a truncated version of a tubulin-like protein, FtsZ, a protein which forms a cytokinetic ring

structure, an essential component of the cell division apparatus (Bramhill and Thompson, 1994). Deletion of *ftsZ-like* leads to production of smaller, irregular magnetic particles with poor alignment (Ding *et al,* 2010). *Magnetospirillum* species produce a full length FtsZ, which most likely plays a role in cell division. Comparison of this full length FtsZ and the FtsZ-like protein has been previously carried out, showing that the two proteins behave in a very similar manner and possibly interact (Muller *et al.,* 2014). These findings show that FtsZ-like is involved in magnetosome formation, but it is unclear if it acts as a scaffolding protein or facilitates the correct division of the organelles during cell division. There is no evidence that the last protein from the operon, MamY, is involved in crystal maturation and is suggested to be involved in vesicle formation, as discussed previously.

Overall, crystal maturation is the most complex step of magnetosome synthesis, with proteins from all four operons, linked to magnetite synthesis, involved. There is also evidence that proteins involved in cellular metabolism play an important role. One example of this is the Cbb3 terminal oxidase complex. When the corresponding operon is deleted, *M. gryphiswaldense* cells not only have a reduction in growth, but also produce much smaller, poorly crystalline magnetosomes (Li *et al*., 2014).

1.5.5. Other Factors Involved in Magnetosome Biogenesis

There are two other proteins, produced from the *mamAB* operon, MamA and MamU, which do not seem to be directly involved in any of the stages of magnetosome synthesis. Deletion of *mamA* has a similar phenotype both in *M. magneticum* and *M. gryphiswaldense* (Komeili *et al.,* 2004; Lohbe *et al*., 2014). The observed phenotype is a reduction in the number of magnetosomes, but not in the size or the overall magnetic response of the cells. MamA has been shown to form homo-oligomers, which could be disrupted by a mutation in its putative tetratricopeptide (TPR) repeat motif (Zeytuni *et al.,* 2011). This motif is known to mediate protein-protein interactions for protein complexes (Zeytuni and Zarivach, 2012). It has been suggested that MamA oligomers localise to the outer-surface of magnetosomes, contributing to the stabilization of the organelle (Komeili *et al.,* 2004; Yamamoto *et al.,* 2010). MamU is the final protein produced in the *mamAB* operon. Interestingly, *mamU* knockout strains of both *M. magneticum* and *M. gryphiswaldense* show no observable phenotype for magnetosome formation (Komeili *et al.,* 2004; Lohbe *et al*., 2014). This suggests that the protein is either not involved in magnetosome synthesis

at all or is only necessary under certain growth conditions, which have not been tested during the knockout experiments.

1.6. Magnetosome Production

Production of magnetite particles by cultivating magnetotactic bacteria is still problematic, but large-scale production of magnetosomes using a chemostat culture has been demonstrated (Liu *et al.,* 2010). One of the challenges is the slow growth of these organisms therefore ways of improving magnetosome productivity have been investigated. The most successful of these is the introduction of multiple copies of one or all of the four operons involved in magnetosome generation in *M. gryphiswaldense* genome (Lohbe *et al.,* 2016). The study demonstrates that the specific duplication of the *mms6* operon results in a size increase of the magnetite crystals (~35%), while the duplication of all the operons thought to be involved in magnetosome synthesis (*mamGFDC, mamAB, mms6* and *mamXY*) results in a 2.2-fold increase of the amount of the crystals produced. A different approach is to produce magnetosomes in a more readily cultivable organism. This has been achieved in *Rhodospirillum rubrum* by genetic integration of the four genetic operons from *M. gryphiswaldense* mentioned above (Kolinko *et al.,* 2014). Interestingly, *E. coli* cells were used for conjugative transfer of the operons into *R. rubrum*, but no discussion is given of the effect the genetic information had on the donor cells, suggesting that the presence of wild-type magnetosome operons in *E. coli* does not result in magnetetosome generation.

1.7. Applications of Magnetosomes

Due to a number of desirable physical properties, such as ferrimagnetism, narrow size distribution, nanoscale size and intramembranous environment, magnetic particles produced by magnetotactic bacteria have been investigated for various applications (Yan *et al*., 2012). The first known example of this is the immobilization of glucose oxidase and uricase on magnetosome membranes, which resulted in an increased activity for both of the enzymes (Matsunaga and Kamiya, 1987). Magnetite particles, produced by magnetotactic bacteria, have been used for a variety of applications (Alphandery, 2014). Some examples of these include novel methods for detection of mRNAs (Sode *et al.,* 1993), single nucleotide polymorphisms (Ota *et al.,* 2003; Tanaka *et al.,* 2003) and biomolecules (Amemiya *et al.,* 2005), DNA extraction (Yoza *et al.,* 2003) and an efficient way of gene vaccine (Tang *et al.,* 2007) or drug delivery (Sun *et al.,* 2008). Furthermore, proteins,

involved in the synthesis of magnetosomes, have been used to improve the chemical synthesis of magnetic particles (Staniland and Rawlings, 2016).

1.8. Aims of the Project

Escherichia coli is one of the most widely used microorganisms for a variety of biotechnological applications (Blount, 2015). Recent work has shown that this host can be used for the production of recombinant proteinaceous organelles, the bacterial microcompartments (Parsons *et al.,* 2010). These compartments can then be engineered to produce other chemicals, such as ethanol (Lawrence *et al.,* 2014). The ability to produce recombinant membranous organelles in *E. coli* could have a number of biotechnological applications especially for recombinant membrane protein production, pathway engineering or as vesicles for specialised vaccine production. Complete recombinant magnetosomes in *E. coli* would also represent a major advance for enhanced nano-magnetic bead production and could help facilitate an easier study of the organelle formation. Furthermore, an improved understanding of the biomineralisation process would have applications in bioremediation, especially for the recovery of precious metals or metals from metal-contaminated sites. The overall aim of the project was therefore to investigate if the genes involved in magnetosome synthesis could be cloned into *E. coli* and to determine what phenotype this would produce within the host. Within this broad aim, specific objectives include the characterisation of the proteins involved in the generation of the membranous invagination and biomineralisation. The ambitious nature of the aim coupled with the large number of genes involved (30 in total) with the process means that this is a technically challenging synthetic biology project involving a broad range of biochemical and imaging techniques.

Chapter 2

Materials and Methods

2.1. Chemicals

Chemical	Supplier
2-Mercaptoethanol	Sigma-Aldrich
5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue	Sigma-Aldrich
Acetic acid	Fischer Scientific
Agar (bacteriological)	Oxoid
Agar LV resin	Agar Scientific
Agarose	Alpha Laboratories
Ampicillin Sodium Salt	Melford
B-PER™	Thermo Fisher Scientific
Biotin	Sigma-Aldrich
Bromophenol Blue	Sigma-Aldrich
$CaCl_2$ x 2 H_2O	Sigma-Aldrich
Cacodylate	Sigma-Aldrich
Chelating Sepharose Fast Flow	GE Healthcare
Chloramphenicol	Sigma-Aldrich
Coomassie brilliant blue	Sigma-Aldrich
$CoSO_4$ x 7 H_2O	Sigma-Aldrich
$CuSO_4$ x 5 H_2O	BDH Laboratory Supplies
D-Ca-pantothenate	Sigma-Aldrich
Dithionite	Sigma-Aldrich
Ethanol	Fischer Scientific
Ethidium Bromide	Fischer Scientific
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
$FeCl_3$ x 6 H_2O	Sigma-Aldrich
$FeSO_4$ x 7 H_2O	Sigma-Aldrich
Folic acid	Sigma-Aldrich
Glycerol	Fisher Scientific
Glycine	Sigma-Aldrich
H_3BO_3	Sigma-Aldrich
HCl	Fisher Scientific
Imidazole	Sigma-Aldrich
Isopropyl β-D-1-thiogalactopyranoside	Melford
$KAl(SO_4)_2$ x 12 H_2O	Sigma-Aldrich

Kanamycin	Sigma-Aldrich
KCl	Sigma-Aldrich
KH_2PO_4	Sigma-Aldrich
L(+)-Tartaric acid	Sigma-Aldrich
Lead citrate	Agar Scientific
Lipoic acid	Sigma-Aldrich
Methanol	Fisher Scientific
$MgSO_4$ x 7 H_2O	Sigma-Aldrich
$MnSO_4$ x H_2O	Sigma-Aldrich
Na_2HPO_4	Sigma-Aldrich
Na_2MoO_4 x 2 H_2O	Sigma-Aldrich
Na_2SeO_3 x 5 H_2O	Sigma-Aldrich
Na-acetate	Sigma-Aldrich
NaCl	Fisher Scientific
$NaNO_3$	Sigma-Aldrich
NaOH	Sigma-Aldrich
$NiCl_2$ x 6 H_2O	Sigma-Aldrich
Nicotinic acid	Sigma-Aldrich
$NiSO_4$	Sigma-Aldrich
Nitrilotriacetic acid	Sigma-Aldrich
p-Aminobenzoic acid	Sigma-Aldrich
Pyridoxine-HCl	Sigma-Aldrich
Quinic acid	Sigma-Aldrich
Resazurin	Sigma-Aldrich
Riboflavin	Sigma-Aldrich
Skimmed milk powder	Oxoid
Sodium dodecyl sulfate	Fisher Scientific
Succinic acid	Sigma-Aldrich
TAE	Fisher Scientific
Thiamine-HCl x 2 H_2O	Sigma-Aldrich
Trichloroacetic acid	Sigma-Aldrich
Tris	Fisher Scientific
Tryptone	Fisher Scientific
Uranyl acetate	Agar Scientific
Vitamin B_{12}	Sigma-Aldrich

Yeast Extract	Fisher Scientific
$ZnSO_4$ x 7 H_2O	Sigma-Aldrich

2.2. Media and Solutions

2.2.1. Sterilisation

2.2.1.1. Autoclaving

Sterilisation for 15 min at 121 °C and 15 psi pressure.

2.2.1.2. Filter Sterilisation

Sterilisation achieved by filtration through a 0.2 µm filter next to a Bunsen burner.

2.2.2. Antibiotic Concentrations

Antibiotic	Concentration		Solution
	Stock	Working	
Ampicillin (Sodium Salt)	100 mg/ml	0.1 mg/ml	1:1 EtOH:ddH_2O
Chloramphenicol	34 mg/ml	0.034 mg/ml	EtOH
Kanamycin	50 mg/ml	0.05 mg/ml	ddH_2O

Ampicillin and kanamycin were sterilised by filter sterilisation.

2.2.3. Growth Media

2.2.3.1. Lysogeny Broth (LB)

Chemical	Concentration in dH_2O
NaCl	5 g/l
Yeast Extract	5 g/l
Tryptone	10 g/l

Sterilised by autoclaving.

2.2.3.2. LB - Agar

LB media with 15 g/l agar (bacteriological) added prior to sterilisation.

Sterilised by autoclaving.

2.2.3.3. Magnetospirillum Media

Ingredient	Concentration in dH_2O
Vitamin solution	10 ml/l
Trace elements	5 ml/l
Fe(III) quinate solution	2 ml/l
Resazurin	0.50 mg/l
KH_2PO_4	0.68 g/l
$NaNO_3$	0.12 g/l
L(+)-Tartaric acid	0.37 g/l
Succinic acid	0.37 g/l
Na-acetate	0.05 g/l

Ingredients were dissolved in the order provided and the pH was adjusted to 6.75 with NaOH.

For anaerobic conditions, the media was dispensed in anoxic vials (up to 2/3 of the volume), sealed with rubber closures and N_2 gas was bubbled via a needle for 5min.

Sterilised by autoclaving.

Filter sterilised Na-thioglycolate was added prior to inoculation to a final concentration of 0.05 g/l.

Small amounts of oxygen are added with Na-thioglycolate and the inoculum, providing micro-anaerobic conditions.

2.2.3.3.1. Trace Element Solution

Chemical	Concentration in dH_2O
Nitrilotriacetic acid	1.5 g/l
$MgSO_4 \times 7\ H_2O$	3 g/l
$MnSO_4 \times H_2O$	0.5 g/l
NaCl	1 g/l
$FeSO_4 \times 7\ H_2O$	0.1 g/l
$CoSO_4 \times 7\ H_2O$	0.18 g/l
$CaCl_2 \times 2\ H_2O$	0.1 g/l
$ZnSO_4 \times 7\ H_2O$	0.18 g/l
$CuSO_4 \times 5\ H_2O$	0.01 g/l
$KAl(SO_4)_2 \times 12\ H_2O$	0.02 g/l
H_3BO_3	0.01 g/l
$Na_2MoO_4 \times 2\ H_2O$	0.01 g/l
$NiCl_2 \times 6\ H_2O$	0.03 g/l
$Na_2SeO_3 \times 5\ H_2O$	0.3 mg/l

Sterilised by autoclaving.

2.2.3.3.2. Vitamin Solution

Chemical	Concentration in dH_2O
Biotin	2 mg/l
Folic acid	2 mg/l
Pyridoxine-HCl	10 mg/l
Thiamine-HCl $\times 2\ H_2O$	5 mg/l
Riboflavin	5 mg/l
Nicotinic acid	5 mg/l
D-Ca-pantothenate	5 mg/l
Vitamin B_{12}	0.1 mg/l
p-Aminobenzoic acid	5 mg/l
Lipoic acid	5 mg/l

Sterilised by autoclaving.

2.2.3.3.3. Fe(III) Quinate Solution

Chemical	Concentration in dH_2O
$FeCl_3$ x 6 H_2O	0.45 g/l
Quinic acid	0.19 g/l

Sterilised by autoclaving.

2.2.4. Media for Protein Work

2.2.4.1. Protein Purification Solutions

Buffer	Concentration of chemicals in dH_2O				
	$NiSO_4$	Tris-HCl (pH 8.0)	NaCl	Imidazole	EDTA
Charge	0.1% w/v	-	-	-	-
Binding	-	20 mM	100 mM	5 mM	-
Wash I	-	20 mM	100 mM	50 mM	-
Wash II	-	20 mM	100 mM	100 mM	-
Elution	-	20 mM	100 mM	400 mM	-
PD10	-	20 mM	100 mM	-	-
Strip	-	20 mM	100 mM	-	100 mM

2.2.4.2. 2x Laemmli Buffer

Chemical	Concentration in dH_2O
Tris-HCl (pH 6.8)	125 mM
Glycerol	0.2 ml/ml
Sodium dodecyl sulfate	0.04 g/l
2-Mercaptoethanol	0.1 ml/ml
Bromophenol Blue	0.4 mg/ml

2.2.4.3. Coomassie Blue Stain

Chemical	Concentration in dH_2O
Trichloroacetic acid	250 ml/l
Coomassie brilliant blue	0.6 g/l
Sodium dodecyl sulfate	0.1 g/l
Tris	0.25 g/l
Glycine	0.15 g/l

2.2.4.4. Western Blotting solutions

2.2.4.4.1. Transfer Buffer

Chemical	Concentration in dH_2O
Tris	3.03 g/l
Glycine	14.41 g/l
Methanol	200 ml/l

2.2.4.4.2. Phosphate Buffered Saline

Chemical	Concentration in dH_2O
NaCl	8.18 g/l
KCl	0.22 g/l
Na_2HPO_4	1.42 g/l
KH_2PO_4	0.27 g/l

2.2.4.4.3. Phosphate-free Solution

Chemical	Concentration in dH_2O
NaCl	8.18 g/l
Tris	6.06 g/l

Ingredients were dissolved in the order provided and the pH was adjusted to 7.5 with HCl.

2.3. Microbiology

2.3.1. Bacterial Strains

Strain	Description	Function
E. coli BL21 (DE3) Star	F^- ompT hsdSB ($r_B^-m_B^-$) gal dcm rne131 (DE3)	Protein production
E. coli JM109	endA1 glnV44 thi-1 relA1 gyrA96 recA1 $mcrB^+$ Δ(lac-proAB) e14- [F' traD36 $proAB^+$ $lacI^q$ lacZΔM15] hsdR17($r_{K0}^-m_K^+$)	Cloning
E. coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ80dlacZΔM15 Δ(lacZYA-argF)U169 hsdR17($r_K^-m_K^+$), λ^-	Cloning
E. coli DH10β	F^- endA1 $deoR^+$ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 ϕ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) StrR λ^-	Cloning

2.3.2. Bacterial Growth

2.3.2.1. LB Liquid Media

Liquid cultures with appropriate antibiotics were inoculated to a starting OD of 0.01 from overnight starter cultures and grown at 37 °C with shaking (160 rpm).

2.3.2.2. LB Agar Plates

Bacteria were spread out on LB-agar plates with appropriate antibiotics and incubated overnight at 37 °C.

2.3.2.3. Overproduction of Individual Proteins for SDS-PAGE

E. coli BL21 (DE3) Star competent cells were transformed with a plasmid containing the gene of interest. 50 ml of LB with appropriate antibiotics was inoculated with 3 colonies from the transformation plate. The culture was grown at 30°C with shaking for 6 hours, protein production was induced by the addition of IPTG (1 M stock, filter-sterilised) to a final concentration of 100 µM. The cultures were then incubated overnight at 19°C with shaking. Cells were harvested by centrifugation at 3320 x g for 20 minutes at 4 °C. Cells

were lysed and fractionated to soluble and insoluble fractions using B-PER™ standard protocol.

2.3.2.4. Overproduction of Cytochrome C Proteins

E. coli BL21 (DE3) Star competent cells, with or without pEC86, were transformed with a plasmid containing the gene of interest. 5 ml of LB with appropriate antibiotics was inoculated with three colonies from the transformation plate and grown overnight at 37 °C with shaking (160 rpm). 1 L of LB with appropriate antibiotics was then inoculated with 1 ml of the overnight culture. The culture was grown at 30°C with shaking for 8 hours, protein production was induced by the addition of IPTG (1M stock, filter-sterilised) to a final concentration of 100 µM. The cultures were then incubated overnight at 19°C with shaking. Cells were harvested by centrifugation at 3320 x g for 20 minutes at 4 °C.

2.3.3. Preparation of Competent Cells

50 ml of LB in a conical baffled flask was inoculated with 3 colonies from a fresh overnight LB-agar plate. Cells were grown at 37°C with shaking to an OD600 of ~ 0.6. Cells were incubated on ice for 1 hour, then centrifuged at 2,700 x g for 10 minutes at 4°C. Pelleted cells were resuspended in 10ml 0.1M $CaCl_2$ (filter sterilised and pre-cooled) stored on ice for 1 hour. Following centrifugation at 2,700 x g for 10 minutes at 4 °C the cells were resuspended in 1ml 0.1 M $CaCl_2$ with 10% glycerol (filter sterilised and pre-cooled). Cells were then aliquoted into 100µl aliquots in sterile 1.5ml eppendorf tubes, flash frozen on dry-ice and stored at -80 °C.

2.3.4. Transformation of Competent Cells

0.2 µl of plasmid DNA was added to 20 µl of competent cells for single transformations. For ligations 5 µl of the ligation was used with 50 µl of competent cells. The cells were then incubated on ice for 30 minutes, heat shocked at 42°C for 1 minute and incubated on ice for a further 2 minutes. 200 µl of LB was then added and the cells were incubated at 37°C for 1 hour. Cells were subsequently plated on LB agar plates with appropriate antibiotics and incubated overnight at 37°C.

2.4. Molecular Biology

2.4.1. Primers

All primers were supplied by Eurofins™.

Name	Gene	Primer	Restriction site
Mg4070nde2	4070	CTTCATATGCTATTACGCCTGATCGTC	NdeI (CATATG)
Mg4070spe	4070	GCTACTAGTCATGTACTGCGGAACAG	SpeI (ACTAGT)
Mg4071nde	4071	CTGCATATGGACATCAACGAAAAGGCAC	NdeI (CATATG)
Mg4071spe2	4071	GTCACTAGTAATAGCATGGATCACTCGTCTC	SpeI (ACTAGT)
MgMmsFnde	mmsF	CTGCATATGAAGAAGTCGAACTGCGCGAC	NdeI (CATATG)
MgMmsFspe	mmsF	GTCACTAGTCAGATCCGGTCGGCCAC	SpeI (ACTAGT)
MgMms6nde	mms6	GTACATATGGTTTGCCCCCCTGGGGTTCC	NdeI (CATATG)
MgMms6spe	mms6	CATACTAGTTCAGGACAGCGCGTCGCACAG	SpeI (ACTAGT)
Mg4074nde	4074	GCTCATATGGGCTTGTGGTTTTGG	NdeI (CATATG)
Mg4074spe	4074	GTCACTAGTCAATCAAGTAGTGCGGGACTG	SpeI (ACTAGT)
MgMamGnde2	mamG	GCTCATATGATCAAGGGCATCGCGGGAG	NdeI (CATATG)
MgMamGspe	mamG	GTCACTAGTTAAGCAGGCTCGGCGGAG	SpeI (ACTAGT)
MgMamFnde	mamF	CTGCATATGGCCGAGACTATTTTGATC	NdeI (CATATG)
MgMamFspe	mamF	GTCACTAGTCAGATCAGGGCGACTAC	SpeI (ACTAGT)
MgMamDnde2	mamD	GCTCATATGCAGGACCTTTTTCTCGCCAAG	NdeI (CATATG)
MgMamDspe	mamD	GTCACTAGTTATTCCTCGCCGACAGCCGCCAG	SpeI (ACTAGT)

MgMamCnde	mamC	GCACATATGAGCTTTCAACTTGCGCCGTAC	NdeI (CATATG)
MgMamCspe2	mamC	GCTACTAGTCAGGCCAATTCTTCCCTCAG	SpeI (ACTAGT)
MgMamHase	mamH	TGCGATTAATATGGAACCTGGCAGATCAGAAGTTG	AseI (ATTAAT)
MgMamHspe	mamH	CGCTACTAGTGATACAGAACACAAGC	SpeI (ACTAGT)
MgMamIATGnde	mamI	GCAGCATATGCCAAGCGTGATTTTCG	NdeI (CATATG)
MgMamIspe	mamI	CGCTACTAGTCAACCATCGATGTTAGG	SpeI (ACTAGT)
MgMamEnde2	mamE	GCTCATATGACCATGTTCAATGGTGATGTGGAAGACG	NdeI (CATATG)
MgMamEspe2	mamE	GCTACTAGTCAAAGAACAATCCAGAACTCTTGGCCATTGC	SpeI (ACTAGT)
MgMamEsolSac	mamE	CATGAGCTCATGGAGCAGATGACAGGCGCACG	SacI (GAGCTC)
MgMamExhoSpe	mamE	GATACTAGTCTCGAGAAGAACAATCCAGAACTCTTGGCCATTGC	XhoI (CTCGAG); SpeI (ACTAGT)
MgMamJase	mamJ	GCATCATATGGCAAAAAACCGGCGTG	NdeI (CATATG)
MgMamJspe	mamJ	GACGACTAGTCCACAGGTCTCTATTTATTC	SpeI (ACTAGT)
MgMamKnde2	mamK	GTGCATATGTGGATTGATCTGTTAGCACG	NdeI (CATATG)
MgMamKspe	mamK	GACGACTAGTCACTGACCGGAAACTGC	SpeI (ACTAGT)
NewMgMamLase	mamL	GTGATTAATATGGTAAGAGTGATCGGATC	AseI (ATTAAT)
MgMamLspe	mamL	GATCACTAGTCCGACGGAGCATGGAATG	SpeI (ACTAGT)
MgMamMnde	mamM	CGACATATGAGGAAGAGCGGTTG	NdeI (CATATG)
MgMamMspe	mamM	GTCACTAGTAAATCCAACCACCTAG	SpeI (ACTAGT)

MgMamMxho	mamM	CAGCTCGAGGTTATCCACCTTGGACAGCATG	XhoI (CTCGAG)
MgMamNnde	mamN	CGTCATATGGTTGGATTTATCACCCTCGCTGTG	NdeI (CATATG)
MgMamNspe	mamN	GCTACTAGTCATCCTGCGAGAACGGCGATGTAC	SpeI (ACTAGT)
MgMamOase	mamO	CGTGATTAATATGATTGAAATTGGCGAGACCATGGGTG	AseI (ATTAAT)
MgMamOspe	mamO	CATACTAGTCACACCGTTGTCAGCATC	SpeI (ACTAGT)
MgMamPnde	mamP	CGACATATGAATAGCAAACTCGTCCTG	NdeI (CATATG)
MgMamPspe	mamP	GTCACTAGTGGCTAATTTATCACGTGG	SpeI (ACTAGT)
MgMamPsolSac	mamP	CATGAGCTCGTTGCTCCTCAGTCGATCAG	SacI (GAGCTC)
MgMamAnde	mamA	CGACATATGTCTAGCAAGCCGTC	NdeI (CATATG)
MgMamAspe	mamA	CAGACTAGTACATCGACTGCTTAG	SpeI (ACTAGT)
MgMamQase	mamQ	GCTGATTAATATGGCAGTAAGCGATGC	AseI (ATTAAT)
MgMamQspe	mamQ	GTGCACTAGTCAATTCTTGGATTCCTG	SpeI (ACTAGT)
MgMamQ-sSac	mamQ	GTCGAGCTCATTCTTGGATTCCTGCGAATG	SacI (GAGCTC)
MgMamRnde	mamR	GCACATATGACCTTTGTTCAGGGCGCCATG	NdeI (CATATG)
MgMamRspe	mamR	GAGACTAGTCATCGGTTCATGTATTCCAC	SpeI (ACTAGT)
MgMamBnde	mamB	GCCGCATATGAAGTTCGAAAATTGCAG	NdeI (CATATG)
MgMamBspe	mamB	CATTACTAGTGATCAGACCCGGACCGT	SpeI (ACTAGT)
MgMamBxho	mamB	GTTCTCGAGGACGAACAGGCGGATATCTC	XhoI (CTCGAG)
MgMamSnde	mamS	GCTCATATGGACTTTCGGCCTGATC	NdeI (CATATG)

MgMamSspe	mamS	CAGACTAGTCACTGCACGGTCATC	SpeI (ACTAGT)
MgMamTnde	mamT	GTTCATATGGGTACGCCAGG	NdeI (CATATG)
MgMamTspe	mamT	GTCACTAGTACCGGCGCTTAC	SpeI (ACTAGT)
MgMamTsolSac	mamT	CATGAGCTCTGGGATGAGCTGTCCCTC	
MgMamUnde	mamU	GTGCATATGCGGATCGCCGCAATC	NdeI (CATATG)
MgMamUspe	mamU	CTATACTAGTGGCGGCGTTATTTCGGAACCAGTATGG	SpeI (ACTAGT)
MgMamYnde	mamY	GCTCATATGTTGATGAACTTTGTCAACAATG	NdeI (CATATG)
MgMamYspe	mamY	CACACTAGTCACGCATCGGAGATGG	SpeI (ACTAGT)
MgMamXnde	mamX	CAGCATATGAACACCAAAGCCGTTGC	NdeI (CATATG)
MgMamXspe	mamX	CAGACTAGTTACCGCTCTTCGGCATCC	SpeI (ACTAGT)
MgMamXsolNdeSac	mamX	GATCATATGGAGCTCGCGCCTCCCATTGTC	SacI (GAGCTC); NdeI (CATATG)
MgMamZnde	mamZ	CTACATATGACCGTGGGCTCTCTG	NdeI (CATATG)
MgMamZspe	mamZ	CAGACTAGTACCATTTAGCCGATTCGCAG	SpeI (ACTAGT)
MgFtsZmNde	ftsZ-like	TGTCATATGATCGGCGTCGGTGG	NdeI (CATATG)
MgFtsZmspe	ftsZ-like	GCTACTAGTCAGGCGATACCGGTGG	SpeI (ACTAGT)
MgMagAnde	magA	GCTCATATGGAACTGCATCACCCCGAAC	NdeI (CATATG)
MgMagAspe	magA	CACACTAGTCAGGCATCGTCGGACCTC	SpeI (ACTAGT)
3a vector BglII FW	-	CAAGATCTCGATCCCGCGAAATTAATACG	BglII (AGATCT)
EcBtuFsig1Rev	btuF	CGTACTAGTATTGAGCTCGGCGGGAGAAAGCGTGATGACG	SpeI (ACTAGT);

			SacI (GAGCTC)
LemA.153.FW	LemA	GAGCATATGAAGAAAAGAGGTAGTACATTG	NdeI (CATATG)
LemA.153.RV	LemA	CATACTAGTCACTTATCGCTCCCAAAATCAAC	SpeI (ACTAGT)
LemA.159.FW	LemA	GAGCATATGGATCTTTTAACATCTATTATACTATC	NdeI (CATATG)
LemA.159.RV	LemA	GAGACTAGTTAAAATTGAATCTTTACATCGGCCTTTTTC	SpeI (ACTAGT)
LemA.501.FW	LemA	GTACATATGACCGCTCAGACGGTTGCCAC	NdeI (CATATG)
LemA.501.RV	LemA	GACACTAGTCAGTTGAAGTTGACCTTAG	SpeI (ACTAGT)
LemA.565.FW	LemA	GTACATATGAGTCTGACCGCTATCGCTTTC	NdeI (CATATG)
LemA.565.RV	LemA	GTAACTAGTCAGCCGAACAGGGCCTTGAG	SpeI (ACTAGT)
COCO2.KpnI.FW	-	CATGGTACCAGATCTTCATATTCATG	KpnI (GGTACC)
COCO2.NcoI.RV	-	CATCCATGGTCCCAGACTAATAATCAG	NcoI (CCATGG)
COCO2.Mut.FW	-	CATCAAGAACAAGTTTAAGCTC	-
COCO2.Mut.RV	-	GAGCTTAAACTTGTTCTTGATG	-

2.4.2. Plasmids

2.4.2.1. Plasmids Obtained From Other Sources

Name	Resistance	Description
pET3a	Ampicillin	pET3a vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pETlac	Ampicillin	pETlac vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pET3a.TBAD	Ampicillin	pET3a.TBAD vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pET3a.pRha	Ampicillin	pET3a.pRha vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pET14b	Ampicillin	pET14b vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pEC86	Chloramphenicol	Cytochrome C maturation plasmid containing the *Ccm* operon from *E. coli* expressed form a constitutive promoter. Obtained from Professor Stuart Fergusons laboratory, University of Oxford.
pZS24	Kanamycin	pZS24 vector plasmid obtained from Dr. Evelyne Deery, University of Kent.
pETcoco2	Ampicillin	pETcoco2 vector plasmid obtained from Dr. Evelyne Deery, University of Kent.

2.4.2.2. Plasmids Constructed as Part of This Work

Name	Resistance	Description
pET.cocoR	Ampicillin	pETcoco2 vector with a removed *SpeI* site in the parB gene.
pET3a.mamH-N	Ampicillin	*mamHIEJKLMN* expressed from a T7 promoter. *mamH* has the ribosome binding site from the pET3a vector, while all other genes have native sites.
pET3a.mamO-U	Ampicillin	*mamOPARQBSTU* expressed from a T7 promoter. *mamO* has the ribosome binding site from the pET3a vector, while all other genes have native sites.

pET3a.mamH-U	Ampicillin	*mamAB* operon expressed from a T7 promoter. *mamH* and *mamO* have ribosome binding sites from the pET3a vector, while all other genes have native sites.
pETlac.mamH-U	Ampicillin	*mamAB* operon expressed from a lactose promoter. *mamH* and *mamO* have ribosome binding sites from the pET3a vector, while all other genes have native sites.
pET3a.4070-74.mamH-U	Ampicillin	*mms6* and *mamAB* operons expressed from a T7 promoter in a single transcript. All genes in the *mms6* operon, *mamH* and *mamO* have ribosome binding sites from the pET3a vector, while all other genes have native sites.
pET3a.mamG	Ampicillin	*mamG* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamF	Ampicillin	*mamF* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamD	Ampicillin	*mamD* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamC	Ampicillin	*mamC* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.4070	Ampicillin	*mg4070* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.4071	Ampicillin	*mg4071* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mmsF	Ampicillin	*mmsF* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mms6	Ampicillin	*mms6* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.4074	Ampicillin	*Mg4074* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamH	Ampicillin	*mamH* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamI	Ampicillin	*mamI* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.

pET3a.mamE	Ampicillin	*mamE* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamJ	Ampicillin	*mamJ* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamK	Ampicillin	*mamK* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamL	Ampicillin	*mamL* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamM	Ampicillin	*mamM* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamN	Ampicillin	*mamN* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamO	Ampicillin	*mamO* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamP	Ampicillin	*mamP* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamA	Ampicillin	*mamA* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamQ	Ampicillin	*mamQ* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamR	Ampicillin	*mamR* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamB	Ampicillin	*mamB* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamS	Ampicillin	*mamS* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamT	Ampicillin	*mamT* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamU	Ampicillin	*mamU* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamY	Ampicillin	*mamY* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamX	Ampicillin	*mamX* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET3a.mamZ	Ampicillin	*mamZ* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.

pET3a.ftsZ-like	Ampicillin	*ftsZ-like* expressed from a T7 promoter with the ribosome binding site from the pET3a vector.
pET14b.mamI	Ampicillin	N-terminus 6xHis tagged MamI produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamE	Ampicillin	N-terminus 6xHis tagged MamE produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamJ	Ampicillin	N-terminus 6xHis tagged MamJ produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamL	Ampicillin	N-terminus 6xHis tagged MamL produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamM	Ampicillin	N-terminus 6xHis tagged MamM produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamN	Ampicillin	N-terminus 6xHis tagged MamN produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamO	Ampicillin	N-terminus 6xHis tagged MamO produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamP	Ampicillin	N-terminus 6xHis tagged MamP produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamA	Ampicillin	N-terminus 6xHis tagged MamA produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamQ	Ampicillin	N-terminus 6xHis tagged MamQ produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamR	Ampicillin	N-terminus 6xHis tagged MamR produced from a T7 promoter, with the ribosome binding site from the pET14b vector.

pET14b.mamB	Ampicillin	N-terminus 6xHis tagged MamB produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamS	Ampicillin	N-terminus 6xHis tagged MamS produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamT	Ampicillin	N-terminus 6xHis tagged MamT produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamU	Ampicillin	N-terminus 6xHis tagged MamU produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.mamM	Ampicillin	C-terminus 6xHis tagged MamM produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.mamB	Ampicillin	C-terminus 6xHis tagged MamB produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET14b.mamE-soluble	Ampicillin	N-terminus 6xHis tagged soluble domain of MamE produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.Sec.S	Ampicillin	Periplasmic targeting sequence from BtuF produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.Sec.S.mamE-soluble	Ampicillin	C-terminus 6xHis tagged soluble domain of MamE with an N-terminus periplasmic targeting sequence from BtuF produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.Sec.S.mamX-soluble	Ampicillin	C-terminus 6xHis tagged soluble domain of MamX with an N-terminus periplasmic targeting sequence from BtuF produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET23b.Sec.S.mamP-soluble	Ampicillin	C-terminus 6xHis tagged soluble domain of MamP with an N-terminus periplasmic targeting sequence from BtuF produced from a T7 promoter, with the ribosome binding site from the pET14b vector.

pET23b.Sec.S.mamT-soluble	Ampicillin	C-terminus 6xHis tagged soluble domain of MamT with an N-terminus periplasmic targeting sequence from BtuF produced from a T7 promoter, with the ribosome binding site from the pET14b vector.
pET3a.4070.71	Ampicillin	*mg4070* and *mg4071* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.4070.71.mF	Ampicillin	*mg4070, mg4071* and *mmsF* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.m6.4074	Ampicillin	*mms6* and *mg4074* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.4070-74	Ampicillin	*mms6* operon expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.FD	Ampicillin	*mamF* and *mamD* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.FDC	Ampicillin	*mamF, mamD* and *mamC* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.YX	Ampicillin	*mamY* and *mamX* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.YXZ	Ampicillin	*mamY, mamX* and *mamZ* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.YXZFtsZ	Ampicillin	*mamYX* operon expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.HI	Ampicillin	*mamH* and *mamI* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.EL	Ampicillin	*mamE* and *mamL* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.EM	Ampicillin	*mamE* and *mamM* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.MN	Ampicillin	*mamM* and *mamN* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.NO	Ampicillin	*mamN* and *mamO* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.

pET3a.OP	Ampicillin	*mamO* and *mamP* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.PA	Ampicillin	*mamP* and *mamA* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.QB	Ampicillin	*mamQ* and *mamB* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.ST	Ampicillin	*mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.HU	Ampicillin	*mamH* and *mamU* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.ELM	Ampicillin	*mamE, mamL* and *mamM* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.HIEL	Ampicillin	*mamH, mamI, mamE* and *mamL* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.IELM	Ampicillin	*mamI, mamE, mamL* and *mamM* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.NOPA	Ampicillin	*mamN, mamO, mamP* and *mamA* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.QBST	Ampicillin	*mamQ, mamB, mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.RQBST	Ampicillin	*mamR, mamQ, mamB, mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.PARQBST	Ampicillin	*mamP, mamA, mamR, mamQ, mamB, mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.NOPARQBST	Ampicillin	*mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.

pET3a.MNOPARQBST	Ampicillin	*mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS* and *mamT* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.TBAD.HI	Ampicillin	*mamH* and *mamI* expressed from an arabinose promoter with ribosome binding sites from the pET3a vector.
pET3a.pRha.HI	Ampicillin	*mamH* and *mamI* expressed from a rhamnose promoter with ribosome binding sites from the pET3a vector.
pET3a.pRha.HIEL	Ampicillin	*mamH, mamI, mamE* and *mamL* expressed from a rhamnose promoter with ribosome binding sites from the pET3a vector.
pZS24.U	Kanamycin	*mamU* expressed from a T7 promoter with ribosome binding sites from the pET3a vector in pZS24.
pZS24.QBSTU	Kanamycin	*mamQ, mamB, mamS, mamT* and *mamU* expressed from a T7 promoter with ribosome binding sites from the pET3a vector in pZS24.
pZS24.MNOPARQBSTU	Kanamycin	*mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS, mamT* and *mamU* expressed from a T7 promoter with ribosome binding sites from the pET3a vector in pZS24.
pET.cocoR.HIEL	Ampicillin	*mamH, mamI, mamE* and *mamL* expressed from a T7lac promoter with ribosome binding sites from the pETcoco2 vector in pETcocoR.
pET.cocoR.HIELMN	Ampicillin	*mamH, mamI, mamE, mamL, mamM,* and *mamN* expressed from a T7lac promoter with ribosome binding sites from the pETcoco2 vector in pETcocoR.
pET.cocoR.HIELMNOP	Ampicillin	*mamH, mamI, mamE, mamL, mamM, mamN, mamO* and *mamP* expressed from a T7lac promoter with ribosome binding sites from the pETcoco2 vector in pETcocoR.
pET.cocoR.HIELMNOPARQBST	Ampicillin	*mamH, mamI, mamE, mamL, mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS* and *mamT* expressed from a T7lac promoter with

		ribosome binding sites from the pETcoco2 vector in pETcocoR.
pET3a.MamQ-solbule	Ampicillin	The soluble domain of MamQ (amino-acids 73-272) produced from a T7 promoter with ribosome a binding site from the pET3a vector.
pET3a.IQ	Ampicillin	*mamI* and *mamQ* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.LQ	Ampicillin	*mamL* and *mamQ* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.QB	Ampicillin	*mamQ* and *mamB* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.ILQ	Ampicillin	*mamI, mamL* and *mamQ* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.LQB	Ampicillin	*mamL, mamQ* and *mamB* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.IQB	Ampicillin	*mamI, mamQ* and *mamB* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.ILQB	Ampicillin	*mamI, mamL, mamQ* and *mamB* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.IQBY	Ampicillin	*mamI, mamQ, mamB* and *mamY* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.LQBY	Ampicillin	*mamL, mamQ, mamB* and *mamY* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.ILQBY	Ampicillin	*mamI, mamL, mamQ, mamB* and *mamY* expressed from a T7 promoter with ribosome binding sites from the pET3a vector.
pET3a.LemA.153	Ampicillin	*lemA* from *Bacillus megaterium* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.

pET3a.LemA.159	Ampicillin	*lemA* from *Clostridium kluyveri* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.LemA.501	Ampicillin	*lemA* from *Brucella melitensis* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.LemA.565	Ampicillin	*lemA* from *Pseudomonas aeruginosa* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*RFP	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to RFP expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*LemA.153s	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to the soluble domain of LemA from *Bacillus megaterium* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*LemA.159s	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to the soluble domain of LemA from *Clostridium kluyveri* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*LemA.501s	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to the soluble domain of LemA from *Brucella melitensis* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*LemA.565s	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to the soluble domain of LemA from *Pseudomonas aeruginosa* expressed from a T7 promoter with a ribosome binding site from the pET3a vector.
pET3a.BamE*MamQs	Ampicillin	The first 21 amino-acids of BamE from *Escherichia coli* fused to the soluble domain of MamQ from *Magnetospirillum gryphiswaldense* expressed from a

		T7 promoter with a ribosome binding site from the pET3a vector.

2.4.3. Gene Synthesis

Synthesis of the DNA corresponding to the BamE outer membrane targeting tag was ordered from Eurofins™.

2.4.4. Fast-start High Fidelity PCR

Standard protocol for FastStart High Fidelity PCR System from Roche™ was used. All DNA used was obtained from DSMZ. Amount of DMSO used was between 0 µl and 3 µl with the annealing temperature between 58°C and 65°C for the amplification of all the singe genes of interest.

2.4.5. Overlap Extension Mutagenesis

Overlap extension mutagenesis was carried out to remove a restriction enzyme site in pETcoco2. The targeted region was the *KpnI/NcoI* site in the vector. Firstly, fast-start high fidelity PCR was carried out, using the following primer pairs: COCO2.KpnI.FW together with COCO2.NcoI.RV and COCO2.Mut.RV together with COCO2.Mut.FW. The Mut primers carry a silent mutation of a thymine to adenine within the desired *SpeI* site. The second PCR was then carried out, using the generated products as template 1 and template 2.

Reagent	Volume
Fast-start High Fidelity buffer 10x	5 µl
FW Primer	2 µl
RV Primer	2 µl
dNTPs	5 µl
Template 1	0.5 µl
Template 2	0.5 µl
H_2O	34 µl
Enzyme	1 µl

The reaction was carried out in a PCR block with the following settings: 2 minutes at 96°C, 20 cycles of 30 seconds at 96°C, 30 seconds at 55°C and 120 seconds at 72°C, followed by a single cycle of 7 minutes at 72°C. The PCR product was then ran on an agarose gel, gel-extracted, digested with restrictases (*KpnI/NcoI*) and ligated into the digested vector (*KpnI/NcoI* site). The correct plasmid was chosen for using restriction enzyme analysis and confirmed by sequencing.

2.4.6. Expand Long Range PCR

Standard protocol for Expand Long Range PCR System (now called Expand Long Template PCR System) from Roche™ was used to amplify the large genetic fragments (*mamH-N* and *mamO-U*). Amount of DMSO used was 4 µl with the annealing temperature of 58°C. DNA used was obtained from DSMZ

2.4.7. Plasmid Purification

QIAprep Spin Miniprep Kit from QIAGEN™ was used to purify all plasmids following the standard protocol, with the optional PB step. All plasmids were eluted in 30 µl of EB. The amount of overnight culture used for pZS24 vectors was increased to 10ml, doubling the volumes of reagents used for cell lysis and neutralisation. For pETcoco2 based vectors, induction of plasmid production prior to purification was carried using the manufacturers' guidelines.

2.4.8. Restriction Digests

Restriction enzymes from Promega™ (*BglII; KpnI; NdeI; PstI; ScaI; SacI; SpeI; XbaI; XhoI*) and NEB (*AseI*) were used in the experiments digesting in the buffer with highest overall activity and at least 50% activity for each of the enzymes. DNA amount used varied form 2 µl for test digests, 5 µl for ligations yielding small plasmids (up to 3 genes in a single vector) to 15 µl for ligations yielding large plasmids (more than 5 genes). Digestions were carried out for at least 1 hour in a 37°C water bath.

2.4.9. Ligations

DNA fragments were ligated in 10 µl reactions for at least 1 hour at room temperature.

Reagent	Volume
2x Ligation Buffer	5 µl
Insert	3.5 µl
Vector	1 µl
T4 Ligase (Promega™)	0.5 µl

2.4.10. DNA Electrophoresis

DNA fragments were separated in a 0.8% agarose-TAE gel run in a TAE buffer. DNA was visualised with ethidium bromide at a final concentration of 0.5 µg/ml. Before loading, the DNA was diluted in 5x DNA loading buffer (Bioline). HyperLadder™ 1kb (Bioline) was used to approximate fragment sizes. DNA gels were visualised under UV light.

2.4.11. Extraction of DNA from DNA Gels

Bands, corresponding to the correct sizes, were excised using a scalpel. QIAquick Gel Extraction Kit from QIAGEN™ was used to extract all DNA following the standard protocol, with the optional additional wash step. All DNA was eluted in 30 µl of EB. For larger constructs (10kb+) the EB was pre-warmed to 50°C.

2.4.12. Sequencing

Sanger sequencing, to confirm correct amplification and insertion of DNA fragments, was carried out by Beckman-Coulter Genomics™ (now part of GENEWIZ™).

2.5. Protein Production and Purification

2.5.1. Protein Extraction Using Sonication

Cell pellets resuspended in binding buffer were sonicated for 5 minutes with a 30 second on 30 second off cycle at an amplitude of 65%.

2.5.2. Purification of Cytochrome Proteins Using IMAC

Following cell lysis using sonication, cell debris was removed by centrifugation at 38,000 g for 20 minutes at 4 °C. 5 ml of Chelating Sepharose Fast Flow resin was loaded in a column and washed with 20 ml of dH_2O. The column was charged with 10 ml of charging buffer and equilibrated with 20 ml binding buffer. The supernatant was applied to the column and then column was then washed with 20 ml of binding buffer, followed by 10 ml of washing buffer I and 10ml of wash buffer II. Proteins of interest were eluted with 15 ml of elution buffer and 1 ml elution fractions were collected. Nickel was then removed with 5ml strip buffer and the column was washed with 10 ml of dH_2O for the re-use of the resin.

2.5.3. Buffer Exchange Using a PD10 Column

Following IMAC 2.5 ml of the elution fractions with the highest protein concentration were loaded onto a PD-10 desalting column equilibrated in PD-10 buffer and allowed to flow through. The protein of interest was eluted by addition of 3.5 ml of PD-10 buffer.

2.5.4. SDS-PAGE

Prior to analysis, the cells were lysed and fractionated into soluble and insoluble fractions using B-PER™ Bacterial Protein Extraction Reagent (ThermoFisher Scientific™). NuPAGE™ 4%-12% Bis-Tris Gels (Thermofisher Scientific™) were run in NuPage MOPS SDS Running buffer (Life Technologies™) using standard procedures. Blue Prestained Protein Standard, Broad Range (11-190 kDa) (NEB™) was used to quantitate the observed protein sizes. The proteins were denatured prior to loading by diluting in Laemmli buffer and incubation at 100^{o}C for 15 minutes in a heat block. The proteins were visualised by staining with Coomasie Blue stain for 30 min following de-staining in dH_2O overnight.

2.5.5. Western Blotting

SDS-PAGE gels were run as described above, using various markers depending on the blot. Nitrocellulose membrane was equilibrated in methanol for 10 seconds, washed with dH_2O and equilibrated in transfer buffer together with the gel. Proteins were blotted onto the nitrocellulose membrane in a cooled gel tank for 1 hour at 100 volts, constant voltage. Non-specific antibody binding sites were blocked by incubation of the membrane in phosphate buffered saline containing 5% (w/v) skimmed milk powder at 4^{o}C overnight. The membrane was subsequently incubated with the primary antibody (mouse anti-His; Sigma-Aldrich™) diluted 1:1000 in PBS for 1 hour under gentle agitation. The membrane was rinsed three times with PBS and subsequently equilibrated in phosphate-free solution for 10 minutes. The secondary antibody (anti-mouse IgG-AP; Promega™) was diluted 1:5000 in phosphate-free solution containing 5% (w/v) skimmed milk powder, the membrane was incubated in this solution for 1 hour under gentle agitation followed by three 10 minute washes in phosphate-free solution. A single tablet of the substrate 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT) was dissolved in 10 ml of distilled water. The membrane was then equilibrated in water for 10 minutes and incubated in the BCIP/NBT solution until sufficient colour has developed. The membrane was then washed in dH_2O, dried and visualised using a camera.

2.5.6. Cytochrome C Spectrometry

UV spectra was obtained using Cary 60 UV-Vis spectrophotometer (Agilent Technologies), using the Cary WinUV Scan Application (version 5.0.0.999) scanning 750 nm to 250 nm at medium scan speed. 1ml PLASTIBRAND UV-cuvettes were used and the path length was 1 cm. 1 mM solution of dithionite in dH_2O was used to reduce the cytochromes.

2.6. Electron Microscopy

2.6.1. Preparation of Whole Cells

Cells were grown for 6h at 30^oC at 180 rpm in a shaking incubator, from multiple colonies in 50 ml of either LB or MTB media and induced with 100 µM IPTG overnight at 19^oC at 180 rpm. Cells were harvested using centrifugation, washed with 10 ml of phosphate buffered saline three times and resuspended in 1 ml of phosphate buffered saline. 5 µl was pipetted on a carbon-coated copper electron microscopy grid and allowed to settle for 5 min followed by the removal of the liquid media by absorption using filter paper and air-drying.

2.6.2. Fixing and Embedding Bacteria in Resin

Cells were grown and harvested as above. The cell pellet was resuspended in 2 ml 2.5% glutaraldehyde in 100 mM cacodylate pH 7.2 in dH_2O and incubated for 2 hours with gentle spinning. Cells were pelleted by centrifugation at 8,000 rpm for 2 minutes and were washed twice with 100 mM cacodylate pH 7.2 in dH_2O. Cells were stained with 1% osmium tetroxide in 100 mM cacodylate pH 7.2 in dH_2O for 2 hours and subsequently washed twice with dH2O. Cells were dehydrated by incubation in an ethanol gradient, 50% ethanol for 10 minutes, 70% ethanol overnight followed by two 10 minute washes in 100% ethanol. Cells were further dehydrated using two wash steps in propylene oxide for 15 minutes. Cell pellets were infiltrated by resuspension in 1 ml of a 1:1 mix of propylene oxide and Agar LV Resin and incubated for 30 minutes with spinning. Cell pellets were embedded by incubation in 100% Agar LV resin for 2 hours twice. The cell pellet was resuspended in fresh resin and transferred to a 0.5 ml mould, centrifuged for 5 minutes at 1,000 rpm to concentrate the cells to the tip of the mould and incubated for 16 hours at 60 °C to polymerise.

2.6.3. Ultra-thin Sectioning and Staining of Embedded Samples

Samples were thin sectioned on a RMC MT-XL ultramicrotome with a diamond knife (diatome 45°). Sections were placed on 300 mesh copper grids (Agar Scientific™).

Grids were stained by incubation in 4.5% uranyl acetate in 1% acetic acid solution for 45 minutes followed by 2 washes in dH_2O. Grids were then stained with 0.1% lead citrate for 8 minutes followed by a wash in dH_2O.

2.6.4. Visualisation of Samples

The samples were visualised using a JEOL-1230 TEM.

Chapter 3

Cloning and Initial Characterisation of the *mamAB* Operon

3.1. Introduction

The bacterial cell is densely packed with different macromolecules, including DNA, RNA and proteins (Spitzer and Poolman, 2013). In order for cells to function efficiently, sub-cellular organisation must be achieved and maintained (Harold, 2005). It is thought that the simplest way of organisation, which is present in all cells, is based on protein-protein interactions within hyperstructures (Norris *et al.*, 2007). These include structures involved in such key processes as DNA repair and replication, protein production, cell division and motility. Some bacteria have also employed more specialised, protein-bound and lipid-bound organelles (Murat *et al.*, 2010b), one of which is the magnetosome (Bazylinski and Frankel, 2004).

Magnetosomes are membranous organelles that produce a magnetic mineral within its lumen (Gorby *et al.,* 1988; Balkwill *et al.*, 2006). These organelles are aligned via an actin-like protein (Komeili *et al.*, 2006) and are thought to help the bacteria locate its preferred environment (Frankel *et al*., 1997). The membranous nature of the organelle allows for the enhanced incorporation of membranous proteins that are associated with the mineralisation process but also acts as a barrier in order to facilitate the accumulation and reduction of iron. Magnetosomes have great potential in biotechnology, not only because of the production of very specific nanomagnets (Bain and Staniland, 2015), but also because there is the potential to exploit the membranous nature of the compartment itself.

E. coli is a model bacterium that has been extensively studied and is used for a variety of biotechnological applications (Blount, 2015). This bacterium does not normally produce any specific membranous organelles. The research described in this chapter sets out to engineer the membranous magnetosome organelle within this organism.

This chapter presents the research that was carried out in order to clone the *mamAB* operon in its native organisation and to investigate the growth conditions needed for the production of magnetosomes. The cloning strategy is complicated as *mamAB* is a large operon (over 16kb), containing 17 genes: *mamH, mamI, mamE, mamJ, mamK, mamL, mamM, mamN, mamO, mamP, mamA, mamQ, mamR, mamB, mamS, mamT* and *mamU*, of these *mamL, mamQ, mamB, mamI, mamE, mamM* and *mamO* have been shown to be sufficient for magnetosome membrane formation in *M. gryphiswaldense* (Raschdorf *et al.,* 2016), with some other members being essential for biomineralisation and establishing magnetism. A plasmid, containing the *mamAB* operon from *M. gryphiswaldense,* was

constructed and expressed in *E. coli* and the cells were visualised using electron microscopy. The construct was then modified and different variants were produced. A variety of growth conditions, which yield magnetosome formation in MTB, were investigated for production of magnetosomes using the various constructs in *E. coli*.

3.2. Results

3.2.1. Choosing the Organism

The first step of the experimental design was choosing the organism to be used as a source for the genetic information. Multiple magnetotactic bacteria were considered and *M. gryphiswaldense* was chosen as it is the most extensively studied organism. Moreover, it has had its genome completely sequenced and has been subject to a large number of genetic studies in order to identify the genes involved in magnetosome formation. Prior to cloning, the codon adaptation index (CAI) for expression in *E. coli* was calculated for the *M. gryphiswaldense* proteins thought to be involved in magnetosome formation (**Supplementary table 1**). This gave an average CAI of 0.66, with the lowest being 0.57 (*mg4074*). When compared to the CAI of *lacZ* from *E. coli* (0.75), these results suggest efficient translation of all the analysed proteins in the organism.

3.2.2. Cloning of the *mamAB* Operon

As discussed in the introduction, the MAI of *M. gryphiswaldense* contains four operons that have been implicated in magnetosome formation with the *mamAB* operon being the largest (over 16kb), containing 17 genes. The *mamAB* operon has been proposed to be sufficient for magnetosome formation, but not crystal maturation (Lohbe *et al.*, 2011). Two different approaches were used to clone this large operon: a non-modular and a modular approach. The latter is discussed in **section 4.2.6**. The non-modular approach employed the use of the Expand Long Template PCR System (Roche), which allows generation of PCR products ranging from 5kb to 20kb with high accuracy. Even though it is well within the range of the system, no product was obtained when a PCR of the whole *mamAB* operon was attempted. As an alternative approach, two PCR segments containing *mamHIEJKLMN* and *mamOPARQBSTU* were amplified and ligated into a pET3a vector yielding pET3a.mamH-N and pET3a.mamO-U respectively (**Figure 3.1.**). These were then combined to produce pET3a.mamH-U (**Figure 3.1.**), reconstructing the whole *mamAB* operon that is expressed from a T7 promoter. Partial sequencing of the plasmid revealed multiple

mutations (**Table 3.1.**), of which the most severe was a frame-shift mutation in *mamE* as *mamJ and mamK* are not important for organelle formation these could be disregarded, while the mutation in *mamB* is predicted to be in a non-essential region (protein surface not involved in interactions). Attempts to obtain clones without the mutation in *mamE* failed. Therefore this clone was used for further experiments taking the mutations into consideration.

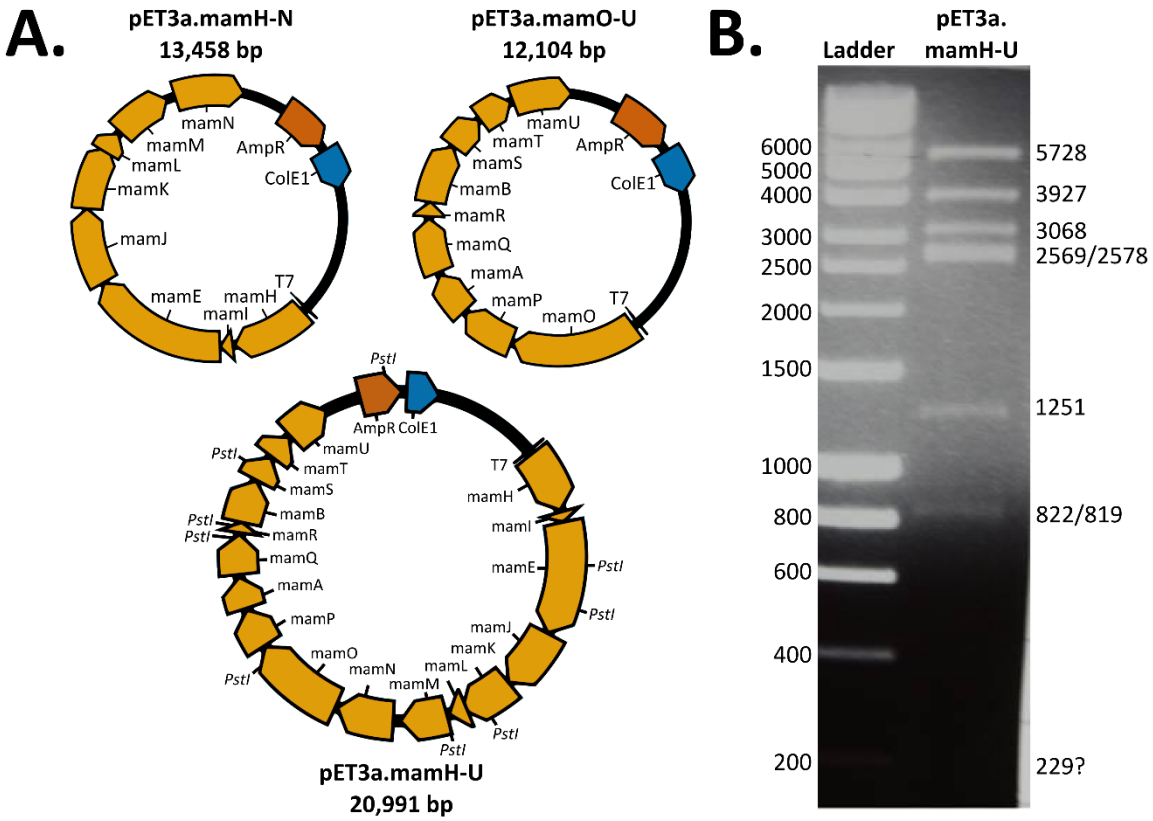


Figure 3.1. Plasmids constructed using long-range PCR.
(A) Vector maps of constructed plasmids. **(B)** *PstI* digest of pET3a.mamH-U showing the correct band pattern. Intensity of the lowest band is too low to be seen in the picture.

Table 3.1. Mutations present in the pET3a.mamH-U construct.

Gene	Mutation
mamE	Threonine^{257} -> Frame Shift
mamJ	Proline^{100} -> Serine Valine^{176} -> Alanine
mamK	Leucine^{25} -> Valine Valine^{111} -> Glycine
mamB	Arginine^{208} -> Histidine

3.2.3. Initial Growth and Microscopy

E. coli BL21 (DE3) Star cells were transformed with pET3a.mamH-U, grown at 37 ^{o}C 180 rpm for 6 hours, induced with 100 µM IPTG and incubated over-night at 18 ^{o}C. The cells were then harvested and imaged either as whole cells or as sections of embedded cells (**Figure 3.2**). No electron-dense particles, which would suggest biomineralisation, could be seen in the whole cell samples but possible membranous invaginations and inclusion bodies were observed in the cell sections. The lack of biomineralisation could be due to a variety of reasons including the use of the wrong growth conditions, poor protein solubility/expression, the absence of certain genes from the construct or mutations present in the construct. Some of these possibilities are addressed in the following sections.

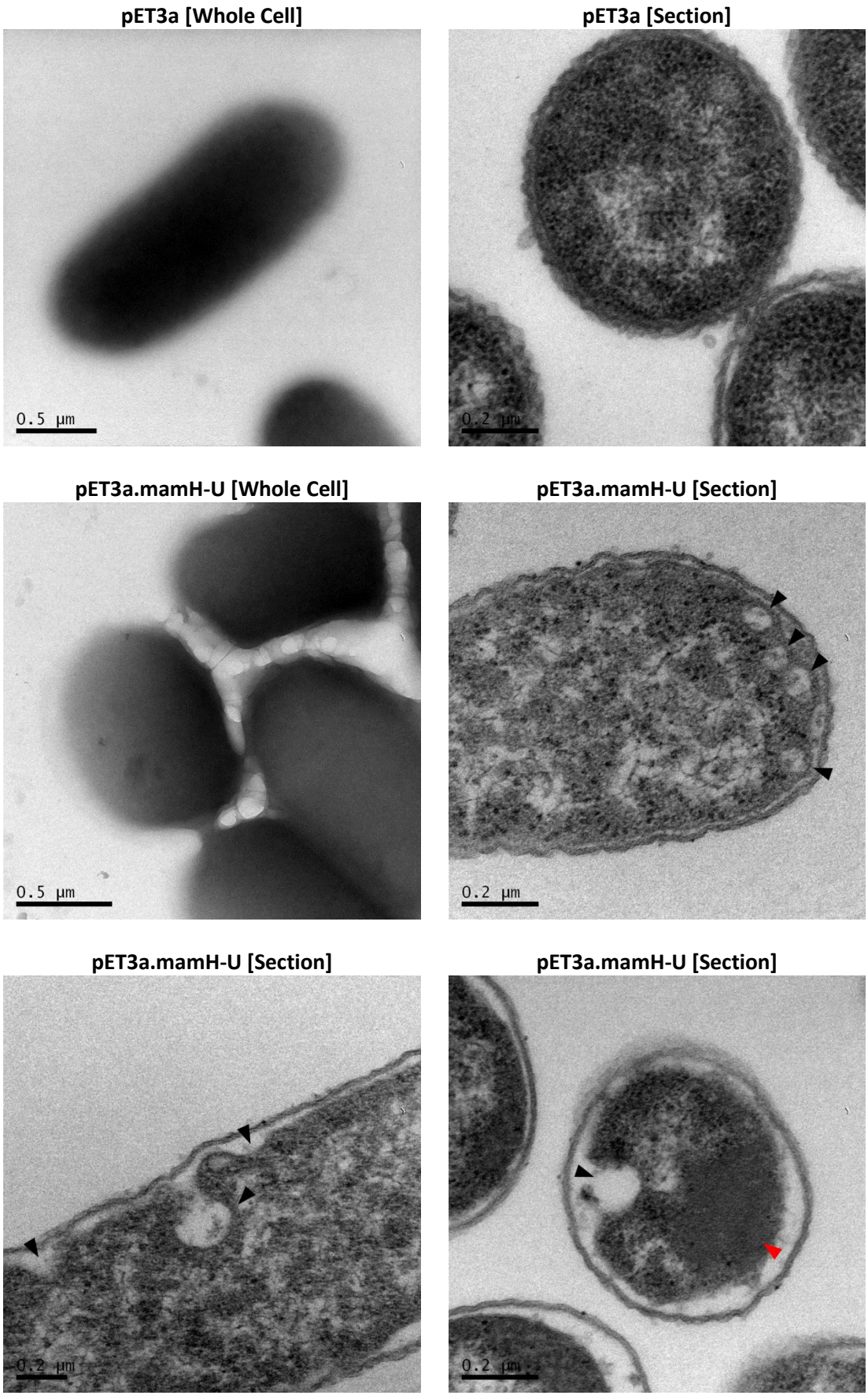
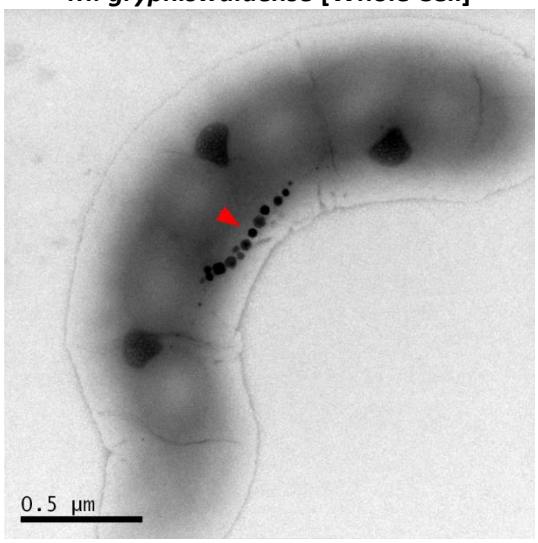


Figure 3.2. TEM of *E. coli* expressing the *mamAB* operon.
Transmission elecron micrographs of *E. coli* BL21 (DE3) cells (**whole cells** or cell **sections**) harboring either an empty **pET3a** control plasmid or **pET3a.mamH-U**. **Black arrows** – possible membrane invaginations; **red arrow** – inclusion body.

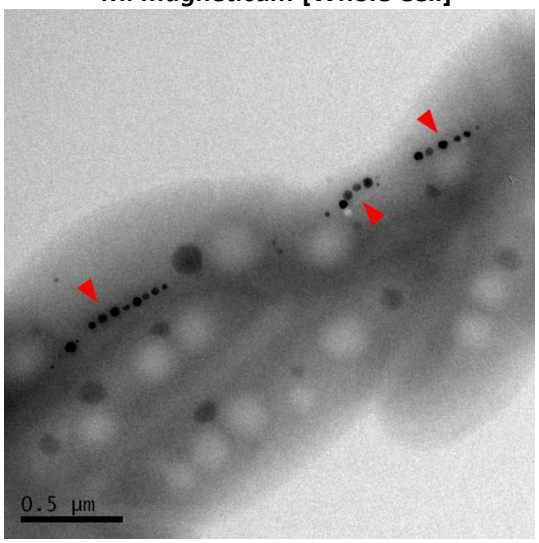
3.2.4. Development of Growth Conditions: Growing Magnetotactic Bacteria

In order to develop media for magnetite production in *E. coli, M. gryphiswaldense* and *M. magneticum,* which both naturally produce magnetosomes, were cultured. *M. magneticum* is a closely related strain to *M. gryphiswaldense* and was chosen to complement the experiment due its much faster growth rate. Cells were cultured as described in materials and methods. Growth in LB media was also tested, but no cell growth was observed. The magnetism of cells was checked by introducing a magnetic field using a neodymium magnet at the side of the growth flask. The cells were observed to migrate towards the magnet. Whole cells of both strains were imaged (**Figure 3.3**). Cell sections were obtained of *M. magneticum* (**Figure 3.3**), but not *M. gryphiswaldense* due to the failure to obtain enough cells for the procedure.

***M. gryphiswaldense* [Whole Cell]**



***M. magneticum* [Whole Cell]**



***M. magneticum* [Section]**

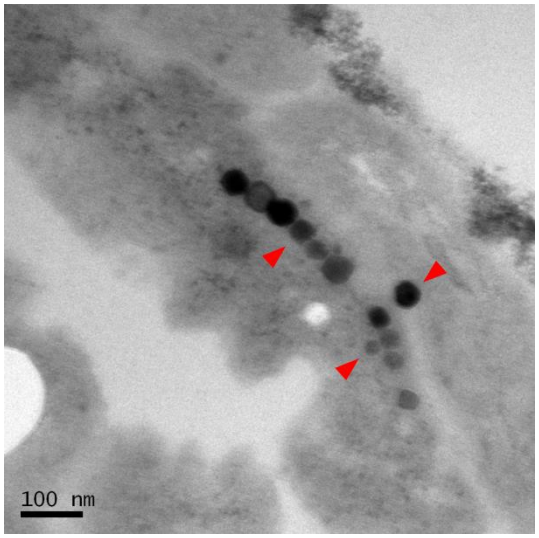


Figure 3.3. Transmission electron microscopy of *Magnetospirillum* species.
Transmission elecron micrographs of either **whole cells** or cell **sections** of *Magnetospirillum* species (***M. gryphiswaldense*** and ***M. magneticum***) grown in MTB media. **Red arrows** – electron dense particles indicitive of magnetosomes

3.2.5. pET3a.4070-74.mamH-U and pETlac.mamH-U

Additional constructs were made for growth studies (**Table 3.2.**). The *mamAB* operon from pET3a.mamH-U was transferred into a pETlac vector, where the T7 promoter had been exchanged for a weaker lactose promoter, yielding pETlac.mamH-U. A plasmid containing two operons from the MAI (pET3a.4070-74.mamH-U) was constructed by inserting the *mms6* operon from pET3a.4070-74 (**Section 4.2.6**) into pET3a.mamH-U, yielding pET3a.4070-74.mamH-U.

Table 3.2. Plasmids constructed for the adjusted growth experiments.

Plasmid	Promoter	Operons Expressed (Chapter reference)
pET3a.mamH-U	T7	*mamAB* (3.2.2)
pETlac.mamH-U	Lactose	*mamAB* (3.2.2)
pET3a.4070-74.mamH-U	T7	*mms6* (4.2.6); *mamAB* (3.2.2)

3.2.6. Adjusted Growth and Microscopy

E. coli BL21 (DE3) cells expressing either pET3a.mamH-U, pETlac.mamH-U, pET3a.4070-74.mamH-U or pET3a (empty vector control) were grown in the same media as MTB for 72h at 30^{o}C under micro-anaerobic conditions. The cells were then embedded, sectioned and imaged as before. Whole cells of pET3a.4070-74.mamH-U were also imaged. The cells looked highly stressed and were miss-shapen (**Figure 3.4. and Figure 3.5.**), which was possibly due to the growth conditions used and the prolonged growth required to obtain enough cells to carry out the procedure. Some observed cells contained electron dense particles that looked similar to those observed in *M. magneticum* (**Figure 3.3**), although some were also present in the control (**Table 3.3.**). When a magnetic field was introduced, the cells did not migrate towards it.

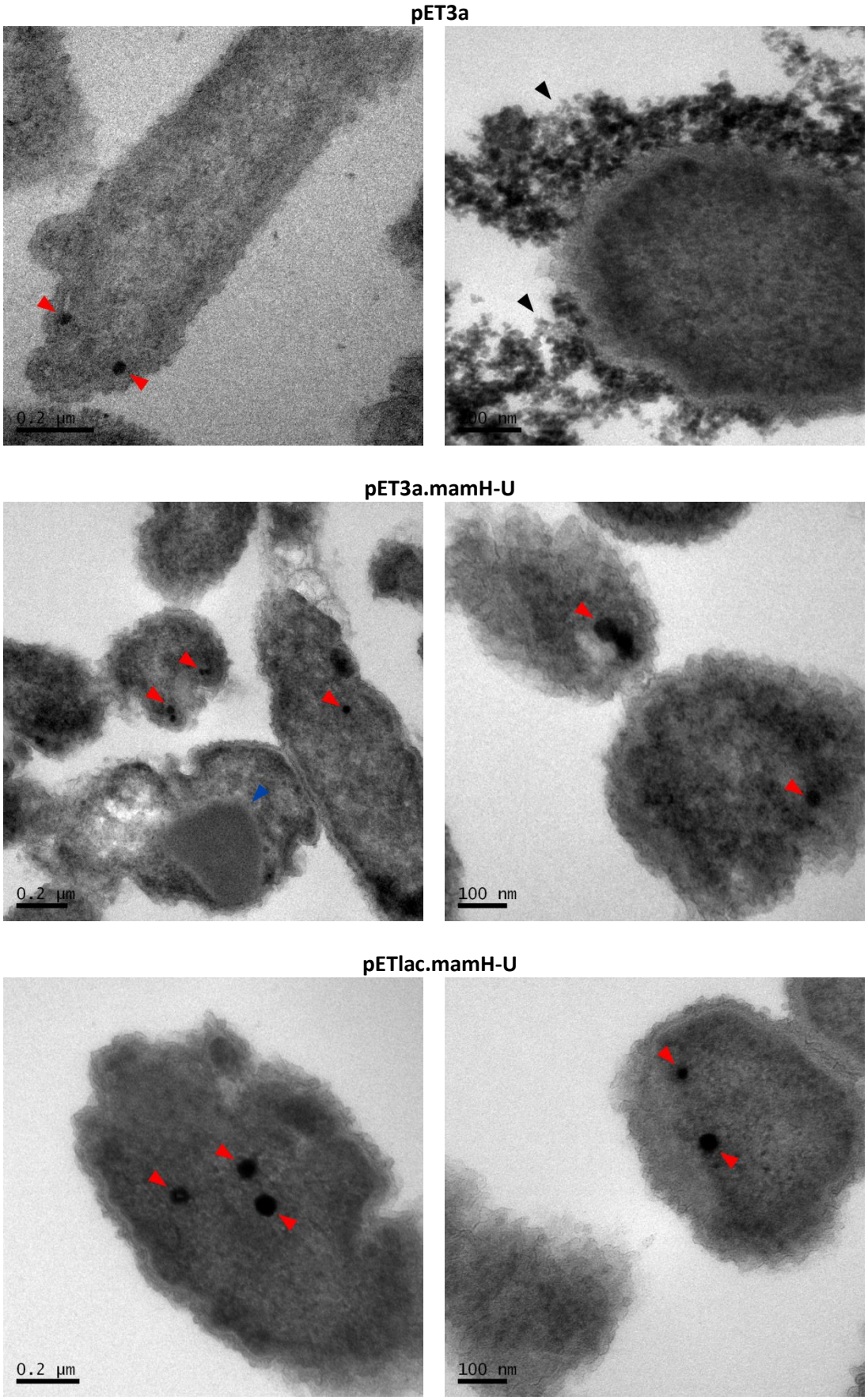


Figure 3.4. TEM of *E. coli* expressing the *mamAB* operon from a variety of operons. Transmission elecron micrographs of sectioned *E. coli* BL21 (DE3) cells harboring either an empty **pET3a** control plasmid, **pET3a.H-U** or **pETlac.H-U**. **Red arrows** – electron dense particles present within cells; **black arrows** – electron dense particles present outside the cells; **blue arrow** – inclusion body.

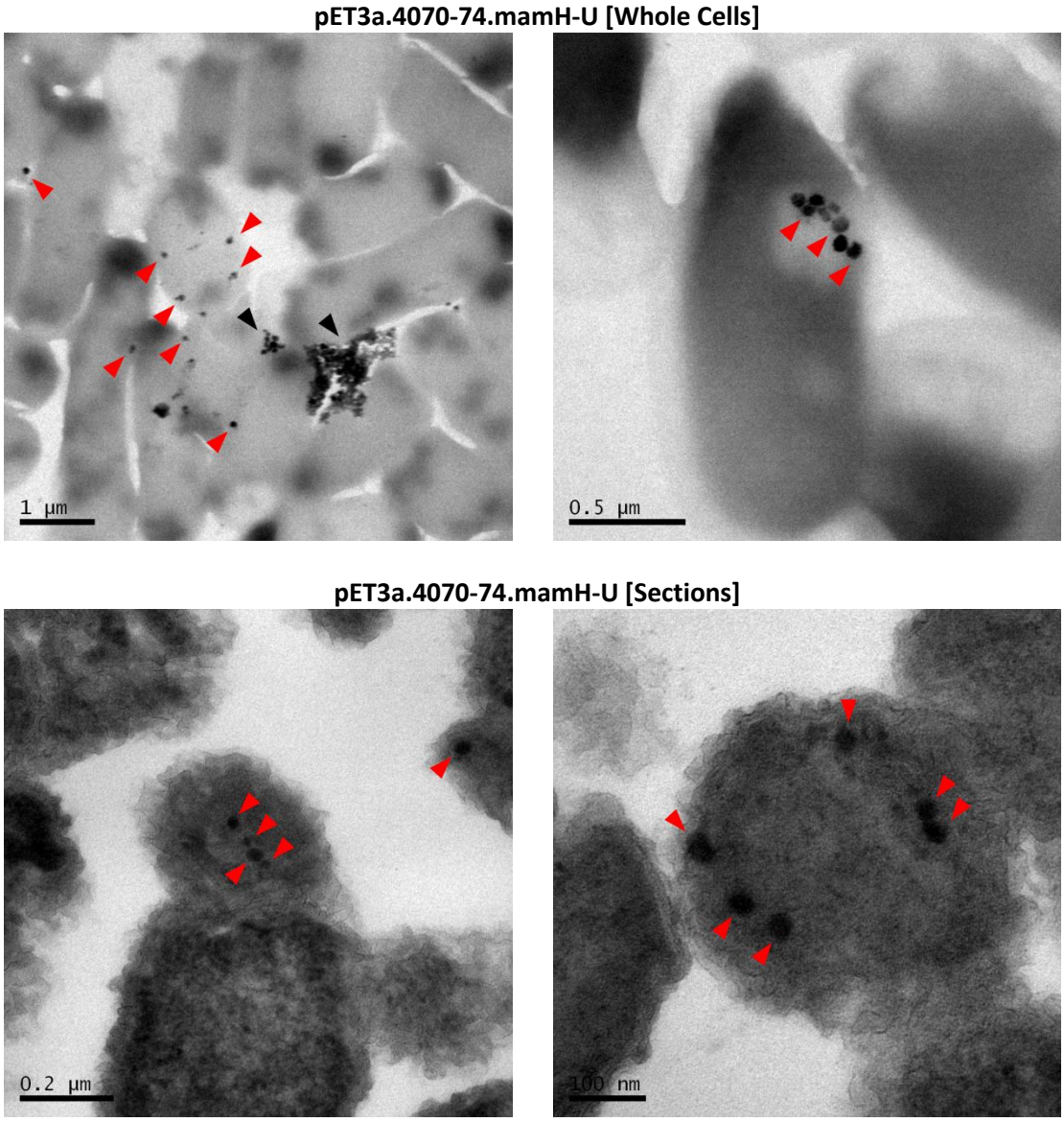


Figure 3.5. TEM of *E. coli* expressing the *mms6* and *mamAB* operons.
Transmission elecron micrographs of either **whole cells** or cell **sections** of *E. coli* BL21 (DE3) cells cells harboring **pET3a.4070-74.mamH-U**. **Red arrows** – electron dense particles present within cells; **black arrows** – electron dense particles present outside the cells.

Table 3.3. Overview of adjusted growth experiments.

Plasmid	Electron-dense particles		Inclusion bodies
	Intracellular*	Extracellular	
pET3a	1 in 50 cells Up to 2 particles/cell observed	+	-
pET3a.mamH-U	1 in 40 cells Up to 2 particles/cell observed	-	+
pETlac.mamH-U	1 in 40 cells Up to 6 particles/cell observed	-	-
pET3a.4070-74.mamH-U	1 in 20 cells Up to 12 particles/cell observed	+	-

* - estimation from electron microscopy images.

3.3. Discussion

In this chapter the cloning and expression of the *M. gryphiswaldense mamAB* operon in *E. coli* is described. Cloning of the whole *mamAB* operon was only partially achieved, as a probable loss-of-function frame-shift mutation was present in one of the key genes, *mamE*. Murat *et al.* (Murat *et al.,* 2010a) have shown that a Δ*mamE* strain of *M. magneticum* is still able to form magnetosome membranes, but not mineralise them. Unfortunately, a construct without this mutation could not be obtained, which might be due to the toxicity of the protein, a point that is discussed briefly in the next chapter.

E. coli cells expressing the *mamAB* operon were subject to TEM studies. When grown under aerobic conditions in LB media membranous invaginations similar, but not identical, to those produced in magnetosomes were observed, although no electron dense particles were observed (**Figure 3.2**). The presence of invaginations, but not particles, could be due to a number of reasons, including the presence of a frame-shift mutation in *mamE*, incorrect growth conditions, poor protein solubility/expression or the absence of certain genes from the construct.

Previously, it has been shown that MTB are only able to produce magnetosomes under specific growth conditions, with iron starvation leading to the formation of magnetosome membranes but not magnetite crystals (Komeili *et al.*, 2004; Lower and Bazylinski, 2013). This suggests that the cellular metabolism plays an important role in the production of the organelles. In the lab, *E. coli* is most commonly grown under aerobic conditions in LB media, which is not chemically defined. MTB are normally found in microaerobic environments (Lower and Bazylinski, 2013), which suggest a preferred anaerobic respiration mechanism. *M. gryphiswaldense* is cultured in media containing a high concentration of nitrate, which is used by the organism as the terminal electron acceptor. *E. coli* is able to use nitrate as a terminal electron acceptor via the nitrate reductase, a heterologous complex that contains a cytochrome-c protein. Under aerobic growth conditions, *E. coli* does not produce the enzymes required for cytochrome-c maturation. These proteins have a covalently attached heme c group and have been shown to be necessary for magnetosome biomineralisation, which suggests that *E. coli* would only be able to produce the minerals under anaerobic growth conditions or if the cytochrome-c maturation machinery is expressed artificially. In order to test this, *E. coli* was grown in the media used to grow MTB (*M. gryphiswaldense* and *M. magneticum*) under microaerobic conditions. Further constructs were made prior

to experiments in the MTB media in order to either try to lower the expression level of the proteins (using the lactose promoter) or to introduce additional proteins suggested to be involved in the process (the *mms6* operon). Electron microscopy experiments were carried out which showed the presence of electron-dense particles in all samples, including the control (**Figure 3.4**). This can be explained in several ways. (1) During media preparation, the metals in the media may have been oxidised or reacted with other salts present in the media to form insoluble, electron-dense particles. (2) *E. coli* might be able to reduce some metals during anaerobic growth, producing metal particles. This process would be very specific to the growth conditions used. Although small, non-defined particles were present in the control, more of these particles were observed in the cells containing the plasmid with the *mamAB* operon. Furthermore, when the construct containing both *mamAB* and *mms6* operons was produced, better defined particles were observed and the total amount of particles per cell was the highest. The particles observed in this construct were also sometimes organised in chain-like structures, which resembled magnetosomes. None of the samples exhibited magnetotaxis, which could be explained by the need for further proteins in order to produce the mature, magnetic crystals.

Recombinant magnetosome production has been shown in *R. rubrum*, an organism that does not normally produce these organelles (Kolinko, 2014). The work concludes that expression of *mamAB* operon is not sufficient for magnetite crystallisation, which was only achieved when *mamAB*, *mamGFDC* and *mms6* operons were inserted into the organism, although expression of *mamAB* only with *mms6* was not explored. Mature crystals were only obtained when all four operons (*mamAB, mamGFDC, mms6* and *mamYX*) were present in the organism.

Overall, the work discussed in this chapter has provided a number of interesting insights into the effects of the *mamAB* operon on the membranous morphology of *E. coli*. Furthermore, the work suggests the possibility of mineral formation when only *mamAB* and *mms6* operons are expressed, although identification of these minerals needs to be carried out, which is being carred out in future work. Due to the mutation present in the operon construct, it was decided to focus on synthetic reconstruction and characterisation of the operons involved in the process, which is discussed in the following chapter.

Chapter 4

Analysis, Cloning and Expression of Magnetosome Proteins: a Modular Approach

4.1. Introduction

Magnetosome formation has been shown to be a complex process which is still not fully understood (Komeili, 2012). Up to 30 proteins (**Table 4.1**) from the MAI are predicted to be directly involved in organelle formation and maintenance (Lohbe *et al.,* 2014). Furthermore, it is clear that proteins involved in other cellular processes, such as redox processes and iron transport, also contribute to the process. For example, the periplasmic nitrate reductase has been shown to be involved in magnetite biomineralisation (Li *et al.*, 2012) and insertion of a ferrous iron transporter into an organism, engineered to produce magnetosomes, improves crystal growth (Kolinko *et al.*, 2014). To date, there has been no comprehensive analysis on how MAI proteins behave in *E. coli*.

E. coli is one of the most commonly used organisms for protein production. A large variety of foreign proteins have been successfully produced in this bacterium. Unfortunately, not all proteins can be produced correctly in *E. coli* and the reasons behind this are not fully understood but likely reflect the ability of the foreign protein to fold within the internal environment. Other possibilities can also include incorrect targeting, lack of suitable chaperones, protease cleavage and poor solubility. Extensive studies have been carried out to characterise the metabolism of *E. coli* (Keseler *et al.*, 2013). It is well established that the organism can grow both aerobically and anaerobically, using a variety of molecules as terminal electron acceptors. Therefore, if the genes, involved in magnetosome formation, are correctly produced in this model bacterium, this respiratory flexibility should allow for the identification of growth conditions where mineral formation is possible.

In this chapter, the proteins from the four operons directly involved in magnetosome synthesis (*mms6, mamGFDC, mamAB* and *mamXY*), are analysed. The analysis includes a bioinformatics approach coupled with the cloning and recombinant expression of the genes in *E. coli*. The chapter also includes the reconstruction of all the operons in a modular manner in vectors compatible with expression in *E. coli*.

Table 4.1. General information about the MAI proteins.

Protein	Operon	Length (bp)	Length (aa)	Molecular Weight (kDa)	pI
MamG	*mamGFDC*	255	84	7.72	8.61
MamF	*mamGFDC*	336	111	12.34	9.14
MamD	*mamGFDC*	945	314	30.26	9.84
MamC	*mamGFDC*	378	125	12.43	5.07
4070	*mms6*	1350	449	48.04	8.17
4071	*mms6*	1044	347	36.38	5.94
MmsF	*mms6*	375	124	13.78	9.26
Mms6	*mms6*	411	136	12.76	9.14
4074	*mms6*	273	90	9.73	9.68
MamH	*mamAB*	1289	428	45.67	7.02
MamI	*mamAB*	234	77	7.16	8.74
MamE	*mamAB*	2319	772	78.05	8.19
MamJ	*mamAB*	1401	466	48.52	4.00
MamK	*mamAB*	1083	360	39.20	5.45
MamL	*mamAB*	372	123	8.59	12.19
MamM	*mamAB*	957	318	34.49	5.88
MamN	*mamAB*	1314	437	46.18	6.66
MamO	*mamAB*	1899	632	65.39	6.50
MamP	*mamAB*	813	270	28.36	7.13
MamA	*mamAB*	654	217	24.01	5.70
MamQ	*mamAB*	819	272	30.03	6.12
MamR	*mamAB*	219	72	8.06	8.31
MamB	*mamAB*	894	297	31.96	5.38
MamS	*mamAB*	543	180	18.72	6.78
MamT	*mamAB*	525	174	18.89	9.65
MamU	*mamAB*	894	297	32.00	9.32
MamY	*mamXY*	909	302	40.90	4.80
MamX	*mamXY*	1851	616	28.22	5.88
MamZ	*mamXY*	810	269	70.55	9.50
FtsZ-like	*mamXY*	1116	371	32.33	4.90

4.2. Results

4.2.1. Bioinformatics

A bioinformatics approach was initially employed to gain some preliminary information concerning the potential function of the proteins associated with magnetosome formation. These programmes included: TMHMM2 (http://www.cbs.dtu.dk/services/TMHMM/) and SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), which were used to investigate protein topology and targeting, and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), which was used to investigate any predictable structural similarity to proteins of known function.

TMHMM2 is currently the best ranked freely available transmembrane prediction program and is able to predict membrane protein topology with very high reliability (Moller *et al.*, 2001). The newest version of the program has been shown to predict 97%-98% of all tested transmembrane helices, although the specificity is lower when signalling peptides are present (Krogh *et al.*, 2001). SignalP 4.1 can be used to predict these signalling peptides and was designed to distinguish between transmembrane helices and signal peptides which share similar properties (Petersen *et al.*, 2011). These methods were used in combination to predict the number of transmembrane helices (**Table 4.2**), protein topology (**Table 4.2** and **Supplementary figure 1**) and the presence of any signalling peptides present (**Table 4.2**) for all of the magnetosome proteins.

I-TASSER (Iterative Threading ASSEmbly Refinement) is currently the best ranked (http://www.predictioncenter.org/casp11/) freely available protein structure and function prediction software (Yang *et al.,* 2015). The software generates a structural model based on the available structures of homologues. It also predicts biological function by analysing the enzyme commission (EC) numbers, gene ontology (GO) vocabulary and the ligand-binding sites of the templates used in the structural prediction and structurally similar proteins. The method was used to generate models for all MAI proteins. The most informative predictions are outlined in **Table 4.3**, with the full table available as supplementary material (**Supplementary table 2**).

Table 4.2. Transmembrane helix (TMHs), largest protein proportion localisation and signalling peptide predictions.

Protein	Operon	TMHs	Largest protein proportion localisation	Signalling peptide
MamG	*mamGFDC*	2	Transmembrane	-
MamF	*mamGFDC*	3	Transmembrane	-
MamD	*mamGFDC*	1	Periplasm	-
MamC	*mamGFDC*	2	Transmembrane	-
4070	*mms6*	2	Cytoplasm	-
4071	*mms6*	1	Periplasm	-
MmsF	*mms6*	3	Transmembrane	-
Mms6	*mms6*	1	Periplasm	-
4074	*mms6*	0	Cytoplasm	-
MamH	*mamAB*	12	Transmembrane	-
MamI	*mamAB*	2	Transmembrane	-
MamE	*mamAB*	1	Periplasm	-
MamJ	*mamAB*	0	Cytoplasm	-
MamK	*mamAB*	0	Cytoplasm	-
MamL	*mamAB*	2	Transmembrane	-
MamM	*mamAB*	3	Periplasm	-
MamN	*mamAB*	11	Transmembrane	-
MamO	*mamAB*	8	Periplasm	-
MamP	*mamAB*	1	Periplasm	-
MamA	*mamAB*	0	Cytoplasm	-
MamQ	*mamAB*	1	Cytoplasm	-
MamR	*mamAB*	0	Cytoplasm	-
MamB	*mamAB*	3	Periplasm	-
MamS	*mamAB*	1	Cytoplasm	-
MamT	*mamAB*	1	Cytoplasm	-
MamU	*mamAB*	0	Cytoplasm	-
MamY	*mamXY*	2	Cytoplasm	-
MamX	*mamXY*	1	Periplasm	-
MamZ	*mamXY*	18	Transmembrane	-
FtsZ-like	*mamXY*	0	Cytoplasm	-

Table 4.3. I-TASSER functional, biological and localisation predictions.
Accuracy refers to the structural prediction confidence by the software (out of 1). Functional prediction confidence (out of 1) is shown in brackets.

Protein	Accuracy	Molecular Function	Biological Process	Cellular Location
MamF	0.35±0.12	Electron transporter (0.15)	Biological adhesion (0.51)	Cytoskeleton (0.42)
		Oxidoreductase (0.15)		
		Heme binding (0.15)	Cellular physiology process (0.51)	
		Structural constituent of the cytoskeleton (0.14)		
MamD	0.39±0.13		Regulation of cell shape (0.35)	Cytoplasm (0.35)
MamC	0.43±0.14	FAD binding (0.71)	Protein tetramerization (0.41)	Organelle membrane (0.41)
		Acyl-CoA dehydrogenase (0.71)	Homo-oligomerization (0.41)	Mitochondrial envelope (0.41)
		Fatty acid binding (0.41)		Mitochondrial matrix (0.36)
4070	0.51±0.15	Protein binding (0.76)	Golgi vesicle transport (0.35)	COPI-coated vesicle (0.35)
			Protein recruitment (0.35)	Golgi-associated vesicle membrane (0.35)
				Vesicle coat (0.35)
4071	0.36±0.12	Protein binding (0.53)	Actin bundling (0.12)	Cytoskeleton (0.12)
MmsF	0.29±0.09	Structural molecule (0.14)		Transmembrane (0.15)
				Actin cytoskeleton (0.14)
Mms6	0.39±0.13	Zinc ion binding (0.41)		Periplasm (0.31)
4074	0.26±0.08	Metalloexopeptidase (0.07)	Proteolysis (0.07)	Peripheral membrane protein (0.07)
		Actin binding (0.07)		Extracellular (0.07)
MamH	0.61±0.14	Solute:proton symporter (0.51)	Transmembrane transport (0.67)	Transmembrane (0.83)
		Cation:sugar symporter (0.51)		Inner membrane (0.76)
				Macromolecular complex (0.54)
MamI	0.34±0.11	Translation termination factor (0.11)	Gene expression (0.57)	Macromolecular complex (0.39)
			Cellular protein metabolism (0.50)	Intracellular non-membrane-bounded organelle (0.39)
				Cytoplasm (0.39)
MamE	0.44±0.14	Serine-type endopeptidase (0.63)	Proteolysis involved in protein catabolism (0.51)	Cell envelope (0.51)

MamE (cont.)	0.44±0.14	Protein binding (0.51)	Response to chemical (0.51)	External encapsulating structure (0.51)
MamJ	0.51±0.15	Enzyme inhibitor (0.48)		Cytoplasm (0.36)
		Protein binding (0.42)		Extracellular (0.31)
		Endopeptidase regulator (0.32)		
		Ion binding (0.32)		
MamK	0.77±0.10	ATP binding (0.98)	Adherens junction organization (0.52)	Cytosol (0.52)
		Identical protein binding (0.52)	Chaperone mediated protein folding (0.52)	Microtubule associated complex (0.52)
		ADP binding (0.52)		Actomyosin, actin portion (0.52)
		ATPase activity (0.52)		
		Structural constituent of cytoskeleton (0.52)		
MamM	0.72±0.11	Wide pore channel activity (0.57)	Siderophore transport (0.57)	Inner membrane (0.72)
		Ferrous iron transmembrane transporter (0.44)	Establishment of protein localization (0.57)	Transmembrane (0.72)
			Ferrous iron transport (0.44)	
MamN	0.52±0.15	Intramolecular oxidoreductase (0.48)		Cell periphery (0.48)
		Oxidoreductase (0.39)		Cytoplasm (0.35)
		Metal cluster binding (0.39)		
		Ion binding (0.39)		
MamO	0.43±0.14	Serine-type endopeptidase (0.55)	Cellular response to topologically incorrect protein (0.60)	Intrinsic component of membrane (0.60)
			Response to misfolded protein (0.60)	Plasma membrane (0.60)
		Protein binding (0.48)	Proteolysis (0.55)	Periplasm (0.40)
MamP	0.65±0.13	Protein binding (0.75)	Cellular response to topologically incorrect protein (0.55)	Intrinsic component of membrane (0.55)
			Proteolysis involved in cellular protein catabolic process (0.55)	Plasma membrane (0.55)
				Cell envelope (0.53)
		Serine-type endopeptidase (0.47)	Response to misfolded protein (0.53)	Periplasm (0.47)
MamA	0.71±0.12	Identical protein binding (0.41)	Copper-induced intracellular protein transport (0.39)	Mitochondrial outer membrane (0.39)

MamQ	0.35±0.12		Fatty acid catabolic process (0.35)	Mitochondrial inner membrane (0.09)
			Fatty acid oxidation (0.35)	Peroxisome (0.09)
MamR	0.57±0.15	DNA binding (0.64)	Transcriptional repression (0.51)	Cytoplasm (0.74)
		Obsolete transcription repressor activity (0.51)	Response to redox state (0.51)	
MamB	0.99±0.04	Ferrous iron transmembrane transporter (0.50)	Ferrous iron transport (0.50)	Integral component of membrane (0.92)
		Lipid binding (0.47)	Protein transport (0.37)	Plasma membrane (0.83)
		Porin (0.37)		Cell outer membrane (0.67)
MamS	0.29±0.09	Nucleic acid binding transcription factor (0.40)	Regulation of gene expression (0.40)	Intracellular (0.40)
		ATP binding (0.36)	Inorganic anion transport (0.40)	
MamT	0.31±0.10	Electron carrier (0.07)	Oxidation-reduction (0.13)	Cytoplasm (0.09)
		Heme binding (0.07)		
MamU	0.90±0.06	Diacylglycerol kinase (0.99)	Signalling pathway (0.99)	Cytoplasm (0.88)
		ATP binding (0.95)	Phospholipid biosynthesis (0.95)	
		Metal ion binding (0.95)	Lipid phosphorylation (0.88)	
		NAD+ kinase (0.94)		
		Lipid kinase (0.88)		
MamX	0.28±0.09	Ion binding (0.60)		Intracellular (0.47)
MamY	0.97±0.05		Regulation of gene expression (0.35)	Integral component of membrane (0.32)
			Signalling (0.32)	
MamZ	0.52±0.15	Hexose:proton symporter (0.57)	Hexose transport (0.57)	Integral component of membrane (0.51)
		2 iron, 2 sulphur cluster binding (0.41)	Response to metal ion (0.46)	Peroxisome (0.41)
		Iron ion binding (0.41)		Plasma membrane (0.41)
		FAD binding (0.41)		
		Electron carrier (0.41)		Macromolecular complex (0.34)
		Protein homodimerization (0.41)		
FtsZ-like	0.64±0.13	GTPase (1.00)	GTP hydrolysis (1.00)	Cytoplasm (1.00)
		GTP binding (1.00)	Protein polymerization (1.00)	Plasma membrane (0.77)
		Identical protein binding (0.79)	Barrier septum assembly (1.00)	Microtubule (0.75)
			Cell cycle (1.00)	

4.2.2. Cloning of All the Individual MAI Genes

Primers were designed for all magnetosome island genes to allow for their amplification by PCR. The primers contained flanking restriction enzyme sites for *NdeI* or *AseI* at the 5' and *SpeI* at the 3'. Once amplified, the DNA was digested using the corresponding restriction enzymes and ligated into a pET3a vector via the *NdeI/SpeI* sites. This allows the over-expression of the gene using the T7 promoter and the over-production of the protein using the optimised ribosome binding site present in the vector. Once cloned, the resultant plasmid was transformed into *E. coli* DH5α to allow amplification of the plasmid. Each individual plasmid was confirmed by restriction enzyme analysis and sequencing. More details on the cloning procedures and construct list can be found in the materials and methods section.

Two different *Magnetospirillum gryphiswaldense* MSR-1 genomes are available on the NCBI data base (http://www.ncbi.nlm.nih.gov/genome) and the sequences of the cloned magnetosome island genes were compared to both. Interestingly, a number of genetic differences were identified in the amplified genes when the sequences of the cloned genes were compared to these two available genomes (**Table 4.4**). For instance, protein 4070 was obtained with two mutations, which are not present in either of the genomes. Multiple amplifications yielded the same product so, therefore, it is assumed that this is a variation present in the genomic DNA used.

Table 4.4. Magnetosome protein sequence variation between available genomes.
Comparison of the magnetosome protein differences in two available *Magnetospirillum gryphiswaldense* genomes and the cloned variant, using MSR-1 genome as the reference.

Protein	Operon	MSR-1 v2 genome variant	Cloned variant
MamG	*mamGFDC*	N/A (frame shift)	N/A (frame shift)
MamD	*mamGFDC*	Val^{227}->Ala	MSR-1 v2
MamC	*mamGFDC*	Glu^{81}->Gly	MSR-1 v2
4070	*mms6*	Identical	Lys^{338}->Arg; Lys^{350}->Arg
Mms6	*mms6*	Duplication of two segments (GKVGA and TKVVAAQGAG)	MSR-1 v2
4074	*mms6*	N/A	MSR-1
MamB	*mamAB*	Arg^{193}->His	MSR-1 v2

4.2.3. Production and SDS-PAGE Analysis of Individual MAI Proteins

In order to investigate the over-expression levels of individual proteins in *E. coli* all the constructs were individually transformed into *E. coli* BL21 (DE3) Star cells. These were then grown, induced and harvested as described in materials and methods. *E. coli* BL21 (DE3) Star cells expressing an empty pET3a vector was used as a control. The resulting whole cell, soluble and insoluble fractions for each strain were analysed using SDS-PAGE (**Figure 4.1, Figure 4.2, Figure 4.3 and Figure 4.4**) and their protein profiles were compared to the control. Only 7 out of 29 analysed constructs showed observable protein overproduction using this method. In order to detect poorly produced proteins, a western blot approach was employed as described in the following section.

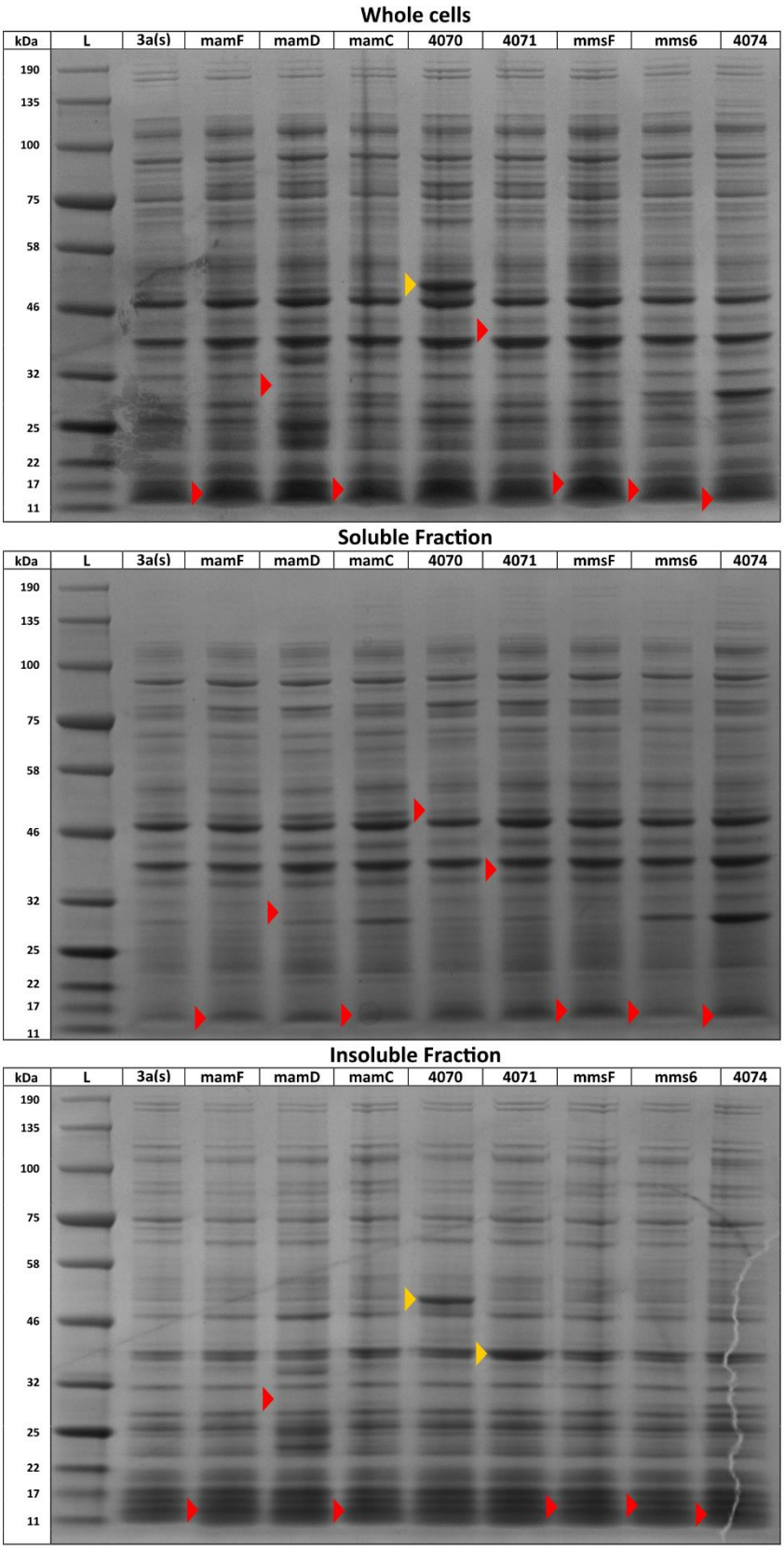


Figure 4.1. SDS-PAGE analysis of proteins from the *mamGFDC* and *mms6* operons.
L is Ladder; 3a(s) is the control sample of cells harbouring a pET3a vector. **Arrows** indicate predicted protein sizes. **Red** indicates no observable over-expression. **Yellow** indicates observable over-expression.

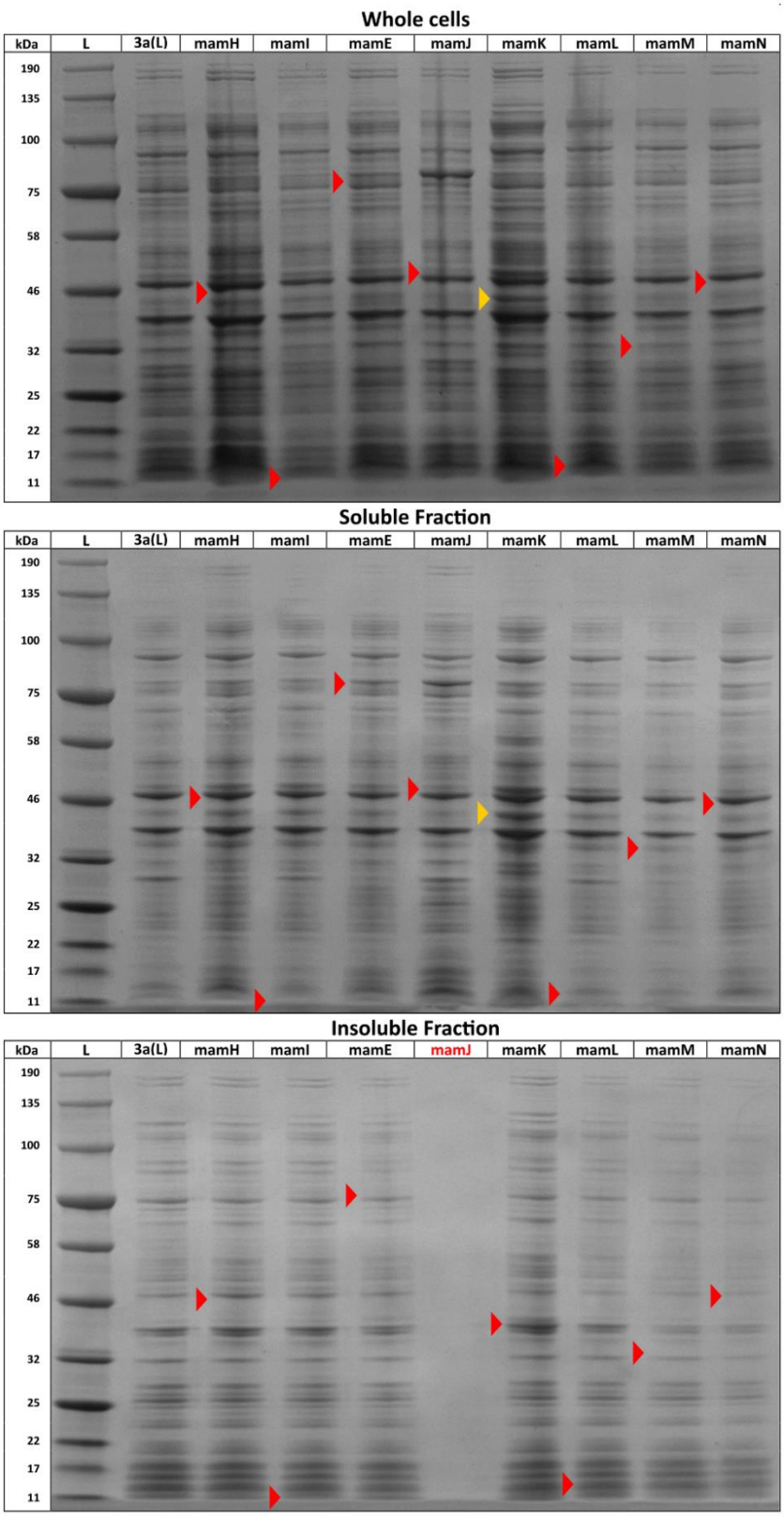


Figure 4.2. SDS-PAGE analysis of the first 8 proteins (except insoluble fraction of mamJ) from the *mamAB* operon.

L is Ladder; 3a(L) is the control sample of cells harbouring a pET3a vector. **Arrows** indicate predicted protein sizes. **Red** indicates no observable over-expression. **Yellow** indicates observable over-expression.

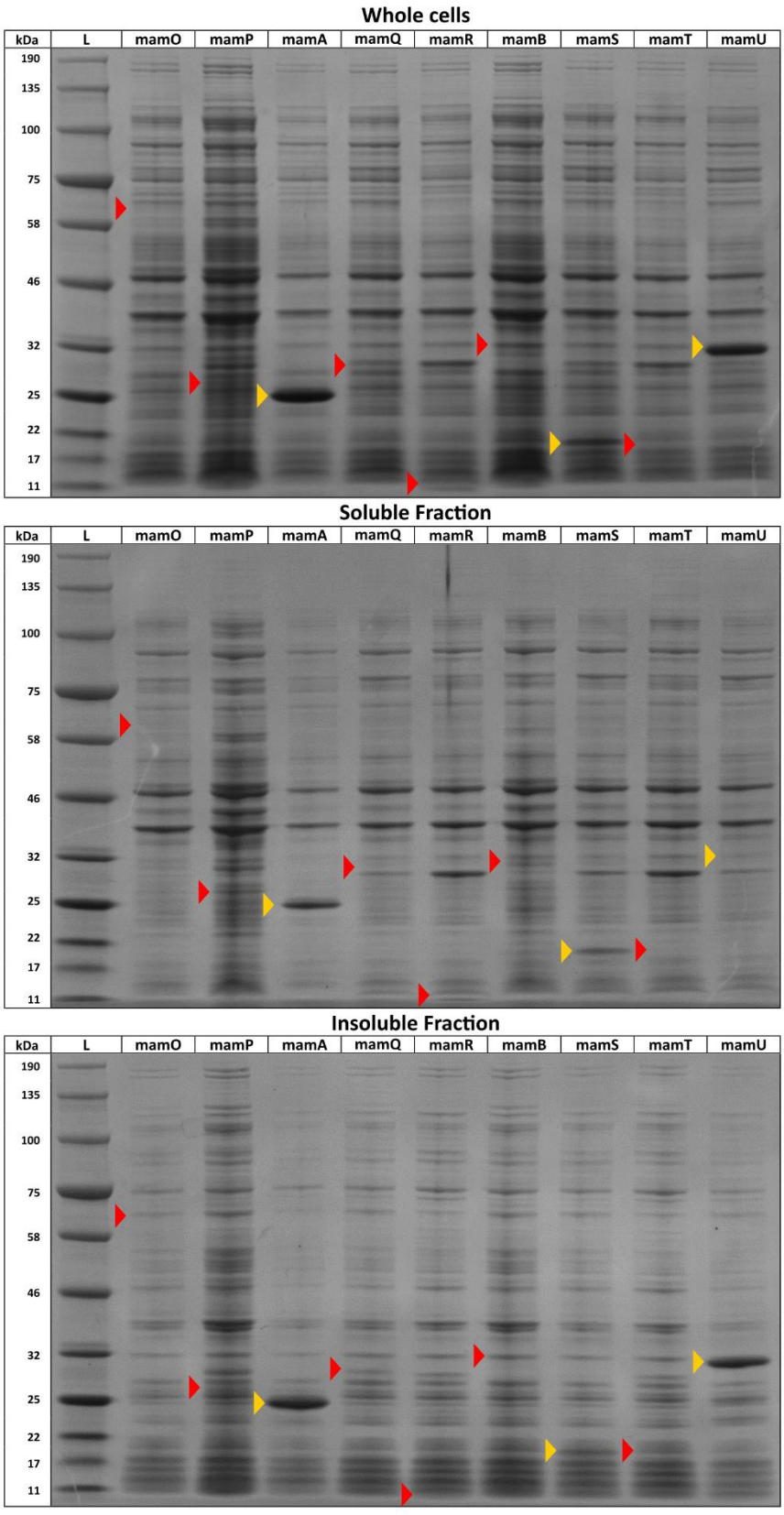


Figure 4.3. SDS-PAGE analysis of the last 9 proteins from the *mamAB* operon.
L is Ladder. **Arrows** indicate predicted protein sizes. **Red** indicates no observable over-expression. **Yellow** indicates observable over-expression.

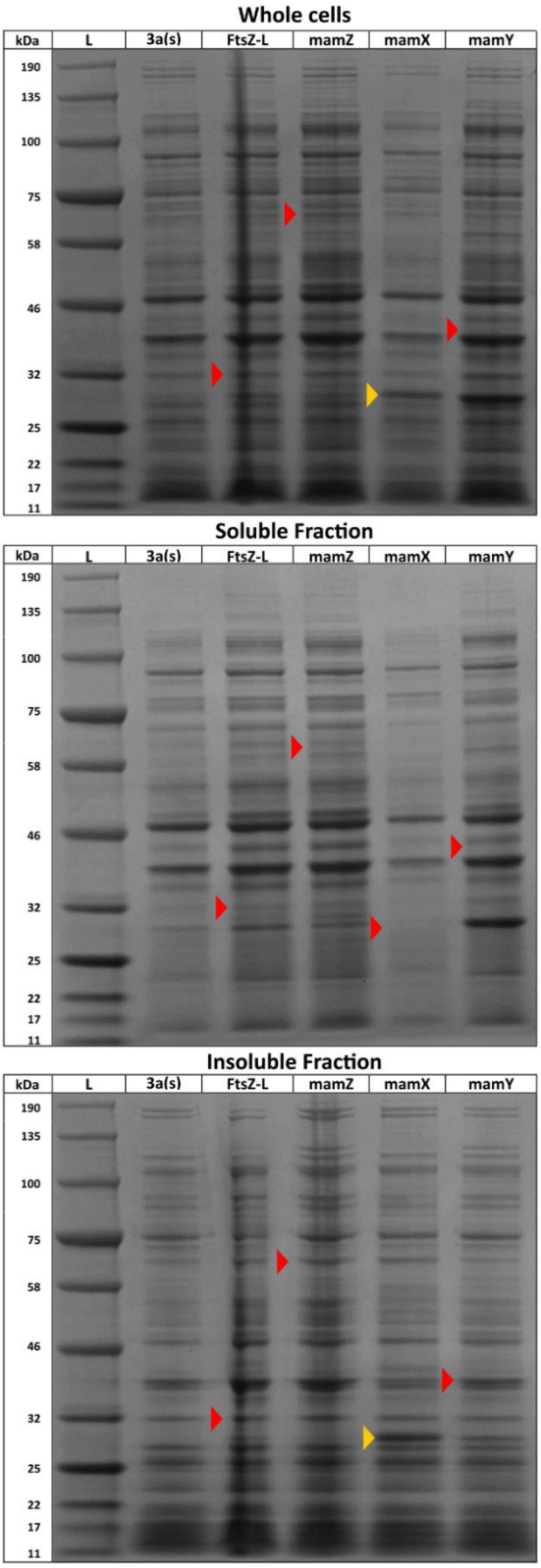


Figure 4.4. SDS-PAGE analysis of proteins from the *mamXY* operon.
L is Ladder; 3a(s) is the control sample of cells harbouring a pET3a vector. **Arrows** indicate predicted protein sizes. **Red** indicates no observable over-expression. **Yellow** indicates observable over-expression.

4.2.4. Western Blotting Analysis of Individual MAI Proteins

In order to use western blotting to analyse the MAI proteins they needed to be tagged with a suitable antigen. An N-terminus hexa-histidine tag was chosen due to the ease of sub-cloning the constructs into an appropriate vector (pET14b) and the availability of a specific antibody against the tag. Key *mamAB* operon proteins (**Table 4.5**) were chosen for analysis due to their roles in the organelle formation process. These were sub-cloned into a pET14b vector (*NdeI/SpeI* site) either from the previously constructed pET3a plasmids (*NdeI/SpeI* segments) or from directly amplified PCR products (*AseI/SpeI* segments). If PCR was used as part of the procedure, the resulting plasmids were sequenced.

Table 4.5. General information about the *mamAB* operon proteins analysed using Western Blotting.

Protein	Vector	Insert	Modification	Molecular Weight (kDa)
MamE	pET14b	pET3a.mamE	N-term 6xHis	80.21
MamJ	pET14b	PCR	N-term 6xHis	50.68
MamL	pET14b	PCR	N-term 6xHis	10.75
MamM	pET14b	pET3a.mamM	N-term 6xHis	36.65
MamN	pET14b	pET3a.mamN	N-term 6xHis	48.34
MamO	pET14b	PCR	N-term 6xHis	67.55
MamP	pET14b	pET3a.mamP	N-term 6xHis	30.52
MamA	pET14b	pET3a.mamA	N-term 6xHis	26.17
MamQ	pET14b	PCR	N-term 6xHis	32.19
MamR	pET14b	pET3a.mamR	N-term 6xHis	10.22
MamB	pET14b	pET3a.mamB	N-term 6xHis	34.12
MamS	pET14b	pET3a.mamS	N-term 6xHis	20.88
MamT	pET14b	pET3a.mamT	N-term 6xHis	21.05
MamU	pET14b	pET3a.mamU	N-term 6xHis	34.16
MamM	pET23b	PCR	C-term 6xHis	34.51
MamB	pET23b	PCR	C-term 6xHis	31.98

The plasmids that were constructed were individually transformed into *E. coli* BL21 (DE3) Star cells. These were then grown, induced and harvested under the same conditions as previously described. A western blot procedure, using an anti-hexa-histidine tag primary antibody, was carried out for each of the strains, as described in materials and methods section. The blots are shown in **Figure 4.5** and **Figure 4.6**. This procedure was followed for all proteins except for MamQ, which, due to a rather unusual membranous phenotype, that was observed during electron microscopy experiments (discussed in the following chapter), was analysed in a range of cell fractions.

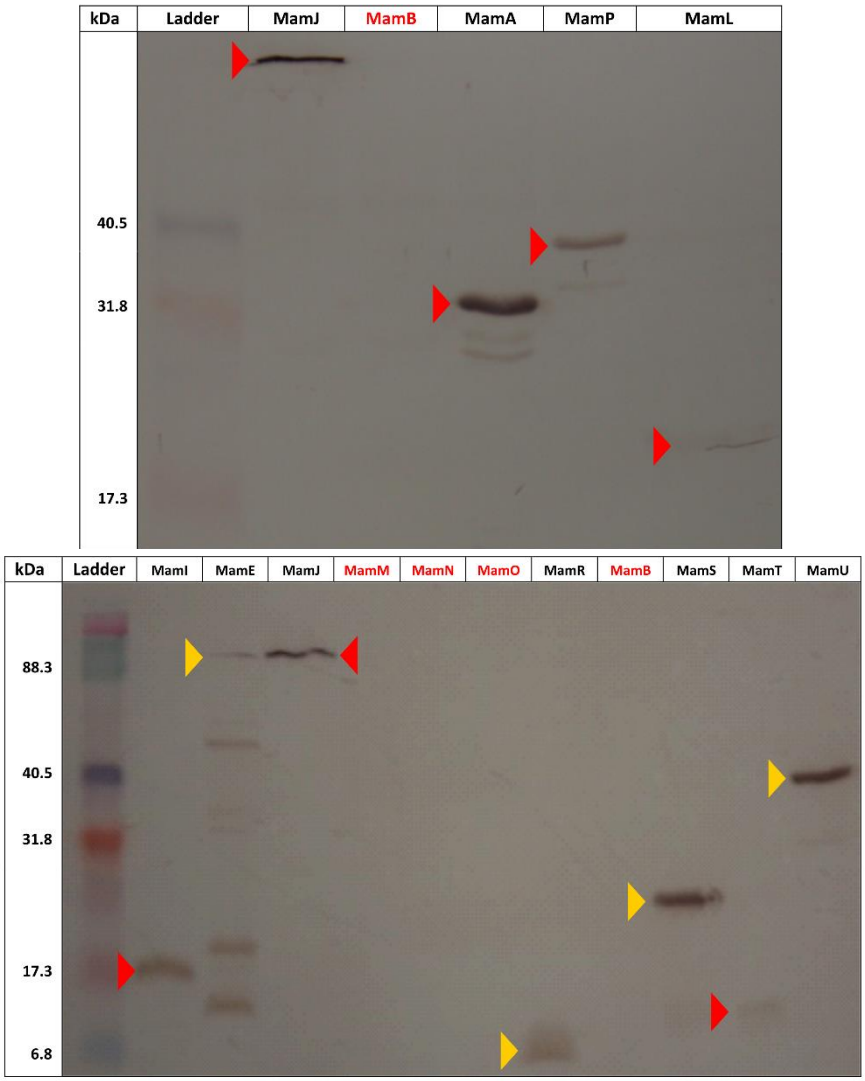


Figure 4.5. Western blot analysis of N-terminus hexa-histidine tagged proteins from the *mamAB* operon.
Arrows indicate identified bands of interest. **Yellow** indicates the band is of predicted size. **Red arrow** indicates the band is of different size from predicted. **Red text** indicated no bands observed.

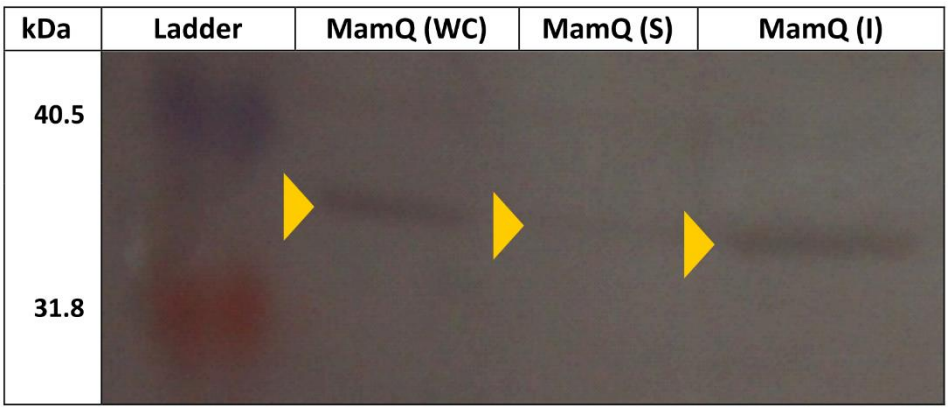


Figure 4.6. Western blot analysis of N-terminus hexa-histidine tagged MamQ protein. WC – whole cells; S – soluble fraction; I – insoluble fraction. **Arrows** indicate identified bands. **Yellow** indicates the band is of predicted size.

From this study of 14 constructs, 11 of the strains showed bands that were immuno-reactive with the anti-His antibody. Of these, only 5 of the constructs produced bands of the expected molecular mass. Furthermore, the strain overproducing MamE revealed a large number of lower sized bands (**Figure 4.5.**) indicative of protein cleavage or degradation.

The presence of an N-terminus hexa-histidine tag, could have interfered with the insertion of membranous proteins into the membrane. To investigate this possibility further, additional constructs of the key predicted membrane proteins, MamM and MamB, were generated with a C-terminus hexa-histidine tag. This was achieved by cloning the PCR products of *mamM* and *mamB* (*NdeI/XhoI* segments) into a pET23b vector (*NdeI/XhoI* site) (**Table 4.5.**). The same western-blot procedure was used to analyse the resultant strains generated by transforming these plasmids into *E. coli* BL21 (DE3) Star cells. The relevant blots are shown in **Figure 4.7**.

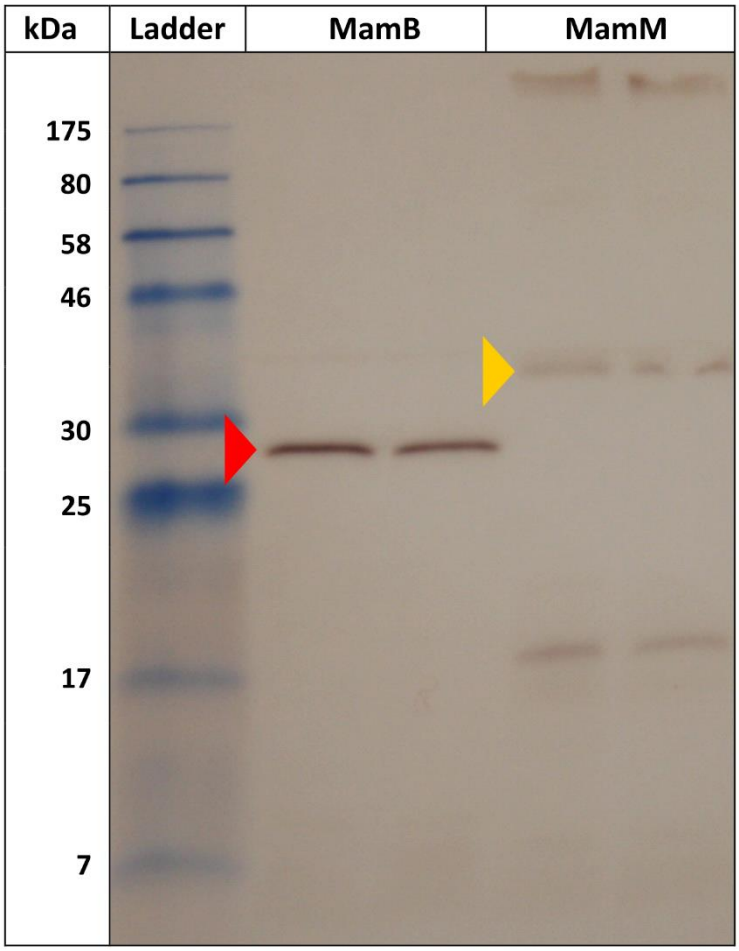


Figure 4.7. Western blot analysis of C-terminus hexa-histidine tagged MamB and MamM proteins. Arrows indicate identified bands of interest. **Yellow** indicates the band is of predicted size. **Red** indicates the band is of different size from predicted.

When a C-terminus hexa-histidine tag was used to detect recombinant MamM and MamB, both proteins could be detected by western analysis. The results confirm that an N-terminus hexa-histidine tag appears to interfere with the correct production of these two membranous proteins.

Overall, it was therefore possible to identify the recombinant production of 12 out of the 14 analysed constructs using western blotting. The 2 proteins with no-observed expression, MamO and MamN, are also both predicted membrane proteins and should also be analysed using C-terminal hexa-histidine tag constructs, but due to time constrains the analysis was not carried out. This analysis suggest that the majority of MAI proteins can be produced recombinantly in *E. coli*, but many are only found at low concentrations and cannot be detected using standard SDS-PAGE procedures.

4.2.5. C Type Cytochrome Proteins

Four cytochrome c proteins have been identified in the MAI, and are characterised by the presence of CXXCH motifs within their sequence. All of these have a suggested role in biomineralisation of magnetite within magnetosomes. The sequence of the cytochrome c domains of these proteins suggests they fall into two different groups (**Figure 4.8 (A)**). Further motif searches were carried out for these proteins and these are outlined in **Figure 4.8 (B)**.

In order to investigate if these cytochromes are processed correctly in *E. coli* it was decided to overproduce the soluble domains of the proteins fused to an affinity tag (hexa-histidine) to allow for easier purification.

The DNA corresponding to predicted soluble domain of MamE$^{\Delta 1\text{-}55}$ (MamE-soluble), was cloned into a pET14b (*NdeI/SpeI* site) after amplification by PCR with primers containing the appropriate restriction enzyme sites (*NdeI/SpeI* segment). The correct plasmids were identified by restriction enzyme analysis and confirmed by sequencing of the relevant clones. *E. coli* BL21 (DE3) Star cells were used to overproduce the protein, which was then purified using nickel affinity chromatography. UV-visible spectrum analysis of purified and reduced protein, together with SDS-PAGE and western blotting results for N-terminus hexa-histidine tagged MamE-soluble are shown in **Figure 4.9**. The protein appeared highly smeared on the western blot suggesting either cleavage by a peptidase or poor denaturation. No α, β or γ peaks, indicative of the presence of a heme group, a characteristic of cytochrome c proteins, were observed. This suggests that no incorporation of heme into the cytochrome had taken place in the cytoplasm of *E. coli*, which is to be expected as the maturation machinery for this process is located in the periplasm of the organism.

In order to investigate how the cytochrome c proteins behave in the periplasm, constructs for the fusion of an N-terminus periplasmic targeting sequence onto the protein together with a C-terminus hexa-histidine tag were made. Briefly, the periplasmic targeting sequence of BtuF (first 32 amino-acids) from *E. coli* was cloned in a pET23b (*NdeI/SacI* site) vector followed by insertion of the soluble domains of the cytochrome c proteins (*SacI/XhoI* site). All constructs were confirmed by restriction enzyme analysis and sequencing. The proteins were overproduced and purified section. As observed previously, no α, β or γ peaks associated with heme were observed for the isolated proteins.

A literature search revealed that *E. coli* only produces the cytochrome c machinery (*Ccm* operon) under very specific growth conditions and hence in order to overproduce cytochrome c proteins, the machinery needs to be constitutively expressed from a plasmid (Arslan *et al.,* 1998). A plasmid containing the *Ccm* operon, contained within a plasmid (pEC86), was obtained from the laboratory of Professor Stuart Ferguson. The periplasmic targeted soluble domains of the cytochrome c proteins were produced in *E. coli* BL21 (DE3) Star, harbouring the pEC86 plasmid, and purified. Purified MamP, MamT and MamX gave UV spectra showing the presence of heme, which could be reduced by the addition of dithionite, suggesting correct heme insertion (**Figure 4.10**). Unfortunately, MamE did not show the presence of any heme prosthetic group, possibly due to the degradation of the protein.

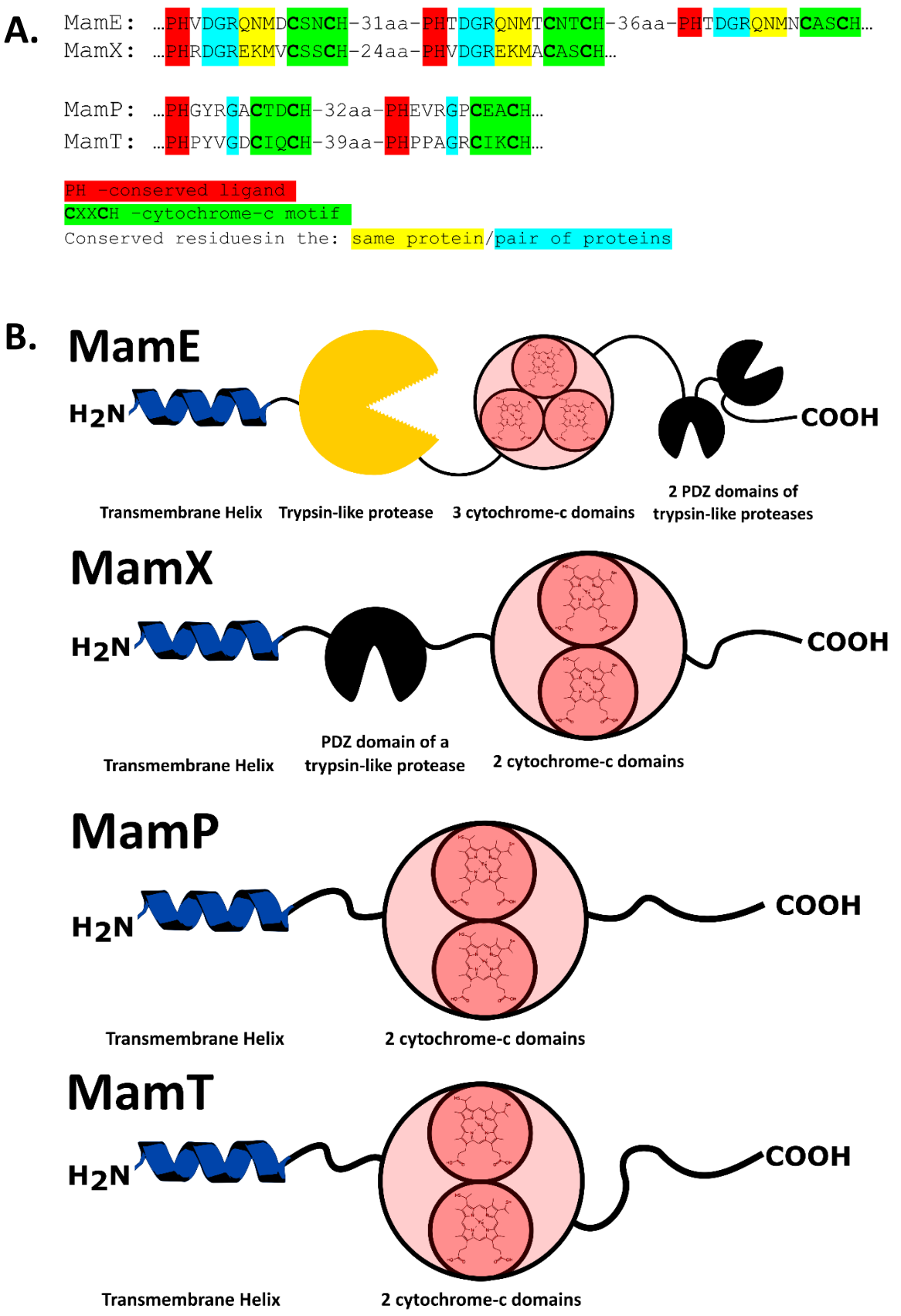
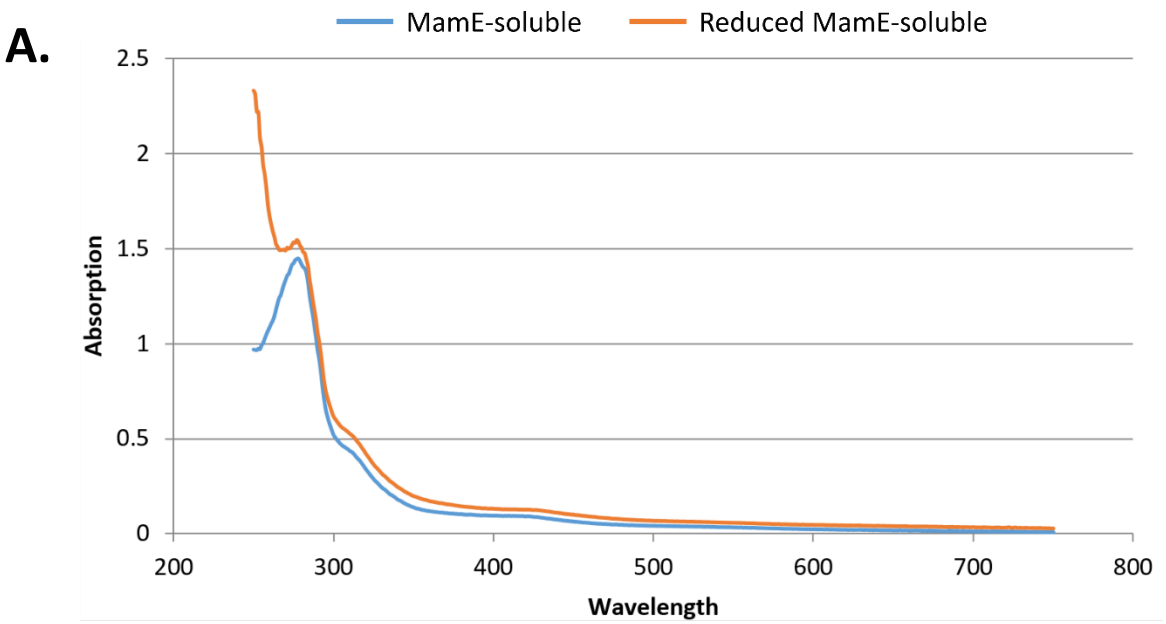
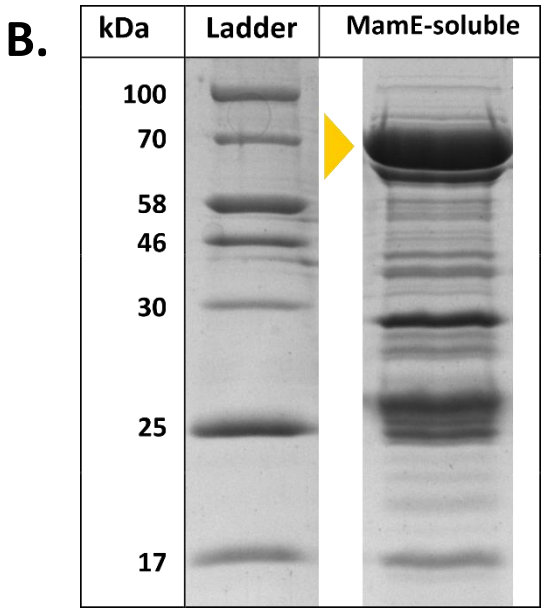


Figure 4.8. Sequence analysis of c type cytochromes from the MAI.
(A) Cytochrome c motifs of MAI proteins. **(B)** Other motifs present in the analysed cytochrome c proteins.





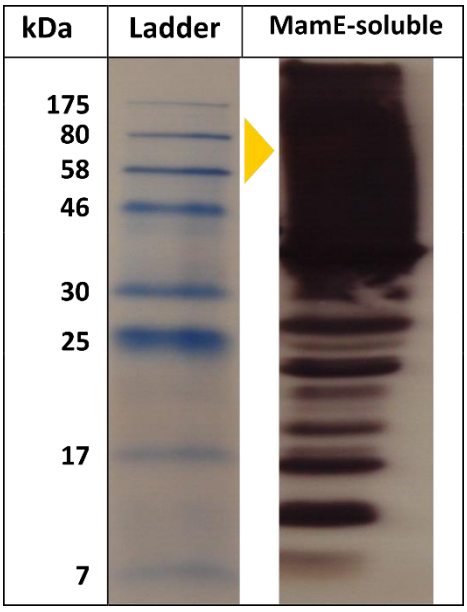


Figure 4.9. Analysis of purified MamE-soluble.
(A) UV spectrum of native and reduced protein. **(B)** SDS-PAGE. **(C)** Western blot. Arrows indicate expected band for MamE-soluble (74.28kDa)

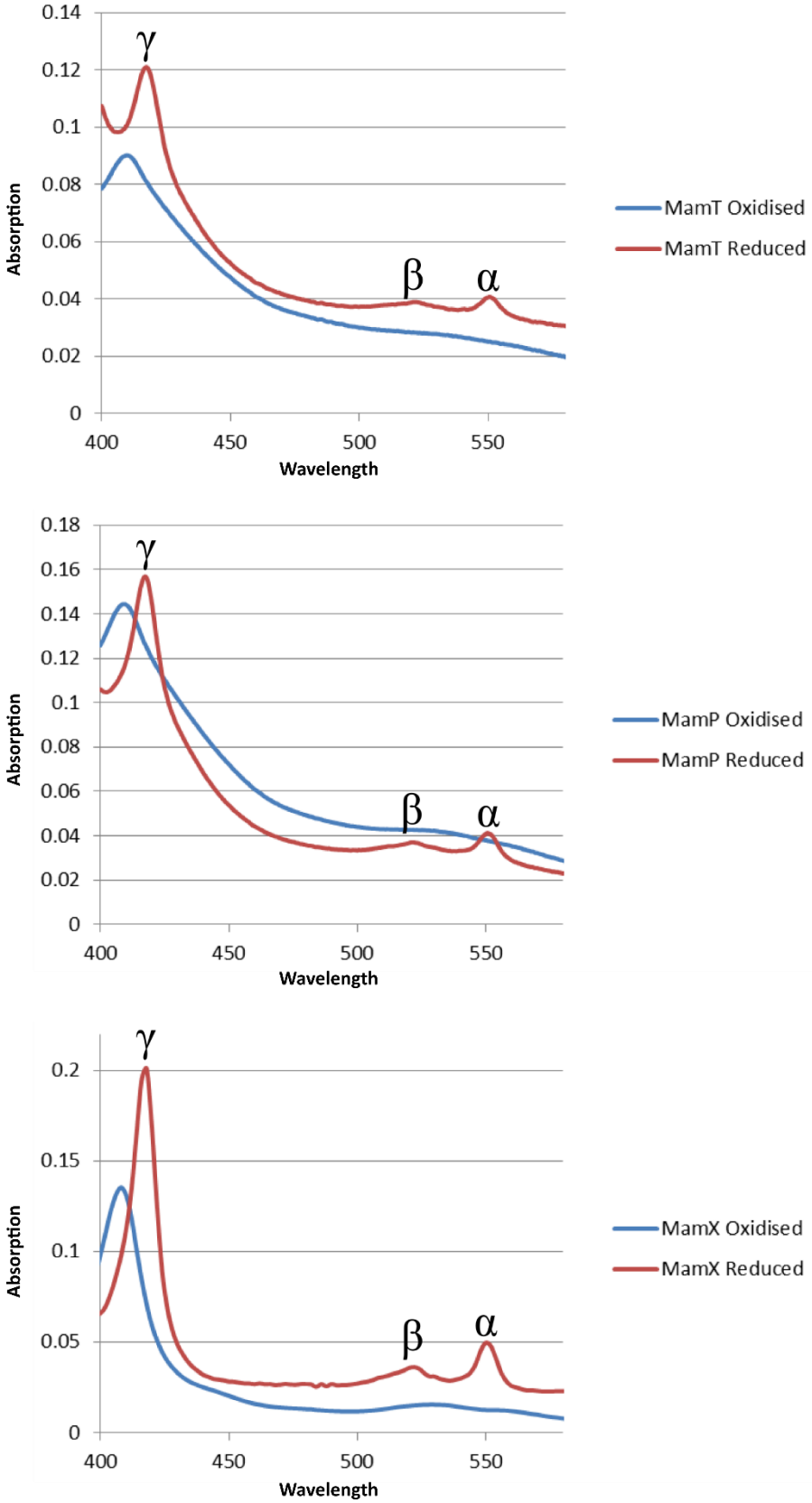


Figure 4.10. UV spectra of 3 different cytochrome c proteins from the MAI.
α, β and γ peaks, indicative of cytochrome c proteins, are highlighted.

4.2.6. Combining Individual Gene Constructs

To produce synthetic variants of all the MAI operons (*mamGFDC, mms6, mamAB* and *mamXY*) a previously developed method, Link and Lock, was employed which relies on having all the genes of interest in the pET3a vector (McGoldrick *et al.,* 2005). Briefly, by using an upstream *XbaI* site on the insert and a downstream *SpeI* site in the vector, together with a downstream restriction site in the vector backbone (usually *ScaI*), it is possible to insert a new gene in the synthetic operon (**Figure 4.11**). During the cloning process the old *XbaI* and *SpeI* sites are destroyed, but new ones are introduced at the 5' and the 3' of the operon respectively. The process can be repeated to introduce further genes. All genes of interest are cloned into the same transcript, expressed from a T7 promoter and have individual ribosome binding sites, which come from the vector. Furthermore, inserts can also be made at the 5' by exchanging the enzymes used for the vector and insert.

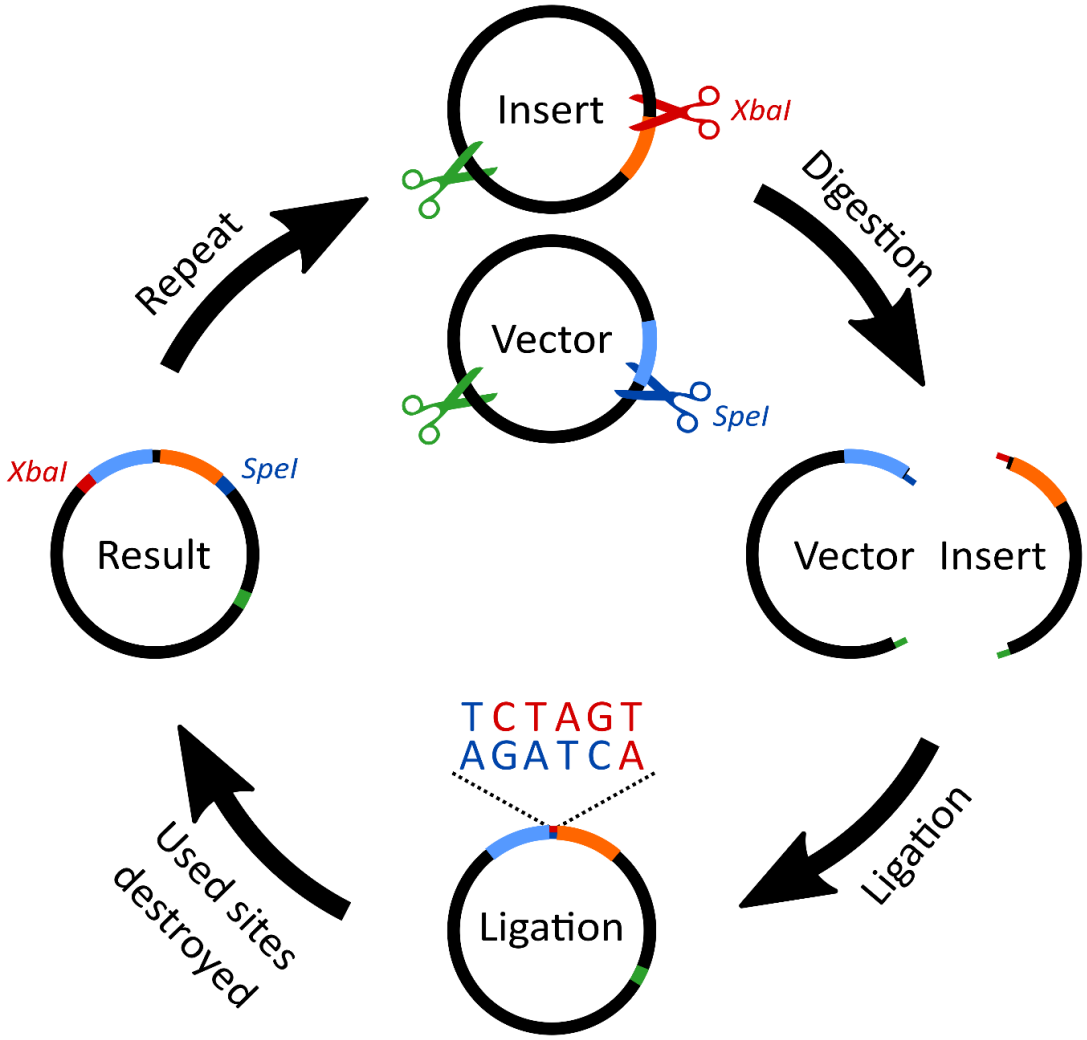


Figure 4.11. Cloning strategy used to create synthetic operons.

This method was used to create a large library of synthetic operons (**Table 4.6**). In this manner it was possible to reconstruct fully the *mms6* and *mamXY* operons. The *mamGFDC* operon was reconstructed without *mamG*, as the gene contained a frame-shift mutation (**Table 4.4**).

Initially, the reconstruction of the *mamAB* operon without *mamH*, *mamJ*, *mamK* and *mamU* was planned, as these genes had previously been shown not to be necessary for organelle formation or biomineralisation (Murat *et al*., 2010a; Lohbe *et al.,* 2016). However, it was not possible to reconstruct the modified operon by employing the method discussed above. The largest construct obtained contained 10 out of 13 genes (**Table 4.6**). No transformants were obtained when attempts were made to add additional genes into the construct. One possible reason for this could be due to the leakiness of the T7 promoter, so alternative promoters were employed as outlined below.

Two different promoters were used, the previously discussed arabinose promoter (pET3a.TBAD) and a rhamnose promoter (pET3a.pRha). Surprisingly, it was even more difficult to obtain any larger constructs using these promoters (**Table 4.7**). Only two or four genes could be combined using the arabinose or rhamnose promoter respectively. Furthermore, transformants containing the pET3a.pRha.HIEL construct had a significant growth defect in that they formed microcolonies (<1mm diameter colonies from 16h growth as opposed to 2mm colonies when transforming the pET3a.HIEL construct).

Table 4.6. Multi-gene constructs made in the pET3a vector.

Operon	Construct	Genes (in construct order)
mms6	pET3a.70	*mg4070*
mms6	pET3a.70.71	*mg4070, mg4071*
mms6	pET3a.70.71.mF	*mg4070, mg4071, mmsF*
mms6	pET3a.m6.74	*mms6, mg4074*
mms6	pET3a.4070-74	*mg4070, mg4071, mmsF, mms6, mg4074*
mamGFDC	pET3a.FD	*mamF, mamD*
mamGFDC	pET3a.FDC	*mamF, mamD, mamC*
mamXY	pET3a.YX	*mamY, mamX*
mamXY	pET3a.YXZ	*mamY, mamX, mamZ*
mamXY	pET3a.YXZFtsZ	*mamY, mamX, mamZ, FtsZ-like*
mamAB	pET3a.HI	*mamH, mamI*
mamAB	pET3a.EL	*mamE, mamL*
mamAB	pET3a.EM	*mamE, mamM*
mamAB	pET3a.MN	*mamM, mamN*
mamAB	pET3a.NO	*mamN, mamO*
mamAB	pET3a.OP	*mamO, mamP*
mamAB	pET3a.PA	*mamP, mamA*
mamAB	pET3a.QB	*mamQ, mamB*
mamAB	pET3a.ST	*mamS, mamT*
mamAB	pET3a.HU	*mamH, mamU*
mamAB	pET3a.ELM	*mamE, mamL, mamM*
mamAB	pET3a.HIEL	*mamH, mamI, mamE, mamL*
mamAB	pET3a.IELM	*mamI, mamE, mamL, mamM*
mamAB	pET3a.NOPA	*mamN, mamO, mamP, mamA*
mamAB	pET3a.QBST	*mamQ, mamB, mamS, mamT*
mamAB	pET3a.RQBST	*mamR, mamQ, mamB, mamS, mamT*
mamAB	pET3a.PARQBST	*mamP, mamA, mamR, mamQ, mamB, mamS, mamT*
mamAB	pET3a.NOPARQBST	*mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS, mamT*
mamAB	pET3a.MNOPARQBST	*mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS, mamT*

Table 4.7. Constructs made in attempt to reconstruct the *mamAB* operon.

Vector	Construct	Genes (in construct order)
pET3a.TBAD	pET3a.TBAD.HI	*mamH, mamI*
pET3a.pRha	pET3a.pRha.HI	*mamH, mamI*
pET3a.pRha	pET3a.pRha.HIEL	*mamH, mamI, mamE, mamL*
pZS24	pZS24.U	*mamU*
pZS24	pZS24.QBSTU	*mamQ, mamB, mamS, mamT, mamU*
pZS24	pZS24.MNOPARQBSTU	*mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS, mamT, mamU*
pET.cocoR	pET.cocoR.HIEL	*mamH, mamI, mamE, mamL*
pET.cocoR	pET.cocoR.HIELMN	*mamH, mamI, mamE, mamL, mamM, mamN*
pET.cocoR	pET.cocoR.HIELMNOP	*mamH, mamI, mamE, mamL, mamM, mamN, mamO, mamP*
pET.cocoR	pET.cocoR.HIELMNOPARQBST	*mamH, mamI, mamE, mamL, mamM, mamN, mamO, mamP, mamA, mamR, mamQ, mamB, mamS, mamT*

As the change in the promoter did not allow larger constructs to be made, a different, low copy number vector (pZS24), was used. This vector contains a kanamycin resistance cassette and utilises a lactose promoter for gene expression. In order to clone multi-gene constructs into this vector, *mamU*, the final gene of the *mamAB* operon from a pET3a vector (*BglII/NheI* site), was first cloned into the pZS24 vector (*BamHI/AvrII* site). In doing so, this strategy also replaced the pZS24 lactose promoter with the T7 promoter from pET3a. Furthermore, the construct obtained allows insertion of other genes from the previously obtained pET3a constructs (*AseI/SpeI* site) to be inserted upstream of *mamU* (*AseI/XbaI* site). This method resulted in the generation of pZS24.MNOPARQBSTU, but, as before, it was not possible to obtain constructs containing any further genes.

The use of another vector, pETcoco2 (Novagen^{TM}), was also investigated. In previous work, this vector was used successfully to clone 25 genes associated with the biosynthesis of cobalamin (Deery *et al.,* 2012). The vector combines the two important properties: a tightly regulated promoter (T7lac) and is maintained at a single copy per cell. Furthermore, using arabinose induction, the copy number can be increased allowing for easy amplification of the plasmid. In order to sub-clone previously constructed plasmids in a modular manner, the *SpeI* site, located in the *parB* coding region of pETcoco2, had to be removed. Overlap extension mutagenesis was used to remove the site (introducing a silent mutation in *parB*), yielding pETcocoR. An operon containing *mamH*, *mamI*, *mamE* and *mamL* from pET3a.HIEL (*XbaI/HindIII* site) was inserted into this vector (*NheI/HindIII* site). The newly constructed plasmid (pETcocoR.HIEL (**Table 4.7**)) allows insertion of further genes (from pET3a constructs) into the same operon, downstream of the final gene (*mamL*), using the *SpeI/BmtI* site (*XbaI/BmtI* site for the insert). Using this method, it was possible to construct pET.cocoR.HIELMNOPARQBST in 3 stages. The new construct contains all the genes that were initially set out to be cloned together, with an additional gene, *mamH*. Unfortunately, the construct could not be transfored into strains containing a T7 polymerase. Further work is going to be carried out in order to attempt to obtain a strain able to express proteins from this plasmid.

4.3. Discussion

In this chapter bioinformatics analysis of the proteins associated with magnetosome formation and the ability of *E. coli* to correctly produce some of these proteins is discussed. Furthermore, the generation of large artificial operons containing multiple genes encoding for these proteins is shown.

Bioinformatics analysis showed a large proportion of the proteins associated with magnetosome formation are predicted to be either membrane anchored (10 out of 30) or transmembrane proteins (11 out of 30). This was expected, as the magnetosome is a membranous organelle and the biomineralisation takes place within the membranous space separated from the cytoplasm by the membrane. Interestingly, no targeting sequences were identified for any of the associated proteins. This points to a novel targeting mechanism for some of the associated proteins, independent of the well characterised bacterial sec and tat targeting systems (Natale *et al.,* 2008). There has been some evidence for this, suggesting MamE to play a key role in the unknown mechanism (Quinlan *et al.,* 2011).

Protein structure prediction is a growing field of interest as with the availability of more and more protein structures together with novel approaches on modelling unknown structures, it is possible to predict un-solved proteins with better accuracy (Dorn *et al.*, 2014). When studying uncharacterised proteins or protein systems, using such predictions can be highly informative as the protein fold can often point to functions not identified using standard protein sequence alignments or reinforce such predictions. These can then be used to design experiments. In order to gain such insight into the proteins involved in magnetosome synthesis I-TASSER, a freely available protein structure and function prediction software was used (Yang *et al.,* 2015). Some of the more interesting observations are discussed.

Protein 4070 was predicted to have a tetratricopeptide (TRP) repeat fold. In eukaryotic cells, proteins containing this fold have been shown to have functions which might be utilised in magnetosome synthesis. These include vesicle formation, contribution to protein complex assemblies and protein import to cellular organelles (Zeytuni and Zarivach, 2012). Notably, six different proteins from the MAI (MamC, 4070, MamI, MamA, MamQ and MamZ) had predicted structural motifs suggesting targeting to such organelles. Furthermore, two of these (MamI and MamZ) and one further protein in the MAI (MamH),

were predicted to be part of macromolecular complexes. TPR-containing proteins can bind a variety of ligands that mostly do not share sequence or secondary structure similarities. Therefore, it is suggested that this protein is either involved in protein targeting to the magnetosome either for transport into the organelle or complex assembly on the peripheral of the organelle.

In previous work, MamK has been established as the actin-like protein forming filamentous structures used to string magnetosomes together (Komeili *et al.*, 2006) with MamJ acting as a cross-linker between the filament and the organelles (Draper *et al.*, 2011). Structural predictions analysed in this chapter suggest further proteins might be involved in these structures. Protein 4071 has predicted structural similarities to spectrin, a protein which is part of the membrane skeleton (Machnicka *et al.*, 2014), and protein 4074 has predicted structural similarities with actin-binding proteins and peptidases, while the cross-linking protein, MamJ, showed predicted regulatory activity against such peptidases. Furthermore, MamF, and its homologue MmsF, were predicted to localise to the cytoskeleton, possibly playing a structural role.

With respect to the biomineralisation process, surprisingly this approach did not identify three of the cytochrome c proteins (MamE, MamP and MamX) as involved in electron transport, only identifying MamT. Interestingly, the structural approach suggested MamN may be involved in the process. This is backed up by a Δ*mamN* phenotype in *M. gryphiswaldense*, which shows magnetosome membrane biogenesis, but not crystal nucleation (Murat *et al.,* 2010b).

Another interesting finding is the high confidence prediction of MamU as a diacylglycerol kinase. This fits well with the protein not being necessary for organelle formation (Murat *et al*., 2010b), but still playing a role in its formation. The predicted product of the enzyme, phosphatidic acid (PA), has been proposed to play a role in membrane curvature, especially in membrane fission (Kooijman *et al.*, 2003). Furthermore, phosphatidic acid is able to bind Fe^{2+} ions, which may assist in biomineralisation and redox control of the metal during the biomineralisation process. This is demonstrated by the ability of PA to protect against iron-dependent oxidation (Dacaranhe and Terao, 2001).

Overall, the bioinformatics approach provided interesting insights into magnetosome synthesis, allowing informed predictions to be made for proteins with unknown functions.

These predictions allow the development of further experiments to better understand the magnetosome biogenesis process.

The construction of a plasmid library, containing all the individual genes (except for *mamG*, due to a frame shift mutation present in the template DNA) from the four MAI operons (*mamGFDC, mms6, mamAB* and *mamXY*), was achieved. Multiple mutations were observed in the cloned variants of the genes, when these were compared to the two available genomes (**Table 4.4.**). This can be explained by the observed high rate of mutations in the MAI (Kolinko *et al.,* 2011), which may be due to the high metabolic demand for the generation of the organelles. This library allows the study of individual genes as well as combinatorial studies, since the plasmids can be easily combined into multi-gene constructs.

The individual gene constructs were then analysed for protein production in *E. coli* BL21 (DE3) Star cells. The proteins were produced using the T7 promoter, utilising a genome integrated T7 RNA polymerase expressed from a lactose promoter, a commonly used method for overproduction of proteins in *E. coli*. Interestingly, SDS-PAGE analysis showed only 7 out of 29 analysed proteins had observable overproduction (**Table 4.8**). In order to carry out a more sensitive, western blot analysis, new constructs with an immuno-sensitive tag were made as no antibodies are available for these proteins. An N-terminus hexa-histidine tag was first used, which allowed identification of 10 out of 14 analysed proteins, with further 2 identified using a C-terminus hexa-histidine tag. Five proteins (MamJ, MamL, MamP, MamB and MamT) identified using this method had sizes which did not directly relate to the predicted protein size. For four of these proteins (MamL, MamP, MamB and MamT) this may be due to the membranous domains predicted to be part of the proteins. It has been reported, that proteins with hydrophobic domains can have abnormal migration during SDS-PAGE (Rath *et al.,* 2009). Considering this abnormality for MamJ, it could be explained by the high proportion of the protein consisting of proline (10.3% of all amino-acids) as proteins rich in this amino-acid have been reported to have an abnormal migration (Boze *et al.,* 2010). This may be due to the curvature induced by this amino-acid. Overall, these results suggest, that MAI proteins are produced, but at low levels and cannot be detected using standard SDS-PAGE procedures. Further analysis should be carried out to confirm this for all MAI proteins suggested to be involved in magnetosome biosynthesis.

Table 4.8. Overview of MAI protein production studies.

Protein	Operon	TMHs	Molecular Weight (kDa)	Observable	
				SDS-PAGE	Western Blot
MamG	*mamGFDC*	2	7.72	N/A	N/A
MamF	*mamGFDC*	3	12.34	-	N/A
MamD	*mamGFDC*	1	30.26	-	N/A
MamC	*mamGFDC*	2	12.43	-	N/A
4070	*mms6*	2	48.04	+	N/A
4071	*mms6*	1	36.38	+	N/A
MmsF	*mms6*	3	13.78	-	N/A
Mms6	*mms6*	1	12.76	-	N/A
4074	*mms6*	0	9.73	-	N/A
MamH	*mamAB*	12	45.67	-	N/A
MamI	*mamAB*	2	7.16	-	N/A
MamE	*mamAB*	1	78.05	-	+
MamJ	*mamAB*	0	48.52	-	+*
MamK	*mamAB*	0	39.20	+	N/A
MamL	*mamAB*	2	8.59	-	+*
MamM	*mamAB*	3	34.49	-	+
MamN	*mamAB*	11	46.18	-	-
MamO	*mamAB*	8	65.39	-	-
MamP	*mamAB*	1	28.36	-	+*
MamA	*mamAB*	0	24.01	+	+
MamQ	*mamAB*	1	30.03	-	+
MamR	*mamAB*	0	8.06	-	+
MamB	*mamAB*	3	31.96	-	+*
MamS	*mamAB*	1	18.72	+	+
MamT	*mamAB*	1	18.89	-	+*
MamU	*mamAB*	0	32.00	+	+
MamY	*mamXY*	2	40.90	-	N/A
MamX	*mamXY*	1	28.22	+	N/A
MamZ	*mamXY*	18	70.55	-	N/A
FtsZ-like	*mamXY*	0	32.33	-	N/A

Multiple proteins suggested to be involved in the generation of the magnetic mineral within the organelle have cytochrome c domains (**Figure 4.8**). These domains contain a covalently attached heme group. It was decided to test if these are processed in *E. coli* BL21 (DE3) Star cells. Since the proteins are produced at low levels, investigation of truncated versions, which only contain the soluble domains, was carried in the hope that higher protein concentrations would be obtained. These domains contained the cytochrome c motif and showed higher levels of expression making them easier to analyse. In order to process the domain under aerobic growth, the protein had to be targeted to the periplasm and the cytochrome c maturation machinery had to be recombinantly produced. This is due to *E. coli* cells only producing this maturation machinery under anaerobic growth conditions, when specific terminal electron acceptors are present (Iobbi-Nivol *et al*., 1994). Our results show *E. coli* is able to correctly process 3 out of 4 of these cytochrome proteins (MamP, MamT and MamX). A functionalised MamE was not obtained, which may be due to high degradation of the protein (**Figure 4.9**). The observed degradation is likely due to a recently reported self-cleavage effect of the protein (Hershey *et al*., 2016b). Overall, these results show that growth conditions need to be taken in consideration if biomineralisation in recombinantly produced organelles is to be achieved. This could be related to the observations from the previous chapter, when the possible biomineralisation could only be observed under anaerobic growth, using nitrate as a terminal electron acceptor.

In an attempt to produce recombinant magnetosomes, combinations of the single gene constructs were combined to represent each of the operons predicted to be involved in magnetosome generation. This was achieved for *mms6* and *mamYX* operons. *mamGFDC* operon was also successfully reconstructed, but without *mamG*, as the gene contained a frame-shift mutation in the genomic DNA used as the template. This would suggest that MamG does not play a key role in the organelle generation process, which has been shown previously (Scheffel *et al.,* 2008), and can be omitted for preliminary studies. Reconstruction of *mamAB* operon proved to be difficult, but using a variety of approaches, a large construct, representing 14 out of 17 genes encoded in the operon, was obtained. The three genes omitted (*mamJ, mamK* and *mamU*), have been shown to not be important in the membranous organelle formation (Murat *et al.,* 2010a; Lohbe *et al.,* 2016), two of which (*mamJ* and *mamK*) are involved in organisation of the organelle within the cell and one (*mamU*) has unknown accessory function (discussed previously). Notably, the most challenging steps, in reconstructing the *mamAB* operon, always included the ligation of a

construct containing *mamE* (data not shown), as often no transformants were observed when a ligation product for such a construct was transformed and if transformants were present, there were less compared to constructs which did not contain *mamE*. This may be due to the protease activity of MamE, which may have toxic effects in the cytoplasm of *E. coli*. Further efforts should be undertaken, to introduce *mamU* into the construct. Growth experiments and electron microscopy analysis should to be carried out on *E. coli* strains expressing the plasmid containing the partially reconstructed *mamAB* operon. Furthermore, expression of all the operons predicted to be involved in magnetosome generation needs to be achieved in a single strain of *E. coli*, either by integrating some of these operons into the genome, expressing from multiple plasmids or combining all the operons in a large construct as was done for vitamin-B12 synthesis.

In conclusion, the work discussed in this chapter shows an analysis of a variety of proteins involved in magnetosome production focusing on their compatibility with the model organism, *E. coli*. The research suggests that the production of magnetosomes might be possible in the organism, but further work needs to be done to achieve this. The effects on the bacterial membrane morphology by some of these proteins is discussed in the next chapter.

Chapter 5

Restructuring *Escherichia coli* Membranes: MAI proteins, LemA Family and Outer-Membrane Targeted Proteins

5.1. Introduction

Until recently it has been thought that one of the defining features of prokaryotic organisms is a lack of cytoplasmic membranous organelles. As multiple membranous organelles have been discovered in prokaryotes so this view is changing (Saier and Bogdanov, 2013). These organelles include chromatophores, nuclear envelopes, magnetosomes, outer membrane vesicles (OMVs) and others. Furthermore, with the development of new bacterial culture approaches (Stewart, 2012) and analytical methods, such as cryo-electron tomography (Li and Jensen, 2009), more novel membranous structures are likely to be discovered.

Membrane restructuring in eukaryotic cells have been studied extensively and a variety of mechanisms employed by these cells have been discovered (Prinz and Hinshaw, 2009). Even though no bacterial homologues of specialised membrane bending proteins, such as clathrin, have been discovered, most membrane restructuring proteins share similar motifs, which may not necessarily relate to sequence identity. One of the most common mechanisms for membrane restructuring is direct insertion of a protein into the lipid bilayer via an amphipathic domain, thereby increasing the area of one leaflet, causing the bilayer to bend. Therefore, proteins with a variety of other functions, that contain this domain, are able to have a drastic effect on the membranes. In other cases, simply having a specific shape allows membrane restructuring. A good example of this is the bacterial chromatophores where the shape of the light harvesting complexes has been shown to induce membrane curvature (Chandler *et al.,* 2009).

As discussed in the introduction, membrane restructuring during magnetosome formation is still not very well understood, although genetic studies have suggested the proteins involved. Four proteins (MamI, MamL, MamQ and MamB) from the *mamAB* operon have been shown to be necessary for the magnetosome membrane formation in *M. magneticum* (Murat *et al.,* 2010a). A more recent study in *M. gryphiswaldense* suggests that a further three proteins (MamE, MamM and MamO) from the *mamAB* are also involved (Raschdorf *et al.,* 2016). Neither of these studies looked at other proteins produced by the other operons (*mamGFDC, mms6* and *mamXY*) involved in magnetosome formation.

In order to better understand the biogenesis of magnetosome membranes it was decided to produce some of these proteins in *E. coli* and analyse the resulting strains using electron microscopy.

5.2. Results

5.2.1. Analysis of Proteins Involved in Magnetosome Membrane Formation

In order to investigate if the potential magnetosome membrane protein had an effect on the *E. coli* cytoplasmic membrane, five proteins were selected for the study: MamI, MamL, MamQ, MamB and MamY. Their effects on the morphology of the bacterial inner membrane were analysed using a transmission electron microscopy approach. Briefly, a pET3a construct encoding each of the proteins (section 4.2.2) or an empty vector control were transformed into *E. coli* BL21 (DE3) Star cells, the cells were grown, harvested, fixed, embedded, sectioned, stained and imaged as described in materials and methods. Electron microscopy revealed that cells producing MamI, MamL or MamB showed no significantly phenotype compared to the control. Electron tomographs of *E. coli* BL21 (DE3) Star cells producing MamQ, MamY and the empty vector control are shown in **Figure 5.1**. A large proportion of cells producing MamY had electron light areas while MamQ producing cells showed a variety of phenotypes, often complex structures, which appeared to be membranous. It was decided to quantitate these effects by cell counting, where any deviation from a standard membrane structure was counted as a membranous phenotype. Overall, the proportion of cells showing such a phenotype were 56.2% for MamY and 48.9% for MamQ compared to 8.6% present in the control cells. The phenotype present in the control cells can be related to the chemical fixation process and is not comparable to the phenotype observed in the experimental strains.

In order to test if the membranous structures observed were dependent on protein expression levels a new construct was made (pET3a.TBAD.MamQ) which utilises an arabinose promoter. The arabinose promoter is known to be a weak promoter compared to the T7 promoter. The phenotypes of *E. coli* BL21 (DE3) Star cells producing MamQ at four different levels (T7 promoter with no induction, T7 promoter with auto-induction (lower strength than IPTG, but over a longer period of time), T7 promoter with IPTG induction (positive control) or arabinose promoter with induction) were then compared. An empty plasmid vector was used as a negative control. Strains producing MamQ with no induction looked comparable to the negative control (Results not shown as identical to **Figure 5.1** 'control' column). Positive control samples looked similar as before, with a large variety of different membranous phenotypes (Results not shown as identical to **Figure 5.1** 'MamQ' column). When expressed using auto-induction (**Figure 5.2**) or arabinose induction (**Figure**

5.3) methods, the cells produced more consistent phenotypes. No complex cellular membranous structures were observed, instead membranous invaginations, which were more defined when using the arabinose promoter, were present. Furthermore, when using the arabinose method, no more than one of these invaginations were present in the cell, as opposed to the multiple invaginations observed using the auto or IPTG induction approaches. Overall, the proportion of cells showing a membranous phenotype were 63.9% for auto-induction method and 27.6% for arabinose induction method.

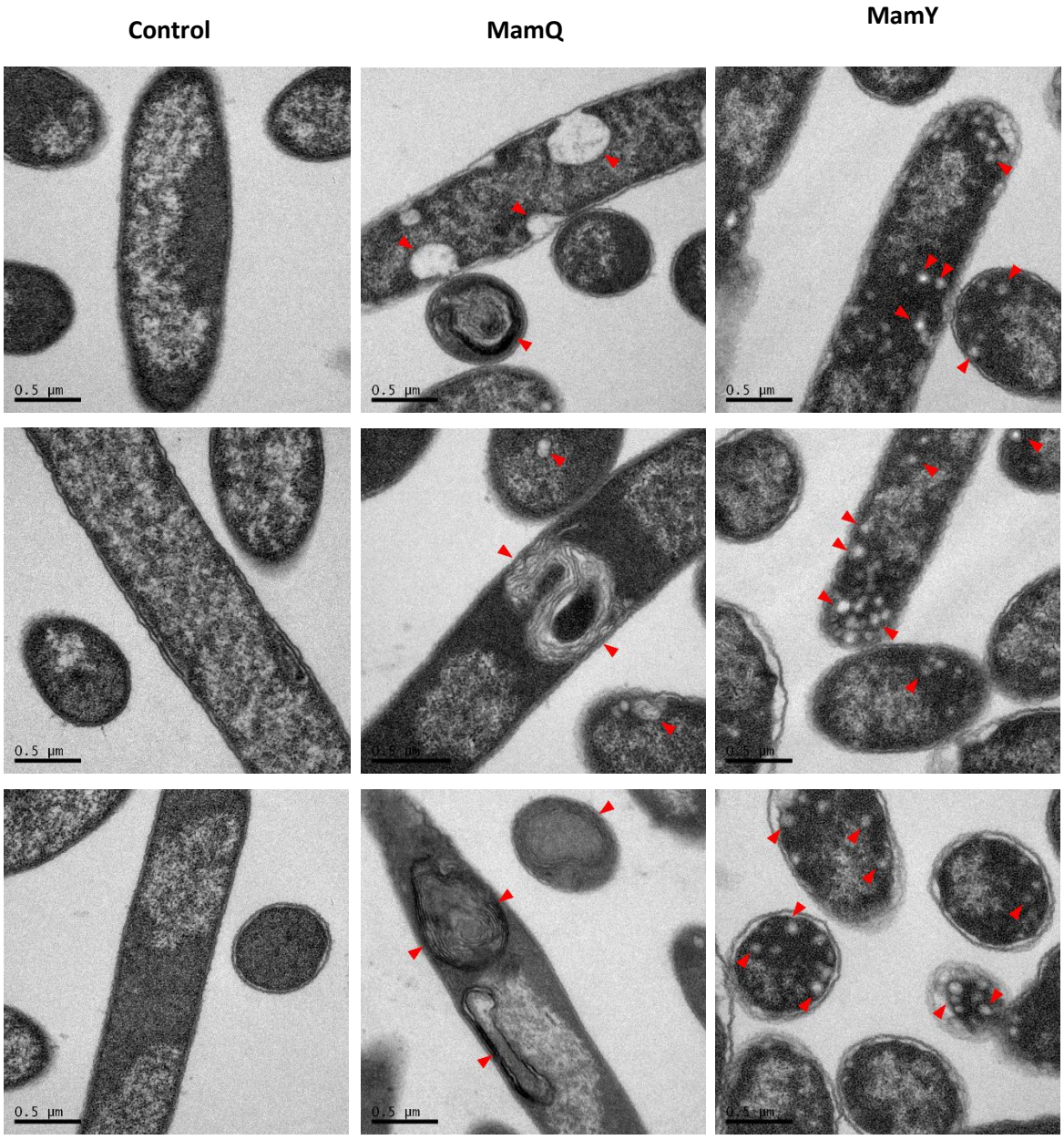


Figure 5.1. TEM of *E. coli* expressing *mamQ* or *mamY*.
Transmission electron micrographs of sectioned *E.coli* BL21 (DE3) Star cells harbouring an empty **control** vector, or a vector encoding **MamQ** or **MamY**. **Red arrows** – points of interest. Scale bar is 0.5µm.

MamQ (auto-induction)

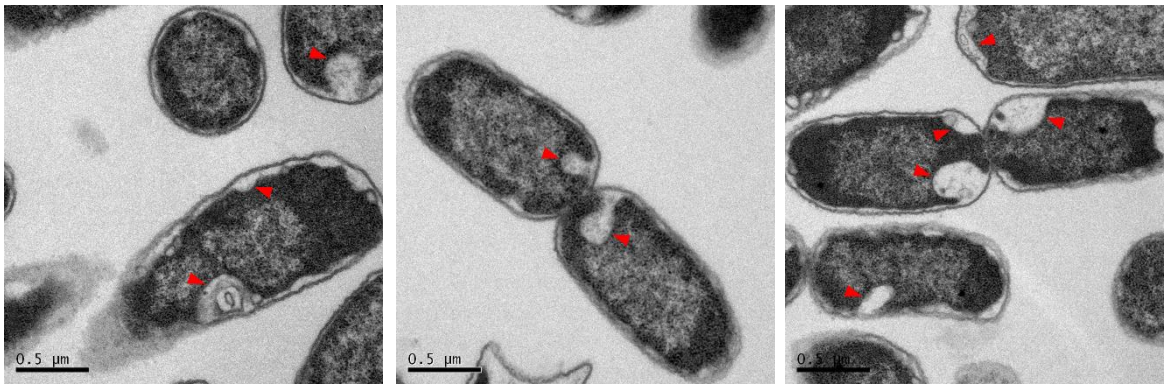


Figure 5.2. TEM of *E. coli* expressing *mamQ* using the auto-induction method.
Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells producing MamQ using the auto-induction method. **Red arrows** – points of interest. Scale bar is 0.5µm.

MamQ (arabinose-induction)

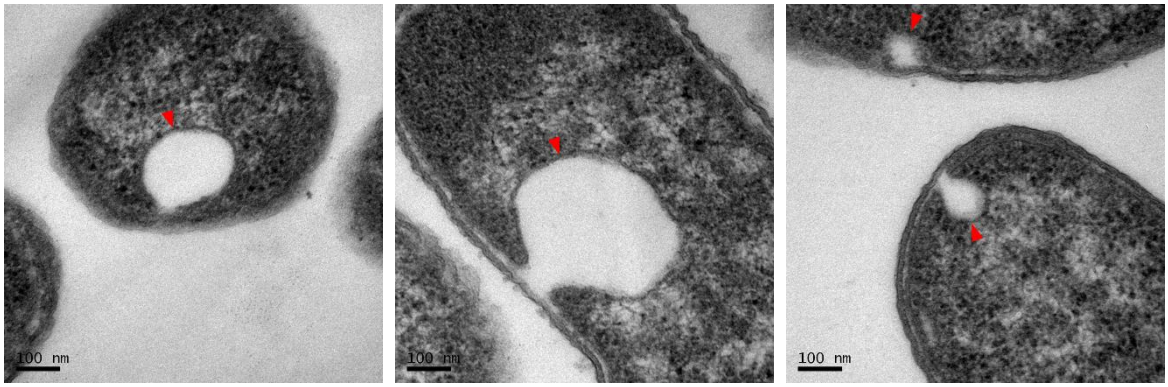


Figure 5.3. TEM of *E. coli* expressing *mamQ* from an arabinose promoter.
Transmission electron micrographs of sectioned *E.coli* BL21 (DE3) Star cells producing MamQ from the arabinose promoter. **Red arrows** – points of interest. Scale bar is 0.1µm.

Cells producing just the soluble domain of MamQ (amino-acids 73-272) were also analysed. These cells had no membranous phenotype although a high proportion of cells contained inclusion bodies (94.2%).

5.2.2. Analysis of MamQ in Combination With Other MAI Proteins

As cells producing MamQ showed a strong membranous phenotype, its effect in combination with other proteins predicted to affect the membrane was investigated. A variety of constructs, using the method discussed in **section 4.2.6**, containing different combinations of the predicted membrane restructuring proteins, were made (**Table 5.1**).

Table 5.1. Multi-gene constructs analysed by electron microscopy.

Operon	Construct	Genes (in construct order)
mamAB	pET3a.IQ	*mamI, mamQ*
mamAB	pET3a.LQ	*mamL, mamQ*
mamAB	pET3a.QB	*mamQ, mamB*
mamAB	pET3a.ILQ	*mamI, mamL, mamQ*
mamAB	pET3a.LQB	*mamL, mamQ, mamB*
mamAB	pET3a.IQB	*mamI, mamQ, mamB*
mamAB	pET3a.ILQB	*mamI, mamL, mamQ, mamB*
mamAB/mamXY	pET3a.IQBY	*mamI, mamQ, mamB, mamY*
mamAB/mamXY	pET3a.LQBY	*mamL, mamQ, mamB, mamY*
mamAB/mamXY	pET3a.ILQBY	*mamI, mamL, mamQ, mamB, mamY*

The constructs were then analysed by electron microscopy as discussed above. Strains expressing constructs containing two (**Figure 5.4**), three (**Figure 5.5**), four or five (**Figure 5.6**) genes varied in their phenotype. No strong phenotype was observed when pET3a.IQBY or pET3a.LQBY were expressed.

As when MamQ was produced on its own using IPTG induction, so did the MamLQ producing cells also showed a variety of membranous phenotypes, including complex structures (**Figure 5.4**). These structures differed when MamIQ was expressed, showing more extreme membrane curvature. On the other hand, MamQB expression gave a result that was more similar to the auto-induction method although no complex membranous structures were observed, but overall the inner membrane of the cells appeared to be curved into the cytoplasm in multiple places throughout the cell. The proportion of cells showing a membranous phenotype were 39.3% for MamLQ, 48.2% for MamIQ and 44.1% for MamQB.

When the three gene constructs were expressed, increased structure could be observed in the membranous invaginations (**Figure 5.5**). For MamILQ and MamLQB, complex

membranous structures were once again observed, but these appeared more structured, showing less variability between the cells. These structures in MamILQ were single-membrane and separated, as opposed to multi-layer structures observed for MamQ, MamLQ or MamLQB. When MamIQB was produced, small vesicular structures could be observed within the cells, sometimes containing multiple membranes. Furthermore, small inclusions could be observed. The proportion of cells showing a membranous phenotype were 51.5% for MamILQ and 39.9% for MamLQB.

Surprisingly, the four gene constructs, MamIQBY and MamLQBY, did not show any specific phenotype, while MamILQB appeared similar to MamIQB, where defined membranous vesicles could be observed within the cell (**Figure 5.6**). Small inclusions could also be observed. The proportion of cells showing a membranous phenotype were 34.2%.

The largest analysed construct, MamILQBY, showed a phenotype similar to MamIQB and MamILQB, where membranous vesicles could be observed, although the number of these was increased. Furthermore, no inclusions could be observed. The proportion of cells showing a membranous phenotype were 32.7%.

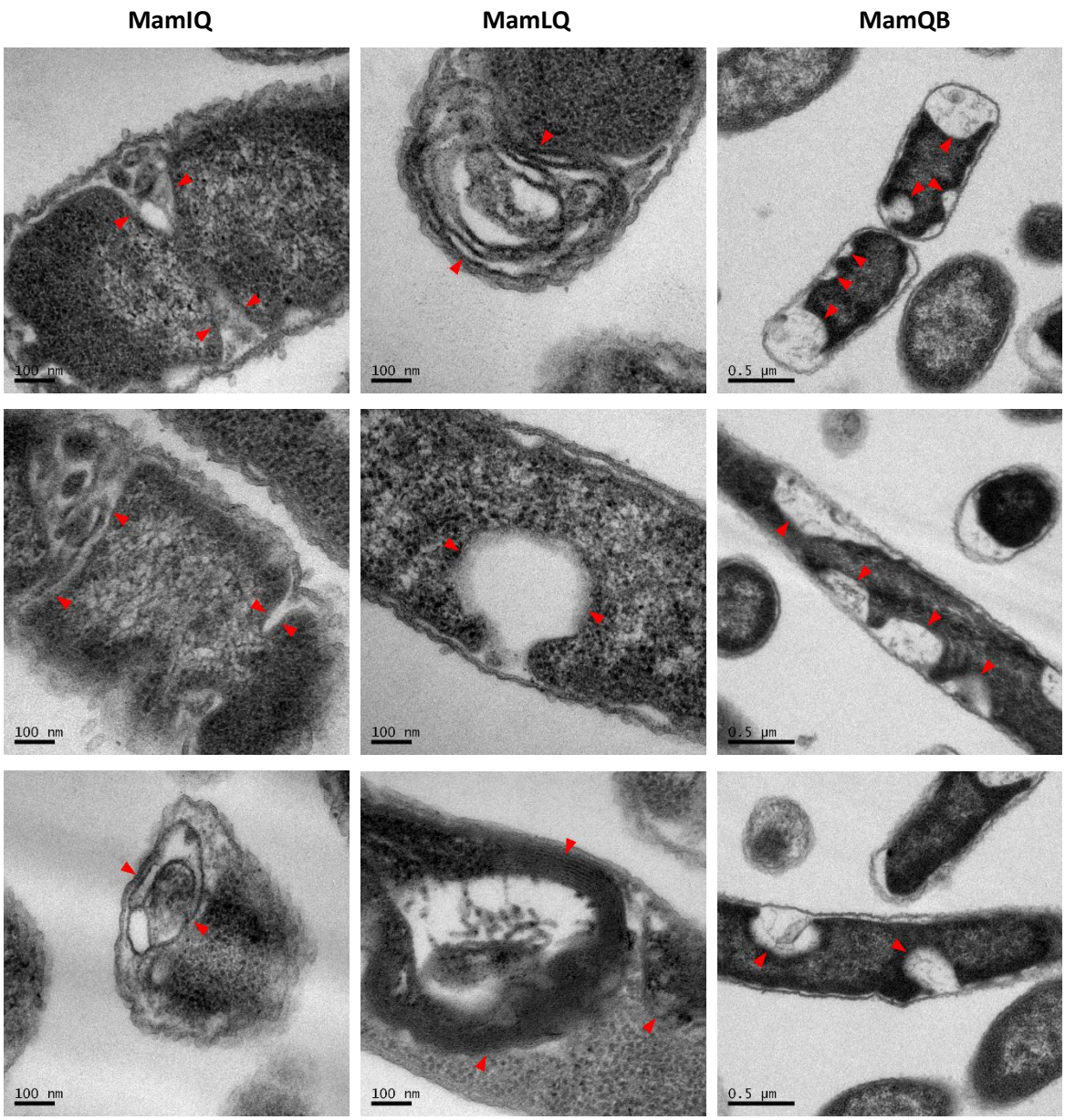


Figure 5.4. TEM of *E. coli* expressing double gene constructs.
Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.IQ (**MamIQ**), pET3a.LQ (**MamLQ**) or pET3a.QB (**MamQB**). **Red arrows** – points of interest. Scale bars: MamIQ and MamLQ is 0.1µm and MamQB is 0.5µm.

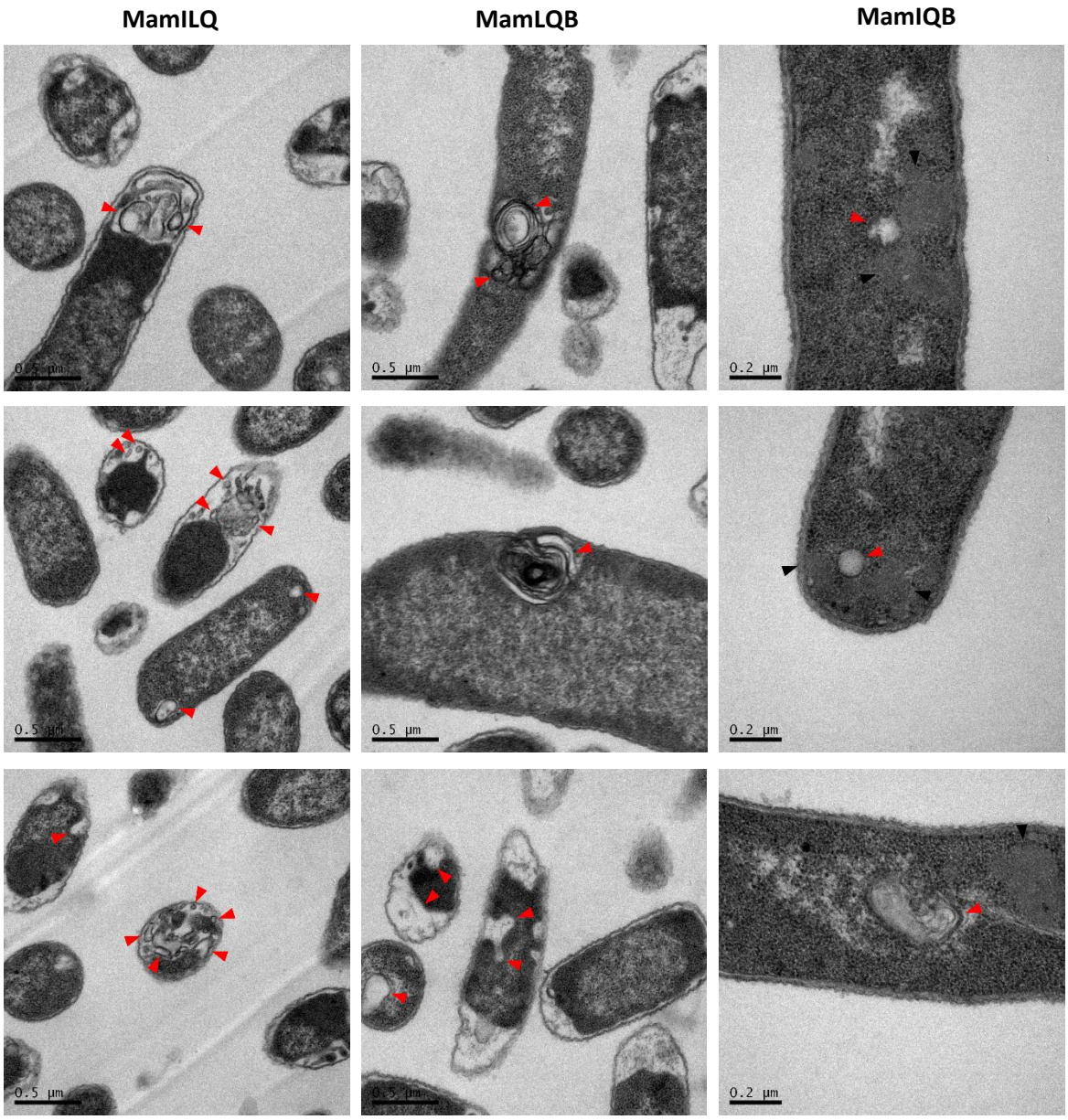


Figure 5.5. TEM of *E. coli* expressing triple gene constructs.
Transmission electron micrographs of sectioned *E.coli* BL21 (DE3) Star cells expressing pET3a.ILQ (**MamILQ**), pET3a.LQB (**MamLQB**) or pET3a.IQB (**MamIQB**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bars: MamILQ and MamLQB is 0.5µm and MamIQB is 0.2µm.

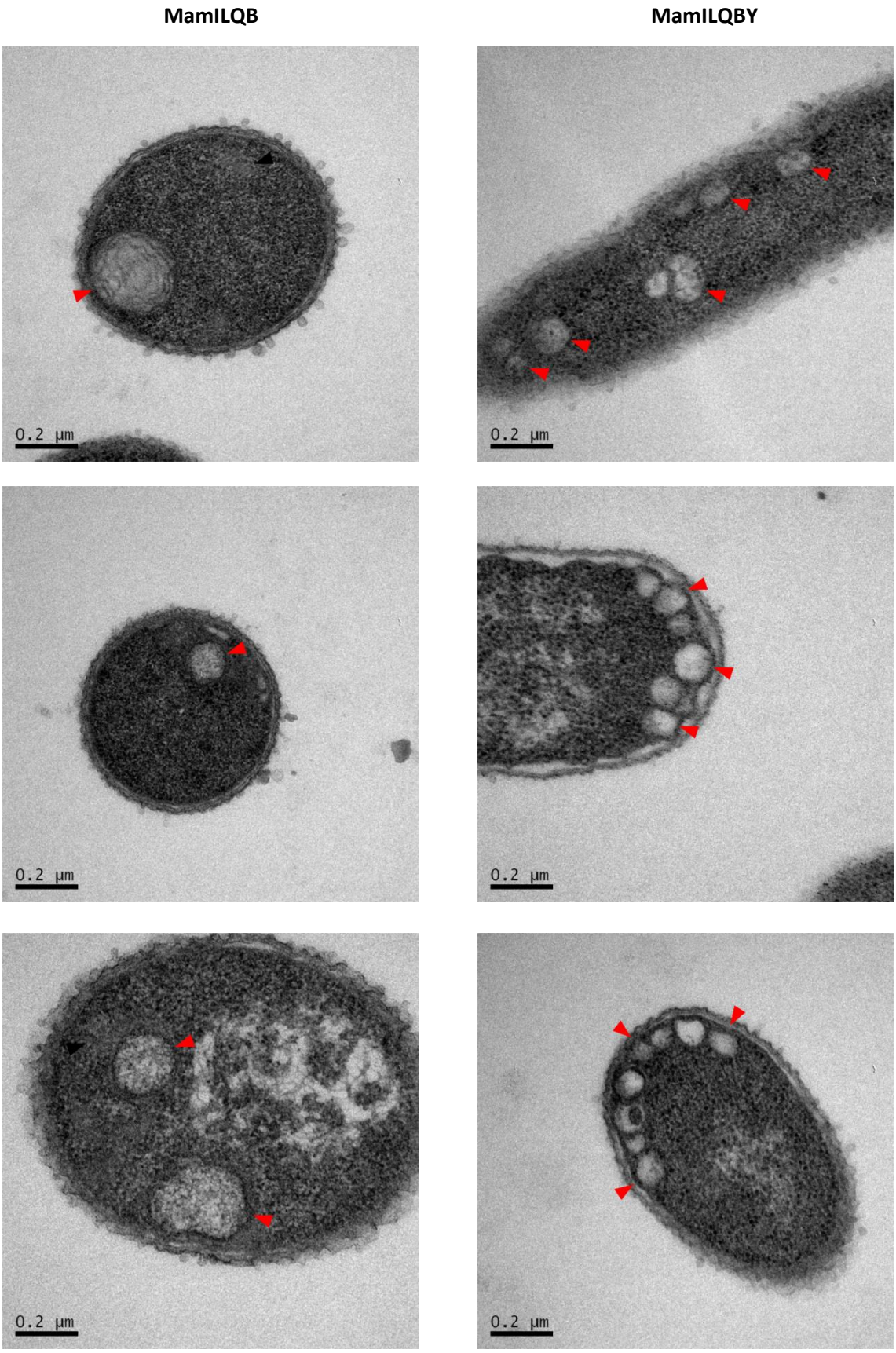


Figure 5.6. TEM of E. coli expressing constructs containing *mamILQB* or *mamILQBY*.
Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.ILQB (**MamILQB**) or pET3a.LQB (**MamILQBY**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

5.2.3. Proteins Related to MamQ

MamQ is a member of the LemA family of proteins, which have no characterised function. Based on Interpro (http://www.ebi.ac.uk/interpro/) there are currently 6604 proteins, including an additional 120 proteins from metagenomes, within the LemA family, the majority of which, 6500, are from bacterial genomes (98 archaeal, 5 eukaryotic and 1 viral). Furthermore, they are wide-spread and present in both Gram-positive and Gram-negative bacteria.

A LemA protein was first characterised in 1996 (Lenz et al., 1996) as an H2-M3-restricted Listeria epitope. Based on sequence, the suggested role at the time was signalling. There is one solved structure of the soluble domain of a LemA-family protein, TM1634 from *Thermatoga maritima* MSB8 (PDB entry: 2ETD), showing a coiled-coil structure. There is no analysis of the structure available to date.

Four different LemA proteins were chosen for comparison with MamQ (**Table 5.2**). These were selected from non-related bacteria (**Figure 5.7**). TMHMM2 and SignalP 4.1 was used to predict protein topology/targeting as before. No targeting sequence was identified for any of the proteins and all of them contained a single TMH. All the selected proteins were cloned into a pET3a vector in the *NdeI/SpeI* site, as before (**Section 4.2.2**).

Table 5.2. LemA proteins chosen for analysis.

Name	Organism	NCBI Accession	Molecular Weight (kDa)	TMHs	Position of soluble domain
LemA.153	*Bacillus megaterium*	ADF40219	21.81	1	Periplasmic
LemA.159	*Clostridium kluyveri*	WP_012102790	21.08	1	Periplasmic
LemA.501	*Brucella melitensis*	WP_006264043	23.24	1	Cytoplasmic
LemA.565	*Pseudomonas aeruginosa*	NP_253059	21.05	1	Cytoplasmic

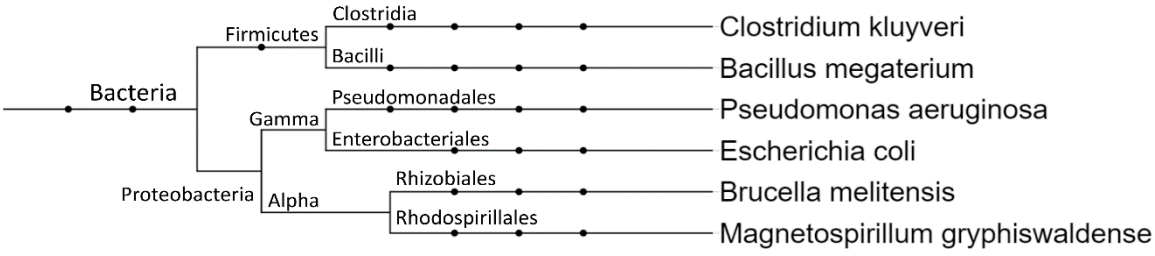


Figure 5.7. **Phylogeny tree of the chosen organisms.**
Phylogeny tree of organism which LemA proteins were selected for and their relationship to *Escherichia coli*. Last common levels are annotated.

5.2.4. Bioinformatic Analysis of LemA Proteins

Clustal Omega (Sievers *et al.,* 2011) multiple sequence alignment program was used to compare 52 different proteins from the LemA family (**Figure 5.8**). The results identified 4 distinct groups, which were analysed for common features in their operon structure (**Supplemental figure 2**). The data collected is analysed in the discussion section.

5.2.5. Expression and SDS-PAGE Analysis of Individual LemA Proteins

In order to compare the over-expression levels of individual LemA proteins in *E. coli* to MamQ, all the constructs were individually transformed into *E. coli* BL21 (DE3) Star cells. These were then grown, induced and harvested as described in the materials and methods. *E. coli* BL21 (DE3) Star cells harbouring an empty pET3a vector was used as a control. The resulting whole cell, soluble and insoluble fractions were analysed using SDS-PAGE (**Figure 5.9**). As with MamQ, no clear over-production could be observed for the LemA proteins analysed. Analysis of LemA.159 showed a stronger band just above 22kDa both in whole cell and the insoluble fraction. This may be the protein of interest, as it was later observed to form inclusion bodies (**Figure 5.10**).

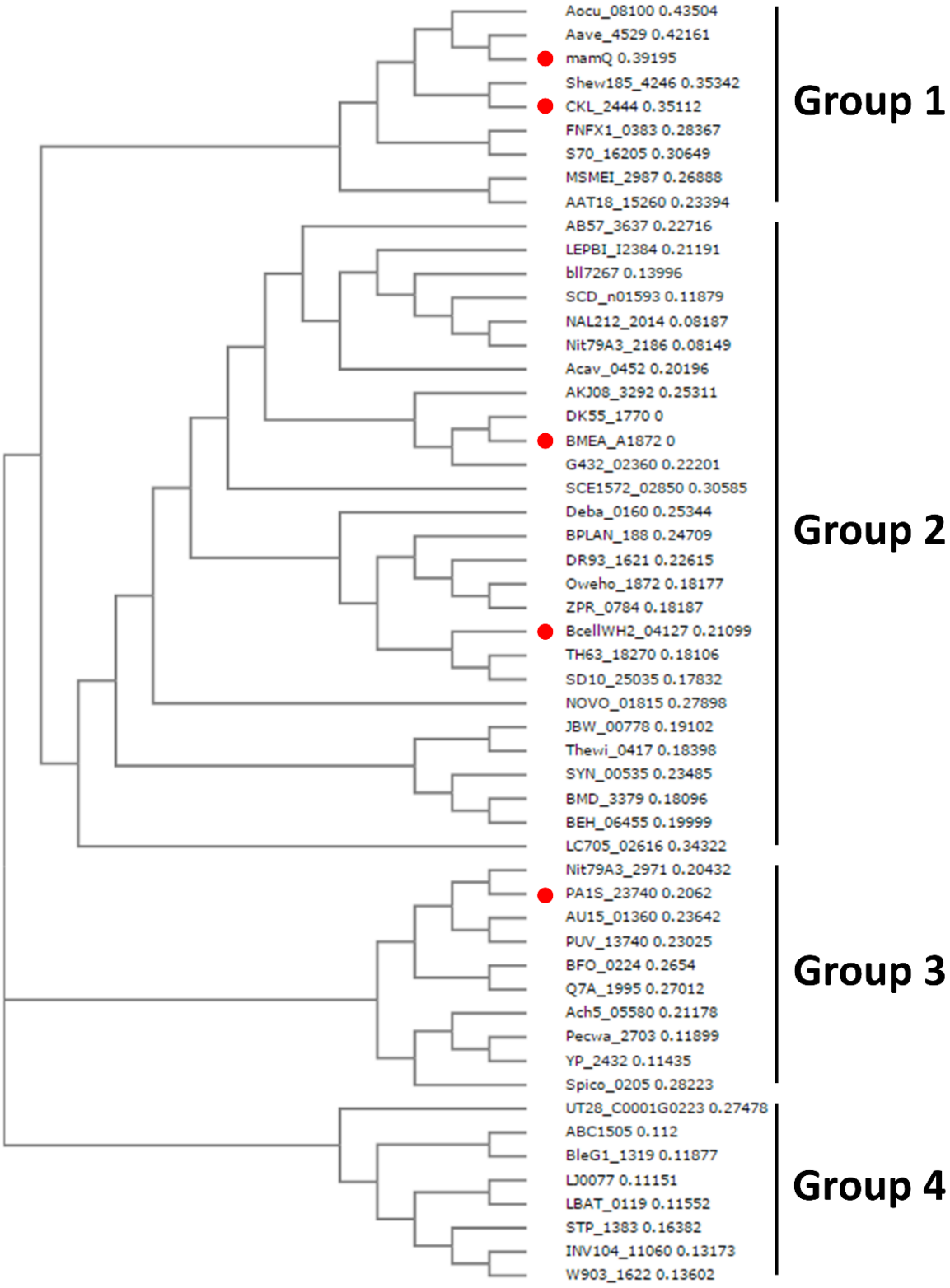


Figure 5.8. Clustal Omega analysis of LemA proteins. ● indicate studied proteins (BMEA_A1872 is LemA.153; CKL_2444 is LemA.159; BcellWH2_04127 is LemA.501; PA1S_23740 is LemA.565)

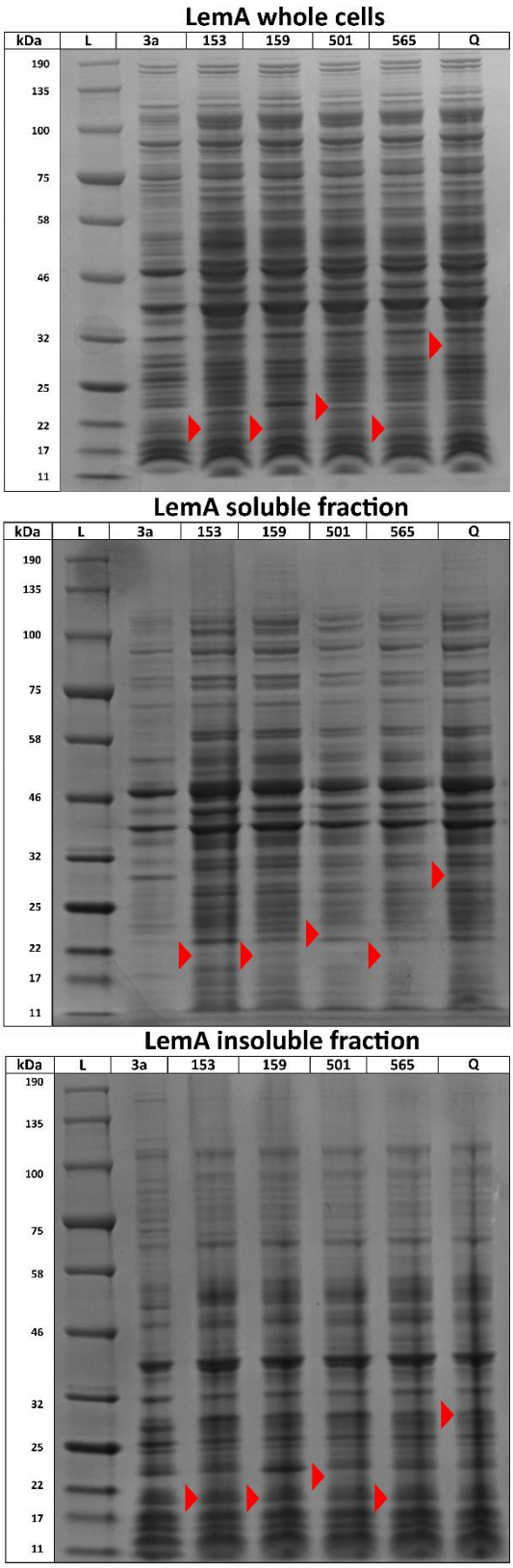


Figure 5.9. SDS-PAGE analysis of *E.coli* expressing different *lemA* genes.
E. coli BL21 (DE3) Star expressing different *lemA* genes or a control vector. 3a is empty vector control; 153 is LemA.153; 159 is LemA.159; 501 is LemA.501; 565 is LemA.565; Q is MamQ. **Red arrows** indicate predicted sizes.

5.2.6. Electron Microscopy Analysis of LemA Proteins

The different LemA producing cells were analysed by electron microscopy as described previously (**Section 5.2.1**). Cells harbouring an empty vector were used as a control. A variety of membranous phenotypes were observed in the cells expressing the different LemA gene variants (**Figure 5.10 and Figure 5.11**). The effect of the membrane was quantified by cell counting in the sections, treating any deviation from standard membrane as a membranous phenotype (**Table 5.3**).

Depending on the LemA protein produced a range of phenotypes were observed. These were defined as: LemA.153 - large (>200 nm) and smaller (<200 nm) membranous vesicles, membranous ruffles and small inclusions (**Figure 5.10**); LemA.159 - small (<100 nm) membranous vesicles and large inclusions (**Figure 5.10**); LemA.501 - membranous ruffles (**Figure 5.11**); LemA.565 - small (<100 nm) vesicles and membranous ruffles (**Figure 5.11**).

Table 5.3. Quantification of LemA phenotype.
Quantitative electron microscopy analysis of thin-sectioned *E. coli* expressing different *lemA* variants.

Sample	Total Count	Cells with a Membranous Phenotype	Percentage
Control	175	15	8.6 %
mamQ	468	229	48.9 %
lemA.153	568	303	53.3 %
lemA.159	400	266	66.5 %
lemA.501	670	488	72.8 %
lemA.565	594	462	77.8 %

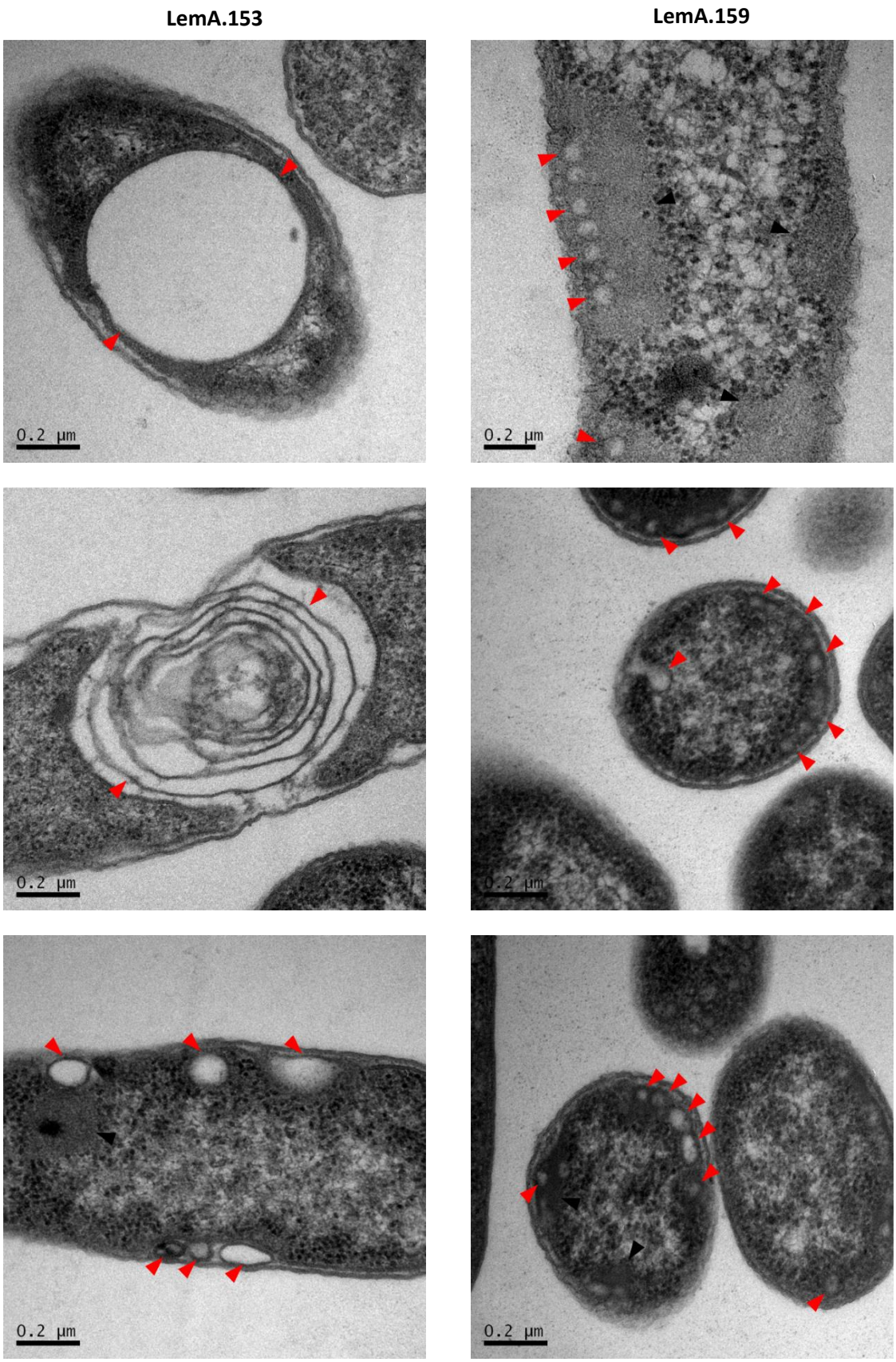


Figure 5.10. TEM of *E. coli* expressing constructs containing *lemA.153* or *lemA.159*. Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.LemA.153 (**LemA.153**) or pET3a.LemA.159 (**LemA.159**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

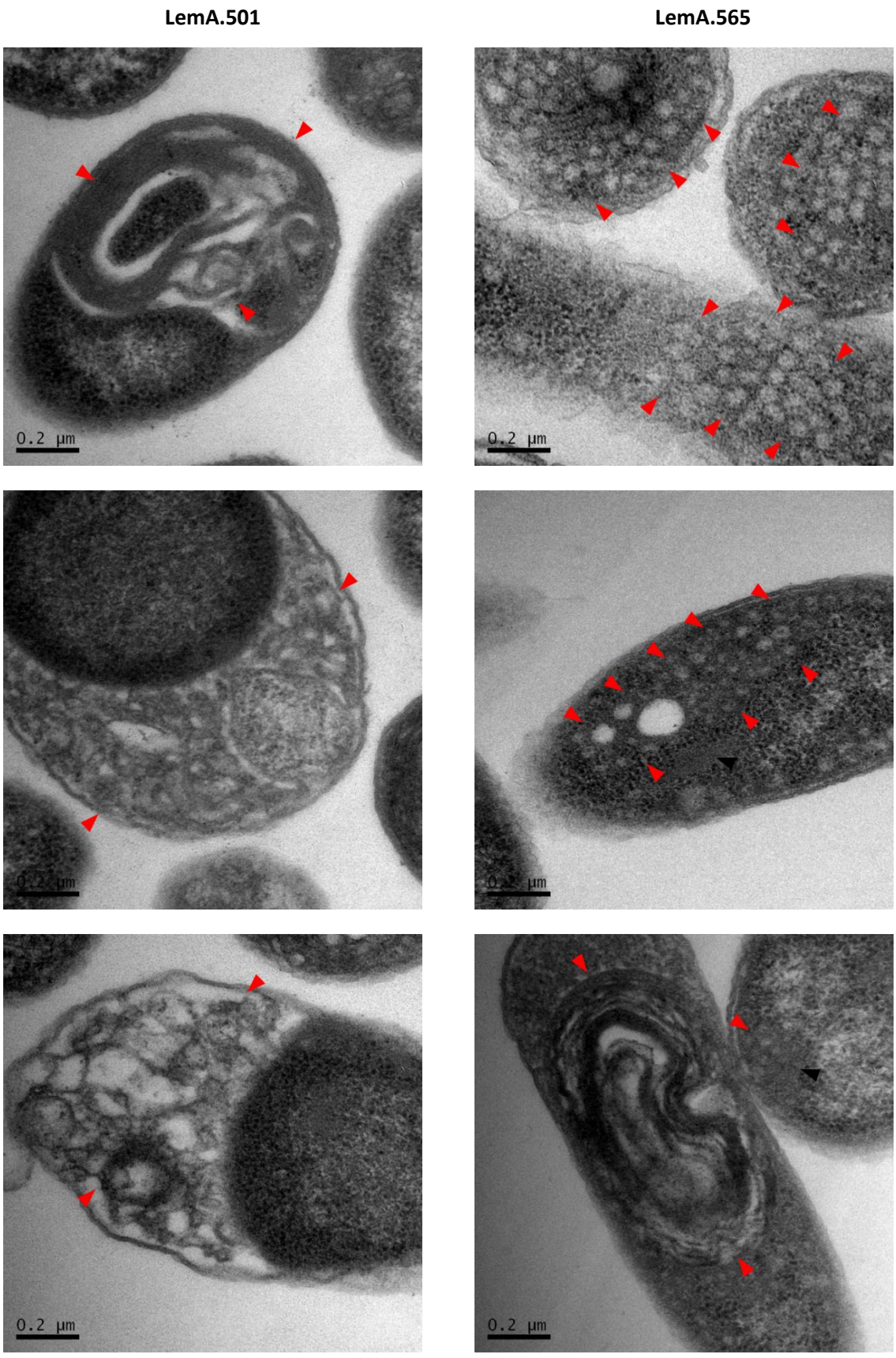


Figure 5.11. TEM of *E. coli* expressing constructs containing *lemA.501* or *lemA.565*. Electron micrographs of sectioned *E.coli* BL21 (DE3) Star cells expressing pET3a.LemA.501 (**LemA.501**) or pET3a.LemA.565 (**LemA.565**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

5.2.7. Targeting LemA Proteins to the Outer Membrane

As LemA proteins had such a strong effect on the inner-membrane it was of interest to know what would happen if these proteins were targeted to the outer-membrane. Outer-membrane proteins are either beta-barrel transmembrane proteins or attached via a lipid modification. To achieve this, the soluble domain of different LemA proteins were fused to an outer-membrane targeting sequence from BamE (the first 21 amino-acids), an outer-membrane lipoprotein, which is part of the outer membrane protein assembly complex (**Table 5.4**). A construct containing an outer-membrane targeted red fluorescent protein (RFP) was also made to be used as a control.

Table 5.4. Constructs used for outer-membrane targeting experiments.

Protein	Domain (amino acids)	Construct	Predicted size (kDa)	
			Expressed	Cleaved
-	-	pET3a.BamE*	4.87	2.95
RFP	N/A	pET3a.BamE*RFP	27.67	25.75
LemA.153	36-161	pET3a.BamE*LemA.153s	20.43	18.51
LemA.159	30-187	pET3a.BamE*LemA.159s	20.12	18.20
LemA.501	38-207	pET3a.BamE*LemA.501s	21.70	19.78
LemA.565	28-190	pET3a.BamE*LemA.565s	20.53	18.61
MamQ	73-272	pET3a.BamE*MamQs	24.39	22.48

The construction of the outer membrane targeting peptide was achieved by first cloning a fragment of synthesised DNA, corresponding to the BamE tag (*AseI/SpeI* segment), into a pET3a vector (*NdeI/SpeI* site). The tag was designed to have an in-frame *NdeI* site at the 3' end of the coding sequence. This site together with the *SpeI* site (*NdeI/SpeI* site) was used to clone the soluble domains of LemA proteins (including MamQ) and the RFP control from PCR products (*NdeI/SpeI* segments).

Cells expressing the different constructs were then analysed by electron microscopy as discussed previously (Section 5.2.1). Controls for the experiment included cells expressing an empty vector, the outer-membrane tag on its own (pET3a.BamE*) and an outer-membrane targeted RFP (pET3a.BamE*RFP). Apart from pET3a.BamE*LemA.159s, which showed high aggregation, an increase in outer membrane vesicle (OMV) formation was observed for all constructs expressing an outer-membrane tagged protein (**Figure 5.12, Figure 5.13 and Figure 5.14**). Cells expressing pET3a.BamE*LemA.153s appeared to have vesicles originating from the inner membrane into the periplasm, as they had similar electron density and structure to the cytoplasm. Furthermore, based on the construct,

these vesicles seemed to differ in size. A large proportion of cells, expressing pET3a.BamE*LemA.501s and pET3a.BamE*LemA.565s, appeared lysed, with a variety of intact membranous structures present within the section (**Figure 5.15**). All tested strains, apart from the empty vector control, showed different levels of protein aggregation. Quantification of outer membrane vesicles undergoing formation (membranous invagination that are visibly still attached to the outer membrane) and aggregation is shown in **Figure 5.16** and protein production analysis is shown in **Figure 5.17**. Interestingly, cells producing BamE*RFP showed the highest levels of inclusion formation and OMV production. No protein overproduction could be observed using the SDS-PAGE method, which may be due to the OMVs not being pelleted with the cells during the centrifugation step or low levels of protein production, as observed for wild-type LemA proteins.

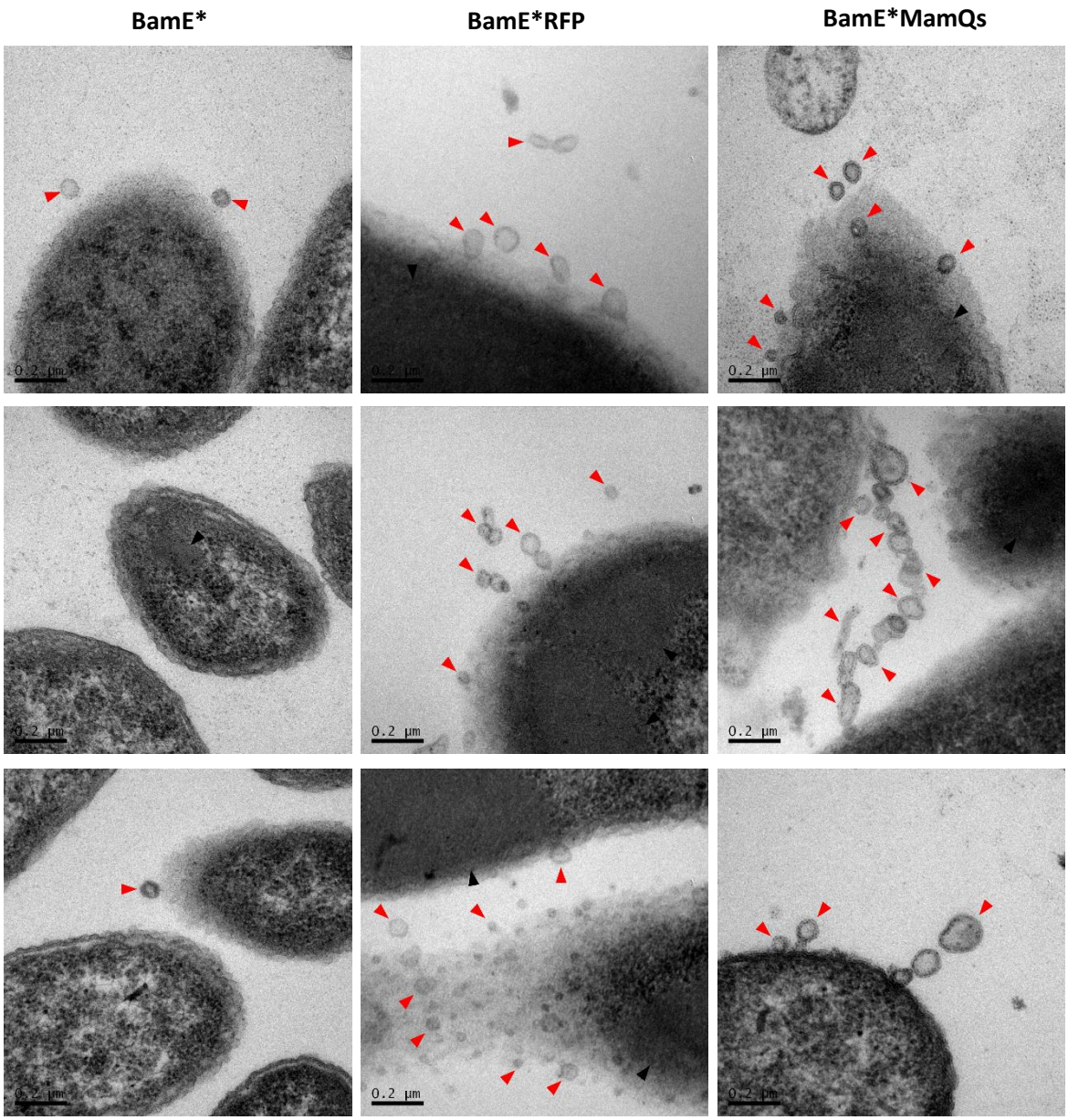


Figure 5.12. TEM of *E. coli* producing outer membrane targeted proteins (1 of 4). Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.BamE* (**BamE***), pET3a.BamE*RFP (**BamE*RFP**) or pET3a.BamE*MamQs (**BamE*MamQs**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

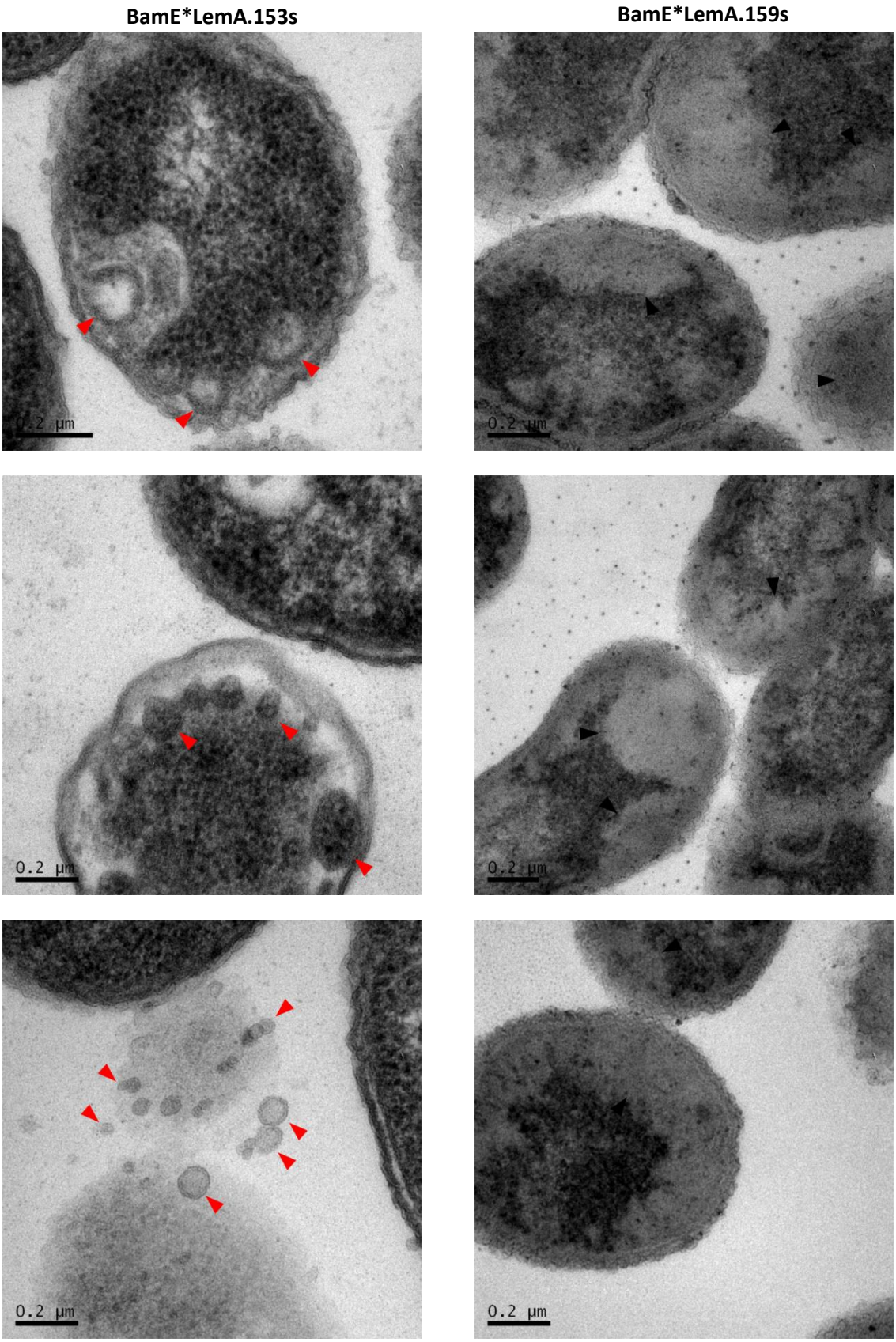


Figure 5.13. TEM of *E. coli* producing outer membrane targeted proteins (2 of 4).
Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.BamE*LemA.153s (**BamE*LemA.153s**) or pET3a.BamE*LemA.159s (**BamE*LemA.159s**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

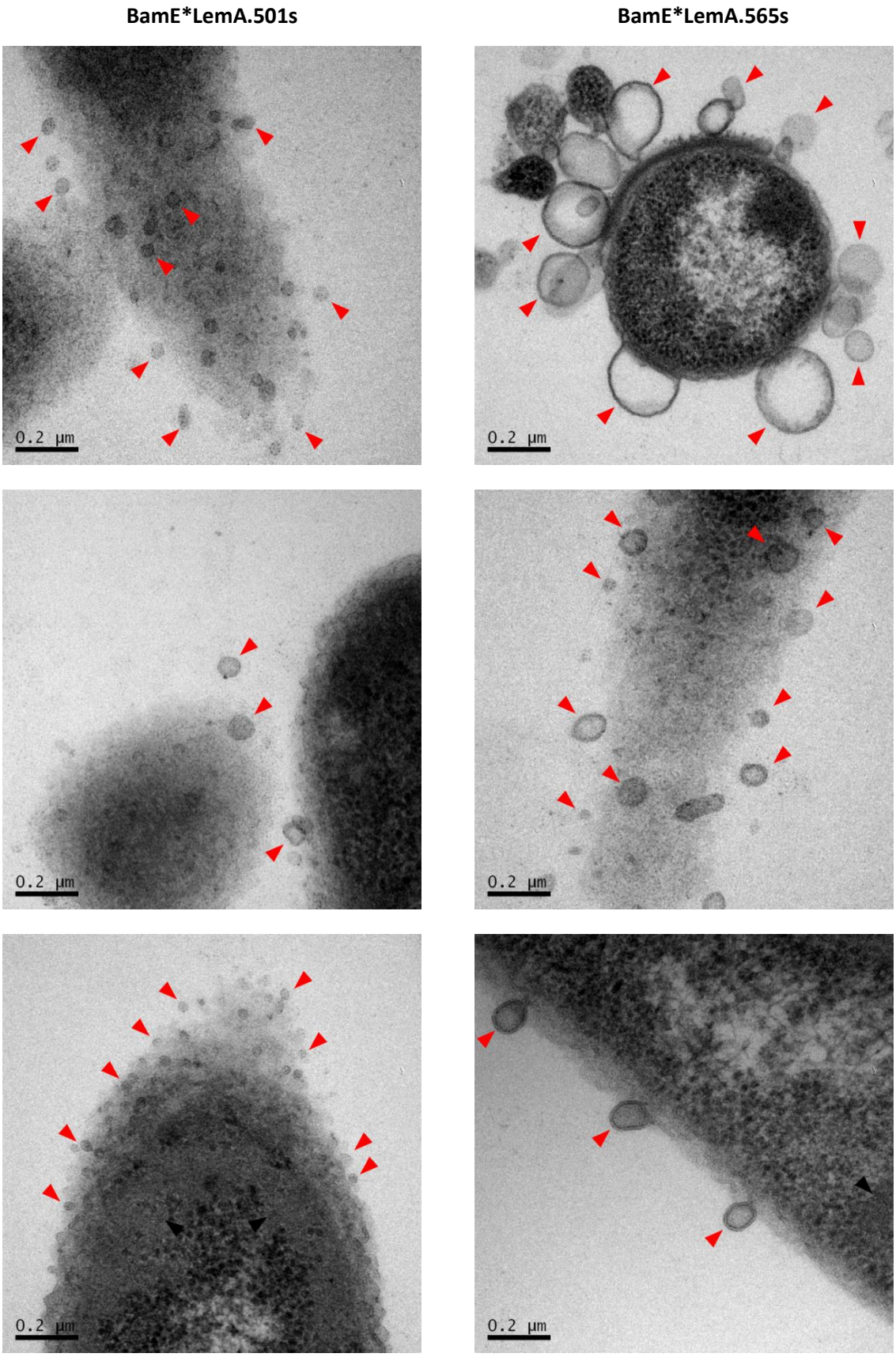


Figure 5.14. TEM of *E. coli* producing outer membrane targeted proteins (3 of 4).
Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.BamE*LemA.501s (**BamE*LemA.501s**) or pET3a.BamE*LemA.565s (**BamE*LemA.565s**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.2µm.

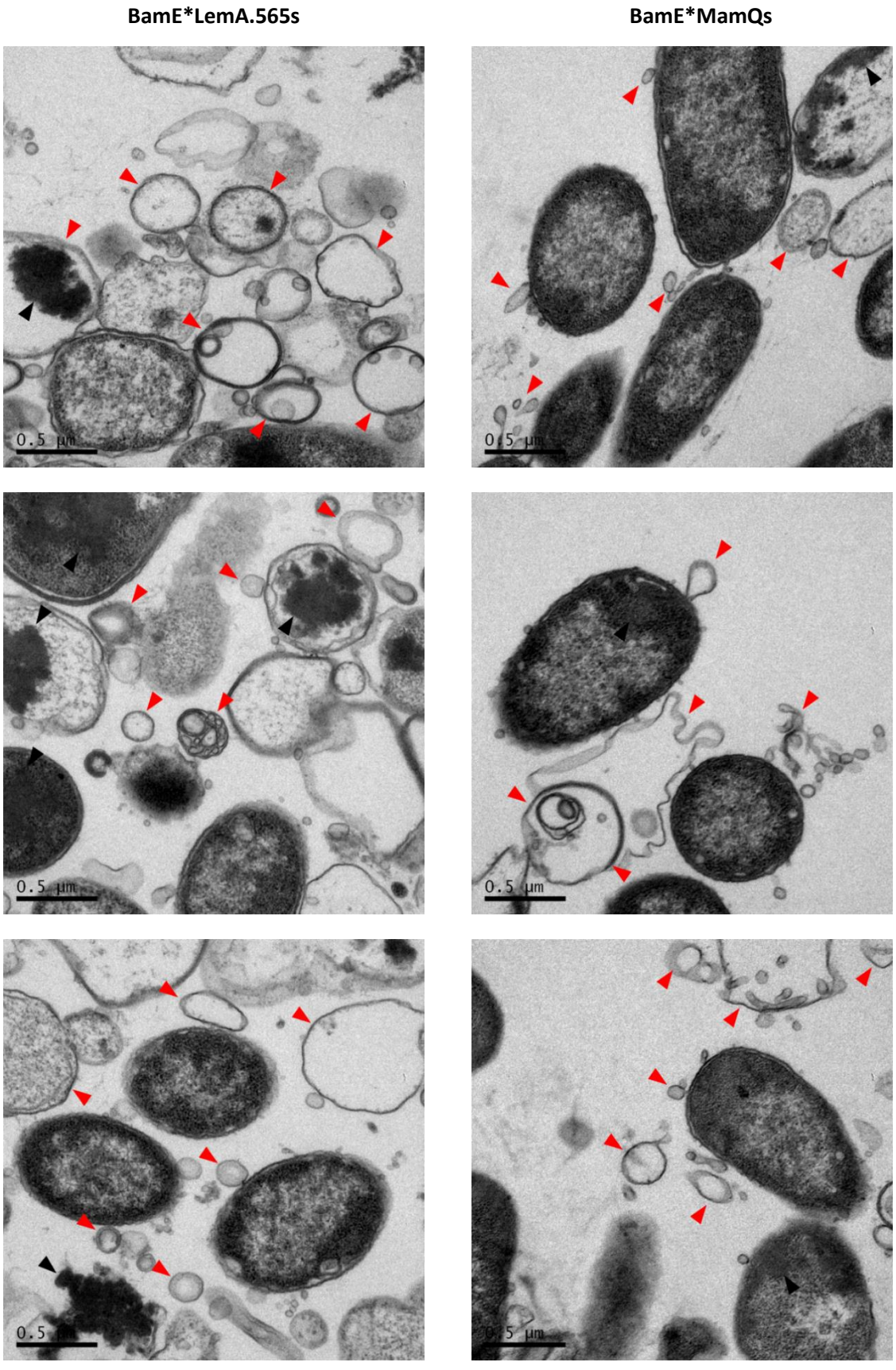


Figure 5.15. TEM of *E. coli* producing outer membrane targeted proteins (4 of 4). Transmission electron micrographs of sectioned *E. coli* BL21 (DE3) Star cells expressing pET3a.BamE*LemA.565s (**BamE*LemA.565s**) or pET3a.BamE*MamQs (**BamE*MamQs**). **Red arrows** – points of interest, **black arrows** – inclusion bodies. Scale bar is 0.5µm.

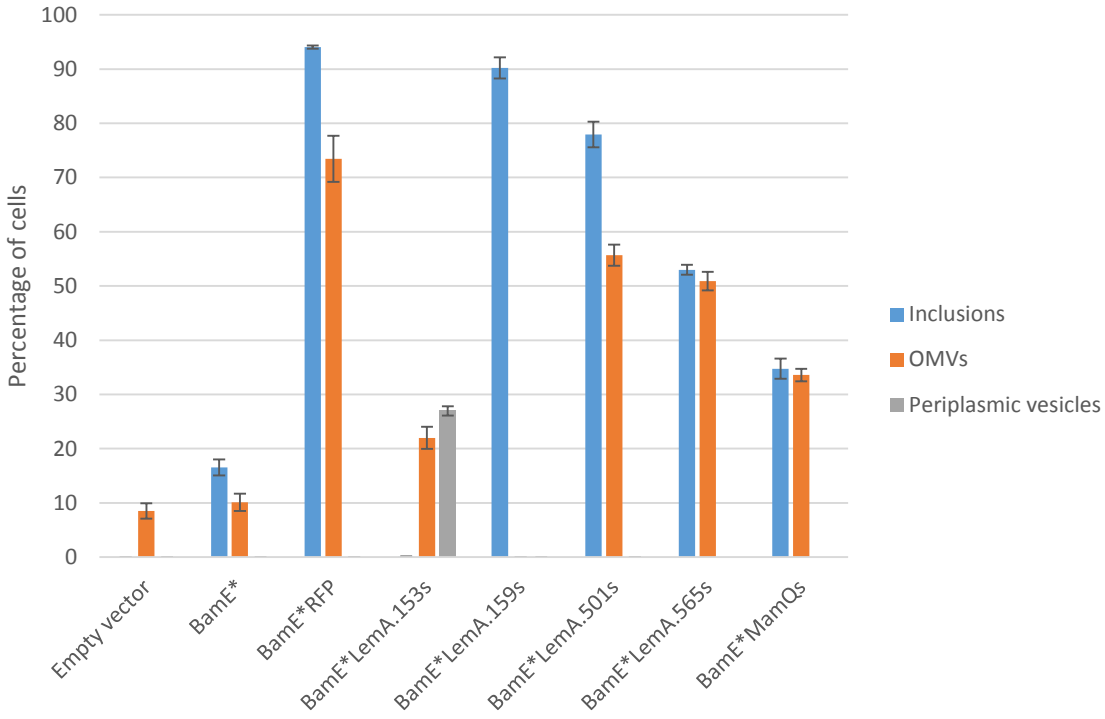


Figure 5.16. Quantification of the phenotypes produced by outer membrane targeted proteins.
Quantification of phenotypes observed in strains producing proteins targetted to the periplasm via the BamE tag.

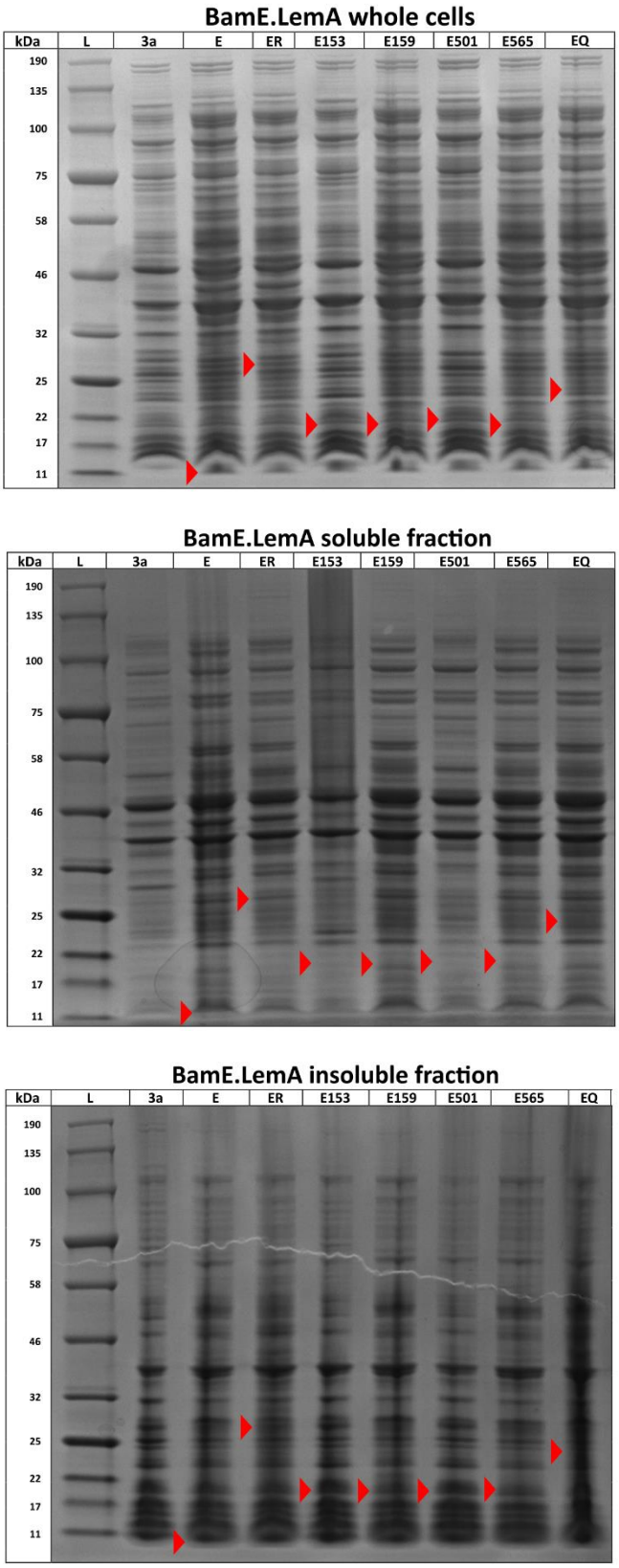


Figure 5.17. SDS-PAGE analysis of *E.coli* producing different outer membrane targeted proteins.

SDS-PAGE analysis of *E. coli* BL21 (DE3) Star cells expressing different proteins tagged with the BamE outer-membrane targeting sequence. 3a is empty vector control; E is BamE*; ER is BamE*RFP; E153 is BamE*LemA.153s; E159 is BamE*LemA.159s; E501 is BamE*LemA.501s; E565 is BamE*LemA.565s; EQ is BamE*MamQs. **Red arrows** indicate predicted sizes.

5.3. Discussion

In this chapter, the effects on the morphology of *E. coli* membrane by different magnetosome proteins and their combinations are shown. The information collected allowed identification of the LemA protein family as a potential membrane restructuring group of proteins, which was confirmed for the four tested members from different organisms. Finally, when an outer-membrane targeting sequence was attached to the soluble domains of these membrane restructuring proteins enhanced OMV formation was observed, although this was also observed when a RFP was targeted using the same tag. A large proportion of cells producing these tagged variants of the proteins also contained inclusion bodies.

Initially, the aim of the research was to characterise the effects of the predicted membrane restructuring proteins from the MAI (MamI, MamL, MamQ, MamB and MamY), that had been previously shown to be necessary or involved in magnetosome membrane formation (Murat *et al.,* 2010a; Tanaka *et al.,* 2010; Lohbe *et al.,* 2016). Furthermore, from the four operons involved in magnetosome biosynthesis in the MAI, only the *mamAB* operon was shown to be necessary for magnetosome membrane formation (Lohbe *et al.,* 2011; Raschdorf *et al.,* 2016). These results suggest that the four proteins from the *mamAB* operon (MamI, MamL, MamQ and MamB) are, at least to some extent, involved in magnetosome membrane biogenesis. Furthermore, a protein from the *mamXY* operon, MamY, was shown to induce liposome tubulation (Tanaka *et al.*, 2010). Because of a more general interest in generating membranous compartments in *E. coli*, it was decided to first study the individual effects of these proteins on the cells. Two of these proteins, MamQ and MamY, had observable phenotypic effects on the cellular membrane. Interestingly, no membrane tubulation was observed when MamY was overproduced, but small vesicular structures were visible. When MamQ was overproduced a variety of membranous phenotypes were observed, ranging from membranous invaginations to a variety of poly-membranous structures. By producing this protein at lower levels, the cells showed a more conserved phenotype, which was a singular membranous invagination of the inner membrane. Since the only analysed protein from the *mamAB* operon, which is the only operon necessary for magnetosome formation, that showed a strong effect on the inner membrane was MamQ, the protein is suggested to play an important role in the initiation of the organelle formation.

When MamQ was co-produced with the other proteins suggested to be necessary for magnetosome membrane formation, different membranous structures were observed. As more of these proteins were produced together, more defined, vesicular, structures were seen. These results provide evidence that all of these proteins are involved in magnetosome membrane formation, as the phenotype observed changes when different combinations of these proteins are produced. Further experiments need to be carried out, such as incorporating MamE, MamM and MamO in the combinatorial experiments, as these proteins have been recently shown to be required for magnetosome membrane formation (Raschdorf *et al.,* 2016).

As MamQ had a strong effect on the membrane of *E. coli* it was logical to look for related proteins in an attempt to identify other membrane restructuring proteins. Protein sequence analysis revealed that MamQ is part of a large family of proteins called the LemA family. The first analysis of a LemA protein was carried out by *Lenz et al.* (Lenz *et al.*, 1996) and at the time it was suggested to be involved in a signalling pathway. Four different LemA family proteins were chosen from a variety of organisms.

A bioinformatics analysis was carried out in order to try to understand the function of the proteins better. A total of 52 different LemA family proteins were compared not only for their amino-acid sequence, but also for similarities in their operon structures. Four distinct groups were identified based on their amino acid sequence (**Figure 5.8**). Group 1, which contained MamQ (● mamQ) and LemA.159 (● CKL_2444), had no similarity in their operon structure. Group 2, which was the biggest group, contained LemA.153 (● BMEA_A1872) and LemA.501 (● BcellWH2_04127). Members of this group were always found in one operon with one or two members of the TPM_Phosphatase family proteins and is discussed further below. Group 3, which contained LemA.565 (● PA1S_23740), were often (4 members out of 9) found in an operon with a peptidase. This group was specific to gram-negative organisms. Members of group 4, which contained none of the proteins reported in this study, were always found in one operon with a member of the Peptidase_M48 or Peptidase_M56 family. Furthermore, group 4 proteins were only found in Gram-positive organisms.

Group 2 proteins appeared always to be co-expressed with a TPM_Phosphatase family protein. The first TPM_Phosphatase (TLP18.3) was initially characterised in the thylakoids of *Arabidopsis thaliana* (Sirpiö *et al.,* 2007) with further structural and biochemical

experiments carried out by Wu *et al.* (Wu *et al.*, 2011). These studies showed that the protein localises to the thylakoid membrane with no further specific localisation and that the protein is up-regulated during high-light stress and is a phosphatase with a variety of substrates. A similar protein was discovered in *Synechocystis* 6803 and was suggested to be conserved in all thylakoid-containing cyanobacteria (Wegener *et al.,* 2011). The studies conclude, that the protein is likely to be involved in the repair of photosystem II in the thylakoids. Our bioinformatics analysis showed that not all organisms containing this family member express a photosynthetic system, which suggests a different function either in the organisms, or altogether. Regarding the proteins co-expression with LemA proteins, the TPM_Phosphatases localise to the highly-curved membrane structures of thylakoids and are over-expressed during high-light stress. Recently, it has been discovered that thylakoid membranes undergo morphological changes during such stress (Fristedt *et al.,* 2009 and Herbastova *et al.,* 2012) and it would be un-wise to rule out the involvement of the TPM_Phosphatase in this change, as the phosphatase could act on the phospholipid bilayer itself. It is also worth noting that at least one of the TPM_Phosphatases co-expressed with group 2 LemA family proteins contain a hydrophobic glycine-rich C-terminus domain, which is not predicted to form a transmembrane helix (in the cases of 2 copies of TPM_Phosphatases co-expressed with LemA, only one of them contained this motif). Such domains have been showed to be involved in RNA binding (Burd and Dreyfuss, 1994), but the highly hydrophobic domain might be interacting with the phospholipid bi-layer in an as of yet uncharacterised manner, possibly inducing curvature.

Group 3 and group 4 members are often co-expressed with peptidases. These identified peptidases were from the Peptidase_M48 and/or Peptidase_M56 families, which have high similarity, and are often identified as HtpX (11/13 from the identified peptidases). This peptidase has been showed to be involved in cellular-stress response (Sakoh *et al.*, 2005) while other members from these families have been shown to be involved in cellular-stress signalling (Zhang *et al.*, 2001). HtpX has been shown to work in conjunction with FtsH in the maintenance of membranous proteins in *E. coli* (Sakoh *et al.*, 2005). Five of the analysed LemA proteins, localised in these groups, had a periplasmic N-terminus belonging to the FtsH_ext (N-terminus) family. This domain has been shown to be involved in protein-protein interactions, both with itself and with other proteins (Akiyama, 1998), suggesting that the LemA proteins with this motif might work in conjunction with membranous protein maintenance.

Overall, the operon analysis suggests LemA proteins to be involved in cell-signalling or membrane protein maintenance. Nevertheless, the lack of consistency in the operons and the missing knowledge about the co-expressed proteins leaves the functions of LemA proteins up for debate.

Production of different LemA proteins in *E. coli* BL21 (DE3) Star cells yielded similar results to the production of MamQ (**Figure 5.9**). No protein overproduction could be observed by SDS-PAGE analysis, but when electron microscopy analysis of the cells producing the different proteins was carried out, a variety of membranous phenotypes could be observed, which varied based on the family member produced. This suggest that the proteins are likely to play different biological roles, as they would localise to different membrane curvature.

As the observed effects were all on the inner-membrane of *E. coli*, it was of interest to investigate how the cells would behave if these proteins were targeted to the outer-membrane. To achieve this, the outer-membrane targeting sequence (the first 21 amino-acids) of an outer-membrane lipoprotein, BamE, was fused to the soluble domains of the LemA proteins and an RFP, as a negative control. As with MamQ and wild-type LemA proteins, no overproduction could be observed by SDS-PAGE analysis. All the hybrids, apart from the BamE*LemA.159s, showed significantly increased OMV formation. Interestingly, the BamE*RFP hybrid showed the largest amount of OMV formation, suggesting that the overproduction of any protein targeted to the outer membrane can enhance this process. A large proportion of lysed cells as well as cells with large OMVs were also observed when BamE*LemA.565s or BamE*MamQs was produced. The increase in size of the vesicles may be the effect of the membrane restructuring by these proteins and the cell lysis can be explained by the decrease in outer-membrane rigidity due to vesiculation, leading to compromised cellular integrity. Questions still remain on how the production of these proteins impact the cellular growth. Furthermore, a detailed characterisation of the produced vesicles needs to be carried out.

Overall, the production of the magnetosome proteins, MamQ and MamY, seems to have an effect on the *E. coli* inner membrane. When combined with other magnetosome proteins this effect can be varied. Furthermore, MamQ belongs to a large family of proteins, which are able to restructure the bacterial membrane in a variety of ways. It was discovered

that the overproduction of proteins, targeted to the outer membrane, can enhance OMV production.

Chapter 6

Discussion

6.1. Discussion

The results presented in this thesis reveal a variety of engineered membranous structures in the model organism *Escherichia coli*. These include a reconstructed recombinant magnetosome, a magnetosome membrane only and other inner or outer membranous structures.

Firstly, in order to produce a recombinant magnetosome, a construct containing the wild-type *mamAB* operon, was made. The initial approach of amplifying the whole operon using a long range DNA polymerase system from Roche™ yielded no product. This was surprising as the size of the operon is well within the amplification range of the system, but even after incorporating a number of troubleshooting approaches, no product was ever obtained. An alternative approach was then designed, which relies on amplifying the operon in two large segments. Interestingly, it was quite difficult to obtain any product even when the amplicon size was reduced to about 9 kilobases for each of the fragments. A large amount of time was spent on optimisation and even then the product yield was inconsistent, which may be due to the poor quality of the template DNA. Alternatively, the manufacturers of the kit may have over-stated the ability of the polymerase. Once the two fragments were combined, DNA sequencing revealed a frame-shift mutation in one of the key genes, *mamE*. It is known that frequent mutations occur in the genomic region where the *mamAB* operon is localised (Kolinko *et al*., 2011). It was decided to try to fix this mutation using a small PCR fragment containing this region. Unfortunately, no colonies were observed after multiple attempts to transform the ligation (data not shown).

At the time of this research little information was available on MamE, therefore it was decided to conduct growth experiments using the construct with the frame-shift mutation. Interestingly, small electron dense particles, which often appeared to be intracellular, could be observed when strains harbouring this construct were grown under conditions used to culture two different *Magnetospirillum* species, which naturally produce magnetosomes. Similar structures were present in the control, but at a lower frequency. When an additional operon, *mms6*, was introduced into the construct a small proportion of cells produced particles which had the appearance of magnetosomes, although no magnetic response was observed. Significantly, such structures were not present in the control, although small electron dense particles were observed. The presence of non-magnetic electron dense particles similar to magnetosomes could be explained by the phenotype observed in *M.*

magneticum cells, expressing different mutated versions of MamE (Quinlan *et al.,* 2011). The work showed that the MamE protease and magnetochrome domains are not necessary for biomineralisation to take place, but the produced minerals were smaller and showed no magnetic response. It is possible that the electron dense particles observed in the cells expressing the *mms6* operon and the *mamAB* operon, with the mutated *mamE*, are non-magnetic iron crystals within pseudo-magnetosomes. If that is the case, the N-terminal domain of MamE might act as a hub for protein complex assembly, allowing localisation of other factors involved in magnetosome assembly.

In order to see if individual magnetosome proteins are produced in *E. coli*, constructs corresponding to each of these, apart for MamG, were made. In the case of MamG the genomic DNA used as the template had a frame-shift mutation in the sequence of *mamG* confirmed by sequencing. Further analysis of this protein was not carried out as it has been shown not to be required for magnetosome synthesis and the production of other proteins from the *mamGFDC* operon can rescue the phenotype produced by Δ*mamG* strains (Scheffel *et al.,* 2008). When the constructs were expressed in *E. coli* BL21 (DE3) Star cells, over-production was only observable for 7 out of the 29 gene products when analysed by SDS-PAGE. This suggests low or no production of these proteins. This is not surprising as most of these proteins are membrane proteins, which usually require extensive optimisation in order to obtain detectable overproduction (Freigassner *et al*, 2009). When some of these proteins were analysed by western blotting, identification of 12 out of 14 proteins was achieved. Two of the proteins which were not identified, MamN and MamO, both have an N-terminus transmembrane helix. This helix may have been compromised by the recombinant addition of the N-terminus hexa-histidine tag, which was used as the antigen for western blotting. Interestingly, multiple analysed proteins showed a variation in the molecular weight observed by SDS-PAGE when compared to their predicted weight. This may be due to protein degradation or the binding of detergent molecules via the hydrophobic domains, thereby obscuring their migration in the gel, as observed previously (Rath *et al.,* 2009). It is suggested that these bands are analysed using a peptide mass fingerprinting approach to identify any degradation. Furthermore, some of the proteins are relatively small (under 10 kDa) and migrate close to the dye front of the gels. In order to resolve these proteins a different SDS-PAGE approach should be employed.

Cytochrome c proteins have been shown to be involved in the production of magnetosomes (Siponen *et al.,* 2013; Lohbe *et al.,* 2016). In this work, the production of

three cytochrome c proteins, involved in magnetosomes formation, and their maturation in the periplasm of *E. coli* is shown. The experiments identify the importance of specific growth conditions required for magnetosome production in this model organism, as the correct processing of these domains can only be achieved under specific growth conditions or when accessory proteins are co-expressed recombinantly. This represents the special requirements to encourage the formation of a covalent bond between a cysteine residue and the vinyl side chain of the porphyrin molecule.

When generating synthetic multi-gene operons, it was found that constructs containing *mamE* had very poor transformation efficiencies. When a pET3a vector was used, the largest combination of genes from the *mamAB* operon that could be combined into a single construct was 10, which was reduced to only 4 when *mamE* was present. This might explain why it was not possible to correct the mutation in *mamE* present in the *mamAB* construct (Chapter 2). These limitations could be partially overcome by a switch to the pETcoco2 cloning system, which has a significantly lower copy number and a more tightly regulated promoter. Constructs containing *mamE* still had lower transformation efficiencies, but transformants could be obtained. It was possible to reconstruct the desired modified *mamAB* operon, containing all the genes, except for *mamJ, mamK* and *mamU*. Unfortunately, due to time limitations and difficulties transforming this plasmid into a strain with a T7 polymerase, growth experiments of cells producing the construct have not been carried out.

Production of one of two individual magnetosome proteins (MamY or MamQ), suggested to be involved in membrane vesicle generation, showed an effect on the inner membrane of *E. coli*. Cells producing MamY appeared to have small vesicular structures, while cells producing MamQ, had a variety of phenotypes, which varied based on the production level. It is important to note that the analysed cells had undergone chemical fixation. Such fixation may introduced various artefacts. Therefore, a control sample and quantification of the phenotype was critical during these experiments. The results show that the effect on the membrane is based on the protein production and not the fixation method. Furthermore, when additional proteins implemented in magnetosome membrane formation were introduced, more defined structures were observed. It is suggested that other electron microscopy sample preparation methods, such as cryofixation, should be used to observe these structures in a more native state. It would also be informative to carry out electron tomography experiments to better define the observed structures.

Interestingly, when SDS-PAGE analysis was carried out for strains producing MamQ or MamY, no clear overproduction was observed. Similar membranous structures to those observed when MamQ was produced, have been reported when the b subunit of F_1F_o ATP synthase was produced, although in this research clear overproduction of the protein is observed by SDS-PAGE analysis (Arechaga *et al.*, 2000). This suggests that comparatively small amounts of MamQ are required to illicit the drastic changes in membrane structure observed, which points to a membrane bending mechanism not dependent on scaffolding or direct bending dependant on the shape of the membrane protein (McMahon and Gallop, 2005). During the bioinformatics analysis, MamQ was predicted to carry out fatty acid oxidation. It is known that phospholipids, containing oxidised fatty acid side-chains, induce membrane curvature (McMahon and Gallop, 2005). A large amount of such fatty acid side-chains incorporated into the phospholipid bilayer would explain the phenotypes observed when MamQ is produced. It is therefore suggested that the analysis of the lipid content of these cells is carried out and compared to that of wild-type cells.

When other proteins, belonging to the same protein family as MamQ, were produced in *E. coli*, a variety of membranous phenotypes were also observed. Interestingly, these varied quite drastically, suggesting different functions for these proteins in their native environments. As suggested above, it would be informative to carry out lipid analysis of the strains producing these proteins as well as use a variety of alternative electron microscopy approaches to analyse the structures produced by these proteins.

While attempting to target membrane restructuring proteins to the outer membrane, it was discovered that the overproduction of any of the tested targeted proteins (including RFP) induced OMV production in *E. coli*. Purification of these vesicles should be carried out to investigate if the targeted proteins are incorporated into the vesicles. Furthermore, the vesicle production enhancement should be tested for a variety of other proteins with different targeting sequences, as only six different proteins with one targeting sequence were used in the experiments reported herein.

Production of various membranous structures in bacteria, as shown in this work, could have a number of biotechnological applications (**Figure 6.1**). Cytoplasmic membranous compartments could be used as a basis for synthetic organelles for mammalian-like protein modifications, as nanobioreactors, in order to enhance synthetic pathway flux or to protect the cell from toxic intermediates, or to improve the energy generation potential of the

bacteria by employing these as synthetic mitochondria-like organelles. Poly-membrane bodies, which are observed when MamQ is produced, could be used to enhance the membrane surface area. This might result in greater production of membranous proteins and enhanced phospholipid production. OMVs have been used as vaccines for over 20 years (Acevedo *et al*., 2014). Such vesicles are highly immunogenic and are excellent carriers for vaccines. Engineering an *in vivo* system, able to produce high amount of such vesicles with specific protein incorporation, might allow for the generation of efficient, low price vaccines. The work discussed, shows a potential for such a system in *E. coli*, whereby overproduction of a protein, targeted to the outer membrane via a specific tag, can induce the formation of such vesicles. These could then be purified and used as potent vaccines, due to their native immunogenicity and the ability to carry a specialised cargo. Other potential applications of OMVs could be the incorporation of biotechnologically relevant pathways. This could potentially increase the flux of the pathways due to close proximity within the OMV, protect the cell from any toxic intermediates produced as well as sequester the product and any intermediates from potential degradation within the cell and increase the substrate availability, as only transport through a single membrane would be required. Furthermore, as it is possible to modify the outside of the outer membrane (Rice *et al.,* 2006), such vesicles could be functionalised on the outside to be used for targeted drug delivery or protein purification. However, an important question still remains regarding the efficiency of incorporation of the targeted proteins into these vesicles and could be a basis for future work.

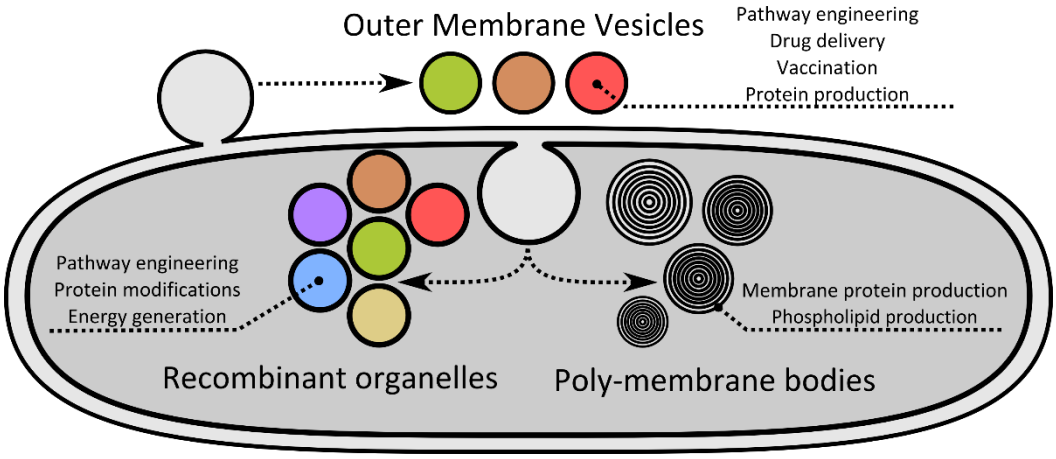


Figure 6.1. Possible application for various membranous structures.
Schematic representation of *E. coli* cell showing various applications of intracellular and extracellular membranous vesicles.

The work discussed contributes to the growing field of synthetic biology exploring a new avenue of membrane engineering in bacteria. A full genetic library of the genes, involved in magnetosome formation, was constructed for analysis in *E. coli*. These were then combined in a modular manner to produce synthetic operons involved in the synthesis of the organelle. This approach allows for easy modification of the genes contained in the constructs for further research into and engineering of the membranous organelle. Analysis of the individual genes led to the identification of a suggested membrane restructuring protein family. Attempts to utilise this protein family for modifying the outer membrane resulted in the discovery of a novel method for increased OMV production. Membranous structures produced by these approaches expand the capabilities for cellular engineering, allowing for new pathways, which may depend on membranous proteins, to be produced in a specialised environment and the development of specialised liposomes.

7. References

Abreu, F., Silva, K. T., Martins, J. L., and Lins, U. (2006). Cell viability in magnetotactic multicellular prokaryotes. *International Microbiology*, **9**, 267–272.

Abreu, F., Martins, J. L., Souza-Silveira, T., Keim, C. N., Lins de Barros, H. G. P., Filho, F. J. G., and Lins, U. (2007). "*Candidatus* magnetoglobus multicellularis", a multicellular, magnetotactic prokaryote from a hypersaline environment. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 1318–1322.

Acevedo, R., Fernandez, S., Zayas, C., Acosta, A., Sarmiento, M. E., Ferro, V. A., Rosenqvist, E., Campa, C., Cardoso, D., Garcia, L., and Perez, J. L. (2014). Bacterial outer membrane vesicles and vaccine applications. *Frontiers in Immunology*, **5**, 1–6.

Akiyama, Y., Kihara, A., Mori, H., Ogura, T., and Ito, K. (1998). Roles of the periplasmic domain of *Escherichia coli* FtsH (HflB) in protein interactions and activity modulation. *Journal of Biological Chemistry*, **273**, 22326–22333.

Alphandery, E. (2014). Applications of magnetosomes synthesized by magnetotactic bacteria in medicine. *Frontiers in Bioengineering and Biotechnology*, **2**, 5.

Amemiya, Y., Tanaka, T., Yoza, B., and Matsunaga, T. (2005). Novel detection system for biomolecules using nano-sized bacterial magnetic particles and magnetic force microscopy. *Journal of Biotechnology*, **120**, 308–314.

Amemiya, Y., Arakaki, A., Staniland, S. S., Tanaka, T., and Matsunaga, T. (2007). Controlled formation of magnetite crystal by partial oxidation of ferrous hydroxide in the presence of recombinant magnetotactic bacterial protein Mms6. *Biomaterials*, **28**, 5381–5389.

Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., De Kruijff, B., Runswick, M. J., and Walker, J. E. (2000). Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F1F(o) ATP synthase. *FEBS Letters*, **482**, 215–219.

Bain, J., and Staniland, S. S. (2015). Bioinspired nanoreactors for the biomineralisation of metallic-based nanoparticles for nanomedicine. *Physical Chemistry Chemical Physics,* **17**, 15508–15521.

Balkwill, D. L., Maratea, D., and Blakemore, R. P. (1980). Ultrastructure of a magnetotactic spirillum. *Journal of Bacteriology*, **141**, 1399–1408.

Bazylinski, D. A., Frankel, R. B., Heywood, B. R., Mann, S., King, J. W., Donaghay, P. L., and Hanson, A. K. (1995). Controlled biomineralization of magnetite (Fe3O4) and greigite (Fe3S4) in a magnetotactic bacterium. *Applied and Environmental Microbiology*, **61**, 3232–3239.

Bazylinski, D. A., Frankel, R. B., and Jannasch, H. W. (1988). Anaerobic magnetite production by a marine, magnetotactic bacterium. *Nature*, **334**, 518–519.

Bazylinski, D. a., Heywood, B. R., Mann, S., and Frankel, R. B. (1993). Fe3O4 and Fe3S4 in a bacterium. *Nature*, **366**, 218.

Bazylinski, D. A., and Frankel, R. B. (2004). Magnetosome formation in prokaryotes. *Nature Reviews Microbiology*, **2**, 217–230.

Bazylinski, D. A., Williams, T. J., Lefevre, C. T., Trubitsyn, D., Fang, J., Beveridge, T. J., Moskowitz, B. M., Ward, B., Schubbe, S., Dubbels, B.L., and Simpson, B. (2013). *Magnetovibrio blakemorei* gen. nov., sp. nov., a magnetotactic bacterium (Alphaproteobacteria: *Rhodospirillaceae*) isolated from a salt marsh. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 1824–1833.

Bellini, S. (2009a). On a unique behaviour of freshwater bacteria. *Chinese Journal of Oceanology and Limnology*, **27**, 3-5.

Bellini, S. (2009b). Further studies on “magnetosensitive bacteria”. Chinese Journal of Oceanology and Limnology, **27**, 6–12.

Blakemore, R. (1975). Magnetotactic bacteria. *Science*, **190**, 377-379.

Blakemore, R. P., Maratea, D., and Wolfe, R. S. (1979). Isolation and pure culture of a freshwater magnetic spirillum in chemically defined medium. *Journal of Bacteriology*, **140**, 720–729.

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *eLife*, **4**, 1–12.

Boze, H., Marlin, T., Durand, D., Perez, J., Vemhet, A., Canon, F., Sarni-Manchado, P., Cheynier, V., and Cabane, B. (2010). Proline-rich salivary proteins have extended conformations. *Biophysical Journal*, **99**, 656–665.

Bramhill, D., and Thompson, C. M. (1994). GTP-dependent polymerization of *Escherichia coli* FtsZ protein to form tubules. *Proceedings of the National Academy of Sciences of the United States of America,* **91**, 5813–5817.

Burd, C.G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**: 616-621.

Butler, R. F. and Banerjee, S. K. (1975). Theoretical single-domain grain size range in magnetite and titanomagnetite. *Journal of Geophysical Research*, **80**, 4049-4058.

Chandler, D. E., Gumbart, J., Stack, J. D., Chipot, C., and Schulten, K. (2009). Membrane curvature induced by aggregates of LH2s and monomeric LH1s. *Biophysical Journal*, **97**, 2978–2984.

Chatfield, C. H., Mulhern, B. J., Viswanathan, V. K., and Cianciotto, N. P. (2012). The major facilitator superfamily-type protein LbtC promotes the utilization of the legiobactin siderophore by *Legionella pneumophila*. *Microbiology*, **158**, 721–735.

Cornejo, E., Subramanian, P., Li, Z., Jensen, G. J. and Komeili, A. (2016). Dynamic remodeling of the magnetosome membrane is triggered by the initiation of biomineralization. *mBio*, **7**, 1–9.

Dacaranhe, C. D., and Terao, J. (2001). Effect of phosphatidic acid and phosphatidylserine on lipid oxidation in beef homogenate during storage and in emulsified sardine oil. *Journal of Food Science*, **66**, 422–427.

DeLong, E. F., Frankel, R. B., and Bazylinski, D. A. (1993). Multiple evolutionary origins of magnetotaxis in bacteria. *Science*, **259**, 803–806.

Diaz-Ricci, J. C., and Kirschvink, J. L. (1992). Magnetic domain state and coercivity predictions for biogenic greigite (Fe3s4) - a comparison of theory with magnetosome observations. *Journal of Geophysical Research-Solid Earth,* **97**, 17309–17315.

Ding, Y., Li, J., Liu, J., Yang, J., Jiang, W., Tian, J., Li, Y., Pan, Y., and Li, J. (2010). Deletion of the *ftsZ - Like* gene results in the production of superparamagnetic magnetite magnetosomes in *Magnetospirillum gryphiswaldense*. *Journal of Bacteriology*, **192**, 1097–1105.

Dorn, M., E Silva, M. B., Buriol, L. S., and Lamb, L. C. (2014). Three-dimensional protein structure prediction: methods and computational strategies. *Computational Biology and Chemistry*, **53**, 251–276.

Draper, O., Byrne, M. E., Li, Z., Keyhani, S., Barrozo, J. C., Jensen, G., and Komeili, A. (2011). MamK, a bacterial actin, forms dynamic filaments in vivo that are regulated by the acidic proteins MamJ and LimJ. *Molecular Microbiology*, **82**, 342–354.

Frankel, R. B., Bazylinski, D. A., Johnson, M. S., and Taylor, B. L. (1997). Magneto-aerotaxis in marine coccoid bacteria. *Biophysical Journal*, **73**, 994–1000.

Frankel, R. B. (2009). The discovery of magnetotactic/magnetosensitive bacteria. *Chinese Journal of Oceanology and Limnology,* **27**, 1.

Freigassner, M., Pichler, H., and Glieder, A. (2009). Tuning microbial hosts for membrane protein production. *Microbial Cell Factories*, **8**, 69.

Fristedt, R., Willig, A., Granath, P., Crèvecoeur, M., Rochaix, J. D., and Vener, A. V. (2009). Phosphorylation of photosystem II controls functional macroscopic folding of photosynthetic membranes in Arabidopsis. *The Plant Cell*, **21**, 3950–3964.

Gorby, A., Beveridge, T. J., and Blakemore, R. P. (1988). Characterisation of the bacterial magnetosome. *Journal of Bacteriology*, **170**, 834–841.

Grunberg, K., Wawer, C., Tebo, B. M. and Schuler, D. (2001). A large gene cluster encoding several magnetosome proteins is conserved in different species of magnetotactic bacteria. *Applied and Environmental Microbiology,* **67**, 4573–4582.

Grunberg, K., Muller, E., Otto, A., Reszka, R., Linder, D., Kube, M., Reinhardt, R., and Schuller, D. (2004). Biochemical and proteomic analysis of the magnetosome membrane in *Magnetospirillum gryphiswaldense*. *Applied and Environmental Microbiology*, **70**, 1040–1050.

Harold, F. M. (2005). Molecules into cells: specifying spatial architecture molecules into cells: specifying spatial architecture. *Microbiology and Molecular Biology Reviews*, **69**, 544–564.

Herbstová, M., Tietz, S., Kinzel, C., Turkina, M. V, and Kirchhoff, H. (2012). Architectural switch in plant photosynthetic membranes induced by light stress. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 20130–5.

Hershey, D. M., Ren, X., Melnyk, R. A., Browne, P. J., Ozyamak, E., Jones, S. R., Chang, M. C. Y., Hurley, J. H., and Komeili, A. (2016). MamO is a repurposed serine protease that promotes magnetite biomineralization through direct transition metal binding in magnetotactic bacteria. *PLoS Biology*, **14**, e1002402.

Iobbi-Nivol, C., Crooke, H., Griffiths, L., Grove, J., Hussain, H., Pommier, J., Mejean, V., and Cole, J.A. (1994). A reassessment of the range of c-type cytochromes synthesized by Escherichia coli K-12. *FEMS Microbiology Letters,* **1**, 89-94.

Jogler, C., and Schuler, D. (2009). Genomics, genetics, and cell biology of magnetosome formation. *Annual Review of Microbiology*, **63**, 501–521.

Jogler, C., Niebler, M., Lin, W., Kube, M., Wanner, G., Kolinko, S., Stief, P., Beck, A. J., de Beer, D., Petersen, N., Pan, Y., Amann, R., Reinhardt, R., and Schüler, D. (2010). Cultivation-independent characterization of "Candidatus *Magnetobacterium bavaricum*" via ultrastructural, geochemical, ecological and metagenomic methods. *Environmental Microbiology*, **12**, 2466–2478.

Keim, C. N., Martins, J. L., Abreu, F., Rosado, A. S., De Barros, H. L., Borojevic, R., Lins, U., and Farina, M. (2004). Multicellular life cycle of magnetotactic prokaryotes. *FEMS Microbiology Letters*, **240**, 203–208.

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta., A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V.,

Collado-Vides, J., Gunsalus, R.P., Paulsen, I. ad Karp, P.D. (2013). EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Research*, **41**, 605–612.

Kolinko, I., Jogler, C., Katzmann, E., and Schüler, D. (2011). Frequent mutations within the genomic magnetosome island of magnetospirillum gryphiswaldense are mediated by RecA. *Journal of Bacteriology*, **193**, 5328–5334.

Kolinko, I., Lohbe, A., Borg, S., Raschdorf, O., Jogler, C., Tu, Q., Posfai, M., Tompa, E., Plitzko, J. M., Brachmann, A., Wanner, G., Muller, R., Zhang, Y. and Schüler, D. (2014). Biosynthesis of magnetic nanostructures in a foreign organism by transfer of bacterial magnetosome gene clusters. *Nature Nanotechnology*, **9**, 193–197.

Komeili, A., Vali, H., Beveridge, T. J., and Newman, D. K. (2004). Magnetosome vesicles are present before magnetite formation, and MamA is required for their activation. *Proceedings of the National Academy of Sciences of the United States of America,* **101**, 3839–3844.

Komeili, A., Li, Z., Newman, D. K., and Jensen, G. J. (2006). Magnetosomes are cell membrane invaginations organized by the actin-like protein MamK. *Science*, **311**, 242–245.

Komeili, A. (2012). Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiology Reviews*, **36**, 232–255.

Kooijman, E. E., Chupin, V., de Kruijff, B., and Burger, K. N. J. (2003). Modulation of membrane curvature by phosphatidic acid and lysophosphatidic acid. *Traffic*, **4**, 162–174.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, **305**, 567–580.

Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.F., Rowe, M.L., Mulvihill, D.P., Prentice, M.B., Howard, M.J., and Warren, M.J. (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synthetic Biology*, **3**, 454–465.

Lefevre, C. T., Abreu, F., Schmidt, M. L., Lins, U., Frankel, R. B., Hedlund, B. P., and Bazylinski, D. A. (2010). Moderately thermophilic magnetotactic bacteria from hot springs in Nevada. *Applied and Environmental Microbiology*, **76**, 3740–3743.

Lefevre, C. T., Frankel, R. B., Posfai, M., Prozorov, T., and Bazylinski, D. A. (2011a). Isolation of obligately alkaliphilic magnetotactic bacteria from extremely alkaline environments. *Environmental Microbiology*, **13**, 2342–2350.

Lefevre, C. T., Menguy, N., Abreu, F., Lins, U., Posfai, M., Prozorov, T., Pignol, D., Frankel, R. B., and Bazylinski, D. A. (2011b). A cultured greigite-producing magnetotactic bacterium in a novel group of sulfate-reducing bacteria. *Science*, **334**, 1720–1723.

Lefèvre, C. T., and Bazylinski, D. A. (2013). Ecology, diversity, and evolution of magnetotactic bacteria. *Microbiology and Molecular Biology Reviews*, **77**, 497–526.

Lenz, L. L., Dere, B., and Bevan, M. J. (1996). Identification of an H2-M3-restricted listeria epitope: implications for antigen presentation by M3. *Immunity*, **5**, 63–72.

Lesuisse, E., Simon-Casteras, M., and Labbe, P. (1998). Siderophore-mediated iron uptake in *Saccharomyces cerevisiae*: The SIT1 gene encodes a ferrioxamine B permease that belongs to the major facilitator superfamily. *Microbiology*, **144**, 3455–3462.

Li, Z., and Jensen, G. J. (2009). Electron cryotomography: a new view into microbial ultrastructure. *Current Opinion in Microbiology*, **12**, 333–340.

Li, Y., Katzmann, E., Borg, S., and Schüler, D. (2012). The periplasmic nitrate reductase nap is required for anaerobic growth and involved in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *Journal of Bacteriology*, **194**, 4847–4856.

Li, Y., Raschdorf, O., Silva, K. T., and Schuler, D. (2014). The terminal oxidase cbb3 functions in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *Journal of Bacteriology*, **196**, 2552–2562.

Lins, U., McCartney, M. R., Farina, M., Frankel, B. B., and Buseck, P. R. (2006). Crystal habits and magnetic microstructures of magnetosomes in coccoid magnetotactic bacteria. *Anais Da Academia Brasileira de Ciencias*, **78**, 463–474.

Liu, Y., Li, G. R., Guo, F. F., Jiang, W., Li, Y., and Li, L. J. (2010). Large-scale production of magnetosomes by chemostat culture of *Magnetospirillum gryphiswaldense* at high cell density. *Microbial Cell Factories*, **9**, 99.

Lohbe, A., Ullrich, S., Katzmann, E., Borg, S., Wanner, G., Richter, M., Voigt, N., Schweder, T., and Schuler, D. (2011). Functional analysis of the magnetosome island in *Magnetospirillum gryphiswaldense:* the *mamAB* operon is sufficient for magnetite biomineralization. *PLoS ONE*, **6**, e25561.

Lohbe, A., Borg, S., Raschdorf, O., Kolinko, I., Tompa, E., Posfai, M., Faivre, D., Baumgartner, J., and Schuler, D. (2014). Genetic dissection of the *mamAB* and *mms6* operons reveals a gene set essential for magnetosome biogenesis in *Magnetospirillum gryphiswaldense*. *Journal of Bacteriology*, **196**, 2658–2669.

Lohbe, A., Kolinko, I., Raschdorf, O., Uebe, R., Borg, S., Brachmann, A., Plitzko, J. M., Muller, R., Zhang, Y., and Schüler, D. (2016). Overproduction of magnetosomes by genomic amplification of biosynthetic gene clusters in a magnetotactic bacterium. *Applied and Environmental Microbiology*, **82**, 3032-3041.

Lower, B.H. and Bazylinski, D.A. (2013). The bacterial magnetosome: a unique prokaryotic organelle. *Journal of Molecular Microbiology and Biotechnology*, **23**, 63-80.

Machnicka, B., Czogalla, A., Hryniewicz-Jankowska, A., Bogusławska, D. M., Grochowalska, R., Heger, E., and Sikorski, A. F. (2014). Spectrins: a structural platform for stabilization and activation of membrane channels, receptors and transporters. *Biochimica et Biophysica Acta*, **1838**, 620–634.

Maratea, D., and Blakemore, R. P. (1981). *Aquaspirillum magnetotacticum* sp. nov., a magnetic spirillum. *International Journal of Systematic Bacteriology*, **31**, 452–455.

Matsunaga T., and Kamiya, S. (1987). Use of magnetic particles isolated from magnetotactic bacteria for enzyme immobilization. *Applied Microbiology and Biotechnology*, **4**, 328-332.

Matsunaga T., Sakaguchi T., and Tadokoro F. (1991). Magnetite formation by a magnetic bacterium capable of growing aerobically. *Applied Microbiology and Biotechnology*, **35**, 651–655.

Matsunaga, T., Okamura, Y., Fukuda, Y., Wahyudi, A. T., Murase, Y., and Takeyama, H. (2005). Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum* sp. strain AMB-1. *DNA Research*, **12**, 157–166.

McGoldrick, H. M., C. A. Roessner, Raux, E., Lawrence, A. D., McLean, K. J., Munro, A. W., Santabarbara, S., Rigby, S. E., Heathcote, P., Scott, A.I. and Warren, M. J. (2005). Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. *The Journal of Biological Chemistry*, **280**, 1086-1094.

McMahon, H. T., and Gallop, J. L. (2005) Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature*, **438**, 590-596.

Meldrum, F. C., Mann, S., Heywood, B. R., Frankel, R. B., and Bazylinski, D. A. (1993). Electron microscopy study of magnetosomes in a cultured coccoid magnetotactic bacterium. *Proceedings of the Royal Society B: Biological Sciences*, **251**, 231–236.

Moller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**, 646–653.

Morillo, V., Abreu, F., Araujo, A. C., De Almeida, L. G. P., Enrich-Prast, A., Farina, M., de Vasconcelos, A. T. R., Bazylinski, D. A., and Lins, U. (2014). Isolation, cultivation and genomic analysis of

magnetosome biomineralization genes of a new genus of South-seeking magnetotactic cocci within the *Alphaproteobacteria*. *Frontiers in Microbiology*, **5**, 1–12.

Muller, F. D., Raschdorf, O., Nudelman, H., Messerer, M., Katzmann, E., Plitzko, J. M., Zarivach, R., and Schüler, D. (2014). The FtsZ-like protein FtsZm of *Magnetospirillum gryphiswaldense* likely interacts with its generic homolog and is required for biomineralization under nitrate deprivation. *Journal of Bacteriology*, **196**, 650–659.

Murat, D., Quinlan, A., Vali, H., and Komeili, A. (2010a). Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5593–5598.

Murat, D., Byrne, M., and Komeili, A. (2010b). Cell biology of prokaryotic organelles. *Cold Spring Harbor Perspectives in Biology*, **2**, 1–18.

Murat, D., Falahati, V., Bertinetti, L., Csencsits, R., Körnig, A., Downing, K., Faivre, D., and Komeili, A. (2012). The magnetosome membrane protein, MmsF, is a major regulator of magnetite biomineralization in *Magnetospirillum magneticum* AMB-1. *Molecular Microbiology*, **85**, 684–699.

Nakazawa, H., Arakaki, A., Narita-Yamada, S., Yashiro, I., Jinno, K., Aoki, N., Tsuruyama, A., Okamura, Y., Tanikawa, S., Fujita, N., Takeyama, H., and Matsunaga, T. (2009). Whole genome sequence of *Desulfovibrio magneticus* strain RS-1 revealed common gene clusters in magnetotactic bacteria. *Genome Research*, **19**, 1801–1808.

Natale, P., Bruser, T., and Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-distinct translocases and mechanisms. *Biochimica et Biophysica Acta*, **1778**, 1735–1756.

Norris, V., den Blaauwen, T., Cabin-Flaman, A., Doi, R. H., Harshey, R., Janniere, L., Jimenez-Sanchez, A., Jin, D. J., Levin, P. A., Mileykovskaya, E., Minsky, A., Saier, M. Jr., and Skarstad, K. (2007). Functional taxonomy of bacterial hyperstructures. *Microbiology and Molecular Biology Reviews*, **71**, 230–253.

Okuda, Y., Denda, K., and Fukumori, Y. (1996). Cloning and sequencing of a gene encoding a new member of the tetratricopeptide protein family from magnetosomes of *Magnetospirillum magnetotacticum*. *Gene*, **171**, 99–102.

Ota, H., Takeyama, H., Nakayama, H., Katoh, T., and Matsunaga, T. (2003). SNP detection in transforming growth factor-β1 gene using bacterial magnetic particles. *Biosensors and Bioelectronics*, **18**, 683–687.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., and Warren, M.J. (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular Cell*, **38**, 305–315.

Paulsen, I. T., and Saier Jr., M. H. (1997). A novel family of ubiquitous heavy metal ion transport proteins. *The Journal of Membrane Biology*, **156**, 99–103.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**, 785–786.

Prinz, W. A, and Hinshaw, J. E. (2009). Membrane-bending proteins. *Critical Reviews in Biochemistry and Molecular Biology*, **44**, 278–291.

Quinlan, A., Murat, D., Vali, H., and Komeili, A. (2011). The HtrA/DegP family protease MamE is a bifunctional protein with roles in magnetosome protein localization and magnetite biomineralization. *Molecular Microbiology*, **80**, 1075–1087.

Raschdorf, O., Muller, F. D., Posfai, M., Plitzko, J. M., and Schuler, D. (2013). The magnetosome proteins MamX, MamZ and MamH are involved in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *Molecular Microbiology*, **89**, 872–886.

Raschdorf, O., Forstner, Y., Kolinko, I., Uebe, R., Plitzko, J. M., and Schüler, D. (2016). Genetic and ultrastructural analysis reveals the key players and initial steps of bacterial magnetosome membrane biogenesis. *PLoS Genetics*, **12**, e1006101.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1760–1765.

Rawlings, A. E., Bramble, J. P., Walker, R., Bain, J., Galloway, J. M., and Staniland, S. S. (2014). Self-assembled MmsF proteinosomes control magnetite nanoparticle formation in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 16094–16099.

Rice, J. J., Schohn, A., Bessette, P. H., Boulware, K. T., and Daugherty, P. S. (2006). Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands. *Protein Science*, **15**, 825–36.

Rodgers, F., Blakemore, R., Blakemore, N., Frankel, R., Bazylinski, D., Maratea, D., and Rodgers, C. (1990). Intercellular structure in a many-celled magnetotactic prokaryote. *Archives of Microbiology*, **154**, 18–22.

Sakaguchi, T., Burgess, J. G., and Matsunaga, T. (1993). Magnetite formation by a sulphate-reducing bacterium. *Nature*, **365**, 47–49.

Sakaguchi, T., Arakaki, A., and Matsunaga, T. (2002). *Desulfovibrio magneticus* sp. nov., a novel sulfate-reducing bacterium that produces intracellular single-domain-sized magnetite particles. *International Journal of Systematic and Evolutionary Microbiology,* **52**, 215–221.

Saier, M. H., and Bogdanov, M. V. (2013). Membranous organelles in bacteria. *Journal of Molecular Microbiology and Biotechnolog*y, **23**, 5–12.

Sakoh, M., Ito, K., and Akiyama, Y. (2005). Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *Journal of Biological Chemistry*, **280**, 33305–33310.

Scheffel, A., Gruska, M., Faivre, D., Linaroudis, A., Plitzko, J. M., and Schuller, D. (2006). An acidic protein aligns magnetosomes along a filamentous structure in magnetotactic bacteria. *Nature*, **440**, 110–114.

Scheffel, A., Gärdes, A., Grünberg, K., Wanner, G., and Schuler, D. (2008). The major magnetosome proteins MamGFDC are not essential for magnetite biomineralization in *Magnetospirillum gryphiswaldense* but regulate the size of magnetosome crystals. Journal of Bacteriology, **190**, 377–386.

Schleifer, K. H., Schüler, D., Spring, S., Weizenegger, M., Amann, R., Ludwig, W., and Kohler M. (1991) The genus *Magnetospirillum* gen. nov., description of *Magnetospirillum gryphiswaldense* sp. nov. and transfer of *Aquaspirillum magnetotacticum* to *Magnetospirillum magnetotacticum* comb. nov. *Systematic and Applied Microbiology*, **14**, 379–385.

Schubbe, S., Williams, T. J., Xie, G., Kiss, H. E., Brettin, T. S., Martinez, D., Ross, C. A., Schuler, D., Cox, B. L., Nealson, K. H., and Bazylinski, D. A. (2009). Complete genome sequence of the chemolithoautotrophic marine magnetotactic coccus strain MC-1. *Applied and Environmental Microbiology*, **75**, 4835–4852.

Schuler, D. (2008). Genetics and cell biology of magnetosome formation in magnetotactic bacteria. *FEMS Microbiology Reviews*, **32**, 654–672.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539.

Simmons, S. L., Sievert, S. M., Frankel, R. B., Bazylinski, D. A., and Edwards, K. J. (2004). Spatiotemporal distribution of marine magnetotactic bacteria in a seasonally stratified coastal salt pond spatiotemporal distribution of marine magnetotactic bacteria in a seasonally stratified coastal salt pond. *Applied and Environmental Microbiology*, **70**, 6230–6239.

Siponen, M. I., Adryanczyk, G., Ginet, N., Arnoux, P., and Pignol, D. (2012). Magnetochrome: a c-type cytochrome domain specific to magnetotatic bacteria. *Biochemical Society Transactions*, **40**, 1319–1323.

Siponen, M. I., Legrand, P., Widdrat, M., Jones, S. R., Zhang, W., Chang, M. C. Y., Faivre, D., Arnoux, P., and Pignol, D. (2013). Structural insight into magnetochrome-mediated magnetite biomineralization. *Nature*, **502**, 681–684.

Sirpiö, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N., and Aro, E. M. (2007). TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. *The Biochemical Journal*, **406**, 415–425.

Smalley, M. D., Marinov, G. K., Bertani, L. E., and DeSalvo, G. (2015). Genome sequence of *Magnetospirillum magnetotacticum* strain MS-1. *Genome Announcements*, **3**, e00233.

Sode, K., Kudo, S., Sakaguchi, T., Nakamura, N., and Matsunaga, T. (1993). Application of bacterial magnetic particles for highly selective mRNA recovery system. *Biotechnology Techniques*, **9**, 688-694.

Spitzer, J., and Poolman, B. (2013). How crowded is the prokaryotic cytoplasm? *FEBS Letters*, **587**, 2094–2098.

Staniland, S. S., and Rawlings, A. E. (2016). Crystallising the function of the magnetosome membrane mineralisation protein Mms6. *Biochemical Society Transactions*, **44**, 883–890.

Stewart, E. J. (2012). Growing unculturable bacteria. *Journal of Bacteriology*, **194**, 4151–4160.

Sun, J. B., Duan, J. H., Dai, S. L., Ren, J., Guo, L., Jiang, W., and Li, Y. (2008). Preparation and anti-tumor efficiency evaluation of doxorubicin-loaded bacterial magnetosomes: magnetic nanoparticles as drug carriers isolated from *Magnetospirillum gryphiswaldense*. *Biotechnology and Bioengineering*, **101**, 1313–1320.

Tanaka, T., Maruyama, K., Yoda, K., Nemoto, E., Udagawa, Y., Nakayama, H., Takeyama, H., and Matsunaga, T. (2003). Development and evaluation of an automated workstation for single nucleotide polymorphism discrimination using bacterial magnetic particles. *Biosensors and Bioelectronics*, **19**, 325–330.

Tanaka, M., Okamura, Y., Arakaki, A., Tanaka, T., Takeyama, H., and Matsunaga, T. (2006). Origin of magnetosome membrane: proteomic analysis of magnetosome membrane and comparison with cytoplasmic membrane. *Proteomics*, **6**, 5234–5247.

Tanaka, M., Arakaki, A., and Matsunaga, T. (2010). Identification and functional characterization of liposome tubulation protein from magnetotactic bacteria. *Molecular Microbiology*, **76**, 480–488.

Tang, Y. S., Wang, D., Zhou, C., Ma, W., Zhang, Y. Q., Liu, B., and Zhang, S. (2012). Bacterial magnetic particles as a novel and efficient gene vaccine delivery system. *Gene Therapy*, **19**, 1187–1195.

Tucker, J. D., Siebert, C. A., Escalante, M., Adams, P. G., Olsen, J. D., Otto, C., Stokes, D.L., and Hunter, C. N. (2010). Membrane invagination in *Rhodobacter sphaeroides* is initiated at curved regions of the cytoplasmic membrane, then forms both budded and fully detached spherical vesicles. *Molecular Microbiology*, **76**, 833–847.

Uebe, R., Junge, K., Henn, V., Poxleitner, G., Katzmann, E., Plitzko, J. M., Zarivach, R., Kasama, T., Wanner, G., Posfai, M., Bottger, T., Matzanke, B., and Schuler, D. (2011). The cation diffusion facilitator proteins MamB and MamM of *Magnetospirillum gryphiswaldense* have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Molecular Microbiology*, **82**, 818–835.

Ullrich, S., Kube, M., Schubbe, S., Reinhardt, R., and Schuler, D. (2005). A hypervariable 130-kilobase genomic region of *Magnetospirillum gryphiswaldense* comprises a magnetosome island which undergoes frequent rearrangements during stationary growth. *Journal of Bacteriology*, **187**, 7176–7184.

Ullrich, S., and Schuler, D. (2010). Cre-lox-based method for generation of large deletions within the genomic magnetosome island of *Magnetospirillum gryphiswaldense*. *Applied and Environmental Microbiology*, **76**, 2439–2444.

Wang, X., Wang, Q., Zhang, W., Wang, Y., Li, L., Wen, T., Zhang, T., Xu, J., Hu., J., Li, S., Liu., J., Jiang, W., Tain, J., Li, Y., Schuler, D., Wang, L., and Li, J. (2014). Complete Genome Sequence of *Magnetospirillum gryphiswaldense* MSR-1. *Genome Announcements*, **2**, 2–3.

Wegener, K. M., Bennewitz, S., Oelmüller, R., and Pakrasi, H. B. (2011). The Psb32 protein aids in repairing photodamaged photosystem II in the cyanobacterium *synechocystis* 6803. *Molecular Plant*, **4**, 1052–1061.

Wu, H.-Y., Liu, M.-S., Lin, T. P., and Cheng, Y. S. (2011). Structural and functional assays of AtTLP18.3 identify its novel acid phosphatase activity in thylakoid lumen. *Plant Physiology*, **157**, 1015–1025.

Yan, L., Zhang, S., Chen, P., Liu, H., Yin, H., and Li, H. (2012). Magnetotactic bacteria, magnetosomes and their application. *Microbiological Research*, **167**, 507–519.

Yang, J., Li, S., Huang, X., Li, J., Li, L., Pan, Y., and Li, Y. (2013). MamX encoded by the *mamXY* operon is involved in control of magnetosome maturation in *Magnetospirillum gryphiswaldense* MSR-1. *BMC Microbiology*, **13**, 203.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, **12**, 7–8.

Yoza, B., Arakaki, A., and Matsunaga, T. (2003). DNA extraction using bacterial magnetic particles modified with hyperbranched polyamidoamine dendrimer. *Journal of Biotechnology*, **101**, 219–228.

Zeytuni, N., Ozyamak, E., Ben-harush, K., Davidov, G., Levin, M., Gat, Y., Moyal, T., Brik, A., Komeili, A., and Zarivach, R. (2011). Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 480–487.

Zeytuni, N., and Zarivach, R. (2012). Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure*, **20**, 397–405.

Zhang, A. H. Z., Hackbarth, C. J., Chansky, K. M., and Chambers, H. F. (2001). A proteolytic transmembrane signalling pathway and resistance to β-lactams in staphylococci. *Science*, **291**, 1962–1965.

Supplementary Data

Supplementary Table 1. CAI of MAI genes from *M. gryphiswaldense* in *E. coli.*

Gene	Length (bp)	CAI
4070	1350	0.599
4071	1044	0.605
mmsF	375	0.634
mms6	411	0.600
4074	396	0.570
mamG	255	0.576
mamF	336	0.762
mamD	945	0.648
mamC	378	0.662
mamH	1287	0.633
mamI	264	0.672
mamE	2319	0.663
mamJ	1281	0.646
mamK	1044	0.706
mamL	372	0.656
mamM	957	0.707
mamN	1170	0.680
mamO	1899	0.653
mamP	813	0.664
mamA	654	0.700
mamQ	819	0.683
mamR	219	0.669
mamB	894	0.725
mamS	543	0.631
mamT	525	0.613
mamU	894	0.601
mamY	1116	0.682
mamX	765	0.688
mamZ	1932	0.702
ftsZ-like	936	0.680
lacZ (E. coli)	3075	0.748

Supplementary Table 2. I-TASSER functional, biological and localisation predictions.

Protein	Accuracy	Molecular Function	Biological Process	Cellular Location
MamG	0.30±0.10	Oxidoreductase (0.43)	Carboxylic acid biosynthesis (0.43)	Membrane (0.08)
		Fatty acid synthase (0.43)	Oxidation-reduction (0.33)	
		Binding (0.33)	Fatty acid metabolism (0.32)	
MamF	0.35±0.12	Electron transporter (0.15)	Biological adhesion (0.51)	Cytoskeleton (0.42)
		Oxidoreductase (0.15)		
		Heme binding (0.15)	Cellular physiology process (0.51)	
		Structural constituent of the cytoskeleton (0.14)		
		Transmembrane signalling receptor (0.12)		
MamD	0.39±0.13	Magnesium ion binding (0.35)	Regulation of cell shape (0.35)	Cytoplasm (0.35)
		N-acetyltransferase (0.35)	Peptidoglycan biosynthesis (0.35)	
		Diphosphorylase (0.35)	Cell wall organisation (0.35)	
			Lipopolysaccharide biosynthesis (0.35)	
MamC	0.43±0.14	Flavin adenine dinucleotide binding (0.71)	Aromatic compound metabolism (0.47)	Organelle membrane (0.41)
			Protein tetramerization (0.41)	
			Butyrate metabolism (0.41)	
		Acyl-CoA dehydrogenase (0.71)	Homo-oligomerization (0.41)	Mitochondrial envelope (0.41)
			Response to corticosteroids (0.41)	
			Short-chain fatty acid catabolism (0.41)	
		Fatty acid binding (0.41)	Nutrient response (0.41)	Mitochondrial matrix (0.36)
			Betaine biosynthesis (0.40)	
			Carboxylic acid biosynthesis (0.40)	
			Carnitine metabolism (0.40)	
4070	0.51±0.15	Protein binding (0.76)	Golgi vesicle transport (0.35)	COPI-coated vesicle (0.35)

			Protein recruitment (0.35)	Golgi-associated vesicle membrane (0.35)
				Vesicle coat (0.35)
4071	0.36±0.12	Protein binding (0.53)	Actin bundling (0.12)	Z disc (0.12)
				Cytoskeleton (0.12)
				Fascia adherens (0.12)
MmsF	0.29±0.09	Structural molecule (0.14)	Ion transport (0.15)	Transmembrane (0.15)
		Fibrinogen binding (0.07)		Actin cytoskeleton (0.14)
		Cation-selective channel (0.07)	Cell adhesion (0.14)	Extracellular (0.07)
				Postsynaptic membrane (0.07)
Mms6	0.39±0.13	Beta-lactamase (0.50)	Response to antibiotic (0.50)	Periplasm (0.31)
		Zinc ion binding (0.41)	Antibiotic catabolism (0.41)	
4074	0.26±0.08	Calcium ion binding (0.13)	DNA catabolism (0.07)	Peripheral membrane protein (0.07)
		Deoxyribonuclease I (0.07)	Apoptosis (0.07)	Photosystem II oxygen evolving complex (0.07)
		Metalloexopeptidase (0.07)	Photosynthesis (0.07)	Extracellular (0.07)
		Actin binding (0.07)	Acetyl-CoA metabolism (0.07)	Nuclear envelope (0.07)
		Dipeptidyl-peptidase (0.07)	Proteolysis (0.07)	
		Transferase (0.07)		
		Dipeptidase (0.07)		
MamH	0.61±0.14	Transferase (0.54)	Transmembrane transport (0.67)	Transmembrane (0.83)
		Galactose transmembrane transporter (0.51)	Valine metabolism (0.54)	
		Solute:proton symporter (0.51)	Isoleucine metabolism (0.54)	Inner membrane (0.76)
		Cation:sugar symporter (0.51)	Drug transport (0.53)	
		Arabinose transmembrane transporter (0.51)	Hexose transport (0.51)	Macromolecular complex (0.54)
		Fucose transmembrane transporter (0.51)	Pentose transport (0.51)	
		Glycerol-3-phosphate transmembrane transporter (0.39)	Hexose metabolism (0.51)	Cytoplasm (0.54)

		Peptide transporter (0.30)	Glycerol metabolism (0.39)	
			Glycerol-3-phosphate transport (0.39)	
			Oligopeptide transport (0.30)	
MamI	0.34±0.11	Structural constituent of ribosome (0.20)	Cellular macromolecule biosynthesis (0.57)	Macromolecular complex (0.39)
		rRNA binding (0.11)	Gene expression (0.57)	Intracellular non-membrane-bounded organelle (0.39)
		Translation termination factor (0.11)	Cellular protein metabolism (0.50)	Cytoplasm (0.39)
MamE	0.44±0.14	Serine-type endopeptidase (0.63)	Proteolysis involved in protein catabolism (0.51)	Cell envelope (0.51)
		Protein binding (0.51)	Response to abiotic stimulus (0.51)	External encapsulating structure (0.51)
			Response to chemical (0.51)	
MamJ	0.51±0.15	Enzyme inhibitor (0.48)	Oxidation-reduction (0.10)	Cytoplasm (0.36)
		Protein binding (0.42)	Fatty acid biosynthesis (0.10)	
		Endopeptidase regulator (0.32)	Neurotransmitter secretion (0.09)	
		Ion binding (0.32)	Inhibition of neurotransmitter uptake (0.09)	Extracellular (0.31)
			Proteolysis (0.09)	
			Pathogenesis (0.09)	
MamK	0.77±0.10	ATP binding (0.98)	Adherens junction organization (0.52)	I band (0.52)
		Protein kinase binding (0.52)	Skeletal muscle thin filament assembly (0.52)	NuA4 histone acetyltransferase complex (0.52)
		Myosin binding (0.52)	Sperm individualization (0.52)	Cytosol (0.52)
		Nitric-oxide synthase binding (0.52)	Nematode larval development (0.52)	Ino80 complex (0.52)
		Identical protein binding (0.52)	Muscle filament sliding (0.52)	Microtubule associated complex (0.52)
		ADP binding (0.52)	Mushroom body development (0.52)	MLL5-L complex (0.52)

		ATPase activity (0.52)	Apoptotic process (0.52)	Striated muscle thin filament (0.52)
		Structural constituent of cytoskeleton (0.52)	Response to lithium ion (0.52)	Intracellular ribonucleoprotein complex (0.52)
		Kinesin binding (0.52)	Chaperone mediated protein folding (0.52)	Stress fiber (0.52)
			Response to mechanical stimulus (0.52)	Actomyosin, actin portion (0.52)
MamL	0.28±0.09	Sphingomyelin phosphodiesterase (0.07)	Cytolysis (0.07)	Extracellular (0.07)
		Triglyceride lipase (0.07)	Pathogenesis (0.07)	
		4-hydroxy-tetrahydrodipicolinate synthase (0.06)	Hemolysis in other organism (0.07)	Cytoplasm (0.06)
			Lipid metabolism (0.07)	
			Diaminopimelate biosynthesis (0.06)	
MamM	0.72±0.11	Wide pore channel activity (0.57)	Multi-organism process (0.58)	Inner membrane (0.72)
			Response to organic substance (0.57)	
		Ferrous iron transmembrane transporter (0.44)	Siderophore transport (0.57)	Transmembrane (0.72)
			Establishment of protein localization (0.57)	
		Zinc efflux active transmembrane transporter (0.44)	Zinc II ion transport (0.44)	Outer membrane (0.50)
			Cadmium ion transport (0.44)	
		Cadmium ion transmembrane transporter (0.44)	Cellular zinc ion homeostasis (0.44)	Intracellular (0.40)
			Cellular cadmium ion homeostasis (0.44)	
		RNA polymerase activity (0.40)	Ferrous iron transport (0.44)	
			Cellular macromolecule biosynthesis (0.40)	
MamN	0.52±0.15	Intramolecular oxidoreductase (0.48)	Generation of precursor metabolites and energy (0.48)	Cell periphery (0.48)
		Oxidoreductase (0.39)	Hexose biosynthesis (0.48)	Cytoplasm (0.35)

		Metal cluster binding (0.39)	Glucose catabolism (0.48)	
		Ion binding (0.39)	Glutamine family amino acid biosynthesis (0.39)	
			Glutamate metabolism (0.39)	
MamO	0.43±0.14	Serine-type endopeptidase (0.55)	Cellular response to topologically incorrect protein (0.60)	Intrinsic component of membrane (0.60)
			Response to misfolded protein (0.60)	Obsolete membrane fraction (0.60)
		Protein binding (0.48)	Proteolysis (0.55)	Plasma membrane (0.60)
				Periplasm (0.40)
MamP	0.65±0.13	Protein binding (0.75)	Cellular response to topologically incorrect protein (0.55)	Obsolete membrane fraction (0.55)
			Proteolysis involved in cellular protein catabolic process (0.55)	Intrinsic component of membrane (0.55)
			Response to misfolded protein (0.53)	Plasma membrane (0.55)
		Serine-type endopeptidase (0.47)	Response to abiotic stimulus (0.53)	Cell envelope (0.53)
				External encapsulating structure (0.53)
				Periplasm (0.47)
MamA	0.71±0.12	Identical protein binding (0.41)	Copper-induced intracellular protein transport (0.39)	Mitochondrial outer membrane (0.39)
		P-P-bond-hydrolysis-driven protein transmembrane transporter activity (0.39)		
MamQ	0.35±0.12	acyl-CoA dehydrogenase (0.37)	Fatty acid catabolic process (0.35)	Mitochondrial inner membrane (0.09)
			Fatty acid oxidation (0.35)	Mitochondrial nucleoid (0.09)
			Aromatic compound catabolism (0.34)	Peroxisome (0.09)
			Xenobiotic catabolism (0.34)	Cytosol (0.09)
MamR	0.57±0.15	DNA binding (0.64)	Negative regulation of transcription (0.51)	Cytoplasm (0.74)

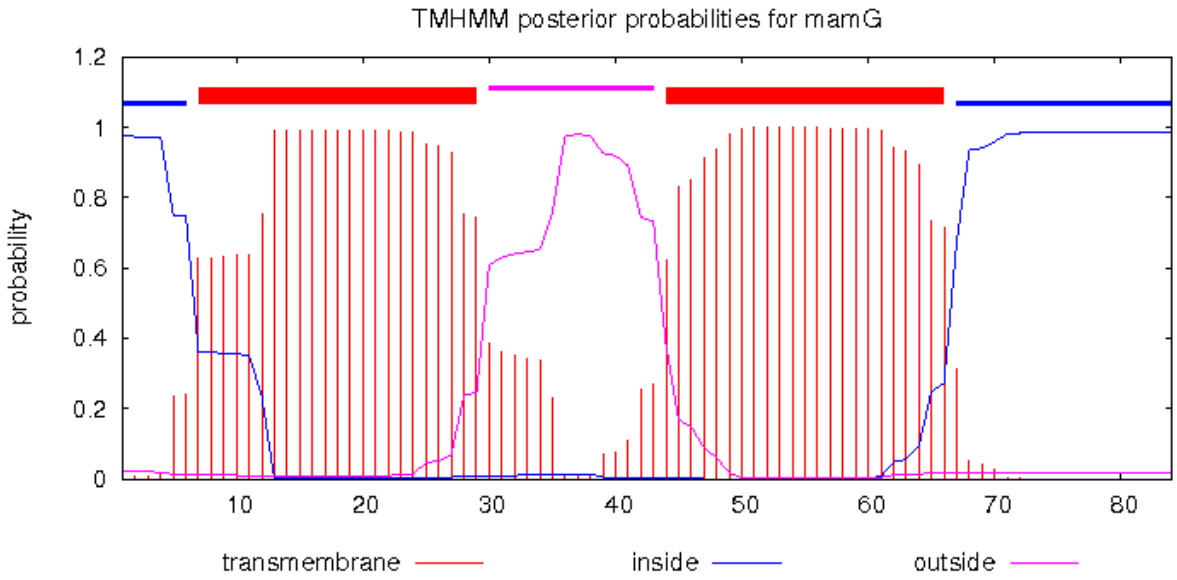
		Obsolete transcription repressor activity (0.51)	Response to redox state (0.51)	
		Translation elongation factor (0.46)	Seleno-cysteine incorporation (0.46)	
		GTP binding (0.46)		
MamB	0.99±0.04	Ferrous iron transmembrane transporter (0.50)	Ferrous iron transport (0.50)	Integral component of membrane (0.92)
			Zinc(II) ion transport (0.50)	
		Zinc efflux permease (0.50)	Cadmium ion transport (0.50)	
			Cellular zinc ion homeostasis (0.50)	Plasma membrane (0.83)
		Cadmium ion transmembrane transporter (0.50)	Cellular cadmium ion homeostasis (0.50)	
			Response to antibiotic (0.47)	
		Lipid binding (0.47)	Protein transport (0.37)	Cell outer membrane (0.67)
			Response to organic cyclic compound (0.37)	
		Porin (0.37)	Enterobactin transport (0.37)	
MamS	0.29±0.09	Nucleic acid binding transcription factor (0.40)	Regulation of gene expression (0.40)	Intracellular (0.40)
		ATP binding (0.36)	Inorganic anion transport (0.40)	
		Molybdenum ion binding (0.32)	regulation of cellular macromolecule biosynthesis (0.40)	ATP-binding cassette (ABC) transporter complex (0.36)
		Molybdate transmembrane-transporting ATPase (0.32)	Regulation of RNA metabolism (0.40)	
MamT	0.31±0.10	DNA binding (0.15)	Oxidation-reduction (0.13)	Cytoplasm (0.09)
		Hydro-lyase (0.09)		
		Cyanate hydratase (0.09)	Cyanate catabolism (0.09)	
		Electron carrier (0.07)		
		Heme binding (0.07)	DNA repair (0.07)	
		Lactaldehyde dehydrogenase (0.07)		
MamU	0.90±0.06	Diacylglycerol kinase (0.99)	Protein kinase C-activating G-protein coupled receptor signalling pathway (0.99)	Cytoplasm (0.88)
		ATP binding (0.95)	Phospholipid biosynthesis (0.95)	

		Metal ion binding (0.95)	Lipid phosphorylation (0.88)	
		NAD+ kinase (0.94)		
		Lipid kinase (0.88)		
MamX	0.28±0.09	Ion binding (0.60)	tRNA metabolic process (0.37)	Intracellular (0.47)
		Adenyl ribonucleotide binding (0.37)	Cellular macromolecule biosynthesis (0.37)	
		Purine ribonucleoside triphosphate binding (0.37)	Gene expression (0.37)	
		ligase activity, forming carbon-oxygen bonds (0.37)	Amino acid activation (0.37)	
			Cellular protein metabolism (0.33)	
MamY	0.97±0.05	Nucleic acid binding transcription factor (0.35)	Regulation of gene expression (0.35)	Host cell plasma membrane (0.32)
			Regulation of cellular macromolecule biosynthesis (0.35)	
			Regulation of RNA metabolism (0.35)	
			Cell communication (0.32)	
		Molecular transducer (0.32)	Cellular response to stimulus (0.32)	Integral component of membrane (0.32)
			Signalling (0.32)	
			Hemolysis in other organism (0.32)	
			Cytolysis (0.32)	
MamZ	0.52±0.15	Hexose:proton symporter (0.57)	Hexose transport (0.57)	Integral component of membrane (0.51)
		Fucose transmembrane transporter (0.57)	Pentose transport (0.57)	
		Galactose transmembrane transporter (0.57)	Hexose metabolism (0.57)	Peroxisome (0.41)
		Arabinose transmembrane transporter (0.57)	Nucleotide catabolism (0.46)	
		2 iron, 2 sulfur cluster binding (0.41)	Purine nucleotide metabolism (0.46)	Extracellular (0.41)
		Iron ion binding (0.41)	Bone remodelling (0.46)	
		Flavin adenine dinucleotide binding (0.41)	Response to metal ion (0.46)	Plasma membrane (0.41)
		Electron carrier (0.41)	Tissue homeostasis (0.46)	

		Protein homodimerization (0.41)	Regulation of epithelial cell differentiation (0.41)	Macromolecular complex (0.34)
		Xanthine dehydrogenase (0.41)	Xanthine catabolism (0.41)	
FtsZ-like	0.64±0.13	GTPase (1.00)	GTP hydrolysis (1.00)	Cytoplasm (1.00)
		GTP binding (1.00)	Protein polymerization (1.00)	
		Identical protein binding (0.79)	Barrier septum assembly (1.00)	Plasma membrane (0.77)
		Magnesium ion binding (0.77)	Cell cycle (1.00)	
			Growth (0.77)	Microtubule (0.75)

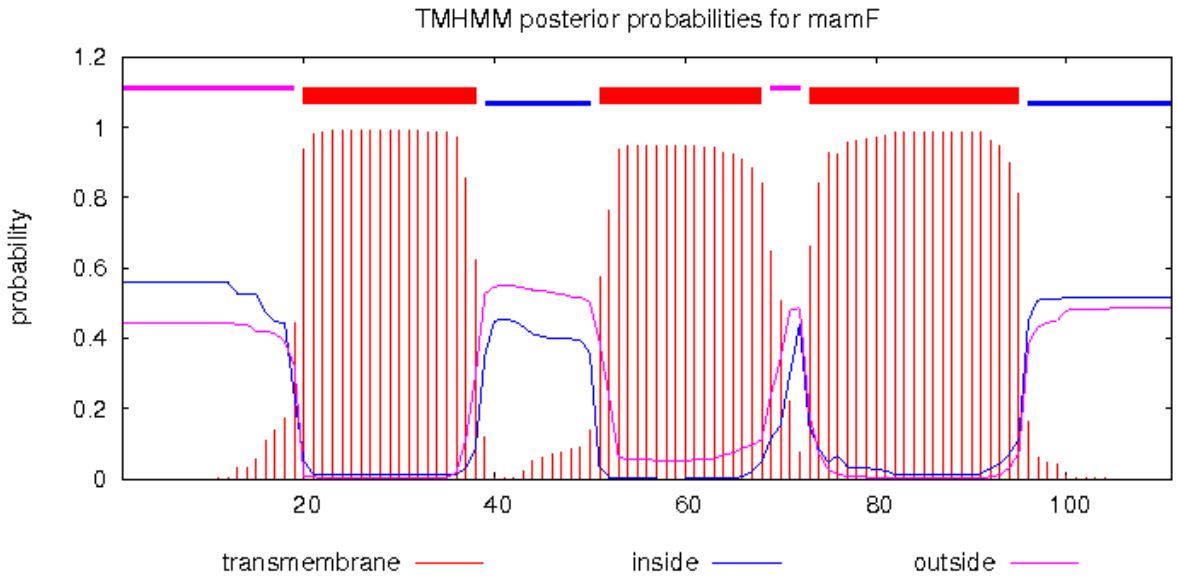
Supplementary Figure 1. TMHMM2 predictions for all studied MAI proteins.

```
# mamG Length: 84
# mamG Number of predicted TMHs:  2
# mamG Exp number of AAs in TMHs: 45.19829
# mamG Exp number, first 60 AAs:  39.56139
# mamG Total prob of N-in:        0.97876
# mamG POSSIBLE N-term signal sequence
mamG    TMHMM2.0        inside       1     6
mamG    TMHMM2.0        TMhelix      7    29
mamG    TMHMM2.0        outside     30    43
mamG    TMHMM2.0        TMhelix     44    66
mamG    TMHMM2.0        inside      67    84
```



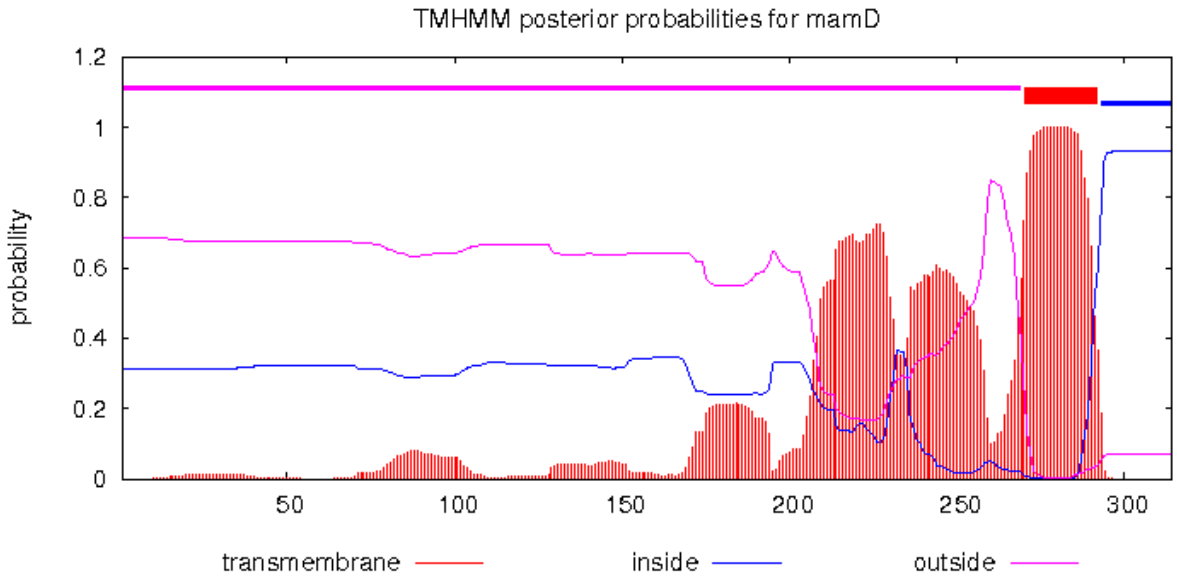
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamF Length: 111
# mamF Number of predicted TMHs:  3
# mamF Exp number of AAs in TMHs: 59.55496
# mamF Exp number, first 60 AAs:  28.7993
# mamF Total prob of N-in:        0.55862
# mamF POSSIBLE N-term signal sequence
mamF    TMHMM2.0        outside      1    19
mamF    TMHMM2.0        TMhelix     20    38
mamF    TMHMM2.0        inside      39    50
mamF    TMHMM2.0        TMhelix     51    68
mamF    TMHMM2.0        outside     69    72
mamF    TMHMM2.0        TMhelix     73    95
mamF    TMHMM2.0        inside      96   111
```



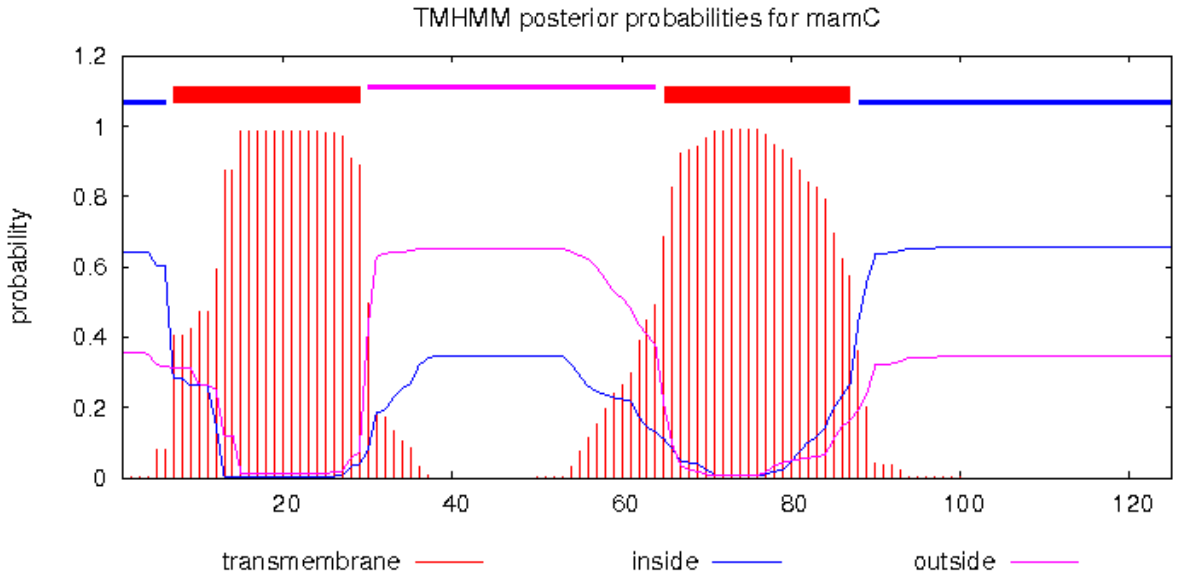
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamD Length: 314
# mamD Number of predicted TMHs:  1
# mamD Exp number of AAs in TMHs: 61.6504
# mamD Exp number, first 60 AAs:  0.31266
# mamD Total prob of N-in:        0.31374
mamD    TMHMM2.0        outside      1   269
mamD    TMHMM2.0        TMhelix    270   292
mamD    TMHMM2.0        inside     293   314
```



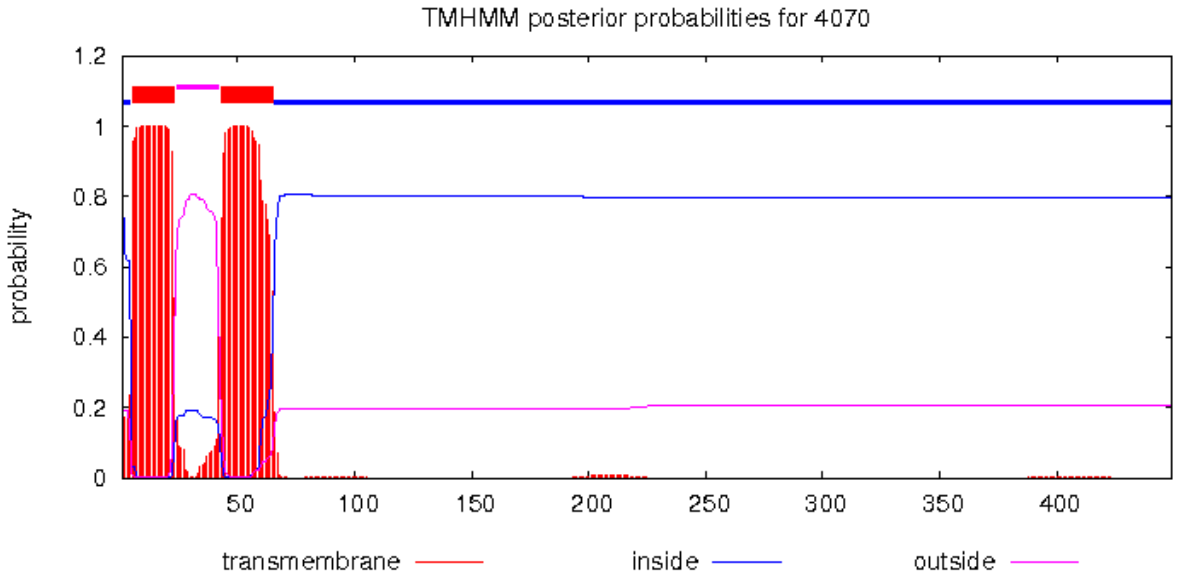
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamC Length: 125
# mamC Number of predicted TMHs:  2
# mamC Exp number of AAs in TMHs: 44.10464
# mamC Exp number, first 60 AAs:  21.57353
# mamC Total prob of N-in:        0.64356
# mamC POSSIBLE N-term signal sequence
mamC    TMHMM2.0        inside       1     6
mamC    TMHMM2.0        TMhelix      7    29
mamC    TMHMM2.0        outside     30    64
mamC    TMHMM2.0        TMhelix     65    87
mamC    TMHMM2.0        inside      88   125
```



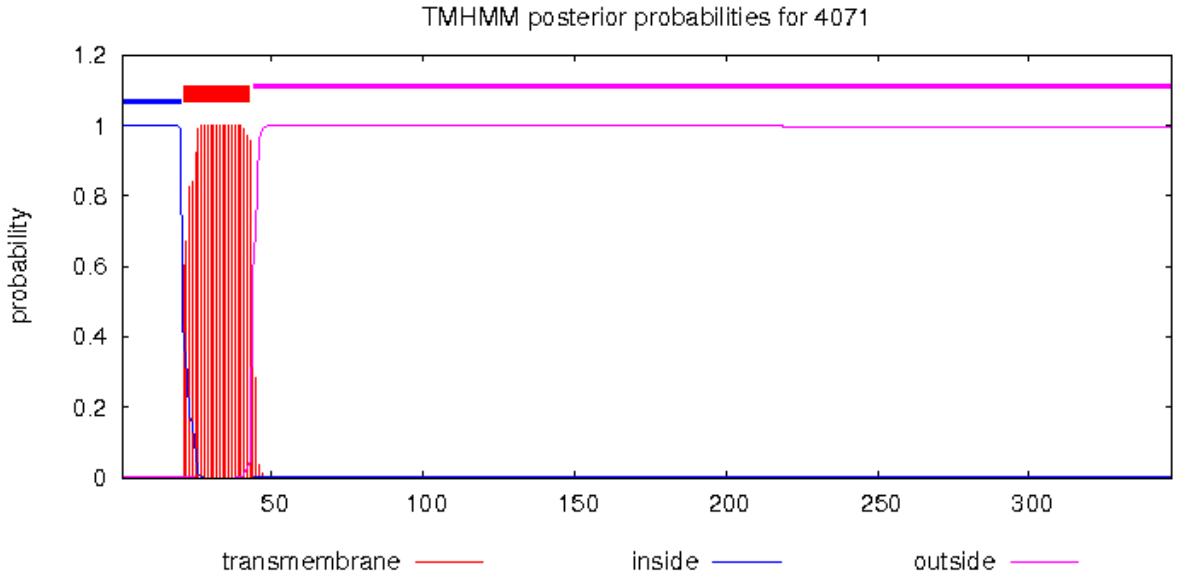
plot in postscript, script for making the plot in gnuplot, data for plot

```
# 4070 Length: 449
# 4070 Number of predicted TMHs:  2
# 4070 Exp number of AAs in TMHs: 41.6469899999999999
# 4070 Exp number, first 60 AAs:  37.64613
# 4070 Total prob of N-in:        0.80623
# 4070 POSSIBLE N-term signal sequence
4070    TMHMM2.0        inside       1     4
4070    TMHMM2.0        TMhelix      5    23
4070    TMHMM2.0        outside     24    42
4070    TMHMM2.0        TMhelix     43    65
4070    TMHMM2.0        inside      66   449
```



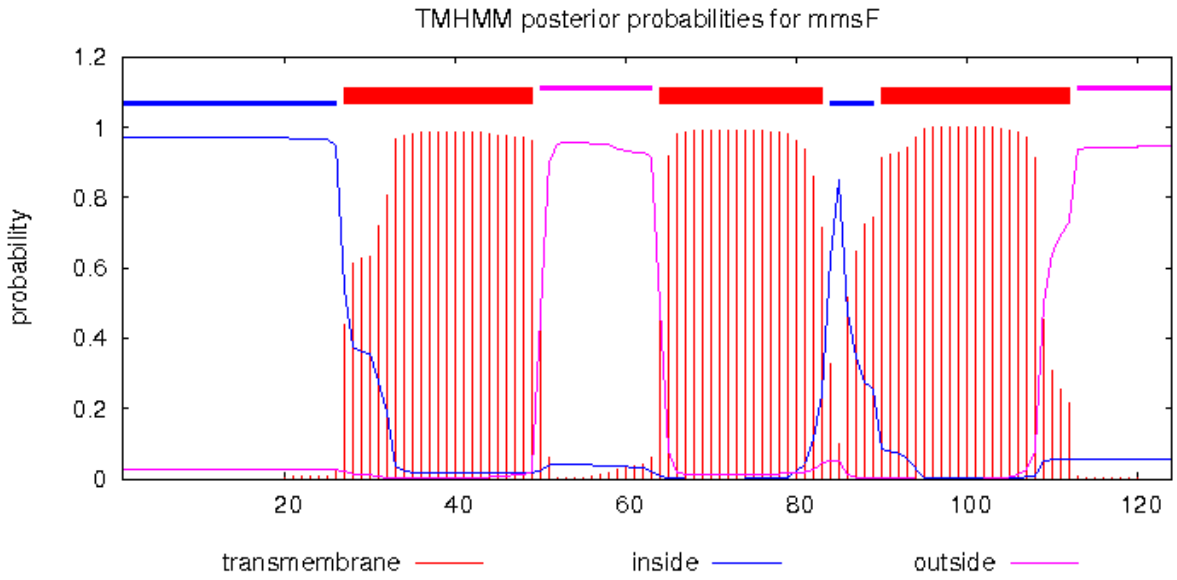
plot in postscript, script for making the plot in gnuplot, data for plot

```
# 4071 Length: 347
# 4071 Number of predicted TMHs:  1
# 4071 Exp number of AAs in TMHs: 22.55247
# 4071 Exp number, first 60 AAs:  22.52011
# 4071 Total prob of N-in:        0.99950
# 4071 POSSIBLE N-term signal sequence
4071    TMHMM2.0       inside       1    20
4071    TMHMM2.0       TMhelix     21    43
4071    TMHMM2.0       outside     44   347
```



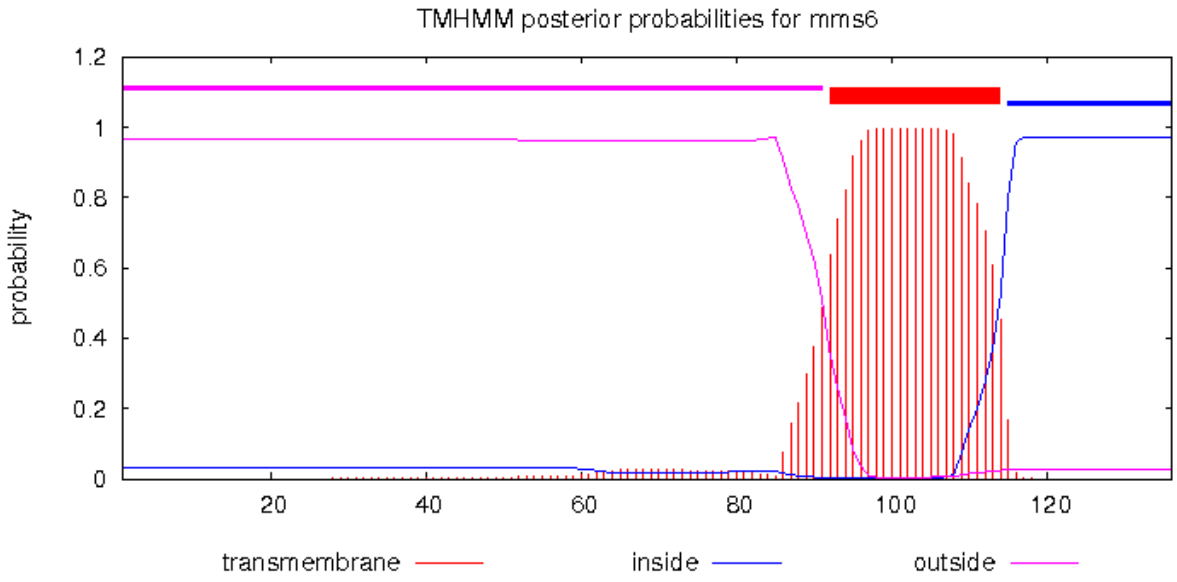
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mmsF Length: 124
# mmsF Number of predicted TMHs:  3
# mmsF Exp number of AAs in TMHs: 62.88052
# mmsF Exp number, first 60 AAs:  21.20792
# mmsF Total prob of N-in:        0.97207
# mmsF POSSIBLE N-term signal sequence
mmsF    TMHMM2.0        inside       1    26
mmsF    TMHMM2.0        TMhelix     27    49
mmsF    TMHMM2.0        outside     50    63
mmsF    TMHMM2.0        TMhelix     64    83
mmsF    TMHMM2.0        inside      84    89
mmsF    TMHMM2.0        TMhelix     90   112
mmsF    TMHMM2.0        outside    113   124
```



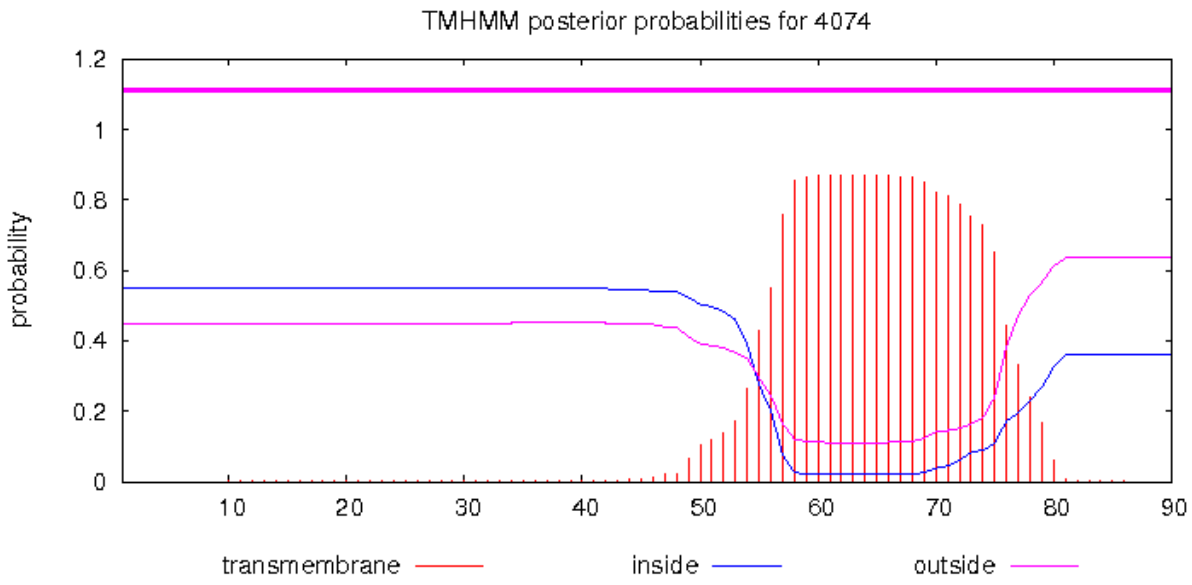
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mms6 Length: 136
# mms6 Number of predicted TMHs:  1
# mms6 Exp number of AAs in TMHs: 22.68297
# mms6 Exp number, first 60 AAs:  0.08179
# mms6 Total prob of N-in:        0.03311
mms6    TMHMM2.0        outside      1    91
mms6    TMHMM2.0        TMhelix     92   114
mms6    TMHMM2.0        inside     115   136
```



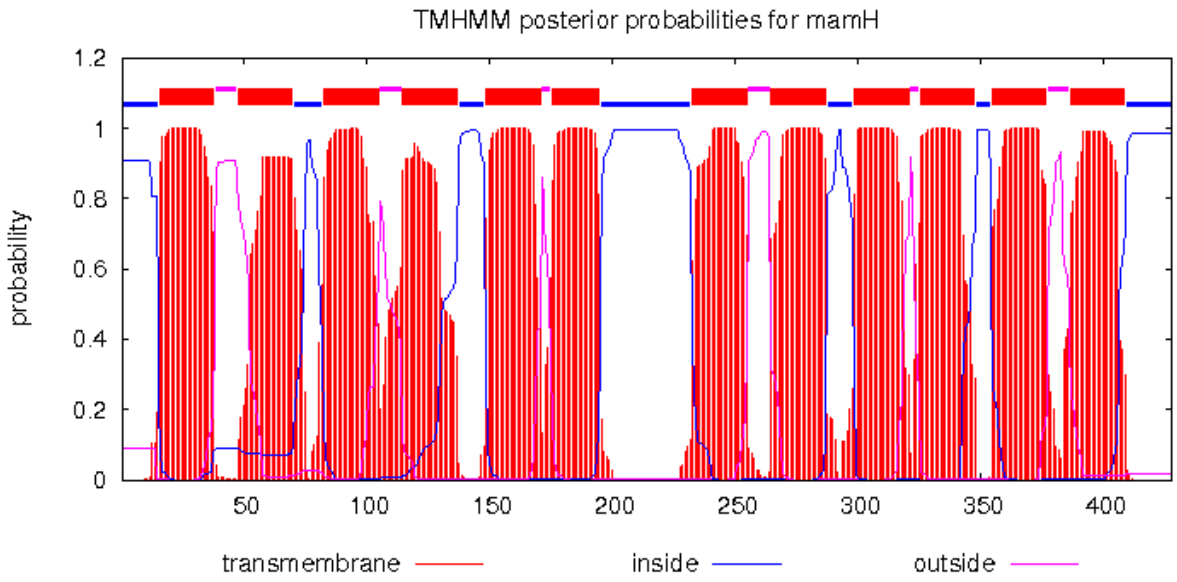
plot in postscript, script for making the plot in gnuplot, data for plot

```
# 4074 Length: 90
# 4074 Number of predicted TMHs:  0
# 4074 Exp number of AAs in TMHs: 18.92606
# 4074 Exp number, first 60 AAs:  5.31624
# 4074 Total prob of N-in:        0.55049
4074    TMHMM2.0        outside      1    90
```



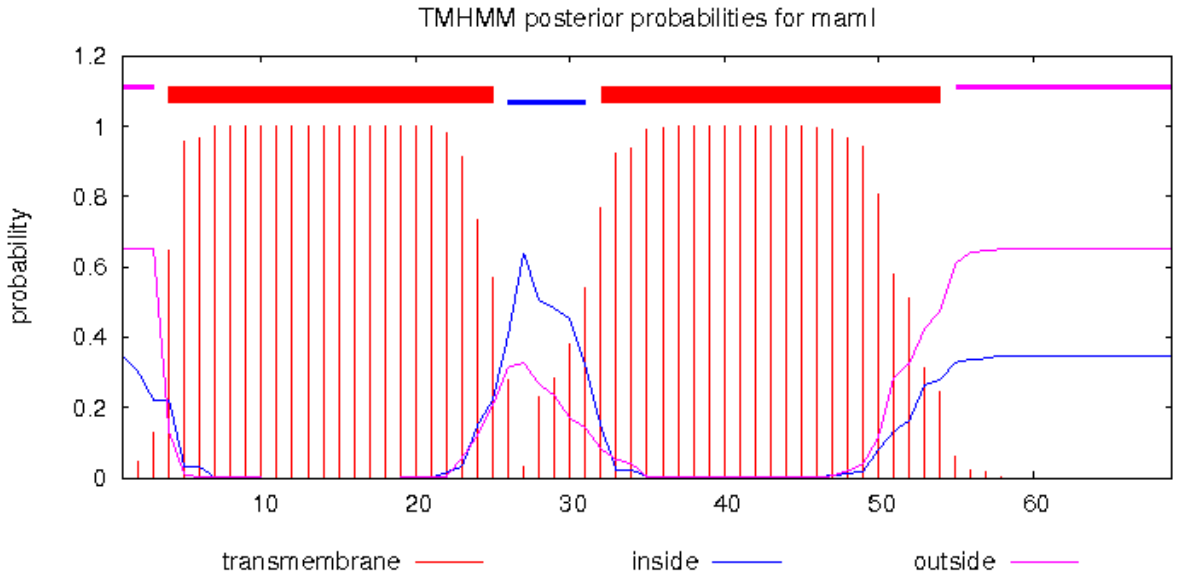
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamH Length: 428
# mamH Number of predicted TMHs:  12
# mamH Exp number of AAs in TMHs: 259.68139
# mamH Exp number, first 60 AAs:  29.99955
# mamH Total prob of N-in:        0.91057
# mamH POSSIBLE N-term signal sequence
mamH    TMHMM2.0        inside       1    15
mamH    TMHMM2.0        TMhelix     16    38
mamH    TMHMM2.0        outside     39    47
mamH    TMHMM2.0        TMhelix     48    70
mamH    TMHMM2.0        inside      71    82
mamH    TMHMM2.0        TMhelix     83   105
mamH    TMHMM2.0        outside    106   114
mamH    TMHMM2.0        TMhelix    115   137
mamH    TMHMM2.0        inside     138   148
mamH    TMHMM2.0        TMhelix    149   171
mamH    TMHMM2.0        outside    172   175
mamH    TMHMM2.0        TMhelix    176   195
mamH    TMHMM2.0        inside     196   232
mamH    TMHMM2.0        TMhelix    233   255
mamH    TMHMM2.0        outside    256   264
mamH    TMHMM2.0        TMhelix    265   287
mamH    TMHMM2.0        inside     288   298
mamH    TMHMM2.0        TMhelix    299   321
mamH    TMHMM2.0        outside    322   325
mamH    TMHMM2.0        TMhelix    326   348
mamH    TMHMM2.0        inside     349   354
mamH    TMHMM2.0        TMhelix    355   377
mamH    TMHMM2.0        outside    378   386
mamH    TMHMM2.0        TMhelix    387   409
mamH    TMHMM2.0        inside     410   428
```



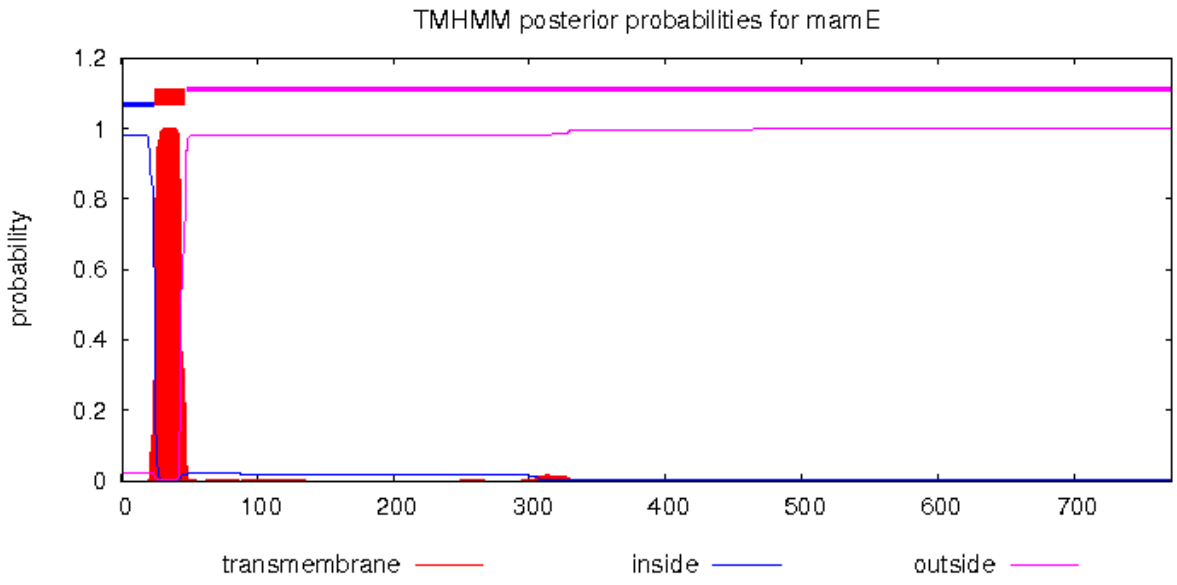
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamI Length: 69
# mamI Number of predicted TMHs:  2
# mamI Exp number of AAs in TMHs: 42.73344
# mamI Exp number, first 60 AAs:  42.73344
# mamI Total prob of N-in:        0.34850
# mamI POSSIBLE N-term signal sequence
mamI    TMHMM2.0        outside      1     3
mamI    TMHMM2.0        TMhelix      4    25
mamI    TMHMM2.0        inside      26    31
mamI    TMHMM2.0        TMhelix     32    54
mamI    TMHMM2.0        outside     55    69
```



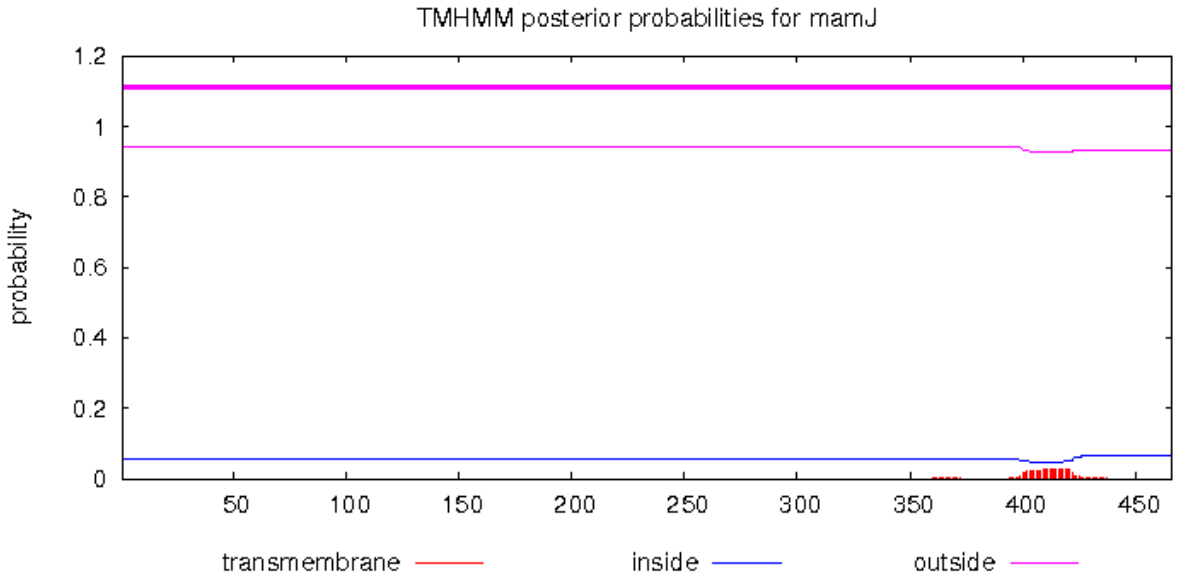
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamE Length: 772
# mamE Number of predicted TMHs:  1
# mamE Exp number of AAs in TMHs: 21.10387
# mamE Exp number, first 60 AAs:  20.58206
# mamE Total prob of N-in:        0.97866
# mamE POSSIBLE N-term signal sequence
mamE    TMHMM2.0       inside       1    24
mamE    TMHMM2.0       TMhelix     25    47
mamE    TMHMM2.0       outside     48   772
```



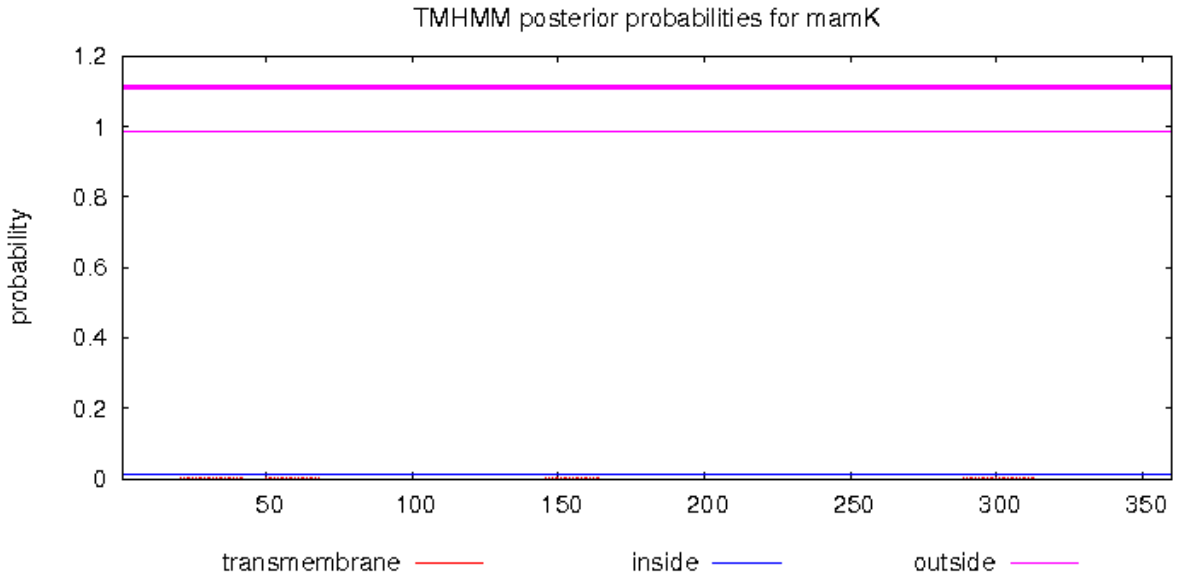
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamJ Length: 466
# mamJ Number of predicted TMHs:  0
# mamJ Exp number of AAs in TMHs: 0.60206
# mamJ Exp number, first 60 AAs:  0.00022
# mamJ Total prob of N-in:        0.05618
mamJ    TMHMM2.0        outside      1   466
```



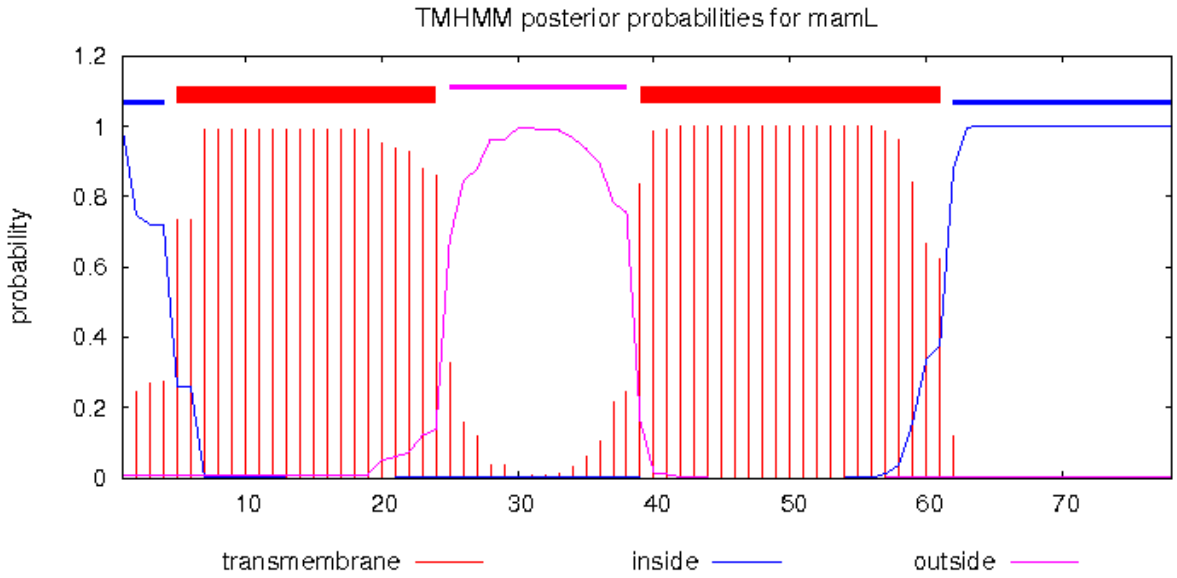
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamK Length: 360
# mamK Number of predicted TMHs:  0
# mamK Exp number of AAs in TMHs: 0.05418
# mamK Exp number, first 60 AAs:  0.02043
# mamK Total prob of N-in:        0.01331
mamK    TMHMM2.0        outside      1   360
```



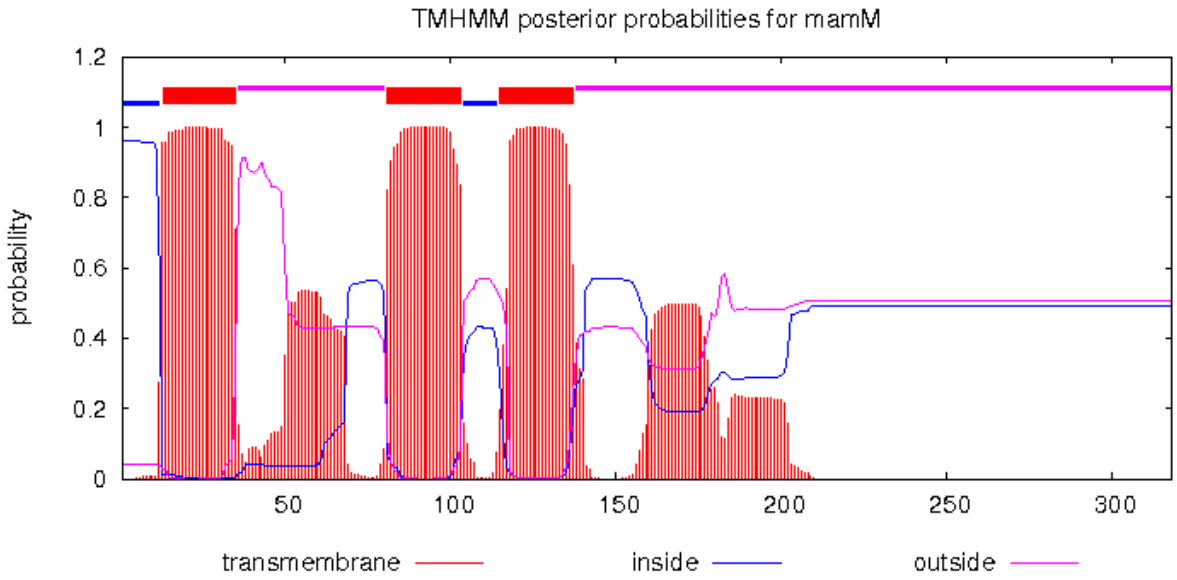
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamL Length: 78
# mamL Number of predicted TMHs:  2
# mamL Exp number of AAs in TMHs: 43.06534
# mamL Exp number, first 60 AAs:  42.32445
# mamL Total prob of N-in:        0.99081
# mamL POSSIBLE N-term signal sequence
mamL    TMHMM2.0        inside       1     4
mamL    TMHMM2.0        TMhelix      5    24
mamL    TMHMM2.0        outside     25    38
mamL    TMHMM2.0        TMhelix     39    61
mamL    TMHMM2.0        inside      62    78
```



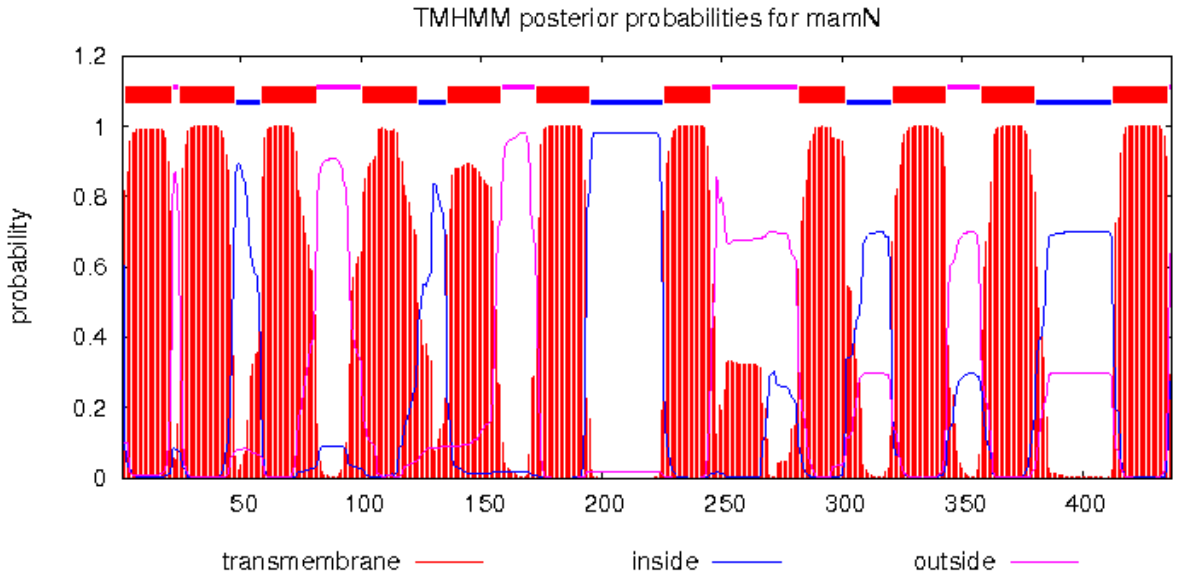
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamM Length: 318
# mamM Number of predicted TMHs:  3
# mamM Exp number of AAs in TMHs: 92.33744
# mamM Exp number, first 60 AAs:  29.70697
# mamM Total prob of N-in:        0.96067
# mamM POSSIBLE N-term signal sequence
mamM    TMHMM2.0        inside       1    12
mamM    TMHMM2.0        TMhelix     13    35
mamM    TMHMM2.0        outside     36    80
mamM    TMHMM2.0        TMhelix     81   103
mamM    TMHMM2.0        inside     104   114
mamM    TMHMM2.0        TMhelix    115   137
mamM    TMHMM2.0        outside    138   318
```



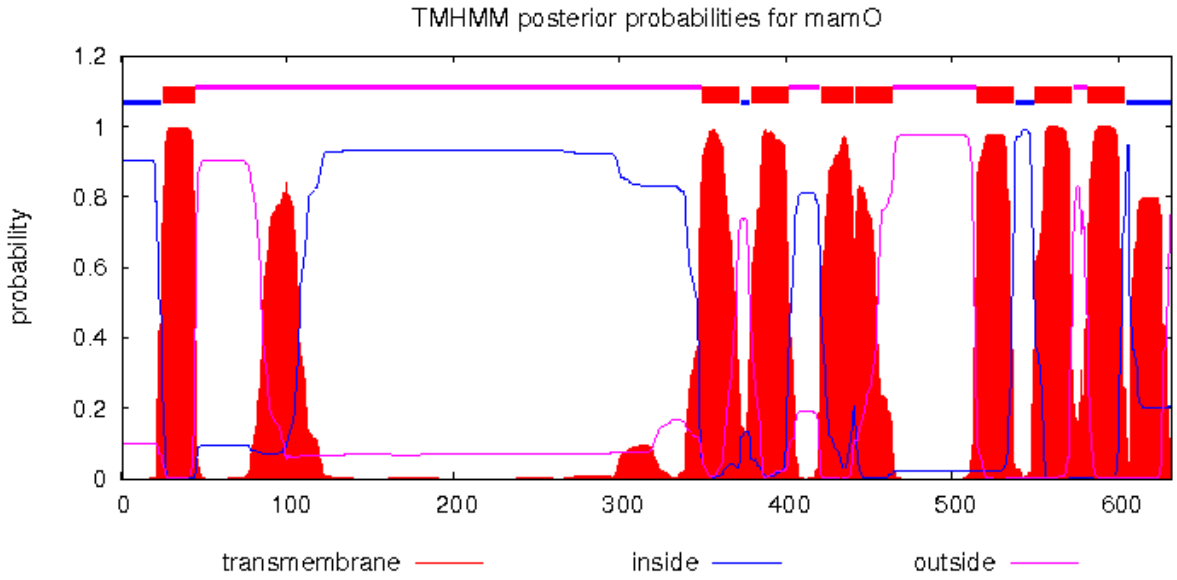
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamN Length: 437
# mamN Number of predicted TMHs:  11
# mamN Exp number of AAs in TMHs: 246.35642
# mamN Exp number, first 60 AAs:  44.44556
# mamN Total prob of N-in:        0.89903
# mamN POSSIBLE N-term signal sequence
mamN    TMHMM2.0        inside       1     1
mamN    TMHMM2.0        TMhelix      2    21
mamN    TMHMM2.0        outside     22    24
mamN    TMHMM2.0        TMhelix     25    47
mamN    TMHMM2.0        inside      48    58
mamN    TMHMM2.0        TMhelix     59    81
mamN    TMHMM2.0        outside     82   100
mamN    TMHMM2.0        TMhelix    101   123
mamN    TMHMM2.0        inside     124   135
mamN    TMHMM2.0        TMhelix    136   158
mamN    TMHMM2.0        outside    159   172
mamN    TMHMM2.0        TMhelix    173   195
mamN    TMHMM2.0        inside     196   225
mamN    TMHMM2.0        TMhelix    226   245
mamN    TMHMM2.0        outside    246   281
mamN    TMHMM2.0        TMhelix    282   301
mamN    TMHMM2.0        inside     302   320
mamN    TMHMM2.0        TMhelix    321   343
mamN    TMHMM2.0        outside    344   357
mamN    TMHMM2.0        TMhelix    358   380
mamN    TMHMM2.0        inside     381   412
mamN    TMHMM2.0        TMhelix    413   435
mamN    TMHMM2.0        outside    436   437
```



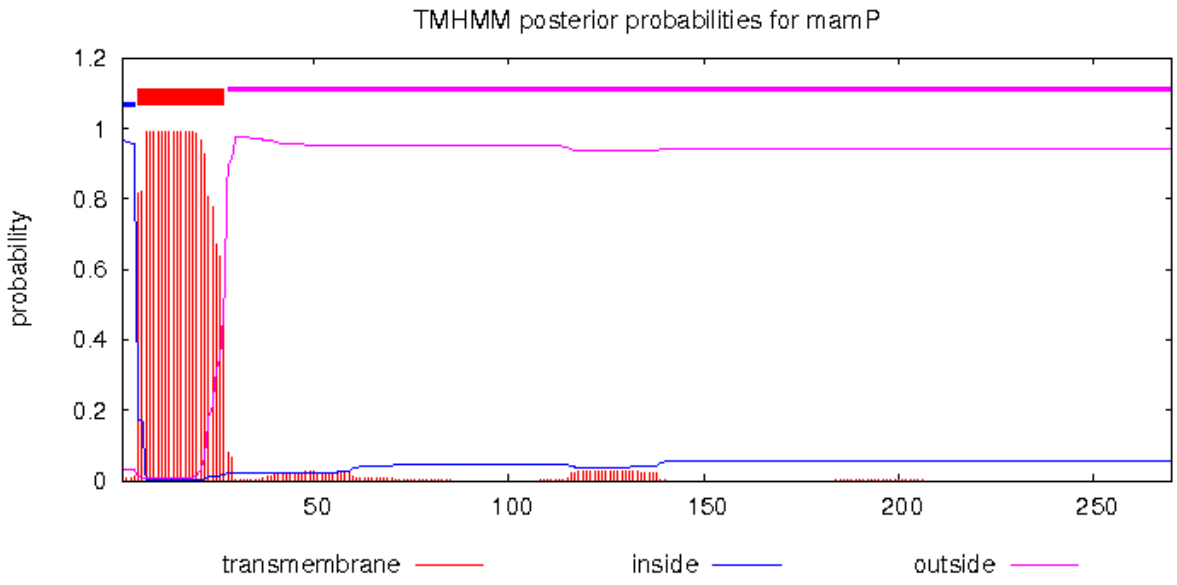
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamO Length: 632
# mamO Number of predicted TMHs:  8
# mamO Exp number of AAs in TMHs: 201.56296
# mamO Exp number, first 60 AAs:  20.92075
# mamO Total prob of N-in:        0.90200
# mamO POSSIBLE N-term signal sequence
mamO    TMHMM2.0        inside       1    24
mamO    TMHMM2.0        TMhelix     25    44
mamO    TMHMM2.0        outside     45   349
mamO    TMHMM2.0        TMhelix    350   372
mamO    TMHMM2.0        inside     373   378
mamO    TMHMM2.0        TMhelix    379   401
mamO    TMHMM2.0        outside    402   420
mamO    TMHMM2.0        TMhelix    421   440
mamO    TMHMM2.0        inside     441   441
mamO    TMHMM2.0        TMhelix    442   464
mamO    TMHMM2.0        outside    465   514
mamO    TMHMM2.0        TMhelix    515   537
mamO    TMHMM2.0        inside     538   549
mamO    TMHMM2.0        TMhelix    550   572
mamO    TMHMM2.0        outside    573   581
mamO    TMHMM2.0        TMhelix    582   604
mamO    TMHMM2.0        inside     605   632
```



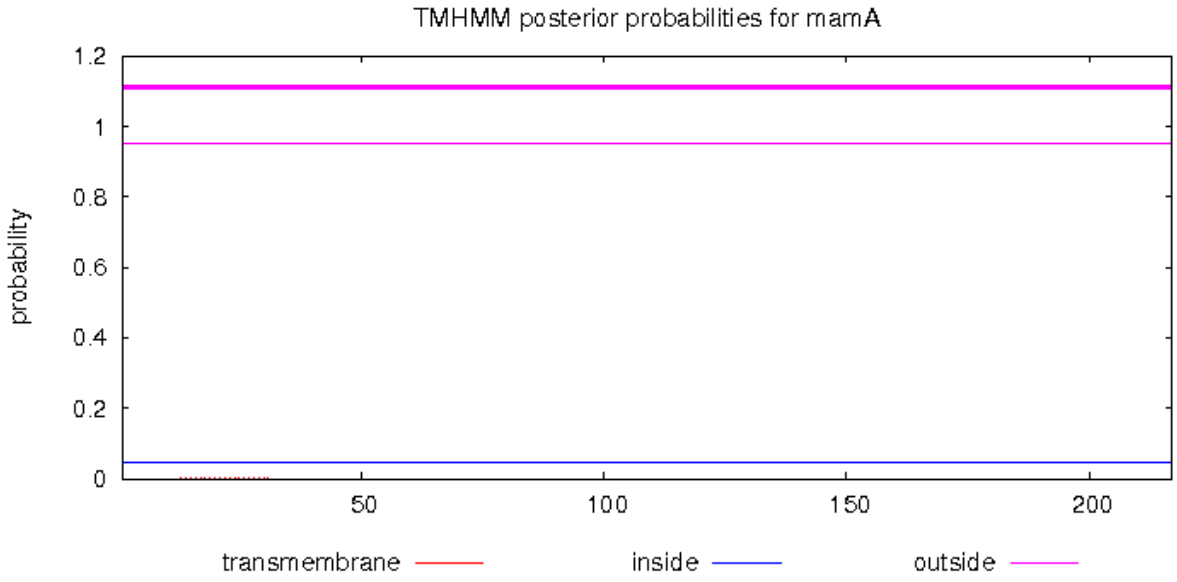
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamP Length: 270
# mamP Number of predicted TMHs:  1
# mamP Exp number of AAs in TMHs: 22.13825
# mamP Exp number, first 60 AAs:  21.45319
# mamP Total prob of N-in:        0.96791
# mamP POSSIBLE N-term signal sequence
mamP    TMHMM2.0        inside       1     4
mamP    TMHMM2.0        TMhelix      5    27
mamP    TMHMM2.0        outside     28   270
```



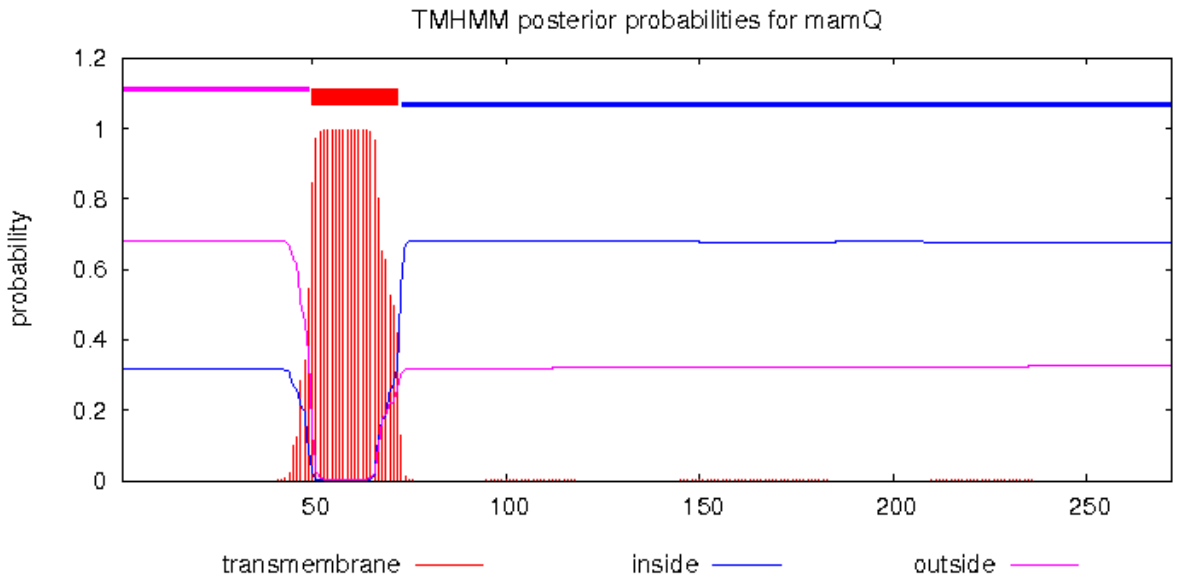
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamA Length: 217
# mamA Number of predicted TMHs:  0
# mamA Exp number of AAs in TMHs: 0.00587999999999999999
# mamA Exp number, first 60 AAs:  0.00393
# mamA Total prob of N-in:        0.04844
mamA    TMHMM2.0        outside      1   217
```



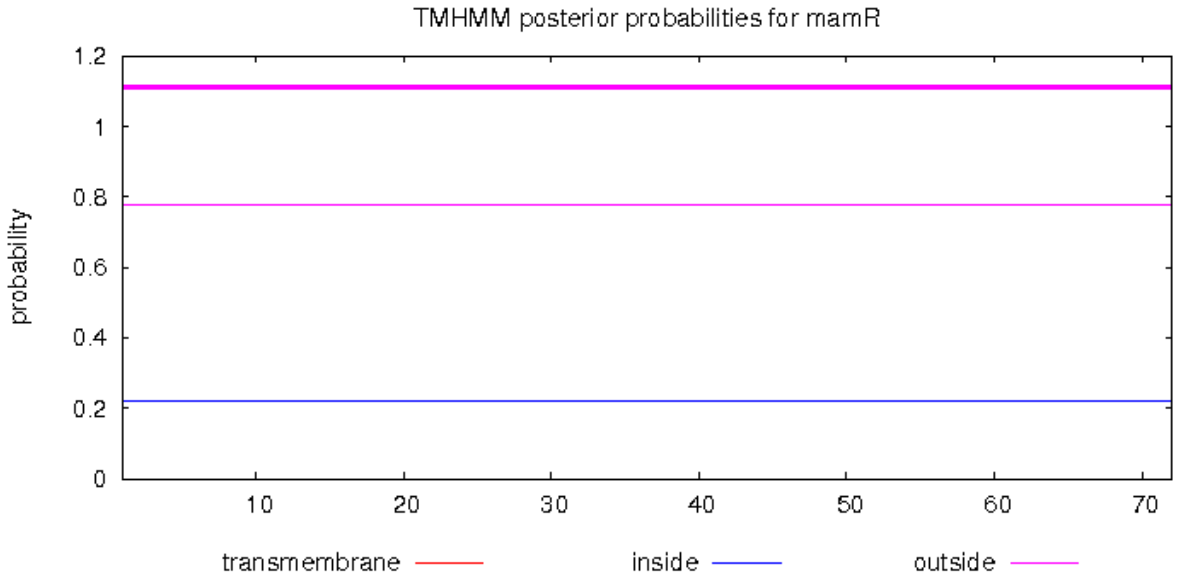
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamQ Length: 272
# mamQ Number of predicted TMHs:  1
# mamQ Exp number of AAs in TMHs: 21.97228
# mamQ Exp number, first 60 AAs:  12.2057
# mamQ Total prob of N-in:        0.31720
# mamQ POSSIBLE N-term signal sequence
mamQ    TMHMM2.0        outside      1    49
mamQ    TMHMM2.0        TMhelix     50    72
mamQ    TMHMM2.0        inside      73   272
```



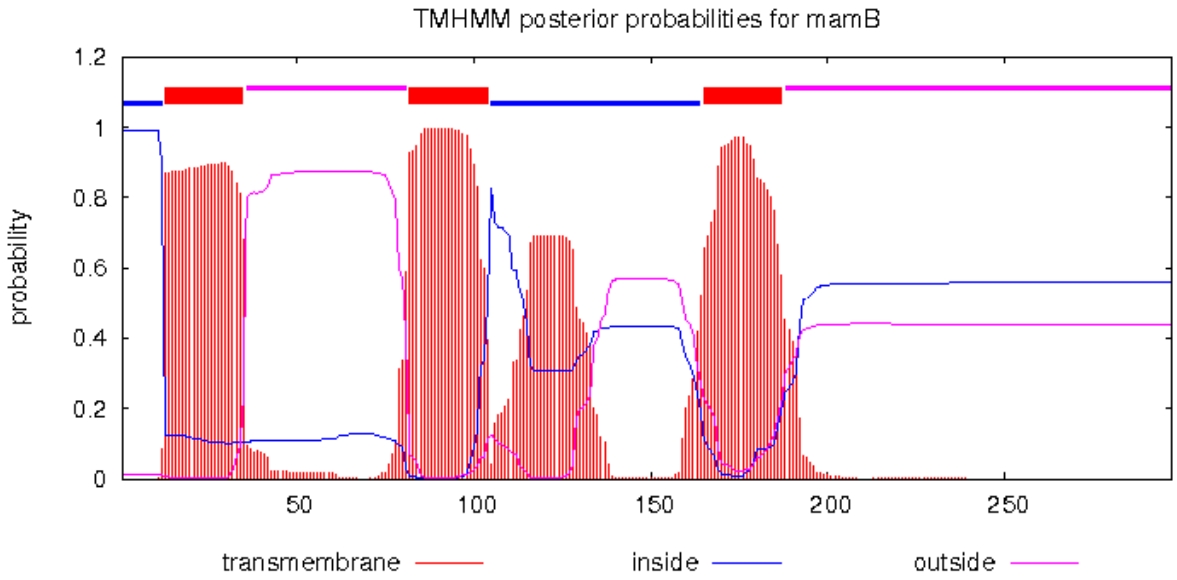
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamR Length: 72
# mamR Number of predicted TMHs:  0
# mamR Exp number of AAs in TMHs: 0.00055
# mamR Exp number, first 60 AAs:  0.00055
# mamR Total prob of N-in:        0.22089
mamR    TMHMM2.0        outside      1    72
```



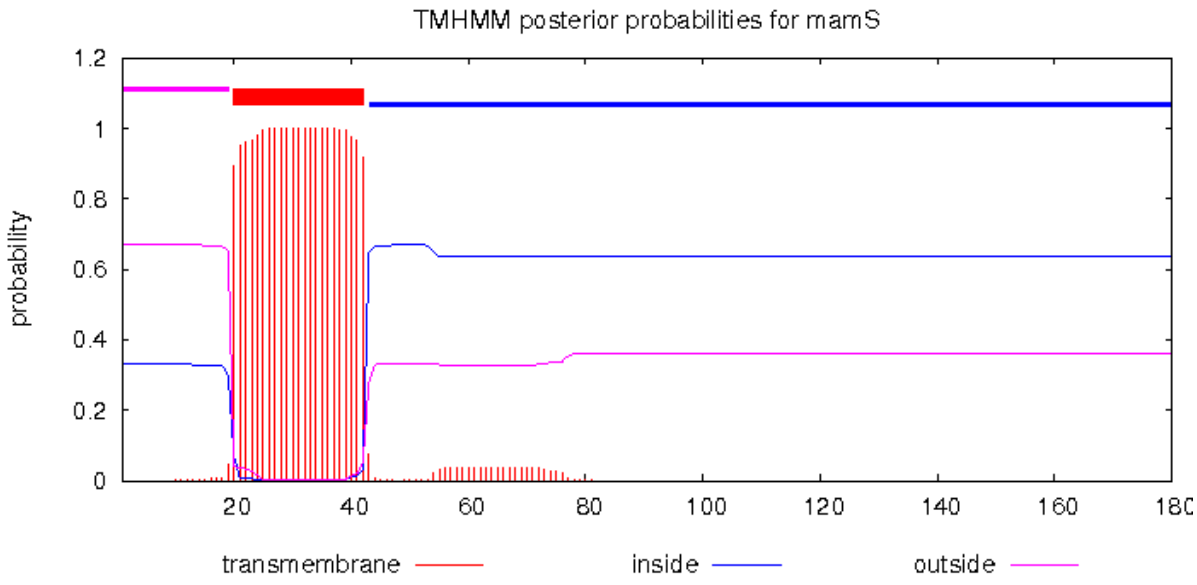
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamB Length: 297
# mamB Number of predicted TMHs:  3
# mamB Exp number of AAs in TMHs: 81.65791
# mamB Exp number, first 60 AAs:  20.96747
# mamB Total prob of N-in:        0.98953
# mamB POSSIBLE N-term signal sequence
mamB    TMHMM2.0        inside       1    12
mamB    TMHMM2.0        TMhelix     13    35
mamB    TMHMM2.0        outside     36    81
mamB    TMHMM2.0        TMhelix     82   104
mamB    TMHMM2.0        inside     105   164
mamB    TMHMM2.0        TMhelix    165   187
mamB    TMHMM2.0        outside    188   297
```



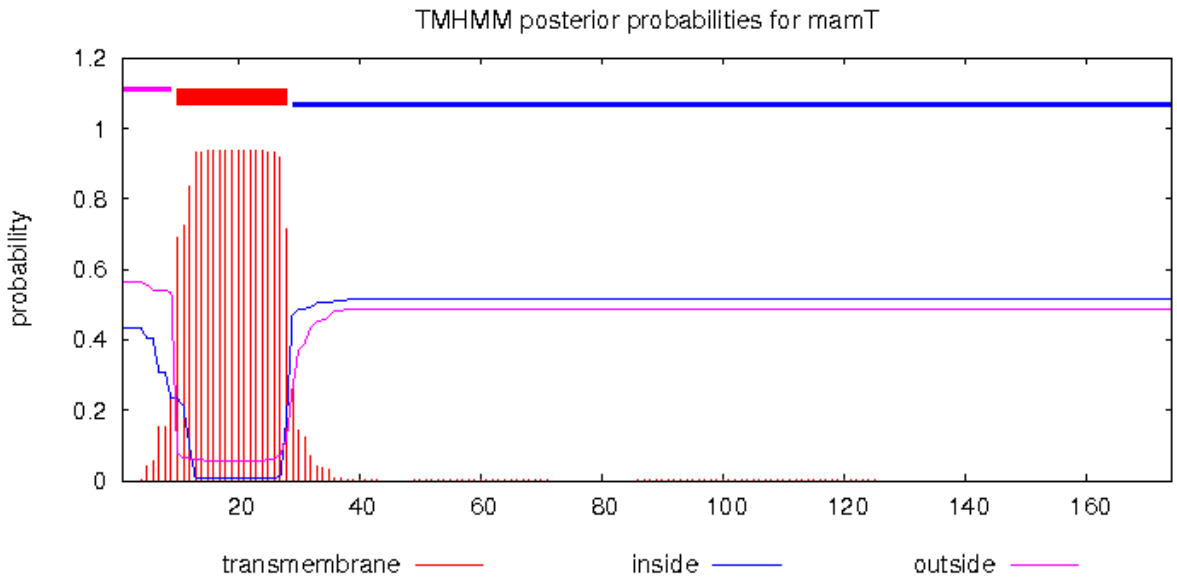
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamS Length: 180
# mamS Number of predicted TMHs:  1
# mamS Exp number of AAs in TMHs: 23.54119
# mamS Exp number, first 60 AAs:  22.98948
# mamS Total prob of N-in:        0.33044
# mamS POSSIBLE N-term signal sequence
mamS    TMHMM2.0        outside      1    19
mamS    TMHMM2.0        TMhelix     20    42
mamS    TMHMM2.0        inside      43   180
```



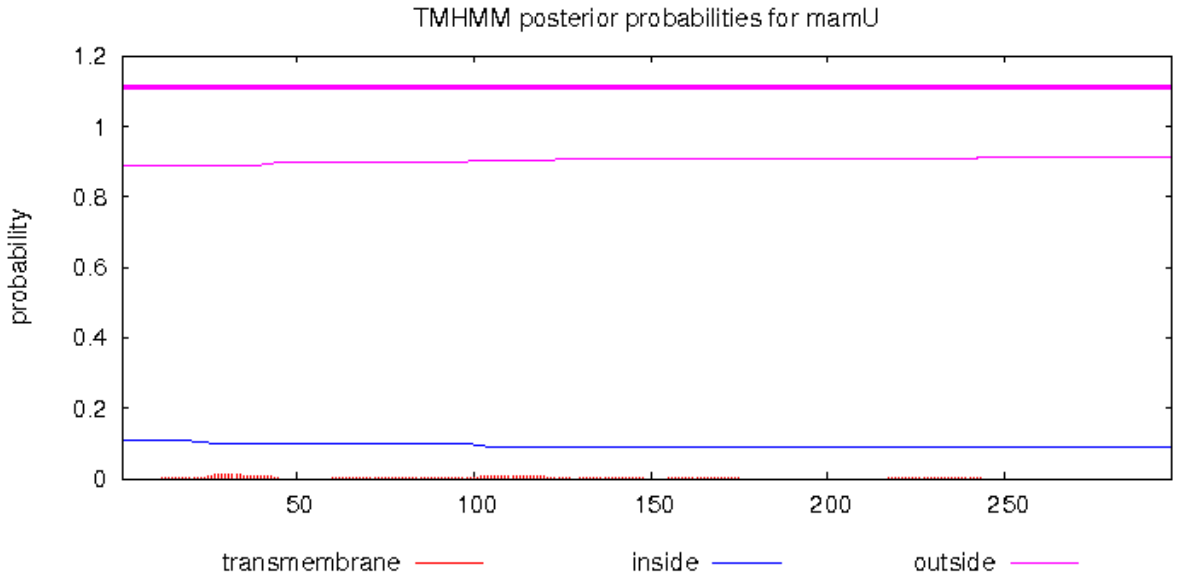
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamT Length: 174
# mamT Number of predicted TMHs:  1
# mamT Exp number of AAs in TMHs: 18.38954
# mamT Exp number, first 60 AAs:  18.37315
# mamT Total prob of N-in:        0.43583
# mamT POSSIBLE N-term signal sequence
mamT    TMHMM2.0        outside      1     9
mamT    TMHMM2.0        TMhelix     10    28
mamT    TMHMM2.0        inside      29   174
```



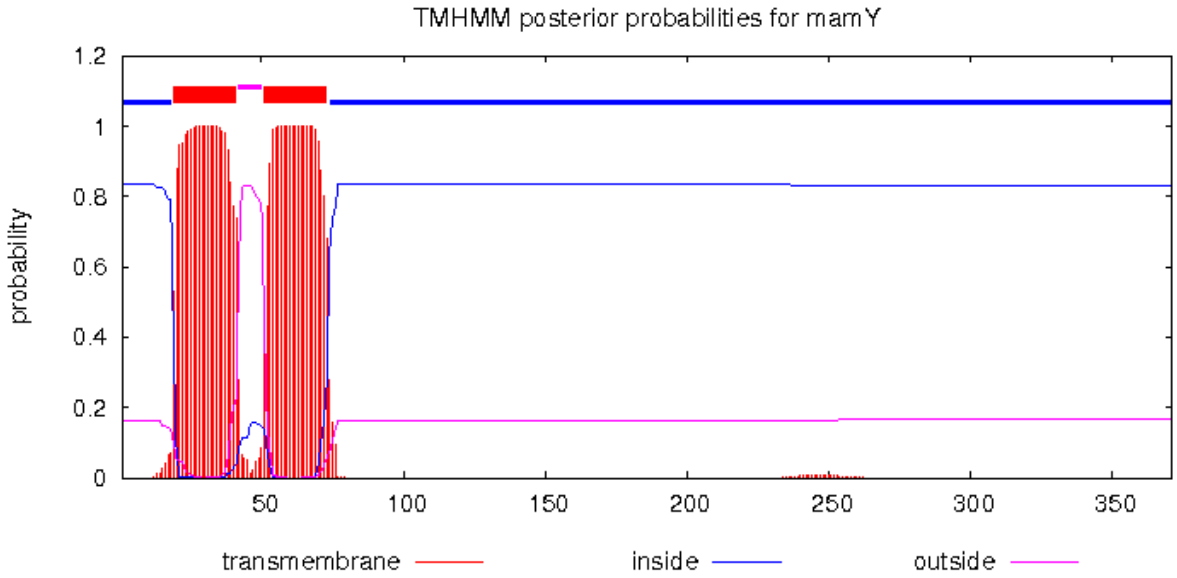
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamU Length: 297
# mamU Number of predicted TMHs:  0
# mamU Exp number of AAs in TMHs: 0.47406
# mamU Exp number, first 60 AAs:  0.19727
# mamU Total prob of N-in:        0.10933
mamU    TMHMM2.0        outside      1   297
```



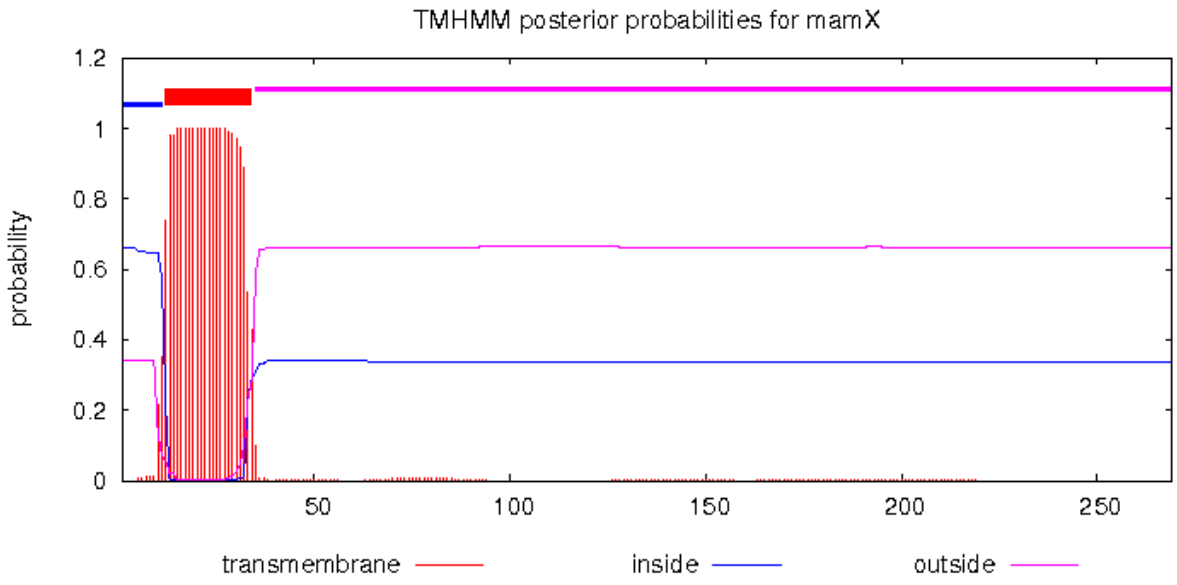
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamY Length: 371
# mamY Number of predicted TMHs:  2
# mamY Exp number of AAs in TMHs: 44.45026
# mamY Exp number, first 60 AAs:  31.50419
# mamY Total prob of N-in:        0.83605
# mamY POSSIBLE N-term signal sequence
mamY    TMHMM2.0        inside       1    18
mamY    TMHMM2.0        TMhelix     19    41
mamY    TMHMM2.0        outside     42    50
mamY    TMHMM2.0        TMhelix     51    73
mamY    TMHMM2.0        inside      74   371
```



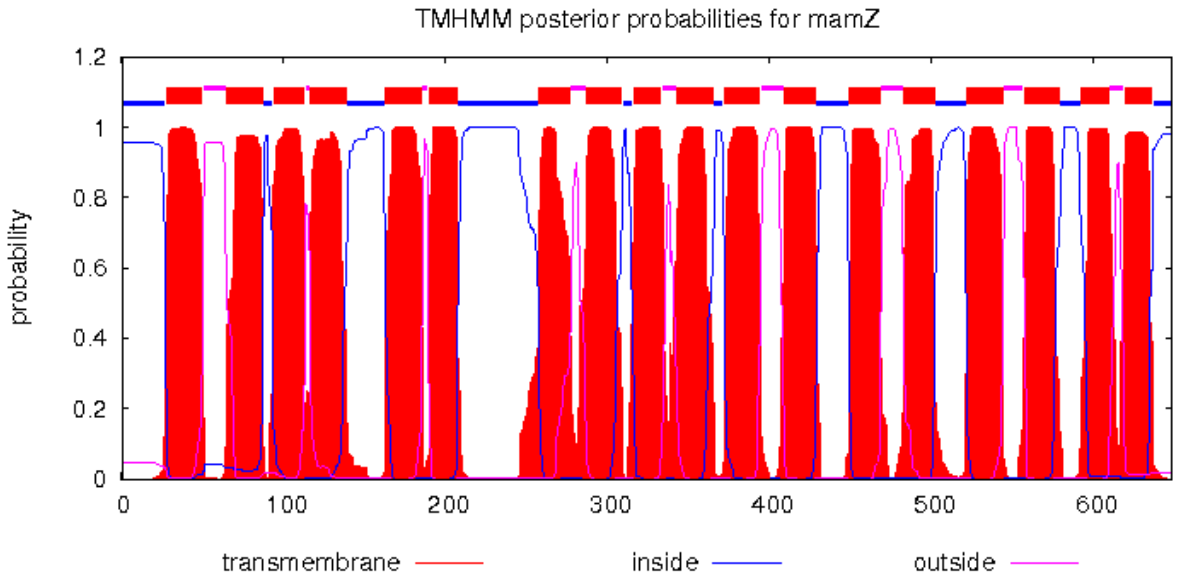
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamX Length: 269
# mamX Number of predicted TMHs:  1
# mamX Exp number of AAs in TMHs: 22.41271
# mamX Exp number, first 60 AAs:  22.16809
# mamX Total prob of N-in:        0.65981
# mamX POSSIBLE N-term signal sequence
mamX    TMHMM2.0        inside       1    11
mamX    TMHMM2.0        TMhelix     12    34
mamX    TMHMM2.0        outside     35   269
```



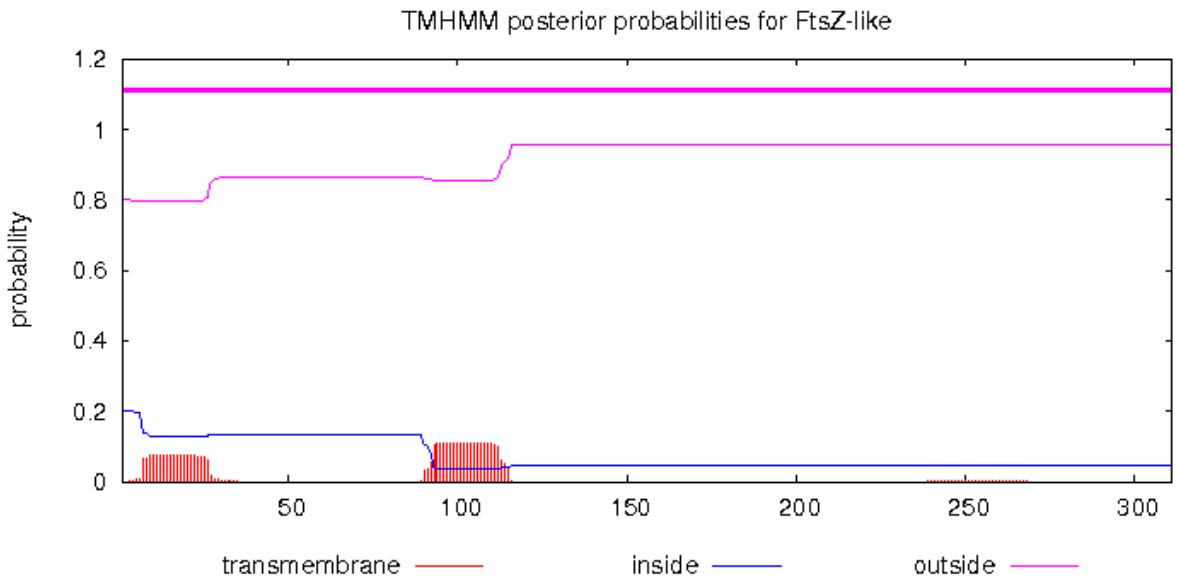
plot in postscript, script for making the plot in gnuplot, data for plot

```
# mamZ Length: 648
# mamZ Number of predicted TMHs:  18
# mamZ Exp number of AAs in TMHs: 373.22384
# mamZ Exp number, first 60 AAs:  22.6034
# mamZ Total prob of N-in:        0.95598
# mamZ POSSIBLE N-term signal sequence
mamZ    TMHMM2.0       inside       1    27
mamZ    TMHMM2.0       TMhelix     28    50
mamZ    TMHMM2.0       outside     51    64
mamZ    TMHMM2.0       TMhelix     65    87
mamZ    TMHMM2.0       inside      88    93
mamZ    TMHMM2.0       TMhelix     94   113
mamZ    TMHMM2.0       outside    114   116
mamZ    TMHMM2.0       TMhelix    117   139
mamZ    TMHMM2.0       inside     140   162
mamZ    TMHMM2.0       TMhelix    163   185
mamZ    TMHMM2.0       outside    186   189
mamZ    TMHMM2.0       TMhelix    190   207
mamZ    TMHMM2.0       inside     208   257
mamZ    TMHMM2.0       TMhelix    258   277
mamZ    TMHMM2.0       outside    278   286
mamZ    TMHMM2.0       TMhelix    287   309
mamZ    TMHMM2.0       inside     310   315
mamZ    TMHMM2.0       TMhelix    316   333
mamZ    TMHMM2.0       outside    334   342
mamZ    TMHMM2.0       TMhelix    343   365
mamZ    TMHMM2.0       inside     366   371
mamZ    TMHMM2.0       TMhelix    372   394
mamZ    TMHMM2.0       outside    395   408
mamZ    TMHMM2.0       TMhelix    409   428
mamZ    TMHMM2.0       inside     429   448
mamZ    TMHMM2.0       TMhelix    449   468
mamZ    TMHMM2.0       outside    469   482
mamZ    TMHMM2.0       TMhelix    483   502
mamZ    TMHMM2.0       inside     503   521
mamZ    TMHMM2.0       TMhelix    522   544
mamZ    TMHMM2.0       outside    545   556
mamZ    TMHMM2.0       TMhelix    557   579
mamZ    TMHMM2.0       inside     580   591
mamZ    TMHMM2.0       TMhelix    592   609
mamZ    TMHMM2.0       outside    610   618
mamZ    TMHMM2.0       TMhelix    619   636
mamZ    TMHMM2.0       inside     637   648
```



plot in postscript, script for making the plot in gnuplot, data for plot

```
# FtsZ-like Length: 311
# FtsZ-like Number of predicted TMHs:  0
# FtsZ-like Exp number of AAs in TMHs: 3.94874
# FtsZ-like Exp number, first 60 AAs:  1.49878
# FtsZ-like Total prob of N-in:        0.19916
FtsZ-like       TMHMM2.0        outside      1    311
```



plot in postscript, script for making the plot in gnuplot, data for plot

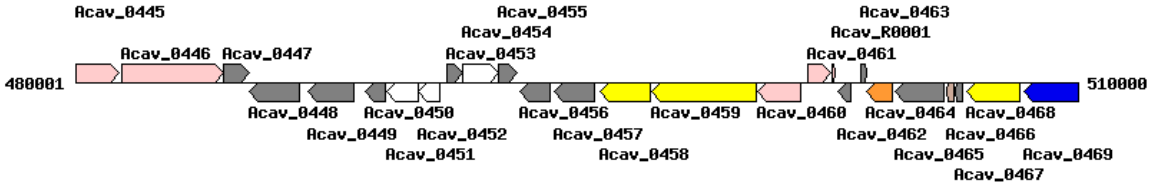
Supplementary Figure 2. Operon analysis of various LemA proteins.

No.	Organism	Pfam						LemA group	TMD*
		TPM_P	[GR-C]	DUF2207	P_M48/56	P_S	DeH		
1	*Acidovorax avenae*	2	1					2	0
2	*Acidovorax citrulli*							1	1
3	*Acinetobacter baumannii* AB0057	2	1					2	0*
4	*Acholeplasma oculi*							1	1
5	*Agrobacterium tumefaciens*			1		1		3	1
6	*Bacillus clausii*				1			4	1
7	*Bacillus lehensis*				1			4	1
8	***Bacillus megaterium***	1	1					2	1
9	*Bacillus endophyticus*	1	1					2	1
10	*Bacteroides cellulosilyticus*	1	1			1	1	2	1
11	*Berkelbacteria bacterium*					1		4	1
12	*Blattabacterium* BPLAN	1	-					2	1
13	*Bradyrhizobium diazoefficiens*	2	1					2	0
14	*Brucella abortus*	2	1	1			1	2	1
15	***Brucella melitensis***	2	1	1			1	2	1
16	***Clostridium kluyveri***							1	1
17	*Desulfarculus baarsii*	1	1					2	1
18	*Francisella cf. novicida*				1			1	1
19	*Lactobacillus johnsonii*				1			4	1
20	*Lactobacillus rhamnosus*	1	1					2	1
21	*Lactobacillus acetotolerans*				1			4	1
22	*Leptospira biflexa*						1	2	0
23	*Marinobacter salaries*				1			3	1
24	*Methylophaga nitratireducenticrescens*				1			3	1
25	*Mycobacterium smegmatis*			1				1	1
26	*Nitrosomonas sp.* AL212	2	1					2	0*
27	*Nitrosomonas* sp. Is79A3	2	1					2/3	0*/1
28	*Owenweeksia hongkongensis*	2	1					2	1
29	*Parachlamydia acanthamoebae*				1			3	1
30	*Pasteurella multocida*	2	1					2	1
31	*Pectobacterium wasabiae*			1			1	3	1
32	*Pelosinus fermentans*	1	1					2	1
33	*Providencia stuartii*				1			1	1+*

34	***Pseudomonas aeruginosa***							3	1
35	*Rhodococcus aetherivorans*			1				1	1
36	*Rickettsiales bacterium*							2	1
37	*Rufibacter* sp. DG31D	2	1					2	0
38	*Shewanella baltica*							1	1
39	*Sorangium cellulosum*	1	1					2	0
40	*Spirosoma radiotolerans*	2	1				1	2	1
41	*Sphaerochaeta coccoides*			1				3	1
42	*Sphingomonas sp.* MM-1	2	1				1	2	0*
43	*Streptococcus parauberis*				1			4	1
44	*Streptococcus pneumoniae*				1			4	1
45	*Streptococcus agalactiae*				1			4	1
46	*Sulfuricella denitrificans*	2	1				1	2	0*
47	*Syntrophus aciditrophicus*	1	1					2	1
48	*Tannerella forsythia*				1			3	1
49	*Thermoanaerobacter wiegelii*							2	1
50	*Vulgatibacter incomptus*	1	1			1		2	0
51	*Yersinia pestis*							3	1
52	*Zunongwangia profunda*	2	1					2	0

TPM_P – TPM_Phosphatase
GR-C – TPM_Phosphatases with glycine-rich C-terminus motif
P_M48/56 – Peptidase_M48/56
P_S – Peptidase_S7/9/15/41/46/49
DeH – Dehydrogenase

1. *Acidovorax avenae*

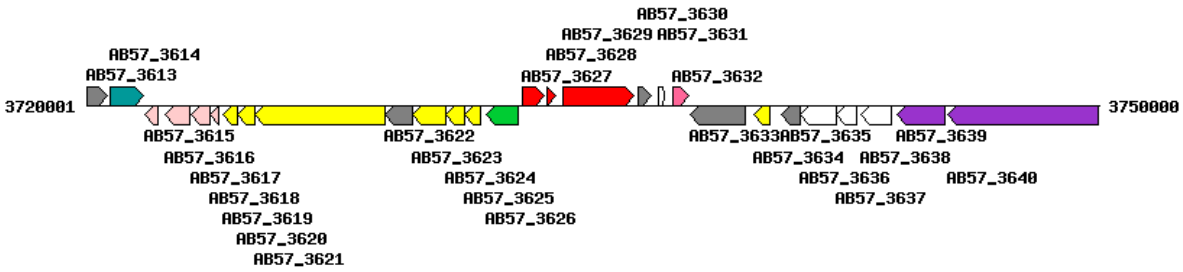


Acav_0452 – LemA (Pfam: LemA; LPAM_2[Prokaryotic lipoprotein-attachment site])
Acav_0451 – protein of unknown function DUF477 (Pfam: TPM_phosphatase; MOLO1)
Acav_0450 – protein of unknown function DUF477 (Pfam: TPM_phosphatase; DUF5130)
Acav_0449 – hypothetical protein (Pfam: NEL[ubiquitin ligase]; TENA_THI-4)
Acav_0448 – hypothetical protein (Pfam: NEL[ubiquitin ligase])

2. *Acidovorax citrulli*

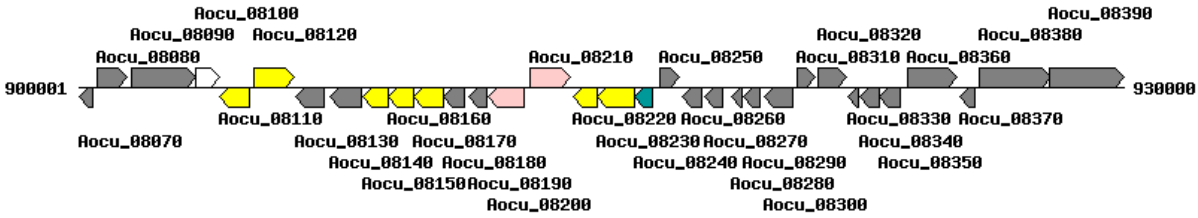
Aave_4529 – LemA (Pfam: LemA; TPR_14)
Aave_4528 – sun protein (Pfam: Methyltr_RsmB-F; NusB; Methyltr_RsmF_N; Methyltransf_25; Methyltransf_24; FtsJ)
Aave_4527 – putative proline rich signal peptide protein (Pfam: DUF4390; Bac_luciferase)
Aave_4526 – multi-sensor signal transduction histidine kinase (Pfam: Response_reg; HTH_8; FleQ; Phage_AlpA; HTH_1)
Aave_4525 – tRNA-Phe

3. *Acinetobacter baumannii* AB0057



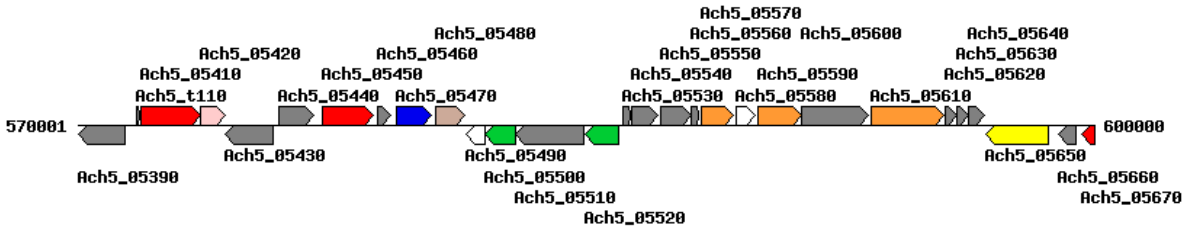
AB57_3638 – hypothetical protein (Pfam: Metal_hydrol; AurF)
AB57_3637 – LemA (Pfam: LemA; DUF1869; Dynein_heavy; HRXXH)
AB57_3636 – conserved hypothetical protein (Pfam: TPM_phosphatase; BCD; DUF5130; SRTM1; Shisa)
AB57_3635 – conserved hypothetical protein (Pfam: TPM_phosphatase; DUF4019)
AB57_3634 – type IV pilin structural subunit (Pfam: Pilin; N_methyl_3; N_methyl_2; N_methyl)
AB57_3633 – O-antigen polymerase family (Pfam: Wzy_C; Wzy_C_2; PglL_A)

4. *Acholeplasma oculi*



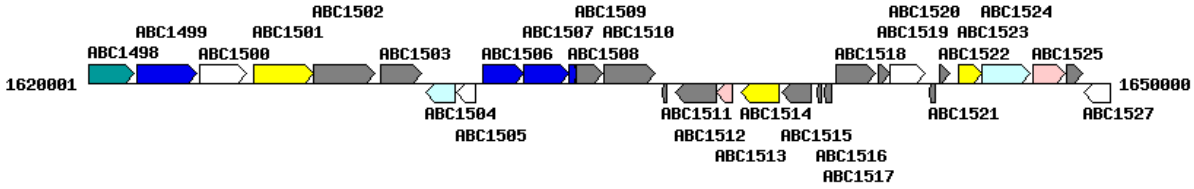
Aocu_08080 – DAK1/DegV-like protein (Pfam: DegV; Cupin_8; DUF1108)
Aocu_08090 – hypothetical protein (Pfam: Adeno_E3_CR1; Sod_Fe_N; DUF1430; **IncA**)
Aocu_08100 – LemA (Pfam: LemA; Pmp3; DUF948; ALIX_LYPXL_bnd; DUF4358)

5. *Agrobacterium tumefaciens*



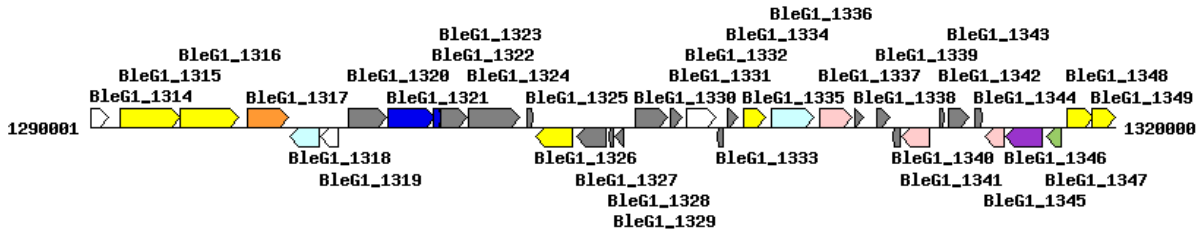
Ach5_05530 – hypothetical protein (Pfam: DUF2007)
Ach5_05540 – methyltransferase (Pfam: MTS; Methyltransf_31; Methyltransf_25; Methyltransf_16; Cons_hypoth95; GidB; Methyltransf_12; NAD_binding_8; Methyltransf_11; Methyltransf_32)
Ach5_05550 – protease sohB [sohB] (Pfam: Peptidase_S49; SDH_sah; CLP_protease; Pribosyltran)
Ach5_05560 – membrane protein (Pfam: HAMP; ParE-like_toxin)
Ach5_05570 – glycyl-tRNA synthetase alpha chain [glyQ] (Pfam: tRNA-synt_2e; N-glycanase_C)
Ach5_05580 – LemA (Pfam: LemA)
Ach5_05590 – 3-phosphoshikimate 1-carboxyvinyltransferase [aroA] (Pfam: EPSP_synthase)
Ach5_05600 – membrane protein (Pfam: DUF2207)
Ach5_05610 – glycyl-tRNA synthetase beta chain [glyS] (Pfam: tRNA_synt_2f; DALR_1; MnmE_helical; AARP2CN)
Ach5_05620 – addiction module protein (Pfam: Gp49)
Ach5_05630 – transcriptional regulator
Ach5_05640 – hypothetical protein (Pfam: DUF523)

6. *Bacillus clausii*



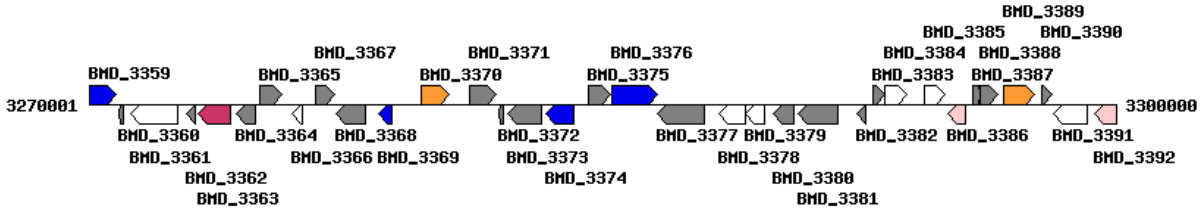
ABC1504 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56; SprT-like; BSP)
ABC1505 – LemA (Pfam: LemA; DUF3584; DASH_Spc34; Vps39_1)

7. *Bacillus lehensis* G1



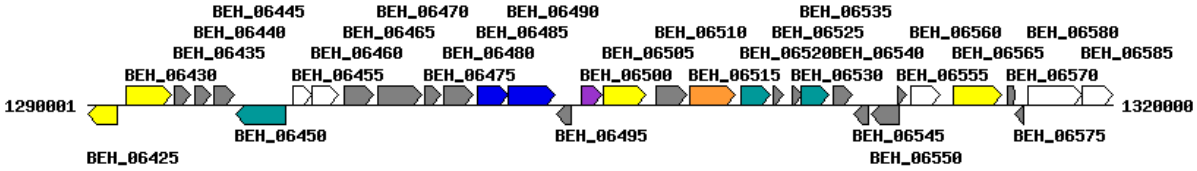
BleG1_1318 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56; SprT-like; Borrelia_P13; BSP; Arteri_Gl)
BleG1_1319 – LemA (Pfam: LemA; DASH_Spc34; DUF3584)

8. *Bacillus megaterium* DSM 319



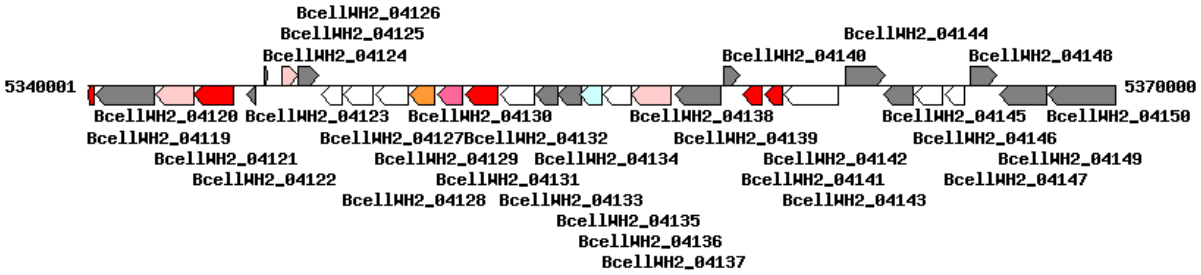
BMD_3382 – conserved hypothetical protein (Pfam: gerPA**[germination]**)
BMD_3381 – glycosyl transferase, family 2 (Pfam: Glycos_transf_2; Glyco_tranf_2_2; Glyco_transf_7C; Glyco_tranf_2_3; Glyco_transf_21; Glyco_tranf_2_4; Glyco_trans_2_3)
BMD_3380 – polysaccharide deacetylase (Pfam: Polysacc_deac_1; Glyco_hydro_57; SpoVS**[sporulation];** DUF2209)
BMD_3379 – LemA (Pfam: LemA; DUF327; DUF948; FadA; DivIC)
BMD_3378 – pomain of unknown function (Pfam: TPM_phosphatase)
BMD_3377 – conserved hypothetical protein (Pfam: DUF2254)

9. *Bacillus endophyticus*



BEH_06455 – LemA (Pfam: LemA; FtsH_ext; FadA; 4HB_MCP_1; DUF4455; Ribosomal_L36e; ABC2_membrane_3)
BEH_06460 – hypothetical protein (Pfam: TPM_phosphatase; Asp4; Amnionless)
BEH_06465 – diacylglycerol kinase (Pfam: DAGK_cat; NAD_kinase)
BEH_06470 – pseudogene
BEH_06475 – hypothetical protein (Pfam: zf-dskA_traR; DUF5133)
BEH_06480 – RpiR family transcriptional regulator (Pfam: SIS; HTH_6; SIS_2; HTH_17)
BEH_06485 – N-acetylmuramic acid-6-phosphate etherase (Pfam: SIS; SIS_2; ANTAR; DUF2682; UBA_4)
BEH_06490 – PTS sugar transporter subunit IIC (Pfam: PTS_EIIC; PTS_EIIB)

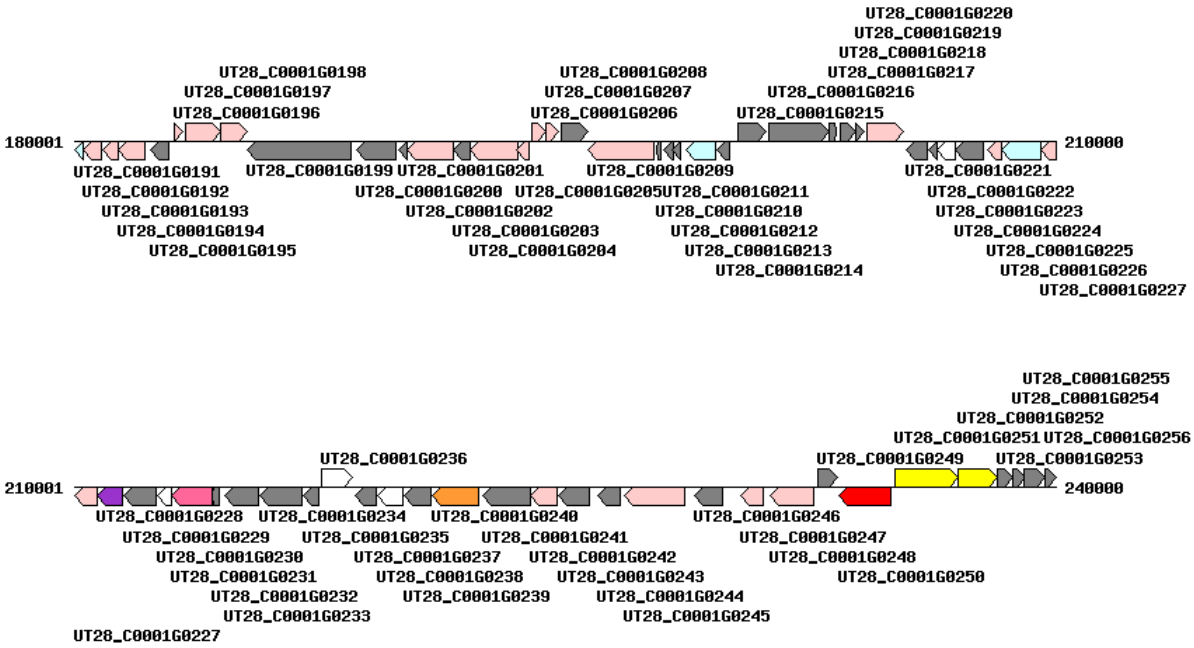
10. *Bacteroides cellulosilyticus*



BcellWH2_04139 – hypothetical protein (Pfam: DUF1593; Glyco_hydro_129; REJ)
BcellWH2_04138 – GTPase Obg [obg] (Pfam: GTP1_OBG; MMR_HSR1; FeoB_N; MMR_HSR1_Xtn; DNA_pol_phi)
BcellWH2_04137 – Laccase domain protein (Pfam: Cu-oxidase_4; DUF4649)
BcellWH2_04136 – lipoprotein signal peptidase (Pfam: Peptidase_A8; DUF2569)
BcellWH2_04135 – B3/4 domain protein (Pfam: B3_4)
BcellWH2_04134 – Murein hydrolase activator NlpD precursor [nlpD_2] (Pfam: Peptidase_M23; PTS_EIIA_1; Biotin_lipoyl_2)
BcellWH2_04133 – PhoH-like protein [ybeZ_1] (Pfam: PhoH; AAA_30; UvrD-helicase; DEAD; IstB_IS21; AAA_19; AAA_22; ResIII; Ribosomal_L4; AAA_25; DUF2075; AAA; Microtub_bd; Sigma54_activ_2; Methyltr_RsmB-F; Kinesin)
BcellWH2_04132 – Phosphoribosylaminoimidazole-succinocarboxamide synthase [purC] (Pfam: SAICAR_synt)
BcellWH2_04131 – C-methyltransferase UbiE [ubiE] (Pfam: Ubie_methyltran; Methyltransf_25; Methyltransf_11; Methyltransf_31; Methyltransf_12; Methyltransf_23; Methyltransf_20; Methyltransf_18; Rsm22; TrmK; UPF0020;

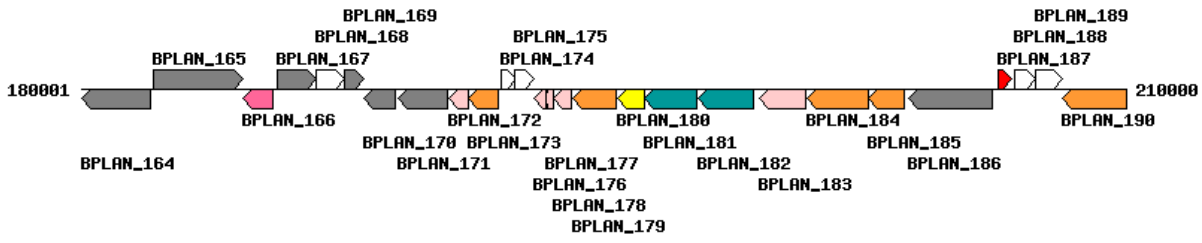
PrmA; MetJ; MTS; Methyltransf_8; PCMT; Methyltransf_3; FtsJ)
BcellWH2_04130 – Shikimate dehydrogenase [aroE] (Pfam: Shikimate_dh_N; 2-Hacid_dh_C)
BcellWH2_04129 – Alpha/beta hydrolase family protein (Pfam: Peptidase_S9; Hydrolase_4; Peptidase_S15; FSH1; DLH; Abhydrolase_2)
BcellWH2_04128 – hypothetical protein (Pfam: TPM_phosphatase; MOLO1; Zn-ribbon_8; zf-ribbon_3; NOB1_Zn_bind; DUF3268; DZR; Ogr_Delta; TF_Zn_Ribbon; zinc_ribbon_15; DNA_RNApol_7kD; Ribosomal_L32p; zinc_ribbon_10)
BcellWH2_04127 – LemA (Pfam: LemA; DivIC; 4HB_MCP_1)

11. *Berkelbacteria bacterium*



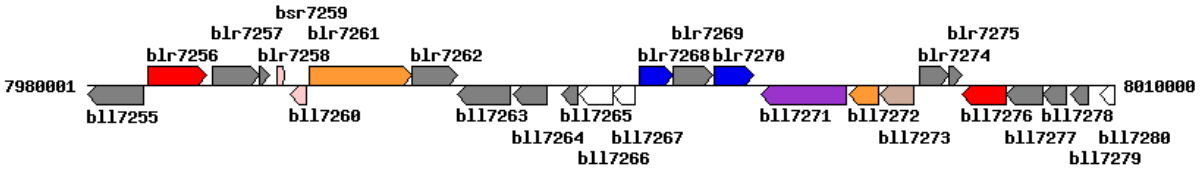
UT28_C0001G0235 – hypothetical protein
UT28_C0001G0234 – hypothetical protein (Pfam: DUF4307; DUF2850; DUF2140)
UT28_C0001G0233 – hypothetical protein (Pfam: PsbP; DUF4307)
UT28_C0001G0232 – hypothetical protein (Pfam: PhdYeFM_antitox; EFG_C)
UT28_C0001G0231 – cysteine desulfurase (Pfam: Aminotran_5; Beta_elim_lyase)
UT28_C0001G0230 – hypothetical protein (Pfam: NifU_N)
UT28_C0001G0229 – hypothetical protein (Pfam: Colicin_V; CAP)
UT28_C0001G0228 – FolD bifunctional protein (Pfam: THF_DHG_CYH_C; THF_DHG_CYH)
UT28_C0001G0227 – tRNA-specific 2-thiouridylase MnmA [mnmA] (Pfam: tRNA_Me_trans; NAD_synthase; QueC; ThiI; DUF3659; Asn_synthase; Arginosuc_synth; LytR_C; ATP_bind_3)
UT28_C0001G0226 – carboxy-terminal-processing protease [ctpA] (Pfam: Peptidase_S41; PDZ; PDZ_2; GRASP55_65[Golgi])
UT28_C0001G0225 – 50S ribosomal protein L9 [rplI] (Pfam: Ribosomal_L9_N; Ribosomal_L9_C)
UT28_C0001G0224 – mucin-2-like protein (Pfam: T2SSG; DctQ)
UT28_C0001G0223 – LemA (Pfam: LemA; DUF3151; HalX; FtsH_ext)
UT28_C0001G0222 – hypothetical protein (Pfam: ARGLU; DUF4889; TRAF_BIRC3_bd)
UT28_C0001G0221 – hypothetical protein (Pfam: PsbP; Rad9_Rad53_bind; PsaN; SSP160; ABC2_membrane_3)

12. *Blattabacterium* BPLAN (*Periplaneta americana*)



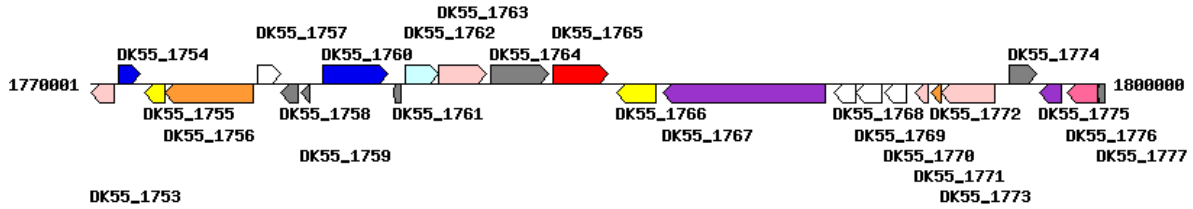
BPLAN_187 – nucleoside-diphosphate kinase (Pfam: NDK; MaoC_dehydratas)
BPLAN_188 – LemA (Pfam: LemA; Fez1; DUF1512; DASH_Spc34; EOS1; DivIC; Exonuc_VII_L; ABC2_membrane_3)
BPLAN_189 – hypothetical protein (Pfam: TPM_phosphatase)

13. *Bradyrhizobium diazoefficiens* USDA 110



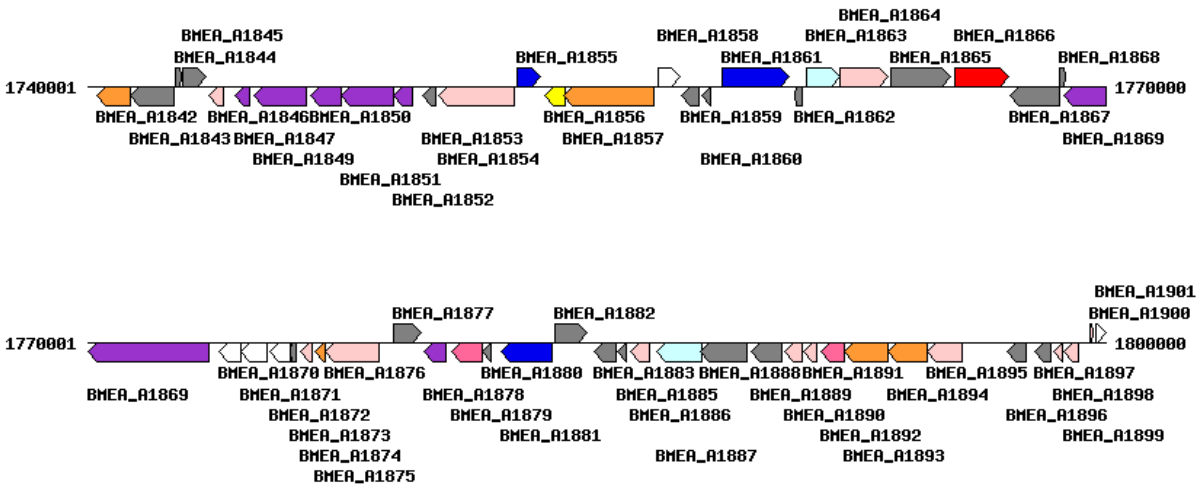
bll7267 – LemA (Pfam: LemA; UPF0693)
bll7266 – hypothetical protein (Pfam: TPM_phosphatase; MOLO1)
bll7265 – hypothetical protein (Pfam: TPM_phosphatase; DUF5130)
bll7264 – hypothetical protein
bll7263 – hypothetical protein

14. *Brucella abortus* bv. 2 86/8/59



DK55_1773 – signal recognition particle protein [ffh] (Pfam: SRP54; SRP_SPB; SRP54_N; CbiA; AAA_33; AAA_30; Zeta_toxin; AAA_19; AAA_31; DnaB_C; AAA_22; APS_kinase; Thymidylate_kin; AAA_16; AAA; 6PF2K; AAA_18; AAA_28; AAA_14; AAA_5; ArsA_ATPase)
DK55_1772 – chorismate mutase (Pfam: CM_2)
DK55_1771 – ribosomal protein S16 [rpsP] (Pfam: Ribosomal_S16; TAF4)
DK55_1770 – LemA (Pfam: LemA; TPR_15; Dehydratase_SU; zf-piccolo)
DK55_1769 – uncharacterized protein (Pfam: TPM_phosphatase; DUF5130; MOLO1; DUF987; DUF1840)
DK55_1768 – putative membrane protein (Pfam: TPM_phosphatase; TM140; DUF2207)
DK55_1767 – bacterial NAD-glutamate dehydrogenase family protein (Pfam: Bac_GDH; ELFV_dehydrog)
DK55_1766 – major Facilitator Superfamily protein (Pfam: ATG22; MFS_1; MFS_1_like)

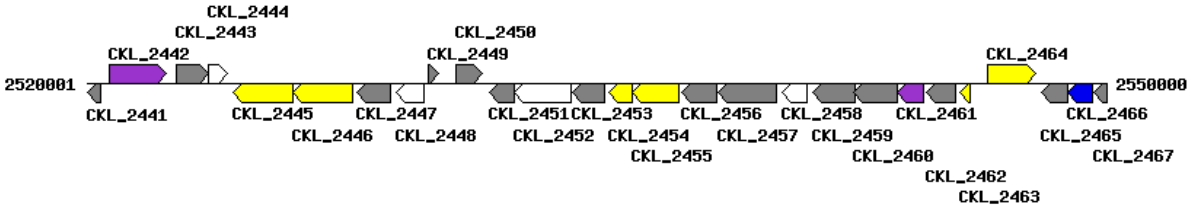
15. *Brucella melitensis* ATCC 23457



BMEA_A1876 – signal recognition particle protein [ffh] (Pfam: SRP54; SRP_SPB; SRP54_N; CbiA; AAA_33; AAA_30; Zeta_toxin; AAA_19; AAA_31; DnaB_C; AAA_22; APS_kinase; AAA_16; AAA; Thymidylate_kin; 6PF2K; AAA_18; AAA_28; AAA_14; ArsA_ATPase; AAA_5)
BMEA_A1875 – chorismate mutase (Pfam: CM_2)

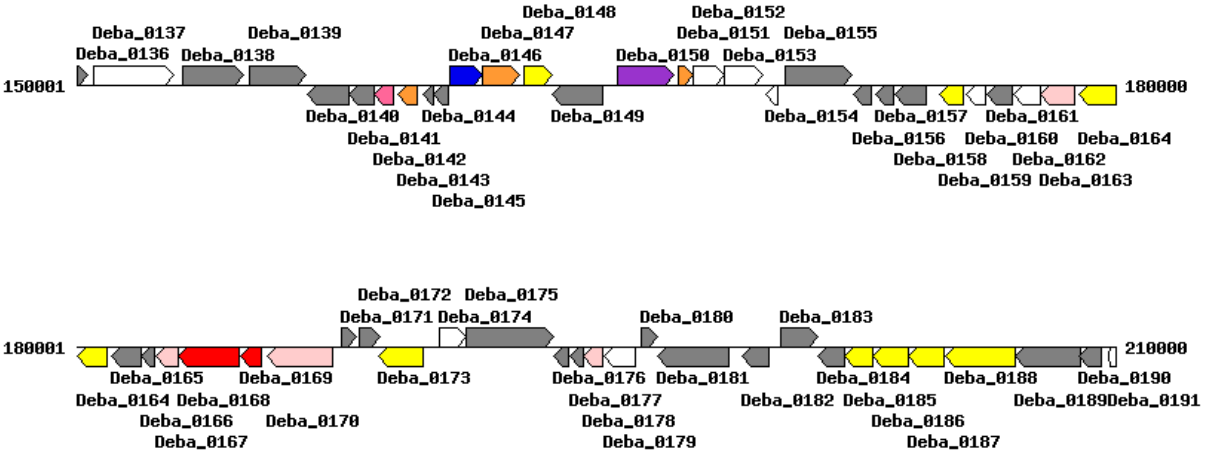
BMEA_A1874 – ribosomal protein S16 [rpsP] (Pfam: Ribosomal_S16)
BMEA_A1873 – hypothetical protein
BMEA_A1872 – LemA (Pfam: LemA; TPR_15; Dehydratase_SU; zf-piccolo)
BMEA_A1871– uncharacterized protein (Pfam: TPM_phosphatase; DUF5130; MOLO1; DUF987; DUF1840)
BMEA_A1870 – putative membrane protein (Pfam: TPM_phosphatase; TM140; DUF2207)
BMEA_A1869 – NAD-glutamate dehydrogenase (Pfam: Bac_GDH; ELFV_dehydrog)

16. *Clostridium kluyveri* DSM 555



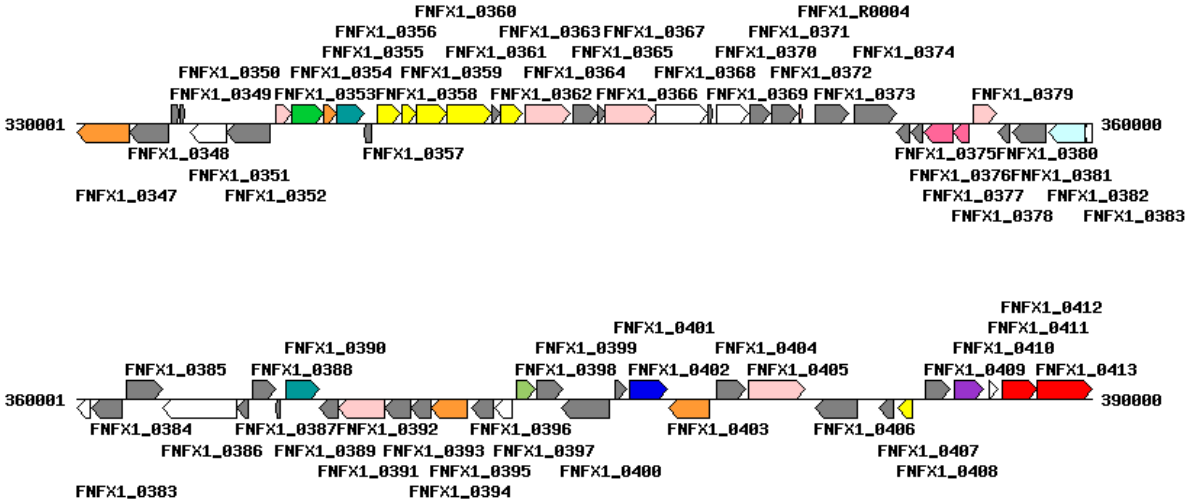
CKL_2442 – hydroxylamine reductase [hcp] (Pfam: Prismane)
CKL_2443 – conserved hypothetical protein (Pfam: DUF3137)
CKL_2444 – LemA (Pfam: LemA; Halogen_Hydrol; Baculo_11_kDa)

17. *Desulfarculus baarsii*



Deba_0170 – chaperone protein DnaK (Pfam: HSP70; MreB_Mbl; FtsA; DMRL_synthase)
Deba_0169 – Phosphoribosylaminoimidazolecarboxamideformyltransferase (Pfam: MGS; AICARFT_IMPCHas)
Deba_0168 – phosphoribosylamine/glycine ligase (Pfam: GARS_A; AIRC; GARS_N; GARS_C; ATPgrasp_Ter; CPSase_L_D2; ATP-grasp; ATP-grasp_5; ATP-grasp_4; ATP-grasp_2; Dala_Dala_lig_C; RimK; ATP-grasp_3; DAGK_cat)
Deba_0167 – Sua5/YciO/YrdC/YwlC family protein (Pfam: Sua5_yciO_yrdC)
Deba_0166 – hypothetical protein
Deba_0165 – metal dependent phosphohydrolase (Pfam: HDOD; HD)
Deba_0164 – methyl-accepting chemotaxis sensory transducer (Pfam: MCPsignal; dCache_1; Laminin_II; KASH_CCD; DUF1395; DUF1664)
Deba_0163 – chaperone DnaJ domain protein (Pfam: DnaJ_C; DnaJ; NIR_SIR_ferr; RRM_3)
Deba_0162 – protein of unknown function DUF477 (Pfam: TPM_phosphatase)
Deba_0161 – putative cytoplasmic protein (Pfam: NTP_transf_2)
Deba_0160 – LemA (Pfam: LemA; DUF4131; DivIC)
Deba_0159 – MgtC/SapB transporter (Pfam: MgtC; TctB)
Deba_0158 – cation diffusion facilitator family transporter (Pfam: Cation_efflux; ZT_dimer)
Deba_0157 – putative cache sensor protein (Pfam: dCache_2; sCache_2; Cache_3-Cache_2; Prim-Pol)
Deba_0156 – putative cache sensor protein (Pfam: dCache_2; sCache_2; Cache_3-Cache_2)

18. *Francisella cf. novicida* Fx1



FNFX1_0384 – hypothetical protein (Pfam: LysR_substrate; HTH_1; PBP_like; PBP_like_2; Phosphonate-bd; Orthopox_F14)

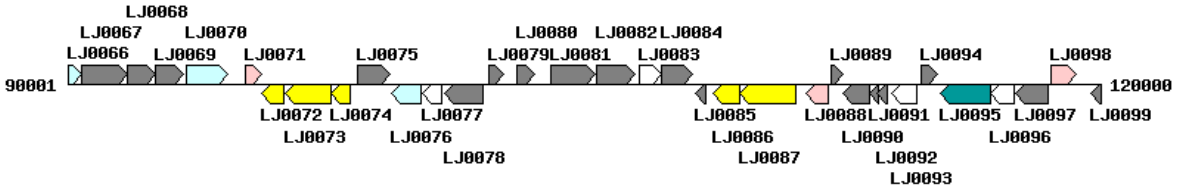
FNFX1_0383 – LemA (Pfam: LemA; DUF4330; DUF4083; Inhibitor_I53; UPF0118; Peptidase_M57; Wzy_C)

FNFX1_0382 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; zf-DHHC; DUF2268; DUF2569; DUF4157; DUF996)

FNFX1_0381 – hypothetical protein (Pfam: N_methyl_2; N_methyl)

FNFX1_0380 – hypothetical protein (Pfam: DUF3287; HrpE; DUF1043; 7TMR-HDED; Spc97_Spc98; DUF4140; ZnuA; DUF4446; Prefoldin_2; DUF342; Lectin_N; Med4; **IncA**)

19. *Lactobacillus johnsonii* NCC 533

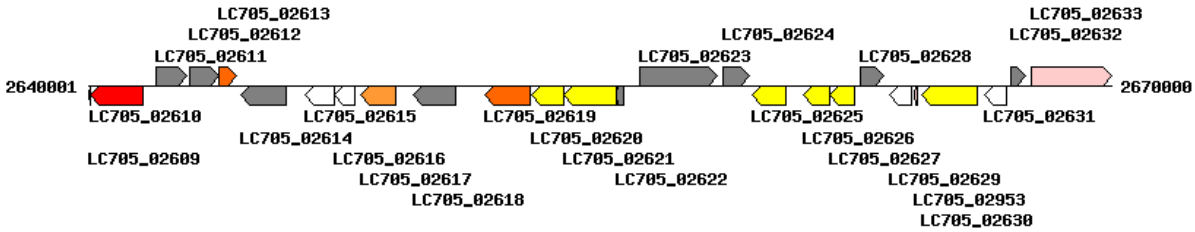


LJ0078 – hypothetical protein (Pfam: CitMHS)

LJ0077 – LemA (Pfam: LemA; DUF2408; FtsH_ext; IncA; DUF342; GAT; DUF3139)

LJ0076 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; Herpes_US9; Cytochrom_B_N_2)

20. *Lactobacillus rhamnosus* Lc 705



LC705_02622 – conserved protein

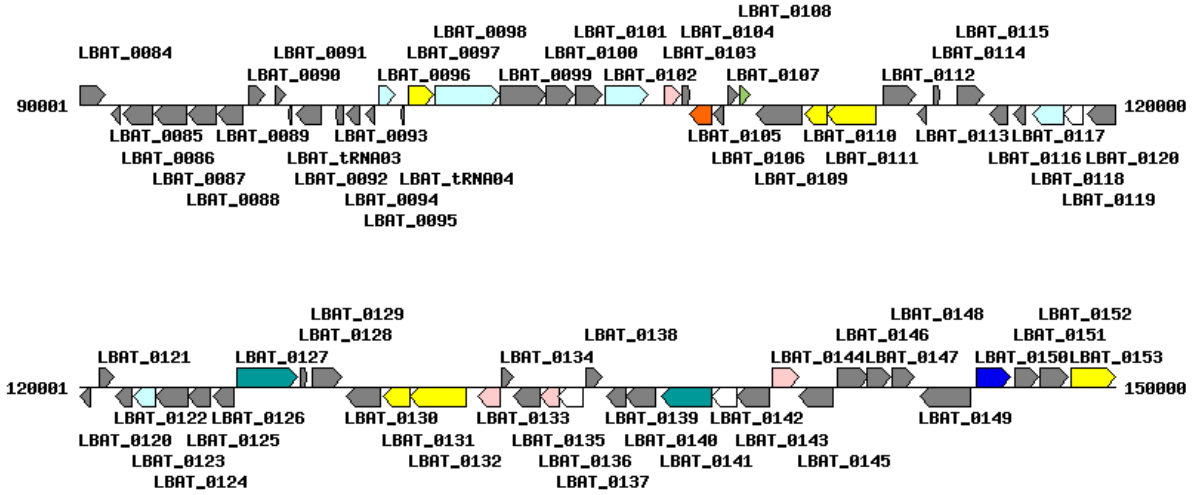
LC705_02621 – ABC transporter, permease protein (Pfam: OpuAC; BPD_transp_1)

LC705_02620 – ABC transporter, ATPase component [proV] (Pfam: ABC_tran; AAA_21; AAA_30; AAA_22; AAA_29; AAA_16; SMC_N; AAA_19; mRNA_decap_C; G-alpha; Zeta_toxin; AAA_24; T2SSE; AAA_18; AAA_28; SbcCD_C; AAA_15; CENP-M; DUF258; AAA_23; AAA_33; AAA)[Osmoprotectant transporter]

LC705_02619 – Glutathione reductase [gshR] (Pfam: Pyr_redox_2; Pyr_redox; Pyr_redox_3; Pyr_redox_dim; FAD_binding_2; NAD_binding_8; DAO; Thi4; 3HCDH_N; HI0933_like; FAD_oxidored; FAD_binding_3; SH3_8; AlaDh_PNT_C; K_oxygenase)

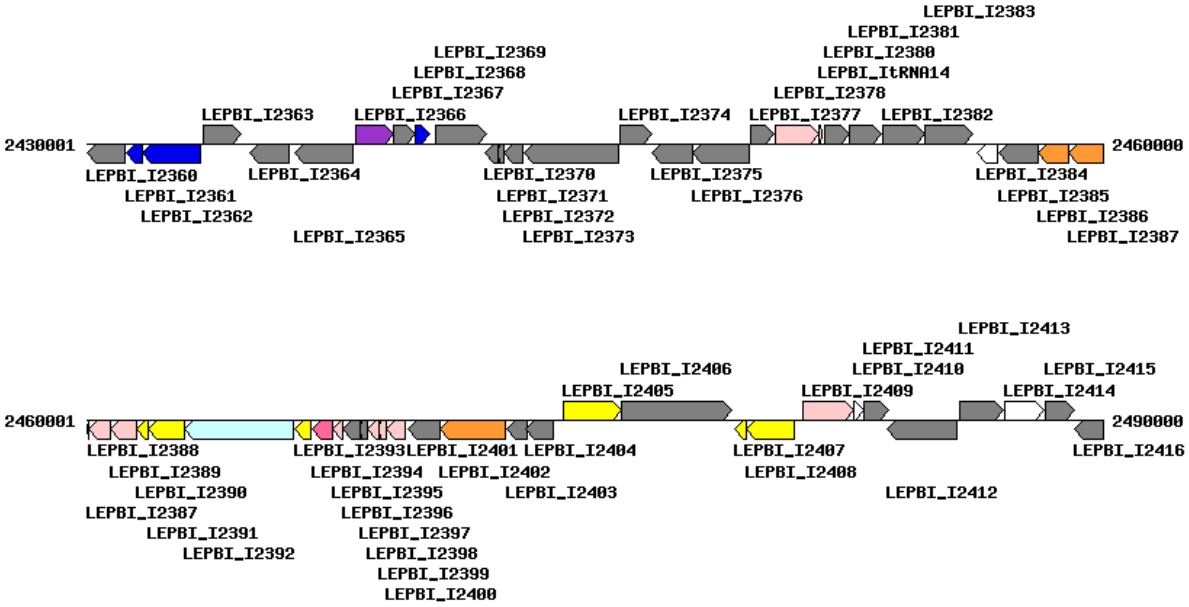
LC705_02618 – Transporter, major facilitator superfamily MFS_1 (Pfam: MFS_1)
LC705_02617 – Tryptophanyl-tRNA synthetase [trpS] (Pfam: tRNA-synt_1b)
LC705_02616 – LemA (Pfam: LemA; DUF327; TPR_21; DUF2613; IncA; DivIC; DUF4363; STT3; UPF0118)
LC705_02615 – Conserved protein (Pfam: TPM_phosphatase; L-fibroin; 60KD_IMP)
LC705_02614 – Na-driven multidrug efflux pump [dinF] (Pfam: MatE; MVIN; Polysacc_synt_C; Glycoprotein_B)

21. *Lactobacillus acetotolerans*



LBAT_0120 – cation transport protein (Pfam: CitMHS)
LBAT_0119 – LemA (Pfam: LemA; DUF3139; ABC2_membrane_3)
LBAT_0118 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; DUF4191; GPI2; Bax1-I)
LBAT_0117 – hypothetical protein (Pfam: SAM35)
LBAT_0116 – hypothetical protein

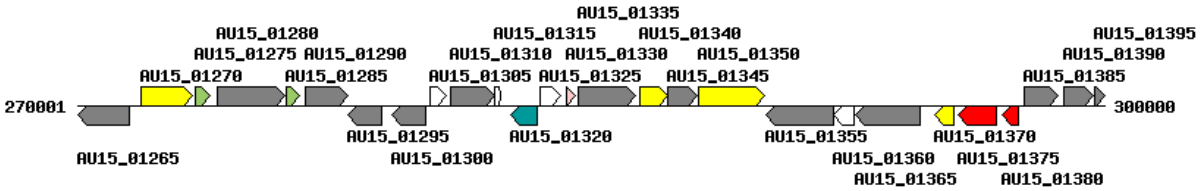
22. *Leptospira biflexa* serovar Patoc Patoc 1



LEPBI_I2404 – putative alpha/beta hydrolase (Pfam: Abhydrolase_6; Hydrolase_4; Abhydrolase_1; Malt_amylase_C; Acyl_transf_2; Ribosomal_S2)
LEPBI_I2403 – putative transcriptional regulator (Pfam: WHG; TetR_N)
LEPBI_I2402 – threonyl-tRNA synthetase (Pfam: tRNA-synt_2b; HGTP_anticodon; tRNA_SAD; TGS)
LEPBI_I2401 – hypothetical protein
LEPBI_I2400 – translation initiation factor IF-3 (Pfam: IF3_C; IF3_N; mIF3)
LEPBI_I2399 – 50S ribosomal protein L35 [rpmI] (Pfam: Ribosomal_L35p)
LEPBI_I2398 – 50S ribosomal protein L20 [rplT] (Pfam: Ribosomal_L20)

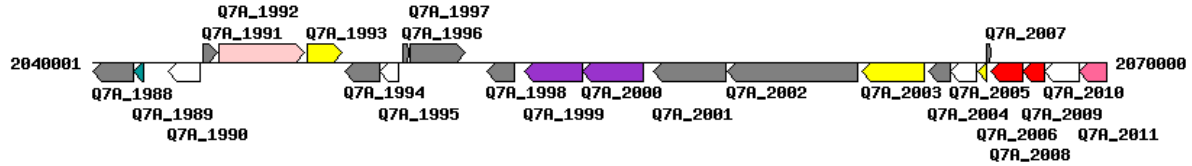
LEPBI_I2397 – hypothetical protein
LEPBI_I2396 – hypothetical protein (Pfam: TolA_bind_tri; Mod_r)
LEPBI_I2395 – cell division protein ZapA (Pfam: ZapA)
LEPBI_I2394 – putative 5-formyltetrahydrofolate cyclo-ligase (Pfam: 5-FTHF_cyc-lig)
LEPBI_I2393 – putative chemotaxis protein CheW-like protein (Pfam: CheW)
LEPBI_I2392 – chemotaxis protein CheA [cheA3] (Pfam: CheW; Hpt; H-kinase_dim; HATPase_c; P2; CheC; CheX)
LEPBI_I2391 – chemotaxis response regulator protein-glutamate methylesterase [cheB3] (Pfam: CheB_methylest; Response_reg)
LEPBI_I2390 – chemotaxis protein CheY [cheY3] (Pfam: Response_reg)
LEPBI_I2389 – chromosome segregation and condensation protein ScpA [scpA] (Pfam: SMC_ScpA)
LEPBI_I2388 – chromosome segregation and condensation protein ScpB [scpB] (Pfam: SMC_ScpB; MarR; HRDC; HTH_IclR)
LEPBI_I2387 – P-protein [pheA] (Pfam: PDT; CM_2; ACT)
LEPBI_I2386 – prephenate dehydrogenase [tyrA] (Pfam: PDH; Semialdhyde_dh)
LEPBI_I2385 – putative signal peptide (Pfam: DUF1343)
LEPBI_I2384 – LemA (Pfam: LemA)

23. *Marinobacter salaries*



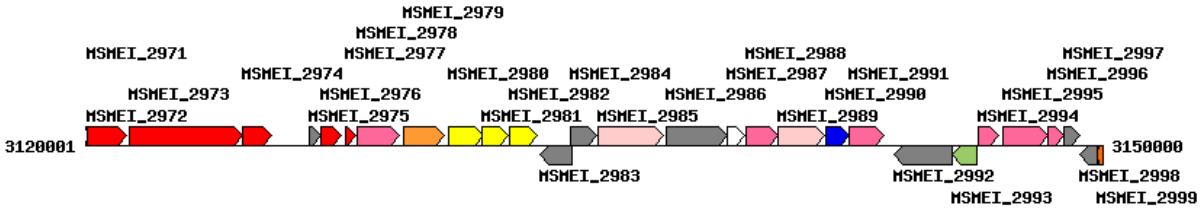
AU15_01365 – pseudogene
AU15_01360 – LemA (Pfam: LemA; DASH_Spc34; Col_cuticle_N; IncA; P12)
AU15_01355 – pseudogene [two ORFs] (peptidase_M48)(heat shock protein HtpX)

24. *Methylophaga nitratireducenticrescens*



Q7A_1995 – LemA (Pfam: LemA; Pro-rich_19)
Q7A_1994 – Zn-dependent protease with chaperone function (Pfam: Peptidase_M48; DUF3357; FtsX; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; SprT-like; Peptidase_A24; Sigma_reg_N)

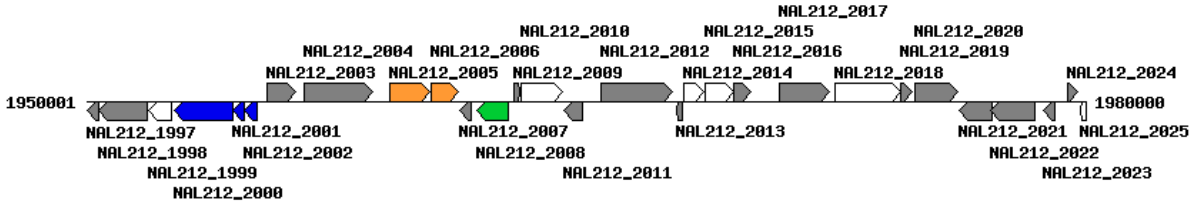
25. *Mycobacterium smegmatis* MC2 155



MSMEI_2984 – YjeF-like protein (Pfam: YhhN; MTTB)
MSMEI_2985 – primosomal protein N priA [priA]
MSMEI_2986 – hypothetical protein (Pfam: DUF2207)
MSMEI_2987 – LemA (Pfam: LemA)
MSMEI_2988 – methionyl-tRNA formyltransferase [fmt] (Pfam: Formyl_trans_N; Formyl_trans_C)
MSMEI_2989 – hypothetical protein (Pfam: Methyltr_RsmB-F; NusB; Methyltransf_25; Methyltransf_11; Methyltransf_15)
MSMEI_2990 – ribulose-phosphate 3-epimerase (Pfam: Ribul_P_3_epim; QRPTase_C; OMPdecase; ThiG; TMP-TENI; IGPS; Staphylcoagulse)

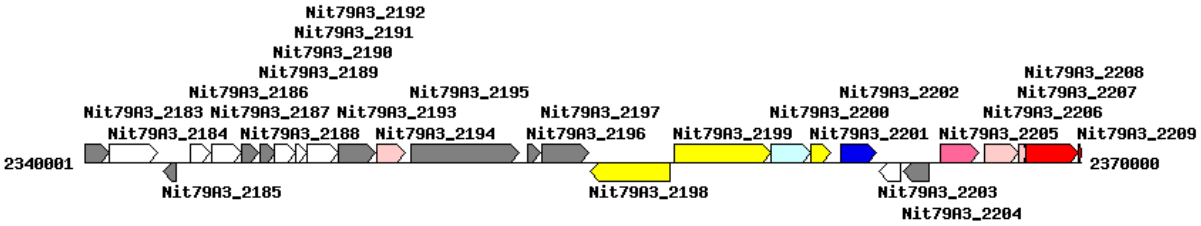
MSMEI_2991 – diaminohydroxyphosphoribosylaminopyrimidine deaminase [ribD] (Pfam: RibD_C; dCMP_cyt_deam_1; MafB19-deam; Ldh_2)

26. *Nitrosomonas sp.* AL212: NAL212_2014 **(Produces similar membranous structures)**

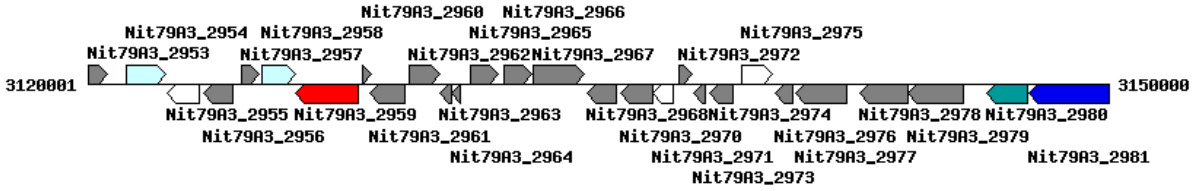


NAL212_2014 – LemA (Pfam: LemA; DUF576; DUF1869)
NAL212_2015 – uncharacterized protein (Pfam: TPM_phosphatase; FUSC_2; MOLO1; Equine_IAV_S2; QLQ)
NAL212_2016 – protein of unknown function DUF477 (Pfam: TPM_phosphatase; DUF3386)

27. *Nitrosomonas* sp. Is79A3 **(Produces similar membranous structures)**

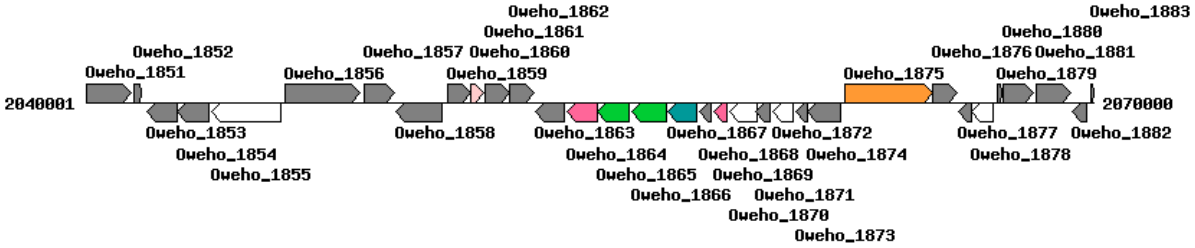


Nit79A3_2186 – LemA (Pfam: LemA; LPAM_2; DUF576; DUF1869)
Nit79A3_2187 – hypothetical protein (Pfam: TPM_phosphatase; DUF4191; MOLO1)
Nit79A3_2188 – hypothetical protein (Pfam: TPM_phosphatase)
Nit79A3_2189 – hydrogenase maturation protease
Nit79A3_2190 – HupE/UreJ protein (Pfam: HupE_UreJ)
Nit79A3_2191 – hydrogenase nickel incorporation protein hypA (Pfam: HypA; DZR; DUF2039; zinc_ribbon_6; zinc-ribbons_6; zf-ribbon_3; DUF4776; tRNA_anti_2; PSD4; zf-ISL3; Ribosomal_L32p; Zn-ribbon_8)
Nit79A3_2192 – hydrogenase accessory protein HypB (Pfam: cobW; MobB; AAA_19; DUF3281; NTPase_1)
Nit79A3_2193 – pseudogene
Nit79A3_2194 – heat shock protein 33 (Pfam: HSP33)
Nit79A3_2195 – hypothetical protein (Pfam: Ctr)
Nit79A3_2196 – response regulator receiver protein (Pfam: Response_reg)
Nit79A3_2197 – integral membrane sensor signal transduction histidine kinase (Pfam: HATPase_c; HATPase_c_2; HATPase_c_3; HisKA)



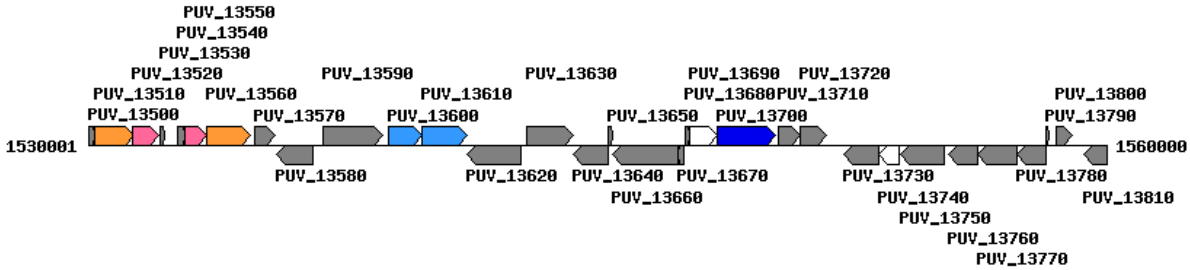
Nit79A3_2971 – LemA (Pfam: LemA; HalX)
Nit79A3_2970 – E3 ubiquitin ligase (Pfam: GIDE; EF-hand_5; EF-hand_1; EF-hand_6; Ribosomal_S4Pg; EF_assoc_2 EF-hand_8; AbiEi_3_N)
Nit79A3_2969 – pseudogene

28. *Owenweeksia hongkongensis*



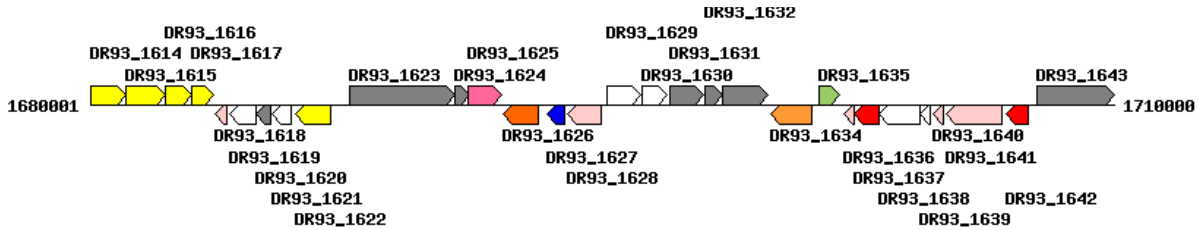
Oweho_1874 – metalloendopeptidase-like membrane protein (Pfam: Peptidase_M23; Amidase_5; Biotin_lipoyl_2)
Oweho_1873 – putative transcriptional regulator (Pfam: MerR_1; MerR; HTH_17; Acetyltransf_11)
Oweho_1872 – LemA (Pfam: LemA; CcmE)
Oweho_1871 – putative membrane protein (Pfam: TPM_phosphatase)
Oweho_1870 – uncharacterized protein (Pfam: TPM_phosphatase)
Oweho_1869 – 6-pyruvoyl-tetrahydropterin synthase (Pfam: PTPS)
Oweho_1868 – S23 ribosomal protein (Pfam: 23S_rRNA_IVP)
Oweho_1867 – malonyl CoA-acyl carrier protein transacylase (Pfam: Acyl_transf_1)
Oweho_1866 – diphosphomevalonate decarboxylase (Pfam: GHMP_kinases_N)
Oweho_1865 – mevalonate kinase (Pfam: GHMP_kinases_C; GHMP_kinases_N)
Oweho_1864 – 4-hydroxybenzoate polyprenyltransferase-like prenyltransferase (Pfam: UbiA; Methyltransf_15)
Oweho_1863 – hypothetical protein

29. *Parachlamydia acanthamoebae*



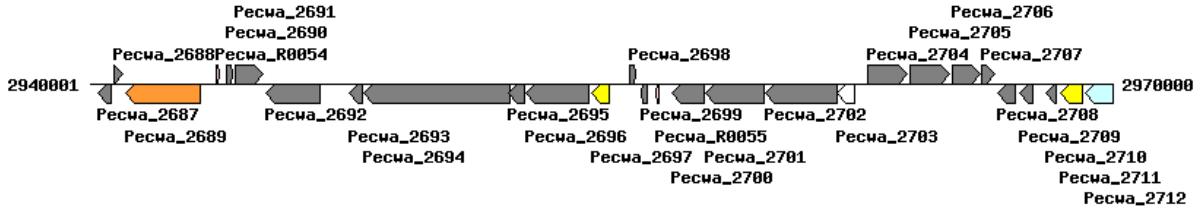
PUV_13780 – hypothetical protein (Pfam: Methyltransf_2; Methyltransf_23)
PUV_13770 – hypothetical protein (Pfam: Glyco_transf_92; Glyco_tranf_2_4)
PUV_13760 – hypothetical protein (Pfam: Glyco_transf_11)
PUV_13750 – hypothetical protein (Pfam: CitMHS)
PUV_13740 – LemA (Pfam: LemA)
PUV_13730 – protease HtpX-like protein [htpX] (Pfam: Peptidase_M48; Yip1[Golgi protein]; Peptidase_M56[Signaling antibiotic resistance in staphylococci])

30. *Pasteurella multocida* ATCC 43137



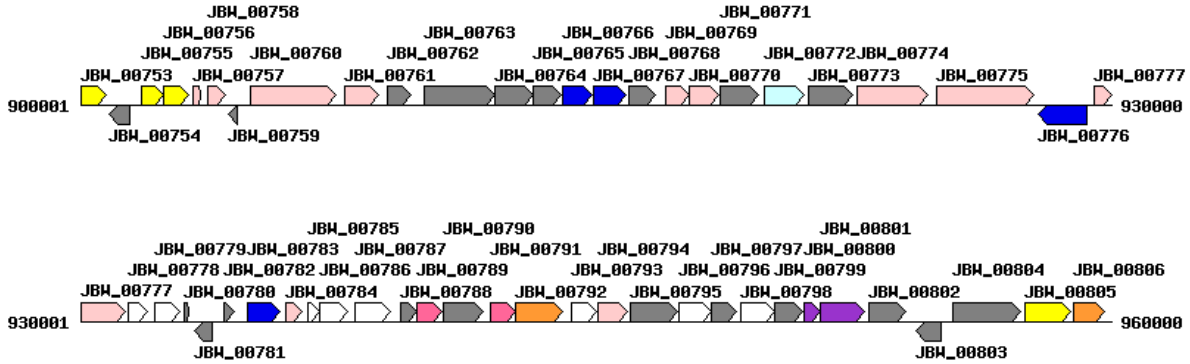
DR93_1622 – OmpA-OmpF porin [ompA] (Pfam: OmpA_membrane; OMP_b-brl; OmpA; BBP2; OprF; Autotransporter)
DR93_1621 – LemA (Pfam: LemA; DUF87; Syntaxin_2; DUF4330; Spore_III_AB; DUF2986)
DR93_1620 – uncharacterized protein (Pfam: TPM_phosphatase; DUF5130)
DR93_1619 – uncharacterized protein (Pfam: TPM_phosphatase; MOLO1)
DR93_1618 – monothiol glutaredoxin (Pfam: Glutaredoxin; GST_N_3; GST_N_2)

31. *Pectobacterium wasabiae*



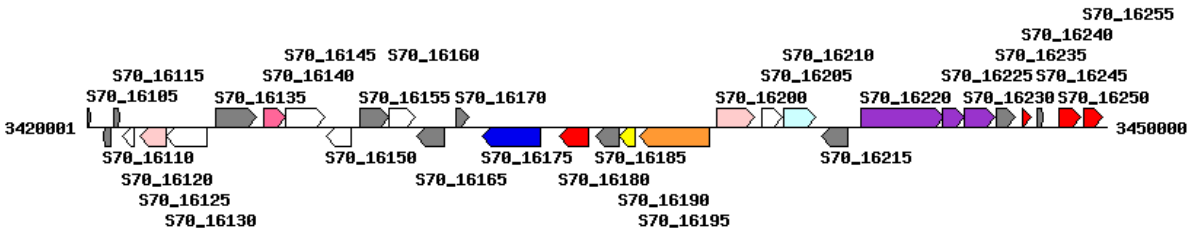
Pecwa_2703 – LemA (Pfam: LemA)
Pecwa_2702 – Protein of unknown function DUF2207 (Pfam: DUF2207)
Pecwa_2701 – conserved hypothetical protein
Pecwa_2700 – D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein (Pfam: 2-Hacid_dh_C; Pyr_redox; Pyr_redox_2; IlvN; ThiF; Shikimate_DH; Oxidored_nitro; FAD_binding_3)

32. *Pelosinus fermentans*



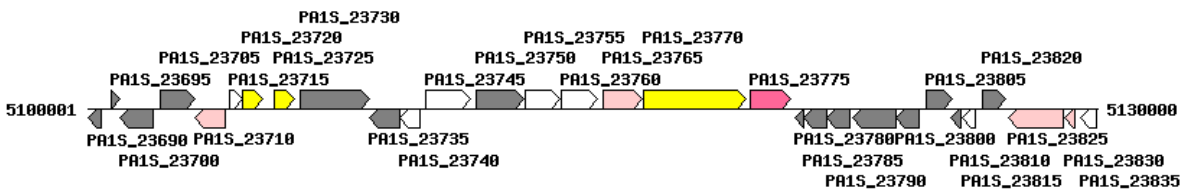
JBW_00777 – UvrABC system protein C (Pfam: UvrC_HhH_N; UVR; GIY-YIG; HHH_2; HHH_5; CUTL)
JBW_00778 – LemA (Pfam: LemA; DivIVA)
JBW_00779 – uncharacterized protein (Pfam: TPM_phosphatase)
JBW_00880 – Rubredoxin domain containing protein (Pfam: Rubredoxin; Vps36-NZF-N; zf-Sec23_Sec24; zinc_ribbon_10; zf-RanBP; PHD_4; ADK_lid)

33. *Providencia stuartii* MRSN 2154



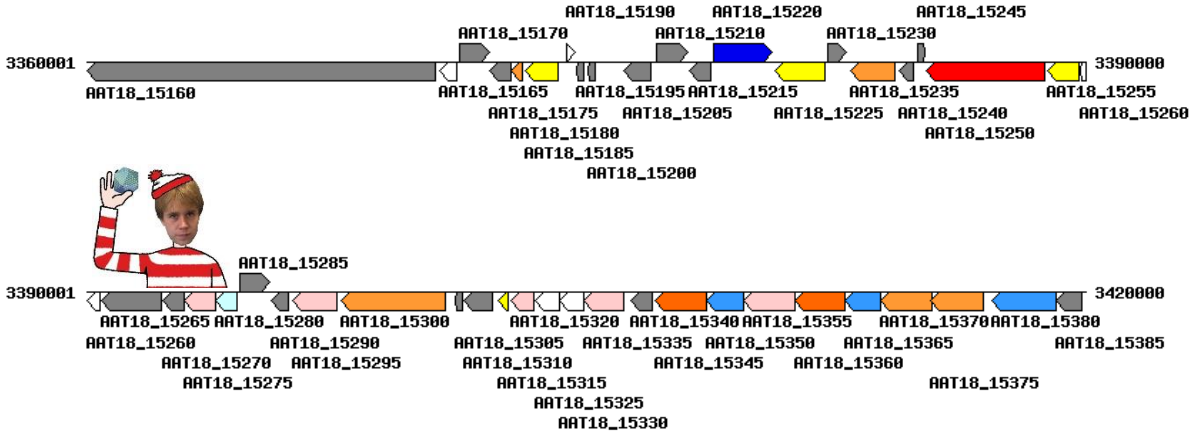
S70_16200 – ATP-binding protein involved in chromosome partitioning (Pfam: ParA; CbiA; AAA_31; Fer4_NifH; MipZ; ArsA_ATPase; AAA_26; CLP1_P; VirC1; CBP_BcsQ; AAA_25; ArgK; SRP54)
S70_16205 – LemA (Pfam: LemA; Anillin_N; DUF4083)
S70_16210 – heat shock protein HtpX (Pfam: Peptidase_M48; DUF4184; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; HisKA_7TM; DUF955; MENTAL)

34. *Pseudomonas aeruginosa* PA1



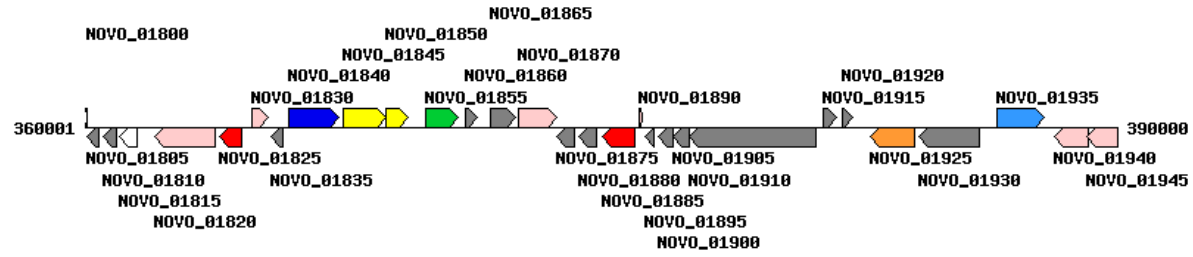
PA1S_23740 – LemA (Pfam: LemA)
PA1S_23735 – hypothetical protein (Pfam: GIDE; EF-hand_1; DUF1771)

35. *Rhodococcus aetherivorans*



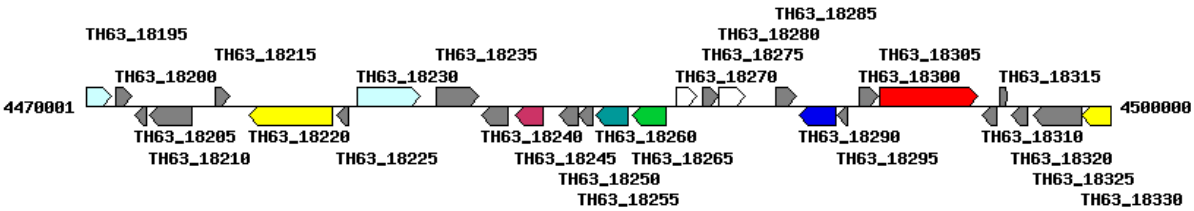
AAT18_15280 – peptidase A8 (Pfam: Peptidase_A8
AAT18_15275 – pseudouridine synthase (Pfam: PseudoU_synth_2; S4)
AAT18_15270 – hypothetical protein
AAT18_15265 – membrane protein precursor (Pfam: DUF2207)
AAT18_15260 – LemA (Pfam: LemA; SP_C-Propep)
AAT18_15255 – protein rarD (Pfam: EamA)
AAT18_15250 – DNA polymerase III subunit alpha (Pfam: DNA_pol3_alpha; PHP; HHH_6; tRNA_anti-codon; DUF655)

36. *Rickettsiales bacterium* Ac37b



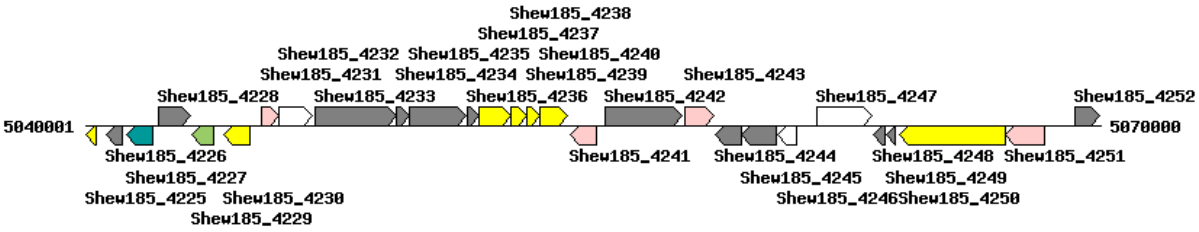
NOVO_01815 – LemA (Pfam: LemA; Yuri_gagarin; HAP1_N)
NOVO_01815 – Transposase
NOVO_01815 – hypothetical protein (Pfam: DDE_3; rve; DDE_1)

37. *Rufibacter* sp. DG31D



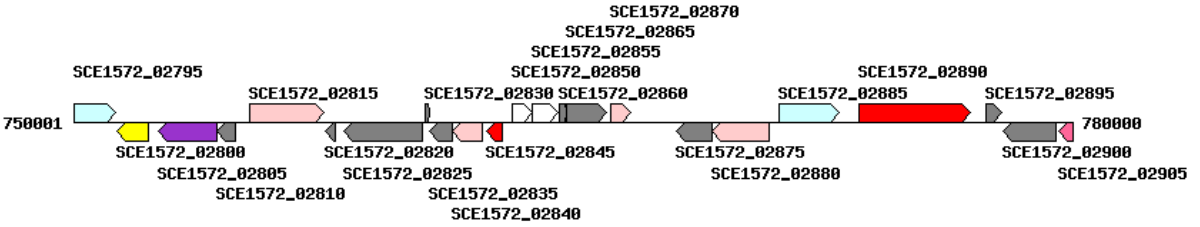
TH63_18270 – LemA (Pfam: LemA)
TH63_18275 – membrane protein (Pfam: TPM_phosphatase; N-glycanase_N)
TH63_18280 – uncharacterized protein (Pfam: TPM_phosphatase)

38. *Shewanella baltica* OS185



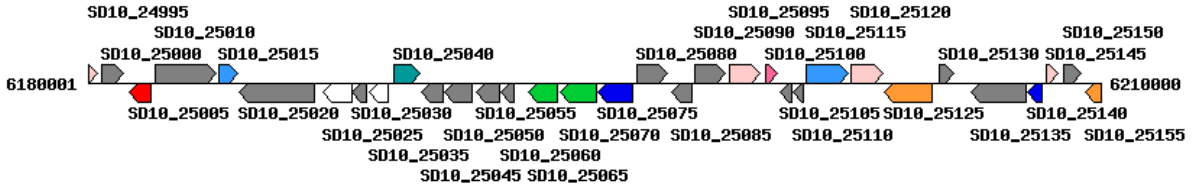
Shew185_4246 – LemA (Pfam: LemA; DUF86; TPR_15; ALIX_LYPXL_bnd; DUF1825)
Shew185_4245 – conserved hypothetical protein (Pfam: DUF3137)
Shew185_4244 – Lytic transglycosylase catalytic (Pfam: SLT; SLT_2; DUF4124; DUF2530)

39. *Sorangium cellulosum* So0157-2



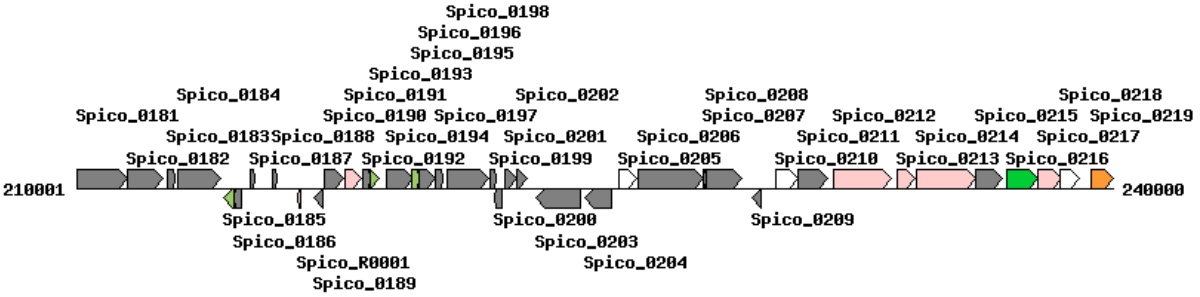
SCE1572_02850 – LemA (Pfam: LemA; BBP1_N)
SCE1572_02855 – hypothetical protein (Pfam: TPM_phosphatase; DUF5130; NHS; RskA)
SCE1572_02860 – hypothetical protein
SCE1572_02865 – acetyl-CoA acetyltransferase (Pfam: Thiolase_N; Thiolase_C; ketoacyl-synt)
SCE1572_02870 – cyclophilin (Pfam: Pro_isomerase)

40. *Spirosoma radiotolerans*



SD10_25035 – LemA (Pfam: LemA; Ribosomal_L36e; Laminin_II)
SD10_25030 – hypothetical protein (Pfam: TPM_phosphatase; Catalase-rel)
SD10_25025 – beta-propeller domains of methanol dehydrogenase (TPM_phosphatase; Cytomega_TRL10; Rifin_STEVOR)
SD10_25020 – hypothetical protein (Pfam: CBM9_1)

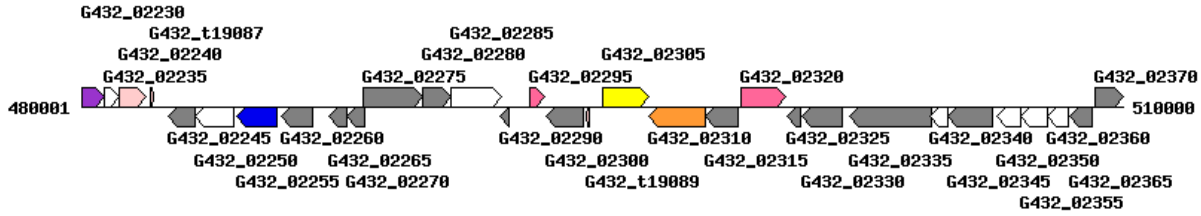
41. *Sphaerochaeta coccoides*



Spico_0205 – LemA (Pfam: LemA; DUF1320; IncA; NusG_II; Complex1_51K)
Spico_0206 – hypothetical protein (Pfam: DUF2207; Acyl_transf_3)

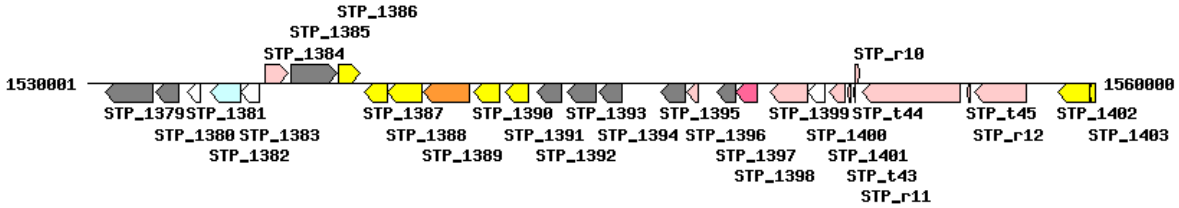
Spico_0207 – hypothetical protein
Spico_0208 – hypothetical protein (Pfam: DUF2804)

42. *Sphingomonas sp.* MM-1



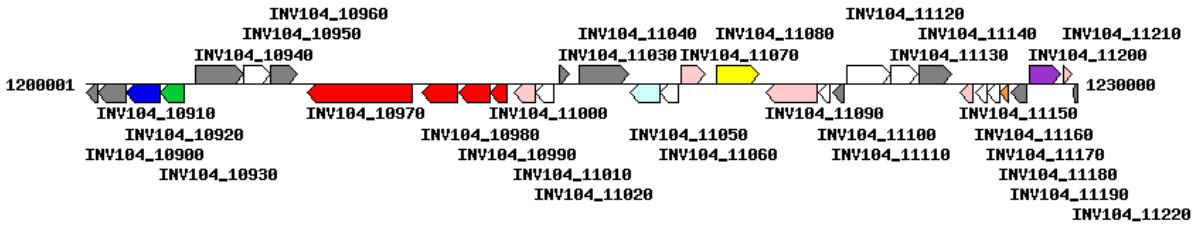
G432_02365 – TetR family transcriptional regulator (Pfam: TetR_N; POTRA_2; HTH_AraC; LacI)
G432_02360 – LemA (Pfam: LemA; EcoEI_R_C; DUF87; EzrA; Laminin_II; DUF1525)
G432_02355 – hypothetical protein (Pfam: TPM_phosphatase)
G432_02350 – hypothetical protein (Pfam: TPM_phosphatase)
G432_02345 – gluconate 2-dehydrogenase (Pfam: Cytochrome_CBB3; Cytochrom_C; Cytochrom_C550; GSu_C4xC__C2xCH)
G432_02340 – putative dehydrogenase subunit (Pfam: Fer2_2; Fer2; Fer2_3; Fer2_4; DHODB_Fe-S_bind)
G432_02335 – aldehyde oxidase and xanthine dehydrogenase molybdopterin binding protein (Pfam: Ald_Xan_dh_C2)
G432_02330 – epocide hydrolase domain-containing protein (Pfam: EHN; Abhydrolase_1)
G432_02325 – hypothetical protein (Pfam: Ribonuc_L-PSP)

43. *Streptococcus parauberis*



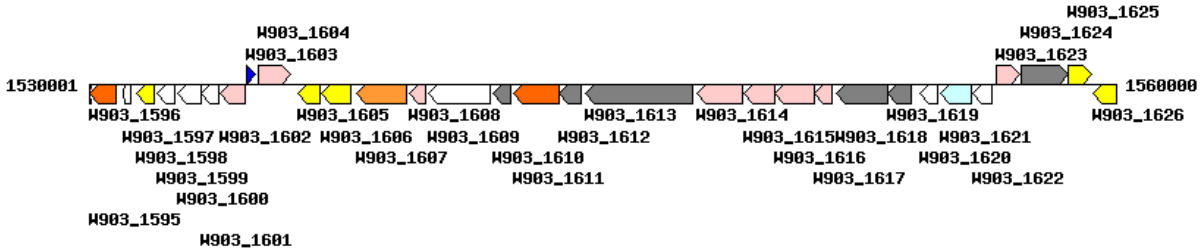
STP_1383 – LemA (Pfam: LemA; Yuri_gagarin; DUF342; Ribosomal_L36e; RuBisCO_large_N; DUF1049; T4SS; Fib_alpha)
STP_1382 – putative protease HtpX-like protein (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; Yip1; DUF799; DUF2162; DUF2964)
STP_1381 – hypothetical protein (Pfam: DUF177)
STP_1380 – response regulator protein (Pfam: Response_reg; Trans_reg_C; QRPTase_C)
STP_1379 – sensor histidine kinase (Pfam: HATPase_c; HisKA; HAMP; HATPase_c_3; HATPase_c_5; DUF3013)

44. *Streptococcus pneumoniae* INV104



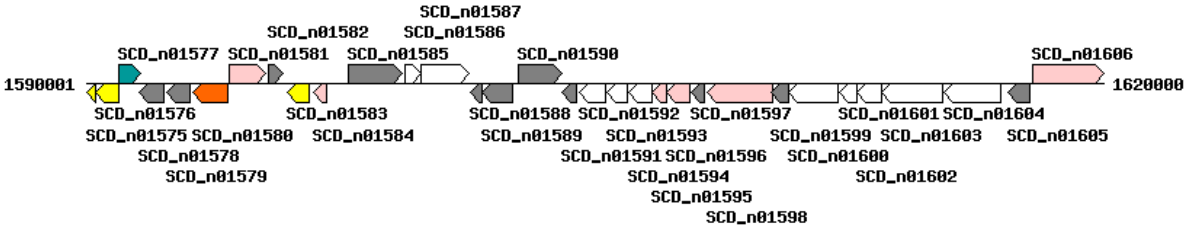
INV104_11060 – LemA (Pfam: LemA; FtsH_ext; P12)
INV104_11050 – putative protease HtpX homolog [htpX] (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; EI24; Reticulon; BPD_transp_2; SecD-TM1)

45. *Streptococcus agalactiae* CNCTC



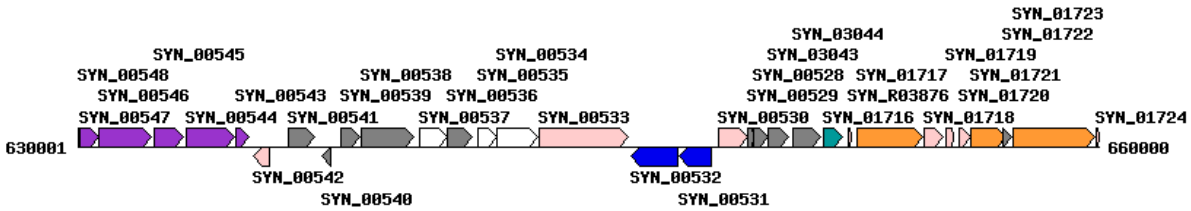
W903_1622 – LemA (Pfam: LemA; RINT1_TIP1; FtsH_ext; pRN1_helical; Ribosomal_L36e; IncA)
W903_1621 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; DUF4191; DUF1218)
W903_1620 – hypothetical protein (Pfam: DUF177)
W903_1619 – putative transcriptional activator CadC (Pfam: Trans_reg_C; Response_reg)
W903_1618 – HAMP domain protein (Pfam: HATPase_c; HisKA; HAMP; HATPase_c_3; HATPase_c_5; bPH_3; YrvL)

46. *Sulfuricella denitrificans*



SCD_n01605 – hypothetical protein (Pfam: Apolipoprotein; ApoLp-III; DUF883)
SCD_n01604 – ubiquinone biosynthesis protein (Pfam: ABC1; APH)
SCD_n01603 – hydrogen dehydrogenase (Pfam: Complex1_51K; 2Fe-2S_thioredx; NADH_4Fe-4S; SLBB)
SCD_n01602 – hydrogen dehydrogenase (Pfam: Fer2_4; Fer4_15; NADH-G_4Fe-4S_3; Fer2; Fer4_13; Fer4; Fer4_7; Fer2_3)
SCD_n01601 – hydrogen dehydrogenase (Pfam: Oxidored_q6)
SCD_n01600 – nickel-dependent hydrogenase (Pfam: NiFeSe_Hases; Complex1_49kDa)
SCD_n01599 – hydrogenase maturation protease
SCD_n01598 – helicase c2 (Pfam: Helicase_C_2; DEAD; ResIII; DEAD_2; Helicase_C; RhoGAP-FF1; PhoH)
SCD_n01597 – hypothetical protein (Pfam: HIRAN)
SCD_n01596 – eptidase M22, glycoprotease (Pfam: Peptidase_M22)
SCD_n01595 – SSU ribosomal protein S18P alanine acetyltransferase (Pfam: Acetyltransf_1; Acetyltransf_10; Acetyltransf_7; FR47; Acetyltransf_4; Acetyltransf_3; Acetyltransf_9)
SCD_n01594 – phage DNA polymerase (Pfam: UDG)
SCD_n01593 – LemA (Pfam: LemA; DUF576; DUF1869; Alanine_zipper; Lipase_bact_N; LPAM_2)
SCD_n01592 – hypothetical protein (Pfam: TPM_phosphatase)
SCD_n01591 – hypothetical protein (Pfam: TPM_phosphatase)

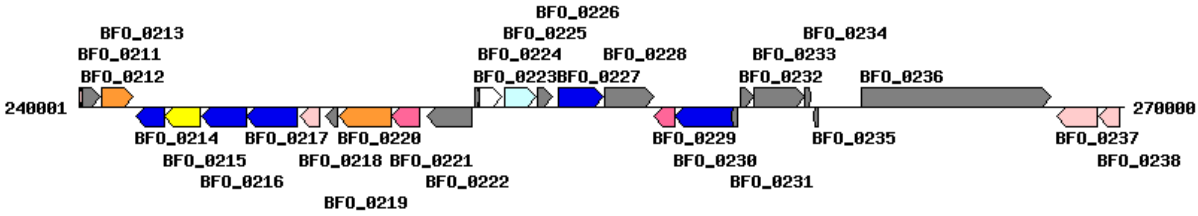
47. *Syntrophus aciditrophicus*



SYN_00539 – phosphohydrolase (Pfam: NUDIX)
SYN_00538 – Fe-S oxidoreductase (Pfam: Radical_SAM; B12-binding)
SYN_00537 – uncharacterized protein (Pfam: TPM_phosphatase)
SYN_00536 – hypothetical cytosolic protein (Pfam: NTP_transf_2)
SYN_00535 – LemA (Pfam: LemA)

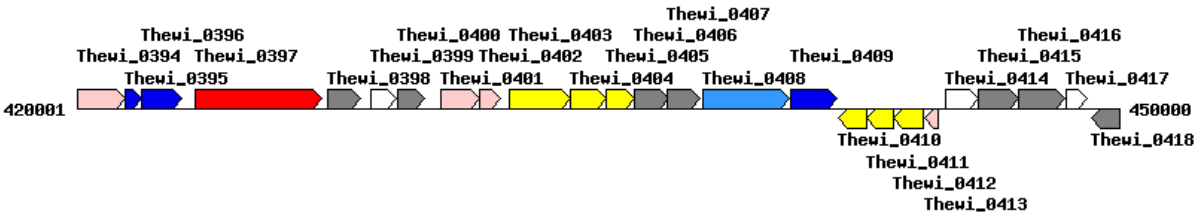
SYN_00534 – phosphoesterase (Pfam: Metallophos; Metallophos_2; AmoA; Metallophos_3)
SYN_00533 – tRNA nucleotidyltransferase (Pfam: PolyA_pol; CBS; PolyA_pol_RNAbd; DHH; DHHA1)

48. *Tannerella forsythia*



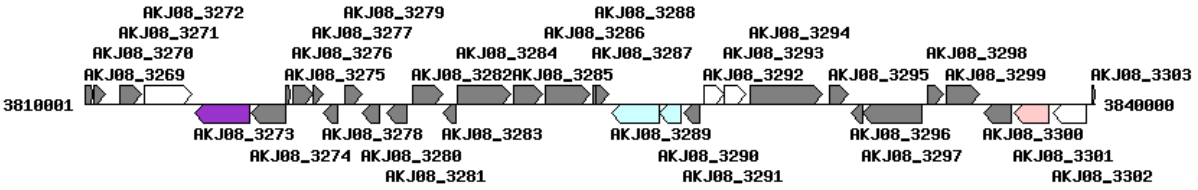
BFO_0223 – hypothetical protein
BFO_0224 – LemA (Pfam: LemA; DUF4083; ABC2_membrane_3; PBC; Vps5; Erp_C; IncA; Wzy_C; DASH_Spc34; Leu_Phe_trans; S1FA; Tmemb_9; DUF4303; DUF4446)
BFO_0225 – heat shock protein HtpX (Pfam: Peptidase_M48; Peptidase_M56 [Signaling antibiotic resistance in staphylococci]; MadM; SprT-like; DUF962)
BFO_0226 – threonine/alanine tRNA ligase (Pfam: tRNA_SAD)
BFO_0227 – phosphoglycerate kinase [pgk] (Pfam: PGK; GST_C)
BFO_0228 – MATE efflux family protein (Pfam: MatE; Polysacc_synt_C; MVIN; Polysacc_synt; UPF0259)

49. *Thermoanaerobacter wiegelii*



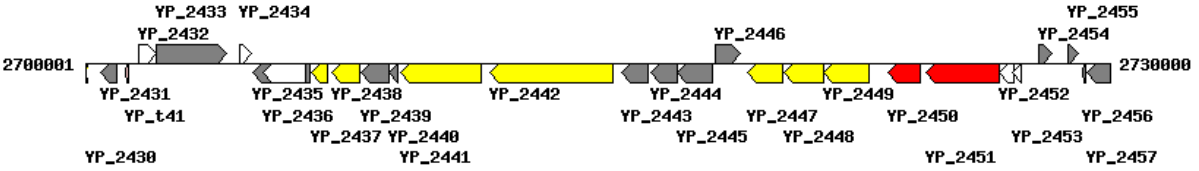
Thewi_0414 – MoxR-like ATPase (Pfam: AAA_3; AAA_5; MCM; AAA; RuvB_N; Sigma54_activat; Mg_chelatase; AAA_18; SKI; AAA_2; Thymidylate_kin; AAA_16; AAA_19)
Thewi_0415 – protein of unknown function (Pfam: DUF58; CstA)
Thewi_0416 – hypothetical protein (Pfam: DUF4129; OAD_gamma; SUR7)
Thewi_0417 – LemA (Pfam: LemA; DUF4837; OmpH; DUF1043; DUF1628; ABC2_membrane_3; Ribosomal_L36e)

50. *Vulgatibacter incomptus*



AKJ08_3292 – LemA (Pfam: LemA; DUF1869; Lipase_bact_N; LPAM_2; LptE)
AKJ08_3293 – uncharacterized protein (Pfam: TPM_phosphatase; HigB-like_toxin)
AKJ08_3294 – hypothetical protein (Pfam: Peptidase_S46; Trypsin_2; Trypsin; Peptidase_S7; DUF31)
AKJ08_3295 – Fatty acid hydroxylase-like protein (Pfam: FA_hydroxylase; Papilloma_E5A)

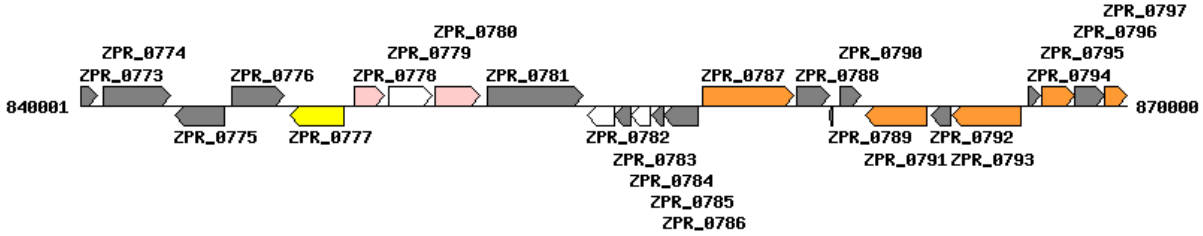
51. *Yersinia pestis* 91001 (biovar Microtus)



YP_2432 – LemA (Pfam: LemA)
YP_2433 – pseudogene
YP_2434 – Uncharacterized Zn-ribbon-containing protein involved in phosphonate metabolism [phnA] (Pfam: PhnA;

PhnA_Zn_Ribbon; zinc-ribbons_6; DNA_RNApol_7kD; UPF0547; TF_Zn_Ribbon; HypA; DUF2318; zinc_ribbon_4; zf-trcl; OrfB_Zn_ribbon; Zn-ribbon_8; zf-ribbon_3; Zn_Tnp_IS1595)

52. *Zunongwangia profunda*



ZPR_0786 – M23 family peptidase (Pfam: Peptidase_M23; Amidase_5; Biotin_lipoyl_2; PTS_EIIA_1)
ZPR_0785 – MerR family transcriptional regulator (Pfam: MerR_1; MerR; Acetyltransf_11)
ZPR_0784 – LemA (Pfam: LemA; NTNH_C)
ZPR_0783 – conserved hypothetical protein (Pfam: TPM_phosphatase; DUF5130; CinA)
ZPR_0782 – uncharacterized protein (Pfam: TPM_phosphatase; DUF4105)