THE UNIVERSITY *of* EDINBURGH

Magnetosome formation in marine vibrio MV-1

Denis Trubitsyn

PhD – The University of Edinburgh – 2010

Declaration

I hereby declare that this thesis was composed by myself and all the work presented in this thesis is my own, except where otherwise stated.

Denis Trubitsyn

Acknowledgements

I am very grateful to my supervisor Dr. Bruce Ward for his kind support and guidance throughout the course of this study. Under his leadership I was able to communicate my individualism and develop my creativity as a scientist.

I would like to thank Professors Vladimir Prokulevich and Yury Fomichev for encouraging me as an undergraduate student to pursue a career in biology.

I am grateful to my family and to all those who have loved and supported me throughout this time.

I would like to thank other members of the institute for their helpful discussions and encouragement.

I am very appreciative to everyone who had involvement in the project and particularly to Dennis Bazylinski for his advice on marine vibrio MV-1 and also Andrew Cronshaw, Juri Rappsilber and Flavia Alves for their help with proteomics.

This project would not be possible without the financial support from The Darwin Trust of Edinburgh and its founder Professor Kenneth Murray.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

aa	Amino acid
ATP	Adenosine-5'-triphosphate
bp	Base pair
BSA	Bovine Serum Albumin
CDF	Cation Diffusion Facilitator
DAP	Diaminopimelate
Da	Dalton
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enchanced Green Fluorescent Protein
IPTG	Isopropyl-β,D-thiogalactoside
IS	Insertion Sequence
kb	Kilobase pair
LC-MS	Liquid Chromatography-Mass Spectrometry
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization-Time Of Flight
MB	Magnetotactic Bacteria
Mb	Megabase pair
MCS	Multiple Cloning Site
NCBI	National Center for Biotechnology Information

ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
RAST	Rapid Annotation using Subsystem Technology
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
TEM	Transmission Electron Microscopy
TPR	Tetratricopeptide Repeat
UV	Ultra Violet
v/v	Volume per volume
w/v	Weight per volume

Abstract

Marine vibrio MV-1 is a magnetotactic bacterium capable of aligning its cell in response to the Earth's magnetic field. This ability is due to the presence of chain-like structures comprising magnetosomes, magnetite particles enclosed in a lipid membrane with associated proteins. Strain MV-1 differs from other, better-characterized strains of magnetotactic bacteria as the cells produce higher amounts of biomagnetite per litre of culture and its magnetosomes are unique in shape.

This study investigates the presence and organisation of a gene cluster termed a "magnetosome island" within the genome of MV-1. In other magnetotactic bacteria this genomic region has been shown to contain many of the genes associated with magnetosome formation but has not been previously investigated for MV-1. One of the conserved fragments of this region was amplified using degenerate primers followed by extension of the known sequence using inverse PCR based technique and constructing plasmid libraries.

Sequencing of the genome of strain MV-1 was accomplished as a part of this study. Significant work was done on comparison of the sequence quality obtained from SOLEXA, 454 and Sanger sequencing technologies. A number of obtained contigs were joined manually and the resulting sequence was automatically annotated using RAST. The obtained genome sequence of 3.6 Mb with a G+C content of 54.3 % was preliminarily analysed and used to search for magnetosome related genes.

This study also analysed proteins associated with the magnetosomes of strain MV-1 using MALDI-TOF, LC-MS and Orbitrap mass spectrometry. These approaches allowed the identification of a number of proteins in the isolated magnetosome

membrane fraction. Some of these proteins have very low similarity with other characterized proteins (either in magnetotactic bacteria or in other organisms). Another significant point is that genes that code for proteins such as MamR, MamK and MmsF were found to be present in several homologous copies within the "magnetosome island" of MV-1. Interestingly, this study shows that all homologous copies of these proteins were identified in the magnetosome membrane fraction.

Generation of knock-out mutants of several specific genes from the "magnetosome island" of strain MV-1 was attempted; constructs were made based on suicide plasmids carrying the *cre-lox* or I-*Sce*I systems. Despite altering numerous experimental conditions it was not possible to obtain conclusive evidence of the isolation of MV-1 transconjugants containing the integrated constructs.

In order to investigate the cell localization of the magnetosome associated protein CAV30779.1, an enhanced green fluorescent protein (EGFP) fusion based construct was generated and transferred into MV-1 cells. The EGFP fluorescent protein fusions within the cells were detected by microscopy.

This study reveals novel information about magnetosome formation in marine vibrio MV-1. The obtained results provide an important foundation for further investigation of this organism and contribute towards broadening the knowledge of the complex process of magnetosome formation in bacteria.

1 Introduction

1.1 Magnetotactic bacteria

The motile bacteria that align and navigate along the Earth's geomagnetic field lines are known as magnetotactic. This behaviour, termed magnetotaxis, is due to a complex phenotype. The presence of intracellular membrane-bound crystals of magnetite (Fe_3O_4) and/or greigite (Fe_3S_4), which are referred as magnetosomes, is obligatory to magnetotactic bacteria.

The original discovery of this group of microorganisms for many years up until recently was believed to be accomplished by Richard P. Blakemore in 1975 and the discovery of magnetotactic bacteria was referred to his original paper in a numerous publications on this group of prokaryotes (Blakemore, 1975). However, in 2007 it became known that magnetotactic microorganisms were originally observed by an Italian researcher, Salvatore Bellini, much earlier in 1958 (Institute of Microbiology, University of Pavia, Italy). Bellini has produced two manuscripts in 1963 describing behaviour of an aquatic bacterium that was swimming in the North Pole direction – the phenomenon he termed "magnetosensitivity". He has also established that this bacterium changed the direction of swimming in response to an artificial magnetic field applied in the laboratory condition and lost this ability when cultured in an iron limited medium (Bellini, 1963a, b). The manuscripts prepared by this researcher were never published in a scientific journal and remained unknown to a majority of a scientific community until translated and examined by Richard Frankel (Frankel, 2009).

The observations made by Richard P. Blakemore in 1975 agree with the data found by Bellini. Blakemore observed populations of coccoid bacteria with the ability to change direction of movement in response to magnetic fields. This movement is based on the presence of a flagella and the presence of an intracellular organelle rich in iron crystals.

There are several common features for all known magnetotactic bacteria (MB). They are Gram-negative bacteria and they produce magnetosomes which allow them to orient in the Earth's magnetic field. They use flagella as a source of motility and they are sensitive to changes in oxygen concentration (Bazylinski and Frankel, 2004).

Described strains of MB vary in many aspects but they are united by the presence of magnetosomes and form a heterogeneous group of prokaryotes though many of them are classified as alphaproteobacteria. The variations in cell morphology are represented by coccoid, rod-shaped, spirilla, vibrio and multicellular microorganisms (Figure 1) (Keim *et al.*, 2004; Schuler, 2002).

MB are common in sediments of freshwater or marine habitats, stratified water columns and wet soils. The occurrence of these bacteria appears to be dependent on the gradients of reduced and oxidized compounds such as sulfur species and oxygen (Frankel and Bazylinski, 2004).

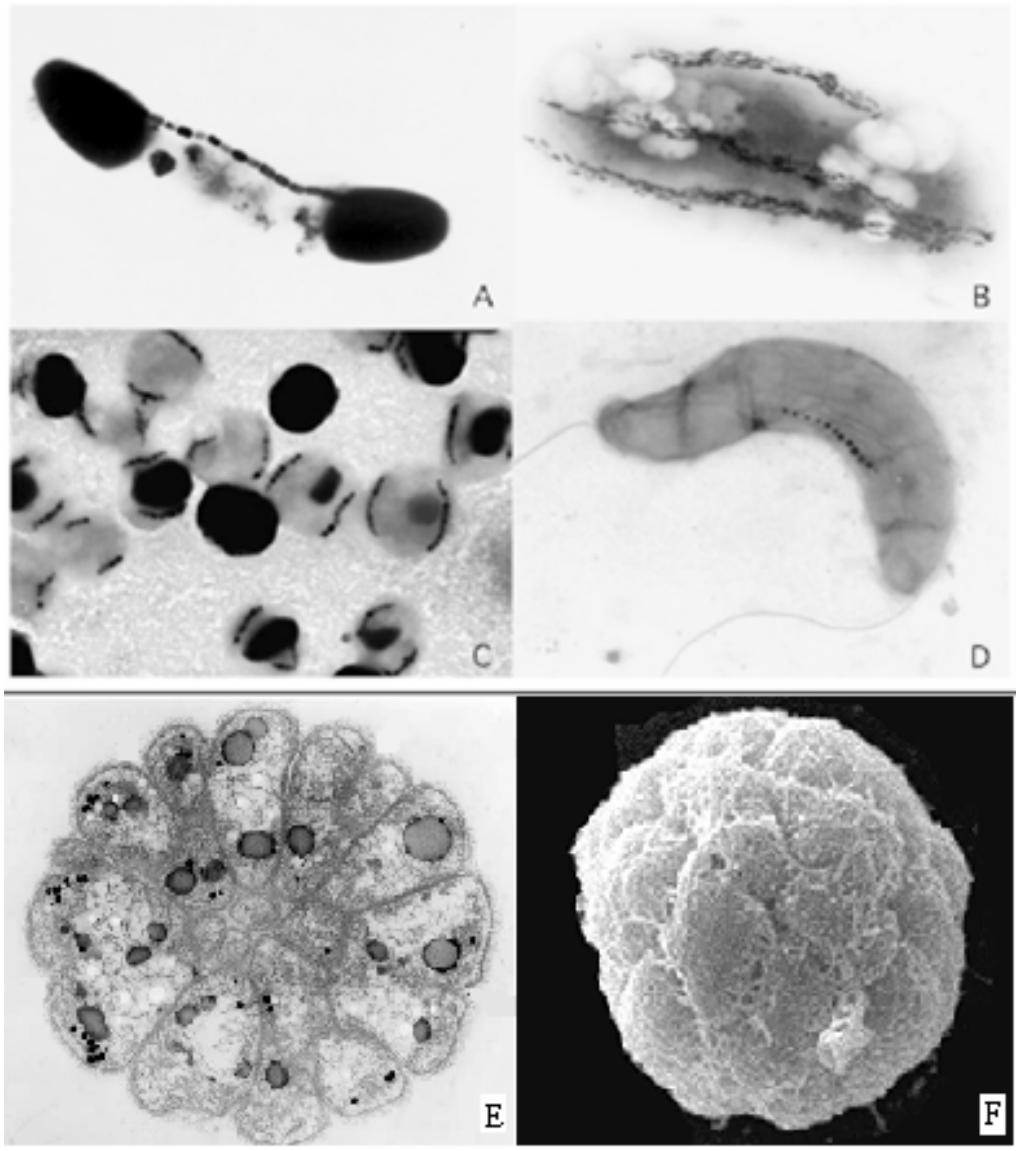


Figure 1. Electron micrographs showing cells of various magnetotactic bacteria and magnetosome crystals. The morphological forms include large rods (A, B), coccoid cells (C), spirilla (D) and multicellular organisms (E, F). These micrographs were adapted from the original publications (Keim *et al.*, 2004; Schuler, 2002).

The research on physiology of MB has shown that there is no evidence of fermentation as a model of energy metabolism. The capability of nitrogen fixation was observed in a number of MB strains. *Magnetospirillum magnetotacticum* can use anaerobic respiration with nitrogen oxides, nitrate or nitrite as a terminal electron acceptor producing nitrous oxide or N_2. The marine vibrio MV-1 has been positively tested for the ability to reduce N_2O to N_2 (Bazylinski *et al.*, 1988). Several strains of MB can oxidize reduced sulfur compounds. The sulfate-reducing strain *Desulfovibrio magneticus* produces sulfide from sulfate. It is interesting that the cultivation of marine strains MV-1, MV-2, *Magnetococcus* MC-1 on media with reduced sulfur compounds allowed investigation of their ability to fixate CO_2 as a source of carbon (Bazylinski and Frankel, 2003).

The analysis of 16S rRNA sequences of several MB led them being classified phylogenetically within the α-subdivision of the Proteobacteria. However, some gregite-producing MB were shown to be associated with the sulfate-reducing bacteria of the δ-subdivision of the Proteobacteria (DeLong *et al.*, 1993)

The presence of magnetosomes helps to isolate MB from the environment. Collected water sediment samples are enriched using permanent or electromagnetic field. Cells move within a capillary towards a magnet making it possible to enrich a sample with cells of MB. An example of such a set up is demonstrated in the next figure (Figure 2).

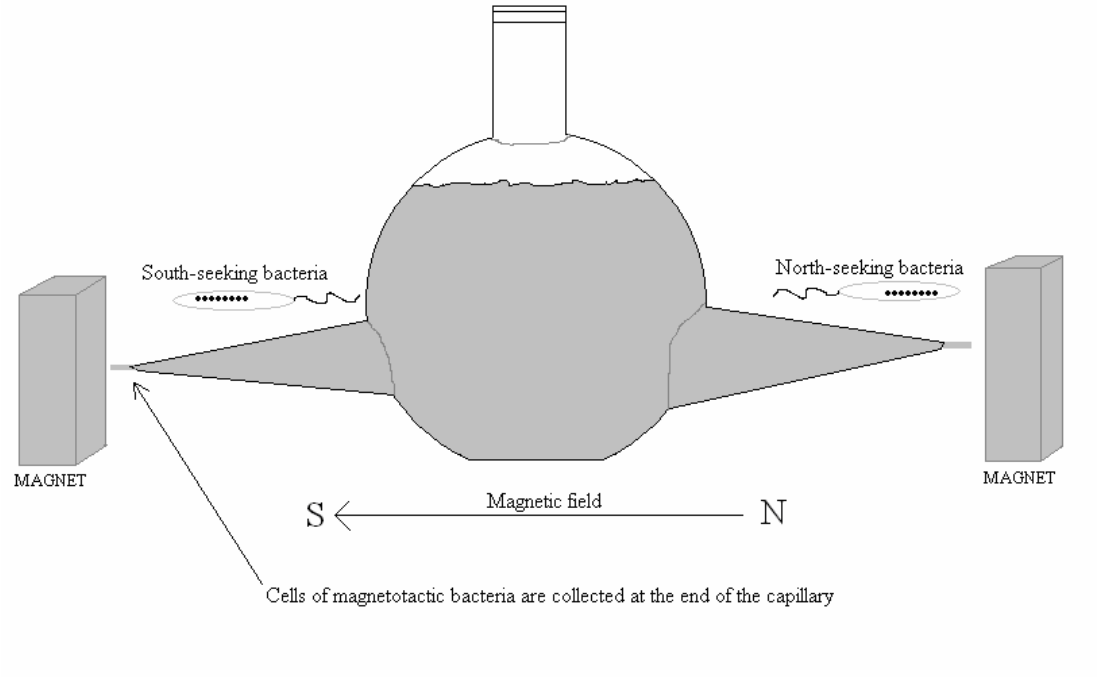


Figure 2. A set-up for enrichment of magnetotactic microorganisms. This schematic diagram demonstrates the principle of the sample enrichment with MB. Strong magnetic field allows separation and collection of the cells containing magnetosomes from the rest of the sample. This diagram was inspired by the photograph in the original publication (Jogler *et al.*, 2009b).

The relatively simple technique of MB sample enrichment allows the isolation of the magnetotactic microorganisms. However in order to carry out most of the research such isolation has to be followed by a development of a method of culturing the isolated strain in the laboratory condition as well as a development of the genetic system in order to carry out further analysis. A recent publication reveals an approach that allowed cloning and sequencing of the fragments of the genome of some uncultivated MB (Jogler *et al.*, 2009b). An uncultured MB can be isolated from the environmental sample as described above in amounts sufficient to extract genomic DNA. These DNA samples are cloned it into a cosmid library followed by a sequencing.

1.2 Magnetosomes

Magnetosomes are unique intracellular structures found in MB. Magnetic mineral crystals in magnetosomes are surrounded by a bilayer of lipids which is 3-4 nm in thickness. The magnetosome membrane is associated with a number of specific transmembrane and some entrapped soluble proteins (Gorby *et al.*, 1988). The magnetosome membrane was previously not thought to be formed as a cytoplasmic membrane invagination, however, more recent studies suggest that magnetosome membranes are formed from the cytoplasmic membrane (Komeili *et al.*, 2006). The proteins that are associated with the magnetosome membrane are described further in section 1.4. Lipid analysis suggest the following data: there are 8 % by weight of neutral lipids and free fatty acids, 30 % of fraction shown to contain glycolipids, sulfolipids and phosphatides and finally 62% of phospholipids which does not appear significantly different from other cell membranes (Gorby *et al.*, 1988).

Chemical analysis of magnetic particles shows that MB usually mineralize either iron oxide crystals, which contain magnetite (Fe_3O_4), or iron sulphide particles which contain crystals of greigite (Fe_3S_4). Other iron sulphide minerals such as mackinawite and a cubic FeS were also found to be produced by MB (Posfai *et al.*, 1998).

The morphology of biologically synthesized magnetic particles is strain specific. The size is almost equal within one strain, however, sizes can range from 35 to 120 nm for different MB (Bazylinski and Frankel, 2004). The localization in the bacterial cell is specific – they usually form a chain or several chains and recent studies suggest that the actin-like protein MamK is involved in this process (Komeili *et al.*, 2006). The role of MamJ in the organization of magnetosomes within the cell has also been reported (Scheffel *et al.*, 2006).

As was shown in cells of magnetotactic vibrio MV-1 magnetosomes are magnetized parallel to the chain axis and localization in chains allows enhancing and combining together the magnetic moments of each particle. This arrangement produces enough main magnetic dipole moment to passively orient the cell parallel to the Earth's magnetic field as a compass needle does (Dunin-Borkowski *et al.*, 1998).

The four main types of crystal morphologies observed so far are shown in figure below (Figure 3). These morphologies include: cuboidal (cubo-octahedral), elongated prismatic (quasi-rectangular), arrowhead-shaped (bullet-shaped) and unusually large elongated prismatic crystals (Bazylinski and Frankel, 2003; Lins et al., 2005).

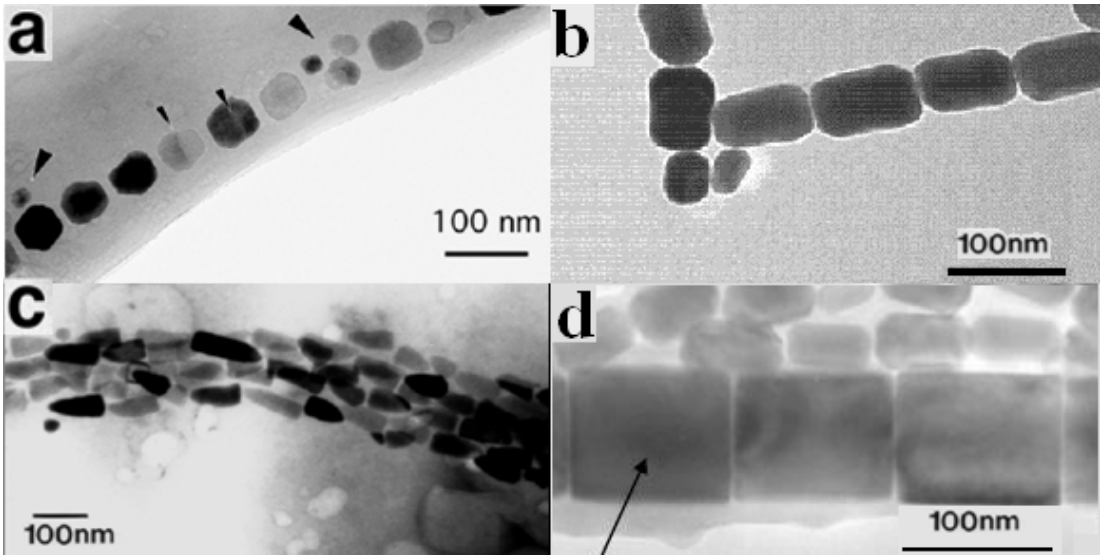


Figure 3. Transmission electron micrographs of magnetite magnetosomes in chains from the cells of various magnetotactic bacteria. (a) Equidimensional cubo-octahedral crystals in *Magnetospirillum magnetotacticum*, small arrows denote twinned crystals and large arrows indicate smaller superpara-magnetic crystals (Bazylinski and Frankel, 2003); (b) Elongated crystals with quasi-rectangular projected shape (two chains of crystals from the marine vibrio strain MV-1; the image was produced in this work); (c) the multiple chains of bullet-shaped anisotropic magnetite crystals within the cell of uncultured magnetotactic bacterium collected from the Pettaquamscutt Estuary (Bazylinski and Frankel, 2003); (d) unusually large magnetosomes and normal size magnetosomes all isolated from Itaipu lagoon in Brazil (Lins *et al.*, 2005).

Although magnetosomes are generally arranged in chains to form magnetic dipole some species of MB do not exhibit such ideal localization. For example several chains of magnetosomes can be formed within one cell of *Magnetobacterium bavaricum* (Spring *et al.*, 1993), whereas some freshwater magnetotactic cocci produce magnetosomes that are not arranged in chains but clustered on the one side of the cell (Cox *et al.*, 2002).

Investigation of the process of magnetosome formation is based not only on scientific interest but also on the potential for biotechnological applications. Biologically synthesized magnetic particles (BSMP) have several advantages compared to those that are produced inorganically. BSMP are uniform, consistent in size and coated with a lipid membrane associated with specific proteins. These features have allowed research on BSMP's applications in different fields.

BSMP can be used in DNA extraction with the modification of membranes with amino-silane. The application of BSMP as drug carriers is based on high ability to disperse in aqueous solutions. Amine groups that are found on BSMP's membrane allow the development of techniques of immobilization of various molecules (e.g. enzymes and antibodies). Genetically modified BSMP which possess fusion proteins on the membrane surface can be applied in immuno/receptor assay. An investigation of an automated method for BSMP's based single nucleotide polymorphism analysis has been reported also. As well as the application for membrane coated BSMP there are potential interest in pure magnetic crystals for magnetic memory production (Matsunaga *et al.*, 2004).

Another potentially interesting application for the magnetosomes may be the development of a method of visualization of the magnetic particles present in the specific areas of the organism (e.g. human body). For example magnetosomes with attached antibodies can be injected in the blood flow and will eventually concentrate in the area of infected or inflamed tissue. These areas can then be visualized using a magnetic resonance measuring instrument. Some preliminary research on such visualization of magnetite particles in organs was accomplished by St. Pierre *et al.* (St. Pierre T. G. *et al.*, 2005).

1.3 The role of magnetosomes

The presence of magnetosomes in bacterial cells is a relatively unique feature of the phenotype. There are several theories that attempt to explain the role of magnetosomes for the microorganism.

1.3.1 Magnetotaxis

The most common and widely accepted explanation is that combined magnetic dipole of the magnetosomes orients and aligns cells to Earth's magnetic field. This alignment allows the movement of the cells towards the most favourable environment – usually within the gradient of oxidized and reduced compounds. The schematic representation of this theory can be found on the diagram below (Figure 4) (Frankel and Bazylinski, 2004).

The model of axial magnetotaxis describes those strains of MB that swim along the magnetic fields with frequent changes of movement direction without changing the cell orientation. This model has been observed for fresh water spirilla. Another model, polar magnetotaxis, which was reported for cocci forms mostly, describes

bacteria that swim in one direction oriented by the magnetic fields (Bazylinski and Frankel, 2004). The model of polar magneto-aerotaxis could be extended to a more complex model named redoxtaxis in habitats in which rapid chemical oxidation of reduced chemical species such as sulfur near the oxic-anoxic transition zone results in separated pools of reductants and oxidants (Spring and Bazylinski, 2000).

One of the disadvantages of magnetotaxis is that the production of magnetosomes appears to require significant amounts of energy and resources and therefore may not appear to be beneficial. However, the comparison of the classic model of chemotaxis and magnetotaxis demonstrates possible overall advantage of the use of magnetosomes for cell orientation. All isolated so far MB were found in aquatic environment. Such environment is a 3D space and therefore an individual cell can move in any direction. The processing of the signals and change of the orientation by reversing the direction of flagella rotation also requires energy and time. It is also important to take into the account that the orientation of the cell after "tumbling" is completely random and might result in some movement in the unfavourable direction until another round of the complex process of reception and change of direction is completed (Berg and Brown, 1972).

In case of magnetotaxis cells are aligned to the single axis and therefore move only in 1D space which makes taxis much more precise and time and energy efficient. Another possible factor that needs to be taken into the account is that in aquatic environment significant movements of the layers of water take place (for example due to waves and wind). Such events can put a cell into a significantly long distance from the favourable oxic-anoxic zone and completely change the orientation of a

nonmagnetic cell. An ability to move whilst being passively aligned to the preferred single axis appears to be beneficial for the survival of MB (Figure 5).

1.3.2 Storage of iron

Another theory suggests that magnetosomes can play a role in iron storage. This may be beneficial for the survival of the microorganisms where the concentration of the iron in the environment is high enough to be toxic and the ability to tolerate high concentrations of iron can be evolutionary beneficial for a population of MB. Another theory is that storage of iron can be used as an energy source, some metabolic intermediate or an electron sink. Even though this theory is not considered as the main model for the role of magnetosomes there is an interesting observation that is worth to point out. A coccid MB isolated from Itaipu lagoon in Brazil contains unusually large magnetosomes that are shown in Figure 3 (d). The magnetic dipole of this microorganism is higher by at least a factor of 10 than that of typical MB. The first explanation of this phenomenon was that the lagoon where this microorganism was isolated is located in the area with abnormally low magnetic field and therefore larger magnetosomes are required to orient cells. However, this theory was not proven to be right because several other MB species were isolated from the same lagoon all containing much smaller magnetosomes. The reason for the synthesis of such unusually larger crystals remains unknown (Lins et al., 2005).

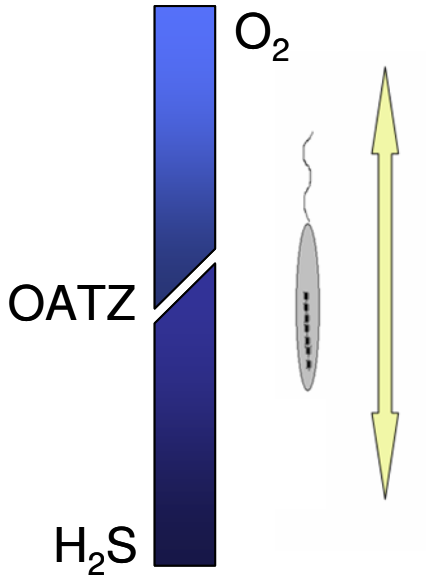


Figure 4. The theory of axial magnetotaxis. This diagram demonstrates a model where magnetotactic bacteria move along the magnetic field depending on the gradients of oxidized and reduced compounds such as oxygen and sulfur species. The favourable conditions are believed to be found in so called "oxic-anoxic transition zone (OATZ)".

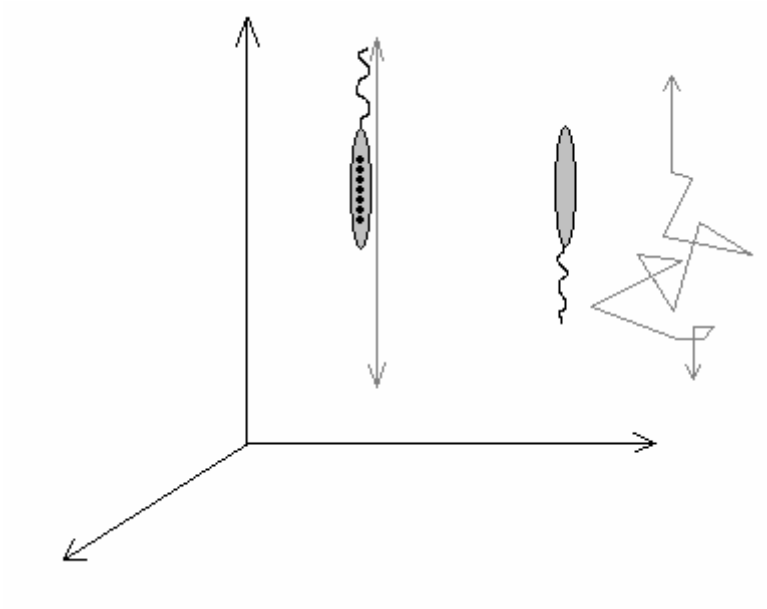


Figure 5. Comparison of magnetotaxis and classical chemotaxis models. This schematic diagram demonstrates the advantage of the cell being aligned to the single axis during the movement.

1.3.3 Magnetotaxis as a beneficial feature for the population

As a speculative theory developed in this work another idea can be proposed. The presence of magnetosomes makes the population of MB unique compared to all other microorganisms present in the specific environment. This feature allows cells of this population to stay close to each other and therefore possibly can help to compete with other organisms. This may also help to change some features of the environment into more favorable. A possible speculative evidence for this idea is the fact that a relatively large amount of inoculum is needed to set up a new culture of MB (for example marine vibrio MV-1) suggesting that only large number of cells can change something in the medium to allow the growth of the culture. An artificially prepared growth medium, however, cannot be directly compared to the conditions of the natural habitat.

1.4 Magnetosome formation

It is important in investigation of the mechanisms of magnetosome formation in MB to describe the steps involved in this process. The full process is not fully understood but the following information shows some specific details of the process. There are four main stages in the process of magnetosome formation: iron uptake by bacterial cell, vesicle formation, iron transport into the vesicle and crystallization of magnetic particle (Bazylinski and Frankel, 2003). Although, the exact order of these steps is poorly understood, a review by Arash Komeili suggests a simple model for this process. Firstly, the biogenesis of the magnetosome compartment occurs by invagination of a cytoplasmic membrane. This may happen concurrently with sorting and targeting of magnetosome membrane proteins to the forming compartment. This is followed by assembly of formed invaginations into a chain-like structure. The final

stage of the proposed model combines all reactions involved in biomineralization of magnetite. This includes iron uptake by the cell, formation of ferrihydrite in the periplasm followed by its transport into the vesicle and partial reduction to form magnetite crystal (Komeili, 2007). The following sections provide some details of the described model.

1.4.1 Iron uptake by MB

Iron uptake seems to occur similarly in MB and in other bacterial cells. The concentration of iron in cells of MB is significantly higher than in species that do not produce magnetosomes and can reach 2-4% of dry weight (Blakemore, 1982). Fe^{2+} is soluble in aquatic solutions and can influx into the cells non-specifically in a slow diffusion like process which was reported for *Magnetospirillum gryphiswaldense.* The iron uptake system in *M. gryphiswaldense* has also a faster energy-dependent transport system for Fe^{3+} that follows Michaelis-Menten kinetics (Schuler and Baeuerlein, 1996). Fe^{3+} is less soluble and can not be uptaken by cell without the help of specific systems. Iron chelators, the molecules that are able to bind and solubilize Fe^{3+} for uptake, named siderophores, have been found to be present in a number of MB. Hydroxamate siderophores were tested positively in *Magnetospirillum magnetotacticum* MS-1 when grown in iron rich medium (20μM) (Paoletti and Blakemore, 1986). The iron sufficient conditions were also found to be obligatory for the production of hydroxamate and catechol siderophores in *Magnetospirillum magneticum* AMB-1 (Calugay *et al.*, 2003). There is no evidence of the presence of siderophores in *M. gryphiswaldense*; however the hydroxamate siderophores were found and preliminary reported in the magnetotactic marine vibrio MV-1 (Bazylinski and Frankel, 2004).

An expression of iron uptake protein FeoB was identified in the cells of *M. magneticum* MS-1 (Taoka *et al.*, 2009). A recent paper shows the significance of a role played by a protein FeoB in uptake of iron by cells of *M. gryphiswaldense* strain MSR-1. The mutant with a deletion of the gene coding for FeoB has shown a reduction in iron uptake. This mutant also formed a lower number of magnetosomes per cell and magnetosomes that were smaller in size compared to those in wild-type cells (Rong *et al.*, 2008).

1.4.2 Magnetosome vesicle formation

The magnetosome vesicle seems to originate from the cytoplasmic membrane (Komeili *et al.*, 2006). The magnetosome vesicle formation in *Magnetospirillum magneticum* AMB-1 involves the protein, known as Mms16. This protein was successfully tested for GTPase activity; moreover its sequence shows the presence of an ATP/GTP binding motif and magnetosome synthesis has been inhibited with addition of the GTPase inhibitor AlF^{4-} It is interesting that small GTPases are involved in priming of vesicles in eukaryotic cells (Okamura *et al.*, 2001).

It has been reported that the homology between MpsA and MpsB proteins found in the magnetosome membrane fraction of *Magnetospirillum magneticum* AMB-1 with the acyl-CoA carboxylase of *Escherichia coli* was identified. This research also refers to the reported data that fatty acyl-CoA stimulates budding of transport vesicles in *E.coli* (Tanaka *et al.*, 2006). All these facts suggest that GTPases and acyl-CoA carboxylase-like proteins are likely to be involved in the process of magnetosome vesicle formation; however there are no further details of this process available.

1.4.3 Iron transport into the magnetosome vesicle

There are several possible models of iron transport into the magnetosome vesicles. The first hypothesis is that iron is transported directly from the periplasm into the formed vesicle as the magnetosome interior is contiguous with the periplasm. Furthermore, there is a thought has been reported that magnetite precursors are found in the periplasm of MB (Ofer *et al.*, 1984). The second, more recent, hypothesis is based on the investigation of the protein MagA associated with the membrane of *Magnetospirillum magneticum* AMB-1. This protein is found to be homologous to proteins exhibiting cation efflux activity, the *E. coli* potassium ion-translocating protein, KefC, and the putative Na^+/H^+-antiporter, NapA, from *Enterococcus hirae*. When this protein was expressed in *E. coli* the ability to accumulate iron in energy-dependant process was shown (Nakamura *et al.*, 1995). The homologues of *magA* gene were found in preliminary analysis to be present in both *M. magnetotacticum* and an unnamed *Magnetococcus* strain MC-1 (Grunberg *et al.*, 2001).

1.4.4 Controlled magnetite biomineralisation

The available data suggests that the size and shape of magnetic particles is controlled by the magnetosome membrane with allocated proteins. The details of this process are not well described. A number of proteins bound to magnetic crystal have been characterized in *M. magneticum* AMB-1. These proteins, known as Mms5, Mms6, Mms7, and Mms13, have similar features in their sequences. Furthermore, the C-terminal regions in these proteins have been shown to contain dense carboxyl and hydroxyl groups that bind ions of iron. Expression of Mms6 in *E. coli* with subsequent purification allowed an important experiment to be done. The presence of this protein was shown to catalyze the formation of magnetite crystals. These crystals

ranged in size from 20 to 30 nm and exhibited morphology similar to that synthesized by *M. magneticum* AMB-1. The crystals synthesized inorganically ranged in size from 1 to 100 nm and were non-homogeneous in shape (Arakaki *et al.*, 2003).

1.4.5 Oxidation and reduction of iron

One of the theories suggests that iron is transported into the magnetosome not from periplasm but from the cytoplasm. In order for Fe^{3+} to pass through the magnetosome membrane it needs to be converted into Fe^{2+}. Iron reduction is thought to be involved in the magnetosome formation process. The active reduction of Fe^{3+} linked to proton translocation has been shown for *M. magnetotacticum* MS-1 twhich was inhibited by the addition of the terminal electron acceptor, sodium azide (Short and Blakemore, 1986).

There is also evidence that the quinone inhibitor, dicumarol, inhibited iron reduction under anaerobic conditions and also reduced growth and magnetite formation in strain AMB-1 (Matsunaga and Tsujimura, 1993). Furthermore, a ferric iron reductase was purified from *M. magnetotacticum* cells and is believed to participate in magnetosome synthesis (Noguchi *et al.*, 1999). Several proteins showing a ferric reductase activity from the cytoplasmic/cytoplasmic membrane fraction have been reported to be purified from cells of *M. gryphiswaldense* strain MSR-1. One ferric reductase was purified in active form and an analysis of its N-terminal sequence showed no homology to other proteins. This might suggest that iron uptake mechanisms in MB may be significantly different to previously described for other bacteria (Xia *et al.*, 2007).

1.4.6 Genes involved in magnetosome formation

In order to understand the process of magnetosome formation it is important to investigate genes that are involved in this complex process. Analysis of the available complete and draft sequences of strains *M. magnetotacticum* AMB-1, MS-1, *M. gryphiswaldense* MSR-1 and *Magnetococcus* MC-1 together with the recent approach of analysis of the genomes of uncultivated strains of MB reveals the presence of the conserved region. This highly conserved region, found within every analysed genome of MB, contains genes involved in magnetosome formation. This region, termed "magnetosome island", was originally discovered in *M. gryphiswaldense* MSR-1 when a deletion of the fragment of the chromosome resulted in the lost of ability to produce magnetosomes (Ullrich *et al.*, 2005). The comparison of the "magnetosome islands" from different MB reveals high levels of similarity between many of the genes; however some genes are highly strain specific and have not been found in either MB or other bacteria. The schematic organization of the magnetosome islands of different MB is shown in Figure 6. It is important to outline that the organization of magnetosome island of marine vibrio MV-1 was not available when similar work was accomplished within this study (see details in section Investigation of the "magnetosome island" in marine vibrio MV-1). In addition, an organisation of the "magnetosome island" in an anaerobic δ-proteobacterium *Desulfovibrio magneticus* strain RS-1 was reported in a recent publication by the Matsunaga research group. The genome sequencing of this strain has revealed the presence of the the cluster of genes flanked by direct repeats and transposases. This cluster contains nine genes conserved with other MB (*mamA, mamB, mamE, mamK, mamQ, mamM, mamO, mamP, mamT*). In addition, 5 ORFs

located on the cryptic plasmid pDMC1 were found to be homologous to the "magnetosome island" of *Magnetococcus* strain MC-1. Interestingly, the obtained sequence lacked several genes that are commonly identified in other MB (*mms6, mamD, mamC, mamF, mamG, mamJ, mamX, mamY*). The absence of these genes may affect the formation of magnetosomes in this strain that are irregular, bullet-shaped (Nakazawa *et al.*, 2009).

Truly remarkable work was carried out by Dorothee Murat *et al.* on the investigation of the specific roles of the genes of "magnetosome island" in *M. magneticum* AMB-1. Generation of deletion mutants led to an investigation into the function of a number of genes that previously had only putatively predicted functions. A "magnetosome island" of this strain that contains 106 ORFs was divided into 14 regions. Generation of deletions for each of these regions revealed that only one of them is crucial for the magnetosome formation. This region contains ORFs: *mamH, mamI, mamE, mamJ, mamK, mamL, mamL, mamM, mamN, mamO, mamP, mamA, mamQ, mamR, mamB, mamS, mamT, mamU* and *mamV*. The only genes with previously genetically analysed function were *mamA, mamJ* and *mamK*. In order to improve knowledge in this area, 14 single non-polar mutations for the remaining genes and one insertion mutation for *mamV* were generated. Double deletions were generated for the homologous copies of *mamR, mamQ* and *mamB* within the chromosome. The method of generation of non-polar deletions was based on the use of SacB mutagenesis. The results of this extensive study showing the specific function of the genes are summarized in Table 1. Additionally, a deletion of *mamV* did not show any significant difference in magnetosome formation to the wild-type cells (Murat *et al.*, 2010).

Figure 6. Organisation of "magnetosome islands" in different MB. This diagram shows schematic organisation of clusters of genes that are believed to be involved in magnetosome formation. MB are shown by the strain names. Two "magnetosome islands" for uncultured MB Fos001 and Fos0002 are also shown. This diagram was adapted from the original publication (Jogler *et al.*, 2009b).

Functions for many of the genes from "magnetosome island" are unknown. However, the role of some of magnetosome related genes have been investigated. Data showing both putative and proven function of such genes is summarized in the table below using the organization of "magnetosome island" of *M. gryphiswaldense* MSR-1 as a model (Table 1). As can be seen from this table there is very little experimental evidence for a specific role of the proteins of this cluster in the magnetosome formation.

The G+C content is significantly distinct between the *mamGFDC* (64.6%), *mms6* (63.8%), *mamAB* (59.5%) clusters and the rest of the chromosome (62.2%) in *M. gryphiswaldense* MSR-1. The presence of IS elements and difference in G+C content between the magnetosome island and the rest of the chromosome in several MB is a possible evidence for their acquisition transition via horizontal gene transfer.

A significant analysis was carried out in order to compare the genes between different MB. The analysis included available partial and completed genomes of *M. magnetotacticum* AMB-1, MS-1, *M. gryphiswaldense* MSR-1 and *Magnetococcus* strain MC-1 (available at National Center for Biotechnology Information, Joint Genome Institute and The Institute for Genomic Research web sites). The authors grouped genes in several classes: specific to "magnetosome island", specific to MB and MB-related (with homologous sequences found in other organisms). The 26 genes are described as MB-specific genes and their homologues were not found in non-magnetotactic bacteria. Eleven of these genes were found to be present in all four strains: *mamD, mamT, mamS, mamF, mamC, mamI, mtxA,* MGR2333,

ORF	Operon	Gene size (bp)	Predicted function	Shown function
mamW		414	Membrane protein	
ORF1	*mms*	693	TPR motive	
ORF2	*mms*	1044	Unknown	
ORF3	*mms*	324	Transmembrane protein	
mms6	*mms*	411	Putative iron-binding protein	Biomineralisation of magnetic particle
mamG	*mam GFDC*	255	Unknown	Size control of magnetic particle
mamF	*mam GFDC*	336	Unknown	
mamD	*mam GFDC*	945	Unknown	
mamC	*mam GFDC*	378	Unknown	
ORF9		837	Putative IdiA	
ORF10		645	Putative transposase	
ORF11		1311	Putative transposase, IS21 family	
ORF12		1528	Hemerythrin-like protein	
ORF13		450	Hemerythrin-like protein	
ORF14		954	Putative transposase	
ORF15		231	Putative transposase	
ORF16		1248	Unknown	
mamH	*mamAB*	1287	Major facilitator superfamily	Not essential
mamI	*mamAB*	234	Unknown	Required for MF and MCF
mamE	*mamAB*	2319	Putative HtrA-like protein	Required for MF
mamJ	*mamAB*	1401	Magnetosome chain assembly	Magnetosome chain assembly

ORF	Operon	Gene size (bp)	Predicted function	Shown function
mamK	*mamAB*	1083	Putative MreB-like protein	Magnetosome chain assembly
mamL	*mamAB*	372	Unknown	Required for MF and MCF
mamM	*mamAB*	957	Putative CDF transporter	Required for MF
mamN	*mamAB*	1314	Unknown	Required for MF
mamO	*mamAB*	1899	Putative HtrA-like protein	Required for MF
mamP	*mamAB*	813	Putative HtrA-like protein	Crystal number and size
mamA	*mamAB*	654	Putative TPR protein	Magnetosome activation
mamQ	*mamAB*	819	Putative LemA-like protein	Required for MF and MCF
mamR	*mamAB*	219	Unknown	Crystal number and size
mamB	*mamAB*	894	Putative CDF transporter	Required for MF and MCF
mamS	*mamAB*	543	Putative membrane protein	Crystal morphology and size
mamT	*mamAB*	525	Putative CytC binding	Crystal growth
mamU	*mamAB*	894	Unknown	Not essential
ORF34		1236	Unknown	
ORF35		762	Putative transposase	
ORF36		315	Putative transposase	
ORF37		849	Putative transposase	
mamZ		1983	Similar to MamH	
mamX		804	Similar to MamE	
mamY		1100	Unknown	Magnetosome membrane tabulation

Table 1. Genes of *M. gryphiswaldense* MSR-1 "magnetosome island". This table summarizes predicted and proven functions in different *Magnetospirillum* of the genes that are believed to be involved in magnetosome formation. MF- Magnetosome Formation, MCF – Magnetosome Compartments Formation. The data was acquired from original publications (Komeili *et al.*, 2004; Komeili *et al.*, 2006; Murat *et al.*, 2010; Richter *et al.*, 2007; Schuebbe *et al.*, 2003; Schuler, 2008; Tanaka *et al.*, 2010).

MGR2349, *mmsF, mamX*. The genes *mtxA*, MGR2333, MGR2349, *mmsF, mamX* are novel candidates for magnetosome formation and they are not located within the 130 kb magnetosome island of *M. gryphiswaldense* MSR-1. Another set of MB-specific genes includes 16 genes that are not found in *Magnetococcus* MC-1 including both previously described (*mms6, mamG, mamR, mamJ, mamL, mamW*) and newly recognized genes (MGR4045, MGR4047, MGR4052, MGR4063, MGR4066, MGR4114, MGR4115, *mamY*, MGR4153, *mmeA*). Finally, a set of genes that are likely to be involved in magnetosome formation were grouped as MB-related because their homologous were found in non-magnetic bacteria. Nine of seventeen MB-related genes were described before and located within "magnetosome island": *mamH, mamK, mamO, mamP, mamQ, mamA, mamB, mamM, mamE* as well as *mamN, mamU* which are described as non-extractable. Eight novel candidates are described as MB-related: MGR0611, MGR1882/1883, MGR4148, MGR3500, MGR0292, MGR0267, MGR0626, MGR1564. Only MGR4148 of this group of novel candidates is located in "magnetosome island". However, almost all of newly described candidate genes for magnetosome formation are organized in clusters (Richter *et al.*, 2007).

Additional data became available with the publication of the organization of "magnetosome island" in marine vibrio MV-1 (Jogler *et al.*, 2009a); however as this research is similar to the one carried out independently in this work it will be described in Investigation of the "magnetosome island" in marine vibrio MV-1 section (Page 83). The functions of these genes are almost unknown and their involvement in magnetosome formation is speculative; however, it makes them good candidates for further research.

1.4.7 Marine vibrio MV-1

Marine vibrio MV-1 (magnetite-containing magnetic vibrio MV-1) is a strain of MB isolated from a salt marsh near Boston, Massachusetts. Dennis Bazylinski (UNLV, Las Vegas) has isolated this strain and investigated its growth in pure culture. Ability to grow and produce magnetite under strictly anaerobic conditions is one of the unique features of this strain. Cells of MV-1 use N_2O as a terminal electron acceptor and produce N_2 when grown in anaerobic conditions or utilize oxygen when cultured microaerophilically (Bazylinski *et al.*, 1988). Cells of MV-1 grow to a relatively high cell density ($A_{595}\approx0.39$) when cultured in a routine laboratory conditions and produce magnetosomes constitutively. The shape of magnetite crystals produced by MV-1 is unique and described as truncated hexa-octahedral magnetite crystals. Better characterized strains of *Magnetospirillum* spp. produce cubo-octahedrons, the form of Fe_3O_4 crystals that is often observed in inorganically synthesized Fe_3O_4 (Dubbels *et al.*, 2004).

The amount of available research data on strain MV-1 was mostly limited to metabolic experiments; isolation of magnetosome membrane fraction and isolation of a spontaneous mutant lacking production of magnetosomes that was shown do be due to the lack of expression of a copper dependent protein involved in iron transport (Dubbels *et al.*, 2004). There was no data on either the presence of "magnetosome island" or identification of the proteins associated with the magnetosome membrane in this strain.

The fact that the knowledge of the magnetosome formation in marine vibrio MV-1 was very limited in addition to unique features of this strain listed above made it an interesting candidate for research in this work.

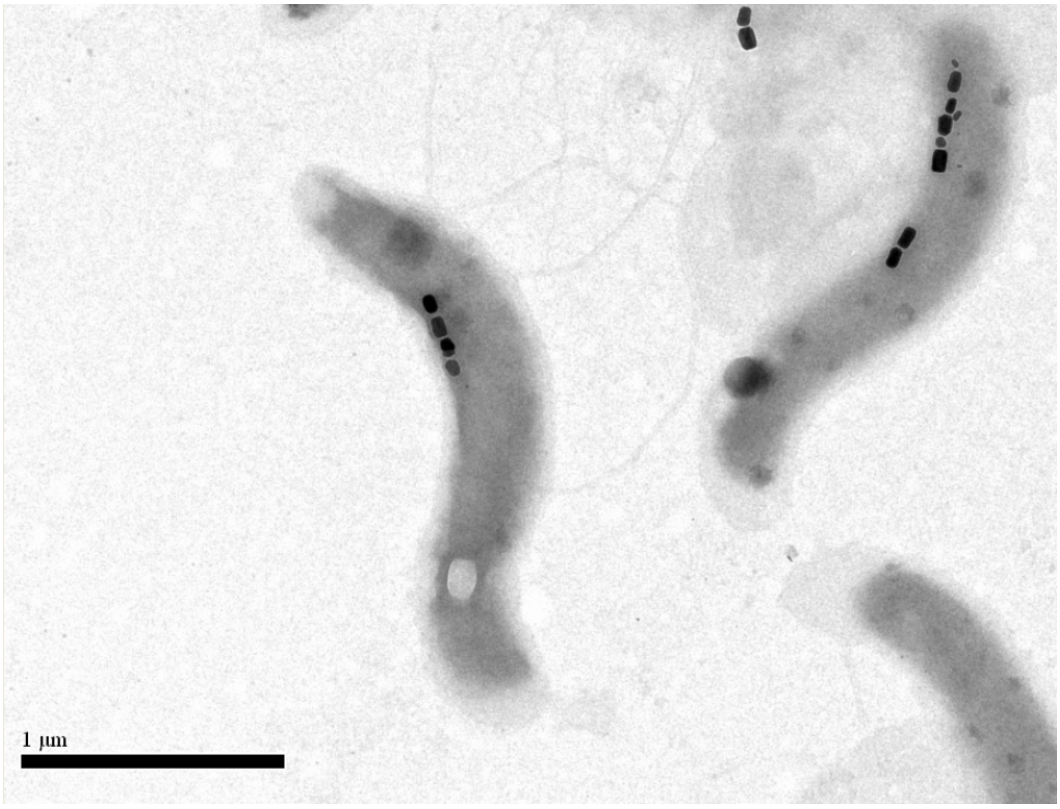


Figure 7. TEM of cells of marine vibrio MV-1. Chains of magnetosomes can be observed within the cells. The bar represents 1 µm. The image was produced in this work.

1.4.8 Research strategies to investigate magnetosome formation

Significant progress has been achieved over the last few years in identification of the genes proposed to be involved in the process of magnetosome formation in different MB. However, investigation of the biological properties and function of the specific genes of this group is still in progress.

There are several major limitations that face researchers who work with MB. MB are difficult to cultivate in laboratory conditions. Only a small number of all isolated MB have been cultivated in the form of pure cultures and they require either anaerobic or microaerophilic conditions. A very recent approach has led to cloning and analysis of chromosomal regions into cosmids of some uncultivated strains in environmental sample. The sufficient amounts of DNA were isolarted from cells collected by the application of magnetic field (Jogler *et al.*, 2009b). Another approach that was reported by Arakaki *et al.* has allowed amplification of the genome of uncultivated MB in sufficient amounts to be used for genome sequencing and other analysis. This approach is based on the multiple displacement amplification using template DNA from just 100 isolated cells of MB (Arakaki *et al.*, 2010).

Genetic systems for MB also need to be developed further to allow a wider analysis.

The analysis of proteins closely associated with the magnetic particle should also help to identify the minimum number of proteins that are required for magnetosome formation. A reverse genetics approach allowed identification of 23 such proteins in *M. gryphiswaldense* MSR-1 and 78 in *M. magneticum* (Jogler and Schuler, 2009). The question remains is what to consider as a protein “closely associated” with the magnetic crystal, because as discussed in section 3.4.1 methods of magnetosome

membrane isolation differ in independent research groups. The magnetosome membrane also appears to contain many proteins that are found in the cytoplasmic membrane of bacteria and therefore may not be specific to the magnetosome membrane (Dubbels *et al.*, 2004). This problem of contamination in the magnetosome membrane with other cell fractions was also reported in the analysis of the magnetosomes isolated from *D. magneticus* RS-1. In this analysis 190 proteins were identified in 41 gel bands of magnetosome membrane fraction. 150 of these proteins were found to be common in *Desulfovibrio* sp. and therefore are not specific to the magnetosome membrane of *D. magneticus* RS-1 (Matsunaga *et al.*, 2009). The development of a standard protocol for isolation of magnetosome membrane fraction appears to be unrealistic but would help to obtain more comparable data for different MB.

The milestone aim for many researchers who work on MB is an ability to produce magnetosomes in heterologous hosts by cloning all required genes. The basis for an alternative approach has been developed in this study that may be used in future work. If the hypothesis that magnetosome formation genes were acquired by not closely related organisms as a result of a horizontal gene transfer is true then the event of such transfer can be selected for in experimental conditions. The set up of an experiment is demonstrated on the diagram below (Figure 8). The idea is that a chosen host organism for example *E. coli* or some closely related *Alphaproteobacterium* is added to a flask together with either cells of a donor MB or its DNA. Large number of cells present in the mixture should be sufficient to compensate the fact that this transfer occurs at low frequency. In case when DNA is used conditions that increase frequency of transformation should be tested

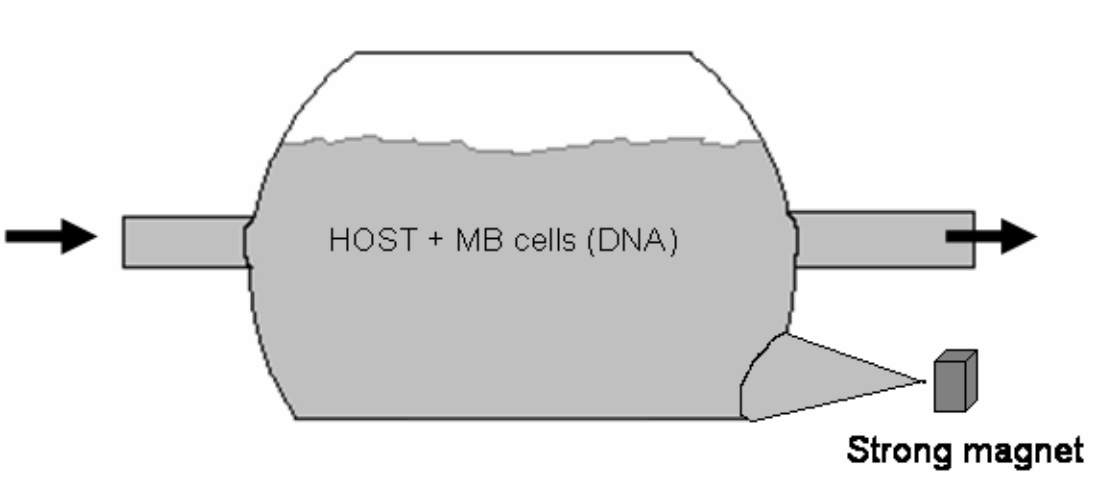


Figure 8. Set up for isolation of heterologous organism that acquired magnetosome formation system. This schematic diagram shows a flask that can be used to mix recipient cells with either cells of donor MB or its DNA. Addition of medium, antibiotics and inoculation can be carried out thorough inlet valve whilst outlet valve can be used to discard old medium containing cells that do not form magnetosomes. Strong magnet is applied to collect magnetite producing recipient cells.

(e.g. competent cells, heat shock). Once the event of transfer and magnetosome formation is considered to be accomplished then cells of MB should be eliminated by addition of an antibiotic that will kill donor cells but will not affect recipient cells. Then the flask should be exposed to strong magnetic field that will collect only recipient cells that contain magnetosomes. The medium containing cells that did not acquire magnetosome formation genes can be flushed through enriching the medium in the flask with magnetotactic cells of the recipient organism. The possible difficulty is that the event of donor MB cells mutating or acquiring antibiotic resistance can happen more often then the event of acquiring of magnetosome formation genes by recipient. In this case it can be suggested to use recipient organism that is resistant to two or more antibiotics or to attempt to use DNA of MB donor rather then living cells. The set up of this experiment allows repetition by adding medium and cells through the inlet and discarding old medium through the outlet aseptically if the frequency of successful transfer does not appear to be sufficiently high to isolate transconjugants or transformants with the amount of cells present in one volume of the flask. Anaerobic or microaerophilic conditions can also be applied if needed. (Jogler and Schuler, 2009)

1.5 Aims of PhD project

The overall aim of this PhD study is to understand more about mechanisms of the magnetosome formation in bacteria using marine vibrio MV-1 as a model organism.

The specific aims of the project were:

Aim 1. To investigate the organization and sequence of the "magnetosome island" in marine vibrio MV-1 and to identify genes the genome that may play a role in magnetosome formation.

Aim 2. To isolate the magnetosome membrane fraction and to identify proteins associated with magnetosomes.

Aim 3. To generate knock-out mutants in order to investigate the role of specific genes from the "magnetosome island" of marine vibrio MV-1.

2 Materials and Methods

2.1 Health and safety

With no exception all reasonable care was taken for the duration of the project to minimize any risk for anyone who could be affected by the processes and materials used in experiments. According to the relevant legislations all work and disposal of live genetically modified as well as wild type organisms and materials was approved before being performed by the departmental board. Each protocol was assessed to avoid any contradiction with the Health and Safety at Work Act and the University of Edinburgh guidelines.

2.2 Bacterial strains and media

The sources of bacterial strains used in this study are shown in table below (Table 2).

2.2.1 *Magnetospirillum gryphiswaldense* medium

Activated charcoal medium contained per litre of distilled water: 2.38 g HEPES, 3 g sodium pyruvate, 0.10 g yeast extract, 3 g soybean peptone, 0.34 g $NaNO_3$, 0.10 g KH_2PO_4, 0.15 g $MgSO_4 \cdot 7H_2O$ and 3 g activated charcoal. The pH was adjusted to 7.0 with NaOH and the medium was sterilized by autoclaving. After autoclaving 500 µM of filter sterilized ferric citrate (500 mM) and 1 mM of filter sterilised 1,4-dithiothreitol (DTT) (1M) were added (Schultheiss and Schuler, 2003).

Inoculated tightly closed flasks were incubated at room temperature over a period of 5 days. Inoculated plates were incubated under microaerobic conditions (5% CO_2, 94% N_2, 1% O_2 - all (v/v)) at 28.4 °C in the MACS VA 500 microaerophilic workstation (DW Scientific).

2.2.2 Marine vibrio MV-1 liquid medium

MV-1 medium contained per 9.6 litres of distilled water: NaCl 157.7 g, $MgCl \cdot 6H_2O$ 33.5 g, Na_2SO_4 26.3 g, $CaCl_2 \cdot 2H_2O$ 3.7 g, KCl 4.5 g, Wolfe's minerals 48 ml, sodium succinate 4.8 g, sodium acetate$\cdot 3H_2O$ 1.92 g, NH_4Cl 2.4 g, casamino acids 4.8 g, 0.2 % (w/v) resazurin 960 µl.

The pH was adjusted to 7.0 with 0.1M NaOH. After autoclaving the medium was bubbled with N_2O overnight. Before subculturing the following filter sterilized solutions were added: phosphate buffer 14.4 ml (see 2.3.2.6), $NaHCO_3$ (0.3 M) 23.2 ml, cysteine (fresh solution) to a final consentration of 1.1 mM (1.92 g in 40 ml of

Strain	Genotype	Source
Magnetospirillum gryphiswaldense	wild type	lab stock
marine magnetotactic vibrio MV-1	wild type	Dennis Bazylinski, UNLV, Las Vegas
Escherichia coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	lab stock
E. coli S-17	TpR SmR recA, thi, pro, hsdR-M^+RP4: 2-Tc:Mu: Km Tn7 λpir	lab stock
E. coli WM3064	: *thrB1*004 *pro thi rpsL hsdS lacZ*_M15 RP4–1360 _(*araBAD)567_dapA*1341::[*erm pir*(wt)]	W. Metcalf, Univ. of Illinois, Urbana

Table 2 Bacterial strains used in this work.

distilled water, pH 6.9-7.0), ferrous sulphate (10 mM) 23.2 ml, vitamin solution 0.55 ml (Dean and Bazylinski, 1999). The composition of the vitamin and Wolfe's minerals solutions are given in the following sections.

Wolfe's minerals solution

Per litre of distilled water were added: $MnCl_2 \cdot 4H_2O$ 0.1 g, $FeSO_4$ 0.3 g, $CoCl_2 \cdot H_2O$ 0.17 g, $ZnSO_4 \cdot 7H_2O$ 0.2 g, $CuCl_2 \cdot 2H_2O$ 0.03 g, $KAl(SO_4) \cdot 12H_2O$ 5 mg, H_3BO_4 5 mg, Na_2MoO_4 0.09 g, $NiSO_4 \cdot 6H_2O$ 0.11 g, $Na_2WO_4 \cdot 2H_2O$ 0.02 g, nitrilotriacetic acid 2.14 g.

Vitamin solution

Per litre of distilled water were added: biotin 2.00 mg, folic acid 2.00 mg, pyridoxine-HCl 10.00 mg, riboflavin 5.00 mg, nicotinic acid 5.00 mg, D-Ca-pantothenate 5.00 mg, vitamin B12 0.10 mg, p-Aminobenzoic acid 5.00 mg, lipoic acid 5.00 mg (Frankel *et al.*, 1997).

2.2.3 Marine vibrio MV-1 agar medium

To produce MV-1 agar medium the following components were added per liter of the liquid medium before autoclaving: 15 g of Bacto Agar, 3 g of activated charcoal, 3 g of peptone, 0.1 g of yeast extract. The amount of iron was increased 10 fold compared to the liquid medium composition. This recipe was adapted by Sabrina Schuebbe (unpublished data) from the method suggested for *Magnetospirillum gryphiswaldense* (Schultheiss and Schuler, 2003).

2.2.4 LB medium

The following solids were suspended in 800 ml of distilled water: 10 g Bacto-Tryptone, 5 g Bacto-Yeast extract and 10 g NaCl. The pH was adjusted to 7.0 with

1M NaOH and the final volume was adjusted to 1 L before sterilization (Bertani, 1951).

2.2.5 SOC medium

A solution containing: 5 g Bacto-Yeast extract 20 g Bacto-Tryptone, 0.5 g NaCl and 5 ml of 500 mM KCl was prepared in 800 ml of distilled water. The pH was adjusted to 7.0 with 1M NaOH and the final volume was adjusted to 970 ml before sterilization. Before use the following sterile solutions were added: 10 ml of 1 M $MgCl_2$ and 20 ml of 1 M glucose (Hanahan, 1983).

2.3 Materials used

2.3.1 Oligonucleotides and plasmids

Oligonucleotide primers used in this project were designed manually and ordered from Invitrogen. The Sigma-Genosys DNA Calculator (Sigma-Aldrich Co.) or Vector NTI (Invitrogen) was used to calculate annealing temperatures, presence of primer dimmers and secondary structures. Oligonucleotide primers are listed in the table below (Table 3).

All plasmids used and generated in this project are listed in the Table 4.

Gene	Direction	Annealing T, ℃	Sequence (5'→3')
Primers for investigation of "magnetosome island"			
mamA	forward	50.5	TCGGTGGCGAAGAAGCTCG
	reverse		ACCTTGCCTTCATTGGGACGCAG
mamM	forward	55	CATGCTGGCCGACGCCATGTATTC
	reverse		CGGGATCGATCCAGGGCAT
mamC	forward	63	GCCGCCGCCCTTGCCAAGAA
	reverse		GTCCCAGGCGTATTTGGCGGCGAC
mamB	forward	50.5	TTCGGMTWCGGMAAYMTCCA
	reverse		CCGATCACCSAGSCGATSAC
mms6	forward	60	ATGGGCGAGATGGARCGCGARG
	reverse		ACGGCTCTTCATATACGCGTAAACCGCCCCGGC
mamK	forward	50	GATCTKGGGACTTCCCRTAC
	reverse		AYGTCDCCAASCTGRYDCC
mamO	forward	48.5	YBTGYNACTTTYGCBTGGGG
	reverse		CCTCCNCCCATAGTCATCATNCCTCC
mamE	forward	55	GATCTCGCCTTGCTGAAR
	reverse		ATTGCTGGGCACGGCRAARCC
mamG	forward	63	GCC CTT GGC GTC GGK GG
	reverse		CTGCGCTTTGCGATTCTTCA
mamP	forward	59.6	GAR GCS CAY TGG CAR GGN ATG GA

Gene	Direction	Annealing T, ℃	Sequence (5'→3')
mamP	reverse	59.6	GGATGGCACTGVGTRCAVGG
mamT	forward	60	TCGGACTGGGACTCTATTGGGA
	reverse		ACAGGCACCTTGACCACAATG
mamJ	forward	65	GAGATTGTTTCGGTGACGGTCCATCCC
	reverse		CCGAAAATTCCACCGAGAAGGTCTTC
ORF12	forward	63	AAGGGCTATAAGGGACCGTTCCGCAGG
	reverse		GCCAACCCCATCAGATTAATCAGACGCTC
mms6 cluster	forward	65	TTTTATCACAGCACGCTCTCCACTC
	reverse		CCTATTCGGCAACCTCATGGTATTGAA
mamGFDC	forward	65	CCTGCGTCGATTTTAAGGGGCA
	reverse		GCCGTGAACCAGCCCAATATTTC
mamN	forward	55	TGGTGACCTACGGGATTTC
	reverse		GCCAGCGCCCACCAAGCC
	reverse 2	55	GAGAATCCCMACCGAGGTC
iPCR	forward	65	AATGGCCTCAGGGACCCGAC
	reverse		TGGGCATCATCGGCGCGCAC
mamP *cre-lox* system	Forward Upstream	64	AGAGCTCGCTCAACAAAAACCAAGAAGAAGTC
	Reverse upstream	75	TTGTGCACCGAGCCGTTTCCCAGCTTTC
	Forward downstream	61	AAGGTACCCCTAAAGGTAACCGTAAGATGATTC

Gene	Direction	Annealing T, ℃	Sequence (5'→3')
mamP cre-lox system	Reverse downstream	69	TTTGTACATTCGTTAGCGCGTTTAAAGTATGCC
ORF2 *cre-lox* system	Forward Upstream	73	AGAGCTCGGAAGGCAATCGTCAACTGATGGC
	Reverse upstream	66	TTGTGCACTTGGCCTACATCAATCAGAACG
	Forward downstream	66	AACCATGGATGCTACAATCCAACCACGTGAG
	Reverse downstream	61	TGAATTCTATCCCCTGTAGCTCTATATCCC
ORF2 I-*secI* system	X Forward Upstream	73	ACTCGAGGGAAGGCAATCGTCAACTGATGGC
	X Reverse upstream	66	TTACTAGTTTGGCCTACATCAATCAGAACG
	Z Forward downstream	61	AAACTAGTATGCTACAATCCAACCACGTGAG
	Z Reverse downstream	71	TTTCGCTCTAGATATCCCCTGTAGCTCTATATCCC
	Forward ORF1	65	GCTCCGTGATCTTGAAAATGG
	Forward within ORF2	65	AGCAGCAATATTTAGCGAACGG
	Reverse ORF3	64	TTTTAGCCATTTTTTACGTTTCCTG
pDT779 EGFP	Forward	70	GGGGTACCATGGCTTATTCGCAAGAGATTTGTCG
	Reverse	70	CCTAGCTAGCAACGGCTTGCAGCGAATGC

Table 3. Oligonucleotide primes designed and used in this work.

Plasmid	Description	Markers	Reference/Source
pBluescript KS+	Used for plasmid libraries	Ampicillin	Stratagene
pBAD24	Used for plasmid libraries	Ampicillin	(Guzman *et al.*, 1995)
pGEM-T Easy	Shuttle vector for PCR products	Ampicillin	Promega
pSAB1	pBBR1MCS 2 + EGFP	Kanamycin	Courtesy of Sabrina Schuebbe (UNLV, Las Vegas)
pCM184	Cre-*lox* based system for mutagenesis	Kanamycin	(Marx and Lidstrom, 2002)
pDTmamTcrelox	pCM184 + fragments upstream and downstream of *mamT*	Kanamycin	This work
pDTORF2crelox	pCM184 + fragments upstream and downstream of ORF2	Kanamycin	This work
pGB909	Suicidal vector with I-secI recognition site	Chloramphenicol	(Patrick *et al.*, 2009)
pDTORF2	pGB909 + fused fragments upstream and downstream of ORF2	Chloramphenicol	This work
pDT779EGFP	pSAB1 + gene coding for protein CAV30779.1 fused with EGFP	Kanamycin	This work

Table 4. Plasmids used in this study.

2.3.2 Stock solutions

The compositions of the simple solutions are listed at appropriate points where the solution is mentioned in the text. The compositions of solutions that require more detailed explanation are listed in this section.

2.3.2.1 0.5 M EDTA

18.61 g of disodium salt of ethylene-diamine-tetraacetic acid was mixed with 80 ml of distilled water. The pH was adjusted to 8.0 using 10M NaOH solution and the total volume increased to 100 ml with distilled water. The solution was sterilized by autoclaving.

2.3.2.2 CTAB/NaCl solution

4.1 g of NaCl was dissolved in 80 ml of distilled water. 10 g of CTAB was added and dissolved by stirring and heating to 65 ºC. The final volume was adjusted to 100 ml with distilled water.

2.3.2.3 20 % (w/v) SDS

A solution of 200 g sodium dodecyl sulphate (SDS) in 800 ml of distilled water was prepared. In order to dissolve crystals completely stirring with heating was used. The final volume was adjusted to 1 L.

2.3.2.4 1M IPTG

A solution of 1.19g Isopropyl-b,D-thiogalactoside (IPTG) in 5 ml of distilled water was prepared. Once sterilized with 0.22 μm filter the solution was stored at -20 ºC. The working solution of 0.1 M IPTG was prepared by diluting 1 M stock.

2.3.2.5 X-gal (20 mg/ml)

A solution of 0.2 g 5-bromo-4-chloro-3-indolyl b-D-galactopyranoside (X-gal) was prepared in 10 ml of dimethyl formamide (DMF). This solution was stored protected from light at -20 ºC.

2.3.2.6 Phosphate buffer

Solutions of 17.42 g K_2HPO_4 (dibasic) in 200 ml of distilled water and 6.81 g KH_2PO_4 (monobasic) were prepared. A 500 ml flask was filled with ≈ 150 ml of dibasic solution and then the pH was adjusted with mononbasic solution until pH 7.0.

2.3.2.7 10X PBS (Phosphate Buffered Saline)

A solution of 80 g NaCl, 2 g KCl, 26.8 g $Na_2HPO_4 \cdot 7H2O$ and 2.4 g KH_2PO_4 in 800 ml of distilled water was prepared. The pH 7.2 was achieved using 1M HCl and the final volume was adjusted to 1 L. The solution was sterilized with autoclaving.

2.3.2.8 TE buffer

A solution of 10 ml of 1 M Tris (pH 8.0) and 2 ml of 0.5 M EDTA (pH 8.0) was prepared in 900 ml of distilled (preferably nuclease free) water was prepared. The final volume was adjusted to 1 L.

2.3.2.9 10X TBE buffer

A solution of 108 g Tris base, 55 g boric acid, 7.4 g EDTA (disodium salt) was prepared in 800 ml of distilled water. The final volume was adjusted to 1 L.

2.3.2.10 6X DNA loading dye

A solution of 25 mg bromophenol blue (0.25%), 25 mg xylene xyanol (0.25%), 4 g sucrose (40%) was prepared to the final volume of 10 ml in distilled water. Where available a commercially produced 6X DNA Loading Dye (Fermentas) was used.

2.3.2.11 Ethidium bromide

A solution of 0.1 g of ethidium bromide in 10 ml of distilled water was prepared and stored in the dark at room temperature.

2.3.2.12 Antibiotic solutions

All antibiotic solutions were prepared either in 1000X or 500X stocks according to the supplier recommendations. The solutions were filter sterilized and stored at -20 ºC.

2.3.2.13 5X SDS electrode (running) buffer

A solution of 45 g Tris base, 216 g glycine and 15 g SDS was prepared in distilled water with final volume adjusted to 3 L. The stock was stored at 4 ºC and diluted with distilled water before use.

2.3.2.14 Coomassie blue staining solution

A solution of 2.5 g Coomassie Brilliant Blue R-250 was mixed with 450 ml of methanol, 100 ml acetic acid and 400 ml distilled water was prepared. After the final volume was adjusted to 1 L any undissolved particles were removed by passing the solution through a Whatman (No. 1) filter paper.

2.3.2.15 Coomassie blue destaining solution

A solution of 140 ml of acetic acid and 100 ml methanol was prepared in 1 L of distilled water. The final volume was adjusted to 2 L and the solution was stored at room temperature.

2.3.2.16 CHCA matrix solution

A solution containing 25 mg α-Cyano-4-hydroxycinnamic acid (CHCA) in 1 ml of 50 % acetonitrile and 0.1 % TFA in proteomics grade water was prepared. If some

crystals were not dissolved a mixture was briefly centrifuged to collect them at the bottom of the tube. The solution was stored at 4 ºC.

2.4 Methods

2.4.1 Strain storage

Magnetospirillum gryphiswaldense and marine vibrio MV-1 were stored in 20% (v/v) glycerol in the liquid medium at -80 ºC. *Escherichia coli* strain DH5α was stored at -80 ºC in 15% glycerol and was recovered by streaking a sample on LB agar and incubating overnight at 37 °C.

2.4.2 *E. coli* strains growth

All *E. coli* strains were grown at 37 ºC in LB with shaking at 200 rpm (New Brunswick Scientific, G45 shaker) or on LB agar plates. The total volume of the flasks for liquid culture was at least 4 times greater than the volume of medium to allow sufficient aeration. Antibiotics and other supplements were added where necessary.

2.4.3 MV-1 growth in liquid culture

For routine subculturing in large volumes 250 ml of inoculum were added to the 10 L growth medium. N_2O was added to a pressure of 1 bar. Culturing in smaller volumes was carried out when necessary. The method of cultivation is discussed in Chapter 3.1 in more detail.

2.4.4 MV-1 growth on agar plates

MV-1 agar medium was poured onto plates and solidified in aerobic conditions. MV-1 inoculum was plated with spreader and then plates were immediately transferred into anaerobic jars. Air was substituted either with N_2O or with 2% oxygen in nitrogen and plates were incubated at room temperature for 5-8 days. The method of cultivation is discussed in Chapter 3.1 in more detail.

2.4.5 Selective enrichment of MB (Race track)

A magnetic method was used to enrich magnetic bacteria from mixed cultures or environment samples. A sterile glass Pasteur pipette was sealed in the gas flame from the narrow side. Then capillary was filled with sterilized medium using a long hypodermic needle. Finally inoculated medium was added and the pipette was sealed with the cotton plug.

A permanent magnet was used to produce a magnetic field oriented along the capillary over a period of 45-90 min. The end of the capillary was aseptically cut and accumulated magnetic cells were transferred to the sterilized medium (Flies *et al.*, 2005).

2.4.6 TEM microscopy

To visualize cells and magnetosome crystals TEM microscopy was used. Bacterial culture was harvested by centrifugation and then washed twice in 1X PBS, followed by a wash in 0.5X PBS and finally once in distilled water with centrifugation steps after each washing. Samples were spotted onto Formvar carbon-coated copper grids and air dried for 20 minutes. Optionally, after 10min samples were dabbed dry with filter paper. Visualization on TEM was carried out using a Philips CM120 BioTwin transmission electron microscope (accelerating voltage: 70kV) in the SBS Electron microscope facility.

2.4.7 Fluorescent microscopy

Bacterial cells were resuspended in the liquid medium containing 2% (w/v) of Low Melting Point Agarose (Gibco BRL). 30 µl of the suspension was applied to the microscope slides and cover glasses were placed on the top. Slides were left to set to

immobilize and then were examined using Axioplan 2 microscope (Zeiss) with ProScan filter wheel (Prior). Images were processed using MetaMorph imagine software.

2.4.8 Plasmid DNA isolation

Plasmid DNA was isolated from 5 ml of overnight bacterial culture using the standard protocol for Wizard® *Plus* SV Minipreps DNA Purification System (Promega) or with a QIAprep Spin miniprep kit (Qiagen). Purified DNA was eluted in 50 µl of distilled water or TE.

2.4.9 Small scale genomic DNA isolation

Genomic DNA from *Magnetospirillum gryphiswaldense* (20 ml of liquid culture A_{600}=0.123) and magnetotactic vibrio MV-1 (20 ml of liquid culture A_{600}=0.222) was isolated with AquaPure Genomic DNA isolation kit (BIO-RAD) according to the protocol for Gram-negative bacteria.

2.4.10 Large scale genomic DNA isolation

Genomic DNA from magnetotactic vibrio MV-1 for the purposes of genome sequencing and plasmid library preparation was isolated by the following method:

0.2 g of wet weight of cells MV-1, obtained by harvesting of 400 ml of liquid culture of strain MV-1 was resuspended in 5.7 ml of TE. Then 0.3 ml of 10% (w/v) SDS and 30 µl of proteinase K (20 mg/ml) were added and thoroughly mixed with the solution. The resultant solution was incubated for 1 h at 37 °C, then 1 ml of 5M NaCl was added and mixed thoroughly. The suspension was mixed with 0.8 ml of CTAB/NaCl and incubated for 10 min at 65 °C. After this 7-8 ml of chloroform/isoamyl alcohol (24:1) was added and the suspension was mixed gently. Then the suspension was centrifuged at 13 000 g for 5 min and the upper fraction

was collected in a fresh tube. An equal amount of mixture of phenol/chloroform/isoamyl alcohol (25:24:1) was added and mixed gently followed by centrifugation for 5 min at 13 000 g. The upper fraction was again collected and treated with phenol/chloroform/isoamyl alcohol mixture as above. Afterwards, the upper fraction was mixed with 0.6 volumes of ice cold isopropanol and incubated for 30 min at - 20 °C. After incubation DNA was pelleted by centrifugation for 5 min at 13 000 g and washed by addition of 5 ml of 70% (v/v) ethanol followed by centrifugation for 5 min at 13 000 g. Finally the pellet containing DNA was dried at room temperature and then dissolved in distilled water and stored at 4 °C. This method was adapted by scaling up 10 times all reagents from the original method (Wilson, 1997).

2.4.11 DNA isolation for PCR screening

A colony was resuspended in 200 µl of distilled H_2O and then centrifuged at 13 000 g for 3 min. The supernatant was discarded and the pellet was resuspended in 100 µl of distilled H_2O and then boiled for 5 min. 1-2 µl of the obtained crude solution containing DNA were used as a template for PCR screening.

2.4.12 Estimation of DNA concentration and purity

In order to obtain a rough estimation of the DNA concentration the sample was run on an agarose gel in several dilutions against a specified amount of DNA marker (1 kb DNA ladder, New England Biolabs). Once the band with the dilution with the brightness similar to the band of the appropriate size on the marker of the band was chosen the concentration of the DNA was calculated.

For more accurate DNA concentration and purity estimation 2 µl of sample was loaded onto the NanoDrop 1000 spectrophotometer (Thermo Scientific).

2.4.13 Agarose gel electrophoresis

DNA was resolved using a 0.8-1% (w/v) agarose gel in 0.5X or 1X TBE buffer. Ethidium bromide (Sigma) was included to a final concentration of 0.5 μg/ml to allow visualization. The gel (15 cm x 15 cm x 0.8 cm) was run in an electrophoresis chamber at 90-130V for up to 1.5 h. DNA was visualized using a UV light box (300-360 nm) and photographed using a digital camera (UVP Laboratory products).

2.4.14 Pulsed Field Gel Electrophoresis (PFGE)

In order to produce genomic DNA preparation in agarose plugs pelleted by centrifugation cells were resuspended in buffer containing 100 mM Tris, 100 mM EDTA, pH 8.0 such that A_{600} was adjusted to approximately 0.7. The suspension was diluted with an equal volume of 2% (w/v) Clean Cut Agarose (BioRad) and the mixture was left to set in plastic moulds. Cells were lysed by transferring plugs into the lysozyme buffer (10 mM Tris pH 7.2, 0.5% (w/v) sodium lauryl sarcosine, 50 mM NaCl, 0.2% (w/v) sodium deoxycholate,) containing 1 mg/ml lysozyme and were incubated for 45 minutes at 37 °C. After removing the buffer plugs were washed by adding 20 ml of washing buffer (20 mM Tris pH 8.0, 50 mM EDTA pH 8.0) for at least 30 min at room temperature. The plugs then were incubated in the buffer containing 0.4 mg/ml proteinase K (50 mM Tris pH 8.0, 50 mM EDTA pH 8.0, 1% (w/v) sodium sarcosine) for 18 hours at 50 °C. After the buffer was removed residual proteinase K was inactivated by addition of 10 ml of TE buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF). At the final stage plugs were washed with 20 ml washing buffer twice for 30 min at room temperature and then stored at 4 °C.

Gel electrophoresis was performed using 1% (w/v) Pulsed Field Certified Agarose (BioRad) gels in 1X TBE buffer. Agarose plugs containing sample DNA were loaded into wells and sealed with molten 1 % (w/v) low melting point agarose. The Contour-clamped Homogenous Electric Field electrophoresis system (CHEF-DRII, Bio-Rad) was used at the following conditions: 14 °C and 6 V/cm, using the following conditions: block 1, pulse time 0.5 – 60 s for 16 hours; block 2, 60 – 120 s for 6 hours. As a marker Lambda Ladder PFG Marker with a size range from 48.5 to 679 kb and Yeast Chromosome PFG Marker with a size range from 225-2,500 kb were used (New England Biolabs). After electrophoresis gels were visualized by staining in 1 µM ethidium bromide solution or using SYBR Green I (Invitrogen) for 30 min at room temperature with slow shaking followed by examination under UV.

2.4.15 Amplification of DNA

The PCR reactions (50 µl) each contained 27 µl distilled water, 10 µl 5x reaction buffer, 8 µl 25 mM $MgCl_2$, 2-6 µl (depending on the estimated degeneracy) primers (100 µM) each, 1 µl dNTP (10 mM), 2 µl template and 0.25 µl Go*Taq* polymerase (Promega). The reactions were then exposed to the following steps and stages. Step 1 at 94 °C for 2 minutes. Step 2 incorporates three stages (X 30): Stage 1 at 94 °C for 30 seconds; Stage 2 at the recommended annealing temperature for 30 seconds; Stage 3 at 72 °C for 60 sec per 1000 bp. Step 3 included an elongation stage at 72 °C for 5 minutes and finally the holding step 4 was used to keep samples at 4 °C until collection. All PCR reactions were carried out using a MBS satellite thermal cycler (Thermo Scientific).

2.4.16 Purification of DNA from agarose gel

DNA fragments were recovered from agarose gel using the UV light visualization; each band was excised with a clean scalpel blade and purified using a QIAGEN Gel Purification Kit, following the manufacturer's recommendations. All reasonable care was taken to minimise exposure of molecules to UV light.

2.4.17 Restriction enzyme digestion of DNA

Reactions were carried out in 30 µl volumes: 3 µl 10X buffer, 5-10 units of restriction enzyme (New England Biolabs), 1 µg DNA and 21 µl dH_20 and incubated in a water bath at 37 °C for 2-4 hours. Restriction enzymes were heat inactivated at 65 °C for 20 minutes where appropriate.

Restriction for PFGE digestion was carried out in agarose plugs containing DNA. Plugs were washed several times 0.3 ml of TE buffer at room temperature. Then 100 µl of restriction buffer was added at 1X concentration and 10-30 U of appropriate restriction enzyme. Finally reactions were incubated at 37 °C for 10-14 h (unless otherwise stated).

2.4.18 Inverse PCR

5 µl genomic DNA (magnetotactic vibrio MV-1) was digested with appropriate restriction enzyme (NEB) in a total volume of 50 µl for 3 hours and then inactivated according to manufacturer recomendations. 10 µl of digested DNA were used in ligation reaction with 20 µl of 10X ligation buffer, 1 µl of T4 DNA ligase (Promega) and 169 µl dH_20 overnight at +4 °C. Then ligated DNA was ethanol precipitated and used in PCR (Ochman *et al.*, 1988).

2.4.19 Preparation of chemically competent cells

1 ml of an overnight culture of *E. coli* was diluted 100 times and incubated at 37 °C with shaking at 200 rpm (New Brunswick Scientific, G45 shaker). Once cells had grown to an A_{600} of c. 0.6 the cells were immediately placed on ice and then were harvested by centrifugation at 7 000 g for 7 min at +4 °C. The supernantant was discarded and the pellet was gently dissolved in 15 ml of 100 mM $MgCl_2$ fer and then centrifuged for 10 min at 5 000 g at +4 °C. The pellet was gently dissolved in 15 ml of 100 mM $CaCl_2$ and incubated on ice for 30 min. Cells were centrifuged for 10 min at 5000 g at +4 °C once again and the pellet was dissolved in 2,5 ml of 100 mM $CaCl_2$ containing 15% glycerol (v/v). After dispensing 100 µl of the prepared cells into cold sterile eppendorf tubes cells were snap frozen in a dry ice-EtOH bath and stored at -80 °C (Sambrook *et al.*, 1989).

2.4.20 Heat shock transformation of chemically competent cells

A tube containing competent cells was thawed on ice for 5 minutes. 1-5 µl of DNA were added to the cells and gently mixed by resuspending. Cells were kept on ice for 20 minutes and then placed into the water bath at exactly 42 °C for 90 sec. Then the tube was transferred back on ice and after 2 min 900 µl of SOC medium was added. Before plating cells were incubated at 37 °C with shaking for 1-1.5 hours to allow selective marker gene expression.

2.4.21 Sequencing of DNA (Sanger)

DNA sequencing was carried by the School of Biological Sciences Sequencing Service, University of Edinburgh using BIGDYE v3.1 sequencing kit (Applied Biosystems), and the sequencing information received was visualized and processed using Vector NTI (Invitrogen) and Bio Edit Sequence Alignment Editor (Tom Hall).

2.4.22 Ligation of DNA

Ligations were performed in 10-30 µl volumes containing vector and insert DNA (at a ratio of 5:1), 1 x T4 DNA ligase buffer and 1 unit of T4 DNA ligase (Promega) and incubated overnight at +4 °C.

2.4.23 Bacterial cell disruption by high pressure

Bacterial cells were resuspended in 50 ml 20 mM HEPES buffer (pH 7.4) and disrupted using One Shot Model (Constant Cell Disruption Systems) set to 30 KPSI (206.8 MPa or 2 109 kg/cm^2). The suspension was passed through the machine 3 times to ensure total cell disruption. The sample was kept on ice at all times and 4.5 ml of 10 mM phenylmethylsulphonyl fluoride (PMSF) was added as a protease inhibitor.

2.4.24 Magnetosome isolation

1 g (wet weight) of MV-1 cells was resuspended in 40 ml 50 mM HEPES - 4mM EDTA and sonicated on ice 6×30 seconds + 10 seconds rest (Sonicator Ultrasonic Processor XL, Hert Systems, output tune 7) unless cells were disrupted by high pressure (see Results and Discussion for more details). 4.5 ml 10 mM phenylmethylsulphonyl fluoride (PMSF) was added as protease inhibitor and magnetosomes were separated with the strong magnet placed on the side of 50 ml tube. Washes with the following magnetic separation at +4°C were repeated 5 times for 10 mM HEPES - 200 mM NaCl, pH 7.4, then magnetosomes were washed once with 10 mM Tris - 200 mM NaCl, pH 8.0 then with 10 mM Tris, pH 8.0 and finally with 10 mM HEPES, pH 7.4. As an adaptation of this method an intermediate sonication step was added. See Results and Discussion section for details. Magnetosome membrane fraction was isolated by addition of 200 µl 20 mM HEPES

(pH 7.2) containing 1 % (w/v) SDS at room temperature with constant slow stirring for 2-3 hours (Gorby *et al.*, 1988; Taoka *et al.*, 2006).

2.4.25 Inner membrane fraction isolation by sucrose gradient centrifugation

After cell disruption either by sonication or French Press cells debris was exposed to the magnetic field in order to separate magnetosome particles from the rest of the debris (see Magnetosome isolation). The sucrose gradient centrifugation was used in order to isolate inner membrane fraction. Step gradients of sucrose were prepared by slowly pouring 2 ml of sucrose solutions starting at 60% with 5% step to 30% in membrane buffer (1mM Tris, 1% EDTA pH 7.0). 2 ml of the cell debris was mixed with sucrose in order to obtain 20 % final concentration of sucrose. After loading the cell debris onto the top of the gradient tubes were equilibrated with 0.01 g accuracy to balance the rotor. The centrifugation was carried out for 10-12 h at 40,000 rpm (240 000 g) using Beckman SW40Ti rotor and centrifuge at 4 °C. After centrifugation red band of cytoplasmic-periplasmic fraction was collected and washed several times with 20 mM HEPES buffer (pH 7.4) (Tanaka *et al.*, 2006).

2.4.26 Protein concentration estimation

The estimation of the concentration of the protein in solution was carried out using a standard assay procedure from BIORAD based on the Bradford method. In order to obtain data for the standard curve Bovine Serum Albumin (BSA) solutions were used in concentrations from 0 to 12 μg/ml and after mixing with the dye-reagent the optical density was measured at 595 nm. The protein concentration was estimated by mixing sample with the dye-reagent followed by measuring its optical density and then plotting the data onto the standard curve.

2.4.27 1D SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

For routine SDS PAGE electrophoresis the following method was used. The Mini Protean II or the Protean II xi (gel dimensions: 70×85×0.5 mm and 183×200×1.5 mm respectively) apparatus (BIO-RAD) was cleaned with ethanol and assembled following the manufacturer's instructions. Single concentration gels were used consisting of 12.5% (v/v) separating gel and 5% (v/v) stacking gel (Laemmli, 1970). 1 volume of SDS (sodium dodecyl sulfate) sample buffer (10% glycerol, 10% SDS, 125 mM Tris-HCl (pH 6.8), 2-mercaptoethanol, bromophenol blue) was added to each membrane proteins sample. Samples were then boiled for 5 min prior to loading. A constant voltage of 200 V for the Mini Protean II for 1 hour and constant current of 35 mA for the Protean II xi for 5 hours was applied to the gel to separate the proteins. In order to estimate band sizes appropriate pre-stained protein marker was added to each gel.

In experiment where extra care was needed to be taken to avoid contamination pre-cast SDS gels were used. The XCell SureLock™ Mini Cell system (Invitrogen) with NuPAGE 4-12% BisTris gels was used according to the recommendations provided by the manufacturer.

2.4.28 Coomassie Blue gel staining

For routine SDS gel visualizations gels were covered with Coomassie blue staining solution and stained for 1-2 h with slow shaking. The staining solution was removed and destaining solution was added to the gel. The gel was destained with slow shaking and destaining solution was changed every several hours until all dye was removed from the background allowing stained bands visualization.

2.4.29 Sample preparation for protein analysis by Mass Spectrometry

SDS-PAGE gel was washed in 200 ml of dH_20 for 15 min 3 times and then stained with 20 ml GelCode Blue stain (Pierce) for 1 hour. Gel was destained in dH_20 for 1-2 hours followed by cutting out gel bands. Each band was incubated in 300 µl 200 mM NH_4HCO_3 in 50% (v/v) acetonitrile at 30 °C for 3 times and then incubated for 1 hour at 30 °C in 300 µl 20 mM DTT, 200 mM NH_4HCO_3 in 50% acetonitrile. Bands were washed 3 times 300 µl 200 mM NH_4HCO_3 in 50% acetonitrile and incubated for 20 min in dark in fresh 100 µl 50 mM iodacetamide, 200 mM NH_4HCO_3 in 50% acetonitrile. Then 3 washes in 500 µl 20 mM NH_4HCO_3 in 50% acetonitrile followed by cutting bands in 2×1 mm pieces and centrifuged at 13 000 rpm (SEC Microcentrifuge) for 2 min. Samples were covered with acetonitrile which was decanted and were dried at room temperature. As a final stage of preparation samples were swelled in 59 µl 50 mM NH_4HCO_3 containing 1 µl of trypsin (Promega) and incubated at 32 °C for 16-24 hours (Cronshaw and Florence, 2002).

2.4.30 Peptide concentration and purification using ZipTip

After protein samples were prepared and digested with trypsin ZipTips (Millipore) were used to concentrate samples prior to masspectrometry. A ZipTip was prepared by aspirating and dispensing 10 µl of wetting solution (100 % acetonitrile) twice followed by aspirating and dispensing 10 µl of equilibrating solution (0.1 % trifluoroacetic acid) twice. Then the peptides from the sample were transferred into the ZipTip by aspirating and dispensing 10-15 times. Once the tip was washed 4-5 times by aspirating and dispensing 10 µl of equilibrating solution sample was

transferred directly onto the MALDI-TOF plate by eluting it with 2-3 µl CHCA matrix solution.

2.4.31 The identification of the proteins by Liquid Chromatography Mass Spectrometry

The Liquid Chromatography Mass Spectrometry (LC-MS) system consisted of an Agilent 1200 Series HPLC (Agilent Technologies) with a Kasil sealed fused silica pre-column (Next Advance) packed to a length of approx. 3cm with Pursuit C18, 5um particle size (Varian) and PicoTip Emitter analytical column PF 360-75-15-N-5 (New Objective) packed to a length of approx. 20cm with Pursuit C18, 5um particle size (Varian).The column was equilibrated with solvent A (0.1% formic acid in 2.5% acetonitrile) and eluted with a linear gradient from 0 to 10% over 6 to 8min; from 8 to 60% over 8 to 35min; from 60 to 100% over 35 to 40min; solvent B (0.1% formic acid, 0.025% TFA in 90% acetonitrile) over 45min at a flow rate of 5ul/min. The LTQ mass spectrometer (ThermoScientific) was fitted with a nanoLC ESI source. These specifications were provided by Andrew Cronshaw (University of Edinburgh) who has carried out instrument operations for this work.

Data dependent acquisition was controlled by Xcalibur software and database searching was achieved using in-house licensed MASCOT software.

2.4.32 The identification of the proteins by Obitrap Mass Spectrometry

An LTQ-Orbitrap mass spectrometer (ThermoElectron) was coupled online to an Agilent 1100 binary nanopump and an HTC PAL autosampler (CTC). To prepare an analytical column with a self-assembled particle frit (Ishihama *et al.*, 2002), C18 material (ReproSil-Pur C18-AQ 3 µm; Dr. Maisch, GmbH) was packed into a spray emitter (75-µm ID, 8-µm opening, 70-mm length; New Objectives) using an air-

pressure pump (Proxeon Biosystems). Mobile phase A consisted of water, 5% acetonitrile, and 0.5% acetic acid; mobile phase B, consisted of acetonitrile and 0.5% acetic acid. The gradient used was 2 hours. The peptides were loaded onto the column at a flow rate of 0.7uL/min and eluted at a flow rate of 0.3uL/min according to the gradient. 0% to 20% buffer B in 75 min and then to 80% B in 13 min for a 2 hours run. FTMS spectra were recorded at 30,000 resolution and the six most intense peaks of the MS scan were selected in the ion trap for MS2, (normal scan, wideband activation, filling 7.5E5 ions for MS scan, 1.5E4 ions for MS2, maximum fill time 150 msec, dynamic exclusion for 60s sec). Raw files were processed using DTA SuperCharge to obtain the peak list.

Searches were conducted using Mascot 2.2 against a database containing MV-1 sequences. The search parameters were: MS accuracy, 6 ppm; MS/MS accuracy, 0.6 Da; enzyme, trypsin; allowed number of missed cleavages, 2; fixed modification, carbamidometylation on Cysteine and variable modification, oxidation on Methionine.

3 Results and Discussion

3.1 Investigation of the cultivation of marine vibrio MV-1 in the laboratory conditions

Strain MV-1 is able to grow either in anaerobic conditions using N_2O as a terminal electron acceptor or in microaerobic conditions. The original experiments of its cultivation were carried out by the laboratory of Dennis Bazylinski (Dean and Bazylinski, 1999). Due to the fact that even minor variations in cultivation techniques can have a major effect on the growth of such microorganisms it is important to describe these in detail.

Growth of MV-1 under N_2O demands gas-tight seals to prevent efflux of N_2O and influx of O_2. Available equipment was adapted and modified to allow cultivation of this strain. Liquid cultures were grown in glass vials (10 and 100 ml in volume) with rubber stoppers. Thicker stoppers were required than the ones used initially. A method of creating an extra seal by pouring molten polypropylene on the top to seal perforations made by needles was developed and used for inoculation (Figure 9). This method has allowed the growth of marine vibrio MV-1 in liquid medium routinely.

In order to obtain large amounts of cells to use in magnetosome isolation strain MV-1 was grown in 10 L cultures. The system of gas tubing with attached filters (Whatman) was designed to allow aseptic addition of the medium components and inoculation. The time of the medium saturation with N_2O was increased from 45 min to 5-8 h to allow sufficient saturation.

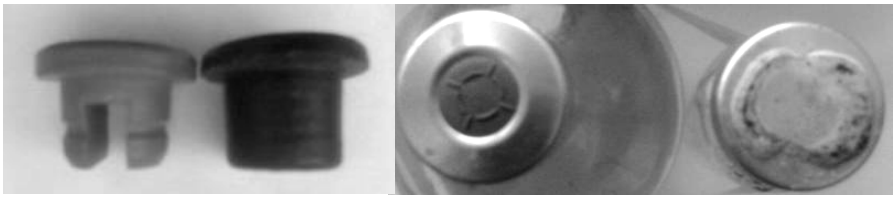


Figure 9. Rubber tops sealed with plastic used to stop gas escaping. This figure demonstrates the difference between thinner and thicker plugs used in this study (on the left) and sealing of the rubber plugs with a molten plastic (on the right).

Figure 10. Large scale culturing of marine vibrio MV-1. The photographs demonstrate a set up developed in this work to grow cells MV-1 in 10 L culture. An end part of rubber tubing used for delivery of N_2O was perforated to increase the number of bubbles to improve saturation. An upper tube set up was used to add medium components and an inoculum and as a gas escape outlet during saturation.

The standard growth curve was produced to estimate the optimal time for cultivation of MV-1. The obtained data is summarized on the graph below (Figure 11). Interestingly, the cultures of MV-1 do not demonstrate any significant stationary growth phase and tend to decrease in cell density rapidly after day 11.

3.1.1 Culturing of strain MV-1 on solid medium

One of the most important techniques in developing of a genetic system for a microorganism is its cultivation on the solid medium in the form of isolated colonies. A set of experiments was carried out in the laboratory of Dennis Bazylinski (UNLV, Las Vegas) by Sabrina Schuebbe (unpublished data) based on the method suggested for *Magnetospirillum gryphiswaldense* (Schultheiss and Schuler, 2003). The main idea behind this method is an addition of activated charcoal, yeast extract, bacterial peptone and an increased concentration of iron sulphate (see Marine vibrio MV-1 agar medium) to obtain growth on solid media. An activated charcoal was suggested to eliminate any toxic effect of oxidative metabolites whilst yeast extract and bacterial peptone should provide more accessible energy sources.

This method was adapted for use in this work. The gas mixture of 2 % (v/v) oxygen in nitrogen was used instead of 1 % (v/v) used in the suggested method. It was discovered that it is essential to use a fresh exponential growth phase culture for inoculation. The colonies appeared on the plate surface after 7 to 10 days on average. The colonies appearance can be described as follows: grey in colour, 1-2 mm in diameter, mostly with smooth edge but some with “rays” originating from the colony probably due to the ability of cells to move on the surface of the wet agar (Figure 12).

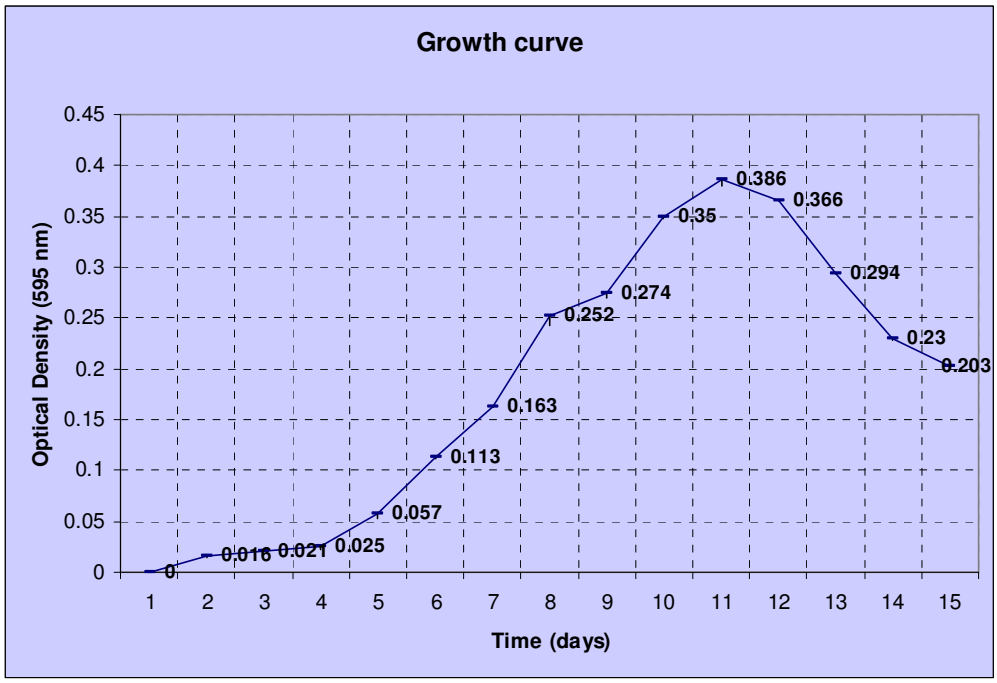


Figure 11. Growth curve for the liquid medium culturing of strain MV-1. This graph shows the growth curve for the strain MV-1 cultured in 90 ml of liquid medium with N_2O as an electron acceptor. Cultures were set up in repetitions of 3 and 1.5 ml of culture was taken every 24 hours to measure the optical density. The absence of contamination was tested by microscopy.

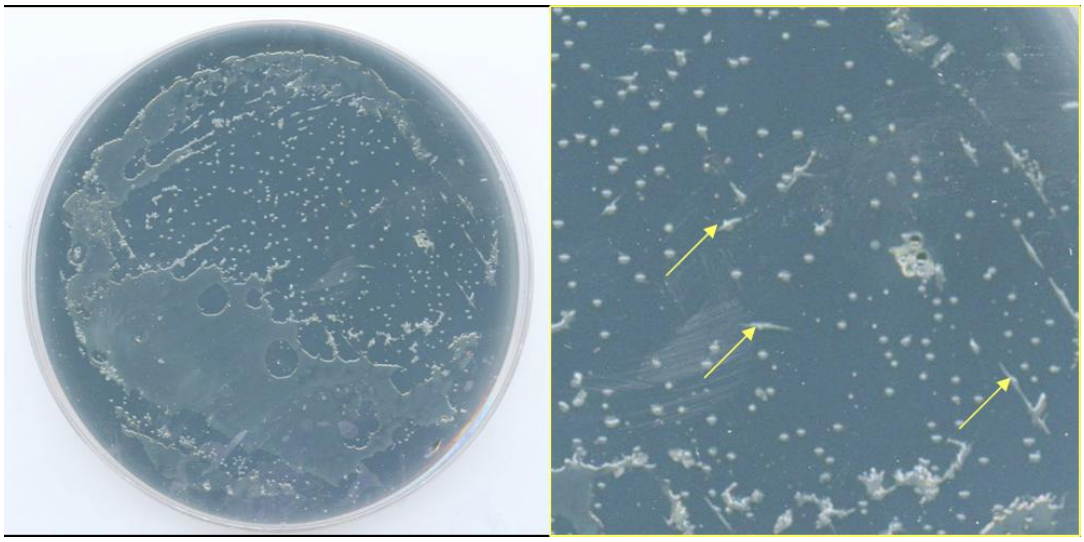


Figure 12. Colonies of strain MV-1 on an agar plate. This photograph shows an agar plate (magnified on the right) with colonies of MV-1. The plate was incubated at room temperature for 10 days in 2% (v/v) oxygen in nitrogen. 20 µl of the exponential growth phase culture were spread on the surface of the plate. Arrows show examples of unusual colonies with "rays".

The reproducibility of the growth on the surface of agar plates was relatively low. It appears that strain MV-1 has low plating efficiency and is very sensitive to some components of the growth medium and any little variations in the method, e.g. different manufacturer of the same component can lead to a significant reduction of growth rate. As a result, only very small, ≈0.5 mm in diameter, colonies appear on plates after 2-3 weeks of cultivation.

As a variation of the culturing in the form of individual colonies strain MV-1 was grown in test tubes filled with agar medium. 15 ml screw top test tubes were used. Different concentrations of agar in the standard liquid medium were tested. The best results were achieved with concentrations between 1 and 1.5 % (v/v). An example of such tube is shown on the photograph below. The production of nitrogen by the cells results in solid agar medium being disintegrated (Figure 13).

It was possible to isolate individual colonies using sterile glass Pasteur pipettes. The capillary at the end of the pipette allowed collection of a column of an agar containing chosen colony.

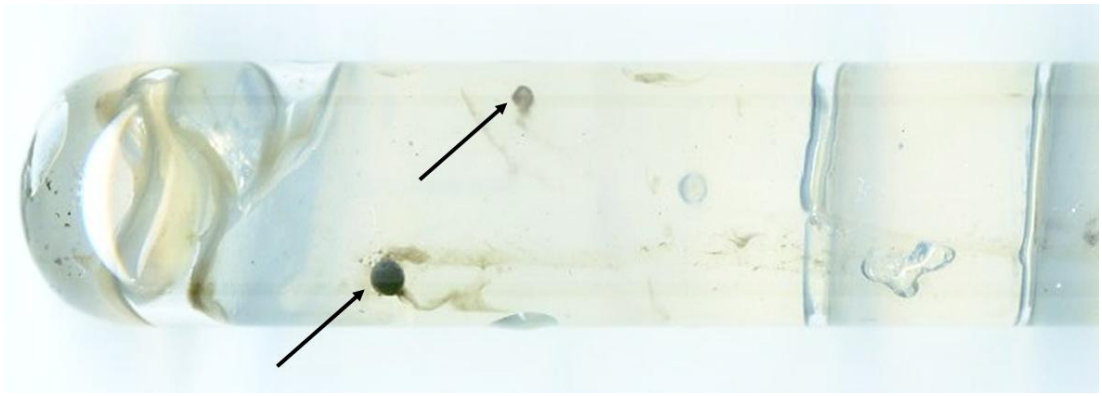


Figure 13. Colonies of strain MV-1 in the agar tubes. Arrows show dark lens shaped colonies of MV-1. Tubes were kept at room temperature for a period of 14 days. 10 µl of the exponential growth phase culture were mixed with medium containing 1.3% agar. The disintegration of agar by the produced nitrogen occurred gradually.

3.2 Investigation of the "magnetosome island" in marine vibrio MV-1

3.2.1 Introduction

This section provides a detailed description of the results and methods used to investigate the presence of the "magnetosome island" in the marine vibrio MV-1. As described before this cluster of genes is believed to be involved in magnetosome formation and has a similar organization in different magnetotactic microorganisms (Grunberg *et al.*, 2001).

In order to investigate the presence of similar genes in the genome of this microorganism it was decided to attempt to amplify the most conserved regions and then extend the obtained sequence using inverse PCR method and variations of it.

3.2.2 Primer design for MV-1 magnetosome formation genes

The availability of complete and draft sequences of strains *M. magnetotacticum* AMB-1, MS-1, *M. gryphiswaldense* MSR-1 and unnamed *Magnetococcus* MC-1 allowed alignment of the fragments most likely to be involved in the magnetosome formation. The "magnetosome formation genes" of these strains showed very high similarity apart from strain MC-1. Strain MC-1 has a significantly different genome GC content compared to other previously characterized magnetotactic bacteria (Table 12, page 136). Sets of primers were designed for the following genes of the *M. gryphiswaldense* MSR-1 magnetosome island: *mamA, mamM, mamC, mamB, mms6, mamK, mamO, mamE, mamG, mamP,* ORF12, *mamN, mamT, mamJ* (Table 3). As an example the procedure for designing primers for gene *mamT* is shown below:

1. Amino-acid sequences were obtained from the National Center for Biotechnology Information BLAST (Basic Local Alignment Search Tool) web site and aligned with the software ClustalW (European Bioinformatics Institute). The sequence of MC-1 sometimes was not used because of low similarity (Figure 14).

2. The conserved regions were chosen and nucleotide sequences of these regions were aligned (Figure 15).

3. Possible candidates for primers were chosen. Primers with a number of degeneracy options at several positions in the sequence were designed to allow annealing to a variety of related sequences. The annealing temperatures, presence of primer dimmers and secondary structures were calculated with the Sigma-Genosys DNA Calculator (Sigma-Aldrich Co.) or Vector NTI (Invitrogen).

CLUSTAL W (1.83) multiple sequence alignment:

```
AMB-1           MSMEAPRRGRRWVSLGMIALLAAIGLGLYWDQLSTPSGITPATSPRRAEGLLLGRLPLPM 60
MS-1            ----------------MIALLAAIGLGLYWDQLSTPSGITPATSPRRAEGLLLGRLPLPM 44
MSR-1           --MGTPGGGRRWMTLISITLLMVVGLGLYWDELSLSAGISPATSPRRAEGLLLGRLPLPM 58
                             *:** .:*******:** .:**:********************


AMB-1           EPSLLSPLERLLEPPLRYKLMTIRHIPPVKPGTGMPHPYVGDCIQCHLMVGGPAAGSQFK 120
MS-1            EPSLLSPLERLLEPPLRYKLMTIRHIPPVKPGTGMPHPYVGDCIQCHLMVGGPAAGSQFK 104
MSR-1           EPSILSPLEHLIEPPLQYKLMTIRHIPPVMPGTGMPHPYVGDCIQCHLMVGGPAAGSQFK 118
                ***:*****:*:****:************ ******************************


AMB-1           TPYGAVLENLSRVRKLGPPILPTSRQPHPPAGRCIKCHDIVVKVPVDKKGGMRWQL 172
MS-1            TPYGAVLENLSRVRKLGPPILPTSRQPHPPAGRCIKCHDIVVKVPVDKKGGMRWQL 160
MSR-1           TPYGAVLENLSRVRKLGPPILPTTRQPHPPAGRCIKCHDIVVKVPVEKKSGIKWLL 170
                ************************:**********************:**.*::* *
```

Figure 14. The alignment of the MamT protein sequences from strains *M. magnetotacticum* AMB-1, MS-1 and *M. gryphiswaldense* MSR-1.

CLUSTAL W (1.83) multiple sequence alignment:

```
AMB-1           GTGAGCATGGAGGCGCCGCGGCGCGGCCGTCGCTGGGTAAGCTTGGGGATGATCGCCTTG 60
MS-1            -------------------------------------------------ATGATCGCCTTG 12
MSR-1           ------ATGGGTACGCCAGGGGGCGGCCGTCGCTGGATGACCTTGATCTCGATCACCTTG 54
                                                                 **** *****

AMB-1           TTGGCGGCGATCGGGCTGGGACTCTATTGGGACCAGCTGTCCACGCCGTCCGGCATAACG 120
MS-1            TTGGCGGCGATCGGGCTGGGACTCTATTGGGACCAGCTGTCCACGCCGTCCGGCATAACG 72
MSR-1           CTGATGGTGGTCGGACTGGGACTCTATTGGGATGAGCTGTCCCTCTCCGCCGGCATCTCC 114
                 **  ** * **** *****************  ********    *  *******  *

AMB-1           CCCGCCACCTCGCCCCGACGGGCGGAGGGGCTGCTGCTGGGGCGTCTACCGCTGCCCATG 180
MS-1            CCCGCCACCTCGCCCCGACGGGCGGAGGGGCTGCTGCTGGGGCGTCTACCGCTGCCCATG 132
MSR-1           CCCGCCACATCGCCCCGTCGGGCGGAGGGGCTTTTGTTGGGGCGGCTGCCCTTGCCCATG 174
                ******** ******** **************  ** ******* ** **  ********

AMB-1           GAGCCATCGCTGCTGTCGCCGTTGGAGCGGCTGCTGGAACCGCCGCTCCGCTACAAGCTG 240
MS-1            GAGCCATCGCTGCTGTCGCCGCTGGAGCGGCTGCTGGAACCGCCGCTCCGCTACAAGCTG 192
MSR-1           GAGCCTTCGATTCTGTCGCCGCTGGAGCATCTCATTGAGCCGCCGCTTCAGTACAAGCTG 234
                ***** *** * ********* ******  **  * ** ******** *  *********

AMB-1           ATGACCATCCGCCATATCCCGCCGGTGAAGCCAGGAACCGGAATGCCGCATCCCTATGTA 300
MS-1            ATGACCATCCGCCATATCCCGCCGGTGAAGCCAGGAACTGGCATGCCGCATCCCTATGTA 252
MSR-1           ATGACCATTCGTCATATCCCGCCGGTAATGCCGGGGACAGGCATGCCCCATCCCTATGTG 294
                ******** ** ************** * *** ** ** ** ***** ***********

AMB-1           GGCGATTGCATCCAGTGCCACCTGATGGTGGGCGGCCCGGCCGCCGGCTCCCAGTTCAAG 360
MS-1            GGCGATTGCATCCAGTGCCACCTGATGGTGGGCGGCCCGGCCGCCGGGTCCCAGTTCAAG 312
MSR-1           GGGGATTGCATCCAATGCCATCTGATGGTCGGTGGCCCCGCTGCCGGATCACAGTTCAAG 354
                ** ************ ***** ******** ** ***** ** ***** ** *********

AMB-1           ACGCCCTATGGCGCCGTTTTGGAAAATCTCTCGCGGGTCCGCAAGCTGGGACCGCCCATT 420
MS-1            ACGCCCTATGGCGCCGTTTTGGAAAATCTGTCGCGGGTCCGCAAGCTGGGACCGCCCATT 372
MSR-1           ACGCCCTATGGCGCCGTACTGGAAAACCTGTCGCGGGTCCGCAAACTGGGGCCTCCCATT 414
                ****************  ******* ** ************** ***** ** ******

AMB-1           CTGCCGACCTCCCGCCAGCCGCATCCGCCGGCCGGACGATGCATCAAATGCCACGACATC 480
MS-1            CTGCCGACCTCCCGCCAGCCGCATCCGCCAGCCGGACGATGCATTAAATGCCACGACATC 432
MSR-1           CTTCCCACGACGCGCCAGCCGCATCCGCCTGCCGGGCGCTGCATTAAGTGCCATGACATT 474
                ** ** **  * ***************** ***** ** ***** ** ***** *****

AMB-1           GTGGTCAAGGTGCCGGTGGACAAGAAAGGCGGCATGAGATGGCAATTATGA 531
MS-1            GTGGTCAAGGTGCCGGTGGACAAGAAAGGCGGCATGAGATGGCAATTATGA 483
MSR-1           GTGGTCAAGGTGCCTGTGGAAAAGAAGTCCGGCATTAAATGGCTGTTGTAA 525
                ************** ***** *****   ****** * *****  ** * *
```

Figure 15. The alignment of the conserved regions of the DNA sequence for gene *mamT* from *M. magnetotacticum* AMB-1, MS-1 and *M. gryphiswaldense* MSR-1for primer design. The possible candidates for forward and reverse primers are highlighted in grey.

3.2.3 Amplification of magnetosome related genes in MV-1

Attempts to amplify specific magnetosome formation genes were made by using PCR with the designed primers and purified chromosomal DNA. PCR products were separated and detected by agarose gel electrophoresis.

As an example of successful amplification of *mamM* gene fragment the agarose gel photograph is shown below (Figure 16). It is interesting that two products of *mamM* fragment were amplified with sizes around 400 and 500 bp. However, the sequencing results returned highly similar sequences around 400 bp for both of them (Figure 17). The amplification of the second product was considered at first as possible evidence of the presence of another copy of *mamM* gene within the chromosome. Further analysis with the availability of the whole genome sequence suggests that this probably was a non-specific annealing of one of the primers.

The similarity with the *mamM* fragment from *M. magnetotacticum* AMB-1, MS-1, and *M. gryphiswaldense* MSR-1 to that of strain MV-1 is very high. An alignment of translated fragments is shown on Figure 18.

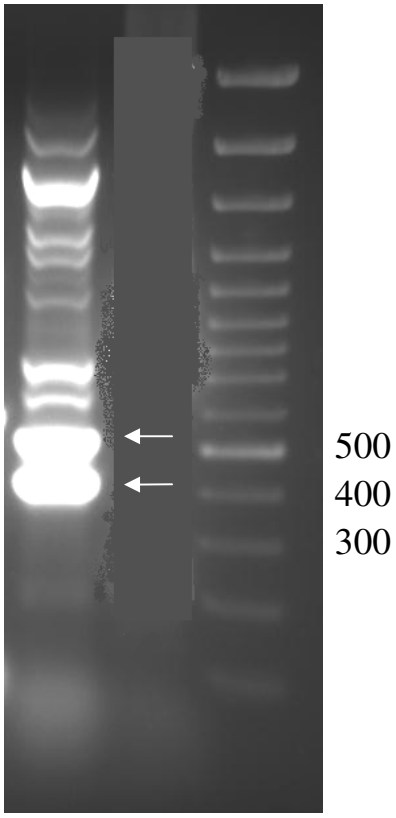


Figure 16. An agarose gel electrophoresis of amplified *mamM* gene fragment from strain MV-1. An agarose gel electrophoresis of amplified *mamM* gene fragment from MV-1. According to the product size expectations the products of interest are shown with arrows. DNA marker (Fermentas 0328) with sizes in base pairs is shown on the right side. The middle lane of this gel is hidden to allow better image interpretation.

```
CLUSTAL W (1.83) multiple sequence alignment

mamM 400 bp       AAAGGGTCGGGTCC-TGAGGCCATTGTAGGCATGACCG-TTTCCGAACAAGCATTGGATA 58
mamM 500 bp       NAAGGGACGGGTCCCNNAGGCCATTGTAGGCATGACCGCTTTCCGAACAAGCATTGGATA 60
                   ***** *******   ********************* *********************

mamM 400 bp       GGGAACATCCTTACGGCCACGGTAAAATCGAATTTGTCTTGTCGTTGTTTGTCAGCGTGA 118
mamM 500 bp       GGGAACANCCTTACGGCCACGGTAAAATCGAATTTGTCTTGTCGTTGTTTGTCAGCGTGA 120
                  ******* ************************************************

mamM 400 bp       TTTTTTTCATCATCGCCGCATATCTTTTGGTTCATGCTATTTTTGTCCTGATGGATCCAT 178
mamM 500 bp       TTTTTTTCATCATCGCCGCATATCTTTTGGTTCATGCTATTTTTGTCCTGATGGATCCAT 180
                  ************************************************************

mamM 400 bp       CGTTGCACCGCGCCCCGCACTTGATTGCGTTGTGGGCGTCGCTGTTGGTCGTGATCGTCA 238
mamM 500 bp       CGTTGCACCGCGCCCCGCACTTGATTGCGTTGTGGGCGTCGCTGTTGGTCGTGATCGTCA 240
                  ************************************************************

mamM 400 bp       ACGTTATTATGTACTTTTATTCGCGCTGCGTGGCCATTGAAACCAACAGCCCGTTGGTGC 298
mamM 500 bp       ACGTTATTATGTACTTTTATTCGCGCTGCGTGGCCATTGAAACCAACAGCCCGTTGGTGC 300
                  ************************************************************

mamM 400 bp       GGACGTTGGCCAAGCATCATCACGGTGACGCGGCGTCGTCAGGGGTGGTCGCACTGGGCA 358
mamM 500 bp       GGACGTTGGCCAAGCATCATCACGGTGACGCGGCGTCGTCAGGGGTGGTCGCACTGGGCA 360
                  ************************************************************

mamM 400 bp       TCATCGGCGCGCACTTTTTCAACATGCCCTGGATCGATCCCGANNACCAANANNNACNCC 418
mamM 500 bp       TCATCGGCGCGCACTTTTTCAACATGCCCTGGATCGATCCCGANNNNNNNNNNNNNNNN- 419
                  ******************************************     * ***  *
```

Figure 17. CLUSTAL W multiple sequence alignment for two amplified fragments of *mamM* gene. This alignment shows that there is no significant difference between sequences of 400 and 500 bp amplified fragments. Mismatches in the beginning of the sequence can be explained by the limitations in sequencing quality.

CLUSTAL W (1.83) multiple sequence alignment

```
AMB-1           TTISSKPLDAEHPYGHGKVEFILSMVVSVVFIGLTGYLLVHAVQILLDESMHRTPHLIVL 60
MS-1            TTISSKPLDAEHPYGHGKVEFILSMVVSVVFIGLTGYLLVHAVQILLDESMHRTPHLIVL 60
MSR-1           TTISSKPLDAEHPYGHGKVEFILSMVVSVVFIVLTGYLLVHAVQILLDESLHRTPHLIVL 60
MV-1            -------LDREHPYGHGKIEFVLSLFVSVIFFIIAAYLLVHAIFVLMDPSLHRAPHLIAL 53
                       ** ********:**:**:.***:*: ::.******: :*:* *:**:****.*

AMB-1           WAALVSVGVNVAMYFYSRCVAIETNSPIIKTMAKHHHGDATASGAVALGIIGAHYLNMPW 120
MS-1            WAALVSVGVNVAMYFYSRCVAIETNSPIIKTMAKHHHGDATASGAVALGIIGAHYLNMPW 120
MSR-1           WAALVSIGVNVGMYFYSRCVAIETNSPLIKTMAKHHHGDATASGAVALGIIGAHYLNMPW 120
MV-1            WASLLVVIVNVIMYFYSRCVAIETNSPLVRTLAKHHHGDAASSGVVALGIIGAHFFNMPW 113
                **:*: : *** ***************:::*:********::**.*********::****

AMB-1           IDPAVAL 127
MS-1            IDPAVAL 127
MSR-1           IDPAVAL 127
MV-1            IDP---- 116
                ***
```

Figure 18. CLUSTAL W alignment of translated fragments of gene *mamM* with all 4 strains of magnetotactic microrganisms. The alignment demonstrates high similarity in the amino acid sequences of these fragments.

In a number of cases multiple products were obtained for sets of primers designed for other magnetosome related genes, which are can be explained by non-specific hybridization. The non-specific hybridization may occur due to the fact that the actual sequence of MV-1 may vary from the sequences of other MB. The adjustment of the PCR conditions, like increasing the annealing temperature and the amount of template DNA allowed more specific alignment and numbers of bands on the gel were decreased. Some of the PCR products obtained were of the size that is expected from the *M. gryphiswaldense* MSR-1 magnetosome island.

The fragments of expected size obtained from amplification were sequenced by Sanger sequencing using The GenePool service (University of Edinburgh). Most of those sequencing returned not recognizable (double sequence) results probably due either to the presence of multiple products within one sample or to the non-specific sequencing primer hybridizations within the one PCR product.

As a result of amplification of fragments with non-specific hybridization the BLAST search returned partial similarity to the following genes of MB: long chain fatty acid CoA ligase (primers for *mamP*), DNA methylase N-4/N-6 (primers for *mamO*), and the putative chemotaxis protein (primers for *mamN*).

3.2.4 Investigation of the "magnetosome island" in strain MV-1 with inverse PCR

Based in the hypothesis that magnetosome formation genes are clustered on the chromosome it was decided to attempt to extend obtained sequence of *mamM* gene both up- and down-stream. In order to achieve this aim inverse PCR experiments were carried out. The MV-1 chromosomal DNA was cut with the AvaI enzyme. This

enzyme was chosen because it is calculated to have recognition sites each 1000-1300 bp in average in all characterized magnetotactic microorganisms. The approach used in the inverse PCR technique is shown schematically on the diagram on page 93.

In a number of cases multiple products were obtained in PCR reactions. This can be explained by the possibility of ligation of several small products into the one circular DNA molecule as well as non-specific hybridization. In order to reduce occurrence of a multiple products the volumes of ligation were increased up to 10 times to increase a chance of a self ligation. Amplified fragments are then sequenced and analysed.

Although these experiments have allowed amplification and sequencing of a short fragment of *mamM* gene and have expanded the known sequence for 153 nucleotides it was decided to attempt a different approach to extend the known sequence.

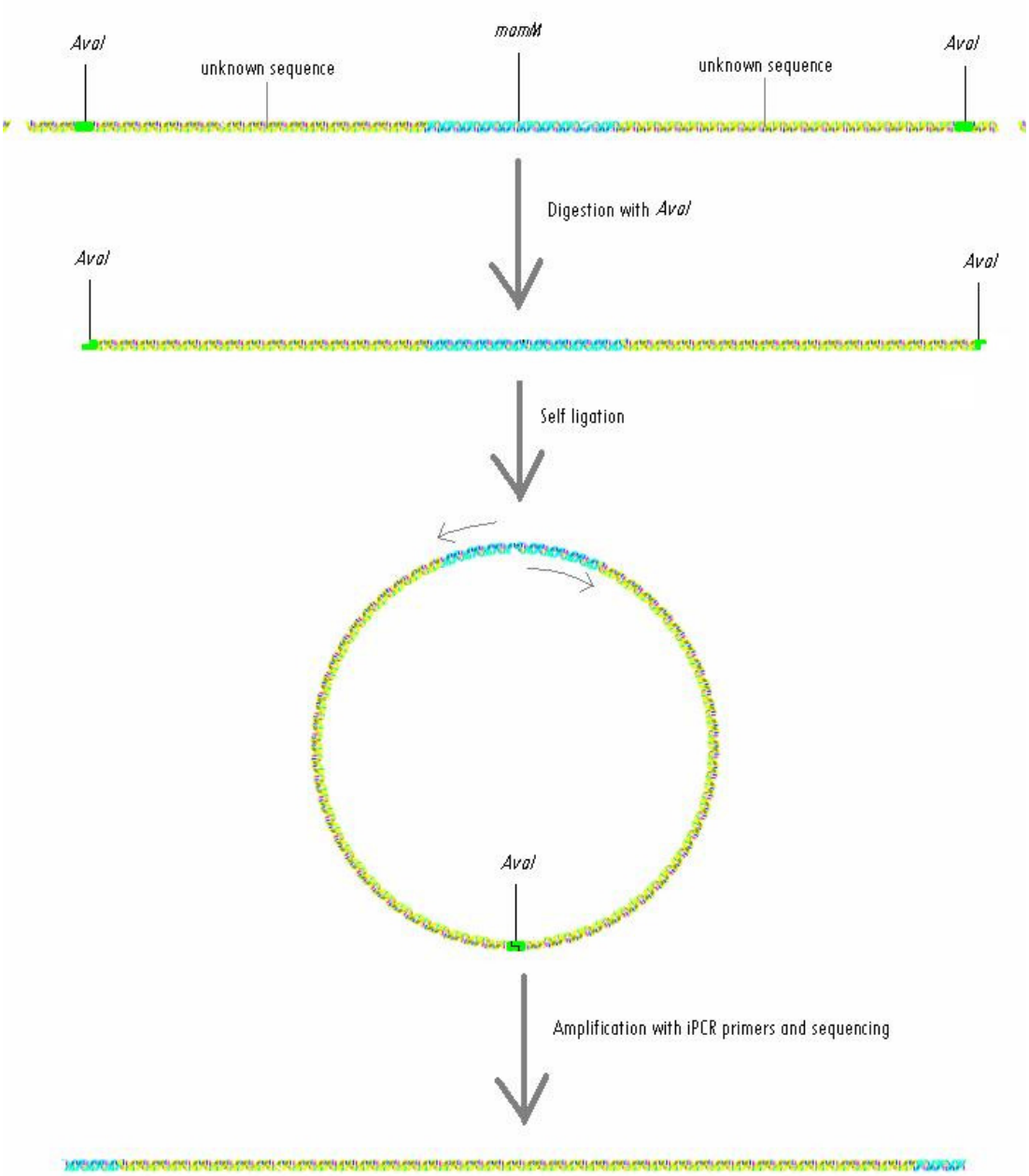


Figure 19. The schematic representation of the inverse PCR technique. This method involved a restriction of the chromosomal DNA with the chosen enzyme followed by the ligation of this fragment on itself. An amplification of the unknown sequence became possible with the primers designed using known sequence so that their 3' ends face unknown regions (Ochman *et al.*, 1988).

3.2.5 Amplification of the flanking sequence using the predicted map of "magnetosome island"

To improve the gained data another attempt at amplification of magnetosome related genes have been carried out based on the theory that these genes are localized close to the *mamM* fragment. PCRs with the sets of primers where one of the primers was either *mamM* forward or *mamM* reverse primer and the second primer was one of the previously designed primers for *mam/mms* genes. For example, the combination of a *mamM* forward primer and a *mamP* reverse primer should allow amplification of a product with the size around 4.5 kb. These PCR experiments resulted in PRC fragments with the same sizes as might be expected based on the *M. gryphiswaldense* MSR-1 magnetosome island gene map. These fragments were not suitable for direct sequencing because a double sequence was obtained in a number of cases. So it was decided to sub-clone these fragments into the vector pGEM-T (Promega). These vectors with directly sub-cloned PCR products were transformed into the *E. coli*. After plasmid purification the cloned fragment was sequenced with the pUC/M13 sequencing primers.

This approach was shown to be efficient and allowed extension of the known sequence from just fragmentary *mamM* gene onto the two more genes downstream as shown in Figure 20.

3.2.6 Investigation of the "magnetosome island" in strain MV-1 with variation of inverse PCR and use of plasmid libraries

Although methods described in the previous sections proved successful in identification of novel sequences of marine vibrio MV-1 magnetosome formation genes it was decided to test another approach to extend the known sequence of this region.

This technique was a variation of the inverse PCR method described above (Natarajan and Boulter, 1995). In the first stage plasmid libraries were prepared by digesting MV-1 chromosomal DNA with an enzyme of choice (EcoRI, BamHI, HindIII and Sau3A were used in this work). Then the digested DNA sample was run on the agarose gel and a smear of DNA with size ranges from 2 to 6 kb was excised (

Figure 21).

Reasonable care was taken to limit exposure to UV light. After gel purification of the excised DNA it was ligated into the vector of choice. In this study vectors pBluescript SK+ (Stratagene) and pBAD24 (Guzman *et al.*, 1995) were used. The only requirements in the choice of the vector are a high copy number and presence of the matching recognition sites in the cloning site with the sites used for library preparation. As a result a sample containing a large number of plasmid molecules with inserted fragments of chromosomal DNA with sizes from 2 to 6 kb was produced. The originally described method suggested an amplification of the produced library by transforming it into *E. coli* strain followed by a plasmid purification; however in this study it was omitted due to sufficient amount of sample being produced. This sample then was used to amplify the unknown sequence as depicted schematically in Figure 22.

Products that were amplified using this method were then subjected to direct sequencing if the PCR reaction yielded a single product. For those reactions containing multiple products bands of interest were excised and purified from an agarose gel and followed by cloning into pGEM-T vector and then subjected to sequencing.

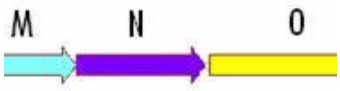


Figure 20. "Magnetosome island" cluster extended to 3 ORFs. This diagram demonstrates schematic organization of the "magnetosome island" genes obtained through extension of known sequence of *mamM* gene.

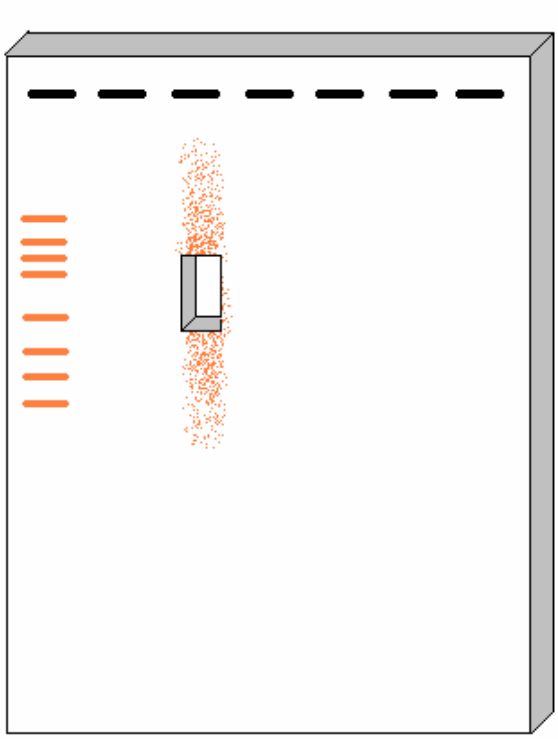


Figure 21. The isolation of fragments of the digested genomic DNA from the agarose gel. This diagram is the schematic illustration of the excision of the part of agarose gel containing fragments of DNA of specific size range by using DNA ladder as a marker (left lane).

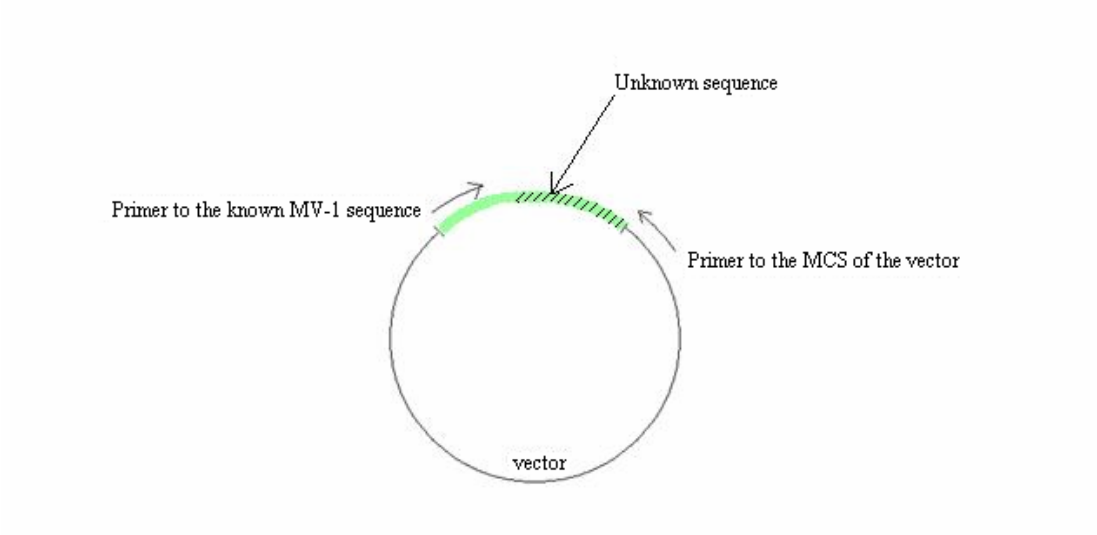


Figure 22. Variation of the inverse PCR method. This diagram shows the schematic representation of the amplification of the unknown sequence of interest. One of the primers is complementary to the known region of the vector for example MCS (Multiple Cloning Site) and another primer is complementary to the known region of the sequence of the insert.

This method showed high efficiency and can be suggested for similar research when other methods for example involving production of the cosmid libraries and DNA probes are not available. The method was relatively easy to use and allowed me to carry out an amplification of the each of the target fragment in a time scale of 2 to 3 days and required only standard reagents and kits.

The sequencing of the amplified products allowed extension of the known sequence of the MV-1 magnetosome island from 3 ORFs previously to 7 (Figure 23).

3.2.7 Discussion

This part of the work provided strong evidence for the presence of "magnetosome island" in marine vibrio MV-1. The organization of the sequenced cluster is similar to that described for previously characterised magnetotactic microorganisms. The correlation of the gene order can be interpreted as evidence of the origin of these genes by horizontal gene transfer (Figure 24).

It is necessary to outline that similar work on investigation of the "magnetosome island" in marine vibrio MV-1 was carried out completely independently by the group of Dennis Bazylinski (UNLV, Las Vegas) (Jogler *et al.*, 2009a). The results of the work of this group became available in public domain in the form of publication in February 2009, whilst this stage of this project was completed by February of 2008.

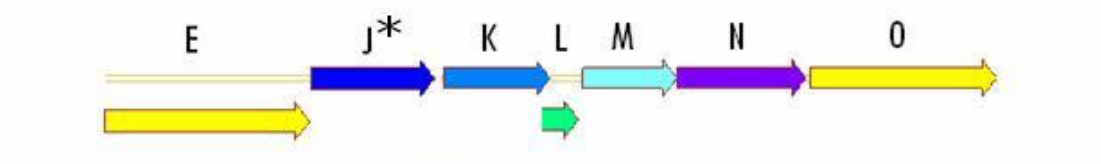


Figure 23. "Magnetosome island" cluster extended to 7 ORFs using variation of the inverse PCR technique. This diagram shows organization of the "*mam*" genes of the stain MV-1 magnetosome formation cluster.

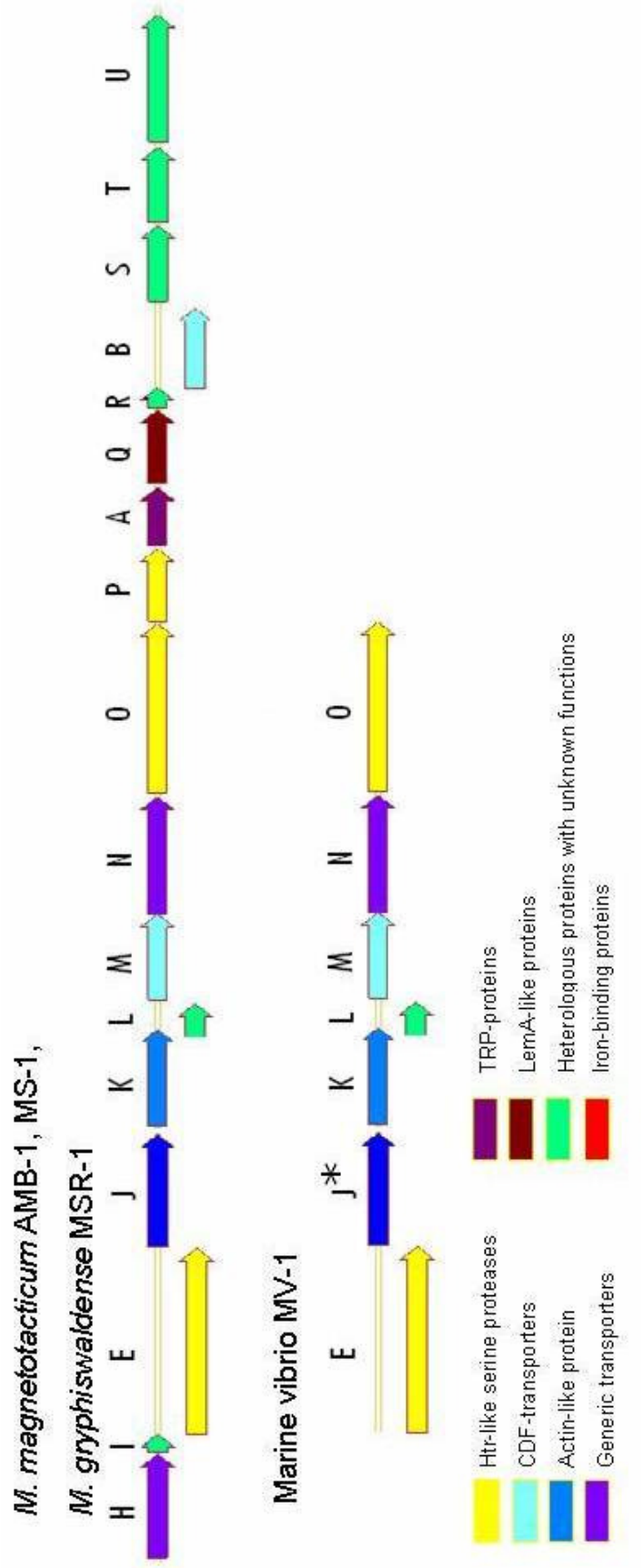


Figure 24. The comparison of the organisation of the known sequence "*mam*-genes" cluster of the "magnetosome island" in marine vibrio MV-1 (partial sequence) and other strains of magnetotactic bacteria. This diagram demonstrates the similarity in the organisation of the chosen region of magnetosome formation related genes.

The gene marked as J* (*mamJ**) has very low similarity to the gene *mamJ* in previously described strains.

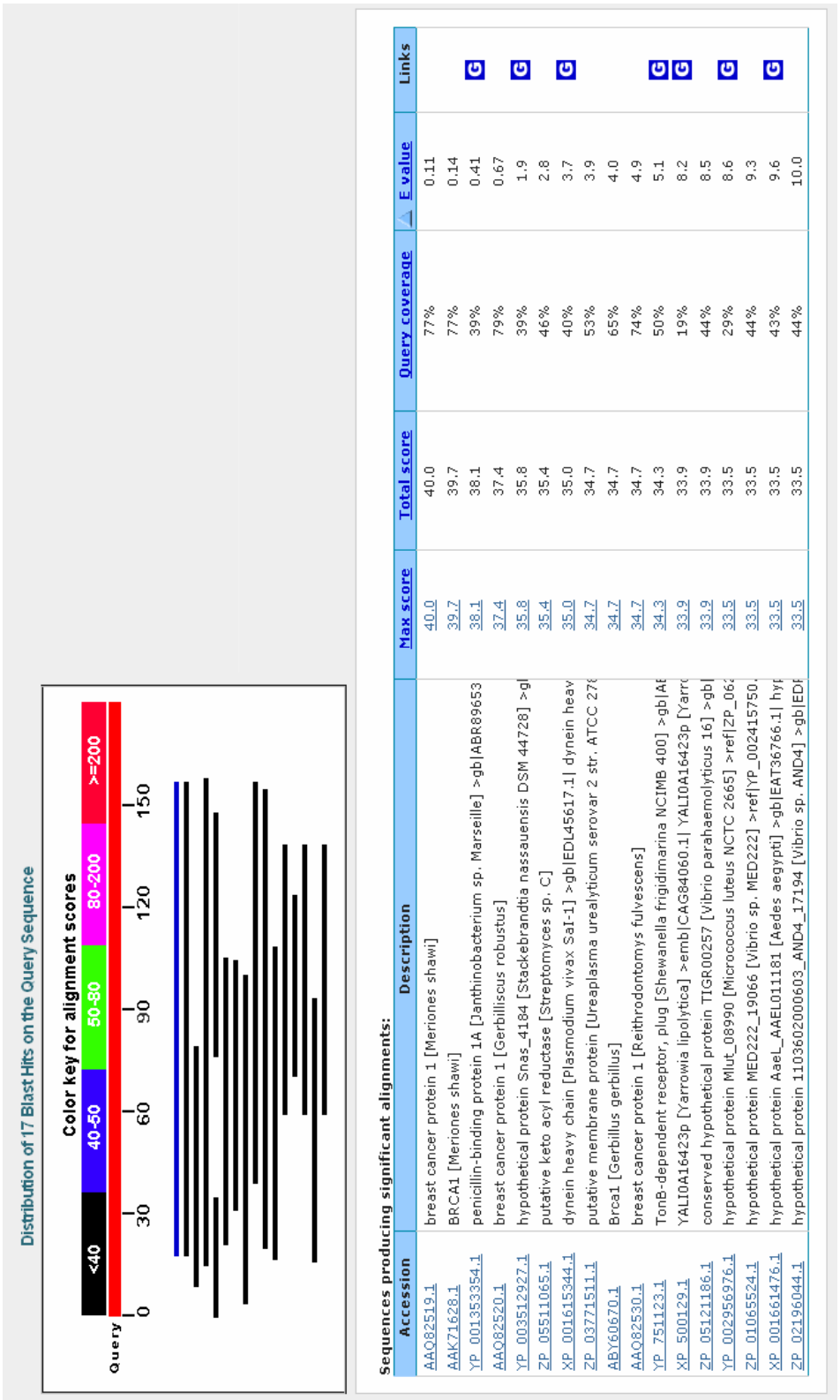
Although most of the analysed genes at this stage of the work have shown high similarity to those reported for other strains of magnetotactic microorganisms the gene called *mamJ** has almost no similarity to *mamJ* gene in *M. gryphiswaldense* MSR-1. The Clustal W alignment of these transcribed genes is shown below (Figure 25).

An attempt to analyse the sequence of this putative protein using BLAST search shows very low similarity with other proteins (Figure 26). This putative protein is a good candidate for further research due to the fact that it highly dissimilar to other described proteins and its function is unknown.

It would have been possible to continue extension of the known sequence of the magnetosome formation cluster in strain MV-1 using methods described in this section or by involving production of cosmid libraries. However, once strong evidence of the presence of the "magnetosome island" in marine vibrio MV-1 was obtained it was decided to proceed with the sequencing of the whole genome. The sequencing of the genome of marine vibrio MV-1 is described in the following section.

```
CLUSTAL 2.0.12 multiple sequence alignment

MV-1    -----------MASGSNKKFLARQKILSGETP-------FVNGVGAQFTEGALTSSPW-- 40
MSR-1   MAKNRRDRGTDLPGDGDQKISTGPEIVSVTVHPSPNLAAAAKPVQGDIWASLLESSPWSA 60
                  :....::*: :  :*:*  .         .: * .::  . * ****

MV-1    ----------------------------VKKAELPSFSKTKLNGSE---------PANTPE 64
MSR-1   NQGGLVETAQPPSAPIRSQDPVPVADLVNRWSQPIWRTAPLAGNAESSEEGVVAPSLTQS 120
                                  *:: . * : .: * *.          *: * .

MV-1    DNASAWG---VSTLPETAPELDSPTLSQASGSD-------------------------WD 96
MSR-1   DSVLAVSDLVIDVQPETDAEVEVSIEPEPALVEPVIEIEAEAAEVEPEPAPVADLVNRWA 180
        *.. * .   :.. *** .*:: .  .:.:  :                         *

MV-1    DAFAMVSPWGAMGEGALVKAREASISPSATVPSTLEVRFKSR-------NLNVPHSNGVK 149
MSR-1   QPIWRTAPLAGNAESSEEGVVAPSLTQSDSVLAVSDLVIDVQPEANAEVEVSIEPEPALV 240
        :.:  .:* .. .*.:   .  .*:: * :* :. :: :. :       ::.:  . .:

MV-1    RRESTVVDNTKGWNPLTTEAQDALDAKTGK------------------------------ 179
MSR-1   EPVIEIEAEAAEVEPEPAPVEPAIEIEAIRVELEPVLIDEVVELVTEFEYSQAESVASAD 300
        .    :  ::   :* .: .: *:: :: :

MV-1    ------------------------------------------------------------
MSR-1   LIANPAPAESSRLAELLDEAAAIAAPAVAVAVEATRQPNKITASVKKRAPVQEVPVEDLL 360


MV-1    ------------------------------------------------------------
MSR-1   GGIFGVAGSAVRGVFTIGGGFVDGVVKGGRLVGSNVVAGTRRLAQTIEVSCGSCSSPKCD 420


MV-1    ------
MSR-1   AEDKNK 426
```

Figure 25. Clustal W alignment of the transcribed sequences of *mamJ* genes. This alignment demonstrates lack of similarity in the transcribed sequences of *mamJ* analogues in marine vibrio MV-1 and *M. gryphiswaldense* MSR-1.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
AAQ82519.1	breast cancer protein 1 [Meriones shawi]	40.0	40.0	77%	0.11	
AAK71628.1	BRCA1 [Meriones shawi]	39.7	39.7	77%	0.14	
YP_001353354.1	penicillin-binding protein 1A [Janthinobacterium sp. Marseille] >gb|ABR89653	38.1	38.1	39%	0.41	G
AAQ82520.1	breast cancer protein 1 [Gerbilliscus robustus]	37.4	37.4	79%	0.67	
YP_003512927.1	hypothetical protein Snas_4184 [Stackebrandtia nassauensis DSM 44728] >gl	35.8	35.8	39%	1.9	G
ZP_05511065.1	putative keto acyl reductase [Streptomyces sp. C]	35.4	35.4	46%	2.8	
XP_001615344.1	dynein heavy chain [Plasmodium vivax Sal-1] >gb|EDL45617.1| dynein heav	35.0	35.0	40%	3.7	G
ZP_03771511.1	putative membrane protein [Ureaplasma urealyticum serovar 2 str. ATCC 27	34.7	34.7	53%	3.9	
ABY60670.1	Brca1 [Gerbillus gerbillus]	34.7	34.7	65%	4.0	
AAQ82530.1	breast cancer protein 1 [Reithrodontomys fulvescens]	34.7	34.7	74%	4.9	
YP_751123.1	TonB-dependent receptor, plug [Shewanella frigidimarina NCIMB 400] >gb|AE	34.3	34.3	50%	5.1	G
XP_500129.1	YALI0A16423p [Yarrowia lipolytica] >emb|CAG84060.1| YALI0A16423p [Yarr	33.9	33.9	19%	8.2	G
ZP_05121186.1	conserved hypothetical protein TIGR00257 [Vibrio parahaemolyticus 16] >gb|	33.9	33.9	44%	8.5	
YP_002956976.1	hypothetical protein Mlut_08990 [Micrococcus luteus NCTC 2665] >ref|ZP_06	33.5	33.5	29%	8.6	G
ZP_01065524.1	hypothetical protein MED222_19066 [Vibrio sp. MED222] >ref|YP_002415750.	33.5	33.5	44%	9.3	
XP_001661476.1	hypothetical protein AaeL_AAEL011181 [Aedes aegypti] >gb|EAT36766.1| hyp	33.5	33.5	43%	9.6	G
ZP_02196044.1	hypothetical protein 1103602000603_AND4_17194 [Vibrio sp. AND4] >gb|EDF	33.5	33.5	44%	10.0	

Figure 26. The summary of the BLAST search for the transcribed sequence of *mamJ gene.** The diagram was adapted from the National Center for Biotechnology Information (NCBI) web site.

3.3 The whole genome sequencing of the marine vibrio MV-1

With the solid evidence of the presence of the magnetosome island in marine vibrio MV-1 gained in the previous part of this work it was decided to proceed with whole genome sequencing. The availability of the genome sequence should significantly add to the understanding of the magnetosome formation process in bacteria generally and this strain in particular.

The size of the genome of marine vibrio MV-1 was estimated using Pulse Field Gel Electrophoresis (PFGE). Chromosomal DNA of MV-1 was digested by SpeI, PmeI and PacI restriction endonucleases. Digested by a specific enzyme fragments were separated on the agarose gel and the distances migrated by each fragment were measured. The molecular weight markers were used to build a size calibration curve. Molecular weights calculated for each of the fragments were summarized allowing the prediction of the size of the genome to be 3.7±0.4 Mb which confirmed an estimation obtained earlier by the group of Dennis Bazylinski (Dean and Bazylinski, 1999).

This section contains detailed explanation of the approaches used for sequencing and the results that were gained with each of the methods. The results of the automated annotation of the genome are also included in this section.

3.3.1 Genome sequencing using SOLEXA technology

The initial sequencing of the genome of strain MV-1 was carried out with Illumina Solexa GAII system (The Gene Pool, University of Edinburgh). The sample containing 40 μg of genomic DNA was prepared and submitted to the Gene Pool service.

The Solexa sequencing has generated 3,674,724 reads that equals 128,615,340 bases of sequence. The sequence was generated in the form of 35 bp reads and then was assembled into contigs using the Velvet *de-novo* assembler (Zerbino and Birney, 2008). Any two fragments were considered as positive for assembly if there was at least 20 bp overlap of the aligned sequences. This assembly resulted in generation of 2300 contigs which is a relatively large number. One of the reasons that have negative effect on the number of successful assemblies is a presence of transposons and various repeats in the bacterial chromosome sequence. Software that performs automated assembly finds more than one place to which the fragment of the sequence can be aligned and as a result does not use such fragment in assembly. The statistics of the contig sizes and numbers is shown on the following diagram (Figure 27).

Another graph is shown to demonstrate a ratio between a high number of short contigs and a small number of long contigs (Figure 28). As it can be seen from this graph the number of the fragments with short length is high. For example there were 545 contigs with length of less than a 100 bases. These contigs are difficult to use in analysis and therefore it was decided to attempt to improve the obtained data using different sequencing techniques.

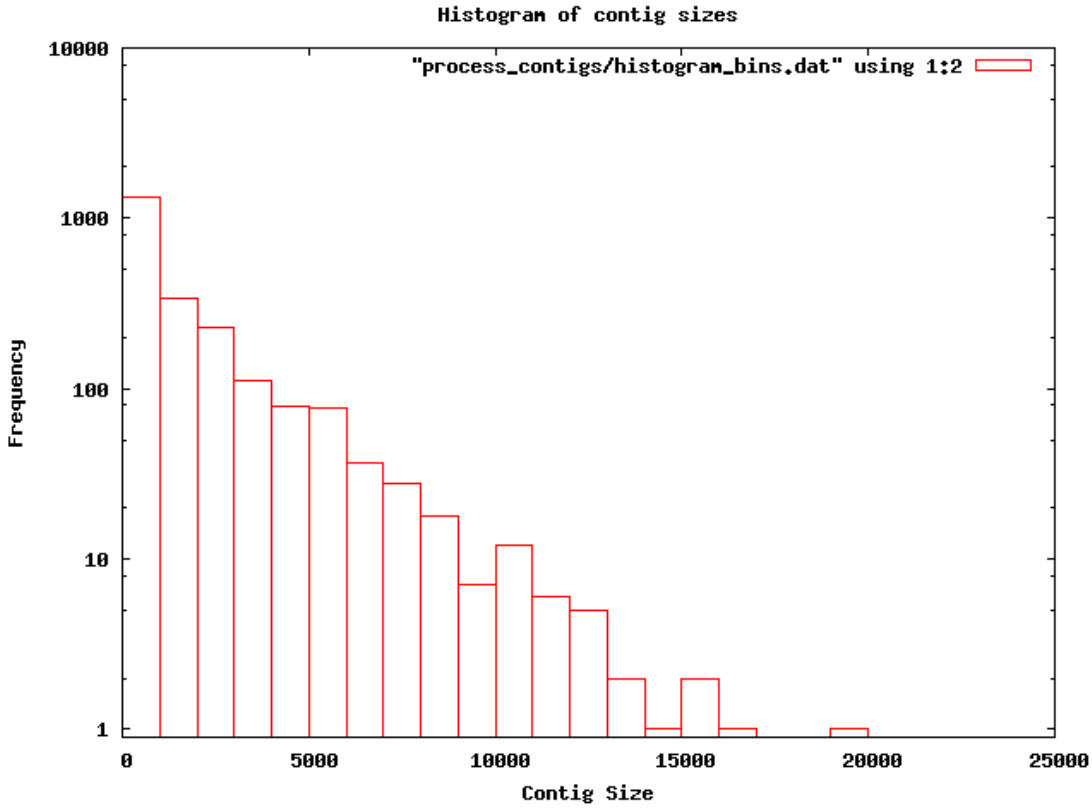


Figure 27. Numbers and sizes of contigs obtained after assembling of SOLEXA generated reads for genome sequencing. This histogram summarizes the number of contigs (vertical axis) of specific sizes in bases (horizontal axis).

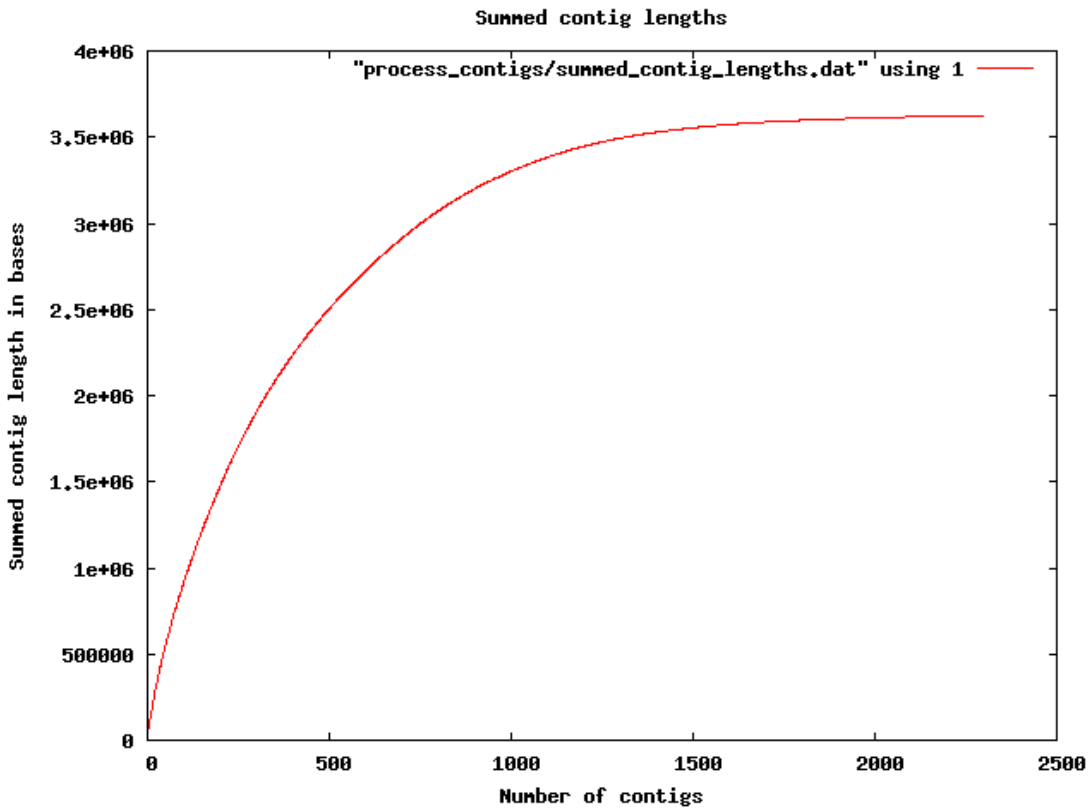


Figure 28. The summed contig length to the number of contigs obtained assembling of reads generated with SOLEXA sequencing. This graph is an additional illustration of the large number of a relatively small contigs produced with SOLEXA sequencing technique.

3.3.2 Genome sequencing using 454 sequencing technology

The genome of strain MV-1 was also sequenced using 454 sequencing technology (Roche 454 FLX). This sequencing technique allows generation of longer reads of 250 bases on average. This sequencing was also carried out at the University of Edinburgh sequencing facilities (The Gene Pool).

This sequencing method has generated 125,028 reads that equals 29,400,978 bases of sequence. The number of generated contigs produced after assembling of reads generated with 454 sequencing is 493. The statistical data obtained in this sequencing analysis is summarized in the table below (Table 5).

The obtained data has significantly decreased the total number of contigs from 2300 to 493 compared to SOLEXA sequencing. The obvious next step was an attempt to combine generated data to improve genome assembly. The combined assembling allowed reduction of the number of contigs from 493 to just 191 which is a very significant improvement. The table below summarizes the number of contigs obtained by each of the method and by the combined assembly (Table 6).

Not only the total number of contigs has been significantly reduced but also the number of large contigs has been increased. Very significant improvement of the assembling can also be observed on the histogram below (Figure 29).

The statistical analysis also confirms an improvement in assembling of the genome. These results are shown in Table 7.

N50 (bases)	19534
Maximal contig size (bases)	96718
Number of bases in contigs	3591137
Number of contigs	493
Number of contigs >=1 kb	297
Number of contigs in N50*	46
Number of bases in contigs >=1kb	3528532
GC Content of contigs (%)	54.3024952821349

Table 5. Summarized statistical data for assembling of reads generated with 454 sequencing of strain MV-1. This table shows statistical summary of contig numbers and sizes obtained with 454 sequencing. *Number of contigs in N50 shows the minimum number of the generated contigs that contain at least half of the number of bases in the sequence.

Assembler input data	Number of contigs
SOLEXA	2300
454	493
SOLEXA + 454	191

Table 6. The number of contigs generated with different assembler input data. This table shows the number of contigs generated by assembling raw data obtained by SOLEXA and 454 sequencing methods and using combined data.

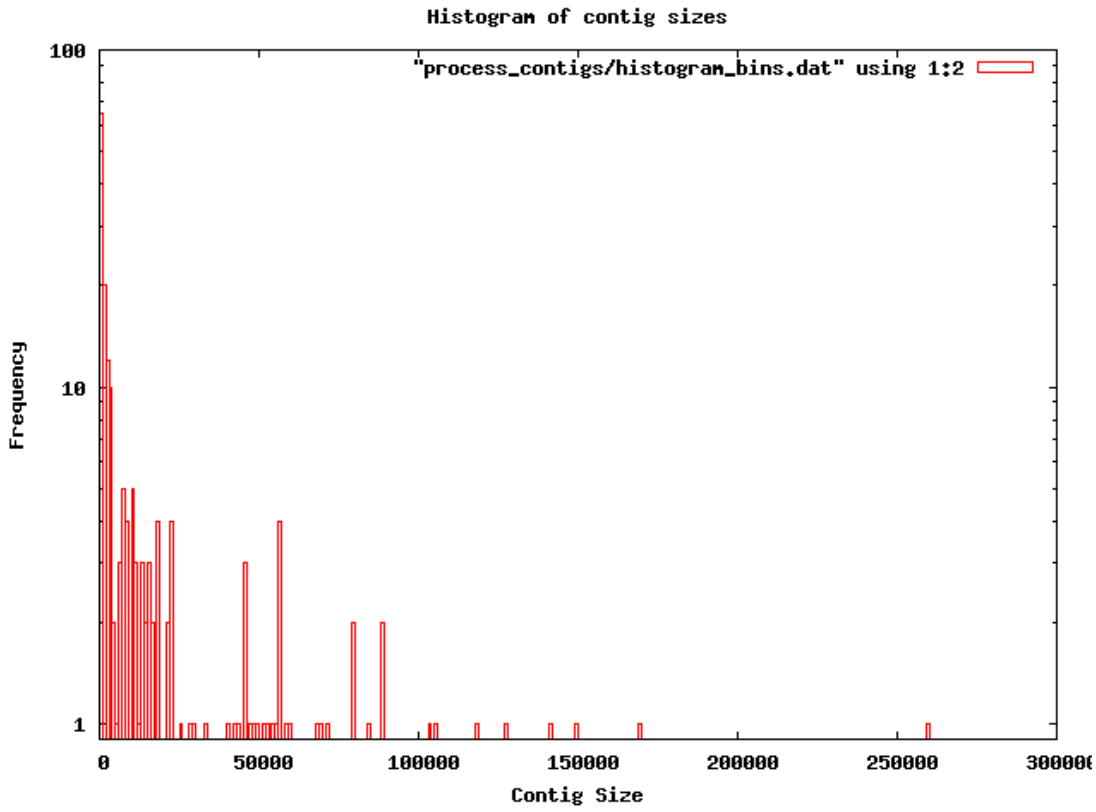


Figure 29. Numbers and sizes of contigs obtained in combined assembling of reads generated with SOLEXA and 454 genome sequencing. This histogram summarizes the number of contigs (vertical axis) of specific sizes in bases (horizontal axis).

N50 (bases)	68827
Maximal contig size (bases)	259427
Number of bases in contigs	3593869
Number of contigs	191
Number of contigs >=1 kb	126
Number of contigs in N50	16
Number of bases in contigs >=1kb	3562840
GC Content of contigs (%)	54.307460845122

Table 7. Summarized statistical data for combined assembling of reads generated with SOLEXA and 454 sequencing of strain MV-1. This table shows statistical summary of contig numbers and sizes obtained with 454 sequencing.

Although the total number of contigs was decreased from 2300 with SOLEXA sequencing alone to 493 with 454 sequencing and followed by combined assembling that resulted in the total number of 191 it was decided to attempt to improve the data by utilization of other sequencing methods.

3.3.3 The generation of the MV-1 genome sequence using Sanger method and plasmid libraries

In order to improve the sequence of the genome by joining contigs it was decided to obtain additional sequence by performing Sanger sequencing of the plasmid libraries obtained in the previous part of this work (See: Investigation of the "magnetosome island" in strain MV-1 with variation of inverse PCR and use of plasmid libraries).

Plasmid libraries produced by EcoRI restriction enzyme digestion of the chromosome were amplified by transformation into *E. coli* cells. Individual colonies were then subcultured into wells of two 96 well plates and incubated overnight with antibiotic (Ampicillin). Obtained cultures were submitted to the University of Edinburgh sequencing service (Gene Pool) for sequencing.

This analysis was expected to generate up to 150 kb of sequence (two plates by 96 wells by up to 800 bases of sequence each). It can be concluded that only 40 % of the total number of sequencing reactions have produces good quality sequence. This can be explained by several reasons. Firstly, there were some unreadable double sequences possibly due to the fact that a colony picked up for subculturing was a mixture of cells containing different plasmids. Secondly, it was possible due to non-

specific annealing of the sequencing primers. Finally, it was possible that some reactions were not successful.

This analysis has allowed to generate some sequence that was used later in the quality control alignments and to manually join contigs.

3.3.4 The manual joining of contigs

To generate manual joins of the contigs obtained in automated assembly several strategies have been used. This part of the genome sequencing was carried out by Bruce Ward and Denis Trubitsyn.

One of the used strategies was to look for alignments of overlapping fragments. Automated assembling software was using 25 bases to consider joining as positive; however when done manually it was possible to outline candidates for joining using much shorter overlapping fragments and then check it with PCR. Primers for PCR were designed at the end of the each of potentially joinable contigs to amplify the fragment across the join. Amplifications were followed by size estimation with an agarose gel electrophoresis and Sanger sequencing.

Another strategy used to join contigs was to look at the Open Reading Frames at the end of the contigs in order to identify those returning BLAST results for the same protein. The potential joins were checked with PCR and sequencing as described in the above paragraph.

This part of the project has allowed production a large number of manual joins and reduction the total number of fragments to 87. The work on joining contigs is not completed and is in progress.

3.3.5 The control of the sequence quality generated by different methods

Some additional 454 and Sanger sequencing of the strain MV-1 chromosome was carried out in collaboration with RIKEN (Japan). 454 sequncing generated in RIKEN provided 615,784 reads that is equal to 238,942,922 bases. The Sanger paired read sequencing (sequencing from the both sides of the fragment) was carried out for 2,622 pairs (5,244 reas). With the availability of this additional data it was decided to carry out the comparison of assembled contigs generated by SOLEXA and 454 sequencing methods using 454 and Sanger sequence generated by the collaborators in RIKEN as references. This analysis had two main purposes: to cross-check quality of the sequence generated in independent sequencing facilities and to do preliminary annotation using an internal BLAST program.

In order to add some automation to the process special piece software was developed in collaboration with Dr. Hongwu Ma. This software allowed the generation of the alignments between those contigs produced by SOLEXA and 454 methods. An example of such comparison and the algorithm of the decision making is shown below (Figure 30).

As it is demonstrated on the figure above the mismatch is identified and then partial sequence generated by the collaborators in Japan is used as a reference to make a decision if an extra base is left in the final sequence. The decision is made towards the variant that is homologues to those generated by 454 or Sanger sequencing (RIKEN, Japan).

In some rare cases it was impossible to find the region required as a reference for comparison or the all sequences contained contradicting sequence which was making

Figure 30. An example of the sequence quality control. This diagram demonstrates an alignment of two fragments of the genome sequence generated by SOLEXA and 454 methods. A point of mismatch is highlighted.

it impossible to make a logical decision. In these situations a fragment was analysed by using the ORF finder (National Center for Biotechnology Information web site). The variant of the sequence that produced a longer ORF and returned a higher score in BLAST analysis was chosen as a final. The diagram below demonstrates such analysis carried out for the fragment shown above (Figure 31).

This analysis was carried out by Denis Trubitsyn and Bruce Ward and has allowed an efficient comparison of the sequences followed by generation of the genomes sequence where a number of mistakes was identified and corrected.

3.3.6 Identification of magnetosome formation related genes within the genome sequence

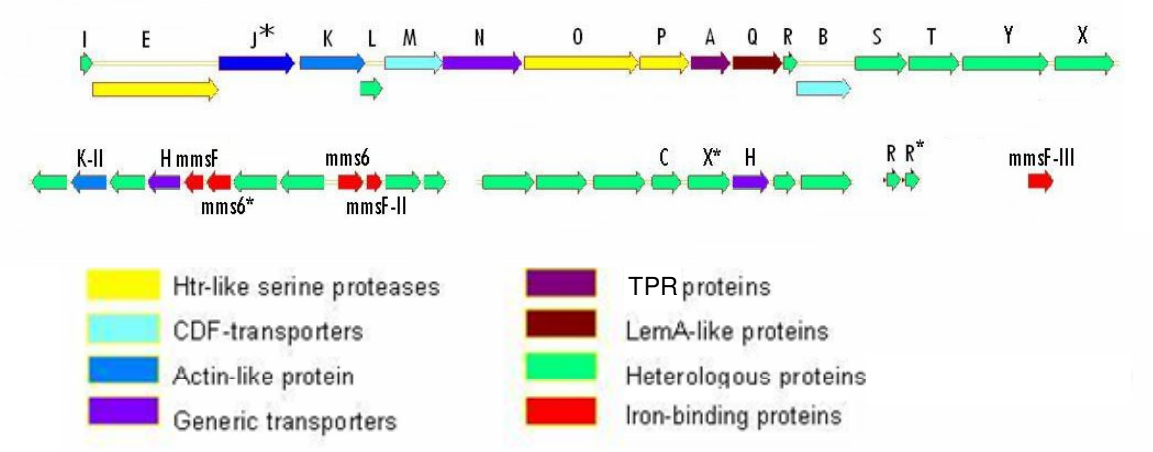
Once the first set of contigs was assembled after SOLEXA sequencing one objective was to identify further genes involved into the production of magnetosomes to those identified before in this work by inverse PCR.

In order to achieve this all 2300 contigs were uploaded as a database in VectorNTI 10 software (Invitrogen). This software was used to search within the database for specific sequences allowing a certain percentage of mismatches. The most conserved regions of the “magnetosome island” genes from other previously sequenced strains of magnetotactic microorganisms were selected and searched for matches within the genome.

Figure 31. The comparison of the Open Reading Frames found in the mismatching regions of the sequence produced by 454 and SOLEXA sequencing methods. This diagram demonstrates the differences in the length of the predicted ORF followed by the BLAST analysis. The fragment produced with SOLEXA (shown on the left) gives a longer predicted ORF and higher BLAST scores compared to the fragment produced by 454 sequencing that is containing an extra Thymine base (shown on the right). This diagram is adapted from the NCBI web-site (ncbi.nlm.nih.gov).

This analysis has allowed the fast and effective identification of "magnetosome island" organization in marine vibrio MV-1. As a result a molecular organization of the "magnetosome island" of marine vibrio MV-1 is shown on the next figure (Figure 32). Table below summarizes all putative proteins that have been identified including their sizes, predicted localization in the cell and BLAST search results (Table 8). Two different pieces of software were used to predict localization of the analysed proteins: the PSORT (Prediction of Protein Localization Sites, version 6.4) and CELLO (subCELlular LOcalization predictive system) (Gardy *et al.*, 2005; Yu *et al.*, 2004).

Interestingly, in some cases results predicted by PSORT and CELLO were different. Such inconsistency can be explained by differences in used algorithms. Developers of CELLO claim that PSORT-B will correctly predict localization of the proteins with an average success rate of 75%, whilst CELLO provides a higher success rate of 89%. It is also interesting, that PSORT-B provides a greater accuracy for proteins with sub-cellular localizations but its algorithm is less reliable for cytoplasmic and periplasmic sequences (Yu *et al.*, 2004). In some cases it was possible to establish the localization based on the putative or shown function. For example, Mms6 is an iron-binding protein involved in the growth of a magnetosome crystal and therefore its localization predicted by CELLO as "periplasmic/extracellular" is wrong. Another example is the incorrect prediction of localisation of proteins similar to Ccm (Table 9). It is established that these proteins are localized in cell membranes and their periplasmic localization predicted by CELLO is incorrect.



Functions of the genes in *M. magneticum* AMB-1:

mamI	Required for MF and MCF	***mamR***	Crystal number and size
mamE	Required for MF	***mamB***	Required for MF and MCF
mamJ	Chain assembly	***mamS***	Crystal morphology and size
mamK	Chain assembly	***mamT***	Crystal growth
mamL	Required for MF and MCF	***mamY***	Magnetosome membrane tabulation
mamM	Required for MF	***mms6***	Biomineralization
mamN	Required for MF	***mamH***	Not essential
mamO	Required for MF	***mamC***	Size control
mamP	Crystal number and size		
mamA	Magnetosome activation		
mamQ	Required for MF and MCF		

Figure 32. Organization of "magnetosome island" in marine vibrio MV-1. This diagram shows schematic organization of genes that are suggested to be involved in magnetosome formation. The colour-coded legend shows predicted functions of the proteins. Functions of the homologous genes shown in *M. magneticum* AMB-1 were taken from the original publication (Murat *et al.*, 2010).

Gene name	Protein size (aa)	Predicted cell localization		BLAST results		
		PSORT	CELLO	Description	E value	Organism name
mamI	68	Unknown	Inner membrane	MamI	2e-17	uncultured bacterium
mamE	705	Unknown	Periplasmic	MamE, putative trypsin-like serine protease	2e-144	uncultured bacterium
*mamJ**	179	Unknown	Extracellular	breast cancer protein 1	0.11	-
mamK	360	Cytoplasmic	Cytoplasmic	MamK, MreB-actin-like	7e-141	uncultured bacterium
mamL	79	Cytoplasmic membrane	Inner membrane	conserved uncharacterized protein	3e-10	uncultured bacterium
mamM	318	Cytoplasmic membrane	Inner membrane	MamM, Cation efflux family	1e-100	uncultured bacterium
mamN	438	Cytoplasmic membrane	Inner membrane	MamN	9e-145	uncultured bacterium
mamO	707	Cytoplasmic membrane	Inner membrane	Trypsin-like serine protease	3e-149	uncultured bacterium
mamP	316	Cytoplasmic membrane	Periplasmic	LemA	2e-66	uncultured bacterium
mamA	215	Cytoplasmic	Cytoplasmic	TPR repeat, MamA	2e-59	uncultured bacterium
mamQ	270	Unknown	Inner membrane	MamQ, LemA family	3e-74	uncultured bacterium

Gene name	Protein size (aa)	Predicted cell localization		BLAST results		
		PSORT	CELLO	Description	E value	Organism name
mamR	83	Unknown	Cytoplasmic	MamR	1e-18	uncultured bacterium
mamB	309	Cytoplasmic membrane	Inner membrane	MamB, Cation efflux family	7e-120	uncultured bacterium
mamS	*184*	Unknown	Extracellular	MamS	1e-41	uncultured bacterium
mamT	178	Unknown	Periplasmic	hypothetical protein amb0976	6e-37	*M. magneticum* AMB-1
mamY	333	Cytoplasmic membrane	Inner membrane	Methyl-accepting chemotaxis protein	6e-19	*M. magnetotacticum* MS-1
mamX	871	Cytoplasmic membrane	Cytoplasmic	tetratricopeptide domain-containing protein	7e-43	*Magnetococcus* sp. MC-1
ORF8	225	Cytoplasmic membrane	Cytoplasmic/ Periplasmic	hypothetical protein An16g04460	1.1	*Aspergillus niger*
Mamk -II	356	Cytoplasmic	Cytoplasmic	MamK, MreB-actin-like	1e-139	uncultured bacterium
ORF6	289	Unknown	Cytoplasmic/ Periplasmic	nuclear transcription factor	1.1	*Ricinus communis*
mamH	436	Cytoplasmic membrane	Inner membrane	MamH, major facilitator superfamily	1e-160	uncultured bacterium

Gene name	Protein size (aa)	Predicted cell localization		BLAST results		
		PSORT	CELLO	Best match description	E value	Organism name
mmsF -I	107	Cytoplasmic membrane	Inner membrane	MamF	7e-34	uncultured bacterium
*mms6**	155	Unknown	Periplasmic/ Extracellular	bacterial magnetic particle specific iron-binding	0.37	*M. magneticum*
ORF2	421	Cytoplasmic membrane	Periplasmic	hypothetical protein amb0908	1e-50	*M. magneticum* AMB-1
ORF1	215	Cytoplasmic membrane	Extracellular	uncharacterized protein	7e-23	uncultured bacterium
mms6	77	Unknown	Cytoplasmic/ Periplasmic	Mms6	2e-06	*M. gryphiswaldense MSR-1*
mmsF -II	107	Cytoplasmic membrane	Inner membrane	MamF	9e-30	uncultured bacterium
ORF3	194	Unknown	Cytoplasmic/ Periplasmic	adhesin	0.002	Mycoplasma hyopneumoniae
ORF4	106	Unknown	Cytoplasmic	hypothetical protein RPE_3722	0.006	*Rhodopseudomonas palustris* Biza53
ORF1	305	Cytoplasmic	Cytoplasmic	conserved uncharacterized protein	3e-125	uncultured bacterium
ORF2	363	Unknown	Cytoplasmic/ Extracellular	conserved uncharacterized protein	5e-85	uncultured bacterium

Gene name	Protein size (aa)	Predicted cell localization		BLAST results		
		PSORT	CELLO	Description	E value	Organism name
ORF3	327	Unknown	Inner membrane/ Periplasmic	Serine/threonine protein kinase	2e-104	uncultured bacterium
mamC	113	Unknown	Cytoplasmic/ Periplasmic	MamC	5e-15	*M. gryphiswaldense MSR-1*
*mamX**	110	Unknown	Periplasmic/ Cytoplasmic	MamX	5e-19	*M. gryphiswaldense MSR-1*
mamH	630	Cytoplasmic membrane	Inner membrane	ferric reductase domain-containing protein	1e-134	*Magnetococcus* sp. MC-1
ORF7	102	Unknown	Cytoplasmic	septum site-determining protein MinC	1.9	*E. albertii* TW07627
ORF8	370	Cytoplasmic membrane	Inner membrane	permease	1e-87	*gamma proteobacterium* HTCC5015
mamR	96	Unknown	Cytoplasmic/ Periplasmic	MamR	2e-06	uncultured bacterium
*mamR**	115	Unknown	Cytoplasmic/ Inner membrane	NM21-3	2e-08	*M. gryphiswaldense*
mmsF-III	88	Cytoplasmic membrane	Inner membrane	MamF	6e-20	uncultured bacterium

Table 8. Magnetosome formation proteins in the genome of marine vibrio MV-1. The inconsistency in prediction of cell localizations by PSORT and CELLO is addressed in the main text. The best match from BLAST search was taken as a description of the protein.

Interestingly, as can be seen from the Table 8, many of the proteins from marine vibrio MV-1 "magnetosome island" find the highest similarity with the recently obtained sequences of an uncultured magnetotactic bacterium (Jogler *et al.*, 2009b).

Another interesting observation is that some of the magnetosome formation genes within the genome of strain MV-1 are present in homologous copies. There are 2 homologues of the gene *mamK*, 2 copies of *mamX*, 3 genes similar to *mamR* and finally, 3 homologues of *mmsF*, one of which is found to be located outside of magnetosome island.

The clustalW alignment of the amino acid sequences of the homologues of MmsF found in strain MV-1 shows significant differences (Figure 33). The flanking sequence around the region encoding MmsF-III was determined in two ways: from analysis of the sequence obtained by Sanger sequence generated in RIKEN and by inverse PCR using a sequence within the *mmsF*-III region. Two independently obtained sequences showed that *mmsF*-III was flanked on one side by a gene encoding an IS21 transposase. The data from the iPCR showed on the otherside a partial pseudogene for an Integration Host Factor lacking the N-terminal region. The fact that MmsF-III is followed by the sequence of a transposase is an additional possible evidence of the acquisition of the magnetosome formation genes through horizontal gene transfer.

3.3.7 An automated annotation of the MV-1 genome using RAST

The latest version of the assembled automatically and manually contigs was uploaded to the RAST (Rapid Annotation using Subsystem Technology) web-site. This resource allows an automated annotation of the prokaryotic genomes and a private storage of the data prior to the publication.

```
CLUSTAL 2.0.12 multiple sequence alignment

MmsF-I     MAKKTVRSKGGFRARMLAIMSYLGILCFVPLMRGRDDEFVYFHARQGLIIWMLGVVGIFS 60
MmsF-II    MAKQVVRRESGIHSYAMGILSYMGVLCLVPLITNRDDEFIHFHAKQGLVIWMWSVLAIMA 60
MmsF-III   -------------------MAYLGTLCFIPLMVTDRDAFVLFHARQGVVLWGWTVVAGFS 41
                              ::*:* **::**:    * *: ***:**:::*   *:. ::

MmsF-I     LYIPGLGKWMFTTSLFFVLVLSIIGVISVFLHRAWKLPMIHTLSTYI 107
MmsF-II    LYMPGLGKFFFSSSAMLIVLASVIGIVSVLFSRAWKLPVIHNISTKI 107
MmsF-III   LFIPGIGGPIFVFSLIGVVGFSVAGIVSVVLRKTWKLPIIYXFVIAI 88
           *::**:*  :*  * : ::  *: *::**.: ::****:*: :   *
```

Figure 33. ClustalW alignment of the amino acid sequences of homologous copies of MmsF protein.

The results of the automated annotation with RAST were downloaded as a FASTA file which then was used as a database for mass spectrometry protein identification (see: Magnetosome membrane proteins identification).

The data obtained with RAST annotation allows an extensive analysis of the genome sequence by comparison with other organisms, prediction of the function of the specific proteins as well as an attempt of reconstructions of the metabolic pathways of the organism.

A summarized data of the identified subsystem features allowed access to the first general overview of the genome of strain MV-1. This data is demonstrated on the next diagram (Figure 34).

There are several interesting observations can be made from analysis of this summarized data. These interesting features of the genome can be used for further analysis of the metabolism of the organism. For example there are no identified proteins that can be involved in photosynthesis which confirms that this organism does not rely on photosynthetic metabolism.

Interestingly this automated analysis has identified 47 features that may be involved in virulence. A closer look at this subsection is shown in the diagram below (Figure 35). As it can be concluded from the diagram below there is a number of interesting features that may be involved in resistance to antibiotics and toxic compounds particularly to those involved in copper homeostasis. It has been shown previously that a copper-dependent system may play a significant role in the production of magnetosomes by marine vibrio MV-1 and particularly its iron uptake (Dubbels et al., 2004).

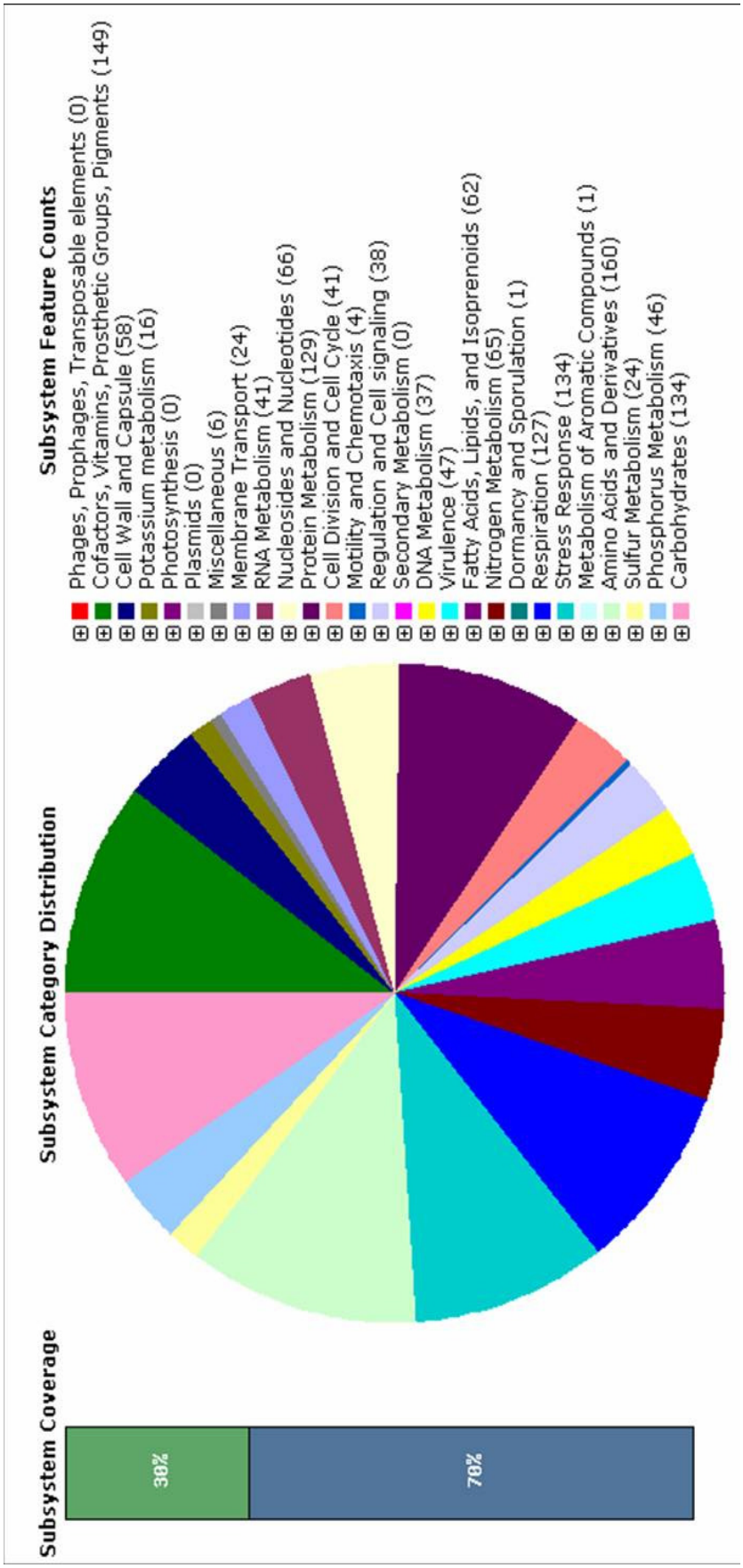


Figure 34. Subsystem category summary generated by RAST annotation. This diagram demonstrates the summary of the identified subsystems for strain MV-1 genome. Numbers of identified features are shown in brackets. Adapted from the RAST web-site.

Figure 35. Features combined under the Virulence subsection. This list provides information on the features that may be involved in virulence-related pathways.

This research suggested that the spontaneous mutants of strain MV-1 do not produce magnetosomes and an absence of expression of one of the genes encoding for copper handling protein involved in high affinity copper transport can cause an absence of biomineralisation.

All 11 proteins from the genome sequence that were combined into the copper metabolism feature were analysed for the predicted localization within the bacterial cell. The results of this analysis are summarized in the table below (Table 9).

The availability of the genome sequence together with its automated annotation can lead to outlining of the new targets for the research of magnetosome formation.

The analysis summarized in the above table can be used to outline additional proteins that may be involved in magnetosome formation. For example a copper chaperone may play a role in avoiding toxic effects of copper ions in copper delivery to the specific copper-dependant enzymes (Harrison *et al.*, 1999).

Another interesting gene found is the genome of a multicopper oxidase. There is research that suggests the involvement of such proteins in copper-dependent uptake

of Fe (II) (Huston *et al.*, 2002). This protein can be a good target for further analysis as any iron acquisition system may play a significant role in magnetosome formation.

Predicted protein function	Contig	Start	Stop	Size (aa)	Predicted cell localization	
					PSORT	CELLO
Cu(I)-responsive transcriptional regulator	MVN31	9983	9567	139	cytoplasm	cytoplasm
Type cbb3 cytochrome oxidase biogenesis protein CcoI Copper-translocating P-type ATPase (EC 3.6.3.4)	MVN56	69880	67418	821	inner membrane	inner membrane
Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5) Copper-translocating P-type ATPase (EC 3.6.3.4)	MVN22	7935	9770	612	inner membrane	inner membrane/ cytoplasm
Copper chaperone	MVN31	7964	7764	67	cytoplasm	cytoplasm/ periplasm
Multicopper oxidase	MVN56	3388	1943	482	inner membrane	cytoplasm
Putative diheme cytochrome c-553	MVN5	954	1727	258	cytoplasm	periplasm
Cytochrome c heme lyase subunit CcmH	MVN12	120972	119746	409	inner membrane	inner membrane
Cytochrome c heme lyase subunit CcmL	MVN12	121496	120969	176	inner membrane	cytoplasm
Cytochrome c heme lyase subunit CcmF	MVN12	124045	122048	666	inner membrane	inner membrane
Cytochrome c-type biogenesis protein CcmG/DsbE, thiol: disulfideoxidoreductase	MVN12	122014	121463	184	inner membrane	inner membrane/ periplasm
Cytochrome c-type biogenesis protein CcmE, heme chaperone	MVN12	124512	124042	157	inner membrane	cytoplasm/ periplasm

Table 9. Proteins combined by involvement into the copper metabolism and their predicted cell localization. This table summarizes 11 proteins that were combined by automated annotation by RAST into the copper homeostasis feature. Predicted cell localizations were analysed by the indicated software.

3.3.8 Transposases in the genome of marine vibrio MV-1

One of the main difficulties during the assembly of the contigs into a single fragment of genome sequence is the presence of numerous transposases and transposase-related genes (TRG). These elements with similar sequence are often found at the end of the automatically assembled contigs. Contig fragments with identical sequence at the ends can not be joined through sequence alignment analysis due to the ambiguity of their possible localization.

It was decided to carry out analysis of the transposases and TRG present in the sequence of marine vibrio MV-1 and compare the acquired data with those available from the genome sequences of other microorganisms. The total number of transposases and TRG in the genome of strain MV-1 can not be calculated with high accuracy at this stage. The reason behind this statement is explained by the fact that transposases found at the ends of the contigs can be actually parts of the same transposase but not yet be joined into the one contig and therefore counted twice. The total numbers of such elements in the available genomes are summarized in the table below (Table 10). All ORFs identified as transposases and "transposase and inactivated derivatives" were taken into the calculation.

Name of the organism	Genome size (Mb)	Predicted number of transposases and TRG genes	Predicted average number of transposases and TRG per megabase
Sinorhizobium meliloti 1021	3.6	73	20.3
Rhodospirillum rubrum ATCC 11170	4.3	16	3.7
Magnetospirillum magneticum AMB-1	4.9	46	9.4
Magnetococcus sp. MC-1	4.6	56	12.2
Desulfovibrio magneticus RS-1	5.4	55	10.2
marine vibrio MV-1	3.6	29	8.1

Table 10. Total numbers of transposases and transposase-related genes found in the genomes of different members of *Proteobacteria* and other magnetotactic microorganisms. This table provides examples of the numbers of transposase-related genes found in the different microorganisms. The number of transposases was divided by the size of the genome to find numbers of transposases and TRG per megabase of sequence to give a better interpretation of the data. The genomes were accessed through Genebank.

As it can observed from the above table the predicted average number of transposases in the genome of marine vibrio MV-1 is slightly higher than that of some other members of the Proteobacteria. For example it is 2 times higher than the number of transposases in the genome of *Rhodospirillum rubrum* ATCC 11170. When compared to other magnetotactic microrganisms the number is similar to that in *M. magneticum* AMB-1 and *Magnetococcus sp.* MC-1.

Once again it is important to outline that the total number of transposases in strain MV-1 will become lower with finishing of the joining of fragments of the genome. For the same reason the number of transposases in another magnetotactic microorganism *M. gryphiswaldense* MSR-1 (not shown in the table) is difficult to take into the account for comparison. The available data from the genome sequencing project for this organism contains 373 unordered fragments and the search for transposases returns a comparatively high number of 206. This number is likely to be significantly lower when the fragments are joined but gives another example of difficulties that face researches attempting to produce sequences of the bacterial genomes.

As was stated above it is difficult to analyse the total number of transposases in the unfinished sequence of the genome. It was decided to search for different types of TRG in the sequence of strain MV-1. This search resulted in identification of 17 different types of TRG. The summarized results of this search are shown in the table below (Table 11).

To summarize, no analysis was carried out on comparison of transposases found in the genome of different MB. These elements with repetitive sequences appear to be

Transposase	Number found in genome of strain MV-1
ISSpo6, Transposase orfB	1
Transposase (class II)	3
Transposase	5
Transposase for insertion sequence element IS406	1
Putative IS1016 Transposase	2
Transposase protein B	1
ISSod2, Transposase OrfA	3
ISxac2 Transposase	1
Transposase, IS110 and IS111A family	1
Transposase and inactivated derivatives	2
Transposase IS3/IS911	2
Transposase, IS4	2
Transposase IS66	1
Probable remnant of a transposase gene protein	1
Transposase, mutator type	1
ISMca6, Transposase, OrfA	1
ISSod11, Transposase	1

Table 11. Different transposase-related genes in the genome of marine vibrio MV-1.

present in a relatively high numbers in MB. This is one of the main difficulties that face researchers working on the genome sequences of this group of organisms. For example the contigs generated in genome sequencing of *M. gryphiswaldense* MSR-1, as mentioned above, contains 206 transposase-related genes. It is interesting to investigate if a relatively high number of transposases in chromosomes of MB can increase their susceptibility to accept the "magnetosome island" through horizontal gene transfer.

3.3.9 Comparison of the genome of marine vibrio MV-1 and other sequenced magnetotactic bacteria

With the availability of automatically annotated genome sequence of marine vibrio MV-1 it was interesting to compare the sequence to other magnetotactic microorganisms to find if the protein sequences show high similarity. At the first stage a comparison of the sizes of the genomes and G+C content was carried out (Table 12).

In order to achieve this goal the genomes were compared using the SEED viewer, a piece of on-line software designed to view and analyse sequences annotated by RAST. The SEED viewer allows to compare genome sequences and to visualize such comparison by producing a coloured diagram with the link between level of similarity and the specific colour. The results of this analysis are shown on the next diagram (Figure 36).

Species	%GC	Size (Mb)	Number of contigs	Number of predicted ORFs
M. gryphiswaldense MSR-1	62.8	4.26	373	4,264
M. magnetotacticum MS-1	64.0	4.5	316	4,925
M. magneticum AMB-1	65.1	4.97	1	4,559
Magnetococcus strain MC-1	54.8	4.7	1	3,716
Desulfovibrio magneticus RS-1	62.8	5.4	1	4,629
marine vibrio MV-1	54.3	3.6	87	3,471

Table 12. The comparison of the sizes and GC content of the genomes of magnetotactic microorganisms. Data for strain MV-1 was generated in this study. Genomes of other magnetotactic microorganisms were accessed through Genebank.

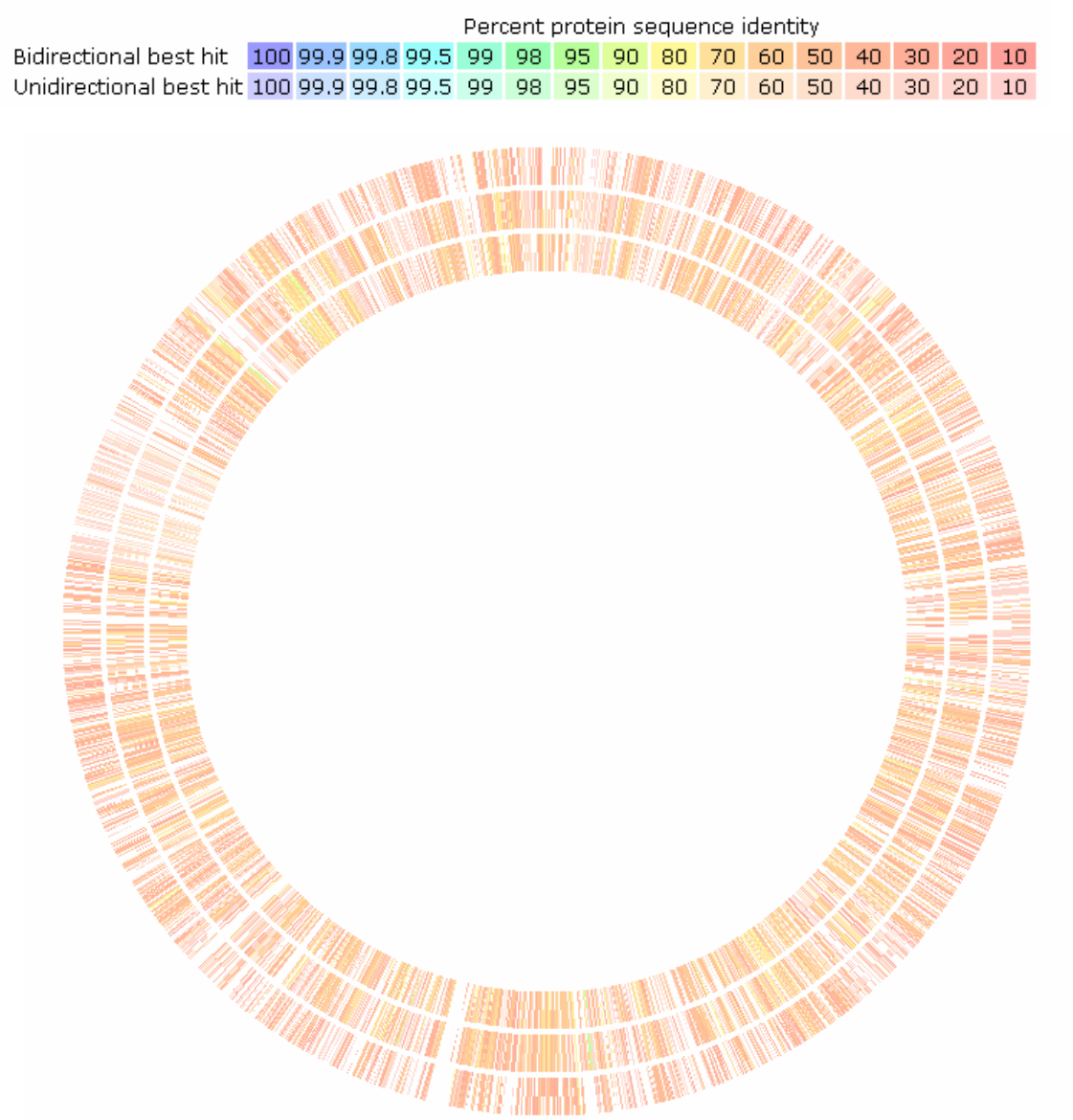


Figure 36. Comparison of the genome sequences of marine vibrio MV-1 and other magnetotactic bacteria. This diagram was produced by SEED viewer to visualize levels of similarity of the genome sequences. The genome of strain MV-1 was compared to genomes of *Magnetococcus sp*. MC-1, *M. gryphiswaldense* MSR-1 and *M. magneticum* AMB-1 and demonstrated as outer, middle and inner circles respectively.

As it can be observed on the above diagram the levels of similarity vary dramatically in each of the comparisons which can be expected for such microorganisms that are different in metabolisms, cell shapes and natural habitats.

This piece of software joins all contigs of unfinished genome automatically into the single sequence fragment and allows production of a Blast Dot Plot for every two compared sequences. This feature is demonstrated on the next diagram by comparison of genomes of marine vibrio MV-1 and *M. magneticum* AMB-1 (Figure 37).

This comparison is another demonstration of the lack of direct similarity of the organization of the genomes with only small number of regions forming short uninterrupted lines. However this analysis is not final due to the fact that contigs of the MV-1 genome were connected automatically and may not represent the real genome arrangement.

Better results were achieved in alignments between individual contigs. For example contigs N2 (*mamAB* cluster) was aligned against the genome of *M. magneticum* AMB-1. The software MUMmer was used to carry out and visualize this alignment (Figure 38) (Kurtz *et al.*, 2004). This alignment shows much higher level of similarity within the chosen conserved region of *mamAB* cluster compared to those achieved in direct comparison of the genomes.

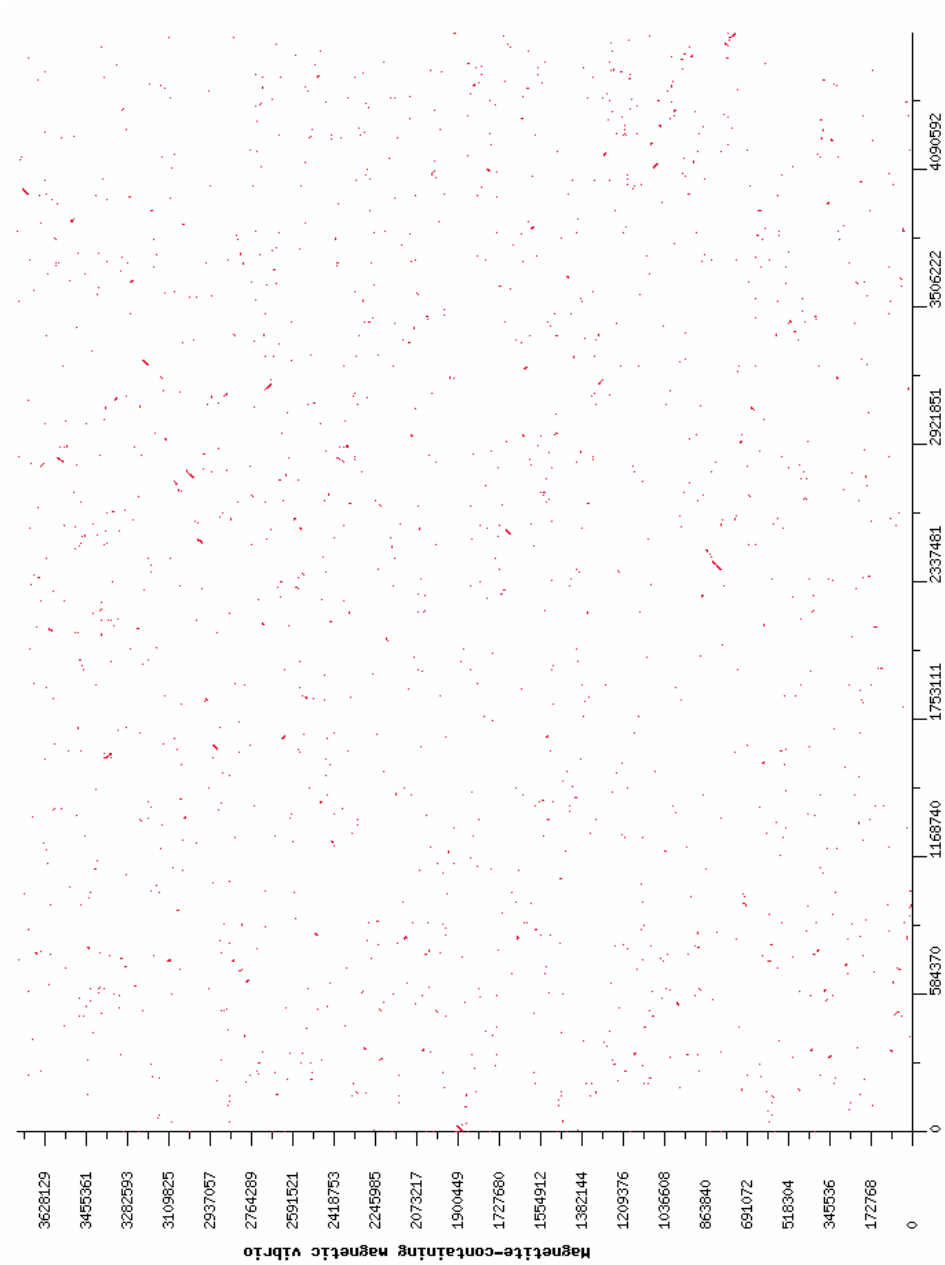


Figure 37. Blast Dot Plot of the genome sequence comparisons of marine vibrio MV-1 and *M. magneticum* AMB-1. This diagram demonstrates the schematic location of the similar sequences. The contigs of the genome of strain MV-1 were automatically joined into a single fragment of sequence. This diagram was produced with the use of SEED viwer

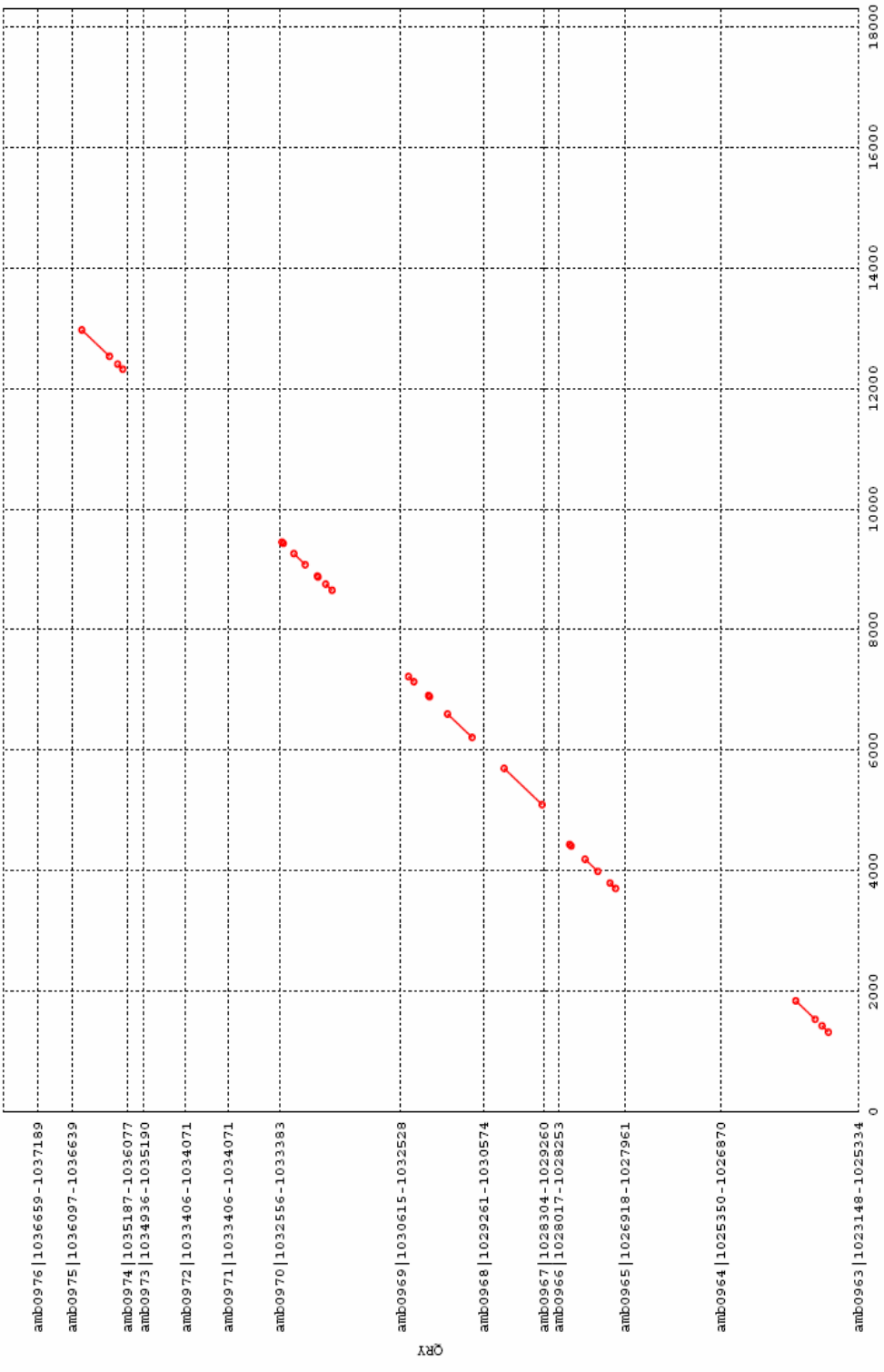


Figure 38. Blast Dot Plot of the contig N2 containing *mamAB* cluster of strain MV-1 and genome of *M. magneticum* AMB-1. This diagram demonstrates the schematic location of the similar sequences. The vertical axis represents the fragment of the genome of *M. magneticum* AMB-1 and the horizontal axis represents contig N2. This diagram was produced with the use of MUMmer.

3.3.10 Future work on the genome of marine vibrio MV-1

In order to complete the work on the genome of strain MV-1 contigs need to be joined together in order to produce one continuous fragment of sequence. This can be achieved by several approaches. The generation of additional sequencing data, preferably with random starting points and longer reads length such as with Sanger sequencing, should allow generation of a sufficient number of fragments that close gaps in the sequence by aligning to both end of the joined contigs. Another strategy is to continue joining of the contigs by trying to amplify by PCR regions across hypothetically joined contigs. This work is being continued, however, it requires testing of a high number of combinations of primers and a significant amount of time.

The generation of the completed sequence needs to be followed by a detailed annotation. With the availability of faster and cheaper methods of sequencing the number of sequenced genomes has grown dramatically over the last few years. The resources of manual annotation are very limited and therefore modern ways of annotation are needed. The automated annotation systems that are available at the moment produce fast annotation, however, results vary from one system to another suggesting that the algorithms used are different and there is no clear universally suggested approach. The widely used automated service systems include RAST (Rapid Annotation using Subsystem Technology), IMG (Integrated Microbial Genome at The Joint Genome Institute) and JCVI (J. Craig Venter Institute). A recent publication compares all three systems by analysis of the annotations produced for the identical sequence of an archaeon (Bakke *et al.*, 2009). Interestingly, there are very significant differences in the predicted gene sequences.

For example an average length of the predicted genes was higher for RAST (967 bp) followed by JCVI (940 bp) and IMG (934 bp). This difference is explained by the fact that RAST more often considers alternative to ATG start codons as a starting point of a gene. The use of alternative start codons by RAST, IMG and JCVI occurred in 39.0, 19.9 and 14.3 percent of the predicted genes respectively. There were also significant differences in association of predicted proteins with EC numbers that are used in universal enzyme classification. All these suggest that automated systems although quick and easy to operate can potentially generate a large number of errors that can lead to inaccurate analysis in the future. The annotation of the genome of MV-1 upon completion in the form of a single fragment by all three systems can be suggested as a future work.

3.4 Magnetosome membrane proteins

3.4.1 Isolation of magnetosome membrane proteins

Proteins that are closely attached to magnetic crystals are likely to be important in the process of magnetosome formation. In order to investigate the composition of these proteins an isolation of magnetosomes from the rest of the cell debris was carried out. There is one main feature that allows a relatively easy method of separating such particles from the rest of the broken bacterial cells – the ability to separate them by applying a magnetic field to the mixture.

Although the main principle behind different methods of separation is the same, there are several variations in methods that are used by different research groups investigating magnetosome membrane proteins. Some of the advantages and disadvantages of the proposed methods will be discussed in the following sections.

3.4.1.1 Bacterial cells disruption

In order to break cells several research groups suggest passing a cell suspension through a French pressure cell. This approach was used in this work and has shown similar results to those previously reported (Grunberg *et al.*, 2004; Tanaka *et al.*, 2006). However, another method of cell disruption was tested in this study: cells were disrupted by a high energy ultrasound (see Materials and methods section for details). The use of this method can be beneficial in situations where a French pressure cell is not available or when it is essential to keep low sample volumes. Another advantage of this method is that it allows addition of protease inhibitor (phenylmethylsulphonyl fluoride) before cell disruption without formation of the foam as was observed when using a French pressure cell.

The possible effects of the ultrasound on the composition of the membranes were investigated. There are suggestions that ultrasound can cause membranes to invert and therefore potentially loose some important proteins however similar findings were shown for the use of a French pressure cell (Futai, 1974). If this effect is taken into account it is important to outline that in this work it was decided to attempt to limit the number of proteins for investigation to those that are closely and firmly attached to the particle and therefore are more likely to be truly involved in the process of magnetosome formation. This obviously does not suggest that these proteins are the only proteins that are involved in the process of formation. The use of sonication therefore will be only beneficial and will help to remove some crude cell inner membrane proteins that tend to attach to the particles. Finally it was shown that protein pattern on the SDS gel is relatively similar to those shown in previous studies and contains a significant number of proteins (see below).

3.4.1.2 Magnetosome isolation

The main principle behind separating magnetosomes from the crude solution of disrupted bacterial cells is to wash magnetosomes several times with different buffers while collecting particles on the side of the tube with magnetic field between changing the buffers. One significant difficulty in this approach is that magnetosomes tend to form aggregates when collected on the side of the tube by the magnetic field (the strong permanent magnet was used). These aggregates can be easily observed by eye in the tube and can be seen on the TEM micrograph (Figure 39).

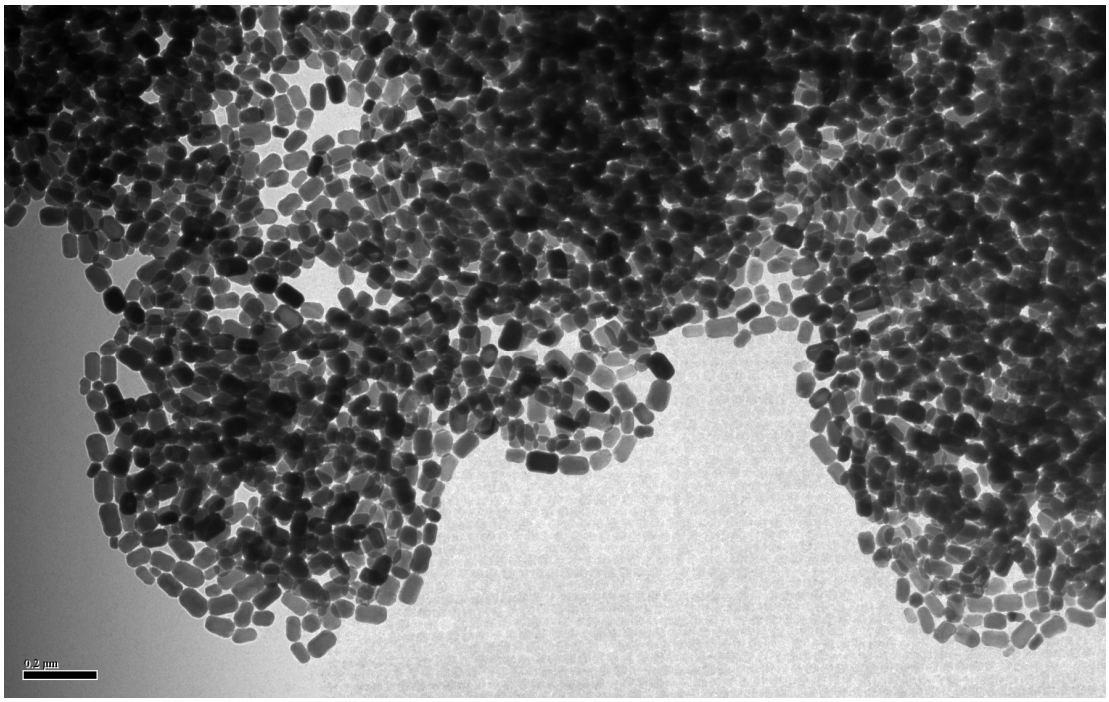


Figure 39. Magnetosome aggregates. The TEM micrograph of magnetosome forming aggregates. The bar represents 0.2 µm. This micrograph was produced in this work.

The reason for this is either the magnetic properties of the crystals or the aggregation of the membranes or both of these factors. These aggregates tend to form more often on the early stages of magnetosome isolation so it is possible that membranes play more important role than magnetic properties. The longer magnetosomes stayed exposed to the magnetic field the stronger these aggregates became and therefore more energy was needed to disaggregate them. On the other hand if magnetosomes were not exposed to magnetic field for long enough then some of the particles had stayed in the buffer suspension and were discarded when the buffer was changed. As a result of this observation it was concluded that the time of exposition to the magnetic field affects the balance between strength of the magnetosome aggregation and amount of sample loss during the isolation.

In order to help separate magnetosomes there are two methods were described before in the literature.

The first method used in the laboratory of Dennis Bazylinski is to simply resuspend magnetosomes in the fresh buffer with the pipette tip and the addition of high concentration of NaCl to the washing buffer (Gorby *et al.*, 1988). This method allows to isolate magnetosomes but requires a relatively high number of washes to achieve a significant difference in protein patterns between magnetosome proteins and cell inner membrane fraction.

The second approach that is used in the laboratory of Tadashi Matsunaga is to segregate magnetosome with weak sonication. In order to do this the flask containing magnetosomes suspended in a fresh buffer is submerged into a sonicating water bath for 5-10 seconds with slow shaking (Tanaka *et al.*, 2006). This approach allows to

segregate magnetosomes more easily. This method was also tested in this study and one significant disadvantage was noticed: if magnetosomes are left to collect on the bottom of the flask for 1 hour as it is suggested in the Tanaka's method than it requires much longer (up to 2-3 minutes periods of weak sonication – possibly due to the differences in the used equipment) to separate magnetosome aggregates. In order to avoid such situations the time periods between washes were decreased to 30 minutes however this lead to a significant loss of material with each wash cycle because not all magnetosomes have been collected by the magnet. As a result the number of washes had to be decreased from 20 to 10 and this approach did not allow to achieve such a significant difference in protein pattern on the gel between magnetosome and inner membrane fractions with method adapted in this study (see below).

After numerous test and comparisons an adapted method of magnetosome isolation was developed in this study. Bacterial cells were disrupted by sonication and then transferred into a 50 ml hand homogenizer (Thomas, No C18856). This homogenizer fits well into the standard Magnetic Separation Stand (Promega); however a self produced stand was used which was a plastic box with a hole on the top to fit the homogenizer and several strong neodymium magnets attached to the side.

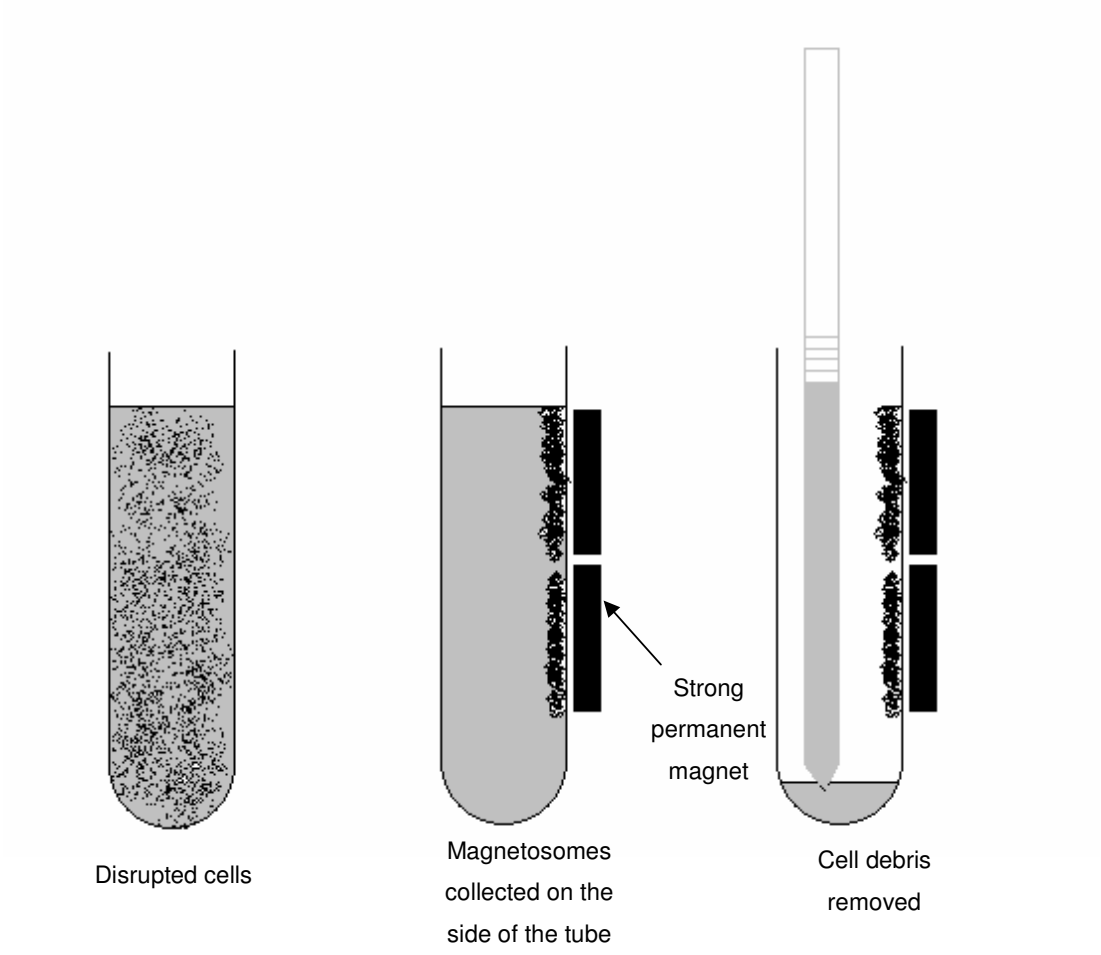


Figure 40. Magnetosome isolation. The homogenizer containing disrupted cells is placed into the magnetic separation stand and cell debris can be removed while magnetosomes are collected at the side of the homogenizer.

After 1-2 hours magnetosomes were collected on the side of the homogenizer cell debris can be removed by aspiration (Fig). The washing buffer containing 200 mM NaCl was added and the suspension was homogenized by repeated movement of the piston. This was sufficient to separate magnetosome aggregates. Then magnetosomes were washed as described in the Materials and methods section (2.4.24). Before the last wash with 10 mM HEPES (pH 7.4) they were treated by sonication for 30 seconds (output tune 5). This step has very significant effect on the whole process of magnetosome isolation and the difference can be observed on the SDS gel (see next section, Figure 43).

In the course of this study a number of variations of the methods were tested in order to optimize the magnetosome isolation. A total number of 42 magnetosome samples were prepared witch required approximately 60 g of wet weight cells harvested from 11 cultures 9.6 L each.

3.4.1.3 Magnetosome membrane proteins fraction isolation

Once magnetosomes were isolated and washed from cell debris, the magnetosome membrane fraction was isolated from the magnetite crystals. The method used in the laboratory of Dennis Bazylinski involves separating membranes from the crystals by addition of an anionic surfactant sodium dodecyl sulphate (SDS). This compound isolates membrane protein fraction very successfully and the evidence of this can be observed on the TEM micrographs (Figure 41). The effect of 1 % SDS on magnetosome membranes on these micrographs is similar to those previously reported for *Magnetospirillum gryphiswaldense* (Grunberg *et al.*, 2004).

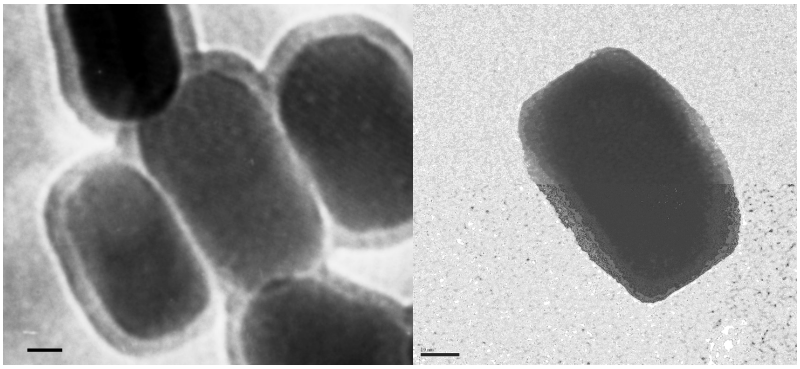


Figure 41. TEM micrograph of MV-1 magnetosomes. Magnetosomes crystals surrounded by a membrane (left) and a crystal after treatment with 1 % (w/v) SDS (right). The bar represents 10 nm. The micrograph was produced in this work.

In order to isolate inner membrane fraction the method of ultra speed centrifugation in sucrose gradient was used (Tanaka *et al.*, 2006). The As a result the red inner membrane fraction was collected and washed from sucrose (Figure 42).

After isolation of the magnetosome membrane fraction and inner membrane fractions the protein concentrations were determined in all samples by the Bradford essay. Equal amounts of protein were separated by SDS PAGE gel electrophoresis (Figure 43).

As it can be seen on this gel the magnetosome membrane fraction that was isolated by the method involving separation of the aggregates by weak sonication (MM*) does not show sufficient difference compared to the inner membrane fraction. The intermediate stage of the magnetosome purification by the method developed in this study (MM**) is shown to demonstrate the effect of the sonication step. Finally, several bands that appear to be the most different were selected and excised from the gel for identification.

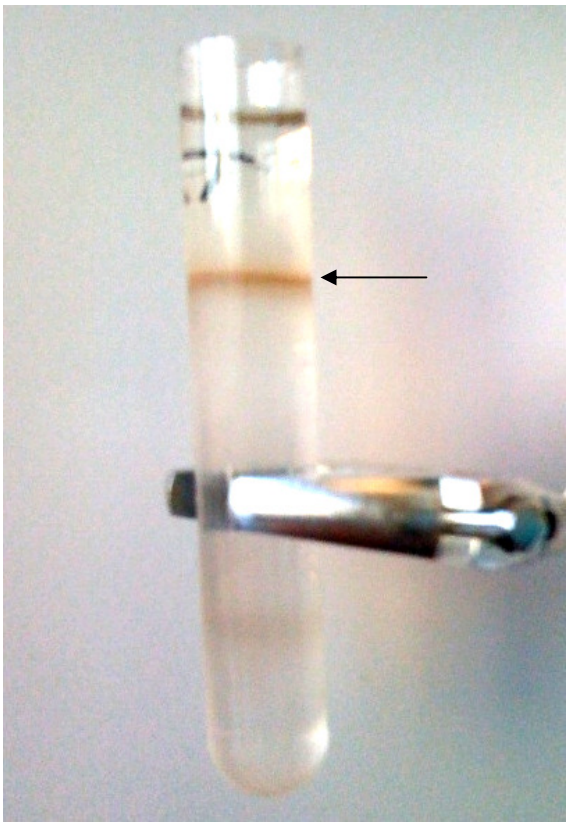


Figure 42. The inner membrane fraction. The photograph shows the inner membrane fraction isolated by the sucrose gradient centrifugation. The arrow shows the red inner membrane fraction.

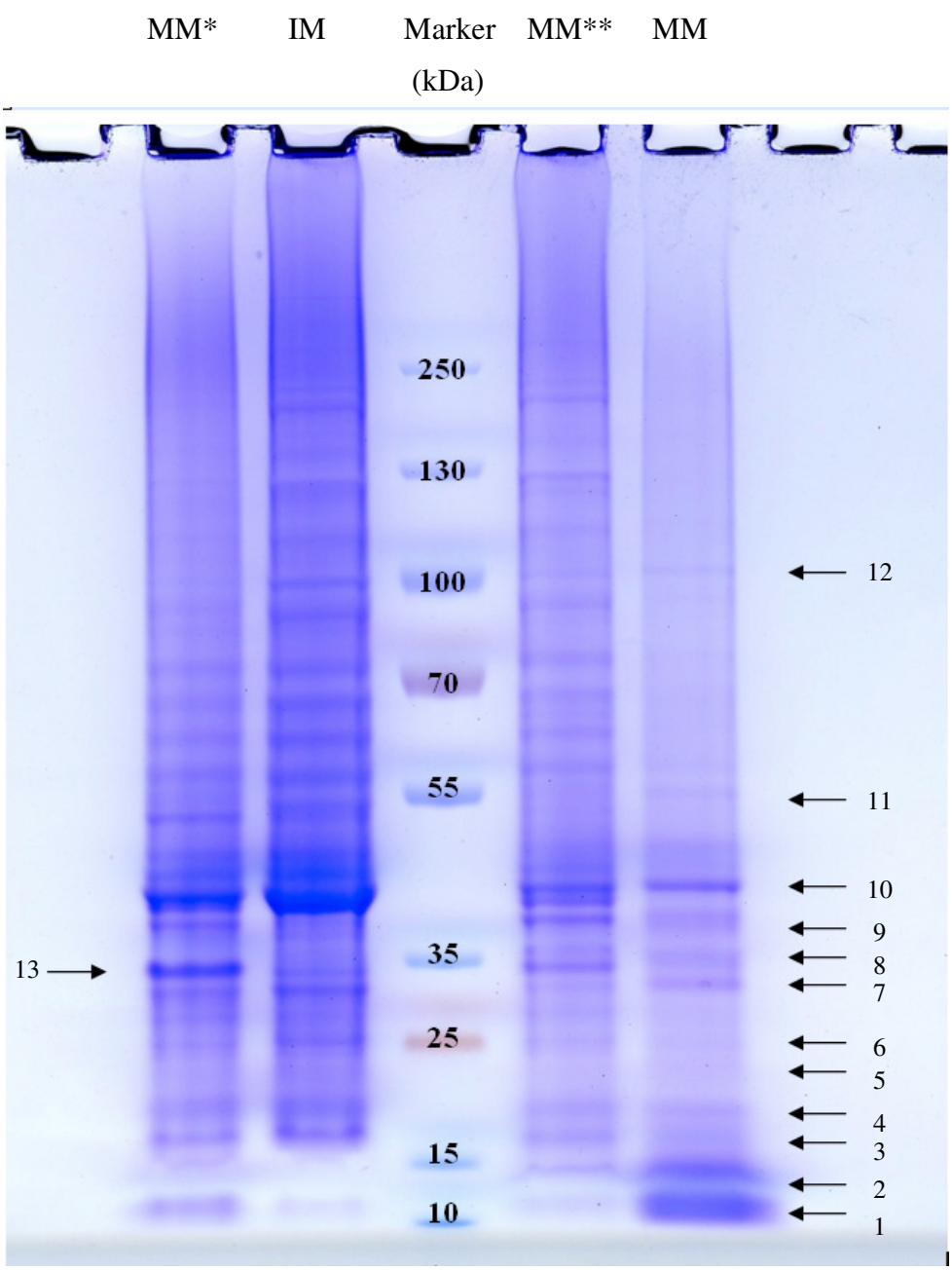


Figure 43. SDS PAGE gel of membrane fractions of marine vibrio MV-1. MM* – magnetosome membrane fraction preparation with weak sonication; IM – inner membrane fraction; MM** – an intermediate stage of magnetosome preparation; MM – final stage of magnetosome preparation. The arrows show protein bands selected for identification. NuPAGE 4-12% BisTris gel (Invitrogen); molecular weight marker PageRuler Plus Prestained Protein Ladder (Fermentas SM1811); 20 microgram of the protein was loaded onto each lane. Numbers with arrows show fragments that were excised and used for protein identification. These numbers directly correspond to the sample numbers used in the following sections.

3.4.2 Magnetosome membrane proteins identification

The masspectrometry (MS) approach was chosen to identify magnetosome membrane proteins. The bands of interest were excised from the SDS gel and subjected to trypsin digestion. Obtained peptides were analysed using several different masspectrometry instruments in order to identify target proteins. The methods and results for each of the method are described in detail in the following sub-sections. It is important to outline that the original analysis on the protein identification was carried out using the protein database produced and annotated in this study, however in order to avoid confusion and allow better data interpretation in the following sections protein identities were obtained by the search against 92 putative proteins from the 107 kb sequence of MV-1 "magnetosome island" deposited by the groups of Dennis Bazylinski and Dirk Schuler (Jogler *et al.*, 2009a).

3.4.2.1 MALDI TOF proteins identification

Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) analysis was used to identify peptides produced after trypsin digestion of the target proteins. The instrument Voyager-DE STR (Applied Biosystems) was used in this work.

One of the limitations of this method is that the amount of peptides in the processed sample has to be reasonably high and therefore sufficient to produce spectra that then can be analysed. In order to isolate a sufficient amount of samples for identification it was decided to test large format gels (20×20 cm), however this did not show any significant concentrating. An approach involving the use of ZipTips (Millipore) to concentrate sample was chosen and it has shown a significant improvement. An example of ZipTip concentrating can be observed on the diagram below (Figure 44).

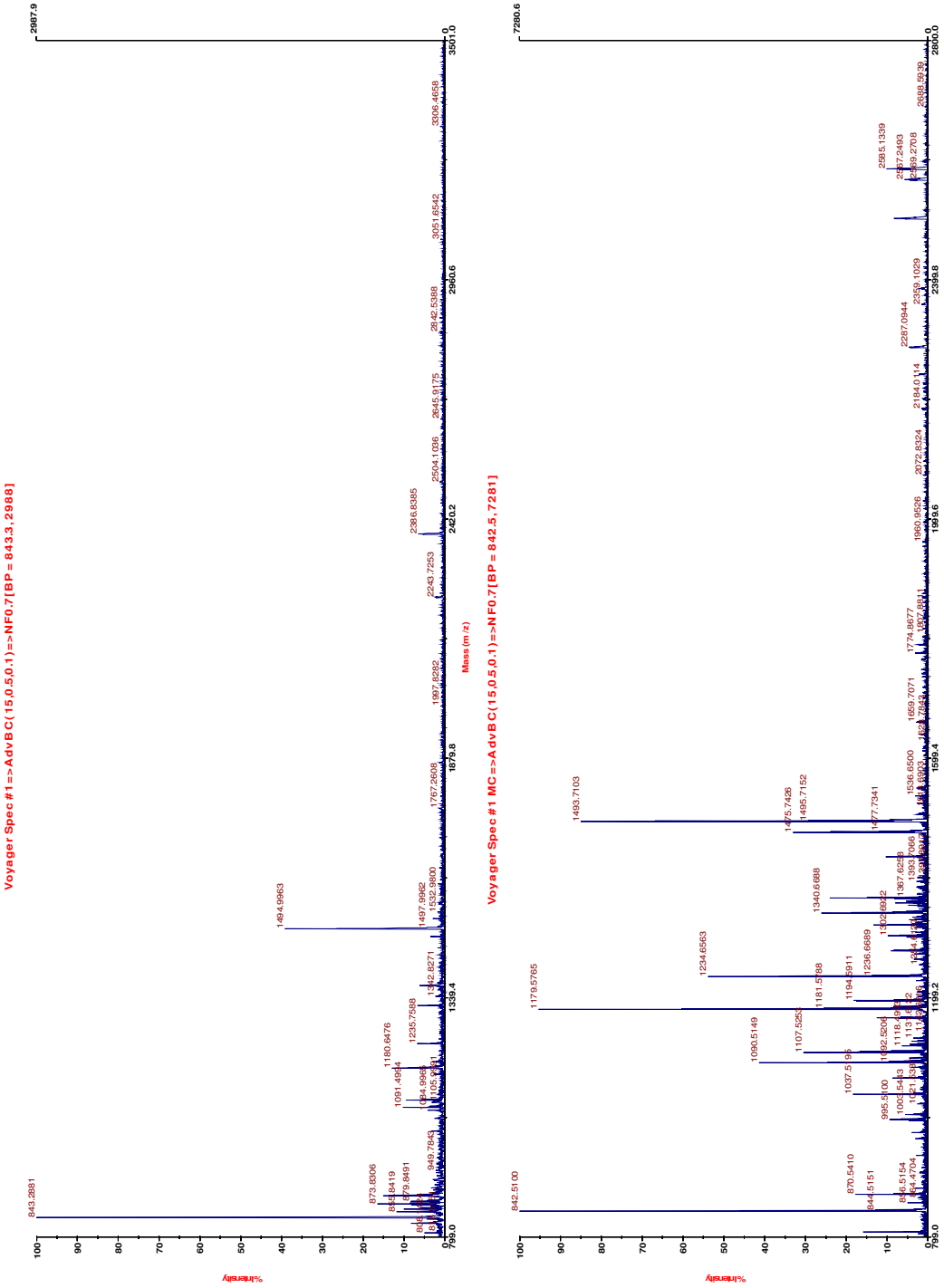


Figure 44. The MS spectra comparison of the peptide sample before and after concentrating. The diagram shows MS spectra before (top) and after (bottom) concentrating with ZipTip for the same sample. A noticeable improvement in signal/noise ratio can be observed.

Another significant difficulty that was faced in this study is that in order to identify matches for proteins analysed by masspectrometry a protein database containing exact sequences of analysed proteins is required. As for strain MV-1 there was no known sequence of magnetosome membrane proteins and therefore the first attempts to identify proteins using databases for other better characterized strains of magnetotactic bacteria did not gain any positive results. The reason for this is that the software that is used to identify proteins from MS data performs a virtual digest of the proteins in the database even small differences in the amino acid sequence result in a major effect on the peptide mass which makes it impossible to match with the generated data.

Once the FASTA database was produced in this work, it was uploaded onto the local MASCOT server and MS-Fit (University of California) and the search was carried out. The identifications were considered as positive if there were observed a minimum of 10 % coverage by matching peptides, a minimum of four independent matching peptides per protein (Tanaka *et al.*, 2006). Another condition of a positive identification was that the analysed spectra were dominated by peaks representing the matching peptides. An example of such spectra for sample № 5 which was identified as MamE can be observed on the next figure (Figure 45).

This is followed by the page demonstrating another example of the successful identification using Mascot for sample № 9 (Figure 46).

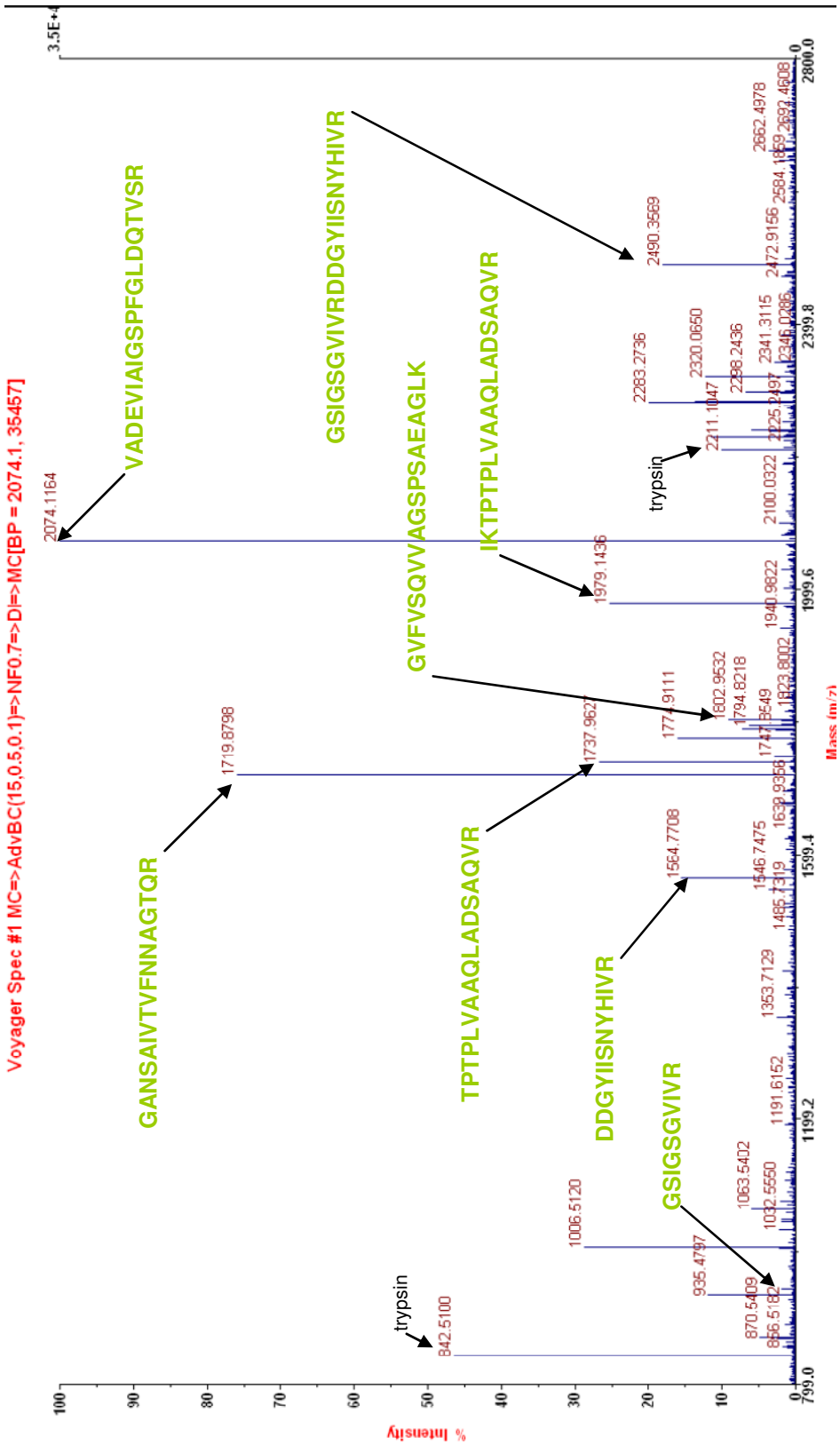


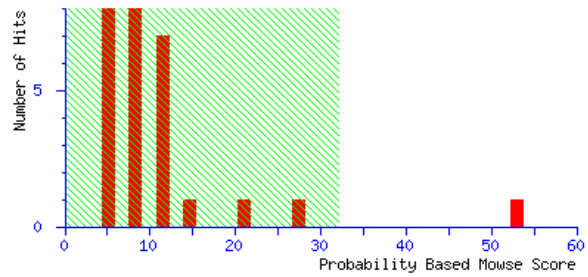
Figure 45. The MS spectra dominated by the signals matching peptide masses of MamE protein. The spectra obtained for sample № 5 is clearly dominated by the peaks that match masses for all peptides identified for the magnetosome membrane protein MamE. Two peaks representing trypsin autolysis products were used for calibration are shown as well.

Mascot Search Results

```
User            : Denis Trubitsyn
Email           : D.Trubitsyn@sms.ed.ac.uk
Search title    :
Database        : MV1  (92 sequences; 31525 residues)
Timestamp       : 8 Feb 2010 at 17:56:39 GMT
Top Score       : 53 for gi|238653837, gi|238653837|emb|CAV30779.1| conserved hypothetical protein
```

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 32 are significant (p<0.05).



gi|238653837 **Mass:** 33940 **Score:** 53 **Expect:** 0.00046 **Queries matched:** 9

gi|238653837|emb|CAV30779.1| conserved hypothetical protein [MV-1]

Match to: **gi|238653837** Score: **53** Expect: **0.00046**
gi|238653837|emb|CAV30779.1| conserved hypothetical protein [MV-1]
Nominal mass (M_r): **33940**; Calculated pI value: **4.88**
Fixed modifications: Carbamidomethyl (C)
Sequence Coverage: **42%**
Matched peptides shown in **Bold Red**

```
     1 MAYSQEICRD YPGAFLFLVD QSRSMHKPFG VDGAGKPVER ATVVAEALNS
    51 TLEEIVNRCM RDEGVSDYFD VGIIGYGKTS RPAFCWQGSL AGRGMVPISE
   101 VADNAIVETK EIETLVRDQI VKETVTVSRW VEPVAAESTP MNGALQLARA
   151 AIQDWIFRHP KSFPPIVINI TDGMANDVSS EEELLNSARR LTSLKTTDGN
   201 VLMVNCHISD NTARPVVFPW NALELPDDTY AKLLFEMSSE MPDRYRSVIC
   251 EIFDRDLSST PAIRGMAFNA DAMALVKLLD IGTRQAFVFS EQPPASAHLH
   301 SLQAV
```

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
10 - 23	1627.7979	1626.7906	1626.8042	-8	0	R.DYPGAFLFLVDQSR.S
41 - 58	1929.0165	1928.0093	1928.0214	-6	0	R.ATVVAEALNSTLEEIVNR.C
59 - 78	2280.9901	2279.9828	2280.0191	-16	1	R.CMRDEGVSDYFDVGIIGYGK.T
79 - 93	1693.7961	1692.7888	1692.8155	-16	0	K.TSRPAFCWQGSLAGR.G
94 - 110	1772.8708	1771.8635	1771.9026	-22	0	R.GMVPISEVADNAIVETK.E
130 - 149	2140.1040	2139.0967	2139.0782	9	0	R.WVEPVAAESTPMNGALQLAR.A
150 - 158	1119.5838	1118.5765	1118.5873	-10	0	R.AAIQDWIFR.H
247 - 255	1138.5663	1137.5590	1137.5488	9	0	R.SVICEIFDR.D
247 - 264	2079.0961	2078.0888	2078.0466	20	1	R.SVICEIFDRDLSSTPAIR.G

Figure 46. Mascot search results for Sample № 9. This page demonstrates an example of Mascot search results for Sample № 9. The probability score, sequence coverage and masses of matched peptides can be observed.

The results of protein identification using MALDI technique are summarized in the table below (Table 13).

The identification of the samples numbered 2, 3, 4, 10, 11 and 12 using MALDI TOF instrument was not successful and another approach was utilized to identify them.

Sample number	Protein name	Mr Da	Number of identified peptides	Coverage (%)
1	CAV30808.1| MamR	9364	5	56
	CAV30776.1| MamC	11056	4	35
5	CAV30818.1| MamE	73406	10	17
6	CAV30810.1| MamA	24282	8	54
7	CAV30789.1| hypothetical protein	20680	6	36
8	CAV30814.1| MamM	35073	5	21
9	CAV30779.1| conserved hypothetical protein	33940	11	52
13	CAV30814.1| MamM	35073	4	17

Table 13. Magnetosome membrane proteins identified with MALDI TOF. The table shows samples that have been identified using MALDI TOF instrument.

3.4.2.2 LC MS/MS proteins identification

It was decided to attempt identification of the remaining samples (№: 2, 3, 4, 10, 11 and 12) using the more sensitive Liquid Chromatography-Mass Spectrometry (LC MS) methods. There was a limitation of the available runs of this instrument in this work and therefore it was decided to analyse samples in combination of 2 to reduce the total number of runs. In an attempt to produce more unambiguous data the samples were combined as 2 + 10; 3 + 11 and 4 + 12. This combination was based on the fact that chosen bands run very distantly from each other and it would be easier to distinguish a genuine match when a molecular weight is taken onto account.

The advantage of this method is its higher sensitivity than of MALDI TOF however in this case this can be considered as its downside. During the preparation of most of the protein samples there is a high probability that proteins will break down and some smaller fragments or will interact with each other and therefore run on the gel to the distances that were not expected. Overall it gives the explanation for the fact that LC MS instrument was able to identify signals from much higher number of proteins that can by observed on the gel. And instead of identifying just the most abundant protein present in the analysed band this instrument has identified several proteins that were present in the mixture. This stage of the project was carried out in collaboration with Andrew Cronshaw (University of Edinburgh). The preparation of the samples and the following data analysis was done in this project whilst actual loading of the samples and machine operating was carried out by Andrew Cronshaw.

As a result of the initial analysis using Mascot the proteins identified with LC MS a number of proteins was identified as positive in cases where probability based

Mowse score was considered as significant (over 11). All these identified proteins are listed in the following table (Table 14).

Sample number	Protein name	Probability Based Mowse Score	Nominal mass (Mr), Da	Number of identified peptides	Coverage (%)
2+10	CAV30776.1|MamC	91	11056	4	36
	CAV30807.1|MamB	81	33587	3	7
	CAV30789.1|hypothetical protein	64	20680	4	13
	CAV30799.1|hypothetical protein	51	23825	5	25
	CAV30791.1|Mms6	46	7543	4	58
	CAV30774.1|MamZ	38	69440	5	5
	CAV30814.1|MamM	37	35073	3	10
	CAV30752.1|MamR-like	36	12504	4	32
	CAV30812.1|MamO	33	74272	5	4
	CAV30796.1|MamH	30	47996	2	4
	CAV30819.1|MamI	30	7449	2	22
	CAV30795.1|MmsF	28	12311	2	11
	CAV30758.1|Sodium: alanine symporter	26	48123	3	6
	CAV30783.1|hypothetical protein	24	19171	5	36
	CAV30792.1|MamD	23	20935	5	25
	CAV30794.1|hypothetical protein	23	14519	4	28
	CAV30813.1|MamN	20	47083	1	2
	CAV30825.1|hypothetical protein	17	27336	1	2
	CAV30790.1|MmsF	16	12032	1	9
	CAV30779.1|conserved hypothetical protein	16	33940	2	4

Sample number	Protein name	Probability Based Mowse Score	Nominal mass (Mr), Da	Number of identified peptides	Coverage (%)
2+10	CAV30815.1|MamL	15	8949	1	8
3+11	CAV30776.1|MamC	74	11056	2	15
	CAV30815.1|MamL	60	8949	2	15
	CAV30807.1|MamB	46	33587	3	7
	CAV30791.1|Mms6	40	7543	2	25
	CAV30790.1|MmsF	36	12032	1	9
	CAV30810.1|MamA	36	24282	3	12
	CAV30812.1|MamO	35	74272	5	5
	CAV30789.1|hypothetical protein	35	20680	2	9
	CAV30799.1|hypothetical protein	32	23825	5	23
	CAV30795.1|MmsF	31	12311	1	9
	CAV30797.1|hypothetical protein	25	30105	2	5
	CAV30792.1|MamD	25	20935	3	14
	CAV30818.1|MamE	23	73406	2	2
	CAV30794.1|hypothetical protein	21	14519	5	33
	CAV30752.1|MamR-like	19	12504	2	13
	CAV30814.1|MamM	17	35073	1	3
4+12	CAV30776.1|MamC	100	11056	4	36
	CAV30791.1|Mms6	43	7543	3	46
	CAV30812.1|MamO	42	74272	2	3
	CAV30815.1|MamL	41	8949	1	8

Sample number	Protein name	Probability Based Mowse Score	Nominal mass (Mr), Da	Number of identified peptides	Coverage (%)
4+12	CAV30807.1|MamB	39	33587	2	6
	CAV30818.1|MamE	38	73406	8	14
	CAV30789.1|hypothetical protein	37	20680	2	12
	CAV30794.1|hypothetical protein	33	14519	1	10
	CAV30814.1|MamM	33	35073	5	18
	CAV30790.1|MmsF	31	12032	2	18
	CAV30796.1|MamH	29	47996	3	6
	CAV30774.1|MamZ	29	69440	1	1
	CAV30799.1|hypothetical protein	20	23825	2	16
	CAV30810.1|MamA	18	24282	3	11
	CAV30795.1|MmsF	18	12311	3	28
	CAV30836.1|conserved hypothetical protein	14	24258	5	16

Table 14. Proteins identified with LC MS analysis. This table illustrates proteins in samples combination 2+10, 3+11 and 4+12 identified using LC MS analysis. The table includes only proteins with Mowse probability score above 11 which is considered as significant for this analysis.

In order to improve the data analysis gained in the LC MS an attempt to compare spectra obtained from MALDI TOF and LC MS was carried out. The identified as positive match masses in LC MS data were compared with the mass lists for each sample received in MALDI TOF analysis. As a result some additional details regarding the presence of a specific protein in the specific band on the gel were obtained. If the mass value Delta between LC MS and MALDI was below 1 then the signal match was considered as strong; between 1 and 3 was considered as possible and for those above 3 was ignored. The results of this analysis are summarized in the table below (Table 15)

Sample number	Protein name	Nominal mass (Mr), Da	Number of matched peptides	Signal match
2	CAV30815.1|MamL	8949	1	possible
	CAV30752.1|MamR-like	12504	1	possible
	CAV30774.1|MamZ	69440	2	possible
	CAV30799.1|hypothetical protein	23825	1	possible
	CAV30819.1|MamI	7449	1	possible
	CAV30790.1|MmsF	12032	1	strong
	CAV30813.1|MamN	47083	1	possible
	CAV30796.1|MamH	47996	1	strong
	CAV30795.1|MmsF	12311		
	CAV30794.1|hypothetical protein	14519	1	strong
	CAV30776.1|MamC	11056	1	strong
3	CAV30815.1|MamL	8949	1	possible
	CAV30752.1|MamR-like	12504	1	possible
	CAV30794.1|hypothetical protein	14519	2	possible
	CAV30795.1|MmsF	12311	2	strong
	CAV30812.1|MamO	74272	1	possible
	CAV30789.1|hypothetical protein	20680	1	strong
	CAV30810.1|MamA	24282	1	possible
	CAV30792.1|MamD	20935	1	strong
	CAV30776.1|MamC	11056	1	possible
	CAV30815.1|MamL	8949	1	strong
	CAV30814.1|MamM	35073	1	possible

Sample number	Protein name	Nominal mass (Mr), Da	Number of matched peptides	Signal match
3	CAV30790.1|MmsF	12032	2	strong
	CAV30810.1|MamA	24282	1	possible
	CAV30789.1|hypothetical protein	20680	2	strong
	CAV30796.1|MamH	48027	1	strong
	CAV30794.1|hypothetical protein	14519	1	possible
10	CAV30814.1|MamM	35073	2	strong
	CAV30789.1|hypothetical protein	20680	1	strong
	CAV30791.1|Mms6	7543	1	possible
11	CAV30797.1|hypothetical protein	30105	1	possible
	CAV30812.1|MamO	74272	2	strong
	CAV30818.1|MamE	73406	1	possible
	CAV30807.1|MamB	33587	3	strong
	CAV30792.1|MamD	20935	2	strong
	CAV30795.1|MmsF	12311	1	strong
	CAV30789.1|hypothetical protein	20680	1	strong
	CAV30799.1|hypothetical protein	23825	1	strong
	CAV30794.1|hypothetical protein	14519	1	possible
	CAV30776.1|MamC	11056	1	possible
	CAV30791.1|Mms6	7543	1	possible
12	CAV30812.1|MamO	74272	1	strong

Sample number	Protein name	Nominal mass (Mr), Da	Number of matched peptides	Signal match
12	CAV30807.1|MamB	33587	1	strong
	CAV30814.1|MamM	35073	2	strong
	CAV30790.1|MmsF	12032	1	strong
	CAV30795.1|MmsF	12311	1	strong
	CAV30789.1|hypothetical protein	20680	1	strong
	CAV30794.1|hypothetical protein	14519	1	possible

Table 15. LC-MS and MALDI TOF data comparison. This table demonstrates proteins that were identified by LC-MS analysis and have got matching signals produced in MALDI TOF analysis. See text for details.

There are several trends that can be observed from the above tables. First of all the fact that not all peaks that were identified in LC MS matched peaks obtained in MALDI TOF experiments can be explained in several ways. It is known that not every peptide can be ionized by the laser and therefore signals for such will not be collected by MALDI TOF. Another reason is that in MALDI TOF experiments signals were collected for those peptides with masses above 800 whilst for LC-MC it was 400; thus is another reason (apart from the better instrument sensitivity) that explains the significantly higher number of peptides identified by LC MS.

Another interesting observation is that some of the proteins were identified in the positions where they are expected to run on the SDS gel. For example proteins MamL (8949 Da) and MamR-like (12512 Da) were identified in samples 2, 3, 4 and 2 ,3 respectively but not in the bands 10, 11 and 12 which are expected to have much higher molecular masses. On the other hand, proteins MamO (74272 Da) and MamB (33587 Da) were only identified in the higher molecular mass bands 11 and 12. However it is worth pointing out that this trend does not work for protein CAV30789.1 (20680 Da) which was identified in the majority of the samples (3, 4, 10, 11 and 12).

3.4.2.3 Protein identification by Orbitrap

The most sensitive instrument was used in order to identify all remaining proteins present in the magnetosome membrane fraction. Due to the limited availability of this instrument it was decided to run a sample that contained a mixture of all proteins. A SDS gel was loaded with 20 μg of the magnetosome membrane fraction and was run to allow proteins to enter the gel. Then the 5 mm in length gel fragment containing the proteins was excised and subjected to a trypsin digest. This stage of

the project was completed in the University of Edinburgh in collaboration with Juri Rappsilber and Flavia Alves.

The results of the magnetosome membrane proteins identification by this method can be observed in Table 16.

Another Orbi trap analysis was carried out for the magnetosome membrane sample produced by the method involving high pressure cell breakage and weak sonication (Tanaka *et al.*, 2006). Although the main list of identified proteins is similar to that produced in the method adapted in this work some new proteins were identified (Table 17).

As it was mentioned above, the attempts to identify magnetosome membrane proteins was carried out using the database containing 92 protein sequences that include the "magnetosome island". In order to provide more detailed analysis the search against the database protein sequences obtained by automated annotation of the whole genome was carried out. This search has resulted in identification of a high number of proteins.

The search using the spectra obtained for magnetosome membrane fraction isolated by the method developed in this work identified 486 proteins (310 matching two and more peptides). The table containing the results of the search is shown in Appendix I (page 217)

The search using the spectra obtained for magnetosome membrane fraction isolated by Tanaka's method identified 463 proteins (283 matching two and more peptides). The table containing the results of the search is shown in Appendix II (page 229).

№	Description	Mass [Da]	Peptides	Score
1	**CAV30812.1| MamO**	**74272**	**11**	**663**
2	**CAV30818.1| MamE**	**73406**	**11**	**488**
3	**CAV30814.1| MamM**	**35073**	**9**	**521**
4	**CAV30753.1| MamR like**	**10576**	**9**	**461**
5	**CAV30792.1| similar to MamD**	**20935**	**8**	**371**
6	**CAV30804.1| MamY**	**36786**	**7**	**489**
7	**CAV30796.1| MamH**	**47996**	**6**	**291**
8	**CAV30752.1| MamR-like**	**12504**	**6**	**277**
9	**CAV30797.1| hypothetical protein**	**30105**	**6**	**268**
10	**CAV30779.1| conserved hypothetical protein**	**33940**	**5**	**210**
11	**CAV30774.1| MamZ**	**69440**	**5**	**183**
12	**CAV30789.1| hypothetical protein**	**20680**	**4**	**251**
13	**CAV30807.1| MamB**	**33587**	**4**	**187**
14	**CAV30808.1| MamR**	**9364**	**4**	**170**
15	**CAV30809.1| MamQ**	**30509**	**4**	**154**
16	**CAV30823.1| Integration host factor, alpha subunit**	**11670**	**3**	**249**
17	**CAV30776.1| MamC**	**11056**	**3**	**222**
18	**CAV30758.1| Sodium:alanine symporter**	**48123**	**3**	**139**
19	**CAV30790.1| MmsF**	**12032**	**3**	**132**
20	**CAV30795.1| MmsF**	**12311**	**3**	**122**
21	**CAV30783.1| hypothetical protein**	**19171**	**3**	**99**
22	**CAV30810.1| MamA**	**24282**	**2**	**130**
23	**CAV30794.1| hypothetical protein**	**14519**	**2**	**120**
24	**CAV30756.1| Fe2+ transport system protein B**	**81030**	**2**	**108**
25	**CAV30813.1| MamN**	**47083**	**2**	**91**
26	**CAV30819.1| MamI**	**7449**	**2**	**77**
27	CAV30798.1| Magnetosome protein MamK-II	39127	1	51
28	CAV30803.1| histidine kinase, TPR motif	97613	1	47
29	CAV30799.1| hypothetical protein	23825	1	30
30	CAV30791.1| Mms6	7543	1	29

Table 16. Magnetosome membrane proteins identified using Orbitrap. This table gives a description of the proteins identified using Orbitrap instrument. The magnetosome membrane fraction was processed as a mixture.

№	Description	Mass [Da]	Peptides	Score
1	**CAV30814.1|MamM**	**35073**	**12**	**690**
2	**CAV30812.1|MamO**	**74272**	**11**	**573**
3	**CAV30789.1|hypothetical protein**	**20680**	**10**	**561**
4	**CAV30818.1|MamE**	**73406**	**9**	**515**
5	**CAV30810.1|MamA**	**24282**	**8**	**398**
6	**CAV30753.1|MamR like**	**10576**	**7**	**446**
7	**CAV30792.1|MamD**	**20935**	**7**	**411**
8	**CAV30796.1|MamH**	**47996**	**6**	**266**
9	**CAV30752.1|MamR-like**	**12504**	**5**	**355**
10	**CAV30794.1|hypothetical protein**	**14519**	**5**	**326**
11	**CAV30807.1|MamB**	**33587**	**5**	**310**
12	**CAV30777.1|Serine/threonine protein kinase**	**36218**	**5**	**266**
13	**CAV30816.1|MamK**	**39624**	**5**	**239**
14	**CAV30797.1|hypothetical protein**	**30105**	**5**	**197**
15	**CAV30804.1|MamY**	**36786**	**4**	**335**
16	**CAV30776.1|MamC**	**11056**	**4**	**311**
17	**CAV30808.1|MamR**	**9364**	**4**	**217**
18	**CAV30790.1|MmsF**	**12032**	**4**	**209**
19	**CAV30758.1|Sodium:alanine symporter**	**48123**	**4**	**204**
20	**CAV30783.1|hypothetical protein**	**19171**	**4**	**175**
21	**CAV30795.1|MmsF**	**12311**	**3**	**113**
22	**CAV30811.1|MamP**	**33831**	**2**	**132**
23	**CAV30819.1|MamI**	**7449**	**2**	**118**
24	**CAV30799.1|hypothetical protein**	**23825**	**2**	**108**
25	**CAV30779.1|conserved hypothetical protein**	**33940**	**2**	**104**
26	**CAV30774.1|MamZ**	**69440**	**2**	**90**
27	**CAV30809.1|MamQ**	**30509**	**2**	**84**
28	**CAV30760.1|conserved hypothetical protein**	**34741**	**2**	**74**
29	CAV30803.1|histidine kinase, TPR motif	97613	1	72
30	CAV30805.1|MamT	19545	1	67
31	CAV30798.1|MamK-II	39127	1	55
32	CAV30791.1|Mms6	7543	1	50
33	CAV30813.1|MamN	47083	1	45

№	Description	Mass [Da]	Peptides	Score
34	CAV30756.1|Fe2+ transport system protein B	81030	1	28
35	CAV30806.1|MamS	19952	1	26

Table 17. Magnetosome membrane proteins isolated by Tanaka's method identified using Orbitrap. This table shows description of proteins identified using Orbitrap instrument. The magnetosome membrane fraction was processed as a mixture. Novel proteins identified in this sample are highlighted in grey.

The identification of a protein in mass spectrometry is based on probability and always involves the decision making on what can be taken as a positive match. The research group of Juri Rappsilber (University of Edinburgh) that was involved in this part of the work has developed a highly accurate method of analysis that allows to consider the identification as positive even if only mass of one peptide was matched (Sennels *et al.*, 2009).

3.4.3 Discussion

The analysis of magnetosome membrane proteins allowed the generation of a significant amount of novel data. There was no knowledge of protein composition of magnetosome for marine vibrio MV-1 apart from the identification one of the two lowest bands on the gel that was previously identified as MamC (Dennis Bazylinski, unpublished data). In this work a total number of 33 proteins was identified to be present in the magnetosome membrane fraction when searched against 92 proteins in a database that includes "magnetosome island" proteins. The analysis of the sample produced by Tanaka's method has identified 35 proteins. The number of proteins identified for the sample produced by the method developed in this work is shown in Figure 47. This diagram suggests that a relatively high number of identified proteins, 24 out of 30 and 27 for Orbitrap and LC MS combined with MALDI TOF respectively, were identified by both methods giving high confidence of their presence in the analysed sample.

Several interesting conclusions can be made from the analysis of the generated data. It has to be pointed out that the identification of the presence of the protein in the magnetosome particle does not necessarily suggest the importance of this protein for magnetosome formation. However, such identifications can help to outline the

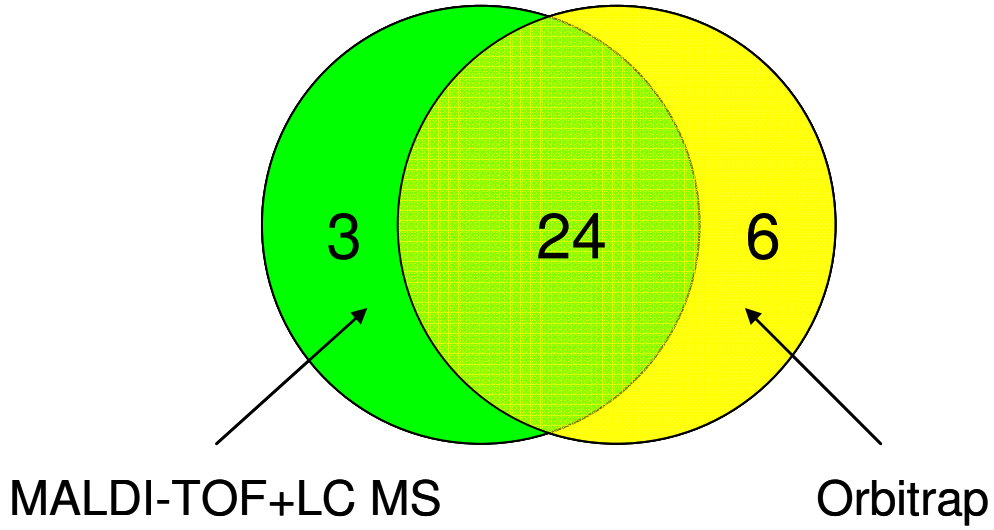


Figure 47. The comparison of numbers of magnetosome membrane proteins identified using MALDI TOF+LC MS and Orbitrap. The Venn diagram shows the number of proteins identified by each of the methods separately and the number of proteins identified by both of the methods.

possible candidates for further research. For example the generation of mutants that lack an expression of a specific gene can show phenotype difference in magnetosome formation.

The presence of several analogues of some of the genes was observed for the "magnetosome island" cluster in marine vibrio MV-1. However, there was no data regarding the involvement of these analogues in the magnetosome formation.

In this study the presence of all 3 proteins that are the products of analogues of *mamR* gene was shown in magnetosomes. This can suggest the possible involvement of MamR protein and its analogues in the process of magnetosome formation.

The investigation of the "magnetosome island" of marine vibrio MV-1 has also revealed the presence of two analogues of *mamK* gene that is shown to code for actin-like protein. In this work MamK-II protein was identified by Orbi trap analysis in both samples (this study and Tanaka's isolation methods). Interestingly, the sample isolated by Tanaka's method has revealed the presence of both MamK and MamK-II.

Another interesting result is the identification of the presence in the magnetosome membrane fraction of several proteins that are products of genes that are specific to marine vibrio MV-1 "magnetosome island". The identified proteins that have very low and often insignificant similarity to those previously described for "magnetosome island" or in some case any described protein are summarized in the table below (Table 18).

№	Description	Mass [Da]	Blast score	E value	Highest similarity
1	CAV30789.1|hypothetical protein	20680	46.2	0.002	adhesin [Mycoplasma hyopneumoniae]
2	CAV30799.1|hypothetical protein	23825	37.4	1.1	hypothetical protein An16g04460 [Aspergillus niger]
3	CAV30783.1|hypothetical protein	19171	36.6	1.3	Magnetosome protein MamG-like [uncultured bacterium]
4	CAV30797.1|hypothetical protein	30105	38.1	1.0	nuclear transcription factor, X-box binding, putative [Ricinus communis]
5	CAV30825.1|hypothetical protein	27336	51.2	1e-04	putative relaxase [Campylobacter jejuni subsp. jejuni IA3902]
6	CAV30794.1|hypothetical protein	14519	37.7	0.35	bacterial magnetic particle specific iron-binding protein [Magnetospirillum magneticum]

Table 18. Identified proteins that are products of the genes highly specific to "magnetosome island" of marine vibrio MV-1. This table shows BLAST search results for some of the proteins identified in this work. The very low BLAST score makes similarity insignificant and proves high specificity of these proteins.

These proteins are very interesting candidates for further research and the generation of knock-out mutants can be suggested also as a next step. Interestingly, in research carried out to identify proteins in magnetosome membrane fractions of other MB, for example, in *D. magneticus* RS-1, 13 out of 190 identified proteins were also shown to have no homology with any known proteins (Matsunaga *et al.*, 2009).

A large number of proteins identified in the analysis of the spectra obtained by Orbitrap suggest that the methods of isolation of membrane fraction do not prevent contamination with other proteins present in the cell. This can be explained by the fact that the magnetosome membrane is formed as an invagination of the cytoplasmic membrane and therefore may contain all membrane associated proteins (Komeili *et al.*, 2006). Another difficulty is that during the cell disruption by either high pressure or sonication membrane vesicles can be easily integrated into each other or inverted and therefore trap non-specific proteins. One of the uses of the obtained data is in annotation of the genome. Identified proteins will eliminate ambiguity showing that an analysed ORF is expressed and translated into a protein.

Another approach can also be suggested as a part of future work. The LC/MS techniques provide good identification of the presence of the protein in the mixture (qualitative proteomics). The main limitation of the most of the available mass spectrometry techniques is the inability to reliably quantify proteins present in the compared samples. One of the relatively new methods in proteomics, called isobaric Tag for Relative and Absolute Quantitation (iTRAQ), allows quantification of the analysed proteins (Zieske, 2006). The idea behind this method is that peptides generated by digestion are chemically tagged at their N-terminus by specific tags one for each sample. The samples are then mixed and fractionated by nanoLC and

analysed by MS/MS. The signals obtained for analysed peptides are used to identify proteins and signals obtained for reporter ions enable relative quantification of the amounts of protein.

It appears to be interesting to attempt to analyse membrane fractions of MB using this method. One of the analysed samples should contain membrane fraction of the cells producing magnetosomes and the other one should contain membrane fraction of the cells that were cultured under conditions that suppress magnetosome production. This experiment should help to accurately outline the list of proteins the concentration of which raises or changes in the cells that produce magnetosome and therefore suggest that they are important for the magnetosome formation.

3.5 Attempts to generate knock-out mutants in marine vibrio MV-1

In order to investigate the role in magnetosome formation of a specific gene in the "magnetosome island" it was decided to attempt to generate mutants. Knock-out mutants where one or more genes are not transcribed should help to outline the minimum number of genes involved in the process of magnetosome formation.

The most important in the process of mutant construction for any microorganism is the developing of a genetic system which involves an ability of a bacterium to grow on solid medium and ability to uptake DNA from an external source. For the marine vibrio MV-1 these techniques were developed in the laboratory of Dennis Bazylinski (UNLV, Las Vegas) by Sabrina Schuebbe (unpublished data) from the method suggested for *Magnetospirillum gryphiswaldense* (Schultheiss and Schuler, 2003). During the collaboration with Bazylinski group these methods were adapted to use in this project in the University of Edinburgh.

3.5.1 Plasmid transfer by conjugation

The main idea of the method is to carry out mating on the surface of the solid medium plates for a period of 8 hours using 1:1 donor:recipient ratio. *E. coli* WM3064 was chosen as a donor strain due to its ability to mobilize plasmids for transfer and its specific growth requirements. This strain is a diaminopimelate (DAP) auxotroph and can be counter-selected on the agar plates when DAP is not present in the medium.

Numerous variations of the tested conditions were attempted. Donor to recipient ratio was changed from 1:1 to 1:2 and 2:1. The donor and recipient were both used in either stationary or exponential growth phases. Mating time was changed from 8 h to

14 h. These experiments allowed optimization of the relatively efficient method of conjugation described below.

Fresh exponential phase liquid culture of MV-1 was used as a recipient whilst an overnight culture of *E. coli* WM3064 was used as a donor. The numbers of cells were equilibrated by measuring optical density at 595 nm. Approximately 10^8 of mixed cells were plated onto the solid MV-1 medium for 8 hours with 2% oxygen in nitrogen (v/v) in an anaerobic jar at room temperature. After incubation, cells were collected from the surface of the plates by pipetting 1 ml of liquid MV-1 medium and using a spreader to rub the surface of the medium. The cell suspension was then transferred into the sterile 1.5 ml tubes and incubated for further 2 hours in an anaerobic jar with nitrous oxide to allow sufficient expression of the antibiotic resistance gene. The growth of the donor *E. coli* cells on the following stages of the experiment was expected to be inhibited by the absence of DAP in the medium.

Finally, 50 µl of the cells suspension were plated in 0.1X, 1X and 10X concentrations onto solid MV-1 medium with appropriate antibiotic in order to select for transconjugants. The colonies were expected to become visible after 7 to 10 days. However, no colonies with sufficient biomass were observed at this stage. Further observations revealed that a lawn of cells has appeared on the plates with 10X concentrated mixture of cells after day 12 and individual colonies appeared after day 14. The size of the colonies was ≈0.3 mm which made it extremely difficult to use an individual colony for subculturing. The MV-1 phenotype of the cells was confirmed by light microscopy.

Due to the low plating efficiency growth of transconjugants was attempted in test tubes filled with liquid medium and 1 % (w/v) agar. Test tubes with rubber tops were used to allow inoculation with a syringe needle. 10 ml of the medium containing an appropriate antibiotic was added to the tube. Once the medium had reached a temperature of 40 ºC, 200 µl of conjugation mixture were added and mixed with the medium by inverting the tube several times. Various concentrations (0.1X, 1X and 10X) of the conjugation mixtures were tested. It was concluded that the use of 1X concentration results in formation of a number of colonies that was convenient to use for further analysis. Tubes were incubated at room temperature for several days. Air from the upper part of the tube was not removed. The presence of oxygen in the upper space most likely led to the formation of a gradient and therefore bacterial cells attempted to move to a certain depth in the medium seeking an optimal condition of oxygen tension. This approach has shown its efficiency and the results of the conjugations are shown in sections 3.5.4 and 3.6.

3.5.2 Gene candidates for construction of knock-out mutants

For a number of genes found in the "magnetosome island" of other magnetotactic microorganisms the predicted function was reported either by comparison of the conserved regions or by generation of mutants (Murat *et al.*, 2010). The predicted functions of such genes were discussed in the relevant part of the Introduction section. In this work it was decided to concentrate on the genes that are either specific to the marine vibrio MV-1 "magnetosome island" or specific to magnetotactic microorganisms in general. Although the protein analysis of the magnetosome membrane has outlined an interesting list of the potential targets these results were not available when the decision had to be made regarding the candidate list. Therefore, the decisions were made based on the available nucleotide sequence but not on the presence of the specific protein in the magnetosome membrane. Four genes were chosen as potential targets for knock-out: *mamP* (similarity with trypsin-like serine protease, unknown role), *mamT* (similarity with cytochrome c, unknown role), *mamY* (unknown role) and ORF2 (unknown role) in *mamK*-II cluster (Figure 48).

3.5.3 Mutants construction using broad host range *cre-lox* system

This part of this work was carried out in the laboratory of Dennis Bazylinski (UNLV, Las Vegas) during the collaborative research. The method of mutant construction based on the use of *cre-lox* system was suggested by Sabrina Schuebbe and adapted from the Research Report published in Biotechniques (Marx and Lidstrom, 2002).

The principle of this method can be summarized in the diagram adopted from the original publication (Figure 49).

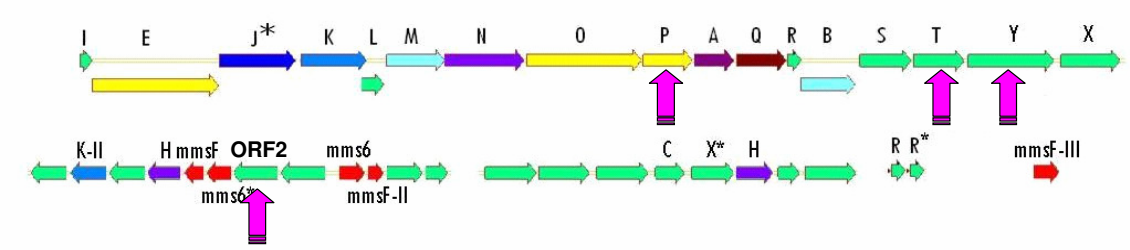


Figure 48. Genes chosen as targets for mutants construction. The diagram shows four genes (shown with arrows) that were chosen for mutant construction.

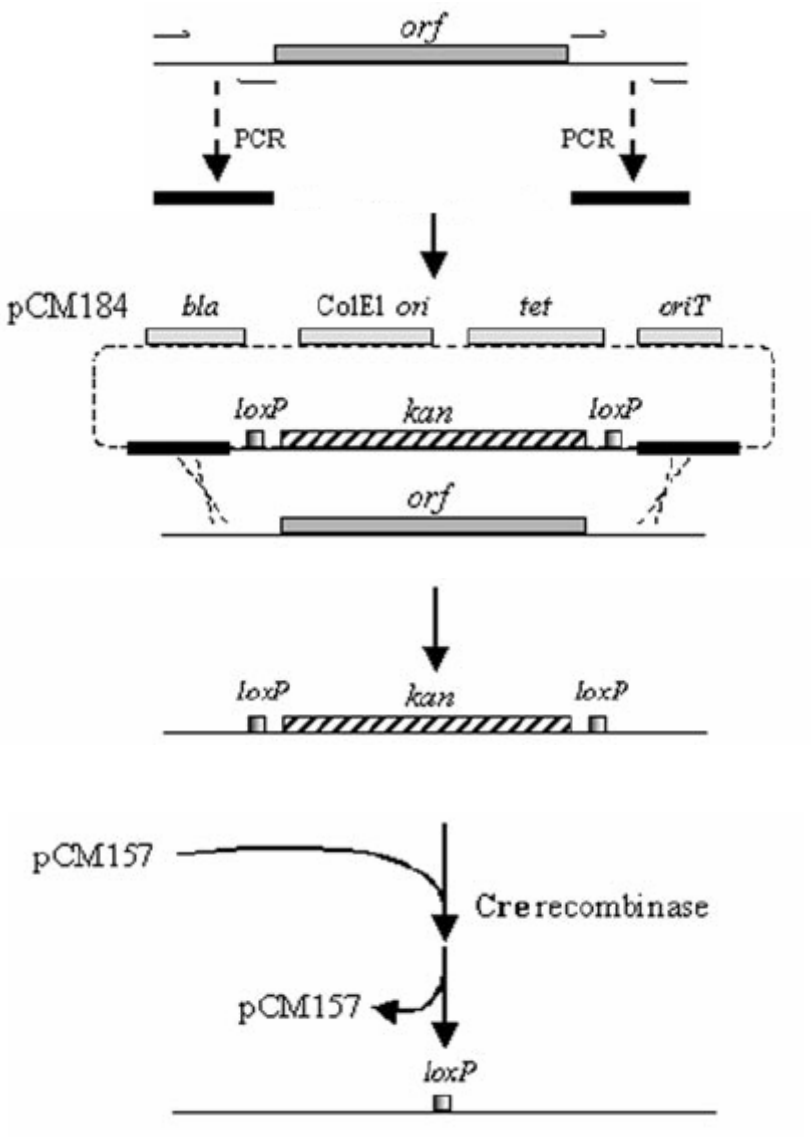


Figure 49. Cre/lox based system of mutant generation. This diagram shows steps of generation of the knock-out mutant (explained in the main text). This diagram is adapted from the original publication (Marx and Lidstrom, 2002). The idea behind this method is that in the first step flanking regions of the target gene are amplified by PCR and then cloned into the pCM184 vector. The obtained construct is then introduced into the cells of the target organism and being unable to replicate the only way resistance cassette can remain in the cell is by a crossing over with the parts of the chromosome. The single crossover results in the whole vector being integrated into the chromosome and in the event of a double crossover the target gene is substituted by the kanamycin resistance cassette in the middle flanked by *loxP* sites. The double recombinants are selected by the lack of resistance to tetracycline. Finally, Cre recombinase is expressed in the cell by introduction of the pCM157 vector which results in excision of the kanamycin resistance cassette.

In this work two constructs for genes *mamT* and ORF2 in *mamK*-II cluster were generated using the above method.

For *mamT* (951 bp) an upstream fragment of 538 bp was amplified with primers containing restriction sites for SacI and ApaI and a downstream fragment of 651 bp with primers containing restriction sites for KpnI and BsrGI.

For ORF2 (1488 bp) an upstream fragment of 592 bp was amplified with primers containing restriction sites for SacI and ApaI and a downstream fragment of 651 bp with primers containing restriction sites for NcoI and EcoRI.

The restriction sites were chosen to comply with the following conditions: an absence of the recognition sites within the sequence of the fragment of interest; a cutting efficiency in the presence of "universal" restriction buffer to give 100 % efficiency and a presence of the site in the multiple cloning sites of the pCM184 vector (Figure 50).

Amplified fragments were then purified from the gel and digested with the appropriate restriction enzymes. The pCM184 vector was also digested with the appropriate enzymes (using a pair of enzymes to clone one fragment at the time) followed by an enzyme inactivation and dephosphorylation of the 5´ phosphate groups to prevent self-ligations. After the amounts of vector and insert were equilibrated the sample was re-purified prior to ligation using a standard Gel Purification Kit (QIAGEN) to avoid the effect of salts present in the previous reactions buffers.

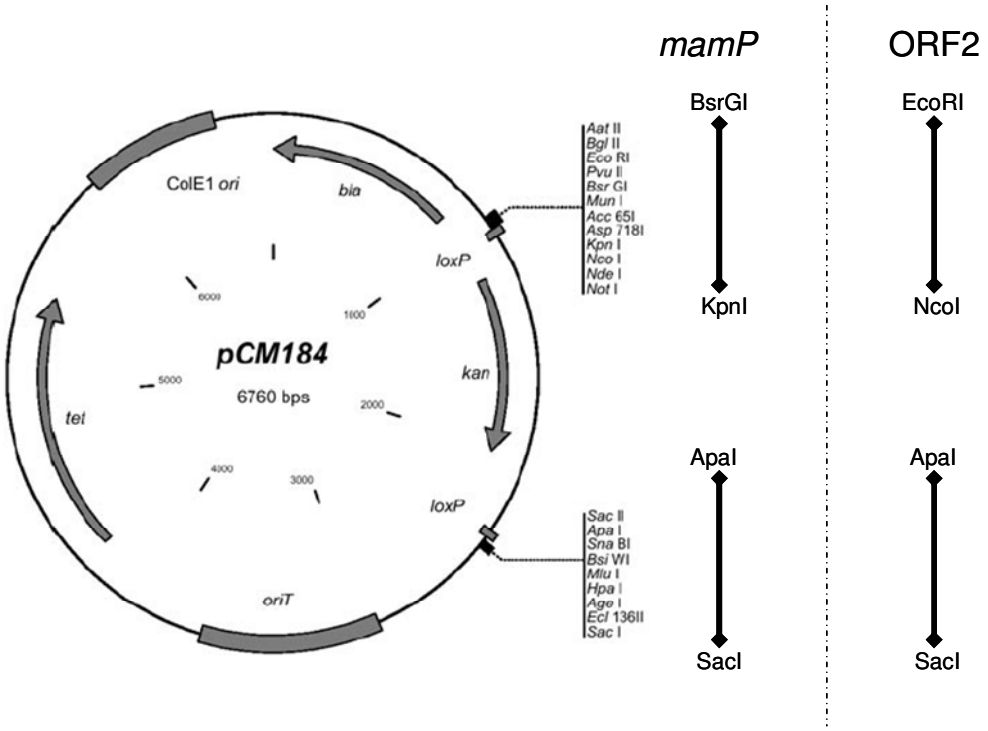


Figure 50. The map of the pCM184 vector and amplified inserts. This diagram demonstrates the map of the vector and inserts fragments with the restriction sites. This diagram is adapted from the original publication (Marx and Lidstrom, 2002).

After ligation the constructs were transferred into the *E. coli* by transformation. The transformants were selected by their resistance to kanamycin. After the colonies were subcultured into liquid medium the plasmids were purified and screened first by digestion with appropriate enzymes followed by separation on an agarose gel and then by amplification of the inserted fragments. In order to prove that the fragments were cloned successfully the PCR reactions were set up not only to amplify individual inserts but also using a pair of the most distant primers in order to amplify both of the inserts and a kanamycin resistance cassette located in the middle. The sizes of the obtained fragments corresponded directly with those estimated theoretically. Obtained constructs for *mamT* and ORF2 were named pDTmamTcrelox and pDTORF2crelox respectively.

All attempts to introduce the generated constructs into the marine vibrio MV-1 cells were carried out by bacterial mating. There were no colonies observed after incubation. There are several possible reasons for this method not being successful. For example when the vector integrates into the chromosome the level of transcription of the antibiotic resistance gene might not be of sufficient level to allow growth on selective medium. Also there might be a very active nuclease system that destroys any non-native DNA introduced into the cells. The last hypothesis was disproved in the following experiments where plasmids were successfully acquired by the cells of strain MV-1.

The proposed method has a significant disadvantage due to the fact that the chance of a double crossover event is much lower than that of a single crossover. This results in

the need to screen very high number of recombinants for the lack of tetracycline resistance. It was decided to try another approach to generate knock-out mutants of strain MV-1 that is described in the next section.

3.5.4 Mutants construction using I-SceI system

Another approach was attempted in order to generate knock-out mutants in strain MV-1. This is an adaptation of the method used in the laboratory of Garry Blakely (University of Edinburgh). The idea behind this method is explained in the diagram adapted from the original publication (Figure 51).

One of the advantages of this method compared with the method described in the previous section is that there is no need to screen a very large number of recombinants in order to isolate one with double recombination. It was decided to attempt a mutant construction using the I-SceI based method in this work.

The gene ORF2 (1488 bp) in the *mamK*-II cluster was selected as a target gene. An upstream flanking region of 592 bp was amplified with primers containing restriction sites for XhoI and SpeI and a downstream region of 651 bp with primers containing restriction sites for SpeI and XbaI. After digestion with SpeI enzyme these fragments were ligated together, followed by PCR amplification of the 1243 bp fused fragment.

The pGB909 vector containing I-SceI site was digested with XhoI and XbaI followed by enzyme inactivation and dephosphorylation of the 5´ phosphate groups to prevent self-ligation. The generation of this construct is shown schematically in Figure 52.

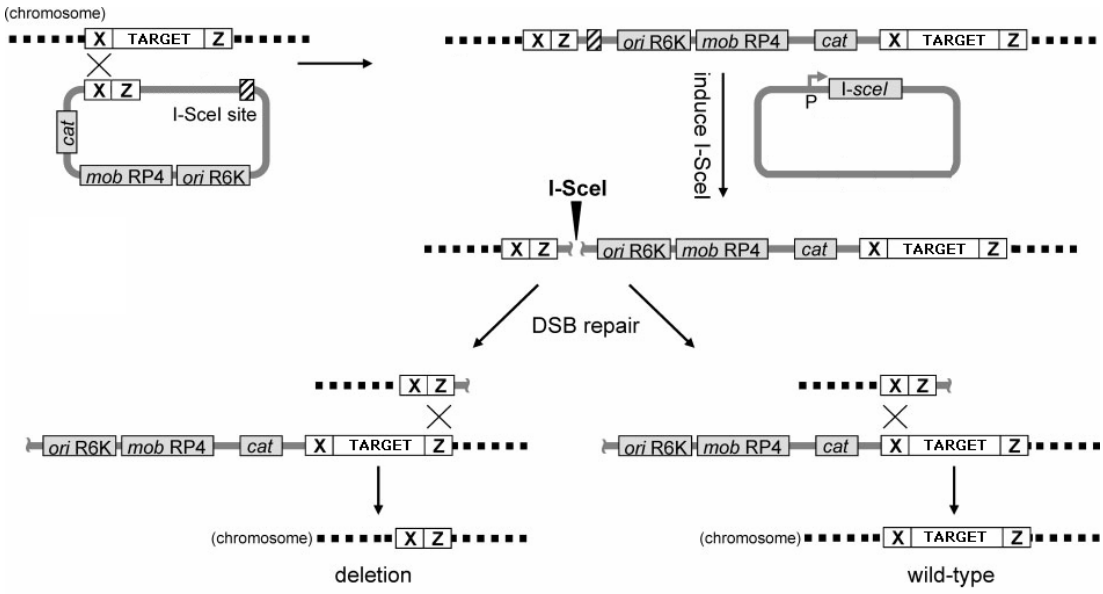


Figure 51. Generation of mutants using I-SecI site. A suicidal vector containing I-SceI site and fused flanking regions of a target gene (X and Z) is introduced into the cell. A single crossover event leads to the vector integration into the chromosome and integrants are selected for resistance to chloramphenicol. A vector containing a promoter controlled I-*sceI* gene is introduced into the cell and an expression I-SceI meganuclease leads to double-strand break (DSB). A resolution of DSB leads either to generation of a deletion or return to wild-type genotype. Adapted from the original publication (Patrick *et al.*, 2009).

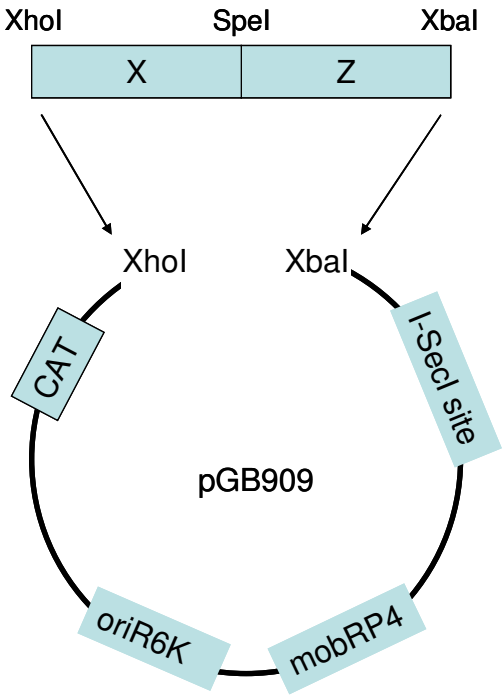


Figure 52. Generation of construct pDTORF2 for mutagenesis. This diagram demonstrates schematic representation of the construction of the suicidal vector for I-SceI based mutagenesis. Up- and down-stream fragments (X and Z) were amplified and fused together followed by ligation into pGB909 vector.

As was described above the amounts of vector and insert were equilibrated. To remove salts present in the restriction digestion buffers both vector and insert DNA were re-purified prior to ligation using a standard Gel Purification Kit (QIAGEN).

After ligation the constructs were transferred into strain *E. coli* S-17 by transformation. The transformants were selected on plates by their resistance to chloramphenicol. Colonies were transferred into liquid cultures, the plasmids were purified and screened – first by digestion with appropriate enzymes followed by separation on the agarose gel and then by amplification of the inserted fragments. The sizes of the obtained fragments corresponded directly with those estimated theoretically. The insert fragment with the size of ≈1243 bp was detected by both restriction and PCR analysis. The final confirmation of the generated construct was obtained by direct sequencing.

Attempts were made to introduce the constructed plasmid (pDTORF2) into MV-1 cells. The method of transfer used is described in the section "Plasmid transfer by conjugation". Very small dark colonies (≈0.3 mm in diameter) were observed on agar plates after 14 days; however the biomass of cells in a single colony was not sufficient to use it for subculturing. Therefore a smear of cells was taken in order to isolate DNA and test for integration of the suicidal plasmid.

Another sample of cells was taken from the tubes with 1 % (v/w) agar in liquid medium. Growth of cells MV-1 after conjugation in such tubes resulted in the formation of a dark coloured biomass of cells. The distinctive dark coloured biomass of cells on the bottom of the tube became visible after 21 days at room temperature

which is 7 days later compared to cells MV-1 carrying plasmids with EGFP (Figure 55).

A column of agar (≈30 µl) containing cells of strain MV-1 were aseptically isolated from the tube using a glass Pasteur pipette (as described in 3.1.1) and used for DNA extraction. The samples of DNA that were obtained from the tubes and surface of the plate were tested by PCR in order to show a site specific integration of the vector pDTORF2. The main idea behind these experiments was to attempt amplification of the fragment using pairs of primers one of which was complementary to the region on the chromosome whilst another one was complementary to the region of the integrated suicidal construct pDTORF2. Two possible points of site specific integration were taken into account in the process of primer design. The schematic representation of the integrated constructs and designed primers are shown below (Figure 53). As a result of PCR a fragment for the primers one of which was complementary to the I-SceI site from pDTORF2 and another primer was complementary to the region of the target gene ORF2 (bottom left on Figure 53). PCR products were separated on an agarose gel. The size of the obtained product (≈1.1 kb) matched the size estimated theoretically. The absence of product in PCR with a genomic DNA of a wild type was used as a negative control.

In order to obtain more conclusive evidence of the isolation of cells MV-1 with an integrated construct pDTORF2 it was decided to attempt an amplification of a region within the previously isolated PCR product. The PCR product was extracted from an agarose gel and used as a template in a PCR. A pair of primers that were previously designed to amplify fragment downstream of the target ORF2 (shown as fragment

"Z" on Figure 53) was used to amplify fragment of 651 bp. Obtained PCR product was separated visualized on an agarose gel. The size of the obtained product matched the size estimated theoretically.

To provide additional evidence of the isolation of the transconjugants of MV-1 with integrated construct pDTORF2 it was decided to carry out further PCR tests, primer pairs such that one was complementary to the chromosome and the other was complementary to a region of the integrated construct (upper left on Figure 53). The first PCR was carried out using a primer complementary to the region of ORF1 and a primer to the fragment Z. The product size of 1373 bp can only be amplified from the DNA of a transconjugant as in wild type there would be ORF2 in between of fragments X and Z which would result in amplification of a larger product ≈2.8 kb in size). The amplification of the product of an estimated size was followed by gel extraction. Extracted product was then used as a template in the PCR that was designed to amplify the product within the template obtained in previous PCR (shown as fragment "X" on Figure 53). The product with a size of 592 bp was successfully amplified confirming the integration of pDTORF2.

Similar PCR analysis was carried out for another region using a primer complementary to the fragment of the integrated construct (shown as "X" on the bottom right of Figure 53) and a primer complementary to the region of the ORF3. A product of an estimated size of 1473 bp was obtained and used as a template as described above to confirm the presence of a specific sequence (fragment "Z", 651) within the amplified region.

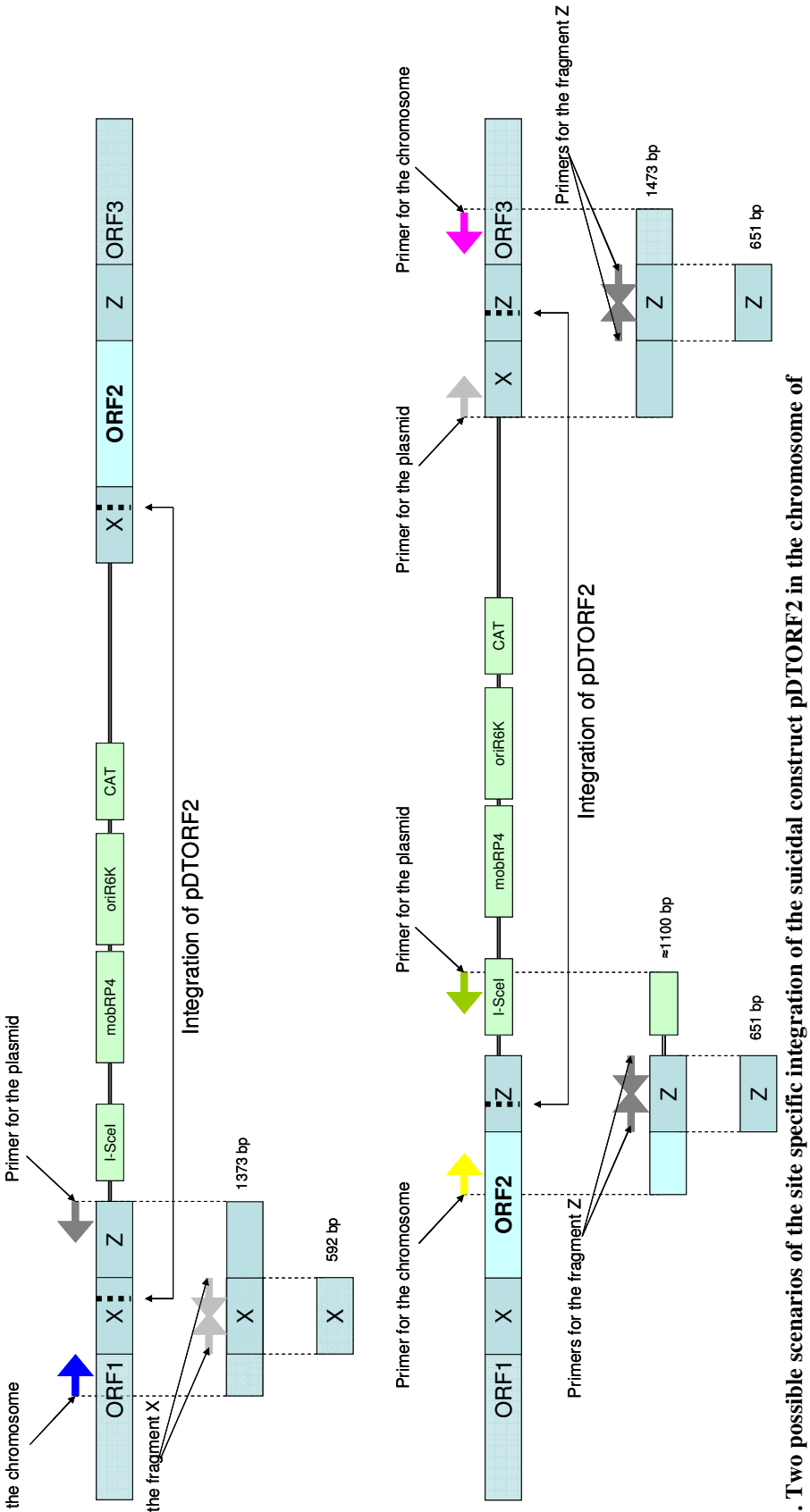


Figure 53. Two possible scenarios of the site specific integration of the suicidal construct pDTORF2 in the chromosome of strain MV-1. This diagram shows two possible outcomes of the single crossover events after introduction of the suicidal construct pDTORF2. Primers used and PCR products obtained provide evidence of the isolation of transconjugants.

To summarize, the isolation of the transconjugants within the smear of colonies was conclusively confirmed using a number of PCR reactions designed to prove a site specific integration of the construct pDTORF2 in the chromosome of strain MV-1.

3.5.4.1 Future work on generation of knock-out mutants

In order to accomplish the generation of a knock-out mutant using I-SceI system the transconjugants need to be cultured in the form of isolated colonies. The method of growth utilized (see section Culturing of strain MV-1 on solid medium) allowed visualization of isolated colonies of transconjugants resistant to chloramphenicol after about 3 weeks. Once an isolated colony is subcultured it can be used as a recipient in a mating experiment to acquire a plasmid expressing I-SceI meganuclease. As described above, the induction of the expression will eventually lead to a formation of a double strand break followed by a repair event. This should result in the generation of knock-out mutant and wild-type revertants in a ratio of 1:1 (Figure 51). The isolated colonies can be tested by PCR followed by a direct sequencing of obtained products. The obtained mutant with a deletion of ORF2 should be analysed for any phenotypic changes and should help define a role of this gene as well as a providing a first method of generation of knock-out mutants in strain MV-1 that can be used to generate mutants of other genes.

3.6 Investigation of the protein localization using fusions with EGFP

In order to investigate localization of the proteins within the cell it was decided to generate fusions with Enhanced Green Fluorescent Protein (EGFP). Sabrina Schuebbe (unpublished data) has started this work with experiments in the laboratory of Dennis Bazylinski (UNLV, Las Vegas) for several of the proteins that are products of genes specific to the "magnetosome island" and has successfully shown localization for several of such proteins. However at the time, there was no mass spectrometry data suggesting what proteins are actually identified as being abundantly present in the magnetosome membranes. The proteomics data obtained in this work allowed to outline a number of proteins that are present in the magnetosome membrane in higher concentrations (see Magnetosome membrane proteins identification).

One of such proteins is CAV30779.1 a protein identified to be present in the magnetosome membrane by all mass spectrometry methods used in this work. According to BLAST results this protein has similarity to von Willebrand factor (vWF) type A domain. There are suggestions in the literature that proteins containing this domain in bacterial cells may be involved in the metal ion and/or protein-binding functions (Ponting *et al.*, 1999). Therefore it was decided that this protein is a good target for the investigation of localization by EGFP fusion.

The plan of generation of a plasmid expressing a fusion between EGFP and CAV30779.1 is shown on the next diagram (Figure 54). The ORF coding for this protein (918 bp) was amplified by PCR with primers containing restriction sites for KpnI and NheI respectively followed by gel purification and digestion with the mentioned enzymes.

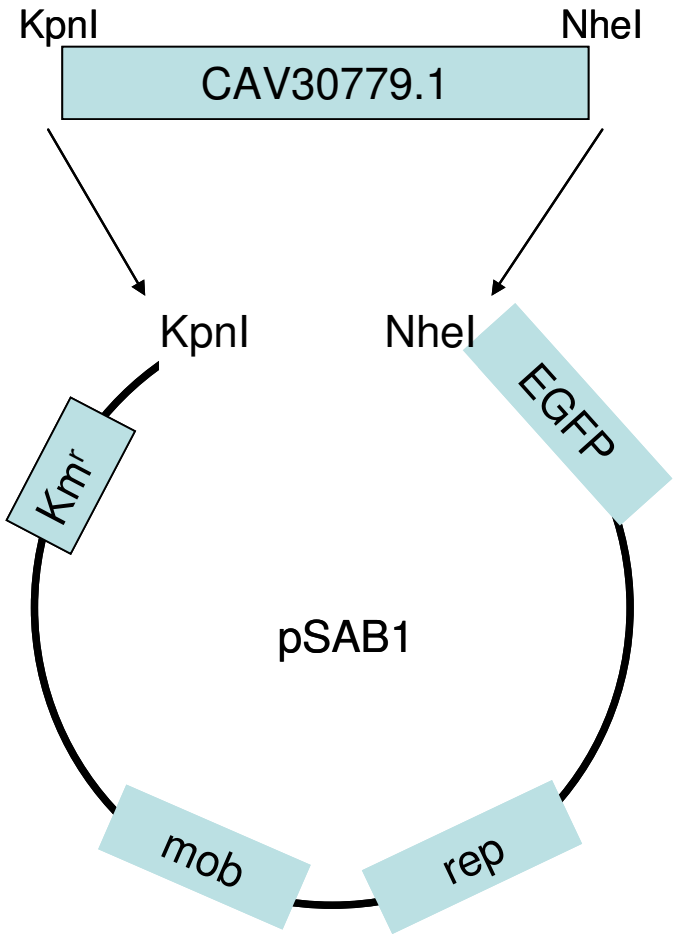


Figure 54. Generating the construct pDT779EGFP for EGFP fusion with CAV30779.1. This diagram is a schematic representation of the construction of the plasmid for EGFP fusion expression. The amplified gene of 918 bp coding for CAV30779.1 protein was ligated into the pSAB1 plasmid.

The pSAB1 vector (courtesy of Sabrina Schuebbe) is pBB1MCS2 vector with an integrated EGFP fragment (≈800 bp) between sites XbaI and HindIII. This plasmid was digested with KpnI and NheI followed by an enzyme inactivation and dephosphorylation of the 5´ phosphate groups to prevent self-ligations. The obtained fragments were ligated and transformed into *E. coli* WM3064 with selection for kanamycin resistance. Transformants were subcultured followed by plasmid isolation to screen for correct construct. The restriction digestion with agarose gel electrophoresis was used to confirm generated construct (named pDT779EGFP) and the sizes of the obtained fragments corresponded directly with those estimated theoretically.

This construct was transferred into the MV-1 cells as explained in section 3.5.1 (Plasmid transfer by conjugation); control plasmid pSAB1 containing a gene expressing native EGFP protein, was also transferred into MV-1 cells to use as a control for fluorescence in cells.

It was possible to obtain the growth of transconjugants in the form of isolated colonies by culturing transconjugants in the tubes with 1 % (v/w) agar in liquid medium. There were two interesting trends to be observed. First, cells of MV-1 have moved whilst the medium was in liquid phase and formed a layer of cells about 3 mm deep from the surface of the agar that became visible after around 7 days (area C, Figure 55). This can be explained by the fact that air was not removed from the tubes and cells probably moved within a gradient of oxygen to an optimal microaerophilic conditions (phenomenon was observed previously by F.B. Ward, personal communication). Second, in tubes with cells carrying plasmids pSAB1 and pDT779EGFP there was a second dark area at the bottom of the tubes (area A,

Figure 55) suggesting that cells are growing successfully whilst far from the oxygen on the top. This, however, was not observed initially for experiments with suicidal vector pDTORF2 (see Mutants construction using I-SceI system). There was also an area with no visible growth (area B, Figure 55).

A possible explanation to the presence of two areas of growth is that cells at the top layer find optimal microaerophilic conditions and use oxygen as an electron acceptor, whilst cells at the bottom of the tubes grow anaerobically using N_2O as a terminal electron acceptor.

Interestingly, it was possible to observe fluorescence emitted by cells in the tube by eye using the Safe Imager Transilluminator (Invitrogen). The tube with cells that acquired plasmid pSAB1 shown significantly higher level of fluorescence compared to the tube with cells carrying plasmid pDT779EGFP (Figure 56). Such fluorescence was observed only for the cells at the top of the tube (area C, Figure 55). It can be explained by the fact that production (folding) of EGFP proteins requires molecular oxygen and therefore can be the reason for the absence of sufficient levels of fluorescence to be visible by eye at the bottom of the tube. The effects of oxygen on the levels of fluorescence of EGFP fusions were shown in experiments carried out with *M. gryphiswaldense* strain MSR-1 (Hansen *et al.*, 2001; Lang and Schuler, 2008).

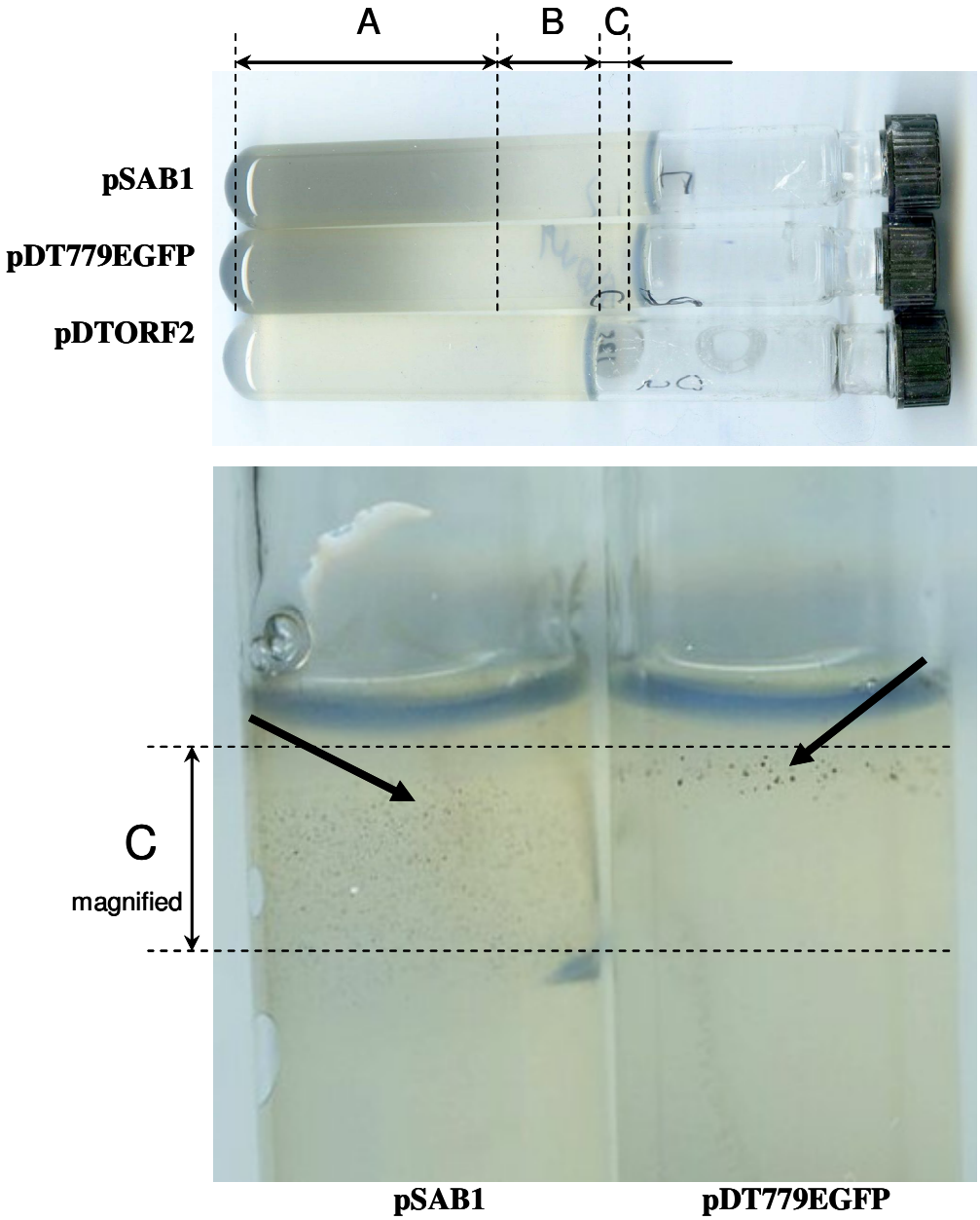
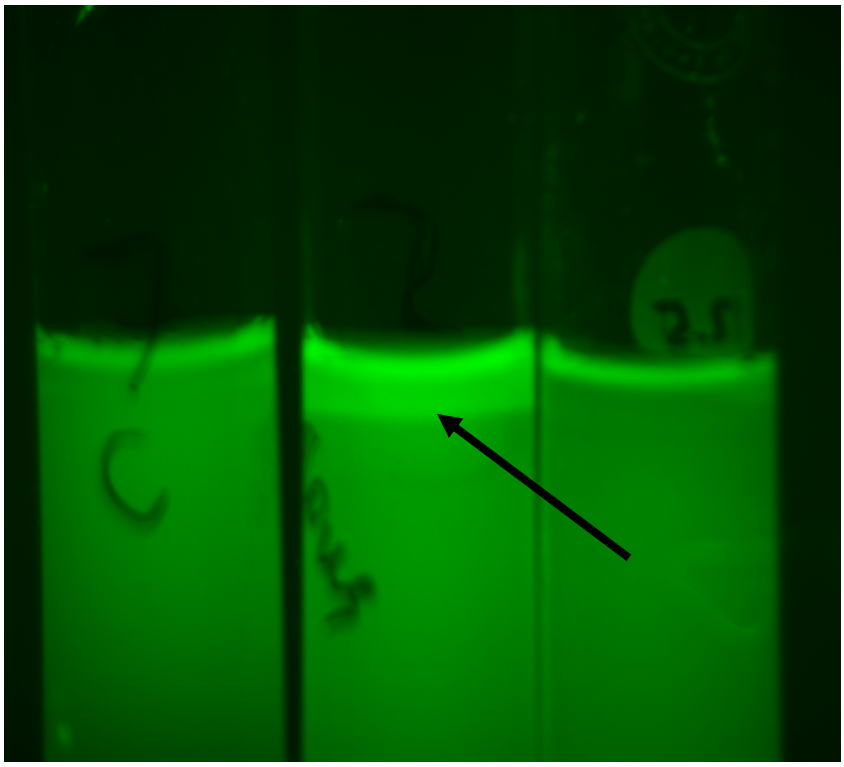
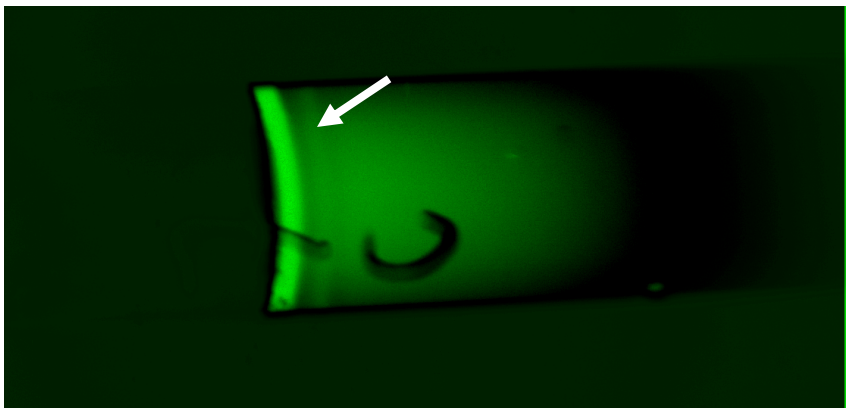


Figure 55. Transconjugants of MV-1 cultured in agar test tubes. These photographs show the growth of MV-1 transconjugants. For cells that acquired plasmids pSAB1and pDT779EGFP the distinctive dark coloured biomass of cells at the bottom of the tubes (A) appeared after 7 days after inoculation, whilst for pDTORF2 it appeared after 14 days. The area with no visible growth is shown as B and area with growing cells (C). The dark colour suggests that cells produce magnetosomes. There are also typical dark brown colonies that became visible after 14 days (shown with arrows) in the area C. The growth and inoculation conditions were identical. The tube with cells carrying pSAB1 plasmid with native EGFP contains larger number and more uniform colonies compared to the cells carrying pDT779EGFP.



pDT779EGFP **pSAB1** **pDTORF2**



pDT779EGFP

Figure 56. Fluorescecent cells in the agar tubes. These photograph demonstrate fluorescence of the cells carrying plasmids expressing EGFP. The Safe Imager Transilluminator (Invitrogen) was adapted for use with the digital camera (UVP Laboratory products). The tube with transconjugants that acquired plasmid pDTORF2 (no EGFP) was used as a negative control. The fluorescence from the cells is not to be confused with fluorescence occurring on the edge of the medium. A bright fluorescent ring can be observed for native EGFP expressed from pSAB1 (black arrow), less intense fluorescence emitted by the cells carrying plasmid pDT779EGFP can be observed on the magnified photograph with increased exposure time (white arrow). Tubes were photographed 10 days after inoculation.

In order to detect fluorescence transconjugants were taken from the layer of cells on the surface of the agar plates and examined by microscopy. The phase contrast microscopy confirmed the strain MV-1 cells phenotype. Further analysis was carried out to detect fluorescence in the cells of MV-1. Fluorescence was detected in both control cells carrying pSAB1 plasmid and expressing native EGFP and in cells carrying plasmid pDT779EGFP expressing EGFP fused with CAV30779.1 protein. The micrographs are shown below (Figure 57).

The distribution within the cell does not appear to be specific to the magnetosome chains. Such general distribution can be explain by the fact that, as described above, protein CAV30779.1 has similarity to von Willebrand factor (vWF) type A domain that can be involved in the transport of iron. And therefore should not necessarily only be found to be associated with a magnetosome chain. In order to confirm that EGFP dissociates from the fusion, SDS PAGE can be used to compare molecular weigh difference between EGFP and EGFP- CAV30779.1 fusion.

It is also interesting that cells expressing the EGFP-CAV30779.1 fusion demonstrate lower viability compared to those expressing just EGFP. This can be seen on the comparison of the number of colonies when cultured in the agar tubes (Figure 55) and it was also detected by microscopy. A possible explanation of this phenomenon is that this protein remains partially or fully functioning whilst in fusion with EGFP and may be over expressed and cause some toxic effect on the cells by transporting higher amounts of iron. To test this hypothesis cells of MV-1 carrying the above plasmids can be cultured without iron source in the medium followed by comparison of the growth rates those expressing EGFP and the fusion protein.

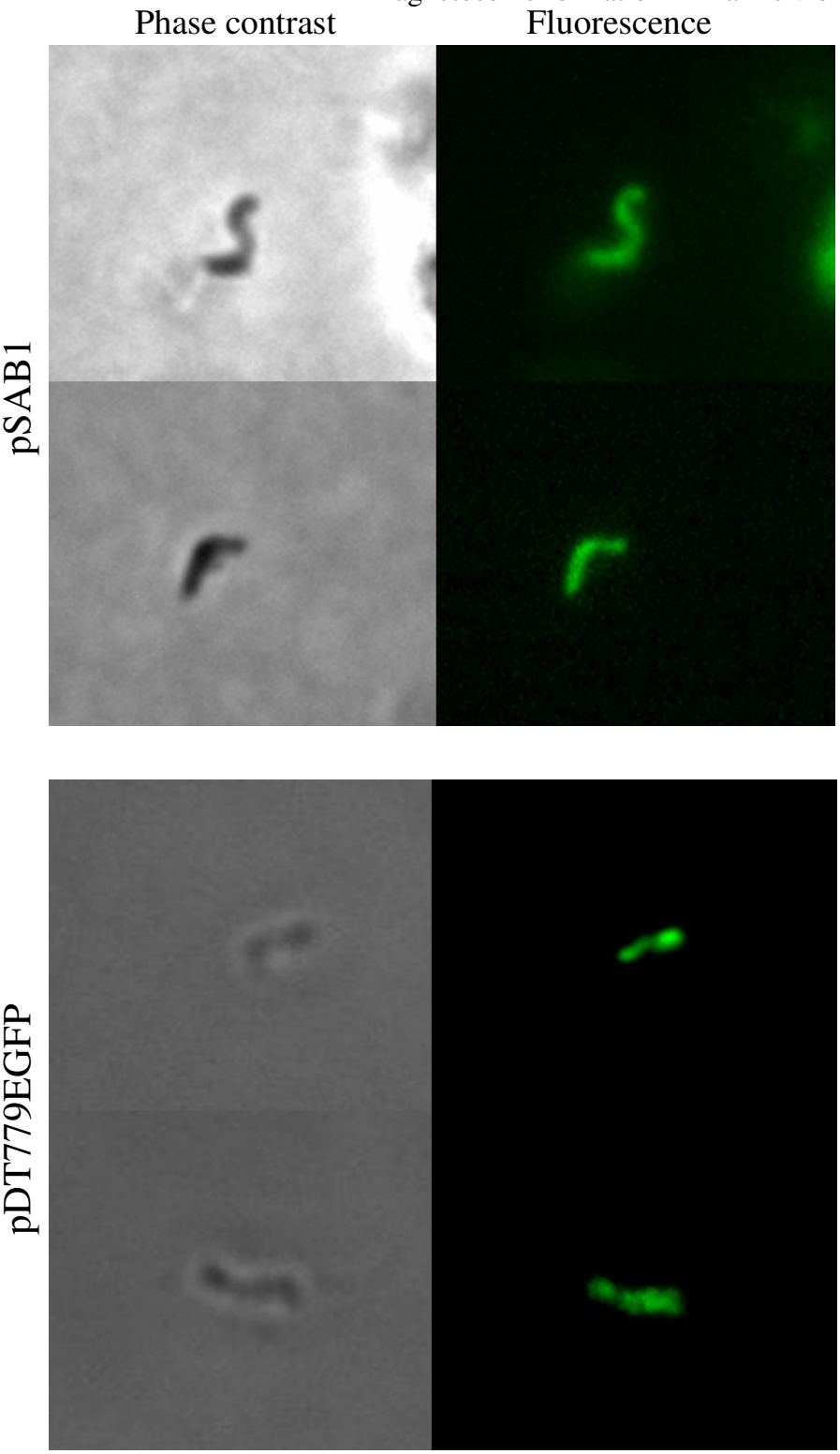


Figure 57. EGFP fluorescence visualized in the cells of MV-1. These micrographs show fluorescence in cells MV-1 after acquiring plasmids pSAB1 and pDT779EGFP respectively. The fluorescence of the EGFP fusion with protein CAV30779.1 shows cytoplasmic localization similar to the general distribution shown for native EGFP but tends to be less consistent.

Another experiment to test the effects of over expression of the fusion between EGFP and CAV30779.1 is to induce the LacZ promoter by addition of lactose or Isopropyl β-D-1-thiogalactopyranoside and measure the viability of the cells.

These experiments demonstrate that it is possible to efficiently generate and transfer into cells of MV-1 plasmid constructs expressing EGFP fusions with proteins of interest. The investigation of cell localization of the proteins that are associated with magnetosome membrane using fusions with fluorescent proteins was shown to be an efficient approach that allowed investigation of the cell localization of a number of proteins for example MamC, MamF, and MamG that were shown to be associated with magnetosome chain in *M. gryphiswaldense* (Lang and Schuler, 2008). The use of this approach can be extended to other proteins specific to the "magnetosome island" of strain MV-1. Similar analysis involving several other genes was carried out by Sabrina Schuebbe (unpublished data).

A very interesting microscopy technique that allows the visualization of single molecule fusion of a fluorescent protein and a protein of interest within the live cells was described in a recently published review. The use of the Ultrahigh-resolution imaging techniques such as Fluorescence Imaging with One Nanometer Accuracy (FIONA). This method is based on a single fluorophore being excited with wide field illumination. The exact centre of the fluorescent spot is then localized by a two-dimensional Gaussian fit (Yildiz and Selvin, 2005). Another technique that can be used is Single-molecule High-Resolution Co-localization (SHREC) that allows visualization of two fluorescent markers within one sample (Churchman *et al.*, 2005). The visualization of freely diffusing molecules using such methods can bring the analysis of magnetosome protein localization to a new level.

4 Conclusions

In the course of this study the sequence and organization of the genes of the "magnetosome island" of marine vibrio MV-1 was established using inverse PCR and whole genome sequencing. The sequencing of the genome was followed by an automated annotation by RAST has generated 3.6 Mb of sequence with 3,741 predicted ORFs. The work on completion of the genome sequence is still in progress. The search for magnetosome formation gene was carried out within the generated sequence. This resulted in identification of 39 genes located within this cluster 27 of which have similarity with magnetosome formation genes present in previously characterized MB. Several homologous copies of the magnetosome formation genes are found in the sequence. There are 2 copies of *mamK* and *mamX* genes, three genes similar to *mamR* and, finally, 3 copies of *mmsF* gene one of which was found to be located outside the "magnetosome island" region.

The magnetosome membrane fraction was isolated by two different methods. The identification of the proteins associated with this fraction was carried out using different mass spectrometry techniques including MALDI-TOF, LC-MS and Orbitrap. The generated spectra was used to search against the database containing genes of the 107 kb fragment of containing "magnetosome island" resulting in identification of 33 and 35 proteins isolated by the method developed in this work and the method developed by Tanaka *et al.* respectively. The search against database generated in automated annotation and containing all predicted proteins has resulted in identification of 486 and 463 proteins isolated method by the method developed in this work and the method developed by Tanaka *et al.* respectively. This suggests that

with the use of sensitive equipment such as Orbitrap it is needed to utilize not only qualitative but also quantitative proteomics approaches in order to outline proteins that dominant in the magnetosome membrane fraction. A universal method of magnetosome membrane isolation used by independent researchers should make the process of protein analysis of the magnetosome membrane more consistent and reproducible.

In order to investigate cell localization of the protein identified by mass spectrometry to be present in the magnetosome membrane fraction a construct expressing a fusion of EGFP and CAV30779.1 was made. The detected fluorescence in cells of MV-1 carrying the generated plasmid did not show any specific localizations that may suggest that this protein is not specifically associated with the magnetosome chain.

In order to generate a knock-out mutant a suicidal construct carrying fused upstream and downstream sequence of the target gene ORF2 in *mamK*-II cluster and I-SceI recognition site was generated. The transfer of this construct by conjugation has allowed isolation of MV-1 cells with the construct being integrated into the chromosome in a specific location. The future work should involve transfer of the plasmid that expresses I-SceI meganuclease that will create a double strand break and induce a recombination event that should result in generation of a deletion of the target gene.

To summarize, this work has generated significant amount of novel data on magnetosome formation in marine vibrio MV-1. This includes identification of the "magnetosome island" and its genes, identification of a homologous copy of mmsF-III outside of "magnetosome island", generation of a genome sequence, identification

of magnetosome membrane proteins, investigation of the EGFP cell localization of a protein from the “magnetosome island” and a significant progress on an attempt to generate a knock-out mutant.

5 References

Arakaki, A., Webb, J., and Matsunaga, T. (2003) A novel protein tightly bound to bacterial magnetic particles in Magnetospirillum magneticum strain AMB-1. *J Biol Chem* **278**: 8745-8750.

Arakaki, A., Shibusawa, M., Hosokawa, M., and Matsunaga, T. (2010) Preparation of genomic DNA from a single species of uncultured magnetotactic bacterium by multiple-displacement amplification. *Appl Environ Microbiol* **76**: 1480-1485.

Bakke, P., Carney, N., Deloache, W., Gearing, M., Ingvorsen, K., Lotz, M., McNair, J., Penumetcha, P., Simpson, S., Voss, L., Win, M., Heyer, L.J., and Campbell, A.M. (2009) Evaluation of three automated genome annotations for Halorhabdus utahensis. *PLoS One* **4**: e6291.

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

Bazylinski, D.A., and Frankel, R.B. (2003) Biologically controlled mineralization in prokaryotes. *Rev Mineral Geochem* **54**: 217-247.

Bazylinski, D.A., and Frankel, R.B. (2004) Magnetosome formation in prokaryotes. *Nat Rev Microbiol* **2**: 217-230.

Bellini, S. (1963a) Su di un particolare comportamento di batteri d'acqua dolce. Salvatore Instituto di Microbiologia dell'Universita di Pavia.

Bellini, S. (1963b) Ulteriori studi sui "batteri magnetosensibili". Salvatore Instituto di Microbiologia dell'Universita di Pavia.

Berg, H.C., and Brown, D.A. (1972) Chemotaxis in Escherichia coli analysed by three-dimensional tracking. *Nature* **239**: 500-504.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol* **62**: 293-300.

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

Blakemore, R.P. (1982) Magnetotactic bacteria. *Annu Rev Microbiol* **36**: 217-238.

Calugay, R.J., Miyashita, H., Okamura, Y., and Matsunaga, T. (2003) Siderophore production by the magnetic bacterium Magnetospirillum magneticum AMB-1. *FEMS Microbiol Lett* **218**: 371-375.

Churchman, L.S., Okten, Z., Rock, R.S., Dawson, J.F., and Spudich, J.A. (2005) Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proc Natl Acad Sci U S A* **102**: 1419-1423.

Cox, B.L., Popa, R., Bazylinski, D.A., Lanoil, B., Douglas, S., Belz, A., Engler, D.L., and Nealson, K.H. (2002) Organization and Elemental Analysis of P-, S-, and Fe-rich Inclusions in a Population of Freshwater Magnetococci. *Geomicrobiology Journal* **19**: 387-406.

Cronshaw, A., and Florence, H. (2002) Identification of proteins by mass spectrometry. Edinburgh: The Unuiversity of Edinburgh.

Dean, A.J., and Bazylinski, D.A. (1999) Genome analysis of several marine, magnetotactic bacterial strains by pulsed-field gel electrophoresis. *Current Microbiology* **39**: 219-225.

DeLong, E.F., Frankel, R.B., and Bazylinski, D.A. (1993) Multiple Evolutionary Origins of Magnetotaxis in Bacteria. *Science* **259**: 803-806.

Dubbels, B.L., DiSpirito, A.A., Morton, J.D., Semrau, J.D., Neto, J.N., and Bazylinski, D.A. (2004) Evidence for a copper-dependent iron transport system in the marine, magnetotactic bacterium strain MV-1. *Microbiology* **150**: 2931-2945.

Dunin-Borkowski, R.E., McCartney, M.R., Frankel, R.B., Bazylinski, D.A., Posfai, M., and Buseck, P.R. (1998) Magnetic microstructure of magnetotactic bacteria by electron holography. *Science* **282**: 1868-1870.

Flies, C.B., Peplies, J., and Schuler, D. (2005) Combined approach for characterization of uncultivated magnetotactic bacteria from various aquatic environments. *Appl Environ Microbiol* **71**: 2723-2731.

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

Frankel, R.B., and Bazylinski, D.A. (2004) Magnetosome Mysteries. *ASM News*: 176-183.

Frankel, R.B. (2009) The discovery of magnetotactic/magnetosensitive bacteria. *Chinese Journal of Oceanology and Limnology* **27**: 1-2.

Futai, M. (1974) Orientation of membrane vesicles from Escherichia coli prepared by different procedures. *J Membr Biol* **15**: 15-28.

Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M., and Brinkman, F.S. (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**: 617-623.

Gorby, Y.A., Beveridge, T.J., and Blakemore, R.P. (1988) Characterization of the bacterial magnetosome membrane. *J Bacteriol* **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. *Appl Environ Microbiol* **67**: 4573-4582.

Grunberg, K., Muller, E.C., Otto, A., Reszka, R., Linder, D., Kube, M., Reinhardt, R., and Schuler, D. (2004) Biochemical and proteomic analysis of the magnetosome membrane in Magnetospirillum gryphiswaldense. *Appl Environ Microbiol* **70**: 1040-1050.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**: 557-580.

Hansen, M.C., Palmer, R.J., Jr., Udsen, C., White, D.C., and Molin, S. (2001) Assessment of GFP fluorescence in cells of Streptococcus gordonii under conditions of low pH and low oxygen concentration. *Microbiology* **147**: 1383-1391.

Harrison, M.D., Jones, C.E., and Dameron, C.T. (1999) Copper chaperones: function, structure and copper-binding properties. *J Biol Inorg Chem* **4**: 145-153.

Huston, W.M., Jennings, M.P., and McEwan, A.G. (2002) The multicopper oxidase of Pseudomonas aeruginosa is a ferroxidase with a central role in iron acquisition. *Mol Microbiol* **45**: 1741-1750.

Ishihama, Y., Rappsilber, J., Andersen, J.S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J Chromatogr A* **979**: 233-239.

Jogler, C., Kube, M., Schubbe, S., Ullrich, S., Teeling, H., Bazylinski, D.A., Reinhardt, R., and Schuler, D. (2009a) Comparative analysis of magnetosome gene clusters in magnetotactic bacteria provides further evidence for horizontal gene transfer. *Environ Microbiol* **11**: 1267-1277.

Jogler, C., Lin, W., Meyerdierks, A., Kube, M., Katzmann, E., Flies, C., Pan, Y., Amann, R., Reinhardt, R., and Schuler, D. (2009b) Toward cloning of the magnetotactic metagenome: identification of magnetosome island gene clusters in uncultivated magnetotactic bacteria from different aquatic sediments. *Appl Environ Microbiol* **75**: 3972-3979.

Jogler, C., and Schuler, D. (2009) Genomics, genetics, and cell biology of magnetosome formation. *Annu Rev Microbiol* **63**: 501-521.

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 Microbiol Lett* **240**: 203-208.

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. *Proc Natl Acad Sci U S A* **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. (2007) Molecular mechanisms of magnetosome formation. *Annu Rev Biochem* **76**: 351-366.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lang, C., and Schuler, D. (2008) Expression of green fluorescent protein fused to magnetosome proteins in microaerophilic magnetotactic bacteria. *Appl Environ Microbiol* **74**: 4944-4953.

Lins, U., McCartney, M.R., Farina, M., Frankel, R.B., and Buseck, P.R. (2005) Habits of magnetosome crystals in coccoid magnetotactic bacteria. *Appl Environ Microbiol* **71**: 4902-4905.

Marx, C.J., and Lidstrom, M.E. (2002) Broad-host-range cre-lox system for antibiotic marker recycling in gram-negative bacteria. *Biotechniques* **33**: 1062-1067.

Matsunaga, T., and Tsujimura, N. (1993) Respiratory inhibitors of a magnetic bacterium Magnetospirillum sp. AMB-1 capable of growing aerobically. *Applied Microbiology and Biotechnology* **39**: 368-371.

Matsunaga, T., Okamura, Y., and Tanaka, M. (2004) Biotechnological application of nano-scale engineered bacterial magnetic particles. *Journal of Materials Chemistry* **14**: 2 0 9 9 – 2 1 0 5.

Matsunaga, T., Nemoto, M., Arakaki, A., and Tanaka, M. (2009) Proteomic analysis of irregular, bullet-shaped magnetosomes in the sulphate-reducing magnetotactic bacterium Desulfovibrio magneticus RS-1. *Proteomics* **9**: 3341-3352.

Murat, D., Quinlan, A., Vali, H., and Komeili, A. (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc Natl Acad Sci U S A* **107**: 5593-5598.

Nakamura, C., Burgess, J.G., Sode, K., and Matsunaga, T. (1995) An iron-regulated gene, magA, encoding an iron transport protein of Magnetospirillum sp. strain AMB-1. *J Biol Chem* **270**: 28392-28396.

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 Res* **19**: 1801-1808.

Natarajan, D., and Boulter, C.A. (1995) Isolation of genomic sequences flanking a retroviral insertion site using a novel PCR-based method. *Gene* **161**: 195-198.

Noguchi, Y., Fujiwara, T., Yoshimatsu, K., and Fukumori, Y. (1999) Iron reductase for magnetite synthesis in the magnetotactic bacterium Magnetospirillum magnetotacticum. *J Bacteriol* **181**: 2142-2147.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**: 621-623.

Ofer, S., Nowik, I., Bauminger, E.R., Papaefthymiou, G.C., Frankel, R.B., and Blakemore, R.P. (1984) Magnetosome dynamics in magnetotactic bacteria. *Biophys J* **46**: 57-64.

Okamura, Y., Takeyama, H., and Matsunaga, T. (2001) A magnetosome-specific GTPase from the magnetic bacterium Magnetospirillum magneticum AMB-1. *J Biol Chem* **276**: 48183-48188.

Paoletti, L.C., and Blakemore, R.P. (1986) Hydroxamate production by Aquaspirillum magnetotacticum. *J Bacteriol* **167**: 73-76.

Patrick, S., Houston, S., Thacker, Z., and Blakely, G.W. (2009) Mutational analysis of genes implicated in LPS and capsular polysaccharide biosynthesis in the opportunistic pathogen Bacteroides fragilis. *Microbiology* **155**: 1039-1049.

Ponting, C.P., Aravind, L., Schultz, J., Bork, P., and Koonin, E.V. (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* **289**: 729-745.

Posfai, M., Buseck, P.R., Bazylinski, D.A., and Frankel, R.B. (1998) Iron sulfides from magnetotactic bacteria: Structure, composition, and phase transitions. *American Mineralogist* **83**: 1469–1481.

Richter, M., Kube, M., Bazylinski, D.A., Lombardot, T., Glockner, F.O., Reinhardt, R., and Schuler, D. (2007) Comparative genome analysis of four magnetotactic bacteria reveals a complex set of group-specific genes implicated in magnetosome biomineralization and function. *J Bacteriol* **189**: 4899-4910.

Rong, C., Huang, Y., Zhang, W., Jiang, W., Li, Y., and Li, J. (2008) Ferrous iron transport protein B gene (feoB1) plays an accessory role in magnetosome formation in Magnetospirillum gryphiswaldense strain MSR-1. *Res Microbiol* **159**: 530-536.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning : a laboratory manual.* Cold Spring Harbor: Cold Sprng Harbor Laboratory Press.

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

Schuebbe, S., Kube, M., Scheffel, A., Wawer, C., Heyen, U., Meyerdierks, A., Madkour, M.H., Mayer, F., Reinhardt, R., and Schueler, D. (2003) Characterization of a spontaneous nonmagnetic mutant of Magnetospirillum gryphiswaldense reveals a large deletion comprising a putative magnetosome island. Unpublished.

Schuler, D., and Baeuerlein, E. (1996) Iron-limited growth and kinetics of iron uptake in Magnetospirillum gryphiswaldense. *Arch Microbiol* **166**: 301-307.

Schuler, D. (2002) The biomineralization of magnetosomes in Magnetospirillum gryphiswaldense. *Int Microbiol* **5**: 209-214.

Schuler, D. (2008) Genetics and cell biology of magnetosome formation in magnetotactic bacteria. *FEMS Microbiol Rev* **32**: 654-672.

Schultheiss, D., and Schuler, D. (2003) Development of a genetic system for Magnetospirillum gryphiswaldense. *Arch Microbiol* **179**: 89-94.

Sennels, L., Bukowski-Wills, J.C., and Rappsilber, J. (2009) Improved results in proteomics by use of local and peptide-class specific false discovery rates. *BMC Bioinformatics* **10**: 179.

Short, K.A., and Blakemore, R.P. (1986) Iron respiration-driven proton translocation in aerobic bacteria. *J Bacteriol* **167**: 729-731.

Spring, S., Amann, R., Ludwig, W., Schleifer, K.H., van Gemerden, H., and Petersen, N. (1993) Dominating Role of an Unusual Magnetotactic Bacterium in the Microaerobic Zone of a Freshwater Sediment. *Appl Environ Microbiol* **59**: 2397-2403.

Spring, S., and Bazylinski, D.A. (2000) Magnetotactic bacteria. In *The prokaryotes: an evolving electronic resource for the microbiological community*. Dworkin, M. (ed.) New York, N.Y.

St. Pierre T. G., Clark P.R., Chua-anusorn W., Fleming A., Pardoe H., Jeffrey G. P., Olynyk J. K., Pootrakul P., Jones S., and Moroz, P. (2005) Non-invasive measurement and imaging of tissue iron oxide nanoparticle concentrations *in vivo* using proton relaxometry. In *Journal of Physics: Conference Series*. Vol. 17, pp. 122-126.

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. *Mol Microbiol*.

Taoka, A., Asada, R., Sasaki, H., Anzawa, K., Wu, L.F., and Fukumori, Y. (2006) Spatial localizations of Mam22 and Mam12 in the magnetosomes of Magnetospirillum magnetotacticum. *J Bacteriol* **188**: 3805-3812.

Taoka, A., Umeyama, C., and Fukumori, Y. (2009) Identification of iron transporters expressed in the magnetotactic bacterium Magnetospirillum magnetotacticum. *Curr Microbiol* **58**: 177-181.

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. *J Bacteriol* **187**: 7176-7184.

Wilson, K. (1997) Preparation of Genomic DNA from Bacteria. *Current Protocols in Molecular Biology* **2.4.1-2.4.5**.

Xia, M., Wei, J., Lei, Y., and Ying, L. (2007) A novel ferric reductase purified from Magnetospirillum gryphiswaldense MSR-1. *Curr Microbiol* **55**: 71-75.

Yildiz, A., and Selvin, P.R. (2005) Fluorescence imaging with one nanometer accuracy: application to molecular motors. *Acc Chem Res* **38**: 574-582.

Yu, C.S., Lin, C.J., and Hwang, J.K. (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* **13**: 1402-1406.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821-829.

Zieske, L.R. (2006) A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* **57**: 1501-1508.

Appendix I

Proteins identified using Orbitrap instrument in the magnetosome membrane fraction prepared by the method developed in this work.

Number	Accession number	Description	Mass [Da]	Peptides	Score
1	MV-43-8	DNA-directed RNA polymerase beta' subunit (EC	155896	44	2192
2	MV-43-7	DNA-directed RNA polymerase beta subunit (EC	154088	34	1665
3	MV-7-23	polyketide synthase, type I	347856	31	1488
4	MV-1-34	Propionyl-CoA carboxylase biotin-containing	80412	30	1440
5	MV-79-53	Cytosol aminopeptidase PepA (EC 3.4.11.1)	52892	23	1222
6	MV-117-5	acriflavin resistance protein	149841	22	1086
7	MV-42-33	Malate dehydrogenase (EC 1.1.1.37)	38732	20	1315
8	MV-65-111	NAD-specific glutamate dehydrogenase (EC	181583	20	899
9	MV-7-25	mixed type I polyketide synthase - peptide	237111	19	812
10	MV-65-4	ATP-dependent RNA helicase	78469	17	675
11	MV-65-134	Porin, Gram-negative type	38577	15	1128
12	MV-42-18	ATP synthase beta chain (EC 3.6.3.14)	50800	15	808
13	MV-1-32	Propionyl-CoA carboxylase carboxyl transferase	56263	15	786
14	MV-3-32	Signal transduction histidine kinase CheA (EC	103432	15	725
15	MV-84-100	Translation elongation factor Tu	43088	15	582
16	MV-26-44	Signal transduction histidine kinase	129548	13	649
17	MV-98-89	YjeF protein, function unknown	54514	13	613
18	MV-7-28	polyketide synthase	265141	13	552
19	MV-63-47	ABC-type protease/lipase transport system,	86035	12	606
20	MV-26-51	Chorismate synthase (EC 4.2.3.5) # AroG	38692	12	551
21	MV-42-30	2-oxoglutarate dehydrogenase E1 component (EC	111107	12	548
22	MV-42-82	5-methyltetrahydrofolate--homocysteine	134485	12	457
23	MV-35-6	Serine protease precursor MucD/AlgY associated	74272	11	663
24	MV-42-98	Adenylylsulfate reductase alpha-subunit (EC	70715	11	534
25	MV-35-2	serine protease	73406	11	488
26	MV-77-100	(S)-2-hydroxy-acid oxidase	142895	11	478
27	MV-7-29	mixed type I polyketide synthase - peptide	221054	11	421
28	MV-98-145	COG0840: Methyl-accepting chemotaxis protein	16861	10	630
29	MV-80-75	Cytochrome c oxidase subunit CcoP (EC 1.9.3.1)	31863	10	549
30	MV-63-46	Adenylate cyclase (EC 4.6.1.1)	41260	10	495
31	MV-7-24	polyketide synthase of type I	124123	10	459
32	MV-104-3	transport protein	32458	10	459
33	MV-42-20	ATP synthase alpha chain (EC 3.6.3.14)	54905	10	456
34	MV-7-45	6-phosphofructokinase (EC 2.7.1.11)	43962	10	445
35	MV-84-104	LSU ribosomal protein L4p (L1e)	22242	9	521
36	MV-60-2	hypothetical protein	10576	9	461
37	MV-7-30	Polyketide synthase	173413	9	440
38	MV-110-13	Putrescine ABC transporter putrescine-binding	40239	9	415
39	MV-87-3	SSU ribosomal protein S4p (S9e)	23556	9	378
40	MV-42-106	transcriptional regulator, Fis family	105437	9	334
41	MV-7-22	hypothetical protein	18414	8	483
42	MV-80-74	Type cbb3 cytochrome oxidase biogenesis protein	55153	8	440
43	MV-38-51	Methyl-accepting chemotaxis protein	59277	8	406
44	MV-84-106	LSU ribosomal protein L2p (L8e)	30077	8	389
45	MV-77-9	Cobalamin biosynthesis protein CbiG /	91161	8	384
46	MV-1-11	COG0206: Cell division GTPase	19676	8	371
47	MV-32-25	Cysteine desulfurase (EC 2.8.1.7)	45960	8	366

Number	Accession number	Description	Mass [Da]	Peptides	Score
48	MV-45-7	NAD kinase (EC 2.7.1.23)	28449	8	346
49	MV-82-14	Pyruvate dehydrogenase E1 component alpha	39094	8	338
50	MV-112-41	OmpA family protein	30024	8	328
51	MV-113-6	Protein-export membrane protein SecD (TC	55977	8	314
52	MV-80-94	Oxidoreductase (flavoprotein)	42371	8	300
53	MV-84-126	SSU ribosomal protein S11p (S14e)	14109	7	504
54	MV-35-14	Methyl-accepting chemotaxis protein	34372	7	489
55	MV-56-26	COG5283: Phage-related tail protein	27936	7	459
56	MV-38-20	Pyrophosphate-energized proton pump (EC	68887	7	418
57	MV-35-4	Cobalt-zinc-cadmium resistance protein	30736	7	400
58	MV-42-136	hypothetical protein	31297	7	378
59	MV-84-109	SSU ribosomal protein S3p (S3e)	25320	7	375
60	MV-17-6	PAS	94353	7	321
61	MV-42-205	Phosphoenolpyruvate carboxykinase [GTP] (EC	67241	7	321
62	MV-116-54	NAD(P) transhydrogenase subunit beta (EC	47910	7	310
63	MV-42-137	hypothetical protein	44908	7	292
64	MV-84-125	SSU ribosomal protein S13p (S18e)	13757	7	288
65	MV-3-1	Flagellin protein FlaA	40038	6	457
66	MV-84-110	LSU ribosomal protein L16p (L10e)	15911	6	345
67	MV-42-29	Dihydrolipoamide succinyltransferase component	44274	6	312
68	MV-16-1	hypothetical protein	34559	6	296
69	MV-1-7	Permease of the major facilitator superfamily	47996	6	291
70	MV-39-8	Cytochrome b subunit of the bc complex	32215	6	288
71	MV-84-122	LSU ribosomal protein L15p (L27Ae)	17425	6	281
72	MV-109-22	OmpA/MotB	17362	6	270
73	MV-63-45	Adenylate cyclase (EC 4.6.1.1)	42118	6	267
74	MV-42-194	Glutamate synthase [NADPH] large chain (EC	169803	6	257
75	MV-117-11	hypothetical protein	38456	6	250
76	MV-98-85	ATP-dependent protease La (EC 3.4.21.53) Type	89811	6	219
77	MV-84-112	SSU ribosomal protein S17p (S11e)	9232	6	190
78	MV-84-108	LSU ribosomal protein L22p (L17e)	14259	5	358
79	MV-56-59	Fructose-bisphosphate aldolase class II (EC	38377	5	298
80	MV-98-127	methyltransferase, putative	25236	5	279
81	MV-65-133	Porin 41 (Por41) precursor	38025	5	267
82	MV-77-92	LSU ribosomal protein L20p	13228	5	264
83	MV-7-3	LSU ribosomal protein L13p (L13Ae)	17479	5	262
84	MV-80-78	Cytochrome c oxidase subunit CcoN (EC 1.9.3.1)	55653	5	261
85	MV-49-8	RND efflux system, inner membrane transporter	72460	5	247
86	MV-39-124	Aconitate hydratase (EC 4.2.1.3)	96023	5	235
87	MV-82-16	Dihydrolipoamide acetyltransferase component of	43683	5	231
88	MV-60-1	hypothetical protein	9572	5	229
89	MV-8-6	HflC protein	32584	5	226
90	MV-63-34	Type I secretion outer membrane protein, TolC	49643	5	224
91	MV-25-22	RND family efflux transporter	111026	5	222
92	MV-112-10	Putative multidrug resistance protein	112662	5	218
93	MV-51-2	Rod shape-determining protein MreB	34371	5	215
94	MV-26-36	hypothetical protein	27610	5	215

Number	Accession number	Description	Mass [Da]	Peptides	Score
95	MV-42-204	transcriptional regulator	16763	5	213
96	MV-68-1	hypothetical protein	33940	5	210
97	MV-42-21	ATP synthase delta chain (EC 3.6.3.14)	19850	5	207
98	MV-111-2	Nicotinate-nucleotide--dimethylbenzimidazole	37587	5	204
99	MV-105-8	sensory box histidine kinase/response	47051	5	199
100	MV-98-88	Cell division trigger factor (EC 5.2.1.8)	53199	5	192
101	MV-43-10	SSU ribosomal protein S7p (S5e)	17682	5	177
102	MV-63-24	Sulfide-quinone reductase	41815	5	165
103	MV-84-120	SSU ribosomal protein S5p (S2e)	21396	4	294
104	MV-80-40	Twin-arginine translocation pathway signal	42934	4	291
105	MV-42-85	Surface lipoprotein	33428	4	267
106	MV-1-14	hypothetical protein	15572	4	251
107	MV-80-43	branched-chain amino acid ABC transporter,	28494	4	217
108	MV-77-5	Adenosylcobinamide-phosphate synthase	36327	4	214
109	MV-77-7	Cobalt-precorrin-8x methylmutase (EC 5.4.1.2)	22749	4	214
110	MV-25-20	putative methyl-accepting chemotaxis protein	75871	4	213
111	MV-111-1	Cobalamin synthase	27333	4	202
112	MV-76-7	hypothetical protein	134765	4	198
113	MV-98-109	ATP-dependent RNA helicase Atu1833	66090	4	196
114	MV-113-7	Protein-export membrane protein SecF (TC	34293	4	193
115	MV-98-45	SSU ribosomal protein S2p (SAe)	28150	4	190
116	MV-35-11	Cobalt-zinc-cadmium resistance protein	33587	4	187
117	MV-84-123	Preprotein translocase secY subunit (TC	48517	4	186
118	MV-69-46	Acetyl-coenzyme A carboxyl transferase alpha	34574	4	177
119	MV-82-62	hypothetical protein	40826	4	176
120	MV-77-8	Cobalt-precorrin-6y C5-methyltransferase (EC	43198	4	175
121	MV-42-148	60 kDa inner membrane insertion protein	65766	4	173
122	MV-84-117	SSU ribosomal protein S8p (S15Ae)	14658	4	172
123	MV-96-19	hypothetical protein	36797	4	171
124	MV-35-10	hypothetical protein	9364	4	170
125	MV-49-9	RND efflux system, inner membrane transporter	40017	4	170
126	MV-39-77	3-hydroxybutyryl-CoA dehydrogenase (EC	31595	4	170
127	MV-42-96	hypothetical protein	31227	4	169
128	MV-88-2	hypothetical protein	15104	4	161
129	MV-71-5	Cytochrome c-type biogenesis protein DsbD,	78137	4	161
130	MV-26-41	acriflavin resistance protein	118680	4	158
131	MV-48-10	Lead, cadmium, zinc and mercury transporting	65170	4	157
132	MV-50-4	AsmA protein	124776	4	155
133	MV-22-54	LSU ribosomal protein L25p	22587	4	153
134	MV-84-115	LSU ribosomal protein L5p (L11e)	20605	4	152
135	MV-98-23	Citrate synthase (si) (EC 2.3.3.1)	18020	4	152
136	MV-22-64	Ubiquinol--cytochrome c reductase, cytochrome B	47463	4	152
137	MV-98-75	NADH-ubiquinone oxidoreductase chain D (EC	44960	4	150
138	MV-42-108	COG0582: Integrase	45130	4	136
139	MV-7-2	SSU ribosomal protein S9p (S16e)	17496	3	215
140	MV-84-113	LSU ribosomal protein L14p (L23e)	13470	3	209
141	MV-43-4	LSU ribosomal protein L1p (L10Ae)	24588	3	194

Number	Accession number	Description	Mass [Da]	Peptides	Score
142	MV-36-12	hypothetical protein	25896	3	189
143	MV-37-1	DNA-binding protein HU-beta, NS1 (HU-1), plays	7020	3	188
144	MV-77-91	LSU ribosomal protein L35p	7610	3	180
145	MV-80-93	hypothetical protein	18963	3	176
146	MV-80-77	Cytochrome c oxidase subunit CcoO (EC 1.9.3.1)	26568	3	172
147	MV-117-56	ATP synthase B chain (EC 3.6.3.14)	21080	3	172
148	MV-105-9	CheY-like receiver	38581	3	171
149	MV-21-3	Outer membrane protein	47069	3	168
150	MV-109-85	COG3143: Chemotaxis protein	21397	3	162
151	MV-84-80	Transcriptional regulator, GntR family	24829	3	159
152	MV-87-13	Methylmalonyl-CoA mutase (EC 5.4.99.2)	77700	3	159
153	MV-77-25	TRAP-type C4-dicarboxylate transport system,	25279	3	158
154	MV-84-129	LSU ribosomal protein L17p	15685	3	157
155	MV-56-27	Molybdopterin-guanine dinucleotide biosynthesis	23528	3	156
156	MV-65-103	Polysaccharide biosynthesis protein CapD	68972	3	155
157	MV-34-10	D-3-phosphoglycerate dehydrogenase (EC	55822	3	154
158	MV-98-71	NADH-ubiquinone oxidoreductase chain H (EC	37970	3	154
159	MV-113-5	Preprotein translocase subunit YajC (TC	15379	3	151
160	MV-98-22	Citrate synthase (si) (EC 2.3.3.1)	32939	3	146
161	MV-16-5	N-acetylneuraminate synthase (EC 2.5.1.56)	32647	3	145
162	MV-110-18	Heat shock protein 60 family chaperone GroEL	58380	3	143
163	MV-22-111	COG1463: ABC-type transport system involved in	33425	3	141
164	MV-98-72	NADH-ubiquinone oxidoreductase chain G (EC	74416	3	140
165	MV-12-1	cd1 nitrite reductase	46078	3	137
166	MV-82-44	GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp	81873	3	136
167	MV-1-5	hypothetical protein	20928	3	135
168	MV-30-5	hypothetical protein	23737	3	134
169	MV-84-118	LSU ribosomal protein L6p (L9e)	19060	3	133
170	MV-84-116	SSU ribosomal protein S14p (S29e) ##	11605	3	133
171	MV-7-52	amino acid ABC transporter, periplasmic amino	36289	3	132
172	MV-42-220	Phosphoenolpyruvate-protein phosphotransferase	63841	3	131
173	MV-94-13	hypothetical protein	67445	3	128
174	MV-117-4	efflux transporter, RND family, MFP subunit	41256	3	123
175	MV-8-5	HflK protein	24486	3	123
176	MV-1-8	hypothetical protein	12311	3	122
177	MV-56-81	COG3258: Cytochrome c	28701	3	115
178	MV-98-44	Translation elongation factor Ts	31580	3	114
179	MV-43-11	Translation elongation factor G	76605	3	114
180	MV-63-2	Alanine dehydrogenase (EC 1.4.1.1)	39130	3	112
181	MV-35-9	COG1704: Uncharacterized conserved protein	25466	3	109
182	MV-65-135	hypothetical protein	17366	3	107
183	MV-96-17	cytochrome c4	20540	3	104
184	MV-96-21	Ferredoxin	28401	3	101
185	MV-42-32	Succinyl-CoA ligase [ADP-forming] beta chain	41424	3	101
186	MV-96-23	Nitrous oxide reductase maturation protein NosF	30770	3	97
187	MV-63-42	Adenylate cyclase (EC 4.6.1.1)	84162	3	95
188	MV-49-10	Membrane fusion protein of RND family multidrug	42175	3	95

Number	Accession number	Description	Mass [Da]	Peptides	Score
189	MV-32-35	hypothetical protein	39405	3	93
190	MV-33-4	Uncharacterized monothiol glutaredoxin	12566	3	93
191	MV-77-83	Translation initiation factor 3	16583	3	89
192	MV-111-3	COG0784: FOG: CheY-like receiver	35846	3	89
193	MV-39-20	hypothetical protein	42715	2	149
194	MV-84-102	LSU ribosomal protein L3p (L3e)	19616	2	147
195	MV-113-65	Type II/IV secretion system secretin RcpA/CpaC,	52191	2	147
196	MV-39-140	Predicted exporter of the RND superfamily	88496	2	133
197	MV-35-8	COG0457: FOG: TPR repeat	24282	2	130
198	MV-39-61	sensor protein fixL(EC:2.7.3.-)	92150	2	130
199	MV-117-52	DNA uptake lipoprotein	30827	2	128
200	MV-22-14	LSU ribosomal protein L21p	15998	2	127
201	MV-98-3	Peptidyl-prolyl cis-trans isomerase ppiD (EC	67515	2	127
202	MV-45-12	Malate dehydrogenase (EC 1.1.1.37)	16857	2	123
203	MV-1-9	hypothetical protein	14519	2	120
204	MV-65-72	protein of unknown function UPF0005	26825	2	120
205	MV-42-19	ATP synthase gamma chain (EC 3.6.3.14)	32526	2	117
206	MV-113-56	Flp pilus assembly protein, pilin Flp	6401	2	116
207	MV-7-27	polyketide synthase of type I	85630	2	115
208	MV-93-32	Transporter, AcrB/D/F family	112214	2	113
209	MV-38-27	Integration host factor alpha subunit	11129	2	112
210	MV-31-4	probable spore coat polysaccharide biosynthesis	61400	2	112
211	MV-42-40	Succinate dehydrogenase iron-sulfur protein (EC	30173	2	109
212	MV-69-10	Ferredoxin	41827	2	109
213	MV-39-138	hypothetical protein	20036	2	108
214	MV-94-6	hypothetical protein	43985	2	108
215	MV-39-29	Acetyl-CoA acetyltransferase (EC 2.3.1.9)	40444	2	106
216	MV-31-2	hypothetical protein	65204	2	106
217	MV-77-4	CobN component of cobalt chelatase involved in	137154	2	106
218	MV-80-7	2-keto-3-deoxy-D-arabino-heptulosonate-7-	50910	2	106
219	MV-22-89	AhpC/TSA family protein	16852	2	104
220	MV-82-15	Pyruvate dehydrogenase E1 component beta	49124	2	102
221	MV-42-223	hypothetical protein	22680	2	102
222	MV-22-63	ubiquinol cytochrome C oxidoreductase,	31765	2	102
223	MV-109-2	hypothetical protein	50032	2	101
224	MV-117-33	Cell division protein MraZ	18083	2	101
225	MV-98-55	Lipoprotein releasing system transmembrane	45307	2	100
226	MV-83-5	amino acid carrier protein	40223	2	99
227	MV-39-114	Cytochrome c heme lyase subunit CcmH	44470	2	99
228	MV-42-41	Succinate dehydrogenase flavoprotein subunit	65281	2	95
229	MV-84-107	SSU ribosomal protein S19p (S15e)	10308	2	95
230	MV-98-120	Polyferredoxin	49587	2	94
231	MV-38-88	D-alanyl-D-alanine carboxypeptidase (EC	38789	2	92
232	MV-79-48	Nucleoside diphosphate kinase (EC 2.7.4.6)	15225	2	92
233	MV-35-5	Arsenical pump membrane protein	47083	2	91
234	MV-42-97	Adenylylsulfate reductase beta-subunit (EC	15790	2	91
235	MV-76-8	type II restriction enzyme	136404	2	91

Number	Accession number	Description	Mass [Da]	Peptides	Score
236	MV-22-53	Ribose-phosphate pyrophosphokinase (EC	34360	2	91
237	MV-77-6	Sirohydrochlorin cobaltochelatase (EC	37400	2	90
238	MV-38-12	Serine hydroxymethyltransferase (EC 2.1.2.1)	45935	2	90
239	MV-50-9	Fumarate hydratase class I, aerobic (EC	59608	2	89
240	MV-65-137	hypothetical protein	20124	2	87
241	MV-84-101	SSU ribosomal protein S10p (S20e)	11622	2	87
242	MV-93-6	DNA topoisomerase I (EC 5.99.1.2)	98938	2	85
243	MV-117-57	ATP synthase B' chain (EC 3.6.3.14)	19826	2	85
244	MV-84-99	COG0566: rRNA methylases	34126	2	85
245	MV-109-62	SSU ribosomal protein S18p	8727	2	84
246	MV-69-66	LSU ribosomal protein L28p	10637	2	84
247	MV-98-86	ATP-dependent Clp protease ATP-binding subunit	46309	2	84
248	MV-96-22	Polyferredoxin NapH (periplasmic nitrate	36705	2	83
249	MV-84-103	LSU ribosomal protein L3p (L3e)	8105	2	83
250	MV-69-45	Protein export cytoplasm protein SecA ATPase	100943	2	83
251	MV-1-13	Magnetosome protein MamF	12032	2	82
252	MV-4-10	Ketol-acid reductoisomerase (EC 1.1.1.86)	37622	2	82
253	MV-116-2	GTP-binding protein EngA	52344	2	81
254	MV-42-208	hypothetical protein	14781	2	81
255	MV-16-6	hypothetical protein	51598	2	80
256	MV-7-17	putative ABC transporter ATP-binding protein	40465	2	80
257	MV-77-80	Glycosyltransferase	44103	2	80
258	MV-36-3	Ferrochelatase, protoheme ferro-lyase (EC	39738	2	80
259	MV-36-40	Chaperone protein DnaK	68706	2	79
260	MV-82-53	hypothetical protein	30779	2	79
261	MV-39-15	Pyruvate-flavodoxin oxidoreductase (EC	131575	2	78
262	MV-109-1	hypothetical protein	18226	2	78
263	MV-35-1	hypothetical protein	7449	2	77
264	MV-42-87	ABC transporter ATP-binding protein	62787	2	77
265	MV-22-104	Cytochrome c oxidase polypeptide II (EC	26230	2	76
266	MV-80-70	COG0318: Acyl-CoA synthetases	64515	2	76
267	MV-16-3	Cobalamin B12-binding:Radical SAM	56612	2	76
268	MV-39-10	Uncharacterized protein SCO1/SenC/PrrC	33714	2	75
269	MV-77-14	Cytochrome oxidase biogenesis protein	21431	2	75
270	MV-33-9	Phosphoribosylaminoimidazole-succinocarboxamide	29115	2	75
271	MV-65-98	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	71879	2	75
272	MV-7-50	ABC-type amino acid transport system, permease	38392	2	74
273	MV-98-21	Glutamyl-tRNA synthetase (EC 6.1.1.17)	51918	2	74
274	MV-21-4	Probable ATP-binding/permease fusion ABC	44965	2	74
275	MV-113-25	Adenylate cyclase (EC 4.6.1.1)	24656	2	73
276	MV-22-25	Carboxyl-terminal protease (EC 3.4.21.102)	49994	2	72
277	MV-90-14	Predicted transcriptional regulator of cysteine	17361	2	71
278	MV-109-76	hypothetical protein	10627	2	71
279	MV-110-22	Dolichol-phosphate mannosyltransferase	28242	2	71
280	MV-42-207	Putative heme iron utilization protein	28510	2	70
281	MV-39-51	1-acyl-sn-glycerol-3-phosphate acyltransferase	29127	2	70
282	MV-42-16	hypothetical protein	15026	2	69

Number	Accession number	Description	Mass [Da]	Peptides	Score
283	**MV-96-16**	**Nitrous-oxide reductase (EC 1.7.99.6)**	**85520**	**2**	**67**
284	**MV-53-22**	**hypothetical protein**	**92623**	**2**	**66**
285	**MV-111-16**	**Multimodular transpeptidase-transglycosylase**	**70008**	**2**	**65**
286	**MV-7-19**	**band 7 protein**	**21443**	**2**	**65**
287	**MV-116-37**	**Permease of the major facilitator superfamily**	**44595**	**2**	**65**
288	**MV-42-53**	**LSU ribosomal protein L19p**	**17035**	**2**	**64**
289	**MV-109-17**	**Cell division protein FtsH (EC 3.4.24.-)**	**53561**	**2**	**64**
290	**MV-116-3**	**Outer membrane protein YfgL, lipoprotein**	**47106**	**2**	**64**
291	**MV-7-109**	**hypothetical protein**	**39547**	**2**	**63**
292	**MV-39-52**	**Protein of unknown function DUF218**	**23009**	**2**	**63**
293	**MV-112-5**	**Phosphatidylserine decarboxylase (EC 4.1.1.65)**	**25143**	**2**	**63**
294	**MV-22-15**	**LSU ribosomal protein L27p**	**8914**	**2**	**62**
295	**MV-98-66**	**NADH-ubiquinone oxidoreductase chain M (EC**	**55959**	**2**	**62**
296	**MV-107-1**	**COG4818: Predicted membrane protein**	**11631**	**2**	**61**
297	**MV-39-11**	**Cytochrome c oxidase (B(O/a)3-type) chain I (EC**	**60978**	**2**	**60**
298	**MV-69-68**	**Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)**	**45726**	**2**	**60**
299	**MV-7-51**	**ABC-type amino acid transport system, permease**	**43664**	**2**	**60**
300	**MV-104-1**	**hypothetical protein**	**36281**	**2**	**59**
301	**MV-26-6**	**Heme A synthase, cytochrome oxidase biogenesis**	**40104**	**2**	**59**
302	**MV-104-4**	**FkbM family methyltransferase**	**31937**	**2**	**58**
303	**MV-111-11**	**NAD-dependent epimerase/dehydratase family**	**35776**	**2**	**57**
304	**MV-42-17**	**ATP synthase epsilon chain (EC 3.6.3.14)**	**14858**	**2**	**57**
305	**MV-30-2**	**hypothetical protein**	**20262**	**2**	**57**
306	**MV-98-77**	**NADH-ubiquinone oxidoreductase chain B (EC**	**21005**	**2**	**57**
307	**MV-26-4**	**methyl-accepting chemotaxis sensory transducer**	**41247**	**2**	**54**
308	**MV-38-19**	**hypothetical protein**	**18245**	**2**	**54**
309	**MV-82-9**	**signal transduction histidine kinase containing**	**81089**	**2**	**53**
310	**MV-38-33**	**hypothetical protein**	**15365**	**2**	**52**
311	MV-80-42	possible branched-chain amino acid ABC	41688	1	102
312	MV-104-2	ABC transporter, multidrug efflux family	35782	1	98
313	MV-32-27	Iron-sulfur cluster regulator IscR	17263	1	96
314	MV-39-28	PhbF	22044	1	94
315	MV-116-52	NAD(P) transhydrogenase alpha subunit (EC	39658	1	90
316	MV-39-126	hypothetical protein	6436	1	82
317	MV-109-100	probable transport system ATP-binding protein	39946	1	78
318	MV-7-18	putative ABC transporter (ATP-binding protein)	61940	1	77
319	MV-109-59	LSU ribosomal protein L9p	13797	1	75
320	MV-65-88	Homolog of E. coli HemY protein	48789	1	73
321	MV-107-2	Integration host factor, alpha subunit	7749	1	72
322	MV-93-2	LSU ribosomal protein L33p	6392	1	72
323	MV-48-11	Lead, cadmium, zinc and mercury transporting	20346	1	71
324	MV-98-65	NADH-ubiquinone oxidoreductase chain N (EC	51707	1	68
325	MV-39-41	Inosine-5'-monophosphate dehydrogenase (EC	16326	1	67
326	MV-22-100	hypothetical protein	51448	1	66
327	MV-109-77	Flagellar biosynthesis protein fliP	29196	1	65
328	MV-71-51	serine phosphatase	95339	1	64
329	MV-49-33	hypothetical protein	15388	1	64

Number	Accession number	Description	Mass [Da]	Peptides	Score
330	MV-63-9	Twin-arginine translocation protein TatC	34686	1	61
331	MV-53-13	hypothetical protein	37442	1	61
332	MV-57-10	Predicted permease	42399	1	61
333	MV-116-53	NAD(P) transhydrogenase alpha subunit (EC	14394	1	60
334	MV-56-49	hypothetical protein	9447	1	59
335	MV-39-116	Cytochrome c-type biogenesis protein CcmG/DsbE,	20347	1	58
336	MV-87-9	Chemotaxis protein methyltransferase CheR (EC	30727	1	57
337	MV-42-118	Acyl-CoA synthetases (AMP-forming)/AMP-acid	22945	1	57
338	MV-109-86	Chemotaxis regulator - transmits chemoreceptor	14522	1	57
339	MV-69-62	peptidase M48, Ste24p	35488	1	57
340	MV-109-49	Flagellar motor rotation protein MotB	39736	1	57
341	MV-34-33	TRAP-type C4-dicarboxylate transport system,	25692	1	56
342	MV-31-8	Probable poly(beta-D-mannuronate) O-acetylase	53544	1	56
343	MV-84-128	DNA-directed RNA polymerase alpha subunit (EC	15890	1	56
344	MV-42-58	COG1396: Predicted transcriptional regulators	16204	1	56
345	MV-98-73	NADH-ubiquinone oxidoreductase chain F (EC	48279	1	56
346	MV-39-115	Cytochrome c heme lyase subunit CcmL	19397	1	55
347	MV-80-47	periplasmic sensor signal transduction	102028	1	55
348	MV-39-9	Cytochrome b subunit of the bc complex	27811	1	55
349	MV-109-11	Transcriptional regulator, GntR family /	52883	1	55
350	MV-63-11	Twin-arginine translocation protein TatA	8378	1	54
351	MV-98-100	OmpA/MotB	38855	1	54
352	MV-38-34	Response regulator consisting of a CheY-like	20631	1	54
353	MV-42-228	Ribosomal protein S6 glutaminyl transferase	32623	1	53
354	MV-98-67	NADH-ubiquinone oxidoreductase chain L (EC	71685	1	53
355	MV-42-31	Succinyl-CoA ligase [ADP-forming] alpha chain	30003	1	53
356	MV-98-70	NADH-ubiquinone oxidoreductase chain I (EC	19406	1	53
357	MV-42-254	Integration host factor beta subunit	10475	1	53
358	MV-79-43	Phosphoribosylformylglycinamidine cyclo-ligase	38724	1	52
359	MV-55-2	two component transcriptional regulator, winged	39636	1	51
360	MV-1-4	Actin-like ATPase	39127	1	51
361	MV-7-47	hypothetical protein	38195	1	51
362	MV-36-5	Uroporphyrinogen III decarboxylase (EC	34842	1	51
363	MV-39-12	Cytochrome c oxidase (B(O/a)3-type) chain II	20900	1	50
364	MV-98-76	NADH-ubiquinone oxidoreductase chain C (EC	24421	1	50
365	MV-43-9	SSU ribosomal protein S12p (S23e)	7844	1	50
366	MV-3-30	Chemotaxis response regulator protein-glutamate	38809	1	50
367	MV-93-26	COG0582: Integrase	47883	1	50
368	MV-43-3	LSU ribosomal protein L11p (L12e)	14943	1	49
369	MV-109-89	hypothetical protein	28211	1	49
370	MV-26-50	Enoyl-[acyl-carrier-protein] reductase [NADH]	29297	1	49
371	MV-4-9	Acetolactate synthase small subunit (EC	19814	1	49
372	MV-42-8	Aspartate aminotransferase (EC 2.6.1.1)	43592	1	49
373	MV-3-17	Flagellar motor rotation protein MotA	28744	1	49
374	MV-39-129	ATP-DEPENDENT DNA HELICASE	97946	1	48
375	MV-84-92	Adenylate cyclase (EC 4.6.1.1)	63576	1	47
376	MV-65-56	hypothetical protein	35111	1	47

Number	Accession number	Description	Mass [Da]	Peptides	Score
377	MV-35-15	TPR domain protein	97613	1	47
378	MV-7-85	hypothetical protein	7641	1	47
379	MV-57-9	COG0795: Predicted permeases	40645	1	47
380	MV-113-22	rhodanese domain protein	16242	1	46
381	MV-109-82	Flagellar motor rotation protein MotB	25783	1	46
382	MV-80-44	hydrophobic amino acid uptake transporter	28821	1	45
383	MV-73-3	methyltransferase FkbM family	32542	1	45
384	MV-18-11	RNA polymerase sigma factor RpoD	77876	1	44
385	MV-65-45	ATP-dependent protease La (EC 3.4.21.53) Type	90996	1	44
386	MV-83-1	Ferrous iron transport protein B	67343	1	44
387	MV-98-119	Periplasmic protein p19 involved in	19329	1	44
388	MV-75-4	Sulfate adenylyltransferase subunit 1 (EC	22282	1	44
389	MV-42-250	hypothetical protein	19030	1	44
390	MV-96-78	COG0840: Methyl-accepting chemotaxis protein	57406	1	43
391	MV-98-121	High-affinity Fe2+/Pb2+ permease	43854	1	43
392	MV-38-37	hypothetical protein	30765	1	43
393	MV-93-9	Acyl-phosphate:glycerol-3-phosphate	21507	1	43
394	MV-109-48	MotA/TolQ/ExbB proton channel family protein,	43544	1	43
395	MV-116-8	Sel1 domain protein repeat-containing protein	27534	1	43
396	MV-36-2	Hypothetical membrane protein, possible	17795	1	43
397	MV-21-15	Ferrous iron transport protein B	8169	1	43
398	MV-36-6	conserved hypothetical protein	30609	1	42
399	MV-65-70	ABC transporter component	25772	1	42
400	MV-84-138	Signal transduction histidine kinase	28101	1	42
401	MV-42-73	COG2199: FOG: GGDEF domain	39580	1	41
402	MV-110-108	transcriptional regulator, HTH family	17010	1	41
403	MV-65-41	YGGT family protein	12609	1	41
404	MV-80-79	Heavy-metal-associated domain (N-terminus) and	27016	1	41
405	MV-42-122	ABC-type multidrug transport system	28446	1	41
406	MV-34-32	TRAP-type C4-dicarboxylate transport system,	22871	1	41
407	MV-109-18	Cell division protein FtsH (EC 3.4.24.-)	15964	1	41
408	MV-98-84	DNA-binding protein HU-beta	9227	1	40
409	MV-109-63	SSU ribosomal protein S6p	17379	1	40
410	MV-65-77	Potassium efflux system KefA protein /	92531	1	40
411	MV-87-22	Rhodanese-related sulfurtransferase	16432	1	40
412	MV-75-13	COG2716: Glycine cleavage system regulatory	19131	1	40
413	MV-65-8	oxidoreductase	37043	1	40
414	MV-33-13	GMP synthase [glutamine-hydrolyzing] (EC	57969	1	40
415	MV-83-4	hypothetical protein	6212	1	40
416	MV-22-99	Heme O synthase, protoheme IX	34789	1	40
417	MV-42-109	Signal transduction histidine kinase	82157	1	39
418	MV-113-60	Flp pilus assembly protein TadB	35090	1	39
419	MV-113-3	hypothetical protein	13188	1	39
420	MV-26-45	Cobalt-zinc-cadmium resistance protein	34159	1	39
421	MV-42-36	sterol desaturase-related protein	31954	1	39
422	MV-56-85	Predicted transporter component	38682	1	39
423	MV-73-11	Lipid A export ATP-binding/permease protein	65204	1	38

Number	Accession number	Description	Mass [Da]	Peptides	Score
424	MV-26-16	Sulfite reduction-associated complex DsrMKJOP	27468	1	38
425	MV-63-3	glutathione-regulated potassium-efflux system	58703	1	38
426	MV-110-37	putative signal-transduction protein with CBS	18067	1	38
427	MV-77-24	TRAP-type C4-dicarboxylate transport system,	56884	1	38
428	MV-109-81	hypothetical protein	123055	1	37
429	MV-96-74	ClpB protein	95483	1	37
430	MV-42-39	probable tungsten-containing aldehyde	61359	1	37
431	MV-49-35	Hypothetical protein YaeJ with similarity to	16162	1	37
432	MV-98-128	ABC-type transport system involved in	17317	1	36
433	MV-94-3	UDP-N-acetylglucosamine 4,6-dehydratase (EC	38428	1	36
434	MV-77-3	CobW GTPase involved in cobalt insertion for	38161	1	36
435	MV-116-4	hypothetical protein	23793	1	36
436	MV-63-16	cyclin-dependent kinase inhibitor 1C (p57,	32931	1	36
437	MV-94-11	N-acetylneuraminate synthase (EC 2.5.1.56)	39622	1	36
438	MV-87-12	hypothetical protein	17554	1	36
439	MV-109-91	Flagellar biosynthesis protein fliL	21505	1	36
440	MV-84-3	Nitrite-sensitive transcriptional repressor	17077	1	36
441	MV-116-44	Cation transport ATPase(EC:3.6.3.8)	99400	1	36
442	MV-38-41	Nitrogen regulation protein ntrY (EC 2.7.13.3)	84378	1	35
443	MV-56-82	Sulfur oxidation protein SoxZ	11819	1	35
444	MV-80-92	glycosyl transferase, family 2	95740	1	35
445	MV-36-56	COG0524: Sugar kinases, ribokinase family	36603	1	35
446	MV-34-3	N-terminal domain of molybdenum-binding	11741	1	35
447	MV-113-64	Type IV pili component	22836	1	35
448	MV-42-45	COG2301: Citrate lyase beta subunit	31954	1	35
449	MV-45-32	hypothetical protein	33010	1	34
450	MV-43-5	LSU ribosomal protein L10p (P0)	17670	1	34
451	MV-26-43	hypothetical protein	7424	1	34
452	MV-84-74	Tetrapyrrole (Corrin-Porphyrin) methylase	31223	1	34
453	MV-34-27	Acetate operon repressor	30262	1	34
454	MV-38-6	Positive regulator of CheA protein activity	21455	1	34
455	MV-22-48	Prolipoprotein diacylglyceryl transferase (EC	29520	1	34
456	MV-98-68	NADH-ubiquinone oxidoreductase chain K (EC	10939	1	33
457	MV-69-8	hypothetical protein	31714	1	33
458	MV-36-10	Protein export cytoplasm chaperone protein	19924	1	33
459	MV-109-52	hypothetical protein	32969	1	33
460	MV-96-49	Adenylosuccinate lyase (EC 4.3.2.2)	52161	1	33
461	MV-79-17	Protein acetyltransferase	98299	1	33
462	MV-65-80	FOG: GGDEF domain	41320	1	32
463	MV-116-75	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	69514	1	32
464	MV-77-16	Poly-beta-hydroxyalkanoate depolymerase	48459	1	32
465	MV-116-69	2-Hydroxychromene-2-carboxylate isomerase	22967	1	32
466	MV-42-56	SSU ribosomal protein S16p	16118	1	31
467	MV-45-36	Pyruvate,phosphate dikinase (EC 2.7.9.1)	95809	1	31
468	MV-42-230	Adenosylhomocysteinase (EC 3.3.1.1)	51755	1	31
469	MV-100-1	PAS/PAC sensor hybrid histidine kinase	68464	1	30
470	MV-1-3	hypothetical protein	16677	1	30

Number	Accession number	Description	Mass [Da]	Peptides	Score
471	MV-77-95	COG2202: FOG: PAS/PAC domain	19997	1	30
472	MV-58-26	RsbR, positive regulator of sigma-B	31839	1	29
473	MV-96-8	Acetyltransferase	21324	1	28
474	MV-42-224	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	22381	1	28
475	MV-65-118	2-octaprenyl-6-methoxyphenol hydroxylase	46564	1	28
476	MV-109-40	NADPH-dependent glyceraldehyde-3-phosphate	36279	1	27
477	MV-31-3	putative oxidoreductase	40558	1	27
478	MV-56-52	Sulfide dehydrogenase [flavocytochrome c]	45906	1	27
479	MV-42-126	Methyl-accepting chemotaxis protein	59011	1	26
480	MV-63-71	COG2206: HD-GYP domain	81034	1	26
481	MV-63-48	Membrane-fusion protein; Multidrug resistance	48487	1	26
482	MV-65-63	hypothetical protein	50128	1	25
483	MV-98-101	Polyhydroxyalkanoic acid synthase	70652	1	25
484	MV-96-62	Aspartokinase (EC 2.7.2.4)	43669	1	25
485	MV-77-98	COG5001: Predicted signal transduction protein	91732	1	25
486	MV-109-60	LSU ribosomal protein L9p	9159	1	25

Appendix II

Proteins identified using Orbitrap instrument in the magnetosome membrane fraction prepared by Tanaka's method.

Number	Accession number	Description	Mass [Da]	Peptides	Score
1	MV-42-33	Malate dehydrogenase (EC 1.1.1.37)	38732	30	1883
2	MV-43-8	DNA-directed RNA polymerase beta' subunit (EC	155896	21	906
3	MV-42-29	Dihydrolipoamide succinyltransferase component	44274	18	1103
4	MV-77-9	Cobalamin biosynthesis protein CbiG /	91161	18	918
5	MV-42-28	Dihydrolipoamide dehydrogenase of	48014	17	1193
6	MV-43-7	DNA-directed RNA polymerase beta subunit (EC	154088	17	628
7	MV-79-53	Cytosol aminopeptidase PepA (EC 3.4.11.1)	52892	16	1011
8	MV-3-32	Signal transduction histidine kinase CheA (EC	103432	14	738
9	MV-84-100	Translation elongation factor Tu	43088	14	675
10	MV-7-45	6-phosphofructokinase (EC 2.7.1.11)	43962	13	691
11	MV-112-41	OmpA family protein	30024	12	539
12	MV-65-111	NAD-specific glutamate dehydrogenase (EC	181583	12	494
13	MV-65-134	Porin, Gram-negative type	38577	11	928
14	MV-35-4	Cobalt-zinc-cadmium resistance protein	30736	11	637
15	MV-35-6	Serine protease precursor MucD/AlgY associated	74272	11	573
16	MV-1-34	Propionyl-CoA carboxylase biotin-containing	80412	11	504
17	MV-7-22	hypothetical protein	18414	10	607
18	MV-8-5	HflK protein	24486	10	411
19	MV-42-30	2-oxoglutarate dehydrogenase E1 component (EC	111107	10	389
20	MV-39-127	Chaperone protein HtpG	69053	10	376
21	MV-35-2	serine protease	73406	9	515
22	MV-77-8	Cobalt-precorrin-6y C5-methyltransferase (EC	43198	9	472
23	MV-76-7	hypothetical protein	134765	9	389
24	MV-43-11	Translation elongation factor G	76605	9	388
25	MV-80-75	Cytochrome c oxidase subunit CcoP (EC 1.9.3.1)	31863	9	361
26	MV-56-26	COG5283: Phage-related tail protein	27936	8	584
27	MV-38-51	Methyl-accepting chemotaxis protein	59277	8	577
28	MV-8-6	HflC protein	32584	8	491
29	MV-65-63	hypothetical protein	50128	8	429
30	MV-32-25	Cysteine desulfurase (EC 2.8.1.7)	45960	8	422
31	MV-26-51	Chorismate synthase (EC 4.2.3.5) # AroG	38692	8	397
32	MV-117-11	hypothetical protein	38456	8	275
33	MV-98-145	COG0840: Methyl-accepting chemotaxis protein	16861	7	581
34	MV-60-2	hypothetical protein	10576	7	446
35	MV-25-20	putative methyl-accepting chemotaxis protein	75871	7	418
36	MV-1-11	COG0206: Cell division GTPase	19676	7	411
37	MV-96-78	COG0840: Methyl-accepting chemotaxis protein	57406	7	392
38	MV-42-205	Phosphoenolpyruvate carboxykinase [GTP] (EC	67241	7	387
39	MV-35-8	COG0457: FOG: TPR repeat	24282	7	366
40	MV-77-100	(S)-2-hydroxy-acid oxidase	142895	7	319
41	MV-98-3	Peptidyl-prolyl cis-trans isomerase ppiD (EC	67515	7	317
42	MV-42-98	Adenylylsulfate reductase alpha-subunit (EC	70715	7	316
43	MV-110-18	Heat shock protein 60 family chaperone GroEL	58380	7	312
44	MV-98-45	SSU ribosomal protein S2p (SAe)	28150	7	287
45	MV-42-20	ATP synthase alpha chain (EC 3.6.3.14)	54905	7	281
46	MV-16-1	hypothetical protein	34559	7	280
47	MV-63-71	COG2206: HD-GYP domain	81034	7	266

Number	Accession number	Description	Mass [Da]	Peptides	Score
48	MV-1-14	hypothetical protein	15572	6	427
49	MV-31-4	probable spore coat polysaccharide biosynthesis	61400	6	353
50	MV-110-19	Heat shock protein 60 family co-chaperone	11053	6	324
51	MV-98-89	YjeF protein, function unknown	54514	6	293
52	MV-1-7	Permease of the major facilitator superfamily	47996	6	266
53	MV-111-3	COG0784: FOG: CheY-like receiver	35846	6	257
54	MV-98-23	Citrate synthase (si) (EC 2.3.3.1)	18020	6	244
55	MV-98-101	Polyhydroxyalkanoic acid synthase	70652	6	244
56	MV-113-6	Protein-export membrane protein SecD (TC	55977	6	239
57	MV-56-58	Transketolase (EC 2.2.1.1)	71051	6	232
58	MV-51-2	Rod shape-determining protein MreB	34371	6	231
59	MV-84-131	Trypsin-like serine protease, typically	50113	6	228
60	MV-42-87	ABC transporter ATP-binding protein	62787	6	225
61	MV-65-104	IMP cyclohydrolase (EC 3.5.4.10) /	56163	6	224
62	MV-63-24	Sulfide-quinone reductase	41815	5	337
63	MV-1-9	hypothetical protein	14519	5	326
64	MV-35-11	Cobalt-zinc-cadmium resistance protein	33587	5	310
65	MV-87-13	Methylmalonyl-CoA mutase (EC 5.4.99.2)	77700	5	296
66	MV-98-22	Citrate synthase (si) (EC 2.3.3.1)	32939	5	293
67	MV-109-78	putative molecular chaperone small heat shock	18261	5	275
68	MV-80-77	Cytochrome c oxidase subunit CcoO (EC 1.9.3.1)	26568	5	274
69	MV-65-8	oxidoreductase	37043	5	264
70	MV-88-2	hypothetical protein	15104	5	257
71	MV-36-12	hypothetical protein	25896	5	255
72	MV-42-18	ATP synthase beta chain (EC 3.6.3.14)	50800	5	253
73	MV-77-6	Sirohydrochlorin cobaltochelatase (EC	37400	5	251
74	MV-4-8	Acetolactate synthase large subunit (EC	64701	5	250
75	MV-35-3	Actin-like ATPase	39870	5	239
76	MV-82-62	hypothetical protein	40826	5	230
77	MV-65-4	ATP-dependent RNA helicase	78469	5	230
78	MV-42-57	Signal recognition particle, subunit Ffh SRP54	50486	5	227
79	MV-69-45	Protein export cytoplasm protein SecA ATPase	100943	5	208
80	MV-42-256	SSU ribosomal protein S1p	61616	5	195
81	MV-111-2	Nicotinate-nucleotide--dimethylbenzimidazole	37587	5	193
82	MV-104-1	hypothetical protein	36281	5	190
83	MV-8-3	ATPase involved in chromosome partitioning	17992	5	185
84	MV-104-3	transport protein	32458	5	177
85	MV-26-36	hypothetical protein	27610	5	171
86	MV-90-1	DNA gyrase subunit B (EC 5.99.1.3)	90662	5	171
87	MV-35-14	Methyl-accepting chemotaxis protein	34372	4	335
88	MV-60-1	hypothetical protein	9572	4	291
89	MV-50-4	AsmA protein	124776	4	258
90	MV-65-133	Porin 41 (Por41) precursor	38025	4	248
91	MV-16-5	N-acetylneuraminate synthase (EC 2.5.1.56)	32647	4	240
92	MV-35-10	hypothetical protein	9364	4	217
93	MV-33-9	Phosphoribosylaminoimidazole-succinocarboxamide	29115	4	214
94	MV-98-143	ATP-dependent Clp protease ATP-binding subunit	87630	4	214

Number	Accession number	Description	Mass [Da]	Peptides	Score
95	MV-1-32	Propionyl-CoA carboxylase carboxyl transferase	56263	4	213
96	MV-42-41	Succinate dehydrogenase flavoprotein subunit	65281	4	212
97	MV-1-13	Magnetosome protein MamF	12032	4	209
98	MV-98-127	methyltransferase, putative	25236	4	208
99	MV-96-17	cytochrome c4	20540	4	203
100	MV-116-52	NAD(P) transhydrogenase alpha subunit (EC	39658	4	192
101	MV-65-98	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	71879	4	191
102	MV-45-7	NAD kinase (EC 2.7.1.23)	28449	4	188
103	MV-116-54	NAD(P) transhydrogenase subunit beta (EC	47910	4	186
104	MV-26-4	methyl-accepting chemotaxis sensory transducer	41247	4	182
105	MV-69-46	Acetyl-coenzyme A carboxyl transferase alpha	34574	4	181
106	MV-65-101	hypothetical protein	57523	4	173
107	MV-98-109	ATP-dependent RNA helicase Atu1833	66090	4	166
108	MV-30-26	hypothetical protein	64631	4	166
109	MV-82-14	Pyruvate dehydrogenase E1 component alpha	39094	4	165
110	MV-42-106	transcriptional regulator, Fis family	105437	4	164
111	MV-96-74	ClpB protein	95483	4	163
112	MV-80-94	Oxidoreductase (flavoprotein)	42371	4	160
113	MV-116-35	Alanyl-tRNA synthetase (EC 6.1.1.7)	94267	4	158
114	MV-3-4	hypothetical protein	38514	4	152
115	MV-98-128	ABC-type transport system involved in	17317	4	152
116	MV-63-47	ABC-type protease/lipase transport system,	86035	4	150
117	MV-42-39	probable tungsten-containing aldehyde	61359	4	142
118	MV-42-204	transcriptional regulator	16763	4	138
119	MV-42-128	TPR repeat protein	53162	4	131
120	MV-117-61	Chromosome partition protein smc	127242	4	119
121	MV-109-17	Cell division protein FtsH (EC 3.4.24.-)	53561	4	116
122	MV-98-21	Glutamyl-tRNA synthetase (EC 6.1.1.17)	51918	3	198
123	MV-34-10	D-3-phosphoglycerate dehydrogenase (EC	55822	3	193
124	MV-110-13	Putrescine ABC transporter putrescine-binding	40239	3	186
125	MV-65-88	Homolog of E. coli HemY protein	48789	3	180
126	MV-37-1	DNA-binding protein HU-beta, NS1 (HU-1), plays	7020	3	174
127	MV-82-18	dihydrolipoamide dehydrogenase(EC:1.8.1.4)	30905	3	173
128	MV-110-43	Fructose-bisphosphate aldolase class I (EC	35932	3	166
129	MV-117-57	ATP synthase B' chain (EC 3.6.3.14)	19826	3	163
130	MV-26-44	Signal transduction histidine kinase	129548	3	158
131	MV-42-136	hypothetical protein	31297	3	153
132	MV-42-21	ATP synthase delta chain (EC 3.6.3.14)	19850	3	153
133	MV-63-46	Adenylate cyclase (EC 4.6.1.1)	41260	3	151
134	MV-65-43	Methyl-accepting chemotaxis protein	44064	3	150
135	MV-77-5	Adenosylcobinamide-phosphate synthase	36327	3	148
136	MV-45-37	Glycyl-tRNA synthetase beta chain (EC	78288	3	146
137	MV-42-246	Acetyl-coenzyme A carboxyl transferase beta	35707	3	145
138	MV-94-13	hypothetical protein	67445	3	143
139	MV-18-11	RNA polymerase sigma factor RpoD	77876	3	141
140	MV-116-75	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	69514	3	138
141	MV-69-10	Ferredoxin	41827	3	135

Number	Accession number	Description	Mass [Da]	Peptides	Score
142	MV-77-80	Glycosyltransferase	44103	3	132
143	MV-39-68	Ethylmalonyl-CoA mutase,	73059	3	132
144	MV-33-20	Inosine-5'-monophosphate dehydrogenase (EC	51681	3	131
145	MV-3-21	Flagellar synthesis regulator FleN	28727	3	128
146	MV-83-5	amino acid carrier protein	40223	3	124
147	MV-49-8	RND efflux system, inner membrane transporter	72460	3	123
148	MV-42-52	3-isopropylmalate dehydratase large subunit (EC	50393	3	123
149	MV-42-161	COG0840: Methyl-accepting chemotaxis protein	53631	3	122
150	MV-63-42	Adenylate cyclase (EC 4.6.1.1)	84162	3	120
151	MV-63-44	NA	45707	3	120
152	MV-82-17	dihydrolipoamide dehydrogenase	19344	3	119
153	MV-42-137	hypothetical protein	44908	3	117
154	MV-96-19	hypothetical protein	36797	3	116
155	MV-6-1	hypothetical protein	44942	3	114
156	MV-1-8	hypothetical protein	12311	3	113
157	MV-39-124	Aconitate hydratase (EC 4.2.1.3)	96023	3	111
158	MV-16-3	Cobalamin B12-binding:Radical SAM	56612	3	111
159	MV-96-23	Nitrous oxide reductase maturation protein NosF	30770	3	111
160	MV-63-11	Twin-arginine translocation protein TatA	8378	3	109
161	MV-1-5	hypothetical protein	20928	3	109
162	MV-56-6	non-motile and phage-resistance protein	109142	3	105
163	MV-94-3	UDP-N-acetylglucosamine 4,6-dehydratase (EC	38428	3	103
164	MV-22-89	AhpC/TSA family protein	16852	3	103
165	MV-109-48	MotA/TolQ/ExbB proton channel family protein,	43544	3	103
166	MV-39-8	Cytochrome b subunit of the bc complex	32215	3	102
167	MV-76-10	ATP-dependent Lon-type protease	75755	3	101
168	MV-84-104	LSU ribosomal protein L4p (L1e)	22242	3	101
169	MV-49-10	Membrane fusion protein of RND family multidrug	42175	3	93
170	MV-80-74	Type cbb3 cytochrome oxidase biogenesis protein	55153	3	91
171	MV-117-12	hypothetical protein	21910	3	90
172	MV-49-46	Multimodular transpeptidase-transglycosylase	91901	3	90
173	MV-98-75	NADH-ubiquinone oxidoreductase chain D (EC	44960	3	88
174	MV-43-5	LSU ribosomal protein L10p (P0)	17670	2	176
175	MV-59-1	hypothetical protein	5888	2	166
176	MV-109-45	Inositol-1-monophosphatase (EC 3.1.3.25)	28756	2	152
177	MV-3-1	Flagellin protein FlaA	40038	2	139
178	MV-98-121	High-affinity Fe2+/Pb2+ permease	43854	2	133
179	MV-35-7	hypothetical protein	33831	2	132
180	MV-50-9	Fumarate hydratase class I, aerobic (EC	59608	2	130
181	MV-80-13	diguanylate cyclase/phosphodiesterase with	87471	2	122
182	MV-94-6	hypothetical protein	43985	2	121
183	MV-4-2	Tryptophan synthase beta chain like (EC	49492	2	121
184	MV-36-40	Chaperone protein DnaK	68706	2	120
185	MV-45-6	Molybdenum cofactor biosynthesis protein A	36954	2	119
186	MV-42-82	5-methyltetrahydrofolate--homocysteine	134485	2	119
187	MV-35-1	hypothetical protein	7449	2	118
188	MV-113-5	Preprotein translocase subunit YajC (TC	15379	2	116

Number	Accession number	Description	Mass [Da]	Peptides	Score
189	MV-54-2	hypothetical protein	60561	2	116
190	MV-63-48	Membrane-fusion protein; Multidrug resistance	48487	2	115
191	MV-109-98	Ferric iron ABC transporter, iron-binding	38920	2	115
192	MV-117-5	acriflavin resistance protein	149841	2	110
193	MV-1-3	hypothetical protein	16677	2	108
194	MV-8-8	HtrA protease/chaperone protein	52134	2	107
195	MV-113-7	Protein-export membrane protein SecF (TC	34293	2	106
196	MV-63-10	Twin-arginine translocation protein TatB	15996	2	106
197	MV-68-1	hypothetical protein	33940	2	104
198	MV-45-12	Malate dehydrogenase (EC 1.1.1.37)	16857	2	103
199	MV-109-11	Transcriptional regulator, GntR family /	52883	2	102
200	MV-26-50	Enoyl-[acyl-carrier-protein] reductase [NADH]	29297	2	102
201	MV-3-30	Chemotaxis response regulator protein-glutamate	38809	2	102
202	MV-39-11	Cytochrome c oxidase (B(O/a)3-type) chain I (EC	60978	2	101
203	MV-96-4	Similar to CDP-glucose 4,6-dehydratase (EC	37930	2	100
204	MV-39-51	1-acyl-sn-glycerol-3-phosphate acyltransferase	29127	2	100
205	MV-42-121	Acetylornithine aminotransferase (EC 2.6.1.11)	42166	2	100
206	MV-109-18	Cell division protein FtsH (EC 3.4.24.-)	15964	2	99
207	MV-77-14	Cytochrome oxidase biogenesis protein	21431	2	97
208	MV-8-2	MRP-like protein (ATP/GTP-binding protein)	21901	2	97
209	MV-39-12	Cytochrome c oxidase (B(O/a)3-type) chain II	20900	2	93
210	MV-84-123	Preprotein translocase secY subunit (TC	48517	2	93
211	MV-80-50	Putative stomatin/prohibitin-family membrane	35046	2	93
212	MV-116-16	Amidophosphoribosyltransferase (EC 2.4.2.14)	53531	2	93
213	MV-65-6	Transcriptional regulator	13945	2	92
214	MV-98-67	NADH-ubiquinone oxidoreductase chain L (EC	71685	2	91
215	MV-109-2	hypothetical protein	50032	2	90
216	MV-65-45	ATP-dependent protease La (EC 3.4.21.53) Type	90996	2	90
217	MV-80-43	branched-chain amino acid ABC transporter,	28494	2	90
218	MV-31-7	Glycosyltransferase	41437	2	90
219	MV-117-56	ATP synthase B chain (EC 3.6.3.14)	21080	2	89
220	MV-36-54	membrane protein, putative	28578	2	89
221	MV-65-103	Polysaccharide biosynthesis protein CapD	68972	2	89
222	MV-98-95	Pyridoxal-5'-phosphate-dependent enzyme, beta	35421	2	89
223	MV-22-63	ubiquinol cytochrome C oxidoreductase,	31765	2	87
224	MV-84-126	SSU ribosomal protein S11p (S14e)	14109	2	86
225	MV-63-16	cyclin-dependent kinase inhibitor 1C (p57,	32931	2	86
226	MV-98-72	NADH-ubiquinone oxidoreductase chain G (EC	74416	2	85
227	MV-105-9	CheY-like receiver	38581	2	85
228	MV-49-44	Cytoplasmic axial filament protein CafA and	97996	2	85
229	MV-22-49	COG1565: Uncharacterized conserved protein	39603	2	85
230	MV-35-9	COG1704: Uncharacterized conserved protein	25466	2	84
231	MV-80-110	Tetraacyldisaccharide 4'-kinase (EC 2.7.1.130)	37248	2	84
232	MV-45-36	Pyruvate,phosphate dikinase (EC 2.7.9.1)	95809	2	83
233	MV-63-45	Adenylate cyclase (EC 4.6.1.1)	42118	2	82
234	MV-22-99	Heme O synthase, protoheme IX	34789	2	82
235	MV-109-67	sensory box histidine kinase/response	99778	2	81

Number	Accession number	Description	Mass [Da]	Peptides	Score
236	MV-42-96	hypothetical protein	31227	2	80
237	MV-90-11	Ubiquinone biosynthesis monooxygenase UbiB	37262	2	80
238	MV-31-2	hypothetical protein	65204	2	78
239	MV-82-44	GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp	81873	2	78
240	MV-38-20	Pyrophosphate-energized proton pump (EC	68887	2	78
241	MV-34-16	TPR repeat-containing protein	51373	2	78
242	MV-69-6	hypothetical protein	48881	2	77
243	MV-116-30	RecA protein	39386	2	77
244	MV-80-7	2-keto-3-deoxy-D-arabino-heptulosonate-7-	50910	2	77
245	MV-80-78	Cytochrome c oxidase subunit CcoN (EC 1.9.3.1)	55653	2	77
246	MV-38-88	D-alanyl-D-alanine carboxypeptidase (EC	38789	2	77
247	MV-25-21	Probable Co/Zn/Cd efflux system membrane fusion	33783	2	73
248	MV-42-140	Pantoate--beta-alanine ligase (EC 6.3.2.1)	31244	2	73
249	MV-69-49	hypothetical protein	57601	2	73
250	MV-96-8	Acetyltransferase	21324	2	72
251	MV-76-8	type II restriction enzyme	136404	2	71
252	MV-53-14	Methyltransferase type 12	26754	2	71
253	MV-42-231	PAS/PAC Sensor Signal Transduction Histidine	67384	2	70
254	MV-73-11	Lipid A export ATP-binding/permease protein	65204	2	69
255	MV-16-4	hypothetical protein	45084	2	68
256	MV-109-3	hypothetical protein	18918	2	66
257	MV-107-1	COG4818: Predicted membrane protein	11631	2	64
258	MV-94-9	hypothetical protein	27263	2	64
259	MV-42-51	3-isopropylmalate dehydratase small subunit (EC	22834	2	64
260	MV-82-10	UDP-galactose-lipid carrier transferase (EC	57255	2	64
261	MV-56-59	Fructose-bisphosphate aldolase class II (EC	38377	2	63
262	MV-96-63	Enoyl-[acyl-carrier-protein] reductase [FMN]	37006	2	63
263	MV-38-25	Phosphate:acyl-ACP acyltransferase PlsX	37680	2	63
264	MV-34-19	hypothetical protein	13545	2	63
265	MV-42-8	Aspartate aminotransferase (EC 2.6.1.1)	43592	2	62
266	MV-69-68	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)	45726	2	61
267	MV-39-114	Cytochrome c heme lyase subunit CcmH	44470	2	61
268	MV-98-147	UDP-glucose dehydrogenase (EC 1.1.1.22)	47359	2	61
269	MV-84-116	SSU ribosomal protein S14p (S29e) ##	11605	2	61
270	MV-39-76	Electron transfer flavoprotein, alpha subunit	31907	2	61
271	MV-79-51	hypothetical protein	11618	2	60
272	MV-84-125	SSU ribosomal protein S13p (S18e)	13757	2	59
273	MV-42-3	Translation elongation factor G-related	28547	2	59
274	MV-110-112	Alanine dehydrogenase (EC 1.4.1.1)	38689	2	59
275	MV-65-145	Thymidylate kinase (EC 2.7.4.9)	76311	2	58
276	MV-7-51	ABC-type amino acid transport system, permease	43664	2	57
277	MV-109-22	OmpA/MotB	17362	2	56
278	MV-77-82	Threonyl-tRNA synthetase (EC 6.1.1.3)	72583	2	56
279	MV-110-52	Glycogen phosphorylase (EC 2.4.1.1)	75789	2	56
280	MV-39-92	Electron transfer flavoprotein-ubiquinone	60222	2	56
281	MV-56-87	hypothetical protein	18270	2	55
282	MV-42-32	Succinyl-CoA ligase [ADP-forming] beta chain	41424	2	52

Number	Accession number	Description	Mass [Da]	Peptides	Score
283	**MV-82-41**	**hypothetical protein**	**25725**	**2**	**52**
284	MV-22-100	hypothetical protein	51448	1	98
285	MV-116-53	NAD(P) transhydrogenase alpha subunit (EC	14394	1	92
286	MV-22-78	Methyl-accepting chemotaxis protein	47399	1	85
287	MV-42-19	ATP synthase gamma chain (EC 3.6.3.14)	32526	1	85
288	MV-83-4	hypothetical protein	6212	1	80
289	MV-36-57	GTP-binding protein TypA/BipA	66662	1	80
290	MV-38-40	Nitrogen assimilation regulatory protein	53131	1	79
291	MV-80-93	hypothetical protein	18963	1	78
292	MV-113-28	chemotaxis protein CheYIII	21125	1	74
293	MV-22-46	FIG018229: hypothetical protein	9535	1	73
294	MV-109-23	tolB protein precursor, periplasmic protein	50260	1	72
295	MV-35-15	TPR domain protein	97613	1	72
296	MV-98-73	NADH-ubiquinone oxidoreductase chain F (EC	48279	1	72
297	MV-34-3	N-terminal domain of molybdenum-binding	11741	1	70
298	MV-22-25	Carboxyl-terminal protease (EC 3.4.21.102)	49994	1	69
299	MV-38-41	Nitrogen regulation protein ntrY (EC 2.7.13.3)	84378	1	68
300	MV-39-61	sensor protein fixL(EC:2.7.3.-)	92150	1	67
301	MV-35-13	hypothetical protein	19545	1	67
302	MV-12-1	cd1 nitrite reductase	46078	1	67
303	MV-82-28	Integral membrane protein CcmA involved in cell	20245	1	66
304	MV-7-18	putative ABC transporter (ATP-binding protein)	61940	1	65
305	MV-111-16	Multimodular transpeptidase-transglycosylase	70008	1	65
306	MV-109-95	Flagellar L-ring protein FlgH	27616	1	64
307	MV-36-3	Ferrochelatase, protoheme ferro-lyase (EC	39738	1	63
308	MV-33-14	hypothetical protein	20578	1	63
309	MV-21-3	Outer membrane protein	47069	1	62
310	MV-3-13	Flagellar M-ring protein fliF	60276	1	62
311	MV-39-116	Cytochrome c-type biogenesis protein CcmG/DsbE,	20347	1	61
312	MV-56-85	Predicted transporter component	38682	1	61
313	MV-98-118	Biotin carboxylase of acetyl-CoA carboxylase	49481	1	59
314	MV-32-24	Iron-sulfur cluster assembly scaffold protein	14378	1	58
315	MV-32-28	Serine acetyltransferase (EC 2.3.1.30)	27054	1	58
316	MV-58-26	RsbR, positive regulator of sigma-B	31839	1	58
317	MV-30-5	hypothetical protein	23737	1	58
318	MV-109-26	Biopolymer transport protein ExbD/TolR	17068	1	57
319	MV-116-64	hypothetical protein	31706	1	57
320	MV-39-117	Cytochrome c heme lyase subunit CcmF	72035	1	57
321	MV-116-8	Sel1 domain protein repeat-containing protein	27534	1	57
322	MV-39-40	SAM-dependent methyltransferase 2, in cluster	27902	1	56
323	MV-43-10	SSU ribosomal protein S7p (S5e)	17682	1	56
324	MV-1-4	Actin-like ATPase	39127	1	55
325	MV-22-111	COG1463: ABC-type transport system involved in	33425	1	54
326	MV-36-1	Transcription termination factor Rho	46800	1	53
327	MV-82-15	Pyruvate dehydrogenase E1 component beta	49124	1	53
328	MV-96-61	3-demethylubiquinone-9 3-methyltransferase (EC	28057	1	52
329	MV-113-27	Excinuclease ABC subunit A	104857	1	52

Number	Accession number	Description	Mass [Da]	Peptides	Score
330	MV-45-32	hypothetical protein	33010	1	52
331	MV-80-48	hypothetical protein	13460	1	51
332	MV-116-4	hypothetical protein	23793	1	51
333	MV-16-6	hypothetical protein	51598	1	51
334	MV-56-77	thioredoxin SoxW	20995	1	50
335	MV-105-2	Phosphonate ABC transporter phosphate-binding	17860	1	49
336	MV-112-11	efflux transporter, RND family, MFP subunit	40167	1	49
337	MV-96-40	ATP-binding protein PhnN; Guanylate kinase (EC	22192	1	49
338	MV-25-22	RND family efflux transporter	111026	1	49
339	MV-39-77	3-hydroxybutyryl-CoA dehydrogenase (EC	31595	1	48
340	MV-109-42	diguanylate cyclase with PAS/PAC sensor	20150	1	48
341	MV-56-44	PUTATIVE SENSOR HISTIDINE KINASE TRANSMEMBRANE	70683	1	47
342	MV-36-20	ATP-dependent hsl protease ATP-binding subunit	48037	1	47
343	MV-7-5	Enoyl-CoA hydratase (EC 4.2.1.17)	29592	1	46
344	MV-35-5	Arsenical pump membrane protein	47083	1	45
345	MV-84-120	SSU ribosomal protein S5p (S2e)	21396	1	45
346	MV-36-39	Chaperone protein DnaJ	41824	1	45
347	MV-26-43	hypothetical protein	7424	1	45
348	MV-112-26	Lipid carrier :	41931	1	45
349	MV-56-81	COG3258: Cytochrome c	28701	1	44
350	MV-112-5	Phosphatidylserine decarboxylase (EC 4.1.1.65)	25143	1	44
351	MV-116-11	hypothetical protein	16861	1	44
352	MV-42-190	protein of unknown function DUF533	21227	1	44
353	MV-65-126	DNA-binding response regulator	26447	1	44
354	MV-98-120	Polyferredoxin	49587	1	43
355	MV-82-16	Dihydrolipoamide acetyltransferase component of	43683	1	43
356	MV-69-67	CheY-like receiver	27507	1	43
357	MV-77-15	Cobyrinic acid A,C-diamide synthase	47925	1	43
358	MV-53-10	GDP-fucose synthetase	35433	1	43
359	MV-42-150	Acetylglutamate kinase (EC 2.7.2.8)	32891	1	42
360	MV-113-33	Signal transduction histidine kinase	83965	1	42
361	MV-84-127	DNA-directed RNA polymerase alpha subunit (EC	21974	1	42
362	MV-42-166	response regulator	54304	1	42
363	MV-7-49	ABC-type polar amino acid transport system,	29086	1	42
364	MV-26-41	acriflavin resistance protein	118680	1	42
365	MV-113-56	Flp pilus assembly protein, pilin Flp	6401	1	42
366	MV-65-14	ABC-type (unclassified) transport system,	29028	1	41
367	MV-110-31	Potassium efflux system KefA protein /	48342	1	41
368	MV-42-11	Excinuclease ABC subunit B	79715	1	41
369	MV-56-33	hypothetical protein	24208	1	41
370	MV-42-153	ATP-dependent RNA helicase RhlE	50632	1	41
371	MV-22-68	hypothetical protein	13138	1	41
372	MV-65-135	hypothetical protein	17366	1	41
373	MV-42-163	hypothetical protein	43499	1	40
374	MV-113-65	Type II/IV secretion system secretin RcpA/CpaC,	52191	1	40
375	MV-45-43	hypothetical protein	69956	1	40
376	MV-42-207	Putative heme iron utilization protein	28510	1	40

Number	Accession number	Description	Mass [Da]	Peptides	Score
377	MV-77-60	TRAP transporter solute receptor, TAXI family	46950	1	40
378	MV-26-37	UspA	31214	1	40
379	MV-96-62	Aspartokinase (EC 2.7.2.4)	43669	1	39
380	MV-80-112	Protein of unknown function DUF374	25659	1	39
381	MV-42-85	Surface lipoprotein	33428	1	39
382	MV-111-1	Cobalamin synthase	27333	1	38
383	MV-8-4	HflK protein	16305	1	38
384	MV-82-53	hypothetical protein	30779	1	38
385	MV-116-10	sulfide-quinone reductase	46540	1	38
386	MV-39-122	ABC transporter involved in cytochrome c	23850	1	38
387	MV-65-90	hypothetical protein	39577	1	38
388	MV-56-46	Phosphoribulokinase (EC 2.7.1.19)	32973	1	38
389	MV-77-77	Nucleoside-diphosphate-sugar epimerases	33117	1	38
390	MV-39-35	hypothetical protein	73124	1	37
391	MV-77-57	hypothetical protein	45343	1	37
392	MV-39-57	Diaminopimelate decarboxylase (EC 4.1.1.20)	45597	1	37
393	MV-112-15	PAS/PAC sensor hybrid histidine kinase	54502	1	37
394	MV-79-33	heat shock protein Hsp20	10718	1	37
395	MV-79-45	Signal transduction histidine kinase	42596	1	37
396	MV-96-24	hypothetical protein	16777	1	37
397	MV-98-32	Outer membrane protein H precursor	21463	1	37
398	MV-65-77	Potassium efflux system KefA protein /	92531	1	36
399	MV-65-62	Argininosuccinate synthase (EC 6.3.4.5)	45188	1	36
400	MV-77-3	CobW GTPase involved in cobalt insertion for	38161	1	36
401	MV-94-10	Pseudaminic acid cytidylyltransferase (EC	56476	1	36
402	MV-112-21	hypothetical protein	76947	1	36
403	MV-42-40	Succinate dehydrogenase iron-sulfur protein (EC	30173	1	35
404	MV-96-2	Glucose-1-phosphate cytidylyltransferase (EC	28884	1	35
405	MV-33-4	Uncharacterized monothiol glutaredoxin	12566	1	35
406	MV-42-151	hypothetical protein	25642	1	35
407	MV-42-148	60 kDa inner membrane insertion protein	65766	1	34
408	MV-39-16	putative cyclase/kinase	61662	1	34
409	MV-84-80	Transcriptional regulator, GntR family	24829	1	34
410	MV-65-36	Translation elongation factor LepA	66634	1	34
411	MV-116-18	DNA repair protein RadA	48642	1	34
412	MV-39-118	Cytochrome c-type biogenesis protein CcmE, heme	16653	1	34
413	MV-116-15	oxidoreductase, short-chain	26996	1	34
414	MV-38-98	hypothetical protein	46857	1	34
415	MV-39-70	isobutyryl-CoA dehydrogenase	60157	1	33
416	MV-42-230	Adenosylhomocysteinase (EC 3.3.1.1)	51755	1	33
417	REV_MV-22-21	LSU m3Psi1915 methyltransferase RlmH	16571	1	33
418	MV-42-58	COG1396: Predicted transcriptional regulators	16204	1	33
419	MV-38-12	Serine hydroxymethyltransferase (EC 2.1.2.1)	45935	1	33
420	MV-71-5	Cytochrome c-type biogenesis protein DsbD,	78137	1	33
421	MV-7-67	ETC complex I subunit conserved region	11114	1	33
422	MV-26-31	universal stress protein	30298	1	33
423	MV-22-85	peptidase C14, caspase catalytic subunit p20	63063	1	33

Number	Accession number	Description	Mass [Da]	Peptides	Score
424	MV-39-10	Uncharacterized protein SCO1/SenC/PrrC	33714	1	32
425	MV-22-103	Cytochrome c oxidase polypeptide I (EC	57332	1	32
426	MV-42-245	Dihydrofolate synthase (EC 6.3.2.12) /	47601	1	32
427	MV-38-19	hypothetical protein	18245	1	32
428	MV-42-126	Methyl-accepting chemotaxis protein	59011	1	31
429	MV-117-13	hypothetical protein	58060	1	31
430	MV-25-42	Signal transduction histidine kinase	46146	1	31
431	MV-110-22	Dolichol-phosphate mannosyltransferase	28242	1	31
432	MV-39-93	COG0457: FOG: TPR repeat	64219	1	31
433	MV-96-5	putative methyltransferase	44433	1	31
434	MV-28-2	filamentation induced by cAMP protein Fic	47995	1	31
435	MV-94-12	hypothetical protein	61950	1	31
436	MV-39-135	peptidase M48, Ste24p	54056	1	31
437	MV-75-13	COG2716: Glycine cleavage system regulatory	19131	1	31
438	MV-56-32	heterodisulfide reductase	46997	1	30
439	MV-80-44	hydrophobic amino acid uptake transporter	28821	1	30
440	MV-84-3	Nitrite-sensitive transcriptional repressor	17077	1	30
441	MV-32-27	Iron-sulfur cluster regulator IscR	17263	1	29
442	MV-77-25	TRAP-type C4-dicarboxylate transport system,	25279	1	29
443	MV-98-86	ATP-dependent Clp protease ATP-binding subunit	46309	1	29
444	MV-73-5	hypothetical protein	48625	1	29
445	MV-30-35	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	56333	1	29
446	MV-42-97	Adenylylsulfate reductase beta-subunit (EC	15790	1	29
447	MV-110-17	glutamine synthetase family protein	51293	1	29
448	MV-69-9	cAMP-binding proteins - catabolite gene	17857	1	29
449	MV-79-43	Phosphoribosylformylglycinamidine cyclo-ligase	38724	1	29
450	MV-98-77	NADH-ubiquinone oxidoreductase chain B (EC	21005	1	28
451	MV-83-1	Ferrous iron transport protein B	67343	1	28
452	MV-3-18	two-component flagellar transcriptional	48958	1	28
453	MV-49-9	RND efflux system, inner membrane transporter	40017	1	28
454	MV-98-85	ATP-dependent protease La (EC 3.4.21.53) Type	89811	1	27
455	MV-84-92	Adenylate cyclase (EC 4.6.1.1)	63576	1	27
456	MV-80-82	Riboflavin kinase (EC 2.7.1.26) / FMN	35064	1	27
457	MV-80-53	Sulfate adenylyltransferase, dissimilatory-type	43357	1	27
458	MV-117-33	Cell division protein MraZ	18083	1	27
459	REV_MV-42-47	response regulator receiver domain protein	-1	1	26
460	MV-36-10	Protein export cytoplasm chaperone protein	19924	1	26
461	MV-22-77	Cytochrome c-type protein NapC	21378	1	26
462	MV-31-3	putative oxidoreductase	40558	1	25
463	MV-42-141	3-methyl-2-oxobutanoate	31215	1	25