Title
Distinct Roles for Two Trypsin-like Proteases in Magnetosome Biomineralization

Author
Hershey, David Michael

Publication Date
2016

Peer reviewed|Thesis/dissertation

Distinct Roles for Two Trypsin-like Proteases in Magnetosome Biomineralization

By

David Michael Hershey

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Arash Komeili, Chair
Professor James H. Hurley
Professor Michiko E. Taga
Professor Andreas Martin

Spring 2016

Abstract

Distinct Roles for Two Trypsin-like Proteases in Magnetosome Biomineralization

by

David Michael Hershey

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Arash Komeili, Chair

The ability of living organisms to transform inorganic elements into insoluble crystalline structures is an underexplored theme in biology. A group of aquatic bacteria, called magnetotactic bacteria, produces chains of nanometer-sized magnetic crystals within their cells, allowing them to align in the earth's geomagnetic fields. Understanding the mechanism for this process has become increasingly important as the demand for customized nanoparticles in medical and industrial applications has grown. In particular, the protein factors required to transform soluble iron into magnetite (Fe_3O_4) are likely to contain novel mechanisms for manipulating insoluble inorganic compounds.

This thesis describes the biochemical and genetic features of two such factors, MamO and MamE. A historical account is provided describing the discovery of magnetotactic bacteria, the development of *Magnetospirillum magneticum* AMB-1 as a model system and the identification of specific genes required to produce magnetite. This previous body of work led to the identification of MamO and MamE as predicted trypsin-like proteases that are required for biomineralization in AMB-1. Genetic, biochemical and structural studies reported here showed that the MamO protease domain is catalytically inactive and incapable of serine protease activity. Instead it has a novel di-histidine motif that participates in direct binding to transition metals. This motif is required for biomineralization *in vivo*, confirming that the MamO protein is a repurposed trypsin-like scaffold that promotes magnetite nucleation by binding directly to iron precursor atoms. Genomic and phylogenetic analysis of related serine proteases in other magnetotactic bacteria showed that the repurposing of trypsin-like proteases has occurred numerous times independently during the evolution of magnetosome formation. Also described is the observation that three biomineralization factors, MamE, MamO and MamP are proteolytically processed in AMB-1. MamE and MamO are both required for these proteolytic events, as are the predicted catalytic residues from MamE. However, consistent with its newly assigned pseudo-protease classification, the predicted MamO active site is dispensable. This suggested that MamE directly processes these targets in a manner that requires MamO. The proteolytic activity of MamE was reconstituted in vitro with a recombinant form of the protein. MamE cleaved a custom peptide substrate based on an *in vivo* cleavage site in MamO with positive cooperativity, and its auto-proteolytic activity could be stimulated by both substrates and peptides that bind to its regulatory domains. These enzymatic properties suggested that a switch-like regulatory mechanism modulated MamE-dependent proteolysis during biomineralization.

This regulatory paradigm was confirmed by showing that both catalytically inactive and constitutively active alleles of *mamE* caused severe biomineralization defects *in vivo.*

Although the genes required for biomineralization were known previously, the molecular mechanisms by which each protein promotes magnetite synthesis had not been explored. The results of these studies define biochemical functions for two of the four factors required for magnetite nucleation in AMB-1. Furthermore, describing the evolutionary repurposing of a trypsin scaffold along with the phylogenetic description of its evolutionary history add broad biological interest to magnetosome research.

CHAPTER 1

Magnetite biomineralization in the genus *Magnetospirillum*

David M. Hershey

ABSTRACT
Biomineralization is the widespread process by which organisms synthesize inorganic crystals. Biomineralization in organisms called magnetotactic bacteria leads to the intracellular production of magnetic nanoparticles called magnetosomes that are organized in chains, allowing the cells to align in magnetic fields. After briefly introducing the concept of biomineralization, this chapter provides a historical account of the discovery of magnetotactic bacteria and early work leading to the identification of a genomic island containing the genetic factors for magnetosome formation. Focusing on the model organism *Magnetospirillum magneticum* AMB-1, the genetic and cellular basis for magnetosome formation in the *Magnetospirillum* genus is reviewed. The current model for magnetosome formation assigns individual genes roles in defined cellular transformations, specifically membrane remodeling, formation of magnetite crystals, maturation of minerals or organization of the magnetosome chain. These genes are discussed individually, emphasizing mechanisms by which each is thought to promote magnetite biosynthesis. Finally, a new direction in magnetosome research in which known genetic factors are assigned biochemical activities is proposed.

MAGNETITE BIOMINERALIZATION IN MAGNETOTACTIC BACTERIA

The phenomenon of biomineralization

Biomineralization is the process by which living organisms create inorganic, crystalline materials with specific biological functions (1). Although the phenomenon has been underexplored at the molecular level, it is nonetheless a common occurrence throughout the tree of life (2). Some of the most notable examples are carried out by multicellular eukaryotes including tooth and bone formation in mammals and the assembly of shells in mollusks (3, 4). Single celled eukaryotes such as coccolithophores and diatoms are also encased in intricate shells that are produced through biomineralization (5, 6). Finally, a number of species in the bacterial kingdom are now known to produce biominerals (7). The growing interest in custom material production through nanosynthesis has inspired a renewed effort to understand the mechanisms behind these processes.

Central to the problem of producing minerals is acquiring the appropriate raw materials. Chemical forms of inorganic elements are often insoluble in water and must be dissolved before they are bioavailable for deposition into minerals (8). Mechanisms for inorganic precursor accumulation in biomineralization can provide fundamental insight into the diversity of strategies available to biological organisms for nutrient acquisition. Mineral producers can often selectively import bioavailable states of inorganic precursors without manipulating them beforehand (9, 10). The accumulation strategies used for biomineralization processes are thought to contribute significantly the global geochemical cycling of inorganic elements (11).

The mineral synthesis phase is thought to involve intimate interactions between biological factors and insoluble forms of inorganic atoms. Dissolved precursor compounds must be transformed into a state that precipitates while being combined with any other components of the mature lattice. Decreasing solubility by altering the redox state is a common strategy involved in the initial mineral formation stage (12-14). Controlling the deposition of new material during mineral growth defines the geometric features of the mature structure. Many minerals appear to

grow as amorphous precursors that become crystalline later in the process (2). In other cases, classical nucleation through super-saturation forms crystals with cellular factors controlling the pattern by limiting growth (15). The precisely defined shapes and sizes of many biominerals suggest that the cellular machinery displays a high degree of control over the crystal maturation process (16).

Understanding the molecular biology underlying biomineralization pathways has direct biotechnological relevance. Materials with functional features on the nanometer size scale, called nanomaterials, are increasingly desirable in a number of medical and industrial applications (17). Many of the current strategies for nanomaterial production require extreme conditions that are expensive and hard to control (18). The biomineralization machinery can produce minerals at ambient temperature and pressure, suggesting that a more elegant and efficient transformation mechanism exists (19). Furthermore, biological organisms control the growth of their associated minerals with a level of precision that cannot yet be replicated *in vitro.* Thus, in addition to understanding fundamental aspects of how living systems interface with inorganic and insoluble compounds, characterizing mechanisms of biomineralization can be used to improve the synthesis of nanomaterials.

Magnetotactic bacteria: discovery and scope

While observing marine sediment samples, Richard Blakemore made the first recognized discovery of magnetotactic bacteria. He noticed that when drops of the sediment were placed under a microscope, a number of bacteria consistently swam toward an edge of the drop corresponding to the magnetic north. He subsequently showed that the direction of migration could be manipulated with a bar magnet and that the migration was not due to a simple "pulling" effect but required swimming by bacteria that were passively aligned in the magnetic fields. Though he could not grow the bacteria in culture, he was able to enrich magnetic cells for electron microscopic studies and elemental analysis. He observed chains of electron dense crystals enriched in iron within the bacterial cells and proposed that these structures allowed the cells to orient in the magnetic field (20). This seminal report marked the beginning of an increasingly large effort to understand a variety of fundamental questions surrounding these so-called magnetotactic bacteria.

Magnetotactic bacteria are now recognized as a large group of aquatic organisms that produce chains of nanometer sized particles called magnetosomes in their cells (21). These particles consist of the magnetic minerals magnetite (Fe_3O_4) or greigite (Fe_3S_4), and they allow the cells to passively align in magnetic fields (22, 23). As aquatic organisms, magnetotactic bacteria live in environments in which redox active electron donors and acceptors used by various respiratory pathways are highly stratified (24). It is thought that each organism has preferred depth zones in the water column corresponding to redox conditions that are optimal for its growth (25). The magneto-aerotaxis model posits that magnetotactic bacteria produce magnetic particles to align and swim along the earth's geomagnetic fields simplifying their search for preferred redox zones (26).

The increasing interest in studying magnetotactic bacteria has lead to more efficient methods for their isolation. Techniques based largely on swimming behavior in artificial magnetic fields have led to the isolation of dozens of strains in pure culture (27). With the increasing number of

cultured representatives has come a broader appreciation for the phylogenetic distribution of magnetotactic bacteria. Isolates from the *Alphaproteobacteria, Gammaproteobacteria,* and *Deltaproteobacteria* orders have now been cultured and analyzed by genome sequencing (28-30). In addition, uncultured representatives from the *Nitrospira* and *Omnitrophica* phyla have been observed in environmental samples (31-33). Both the phylogenetic distribution and the physiological strategies employed by these bacteria are surprisingly broad.

The cellular, genetic and evolutionary basis for magnetosome formation

The earliest cultured magnetotactic bacterium was the *Alphaproteobacterium Magnetospirillum magnetotacticum* MS-1 (34). Ultra-structural studies with this strain established many of the early paradigms now recognized as the cellular basis of magnetosome formation. The cells are spiral-shaped and produce flagella at the cell poles for swimming (34, 35). Each cell encloses numerous electron dense particles rich in iron that are organized in a chain along the cell axis. These chains seem to be associated with the bacterial inner membrane, and a membrane like layer appears to surround each individual particle (35). Biochemical fractionation of the electron dense particles from cells subsequently confirmed that the surrounding material is indeed a lipid bilayer enriched in specific protein components (36). The "electron-dense particles with their enveloping layer(s)" in MS-1 were named magnetosomes, a term that now encompasses the diversity of magnetic nanoparticles produced in the cells of bacteria (35).

Though MS-1 was used in early investigations of the cellular basis for magnetosome formation, it was not until a pair of closely related *Alphaproteobacteria, Magnetospirillum magneticum* AMB-1 and *M. gryphiswaldense* MSR-1, were isolated that progress in understanding the molecular basis for magnetosome formation began in earnest (37, 38). A set of landmark studies with MSR-1 enabled the identification of a genomic region now known to encode the critical genetic factors for magnetosome formation. A class of spontaneous non-magnetic mutants common in MSR-1 cultures was shown to result from the loss of a specific 80kb genomic region of the genome (39). Among other features, this region contained two operons, called the *mamAB* and *mamDC* clusters, known to encode proteins that were identified in biochemical enrichments for magnetosome membrane associated proteins (40-42). Similar genomic islands were detected in *M. magneticum* AMB-1 and the related strain *Magnetococcus marinus* MC-1, indicating that the feature was conserved in a number of magnetotactic organisms (43, 44). This genomic feature, known as the Magnetosome Island (MAI), has been identified in all magnetotactic bacteria analyzed to date.

As the number and diversity of magnetotactic isolates has increased the time and cost associated with genome sequencing has decreased. MAI sequences have been reported for a diverse set of strains, leading to the identification of both conserved and clade specific genes (30). One of the more notable features associated with MAIs is that the genomic regions surrounding them contain repetitive motifs, insertions sequences or transposases suggesting they are mobile elements that are readily transferred horizontally (45-47). This idea was further supported by the rapid loss of the MAI during growth and the polyphyletic nature of the magnetosomes formation trait within the *Proteobacteria* (48). A seminal set of analyses comparing the topology of phylogenetic trees for housekeeping sequences to trees of concatenated core genes in the MAI showed no evidence for horizontal transfer of the core factors among magnetotactic isolates (49-51). These studies demonstrate that magnetosome formation is an ancient process that likely

predates the radiation of the *Proteobacteria*. Furthermore, there is no indication that any of the extant strains acquired the trait via horizontal transfer.

MOLECULAR GENETICS OF MAGNETOSOME BIOGENESIS

The model strain Magnetospirillum magneticum *AMB-1*
Magnetospirillum magneticum AMB-1 is a spirillum shaped magnetotactic bacterium from the *Rhodospirillales* order of *Alphaproteobacteria* that was isolated from freshwater sediments in Tokyo (37, 52). A few key physiological features have led to the widespread use of this strain, and its close relative *M. gryphiswaldense* MSR-1, as models for studying magnetosome formation. First, unlike previous isolates, AMB-1 is aerotolerant. It can grow in atmospheric oxygen concentrations and can form single colonies on agar plates, though magnetosome production is only observed at low oxygen concentrations. Additionally, cultures of AMB-1 reach higher cell densities than previous isolates (37). These features allow the strain to be grown in significantly high numbers for biochemical analysis and to be plated for single colonies during genetic manipulation.

Genetic analysis with AMB-1 began almost immediately. Early studies took advantage of the ability to introduce transposable elements by performing loss-of-function screens for magnetosome synthesis genes (53). In 2004, the *mamA* gene, encoding a tetratricopeptide repeat (TPR) protein known to be enriched in magnetosome extracts, was the first gene disrupted by targeted deletion in a magnetotactic bacterium (54). The Δ*mamA* strain produced fewer magnetite particles during recovery from iron starvation, now considered to be a modest biomineralization defect. In addition, a GFP fusion to MamA localized in a line along the cell length, demonstrating for the first time that specific proteins were targeted to the magnetosomes in live cells (54). More importantly, the molecular techniques used to delete *mamA* and characterize its localization pattern in AMB-1 inspired a set new approaches that produced significant gains in understanding the genetic and cell biological basis for making magnetosomes.

One of the earliest processes to be interrogated with these newly developed molecular techniques was the ultra-structural basis for the linear organization of the magnetosome chain. Electron cryotomography (ECT) is a technique that provides three-dimensional reconstructions of frozen samples using a series of transmission electron microscopic (TEM) images at taken different incident angles by tilting the stage of the microscope (55). An ECT reconstruction of a whole AMB-1 cell revealed a number of previously unrecognized features of magnetosomes. Consistent with biochemical fractionations, a membrane structure surrounded each magnetite crystal. However, these membranes are not autonomous but are invaginations of the cytoplasmic membrane. Furthermore, bundles of filament like structures appear to connect individual magnetosomes suggesting the presence of a cytoskeleton organizing the magnetosomes. By imaging a targeted deletion of the gene encoding the actin-like protein MamK, these filaments were shown be composed of MamK and to be required for organizing the magnetosomes into a chain (56).

Magnetosome formation is not essential for viability of the cells and the magnetic response of each culture can be measured quickly using a bar magnet (39, 47). When combined with the now

well-established techniques for creating gene deletions, biomineralization in AMB-1 is one of the more elegant genetic models available for studying the molecular basis of a biological phenomenon. Genes can be quickly deleted, and plasmids harboring customized alleles of the deleted gene can be assessed for complementation by quantifying the magnetic response. The system is ideal for understanding the biochemical basis for biomineralization because insights from studying protein function *in vitro* can be directly tested for *in vivo* relevance with complementation attempts using site-directed mutants. After nearly four decades since the initial discovery of magnetotactic bacteria, the field was poised for to define the genetic basis for biomineralization in these fascinating organisms.

Genetic dissection of the MAI

The MAI in AMB-1 consists of 106 predicted coding sequences (43). It was divided into 14 smaller regions (named R1-R14) based on predicted operon structure, and each region was deleted separately. The results of this groundbreaking experiment showed that most of the MAI does not contribute to magnetosome formation under laboratory growth conditions. Deleting only three of the fourteen sub-clusters caused magnetosome formation defects. R2 contains a number of hypothetical proteins, and the small crystal phenotype in the initial ΔR2 strain was subsequently shown to be due to an unknown secondary mutation (unpublished results). R3 contains the *mms6* and *mamCDF* clusters, encoding numerous proteins previously shown to bind tightly to isolated magnetosome crystals. Finally, the *mamAB* region (R5) is required for producing both magnetic particles and magnetosome membrane invaginations, indicating that it contained the critical genes for magnetosome formation (47).

The *mamAB* region contains 18 predicted protein-coding genes. To understand the contributions of each of these genes to magnetosome formation, they were deleted individually. The resulting mutants were scored for magnetic response followed by TEM imaging to assess nanoparticle and magnetosome membrane formation. Groups of mutants with no obvious phenotype, small magnetite crystals, no detectable crystals but intact magnetosome membrane and no detectable magnetosome membranes were identified. The phenotypes for these individual gene deletions outlined a developmental process for magnetosome formation characterized by distinct morphological transformations. A step-wise model was proposed for magnetosome biogenesis in which specific sets of genes control membrane invagination, magnetite crystal nucleation and crystal growth (47).

Conclusions from the Murat study provided an invaluable set of insights about magnetosome formation. In addition to unambiguously identifying the core factors, it linked specific genes to known morphological transformations, giving each biogenesis factor a defined cellular function. Analogous experiments performed in *M. gryphiswaldense* MSR-1 confirmed the principles uncovered in AMB-1, suggesting that the functional assignments are conserved in a number of related strains (57, 58). Insights from the step-wise genetic model for magnetosome formation proposed in the Murat study led to a new set of questions for magnetosome research The frontier shifted from trying to identify which genes participated in the process to investigating the mechanisms through which individual gene products promoted each newly identified developmental stage.

Membrane biogenesis and chain alignment

An important class of deletions characterized in the Murat study showed morphologies that were indistinguishable from spontaneous deletions of the entire MAI. These genes, *mamB, mamI, mamL* and *mamQ,* appear to be required for the formation of the magnetosome membrane. Although these are the only genes required for membrane formation, they are not sufficient for membrane invagination in the ΔMAI strain (47). Their annotated functions do not suggest an obvious mechanism for membrane remodeling. *mamI* and *mamL* encode small proteins each with a single transmembrane (TM) helix. *mamQ* shows slight homology to LemA, an uncharacterized protein presented by the major histocompatibility complex during *Listeria* infection (59). *mamB* encodes a predicted cation diffusion family (CDF) transporter (60). Currently, none of these proteins have been characterized biochemically, and the mechanism for membrane invagination is unclear. Because these factors are required for the initial membrane remodeling step, potential roles in the subsequent biomineralization process cannot be ruled out.

Another class of genes participates in organizing the magnetosome chain within the cell. MamK is a bacterial actin-like protein that forms filaments required for the linear organization of magnetosome chains in AMB-1 (56). Purified MamK forms filaments in an ATP-dependent manner*,* confirming its predicted actin-like activity (61, 62). MamJ is an acidic protein with conflicting roles in AMB-1 and MSR-1. Δ*mamJ* cells have magnetosomes clustered at the cell pole in MSR-1, suggesting that MamJ participates in chain maintenance (63). However, Δ*mamJ* cells have only a subtle phenotype in AMB-1 in which the chain remains linear but has more gaps between individual magnetosomes (64). Furthermore, the *mamJ* deletion abolishes the fluorescence recovery after photobleaching (FRAP) of MamK-GFP filaments in these cells (64). The MamJ protein was reported to interact with MamK from MSR-1 in two-hybrid experiments, suggesting that the two proteins participate in the same maintenance mechanism (65). The current model is that MamJ modulates the filament architecture of MamK, but no definitive experimental evidence supports this model. Neither mutant has detectable defects in magnetite synthesis, indicting that the do not participate in biomineralization.

Roles for some genes in the *mamAB* region remain unclear. As described above, Δ*mamA* cells produce magnetite more slowly than wild-type upon addition of iron to iron starved cells (54). No phenotype has been detected in cells cultured under standard growth conditions and the exact role of MamA remains unclear (47). *mamU* and *mamV* show no observable defects in any aspect of magnetosome formation, and Δ*mamH* shows only an extremely modest magnetic response defect in both AMB-1 and MSR-1 (47, 66). It is important to note that phenotypes for the large majority of genes in the MAI that appear dispensable for magnetosome formation may not have been detected under the experimental conditions. Some of these genes may play roles similar to the reactivation from iron starvation seen in the Δ*mamA* strain or their activities may be redundant with other factors encoded in the island. Perhaps a more intriguing explanation is that their phenotypes would appear under different growth conditions. It is unlikely that the standard AMB-1 growth conditions represent the diversity of environmental conditions experienced in the environment. Perhaps if the cells were grown with different iron sources, pH or redox conditions some of these genes would become important.

BIOMINERALIZATION FACTORS

Challenges for magnetite biomineralization

The bioavailability of iron is dictated by a variety of abiotic transitions that occur under ambient environmental conditions (67). Ferric iron (Fe^{3+}) is insoluble in aqueous solution at neutral pH, but can be dissolved in highly acidic solutions (68). While the ferrous (Fe^{2+}) state is highly soluble, it is rapidly oxidized to Fe^{3+} when oxygen is present (68). Thus, upon introduction to environments that contain oxygen, un-chelated iron becomes oxidized and precipitates as a solid (69). These insoluble materials are extremely diverse, with a number of amorphous or crystalline compounds produced depending on the iron concentration, redox potential, pH, temperature and other factors (70). The complexity of abiotic iron transitions presents a number of challenges to magnetosome formation.

Iron acquisition is a major growth limitation for bacteria in a range of environmental conditions, from aquatic habitats to pathogenic situations in human hosts (67). Bacteria have machinery to directly import both ferric and ferrous iron, but these solutes are often too dilute for direct import (71, 72). Furthermore, insoluble ferric forms of dominate many environments, forcing bacteria to use other strategies (67). When iron concentrations are too low to import directly bacteria secrete small molecule siderophores that chelate ferric iron with high affinity (73). These compounds can even dissolve individual atoms from insoluble species (73). The proper deployment of these strategies is likely to be important for acquiring the large amounts of iron needed for magnetosome synthesis.

Magnetite is a mixed valence mineral, meaning that the iron atoms in the lattice do not have the same redox state. The chemical formula could more accurately be described as $Fe(II)Fe(III)_2O_4$ (14). Small amounts of magnetite can be made *in vitro* by adding an oxidant to precipitate ferrous iron solutions at high pH, but the sizes and number of nuclei are hard to control (18, 74). *In vivo*, redox and pH conditions are likely to be critical for producing magnetite while preventing the formation of numerous other iron-oxide precipitates. In addition, conditions in the lumen must restrict the process to the formation of a single crystal that grows into a large, precisely shaped particle. The basic chemical properties of iron suggest that biomineralization factors in magnetotactic bacteria use sophisticated mechanisms to control the conditions in each magnetosome compartment to make magnetite.

The initiation of biomineralization

Much like the genes required for membrane formation, deletions of *mamE, mamM, mamN,* or *mamO* produce cells with no magnetic response. However, when these cells are sectioned and imaged by TEM, they retain the ability to produce chains of membranes that appear to be empty magnetosome membranes (47). Thus, these genes are likely to be required for making magnetite within the magnetosome lumen. Though they are required for the initial nucleation step, they are likely to also contribute to the subsequent process of crystal growth. In contrast to the genes contributing to membrane formation, the functional predictions for the early biomineralization factors implicate specific biochemical activities that are required at this stage of magnetosome formation.

MamE is a predicted member of the HtrA family of trypsin-like proteases, a ubiquitous family of enzymes that participate in various aspects of protein quality control (75). They have domain structures consisting of an N-terminal trypsin-like domain and one or two C-terminal postsynaptic density 95/discs large/zonula occludens-1 (PDZ) domains (76, 77). MamE's domain

structure is similar, but two small c-type cytochrome domains (discussed below) are inserted between the trypsin and two PDZ domains (78). In addition to being required for initiating biomineralization, the *mamE* deletion seems to have an additional protein-sorting defect. GFP fusions to proteins normally localized at the magnetosome are dispersed throughout the cytoplasmic membrane in Δ*mamE* cells (47). Disrupting its protease activity using an allele with the predicted His-Asp-Ser catalytic triad mutated to alanines, produces cells with protein localization restored but severely stunted magnetite crystals (78). Thus, MamE has multiple roles in magnetosome formation. It has a protease-independent role in protein sorting and a protease-dependent crystal maturation function.

MamO also has a predicted serine protein protease domain. Instead of His-Asp-Ser, its catalytic triad is predicted to be His-Asp-Thr. Much like *mamE*, an allele of *mamO* with these three residues muted to alanine showed severe biomineralization defects, suggesting that it has a protease dependent role in magnetite crystal maturation (78). The proteolytic targets of the two serine proteases were not known before the work presented here, and the mechanism by which proteolysis could promote magnetite synthesis is not intuitively obvious. The C-terminus of MamO is a predicted TauE-like integral membrane transporter (78). TauE family members have been implicated in secreting various organosulfur compounds, but their specific cargoes and mechanism of action are not known (79).

Like the membrane biogenesis factor MamB, MamM is a member of CDF ion transporter family (60). CDF transporters are known to contribute to heavy metal tolerance by using the proton motive force to export metal ions out of the cytoplasm (80-82). The accumulation of MamB in MSR-1 requires MamM, and the two proteins have been shown to interact in both two-hybrid experiments and co-immunoprecipitations (83). Similar to other CDF family members, each protein has an N-terminal integral membrane region with six predicted TM helices and a C-terminal cytosolic domain (84). Given that the both proteins are CDF family members and form a complex, it is unclear why *mamB* deletions cannot invaginate magnetosome membranes, while *mamM* deletions stall at the later crystal nucleation stage (47). The crystal structure of MamM's soluble domain has been determined. Soaking experiments using zinc identified metal binding residues that caused crystal maturation defects when mutated (85). A thorough understanding of the protein's role in promoting crystal nucleation will require characterization of the full-length transporter.

MamN is predicted member of the NhaD family of integral membrane transporters. These proteins have been shown to mediate the proton motive force-dependent export of anions from the cytoplasm (86, 87). NhaD proteins contain 12 TM helices without significant soluble regions, and their mechanism of action remains unclear (87). The NhaD family is part of the ion transporter superfamily that contains representatives from throughout the tree of life (88). Interestingly, three of the four proteins required to initiate biomineralization have domains predicted to export ions from the cytoplasm. Since the magnetosome lumen is thought to be continuous with the periplasm, manipulating the solute environment of the magnetosome through ion export likely represents a critical biochemical function in magnetite synthesis (56). Because soluble iron accumulates in the cytoplasm in the ferric (Fe^{2+}) form, it seems likely that export of Fe^{2+} from the cytoplasm could produce high iron concentrations that favor crystal formation in the magnetosome lumen (67). However, *mamM, MamN* and *mamO* do not show genetic

redundancy in promoting crystal nucleation, suggesting that they likely transport have different mechanisms for manipulating the solute environment within the lumen (47).

Growth and maturation of magnetite crystals

The most abundant class of magnetosome formation mutants produces small or irregularly shaped nanoparticles within the magnetosome membrane. Many of these mutants maintain the ability to turn in a magnetic field but respond more weakly (47). Because the defects occur after the initiation of biomineralization, it is assumed that these factors control the subsequent development of magnetite crystals. However, TEM analysis of these strains indicates that their phenotypes vary considerably. Some mutants display magnetite crystals that are smaller, elongated in one direction or even distributed across a wider range of sizes (47, 57, 58, 89). For other mutants, the cells do not respond to a magnetic field, but non-crystalline electron dense material is observed by TEM (47). Despite the phenotypic differences, genes assigned roles in crystal growth and maturation are not required to precipitate insoluble material within the magnetosome lumen.

mamE, mamP, mamT and *mamX,* contain domains with CXXCH motifs in their predicted protein sequences (90). Such motifs are usually sites for the covalent attachment of heme prosthetic groups in *c*-type cytochromes, suggesting that these proteins participate in redox reactions that promote biomineralization (91). Purified forms of the proteins display spectral signatures associated with *c*-type cytochromes, and the domain was named "magnetochrome", due to its predominance in magnetosome island genes (92). The structure of the MamP protein from strain MO-1 was solved using X-ray crystallography. As predicted from its sequence, the protein contains a PDZ domain and two *c*-type cytochrome domains with a novel fold. The hemes in each of MamP's magnetochrome domains are highly solvent exposed and the protein can directly oxidize Fe^{2+} to Fe^{3+}, producing an insoluble iron species (93). However, the physiological relevance of this activity remains unclear.

Phenotypes caused by deleting magnetochrome proteins are consistent with a diverse set of roles in magnetite crystal development. Δ*mamT* cells produce small crystals, and Δ*mamX* cells do not display a biomineralization defect in AMB-1, but a slight defect has been reported in MSR-1 (47, 66). Although Δ*mamE* cells are unable to produce magnetite entirely, cells containing an allele with the CXXCH motifs disrupted produce small crystals, showing that MamE's magnetochrome motifs are required only for crystal maturation (78). Δ*mamP* cells produce an abundance of amorphous precipitates with occasional magnetite crystals that can be as large or larger that those observed in wild-type (47). This lack of uniformity in particle size and morphology suggests that the defects in Δ*mamP* cells may be due to an imbalance of iron or redox allocation and not a direct effect on crystal growth. Though deleting magnetochrome factors causes a variety of crystal maturation defects, the potential redundancy among these factors has not been explored.

In vitro iron oxidation by MamP suggests that magnetochrome proteins may manipulate the redox state of precursor atoms incorporated into the mixed-valence magnetite crystal (90). However, the sources of Fe^{2+} and Fe^{3+} atoms are not clear. AMB-1 expresses predicted Fe^{2+} transporters and produces siderophores for importing iron into the cytoplasm (94, 95). The predominance of predicted solute exporters within the *mamAB* region that are required to initiate biomineralization supports the idea that Fe^{2+} accumulates in the cytoplasm and is pumped into

the magnetosome lumen, but experimental support for this idea is lacking(47) . Furthermore, the terminal oxidant for a potential iron oxidation pathway is not known. One possibility is that magnetochrome domains are re-oxidized for continuous iron oxidation by donating electrons to components of the respiratory chain. In this way, they would link the oxidation of iron to respiratory reduction of oxygen or nitrate. Physiological data about the electron balance processes occurring during biomineralization will be required to clarify magnetochrome-based electron shuttling pathway(s).

One of the earliest sets of magnetosome associated proteins identified is also now known promote proper growth of magnetite crystals. The *mms* proteins were discovered as factors that resisted extraction from biochemically isolated magnetosome crystals, indicating that they were tightly bound the magnetite surface (41). The genes encoding these factors are contained in two adjacent operons that make up the R3 region in AMB-1, whose deletion produces small, elongated crystals (47). Subsequent dissections of this region have shown that the *mamCDF* operon contributes only slightly to this defect, but deleting the *mms6* operon has nearly as severe a defect as the entire R3 deletion (89). Individual gene deletions of the operon showed that *mmsF* is likely the key genetic factor to this elongated crystal phenotype (89). Like many of the other island regions, the role of redundancy in masking phenotypes from individual factors has not been examined comprehensively.

The proteins encoded by *mms6* and *mmsF* have both been studied *in vitro.* Purified Mms6 binds directly to iron atoms and can increase the size and uniformity of magnetite particles formed by *in vitro* iron precipitation (41, 96-98). Its acidic C-terminal tail, which is sufficient for these activities, has been used to manipulate the nano-scale patterns of iron-oxide precipitation (99, 100). Purified MmsF was reported to form protein-based vesicles that can also enhance the size and uniformity of magnetite precipitated *in vitro* (101). These activities combined with their tight association with the magnetite surface have led to the suggestion of crystal templating by the Mms proteins (102). This model predicts that the proteins bind to specific faces of growing crystals, enhancing or inhibiting growth in specific directions to control the shape of the mature particle.

The *in vitro* activities identified for Mms6 and MmsF have obvious biotechnological implications, but their physiological relevance remains questionable. Both proteins contain strongly predicted TM helices, and GFP-MmsF localizes to the cytoplasmic membrane in ΔMAI cells (89). However, both biochemical assays use protein preparations lacking detergents, indicating that the protein "assemblies" seen in the preparations are likely to be non-physiological aggregates. Furthermore, the C-terminal portion of Mms6 used for numerous studies on mineral patterning, is predicted to face the cytoplasmic side of the membrane. Such a topology would be inconsistent with this motif having any interaction with growing minerals in the lumen. A more rigorous approach emphasizing the protein's context within the membrane and the genetic effects of disrupting specific motifs will be required to clarify the relevance of the patterning model.

A final set of crystal maturation genes has no predicted mechanism of action. MamR is a small protein with a predicted helix-turn-helix motif, and MamS does not show homology to any known protein families. Both genes are contained in the *mamAB* region of the MAI. Δ*mamR*

cells have small crystals, while Δ*mamS* cells produce chains of amorphous precipitates (47). The MAI encodes a protein with predicted homology to the filament forming cell division factor FtsZ (57, 103). Deleting the *ftsZ-like* gene in MSR-1 produces cells with abnormally small magnetite crystals, however no phenotype is observed in AMB-1. Like other tubulin homologs, purified FtsZ-like formed filaments in a GTP-dependent manner (103). The functional predictions for these three proteins do not suggest obvious mechanisms by which they promote biomineralization.

Biomineralization factors outside the MA

Producing magnetosomes is a specialized process that is not required for the viability of the cells. However, it seems to interface with a number of central activities that are occurring during normal growth. Membrane remodeling, iron acquisition, and redox homeostasis are not unique challenges for making magnetosomes. Understanding how magnetosome formation is integrated into the broader cellular landscape can be difficult to study genetically because many of those processes are essential for growth. Nonetheless, a number of biomineralization mutants have been identified in genes that are not contained in the MAI, providing hints about how the process is linked to central metabolism.

A critical link to broad cellular processes seems to be the mechanism of iron accumulation in the magnetosome. One of the earliest transposon-insertion mutants in AMB-1 that led to a loss of biomineralization was in a predicted iron transporter named *magA* (53, 104). Indeed, recombinant MagA could transport iron in an ATP-dependent fashion when reconstituted into lipid vesicles, and its heterologous expression in mammalian cells led to an over-accumulation of iron in the cytoplasm (105, 106). However, a recent study found that creating unmarked deletions of the *magA* gene in either MSR-1 or AMB-1 had no effect on biomineralization (107). It is unclear what caused the phenotype of the initial transposon mutation. Since a complementation attempt was not reported, is possible that the transposon insertion had polar effects on adjacent genes or even that a spontaneous deletion of the MAI had occurred.

A number of known iron homeostasis genes have been targeted directly to determine their potential roles in magnetosome formation. Deleting the ferric uptake repressor (*fur*), which regulates a number of iron responsive genes, causes the production of smaller magnetite crystals in MSR-1 (108, 109). Interestingly, none of the *mamAB* genes had altered transcriptional responses in the Δ*fur* strain, suggesting that the defect is due to a more general disruption of iron homeostasis (108, 109). In fact, a number of studies looking at a variety of growth conditions an iron sources have been unable to detect major transcriptional signals controlling expression of genes in the MAI (95, 110). The lack of strong transcriptional regulation suggests that magnetosome biogenesis genes encode a network of biochemical activities capable of regulating the process post-transcriptionally.

MSR-1 encodes two homologs of the ferrous iron transporter *feoB,* with *feoB1* contained in the MAI and *feoB2* located elsewhere in the genome (111). Δ*feoB1* cells produce fewer magnetite crystals that are smaller than wild-type (111). Δ*feoB2* cells show no significant biomineralization defect, but *feoB2* has an additive effect when it is deleted in a Δ*feoB1* background (112). It seems that the ferrous iron import by FeoB1 is critical to biomineralization, but that FeoB2 contributes slightly under conditions in which FeoB1 is absent. *ΔfeoB1ΔfeoB2* cells can still produce

magnetite crystals indicating that their presumed iron transport activity is not absolutely required for magnetite synthesis (112). Importantly, the *feoB1* gene is contained in the R13 section of the MAI of AMB-1 whose deletion showed no biomineralization phenotype (47). The basis of this inconsistent requirement for the *feoB1* gene is not clear, but the different conditions used for growing AMB-1 may mask the importance of the FeoB proteins.

Both MSR-1 and AMB-1 produce magnetosomes only under microaerobic or anaerobic atmosphere (37). To support growth under these conditions, nitrate must be supplemented in the medium as an alternative electron acceptor. MSR-1 contains a complete pathway for the denitrification of nitrate to nitrogen gas through the successive action of the nitrate reductase (Nap), nitrite reductase (Nir), nitric oxide reductase (Nor), and nitrous oxide reductase (Nos) enzymes. Through deletion analysis, each of these enzymes was shown to be required for anaerobic growth with nitrate as the electron acceptor, with the exception of Nos. Δ*nos* cells can grow slowly under anaerobic conditions. AMB-1 contains genes predicted to encode each of the denitrification components, suggesting that it can also perform denitrification (113).

Both the *nap* and *nirS* enzymes appear to contribute to biomineralization. Deletion of *nap,* encoding the periplasmic nitrate reductase, produces cells that grow normally under microaerobic conditions but produce fewer magnetite particles that are smaller and less uniform than wild type. Δ*nap* cells consume added nitrate without improvement of the biomineralization defect, suggesting that the phenotype is not due to effects on the downstream enzymes (113). Deletion of *nirS*, encoding the nitrite reductase, shows both biomineralization and growth defects under microaerobic and anaerobic conditions (114). It is not clear whether these defects are because of a direct link to biomineralization or a nonspecific effect from improper growth or accumulation of nitrite. Δ*nor* and Δ*nos* do not show significant biomineralization defects (113). Finally, deletion of a cbb_3-type cytochrome oxidase, predicted to catalyze the reduction of molecular oxygen under both aerobic and microaerobic conditions, was also shown to inhibit growth and magnetite crystal maturation in MSR-1 (115).

These results further emphasize the link between central respiratory processes and the maturation of magnetite crystals. Enzymes involved in reducing both terminal electron acceptors utilized by the *Magnetospirilla* during biomineralization are required for complete biomineralization, suggesting that these established electron transfer chains are used during magnetosome biogenesis as electron sinks. It is tempting to speculate that the magnetochrome proteins are involved in shuttling electrons from the reduction of ferrous iron into the respiratory chains. The finding that biomineralization defects caused by deleting *mamX* are reminiscent of a *nap* deletion but exacerbated in a Δ*nap* background supports this idea (66). Perhaps, the various magnetochrome proteins provide a redundant network of electron carriers to link iron reduction and respiratory metabolism.

A recent study showed that magnetosome formation can occur in a heterologous host. A set of engineered cassettes containing the *mamAB, mamCDF,* and *mms* operons from MSR-1 is able to allow for magnetosome formation when transferred to *Rhodospirillum rubrum.* Much like in native *Magnetospirillum* hosts, magnetosomes were only produced under conditions in which oxygen was limited or absent (116). Thus, the physiological conditions required for magnetosome synthesis were identical suggesting that the respiration of nitrate under low oxygen

tension is also required in the heterologous host. However, it is not clear if the pathway for electron dissipation during iron oxidation is identical. Because *Magnetospirillum* and *Rhodospirillum* are quite closely related, their respiratory complexes may be similar enough to interface with magnetosome synthesis in the same way (52). As the field moves forward, heterologous expression presents an interesting approach to understand links to central cellular processes by determining host factors required to support magnetosome formation.

CONCLUSIONS AND FUTURE DIRECTIONS

Genetic analysis of magnetosome formation sparked a paradigm shift in how the process is dissected and understood. Any gene in MSR-1 or AMB-1 can be deleted and quickly screened for a biomineralization defect. The phenotypes can then be clarified further through EM-based imaging. In recent years, this pipeline has lead to the identification and characterization of dozens of magnetosome biosynthesis mutants. In the summary above, each of these genes is described along with the specific morphological defects observed along with any predicted mechanism posited based on its annotated function. The list of genes involved at each stage of the process continues to grow.

The challenge is now to define the biochemical mechanisms employed by each gene product in order to define the process in more detail. For instance, at least five predicted solute transporters (MamB, MamM, MamN, MamO and FeoB1) contribute to various steps in the biogenesis process. These proteins are likely maintaining the proper solute environment in the magnetosome lumen, but the specific details are lacking. Functional annotations cannot accurately predict what specific solutes are being transported, what mechanism is being employed or which direction compounds move across the membrane. Biochemical experiments that examine the activities of each transporter would help clarify how the solute environment within the lumen is manipulated during biomineralization.

The goal of the work presented in this thesis is to understand the biochemical activities of protein factors required for magnetite nucleation. Specifically, the serine proteases MamE and MamO were dissected in detail. Biochemical, and genetic experiments define a pathway in which MamO activates MamE in a non-proteolytic manner to cleave a multiple biomineralization factors. Structural studies with MamO show that it is a degenerate protease that has been repurposed to bind metal ions, promoting the nucleation of magnetite crystals *in vivo*. The *in vitro* reconstitution of MamE-dependent proteolysis defines a switch-like behavior through which its proteolytic activity regulates the maturation of magnetite crystals. In addition to defining the activities of MamE and MamO, the work here provides a framework for combining *in vitro* analysis and genetics to link specific biochemical functions to the crystal transformation process occurring in the cells.

In general, biochemical approaches will provide a more detailed view of the biomineralization process, by advancing the model from a list of genes involved in magnetite formation to a list of defined biochemical activities. Hopefully, future studies will continue to address the mechanisms by which magnetotactic bacteria manipulate iron-based minerals with such precision. Protein-metal interfaces like those characterized in MamO, are likely to be uncovered, advancing the interests of materials synthesis for useful applications. Doing this will require a change in approach, emphasizing detailed characterization of known factors over collecting additional

mutants. By defining the process at the individual protein level, the field will be poised for another revolution in understanding the molecular basis for magnetite biosynthesis.

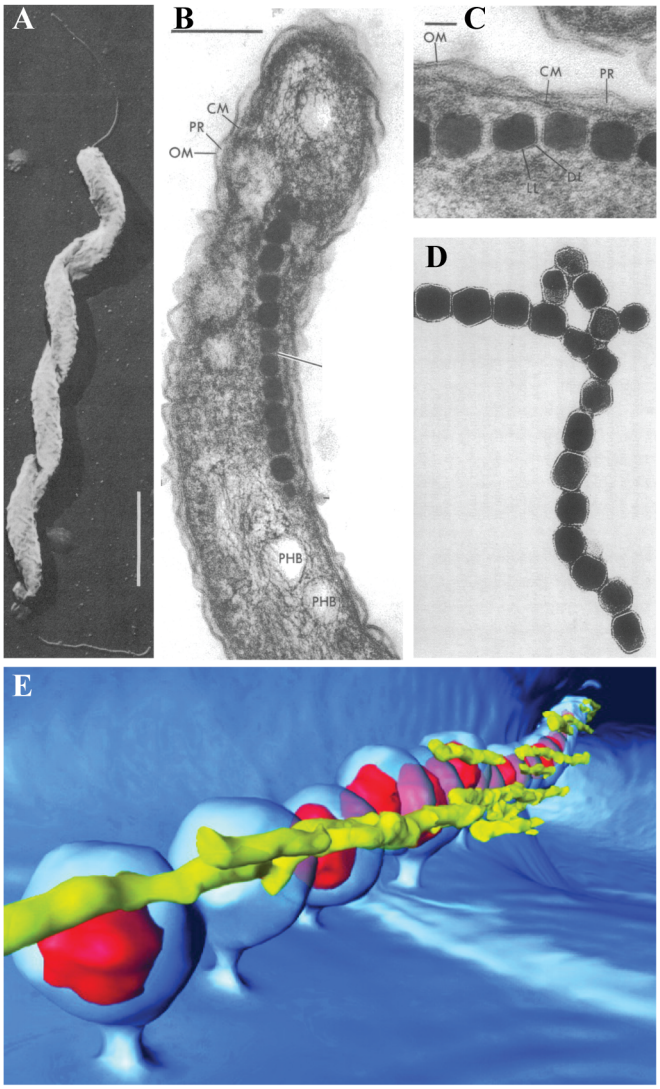


Figure 1-1. Historical representation of the cellular basis for magnetosome formation. (a) SEM analysis of MS-1 cells. (b) TEM imaged of sectioned MS-1 cell showing the close proximity of the magnetite particles to the cell envelope. (c) Close-up view of the association of magnetosomes with the cell envelope in MS-1. OM: outer membrane; PR: periplasmic space; CM: cytoplasmic membrane. (d) TEM of magnetosomes purified from lysed MS-1 cells. A membrane structure appears to surround each magnetite crystal. (e) Three dimensional represnetaion of the magnetosomes in an AMB-1 cell derived from an ECT reconstruction. The magnetite crystals are surrounded by cytoplasmic membrane invaginations and organized in a chain by the filamentous protein MamK. Panels a,b and c from Ref. (35). Panel d from Ref (36). Panel e from Ref (117).

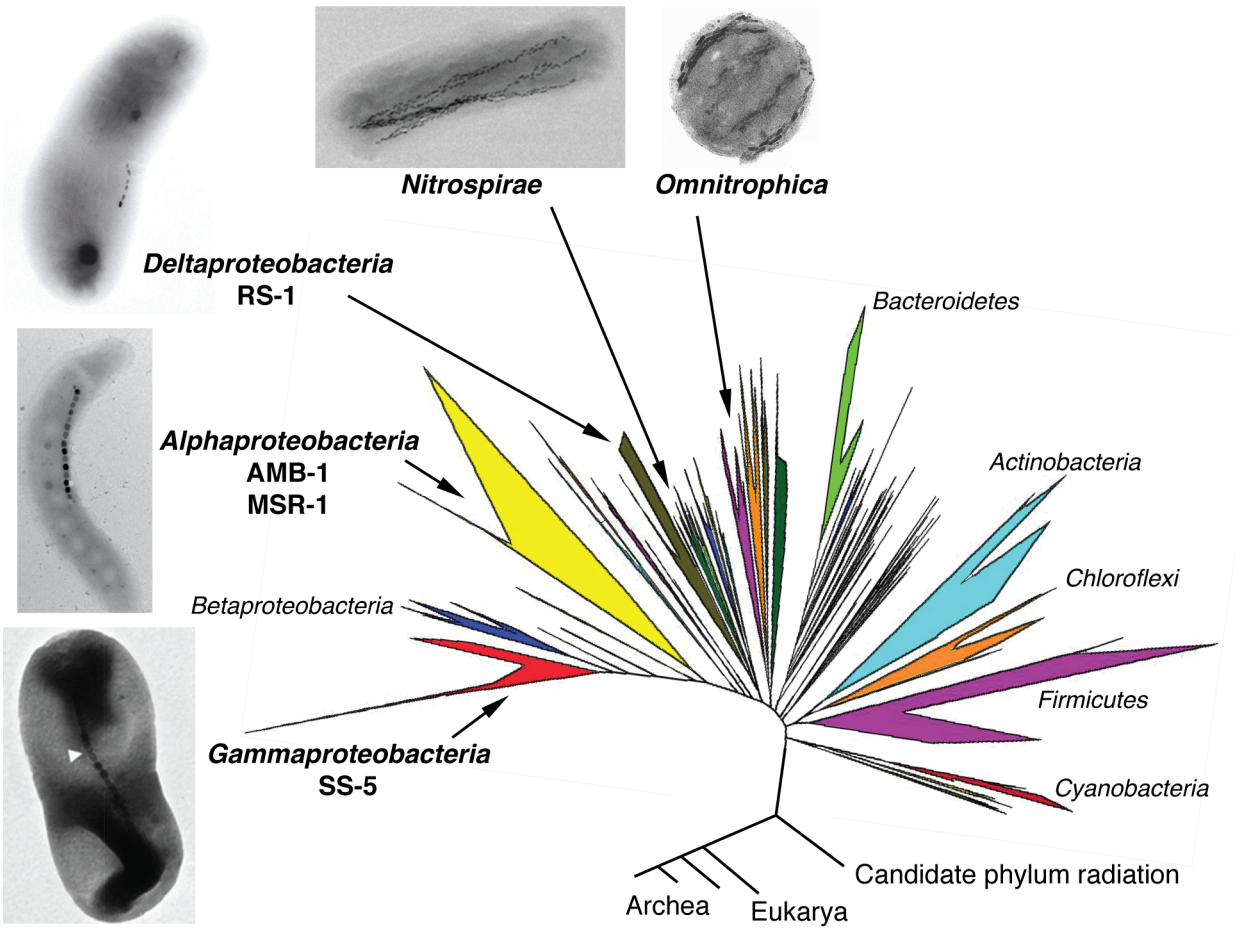


Figure 1-2. Phylogenetic distribution of magnetosome formation. Phyla containing magnetotactic representatives are shown in bold and are identified within the tree of life adapted from Ref. (118). Images from Refs.
(29, 33, 66, 119, 120).

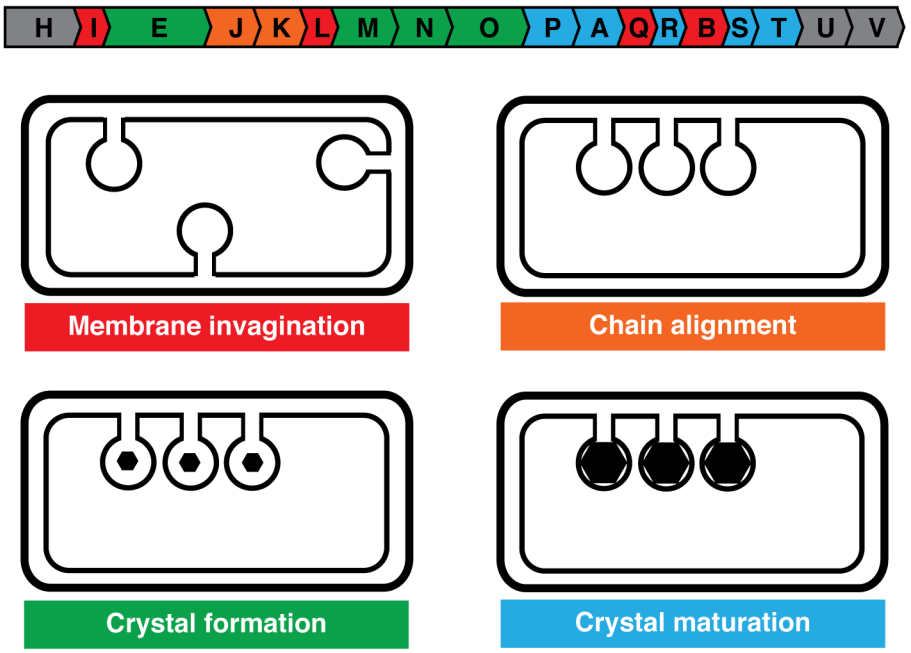


Figure 1-3. Overview of the step-wise genetic model for magnetosome formation in AMB-1. The genomic organization of the *mamAB* region is shown with the colors of each gene representing the stage at which the process has stalled.

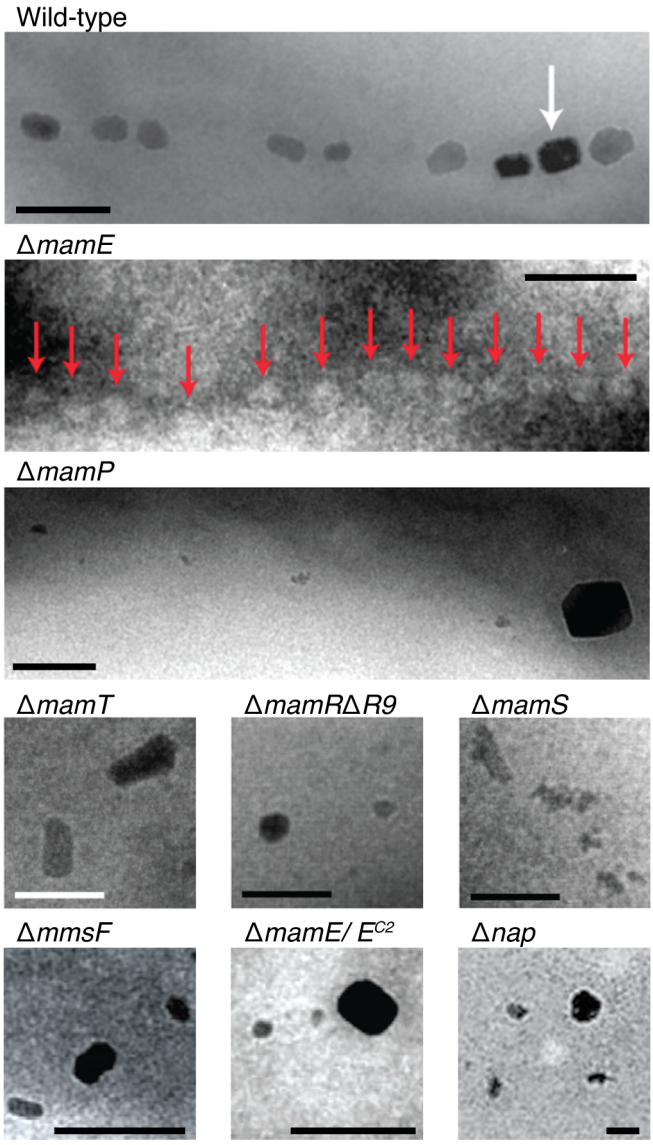


Figure 1-4. Morphological diversity of mutants in later stage of the biomineralization process. Each of the crystal maturation mutants has a distinct defect with respect to the size and morphology of the biominerals. The scale marker in each panel represents 50nm. Images derived from Refs. (7, 47, 78, 89, 113).

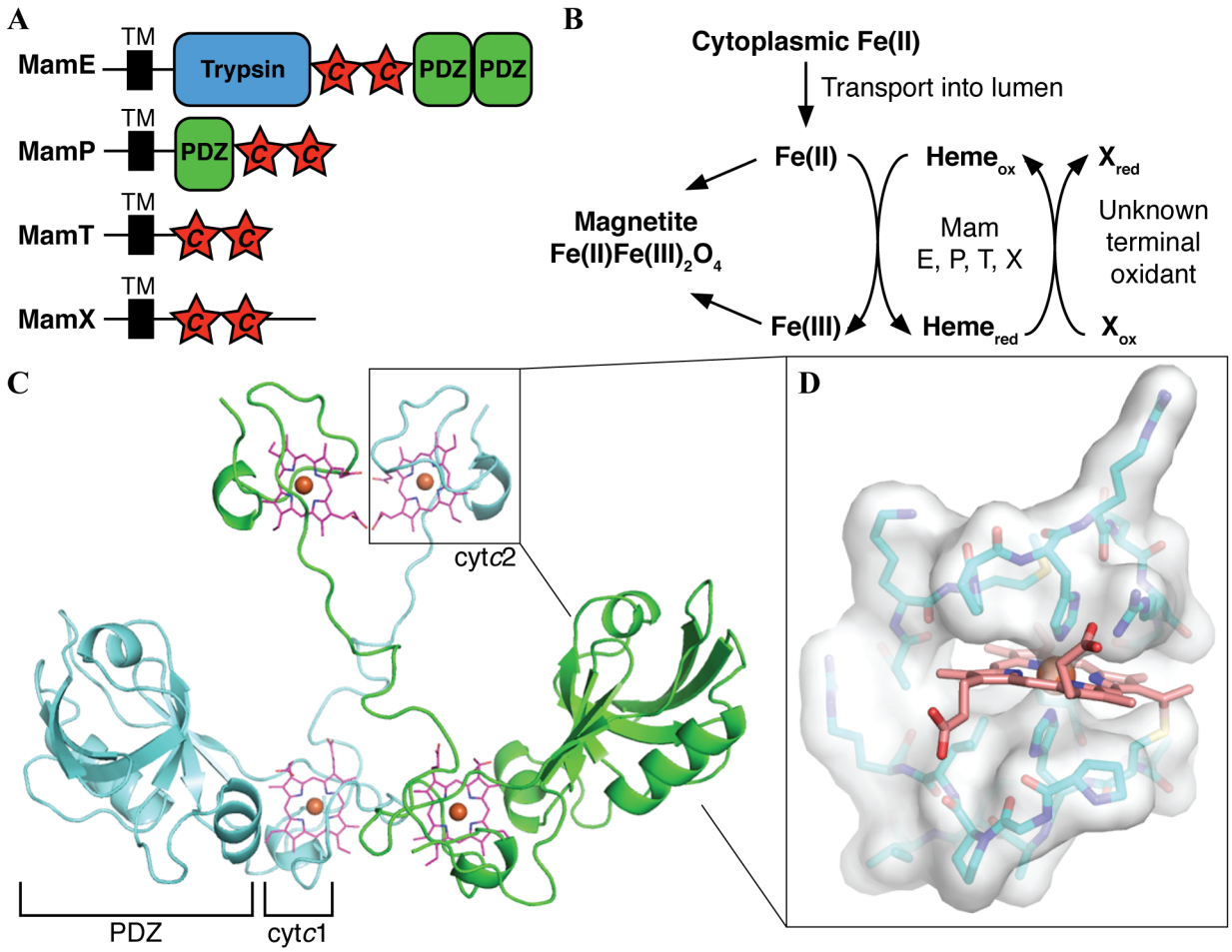


Figure 1-5. Overview of the magnetochrome domain. (a) Domain structure for the magnetochrome containing proteins in AMB-1. (b) Model for magnetochrome mediated biomineralization through iron oxidation. (c) Overview of the structure of the structure of MamP from MO-1 determined in Ref. (93). (d) Inset from Panel c showing view of the magnetochrome fold with an octahedral coordination sphere of the iron atom including two protein derived histidines residues.

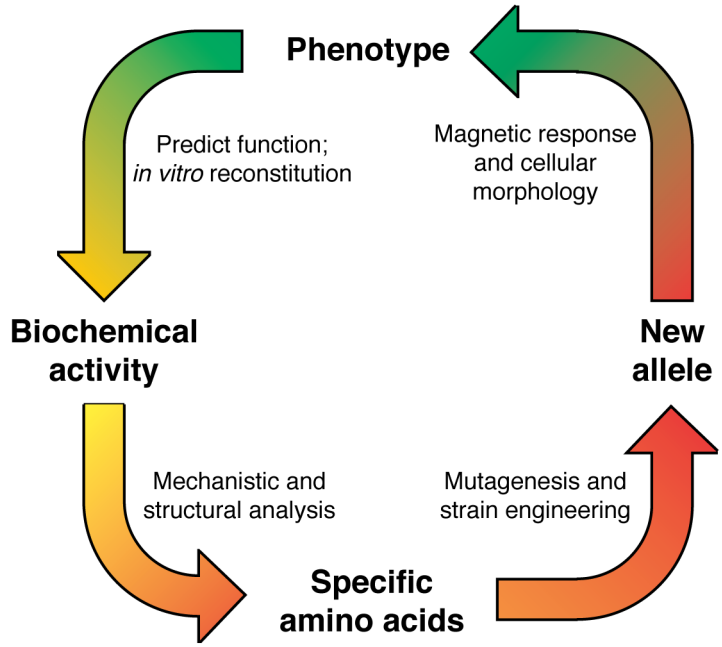


Figure 1-6. Overview of the proposed iterative approach combining biochemistry and reverese genetics to characterize the mechanisms of individual magnetosome biogenesis factors. The approach formed the basis of the experiments described in Chapters 2 and 3 of this thesis.

CHAPTER 2

MamO is a repurposed serine protease that promotes magnetite biomineralization through direct transition metal binding in magnetotactic bacteria

David M. Hershey[1], Xuefeng Ren[2,3], Ryan A. Melnyk[5], Patrick J. Browne[1], Ertan Ozyamak[1], Stephanie R. Jones[4], Michelle C. Y. Chang[2,4,5], James H. Hurley[2,3,6] and Arash Komeili[1,2,3]

[1] Department of Plant & Microbial Biology, [2] Department of Molecular & Cellular Biology, [3] California Institute for Quantitative Biosciences, [4] Department of Chemistry, University of California, Berkeley, CA 94720, USA;
[5] Physical Biosciences Division, [6] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ABSTRACT
Many living organisms transform inorganic atoms into highly ordered crystalline materials. An elegant example of such biomineralization processes is the production of nano-scale magnetic crystals in magnetotactic bacteria. Previous studies implicated the involvement of two putative serine proteases, MamE and MamO, during the early stages of magnetite formation in *Magnetospirillum magneticum* AMB-1. Here, using genetic analysis and X-ray crystallography we show that MamO has a degenerate active site, rendering it incapable of protease activity. Instead, MamO promotes magnetosome formation through two genetically distinct, non-catalytic activities: activation of MamE-dependent proteolysis of biomineralization factors and direct binding to transition metal ions. By solving the structure of the protease domain bound to a metal ion, we identify a surface exposed di-histidine motif in MamO that contributes to metal binding and show that it is required to initiate biomineralization *in vivo*. Finally, we find that pseudo-proteases are widespread in magnetotactic bacteria and that they have evolved independently in three separate taxa. Our results highlight the versatility of protein scaffolds in accommodating new biochemical activities and provide unprecedented insight into the earliest stages of biomineralization.

INTRODUCTION
Biomineralization is the widespread phenomenon by which living organisms transform inorganic atoms into highly ordered, crystalline structures. Controlling the size and shape of such materials requires specialized protein machinery that can define the nano-scale trajectory of crystal growth (3). Incorporating biochemical principles uncovered from studying biomineralization has the potential to improve the design and synthesis of nanomaterials *in vitro* (19). In addition to the well-known examples of tooth, bone, and shell production by multicellular eukaryotes, a number of bacteria have the ability to biomineralize small magnetic crystals within subcellular compartments called magnetosomes (20, 35). These particles allow the cells to passively align in the earth's magnetic field, facilitating the search for their preferred oxygen environments (21). Although these magnetotactic bacteria have drawn longstanding interest due to their ability to manipulate transition metals, the biochemical details of how they transform iron into magnetite (Fe_3O_4) remain poorly understood.

Magnetotactic organisms are phylogenetically diverse. Nearly all isolates come from the α-, δ- or γ- classes of *Proteobacteria,* but representatives from the *Nitrospirae* and *Omnitrophica* phyla have recently been identified (121). The genes responsible for making magnetosomes are often contained in a genomic region called the magnetosome island (MAI) (39, 43, 44, 47, 57). Comparative genomic and phylogenetic studies have identified a set of core genes that appears to have been assembled a single time and inherited vertically, indicating that magnetosome formation likely predates the divergence of the *Proteobacteria* (30, 51). The MAI seems to have formed by incorporating elements from other cellular processes, as the majority of the core factors have homology to ancient and widespread protein domains (42, 122). Uncovering the biochemical functions encoded in the MAI in relation to its evolutionary history provides a unique opportunity to understand how new cellular processes evolve.

Due to the availability of genetic systems, α-*Proteobacteria* like *Magnetospirillum magneticum* AMB-1 are used as models for studying the molecular biology of magnetosome formation (7).

AMB-1 contains 15-20 magnetite crystals, each formed within a cytoplasmic membrane invagination and organized in a chain spanning the length of the cell (54, 56). By making deletions within the MAI and characterizing the ultrastructure of the mutant cells, specific genes have been assigned roles in various stages of magnetosome formation (47, 57, 58, 102). Genes whose deletions produced empty magnetosome compartments or compartments with abnormally small magnetite crystals were termed biomineralization factors. It is the proteins encoded by these genes that hold the secrets of how magnetotactic organisms interface with solid magnetite.

Two genes in the MAI, *mamE* and *mamO*, are homologous to the HtrA proteases, a ubiquitous family of trypsin-like enzymes that functions using His-Asp-Ser catalytic triads (75). An additional pair of genes, called *limE* and *limO,* with homology to the protease domains of *mamE* and *mamO* exists in a secondary genomic region termed R9 (47). Disrupting *mamE* or *mamO* causes cells to produce empty magnetosome membranes, but removing R9 has no effect, showing that *mamE* and *mamO* are required for biomineralization and that *limE* and *limO* are not (Figure 2-1) (47, 123). Adding variants of *mamE* or *mamO* with all three predicted active site residues mutated to alanine could not restore normal biomineralization in the Δ*mamO*Δ*R9* or Δ*mamE*Δ*R9* strains but complemented single Δ*mamO* or Δ*mamE* deletions (78). These genetic analyses show that *mamE* and *mamO* are required for the initiation of magnetite biomineralization. Furthermore, *limE* and *limO* are partially redundant in that they can cross-complement the active site dependent crystal maturation defects of their respective orthologs.

Here we use a combination of *in vivo* and *in vitro* approaches to reveal an unexpected dual role for MamO. It promotes MamE-dependent proteolysis of three biomineralization factors through the use of its C-terminal transporter domain. Separately, the protease domain has lost the ability to carry out catalysis and has instead been repurposed to bind transition metal ions. Two surface-exposed histidine residues that contribute to this metal-binding function are required for initiating magnetite biosynthesis *in vivo*. Bioinformatic analysis shows that similar pseudo-proteases evolved independently in the three major taxa of magnetotactic organisms, highlighting a unique evolutionary mechanism behind microbial nanoparticle synthesis.

MATERIALS & METHODS

Strains, plasmids and culture conditions

The strains and plasmids used in this study are listed in Tables 2-4 and 2-5, respectively. For general maintenance and genetic manipulation, *M. magneticum* AMB-1 was grown in MG medium supplemented with ferric malate (30μM). 0.7% agar was used in plates and kanamycin was used for antibiotic selection at a concentration of 7 μg/mL (solid) or 10μg/mL (liquid). For sucrose counter-selection, MG plates contained 2% sucrose. Cultures for magnetic response measurements, western blotting or transmission electron microscopy (TEM) were grown in 10 mL MG medium containing 25mM HEPES buffer (pH 7.2) and ferric malate under a 10% oxygen atmosphere. For comparing the temperature dependence of magnetic response, the strains were treated as above except that they were grown under anaerobic atmosphere. The magnetic response of each culture was assessed using the Coefficient of Magnetism (C_{mag}), which was measured as described (56).

Genetic manipulation

For complementation of deletion mutants, we used a modified form of pAK253(47). This plasmid contains a neutral region of the AMB-1 genome and integrates as a single copy at this site. Each allele is inserted under the control of the *mamAB* promoter, allowing the constitutive expression of each protein. To create the Δ*mamE*Δ*mamO*Δ*R9* (Δ*E*Δ*O*) strain, plasmid pAK243 (*sacB* based counterselection system for deleting *mamE*) was transformed into strain AK94 (Δ*mamO*Δ*R9*), and initial integrations were selected using kanamycin. The resulting strains were grown to stationary phase in MG medium without antibiotic selection and plated on sucrose for counter selection. Deletion of *mamE* was confirmed by antibiotic sensitivity, PCR analysis and complementation of the magnetic phenotype upon reintroduction of both *mamE* and *mamO*.

Protein analysis
Cultures of AMB-1 were grown to late-log phase and harvested by centrifugation at 6k x *g*. The resulting pellets were resuspended in PBS. The cell suspensions were mixed with an equal volume of 4x SDS Loading buffer and heated for 10 min at 75°C. The lysates were separated on SDS-PAGE and transferred to PVDF. The membranes were blotted and visualized using standard western blotting techniques. Polyclonal antibodies to MamE and MamP were raised in rabbits against recombinant forms of the soluble portion of each protein (124). The monoclonal anti-FLAG antibody was purchased from Sigma.

Transmission electron microscopy (TEM)
For TEM, cultures were grown to late-log phase. 1mL from each culture was pelleted at 16k x *g*, and the pellet was resuspended in the residual medium. Cell suspensions were spotted on formvar coated copper grids, rinsed, dabbed dry and stored at room temperature until imaging. Imaging was performed with a FEI Tecnai 12 TEM at an accelerating voltage of 120kV. For each strain, 15-20 cells totaling >200 crystals were analyzed.

Protein expression
The protease domain of *Mm*MamO (resdiues 45-261) was cloned into mcsII of pETDuet to create pAK876 for expression without a tag. BL21 Codon Plus cells transformed with pAK876 were grown at 37°C in 2xYT with carbenicillin (100 μg/mL) and chloramphenicol (25 μg/mL) until the OD_{600} reached 1.0. The cultures were then equilibrated at 20°C for 30 min, induced with 0.125mM IPTG and grown overnight. Cells were harvested by centrifugation, resuspended in Buffer A (25mM Tris-HCl pH7.4, 400mM NaCl, 5mM Imidazole, 10% glycerol) supplemented with 1μg/mL pepstatin A, 1μg/mL leupeptin and 0.5mM DTT, frozen in liquid nitrogen and stored at -80°C.

Protein purification
For crystallography, the frozen cell suspension was thawed on ice, lysed by sonication and clarified at 13,000 x *g* for 30 min. The supernatant was loaded on a 3 mL Ni-NTA column, which was then washed with 5 column volumes of Buffer B (25mM Tris-HCl pH7.4, 500mM NaCl, 10mM Imidazole, 10% glycerol). The protein binds to Ni-NTA with its native metal binding site. After elution with 1 column volume Buffer C (25mM Tris-HCl pH7.4, 250mM NaCl, 250mM Imidazole, 10% glycerol), the sample was dialyzed overnight against Buffer D (25mM Tris-HCl pH8.0, 50mM NaCl, 10% glycerol, 0.5mM EDTA) and loaded onto a 1mL HiTrap CaptoQImpRes column. The column was developed to Buffer E (25mM Tris-HCl pH8.0, 1M NaCl, 10% glycerol), and the peak fractions containing the MamO protease domain were

pooled, exchanged into Storage Buffer (25mM Tris-HCl pH7.4, 300mM NaCl, 10% glycerol), concentrated to ~20mg/mL and frozen in liquid nitrogen for storage at -80°C.

For tmFRET, the expression was performed as above except that the various mutants of the *Mm*MamO protease domains were expressed as fusions to a C-terminal strepII tag. The lysate was prepared as above and loaded onto a 1mL StrepTrap HP column, which was washed with 5 column volumes of Buffer F (25mM Tris-HCl pH7.4, 250mM NaCl, 10% glycerol) and eluted with 3 column volumes of Buffer F supplemented with 2.5mM desthiobiotin. The eluate was concentrated and loaded onto a 16/60 Superdex 200 column and developed in Storage Buffer. The peak fractions were pooled, concentrated and used immediately for fluorescent labeling. The purity of all proteins was confirmed by SDS-PAGE.

Crystallization and structure determination
Frozen aliquots of untagged MamO protease domain were thawed on ice and exchanged to Buffer G (20mM Tris-HCl pH7.4, 150mM NaCl, 5% glycerol, 0.5mM DTT) while adjusting the protein concentration to 5mg/mL in a 10kDa cutoff Amicon ultrafilter. Crystals grew in the hanging drop vapor diffusion format after mixing the protein with an equal volume of well solution (50mM Na-Acetate pH4.6, 3.6M NH_4Cl, 5% glycerol) and equilibrating against 1mL well solution at 18°C. Cubic crystals of MamO appeared in 1-2 days and grew to their full size after 4-6 weeks. Each crystal was cryo-protected with a solution of 50mM Na-Acetate pH 4.6, 3.6M NH_4Cl and 22% glycerol before plunge-freezing in liquid N_2.

For nickel soaking, $NiCl_2$ was added directly to drops containing fully-grown crystals for a 10mM final concentration. The wells were re-sealed and the crystals soaked overnight. At harvest, each crystal was passed through three 1 min soaks in a buffer containing 50mM Tris-HCl pH 8.0, 3.6M NH_4Cl, 22% glycerol, 10mM $NiCl_2$ before plunge-freezing.

Diffraction data was collected on beamline 8.3.1 at the Advance Light Source (Lawrence Berkeley National Laboratory, Berkeley CA). Indexing and scaling was performed using HKL2000 (125). The apo- structure was solved by molecular replacement in Phaser with the protease domain of *Ec*DegP (1KY9) as a search model (76). The Ni-bound structure was solved using the apo-MamO structure as a molecular replacement search model. Model building and refinement were carried out with alternating cycles of COOT (126) and phenix.refine (127). Placement of the Ni ion was confirmed with an anomalous map from SAD data collected at the Ni absorption edge.

Fluorescein-5-maleimide (F-5-M) labeling
For fluorescent labeling, variants of the MamO protease domain were diluted to 50μM in 1mL of Buffer H (50mM NaPhosphate pH7.2, 150mM NaCl, 10% glycerol). Fluorescein-5-maleimide (dissolved at 50mM in DMSO) was added to a final concentration of 1mM, and the reaction was incubated overnight at 4°C. The reaction was quenched with DTT, exchanged into Buffer D using a PD-10 column, and loaded on a 1mL HiTrap Q FF. The column was developed to Buffer E, and each protein eluted as a single fluorescent peak. The peak fractions were pooled, treated with 1mM EDTA to remove any trace metal, exchanged extensively into Chelex-treated Storage Buffer and frozen in small aliquots. Labeling efficiency was measured based on A_{492}/A_{280} and was between 50-55% for all preparations.

tmFRET

For metal binding experiments all buffers were prepared in acid-washed glassware and treated with chelex resin. Each protein was diluted to 60-80nM in fluorescence buffer (10mM Tris-HCl pH 8.0, 150mM NaCl). Metal solutions were prepared at 10X concentration in chelex treated H_2O. Protein was dispensed into 96-well plates, and the plates were scanned for fluorescence emission in a Tecan Infinite 200 plate reader in top-read mode (Ex: 492nm; Em: 505-570nm). Metal solutions were then diluted into each well at the appropriate concentration and the plate was re-scanned. Due to the spontaneous oxidation of ferrous iron in ambient atmosphere, all iron binding experiments were performed in the absence of oxygen. For iron binding, all solutions were prepared using anoxic liquids. The samples were prepared under anoxic atmosphere in a clear-bottom 96-well plate. To prevent introduction of oxygen, the wells were sealed by covering the plate with a black adhesive cover. The plate was then removed from the anaerobic chamber and scanned using bottom-read mode.

Each measurement was performed on independently prepared solutions in quadruplicate. The metal quenched fluorescence spectrum from each well was normalized to the fluorescence before metal addition. F_{metal}/F represents the normalized fluorescence averaged from an 11nm window around the peak. After plotting F_{metal}/F as a function of metal concentration, each curve was fit to a two site model (below), where K_{d2} and E represent the dissociation constant and FRET efficiency for metal binding by MamO, respectively. K_{d1} represents a nonspecific, solution-based quenching component (128).

$$\frac{F_{metal}}{F} = \left(1 - \frac{E}{1 + \frac{Kd_2}{[metal]}}\right)\left(\frac{1}{1 + \frac{[metal]}{Kd_1}}\right)$$

Ni-NTA binding assays

StrepII tagged forms of the MamO protease domain were purified as described for the tmFRET experiments with the omission of the F-5-M labeling and ion exchange steps. Proteins were diluted to ~15µM in Column Buffer (25mM Tris-HCl pH7.4, 300mM NaCl, 5mM Imidazole, 10% glycerol) and loaded onto a 0.5mL Ni-NTA column equilibrated in column buffer. The column was washed with 10 column volumes of Column Buffer and eluted with 2 column volumes Buffer C. Binding was assessed by separating the fractions on 12% SDS-PAGE and staining with Coomassie Blue.

Phylogenetic analysis

To understand the phylogeny of Mam proteases, we took a broad approach, characterizing their location within the trypsin-like protease family. The *Trypsin_2* Pfam (PF13365) was used to generate a hidden Markov model (HMM) that was used to search a database composed of protein sequences from ~2100 bacterial and archaeal genomes along with the Mam proteases (129). This search was performed using hmmscan (HMMER3.1b1, hmmer.janelia.org), retaining all hits with an e-value for the entire sequence less than 10^{-5}, identifying 6104 proteins. The hits were clustered using CD-HIT with a sequence similarity cut-off of 0.8, yielding 3431 sequences (130). These were aligned using MUSCLE (v3.8 .31) with the maxiters parameter set to 2 (131). The

resulting alignment was trimmed using Gblocks, using parameters appropriate for divergent datasets as described by Sassera *et al*(132, 133). This alignment was then used to generate a phylogeny using FastTree 2 (v2.1.7) with the default settings (134). However, this alignment contained only 6 informative positions. To improve the quality of the alignment, we iteratively removed long branches from the tree and regenerated the alignment. After removing 90 taxa over 4 iterations, we settled on an alignment with 18 conserved positions.

Two clear branches emerged from this analysis. One branch ('the Deg branch') contained DegP, DegQ, and DegS sequences from the γ –*Proteobacteria*, in addition to most of the MamE sequences. The bottom branch contained the YdgD sequence from *E. coli*. We extracted the sequences from the Deg branch (843 sequences) and generated an alignment (46 positions) and tree using the methods described above (Figure 2-14A). In this tree, the MamE sequences appeared to be closely related, but their phylogenetic relationships were ambiguous. Additionally, we realized that a subset of δ-proteobacterial trypsin-like sequences were falsely excluded from the MamE branch due to substitutions in the catalytic triad that obscured the phylogenetic signal in the short alignment. Additional phylogenetic testing confirmed the relationship of these 4 sequences to the canonical MamE sequences, so they were merged into the MamE branch.

We extracted the MamE sequences and aligned them using 5 DegS sequences from the γ –*Proteobacteria*, which appeared to be closely related based on the phylogeny in Figure 2-14B. This alignment was much larger (172 positions) and the resulting tree from FastTree 2 gave well-supported interior nodes for the MamE branch. We used two other phylogenetic methods to test this phylogeny. First, we used ProtTest 3.0 to select the best substitution matrix (in this case, the WAG model) and performed 100 independent inferences with 300 bootstraps in RAxML (135, 136). Secondly, we used the non-parametric Monte Carlo Markov Chain algorithm PhyloBayes 3 to generate a tree not based on prior assumptions about the site-specific evolution of the MamE sequences (137). A summary tree integrating the results from PhyloBayes and RAxML is depicted in Figure 2-14B. Additionally, we rooted the MamE branch to 5 closely-related sequences from the Clostridiales and this too strongly supported the phylogeny in Figure 2-14B according to both PhyloBayes and RAxML.

RESULTS

The catalytic triad in MamO is not required for magnetite nucleation

Trypsin-like proteases utilize a histidine-aspartate pair to deprotonate the hydroxyl group of a serine residue, providing a nucleophile for catalysis. MamO is unusual in that it contains a threonine instead of serine as the predicted nucleophile. To further clarify the role of MamO's unusual active site we focused our initial efforts on assessing the contribution of each putative catalytic residue to biomineralization. We performed these initial studies using the Δ*mamO*Δ*R9* strain (referred to as Δ*O*Δ*R9*) to avoid cross-complementation from *limO*. Consistent with our previous findings, individual H116A and D149A mutations in MamO severely reduced the cells' ability to turn in a magnetic field (Figure 2-2A). Surprisingly, a T225A mutation had no defects in magnetic response (Figure 2-2A). Using transmission electron microscopy (TEM) we confirmed that $mamO^{H116A}$ and $mamO^{D149A}$ cells have small magnetite crystals while $mamO^{T225A}$

crystals are indistinguishable from wild-type, mirroring the bulk magnetic response measurements (Figure 2-2B-D).

Given that the predicted nucleophile is dispensable for magnetosome formation, we found it curious that the other two catalytic triad mutations disrupted crystal maturation. Upon further examination, we found that the phenotypes associated with the $mamO^{H116A}$ and $mamO^{D149A}$ alleles were actually temperature dependent. Growing wild-type AMB-1 at room temperature instead of the standard 30°C did not dramatically alter the magnetic response. However, both the $mamO^{H116A}$ and $mamO^{D149A}$ alleles displayed improved complementation at the lower temperature. In particular, the $mamO^{D149A}$ mutant restored a nearly wild-type magnetic response to the Δ*O*Δ*R9* strain under these conditions (Figure 2-3). Our results suggest that protease activity from MamO is not required for biomineralization. None of the three predicted catalytic residues is required for magnetite nucleation, and, although two of the three contribute to crystal maturation, these effects are conditional suggesting that they are not central to the biomineralization process.

Proteolytic processing of biomineralization factors

During the course of our experiments, we examined the MamO variants by western blotting. Although no changes in overall protein abundance were present, we noticed that each MamO variant was proteolytically processed, having both a full-length and shorter form (Figure 2-5). This finding led us to examine whether other magnetosome proteins are similarly proteolyzed. We found that MamE and another biomineralization factor, MamP, are also proteolytic targets in AMB-1 cells by using antibodies targeted to each protein (Figure 2-4A). MamP is a *c*-type cytochrome, and its iron oxidase activity is required for the proper maturation of magnetite crystals (93, 124). Since these three biomineralization factors exist in both full-length and shorter forms, it is likely that proteolytic maturation plays a role in their function.

To examine the potential involvement of *mamE* and *mamO* in promoting these proteolytic events, we analyzed processing of MamO, MamE, and MamP in various genetic backgrounds. In both the Δ*O*Δ*R9* and Δ*E*Δ*limE* strains we observed only the full-length form of MamP. Similar analyses showed that MamO is required for the proteolysis of MamE and that MamE is required for the proteolysis of MamO. Thus, processing of each target requires both putative proteases (Figures 2-4A and 2-5A). Addition of the $mamE^{WT}$ allele restored processing of each target in the Δ*E*Δ*limE* strain, but the catalytically inactive form ($mamE^{PD}$) did not, suggesting that MamE participates in proteolysis directly. Surprisingly, both the $mamO^{WT}$ and $mamO^{H116A}$ alleles restored processing in the Δ*O*Δ*R9* strain (Figures 2-4A and 2-5). Therefore, the presence of MamO, but not its catalytic triad, appears to be required to promote the activity of MamE. While these results strongly suggest that MamE directly cleaves all three biomineralization factors, we cannot rule out the possibility that its activity contributes to a more complex targeting process.

Given that HtrA proteases are often regulated by the formation higher order oligomers, an attractive model could be that MamO activates MamE through an interaction involving both protease domains. To test this idea, we exploited the partial redundancy from genes in the R9 region. In addition to its N-terminal protease domain, MamO has a predicted TauE-like transporter domain on its C-terminus. *limO,* the partial duplication of *mamO* in R9, is 98% identical to the protease domain of *mamO*, but does not contain a C-terminal TauE domain

(Figure 2-6). We confirmed that alleles with individual point mutations in the protease domain that reduce bioimineralization in the Δ*O*Δ*R9* background do not have defects in the Δ*mamO* strain, reinforcing that LimO is a functionally redundant copy of the MamO protease domain (Figure 2-6). In contrast, processing of MamE and MamP is disrupted in both the Δ*mamO* and Δ*O*Δ*R9* strains, showing that a functional protease domain is insufficient to activate MamE and that activation requires the TauE domain (Figure 2-4B).

Structural analysis of the MamO protease domain

In the trypsin family, loop L1 contains both the nucleophilic serine and oxyanion hole, required for creating the acyl-enzyme intermediate and stabilizing the oxyanion, respectively (Figure 2-7) (138). In addition to the threonine substitution, the entire L1 loop in MamO differs significantly from the trypsin family consensus, once again suggesting that MamO might not be capable of protease activity (Figure 2-7). To explore this possibility further, we determined the crystal structure of the protease domain (Table 2-1). It crystallizes as a monomer with the chymotrypsin fold and the catalytic residues properly placed (Figure 2-7).

Loop L1 of MamO adopts the inactive conformation seen in many HtrA proteases in which the main chain carbonyl of residue 192 (chymotrypsin numbering; W222 in MamO) prevents access to the oxyanion hole (Figure 2-8A). In other HtrAs, the inactive form is thought to be in equilibrium with an active conformation where the main chain flips approximately 180° opening the oxyanion hole (Figure 2-8A) (139, 140). Switching to the active state forces residue 193 into a configuration that is strongly disfavored for non-glycine residues. Although glycine is highly conserved at position 193 in the trypsin family and critical for catalysis, MamO contains a glutamate (E223) at this position (Figure 2-7B)(141). Therefore, the active configuration of MamO would contain a strong steric clash between the E223 side chain and the main chain carbonyl of W222.

To illustrate this steric constraint, we examined a set of trypsin-like protease structures and analyzed the configuration of residue 193. We plotted ϕ and Ψ values for this position on a Ramachandran plot along with the favored and allowed geometries for glycine and non-glycine residues (Table 2-2) (142). As expected, the main-chain torsion angles at position 193 form two groups, distinguished by an approximately180° shift in ϕ angle. The groups correspond to active and inactive forms of loop L1. The active configuration is strongly disfavored unless glycine occupies position 193, indicating that the E223 side-chain in MamO likely prevents the formation of an oxyanion hole (Figure 2-8B).

While refining the MamO structure, we observed a peptide bound near the predicted peptide-binding groove (Figure 2-9A). We suspect its source is the flexible N-terminal region from a neighboring MamO in the crystal that is not built into the model. However, due to the modest resolution, we cannot confirm the sequence. Despite this uncertainty, the peptide has an interesting mode of binding. Similar to other trypsin-like proteases, the peptide enters the binding cleft parallel to loop L2, splitting the two β-barrels of the chymotrypsin fold. However, the bulky side-chain of W222 in MamO seems to block the exit path between loops LA and LD, forcing the peptide away from the catalytic center (Figure 2-9B-D). Overall, the structural features of the L1 loop appear incompatible with protease activity: E223 provides an energetic barrier to catalysis while W222 serves as a physical barrier to productive substrate binding.

Direct transition metal binding by MamO

Given the results of the genetic studies and structural analysis, we conclude that MamO does not act as a protease during biomineralization. This forced us to consider the possibility that the protease domain promotes magnetite formation using a function not predicted from its primary sequence. While purifying MamO, we serendipitously observed that it consistently bound to immobilized metal affinity columns. Knowing that mutations in MamO cause defects in AMB-1's ability to transform iron, a metal, into magnetite, we speculated that direct metal binding played a role in biomineralization.

Although the MamO crystals grow in acidic conditions that disfavor metal binding, they could be soaked at pH 8.0 without affecting diffraction. We solved another structure of the protease domain using crystals soaked in $NiCl_2$ at pH 8.0 and identified a metal binding site. Overall, the conformation of the protease domain is highly similar to the original structure (root-mean-square deviation of 0.17Å over 184 residues), and it contains the unidentified peptide. Additionally, a single Ni^{2+} ion binds between loop LC and helix 2, with H148 and H263 directly coordinating the metal (Figure 2-10A-C). We confirmed the placement of the ligand at this site using single wavelength anomalous dispersion (SAD) data collected at the Ni absorption edge (Figure 2-10A and B). MamO's di-histidine motif is highly reminiscent of the Zn^{2+} binding site in another trypsin-like protease, kallikrein-3, where Zn^{2+}attenuates protease activity by altering the position of catalytic triad residues H57 and D102 (Figure 2-11) (143, 144).

To confirm the putative metal binding site from the MamO structure, we used transition metal ion Förster Resonance Energy Transfer (tmFRET) to assay binding in solution (145). This technique measures fluorescence quenching of a cysteine-conjugated fluorophore by a metal bound at a nearby site. Guided by the structure, we targeted Q258 because of its optimal geometry relative to the metal. Adding a number of transition metals, including iron, to a purified Q258C mutant protease domain that was labeled with fluorescein-5-maleimide caused strong fluorescence quenching (Figures 2-12 and 2-13). Although we expect Fe^{2+} to be the physiological ligand, its propensity to oxidize in ambient atmosphere added significant error to the measurements. Instead, the quenching properties and resistance to oxidation of Ni^{2+} were most suitable for detailed analysis. MamO bound to Ni^{2+} with 2.5μM affinity, compared to 1.1μM in an H148A/H263A mutant. Additionally, the FRET efficiency was significantly lower in the mutant protein, demonstrating that disrupting these residues changes the metal binding properties of MamO (Figures 2-10D and 2-13).

Although the H148A/H263A mutant displays altered behavior in the tmFRET assay, it retains the ability to bind metals (Figure 2-13). We could not identify any other metal ions in our Ni^{2+} soaked crystals, but the two histidines we identified did not appear to be the only metal binding residues in MamO. Consistent with this, both the wild-type and H148A/H1263A forms of MamO bound Ni-NTA resin, confirming that despite its altered binding geometry the mutant still binds to metal (Figure 2-13). Taken together, our biochemical and structural investigations show that MamO binds to transition metal ions using H148 and H263, but that metal binding is not restricted to this motif.

Because disrupting H148 and H263 altered the metal binding behavior *in vitro,* we predicted that these residues were important for MamO dependent biomineralization. Indeed, Δ*O*Δ*R9* strains with $mamO^{H148A}$ or $mamO^{H263A}$ alleles had no magnetic response and failed to produce electron dense particles, implying that proper metal binding is required to initiate biomineralization (Figure 2-10E and F). Both mutants displayed normal stability and proteolytic processing, and the biomineralization defects were not conditional, as lowering the growth temperature did not allow for a magnetic response (Figures 2-3 and 2-5). Additionally, both alleles restored a magnetic response in the single Δ*mamO* background, showing that *limO* can provide the required metal binding activity independent of co-translation with the TauE domain (Figure 2-6). These are the most disruptive point mutations we have observed in MamO, and they completely recapitulate the phenotype of a *mamO* deletion. We conclude that H148 and H263 contribute to a metal binding function that is required for magnetite nucleation *in vivo.*

Convergent evolution of pseudo-proteases in magnetotactic bacteria

While the finding that MamO has lost its protease activity to become a metal binding protein in AMB-1 was quite intriguing, we wanted to know whether this mechanism was conserved in other organisms. Due to the fact that magnetotactic *Nitrospirae* and *Omnitrophica* have not been isolated in culture, we focused our analysis on the *Proteobacteria* for which numerous representatives are available in pure-culture. Examining available whole-genome sequences indicated that all magnetotactic strains from the α-, δ- and γ-*Proteobacteria* contain two predicted trypsin-like proteases in their MAIs (Table 2-3). We attempted to understand the evolutionary history of these proteins by including them in a large phylogenetic tree of the bacterial trypsin-2 superfamily (Materials and Methods). Within this tree, the MamO sequences from each magnetotactic α-*Proteobacterium* form a distinct, monophyletic clade (Figure 2-14A). Each protein has a degenerate catalytic triad, a non-glycine residue at position 193 and a bulky tryptophan in the L1 loop. The metal binding positions in LC and helix 2 also appear conserved. Although one strain has an aspartate at position 263, aspartates are also common metal coordinating residues, and we predict that this H-D motif can also participate in metal binding. We conclude that the MamO family evolved specifically in α-*Proteobacterial* magnetotactic organisms to be metal binding pseudo-proteases (Figure 2-15).

In addition to the MamO proteins, we identified a second magnetosome-specific clade in the trypsin tree (Figure 2-14A). We named this group the MamE clade because it contains the MamE sequence from the MAI of each representative of the α-*Proteobacteria.* This group also features two predicted trypsin-like proteases from the MAI of each species of the δ- and γ-*Proteobacteria,* re-named MamE1 and MamE2 here (Table 2-3). The sequence phylogeny indicates that the δ- and γ -*Proteobacteria* both experienced recent duplications of their MamE's independently. In fact, there appears to have been two rounds of duplication in the δ- family. Strikingly, each duplication event has led to a degenerate catalytic triad in one of the sequences, showing that duplication and loss of function occurred three separate times in the MamE family (Figures 2-14B and 2-16). While these inactive MamEs do not have the di-histidine motif identified in MamO, we cannot rule out the possibility that they bind metals by another mechanism. Regardless, our results imply the existence of selective pressure for pairing active and inactive proteases as each major clade of magnetotactic bacteria has evolved this feature independently.

DISCUSSION

Magnetotactic bacteria control the growth of their associated magnetite crystals with a level of precision that cannot be replicated *in vitro*. The molecular details of how they perform this task can reveal novel bioinorganic interfaces and be exploited for improved synthesis of nanomaterials. Genetic analysis has shown that magnetite biomineralization is surprisingly complex. It requires over 15 factors in AMB-1, nearly all of which are predicted integral membrane proteins (19, 89, 102). A subset of these is required for the initial crystallization of iron within the magnetosome compartment (47). While the key players for this step are known, their biochemical functions have only been inferred from sequence homology, leaving the mechanism of how magnetite biosynthesis begins a mystery.

Here we examined two magnetite nucleation factors, the putative HtrA proteases MamE and MamO. We find that the presence of both MamE and MamO is required for the proteolysis of three biomineralization factors, MamE, MamO and MamP. These events depend on an intact catalytic triad from MamE but not MamO, indicating that MamO activates MamE in a non-catalytic manner. Thus far, we have not been able to detect a physical interaction between MamE and MamO, but we have found that the C-terminal TauE-like transporter domain of MamO is required for activation. The putative ion transport activity of this domain could be the feature that promotes proteolysis. This model is attractive because it does not require a direct interaction between the two proteins. Indirect evidence suggests that TauE family proteins transport sulfite or sulfur containing organic ions, leading us to speculate that the concentrations of specific solutes in the magnetosome might control MamE's activity (79, 146).

Separately, we discovered that a metal-binding function in the protease domain of MamO is required for the initiation of magnetite biomineralization. In our structure, H148 and H263 directly coordinate a single metal ion. Disrupting these residues alters the binding behavior in a modified tmFRET assay, but the effect is unusual in that it lowers the FRET efficiency while slightly increasing the overall affinity for metal. Using the Förster equation and reported radius for Ni^{2+}, we calculated that the fluorophore to metal distance changes from 12.8Å in wild-type to 15.2Å in the H148A/H263A mutant, suggesting that the binding geometry changes in a way that allows metal binding in the same vicinity (128). We favor the explanation that the di-histidine motif identified here is part of a more complex metal coordinating network. Our soaking strategy may have missed additional sites that are inaccessible in our crystal form, and separate attempts at characterizing a fully metal-bound state using co-crystallization have been unsuccessful. Nevertheless, we identified di-histidine motif using the structure that contributes to an unexpected transition metal-binding activity in MamO.

Despite the presence of other binding features, metal binding through H148 and H263 is absolutely required *in vivo* as disruption of either residue completely abolishes magnetite formation. Our structural analysis shows that MamO has lost its ability to perform proteolysis altogether, supporting the idea that metal binding is now the central function of the protease domain. Consistent with this, T225, the predicted catalytic nucleophile, is completely dispensable for biomineralization. Though disrupting H116 and D149 in the predicted catalytic triad causes conditional crystal maturation defects, magnetite nucleation is not affected. Interestingly, H116 and D149 participate in a hydrogen bond on the opposite face of loop LC from the H148/H263 metal binding motif, suggesting that the conditional phenotypes could be

due to temperature-dependent flexibility near the metal binding site (Figure 2-11). A potential link between the two motifs is consistent with the reported inhibition of protease activity through a highly analogous metal binding site in the kallikrein family that rearranges the H-D catalytic pair (Figure 2-11).

While templating of magnetite growth via an interaction between biomineralization factors and the mineral surface has been proposed, our findings with MamO emphasize that direct interactions with individual solute ions also play a role (41, 89, 96). One of the most fascinating aspects of MamO's metal ion interaction is that the H148A and H263A forms of MamO maintain the ability to bind metals but cannot support any magnetite biosynthesis *in vivo*. It appears that binding is insufficient and that the precise coordination geometry must be maintained, leading us to speculate that MamO directly promotes nucleation by guiding individual iron atoms into the magnetite lattice. This model is consistent with the phenotypes observed *in vivo,* the modest binding affinity and the surface exposed nature of the simple di-histidine motif. Additionally, it agrees with topological predictions for MamO placing the protease domain in periplasm, which is continuous with the magnetosome lumen in AMB-1 (56). More broadly, our results define an unexpected mechanism for MamO in biomineralization. It appears to have lost the ability to perform serine protease activity and instead performs two non-catalytic functions: direct metal binding to promote magnetite nucleation and activation of MamE's proteolytic activity (Figure 2-17A).

In addition to the surprising mechanism for MamO in AMB-1, we uncovered a fascinating evolutionary expansion of the trypsin family within magnetotactic bacteria. Our analysis suggests that the ancestral MAI contained a single trypsin-like protease homologous to MamE. The δ- and γ –*Proteobacteria* experienced independent duplications of this ancestral enzyme, while the α-*Proteobacteria* appear to have acquired a second, distantly related trypsin-like protease. Despite these different origins, having two redundant proteases seems to have allowed one copy to lose its catalytic ability in all three clades (Figure 2-17B). In α-*Proteobacteria,* MamO specialized to promote biomineralization through the two non-catalytic activities identified here. While the pseudo-proteases in the other clades remain uninvestigated, the fact that inactive copies are retained strongly suggests that they also play important non-catalytic roles. The pathway that led to convergent evolution of pseudo-proteases in magnetotactic organisms highlights the critical role duplication and redundancy play in facilitating diversification of protein function (147).

Perhaps more intriguing is the fact that MamO's metal binding motif is placed at the same site on the chymotrypsin fold as the highly analogous zinc-binding site seen in the distantly related kallikreins (143, 144). This hints toward the possibility that the ability to bind metals may be a latent biochemical function carried within the fold. Such activities are absent in specific evolutionary states of a protein but can quickly surface under selective pressure (148, 149). Consistent with this, trypsin-like proteases utilize catalytic residues on loops that are well separated from the core, a property termed fold polarity that correlates with the capacity for functional diversification (150, 151). Perhaps, neofunctionalization of the trypsin scaffold within magnetotactic organisms is due to an inherent stability and adaptability in the fold that makes it a useful building block for biochemical innovation.

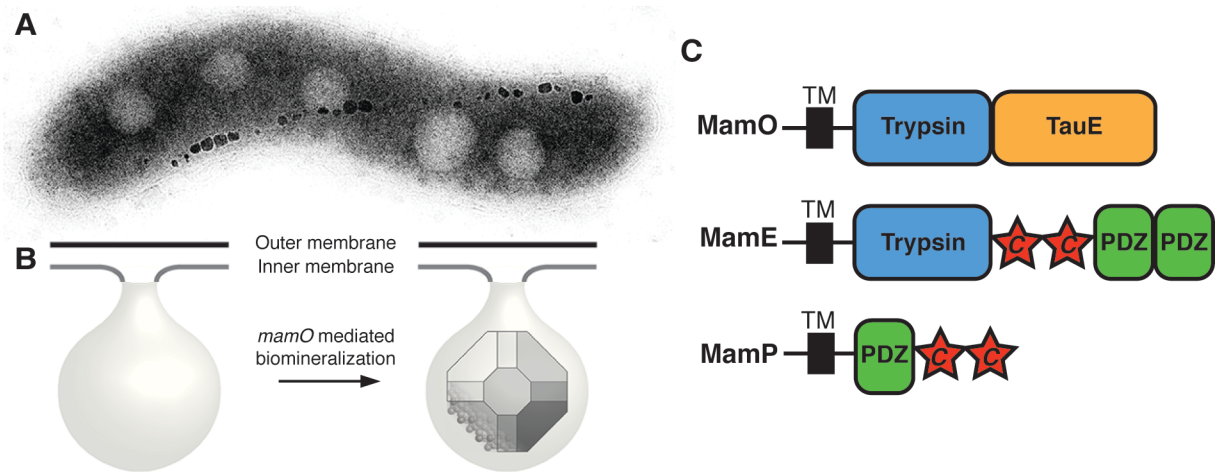


Figure 2-1. MamO promotes the nucleation of magnetosome crystals. (a) TEM micrograph of a wild-type AMB-1 cell. The electron dense particles make up a magnetosome chain. (b) Cellular organization of the magnetosome compartment. MamO promotes the nucleation of magnetite within inner membrane invaginations. (c) Domain structure of the three biomineralization factors discussed in the text. "*c*" represents a CXXCH *c*-type cytochrome motif.

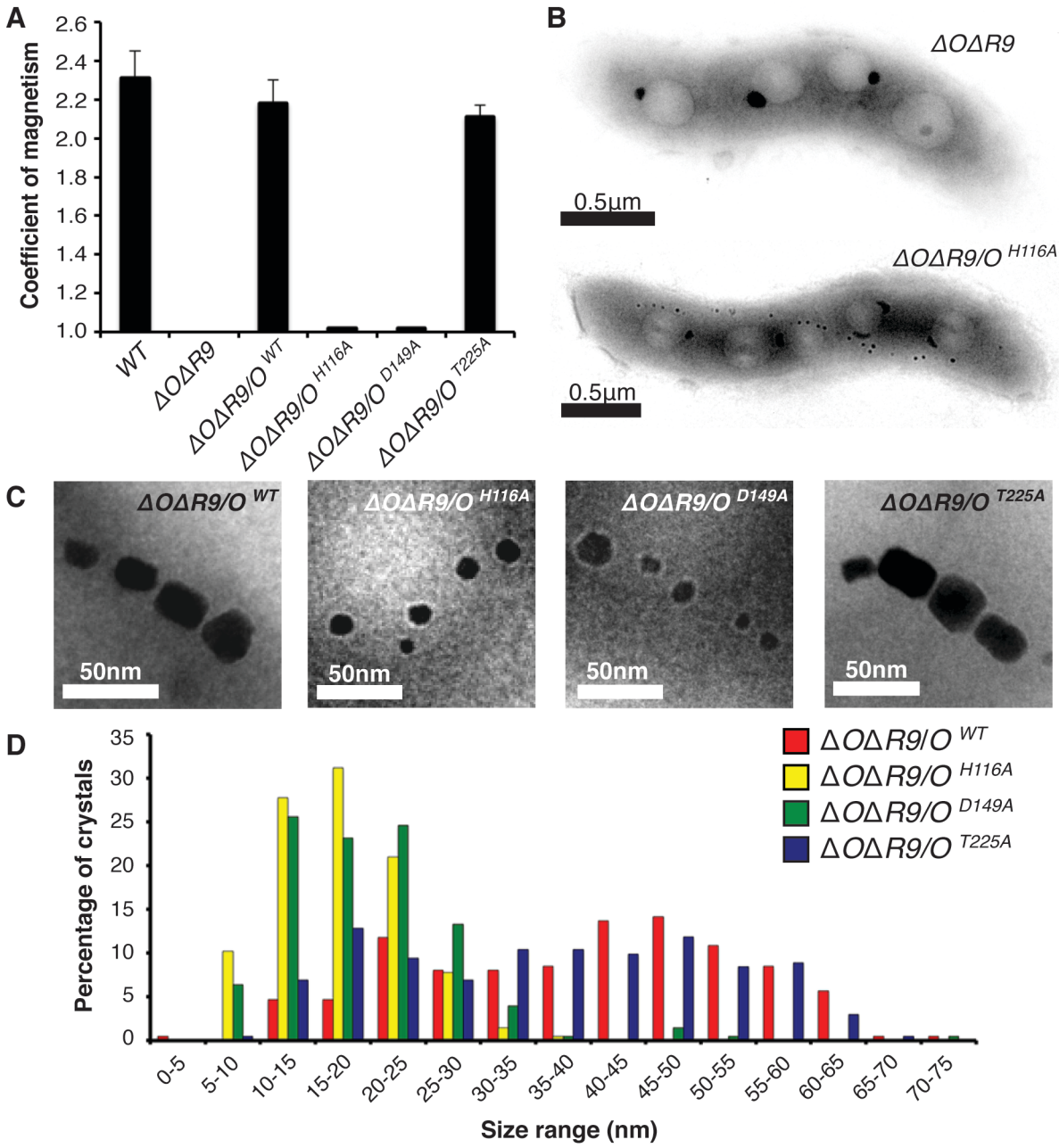


Figure 2-2. Dissection of the MamO catalytic triad. (a) Magnetic response of cultures with the indicated *mamO* alleles. The coefficient of magnetism is an optical density based method assessing cells' ability to turn in a magnetic field. Biological replicates represent independent cultures of each strain. Each measurement represents the average and standard deviation from three independent experiments. (b) TEM of whole AMB-1 cells with the indicated genetic backgrounds. (c) Magnetite crystals from cells with the indicated *mamO* alleles. (d) Crystal size distributions for the indicated *mamO* alleles as assessed by TEM. n > 200 for each strain.

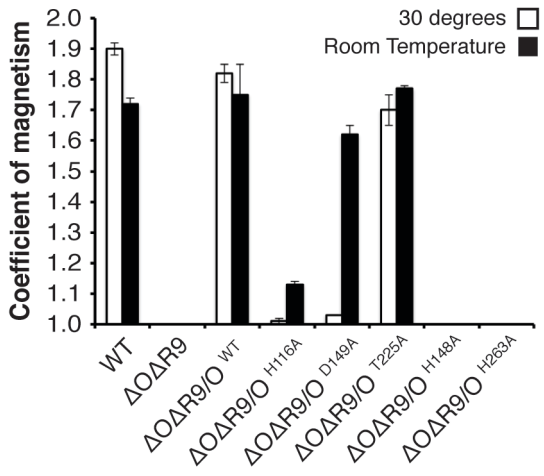


Figure 2-3. Temperature dependence of magnetic response for *mamO* alleles in this study.

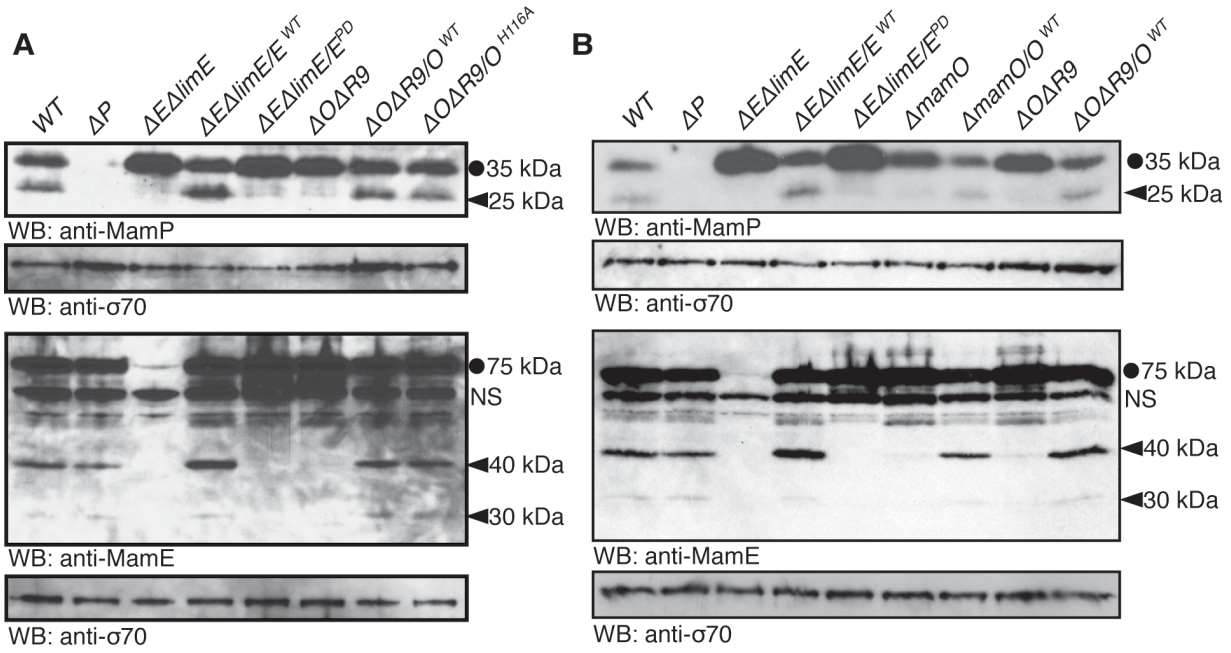


Figure 2-4. Genetic requirements for proteolytic processing of biomineralization factors. Full-length proteins are marked with a circle and proteolytic fragments with a carat. *mamE*^{PD} refers to the previously described allele with all three catalytic triad residues mutated to alanine. A nonspecific interaction with the anti-MamE antibody is marked with "NS". (**A**) Proteolysis of MamE and MamP depends on both MamE and MamO, The MamE active site is required, but the MamO active site is dispensable. (**B**) Efficient proteolysis of MamE and MamP requires the TauE domain of MamO. In both the Δ*O*Δ*R9* strain and the Δ*mamO* strain (which contains *limO*) proteolysis of the two targets is minimal.

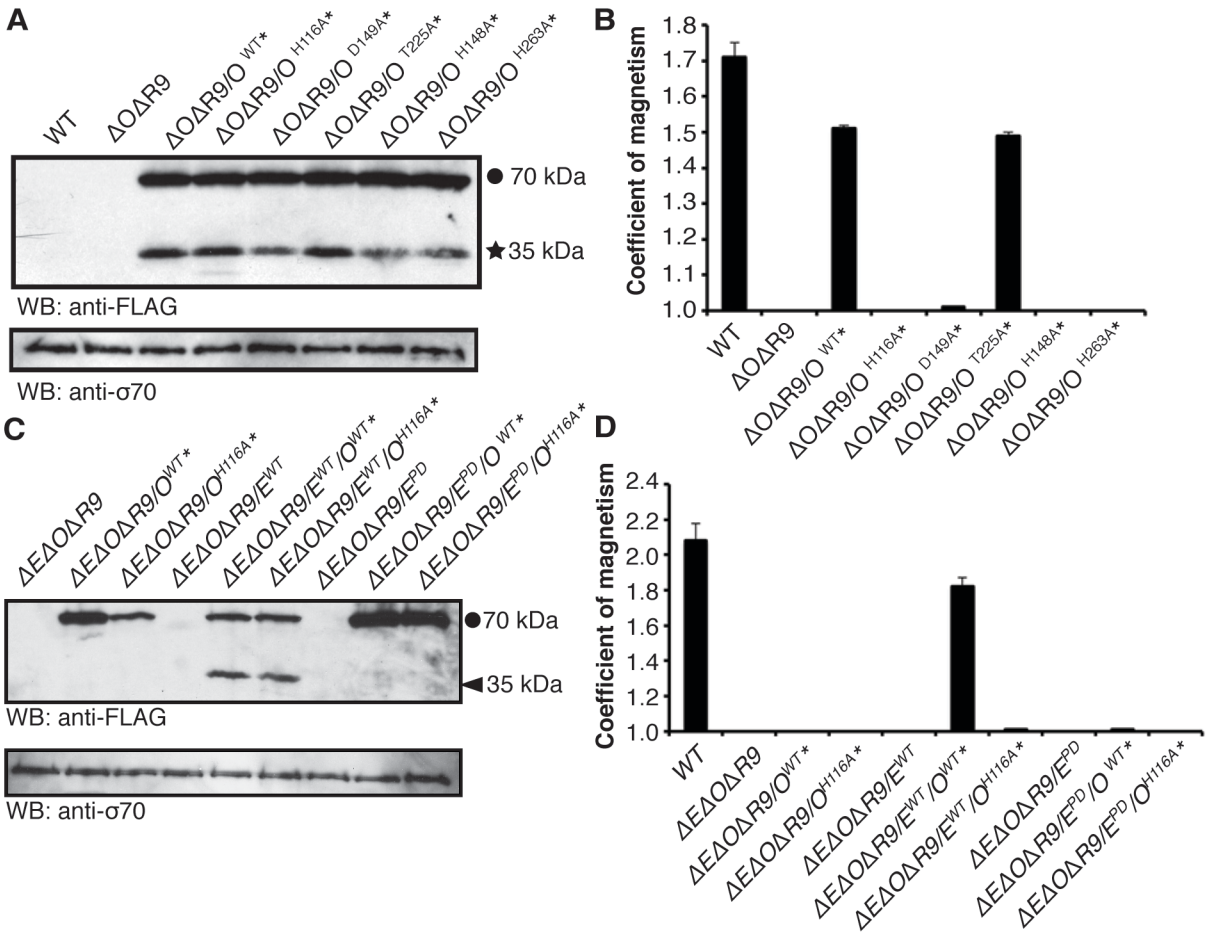


Figure 2-5. Characterization of strains from Figure 2-4. Stars indicate N-terminally 3xFLAG-tagged alleles throughout the figure. In the histograms, each measurement represents the average of three biological replicates. Error bars represent the standard deviation of the replicates. (a) Processing of MamO alleles used in this study in the Δ*O*Δ*R9* background. (b) Magnetic response of the strains in Panel a. (c) Proteolytic processing of MamO requires the MamE active site. (d) Magnetic response of the strains from Panel c.

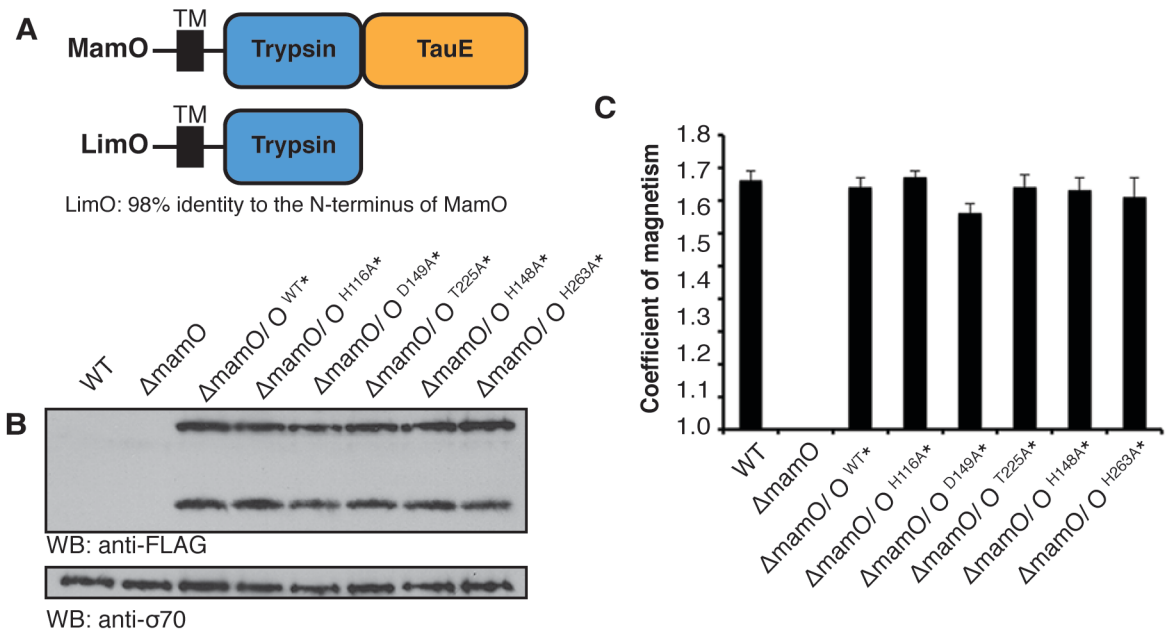


Figure 2-6. Cross-complementation of MamO protease domain mutants by *LimO*. (a) *limO* is contained within a partially duplicated region of the *mamAB* cluster termed R9. While *mamO* is predicted to have a trypsin like-protease domain and a TauE-like transporter domain, *limO* has only a predicted trypsin-like domain with 98% identity to the N-terminus of *mamO*. *LimO* contains all of the critical residues identified in MamO in this study. (b) MamO alleles are proteolytically processed identically in the single Δ*mamO* strain as they are in the Δ*O*Δ*R9* background. (c) All of the *mamO* alleles examined in this work restore wild-type biomineralization in the single Δ*mamO* background, showing that *limO* encodes a fully functional copy of the *mamO* protease domain.

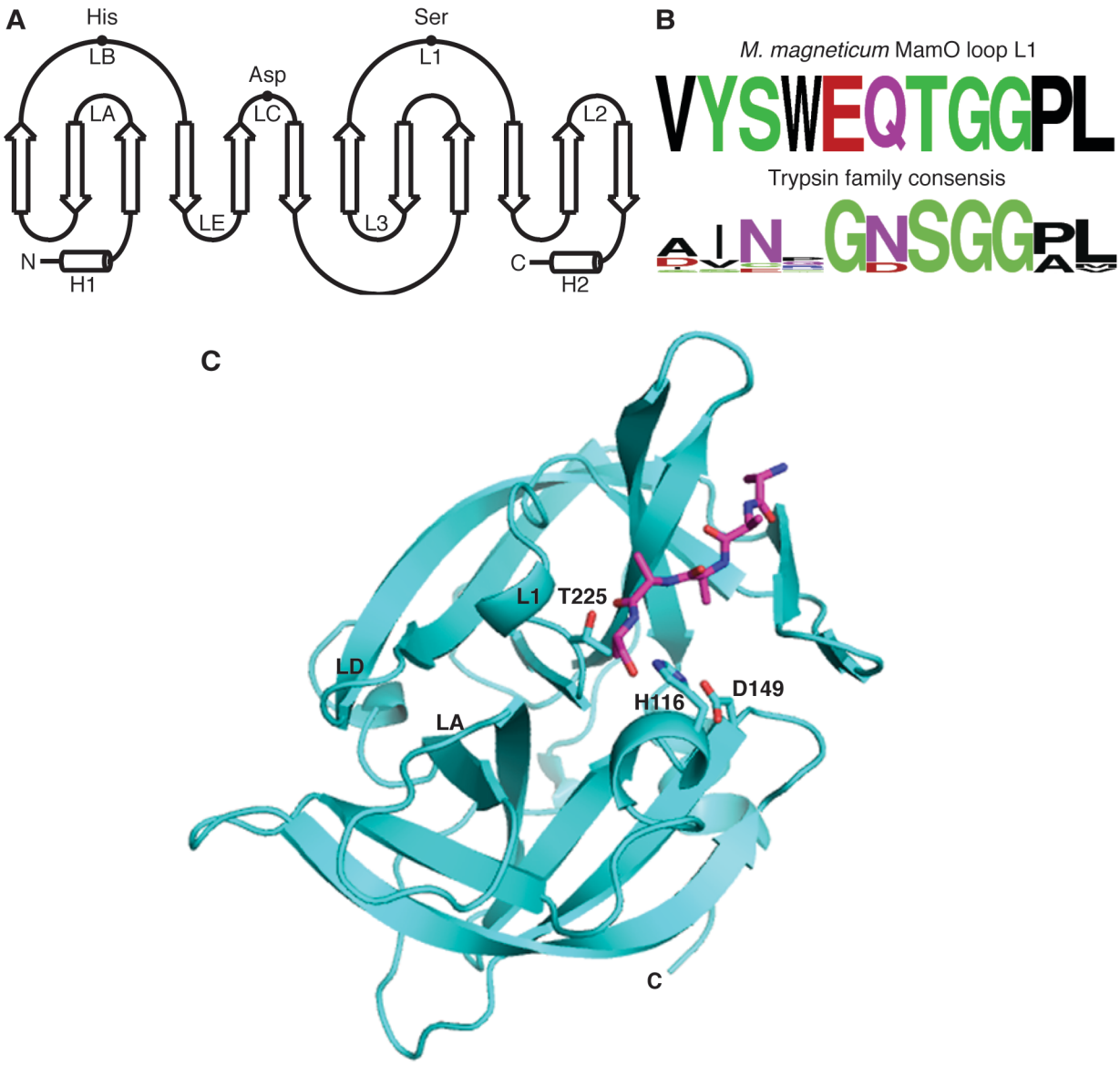


Figure 2-7. Structure of the MamO protease domain. (a) Schematic of the chymotrypsin fold with the loops and catalytic residues indicated. (b) Comparison of the L1 loop in MamO to the trypsin family consensus. (c) Overall structure of the MamO protease domain solved to 2.6Å. The catalytic residues and bound peptides are show in stick representation.

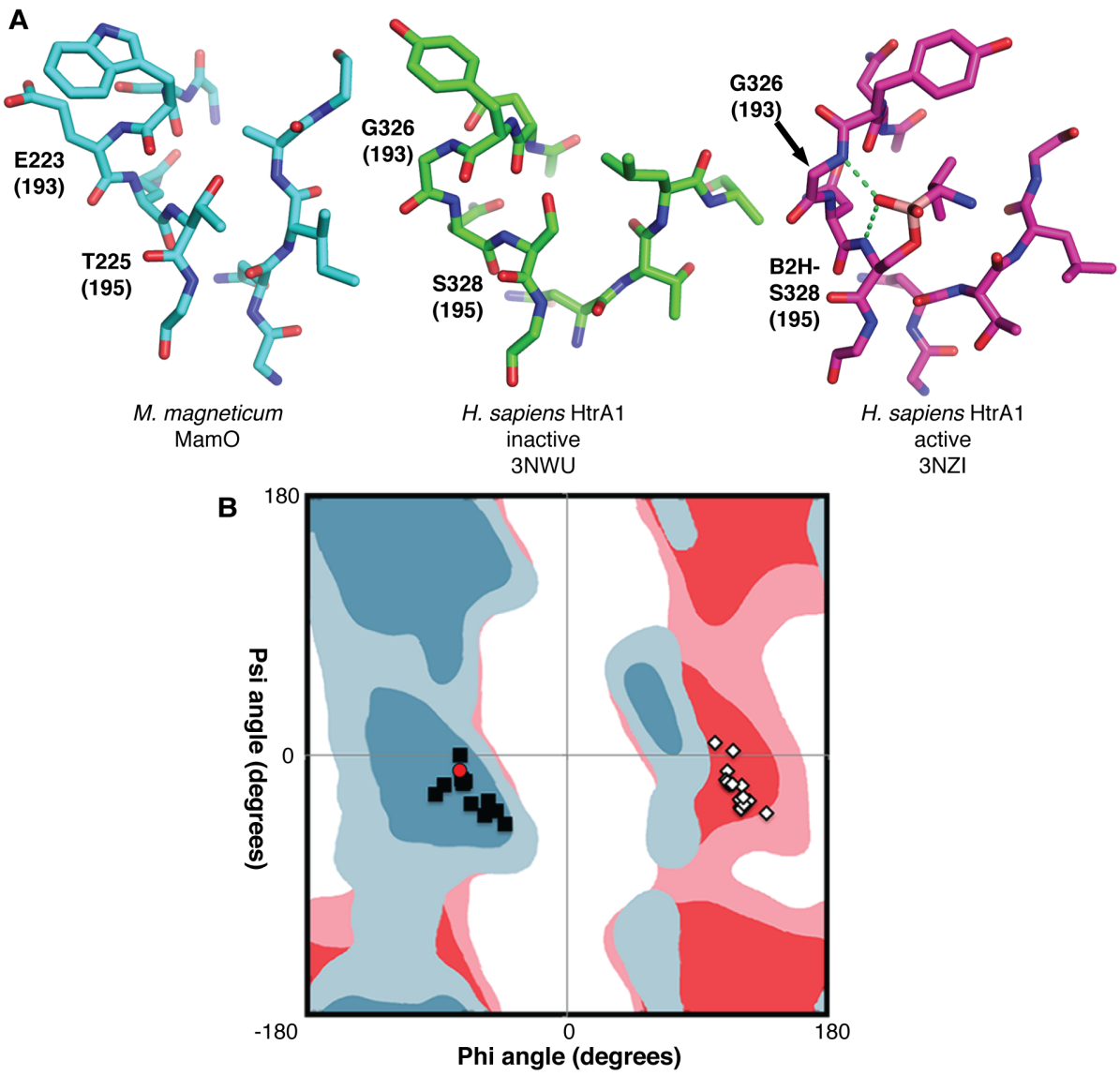


Figure 2-8. Steric constraints on the MamO active site. (a) Comparison of active site structures, showing MamO in the inactive conformation. Each residue's chymotrypsin numbering position is in parentheses. Dashes represent hydrogen bonds contributed by the oxyanion hole. (b) Ramachandran plot showing favored (dark shades) and allowed (light shades) geometries for non-glycine residues (blue) and glycine (red). The configuration at residue 193 for a set of trypsin-like structures is plotted. Black squares: inactive conformation; white diamonds: active conformation; red circle: MamO. The active conformation is disallowed for MamO.

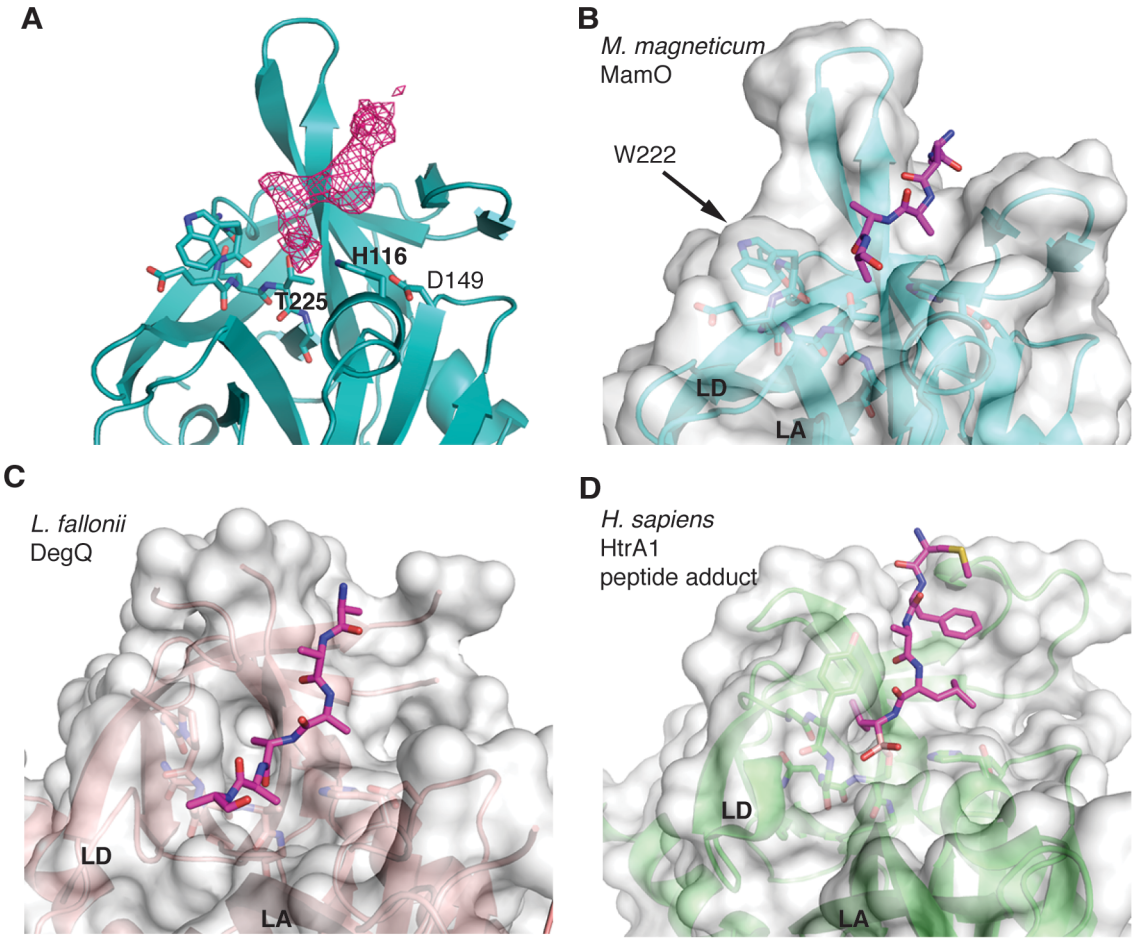


Figure 2-9. W222 in MamO forces the bound peptide away from the active site. (a) $F_o - F_c$ omit map contoured at 3σ showing missing density for the peptide in the MamO crystal structure. (b-d) Comparison of peptide binding pockets in HtrA proteases. W222 in MamO blocks the normal exit path between loops LA and LD and pushes the peptide away from the active site. PDB codes for each panel are as follows b: 5HM9; c: 3PV3; d: 3NZI. Loops LA and LD are marked.

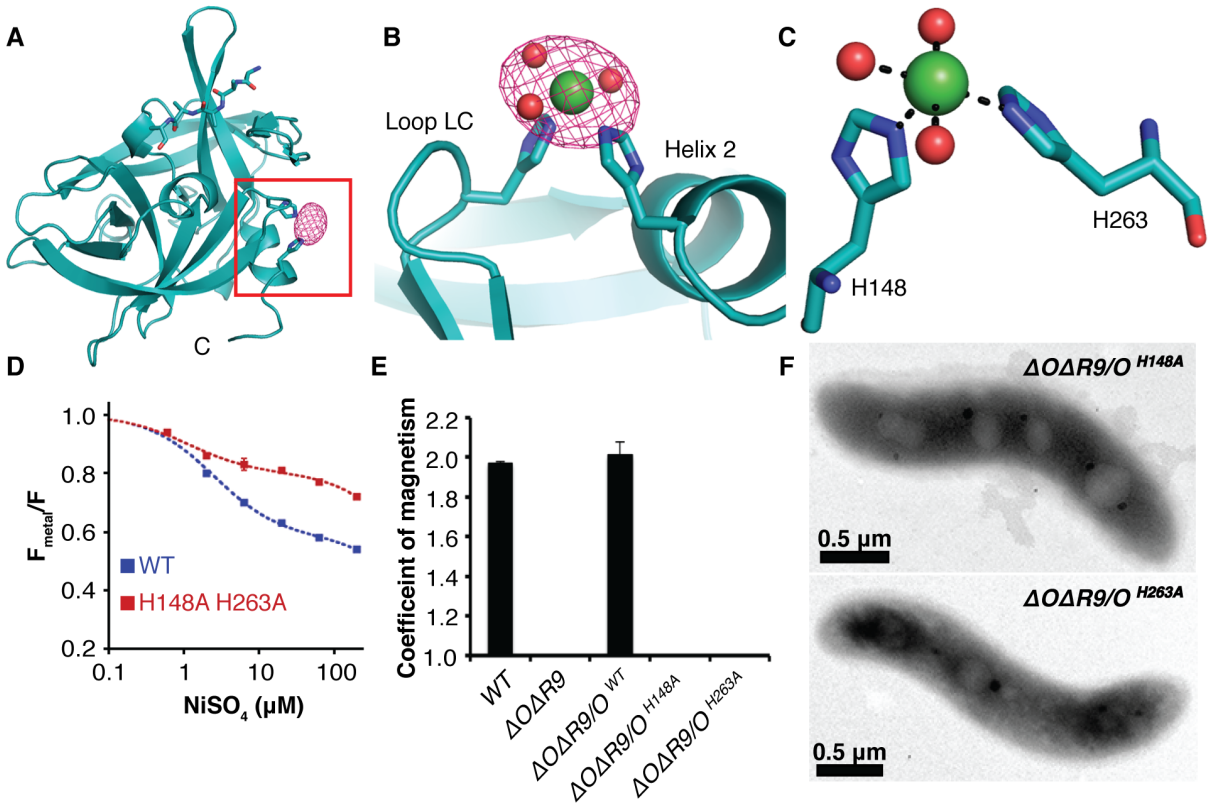


Figure 2-10. Characterization of the MamO metal binding site. (a) Anomalous map contoured at 5σ showing the placement of Ni^{2+} in MamO. (b) Ni^{2+} ion bound between loop LD and helix 2. (c) H148, H263, and three ordered water molecules participate in metal binding. (d) tmFRET analysis of Ni^{2+} binding by MamO labeled at Q258C. Each measurement represents the average from four replicates. Error bars represent the standard deviation of the replicates. (e) Magnetic response of strains with disrupted metal coordination sites. Error bars represent the standard deviation from three cultures. Each measurement represents the average of three biological replicates. Error bars represent the standard deviation of the replicates. (f) TEM analysis showing that the *mamO*^{H148A} and *mamO*^{H263A} strains lack detectable minerals.

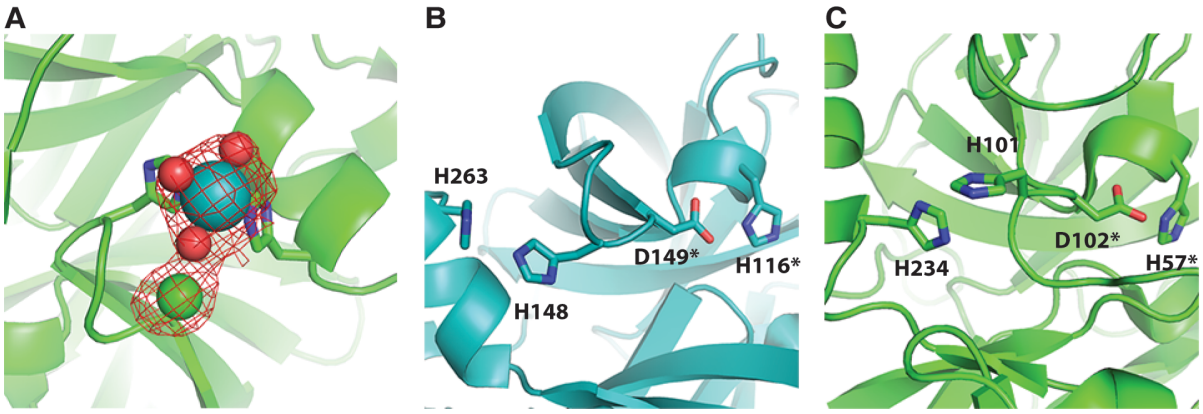


Figure 2-11. Similarity of the MamO metal binding site to equine kallikrein-3. (a) F_O-F_C omit map showing the bound Ni^{2+} ion in MamO. Blue: Ni^{2+}; red: H_2O; green: Cl^-. (b, c) Comparison of metal binding sites in MamO and equine kallikrein-3. Coordinates of the zinc-bound structure reported in Carvahlo et al.(143) were not deposited in the PDB.

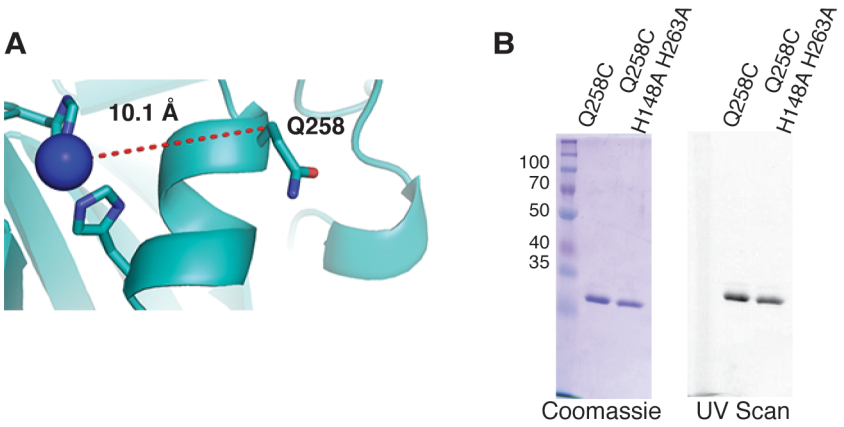
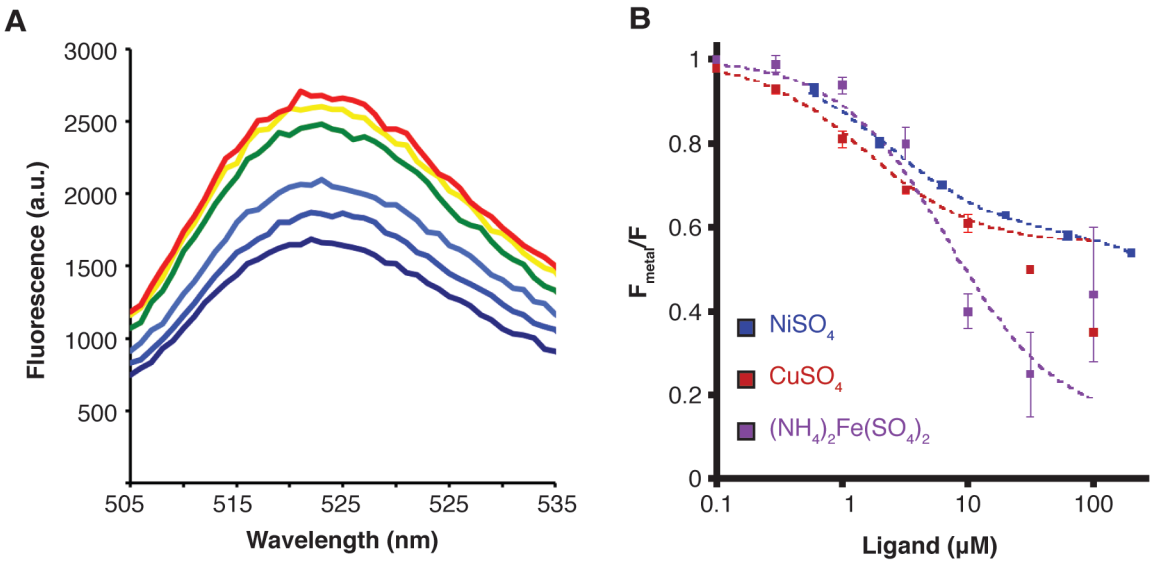


Figure 2-12. Fluorescein-5-maleimide labeling of MamO^{Q258C}. (a) The fluorescent labeling site in MamO was chosen based on the optimum FRET distance from Taraska et al. (145) (b) Purification and fluoresceint labeling of MamO^{Q258C} and MamO$^{Q258C\ H148A\ H263A}$.



C

Ligand	Protein	Kd_2(μM)	E_{FRET}(%)	Kd_1(mM)
$NiSO_4$	Q258C	2.5 ± 0.3	41.6 ± 1.7	2.2 ± 1.1
$(NH_4)_2Fe(II)(SO_4)_2$	Q258C	7.1 ± 0.0	86.6 ± 21.0	2.53 x 10^{14}
$CuSO_4$	Q258C	1.6 ± 0.0	44.0 ± 21.0	1.63 x 10^{21}
$NiSO_4$	Q258C H148A H263A	1.1 ± 0.2	20.0 ± 0.1	1.7 ± 0.3

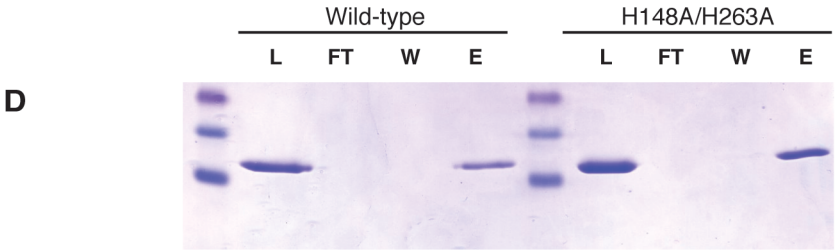


Figure 2-13. tmFRET analysis of metal binding. (a) Fluorescence quenching of MamO Q258C labeled with fluorescein-5-maleimide in the presence of increasing concentrations of $NiSO_4$. (b) Binding of various transition metals to labeled MamO. Error bars represent the standard deviation from four independent measurements. The dotted lines are fits to the binding equation described in Materials and Methods. (c) Binding constants from tmFRET experiments. (d) Ni-NTA affinity assays with purified protease domains.

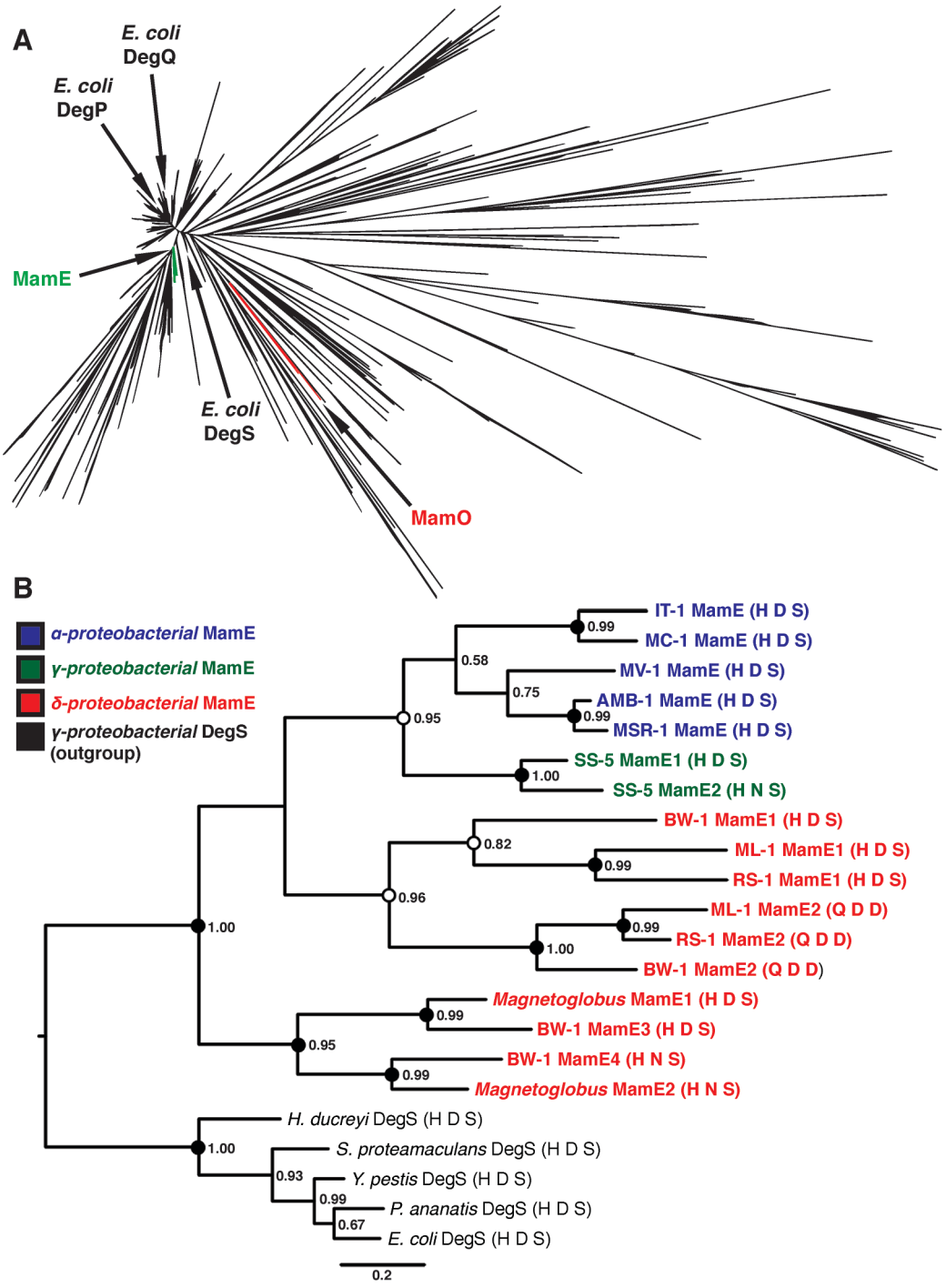


Figure 2-14. Phylogenetic analysis of magnetotactic trypsin-like proteins. (a) Phylogeny of the "Deg" branch created from the trimmed Trypsin-2 alignment. MamE and MamO clades and the three *E. coli* HtrA proteases are marked. (b) Phylogeny of the MamE clade of HtrA proteases. Numbers represent the posterior probability determined by PhyloBayes. Circles represent the degree of support from 300 bootstrap replicates in RAxML. Black: >90% support; white: >80% support. MamE sequences are colored based on the class of the associated organism, and the catalytic triad residues are shown in parentheses after each name.

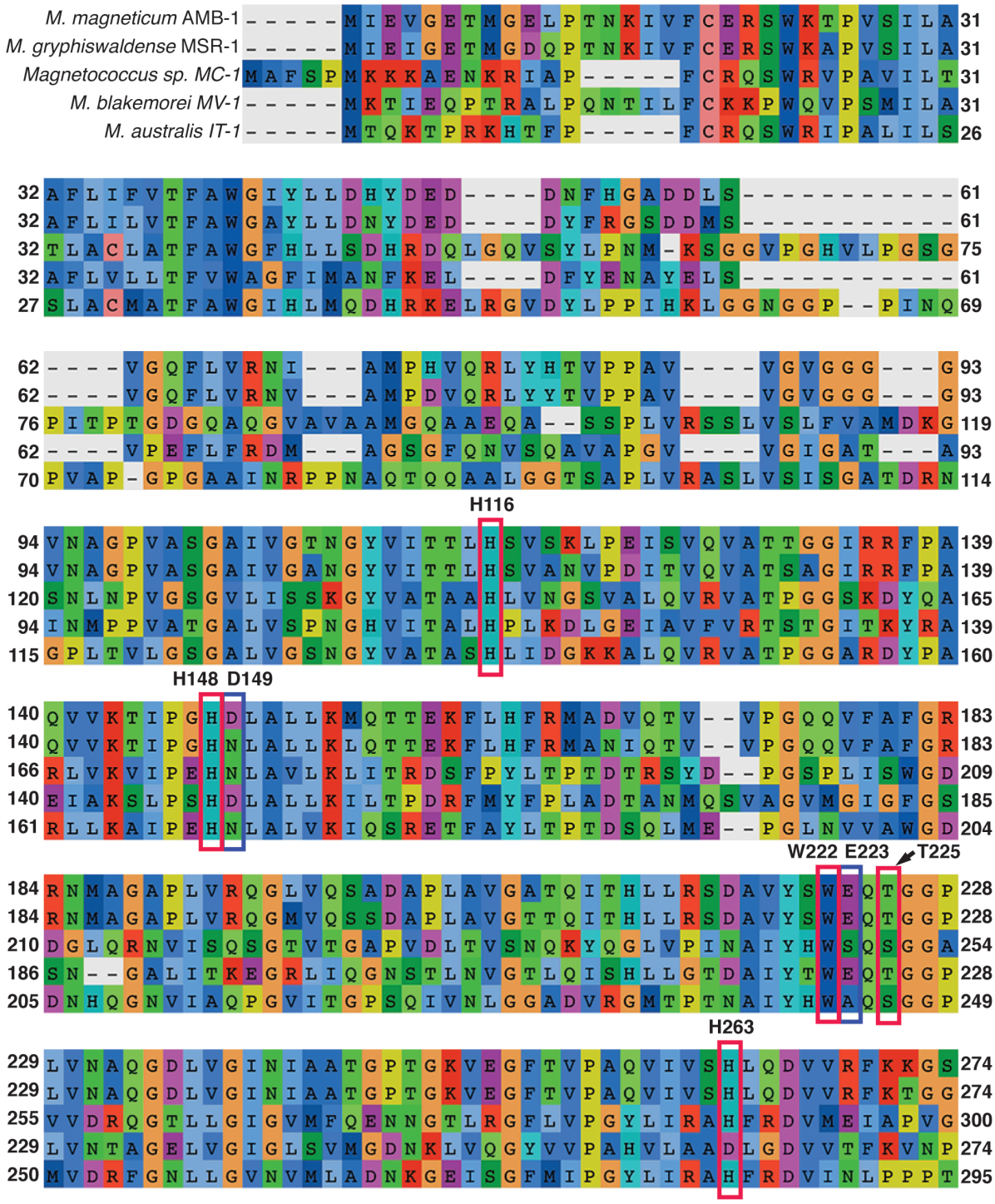


Figure 2-15. Alignment of the MamO family. The conservation of critical residues discussed in the text is indicated with colored boxes.

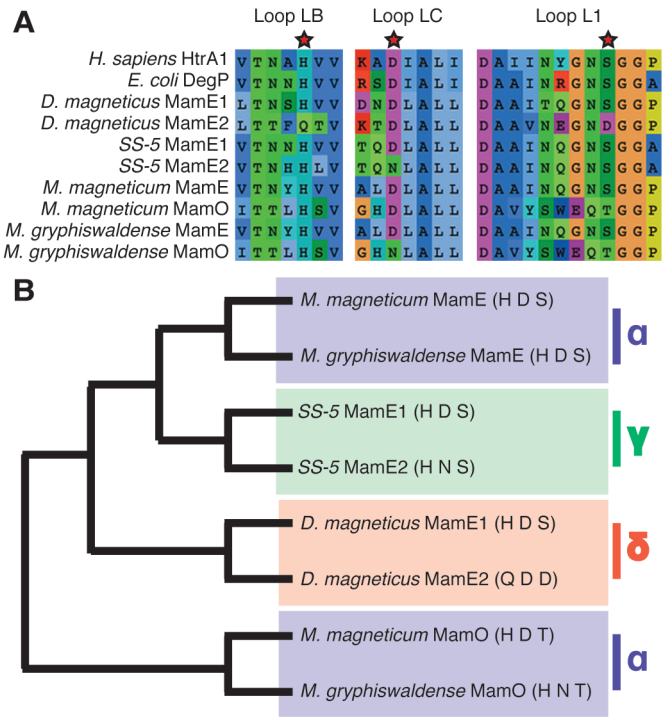


Figure 2-16. Sample analysis of representative trypsin-like sequences from magnetotactic bacteria. (a) Alignment of the catalytic loops from a set of trypsin like-sequences. The trypsin sequences from four magnetotactic organisms were aligned with two canonical HtrAs, *H. sapiens* HtrA1 and *E. coli* DegP. Positions of catalytic triad residues are marked with a star. (b) Phylogeny of the sequences inferred from the detailed analysis shown if Figure 2-14. The identities of the catalytic triad residues are shown in parentheses after each protein name. Boxes represent the class level taxonomy of the organism within the *Proteobacteria.*

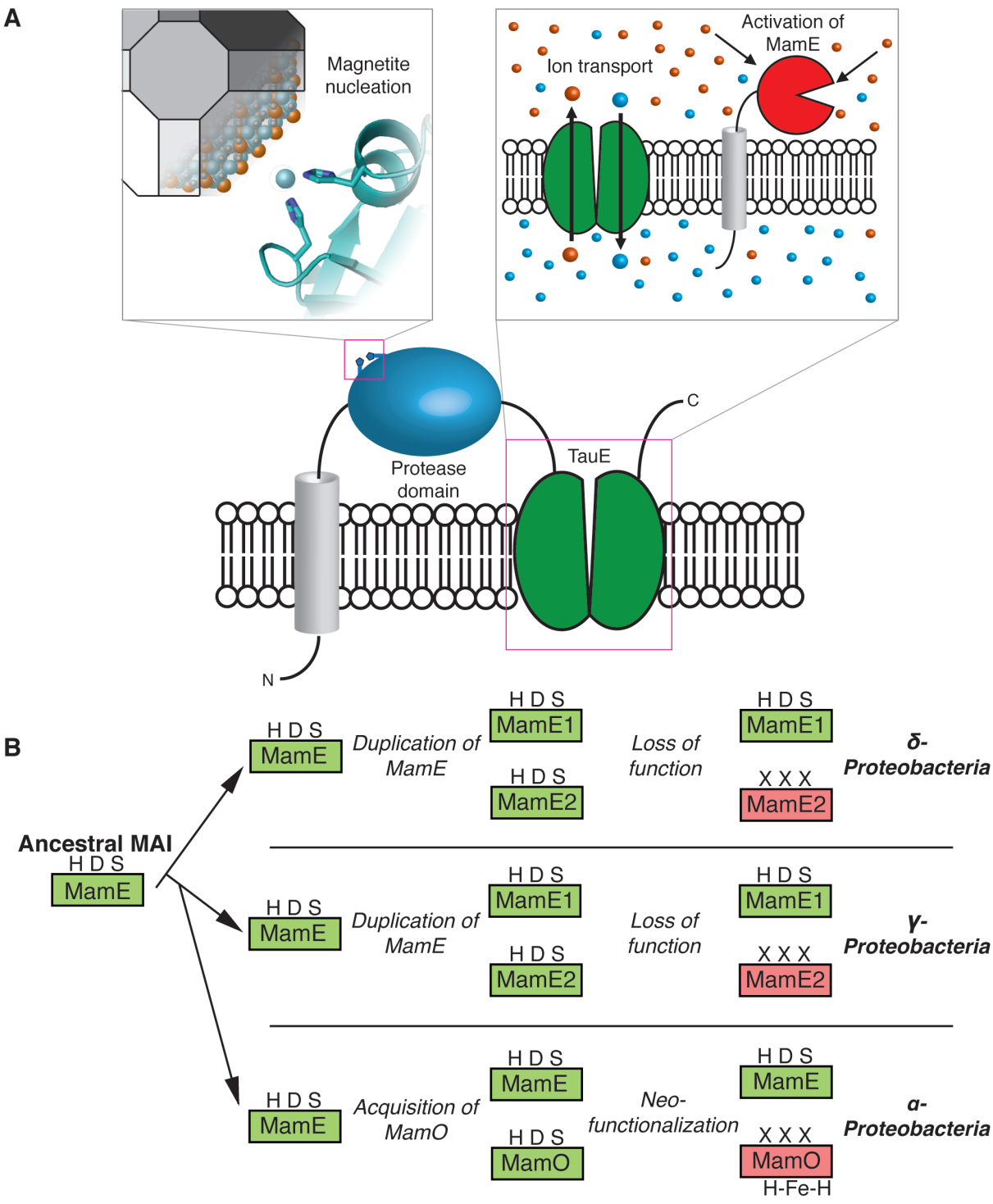


Figure 2-17. MamO in magnetosome formation and evolution. (a) A dual role in biomineralization. Distinct regions of the protein contribute to each activity separately. The protease domain promotes nucleation by binding iron and the TauE domain manipulates solute conditions that regulate MamE's activity. (b) Specialization of the trypsin-like protease family in magnetotactic bacteria through gene duplication and subsequent neofunctionalization.

	MamO (apo): 5HM9	MamO (Ni): 5HMA
Beamline	8.3.1 at ALS	8.3.1 at ALS
Data collection		
Space Group	P432	P432
Cell Dimensions		
a, b, c (Å)	130.22, 130.22, 130.22	129.16, 129.16, 129.16
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	1.116	1.116
Resolution (Å)	50.00-2.60 (2.64-2.60)	50.00-2.30 (2.34-2.30)
No. of Reflections	12179	16999
Completeness (%)	99.8 (98.5)	99.9 (100)
Redundancy	15.2 (12.3)	12.6 (13.1)
R_{sym} (%)	9.8 (69.5)	8.5 (89.7)
$CC_{1/2}$ (last shell)	0.577	0.905
$\langle I \rangle / \langle \sigma(I) \rangle$	79.3 (6.9)	157.7 (6.3)
Refinement		
Resolution (Å)	43.4-2.60 (2.69-2.60)	43.0-2.30 (2.38-2.30)
R_{work}/R_{free}	20.3/23.2 (27.0/33.0)	18.5/20.9 (24.4/29.8)
TLS Groups	0	3
Average *B*-factor (Å)		
Protein	55.9	63.2
Peptide	77.8	98.2
Ni^{2+} (occupancy)		96.0 (1.0)
R.m.s deviation from ideality		
Bond length (Å)	0.009	0.008
Bond angle (°)	1.12	1.05
Ramachandran Plot		
Favored (%)	96.8	97.9
Allowed (%)	3.2	2.1
Outliers (%)	0	0

Table 2-1. Summary of crystallographic data.

PDB	Protein	Chain	Residue Number (PDB)	φ_{193}	ψ_{193}	State	Reference
3NUM	*H. sapiens* HtrA1	A	326	-65.86	-33.17	Inactive	(152)
3NWU	*H. sapiens* HtrA1	A	326	-72.36	-16.53	Inactive	(152)
3NWU	*H. sapiens* HtrA1	B	326	-71.58	-19.3	Inactive	(152)
3NWU	*H. sapiens* HtrA1	C	326	-70.21	-18.01	Inactive	(152)
3MH5	*E. coli* DegP	A	208	-73.41	-0.05	Inactive	(153)
3MH5	*E. coli* DegP	B	208	-90.57	-27.05	Inactive	(153)
3TJN	*H. sapiens* HtrA1	A	326	-84.24	-20.3	Inactive	(154)
3TJN	*H. sapiens* HtrA1	B	326	137	39.78	Active	(154)
3PV2	*L. fallonii* DegQ	A	191	-48.53	-38.25	Inactive	(140)
3PV2	*L. fallonii* DegQ	B	191	-42.8	-47.42	Inactive	(140)
3PV2	*L. fallonii* DegQ	C	191	-53.88	-31.55	Inactive	(140)
3PV2	*L. fallonii* DegQ	D	191	-56.39	-41.34	Inactive	(140)
5HM9	*M. magneticum* MamO	A	179	-73.78	-10.65	Inactive	This work
1CGI	*B. taurus* Chymotrypsin	E	193	101.56	8.71	Active	(155)
3PV3	*L. fallonii* DegQ	A	191	114.28	2.93	Active	(140)
3PV3	*L. fallonii* DegQ	B	191	119.11	-35.67	Active	(140)
3PV3	*L. fallonii* DegQ	C	191	119.98	-20.89	Active	(140)
3PV3	*L. fallonii* DegQ	D	191	109.9	-10.84	Active	(140)
3K6Y	*M. tuberculosis* MarP	A	341	109.05	-16.85	Active	(156)
2RCE	*E. coli* DegS	G	199	113.41	-19.91	Active	Unpublished
2RCE	*E. coli* DegS	H	199	110.12	-18.32	Active	Unpublished
2RCE	*E. coli* DegS	I	199	111.79	-19.76	Active	Unpublished
3OTP	*E. coli* DegP	A	208	120.1	-32.64	Active	(157)
3OTP	*E. coli* DegP	B	208	118.36	-30.42	Active	(157)
3OTP	*E. coli* DegP	C	208	119.61	-36.86	Active	(157)
3OTP	*E. coli* DegP	D	208	125.09	-31.18	Active	(157)
3OTP	*E. coli* DegP	E	208	121.66	-33.92	Active	(157)
3OTP	*E. coli* DegP	F	208	120.96	-28.68	Active	(157)

Table 2-2. L1 loop conformation in a set of trypsin-like protease structures. The main chain configuration at position 193 for all chains in the asymmetric unit of each structure was used to generate the table.

Lefèvre et al. name	Organism	Accession	Uniprot	Triad	Suggested name
MamEO	RS-1	YP_002955485	C4XPP9	H D S	MamE1
MamE-Nter	RS-1	YP_002955484	C4XPP8	Q D D	MamE2
MamEO	BW-1	CCO06679	L0R463	H D S	MamE1
MamE-Nter-1	BW-1	CC06680	L0R577	Q D D	MamE2
MamE-Nter-3	BW-1	CCO06714	L0R4A1	H D S	MamE3
MamE-Nter-2	BW-1	CCO06706	L0R6S4	H N S	MamE4
MamO	*Ca. M. multicellularis*	ETR64737	F4ZYU1	H D S	MamE1
MamE	*Ca. M. multicellularis*	ETR64747	F4ZYV1	H N S	MamE2
MamEO-Nter	ML-1	AFZ77018	U5IHV7	H D S	MamE1
MamEO-Cter	ML-1	AFZ77020	U5IGM4	Q D D	MamE2
MamE'	SS-5	AHY02427	A0A023UKD7	H D S	MamE1
MamE	SS-5	AFX88986	K7Y606	H N S	MamE2

Table 2-3. Suggested renaming of the MamE clade. The naming in a previous comparative genomic study is unconventional due to confusion over analogy to the MAI in *α-Proteobacteria*(30). Our phylogenetic analysis clarifies the ancestry and allows us to use accepted nomenclature.

Strain	Organism	Description	Source
AK30	*M. magneticum* AMB-1	Wild-type AMB-1	(47)
AK69	*M. magneticum* AMB-1	Δ*mamP*	(47)
AK96	*M. magneticum* AMB-1	Δ*mamE* Δ*limE*	(78)
AK94	*M. magneticum* AMB-1	Δ*mamO* Δ*R9*	(78)
AK205	*M. magneticum* AMB-1	Δ*mamE* Δ*mamO* Δ*R9; mamE* deleted from AK94 using pAK241	This work
AK206	*M. magneticum* AMB-1	Δ*mamE* Δ*mamO* Δ*R9* + $mamE^{WT}$*;* allele reintroduced with pAK831	This work
AK207	*M. magneticum* AMB-1	*ΔmamE ΔmamO* Δ*R9* + $mamE^{PD}$*;* allele reintroduced with pAK832	This work
BL21 CodonPlus	*E. coli*	protein expression strain; Cm^R	Agilent
DH5α (*λpir*)	*E. coli*	standard cloning strain	(47)
WM3064	*E. coli*	mating strain; DAP auxotroph used for plasmid transfer	(47)

Table 2-4. Strains used in Chapter 2.

Plasmid	Description	Source
pAK241	Suicide plasmid with fused upsteam and downstream regions to delete *mamE*	(47)
pAK253	Plasmid for integration into neutral site in AMB-1 genome; *tac* promoter	(47)
pAK605	pAK253 with *tac* promoter replaced by *mamAB* promoter (200bp upstream of *mamH)*	This work
pAK619	pAK605 containing *mamE*	This work
pAK620	pAK605 containing $mamE^{PD}$ (H188A T211A S297A mutant)	This work
pAK831	pAK241 with the entire *mamE* region inserted, for reintroduction of *mamE* at native locus	This work
pAK832	pAK831 with the S297A mutation	This work
pAK845	pAK605 containing *mamO*	This work
pAK856	pAK845 with the H116A mutation in *mamO*	This work
pAK885	pAK845 with the D149A mutation in *mamO*	This work
pAK886	pAK845 with the T225A mutation in *mamO*	This work
pAK887	pAK845 with the H148A mutation in *mamO*	This work
pAK888	pAK845 with the H263A mutation in *mamO*	This work
pAK823	pAK605 containing M2(3xFLAG)-tagged mamO; the M2 sequence was inserted upstream of *mamO* inframe with a BamHI linker	This work
pAK824	pAK823 with the H116A mutation in *mamO*	This work
pAK828	pAK823 with the D149A mutation in *mamO*	This work
pAK825	pAK823 with the T225A mutation in *mamO*	This work
pAK890	pAK823 with the H148A mutation in *mamO*	This work
pAK891	pAK823 with the H263A mutation in *mamO*	This work
pAK876	MamO protease domain (residues 46-281) inserted into mcsII of pETDuet using NdeI/BglII; for expression of untagged MamOP in E. coli under the control of T7 promoter	This work
pAK877	MamO protease domain (residues 46-281) with GSAWSHPQFEK (strep tag) fused to the C-terminus inserted into mcsII of pETDuet using NdeI/BglII; for expression of MamOP-strep in *E. coli* under the control of T7 promoter	This work
pAK889	pAK877 with Q258C mutation introduced in the MamO protease domain	This work
pAK899	pAK877 with Q258C H148A H263A mutations introduced in the MamO protease domain	This work

Table 2-5. Plasmids used in Chapter 2.

CHAPTER 3

Magnetite biomineralization in *Magnetospirillum magneticum* is regulated by a switch-like behavior in the HtrA protease MamE

David M. Hershey[1], Patrick J. Browne[1], Anthony T. Iavarone[2,3], Joan Teyra[5], Eun H. Lee[4], Sachdev S. Sidhu[5], and Arash Komeili[1,2,4]

[1] *Department of Plant and Microbial Biology,* [2] *California Institute for Quantitative Biosciences,* [3] *QB3/Chemistry Mass Spectrometry Facility,* [4] *Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA*
[5] *Department of Molecular Genetics, Terrance Donnelly Centre for Cellular and Biomedical Research, University of Toronto, Toronto, Canada ON M5S 3E1*

ABSTRACT
Magnetotactic bacteria are aquatic organisms that produce subcellular magnetic crystals in order to orient in the earth's geomagnetic field. The genetic basis for this process has been dissected in the model magnetotactic bacterium *Magnetospirillum magneticum* AMB-1, leading to the identification of biomineralization genes that control the formation and growth of magnetite crystals. One such factor, MamE, is a predicted member of the HtrA family of serine proteases, a widespread enzyme family that plays important roles in protein turnover and quality control. MamE was recently shown to promote the proteolytic processing of itself and two other biomineralization factors *in vivo*. Here, we have studied MamE-dependent proteolysis in detail. We analyzed the *in vivo* processing patterns of three proteolytic targets and used this information to reconstitute proteolysis with a purified form of MamE. MamE cleaves a custom peptide substrate with positive cooperativity, and its auto-proteolysis can be stimulated with exogenous substrates or peptides that bind to either of its PDZ domains. A constitutively active form of the protease causes biomineralization defects, showing that proper regulation of its activity is required during biomineralization *in vivo*. Our results demonstrate for the first time that MamE is a bonafide HtrA protease and that its activity is consistent with the previously proposed checkpoint model for biomineralization.

INTRODUCTION
Interactions between living organisms and insoluble inorganic compounds are an underexplored aspect of biology. Magnetotactic bacteria assemble magnetic crystals called magnetosomes into chains within their cells allowing them to passively align and navigate along magnetic fields(20, 35). Genetic analyses have shown that magnetosome formation is a complex developmental process controlled by a conserved set of genes(47, 57, 58, 116, 158). A subset of these, called biomineralization factors, whose deletions disrupt or eliminate magnetite crystal formation, have drawn increasing interest for their potential utility in biomedical applications(159). Understanding the biochemical mechanisms employed by these gene products can provide novel strategies for manipulating transition metal-based nanomaterials *in vitro*(98, 101, 160).

Two predicted trypsin-like proteases, MamE and MamO, are required to produce magnetite in the model magnetotactic organism, *Magnetospirillum magneticum* AMB-1(47). Disrupting either gene abolishes the formation of magnetite crystals without affecting the development of their surrounding membrane compartment, indicating that both proteins are required for magnetite nucleation within the magnetosome lumen(47, 56, 123). Furthermore, deleting *mamE* causes a number of proteins that normally localize at the magnetosomes to become dispersed throughout the cytoplasmic membrane. Adding either wild-type or catalytically inactive (E^{PD}) alleles of *mamE* complements the protein localization defect, but only the wild-type allele restores normal magnetite biosynthesis(78).

Cells with the E^{PD} allele show an intermediate biomineralization phenotype in which they produce small magnetite particles. While wild-type AMB-1 has a distribution of crystal sizes centered at 50-60nm in diameter, the size distribution in the E^{PD} cells is centered at ~20nm. Interestingly, ~97% of the crystals in the E^{PD} strain are smaller than 35nm, the point above which magnetite particles become paramagnetic and can hold a stable magnetic dipole. The correlation between mineral sizes in the E^{PD} strain and the superparamagnetic to paramagnetic transition

point lead to the speculation that MamE's putative protease activity regulates the transition to a magnetotactic lifestyle. This so-called checkpoint model predicts that cells produce small superparamagnetic crystals until an unknown signal activates MamE, promoting maturation to paramagnetic particles(78).

MamE is a member of the HtrA/DegP family of trypsin-like proteases, a ubiquitous family of enzymes that controls various aspects of protein quality control(75). The family is characterized by domain structures consisting of an N-terminal trypsin-like domain and one or two C-terminal postsynaptic density 95/discs large/zonula occludens-1 (PDZ) domains(76, 77, 161). Like many proteases, intricate regulatory mechanisms are a hallmark of HtrA catalysis(139, 153). Based on structural and mechanistic investigations, the PDZ domains play several roles in regulating proteolysis including promoting assembly and activating the protease domain by binding to extended peptide motifs(140, 157, 162, 163). MamE has an unusual domain structure in which a tandem *c*-type cytochrome motif has been inserted between the protease and two PDZ domains (Fig. 1A)(78). Unfortunately, studies aimed at understanding the catalytic activity of MamE and its regulation have been hindered by an inability to obtain recombinant protein.

Recently, MamE was found to promote the *in vivo* proteolytic processing of itself, MamO, and another biomineralization factor named MamP in a manner that required the predicted MamE active site(93, 124, 164). Although MamO was also required for these proteolytic events, this effect did not require the predicted MamO active site. Subsequent structural analysis showed that MamO's protease domain was locked in an inactive state and incapable of catalysis, suggesting that it played a non-catalytic role in activating MamE(164). Despite the genetic evidence for MamE's role in proteolysis, its activity has not been confirmed directly using purified components. Here, we have characterized MamE-dependent proteolysis in detail and identified a number of regulatory mechanisms. Developing a method to purify MamE and analyzing *in vivo* proteolytic patterns of each target facilitated reconstitution of MamE-dependent proteolysis *in vitro*. Detailed analysis of its catalytic activity suggests a switch-like model in which the basal state of the protein is an inactive form that can be turned on through a number of routes. A constitutively active allele of MamE disrupts biomineralization, confirming that properly regulated proteolysis is critical to magnetosome formation.

MATERIALS & METHODS

Strains, plasmids and growth conditions

The strains and plasmids used in this study are described in Tables 3-1 and 3-2, respectively. AMB-1 was maintained in MG medium supplemented with kanamycin when necessary as previously described (47). Magnetic response was measured using the coefficient of magnetism as previously described (47). Standard molecular biology techniques were used for plasmid manipulation. *E. coli* strains were grown in LB medium supplemented with appropriate antibiotics. Plasmids were maintained in *E. coli* strain DH5α *λpir*. *E. coli* strain WM3064 was used for plasmid conjugations as described previously(47).

Immunoblotting

Whole-cell lysates of AMB-1 strains were prepared from 10mL cultures and analyzed as described previously(164). For analysis of the auto-cleavage reaction products, a 1 in 10 dilution

of each time-point was separated on a 12% acrylamide gel for immunoblotting. The MamE and MamP antibodies have been described previously(164). The anti-6xHis (Sigma), anti-FLAG (Sigma), anti-σ70 (Thermo Fisher) and anti-strep (Qiagen) antibodies were purchased from commercial sources.

Fractionation of MamO fragments
A strain with the genetic background Δ*O*Δ*R9*/*FLAG-O* was cultured without shaking at 30 °C in 2 L screw-capped flasks that were filled to the top with MG medium. The cells were harvested by spinning at 5k x g for 15 min, re-suspending in cold 20mL 25mM Tris-HCl pH7.4, re-spinning at 8k x g for 10 min and freezing the resulting pellet at -80°C until use. Cell pellets were thawed on ice and re-suspended in 5mL lysis buffer A (10mM Tris-HCl pH 8.0, 50mM NaCl, 1mM EDTA). Pepstatin A and leupeptin were added to a final concentration of 2μg/mL and PMSF was added to 2mM. Lysozyme was added from 50mg/mL stock to a final concentration of 0.5mg/mL and the cells were incubated at room temperature for 15 min. 15mL of lysis buffer B (20mM HEPES-KOH pH7.5, 50mM NaCl, 1.25mM $CaCl_2$) was added along with DTT to 2mM and DNAse to 5μg/mL and the suspension was incubated for 15min at 4°C with agitation. The cells were sonicated twice for 10 seconds and the suspension was spun at 8k x g for 10min to isolate the magnetite-associated material.

The resulting pellet was re-suspended in 5.5 mL solubilization buffer (20mM BisTris-HCl pH7.0, 75mM NaCl, 10% glycerol) and CHAPS was added to 1% from a 10% stock solution. The suspension was incubated at room temperature for 15min with agitation followed by an incubation for 15min at 4°C with agitation. The suspension was spun at 16k x g for 15min. The resulting pellet was re-suspended in 5.5 mL solubilization buffer and the detergent extraction was repeated with 1% FosCholine. The FosCholine-soluble material was loaded on a 1mL HiTrap Q FF column (GE Healthcare) that had been equilibrated with solubilization buffer containing 0.03% *n*-dodecyl β-D-maltoside (DDM). The column was washed with 10mL of solubilization buffer with 0.03% DDM and eluted with 4mL of buffer Q1 (20mM BisTris-HCl pH7.0, 275mM NaCl, 10% glycerol) with 0.03% DDM followed by 4mL of buffer Q2 (20mM BisTris-HCl pH7.0, 400mM NaCl, 10% glycerol) with 0.03% DDM. The Q1 fraction was added to 50μL anti-FLAG M2 resin (Sigma) and incubated at 4°C with agitation for 3 hrs. The resin was isolated by spinning at 4k x g and washed with sequential 1mL washes of buffer Q2, buffer Q1 and solubilization buffer each containing 0.03% DDM. Bound proteins were eluted by 3 washes with 50μL of 0.2M glycine pH 2.8, which were pooled with 50μL of 1M Tris-HCl pH8.0.

Preparation of trypsin digests for liquid chromatography-mass spectrometry (LC-MS)
The concentrated FLAG elution fraction was separated on a 12% acrylamide gel and stained with colloidal Coomassie Blue. A ~3 x 10mm section of the gel corresponding to the processed MamO band was excised from the gel and chopped into small pieces. These were washed with 100mM NH_4HCO_3 followed by reduction and alkylation of cysteines with DTT and iodoacetamide. The gel pieces were then dehydrated by washing with increasing concentrations of acetonitrile in 100mM NH_4HCO_3 and dried under vacuum. A 0.1mg/mL solution of trypsin was used to re-swell the gel pieces, and they were incubated overnight at 37°C. The resulting peptides were extracted from the gel slices with successive washes of 0.1% formic acid solutions

containing increasing concentrations of acetonitrile. The extracts were pooled in a fresh tube, concentrated under vacuum to remove the organic phase and stored at 4°C until analysis.

LC-MS

Trypsin-digested protein samples were analyzed using a Thermo Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ-Orbitrap-XL mass spectrometer equipped with a nanoelectrospray ionization (nanoESI) source (Thermo Fisher Scientific, Waltham, MA). The LC was equipped with a C18 analytical column (Acclaim® PepMap RSLC, 150 mm length × 0.075 mm inner diameter, 2 μm particles, 100 Å pores, Thermo) and a 1-μL sample loop. Acetonitrile (Fisher Optima grade, 99.9%), formic acid (1-mL ampules, 99+%, Thermo Pierce), and water purified to a resistivity of 18.2 MΩ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). The elution program consisted of isocratic flow at 2% B for 4 min, a linear gradient to 30% B over 38 min, isocratic flow at 95% B for 6 min, and isocratic flow at 2% B for 12 min, at a flow rate of 300 nL/min.

Full-scan mass spectra were acquired in the positive ion mode over the range m/z = 350 to 1800 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 60,000 (at m/z = 400, measured at full width at half-maximum peak height, FWHM), which provided isotopic resolution for singly and multiply charged peptide ions. Thus, an ion's mass and charge could be determined independently, i.e., the charge state was determined from the reciprocal of the spacing between adjacent isotope peaks in the m/z spectrum. In the data-dependent mode, the eight most intense ions exceeding an intensity threshold of 50,000 counts were selected from each full-scan mass spectrum for tandem mass spectrometry (MS/MS) analysis using collision-induced dissociation (CID). MS/MS spectra were acquired using the linear ion trap, in centroid format, with the following parameters: isolation width 3 m/z units, normalized collision energy 30%, default charge state 3+, activation Q 0.25, and activation time 30 ms. Real-time charge state screening was enabled to exclude unassigned and 1+ charge states from MS/MS analysis. Real-time dynamic exclusion was enabled to preclude re-selection of previously analyzed precursor ions, with the following parameters: repeat count 2, repeat duration 10 s, exclusion list size 500, exclusion duration 90 s, and exclusion mass width 20 ppm. Data acquisition was controlled using Xcalibur software (version 2.0.7, Thermo). Raw data were searched against the *Magnetospirillum magneticum* AMB-1 FASTA protein database using Proteome Discoverer software (version 1.3, SEQUEST algorithm, Thermo). Peptide identifications were validated by manual inspection of MS/MS spectra, i.e., to check for the presence of y-type and b-type fragment ions[1] that identify the peptide sequences(165).

Expression and purification of MamE

pAK825 or pAK964 was transferred to C43 cells (Lucigen) that had been previously transformed with the pEC86 heme-loading plasmid(166). The transformed cells were maintained at 30°C due to toxicity of the construct at 37°C. An overnight liquid culture was inoculated into 600mL 2xYT medium supplemented with the appropriate antibiotics. The cultures were grown at 30°C until the OD_{600} reached ~0.5, at which point the culture was transferred to 20°C. After a 30 min equilibration, the culture was induced with 35μM IPTG. Expression was performed for 12.5-13 hrs at 20°C with shaking at 200 rpm.

Cells were harvested by immediately chilling the cultures on ice and spinning at 6k x g for 10 min. The resulting pellet was re-suspended in 50mL of cold osmotic shock buffer (50mM NaPhosphate pH8.0, 1mM EDTA, 20% sucrose). Leupeptin (1.5µg/mL), pepstatin A (1.5µg/mL) and lysozyme (0.5mg/mL) were added, and the suspension was rocked at room temperature for 15 min. An equal volume of ice-cold H_2O was added and the suspension was rocked on ice for 15 min before spinning at 8k x g for 10 min to remove debris.

The resulting supernatant was added to 3 mL Ni-NTA resin (Qiagen) and supplemented with NaCl (150mM), DNAseI (5µg/mL), NP-40 (0.1%) and $MgCl_2$ (2.5mM). The slurry was rocked at 4°C for 30 min and the beads were allowed to settle. After decanting the upper phase, the slurry was poured into a column, washed with 10 column volumes of Ni wash buffer (25mM Tris-HCl pH7.4, 250mM NaCl, 10mM imidazole, 10% glycerol) and the bound proteins eluted with Ni elution buffer (25mM Tris-HCl pH7.4, 250mM NaCl, 250mM imidazole, 10% glycerol). The Ni-NTA eluent was loaded onto a 1mL StrepTrap column (GE Healthcare), which was then washed with 5mL strep wash buffer (25mM Tris-HCl pH7.4, 250mM NaCl, 10% glycerol). Bound proteins were eluted in strep wash buffer containing 2.5mM desthiobiotin. The purified protein was concentrated in 50kDa cutoff ultrafilter while simultaneously removing the desthiobiotin by repeated dilution and concentration with strep wash buffer. The concentration was determined by the Bradford method using bovine serum albumin to prepare a standard curve. Aliquots were frozen in liquid N_2 and stored at -80°C until use.

Analysis of MamO1 peptide cleavage

A custom peptide with the sequence 5-carboxymethylfluorescein-Thr-Gln-Thr-Val-Ala-Ala-Gly-Ser-Lys(CPQ2)-D-Arg-D-Arg was obtained commercially (CPC Scientific). The peptide was dissolved in DMSO and stored at -20°C. 5X substrate solutions with various concentrations of the MamO1 peptide were prepared in assay buffer (50mM Tris-HCl pH8.0) containing 0.05% NP-40 and 1.6% DMSO. To initiate the reaction, 10µL samples of the substrate mix were added to 40µL of MamE protein solution that had been diluted to 125nM in assay buffer in a 96-well plate. The fluorescence was scanned (excitation: 485nm; emission 538nm) every 5 min for 2 hrs in a Tecan plate reader.

The slope was determined from the linear portion of each reaction. Cleavage rates were calculated by making a standard curve from a MamO1 cleavage reaction that had been incubated for 24hrs to allow for complete hydrolysis. Specific activities were determined by normalizing these cleavage rates to the enzyme concentration. Rates were plotted as a function of peptide concentration and fit to the Hill form of the Michaelis-Menten equation using the Kaleidagraph software package:

$$V = \frac{V_{max}[O1]^n}{{K_m}^n + [O1]^n}$$

Analysis of MamE auto-proteolysis

25µL reactions were prepared by adding 1µL of activating peptide dissolved in DMSO at the appropriate concentration to 24µL of MamE diluted to 2µM in assay buffer. The reactions were incubated at 30°C, and 8µL aliquots were removed at the appropriate times. Each aliquot was

quenched immediately by mixing with SDS sample buffer. Equal volumes of each aliquot were separated on a 12% acrylamide SDS-PAGE gel and stained with Coomassie Blue to visualize the processing pattern.

Expression and purification of PDZ domains
For all three PDZ domain constructs (EP1, EP2 and EP12), the appropriate plasmids for expression as N-terminal 6xHis-MBP-TEV fusions were transformed into BL21 Codon Plus cells. Cultures were grown in 2xYT at 37°C until the OD_{600} reached ~0.8 at which point they were transferred to 20°C for 30min followed by induction with 0.1mM IPTG and expression overnight. The cells were harvested by centrifugation, resuspended in resuspension buffer (25mM Tris-HCl pH7.4, 800mM NaCl, 10mM imidazole, 10% glycerol) and frozen at -80°C until use.

For protein purification, the cells were thawed on ice and sonicated for three 30-second cycles. The lysate was clarified by spinning at 13k x g for 30 min. The resulting supernatant was loaded on a 3mL Ni-NTA column that had been equilibrated in resuspension buffer. After washing with 10 column volumes of resuspension buffer and 10 column volumes of wash buffer 2 (25mM Tris-HCl pH7.4, 400mM NaCl, 25mM imidazole, 10% glycerol), bound proteins were eluted with Ni elution buffer.

For the purification of the EP1 and EP2 proteins, the elution fractions were dialyzed overnight against AEX buffer A (25mM BisTris-HCl pH7.0, 75mM NaCl and 10% glycerol). The desalted protein was passed through a 1mL HiTrap QFF column (GE Healthcare) and the flow-through was concentrated in a 50kDa cutoff ultrafilter, injected onto a16/60 Superdex 200 column and developed in Storage Buffer. Each protein eluted as a single symmetrical peak. The peak fractions were concentrated in a 50kDa ultrafilter and small aliquots were frozen in liquid N_2 and stored at -80°C for use in the phage display experiments.

For the purification of the EP12 protein, the elution fraction was dialyzed overnight against digest buffer (50mM NaPhosphate pH8.0, 75mM NaCl, 5mM imidazole, 10% glycerol) in the presence of 6xHis tagged TEV protease to remove the 6xHis-MBP tag. The resulting sample was passed through a 3mL Ni-NTA column that had been equilibrated in digest buffer. The flow-through fraction was concentrated in a 10kDa ultrafilter, passed through a 1mL HiTrap SP FF column and concentrated again before injection on a16/60 Superdex 200 column that was developed in storage buffer. The protein eluted as a single symmetrical peak. The peak fractions were concentrated in a 10kDa ultrafilter and small aliquots were frozen in liquid N_2 and stored at -80°C for use in the fluorescence anisotropy experiments.

Phage display
C-terminally and N-terminally displayed peptide libraries were used to assess the peptide binding preferences of MamE PDZ1 and PDZ2. The C-terminal peptide library consisted of random decapeptides constructed using 10 consecutive NNK degenerate codons encoding for all 20 natural amino acids and fused to the C terminus of a mutant M13 bacteriophage major coat protein (2×10^{10} unique members)(159, 167). The N-terminal peptide library consisted of random hexadecapeptides constructed using 16 consecutive mixes of 19 codon trimers (cysteine

and STOP codons were excluded) and fused to the N terminus phage coat protein (2.4×10^{11} unique members)(168).

The phage display selections followed an established protocol used previously for the PDZ human domains(169, 170). Briefly, each library was separately cycled through rounds of binding selection against each immobilized MBP-PDZ fusion protein on 96-well MaxiSorp immunoplates (Nalge Nunc). After each round, phage were propagated in *E. coli* XL1-blue cells (Stratagene) supplemented with M13-KO7 helper phage (New England Biolabs) to facilitate phage production. After four rounds of selection, phages from individual clones were analyzed in a phage enzyme-linked immunosorbent assay (ELISA). Phages that bound to the MBP-PDZ fusion but not a control MBP were subjected to DNA sequence analysis.

Fluorescence anisotropy

Peptides with the following sequences were synthesized commercially with a fluorescein-aminocaproic acid group fused to the N-terminus: WSQEMEDWFWQMPLSG (PDZ1*) and MEDYGIFMTSPEGPWA (PDZ2*). Each peptide was diluted to a concentration of 40nM in 25mM Tris-HCl pH7.4 containing 0.25mg/mL bovine serine albumin. A dilution series of EP12 protein was prepared in storage buffer. 6µL of the ligand solution was added to 18µL of the appropriate protein solution in a 384-well plate, and the mixture was allowed to equilibrate at room temperature for 15 min. Polarization measurements were made at 535nm using a Perkin Elmer Victor 3V 1420 plate reader. The resulting anisotropy values were plotted as a function of protein concentration and fit to a single site binding model using the Kaleidagraph software package:

$$FA = \frac{B_{max} * [\mathrm{EP12}]}{K_{D,app} + [\mathrm{EP12}]}$$

RESULTS

Mapping MamE-dependent cleavage patterns in vivo

To learn more about the context of MamE-dependent proteolysis, we attempted to map the cleavage patterns of MamE, MamP and MamO *in vivo* using an epitope tagging approach. Strains with N- or C-terminally 3xFLAG-tagged alleles of each gene were added to their respective deletion strains. Each of the tagged alleles complemented the biomineralization defects of the deletions with the exception of the C-terminally tagged *mamO* allele (*O-FLAG*). Analysis of cell extracts from these strains by Western blotting was used to assess the processing patterns for each proteolytic target.

For MamE, a number of N-terminal proteolytic fragments but no C-terminal fragments are observed (Figure 3-1B), indicating that short segments are sequentially removed from the C-terminus. Interestingly, this seems to culminate in a stable ~45kDa fragment that corresponds to the MamE protease domain and the N-terminal membrane helix separated from the rest of protein. For MamP, one N-terminal band and two C-terminal bands are observed that correspond to a full-length protein and a protein truncated by approximately 10kDa at the N-terminus, indicating the removal of the membrane anchor from the predicted soluble domains (Figure 3-1B). For MamO, a full-length and a shorter band are observed for both the N- and C-terminally

tagged proteins (Figure 3-1B). The pattern predicts that MamE-dependent proteolysis separates the N-terminal protease and C-terminal TauE domains of MamO (Figure 3-1A).

In many MamE and MamO blots, there are small (i.e. ~20 kDa) bands near the bottom of the gel when analyzing the N-terminally tagged proteins. Although these bands appear in many experiments, the levels and even the presence of these two signals are inconsistent. This potentially suggests that short segments are removed from the N-terminus of each protein to produce unstable fragments that are quickly degraded. Due to the inconsistency in these bands we have not focused on them in our analysis of the processing pattern.

Identification of a putative cleavage motif in MamO

Analysis of the *in vivo* processing patterns strongly indicated that MamE processes itself, MamO and MamP(164). We reasoned that one strategy for reconstituting its proteolytic activity *in vitro* could be to design substrates from motifs that are cleaved in a MamE-dependent fashion. Based on epitope tagging, MamO is cleaved between the protease and TauE domains, which suggests that there is a MamE-dependent cleavage site at the mature C-terminus of the protease domain (Figure 3-2B). Cell pellets from 2L cultures of the Δ*O*Δ*R9*/*FLAG-O* strain were used for biochemical fractionation. Enzymatic lysis and sonication followed by a low speed (8000 x g) spin were used to isolate material associated with the dense magnetite particles. A number of detergents were tested for their ability to dissolve the MamO fragments. Most classes of detergents are ineffective or only partially effective in the initial solubilization step. Lipid-like zwitterionic detergents including lauryldimethylamine oxide (LDAO) and FosCholine-12 extract the fragments from the membrane, but they disrupt binding to the α-FLAG affinity resin. However, once the initial extraction step is complete, the detergent requirements to maintain solubility become less stringent.

Based on the solubility information, the low speed pellet was pre-washed with CHAPS before extracting the MamO fragments from the membranes with FosCholine-12. In order to facilitate binding to the affinity resin, the FosCholine soluble material was loaded on an anion exchange column and exchanged to the detergent DDM by extensive washing before eluting with salt. This fraction was then used as the input for an α-FLAG affinity isolation to yield a final fraction enriched in N-terminal fragments of MamO (Figure 3-2A and B).

The concentrated α-FLAG elution was separated on an SDS-PAGE gel, stained with colloidal Coomassie Blue and the region around 35kDa was excised. After performing in-gel trypsin digestion, peptides were extracted and concentrated for LC-MS/MS analysis. A number of peptides from the MamO sequence are consistently detected, and, as expected, they map almost exclusively to the protease domain of MamO (Figure 3-2C). In all of the samples, the protein sequence coverage drops off sharply in the linker between the protease and TauE domains (Figure 3-2C). A peptide with the sequence GSATAPGQPQTQTV is routinely detected at the C-terminal edge of the peptide coverage (Figure 3-2D). This peptide results from a predicted tryptic cleavage on the N-terminus but has a non-tryptic C-terminus, which suggests that it contains the C-terminal sequence of the mature MamO protease domain.

Purification of MamE

In order to reconstitute its proteolytic activity *in vitro*, we developed a method to express and purify appreciable amounts of MamE. In the expression construct, the N-terminal membrane anchor is replaced with the OmpA signal peptide to produce a soluble protein that can still undergo heme loading in the periplasm(171). Initial fractionations with an N-terminally 6xHis-tagged form of the protein had significant contamination due to what appeared to be truncated fragments caused by auto-cleavage during expression. To eliminate this problem, a strep tag was added on the C-terminus to allow for a sequential affinity isolation of full-length protein. Finally, MamE has a predicted region of 60-70 disordered residues downstream of the N-terminal membrane anchor and upstream of the trypsin-like domain. Removing this region dramatically improved the solubility.

The expression conditions also had to be carefully optimized in order for MamE to accumulate. Maximum expression occurs with low levels of IPTG induction in C43 cells at 20 °C. The precise details of the expression and purification procedure are described in the Materials and Methods. It allows for the soluble region of MamE as well as a catalytically inactive form ($MamE^{S297A}$) to be purified with yields approaching ~1mg/L. Importantly, both preparations appear red and display absorbance spectra characteristic of *c*-type cytochromes (Figure 3-3).

Direct proteolysis of the MamO cleavage motif

We designed a fluorogenic peptide (peptide MamO1) to test as a substrate for purified MamE. The substrate contains 8 residues that make up the putative cleavage motif identified in MamO (Figure 3-2) flanked by an upstream fluorophore and a downstream fluorescence quencher. Normally, the peptide has low fluorescence due to interaction between the fluorophore and quencher. If the peptide is cleaved, the two fragments will separate, resulting in an increase in fluorescence. Upon addition of the O1 peptide to purified MamE, there is a linear increase in fluorescence. Importantly, the $MamE^{S297A}$ protein does not alter the fluorescence indicating the signal is due to serine protease activity from MamE (Figure 3-4A). MamE hydrolyzes the O1 peptide with a k_{cat} of 0.64 ± 0.03 min^{-1} and a K_M of 6.1 ± 0.5 μM. As for other HtrA proteases, the reaction is positively cooperative, displaying a Hill coefficient of 1.5 ± 0.1 (Figure 3-4B) (139, 172). These values are similar to those reported for cleavage of peptide substrate by other trypsin-like proteases and confirm that MamE can efficiently cleave the motif identified in MamO (173). Combined with the *in vivo* analysis these results confirm that MamO is a direct proteolytic target of MamE.

Reconstitution of MamE auto-proteolysis

Analysis of MamE processing in AMB-1 along with the extensive auto-proteolysis during its expression in *E. coli* suggested that MamE is capable of auto-proteolysis. However, purified MamE is relatively stable such that after an hour of incubation at 30°C, nearly all of the protein remains intact (Figure 3-5A). The positive cooperativity observed for the steady-state kinetics of O1 peptide cleavage indicated that MamE's catalytic activity could be stimulated by substrates. This mode of regulation might also lead to peptide-induced activation of auto-cleavage. Indeed, the MamO1 peptide stimulates degradation of full-length MamE in a dose-dependent manner, confirming that MamE's activity can be stimulated by the presence of substrate (Figure 3-5A).

Taking advantage of the distinct tags used to purify MamE, we examined the auto-cleavage fragmentation pattern by Western blotting. Numerous truncated proteins are detected by blotting

for the N-terminal 6xHis tag, the smallest of which is a ~27 kDa fragment presumed to be the protease domain (Figure 3-5B). In contrast, only the full-length protein can be seen when blotting for the C-terminal strep tag (Figure 3-5C). The pattern indicates that the reaction proceeds via sequential removal of small fragments from the C-terminus. Furthermore, it matches the pattern seen by examining epitope-tagged alleles of MamE expressed *in vivo* and confirms the successful reconstitution MamE-dependent proteolysis *in vitro.*

Activation through the PDZ domains

PDZ domains of other HtrA proteases regulate proteolytic activity by binding to extended peptide motifs(153). Phage display has been a productive approach for identifying peptide ligands that bind to PDZ domains(174-177). Each of the MamE PDZ domains was purified and used as bait in phage display selections. Both bait proteins showed phage enrichment for specific particles with a library displaying peptides on the N-terminus of the coat protein, but no enrichment was observed in libraries displaying C-terminal fusions. This suggests that, unlike those associated with other HtrA proteases, the MamE PDZ domains do not display a preference for C-terminal peptides(178). Interestingly, phage selections for both domains showed a strong preference for internal regions (Tables 3-3 and 3-4). However, a single clone dominated both pools making the identification of consensus motifs hard to interpret.

Peptides corresponding to the sequence that dominated each selection were synthesized and labeled with a fluorescent dye. In addition, the C-terminal region of MamE containing only the PDZ domains (EP12) was purified and used to test for direct binding to the phage-derived ligands. Fluorescence anisotropy experiments demonstrate that both phage-derived peptides bind to the C-terminus of MamE (Figure 3-6A). The PDZ1 peptide shows ~10-fold tighter binding than the PDZ2 peptide, but both affinities are comparable to those seen for other PDZ domains(177). Addition of either peptide to full-length MamE, results in a dose-dependent activation of auto-processing (Figure 3-6B). Importantly, the activation threshold for the PDZ2 peptide is higher than the PDZ1 peptide, mirroring the equilibrium binding data. These results show that, like other HtrA proteases, MamE's protease activity can be regulated through peptide binding to its PDZ domains.

Misregulation of MamE disrupts biomineralization in vivo

MamE's low basal activity and its stimulation by various peptides define a switch-like behavior that toggles a high activity and a low activity state of the protein. This *in vitro* behavior suggests that modulation of the two states is important during biomineralization *in vivo.* To date, we have not been able to identify growth conditions that prevent *in vivo* processing of MamE, MamO or MamP. Thus, we utilized a genetic approach to examine whether proper regulation of MamE-dependent proteolysis was required for biomineralization.

Work with the model HtrA protease DegS has shown that mutating residue 192 (chymotrypsin numbering) in the oxyanion hole to proline increases basal cleavage rates by shifting the allosteric equilibrium toward the active state (138, 139). We introduced an allele with the corresponding mutation of MamE (Q294P) into the *mamE* null strain (Figure 3-7A). This strain displayed significantly less MamE and MamP as assessed by western blotting, indicating increased proteolytic activity from MamE. Additionally, the $mamE^{Q294P}$ allele partially circumvents the requirement for MamO in promoting MamE-dependent proteolysis. MamE and

MamP appeared to be processed when the *mamE*^{Q294P} allele was introduced into a strain lacking both *mamE* and *mamO*, though it did not restore processing of MamE to wild-type levels (Figure 3-7B and C). These results indicate that *mamE*^{Q294P} produces a constitutively active, unregulated protease *in vivo*.

We next examined this allele for its ability to complement the biomineralization defects seen in the *mamE* null strain. While wild-type *mamE*^{WT} allele completely restores the magnetic response, the *mamE*^{Q294P} strain has a significantly lower response (Figure 3-7D). Thus, both the inactive (*mamE*^{PD}) and constitutively active forms (*mamE*^{Q294P}) of the protease disrupt biomineralization *in vivo*. The magnetic response of the *mamE*^{Q294P} strain is higher than the negligible signal measured in the *mamE*^{PD} strain, indicating that the biomineralization process is stalled at a different stage when MamE's activity is unregulated. These results demonstrate that complete biomineralization of magnetosome crystals relies not only on the occurrence of MamE-dependent proteolysis but also careful regulation of the activity.

DISCUSSION

Biochemical principles underlying the biomineralization of magnetite by magnetotactic bacteria represent a model for understanding how biological molecules manipulate inorganic compounds(19). Recent advances describing the genetic basis for this process have paved the way for mechanistic studies of the factors that promote mineral formation(47). The HtrA protease MamE has emerged as a central biomineralization factor in the model organism *Magnetospirillum magneticum* AMB-1. In addition to promoting crystal nucleation and protein localization to the magnetosome, MamE regulates a transition from small superparamagnetic crystals to full-sized paramagnetic particles. This crystal maturation phenotype was linked to MamE's putative protease activity suggesting a model where its catalytic activity controls crystal maturation(78).

Here, we have studied MamE's serine protease activity in detail. We mapped the proteolytic patterns of three *in vivo* targets both at the domain level and, in one case, at the individual residue level. Using this information, we reconstituted a number of aspects of MamE-dependent proteolysis with purified components. We show that MamE directly cleaves a motif from the linker between MamO's protease and TauE domains in a positively cooperative fashion. Furthermore, we show that purified MamE has low basal activity, but that it can be activated in a number of ways including the presence of substrate and peptide binding to either of its PDZ domains. This behavior is consistent with a switch-like mode of regulation in which the protease requires activation by environmental cues. These results also show for the first time that MamE is a serine protease that is capable of degrading itself and other biomineralization factors.

Nearly all studied members of the HtrA family behave as trimers or multiples thereof(76, 77, 140, 161, 179, 180). In other systems with two PDZ domains, the first PDZ seems to regulate protease activity directly while the second is thought to mediate rearrangements of core trimers into higher order oligomers(140, 157, 163). Peptide binding to either the first or the second PDZ domain of MamE can activate proteolysis, although the activation through PDZ2 is much weaker. Furthermore, the protein behaves as a monomer as indicated by gel filtration. Transitions between a monomer and higher order assemblies are rare in the HtrA family, but the positive

cooperativity observed for MamE suggests that the active form is indeed a larger assembly(181). The protein production method described here should enable future structural studies aimed at understanding the potential for novel assembly behavior as well as an unusual regulatory role for the second PDZ domain.

The switch like activation of MamE's activity suggested that allosteric regulation of its protease activity was required for proper crystal maturation. Using a mutation reported to stabilize active forms of other HtrA proteases, we showed that, like the catalytically inactive form, a constitutively active form of MamE had defects in crystal maturation(139). These results confirm that both the active and inactive states are important during the process of magnetosome formation. Similar experiments with DegP in *E. coli* indicated that the proper balance between active and inactive forms is required for fitness during heat stress(182). Our results show that, in addition to maintaining fitness during stress, this mode of regulation can be used to control a developmental process.

Taken together, our results are consistent with the checkpoint model for MamE-dependent proteolysis in regulating the maturation of magnetite crystals. However, the detailed mechanism by which that activity promotes crystal growth remains unclear. One possible mechanism could be by controlling the size of the surrounding membrane. A recent study demonstrated a link between the growth of magnetosome membrane compartments and growth of the magnetite crystals within. This finding led to the proposal that there is a checkpoint regulating a second stage of membrane growth after the onset of biomineralization(183). Although *mamE* deletions have intact magnetosome membranes, the sizes of the membranes have not been quantified. It is tempting to speculate that proteolysis controls this switch by linking membrane growth to crystal growth. In this scenario, crystal nucleation in the E^{PD} cells would not lead to membrane growth, while the E^{QP} cells would initiate membrane growth before crystals had grown sufficiently, leading to stunted particles in both cases (Figure 3-8).

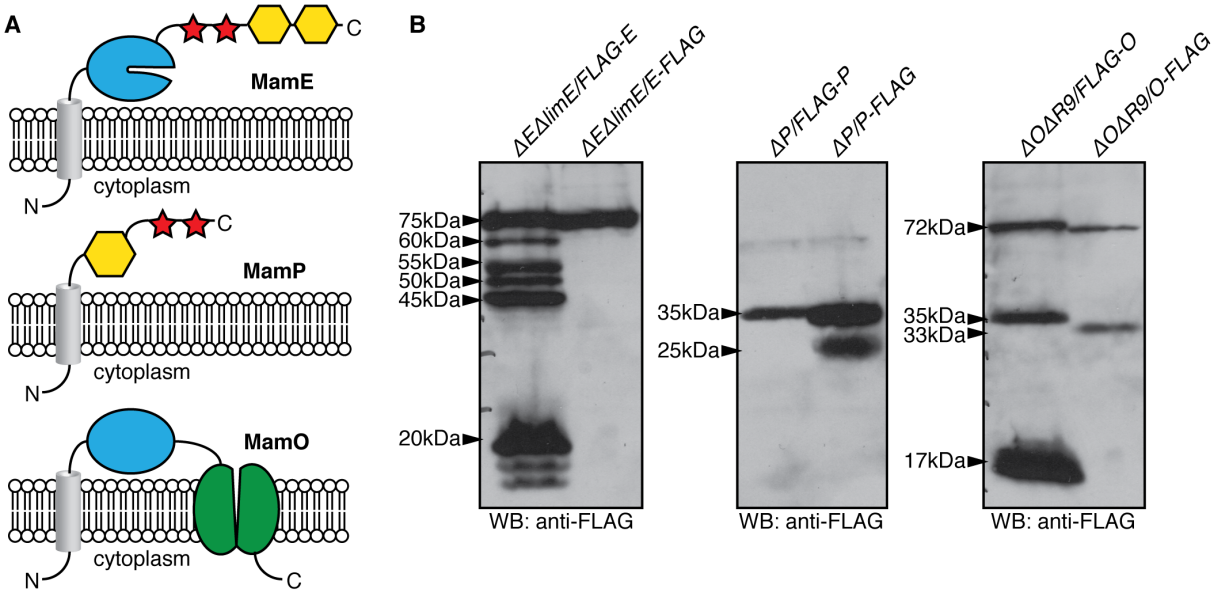


Figure 3-1. *In vivo* proteolytic processing of MamE, MamP and MamO. (a) Predicted domain structures of the three proteolytic targets. Grey cylinder: transmembrane helix; blue: trypsin-like domain; red: *c*-type cytochrome; yellow: PDZ domain; green: TauE domain. (b) Proteolytic processing patterns observed through epitope tagging. The inferred sizes of each fragment are indicated. The fragments were observed in each of at least four independent experiments with the exception of the 20kDa N-terminal fragment of MamE and the 17kDa N-terminal fragment of MamO, which varied dramatically between experiments.

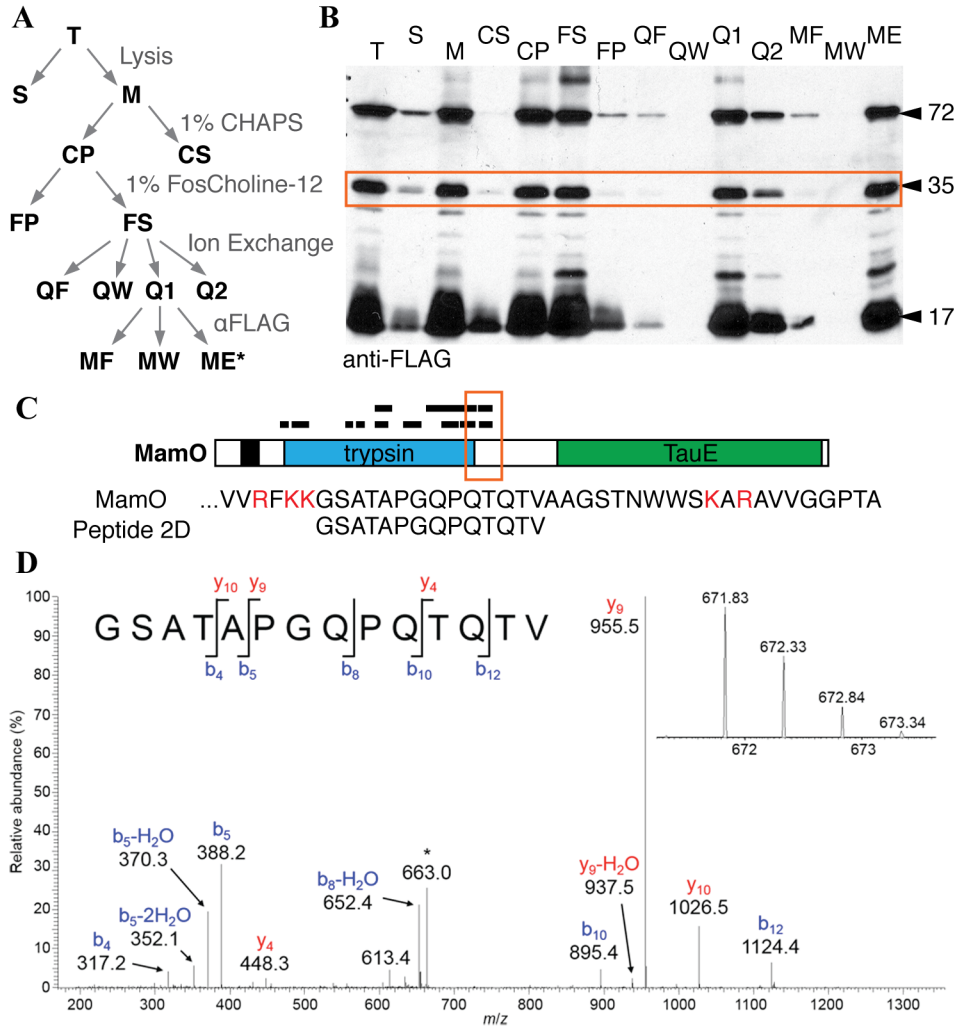


Figure 3-2. Biochemical fractionation to enrich N-terminal MamO fragments. (a) Schematic of the enrichment procedure. (b) Western blot of each fraction from Panel a. The predicted protease domain fragment in marked with a red box. (c) Peptides identified in a representative LC-MS/MS analysis. The red letters in the MamO sequence represent predicted tryptic cleavage sites. The coverage pattern is characteristic of analyses for three separate preparations. (d) Tandem mass spectrum from collision-induced dissociation of the $[M+2H]^{2+}$ ion of the peptide, GSATAPGQPQTQTV, corresponding to amino acid residues 273-286 of MamO. The inset shows detail for the isotopically resolved, unfragmented peptide precursor ion. The fragment ion at m/z = 663 (denoted by the asterisk) is due to precursor ion that has undergone neutral loss of a molecule of water. This peptide was detected in each of three biological replicate experiments.

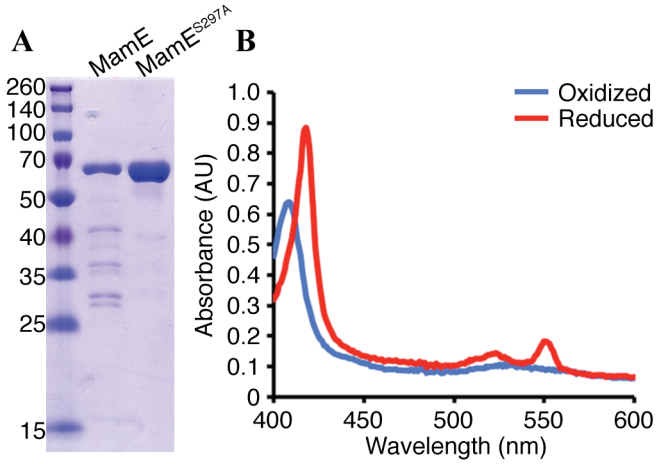


Figure 3-3. Purification of MamE. (a) MamE and MamE^{S297A} (residues 108-728) were purified as a fusion to an N-terminal 6xHis and C-terminal strep tag. (b) Absorbance spectrum of MamE^{S297A} in the oxidized (blue) and dithionite reduced (red) forms.

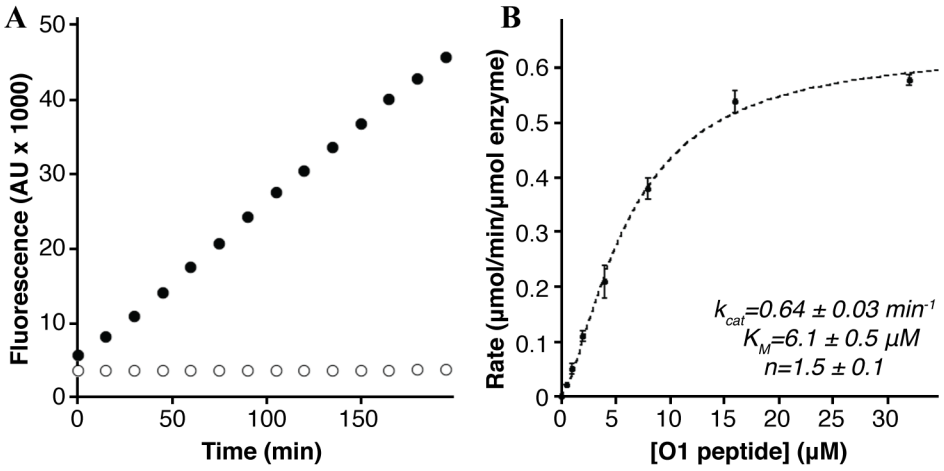


Figure 3-4. Cleavage of the MamO1 fluorogenic substrate. (a) Linear increase in fluorescence upon addition of 20µM MamO1 peptide to 200nM MamE (black circles). No increase is seen with peptide addition to MamE^{S297A} (white circles). (b) Steady-state kinetics of O1 cleavage by MamE. The dotted line represents a fit to the Hill form of the Michelis-Menten equation. Error bars represent the standard deviation from three technical replicates. The plot is characteristic of the data seen in five biological replicates.

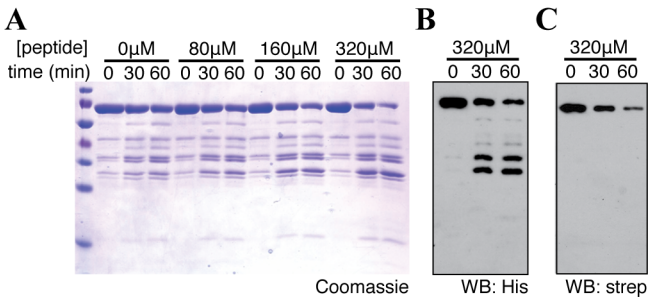


Figure 3-5. Reconstitution of MamE auto-processing. (a) The MamO1 substrate induces auto-cleavage of MamE. The processing pattern was assessed by Western blots (b, c) of the reaction containing 320μM O1 for the indicated tag on MamE. The experiment is representative of three biological replicates.

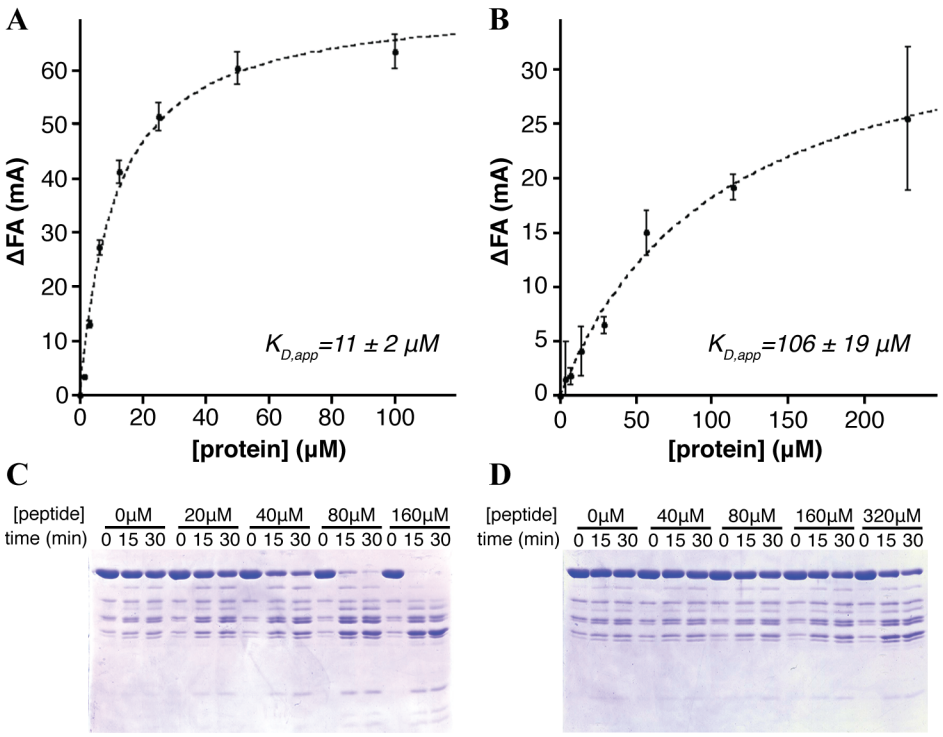


Figure 3-6. Peptide binding to the PDZ domains activates MamE. (a) Fluorescence anisotropy showing binding of the EPDZ1* peptide to the EP12 protein. The error bars represent the standard deviation from three technical replicates. The dotted line represents a fit of the data to a single-site binding model. (b) Binding of the EPDZ2* peptide to the EP12 protein. (c) Activation of MamE auto-cleavage by the EPDZ1* peptide. (d) Activation of MamE auto-cleavage by the EPDZ2* peptide. Each experiment was repeated a minimum of three times.

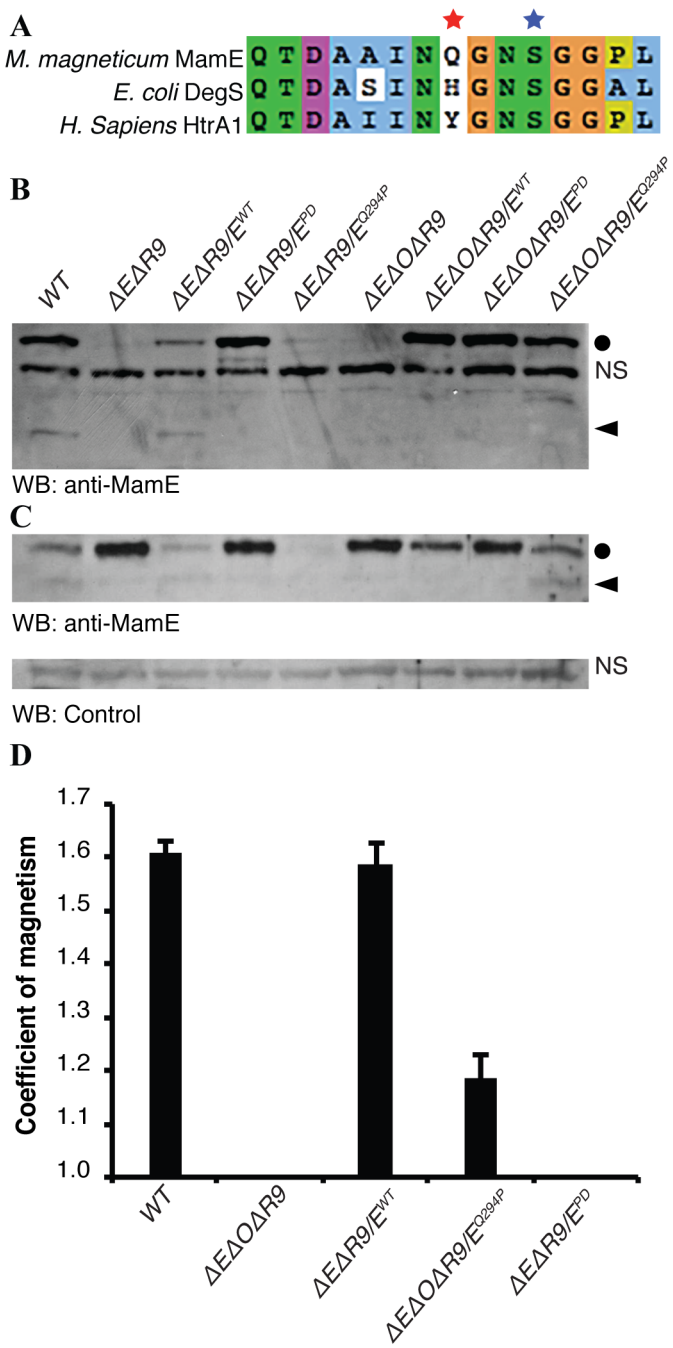


Figure 3-7. A constitutively active form of MamE disrupts biomineralization. (a) Alignment of HtrA proteases. The red star marks position 192 in the oxyanion hole (residue 198 in DegS), and the blue star marks the catalytic serine nucleophile. (b, c) Immunoblot analysis of AMB-1 lysates probed for MamE (b) and MamP (c). Circles mark full-length proteins and carats mark proteolytic fragments. NS marks nonspecific bands reacting with each antibody preparation. (d) Magnetic response of AMB-1 cultures with the indicated genetic background. Biological replicates represent independent cultures of each strain and each measurement represents the average and standard deviation from three independent experiments.

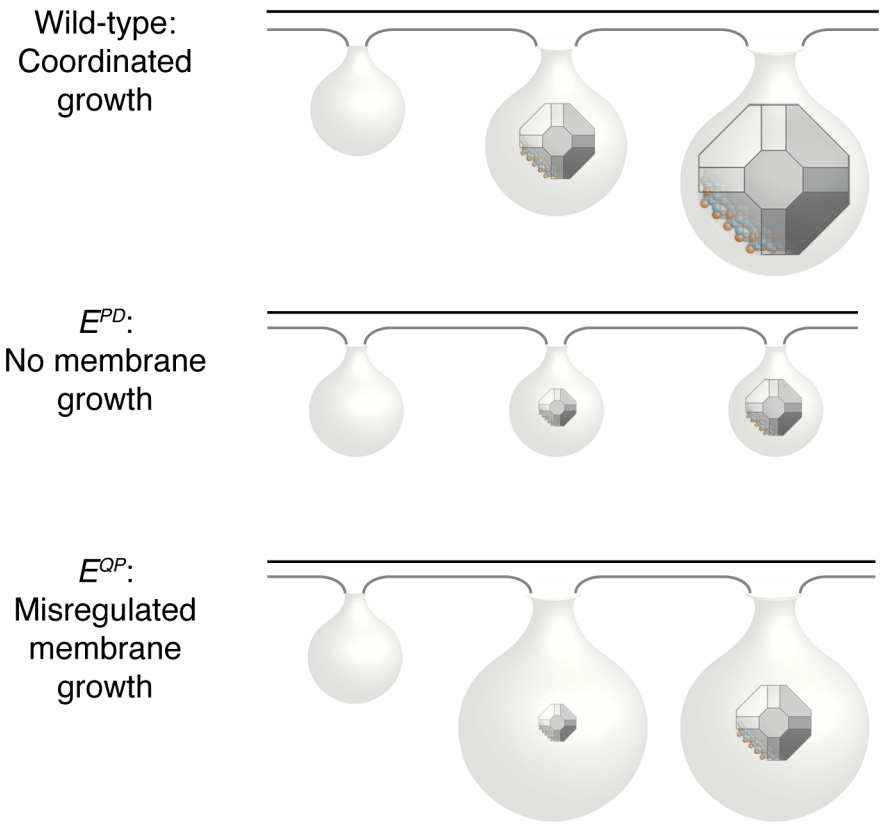


Figure 3-8. Model for MamE dependent membrane remodeling to promote crystal maturation. The model illustrates the predicted membrane remodeling defects in the inactive and misregulated *mamE* alleles that lead to the crystal maturation defects in Figure 3-7.

Strain	Organism	Description	Source
AK30	*M. magneticum* AMB-1	Wild-type AMB-1	(47)
AK69	*M. magneticum* AMB-1	Δ*mamP*	(47)
AK96	*M. magneticum* AMB-1	Δ*mamE* Δ*limE*	(78)
AK94	*M. magneticum* AMB-1	Δ*mamO* Δ*R9*	(78)
AK205	*M. magneticum* AMB-1	Δ*mamE* Δ*mamO* Δ*R9*	(164)
C43	*E. coli*	protein expression strain	Lucigen
BL21 CodonPlus	*E. coli*	protein expression strain; Cm^R	Agilent
DH5α (*λpir*)	*E. coli*	standard cloning strain	(47)
WM3064	*E. coli*	mating strain; DAP auxotroph used for plasmid transfer	(47)

Table 3-1. Strains used in Chapter 3.

Plasmid	Description	Source
pAK605	Plasmid for expressing protein under control of the *mamH* promoter; non-replicative in AMB-1: integrates upstream of amb0397	(164)
pAK701	pAK605 containing N-terminally 3xFLAG tagged *mamE*	This work
pAK702	pAK605 containing C-terminally 3xFLAG tagged *mamE*	This work
pAK787	pAK605 containing N-terminally 3xFLAG tagged *mamP*	This work
pAK788	pAK605 containing C-terminally 3xFLAG tagged *mamP*	This work
pAK823	pAK605 containing N-terminally 3xFLAG tagged *mamO*	(164)
pAK1003	pAK605 containing C-terminally 3xFLAG tagged *mamO*	This work
pAK1005	pLBM4; LIC compatible vector for expressing fusions to the OmpA signal peptide under the control of the *lac* promoter	(171)
pAK1004	pEC86; plasmid constitutively expressing the *E. coli ccm* genes for *c*-type cytochrome maturation	(166)
pAK825	pLBM4 containing 6xHis-TEV-MamE(resi 108-728)-strep fused to the OmpA signal peptide	This work
pAK964	pAK825 with the S297A mutation in MamE	This work
pAK619	pAK605 containing *mamE*	(164)
pAK620	pAK605 containing $mamE^{PD}$ (H188A T211A S297A mutant)	(164)
pAK999	pAK605 containing $mamE^{QP}$ (Q294P mutant)	This work
pAK1000	pET based vector expressing the first PDZ domain of MamE (resi 489-588) fused to 6xHis-MBP-TEV at the N-terminus	This work
pAK1001	pET based vector expressing the second PDZ domain of MamE (resi 634-728) fused to 6xHis-MBP-TEV at the N-terminus	This work
pAK1002	pET based vector expressing the first and second PDZ domains of MamE (resi 489-728) fused to 6xHis-MBP-TEV at the N-terminus	This work

Table 3-2. Plasmids used in Chapter 3.

Peptide Sequence																						Count
-	-	W	S	Q	E	M	**E**	D	W	F	**W**	Q	M	P	L	S	G	-	-	-	-	37
-	-	-	-	-	-	F	**N**	Y	E	Q	**W**	L	Q	A	E	Y	L	Q	E	H	-	3
-	-	-	-	-	D	D	**E**	W	W	M	**W**	V	E	Q	K	L	H	E	A	T	-	2
-	-	-	-	P	E	T	**Q**	Y	W	L	**W**	L	M	E	L	D	S	Q	G	-	-	2
M	L	Y	N	F	F	E	**Q**	G	W	W	**W**	Y	F	S	W	-	-	-	-	-	-	2
E	D	Q	S	K	L	D	**Q**	Y	S	K	**W**	M	L	M	L	-	-	-	-	-	-	1
-	-	I	D	T	L	V	**E**	H	H	W	**W**	G	T	V	T	T	F	-	-	-	-	1
-	-	-	-	-	-	P	**E**	F	D	E	**W**	F	E	Q	R	Y	E	E	M	E	K	1
-	-	-	-	-	V	F	**D**	Y	W	Q	**W**	V	M	E	D	T	D	G	M	I	-	1
-	-	W	E	W	P	D	**D**	A	W	Q	**F**	L	T	R	M	S	T	-	-	-	-	1
-	-	-	-	-	-	F	**D**	Y	W	E	**Y**	I	S	H	A	E	Q	P	S	D	L	1

Table 3-3. Peptide sequences from the phage display selections using MamE PDZ1. The sequences are aligned based on the motif **ψ**xxx**Ω**, where ψ is a hydrophilic residue and Ω is an aromatic residue. Bold type is used to mark the hydrophilic and aromatic residues around which the sequences were aligned.

Peptide Sequence																											Count
-	-	-	-	-	-	-	-	-	-	M	E	D	Y	**G**	I	F	M	T	S	P	E	G	P	W	A	22
-	G	Q	I	E	P	T	W	M	W	D	M	Y	**G**	F	K	L	-	-	-	-	-	-	-	-		2
S	Y	V	P	G	E	W	Q	G	L	E	S	M	**G**	I	V	M	-	-	-	-	-	-	-	-		2
-	A	Q	E	A	D	Y	P	A	L	Y	Q	L	**G**	F	I	P	-	-	-	-	-	-	-	-		1
-	-	E	H	W	Q	D	Y	S	F	E	S	L	**G**	I	Y	I	-	-	-	-	-	-	-	-		1
-	E	S	P	V	E	W	D	Y	L	E	S	F	**G**	L	V	I	-	-	-	-	-	-	-	-		1
-	-	-	-	-	-	E	W	S	F	Q	D	M	**G**	F	M	L	S	Y	D	Q	M	G	-	-		1
-	-	-	-	I	A	P	D	L	W	Y	N	W	**G**	I	M	W	R	D	G	-	-	-	-	-		1
-	-	I	S	Q	E	P	L	Y	I	E	Q	L	**G**	M	M	V	F	-	-	-	-	-	-	-		1
-	-	-	-	-	-	N	E	M	D	N	L	M	**G**	M	I	F	M	S	P		E	-	-	-		1
-	-	-	-	-	-	N	H	A	Y	Q	D	F	**G**	F	V	V	S	E	L	Q	E	-	-	-		1
-	-	P	F	M	P	L	E	Y	W	Q	F	M	**G**	I	V	F	T	-	-	-	-	-	-	-		1
-	-	-	-	P	G	T	F	E	D	A	T	L	**G**	F	T	W	F	H	D	-	-	-	-	-		1
-	Q	N	N	Q	D	M	M	F	Q	P	E	L	**G**	I	W	F	-	-	-	-	-	-	-	-		1
-	-	-	-	R	M	P	E	S	M	R	D	M	**G**	F	S	I	L	M	A	-	-	-	-	-		1
-	-	-	-	-	-	Y	H	G	V	N	E	L	**G**	L	M	M	M	D	Y	I	P	-	-	-		1

Table 3-4. Peptide sequences from the phage display selections using MamE PDZ2. The sequences are aligned based on the motif ψΦ**G**ΦΦΦ, where ψ is a hydrophobic residue and Φ is a hydrophilic residue. Bold type is used to mark the glycine residue around which the sequences were aligned.

CHAPTER 4

Conclusions and perspectives on the roles of trypsin-like proteases in magnetite biomineralization

David M. Hershey

ABSTRACT
The findings described in Chapters 2 and 3 of this thesis identify a multiple independent activities for both MamE and MamO and the phenotypic effects of disrupting them. This chapter reviews each individual function known for the two proteins and attempts to synthesize this information into a concerted mechanism for how they might regulate biomineralization. A model in which MamE acts as a central regulator controlling numerous stages of magnetosome formation by sensing the completion of specific biogenesis events before remodeling the magnetosome membrane accordingly is proposed. Limitations to the current model are described along with the suggestion that physiological features of the *Magnetospirillum* genus should be clarified to understand the cellular context in which biomineralization factors operate. Finally, insights into how magnetosome formation evolved are discussed with an emphasis on the power of biochemical experiments for developing evolutionary hypotheses.

The increasing complexity of magnetite biomineralization
Recent advances in making targeted gene deletions have allowed the identification of genes required for magnetosome formation (47, 58). A major outcome from this era of magnetosome research was the finding that biomineralization in magnetotactic bacteria is surprisingly complex. It requires the presence of over a dozen genes, including both MAI specific factors and factors central to cellular growth and maintenance. Because bacteria have historically been used as simplified models to understand more elaborate processes that occur in eukaryotes, the level of complexity required to make magnetite may have been unexpected. However, as the genes required to biomineralization are being dissected in mechanistic detail, appreciation for the underlying complexity has only grown.

Central to the newest advances in understanding the mechanisms of magnetosome formation has been the link between *in vitro* biochemical studies and reverse genetics. New alleles have been generated by defining specific amino acid residues required for newly identified biochemical activities. Introducing these custom alleles into gene deletion strains can be used to assess the contribution of a single biochemical activity to the magnetosome formation process *in vivo*. The results of such studies show that alleles with missense mutations often have intermediate phenotypes that are less severe than the corresponding gene deletion. Thus, in addition to requiring a surprisingly large set of genes, some magnetite biosynthesis factors appear to have more than one activity (78, 164, 184). These findings imply a more intricate route for magnetosome formation than would be predicted from the step-wise model proposed based on individual gene deletions.

MamE
MamE is likely the most complicated biomineralization factor studied to date. The *mamE* deletion has two recognizable phenotypes. Δ*mamE* cells contain membrane invaginations that contain no biominerals, and GFP-fusions to proteins normally targeted to the magnetosome are mislocalized (47). The $mamE^{PD}$ allele, which lacks protease activity, can restore both proper protein localization and the initiation of biomineralization, but the magnetite crystals produced in this strain do not mature properly. The $mamE^{C2}$ allele, in which the heme attachment site in each magnetochrome domain is disrupted, also restores the biomineralization of small magnetite crystals. Interestingly, the $mamE^{C2}$ strain produces larger crystals that are distributed across a larger range of diameters that the $mamE^{PD}$ strain, indicating that the defects are distinct (78).

Biochemical characterization has clarified some of the molecular events causing the phenotypes associated with disrupting *mamE*. MamE, MamO and MamP are proteolytically processed in wild-type AMB-1, but not in the $mamE^{PD}$ strain (164). *In vitro* analysis with a purified form of MamE demonstrated that it is indeed an HtrA protease capable of directly proteolyzing itself and other biomineralization factors (184). Because the proteolytic targets identified to date were discovered using a candidate approach, it is likely that additional MamE-dependent proteolytic events await identification. Two studies have now independently confirmed that recombinant MamE is heme modified and a bonifide *c*-type cytochrome, though no redox effects on proteolytic rate have been observed (92, 184). Together, the current data indicates that the small crystals observed in the $mamE^{PD}$ strain are due to a deficiency in proteolytic processing, that defects in the $mamE^{C2}$ strain are due to a disruption in electron flow and that the two effects are independent of each other.

How MamE's protease activity is regulated also seems to be important for its role in magnetite crystal maturation. *In vitro* proteolysis assays demonstrated a regulatory paradigm in which the basal proteolytic rate of MamE is low but can be increased by either exogenous substrates or ligands that bind to its PDZ domains. These allosteric effects on MamE led to a model in which specific conditions within the magnetosome activate proteolytic activity, promoting the later stages of biomineralization once the appropriate inputs are detected. The fact that both inactive ($mamE^{PD}$) and over-active ($mamE^{Q297P}$) alleles cause disrupt the cells' to mature magnetite crystals properly, confirms that MamE's protease activity alone is insufficient and that the proper regulation of proteolysis is critical (184). A complete understanding of the MamE-dependent pathway for regulating crystal maturation will await the identification of the signal(s) that modulate protease activity *in vivo.*

The protease-independent mechanisms by which MamE promotes protein sorting and crystal nucleation remain unclear, but it is tempting to speculate that the two functions are related. MamA-GFP localizes as fluorescent puncta within disperse cytoplasmic background in Δ*mamE* cells, distinct from the strictly disperse pattern observed in ΔR5 cells (47). A subsequent study examining GFP-MamI and MamC-GFP in the Δ*mamE*ΔR9 strain in which both *limE* and *mamE* are absent, also concluded that *mamE* was required for proper localization of magnetosome proteins (78). Because the partial redundancy between *mamE* and *limE* was not discovered until after the report that Δ*mamE* cells contain empty magnetosome membranes, membrane biogenesis has not been examined in cells lacking both genes. The similar protein localization patterns in the Δ*mamE*ΔR9 and ΔMAI cells suggest that an inability to form the magnetosome membrane could be the true defect.

A potential role in membrane remodeling could unify the seemingly separate protein localization and biomineralization defects in the Δ*mamE* strain. A recently reported correlation between crystal diameters and the sizes of the surrounding membranes suggests the presence of a regulatory pathway that initiates a second phase of membrane remodeling after the initiation of biomineralization (183). Although this putative checkpoint has not been confirmed in a mutant unable perform biomineralization, the possibility that the membranes mature with the crystals is an intriguing one. The switch-like proteolytic activity of MamE is a prime candidate to regulate the transition into the second growth phase. A model in which MamE has both protease-

independent membrane remodeling and protease-dependent membrane growth activities would clarify the complex set phenotypes associated with various forms of the protein.

MamO

Much like MamE, molecular and biochemical analysis of MamO has uncovered a number of distinct functions for the protein. An early study noted that mutating its predicted active site caused an intermediate phenotype in which the cells produced small magnetite crystals, suggesting that MamO acts as a protease to promote biomineralization. However, structural and genetic studies subsequently showed that MamO's protease domain has lost the ability to carry out the proteolytic reaction. Despite the fact that MamO is not a protease, it contributes by promoting the proteolytic activity of MamE in targeting a number of biomineralization factors., Proteolytic processing is disrupted in the absence of the TauE domain, showing that MamO's protease domain is insufficient its proteolytic activation function (164). Missense mutations in MamO that cannot activate MamE have not been identified and the mechanistic basis for this activity is not clear.

Direct metal binding is now thought to be the major function of MamO's protease domain. The recombinant protein binds to a number of transition metals *in vitro* and a crystal structure of a metal bound form identified two histidines directly coordinating a metal ion. The close spatial proximity of the metal binding motif to the predicted catalytic triad residues suggests that the crystal maturation phenotypes observed in earlier studies were actually caused by altered metal binding activity and not a disruption of proteolytic activity. Interestingly, mutating both residues coordinating the metal ion in the structure does not eliminate metal binding, indicating that another unidentified site exists. However, mutating either residue leads to a complete loss of biomineralization *in vivo*, leading to the model that MamO binds to iron atoms to promote magnetite nucleation. Proteolytic processing by MamE is not disrupted in these metal binding mutants, showing that metal binding and MamE activation are distinct functions of MamO (164).

Finally, the putative transporter activity of MamO's TauE domain has not been explored in detail. Genetic studies demonstrated that this domain is required for the initiation of biomineralization and for MamE activation. It has been suggested that ion transport by MamO might alter the solute environment of the magnetosome lumen to activate MamE, but this possibility has not been examined directly (164). Due to the lack of biochemical or structural information about the TauE family, alleles of MamO with disrupted transport activity have not been made. Understanding the transport mechanism and solute specificity of MamO would help clarify the potential link to its MamE activation function. In addition, it would provide insight into the solute content of the magnetosome lumen during biomineralization, about which surprisingly little is known.

MamE as a central regulator of magnetosome formation

Mechanistic analyses of the two trypsin-like proteases required for biomineralization have identified multiple molecular functions for each protein. MamO has two domains with distinct roles in the same process. Both the metal-binding function of the protease domain and the putative transport activity of the TauE domain are required for the initiation of biomineralization. Given that MamE-dependent proteolysis separates these domains into two separate proteolytic fragments, why they are synthesized as a single polypeptide is not clear. One possibility is that

the C-terminal TauE domain uses the N-terminal protease domain for targeting to the magnetosome, and that the domains are separated upon proper insertion in the membrane. This might explain the finding that the TauE domain is required for MamE-dependent proteolysis because it would allow MamE to ensure the integrity of a critical biomineralization factor before initiating its protease-dependent crystal maturation program.

The genetic, cell biological and biochemical experiments performed to date indicate that MamE is a central regulator throughout the magnetosome formation process. Defects in protein organization, magnetite nucleation and crystal growth have been identified in *mamE* deletions. MamE has two PDZ domains, and these domains often bind directly to biological membranes (185). Perhaps, the two phases of membrane remodeling are both mediated by interactions between MamE's PDZ domains and the magnetosome membrane. Protease-dependent changes in the PDZ-membrane interaction could promote for crystal maturation by regulating the second membrane growth phase. Controlling both the formation of membranes competent to initiate biomineralization and the subsequent maturation of the minerals by modulating the same protein would allow the processes to be separated in time. Much like the targeting dependent-proteolysis model proposed above, MamE would ensure that specific events have occurred before initiating the next phase of magnetosome formation.

New approaches for understanding biomineralization

In the case of both MamE and MamO, a similar gap in knowledge is limiting the scope of their proposed mechanisms. The combination of genetic experiments with biochemical and structural studies of the recombinant protein led to a model in which MamO directly promotes magnetite nucleation by interacting with iron atoms. A major piece of data supporting this model is the ability of the protein to directly bind divalent transition metals with low micromolar K_Ds (164). However, these affinities are modest compared to other metal binding proteins. Low affinity binding might be relevant in the magnetosome lumen where iron concentrations are assumed to be high, but the actual concentrations are unknown. More broadly, solute identities and concentrations in the magnetosome have simply not been characterized. Ideally, magnetite nucleation by MamO would be reconstituted *in vitro* to define specific details of how the nucleation process works. However, because the phase transitions of iron are extremely sensitive to solute parameters, designing and interpreting these experiments is difficult without any knowledge of the solution in which the protein operates.

Understanding the signals that modulate MamE's protease activity would also benefit from knowledge of the solute contents in the magnetosome lumen. A number of allosteric activation mechanisms have been identified for its enzymatic activity *in vitro,* but the specific *in vivo* signals are unknown. Because MamO's TauE transporter domain is required for activation MamE, one possibility is that a specific solute modulates MamE's activity (164). The auto-cleavage activation assay developed for the recombinant protease could easily be used to measure the effects of specific solute conditions on enzymatic rate (184). However, the physiological relevance of any potential activating solutes would be unclear without additional knowledge of the ionic environment of the magnetosome lumen during biomineralization.

Biochemical characterization of the transporters required for biomineralization would help to characterize the solute conditions inside the magnetosome. MamM, MamN and MamO are all

predicted to encode proteins that export ions from the cytoplasm (47). By purifying these proteins and reconstituting their transport function in lipid vesicles, it would be possible to study their mode of action. Parameters like the identity of the transported ion(s), the driving force for the transport reaction and the rate of transport can be determined using these assays. Structural and mechanistic analysis of each protein can identify residues to disrupt by making new alleles interrogating the contribution of specific aspects of their mechanism to the biomineralization process. Knowing which ions are moving in what direction across the magnetosome membrane would provide invaluable hints about what is inside. Perhaps transporters alter more than just the iron pool, changing the pH through proton antiport or importing elements such as phosphorous and sulfur that affect iron solubility.

More broadly, many of the current limitations derive from a lack of knowledge about key metabolic processes contributing to biomineralization. For instance, it has been assumed that iron in magnetite is derived from the cytoplasmic ferrous iron pool, but this has not been determined experimentally. Uptake assays and pulse-chase experiments aimed at identifying which iron sources enter the cells, what acquisition mechanisms contribute, and the cellular compartments involved are needed to understand precursor accumulation phase for magnetite biomineralization. Separately, though respiratory terminal electron acceptors seem to contribute to bioimneralization based on knockouts of critical metabolic enzymes, electron flow in and out of the magnetosome is anything but clear (113-115). Experiments measuring electron and mass balance could determine whether consumption of terminal oxidants outpaces cellular growth when biomineralization occurs. As detailed mechanistic analyses of biomineralization factors become common, more basic physiological information is needed to understand the cellular and metabolic context of their isolated activities.

Insight into the evolutionary origins of magnetosome formation

In addition to clarifying part of the biomineralization pathway in AMB-1, the biochemical experiments performed with MamO lead to broad insights into the evolution of magnetosome formation. Genomic analyses comparing the MAI from *Delta*- and *Gammaproteobacteria* to those in the *Alphaproteobacteria* identified a set trypsin-like proteases in each strain assumed to be MamE and MamO (30). However, phylogenetic analysis inspired by the discovery that MamO is a degenerate protease showed that these were mis-annotated. MamO appears to be a unique family of trypsin-like protease that occurs only in the MAIs of magnetotactic *Alphaproteobacteria*. Although blast searches had identified HtrA proteases as the closest relatives of MamO, the phylogeny of the bacterial trypsin-2 family shows that there are not closely related proteins to the MamO family in sequence databases. Furthermore, this analysis showed that the trypsin-like proteases from each of the *Gamma*- and *Deltraproteobacterial* MAIs are both homologs of MamE from AMB-1 and that these strains do not contain a MamO (164).

The phylogenic placement of the two proteases complicates the evolutionary model for inheritance of the MAI. Within the MamE clade, the sequences cluster according the phylogenetic placement of each strain, clearly demonstrating none of the extant *mamE* genes are derived from horizontal transfer. This finding confirms the idea that MamE was part of an ancient core of magnetosome factors that was inherited vertically throughout the divergence of the *Proteobacteria*. Due to the lack of closely related representatives outside the magnetotactic

Alphaproteobacteria, it is likely MamO was horizontally transferred to these strains as an accessory factor to supplement the core genes. Although a core set of factors encoding ability to make magnetosomes was clearly inherited vertically, horizontal gene transfer may have been important in assembling sets of accessory factors that tailor the process to physiology the extant strains.

Analysis of the sequences in the MamE clade shows that instead of a single, ancient duplication leading to the two copies of the protein in the *Gamma-* and *Deltaproteobacteria,* at least three independent duplications occurred after the clades diverged. Insight into the function of the two MamE copies was informed by the findings with MamO. Analyzing the predicted active site residues from sequences in the MamE clade strongly indicates that, much like MamO, one copy of MamE in each strain is a degenerate protease (164). Thus, all magnetotactic bacteria contain a degenerate protease, but this feature is not ancestral but the result of convergent evolution in the three *Proteobacterial* groups. This finding suggests that each group experienced a similar evolutionary pressure that drove the same adaptive response. Perhaps changes in the earth's atmosphere limited the availability of iron, driving similar innovations that allowed these organisms to continue to produce magnetosomes. It is likely that a wide range of different oxidative conditions continually altered the geochemical state of iron over the 2-3 billion years since radiation of the *Proteobacteria* (186).

In the genomic era, inferences about the functions of genes identified in genetic screens or sequencing initiatives drive many of the predictions about how physiological processes occur. In the case of MamO, these assumptions were completely incapable of predicting its true mechanistic role in magnetosome formation. Biochemical characterization provided functional insights that led to the discovery of an evolutionary phenomenon on a much larger scale. It is an empowering reminder that there is no substitute for experimental analysis of molecular mechanism for understanding how a cellular process occurs and that insights from these studies can still identify wide-ranging conclusions about biology. Within the magnetosome community, using biochemistry to characterize individual biomineralization factors is clearly the most effective way to fill in gaps in current the step-wise genetic model. More broadly, the existing genomic information allows these experiments to clarify the history of magnetosome evolution and potentially define broad mechanisms for how cellular sub-structures evolve.

REFERENCES

1. Lowenstam, H. A. (1981) Minerals formed by organisms. *Science*. **211**, 1126–1131
2. Weiner, S. (2003) An Overview of Biomineralization Processes and the Problem of the Vital Effect. *Reviews in Mineralogy and Geochemistry*. **54**, 1–29
3. Weiner, S. (2008) Biomineralization: A structural perspective. *Journal of Structural Biology*. **163**, 229–234
4. Addadi, L., Joester, D., Nudelman, F., and Weiner, S. (2006) Mollusk Shell Formation: A Source of New Concepts for Understanding Biomineralization Processes. *Chem. Eur. J.* **12**, 980–987
5. Kröger, N., and Poulsen, N. (2008) Diatoms—From Cell Wall Biogenesis to Nanotechnology. *Annu. Rev. Genet.* **42**, 83–107
6. Brownlee, C., Wheeler, G. L., and Taylor, A. R. (2015) Coccolithophore biomineralization: New questions, new answers. *Seminars in Cell and Developmental Biology*. **46**, 11–16
7. Rahn-Lee, L., and Komeili, A. (2013) The magnetosome model: insights into the mechanisms of bacterial biomineralization. *Front Microbiol*. 10.3389/fmicb.2013.00352/abstract
8. Monroe, M., and Abrams, K. (1984) A perspective on solubility rules. *J. Chem. Educ.* **61**, 885–1
9. Hildebrand, M., Volcani, B. E., Gassmann, W., and Schroeder, J. I. (1997) A gene family of silicon transporters. *Nature*. **385**, 688–689
10. Holtz, L.-M., Thoms, S., Langer, G., and Wolf-Gladrow, D. A. (2013) Substrate supply for calcite precipitation in *Emiliania huxleyi*: assessment of different model approaches. *J. Phycol.* **49**, 417–426
11. Van Cappellen, P. (2003) Biomineralization and Global Biogeochemical Cycles. *Reviews in Mineralogy and Geochemistry*. **54**, 357–381
12. Sweeney, R. Y., Mao, C., Gao, X., Burt, J. L., Belcher, A. M., Georgiou, G., and Iverson, B. L. (2004) Bacterial Biosynthesis of Cadmium Sulfide Nanocrystals. *Chemistry & Biology*. **11**, 1553–1559
13. Zhan, G., Li, D., and Zhang, L. (2012) Aerobic bioreduction of nickel(II) to elemental nickel with concomitant biomineralization. *Appl Microbiol Biotechnol*. **96**, 273–281
14. Faivre, D., and Godec, T. U. (2015) From Bacteria to Mollusks: The Principles Underlying the Biomineralization of Iron Oxide Materials. *Angew. Chem.* **127**, 4810–4829
15. Murshed, M., Harmey, D., Millán, J. L., McKee, M. D., and Karsenty, G. (2005) Unique coexpression in osteoblasts of broadly expressed genes accounts for the spatial restriction of ECM mineralization to bone. *Genes & Development*. **19**, 1093–1104
16. Schüler, D. (2008) Genetics and cell biology of magnetosome formation in magnetotactic bacteria. *FEMS Microbiology Reviews*. **32**, 654–672
17. Salata, O. V. (2004) Applications of nanoparticles in biology and medicine. *J Nanobiotechnol*. **2**, 3–6
18. Baumgartner, J., Bertinetti, L., Widdrat, M., Hirt, A. M., and Faivre, D. (2013) Formation of Magnetite Nanoparticles at Low Temperature: From Superparamagnetic to Stable Single Domain Particles. *PLoS ONE*. **8**, e57070–6
19. Komeili, A. (2011) Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiology Reviews*. **36**, 232–255

20. Blakemore, R. (1975) Magnetotactic Bacteria. *Science*. **190**, 377–379
21. Lefèvre, C. T., Bennet, M., Landau, L., Vach, P., Pignol, D., Bazylinski, D. A., Frankel, R. B., Klumpp, S., and Faivre, D. (2014) Diversity of Magneto-Aerotactic Behaviors and Oxygen Sensing Mechanisms in Cultured Magnetotactic Bacteria. *Biophysj*. **107**, 527–538
22. Frankel, R. B., Blakemore, R. P., and Wolfe, R. S. (1979) Magnetite in Freshwater Magnetotactic Bacteria. *Science*. **203**, 1355–1356
23. Lefèvre, C. T., Menguy, N., Abreu, F., Lins, U., Pósfai, M., Prozorov, T., Pignol, D., Frankel, R. B., and Bazylinski, D. A. (2011) A Cultured Greigite-Producing Magnetotactic Bacterium in a Novel Group of Sulfate-Reducing Bacteria. *Science*. **334**, 1720–1723
24. Bazylizinki, D. A., Heywood, B. R., Mann, S., and Frankel, R. B. (1993) Fe_3O_4 and Fe_3S_4 in a bacterium. *Nature*. **366**, 218–218
25. Bazylinski, D. A., Frankel, R. B., Heywood, B. R., Mann, S., King, J. W., Donaghay, P. L., and Hanson, A. K. (1995) Controlled Biomineralization of Magnetite (Fe_3O_4) and Greigite (Fe_3S_4) in a Magnetotactic Bacterium. *Appl Environ Microbiol*. **61**, 3232–3239
26. Frankel, R. B., Bazylinski, D. A., Johnson, M. S., and Taylor, B. L. (1997) Magneto-aerotaxis in marine coccoid bacteria. *Biophysical Journal*. **73**, 994–1000
27. Lefevre, C. T., and Bazylinski, D. A. (2013) Ecology, Diversity, and Evolution of Magnetotactic Bacteria. *Microbiology and Molecular Biology Reviews*. **77**, 497–526
28. Matsunaga, T., Okamura, Y., Fukuda, Y., Wahyudi, A. T., Murase, Y., and Takeyama, H. (2005) Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum sp.* strain AMB-1. *DNA Res.* **12**, 157–166
29. Lefèvre, C. T., Viloria, N., Schmidt, M. L., Pósfai, M., Frankel, R. B., and Bazylinski, D. A. (2011) Novel magnetite-producing magnetotactic bacteria belonging to the *Gammaproteobacteria*. *The ISME Journal*. **6**, 440–450
30. Lefèvre, C. T., Trubitsyn, D., Abreu, F., Kolinko, S., Jogler, C., de Almeida, L. G. P., de Vasconcelos, A. T. R., Kube, M., Reinhardt, R., Lins, U., Pignol, D., Schüler, D., Bazylinski, D. A., and Ginet, N. (2013) Comparative genomic analysis of magnetotactic bacteria from the *Deltaproteobacteria* provides new insights into magnetite and greigite magnetosome genes required for magnetotaxis. *Environ Microbiol*. **15**, 2712–2735
31. Vali, H., Förster, O., Amarantidis, G., and Petersen, N. (1987) Magnetotactic bacteria and their magnetofossils in sediments. *Earth and Planetary Science Letters*. **86**, 389–400
32. 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
33. Kolinko, S., Jogler, C., Katzmann, E., Wanner, G., Peplies, J., and Schüler, D. (2011) Single-cell analysis reveals a novel uncultivated magnetotactic bacterium within the candidate division OP3. *Environ Microbiol*. **14**, 1709–1721
34. Blakemore, R. P., Maratea, D., and Wolfe, R. S. (1979) Isolation and Pure Culture of a Freshwater Magnetic Spirillum in Chemically Defined Medium. *J Bacteriol*. **140**, 720–729
35. Balkwill, D. L., Maratea, D., and Blakemore, R. P. (1980) Ultrastructure of a magnetotactic spirillum. *J Bacteriol*. **141**, 1399–1408
36. Gorby, Y. A., Beveridge, T. J., and Blakemore, R. P. (1988) Characterization of the Bacterial Magnetosome Membrane. *J Bacteriol*. **170**, 834–841

37. Matsunaga, T., Sakaguchi, T., and Tadakoro, F. (1991) Magnetite formation by a magnetic bacterium capable of growing aerobically. *Appl Microbiol Biotechnol*. **35**, 1–5
38. Schuler, D., and Köhler, M. (1992) The isolation of a new magnetic spirillum. *Zentralblatt fr Mikrobiologie : Landwirtschaft, Technologie, Umweltschutz*. **147**, 150–151
39. Schubbe, S., Kube, M., Scheffel, A., Wawer, C., Heyen, U., Meyerdierks, A., Madkour, M. H., Mayer, F., Reinhardt, R., and Schuler, D. (2003) Characterization of a Spontaneous Nonmagnetic Mutant of *Magnetospirillum gryphiswaldense* Reveals a Large Deletion Comprising a Putative Magnetosome Island. *J Bacteriol*. **185**, 5779–5790
40. 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
41. 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
42. 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
43. Fukuda, Y., Okamura, Y., Takeyama, H., and Matsunaga, T. (2006) Dynamic analysis of a genomic island in *Magnetospirillum sp.* strain AMB-1 reveals how magnetosome synthesis developed. *FEBS Letters*. **580**, 801–812
44. Schubbe, S., Williams, T. J., Xie, G., Kiss, H. E., Brettin, T. S., Martinez, D., Ross, C. A., Schuler, D., Cox, B. L., Nealson, K. H., and Bazylinski, D. A. (2009) Complete Genome Sequence of the Chemolithoautotrophic Marine Magnetotactic Coccus Strain MC-1. *Appl Environ Microbiol*. **75**, 4835–4852
45. 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
46. 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
47. 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
48. DeLong, E. F., Frankel, R. B., and Bazylinski, D. A. (1993) Multiple Evolutionary Origins of Magnetotaxis in Bacteria. *Science*. **259**, 803–806
49. Abreu, F., Cantão, M. E., Nicolás, M. F., Barcellos, F. G., Morillo, V., Almeida, L. G., do Nascimento, F. F., Lefèvre, C. T., Bazylinski, D. A., de Vasconcelos, A. T. R., and Lins, U. (2011) Common ancestry of iron oxide- and iron-sulfide- based biomineralization in magnetotactic bacteria. *The ISME Journal*. **5**, 1634–1640
50. Jogler, C., Wanner, G., Kolinko, S., Niebler, M., Amann, R., Petersen, N., Kube, M., Reinhardt, R., and Schuler, D. (2011) Conservation of proteobacterial magnetosome

genes and structures in an uncultivated member of the deep-branching *Nitrospira* phylum. *Proc Natl Acad Sci U S A*. **108**, 1134–1139
51. Lefèvre, C. T., Trubitsyn, D., Abreu, F., Kolinko, S., de Almeida, L. G. P., de Vasconcelos, A. T. R., Lins, U., Schüler, D., Ginet, N., Pignol, D., and Bazylinski, D. A. (2013) Monophyletic origin of magnetotaxis and the first magnetosomes. *Environ Microbiol*. **15**, 2267–2274
52. Williams, K. P., Sobral, B. W., and Dickerman, A. W. (2007) A Robust Species Tree for the *Alphaproteobacteria*. *J Bacteriol*. **189**, 4578–4586
53. Matsunaga, T., Nakamura, C., Burgess, J. G., and Sode, K. (1992) Gene Transfer in Magnetic Bacteria: Transposon Mutagenesis and Cloning of Genomic DNA Fragments Required for Magnetosome Synthesis. *J Bacteriol*
54. 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
55. Oikonomou, C. M., and Jensen, G. J. (2016) A new view into prokaryotic cell biology from electron cryotomography. *Nat Rev Mol Cell Biol*. **14**, 205–220
56. 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
57. Lohße, A., Ullrich, S., Katzmann, E., Borg, S., Wanner, G., Richter, M., Voigt, B., Schweder, T., and Schüler, D. (2011) Functional Analysis of the Magnetosome Island in *Magnetospirillum gryphiswaldense*: The mamAB Operon Is Sufficient for Magnetite Biomineralization. *PLoS ONE*. **6**, e25561–10
58. Lohße, A., Borg, S., Raschdorf, O., Kolinko, I., Tompa, E., Posfai, M., Faivre, D., Baumgartner, J., and Schuler, D. (2014) Genetic Dissection of the mamAB and mms6 Operons Reveals a Gene Set Essential for Magnetosome Biogenesis in *Magnetospirillum gryphiswaldense*. *J Bacteriol*. **196**, 2658–2669
59. D'Orazio, S. E. F., Velasquez, M., Roan, N. R., Naveiras-Torres, O., and Starnbach, M. N. (2003) The *Listeria monocytogenes lemA* gene product is not required for intracellular infection or to activate fMIGWII-specific T cells. *Infect Immun*. **71**, 6721–6727
60. Paulsen, I. T., and Saier, M. H. (1997) A novel family of ubiquitous heavy metal ion transport proteins. *J. Membr. Biol.* **156**, 99–103
61. Sonkaria, S., Fuentes, G., Verma, C., Narang, R., Khare, V., Fischer, A., and Faivre, D. (2012) Insight into the Assembly Properties and Functional Organisation of the Magnetotactic Bacterial Actin-like Homolog, MamK. *PLoS ONE*. **7**, e34189–12
62. Ozyamak, E., Kollman, J., Agard, D. A., and Komeili, A. (2013) The Bacterial Actin MamK: IN VITRO ASSEMBLY BEHAVIOR AND FILAMENT ARCHITECTURE. *J Biol Chem*. **288**, 4265–4277
63. Scheffel, A., Gruska, M., Faivre, D., Linaroudis, A., Plitzko, J. M., and Schüler, D. (2005) An acidic protein aligns magnetosomes along a filamentous structure in magnetotactic bacteria. *Nature*. **440**, 110–114
64. Draper, O., Byrne, M. E., Li, Z., Keyhani, S., Barrozo, J. C., Jensen, G., and Komeili, A. (2011) MamK, a bacterial actin, forms dynamic filaments in vivo that are regulated by the acidic proteins MamJ and LimJ. *Mol Microbiol*. **82**, 342–354
65. Scheffel, A., and Schuler, D. (2007) The Acidic Repetitive Domain of the

Magnetospirillum gryphiswaldense MamJ Protein Displays Hypervariability but Is Not Required for Magnetosome Chain Assembly. *J Bacteriol*. **189**, 6437–6446
66. Raschdorf, O., Müller, F. D., Pósfai, M., Plitzko, J. M., and Schüler, D. (2013) The magnetosome proteins MamX, MamZ and MamH are involved in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *Mol Microbiol*. **89**, 872–886
67. Andrews, S. C., Robinson, A. K., and Rodríguez-Quiñones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiology Reviews*. **27**, 215–237
68. Cornell, R. M., and Schwertmann, U. (2003) *The Iron Oxides*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, 10.1002/3527602097
69. Weber, K. A., Achenbach, L. A., and Coates, J. D. (2006) Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat Rev Micro*. **4**, 752–764
70. Génin, J.-M. R., Ruby, C., Géhin, A., and Refait, P. (2006) Synthesis of green rusts by oxidation of Fe(OH)2, their products of oxidation and reduction of ferric oxyhydroxides; –pH Pourbaix diagrams. *Comptes Rendus Geoscience*. **338**, 433–446
71. Angerer, A., Gaisser, S., and Braun, V. (1990) Nucleotide sequences of the sfuA, sfuB, and sfuC genes of *Serratia marcescens* suggest a periplasmic-binding-protein-dependent iron transport mechanism. *J Bacteriol*. **172**, 572–578
72. Velayudhan, J., Hughes, N. J., McColm, A. A., Bagshaw, J., Clayton, C. L., Andrews, S. C., and Kelly, D. J. (2000) Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. *Mol Microbiol*. **37**, 274–286
73. Saha, R., Saha, N., Donofrio, R. S., and Bestervelt, L. L. (2012) Microbial siderophores: a mini review. *J. Basic Microbiol*. **53**, 303–317
74. Sugimoto, T., and Matijević, E. (1980) Formation of uniform spherical magnetite particles by crystallization from ferrous hydroxide gels. *Journal of Colloid and Interface Science*. **74**, 227–243
75. Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M. (2011) HtrA proteases: regulated proteolysis in protein quality control. *Nat Rev Mol Cell Biol*. **12**, 152–162
76. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., and Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature*. **416**, 455–459
77. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M., and Clausen, T. (2004) Crystal Structure of the DegS Stress Sensor: How a PDZ Domain Recognizes Misfolded Protein and Activates a Protease. *Cell*. **117**, 483–494
78. Quinlan, A., Murat, D., Vali, H., and Komeili, A. (2011) The HtrA/DegP family protease MamE is a bifunctional protein with roles in magnetosome protein localization and magnetite biomineralization. *Mol Microbiol*. **80**, 1075–1087
79. Weinitschke, S., Denger, K., Cook, A. M., and Smits, T. H. M. (2007) The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. *Microbiology*. **153**, 3055–3060
80. Grass, G., Fan, B., Rosen, B. P., Franke, S., Nies, D. H., and Rensing, C. (2001) ZitB (YbgR), a Member of the Cation Diffusion Facilitator Family, Is an Additional Zinc Transporter in *Escherichia coli*. *J Bacteriol*. **183**, 4664–4667
81. Guffanti, A. A., Wei, Y., Rood, S. V., and Krulwich, T. A. (2002) An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K+ and H. *Mol Microbiol*. **45**, 145–153

82. Chao, Y., and Fu, D. (2004) Kinetic Study of the Antiport Mechanism of an *Escherichia coli* Zinc Transporter, ZitB. *J Biol Chem*. **279**, 12043–12050
83. Uebe, R., Junge, K., Henn, V., Poxleitner, G., Katzmann, E., Plitzko, J. M., Zarivach, R., Kasama, T., Wanner, G., Pósfai, M., Böttger, L., Matzanke, B., and Schüler, D. (2011) The cation diffusion facilitator proteins MamB and MamM of *Magnetospirillum gryphiswaldense* have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Mol Microbiol*. **82**, 818–835
84. Lu, M., Chai, J., and Fu, D. (2009) Structural basis for autoregulation of the zinc transporter YiiP. *Nat Struct Mol Biol*. **16**, 1063–1067
85. Zeytuni, N., Uebe, R., Maes, M., Davidov, G., Baram, M., Raschdorf, O., Nadav-Tsubery, M., Kolusheva, S., Bitton, R., Goobes, G., Friedler, A., Miller, Y., Schüler, D., and Zarivach, R. (2014) Cation Diffusion Facilitators Transport Initiation and Regulation Is Mediated by Cation Induced Conformational Changes of the Cytoplasmic Domain. *PLoS ONE*. **9**, e92141–14
86. Nozaki, K., Kuroda, T., Mizushima, T., and Tsuchiya, T. (1998) A new Na+/H+ antiporter, NhaD, of *Vibrio parahaemolyticus*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1369**, 213–220
87. Habibian, R., Dzioba, J., Barrett, J., Galperin, M. Y., Loewen, P. C., and Dibrov, P. (2005) Functional Analysis of Conserved Polar Residues in Vc-NhaD, Na+/H+ Antiporter of *Vibrio cholerae*. *J Biol Chem*. **280**, 39637–39643
88. Prakash, S., Cooper, G., Singhi, S., and Saier, M. H., Jr. (2003) The ion transporter superfamily. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1618**, 79–92
89. Murat, D., Falahati, V., Bertinetti, L., Csencsits, R., Körnig, A., Downing, K., Faivre, D., and Komeili, A. (2012) The magnetosome membrane protein, MmsF, is a major regulator of magnetite biomineralization in *Magnetospirillum magneticum* AMB-1. *Mol Microbiol*. **85**, 684–699
90. Pignol, P. A. A. D. (2014) Structure and evolution of the magnetochrome domains: no longer alone. 10.3389/fmicb.2014.00117/abstract
91. Stevens, J. M., Daltrop, O., Allen, J. W. A., and Ferguson, S. J. (2004) C-type Cytochrome Formation: Chemical and Biological Enigmas. *Acc. Chem. Res.* **37**, 999–1007
92. Siponen, M. I., Adryanczyk, G., Ginet, N., Arnoux, P., and Pignol, D. (2012) Magnetochrome: a c-type cytochrome domain specific to magnetotatic bacteria. *Biochm. Soc. Trans.* **40**, 1319–1323
93. Siponen, M. I., Legrand, P., Widdrat, M., Jones, S. R., Zhang, W.-J., Chang, M. C. Y., Faivre, D., Arnoux, P., and Pignol, D. (2014) Structural insight into magnetochrome-mediated magnetite biomineralization. *Nature*. **502**, 681–684
94. Calugay, R. J., Okamura, Y., Wahyudi, A. T., Takeyama, H., and Matsunaga, T. (2004) Siderophore production of a periplasmic transport binding protein kinase gene defective mutant of *Magnetospirillum magneticum* AMB-1. *Biochemical and Biophysical Research Communications*. **323**, 852–857
95. Suzuki, T., Okamura, Y., Calugay, R. J., Takeyama, H., and Matsunaga, T. (2006) Global Gene Expression Analysis of Iron-Inducible Genes in *Magnetospirillum magneticum* AMB-1. *J Bacteriol*. **188**, 2275–2279
96. Amemiya, Y., Arakaki, A., Staniland, S. S., Tanaka, T., and Matsunaga, T. (2007) Controlled formation of magnetite crystal by partial oxidation of ferrous hydroxide in

the presence of recombinant magnetotactic bacterial protein Mms6. *Biomaterials*. **28**, 5381–5389
97. Wang, L., Prozorov, T., Palo, P. E., Liu, X., Vaknin, D., Prozorov, R., Mallapragada, S., and Nilsen-Hamilton, M. (2012) Self-Assembly and Biphasic Iron-Binding Characteristics of Mms6, A Bacterial Protein That Promotes the Formation of Superparamagnetic Magnetite Nanoparticles of Uniform Size and Shape. *Biomacromolecules*. **13**, 98–105
98. Kashyap, S., Woehl, T. J., Liu, X., Mallapragada, S. K., and Prozorov, T. (2014) Nucleation of Iron Oxide Nanoparticles Mediated by Mms6 Protein in Situ. *ACS Nano*. **8**, 9097–9106
99. Arakaki, A., Masuda, F., Amemiya, Y., Tanaka, T., and Matsunaga, T. (2010) Control of the morphology and size of magnetite particles with peptides mimicking the Mms6 protein from magnetotactic bacteria. *Journal of Colloid and Interface Science*. **343**, 65–70
100. Bird, S. M., Rawlings, A. E., Galloway, J. M., and Staniland, S. S. (2016) Using a biomimetic membrane surface experiment to investigate the activity of the magnetite biomineralisation protein Mms6. *RSC Advances*. **6**, 7356–7363
101. Rawlings, A. E., Bramble, J. P., Walker, R., Bain, J., Galloway, J. M., and Staniland, S. S. (2014) Self-assembled MmsF proteinosomes control magnetite nanoparticle formation in vitro. *Proc Natl Acad Sci U S A*. **111**, 16094–16099
102. Arakaki, A., Yamagishi, A., Fukuyo, A., Tanaka, M., and Matsunaga, T. (2014) Co-ordinated functions of Mms proteins define the surface structure of cubo-octahedral magnetite crystals in magnetotactic bacteria. *Mol Microbiol*. **93**, 554–567
103. Ding, Y., Li, J., Liu, J., Yang, J., Jiang, W., Tian, J., Li, Y., Pan, Y., and Li, J. (2010) Deletion of the ftsZ-Like Gene Results in the Production of Superparamagnetic Magnetite Magnetosomes in *Magnetospirillum gryphiswaldense*. *J Bacteriol*. **192**, 1097–1105
104. Nakamura, C., Kikuchi, T., Burgess, J. G., and Matsunaga, T. (1995) Iron-regulated expression and membrane localization of the magA protein in Magnetospirillum sp. strain AMB-1. *J. Biochem.* **118**, 23–27
105. 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
106. Zurkiya, O., Chan, A. W. S., and Hu, X. (2008) MagA is sufficient for producing magnetic nanoparticles in mammalian cells, making it an MRI reporter. *Magn. Reson. Med.* **59**, 1225–1231
107. Uebe, R., Henn, V., and Schuler, D. (2012) The MagA Protein of *Magnetospirilla* Is Not Involved in Bacterial Magnetite Biomineralization. *J Bacteriol*. **194**, 1018–1023
108. Uebe, R., Voigt, B., Schweder, T., Albrecht, D., Katzmann, E., Lang, C., Bottger, L., Matzanke, B., and Schuler, D. (2010) Deletion of a fur-Like Gene Affects Iron Homeostasis and Magnetosome Formation in *Magnetospirillum gryphiswaldense*. *J Bacteriol*. **192**, 4192–4204
109. Qi, L., Li, J., Zhang, W., Liu, J., Rong, C., Li, Y., and Wu, L. (2012) Fur in *Magnetospirillum gryphiswaldense* Influences Magnetosomes Formation and Directly Regulates the Genes Involved in Iron and Oxygen Metabolism. *PLoS ONE*. **7**, e29572–9
110. Schubbe, S., Wurdemann, C., Peplies, J., Heyen, U., Wawer, C., Glockner, F. O., and

Schuler, D. (2006) Transcriptional Organization and Regulation of Magnetosome Operons in *Magnetospirillum gryphiswaldense*. *Appl Environ Microbiol*. **72**, 5757–5765
111. 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. *Research in Microbiology*. **159**, 530–536
112. Rong, C., Zhang, C., Zhang, Y., Qi, L., Yang, J., Guan, G., Li, Y., and Li, J. (2012) FeoB2 Functions in Magnetosome Formation and Oxidative Stress Protection in *Magnetospirillum gryphiswaldense* Strain MSR-1. *J Bacteriol*. **194**, 3972–3976
113. Li, Y., Katzmann, E., Borg, S., and Schuler, D. (2012) The Periplasmic Nitrate Reductase Nap Is Required for Anaerobic Growth and Involved in Redox Control of Magnetite Biomineralization in *Magnetospirillum gryphiswaldense*. *J Bacteriol*. **194**, 4847–4856
114. Li, Y., Bali, S., Borg, S., Katzmann, E., Ferguson, S. J., and Schuler, D. (2013) Cytochrome cd1 Nitrite Reductase NirS Is Involved in Anaerobic Magnetite Biomineralization in *Magnetospirillum gryphiswaldense* and Requires NirN for Proper d1 Heme Assembly. *J Bacteriol*. **195**, 4297–4309
115. Li, Y., Raschdorf, O., Silva, K. T., and Schuler, D. (2014) The Terminal Oxidase cbb3 Functions in Redox Control of Magnetite Biomineralization in *Magnetospirillum gryphiswaldense*. *J Bacteriol*. **196**, 2552–2562
116. Kolinko, I., Lohße, A., Borg, S., Raschdorf, O., Jogler, C., Tu, Q., Pósfai, M., Tompa, É., Plitzko, J. M., Brachmann, A., Wanner, G., Müller, R., Zhang, Y., and Schüler, D. (2014) Biosynthesis of magnetic nanostructures in a foreign organism by transfer of bacterial magnetosome gene clusters. *Nature Nanotechnology*. **9**, 193–197
117. Murat, D., Byrne, M., and Komeili, A. (2010) Cell Biology of Prokaryotic Organelles. *Cold Spring Harbor Perspectives in Biology*. **2**, a000422–a000422
118. Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., and Banfield, J. F. (2016) A new view of the tree of life. *Nature Microbiology*. **1**, 1–6
119. Byrne, M. E., Ball, D. A., Guerquin-Kern, J.-L., Rouiller, I., Wu, T.-D., Downing, K. H., Vali, H., and Komeili, A. (2010) *Desulfovibrio magneticus* RS-1 contains an iron- and phosphorus-rich organelle distinct from its bullet-shaped magnetosomes. *Proc Natl Acad Sci U S A*. **107**, 12263–12268
120. Lin, W., Deng, A., Wang, Z., Li, Y., Wen, T., Wu, L.-F., Wu, M., and Pan, Y. (2014) Genomic insights into the uncultured genus "*Candidatus Magnetobacterium*" in the phylum *Nitrospirae*. *The ISME Journal*. **8**, 2463–2477
121. Lefevre, C. T., and Bazylinski, D. A. (2013) Ecology, Diversity, and Evolution of Magnetotactic Bacteria. *Microbiology and Molecular Biology Reviews*. **77**, 497–526
122. Zarivach, R. (2014) Structure prediction of magnetosome-associated proteins. *Front Microbiol*. 10.3389/fmicb.2014.00009/abstract
123. Yang, W., Li, R., Peng, T., Zhang, Y., Jiang, W., Li, Y., and Li, J. (2010) *mamO* and *mamE* genes are essential for magnetosome crystal biomineralization in *Magnetospirillum gryphiswaldense* MSR-1. *Res Microbiol*. **161**, 701–705
124. Jones, S. R., Wilson, T. D., Brown, M. E., Rahn-Lee, L., Yu, Y., Fredriksen, L. L., Ozyamak, E., Komeili, A., and Chang, M. C. Y. (2015) Genetic and biochemical

investigations of the role of MamP in redox control of iron biomineralization in *Magnetospirillum magneticum*. *Proc Natl Acad Sci U S A*. **112**, 3904–3909
125. Otwinoski, Z., and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Meth Enzymol*. **276**, 307–326
126. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. **66**, 486–501
127. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. **66**, 213–221
128. Yu, X., Wu, X., Bermejo, G. A., Brooks, B. R., and Taraska, J. W. (2013) Accurate High-Throughput Structure Mapping and Prediction with Transition Metal Ion FRET. *Structure*. **21**, 9–19
129. Eddy, S. R. (1998) Profile hidden Markov models. *Bioinformatics*. **14**, 755–763
130. Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **22**, 1658–1659
131. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**, 1792–1797
132. Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol*. **17**, 540–552
133. Sassera, D., Lo, N., Epis, S., D'Auria, G., Montagna, M., Comandatore, F., Horner, D., Pereto, J., Luciano, A. M., Franciosi, F., Ferri, E., Crotti, E., Bazzocchi, C., Daffonchio, D., Sacchi, L., Moya, A., Latorre, A., and Bandi, C. (2011) Phylogenomic Evidence for the Presence of a Flagellum and cbb3 Oxidase in the Free-Living Mitochondrial Ancestor. *Mol Biol Evol*. **28**, 3285–3296
134. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*. **5**, e9940
135. Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. **21**, 2104–2105
136. Stamatakis, A. (2014) RAxML Version 8: a Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*. **30**, 1312–1313
137. Lartillot, N., Lepage, T., and Blanquart, S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. **25**, 2286–2288
138. Perona, J. J., and Craik, C. S. (1997) Evolutionary Divergence of Substrate Specificity within the Chymotrypsin-like Serine Protease Fold. *J Biol Chem*. **272**, 29987–29990
139. Sohn, J., Grant, R. A., and Sauer, R. T. (2010) Allostery Is an Intrinsic Property of the Protease Domain of DegS: IMPLICATIONS FOR ENZYME FUNCTION AND EVOLUTION. *J Biol Chem*. **285**, 34039–34047
140. Wrase, R., Scott, H., Hilgenfeld, R., and Hansen, G. (2011) The *Legionella* HtrA homologue DegQ is aself-compartmentizing protease thatforms large 12-meric assemblies. *Proc Natl Acad Sci U S A*. **108**, 10490–10495
141. Bobofchak, K. M., Pineda, A. O., Mathews, F. S., and Di Cera, E. (2005) Energetic and Structural Consequences of Perturbing Gly-193 in the Oxyanion Hole of Serine

Proteases. *J Biol Chem*. **280**, 25644–25650
142. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I. W., Prisant, M. G., and Richardson, J. S. (2003) Structure Validation by Cα Geometry: ϕ, ψ, and Cβ Deviation. *Proteins*. **50**, 437–450
143. Carvalho, A. L., Sanz, L., Barettino, D., Romero, A., Calvete, J. J., and Romão, M. J. (2002) Crystal Structure of a Prostate Kallikrein Isolated from Stallion Seminal Plasma: A Homologue of Human PSA. *J Mol Biol*. **322**, 325–337
144. Goettig, P., Magdolen, V., and Brandstetter, H. (2010) Natural and synthetic inhibitors of kallikrein-related peptidases (KLKs). *Biochimie*. **92**, 1546–1567
145. Taraska, J. W., Puljung, M. C., Olivier, N. B., Flynn, G. E., and Zagotta, W. N. (2009) Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nature Methods*. **6**, 532–537
146. Mayer, J., Denger, K., Hollemeyer, K., Schleheck, D., and Cook, A. M. (2012) (R)-Cysteate-nitrogen assimilation by *Cupriavidus necator* H16 with excretion of 3-sulfolactate: a patchwork pathway. *Arch Microbiol*. **194**, 949–957
147. Espinosa-Cantú, A., Ascencio, D., Barona-Gómez, F., and DeLuna, A. (2015) Gene duplication and the evolution of moonlighting proteins. *Front. Genet.* **6**, 1–7
148. Morrone, D., Hillwig, M. L., Mead, M. E., Lowry, L., Fulton, D. B., and Peters, R. J. (2011) Evident and latent plasticity across the rice diterpene synthase family with potential implications for the evolution of diterpenoid metabolism in the cereals. *Biochem. J.* **435**, 589–595
149. Ben-David, M., Wieczorek, G., Elias, M., Silman, I., Sussman, J. L., and Tawfik, D. S. (2013) Catalytic Metal Ion Rearrangements Underline Promiscuity and Evolvability of a Metalloenzyme. *J Mol Biol*. **425**, 1028–1038
150. Dellus-Gur, E., Toth-Petroczy, A., Elias, M., and Tawfik, D. S. (2013) What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-offs. *J Mol Biol*. **425**, 2609–2621
151. Toth-Petroczy, A., and Tawfik, D. S. (2014) The robustness and innovability of protein folds. *Current Opinion in Structural Biology*. **26**, 131–138
152. Truebestein, L., Tennstaedt, A., Mönig, T., Krojer, T., Canellas, F., Kaiser, M., Clausen, T., and Ehrmann, M. (2011) Substrate-induced remodeling of the active site regulates human HTRA1 activity. *Nat Struct Mol Biol*. **18**, 386–388
153. Krojer, T., Sawa, J., Huber, R., and Clausen, T. (2010) HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues. *Nat Struct Mol Biol*. **17**, 844–852
154. Eigenbrot, C., Ultsch, M., Lipari, M. T., Moran, P., Lin, S. J., Ganesan, R., Quan, C., Tom, J., Sandoval, W., van Lookeren Campagne, M., and Kirchhofer, D. (2012) Structural and Functional Analysis of HtrA1 and Its Subdomains. *Structure*. **20**, 1040–1050
155. Hecht, H. J., Szardenings, M., Collins, J., and Schomburg, D. (1991) Three-dimensional Structure of the Complexes between Bovine Chymotrypsinogen A and Two Recombinant Variants of Human Pancreatic Secretory Trypsin Inhibitor (Kazal-type). *J Mol Biol*. **220**, 711–722
156. Biswas, T., Small, J., Vandal, O., Odaira, T., Deng, H., Ehrt, S., and Tsodikov, O. V. (2010) Structural Insight into Serine Protease Rv3671c that Protects *M. tuberculosis* from Oxidative and Acidic Stress. *Structure*. **18**, 1353–1363

157. Kim, S., Grant, R. A., and Sauer, R. T. (2011) Covalent Linkage of Distinct Substrate Degrons Controls Assembly and Disassembly of DegP Proteolytic Cages. *Cell*. **145**, 67–78
158. Rahn-Lee, L., Byrne, M. E., Zhang, M., Le Sage, D., Glenn, D. R., Milbourne, T., Walsworth, R. L., Vali, H., and Komeili, A. (2015) A Genetic Strategy for Probing the Functional Diversity of Magnetosome Formation. *PLoS Genet*. **11**, e1004811–18
159. Laura, R. P., Witt, A. S., Held, H. A., Gerstner, R., Deshayes, K., Koehler, M. F. T., Kosik, K. S., Sidhu, S. S., and Lasky, L. A. (2002) The Erbin PDZ Domain Binds with High Affinity and Specificity to the Carboxyl Termini of -Catenin and ARVCF. *J Biol Chem*. **277**, 12906–12914
160. Wolff, A., Frese, K., Wißbrock, M., Eckstädt, K., Ennen, I., Hetaba, W., Löffler, S., Regtmeier, A., Thomas, P., Sewald, N., Schattschneider, P., and Hütten, A. (2012) Influence of the synthetic polypeptide c25-mms6 on cobalt ferrite nanoparticle formation. *J Nanopart Res*. **14**, 1161–11
161. Li, W., Srinivasula, S. M., Chai, J., Li, P., Wu, J.-W., Zhang, Z., Alnemri, E. S., and Shi, Y. (2002) Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nat. Struct Biol.* **9**, 436–441
162. Sohn, J., Grant, R. A., and Sauer, R. T. (2007) Allosteric Activation of DegS, a Stress Sensor PDZ Protease. *Cell*. **131**, 572–583
163. Krojer, T., Sawa, J., Schäfer, E., Saibil, H. R., Ehrmann, M., and Clausen, T. (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature*. **453**, 885–890
164. Hershey, D. M., Ren, X., Melnyk, R. A., Browne, P. J., Ozyamak, E., Jones, S. R., Chang, M. C. Y., Hurley, J. H., and Komeili, A. (2016) MamO Is a Repurposed Serine Protease that Promotes Magnetite Biomineralization through Direct Transition Metal Binding in Magnetotactic Bacteria. *PLoS Biol*. **14**, e1002402
165. Roepstroff, P. (1984) Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrom*. **11**, 601
166. Arslan, E., Schulz, H., Zufferey, R., Kunzler, P., and Thony-Meyer, L. (1998) Overproduction of the *Bradyrhizobium japonicum* c-Type Cytochrome Subunits of the cbb3 Oxidase in Escherichia coli. *Biochemical and Biophysical Research Communications*
167. Held, H. A., and Sidhu, S. S. (2004) Comprehensive Mutational Analysis of the M13 Major Coat Protein: Improved Scaffolds for C-terminal Phage Display. *J Mol Biol*. **340**, 587–597
168. Arita, Y., Allen, S., Chen, G., Zhang, W., Wang, Y., Owen, A. J., Dentinger, P., and Sidhu, S. S. (2016) Rapid isolation of peptidic inhibitors of the solute carrier family transporters OATP1B1 and OATP1B3 by cell-based phage display selections. *Biochemical and Biophysical Research Communications*. 10.1016/j.bbrc.2016.01.050
169. Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S. S. (2007) Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat Protoc*. **2**, 1368–1386
170. Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J.-H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D., and Sidhu, S. S. (2008) A Specificity Map for the PDZ Domain Family. *PLoS Biol*. **6**, e239–17

171. Londer, Y. Y., Pokkuluri, P. R., Orshonsky, V., Orshonsky, L., and Schiffer, M. (2006) Heterologous expression of dodecaheme “nanowire” cytochromes c from *Geobacter sulfurreducens*. *Protein Expression and Purification*. **47**, 241–248
172. Mauldin, R. V., and Sauer, R. T. (2012) Allosteric regulation of DegS protease subunits through a shared energy landscape. *Nature Chemical Biology*. **9**, 90–96
173. Small, J. L., O'Donoghue, A. J., Boritsch, E. C., Tsodikov, O. V., Knudsen, G. M., Vandal, O., Craik, C. S., and Ehrt, S. (2013) Substrate Specificity of MarP, a Periplasmic Protease Required for Resistance to Acid and Oxidative Stress in *Mycobacterium tuberculosis*. *J Biol Chem*. **288**, 12489–12499
174. Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A., and Sidhu, S. S. (2000) Analysis of PDZ Domain-Ligand Interactions Using Carboxyl-terminal Phage Display. *J Biol Chem*. **275**, 21486–21491
175. Zhang, Y., Appleton, B. A., Wu, P., Wiesmann, C., and Sidhu, S. S. (2007) Structural and functional analysis of the ligand specificity of the HtrA2/Omi PDZ domain. *Protein Sci.* **16**, 1738–1750
176. Runyon, S. T., Zhang, Y., Appleton, B. A., Sazinsky, S. L., Wu, P., Pan, B., Wiesmann, C., Skelton, N. J., and Sidhu, S. S. (2007) Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. *Protein Sci.* **16**, 2454–2471
177. Ivarsson, Y., Arnold, R., McLaughlin, M., Nim, S., Joshi, R., Ray, D., Liu, B., Teyra, J., Pawson, T., Moffat, J., Li, S. S. C., Sidhu, S. S., and Kim, P. M. (2014) Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc Natl Acad Sci U S A*. **111**, 2542–2547
178. Lee, H.-J., and Zheng, J. J. (2010) PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal*. **8**, 8–18
179. MohamedMohaideen, N. N., Palaninathan, S. K., Morin, P. M., Williams, B. J., Braunstein, M., Tichy, S. E., Locker, J., Russell, D. H., Jacobs, W. R., and Sacchettini, J. C. (2008) Structure and Function of the Virulence-Associated High-Temperature Requirement A of *Mycobacterium tuberculosis*. *Biochemistry*. **47**, 6092–6102
180. Truebestein, L., Tennstaedt, A., Mönig, T., Krojer, T., Canellas, F., Kaiser, M., Clausen, T., and Ehrmann, M. (2011) Substrate-induced remodeling of the active site regulates human HtrA1 activity. *Nat Rev Mol Cell Biol*. **18**, 386–388
181. Kley, J., Schmidt, B., Boyanov, B., Stolt-Bergner, P. C., Kirk, R., Ehrmann, M., Knopf, R. R., Naveh, L., Adam, Z., and Clausen, T. (2011) Structural adaptation of the plant protease Deg1 to repair photosystem II during light exposure. *Nat Rev Mol Cell Biol*. **18**, 728–731
182. Kim, S., and Sauer, R. T. (2014) Distinct regulatory mechanisms balance DegP proteolysis to maintain cellular fitness during heat stress. *Genes & Development*. **28**, 902–911
183. Cornejo, E., Subramanian, P., Li, Z., Jensen, G. J., and Komeili, A. (2016) Dynamic Remodeling of the Magnetosome Membrane Is Triggered by the Initiation of Biomineralization. *mBio*. **7**, e01898–15–9
184. Hershey, D. M., Browne, P. J., Iavarone, A. T., Teyra, J., Lee, E. H., Sidhu, S. S., and Komeili, A. (2016) *Magnetite biomineralization in Magnetospirillum magneticum is regulated by a switch-like behavior in the HtrA protease MamE*, 10.1101/047555
185. Chen, Y., Sheng, R., Källberg, M., Silkov, A., Tun, M. P., Bhardwaj, N., Kurilova, S., Hall, R. A., Honig, B., Lu, H., and Cho, W. (2012) Genome-wide Functional Annotation

of Dual-Specificity Protein- and Lipid-Binding Modules that Regulate Protein Interactions. *Molecular Cell*. **46**, 226–237

186. Cavalier-Smith, T. (2006) Cell evolution and Earth history: stasis and revolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **361**, 969–1006