

## Structure and function of a 9.6 megadalton bacterial iron storage compartment

T. W. Giessen<sup>1,2,3\*</sup>, B. J. Orlando<sup>4</sup>, A. A. Verdegaal<sup>1,2,†</sup>, M. G. Chambers<sup>4,‡</sup>, J. Gardener<sup>5</sup>, D. C. Bell<sup>5,6</sup>, G. Birrane<sup>7</sup>, M. Liao<sup>4\*</sup>, P. A. Silver<sup>1,2\*</sup>

### Affiliations:

<sup>1</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

<sup>2</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.

<sup>3</sup>Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

<sup>4</sup>Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

<sup>5</sup>Center for Nanoscale Systems, Harvard University, Cambridge, MA 02138, USA.

<sup>6</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02115, USA.

<sup>7</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, 02215, USA.

\*Correspondence to: [tgiessen@umich.edu](mailto:tgiessen@umich.edu), [maofu\\_liao@hms.harvard.edu](mailto:maofu_liao@hms.harvard.edu), [pamela\\_silver@hms.harvard.edu](mailto:pamela_silver@hms.harvard.edu)

†Current address: Program in Biological and Biomedical Sciences, Yale University, New Haven, CT 06520, USA.

‡Current address: Center for Structural Biology, Vanderbilt University, Nashville TN 37240, USA.

**Abstract:** Iron storage proteins are essential for maintaining intracellular iron homeostasis and redox balance. Iron is generally stored in a soluble and bioavailable form inside ferritin protein compartments. However, some organisms do not encode ferritins and thus rely on alternative storage strategies. Encapsulins, a class of protein-based organelles, have recently been implicated in microbial iron and redox metabolism. Here, we report the structural and mechanistic characterization of a 42 nm two-component encapsulin-based iron storage compartment from *Quasibacillus thermotolerans*. Using cryo-electron microscopy and x-ray crystallography, we reveal the assembly principles of a thermostable T = 4 shell topology and its catalytic ferroxidase cargo. We show that the cargo-loaded compartment has an exceptionally large iron storage capacity storing over 23,000 iron atoms. These results form the basis for understanding alternate microbial strategies for dealing with the essential element iron.

Iron is essential to all organisms on earth. However, the same properties that make iron useful for cellular metabolism can result in toxicity under aerobic conditions (1). Ferrous iron ( $\text{Fe}^{2+}$ ) is easily oxidized to insoluble ferric iron ( $\text{Fe}^{3+}$ ) resulting in the formation of harmful precipitates and reactive oxygen species (ROS) (2). Cells have evolved to cope with these problems by strictly controlling the intracellular concentration and reactivity of free iron (3). Ferritin proteins are used as the main iron storage system by animals, plants and most microbes (4). The main ferritin-like proteins involved in iron storage are ferritin (Ftn), bacterioferritin (Bfr) and DNA-binding proteins from starved cells (Dps) all able to oxidize  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$  via a ferroxidase activity (5). While Ftn and Bfr are primarily used as a dynamic iron storage (6), the main function of Dps proteins is to counteract oxidative stress (7). Ferritins (Ftn and Bfr) assemble into 24 subunit protein compartments up to 12 nm in diameter able to store 2,000 to 4,000 Fe atoms in their interior (8, 9). However, some organisms do not encode ferritin genes in their genomes and their iron storage systems have remained elusive.

A newly discovered class of protein organelles called encapsulin nanocompartments are implicated in microbial iron and redox metabolism and have so far only been shown to be involved in oxidative stress response (10-13). Previously reported encapsulins share an HK97 phage-like fold and self-assemble from a single capsid protein into icosahedral compartments between 24 and 32 nm in diameter with triangulation numbers of  $T = 1$  (60 subunits) and  $T = 3$  (180 subunits), respectively (11, 12, 14). Their key feature is the ability to specifically encapsulate cargo proteins (Fig. 1A). Encapsulation is mediated by short C-terminal sequences referred to as targeting peptides (TPs) (12, 15). Genes encoding encapsulin shell proteins and dedicated cargo proteins are organized in co-regulated operons (10, 12). We have identified a novel type of encapsulin operon involved in iron metabolism in a range of Firmicutes we term the Iron-Mineralizing Encapsulin-Associated Firmicute (IMEF) system (10).

Here, we report the structural and mechanistic characterization of the IMEF system found in *Quasibacillus thermotolerans* (*Qs*), an organism that does not encode any ferritins in its genome. We show that this encapsulin-based system self-assembles into a thermostable 42 nm 9.6 MDa protein compartment with a novel  $T = 4$  topology able to mineralize and store an exceptionally large quantity of iron.

IMEF systems are found in Firmicute genomes and their operon organization indicates a function in dynamic iron storage. We carried out BLASTp searches using IMEF cargo proteins as queries and identified 71 operons in a range of Firmicutes including *Qs* (Fig. S1). The core operon consists of the encapsulin capsid protein and the IMEF cargo protein with 70% of operons also encoding a 2Fe-2S ferredoxin homologous to bacterioferritin-associated ferredoxins (Bfd). Bfd proteins are involved in the mobilization of iron under iron-limited conditions (16). In addition, 31% of operons are associated with proteins similar to ferroxidases involved in catalyzing the insertion of ferrous iron into protoporphyrins (17). The majority of IMEF-encoding genomes do not contain any ferritin or bacterioferritin genes (Table S1). Most IMEF genomes do however contain Dps-encoding genes. Overall, the operon organization of IMEF systems and the lack of other known primary iron storage proteins indicate a function for IMEF systems in dynamic iron storage similar to that of Ftn and Bfr.

Using a recombinant system, we produced homogeneous IMEF cargo-loaded encapsulins (Fig. S2). Through single-particle cryo-EM analysis, we determined the structure of the *Qs* IMEF encapsulin shell at an overall resolution of 3.85 Å (Fig. S3A). The IMEF encapsulin self-assembles into a 240 subunit icosahedral compartment with a diameter of 42 nm (Fig. 1B and

Fig. S3A,B). The IMEF compartment is substantially larger than previously reported encapsulins and possesses a triangulation number of  $T = 4$  instead of  $T = 1$  or  $T = 3$  (Fig. S4). The shell is composed of 12 pentameric and 30 hexameric capsomers occupying icosahedral vertices and faces, respectively. The 5-fold symmetry axes are located at the pentameric vertices while 3-fold symmetry axes are present at all interfaces where 3 hexameric capsomers meet. The center of each hexameric capsomer corresponds to an icosahedral edge possessing 2-fold symmetry. The icosahedral asymmetric unit consists of one pentameric and 3 hexameric monomers (Fig. 1B and Fig. S3C). Symmetrically arranged lower resolution density (ca. 10 Å) representing the IMEF cargo is visible in the compartment interior (Fig. 1B and Fig. S3D). 42 distinct densities, one for each capsomer of the  $T = 4$  structure, can be observed. No connection of cargo and shell density is visible, likely due to the flexibility of the 37 amino acid linker preceding the IMEF targeting peptide that directs and anchors the IMEF cargo to the shell interior. The distance between the shell and cargo densities is 4.5 nm which can be bridged by the 37 amino acid linker.

The 4 capsid proteins of the asymmetric unit adopt different conformations with significant differences found in the E-loop and A-domain (Fig. 1C). E-loops are located at capsomer interfaces and their relative orientation plays a key role in determining the overall topology and triangulation number of encapsulin compartments as evidenced by comparison of the IMEF  $T = 4$  monomer with  $T = 1$ ,  $T = 3$  and  $T = 7$  capsid proteins. A-domain loops form compartment pores and are likely adapted to the particular function of a given encapsulin explaining the observed conformational diversity. In addition, local resolution maps indicate that E-loops and A-domain loops represent the most flexible parts of the shell (Fig. S5).

The IMEF compartment possesses a non-covalent chainmail topology and is highly thermostable. E-loops and P-domains of neighboring capsid monomers arrange head to tail to form interlocking concatenated rings resulting in a non-covalent chainmail topology (Fig. 2A) (18). In contrast to HK97 where an isopeptide bond covalently links E-loops and P-domains (19), the IMEF encapsulin uses non-covalent interactions. At each 3-fold pore, E-loops connect with two neighboring P-domains including the G-loop conserved in  $T = 4$  encapsulins and their interfaces contain complementary electrostatic as well as aromatic and potential anion- $\pi$  interactions (Fig. 2B and Fig. S6) (20). The IMEF cargo and shell protein are both highly thermostable with melting temperatures of 80.6 and 86.6°C, respectively (Fig. 2C). A stabilizing effect is observed for the cargo-loaded compartment (88.9°C). Compartments isolated from high iron conditions show even greater thermal stability (91.8°C) likely due to the internal cavity being stabilized by mineralized material.

The IMEF encapsulin shell contains negatively charged pores at the 3- and 5-fold symmetry axes. The surface view of the intact shell (Fig. S7) shows a tight packing with pores at the 3- and 5-fold symmetry axes and at the interface between two hexameric and one pentameric capsomer (pseudo 3-fold) representing the only conduits to the interior. All pores are negatively charged on both the exterior and interior surface due to the presence of conserved aspartate, glutamate and asparagine residues (Fig. 2D, 2E, Fig. S8 and Fig. S9). The 3-fold pore forms the largest channel to the IMEF compartment interior and is 7.2 Å wide at its narrowest point, substantially larger than previously reported encapsulin pores (Fig. S10). Extra cryo-EM density is observed at the center of both the 3-fold and 5-fold pores potentially resulting from bound positively charged ions. The 2-fold symmetry axes at the center of hexameric capsomers also represent potential channels, as observed in  $T = 3$  systems (21), but the conformation of two asparagine side chains prevents the formation of a 2-fold opening in the  $T = 4$  shell leading to a closed pore (Fig. 2F).

This observation combined with the flexibility observed for loops around the 2- and 5-fold symmetry axes in local resolution maps (Fig. S5) indicates gated pores in encapsulins that may regulate ion flux to the compartment interior, similar to some ferritins (22).

Sequence and x-ray structure analysis show that the IMEF cargo represents a distinct class of ferritin-like protein (Flp) with an unusual ferroxidase center. Phylogenetic analysis revealed that the IMEF cargo protein is a member of the Flp superfamily and is most closely related to Dps proteins (Fig. 3A and Table S3) but no known ferroxidase motifs could be detected at the sequence level (5). All IMEF proteins share a conserved C-terminal TP (Fig. 3B). We determined the x-ray crystal structure of the IMEF cargo to a final resolution of 1.72 Å (Fig. 3C and Table S4). The cargo adopts a four-helix bundle fold characteristic of other members of the Flp superfamily and forms a dimer with two Fe atoms bound at the subunit interface creating a ferroxidase site based on an alternative ferroxidase sequence motif (Fig. 3D, Fig. S11 and Fig. S12). This leads to a combined molecular weight of the fully cargo-loaded IMEF compartment of 9.6 MDa ( $42 \times$  cargo dimer [22.6 kDa] +  $240 \times$  capsid protein, [32.2 kDa]). Removal of the 13 C-terminal residues results in empty encapsulin shells confirming that the IMEF TP is necessary for cargo encapsulation (Fig. 3E).

Additional cryo-EM density around the 2- and 5-fold symmetry axes reveals TP-binding sites and illuminates cargo-shell co-assembly. Through analysis of the  $T = 4$  cryo-EM map, additional densities were identified that could not be explained by the encapsulin capsid protein (Fig. 3F). These densities represent bound TPs anchoring IMEF cargo to the interior surface of the compartment. Even though only 42 cargo densities are observed, TP densities can be found at all 240 capsid monomers indicating averaging during cryo-EM reconstruction. Strong TP density is observed for all 180 monomers that are part of 2-fold symmetrical hexameric capsomers (Fig. 3F) while substantially weaker density is found for TPs bound to the 60 pentameric monomers (Fig. S13A) thus revealing higher occupancy and preferential TP binding around 2-fold symmetry axes which can be explained by different binding site conformations (Fig. S13) and higher local shell mobility (Fig. S5). The main TP binding sites surrounding the 2-fold symmetry axes are formed by conserved residues of the P-domain and N-terminal helix (Fig. S9) similar to the *T. maritima*  $T = 1$  encapsulin system (12). Thus, the presence of the N-terminal helix and the resulting binding site generally underpin encapsulins' ability to interact with TPs and encapsulate cargo proteins. The TP residues TVGSLIQ were tentatively built and refined into the additional density present at hexameric capsomers producing a model with good geometry (Fig. S13C). The TP binds to a surface groove based on shape complementarity and two key ionic interactions with highly conserved positively charged residues locking the TP in place (Fig. S13C).

Heterologous expression of the IMEF core operon in *E. coli* leads to *in vivo* formation of large Fe- and P-rich electron-dense particles. Thin section negative stain transmission electron microscopy (TEM) of *E. coli* cells grown in Fe-rich (4 mM) medium and expressing the *Qs* IMEF core operon results in the formation of clusters of large intracellular electron-dense particles (Fig. 4A and Fig. S14). Scanning TEM-energy-dispersive x-ray spectroscopy (EDS) revealed that these particles primarily contain uniformly distributed Fe, P and O with an estimated Fe:P ratio near 1 (Fig. 4B). Selected area electron diffraction (SAED) further indicates that this mineralized material is amorphous (Fig. S15), similar to bacterioferritin systems (23).

The IMEF encapsulin mineralizes up to 30 nm Fe-rich cores in its interior with up to 23,000 Fe atoms stored per particle. IMEF encapsulins purified from *E. coli* grown under high Fe conditions contain electron dense cores visible in unstained samples with an average diameter of

23 nm (Fig. 4C,D and Fig. S16). The largest observed particles are up to 30 nm in diameter. The theoretical size limit imposed by the  $T = 4$  encapsulin protein shell is 36 nm and particles close to this size are observed in thin-sections of *Geobacillus* natively encoding the IMEF system (Fig. S17). EDS analysis of particles isolated from *E. coli* and comparison with standards indicate a very similar elemental composition and elemental distribution as observed for thin section samples with a Fe:P ratio of 1:1.1 (Fig. S18). To determine the number of iron atoms stored per particle, we carried out electron energy loss spectroscopy (EELS) on purified Fe-loaded compartments (Fig. 4E,F and Fig. S19). The highest observed number of stored Fe per particle was 23,293 (23.6 nm) (Table S5). Extrapolating to the maximum theoretical particle diameter of 36 nm and the highest density observed (3.40 Fe atoms/nm<sup>3</sup>) leads to a maximum number of Fe atoms that can be stored by the IMEF system of around 83,000 (Table S5).

To learn more about the mechanism of iron mineralization, we assayed peroxidase and ferroxidase activity using H<sub>2</sub>O<sub>2</sub> and O<sub>2</sub> as the oxidants, respectively. No peroxidase activity was detected. For the IMEF cargo alone, a sigmoidal ferroxidase iron oxidation curve was observed indicative of autocatalytic Fe oxidation taking place at newly formed mineral surfaces (24, 25). However, assaying the cargo-loaded encapsulin results in a typical hyperbolic enzyme catalysis curve. These observations imply that the encapsulin shell controls the flux of iron to the interior of the compartment thus preventing uncontrolled autocatalytic mineralization (Fig. S20).

Our structural model and functional analysis of the IMEF encapsulin system reveal an alternative way to store large amounts of Fe independent of ferritins. The IMEF system can in principle store more than 20 times more Fe than Ftn or Bfr systems. In contrast to ferritin systems, IMEF encapsulins are two-component systems with the catalytic activity separated from the protein shell. The IMEF cargo protein is flexibly tethered and primarily localizes 4.5 nm away from the capsid interior. This suggests that once iron has been channeled to the encapsulin interior via pores, it diffuses to the ferroxidase active site of the IMEF cargo, making it necessary to strictly control interior iron concentration to prevent runaway mineralization. In sum, we have elucidated the structure and mechanism of the largest iron storage complex to date indicating that alternative systems exist across nature to address the critical problem of safe and dynamic iron storage.

## References:

1. M. Sanchez, L. Sabio, N. Galvez, M. Capdevila, J. M. Dominguez-Vera, Iron chemistry at the service of life. *IUBMB Life* **69**, 382-388 (2017).
2. S. J. Dixon, B. R. Stockwell, The role of iron and reactive oxygen species in cell death. *Nat Chem Biol* **10**, 9-17 (2014).
3. R. Crichton, *Inorganic Biochemistry of Iron Metabolism: From Molecular Mechanisms to Clinical Consequences*. (ed. Second Edition, 2002).
4. P. Arosio, L. Elia, M. Poli, Ferritin, cellular iron storage and regulation. *IUBMB Life* **69**, 414-422 (2017).
5. S. C. Andrews, The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta* **1800**, 691-705 (2010).
6. K. Honarmand Ebrahimi, P. L. Hagedoorn, W. R. Hagen, Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin. *Chem Rev* **115**, 295-326 (2015).
7. E. Chiancone, P. Ceci, A. Ilari, F. Ribacchi, S. Stefanini, Iron and proteins for iron storage and detoxification. *Biometals* **17**, 197-202 (2004).
8. S. C. Andrews, Iron storage in bacteria. *Adv Microb Physiol* **40**, 281-351 (1998).
9. P. M. Harrison, P. Arosio, The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta* **1275**, 161-203 (1996).
10. T. W. Giessen, P. A. Silver, Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**, 17029 (2017).
11. C. A. McHugh *et al.*, A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J* **33**, 1896-1911 (2014).
12. M. Sutter *et al.*, Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol* **15**, 939-947 (2008).
13. D. He *et al.*, Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife* **5**, (2016).
14. F. Akita *et al.*, The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J Mol Biol* **368**, 1469-1483 (2007).
15. A. Tamura *et al.*, Packaging guest proteins into the encapsulin nanocompartment from *Rhodococcus erythropolis* N771. *Biotechnol Bioeng* **112**, 13-20 (2015).
16. H. Yao *et al.*, The structure of the BfrB-Bfd complex reveals protein-protein interactions enabling iron release from bacterioferritin. *J Am Chem Soc* **134**, 13470-13481 (2012).
17. H. A. Dailey *et al.*, Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters. *Cell Mol Life Sci* **57**, 1909-1926 (2000).
18. X. Zhang *et al.*, A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 Å resolution. *Elife* **2**, e01299 (2013).
19. R. L. Duda, Protein chainmail: catenated protein in viral capsids. *Cell* **94**, 55-60 (1998).
20. V. Philip *et al.*, A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-π pairs. *Biochemistry* **50**, 2939-2950 (2011).
21. R. J. Nichols, C. Cassidy-Amstutz, T. Chaijarasphong, D. F. Savage, Encapsulins: molecular biology of the shell. *Crit Rev Biochem Mol Biol* **52**, 583-594 (2017).
22. E. C. Theil, X. S. Liu, T. Tosha, Gated Pores in the Ferritin Protein Nanocage. *Inorganica Chim Acta* **361**, 868-874 (2008).

23. S. C. Andrews *et al.*, Overproduction, purification and characterization of the bacterioferritin of *Escherichia coli* and a C-terminally extended variant. *Eur J Biochem* **213**, 329-338 (1993).
24. F. Bou-Abdallah, P. Santambrogio, S. Levi, P. Arosio, N. D. Chasteen, Unique iron binding and oxidation properties of human mitochondrial ferritin: a comparative analysis with Human H-chain ferritin. *J Mol Biol* **347**, 543-554 (2005).
25. S. Sun, N. D. Chasteen, Ferroxidase kinetics of horse spleen apoferritin. *J Biol Chem* **267**, 25160-25166 (1992).

### References only cited in the supplementary materials:

26. J. A. Gerlt *et al.*, Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854**, 1019-1037 (2015).
27. J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174-181 (2015).
28. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738 (2010).
29. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
30. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
31. H. Ru *et al.*, Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell* **163**, 1138-1152 (2015).
32. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).
33. J. Frank *et al.*, SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199 (1996).
34. L. Swint-Kruse, C. S. Brown, Resmap: automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics* **21**, 3327-3328 (2005).
35. Y. Zhang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
36. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
37. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
38. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
39. C. Grimm, A. Chari, K. Reuter, U. Fischer, A crystallization screen based on alternative polymeric precipitants. *Acta Crystallogr D Biol Crystallogr* **66**, 685-697 (2010).
40. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
41. M. Sammito *et al.*, ARCIMBOLDO\_LITE: single-workstation implementation and use. *Acta Crystallogr D Biol Crystallogr* **71**, 1921-1930 (2015).
42. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

43. A. Thorn, G. M. Sheldrick, Extending molecular-replacement solutions with SHELXE. *Acta Crystallogr D Biol Crystallogr* **69**, 2251-2256 (2013).
44. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
45. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
46. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
47. J. Pesek, R. Buchler, R. Albrecht, W. Boland, K. Zeth, Structure and mechanism of iron translocation by a Dps protein from *Microbacterium arborescens*. *J Biol Chem* **286**, 34872-34882 (2011).
48. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
49. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

**Acknowledgments:** We thank M. Ericsson, P. Coughlin and L. Trakimas for technical support (HMS Electron Microscopy Facility). **Funding:** This work was supported by a Leopoldina Research Fellowship (LPDS 2014-05) from the German National Academy of Sciences Leopoldina (T.W.G), the Wyss Institute for Biologically Inspired Engineering (T.W.G and P.A.S.) and the Gordon and Betty Moore Foundation (T.W.G. and P.A.S., Grant Number 5506). **Authors contributions:** T.W.G. and P.A.S. conceived and designed the study. T.W.G. carried out all cloning and performed biochemical and negative staining TEM experiments. B.J.O. and M. L. performed cryo-EM image processing, data analysis and model building. M.G.C. collected cryo-EM data sets. A.A.V. purified and crystalized the IMEF cargo protein and cultivated *Geobacillus*. J.G. and D.C.B. performed and analyzed EDS and EELS experiments. G.B. helped in crystallization experiments and solved the IMEF x-ray structure. T.W.G. and P.A.S. wrote the manuscript. B.J.O. and M.L. helped in editing the manuscript. **Competing interests:** None declared. **Data and materials availability:** A cryo-EM density map of the cargo-loaded IMEF encapsulin has been deposited in the Electron Microscopy Data Bank under the accession number 9383. The corresponding atomic coordinates for the atomic model have been deposited in the Protein Data Bank (accession number: 6NJ8). Atomic coordinates for the IMEF cargo protein have been deposited in the Protein Data Bank under accession number 6N63. Correspondence and requests for materials should be addressed to the corresponding authors.

### Supplementary Materials:

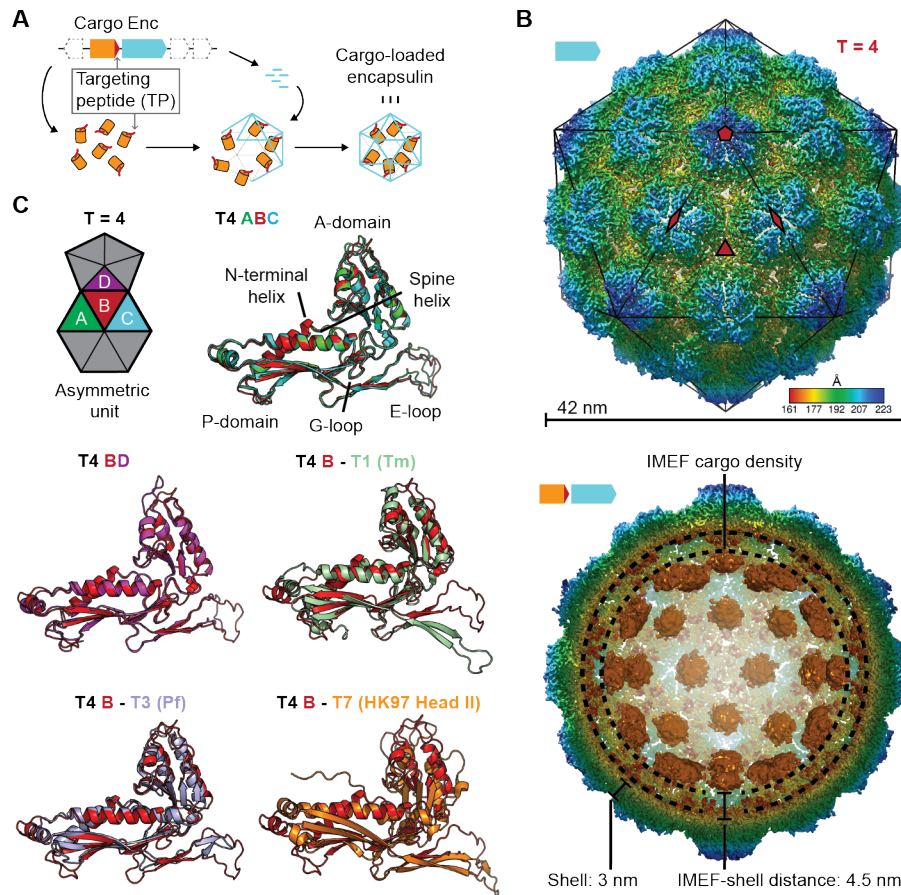
Materials and Methods

Fig. S1 – S20

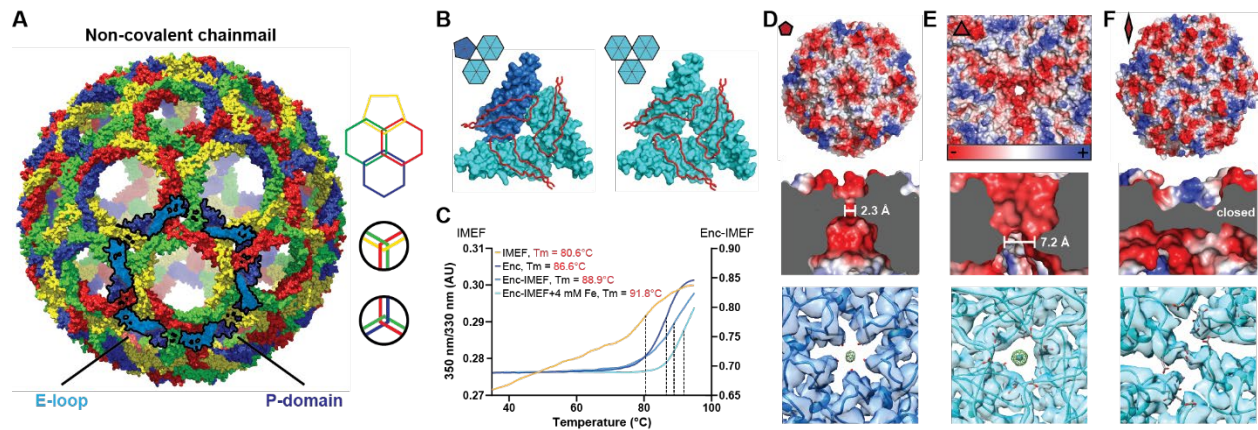
Table S1 – S5

References (26 – 49)

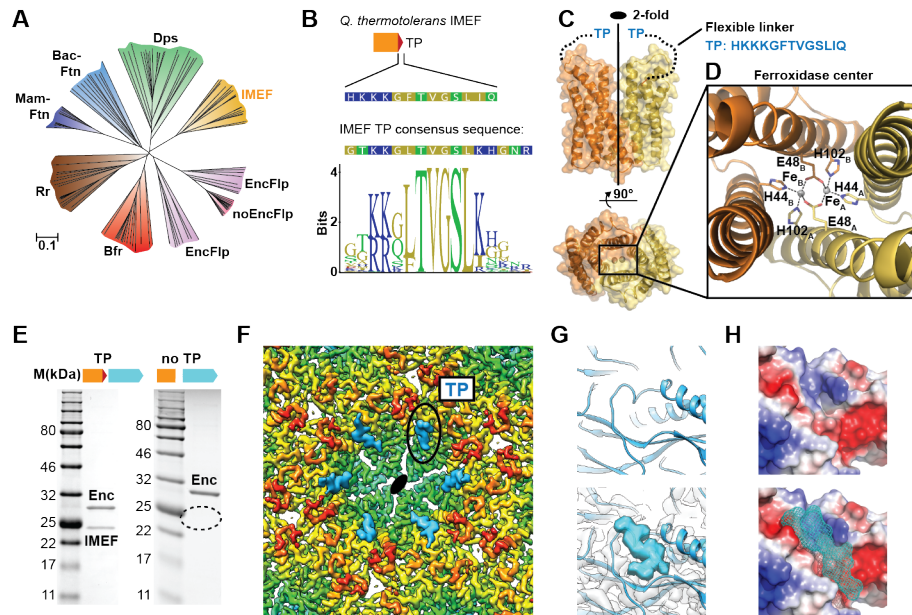




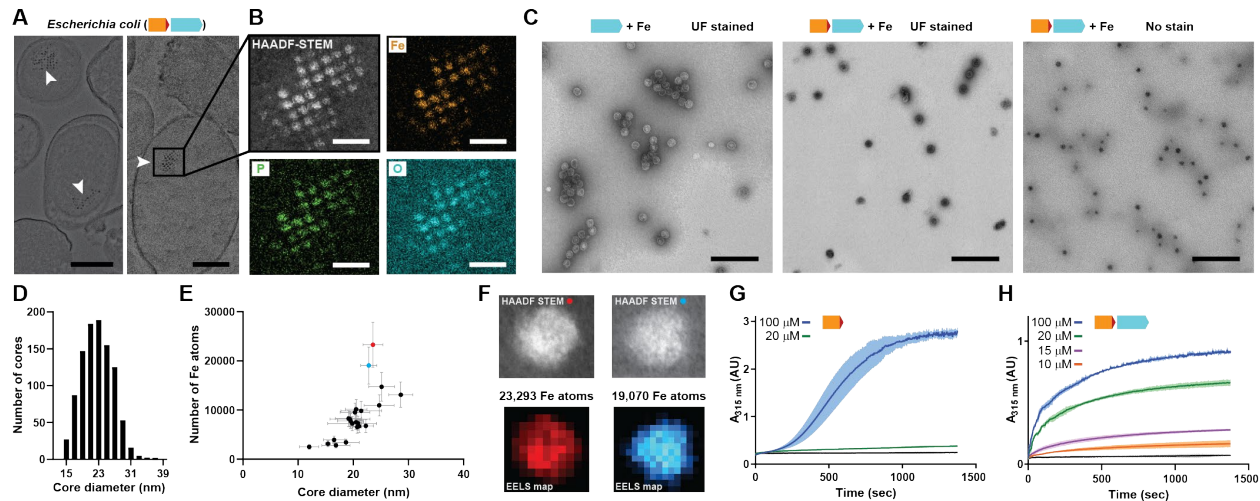
**Fig. 1. Overall architecture of the cargo-loaded T = 4 encapsulin.** (A) Schematic diagram of a core encapsulin operon and targeting peptide (TP)-dependent cargo encapsulation. (B) Surface view of the cryo-EM map of the *Qs* T = 4 encapsulin shell (top) and inside view of cargo-loaded encapsulin (bottom). 5-, 3- and 2-fold symmetry axes are indicated by red symbols. The overall icosahedral symmetry is highlighted by black lines representing icosahedral facets. Cargo-densities are shown in orange while the shell is radially colored. To depict the complete cargo-loaded compartment, a 10 Å filtered map highlighting the cargo was combined with the 3.85 Å map of the shell. (C) Asymmetric unit of the T4 encapsulin shell and structural alignment of the four unique T4 shell monomers with one another and with the *T. maritima* (Tm) T = 1 monomer (3DKT), the *P. furiosus* (Pf) T = 3 monomer (2E0Z) and the HK97 bacteriophage Head II T = 7 monomer (2FT1).



**Fig. 2. Non-covalent chainmail topology, thermal stability and pores of the  $T = 4$  encapsulin shell.** (A) Chainmail network mediated by E-loop and P-domain interactions. Only E-loops and P-domains are shown. E-loops and P-domains of the outlined ring belonging to the same monomer are located next to one another and are shown in light and dark blue, respectively. (B) Extended E-loop interactions interlock neighboring capsid monomers at the two unique three-fold interfaces. Each E-loop interacts with two P-domains. (C) Representative thermal unfolding curves for  $Q_s$   $T = 4$  encapsulin components determined via differential scanning fluorimetry.  $T_m$ : midpoint of the thermal unfolding curve. (D, E and F) Electrostatic surface representation of the 5-fold (D) and 3-fold (E)  $T = 4$  shell pores and the 2-fold symmetry axis (F). Outside views showing negatively charged pores (top) with no pore opening observed at the two-fold symmetry axis, cutaway side view highlighting the narrowest point of the pores (middle) and cryo-EM maps with fitted monomer models in ribbon representation (bottom). Additional cryo-EM density is observed at the center of both pores in interaction distance with the side chains of pore residues (5-fold: Asn200, 3-fold: Asp9, Asp71, Glu251 and Glu252, shown in stick representation).



**Fig. 3. Structure and analysis of the IMEF cargo protein and TP-mediated cargo-shell co-assembly.** (A) Neighbor-joining phylogeny (cladogram) of protein classes involved in iron metabolism that are part of the Fln superfamily. Scale bar: amino acid substitutions per site. EncFln: Flns found within encapsulin operons containing TPs, noEncFln: Flns found outside encapsulin operons not containing TPs, Rr: rubrerythrins, Mam-Ftn: mammalian ferritins, Bac-Ftn: bacterial ferritins. (B) TP sequence of the *Qs* IMEF cargo protein and TP sequence logo highlighting strong sequence conservation. (C) X-ray crystal structure of the *Qs* IMEF cargo. (D) Di-iron ferroxidase active site of the IMEF cargo. The iron-coordinating residues are shown in stick representation. (E) SDS-PAGE gels of purified encapsulins showing that co-purification is dependent on the presence of the TP. (F) Cryo-EM map interior view of the 2-fold symmetry axis with TP density shown in cyan. (G) Close-up of additional cryo-EM density observed around the 2-fold symmetry axis. (H) Electrostatic surface representation of the TP binding site without (top) and with (bottom) TP. The 7 C-terminal IMEF residues are shown as a surface mesh.



**Fig. 4. Mineralization of large iron-rich particles by the T = 4 encapsulin.** (A) Thin section micrographs of *E. coli* heterologously expressing the *Qs* IMEF core operon. Electron-dense particles often cluster together in regular arrays. Scale bars: 500 nm (left), 400 nm (right). (B) Close-up high angle angular dark field (HAADF) scanning TEM and EDS maps of a cluster of particles showing Fe, P and O as the main particle constituents. Scale bars: 100 nm. (C) Micrographs of uranyl formate (UF)-stained encapsulins produced in and isolated from *E. coli* grown in high iron media expressing the capsid protein alone (left) or the core operon (middle and right). Without UF stain, electron-dense particles are clearly visible (right). Scale bars: 250 nm. (D) Size distribution of electron-dense particles in unstained micrographs. (E) Electron energy loss spectroscopy (EELS) of 22 select cores. (F) HAADF-STEM micrographs and EELS maps of the two highlighted cores from (E). (G) *In vitro* ferroxidase assay of purified IMEF cargo at different  $Fe^{2+}$  concentrations. Mean values resulting from technical triplicates and error bands using standard deviation are shown. (H) Ferroxidase assay of cargo-loaded T = 4 encapsulin at different  $Fe^{2+}$  concentrations.