Supplementary Material.

Methods

Construction of mutant strains.

The *encaA*::*pCR2.1* (EH410) disruption strain was constructed by recombination of the suicide plasmid pCR2.1 containing a 396-bp homology region into the *MXAN_3556* (*encA*) gene. The construct pCR2.1-3556d was created by amplification of the homology fragment using the upstream primer CM380 (5' - CCTCTGGAGCCACATTTCAT - 3') and the downstream primer CM381 (5' - CGTCGCCGTAGAAGATGAGT - 3'). The resulting fragment was cloned using the pCR2.1-TOPO kit according to the manufacturer's instructions (Invitrogen). After sequence confirmation, 1 μg of the plasmid pCR2.1-3556d was introduced into electrocompetent *M. xanthus* DK1622 cells, and transformants plated in top agar (0.5%) on CTT + 1.5% agar plates supplemented with 40 μg/mL kanamycin. The resulting insertion disruption strain was confirmed by PCR and kanamycin resistance phenotype.

The in-frame deletion strain Δ*encA* (EH411) was constructed using an allelic exchange by homologous recombination strategy, with plasmid pBJ113 as described previously (Ueki et al, 1996). Briefly, an insert lacking the *MXAN_3556* coding region was created by splicing by overlap extension (Heckman & Pease, 2007), with A primer CM444 (5' – GGCCCGAATTCGCATGTTGGACAGTGGTCAG – 3'), B primer CM445 (5' – ACCGGAACTAGCGACGTTCGGCATGTCCAAGGAAGTCAG – 3'), C primer CM446 (5' – CTGACTTCCTTGGACATGCCGAACGTCGCTAGTTCCGGT – 3'), and D primer CM447 (5' – GGCCCAAGCTTAGGGCTTGGTGATGTAGTCG – 3')

adapted from a previous study (Kim et al, 2009). The insert was cloned into the pBJ113 vector using EcoRI and HindIII, and confirmed by nucleic acid sequencing. The resulting plasmid pBJ113-encdel was electroporated into wild-type DK1622 *M. xanthus* and integrants were selected on CTT agar plates containing 40 μg/mL kanamycin. Integrants were then cultured in CTT containing 2.5% galactose to select for loss of the plasmid (*galK* negative selection). Final clones were analyzed by PCR screening with the control primers CM448 (5' – ACAGTGACTTGGACGACGTG - 3') and CM449 (5' – TGATCTTGATGAACGGCTTG - 3'), followed by nucleic acid sequencing of the region to confirm deletion of *MXAN_3556*.

Quantitation of internal proteins.

Gels were stained with Simply Blue Safe Stain (Invitrogen), scanned with a GelLogic 100 densitometer (Kodak), and quantitated using ImageJ (Schneider et al, 2012). Serial dilutions of the sample were run in neighboring lanes so that bands of approximately equal density in the linear range could be directly compared between EncA and the minor proteins.

Fig. S1. ClustalW multiple sequence alignment of selected encapsulin proteins and HK97 gp5* (Thompson et al, 1994). Bl, *Brevibacterium linens* linocin M18 (gi1279217); Mt, *Mycobacterium tuberculosis* CFP29 (gi15607938); Tm, *Thermotoga maritima* encapsulin TM0785 (gi15643548); Mx, *Myxococcus xanthus* encapsulin MXAN_3556 (gi108462612); gp5*, HK97 gp5* capsid protein from Prohead II, lacking N-terminal delta region of full-length gp5 protein (gi1262530).

Fig. S2. EncB (MXAN_3557) bioinformatic analysis.

EncB ferritin-like domain is marked in green (approximately amino acids 9-66) and is only long enough to contain two alpha-helices of the length of those found in ferritin-like domains (which have four alpha-helices). Di-iron binding motifs (ExxH) are marked in red (amino acids 52-55 and 58-61) and the EncA anchor region in grey. The high number of prolines (underlined) of the sequence after the ferritin-like domain (amino acids 67-141) and the high number of glycine and serines after the EncA anchor region (amino acids 150-158), suggest that these regions of the protein are disordered.

Note that Pfam search (Finn et al, 2014) suggests that EncB possesses a Rubrerythrin domain that extends from aminoacids 20 to 80. The Pfam consensus Rubrerythrin domain

(Rub Consensus) and its alignment with EncB (MATCH shows the amino acid identities and similarities) are shown. Sequences aligned correspond to amino acids 20-80 of EncB and 8-70 of the consensus sequence; these regions are 26% identical and 63% similar. Secondary structure prediction is based on the results of four programs: JPred (Cuff & Barton, 2000), Porter (Pollastri & McLysaght, 2005), PsiPred (McGuffin et al, 2000) and Predict Protein (Rost et al, 2004). If the Rubrerythrin consensus sequence is aligned to its most similar sequence that has a solved structure (a bacterioferritin from *Blastochloris viridis*, pdb 4AM5), the alpha-helices would comprise aminoacids 1-27; 29-58; 80-110 & 114-137.

Figure S3. EncC (MXAN_4464) bioinformatic analysis.

EncC ferritin-like domain is marked in green (approximately amino acids 17-74), and is only long enough to contain two alpha-helices of the length of those found in ferritin-like domains (which have four alpha-helices). The di-iron binding motif (ExxH) is marked in red (amino acids 61-64) and the EncA anchor region in grey.

Note that Pfam search (Finn et al, 2014) suggests that EncC possesses a Rubrerythrin domain that extends from aminoacids 26 to 90. The Pfam consensus Rubrerythrin domain (Rub Consensus) and its alignment with EncC and its alignment with EncC (MATCH shows the amino acid identities and similarities) are shown. Sequences aligned

correspond to amino acids 26-90 of EncC and 5-73 of the consensus sequence; these regions are 22% identical and 69% similar.

Since EncC is labeled as a ferritin in the NCBI Database of Reference Sequences ((Pruitt et al, 2009) accession code WP_020478352.1), it can be hypothesized that EncC contains the four alpha-helices of the ferritin-like domain (of which the secondary prediction programs can only detect helices 1 to 3). However, this would not allow the helices to be arranged in the expected up-down-down-up four-helix bundle characteristic of the ferritin-like domain.

Secondary structure prediction is based on the results of four programs: JPred (Cuff & Barton, 2000), Porter (Pollastri & McLysaght, 2005), PsiPred (McGuffin et al, 2000) and Predict Protein (Rost et al, 2004). If the Rubrerythrin consensus sequence is aligned to its most similar sequence that has a solved structure (a bacterioferritin from *Blastochloris viridis*, pdb 4AM5), the alpha-helices would comprise aminoacids 1-27; 29-58; 80-110 & 114-137.

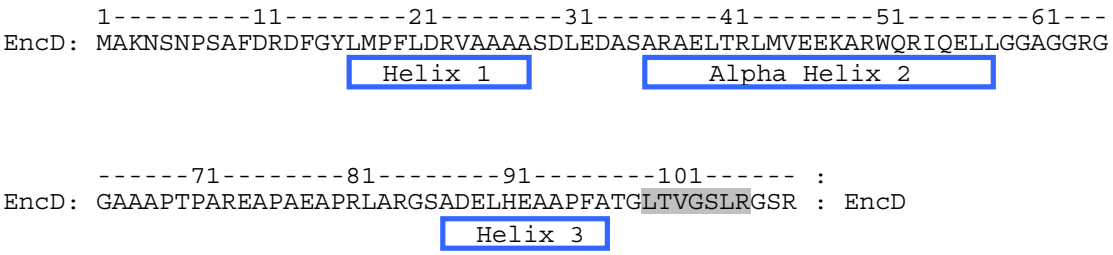

Figure S4. EncD (MXAN_2410) bioinformatic analysis.

Searches using protein blast (Johnson et al, 2008), Pfam (Finn et al, 2014) and HHPRed (Soding et al, 2005) did not find a significant match for EncD.

Secondary structure prediction is based on the results of four programs: JPred (Cuff & Barton, 2000), Porter (Pollastri & McLysaght, 2005), PsiPred (McGuffin et al, 2000) and Predict Protein (Rost et al, 2004).

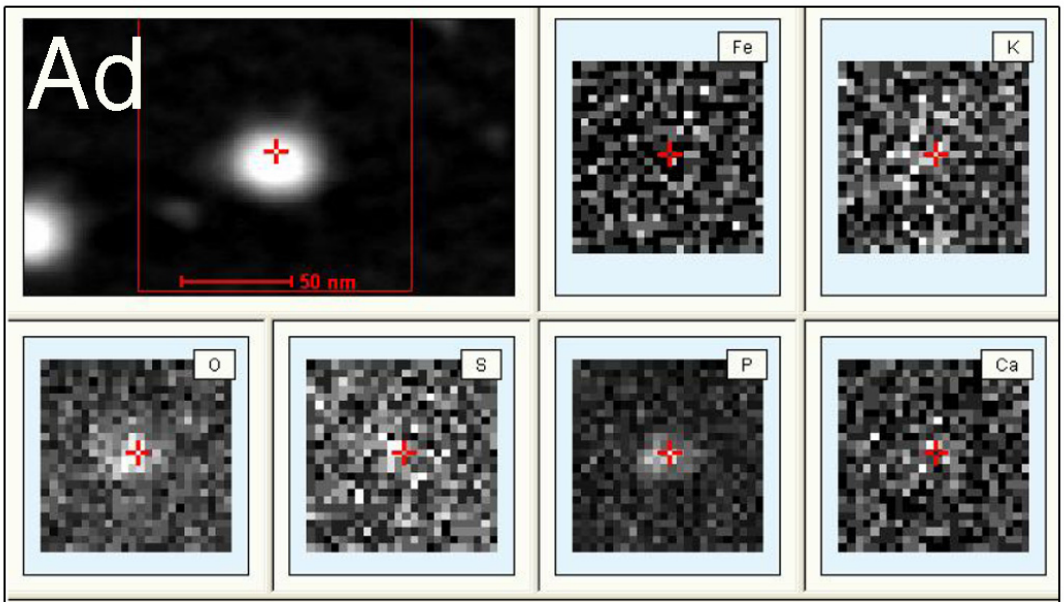

Fig. S5. **EDX analysis of adenovirus control particles.** The large panel in the upper left corner shows the STEM image of the electron density of the unstained particles with the red box demarking the sampled area. The small panels show maps for selected elements that were analyzed in the sampled area, where white pixels indicate high counts and black pixels low counts. Elements are Fe, iron; K, potassium, O, oxygen; S, sulfur; P, phosphorus; Ca, calcium.

- *Supplemental Table 1: Statistical Summary of Experiments Performed.*
-
- *Total number of purifications:*
    - *At least 27 independent isolations of the native nanocompartments were performed. Found to be consistent by negative staining EM and SDS-PAGE.*
    - *2 independent isolations of EncA shells from E coli were performed. SDS-PAGE and Western-blot confirmed that they were composed of >95% EncA.*
    -
- *The proteins forming the encapsulin nanocompartments were identified by two approaches from two independent isolations: MALDI/TOF mass spectrometry and Edman degradation analysis. During optimization of the purification protocol, the same proteins were found by MALDI/TOF, resulting in 4-5 identifications of the encapsulin nanocompartment components.*
-
- *EDX elemental analysis was performed on 6 encapsulin particles. Iron was consistently found at high levels in the core of the encapsulin particles and it was not present in the proteinaceous rim of the particles (corresponding to the EncA shell) nor in the grid background.*
-
- *ICP-MS: measurements consisted of 5 independent isolations done in triplicate.*
-
- *EM:*
    - *The same preparation of EncA shells was used (on separate occasions) to collect first the CM200 data, and then the Polara data, which gave a reconstruction to 4.6Å. The FSC curve shown in Figure S6 attests to the internal consistency of these data.*
    - *Two independent preparations of native nanocompartments were studied by cryo-EM/cryo-ET and by STEM. They were consistent except that the first preparation contained ~6% of the core-lacking particles while the second contained < 1%.*
- *To quantify yields of dense-cored nanocompartments, equal samples of the respective isolates were applied for 15 s to carbon-coated 400 mesh copper grids and the numbers of particles per area counted. At least three independent fields were analyzed in each experiment for two independent isolations for each condition.*
-
- *Peroxide stress assays were repeated three times with independent cultures. Each experiment included three replicates for each sample tested. Further, the disruption and deletion strains used in these experiments were generated completely independently.*

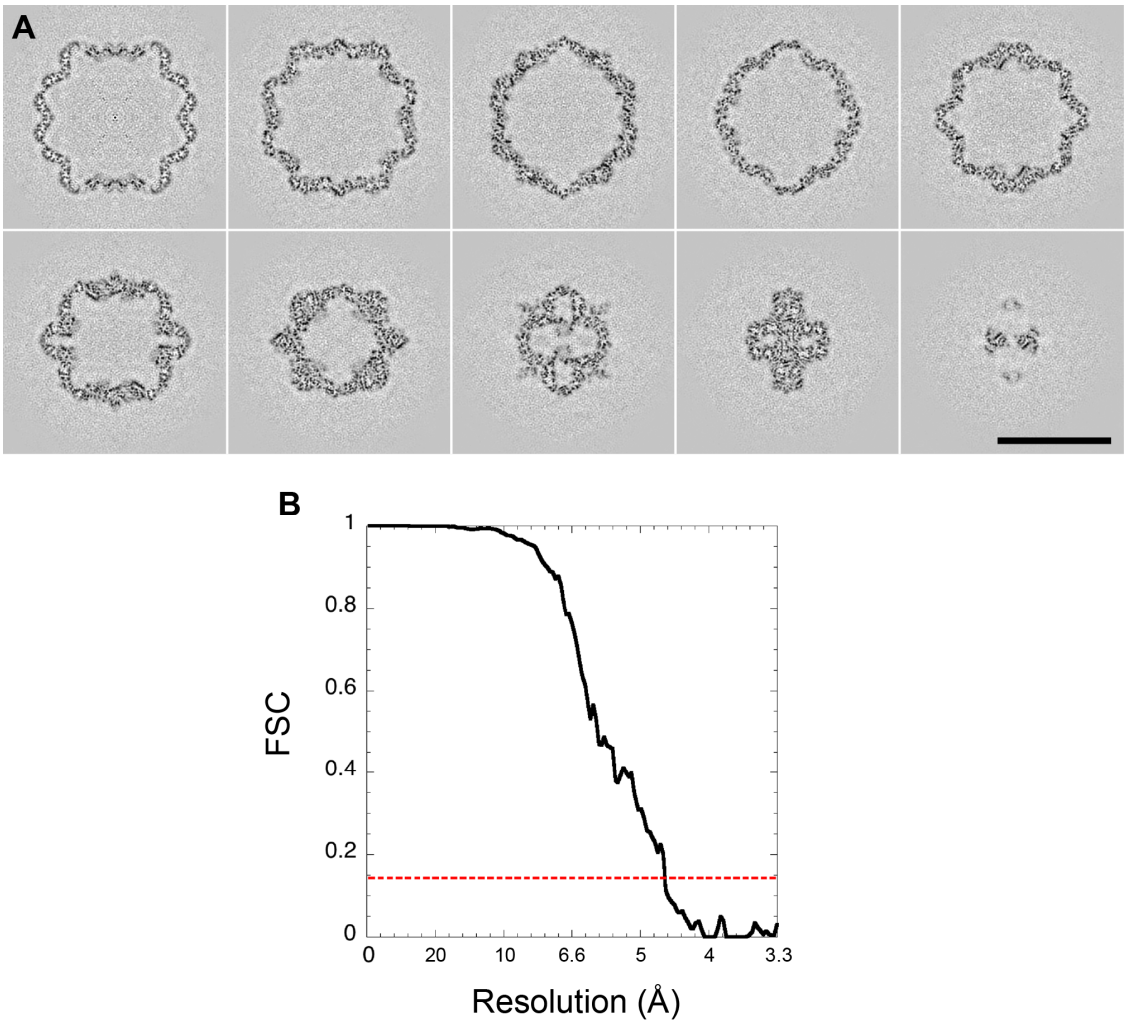

Fig. S6. Cryo-EM reconstruction of EncA T=3 shell. A, slices (~1 Å thick and ~16 Å apart) through the reconstruction. Bar, 25 nm. B, Fourier shell correlation (FSC) resolution curve of the reconstruction. Since the curve was generated by comparing two reconstructions independently calculated from half-data sets, a 0.143 cut-off was used, giving an estimated resolution of 4.6 Å.

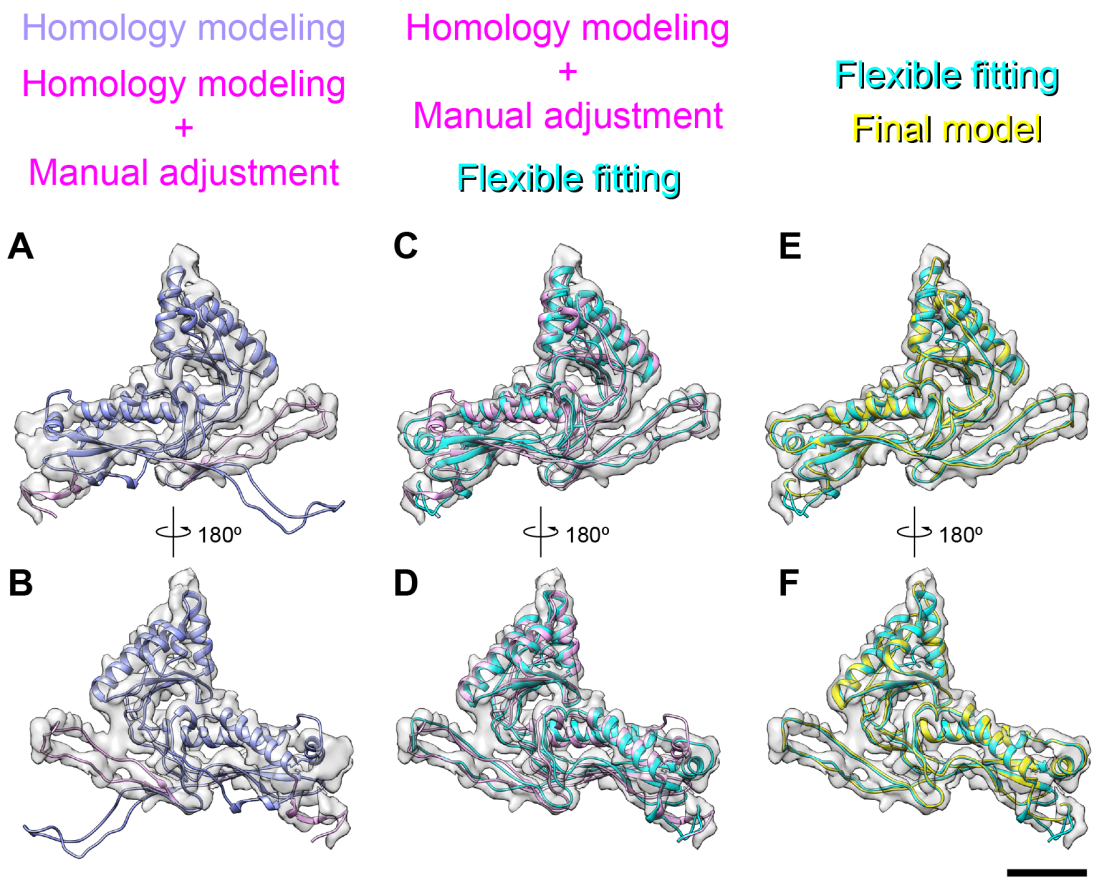

Fig. S7. Atomic modeling of an EncA subunit. Fit of an EncA molecule into the EM density map after each of the steps performed to calculate an atomic model of EncA. See Materials and Methods for details. Bar, 2 nm.

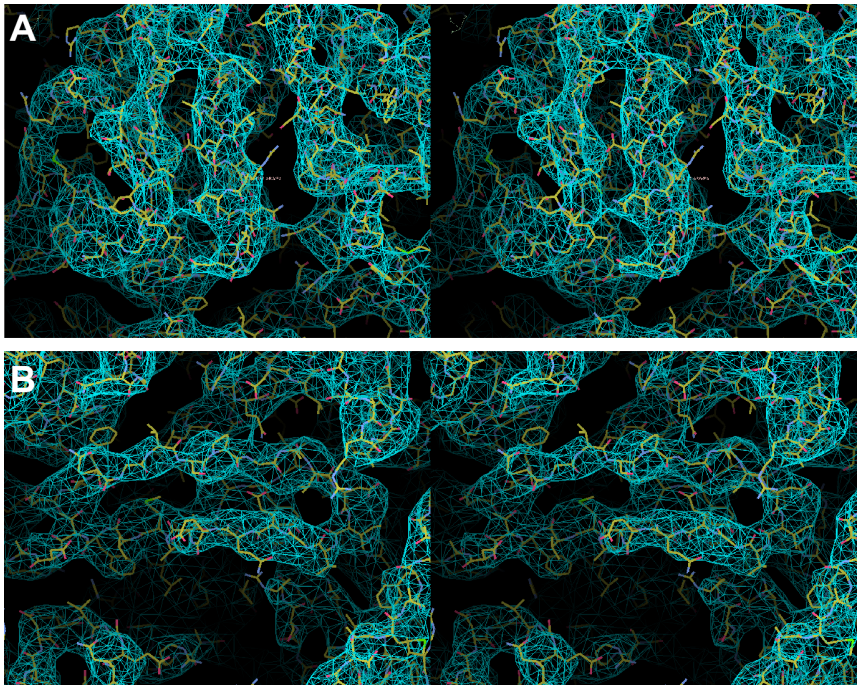

Fig. S8. Fit of EncA atomic model into the EM density. Stereo pair images of an alpha-helix (A) and the E-loop (B) with the corresponding atomic models.

Table S2: Refinement statistics	
Number of "structure factors"	9,859,810
Cell	771.400, 771.400, 771.400 Å
	90, 90, 90º
Polypeptide chains in icosahedral a.u.	3
Number of residues (icosahedral a.u.)	831
Number of atoms in full assembly	387,840
Resolution range (Å)	771.4 – 4.6
R factor	34.9
Ramachandran plot:	
Residues in allowed regions:	100%
Residues in most favored regions:	84.36%
rmsd bond lengths (Å)	0.006
rmsd bond angles (º)	1.758
Average B factor, all atoms ($Å^2$)	47.49

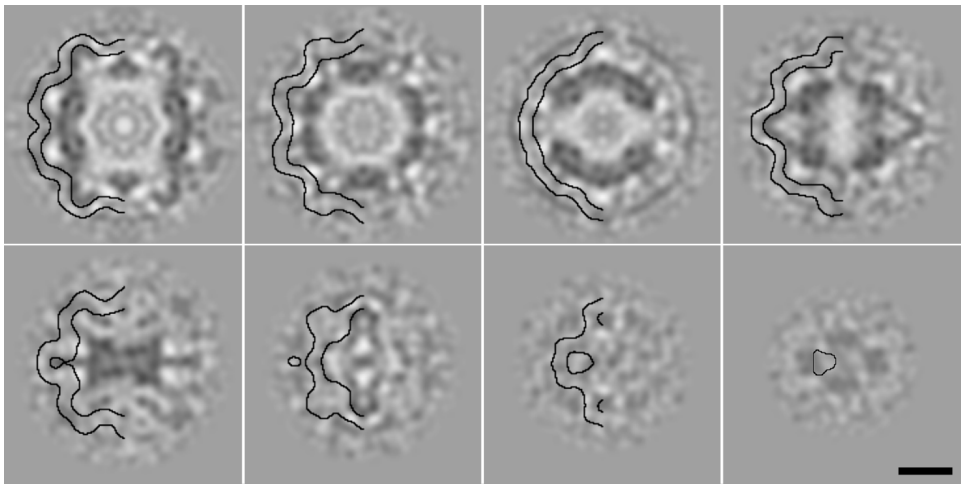

Fig. S9. 3D map of the internal proteins of native encapsulin nanocompartments. The map was generated by subtracting an EncA T=3-shell reconstruction from the reconstruction of native encapsulin nanocompartments lacking the electron-dense content of iron-containing granules. Sections through the difference map are shown (~2.5 Å thick and 22 Å apart). Contours delineating the outer and inner surfaces of the T=3 shell have been superimposed in half of each section to aid in identifying the locations of the densities. Bar, 10 nm.

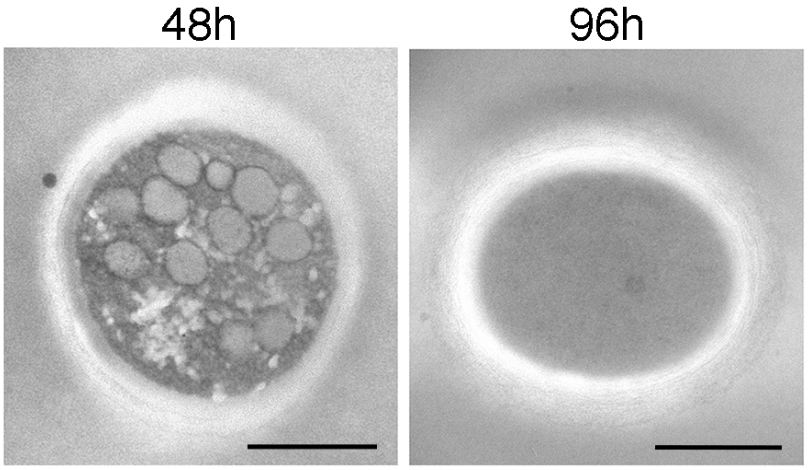

Fig. S10. Cross sections of isolated unstained *M. xanthus* spores after 48 and 96 h. Both spores are devoid of the encapsulin-associated iron clusters that are visible in vegetative cells upon starvation. Note the occurrence and subsequent disappearance of intra-cellular lipid bodies that are important for developmental differentiation (Hoiczyk et al, 2009). Scale bars, 0.5 μm.

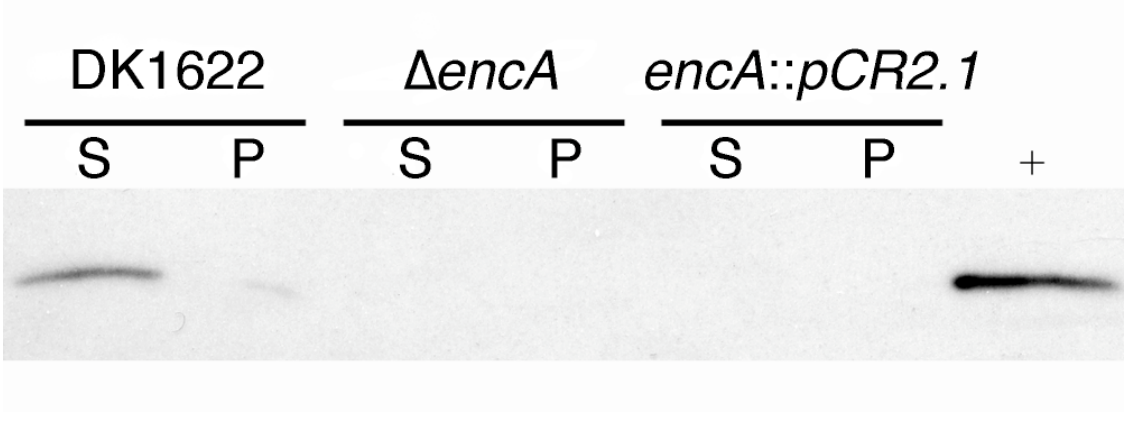

Fig. S11. Mutant strains *encA*::*pCR2.1* (EH410) and Δ*encA* (EH411) do not produce encapsulin. Western blot with antibody G3576 targeting encapsulin from wild-type cells and mutant strains. Cells were harvested and lysed, then centrifuged to separate soluble and insoluble material. S, supernatant; P, pellet; +, positive control of purified encapsulin protein.

References

Cuff JA, Barton GJ (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **40:** 502-511

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-230

Heckman KL, Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2:** 924-932

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36:** W5-9

Kim D, Chung J, Hyun H, Lee C, Lee K, Cho K (2009) Operon required for fruiting body development in Myxococcus xanthus. *J Microbiol Biotechnol* **19:** 1288-1294

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16:** 404-405

Pollastri G, McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* **21:** 1719-1720

Pruitt KD, Tatusova T, Klimke W, Maglott DR (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* **37:** D32-36

Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Res* **32:** W321-326

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671-675

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33:** W244-248

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680

Ueki T, Inouye S, Inouye M (1996) Positive-negative KG cassettes for construction of multi-gene deletions using a single drug marker. *Gene* **183:** 153-157