

36 **Abstract**

37 Large protein complexes play key roles in mediating biological processes in the cell. Little
38 structural information is known on the protein complex mediators governing persistence in the
39 host for *Mycobacterium tuberculosis* (*Mtb*). We applied the ‘shotgun EM’ method for the
40 structural characterisation of protein complexes produced after exposure to stationary phase
41 stress for the model Mycobacterium, *M smegmatis* (*Msm*). We identified glutamine synthetase
42 I, essential for *Mtb* virulence, in addition to bacterioferritin, critical for *Mtb* iron regulation,
43 and encapsulin, which produces a cage-like structure to enclose target proteins. Further
44 investigation found that encapsulin carries dye-decolourising peroxidase (DyP), a potent
45 protein antioxidant, as the primary cargo during stationary phase stress. Our ‘proof-of-concept’
46 application of this method offers insight into identifying potential key-mediators in *Mtb*
47 persistence.

48

49 **Introduction**

50 Protein machinery allows for the vast biochemical capabilities of cells, driving replication,
51 repair, and response to environmental stresses. The pathogenic bacteria, *Mycobacterium*
52 *tuberculosis* (*Mtb*), relies on a range of critical protein machinery to evade and utilise the host
53 immune response (1). Although some studies have shown potential key mediators in *Mtb*
54 survival in the host (2–4) there is a lack of structural information to guide drug discovery or
55 elucidate biological function.

56

57 Structural data provides compelling evidence for the existence of protein complexes which are
58 physiologically relevant to the cell. In addition, this data offers valuable information on subunit
59 composition and mechanism of interaction, which yields biological insight governing its
60 potential actions (5). Single-particle transmission electron microscopy (TEM) is a powerful

61 method to reconstruct large protein complexes. The method has been successfully used to solve
62 the structures of protein complexes in a range of organisms from homogenous (6) as well as
63 heterogenous (7–10) samples. Combining 3D reconstruction of protein complexes from
64 negative stain EM images with information obtained from mass spectrometry data (‘shotgun
65 EM’) (10) allows for a potentially high-throughput approach to finding complexes which could
66 play a key role in *Mtb* pathogenesis.

67

68 Although the ‘shotgun EM’ approach offers a faster method for solving the structures of a range
69 of protein complexes without the need for extensive purification (11), the approach still needs
70 to be tested and adapted for the organism of application (9,10,12). Here, we present our
71 ‘shotgun EM’ methodology for the purification and TEM 3D reconstruction of Mycobacterial
72 protein complexes from the model organism, *M smegmatis* (*Msm*), after exposure to stationary
73 phase stress which is known to induce a protective effect against subsequent exposure to
74 oxidative stress (13). We identified three protein complexes (glutamine synthetase I (GSI) (E.C
75 6.3.1.2), bacterioferritin (BrfA) (E.C 1.16.3.1), and encapsulin); the initial identity assigned by
76 mass spectrometry, as well as the 3D model and subunit composition, was validated by fitting
77 of homologous crystallographic structures into the low-resolution density. We demonstrate that
78 encapsulin encloses dye-decolourising type peroxidase (DyP) (E.C 1.11.17), a potent
79 enzymatic anti-oxidant (14), as the main cargo during stationary phase stress. Furthermore,
80 analysis of our encapsulated DyP shows that it binds on the encapsulin three-fold axis,
81 validating the relationship between cargo binding and substrate access *in vivo*.

82

83 **Results**

84 **Partial fractionation in combination with *in silico* purification yields three-dimensional**
85 **structures.** We induced production of protein complexes potentially involved in the oxidative

86 stress response by growing *Msm* culture to the end of stationary phase (13). As an initial partial
87 fractionation step, we applied ammonium sulfate precipitation to cell lysate. We noted that the
88 EM micrographs displayed a large amount of soluble aggregated protein. Thus, we opted to
89 use a manual picking strategy to limit the amount of “junk” particles in the dataset and improve
90 classification of true (non-aggregated) protein complexes. However, the resolution of
91 separation was too low to build models after applying *in silico* purification by particle picking
92 and classification (**SI Figure 1**). We thus opted for either partial fractionation by anion
93 exchange or sucrose-cushioning. We applied a size cut-off (>100 kDa) when concentrating
94 obtained peaks for EM analysis. Manually picked particles were used to generate separate EM
95 models using the symmetry information gathered from the average in the particle stack (**Figure**
96 **1**).

97
98 **Identification by mass spectrometry and verification by model fitting.** We succeeded in
99 unambiguous identification of our models through a combination of information from mass
100 spectrometry and fitting of homologous structures into our low-resolution models (~25 Å)
101 (**Figure 2**). We first used the EM models to estimate molecular weight, which was roughly size
102 matched to bands in a native PAGE gel of anion exchange fractions. Analysis by liquid
103 chromatography tandem mass spectrometry (LC–MS/MS) showed that peptides for glutamine
104 synthetase I (GSI) and 29 kDa antigen Cfp29 (encapsulin) showed respective highest number
105 of unique peptides and coverage (**SI Table 1**).

106
107 We noted extra density present in the encapsulin particles; furthermore, it is known that
108 recombinant *Mtb* encapsulin can contain three types of cargo protein: dye-decolourising type
109 peroxidase (DyP), bacterioferritin B (BrfB), or 7,8-dihydroneopterin aldolase (FolB) (15).
110 However, we found no plausible hits which could identify the extra density in the LC–MS/MS

111 data (**SI Table 1**). We looked at both GSI and encapsulin in the Mycobrowser
112 (<https://mycobrowser.epfl.ch/>) database (16) and found that they are both membrane associated.
113 Hence to find a potential cargo lead for encapsulin, we isolated the membrane fraction of *Msm*,
114 and ran the resolubilised material on either blue or clear native PAGE; to not introduce bias,
115 all bands were cut out and analysed by LC–MS/MS (**Figure 2a**; **SI Table 2**). Both GSI and
116 encapsulin (Cfp29) were the major peptides found in different blue native bands (**Figure 2a**).
117 Bacterioferritin (BrfA) was also found as a minor peptide in two clear native bands (**SI Table**
118 **2**). The only major peptide which was co-found exclusively with Cfp29 was DyP (**Figure 2a**;
119 **SI Table 2**). Thus, it seemed likely that this was the potential cargo protein for *Msm* encapsulin.
120 To confirm the result, we ran SDS-PAGE on gel filtration fractions confirmed by EM analysis
121 of micrographs to harbour encapsulin particles and cargo (**SI Figure 2**). We cut out a band
122 which corresponded to the approximate size of DyP (~40 kDa) and analysed it by mass
123 spectrometry. Hits which did not match the mass of the band calculated using a standard
124 calibration curve for the gel were excluded from analysis. DyP was found with 6.4% coverage
125 (**SI Table 3**). None of the other known cargo proteins were found.

126

127 To verify protein identities assigned by LC–MS/MS, we searched the Protein Data Bank (PDB)
128 (17) for available solved structures which matched the symmetry of our low-resolution models.
129 We fitted hits manually at first, and optimised by local fitting refinement and checked by cross-
130 correlation; GSI (1hto) (0.8795) (18), encapsulin (3dkt) (0.7319) (19), and bacterioferritin
131 (3uno) (0.9361) (20). Fitted crystallographic structures also showed good correspondence to
132 the symmetry of our models (**Figure 2b**), validating the subunit composition.

133

134 **Binding of DyP cargo on encapsulin three-fold axis and possible export.** It is known that
135 encapsulins function to enclose target proteins via a unique C-terminal extension on the

136 encapsulated protein (15, 21). Additionally, both *Msm* DyP and BrfB, but not *Msm* FolB or
137 BrfA, contain C-terminal extensions (15, 22) (**SI Figure 3**). However, only DyP was identified
138 in our previous LC–MS/MS results. Inspection of *Msm* encapsulin particles showed that BrfB
139 could potentially be present (**Figure 3a**).

140

141 Since all captured proteins are membrane associated, and there is previous indication that *Mtb*
142 Cfp29 (encapsulin) can be exported (23), we partially purified cell-culture filtrate by anion
143 exchange. We detected the presence of encapsulin and cargo by EM micrographs, but failed to
144 detect an SDS-PAGE band corresponding to the expected size of BrfB (**SI Figure 4**). Thus, it
145 seemed unlikely that BrfB is the main encapsulin cargo. Furthermore, it has been proposed that
146 any bound cargo lies at a specific symmetry axis on encapsulin, corresponding to the symmetry
147 of the bound protein (e.g two-fold, three-fold, or five-fold axis) (21) (**SI Figure 3**). DyP could
148 potentially bind on the encapsulin three-fold axis (**SI Figure 3**). To address this question, we
149 manually removed empty/deformed particles or those which may contain BrfB, and
150 reconstructed the encapsulated cargo by masking out the shell. The obtained model is highly
151 consistent with previously solved peroxidase structures, showing a hexameric subunit
152 arrangement, with the C-terminal extension lying on the three-fold axis (**Figure 3b**) (14,21).
153 By docking our model back into *Msm* encapsulin density, we found that the three-fold axis
154 showed good alignment with the tethering C-terminal extension found in *T. maritima*
155 encapsulin crystal structure; only one hexamer can be accommodated in the lumen (**Figure 3c**).
156 To confirm the possibility of DyP binding at a specific symmetry axis, we applied C3 symmetry
157 to unmasked 2D projections; this revealed density clustered to the encapsulin three-fold axis,
158 which is likely to be DyP (**Figure 3d**). Thus, DyP was confirmed to be the primary cargo *in*
159 *vivo* of *Msm* encapsulin under stationary phase stress and there is some evidence that it binds

160 specifically to the three-fold axis, a result which has also been found for *B. linens* encapsulin
161 (24).

162

163 **Discussion**

164 We obtained three protein complexes (GSI, bacterioferritin, and encapsulin) by ‘shotgun EM’
165 after *Msm* was exposed to stationary phase stress. Identification of these initially unknown
166 protein complexes proved to be particularly challenging and relied on a combination of LC–
167 MS/MS of native PAGE gel bands and fitting of homologues crystal structures. Our LC–
168 MS/MS results were confirmed by multiple partial fractionation strategies. However, since it
169 is known that peptide abundance does not strictly correlate to protein abundance (25), and
170 because the number of unique identified peptides for all three proteins was low, we opted to
171 verify the mass spectrometry results by finding homologous crystal structures in the PDB to fit
172 into our low-resolution models. The crystal structure homologues matched the mass
173 spectrometry data, lending confidence to the result. One drawback of this approach is that it
174 relies on both the availability of a homologue in the PDB and the conservation of the quaternary
175 structure. Use of higher resolution structures could potentially eliminate the need for an
176 existing PDB structure to verify the mass spectrometry results.

177

178 Under stationary phase stress, *Msm* encapsulin appears to primarily enclose DyP, a potent
179 protein antioxidant (14). The main role of BrfB, another potential encapsulated protein, is to
180 relieve iron toxicity (26). EM models provide a good indication of the relative abundance of
181 binding partners in a protein complex; more abundant members will dominate the density
182 (7,27). We noted that the majority of particles appeared to contain DyP rather than BrfB cargo;
183 our reconstruction showed a typical peroxidase structure, which was unlikely to have been
184 biased by any remaining BrfB particles. The production of encapsulated DyP may have

185 functional significance in *Msm*, as one of the mechanisms in which *Msm* has increased
186 resistance to oxidative stress after growth in stationary phase (13). Furthermore, we found that
187 *in vivo* DyP is bound on the encapsulin three-fold axis, implying that a specific pore is involved
188 in the entry of the hydrogen peroxide substrate. Since it is also possible for BrfB to be bound
189 at the three-fold axis, and reaction of hydrogen peroxide with iron can produce harmful
190 hydroxyl radicals via the Fenton reaction (28), it remains to be further investigated how
191 encapsulin regulates substrate entry.

192

193 GSI is known to be essential in *Mtb* virulence (29) and growth in macrophages (30), while
194 bacterioferritin (BrfA) is crucial for iron regulation (31). We confirmed that all three proteins
195 are membrane-associated by isolating the membrane fraction and analysing solubilised proteins
196 by LC-MS/MS. This led us to examine the cell-culture filtrate and we found some evidence
197 for the presence of encapsulin and DyP cargo. This leads to an interesting possibility that these
198 proteins could potentially be exported during stationary phase stress, as suggested by previous
199 investigators (23,32). Thus, we have isolated three protein complexes by the new ‘shotgun EM’
200 methodology which can all be implicated in *Msm* response to stationary phase stress, and also
201 play important roles in *Mtb* persistence in the host. Our proof-of-concept study has the potential
202 to be adapted to find more protein complexes involved in *Mtb* pathogenesis.

203

204 **Experimental Procedures**

205

206 **Bacterial Growth.** Culture (Middlebrook 7H9 media supplemented with 0.2% glucose, 0.2%
207 glycerol, and 0.05% Tween- 80) of *Msm* groELΔC (33) was inoculated with 1/100 starter
208 culture and grown at 37°C with shaking (120 rpm) to the end of stationary phase (~4–5 days).

209 Cells were harvested through centrifugation (4000g for 30 minutes) at 4°C. The pellet was
210 stored at -80°C.

211

212 **Cell Lysis and Ammonium Sulphate Precipitation.** The pellet was thawed and resuspended
213 in 25 mL of lysis buffer (50 mM Tris-HCl, 300 mM NaCl, pH 7.2) with protease inhibitor
214 cocktail. Cells were lysed through 4 x (15 seconds on, 15 seconds off for 4 minutes) on ice.
215 Cell debris was pelleted by centrifugation (20,000g for 1 hour) at 4°C.

216 Ammonium sulphate cuts were completed on the filtered (0.45 µm) supernatant (<40%,
217 40–50%, 50–60%, and >60%). For each cut, the ammonium sulphate was added slowly on ice
218 with continual stirring and incubated for 30 minutes before centrifuging at 9000g for 15
219 minutes. Pellets were clarified by re-suspending in 20 mL of gel filtration buffer (50 mM Tris-
220 HCl, 200 mM NaCl, pH 8.0) and centrifuged at 20,000g for 10 minutes at 4°C. The ammonium
221 sulphate cuts were then buffer exchanged to gel filtration buffer using an Amicon® spin-filter
222 with a 100 kDa cut-off (Merck, Darmstadt, Germany).

223

224 **Anion Exchange.** 20 mL HiPrep Q FF 16/10 column (GE Healthcare Life Sciences,
225 Massachusetts, USA) was equilibrated with 5–10 column volumes of start buffer (20 mM Tris-
226 HCl, 20 mM NaCl, pH 8.0). Samples were then eluted with 0.5 M NaCl (3 CV), before applying
227 a gradient of 0.5 – 1 M NaCl (19.5 CV). The flow rate was 5 mL/min with 60 fractions collected.
228 Fractions were stored at 4°C.

229

230 **Gel Filtration.** The PWXL5000 column (Tosoh Biosciences, Tokyo, Japan) was equilibrated
231 with gel filtration buffer and run at a flow rate of 0.5 mL/min for 1 column volume. Fractions
232 were stored at 4°C.

233

234 **Sucrose Cushioning.** The method was adapted from Peyret (2015) (34) with the following
235 modifications: a double sucrose cushion consisting of 25% (top layer) and 70% (bottom layer)
236 sucrose made in sodium phosphate buffer (pH 7.4). The sample was spun at 170,462g for 5
237 hours. The layer just above the 70% cushion was extracted and buffer exchanged to gel
238 filtration buffer using an Amicon® spin-filter with a 100 kDa cut-off (Merck, Darmstadt,
239 Germany).

240

241 **Membrane Preparation and Electrophoresis.** 2 L of *Msm* culture was grown as described
242 and membranes prepared for electrophoresis as described previously (35,36). For clear native
243 PAGE, standard continuous Tris-Glycine (pH 8.8) system was used.

244

245 **Negative Stain Electron Microscopy.** Samples were pipetted onto a glow-discharged (in air)
246 copper grid and washed/stained with 5 rounds of 2% uranyl acetate before being air-dried.
247 Images were taken using the Tecnai F20 transmission electron microscope (Phillips/FEI,
248 Eindhoven, The Netherlands) fitted with a CCD camera (4k x 4k) (GATAN US4000 Ultrascan,
249 USA) at 200 kV under normal dose conditions with a defocus of 2.00 μm at the appropriate
250 magnification. The sampling rate was 2.11 or 3.84 $\text{\AA}/\text{pixel}$.

251

252 **Class Averages.** Class averages were produced in Appion (37) from manually picked particles.
253 Briefly, the Contrast Transfer Function (CTF) was estimated using ACE2 (38) and poor images
254 excluded based on the presence of astigmatism, bad staining, or noticeable microscope drift. A
255 stack was created with CTF correction (ACE2 Phaseflip of whole image) (38). A Spider
256 reference-free alignment (39) was completed, averaging all particles in the stack. Afterwards,
257 Spider Coran classification (39) was completed using appropriate settings. Either K-means or
258 hierarchical clustering was completed using selected eigen images.

259

260 **Reconstruction.** All reconstructions were completed in the Appion pipeline (37). Particles
261 were binned by a factor of 2 for a sampling of 2.11 Å/pixel. The appropriate number of classes
262 was used to complete an initial reconstruction using EMAN Common Lines (40) with the
263 appropriate symmetry imposed. The model was then refined using EMAN model refinement
264 (40) for 26 iterations with the appropriate symmetry imposed; 20 iterations was used for GSI.
265 Angular sampling was as follows: 5 iterations of 10°, 5 iterations of 8°, 10 iterations of 5°, and
266 6 iterations of 3°. For GSI, the angular sampling was 20 iterations of 5°.

267

268 **Reconstruction of Encapsulated Dye-Decolourising Peroxidase.** A sub-stack was created in
269 which particles that were broken, deformed, or may contain BrfB were deleted. This left 207
270 particles. Encapsulin was masked out using a rectangular box with a Gaussian drop-off in
271 intensity. Class averages were produced as described previously using hierarchical clustering.
272 An initial model was created using EMAN Common Lines (40) with C3 symmetry imposed.
273 This model was refined using EMAN projection-matching (40) with D3 symmetry imposed for
274 26 iterations. For refinement, a 15 Å low pass filter was used and a mask radius of 70 Å applied.
275 Angular sampling rate was used as described previously.

276

277 **Mass Spectrometry.** Samples were sent for MS either to the Blackburn Group (in-gel native
278 PAGE LC-MS/MS) (University of Cape Town, South Africa) or to the Yale MS & Proteomics
279 Resource (in gel SDS-PAGE LC-MS/MS) (Yale School of Medicine, New Haven, USA).
280 Samples were digested with trypsin and analysed on an LTQ Orbitrap (ThermoScientific,
281 Massachusetts, USA). MS/MS spectra were searched using the Mascot algorithm (41). Peaks
282 with a charge state of +2 or +3 were located first using a signal-to-noise ratio of >1.2. Potential
283 peaks were screened against the NCBI nr or SWISS-PROT (42) databases.

284

285 **Bioinformatics.** Obtained EM models were imported into UCSF-Chimera (43) and set to the
286 correct voxel size. Crystal structural homologues were manually docked into the low-resolution
287 EM maps and the fit refined using the ‘Fit in Map’ function. Fits were checked by first making
288 a low resolution map of the crystal structure (‘molmap’) and then applying ‘measure
289 correlation’. For MW estimates, protein mass (in Da) was calculated for the estimated lower
290 and upper contour level limits using the following calculation: $825 * V$, where V is the volume
291 (in nm³) of the model density at the specific contour level. See Erickson (2009) (44) for details
292 on the calculation.

293

294 **Data availability.** Data supporting the findings of this manuscript are available from the
295 corresponding author upon reasonable request. The maps have been deposited in the Electron
296 Microscopy Data Bank (<http://www.ebi.ac.uk/pdbe/emdb/>) (45) : *Mycobacterium smegmatis*
297 encapsulin: EMD-4175; *Mycobacterium smegmatis* DyP type peroxidase: EMD-10004;
298 *Mycobacterium smegmatis* with DyP type peroxidase bound on 3-fold axis: EMD-10008;
299 *Mycobacterium smegmatis* glutamine synthetase: EMD-4186; *Mycobacterium smegmatis*
300 bacterioferritin: EMD-10005. Protein sequences are available from Mycobrowser
301 (<https://mycobrowser.epfl.ch/>) (16): *M smegmatis* glutamine synthetase I (MSMEG_3828),
302 bacterioferritin (MSMEG_3564), encapsulin (MSMEG_5830), and dye-decolourising type
303 peroxidase (MSMEG_5829).

304

305 **Supplementary Methods**

306 **Phylogenetic Analysis.** Alignments of protein sequences were produced in UCSF-Chimera
307 (43) and exported to MEGA6 (49) for phylogenetic analysis. Alignment of DNA sequences

308 was completed in MEGA6 using MUSCLE (50) with default parameters. For protein sequences,
309 a neighbour joining-tree was produced using p-distance to model amino acid substitution; the
310 rate of substitution was assumed to be uniform and the pattern among lineages homogenous;
311 gaps or missing data were deleted in the analysis. For DNA sequences, a minimal evolution
312 tree was constructed using p-distance to model nucleotide substitutions; only transitions were
313 included while the rate of substitution was assumed to be uniform and homogenous across
314 lineages; gaps or missing data were deleted from the analysis. Trees were bootstrapped using
315 1000 replicates.

316

317 **References**

318 1. Zhai, W., Wu, F., Zhang, Y., Fu, Y., Liu, Z. (2019) The Immune Escape Mechanisms of
319 *Mycobacterium Tuberculosis*. *Int. J. Mol.* **20(2)**, 340.

320

321 2. Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., Zhang, L., Gao, C., He, Y., Li,
322 Y., Huang, F, *et al.* (2010) Global Protein-Protein Interaction Network in the Human Pathogen
323 *Mycobacterium tuberculosis* H37Rv. *J. Proteome. Res.* **9**, 6665-77.

324

325 3. Huo, T., Liu, W., Guo, Y., Yang, C., Lin, J., Rao, Z. (2015) Prediction of host - pathogen
326 protein interactions between *Mycobacterium tuberculosis* and *Homo sapiens* using sequence
327 motifs. *BMC Bioinformatics.* **16**, 100.

328

329 4. Sun, J., Yang, L.L., Chen, X., Kong, D.X., Liu, R. (2018) Integrating Multifaceted
330 Information to Predict *Mycobacterium tuberculosis*-Human Protein-Protein Interactions. *J.*
331 *Proteome. Res.* **17**, 3810-23.

332

- 333 5. Edwards, A.M., Kus, B., Jansen, R., Greenbaum, D., Greenblatt, J., Gerstein, M. (2002)
334 Bridging structural biology and genomics: assessing protein interaction data with known
335 complexes. *Trends Genet.* **18(10)**, 529-36.
336
- 337 6. Han, B.G., Dong, M., Liu, H., Camp, L., Geller, J., Singer, M., Hazen, T.C., Choi, M.,
338 Witkowska, E., Ball, D.A., Typke, D., *et al.* (2009) Survey of large protein complexes in *D.*
339 *vulgaris* reveals great structural diversity. *PNAS U.S.A.* **106(39)**, 16580-5.
340
- 341 7. Aloy, P., Böttcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C.,
342 Bork, P., Superti-Furga, G., Serrano, L., Russell, R.B. (2004) Structure-Based Assembly of
343 Protein Complexes in Yeast. *Science.* **303**, 2026-9.
344
- 345 8. Maco, B., Ross, I.L., Landsberg, M.J., Mouradov, D., Saunders, N.F., Hankamer, B., Kobe,
346 B. (2011) Proteomic and electron microscopy survey of large assemblies in macrophage
347 cytoplasm. *Mol. Cell. Proteom.* **10(6)**, M111.008763.
348
- 349 9. Kastritis, P.L., O'Reilly, F.J., Bock, T., Li, Y., Rogon, M.Z., Buczak, K., Romanov, N., Betts,
350 M.J., Bui, K.H., Hagen, W.J., Hennrich, M.L., *et al.* (2017) Capturing protein communities by
351 structural proteomics in a thermophilic eukaryote. *Mol. Syst. Biol.* **13**, 936.
352
- 353 10. Verbeke, E.J., Mallam, A.L., Drew, K., Marcotte, E.M., Taylor, D.W. (2018) Classification
354 of Single Particles from Human Cell Extract Reveals Distinct Structures. *Cell Rep.* **24**, 259-68.
355
- 356 11. Kyrilidis, F.L., Meister, A. & Kastritis, P.L. (2019) Integrative biology of native cell extracts:
357 a new era for structural characterization of life processes. *Biol. Chem.* **400(7)**, 831-46.

358

359 12. Yi, X., Verbeke, E.J., Chang, Y., Dickinson, D.J., Taylor, D.W. (2019) Electron
360 microscopy snapshots of single particles from single cells. *J. Biol. Chem.* **294(5)**, 1602-8.

361

362 13. Smeulders, MJ, Keer, J, Speight, RA, Williams, HD. (1999) Adaptation of *Mycobacterium*
363 *smegmatis* to Stationary Phase. *J. Bacteriol.* **181(1)**, 270-83.

364

365 14. Zubieta, C., Krishna, S.S., Kapoor, M., Kozbial, P., McMullan, D., Axelrod, H.L., Miller,
366 M.D., Abdubek, P., Ambing, E., Astakhova, T., Carlton, D., *et al.* (2007) Crystal structures of
367 two novel dye-decolorizing peroxidases reveal a β -barrel fold with a conserved heme-binding
368 motif. *Proteins.* **69**, 223-33.

369

370 15. Contreras, H., Joens, M.S., McMath, L.M., Le, V.P., Tullius, M.V., Kimmey, J.M., Bionghi,
371 N., Horwitz, M.A., Fitzpatrick, J.A., Goulding, C.W. (2014) Characterization of a
372 *Mycobacterium tuberculosis* Nanocompartment and Its Potential Cargo Proteins. *J. Biol. Chem.*
373 **289(26)**, 18279-89.

374

375 16. Kapopoulou, A., Lew, J.M. & Cole, S.T. (2011) The MycoBrowser portal: a comprehensive
376 and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb).* **91(1)**, 8-
377 13.

378

379 17. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov,
380 I.N., Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28(1)**, 235-42.

381

- 382 18. [dataset] Gill, H.S., Eisenberg, D., TB Structural Genomics Consortium (TBSGC). (2002)
383 Crystallographic structure of a relaxed glutamine synthetase from *Mycobacterium tuberculosis*.
384 Protein Data Bank. 1HTO.
385
- 386 19. [dataset] Sutter, M., Boehringer, D., Gutmann, S., Weber-Ban, E., Ban, N. (2008) Crystal
387 structure of *Thermotoga maritima* encapsulin. Protein Data Bank. 3DKT.
388
- 389 20. [dataset] McMath, L.M., Contreras, H., Goulding, C.W., TB Structural Genomics
390 Consortium (TBSGC). (2012) *Mycobacterium tuberculosis* ferritin homolog, BfrB. 3UNO.
391
- 392 21. Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M.J.,
393 Stetter, K.O., Weber-Ban, E., Ban, N. (2008) Structural basis of enzyme encapsulation into a
394 bacterial nanocompartment. *Nat. Struct. Biol.* **15(9)**, 939-47.
395
- 396 22. Khare, G., Gupta, V., Nangpal, P., Gupta, R.K., Sauter, N.K., Tyagi, A.K. (2011) Ferritin
397 Structure from *Mycobacterium tuberculosis*: Comparative Study with Homologues Identifies
398 Extended C-Terminus Involved in Ferroxidase Activity. *PloS One.* **6(4)**, e18570.
399
- 400 23. Rosenkrands, I., Rasmussen, P.B., Carnio, M., Jacobsen, S., Theisen, M., Andersen, P.
401 (1998) Identification and Characterization of a 29-Kilodalton Protein from *Mycobacterium*
402 *tuberculosis* Culture Filtrate Recognized by Mouse Memory Effector Cells. *Infect. Immun.*
403 **66(6)**, 2728-35.
404
- 405 24. Putri, R.M., Allende-Ballesteros, C., Luque, D., Klem, R., Rousou, K.A., Rurup, W.F., Koay,
406 M.S.T., Caston, J.R., Cornelissen, J.J.L.M. (2017) Structural Characterization of Native and

- 407 Modified Encapsulins as Nanoplatforms for *in Vitro* Catalysis and Cellular Uptake. *ACS Nano*.
408 **11**, 12796-804.
- 409
- 410 25. Cottrell, J.S. (2011) Protein identification using MS/MS data. *J. Proteom.* **74**, 1842- 51.
411
- 412 26. Khare, G., Nangpal, P. & Tyagi, A.K. (2017) Differential Roles of Iron Storage Proteins in
413 Maintaining the Iron Homeostasis in *Mycobacterium tuberculosis*. *PLoS One*. **12(1)**, e0169545.
414
- 415 27. De Zorzi, R., Mi, W., Liao, M., Walz, T. (2016) Single-particle electron microscopy in the
416 study of membrane protein structure. *Microscopy*. **65(1)**, 81-96.
417
- 418 28. Smith, J.L. (2004) The Physiological Role of Ferritin-Like Compounds in Bacteria. *Crit.*
419 *Rev. Microbiol.* **30**, 173-85.
420
- 421 29. Harth, G., Masleš-Galić, S., Tullius, M.V., Horwitz, M.A. (2005) All four *Mycobacterium*
422 *tuberculosis* *glnA* genes encode glutamine synthetase activities but only GlnA1 is abundantly
423 expressed and essential for bacterial homeostasis. *Mol. Microbiol.* **58(4)**, 1157-72.
424
- 425 30. Tullius, M.V., Harth, G. & Horwitz, M.A. (2003) Glutamine Synthetase GlnA1 Is Essential
426 for Growth of *Mycobacterium tuberculosis* in Human THP-1 Macrophages and Guinea Pigs.
427 *Infect. Immun.* **71(7)**, 3927-36.
428
- 429 31. Zondervan, N.A., van Dam, J.C.J., Schaap, P.J., Martins dos Santos, V.A.P., Suarez-Diez,
430 M. (2018) Regulation of Three Virulence Strategies of *Mycobacterium tuberculosis*: A Success
431 Story. *Int. J. Mol. Sci.* **19**, 347.

432

433 32. de Souza, G.A., Leversen, N.A., Målen, H., Wiker, H.G. (2011) Bacterial proteins with
434 cleaved or uncleaved signal peptides of the general secretory pathway. *J. Proteom.* **75**, 502-10.

435

436 33. Noens, E.E., Williams, C., Anandhakrishnan, M., Poulsen, C., Ehebauer, M.T., Wilmanns,
437 M. (2011) Improved mycobacterial protein production using a *Mycobacterium smegmatis*
438 groEL1ΔC expression strain. *BMC Biotechnol.* **11**, 27.

439

440 34. Peyret, H. (2015) A protocol for the gentle purification of virus-like particles produced in
441 plants. *J. Virol. Methods.* **225**, 59-63.

442

443 35. Wittig, I., Braun, H.P., Schägger, H. (2006) Blue native PAGE. *Nat. Protoc.* **1**, 418-28.

444

445 36. Zheng, J, Wei, C, Zhao, L, Liu, L, Leng, W, Li, W, Jin, Q. (2011) Combining blue native
446 polyacrylamide gel electrophoresis with liquid chromatography tandem mass spectrometry as
447 an effective strategy for analyzing potential membrane protein complexes of *Mycobacterium*
448 *bovis* bacillus Calmette-Guérin. *BMC Genomics.* **12**, 40.

449

450 37. Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C.,
451 Irving, C., Mulder, A., Lau, P.W., Lyumkis, D., *et al.* (2009) Appion: an integrated, database-
452 driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166(1)**, 95-102.

453

454 38. Carragher, B. & Potter, C. (2009) ACE2. URL:
455 <http://emg.nysbc.org/redmine/projects/software/wiki/ACE2> [Last accessed February 2018]

456

- 457 39. Frank J., Radermacher M., Penczek P., Zhu J., Li Y., Ladjadj M., Leith A. (1996) SPIDER
458 and WEB: processing and visualization of images in 3D electron microscopy and related fields.
459 *J. Struct. Biol.* **116(1)**, 190-9.
- 460
- 461 40. Ludtke S.J., Baldwin P.R. & Chiu W. (1999) EMAN: semiautomated software for high-
462 resolution single-particle reconstructions. *J. Struct. Biol.* **128(1)**, 82-97.
- 463
- 464 41. Hirose, M., Hoshida, M., Ishikawa, M., Toya, T. (1993) MASCOT: multiple alignment
465 system for protein sequences based on three-way dynamic programming. *Comput. Appl. Biosci.*
466 **9(2)**, 161-7.
- 467
- 468 42. Bairoch, A. & Apweiler, R. (2000) The SWISS-PROT protein sequence database and its
469 supplement TrEMBL in 2000. *Nucleic Acids Res.* **28(1)**, 45-8.
- 470
- 471 43. Petterson, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C.,
472 Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and
473 analysis. *J. Comput. Chem.* **25(13)**, 1605-12.
- 474
- 475 44. Erickson, H.P. (2009) Size and Shape of Protein Molecules at the Nanometer Level
476 Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online.*
477 **11**, 32-51.
- 478
- 479 45. Tagari, M., Newman, R., Chagoyen, M., Carazo, J.M., Henrick, K. (2002) New electron
480 microscopy database and deposition system. *Trends Biochem. Sci.* **27**, 589.
- 481

482 46. [dataset] Joint Center for Structural Genomics (JCSG). (2006) Crystal structure of a dye-
483 decolorizing peroxidase (DyP) from *Bacteroides thetaiotaomicron* VPI-5482 at 1.6 Å
484 resolution. Protein Data Bank. 2GVK.

485

486 47. [dataset] Sutter M., Boehringer D., Gutmann S., Gunther S., Prangishvili D., Loessner M.J.,
487 Stetter K.O., Weber-Ban E., Ban N (2009). Structure of hexameric DyP from *Brevibacterium*
488 *linens*. Electron Microscopy Data Bank. EMD-1530.

489

490 **Supplementary Information References**

491 48. [dataset] Goulding, C.W., Apostol, M.I., Sawaya, M.R., Phillips, M., Parseghian, A.,
492 Eisenberg, D., TB Structural Genomics Consortium (TBSGC). (2004) 7,8-Dihydroneopterin
493 Aldolase Complexed with Product From *Mycobacterium Tuberculosis*. Protein Data Bank.
494 1NBU.

495

496 49. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular
497 Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30(12)**, 2725-9.

498

499 50. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
500 throughput. *Nucleic Acids Res.* **32(5)**, 1792-7.

501

502 **Acknowledgments**

503 We acknowledge Mohammed Jaffer and Brandon Weber for technical assistance and the
504 Blackburn Group, University of Cape Town and the Yale MS & Proteomics Resource, Yale
505 School of Medicine, New Haven, USA for mass-spectrometry measurements. Thanks to A.
506 Mulelu for helpful discussions. This work was funded by a Research Career Advancement

507 Fellowship from the National Research Foundation of South Africa (92556) and GCRF-
508 START (UK Science and Technology Facilities Council grant ref. ST/R002754/1).

509

510 **Author Contributions**

511 AMK and JDW conceived the project; AMK performed the experiments and wrote the
512 manuscript; JDW supervised the project.

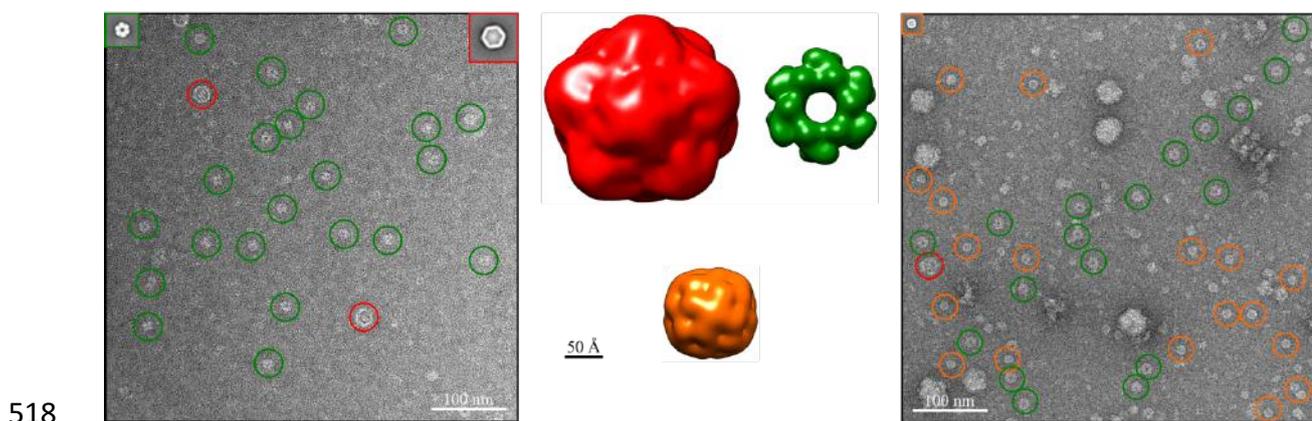
513

514 **Competing interests**

515 The authors declare no competing interests

516

517 **Figures**

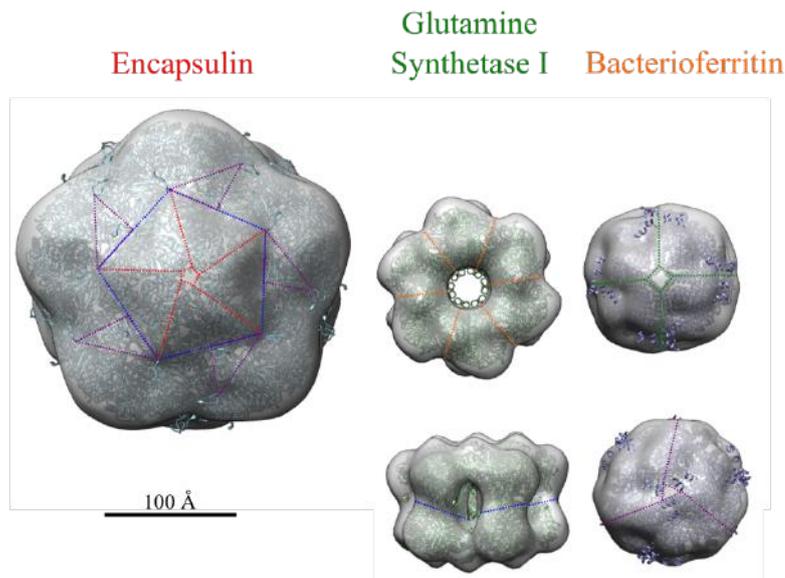


519 **Figure 1. Partial fractionation and *in silico* purification.** Three protein complexes were partially purified from
520 anion exchange chromatography (left) or sucrose cushioning (right). Particles were further purified *in silico*
521 through particle picking and classification. This led to reconstructions of encapsulin (cfp29) (red), glutamine
522 synthetase I (GSI) (green), and bacterioferritin (BrfA) (orange). Particles for encapsulin and GSI were also found
523 in sucrose density purification, but the particles were of too poor quality for averaging and further reconstruction.
524

a

Protein ID	Protein Name	# Unique Peptides	Clear Res 4	Blue Res 3	Blue Res 4	Blue Res 5	MW (kDa)	% Coverage
A0R079	Glutamine Synthetase I	6	0	0	0	6	53.591	16.5
A0R4H0	29 kDa antigen Cfp29	4	0	2	4	0	28.730	15.8
A0QXC0	ABC Transporter Substrate Binding Protein	4	4	0	0	0	40.502	12.9
A0R4G9	Dyp Type Peroxidase	2	0	0	2	0	37.216	10.8
A0QY79	Bacterioferritin	2	2	0	0	0	18.482	10.6

b



525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550

Figure 2. Mass spectrometry and model fitting. (a) We analysed bands cut out from clear and blue native gels of solubilised membrane-bound protein and analysed them by LC-MS/MS. Single MS/MS peptide hits, or those which were likely to be degraded, or present in blank runs were manually removed from analysis. For full data see SI Table 2. (b) We searched for crystal structures with the same symmetries in the PDB. Encapsulin from *Thermotoga maritima* (3dkt) (19), and glutamine synthetase I (1hto) (18) and bacterioferritin B (3uno) (20) from *Mycobacterium tuberculosis* were found and docked into the appropriate density. Crystal structures have good correspondence to the density and symmetry axis on each low-resolution structure. The fit was evaluated by cross-correlation. The 6-fold (orange), 4-fold (green), 5-fold (red), 3-fold (purple), and 2-fold (blue) symmetry axes are shown as appropriate for each structure.

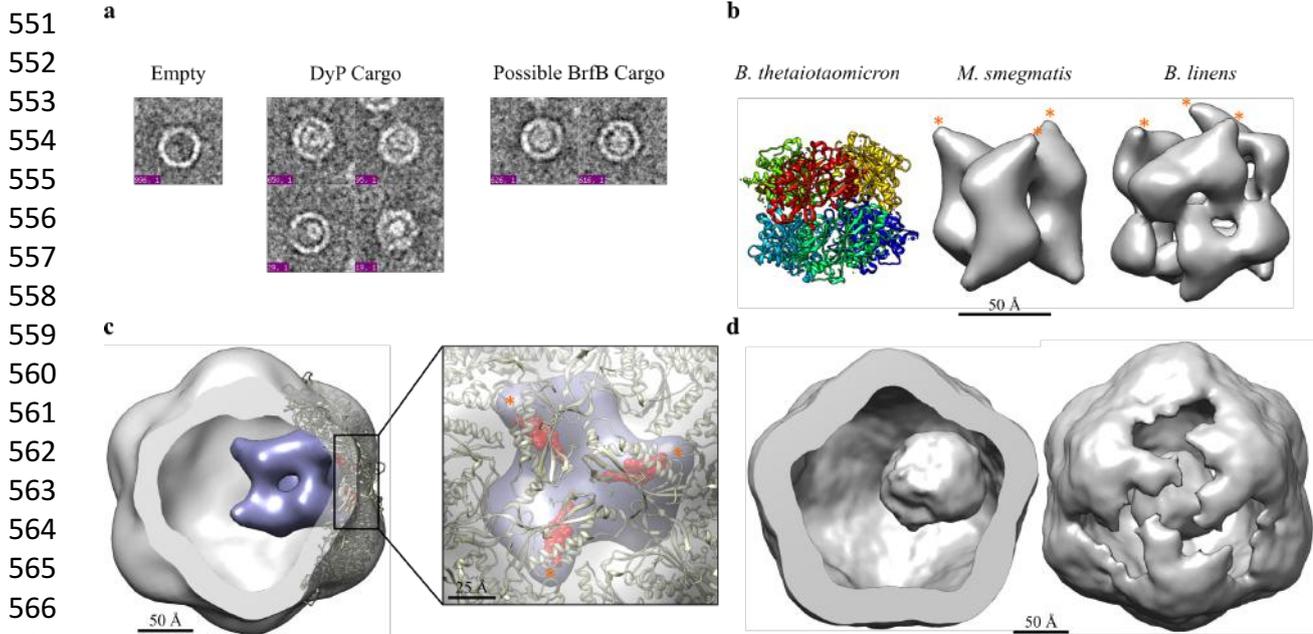
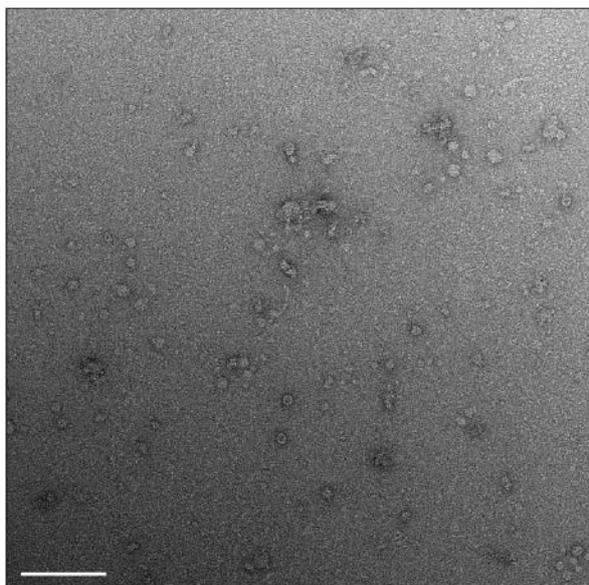


Figure 3. Identification of the primary cargo of *M. smegmatis* encapsulin in response to stationary phase stress. (a) Encapsulin particles can either be empty or contain either DyP-type peroxidase or bacterioferritin B cargo. Bacterioferritin B appears square-like, owing to its octahedral symmetry, while the 3- and 2-fold symmetries of DyP can be seen depending on the particle orientation. (b) Preliminary reconstruction of encapsulated DyP (middle) compared to other solved DyP crystal (left, pdb code 2gvk) (46) or low-resolution EM (right, emd-1530) (47) structures. The position of the C-terminal extension is starred (orange). Note that the C-terminal extension for *B. thetaiotaomicron* DyP was not built into the crystal structure. (c) Model of DyP (purple) docked into the 3-fold axis of encapsulin (grey) fills a substantial part of the lumen. It is clear from the docking that only a single DyP hexamer is likely to be accommodated. The C-terminal extension (red) from the *T. maritima* encapsulin crystal structure (brown) (pdb code 3dkt) (19) aligns well with the location of the missing DyP extension (starred, orange). (d) Application of three-fold symmetry reveals DyP on the three-fold symmetry axis (left), corresponding to the expected position based on the docking model in (c). The density of encapsulin has been partially stripped (right) to better show DyP on the three-fold axis.

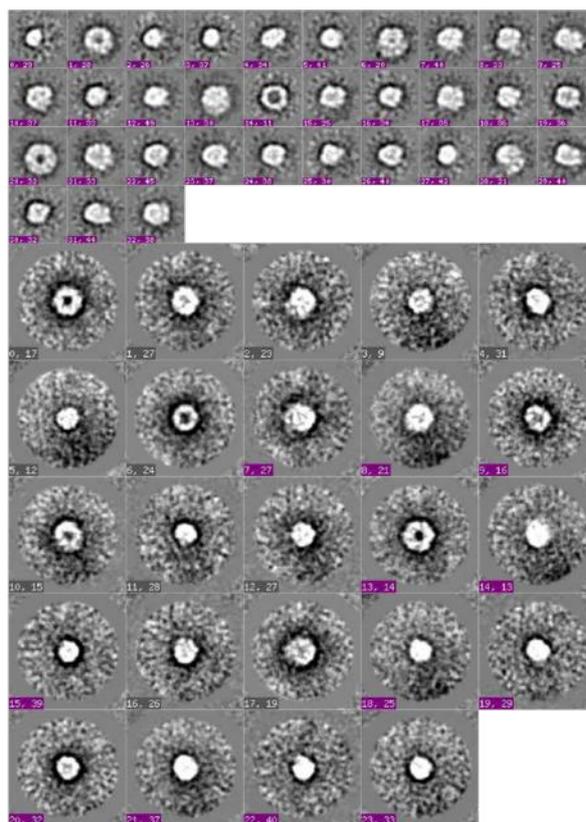
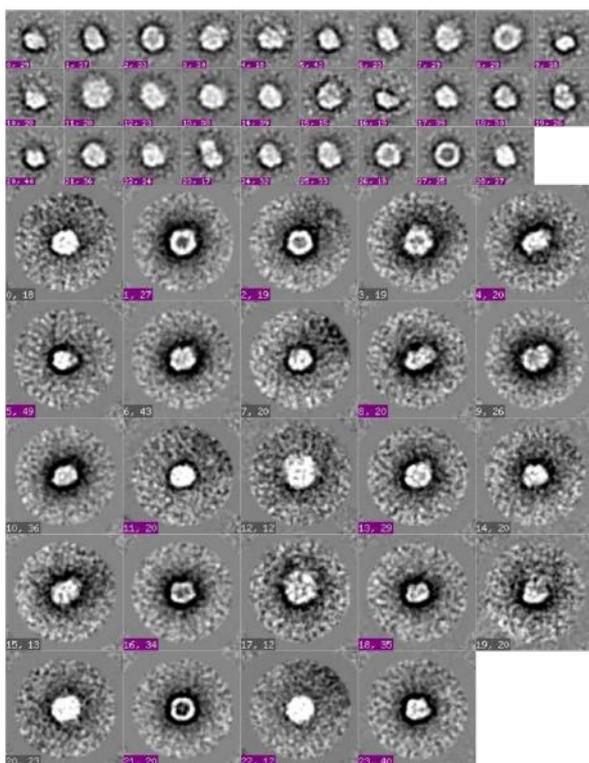
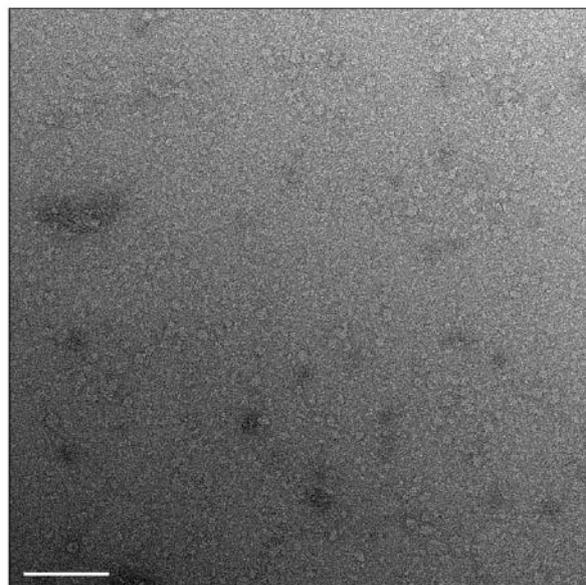
600 **Supplementary Figures**

601

a <40%



b 40–50%



602

603

604

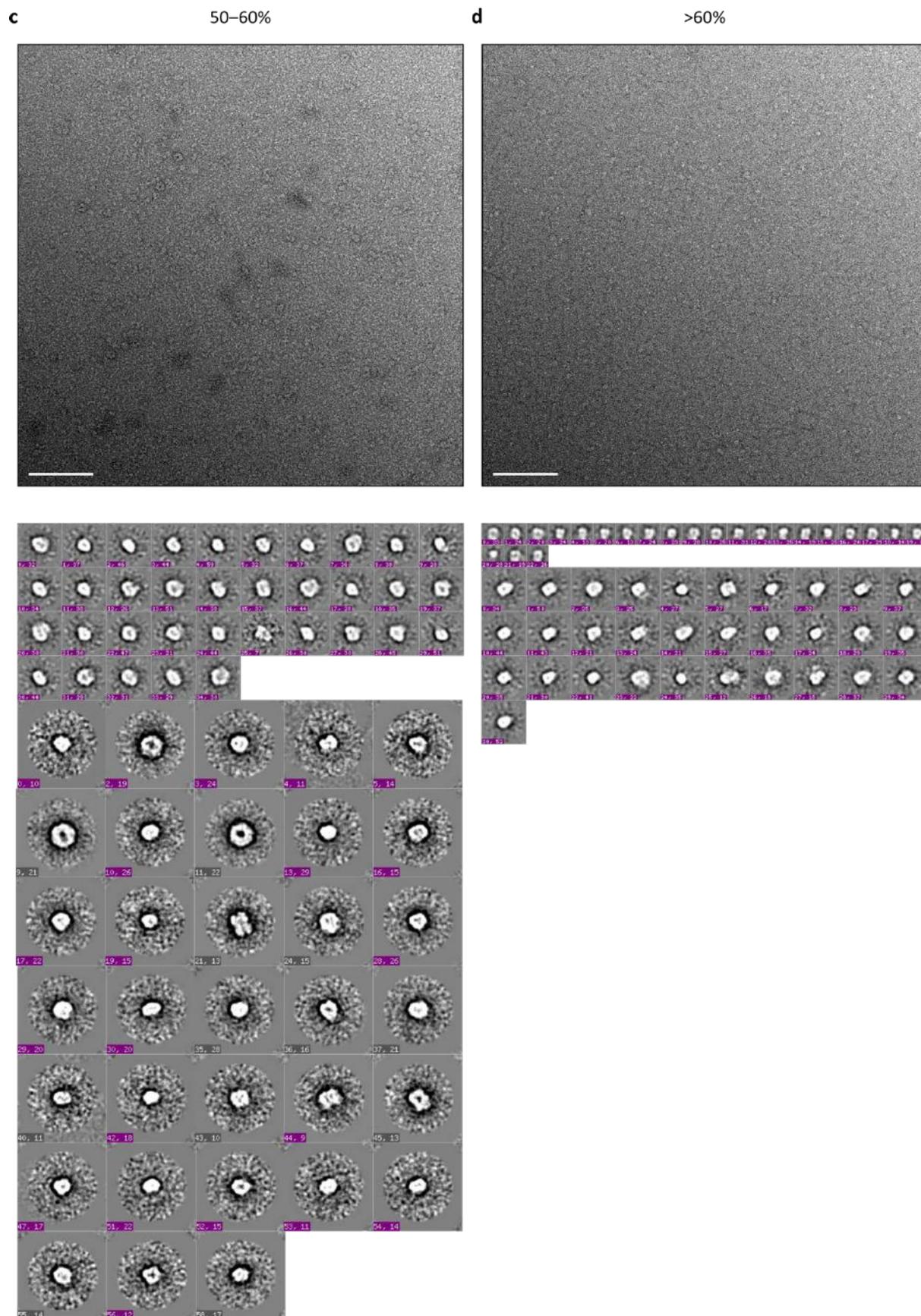
605

606

607

608

609



610
611
612
613

614 **Figure S1 (previous page). Diversity of protein complexes in *M smegmatis*.** Cell lysate was fractionated by a)
615 <40%, b) 40–50%, c) 50–60%, and d) >60% ammonium sulphate cuts (top row). Particles were picked and
616 assigned to class averages using multivariate statistics through the processing pipeline Appion (bottom row)
617 (37) Images were taken at x50,000 magnification at a defocus of 2.00 μm using an F20 Tecnai TEM. Scale bars
618 (white) show 100 nm.

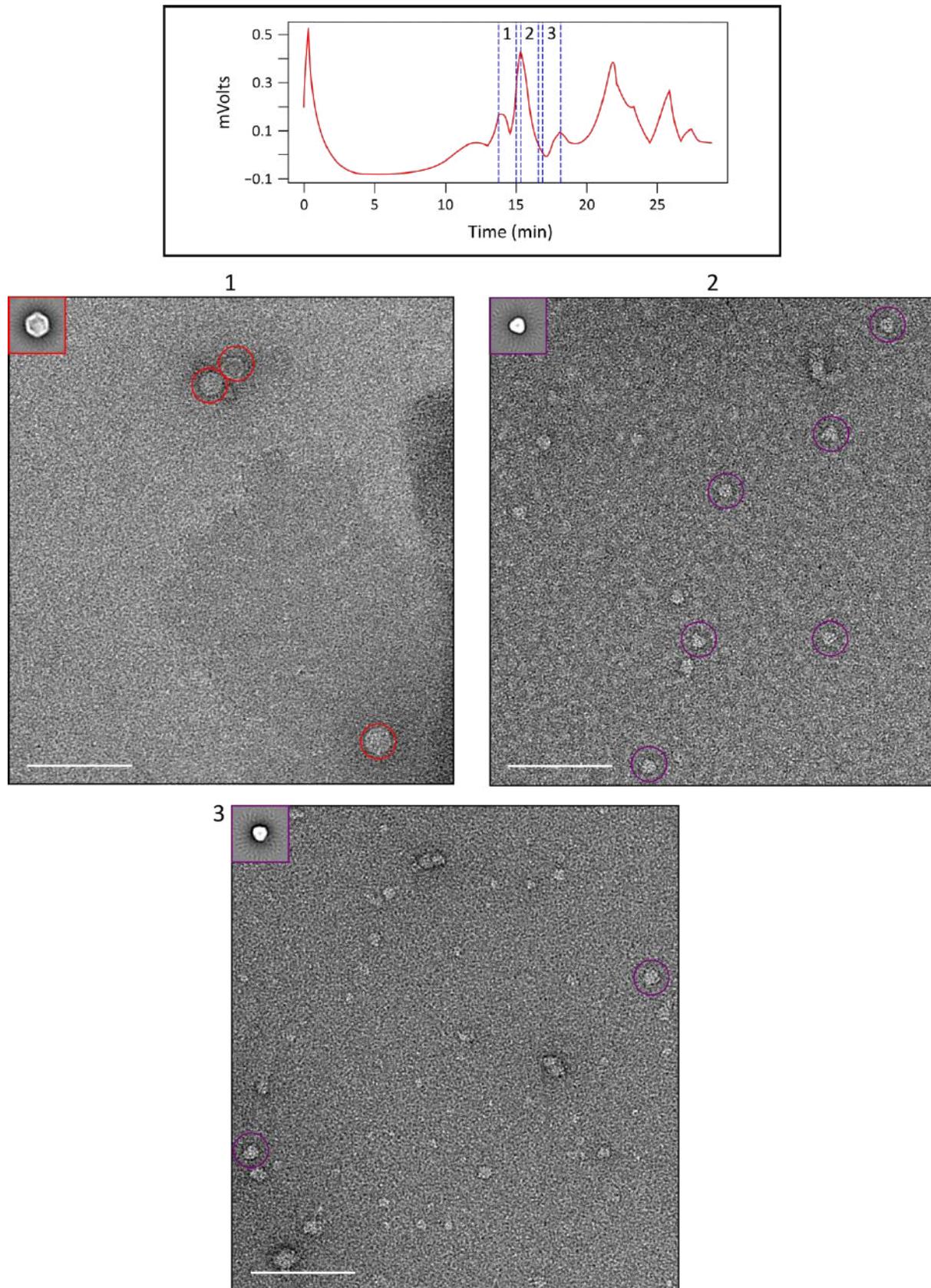
619

620

621

622

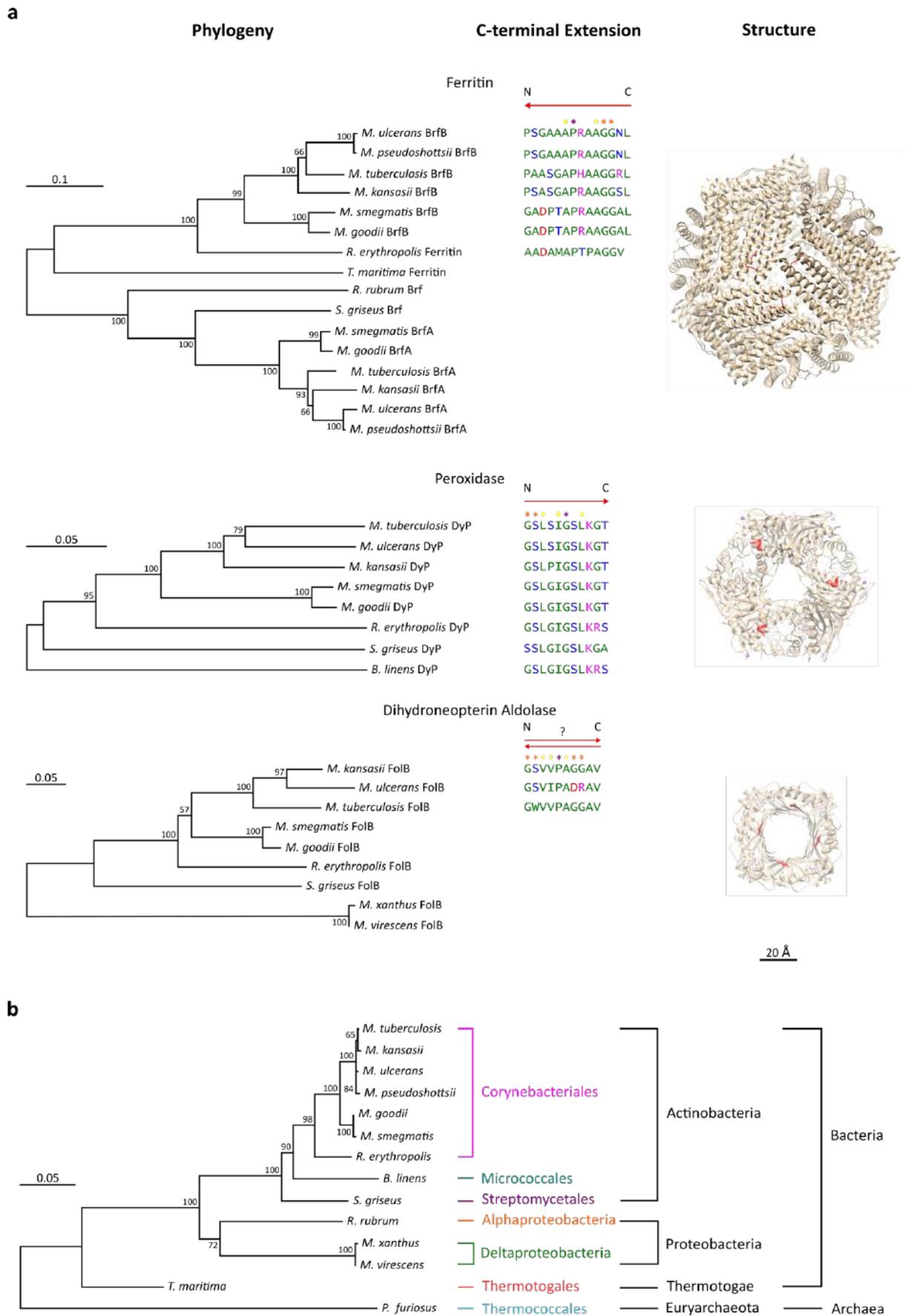
623



624
625
626
627
628
629

Figure S2. Fractionation using gel filtration after anion exchange. Three fractions (1–3) were examined from gel filtration of peak 1 (fractions #15–19) from anion exchange. Fraction 1 (#44–48) contained the presence of Encapsulin (red, circled), while fractions 2 (#49–53) and 3 (#54–58) contained a triangle-shaped average protein complex (purple, circled). Reconstruction of this complex was not pursued due to the presence of preferred

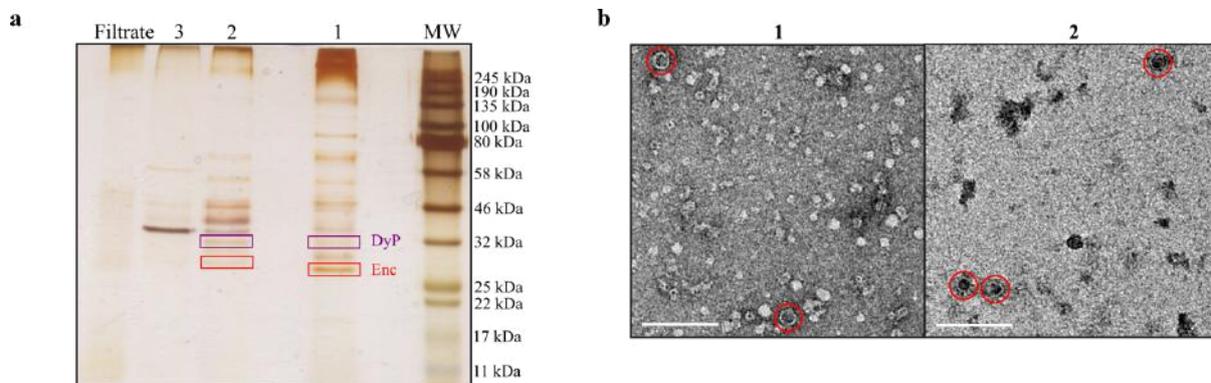
630 orientation. The average for each particle is given in the top left-hand corner. The white scale bar shows 100 nm.
631 Negative stain electron micrographs were taken at a magnification of x50,000 with a defocus of 2.00 μm on an
632 F20 Tecnai TEM.
633



634
635

636 **Figure S3 (previous page). Cargo proteins of encapsulin.** **a)** The phylogeny, C-terminal extension, and structure
637 are given for the three known cargo proteins of *Mtb*. Binding of the cargo protein to the inside of encapsulin is
638 determined by the C-terminal extension, which is dominated by non-polar amino acids (green) with interspersed
639 with mostly conserved polar (blue), positively charged (pink), or negatively charged (red) amino acids. The
640 direction of binding is determined by two N- or C-terminal residues (orange star) while a central residue (purple
641 star) separating two hydrophobic residues (yellow star) aids in positioning (21). The direction of binding for the
642 7,8-dihydroneopterin aldolase cargo is ambiguous. Binding of the C-terminal extension (red) is hypothesised to
643 occur along either the 3-fold or 4-fold axis of the cargo protein. Ferritin cargo protein may also bind along its 2-
644 fold axis (not shown). Note that the C-terminal extension is only visible for *Mtb* ferritin (pdb code 3uno (20)) and
645 was not built into the crystal structure of peroxidase (pdb code 2gvk (46)) and was cleaved from 7,8-
646 dihydroneopterin aldolase (pdb code 1nbu (48)). Also note that the peroxidase shown is from the closest
647 structural homolog, *Bacteroides thetaiotaomicron*. **b)** Phylogenetic relationship between organisms that
648 harbour known and putative encapsulin and cargo proteins, based on 16S rRNA gene sequence. While the
649 peroxidase cargo is found in the Actinobacteria phylum, the ferritin cargo is restricted to the Corneybacteriales
650 order, and the 7,8-dihydroneopterin aldolase cargo is specific to slow-growing Mycobacteria. For the
651 phylogenetic trees, scale bars show amino acid or nucleotide substitutions.

652
653
654



655 **Figure S4. Analysis of cell culture filtrate.** **(a)** Cell culture filtrate was separated by anion exchange which
656 yielded three fractions (1–3). A silver-stained 8–15% gradient SDS-PAGE gel showed that fractions 1 and 2
657 contained the predominant amount of encapsulin. A mass corresponding to that of bacterioferritin B (20 kDa) was
658 not visible. Since bacterioferritin B is present in 24 copies while DyP has 6 copies per biological unit, based on
659 their respective symmetries, and DyP is clearly present in the gel, this suggests that DyP is encapsulated at a
660 greater rate than bacterioferritin B. **(b)** We examined the fractions on an EM micrograph which revealed the
661 presence of encapsulin (red circle). Images were taken at a magnification of x53,000 on the T20 Technai TEM.
662 Scale bars show 100 nm.

663
664
665