bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment revealed by Cryo-EM and hydrogen/deuterium exchange mass spectrometry

Jennifer Ross[1#], Zak McIver[2], Thomas Lambert[1], Cecilia Piergentili[2], Kelly J. Gallagher[1], Jasmine Emma Bird[2], Faye L. Cruickshank[1], Efrain Zarazúa-Arvizu[3], Louise E. Horsfall[3], Kevin J. Waldron[5], Marcus D. Wilson[4], C. Logan Mackay[1], Arnaud Baslé[5], David J. Clarke[1]*, Jon Marles-Wright[2]*

Running title: Hoch-Enc Structure

[1]EaStCHEM School of Chemistry, The University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ.
[2]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU.
[3]School of Biological Sciences, The University of Edinburgh, Edinburgh, EH9 3BF.
[4] Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Kings Buildings, Mayfield Road, Edinburgh, EH9 3JR, UK
[5]Newcastle University Biosciences Institute, Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH.

[#]*Current address: School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD.*

* To whom correspondence should be addressed:
Jon Marles-Wright, +44(0)191 2084855, Jon.marles-wright1@ncl.ac.uk
David J. Clarke, +44(0)131 650 4808, Dave.clarke@ed.ac.uk

Keywords:
Encapsulin, encapsulated ferritin, ferritin, ferroxidase, *Haliangium ochraceum*, cryo-EM, hydrogen-deuterium exchange (HDX), mass spectrometry (MS), metalloprotein, protein structure.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Encapsulins (Enc) are protein nanocompartments which house various cargo enzymes, including a family of decameric ferritin-like proteins. Previously, we elucidated the structure and activity of these ferritin-like proteins (EncFtn) and demonstrated that they must be encapsulated in a nanocompartment for iron storage. Here, we study a recombinant *Haliangium ochraceum* Enc:EncFtn complex using electron cryo-microscopy (Cryo-EM) and hydrogen/deuterium exchange mass spectrometry (HDX-MS) to gain insight into the structural relationship between Enc and EncFtn. An asymmetric single particle reconstruction reveals four EncFtn decamers in a tetrahedral arrangement within the encapsulin nanocompartment. This leads to a symmetry mismatch between the EncFtn cargo and the icosahedral encapsulin shell. The EncFtn decamers are offset from the interior face of the encapsulin shell and are resolved at a much lower overall resolution in the final reconstruction. This flexibility, and the fixed number of EncFtn decamers sequestered within, implies that the loading of the encapsulin nanocompartment is limited by the steric effect of both engaged and free encapsulin localization sequences. Using a combination of focused refinements and HDX-MS, we observed dynamic behavior of the major five-fold pore, and show the pore opening via the movement of the encapsulin A-domain. These data can accelerate efforts to engineer the sequestration of heterologous cargo proteins and to alter the permeability of the encapsulin shell via pore modifications.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Cellular metabolism and reaction pathways can produce toxic by-products which can damage proteins, DNA, and lipids, or become involved in potentially harmful side-reactions. Eukaryotes use membrane-bound organelles, such as lysosomes, to prevent this damage by housing dangerous reactions in chemically privileged environments. In a similar manner, prokaryotes use large protein-based compartments to sequester such reactions and act as barrier from the cytosol (Diekmann and Pereira-Leal, 2013; Kerfeld et al., 2010). Prokaryotes use a variety of compartments such as carboxysomes, which are used for carbon dioxide fixation; and ferritins, for iron oxidation and storage (Andrews, 2010; Chiancone et al., 2004; Yeates et al., 2008).

One nanocompartment utilized by prokaryotes is the encapsulin system (Akita et al., 2007; Prangishvili et al., 2008). Encapsulin (Enc) nanocompartments are large, hollow icosahedral complexes which range in size from 20 nm to 42 nm (Giessen et al., 2019; He et al., 2016; Prangishvili et al., 2008). Encapsulin proteins are structurally related to the viral capsid protein (gp5) of the HK97 bacteriophage and self-assemble from a single monomer into one of three forms: 60 subunits (T = 1 capsid symmetry), 180 subunits (T = 3 capsid symmetry) or 240 subunits (T = 4 symmetry)(Akita et al., 2007; Giessen et al., 2019; McHugh et al., 2014; Prangishvili et al., 2008). Encapsulins share a common feature of housing a cargo enzyme, such as ferritin-like proteins (encapsulated ferritins, EncFtn), iron-mineralizing encapsulin-associated firmicute (IMEF), or dye-decolorizing peroxidases(Giessen et al., 2019; Giessen and Silver, 2017). Cargo enzymes are directed inside the encapsulin nanocompartment by a terminal localization sequence (LS) which binds to the interior face of the encapsulin(Prangishvili et al., 2008; Tamura et al., 2015). Encapsulins and their cargo proteins are found throughout the bacterial and archaeal domains in species inhabiting a range of environmental niches; consequently, the proteins are stable in diverse physical conditions(Akita et al., 2007; He et al., 2019; Rahmanpour and Bugg, 2013; Snijder et al., 2016). For these reasons, the encapsulins have attracted considerable interest for biotechnological applications, through their ability to separate potentially hazardous heterologous reactions from the native cytosol(Jenkins and Lutz, 2021; Lau et al., 2018).

The EncFtn cargo proteins are of particular interest, as they differ from their classical ferritin relatives. Although both proteins oxidize iron using a conserved catalytically active ferroxidase center (FOC), they have remarkably different structural architectures. Classical ferritins oxidize ferrous iron, Fe(II), into a mineral ferric form, Fe(III), which is then stored within a 24-meric 12 nm nanocage (Andrews, 2010; Recalcati et al., 2017). In contrast to this, the EncFtn proteins have an annular structure formed from a pentamer of dimers with the FOC active sites located at a dimer interface (He et al., 2019, 2016). EncFtn oxidizes iron in a similar manner to other ferritins, but due to its open structure, it must be associated with an encapsulin nanocage to act as an iron store(He et al., 2016). Together the encapsulin EncFtn complex (Enc:EncFtn) can perform both the oxidation and storage functions of classical ferritins. However, due to its increased size when compared to classical ferritins, the Enc:EncFtn complex has the potential to house significantly greater quantities of iron, and has been described as an iron megastore(McHugh et al., 2014).

Although there have been several structural studies on encapsulins, a number of key questions remain unanswered. Most notably for the EncFtn containing encapsulin nanocompartments, the structural relationship between the encapsulin shell and the EncFtn cargo protein is unknown. Our previous studies have focused primarily on an

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

encapsulated ferritin system from *Rhodosoprillum rubrum,* where we determined the crystal structure of EncFtn and established its activity as a ferroxidase enzyme(He et al., 2016). Further studies on EncFtn systems revealed a shared structural organization for bacterial and archaeal EncFtn complexes(He et al., 2019). Recombinantly produced encapsulin from *R. rubrum* was found to assemble into a shell with a diameter of 24 nm, consistent with a $T = 1$ icosahedral arrangement. When co-expressed with its respective encapsulated ferritin protein, the encapsulin recruited its cargo protein to the nanocompartment lumen. Previously, we proposed a model for the Enc:EncFtn complex structure based on the unique structure adopted by the EncFtn protein and the orientation of the ferroxidase center(He et al., 2016). In our model, the D5 symmetric decamer of EncFtn is aligned with the pentameric vertices on the encapsulin shell, such that the inner circumference of EncFtn is aligned with the 5-fold pore of the encapsulin nanocompartment. This would result in one EncFtn being positioned at each of the internal five-fold vertices, with a total of 12 EncFtn decamers per encapsulin nanocage.

Herein, we investigate the structure of the Enc:EncFtn nanocompartment from the halophilic bacterium *Haliangium ochraceum,* to gain a better understanding of the complex. Using the complementary structural biology techniques of cryogenic electron microscopy (cryo-EM), native mass spectrometry (MS), and hydrogen-deuterium exchange (HDX) MS, we study the structural relationship between Enc and EncFtn.

We present the first cryo-EM structure of the Enc:EncFtn nanocompartment, which reveals the stoichiometry of cargo loading and structural arrangement of the EncFtn proteins within the encapsulin shell. This model affords insight into the loading capacity and dynamics of the Enc:EncFtn nanocompartment system.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

Recombinant *Haliangium ochraceum* encapsulin complexes form regular nanocompartments that recruit active EncFtn cargoes

In order to gain an understanding of the relationship between encapsulin nanocompartments and their EncFtn cargoes, the *Haliangium ochraceum* Enc:EncFtn nanocompartment was chosen as our model system. This was primarily due to the high yields and ease of purification when compared to the *R. rubrum* homologue.

Constructs for the production of empty and EncFtn-loaded encapsulin nanocompartments were produced for recombinant protein expression in *Escherichia coli*. The protein complexes were purified by heat treatment, followed by anion exchange and size-exclusion chromatography (**Figure 1A and Figure 1 – figure supplement 1Ai, Aii, Bi** and **Bii**).

The molecular masses of the protein constituents of the Empty-Enc and Loaded-Enc assemblies were determined by LC-MS (**Table 1**). MS analysis of the Empty-Enc assembly revealed a single charge state distribution corresponding to a monomer of the encapsulin protein. MS analysis of the Loaded-Enc revealed three charge state distributions present, with deconvoluted masses consistent with the encapsulin protein monomer, a monomer of EncFtn and a dimer of EncFtn (**Figure 1 – Figure supplement 1Aiii** and **Biii**). These results indicate that the Loaded-Enc sample contained both the encapsulin and EncFtn cargo proteins, whilst Empty-Enc has only the encapsulin protein.

The assembly of the purified encapsulin nanocompartments was confirmed through visualization by negative stain transmission electron microscopy (TEM) (**Figure 1B**). Both empty and EncFtn loaded encapsulins assembled into regular nanocompartments, with an average diameter of approximately 21 nm, consistent with other *T*=1 type encapsulins (**Figure 1Aiv** and **1 Biv**)(Prangishvili et al., 2008; Putri et al., 2017; Xiong et al., 2020). The solution behavior of both complexes is consistent with the TEM observations, with both complexes eluting from a size-exclusion column at the same volume, indicating similar hydrodynamic radii (**Figure 1 – Figure Supplement 1Ai** and **Bi).** The micrographs of the EncFtn loaded encapsulin reveal a regular internal density visible within the nanocompartment, suggesting that the EncFtn cargo has been encapsulated in an organized manner.

Ferroxidase assays confirmed the ability of the EncFtn loaded encapsulin nanocompartment to convert Fe(II) to Fe(III) (**Figure 1C**). This result is consistent with our previous observations for the *Rhodospirillum rubrum* Enc:EncFtn encapsulin complex(He et al., 2016). The empty encapsulin, which lacks the EncFtn cargo is enzymatically inactive.

Taken together, these data demonstrate that functionally active EncFtn has been successfully loaded into the encapsulin nanocompartment during expression in the heterologous *E. coli* host.

The cryo-EM structure of the encapsulin EncFtn complex

Motivated by the apparent interior density in the EncFtn loaded encapsulin complex, we performed single particle cryo-EM of the Enc:EncFtn complex. Consistent with previously published X-ray crystallographic and cryo-EM derived encapsulin models, an initial reconstruction was produced with imposed I1 symmetry (**Figure 2, Figure 2 – figure supplement 1**, and **Figure 2 – figure supplement 2**) (Akita et al., 2007; Giessen et al., 2019; Nichols et al., 2020; Prangishvili et al., 2008; Xiong et al., 2020). This resulted in a reconstruction with a global resolution of 2.4 Å as determined by the gold-standard fourier shell correlation at 0.143 (GS-FSC) (**Figure 2 – figure**

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

supplement 1D). The reconstruction displays a *T* = 1 icosahedral arrangement of sixty encapsulin monomers, with clearly resolved secondary structure elements with small pores at, or close to, the 2-, 3-, and 5-fold symmetry axes where protein monomers interact with each other (**Figure 2A**). Regions of the reconstruction around the icosahedral five-fold axes displayed a lower resolution than the other regions of the structure in a local resolution map (**Figure 2B**). This is consistent with observations of local resolution maps for the *Quasibacillus thermotolerans* encapsulin reconstruction(Giessen et al., 2019). The monomer of the encapsulin nanocompartment from the reconstruction displays a HK97-fold typical for encapsulins (**Figure 2C**). The orientation of the E-loop (extended loop) of the HK97 phage-like fold determines the topology of the nanocompartments. The *T* = 1 Family 1 encapsulins have their E-loop shifted away compared to the E-loops of those from the HK97 bacteriophage and other encapsulins (such as, the *T* = 1 Family II Enc from *Synechococcus elongatus*; the *T* = 3 Enc from *Pyrococcus furiosus* and the *T* = 4 from *Quasibacillus thermotolerans*). Loaded-Enc has an E-loop orientation similar to that of the *T* = 1 Family 1 Enc from *Thermotoga maritima* (**Figure 2C**).

With the imposition of I1 symmetry on the reconstruction, the EncFtn cargo is not visible. This suggests that the organization of the EncFtn protein within the encapsulin shell does not conform to icosahedral symmetry and is rotationally averaged through our symmetry-imposed processing (**Figure 2 - figure supplement 1C**).

The encapsulin nanocompartment recruits four EncFtn decamers to its lumen

To gain insight into the structural relationship between the encapsulin shell and its EncFtn cargo protein, a reconstruction was produced with no imposed symmetry averaging (**Figure 3 and Figure 2 – figure supplement 2**). The resolution of the final C1 map is 3.1 Å at the 0.143 FSC threshold (**Figure 3A and Figure 3 - Figure Supplement 1A**). Interestingly, the EM map revealed four distinct densities within the encapsulin nanocompartment lumen, which are consistent in size and shape with four EncFtn decamers (**Figure 3** and **Figure 3 - figure supplement 1B**). This is significantly fewer than the twelve EncFtn decamers in our proposed *T*=1 Enc:EncFtn nanocompartment model(He et al., 2016). Interestingly, despite extensive 3D classification the observed internal density did not improve, suggesting it is flexibly tethered to the inner face of the encapsulin shell.

A local resolution map calculated for the asymmetric reconstruction indicates a degree of flexibility at the pentameric pores compared to the trimeric pores of the encapsulin shell (**Figure 3A**). The interior of the nanocompartment shows a significant falloff in resolution from the inner face of the encapsulin shell to the EncFtn densities. This is consistent with the tethering of the EncFtn to the encapsulin nanocompartment via its localization sequence, but with some degree of conformational freedom of the EncFtn decamer with respect to the encapsulin shell. There is also some fuzzy interior density which is consistent with the untethered localization sequences.

The EncFtn decamers are located approximately 3 nm away from the encapsulin interior wall, which corresponds to the linker region between the main EncFtn domain and the localization sequence on the EncFtn C-terminus (**Figure 3 – figure supplement 1C**). The extended localization sequence of the EncFtn protein acts to offset it from the inner face of the encapsulin shell, an observation consistent with previous reports of the IMEF encapsulin complex from *Q. thermotolerans*(Giessen et al., 2019). Due to the dynamic nature of the EncFtn within the encapsulin it was not possible to trace the path of the localization sequence to its binding site.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The four densities within the encapsulin nanocompartment are discrete, and thus permit the docking of the *H. ochraceum* EncFtn crystal structure in this region (PDB: 5N5F)(He et al., 2019) (**Figure 4B** and **Figure 4 - figure supplement 1C**). The decameric, annular crystal structure fits well into the EncFtn density; although at the observed resolution, it is not possible to fix the rotational alignment around the 5-fold symmetry axis of the EncFtn. Interestingly, the four EncFtn decamers are in a tetrahedral arrangement within the encapsulin nanocompartment, whereas the five-fold axes of the EncFtn complex are aligned with the three-fold tetrahedral axes (**Figure 3B**). This results in a double symmetry mismatch between the icosahedral shell and the EncFtn decamers in the complex.

This symmetry mismatch is particularly interesting, as the pores of the encapsulin nanocompartment allow substrate access to the nanocompartment interior and subsequently EncFtn(Williams et al., 2018). As there is no common symmetry between the pores in the encapsulin shell and the EncFtn decamers, the latter are found in non-equivalent environments in terms of iron availability. Analysis of the relationship between the EncFtn decamers and the inner face of the encapsulin shell reveals two distinct EncFtn environments (**Figure 3 - figure supplement 2**). The first EncFtn environment is shared by two EncFtn decamers and is in line with the five-fold pore of the encapsulin nanocompartment (EncFtn 1 and 2 in **Figure 3 - figure supplement 2**). The shared symmetry of the Enc nanocompartment five-fold pores and of the EncFtn D5 annular structure in these positions is consistent with our previously proposed structural hypothesis for the Enc:EncFtn relationship(He et al., 2016). However, the symmetry-breaking tetrahedral arrangement of the EncFtn decamers creates a second environment shared by the remaining two EncFtn decamers, where they are offset between five-fold and three-fold axes of the icosahedral encapsulin shell (EncFtn 3 and 4 in **Figure 3 - figure supplement 2**).

With the observation that an encapsulin nanocompartment can house four EncFtn decamers, which each have a calculated volume of around 100 nm^3, the 'free' volume within an encapsulin nanocompartment is around 3800 nm^3. This is significantly larger than the classical ferritin lumen, which has interior volume of 270 nm^3. This is in line with our previous biochemical studies, which showed that for our experimental system that the Enc:EncFtn complex can hold approximately 4000 Fe(III) ions, whilst classical ferritin only has a capacity of 500 iron ions(He et al., 2016).

Structural dynamics in the pentameric vertices of the encapsulin shell

To further investigate the apparent conformational flexibility of the encapsulin shell at the pentameric vertices in the reconstructions, we performed symmetry expansion on the I1 refined particle set, followed by masked 3D-classification without alignment centered on the vertex. A number of distinct conformations were revealed, and the most extreme of these were subjected to 3D refinement with local searches (**Figure 4**, **Figure 4 – figure supplement 1**, and **Figure 2 – figure supplement 2**). This resulted in an 'open' pentamer conformation of 2.4 Å resolution and a 'closed' conformation of 2.3 Å, allowing for fitting of residue side chains and the modelling of some water molecules (**Figure 4 – figure supplement 2**).

The 'open' conformation, has a five-fold pore with an aperture diameter of approximately 24 Å, while in the 'closed' conformation the aperture is reduced to 9 Å diameter. To understand the structural changes taking place in the transition between these conformations, an atomic model of the encapsulin protein was refined against both maps (**Table 6**). The two models show a significant movement in the A-domain, with a pivoting around the hinge points connecting this domain to the P-domain. This

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

opens the pore like an iris (**Figure 4**). In the open conformation, the pore loop region (residues 182 - 189) is not well defined in the density; however, it is tightly locked in the closed conformation, with Asp186 forming the outer boundary of the pore and Tyr188 and Lys192 forming the inner bounds (**Figure 4 - figure supplement 3**). The tyrosine is well conserved among the family 1 *T* = 1 encapsulins, while the lysine is substituted for a glutamine in the *R. rubrum* encapsulin (**Figure 4 – figure supplement 4**). The family 2 *T* = 1 encapsulin from *Synechococcus elongatus* has a five-residue sequence insertion in this region, which forms an extended linker between secondary structure elements, rather than a distinct loop within the pore.

In the *H. ochraceum* encapsulin the five-fold pore has a negative charge on the exterior of the encapsulin shell and positive charge on the interior in both the open and closed conformations (**Figure 4 – figure supplement 4** and **Figure 4 – figure supplement 5**). The closed conformation is consistent with observations from the crystal structure of the *T. maritima* encapsulin (Prangishvili et al., 2008) and high resolution cryo-EM structures of other encapsulins(Giessen et al., 2019; Nichols et al., 2020). However, this is the first time that an 'open' pore-conformation has been observed in an encapsulin protein. This observation has important implications for efforts to engineer the pores of encapsulin nanocages. Where early efforts to widen the five-fold pores have demonstrated an increase in mass-transport of model substrates across the encapsulin shell(Williams et al., 2018), more recent efforts to modify the pores have shown that the shell does not act as a strong barrier to the passage of the small lanthanide substrates tested. Our results provide an explanation for these observations, where a dynamic and flexible pore would not act as a barrier to the passage of small ligands, such as divalent cations, across the shell. They would also be able to accommodate a wide range of sizes of potential ligands for engineered nanocages.

Additionally, these focused refinements of the pentameric subunits allowed us to build and sequence the '$_{117}$GSLGIGSLR$_{125}$' peptide from the EncFtn protein (**Figure 5**). This region of the localization sequence forms a network of hydrophobic interactions with the inner face of the P-domain of a single encapsulin monomer, with further stabilization by a number of water-mediated backbone contacts. The core GxLGIxxL motif found in this region of the localization sequence is conserved between the *H. ochraceum* EncFtn and other proteins in the family and is observed in the crystal structure of the *T. maritima* encapsulin(Prangishvili et al., 2008).

Dynamics of the 5-fold encapsulin pore through hydrogen/deuterium exchange mass spectrometry.

To further investigate the dynamic nature of the five-fold pore of the encapsulin shell and the docking of the EncFtn localization sequence to the interior of the nanocompartment, hydrogen/deuterium exchange mass spectrometry (HDX-MS) was performed. The extent of backbone-amide hydrogen exchange was determined at seven time points (0 seconds, 10 seconds, 30 seconds, 5 minutes, 30 minutes, 4 hours, and 24 hours) for both Empty-Enc and Loaded-Enc. By calculating the rate of hydrogen exchange throughout the protein, regions that differ in solvent exposure and/or dynamics can be detected.

HDX-MS analysis of the encapsulin nanocompartment resulted in 40 pepsin peptides, which constituted a protein sequence coverage of 85%, with peptide redundancy of 2.28 (**Table 7**, **Table 8** and **Figure 6 - figure supplement 1**). The encapsulin nanocompartment displayed variable exchange rates throughout the protein sequence and regions of the protein displaying elevated HD exchange rates

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

were clearly evident. Overlaying these local HD exchange rates onto the reconstruction of the encapsulin structure revealed that the regions of high exchange were located around pentameric vertices. (**Figure 6**). This was most notable with the peptide spanning the region between amino acids 180-196, which includes the five-fold pore loop (**Figure 6 - figure supplement 2**). In contrast, lower rates of HDX are observed at the 2-fold interface and the potential 3-fold pore (**Figure 6**). These findings are in agreement with our cryo-EM structural analyses and support the proposed conformational flexibility at the 5-fold pore.

Comparison of the HD exchange rates of Empty-Enc and Loaded-Enc revealed similar exchange profiles throughout the encapsulin protein sequence, suggesting that cargo loading has little effect on the overall architecture and dynamics of the assembled nanocompartment shell (**Table 7**, **Table 8**, and **Figure 6 - figure supplement 3**). However, after prolonged exchange times (4 hours), the Loaded-Enc exhibited areas with a modest reduction in exchange when compared to the Empty-Enc. Notably, several peptides in the N-terminal region displayed reduced exchange rates in Loaded-Enc; for example, the peptide covering amino acids 21-37 displayed almost twelve percent reduction. Mapping the position of this region onto our encapsulin reconstruction highlights that this peptide is located on the interior face of the nanocompartment and included the proposed binding site for the localization sequence of EncFtn (**Figure 5**). In our model, the maximum engagement of the localization sequence binding sites would be only one third of the potential sixty sites, with only twenty localization sequences from the EncFtn oriented towards the inner face of the encapsulin shell. A reduction in exchange across in this region may be a consequence of shielding by the extended EncFtn localization sequence.

Discussion

Our cryo-EM reconstruction of an Enc:EncFtn nanocompartment complex reveals key areas of divergence from a true icosahedral complex with important functional consequences. The asymmetric reconstruction showed that the encapsulin nanocompartment sequesters four decamers of EncFtn within its lumen. In our recombinant system, with EncFtn produced in excess, this likely represents a maximum loading capacity for the Enc:EncFtn nanocompartments. The symmetry breaking tetrameric arrangement of the EncFtn decamers within the encapsulin shell leads to two distinct environments for EncFtn, with two decamers aligned at the five-fold symmetry axes, and the remaining two residing between three- and five-fold axes (**Figure 3 – Figure Supplement 2**). The consistent observation of a gap between the encapsulin shell and encapsulin cargo proteins implies that the loading and capacity of encapsulins is clearly limited by the steric effect of the offset of the cargo protein from the shell, and from the unengaged localization sequences at the core of the nanocompartment. This has implications for efforts to target heterologous proteins to the encapsulin nanocage; effectively setting a limit on the volume of protein that can be accommodated within, which is much lower than the total volume of the lumen of the nanocage.

These observations have functional implications for the oxidation and storage of iron within the Enc:EncFtn nanocompartment. The EncFtn decamers are in non-equivalent positions, and thus have different relationships to the pores of the encapsulin shell. Therefore, if the pores limit the diffusion of substrates, the EncFtn decamers would be subjected to different chemical environments. Furthermore, both

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the engaged and unengaged localization sequences present a 'soft' steric barrier to the diffusion of substrates. It is notable that the ferroxidase activity of the Enc:EncFtn complex is significantly higher than the isolated EncFtn protein. While it is not possible to make mechanistic conclusions from our model, the complex interactions with the components of the encapsulin nanocompartment clearly enhance the iron oxidation activity of the EncFtn protein.

Our data were collected on apo-Enc:EncFtn complexes, and thus is it not possible to infer the nature of the iron mineralization pathway within the encapsulin nanocage. Further, careful work must be performed to titrate iron into the complex prior to structural analysis to gain insight into the flow of metal ions from the exterior to the interior of the encapsulin and to determine if metalation influences the conformational flexibility of the EncFtn within the encapsulin nanocage, as we have demonstrated for isolated EncFtn proteins(He et al., 2019, 2016; Piergentili et al., 2020). Finally, the nature of the iron mineral and its localization within the encapsulin nanocage is still to be determined.

These cryo-EM and HDX-MS data illustrating a highly dynamic five-fold pore in the encapsulin shell have major implications for efforts to engineer recombinant encapsulins for improved access for both native and non-native substrates. The limitations of previously published studies where the five-fold pore is modified for altered substrate access can be explained by a highly dynamic pore structure that is not particularly discriminatory for small molecules. Our work suggests new hypotheses for engineering pore selectivity, through modifications to the hinge regions between the P- and A-domains, which are responsible for the opening of the pore.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Materials and Methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
strain, strain background (*Escherichia coli*)	BL21(DE3)	Cambridge Biosciences	60401	chemically competent
recombinant DNA reagent	CIDAR MoClo Kit	Addgene	Kit # 1000000059	vectors for protein expression in *E. coli*
sequence-based reagent	gBlocks used in this study	Integrated DNA Technologies (IDT)		
commercial assay or kit	Pierce™ BCA Assay kit	Thermo Fisher Scientific	23225	
chemical compound, drug	Lysozyme	Sigma-Aldrich	L6876	
chemical compound, drug	Benzonase	Sigma-Aldrich	E1014	
chemical compound, drug	Tris	Roche	1078976001	Used for casting SDS-PAGE gels
chemical compound, drug	Glycerol	VWR	24387.326	Used for casting SDS-PAGE gels
chemical compound, drug	Sodium dodecyl sulfate	Fisher bioreagents	BP2436-1	Used for casting SDS-PAGE gels

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

chemical compound, drug	Ammonium persulfate	Sigma-Aldrich	A3678-100G	Used for casting SDS-PAGE gels
chemical compound, drug	TEMED	Severn Biotech	20-3000-01	Used for casting SDS-PAGE gels
chemical compound, drug	Bis-Acrylamide 30%	Severn Biotech	20-2100-10	Used for casting SDS-PAGE gels
chemical compound, drug	Uranyl acetate	Agar Scientific	AGR1260A	
chemical compound, drug	$FeSO_4.7H_2O$; Fe(II)	Sigma-Aldrich	1270355	
chemical compound, drug	Deuterium oxide	Sigma Aldrich	7789-20-0	
chemical compound, drug	Potassium phosphate monobasic	Sigma Aldrich	7778-77-0	
chemical compound, drug	Potassium phosphate dibasic	Sigma Aldrich	7758-11-4	
chemical compound, drug	Formic acid 98 - 100% (HPLC grade)	Sigma Aldrich	64-18-6	
software, algorithm	Prism 9	GraphPad		
software, algorithm	Origin 2016	OriginLab		

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

software, algorithm	Fiji	Schindelin et al., 2012		
software, algorithm	MassLynx	Waters	Version 4.1	
software, algorithm	UCSF Chimera	Pettersen et al., 2004		
software, algorithm	UCSF ChimeraX	Goddard et al., 2018		
software, algorithm	Relion	Scheres, 2012	Version 3.1	
software, algorithm	MotionCor2	Zheng et al., 2017		
software, algorithm	CTFFIND4	Rohou and Grigorieff, 2015		
software, algorithm	Phaser	McCoy et al., 2007(McCoy et al., 2007)		
software, algorithm	Phenix.autobuild	Adams et al., 2019(Liebschner et al., 2019)		
software, algorithm	Phenix.refine	Afonine et al., 2018(Afonine et al., 2018b)		
software, algorithm	Coot	Emsley et al., 2010(Emsley et al., 2010)		
software, algorithm	Proteinlynx Global Server 3.0.3	Waters		
software, algorithm	Dynamx 3.0	Waters		

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

other	200 mesh copper grids with formvar/carbon support	TAAB	F077/025	
other	200 mesh gold grids with 2/2 holey carbon	Quantifoil (purchased from Agar Scientific)	AGS173-8	
other	HiTrap Q FF anion; anion exchange column	Cytiva		1 mL
other	HiPrep™ 16/60 Sephacryl® S-400 HR	Cytiva		

Cloning of encapsulin expression constructs

The *Haliangium ochraceum* encapsulin and encapsulated ferritin protein expression constructs were based on the Hoch_3836 and Hoch_3837 genes and were codon optimized for expression in *Escherichia coli* and synthesized as CIDAR MoClo compatible gBlocks by Integrated DNA Technologies (IDT) (**Table 2**). The gBlocks were assembled into a Level 0 CIDAR MoClo storage vector(Iverson et al., 2015), DVA_CD, for subsequent use. The coding sequences for the encapsulin and the EncFtn were assembled into expression cassettes in the level 1 backbones DVK_AE and DVK_EF respectively, each with T7 promoter and transcription terminator parts. The resulting expression cassettes were then combined into the DVA_AF backbone to produce a co-expression plasmid. All assembled plasmids were sequence verified by Sanger sequencing by Eurofins Genomics. The protein sequences for each construct are listed in **Table 3**.

Protein expression

The Empty-Enc and Loaded-Enc expression plasmids were transformed into *E. coli* BL21(DE3) cells and grown overnight at 37 °C on LB-agar plates containing appropriate selection antibiotics (kanamycin for Empty-Enc and ampicillin for Loaded-Enc). A single colony of cells was added to 1 L of autoinduction media(Studier, 2005) supplemented (**Table 4**) with appropriate antibiotic and grown for 38 hours at 37 °C with shaking at 200 rpm. Cells were harvested by centrifugation at 12,000 ´ *g*.

Encapsulin nanocompartment purification

E. coli cell pellets expressing the empty-Enc and loaded-Enc constructs were resuspended in 10 × v/w of lysis buffer (20 mM HEPES, pH 8; 2 mM $MgCl_2$; 1 mg/ml lysozyme, and benzonase, 12.5 – 25 units/mL). Cells were lysed by sonication whilst on ice; sonication was carried out in six 1-minute cycles (30 seconds sonication, 30 seconds rest). The lysate was clarified by centrifugation at 20,000 ´ g for 1 hour, 4 °C.

The supernatant from cell lysis was heated to 85 °C for 10 minutes in a water bath and transferred to a 4 °C ice bath for 10 minutes. The supernatant was then collected after centrifugation at 10,000 g for 1 hour.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Anion exchange chromatography of the clarified supernatant was performed using a 1 mL HiTrap Q Sepharose FF column from Cytiva on an ÄKTA™ start. The column and ÄKTA™ start system were equilibrated with QA buffer (20 mM HEPES, pH 8.0) and the protein sample was loaded. Unbound proteins were removed by washing with QA buffer. Bound proteins were eluted by QB buffer (20 mM HEPES, pH 8.0, 1 M NaCl) over a linear gradient of 0-100% QB over 15 column volumes. Flow-through fractions containing the sample were subjected to SDS-PAGE to identify those containing the protein of interest. These fractions were pooled and concentrated using centrifugal concentrators with a 30 kDa nominal molecular weight cut off (Vivaspin).

Pooled and concentrated samples from the anion exchange step were loaded on a gel filtration column (Sephacryl 400, Cytiva) equilibrated with SEC buffer (20 mM HEPES, pH 8.0, 150 mM NaCl). Fractions eluting from the column containing the desired protein, as identified by SDS-PAGE were pooled and concentrated as above. Protein aliquots were flash cooled in liquid nitrogen and stored at -80 °C (**Figure 1 – figure supplement 1Ai**, **Aii**, **Bi and Bii**). (He et al., 2019, 2016).

Negative stain TEM

Purified encapsulin nanocompartments were initially imaged by negative stain TEM. Continuous carbon/formvar coated copper grids (200 mesh) were glow-discharged for 30 seconds using a Pelco glow discharge system. 5 µL Enc was pipetted onto the glow-discharged grids and excess liquid was removed after 30 seconds with Whatman filter paper (grade 1, diameter 24.0 cm). The grids were washed with distilled water three times, followed by staining with 2 % uranyl acetate for 5 seconds. Grids were left to air dry and then imaged with a JEOL JEM-1400 transmission electron microscope. Images were collected with a Gatan CCD OneView camera and analyzed using FIJI(Schindelin et al., 2012).

Ferroxidase activity assay

The enzymatic activity of Empty-Enc and Loaded-Enc were assessed by ferroxidase assay, as previously described (Piergentili, 2020). Fe(II) samples were prepared by dissolving $FeSO_4.7H_2O$ in HCl 0.1 % (v/v) under anaerobic conditions. Protein samples were diluted anaerobically in Buffer GF (20 mM HEPES, pH 8.0, 150 mM NaCl) to a final encapsulin monomer concentration of 9 µM to allow comparison between experiments.

Iron and protein aliquots were added aerobically to a quartz cuvette (Hellma) resulting in a final concentration of 100 µM iron and 15 µM (Loaded-Enc), or 9 µM (Empty-Enc). The cuvette was placed in a UV-visible spectrophotometer (PerkinElmer Lambda 35) and the reaction sample was incubated at 21 °C for 50 s to stabilise. Absorbance at 315 nm was then recorded every second for 1450 s using the Time-Drive software. A control experiment was conducted by monitoring the background oxidation by atmospheric oxygen of 100 µM $FeSO_4*7H_2O$ in the absence of the enzyme. Each experiment was carried out in three or more technical replicates, with replicate means and standard deviations calculated on the time zero-subtracted progress curves.

Liquid Chromatography Mass Spectrometry

LC-MS experiments were performed on a Synapt G2 Q-ToF instrument (Waters Corp., Manchester, UK) and an Acquity UPLC equipped with a reverse phase C4 Aeris Widepore 50 × 2.1 mm HPLC column (Phenomenex, CA, USA). Mobile phases of; A= water + 0.1% formic acid, and B=acetonitrile + 0.1% formic acid were used on a ten-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

minute gradient from 5% B to 95% B. Samples were analysed at ~2 μM, and data analysis was performed using MassLynx v4.1 and MaxEnt deconvolution.

Cryo-EM data collection and analysis

Sample vitrification

Holey grids (gold, 200 mesh, r 2/2 by Quantifoil) were glow-discharged for 30 seconds using a Pelco glow discharge system. The grids were then mounted into a FEI vitrobot and 4 μL of encapsulin sample (3 mg/mL) was applied. Grids were then blotted (100% humidity, 8 °C, blot force -5, wait time 10 seconds and blot time of 3 seconds) with Whatman filter paper (grade 1) and flash cooled in liquid ethane, cooled with liquid nitrogen.

Cryo-EM data collection

Cryo-EM grid screening was performed on a FEI F20 microscope equipped with a FEG electron source (200 kV) and a TVIPS F816 CMOS detector at the University of Edinburgh. The dataset used for single particle reconstruction was obtained at eBIC on a FEI Titan Krios microscope equipped with Gatan K3 camera (data collection settings are shown in **Table 4**). Alignments, grid transfer and imaging set up was performed by the eBIC local contact Dr Yun Song.

Single particle reconstruction

All processing steps were performed with the Relion 3.1 software package (Nakane et al., 2018). Super-resolution movies were binned (2 × 2 pixels for I1 reconstructions; 3 × 3 for C1 reconstructions) and motion corrected using MotionCor2(Zheng et al., 2017). Defocus values of imaged (summed movies) were determined by CTFFIND4 (Rohou and Grigorieff, 2015) and those with poor CTF fits, or bad ice, were manually discarded. A template for autopicking was created using 2D classes from manually picked particles. Autopicked particles were extracted (using box sizes of 576 pixels (376 Å) and 512 pixels (501 Å) for icosahedral and C1 processing respectively) and subjected to three rounds of 2D classification to remove bad particles. An initial 3D model was created from particles selected from 2D classes. 3D classes were generated both with and without icosahedral symmetry imposed ('I1' and 'C1' symmetry). The best class from each was taken forward for 3D refinement and then CTF refinement followed by further rounds of 3D refinement and Bayesian polishing. After a final round of 3D refinement, postprocessing was performed using a soft spherical mask. Local resolution estimation was performed in Relion3.1. The data processing and refinement pipeline is shown in **Figure 2 - figure supplement 2** with data processing and refinement statistics in **Table 5**.

Motivated by the apparent flexibility of the 5-fold pores of the encapsulin shell, the 5-fold pore pentamer was subjected to symmetry expansion and focused classification. The I1 particle set was expanded using the sym_expand job in Relion3.1 and masked 3D classification without alignment was performed focused on the five-fold symmetry axis. Of the five classes produced the two most highly populated and distinct were taken forward for masked 3D refinement with local searches only, these two classes represent the 'open' and 'closed' conformation of the five-fold pore. The mask used in these steps was produced in Chimera using the molmap command from a docked model of a pentamer of the encapsulin protein. To ensure the box-size and pixel-size of the mask were correct, they were resampled onto the icosahedral map using the vop resample command.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Model building and refinement

An initial homology model of the encapsulin nanocompartment monomer was generated using Phyre 2.0(Kelly et al., 2015) based on the *T. maritima* structure. This was docked into the open and closed maps using ChimeraX(Pettersen et al., 2021) and expanded to a full pentamer model. The model was then fit to the map through an iterative process of automated model refinement with phenix real-space refinement(Afonine et al., 2018b) and manual model building in Coot(Emsley et al., 2010), waters were added using phenix.douse and validated in in Coot. The resulting models and maps were validated using Molprobity(Chen et al., 2010), phenix.mtriage(Afonine et al., 2018a), and EM ringer(Barad et al., 2015) (**Table 6**). Models and maps were visualized using ChimeraX.

Encapsulin sequence analysis

Encapsulin sequences were obtained from the Kyoto Encyclopedia of Genes and Genomes (www.kegg.jp). Sequence alignments were performed using Clustal Omega(Sievers and Higgins, 2014) and visualized using ESPript(Gouet et al., 2003).

Hydrogen/Deuterium Exchange Mass Spectrometry

Hydrogen/deuterium exchange mass spectrometry (HDX-MS) was performed on a Synapt G2 MS system coupled to an ACQUITY UPLC M-Class UPLC with the HDX manager module (Waters Corporation, Manchester, UK)(Masson et al., 2019). For improved reliability and precision, a custom-built Leap automated platform was zutilized in all sample preparation and injections. Prior to HDX-MS analysis, three buffer solutions were prepared – (i) equilibration buffer (4.7 mM K_2HPO_4, 0.3 mM KH_2PO_4 in H_2O), adjusted to pH 8.0 with formic acid; (ii) labelling buffer (4.7 mM K_2HPO_4, 0.3 mM KH_2PO_4 in D_2O), adjusted to pH 8.0 with DCl and quench buffer (50 mM K_2HPO_4, 50 mM KH_2PO_4 in H_2O) adjusted to pH 2.3 with formic acid. Protein samples were diluted in equilibration buffer to final stock concentration of 42 mM. The time course experiments consisted of 7 timepoints: T0 (0 minute; undeuterated control), T1 (20 seconds), T2 (30 seconds), T3 (2 minutes), T4 (5 minutes) and T5 (30 minutes) T6 (4 hours) and T7 (24 hours) with each timepoint being performed in triplicate. Sample preparation consisted of 5 µL protein solution, 57 µL equilibrium buffer (T0) or labelling buffer (T1-7). The final concentration of deuterium during the labelling step was 91.2 %. Exchange was allowed to proceed at 4 °C. To arrest the exchange reaction, 50 µL of quench buffer was added to this initial solution just prior to sample injection.

After injection, samples underwent proteolytic digestion on a 2.1 x 30 mm Waters Enzymate BEH pepsin column for 3 minutes at 200 µL/min. After digestion, the peptide digest was loaded on to a 2.1 x 5.0 mm Acquity BEH C18 VanGuard 1.7µm C18 Trapping column to pre-concentrate the sample for 3 minutes at 200 µL/min. Following trapping, the digests were separated through a 2.1 x 5.0 mm Acquity BEH 1.7 µm analytical column prior to MS/MS (MS^e) analysis via the Water Synapt G2 MS system running MassLynx v4.1 software (Waters Corporation, Manchester, UK). The separation gradient was 5-95 % acetonitrile with 0.1 % formic acid over 12 minutes at 40 µL/min. Both the trapping and LC separation were performed at 1 °C to minimize back exchange. Post-processingwas performed using Proteinlynx Global Server 3.0.3 and Dynamx 3.0 software to determine the average deuterium uptake for each peptide at each time point. For comparative analyses, the relative fractional uptake was

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

599 determined by dividing the observed deuterium uptake by the number of available
600 amide exchangers on each peptide.

Tables

Table 1: Protein masses and assignments obtained by LC-MS

Sample	Protein in Sample	Observed Average Mass (Da)	Assignment	Theoretical Average Mass (Da)
Empty-Enc	Encapsulin from *H. ochraceum*	28969.67 ± 0.23	Hoch-Enc monomer	28968.84*
Loaded-Enc	Encapsulin from *H. ochraceum*	28813.24 ± 0.89	Hoch-Enc monomer	28812.66*
	EncFtn from *H. ochraceum*	14667.42 ± 0.22	Hoch-EncFtn monomer without Met	14668.20
		29334.50 ± 0.64	Hoch-EncFtn dimer without Met	29336.40

The starting methionine residue is not always retained, and this has been indicated in "Assignment". Errors generated by MassLynx v4.1.

* The difference in theoretical mass for the two *H. ochraceum* encapsulin proteins is due to an additional C-terminal arginine residue in the Empty-Enc construct which is a cloning artefact from the MoClo kit used.

Table 2: gBlocks used in this study

Construct	Sequence
Empty-Enc	GGCGAAGACAT**AATG**GATCTGCTGAAACGTCATCTGGCACCGATTGTTCCGGATGCATGGTCAGCAATTGATGAAGAAGCCAAAGAAATTTTTCA GGGCCATCTGGCAGGTCGTAAACTGGTTGATTTTCGTGGTCCGTTTGGTTGGGAATATGCAGCAGTTAATACCGGTGAACTGCGTCCGATTGATG ATACACCGGAAGATGTTGATATGAAACTGCGTCAGGTTCAGCCGCTGGCCGAAGTTCGTGTGCCGTTTACCCTGGATGTTACCGAACTGGATAGC GTTGCACGTGGTGCAACCAATCCGGATCTGGATGATGTTGCCCGTGCAGCAGAACGTATGGTTGAAGCAGAAGATAGCGCAATTTTTCATGGTTG GGCACAGGCAGGTATTAAAGGTATTGTTGATAGCACACCGCATGAAGCACTGGCAGTTGCAAGCGTTAGCGATTTTCCGCGTGCAGTTCTGAGCG CAGCAGATACACTGCGTAAAGCCGGTGTTACCGGTCCGTATGCACTGGTTCTGGGTCCGAAAGCCTATGATGACCTGTTTGCAGCAACCCAGGAT GGTTATCCGGTTGCAAAACAGGTGCAGCGTCTGGTTGTTGATGGTCCGCTGGTTCGTGCAAATGCCCTGGCAGGCGCACTGGTTATGAGCATGCG TGGTGGTGATTATGAACTGACCGTTGGTCAGGATCTGAGCATTGGTTATGCATTTCATGATCGTAGCAAAGTGGAACTGTTTGTGGCAGAAAGTT TTACCTTTCGTGTTCTGGAACCGGGTGCAGCCGTTCATCTGCGTTATGCATAA**AGGT**ATGTCTTCGTC
Loaded-Enc	Encapsulin: as above
	EncFtn: GGCGAAGACAT**AATG**AGCAGCGAACAGCTGCATGAACCGGCAGAACTGCTGAGCGAAGAAACCAAAAACATGCATCGTGCACTGGTTACCCTGATTGAAG AACTGGAAGCAGTTGATTGGTATCAGCAGCGTGCAGATGCCTGTAGCGAACCGGGTCTGCATGATGTTCTGATTCATAACAAAAACGAAGAGGTGGAACA TGCAATGATGACCCTGGAATGGATTCGTCGTCGTAGTCCGGTTTTTGATGCACACATGCGTACCTACCTGTTTACCGAACGTCCGATTCTGGAATTAGAA GAAGAAGATACCGGTAGCAGCAGCAGCGTTGCGGCAAGCCCGACCAGCGCACCGAGTCATGGTAGCTTAGGTATTGGTAGCCTGCGTCAAGAAGGTAAAG AAGATTAA**AGGT**ATGTCTTCCGG

BbsI restriction enzyme recognition sites are underlined, and overhang sequences are shown in bold.

Table 3: Protein constructs used in this study.

Protein Construct	Protein	Sequence	Amino acids	Average Molecular weight (Da)	pI	Extinction coefficient (M^{-1} cm^{-1})
Empty-Enc	Encapsulin shell protein from *H. ochraceum*	MDLLKRHLAPIVPDAWSAIDEEAKEIFQGHLAGRKLVDFRG PFGWEYAAVNTGELRPIDDTPEDVDMKLRQVQPLAEVRVPF TLDVTELDSVARGATNPDLDDVARAAERMVEAEDSAIFHGW AQAGIKGIVDSTPHEALAVASVSDFPRAVLSAADTLRKAGV TGPYALVLGPKAYDDLFAATQDGYPVAKQVQRLVVDGPLVR ANALAGALVMSMRGGDYELTVGQDLSIGYAFHDRSKVELFV AESFTFRVLEPGAAVHLRYA**R**	267	28968.84	4.88	26930
Loaded-Enc	Encapsulin shell protein from *H. ochraceum*	MDLLKRHLAPIVPDAWSAIDEEAKEIFQGHLAGRKLVDFRG PFGWEYAAVNTGELRPIDDTPEDVDMKLRQVQPLAEVRVPF TLDVTELDSVARGATNPDLDDVARAAERMVEAEDSAIFHGW AQAGIKGIVDSTPHEALAVASVSDFPRAVLSAADTLRKAGV TGPYALVLGPKAYDDLFAATQDGYPVAKQVQRLVVDGPLVR ANALAGALVMSMRGGDYELTVGQDLSIGYAFHDRSKVELFV AESFTFRVLEPGAAVHLRYA	266	28812.66	4.82	26930
	Encapsulated ferritin from *H. ochraceum*	MSSEQLHEPAELLSEETKNMHRALVTLIEELEAVDWYQQRA DACSEPGLHDVLIHNKNEEVEHAMMTLEWIRRRSPVFDAHM RTYLFTERPILELEEEDTGSSSSVAASPTSAPSHGSLGIGS LRQEGKED	131	14799.39	4.70	13980

The additional Arg residue (bold, underlined) in the Hoch-Enc construct is a cloning artefact from the MoClo kit used.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

620

621 **Table 4: Autoinduction media components**

622

Media Reagents	Concentration
Tryptone	1 % (w/v)
Yeast Extract	0.5 % (w/v)
Glycerol	0.5 % (v/v)
Glucose	0.05 % (w/v)
α-D-lactose	0.2 % (w/v)
$(NH_4)_2SO_4$	25 mM
KH_2PO_4	50 mM
$Na_2PO_4.7H_2O$	50 mM
Magnesium Sulfate	2 mM

623

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

624 **Table 5: Cryo-EM data collection and processing**
625

	Loaded-Enc Icosahedral Reconstruction	Loaded-Enc C1 Reconstruction
Data Collection		
Microscope	FEI Titan Krios (300 kV)	
Detector	Gatan K3	
Acquisition Mode	Super resolution	
Pixel Size	0.326 Å/pix (super-resolution) 0.652 Å/pix (physical)	
Total dose	40.509 $e^-/Å^2$	
Fractional dose	40 frames during a 1 s exposure	
Defocus range	0.7 – 2.5 µm (0.3 µm steps)	
Micrograph 'movies' recorded	8109	
Movies DOI	TBC	
Reconstruction		
Alignment Software	MotionCor2	
Dose weighting	yes	
CTF fitting software	CTFFIND4	
Correction	full	
Particle picking method/software	Relion – templated from manual picked 2D classes	
Particles picked	340,403	309,404
Particles in 2D classes	304,741	287,914
Particles used in final 3D reconstruction	254,134	117,879
Alignment software	Relion	Relion
Reconstruction software	Relion	Relion
Box size	576 pixels (376 Å)	512 pixels (501 Å)
Voxel size (Å)	0.652	0.978
Symmetry	I1	C1
Map resolution (GS-FSC 0.143) (Å)	2.4	3.1
Sharpening B-factor ($Å^2$)	-97	-66
EMDB ID	tbc	tbc

626
627

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

628

629 **Table 6: Model building and refinement of the 'open' and 'closed' pentamer**
630 **conformations of the encapsulin nanocompartment.**
631

Symmetry expansion of I1 dataset		
	Closed Enc pentamer	**Open Enc pentamer**
Reconstruction		
Number of asymmetric units in final model	4,734,551	4,987,966
Map resolution (GS-FSC 0.143) (Å)	2.3	2.4
Map Sharpening B-factor (Å^2)	-93	-117
Coordinate refinement		
Software	Phenix	Phenix
Refinement algorithm	Real space	Real space
Resolution cutoff (Å)	2.3	2.4
$FSC_{model\text{-}vs\text{-}map}$= 0.5 (Å)	2.4	2.5
Model		
Number of amino acid residues	1375	1375
Bond length outliers	0	0
Bond angle outliers	0	0
RMS deviations		
Bonds (Å)	0.006	0.004
Angles (˚)	0.699	0.632
Validation		
Molprobity score	1.25	1.64
Clash score	2.88	7.83
Rotamer outliers (%)	2.2	0
C_β outliers (%)	0	0
CaBLAM outliers (%)	1.1	1.5
Ramachandaran (%) Favoured Outlier	 97.05 2.95	 96.68 3.32
Model vs Data CC (mask)	0.89	0.78
EM Ringer score	5.66	4.88
PDB ID	tbc	tbc

632

Table 7: Recorded uptake of deuterium for each peptide and timepoint observed in HDX-MS of Loaded-Enc. For each peptic peptide observed in the HDX-MS analysis, the deuterium uptake (in Da) and standard deviation (SD) of triplicate data is shown for each timepoint (10 s, 30 s, 2 mins, 5 mins, 4 hours and 24 hours) for each peptide. The number of exchangeable backbone amide hydrogens is also stated (exchangers).

		Loaded-Enc deuterium uptake (Da)															
Sequence	**Exchangers**	**Start**	**End**	**10s**	**10s SD**	**30s**	**30s SD**	**2 mins**	**2 mins SD**	**5 mins**	**5 mins SD**	**30 mins**	**30 mins SD**	**4 hrs**	**4 hrs SD**	**24 hrs**	**24 hrs SD**
LKRHLAPIVPDAW	10	3	15	0.34	0.08	0.31	0.10	0.36	0.29	0.36	0.15	0.38	0.08	0.55	0.08	NaN	NaN
LKRHLAPIVPDAWSA	12	3	17	0.45	0.06	0.39	0.08	0.36	0.09	0.41	0.06	0.56	0.11	0.89	0.05	NaN	NaN
EAKEIFQGHLAGRKLVD	16	21	37	0.62	0.06	0.55	0.10	0.59	0.10	0.65	0.08	0.90	0.14	1.66	0.06	1.42	0.19
EAKEIFQGHLAGRKLVDF	17	21	38	0.62	0.03	0.56	0.06	0.62	0.06	0.79	0.05	1.08	0.22	1.90	0.03	1.52	0.16
IFQGHLAGRKLVD	12	25	37	0.34	0.08	0.32	0.08	0.31	0.11	0.39	0.09	0.43	0.09	0.95	0.13	0.69	0.13
IFQGHLAGRKLVDF	13	25	38	0.40	0.05	0.38	0.05	0.44	0.06	0.61	0.08	0.79	0.20	1.58	0.10	NaN	NaN
FRGPFGWEY	7	38	46	0.82	0.06	0.70	0.28	0.93	0.10	0.97	0.15	1.45	0.29	2.30	0.07	1.89	0.03
YAAVNTGELRPIDDTPED	15	46	63	1.83	0.13	1.72	0.44	2.32	0.18	2.54	0.10	3.11	0.57	5.13	1.70	4.05	0.34
YAAVNTGELRPIDDTPEDVD	17	46	65	1.92	0.11	1.89	0.45	2.53	0.15	2.75	0.15	3.37	0.66	5.40	0.20	4.52	0.50
AAVNTGELRPIDDTPED	14	47	63	1.81	0.05	1.66	0.45	2.41	0.13	2.54	0.14	3.04	0.47	4.63	0.18	3.68	0.27
LRPIDDTPED	7	54	63	0.97	0.08	0.97	0.29	1.20	0.15	1.29	0.05	1.70	0.29	2.48	0.11	1.94	0.08
LRPIDDTPEDVD	9	54	65	1.13	0.12	0.99	0.31	1.35	0.18	1.53	0.14	1.83	0.28	2.87	0.12	2.59	0.24
MKLRQVQPLAE	9	66	76	0.82	0.12	0.73	0.23	0.67	0.15	0.75	0.23	1.01	0.14	1.29	0.08	1.17	0.23
KLRQVQPLAE	8	67	76	0.88	0.09	0.78	0.20	0.81	0.12	0.87	0.09	1.01	0.12	1.28	0.06	1.00	0.15
RQVQPLAE	6	69	76	0.69	0.11	0.64	0.18	0.71	0.11	0.63	0.14	0.85	0.10	1.07	0.05	0.88	0.09
VRVPFTL	5	77	83	0.12	0.04	0.13	0.04	0.06	0.08	0.09	0.08	0.15	0.05	0.24	0.03	NaN	NaN
VRVPFTLD	6	77	84	0.26	0.03	0.29	0.04	0.27	0.06	0.33	0.04	0.40	0.08	0.64	0.03	0.59	0.10
LDSVARGATNPDLDD	13	88	102	1.50	0.14	1.21	0.28	1.36	0.20	1.79	0.06	2.04	0.38	3.43	0.15	NaN	NaN
DSVARGATNPDL	10	89	100	1.25	0.04	1.12	0.12	1.25	0.07	1.39	0.10	1.69	0.29	3.12	0.17	2.27	0.23
DSVARGATNPDLDD	12	89	102	1.52	0.03	1.45	0.18	1.60	0.12	1.72	0.03	2.05	0.32	3.34	0.13	2.42	0.29
AEDSAIFHG	8	113	121	0.15	0.05	0.14	0.09	0.05	0.07	0.14	0.10	0.16	0.05	0.17	0.05	-0.02	0.04
HGWAQAGIKGIVDSTPHEAL	18	120	139	2.27	0.08	2.15	0.51	2.54	0.23	2.84	0.10	3.25	0.41	6.04	0.09	5.14	0.58
WAQAGIKGIVDSTPHEAL	16	122	139	1.70	0.05	1.63	0.33	2.01	0.14	2.22	0.08	2.56	0.40	4.91	0.11	4.20	0.32
IVDSTPHEAL	8	130	139	0.68	0.07	0.72	0.12	0.86	0.08	0.97	0.04	1.07	0.19	2.39	0.12	1.85	0.21
AVASVSDFPRAVL	11	140	152	1.40	0.06	1.30	0.37	1.69	0.13	1.84	0.06	1.98	0.26	3.08	0.10	2.31	0.28

SAADTLRKAGVTGPYA	14	153	168	0.89	0.04	0.85	0.15	1.07	0.09	1.29	0.13	1.86	0.37	2.89	0.06	2.92	0.27
SAADTLRKAGVTGPYAL	15	153	169	0.85	0.07	0.82	0.14	1.08	0.09	1.25	0.08	1.75	0.44	2.85	0.08	2.92	0.28
LVLGPKAYDDL	9	169	179	0.47	0.06	0.44	0.11	0.59	0.08	0.62	0.05	0.75	0.10	1.27	0.06	1.04	0.15
FAATQDGYPVAKQVQRL	15	180	196	3.03	0.10	3.28	0.17	3.86	0.19	4.31	0.18	5.28	0.92	8.30	0.33	6.92	0.53
VVDGPLVRA	7	197	205	0.62	0.07	0.63	0.17	1.08	0.07	1.26	0.12	1.58	0.30	1.96	0.07	2.00	0.18
VVDGPLVRANA	9	197	207	1.08	0.06	1.20	0.13	1.70	0.09	2.02	0.08	2.59	0.18	3.26	0.10	2.84	0.19
VVDGPLVRANAL	10	197	208	1.09	0.04	1.02	0.28	1.83	0.18	2.01	0.19	2.75	0.21	3.84	0.10	3.53	0.31
VVDGPLVRANALAGA	13	197	211	1.58	0.06	1.45	0.44	2.50	0.27	2.86	0.09	3.76	0.32	5.20	0.15	5.17	0.29
VVDGPLVRANALAGAL	14	197	212	1.56	0.08	1.55	0.39	2.52	0.19	2.89	0.08	3.36	0.59	5.28	0.17	5.13	0.49
NALAGAL	6	206	212	0.46	0.05	0.50	0.14	0.70	0.09	0.77	0.06	1.10	0.18	1.92	0.07	2.14	0.07
LSIGYAFHDRSKVEL	14	229	243	0.28	0.13	0.31	0.13	0.38	0.15	0.34	0.16	0.93	0.23	1.85	0.12	1.59	0.19
YAFHDRSKVEL	10	233	243	0.48	0.07	0.41	0.11	0.39	0.11	0.42	0.20	0.75	0.12	1.18	0.29	0.95	0.30
FVAESFT	6	244	250	0.16	0.02	0.27	0.75	0.13	0.03	0.12	0.04	0.18	0.03	0.12	0.02	NaN	NaN
FRVLEPGAA	7	251	259	0.32	0.06	0.33	0.06	0.44	0.07	0.57	0.08	0.88	0.28	1.45	0.05	1.09	0.11

638
639

Table 8: Deuterium uptake for each peptide and timepoint observed in HDX-MS of Empty-Enc.

The deuterium uptake (in Da) is shown for each timepoint (10 s, 30 s, 2 mins, 5 mins, 4 hours and 24 hours) for each peptic peptide from Empty-Enc observed in HDX-MS. HDX-MS experiments were performed in triplicate and the standard deviation (SD) is shown for each time point. The number of exchangeable backbone amide hydrogens in each peptide is also shown (exchangers).

		Empty-Enc deuterium uptake (Da)															
Sequence	**Exchangers**	**Start**	**End**	**10s**	**10s SD**	**30s**	**30s SD**	**2 mins**	**2 mins SD**	**5 mins**	**5 mins SD**	**30 mins**	**30 mins SD**	**4 hrs**	**4 hrs SD**	**24 hrs**	**24 hrs SD**
LKRHLAPIVPDAW	10	3	15	0.34	0.05	0.35	0.02	0.38	0.02	0.40	0.03	0.40	0.07	1.36	0.23	1.48	0.03
LKRHLAPIVPDAWSA	12	3	17	0.40	0.05	0.42	0.04	0.47	0.05	0.55	0.10	0.56	0.15	2.25	0.33	2.57	0.06
EAKEIFQGHLAGRKLVD	16	21	37	0.71	0.08	0.75	0.05	0.85	0.02	0.98	0.03	1.10	0.18	3.53	0.24	3.31	0.06
EAKEIFQGHLAGRKLVDF	17	21	38	0.65	0.05	0.68	0.04	0.84	0.01	1.07	0.04	1.21	0.25	3.58	0.20	3.29	0.06
IFQGHLAGRKLVD	12	25	37	0.37	0.04	0.40	0.03	0.43	0.03	0.47	0.09	0.55	0.09	1.81	0.28	1.60	0.06
IFQGHLAGRKLVDF	13	25	38	0.46	0.04	0.49	0.04	0.60	0.03	0.77	0.07	0.93	0.19	2.20	0.19	2.05	0.06
FRGPFGWEY	7	38	46	0.92	0.08	0.99	0.03	1.14	0.03	1.23	0.08	1.61	0.07	2.90	0.17	3.01	0.03
YAAVNTGELRPIDDTPED	15	46	63	1.96	0.15	2.15	0.11	2.63	0.08	2.93	0.10	3.45	0.51	5.76	0.20	5.71	0.17
YAAVNTGELRPIDDTPEDVD	17	46	65	2.07	0.12	2.24	0.10	2.76	0.10	3.12	0.12	3.51	0.53	6.70	0.22	6.61	0.15
AAVNTGELRPIDDTPED	14	47	63	1.90	0.12	2.14	0.07	2.60	0.04	2.94	0.10	3.25	0.44	5.25	0.18	5.18	0.17
LRPIDDTPED	7	54	63	1.07	0.05	1.22	0.01	1.41	0.04	1.54	0.09	1.65	0.25	2.42	0.14	2.36	0.03
LRPIDDTPEDVD	9	54	65	1.14	0.07	1.23	0.06	1.44	0.02	1.69	0.09	1.79	0.40	3.25	0.07	3.24	0.06
MKLRQVQPLAE	9	66	76	0.82	0.11	0.86	0.05	0.90	0.05	0.77	0.12	0.76	0.09	1.62	0.16	1.56	0.08
KLRQVQPLAE	8	67	76	0.79	0.08	0.84	0.06	0.88	0.07	0.85	0.08	0.88	0.10	1.60	0.20	1.48	0.08
RQVQPLAE	6	69	76	0.64	0.08	0.69	0.05	0.75	0.06	0.71	0.06	0.70	0.08	1.20	0.13	1.12	0.06
VRVPFTL	5	77	83	0.10	0.03	0.11	0.03	0.11	0.04	0.11	0.06	0.09	0.07	0.63	0.12	0.70	0.03
VRVPFTLD	6	77	84	0.24	0.03	0.24	0.02	0.30	0.02	0.35	0.04	0.42	0.09	1.01	0.08	1.04	0.03
LDSVARGATNPDLDD	13	88	102	1.37	0.16	1.48	0.11	1.64	0.08	1.55	0.21	1.61	0.49	3.68	0.29	3.53	0.14
DSVARGATNPDL	10	89	100	1.27	0.07	1.33	0.03	1.52	0.04	1.54	0.06	1.61	0.33	3.24	0.08	3.15	0.08
DSVARGATNPDLDD	12	89	102	1.48	0.12	1.59	0.04	1.80	0.02	1.95	0.08	2.13	0.27	3.82	0.16	3.61	0.12
AEDSAIFHG	8	113	121	0.12	0.04	0.13	0.05	0.12	0.04	0.11	0.08	0.09	0.08	0.18	0.11	0.13	0.03
HGWAQAGIKGIVDSTPHEAL	18	120	139	2.36	0.19	2.51	0.07	2.86	0.07	3.07	0.13	3.39	0.40	6.66	0.14	6.57	0.19
WAQAGIKGIVDSTPHEAL	16	122	139	1.76	0.18	1.89	0.14	2.19	0.14	2.44	0.15	2.74	0.42	5.65	0.19	5.45	0.19
IVDSTPHEAL	8	130	139	0.72	0.08	0.80	0.08	0.93	0.09	1.02	0.04	1.08	0.12	2.65	0.17	2.59	0.07

AVASVSDFPRAVL	11	140	152	1.52	0.08	1.66	0.04	1.94	0.05	2.09	0.04	2.19	0.19	3.42	0.11	3.38	0.10
SAADTLRKAGVTGPYA	14	153	168	0.93	0.07	1.05	0.05	1.32	0.04	1.53	0.06	1.71	0.31	3.39	0.10	3.37	0.06
SAADTLRKAGVTGPYAL	15	153	169	0.89	0.07	1.00	0.05	1.26	0.03	1.51	0.05	1.72	0.28	3.48	0.11	3.43	0.06
LVLGPKAYDDL	9	169	179	0.58	0.05	0.60	0.04	0.69	0.03	0.80	0.05	0.88	0.18	1.97	0.09	1.98	0.05
FAATQDGYPVAKQVQRL	15	180	196	3.62	0.16	3.96	0.19	4.97	0.10	5.40	0.04	6.34	0.09	8.32	0.12	8.39	0.10
VVDGPLVRA	7	197	205	0.80	0.06	0.93	0.06	1.39	0.06	1.68	0.04	1.83	0.21	2.58	0.15	2.70	0.06
VVDGPLVRANA	9	197	207	1.27	0.07	1.53	0.08	2.15	0.08	2.44	0.04	2.85	0.07	3.86	0.10	3.88	0.07
VVDGPLVRANAL	10	197	208	1.28	0.09	1.51	0.09	2.16	0.10	2.62	0.07	3.22	0.06	4.59	0.11	4.55	0.12
VVDGPLVRANALAGA	13	197	211	2.03	0.09	2.37	0.08	3.15	0.09	3.59	0.20	4.06	0.52	6.15	0.20	6.24	0.13
VVDGPLVRANALAGAL	14	197	212	2.08	0.10	2.41	0.10	3.24	0.06	3.58	0.10	4.26	0.10	6.39	0.18	6.41	0.16
NALAGAL	6	206	212	0.72	0.04	0.81	0.04	0.93	0.04	1.02	0.04	1.26	0.25	2.28	0.11	2.34	0.09
LSIGYAFHDRSKVEL	14	229	243	0.53	0.05	0.64	0.06	0.86	0.07	0.87	0.10	0.89	0.14	1.90	0.23	1.55	0.08
YAFHDRSKVEL	10	233	243	0.44	0.05	0.51	0.05	0.56	0.05	0.62	0.06	0.68	0.11	1.16	0.26	0.93	0.07
FVAESFT	6	244	250	0.09	0.01	0.09	0.01	0.10	0.01	0.11	0.03	0.07	0.02	0.12	0.03	0.10	0.01
FRVLEPGAA	7	251	259	0.29	0.03	0.38	0.55	0.50	0.03	0.72	0.04	0.89	0.20	1.46	0.10	1.36	0.03

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Afonine P V., Klaholz BP, Moriarty NW, Poon BK, Sobolev O V., Terwilliger TC, Adams PD, Urzhumtsev A, IUCr. 2018a. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr Sect D Struct Biol* **74**:814–840. doi:10.1107/S2059798318009324

Afonine P V., Poon BK, Read RJ, Sobolev O V., Terwilliger TC, Urzhumtsev A, Adams PD. 2018b. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr Sect D Struct Biol* **74**:531–544. doi:10.1107/S2059798318006551

Akita F, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A. 2007. 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**:1469–83. doi:10.1016/j.jmb.2007.02.075

Andrews SC. 2010. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta* **1800**:691–705. doi:10.1016/j.bbagen.2010.05.010

Barad BA, Echols N, Wang RY-R, Cheng Y, DiMaio F, Adams PD, Fraser JS. 2015. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* **12**:943–6. doi:10.1038/nmeth.3541

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**:12–21. doi:10.1107/S0907444909042073

Chiancone E, Ceci P, Ilari A, Ribacchi F, Stefanini S. 2004. Iron and proteins for iron storage and detoxification. *BioMetals* **17**:197–202. doi:10.1023/B:BIOM.0000027692.24395.76

Diekmann Y, Pereira-Leal JB. 2013. Evolution of intracellular compartmentalization. *Biochem J* **449**:319–331. doi:10.1042/BJ20120957

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**:486–501. doi:10.1107/S0907444910007493

Giessen TW, Orlando BJ, Verdegaal AA, Chambers MG, Gardener J, Bell DC, Birrane G, Liao M, Silver PA. 2019. Large protein organelles form a new iron sequestration system with high storage capacity. *Elife* **8**:1–23. doi:10.7554/eLife.46070

Giessen TW, Silver PA. 2017. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**:17029. doi:10.1038/nmicrobiol.2017.29

Gouet P, Robert X, Courcelle E. 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**:3320–3. doi:10.1093/nar/gkg556

He D, Hughes S, Vanden-Hehir S, Georgiev A, Altenbach K, Tarrant E, Mackay CL, Waldron KJ, Clarke DJ, Marles-Wright J. 2016. Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife* **5**:e18972. doi:10.7554/eLife.18972

He D, Piergentili C, Ross J, Tarrant E, Tuck LR, Mackay CL, McIver Z, Waldron KJ, Clarke DJ, Marles-Wright J. 2019. Conservation of the structural and functional architecture of encapsulated ferritins in bacteria and archaea. *Biochem J*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

476:975–989. doi:10.1042/bcj20180922

Iverson S V., Haddock TL, Beal J, Densmore DM. 2015. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enables Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth Biol* **5**:151019092657002. doi:10.1021/acssynbio.5b00124

Jenkins MC, Lutz S. 2021. Encapsulin Nanocontainers as Versatile Scaffolds for the Development of Artificial Metabolons. *ACS Synth Biol* acssynbio.0c00636. doi:10.1021/acssynbio.0c00636

Kelly LA, Mezulis S, Yates C, Wass M, Sternberg M. 2015. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc* **10**:845–858. doi:10.1038/nprot.2015-053

Kerfeld CA, Heinhorst S, Cannon GC. 2010. Bacterial Microcompartments.

Laguerre M, Saux M, Dubost JP, Carpy A. 1997. MLPP: A program for the calculation of molecular lipophilicity potential in proteinsPharmaceutical Sciences. John Wiley & Sons, Ltd. pp. 217–222. doi:10.1111/j.2042-7158.1997.tb00257.x

Lau YH, Giessen TW, Altenburg WJ, Silver PA. 2018. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat Commun* **9**:1311. doi:10.1038/s41467-018-03768-x

Liebschner D, Afonine P V., Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev O V., Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD, IUCr. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr Sect D Struct Biol* **75**:861–877. doi:10.1107/S2059798319011471

Masson GR, Burke JE, Ahn NG, Anand GS, Borchers C, Brier S, Bou-Assaf GM, Engen JR, Englander SW, Faber J, Garlish R, Griffin PR, Gross ML, Guttman M, Hamuro Y, Heck AJR, Houde D, Iacob RE, Jørgensen TJD, Kaltashov IA, Klinman JP, Konermann L, Man P, Mayne L, Pascal BD, Reichmann D, Skehel M, Snijder J, Strutzenberg TS, Underbakke ES, Wagner C, Wales TE, Walters BT, Weis DD, Wilson DJ, Wintrode PL, Zhang Z, Zheng J, Schriemer DC, Rand KD. 2019. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat Methods* **16**:595–602. doi:10.1038/s41592-019-0459-y

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**:658–674. doi:10.1107/S0021889807021206

McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, Winkler DC, Lam AS, Wall JS, Steven AC, Hoiczyk E. 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J* **33**:1896–1911. doi:10.15252/embj.201488566

Nakane T, Lindahl E, Zivanov J, Hagen WJJH, Scheres SHWH, Kimanius D, Forsberg BBO, Nakane T, Forsberg BBO, Kimanius D, Hagen WJJH, Lindahl E, Scheres SHWH. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**:1–38. doi:10.7554/elife.42166

Nichols RJ, LaFrance B, Phillips NR, Oltrogge LM, Valentin-Alvarado LE, Bischoff AJ, Nogales E, Savage DF. 2020. Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. *bioRxiv*. doi:10.1101/2020.05.24.113720

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

754 Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH,
755 Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers,
756 educators, and developers. *Protein Sci* **30**:70–82. doi:10.1002/pro.3943
757 Piergentili C, Ross J, He D, Gallagher KJ, Stanley WA, Adam L, Mackay CL, Baslé
758 A, Waldron KJ, Clarke DJ, Marles-Wright J. 2020. Dissecting the structural and
759 functional roles of a putative metal entry site in encapsulated ferritins. *J Biol*
760 *Chem*. doi:10.1074/jbc.RA120.014502
761 Prangishvili D, Sutter M, Boehringer D, Gutmann S, Gu S, Loessner MJ, Stetter KO,
762 Weber-ban E, Ban N, Brevibacterium M. 2008. Structural basis of enzyme
763 encapsulation into a bacterial nanocompartment **15**:939–947.
764 doi:10.1038/nsmb.1473
765 Putri RM, Allende-Ballestero C, Luque D, Klem R, Rousou K-A, Liu A, Traulsen CH-
766 H, Rurup WF, Koay MST, Castón JR, Cornelissen JJLM. 2017. Structural
767 Characterization of Native and Modified Encapsulins as Nanoplatforms for *in*
768 *Vitro* Catalysis and Cellular Uptake. *ACS Nano* **11**:12796–12804.
769 doi:10.1021/acsnano.7b07669
770 Rahmanpour R, Bugg TDH. 2013. Assembly in vitro of Rhodococcus jostii RHA1
771 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J*
772 **280**:2097–2104. doi:10.1111/febs.12234
773 Recalcati S, Gammella E, Buratti P, Cairo G. 2017. Molecular regulation of cellular
774 iron balance. *IUBMB Life* **69**:389–398. doi:10.1002/iub.1628
775 Rohou A, Grigorieff N. 2015. CTFFIND4: Fast and accurate defocus estimation from
776 electron micrographs. *J Struct Biol* **192**:216–21. doi:10.1016/j.jsb.2015.08.008
777 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T,
778 Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ,
779 Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source
780 platform for biological-image analysis. *Nat Methods* **9**:676–682.
781 doi:10.1038/nmeth.2019
782 Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large
783 numbers of sequences. *Methods Mol Biol* **1079**:105–16. doi:10.1007/978-1-
784 62703-646-7_6
785 Snijder J, Kononova O, Barbu IM, Uetrecht C, Rurup WF, Burnley RJ, Koay MST,
786 Cornelissen JJLM, Roos WH, Barsegov V, Wuite GJL, Heck AJR. 2016.
787 Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded
788 Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **17**:2522–9.
789 doi:10.1021/acs.biomac.6b00469
790 Studier F. 2005. Protein production by auto-induction in high-density shaking
791 cultures. *Protein Expr Purif* **41**:207–234. doi:10.1016/j.pep.2005.01.016
792 Tamura A, Fukutani Y, Takami T, Fujii M, Nakaguchi Y, Murakami Y, Noguchi K,
793 Yohda M, Odaka M. 2015. Packaging guest proteins into the encapsulin
794 nanocompartment from Rhodococcus erythropolis N771. *Biotechnol Bioeng*
795 **112**:13–20. doi:10.1002/bit.25322
796 Williams EM, Jung SM, Coffman JL, Lutz S. 2018. Pore Engineering for Enhanced
797 Mass Transport in Encapsulin Nanocompartments. *ACS Synth Biol* **7**:2514–
798 2517. doi:10.1021/acssynbio.8b00295
799 Xiong X, Sun C, Vago FS, Klose T, Zhu J, Jiang W. 2020. Cryo-EM structure of
800 heterologous protein complex loaded thermotoga maritima encapsulin capsid.
801 *Biomolecules* **10**:1–13. doi:10.3390/biom10091342
802 Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. 2008. Protein-based
803 organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

804 *Microbiol* **6**:681–691. doi:10.1038/nrmicro1913
805 Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017.
806 MotionCor2: anisotropic correction of beam-induced motion for improved cryo-
807 electron microscopy. *Nat Methods* **14**:331–332. doi:10.1038/nmeth.4193
808
809

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Funding

This work was supported a Royal Society Research Grant awarded to JMW [RG130585] and a BBSRC New Investigator Grant to JMW and DJC [BB/N005570/1]. CP was funded by the BBSRC New Investigator Grant [BB/N005570/1]. JMW is funded by Newcastle University. DJC and JR are funded by the University of Edinburgh. JR was funded by a BBSRC EastBio DTP studentship [BB/M010996/1]. ZM is funded by a BBSRC NLD DTP studentship [BB/M011186/1]. JEB was funded by DSTL via Imperial College. EZA was funded by an IBioIC PhD studentship with Fujifilm Diosynth Biotechnologies. KJW was funded by the Biotechnology and Biological Sciences Research Council (BB/S006818/1)

Equipment for Transmission Electron Microscopy was funded through the BBSRC 17ALERT call [BB/R013942/1]. MDW's work is supported by the Wellcome Trust and Royal Society (210493), Medical Research Council (T029471/1), and the University of Edinburgh. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149).

Acknowledgements

We acknowledge Diamond Light Source for access and support of the cryo-EM facilities at the UK's national Electron Bio-imaging Centre (eBIC) [under proposal EM16637], funded by the Wellcome Trust, MRC and BBRSC. We thank Dr Yun Song for their assistance at eBIC with sample loading and data collection setup. This research made use of the Rocket High Performance Computing service at Newcastle University. We thank Dr Karen Bower for assistance with the use of the Rocket service. Grid screening was performed in the Edinburgh cryo-EM facility in School of Biological Sciences at the University of Edinburgh. The cryo-EM facility was set up with funding from the Wellcome Trust (087658/Z/08/Z) and SULSA.
and is supported by the Wellcome Centre for Cell Biology.

Author Contributions

Conceptualization: JR, JMW, DJC
Methodology: JR, ZM, TL, CP, AMBL, JMW, DJC
Validation: JR, ZM, TL, JMW, DJC
Formal analysis: JR, ZM, TL, JMW, DJC
Investigation: JR, ZM, KG, JEB, CP, EZA
Resources: KG, FC, CLM, JEB, AMBL, EZA, MDW
Data curation: JR, TL, DJC, AMBL, JMW
Writing – original draft preparation: JR, DJC, JMW
Writing – review and editing: JR, ZM, CP, AMBL, MDW, KJW, JMW, DJC
Visualization: JR, ZM, TL, DJC, JMW
Supervision: JMW, DJC
Funding acquisition: MDW, LEH, KJW, DJC, JMW

Figures

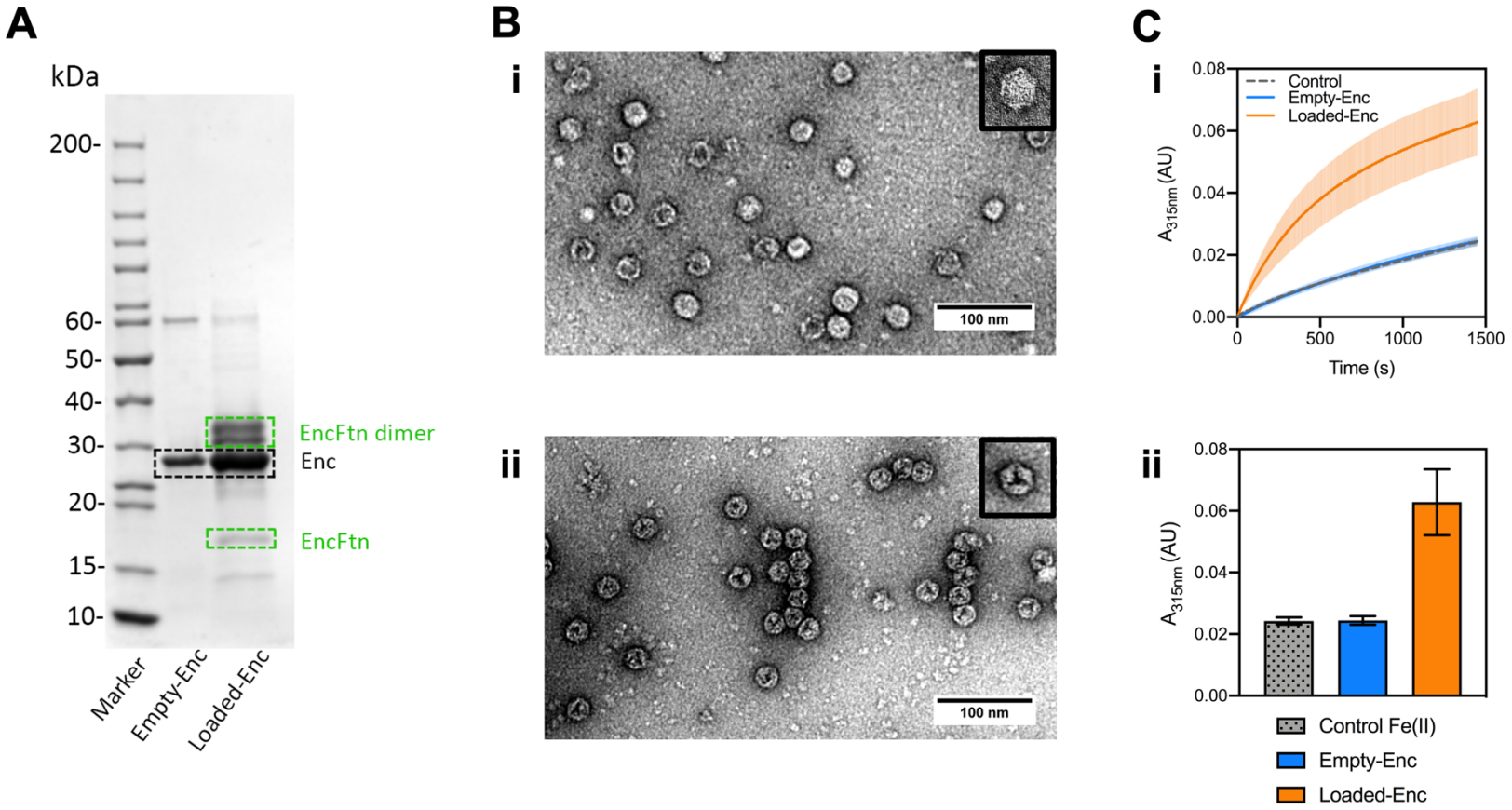

Figure 1. Validation of the Assembly and Activity of Loaded-Enc and Empty-Enc.

A: SDS-PAGE of purified Empty-Enc and Loaded-Enc. Proteins resolved by 15% acrylamide SDS-PAGE and stained with Coomassie blue stain. Encapsulin bands are near the 30 kDa marker and highlighted by a black dashed box. The EncFtn cargo of Loaded-Enc appears as both a monomer and dimer. The purified EncFtn monomer is highlighted by a box labelled 'EncFtn'. Two overlapping bands at approximately 35 kDa are highlighted by a green dashed box and labelled as 'EncFtn dimer'. **B**: Negative stain transmission micrographs of Empty-Enc (**i**) and Loaded-Enc (**ii**) displaying individual particles for each complex. One nanocompartment of Empty-Enc and Loaded-Enc is shown in the upper right corner of each micrograph, with a hexagonal 2D geometry observed. **C: i**: Ferroxidase activity of Loaded-Enc compared to Empty-Enc. Protein samples were mixed with 100 μM $FeSO_4.{}_2 7H_2O$. Following an incubation period at room temperature of 50 seconds absorbance at 315 nm was measured over a time-course of 1450 seconds. Control reference established using enzyme-free reaction as a measure of background iron oxidation. Lines represent the mean of three technical repeats, error bars represent the standard deviation from the mean. **ii**: End point ferroxidase assay comparison. Ferroxidase activity shown by the total increase in $A_{315\ nm}$ at the end point of the assay. Bars represent the mean of three technical repeats, error bars represent the standard deviation from the mean.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

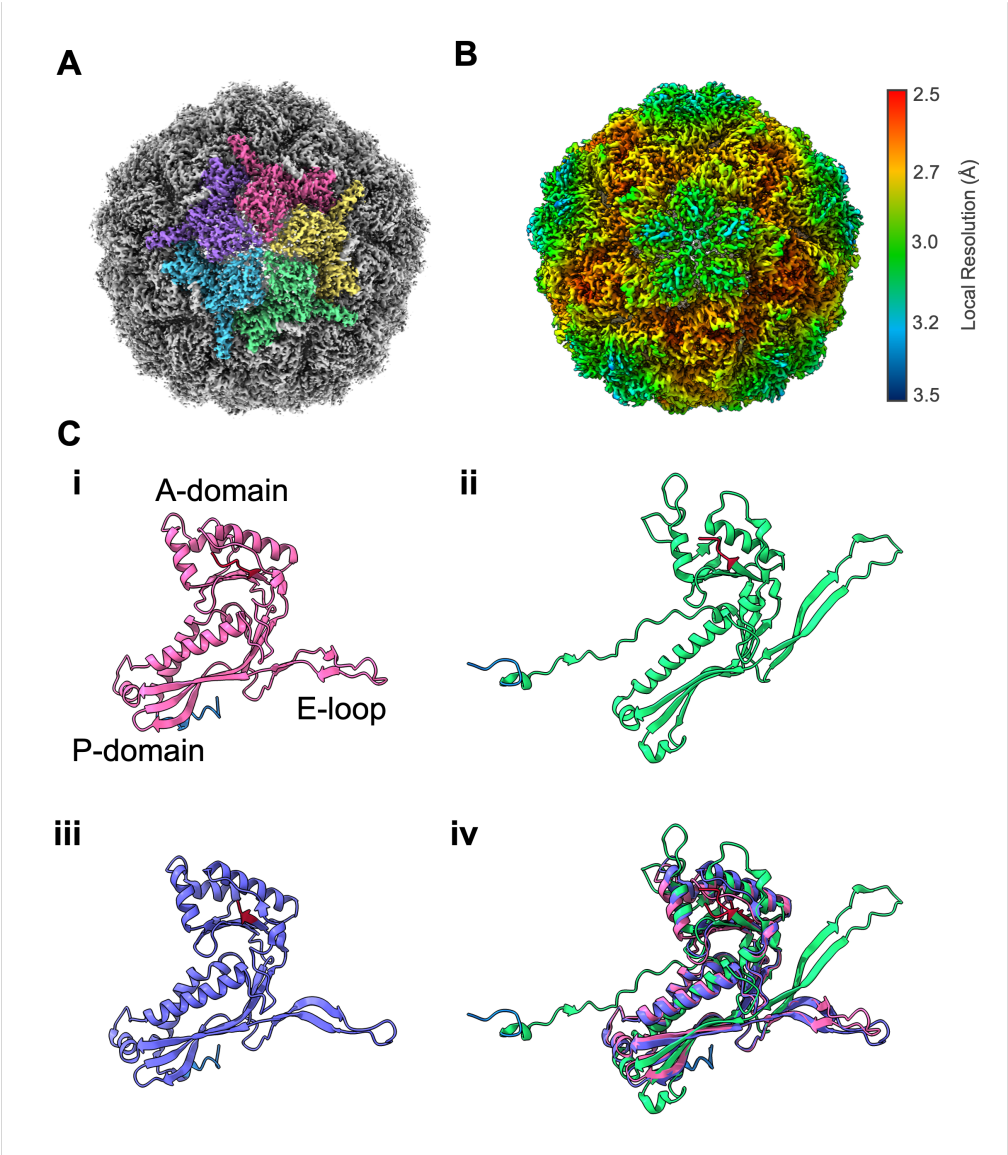

Figure 2: Architecture of the *H. ochraceum* encapsulin nanocompartment shell
Visualization of the electronic potential map of the *H. ochraceum* encapsulin from an icosahedrally averaged single particle reconstruction. **A**: The exterior of the encapsulin shell visualized at 2.4 Å resolution. Five subunits of the encapsulin nanocompartment shell have been colored to highlight the 5-fold axis. **B:** Icosahedral EM map of Loaded-Enc sharpened by local resolution estimate and colored by local resolution. The estimated resolution varies across the exterior of the encapsulin nanocompartment with the lowest resolution at the 5-fold pores. Color key of resolution mapping is shown on the right-hand side of the figure. **C**: The shared phage-like fold in the HK97 bacteriophage capsid and encapsulin proteins. Monomeric subunits of the *H. ochraceum* encapsulin protein modelled from our reconstruction are shown (pink, **C i**), with comparisons to the HK97 bacteriophage Head II *T*=7 monomer (green, **C ii**, PDB: 2FT1), and the *T. maritima T*=1 monomer (purple, **C iii**, PDB: 3DKT). The N-terminus of each monomer is highlighted in blue and the C-termius in red **C iv**: Comparison of the structures of **i** (pink), **ii** (green) and **iii** (purple) showing similar A- and P-domains. A noticeable difference between the monomers is that the E-loop of **ii** is shifted away from the others, which is typical of icosahedral encapsulin structures with a *T*-number greater than 1**.**

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

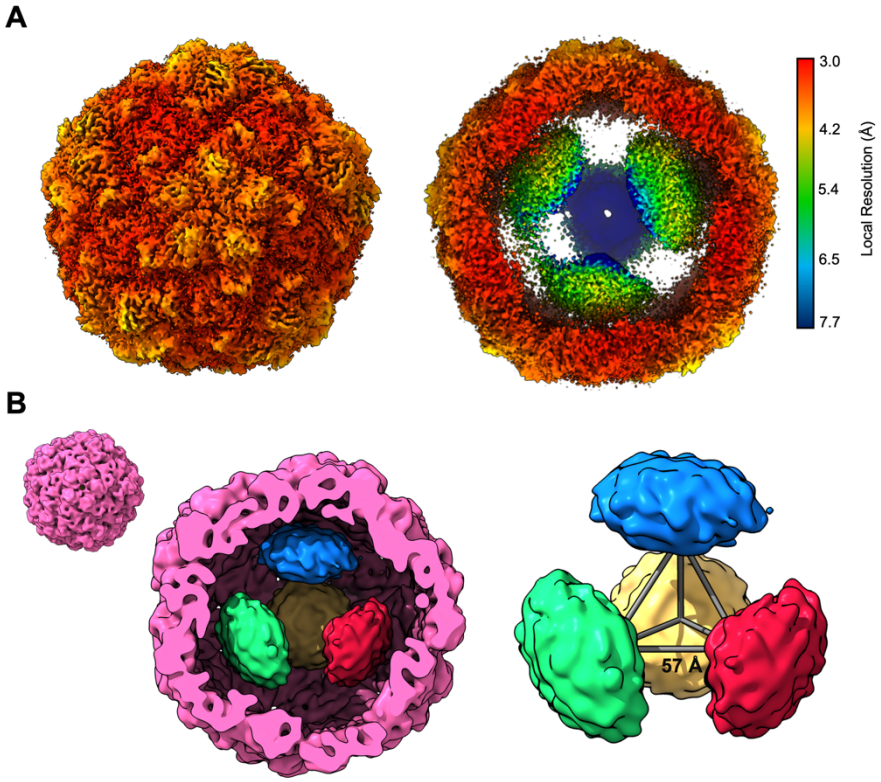

Figure 3: Asymmetric reconstruction of the *H. ochraceum* encapsulin complex reveals a tetrahedral arrangement of EncFtn within the encapsulin nanocompartment.

A: Electronic potential map of the asymmetric reconstruction of the *H. ochraceum* Enc:EncFtn complex. The map is colored by local resolution, highlighting the internal organization of EncFtn within Loaded-Enc. The interior of the nanocompartment holds four EncFtn, which are of significantly lower resolution than the shell. The color key for the local resolution is shown on the right side of **A**. **B**: Gaussian smoothed C1 map showing the four discrete EncFtn densities (red, green, yellow and blue), consistent with the size of a decameric EncFtn comples, within the encapsulin nanocompartment (pink). The four EncFtn are in a tetrahedral arrangement highlighted by grey lines connecting their centers of mass. The average distance between each EncFtn decamer is 57 Å.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.15.439977; this version posted April 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

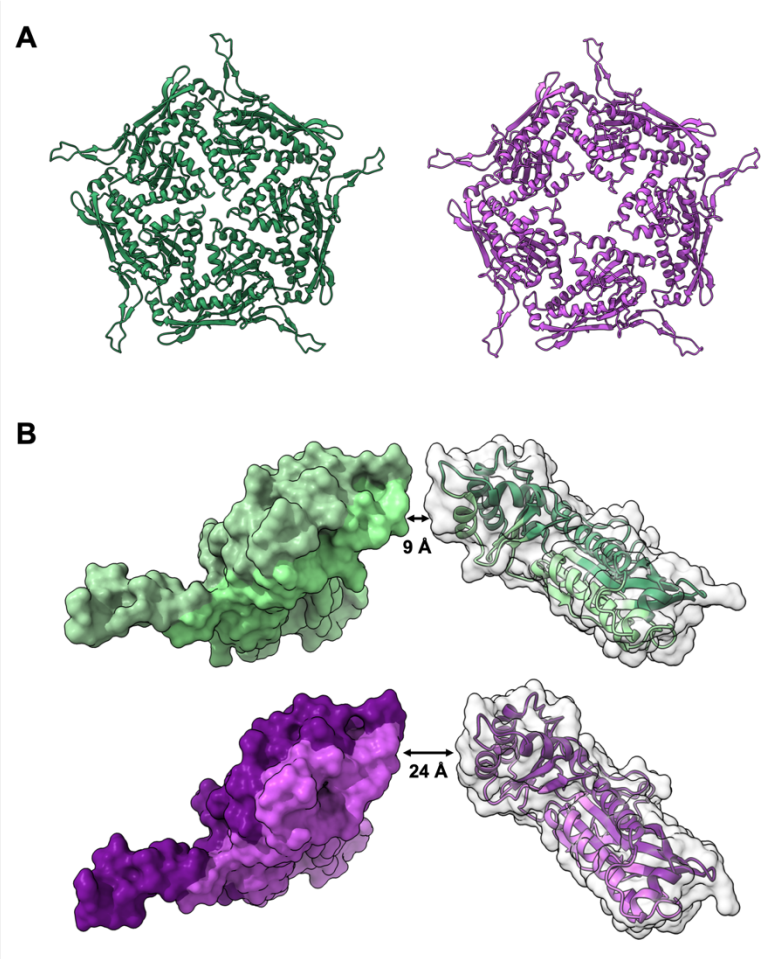

Figure 4: Conformations of the dynamic 5-fold pore of the *H. ochraceum* encapsulin shell.
Masked 3D refinements centered around the 5-fold pore of the icosahedral reconstruction were performed after symmetry expansion of the asymmetric units. **A**: Cartoon representations of the 'closed' (green) and 'open' (purple) conformations of the Enc shell pentamer. The closed conformation has a 5-fold pore diameter of 9 Å whilst the equivalent diameter of the open conformation is 24 Å. **B**: Side-on view of the 5-fold pore highlighting its increased diameter in the open conformation (purple) compared to the closed conformation (green). The interior face of encapsulin for the closed and open conformations is shown in light green and light purple respectively.

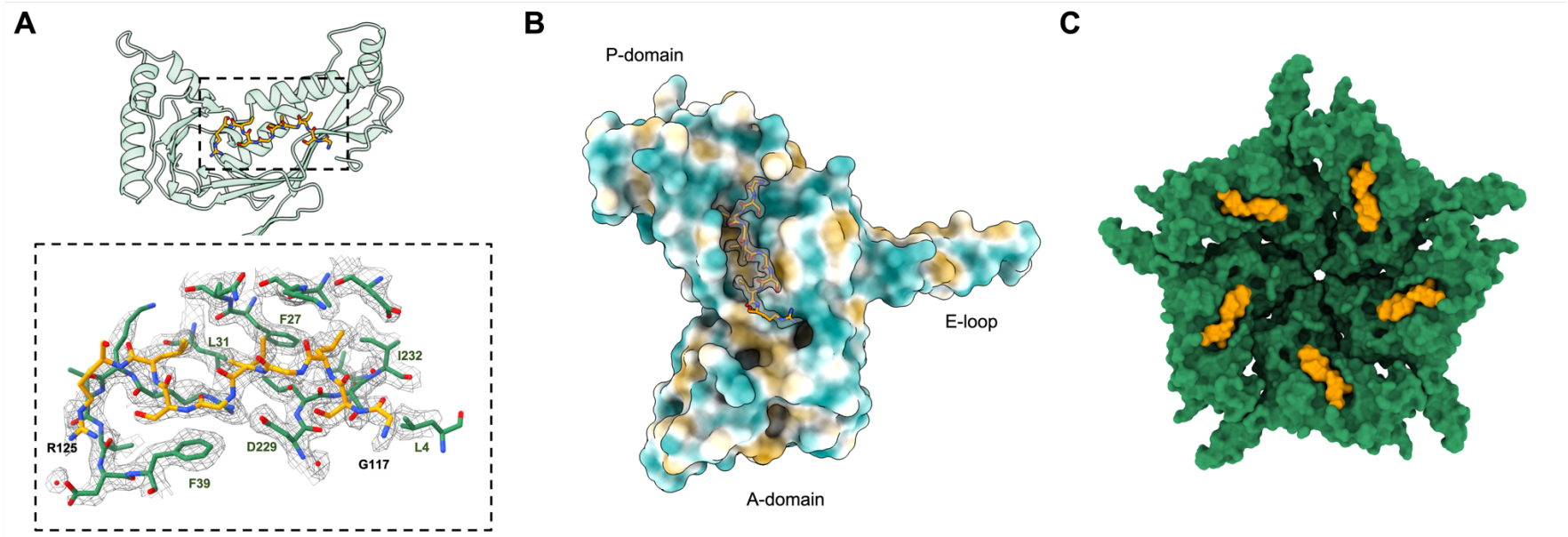

Figure 5: The closed conformation of the *H. ochraceum* five-fold pore allows docking and sequencing of the EncFtn localization sequence to the interior face of the encapsulin monomer.
A: Binding of localization sequence residues from the EncFtn to the interior wall of the encapsulin nanocompartment monomer (green residues and transparent cartoon) and the localization sequence (yellow sticks). Lower panel: Hydrophobic residues from the interior face of the encapsulin form the binding pocket for the localization sequence. The first and last residues modelled for the LS have been labelled in black, and key residues from encapsulin have been labelled in green. **B:** The spatial relationship between an encapsulin monomer and the localization sequence (shown in gold with its EM map density as in **A**). The Enc monomer has been colored by molecular lipophilicity potential (Laguerre et al., 1997) which ranges from dark cyan (corresponding to the most hydrophilic) to white to dark gold (most lipophilic). This highlights the hydrophobic pocket on the interior of the encapsulin nanocompartment where the localization sequenced binds. **C**: The 'closed' conformation pentamer (green) with a localization sequence (yellow) shown on each monomer.

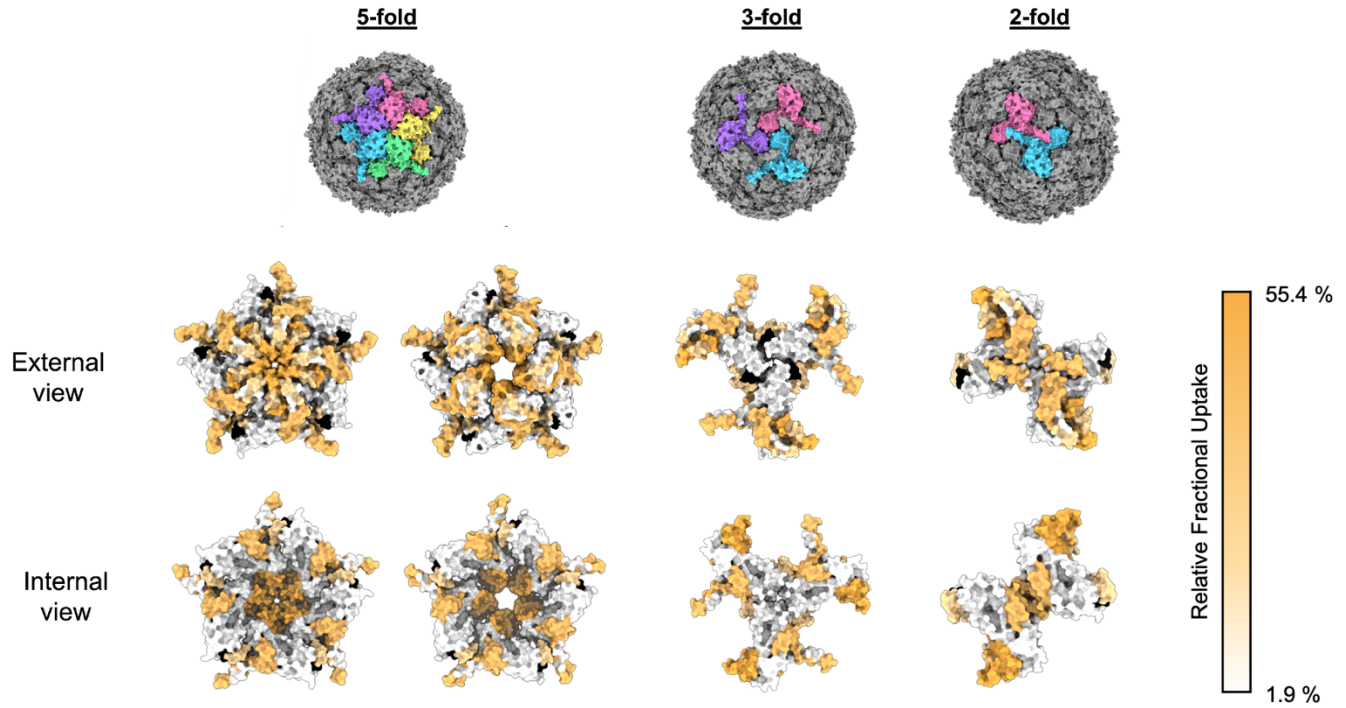

Figure 6: HDX-MS fractional uptake mapped to the symmetry axes of the icosahedral reconstruction of the *H. ochraceum* encapsulin complex.
From top to bottom: The top row depicts molecular surface models of the *H. ochraceum* encapsulin in three orientations to show the 5-fold, 3-fold and 2-fold symmetry axes. Monomers have been colored individually (pink, yellow, green, blue and purple) to highlight the symmetry axes. The final two rows display the relative fractional uptake of each peptide from HDX experiments displayed on the 5-fold, 3-fold and 2-fold pores of the encapsulin nanocompartment. Peptide fractional uptake is show on a white-to-orange color scale with a color key show on the right-hand side of the figure. Areas colored black correspond to no peptide coverage.

Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment revealed by Cryo-EM and hydrogen/deuterium exchange mass spectrometry

Supplementary Figures

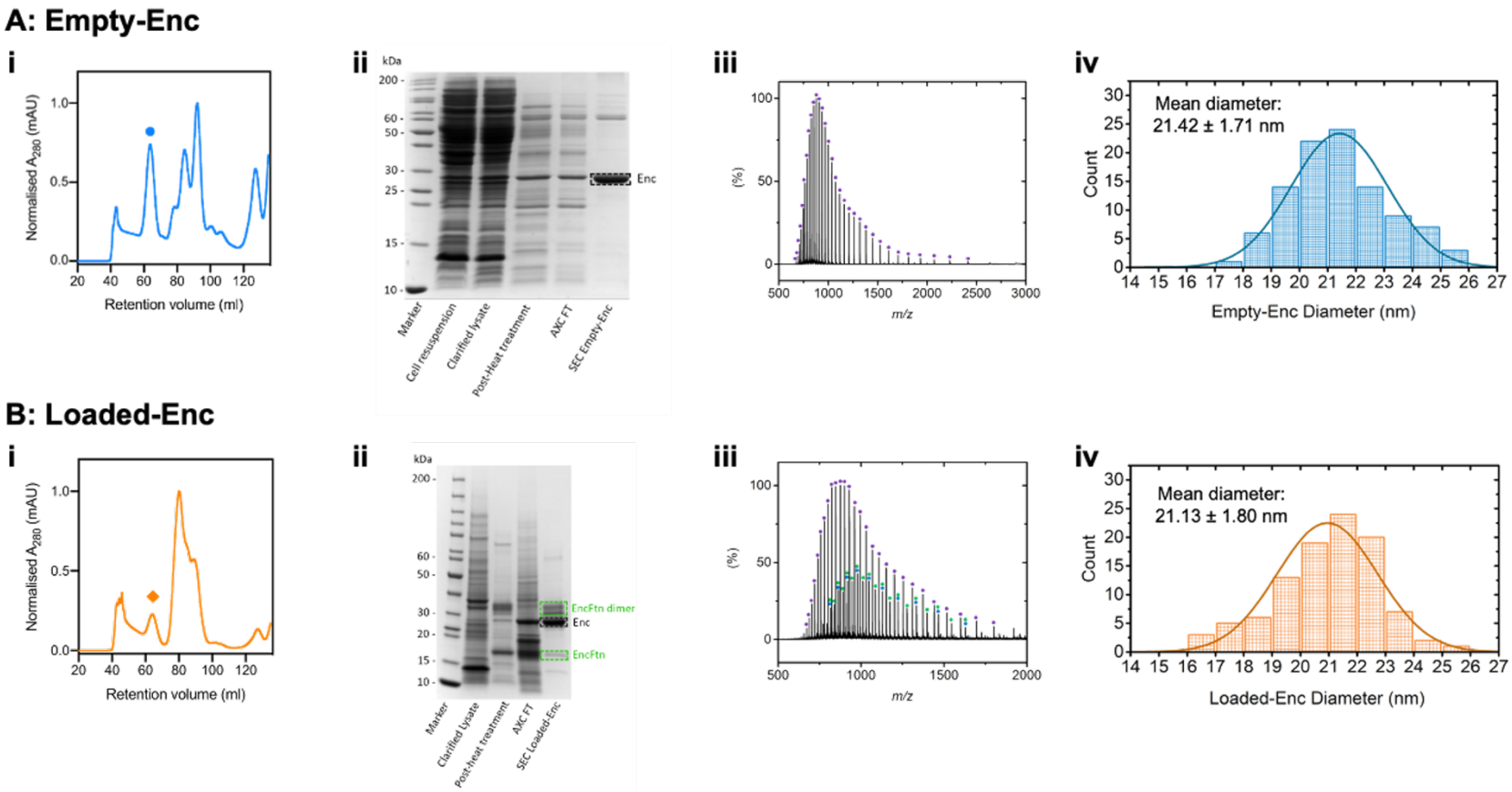

Figure 1 – figure supplement 1: Purification of recombinant *H. ochraceum* encapsulin protein complexes.
A: i. Size exclusion chromatogram traces of Empty-Enc with normalised A_{280} values. Empty-Enc (●) eluted at a retention volume of 65 ml. Other peaks represent contaminants and encapsulin monomers. **ii.** SDS-PAGE of Empty-Enc samples at different stages of purification. Proteins were resolved by 15% acrylamide SDS-PAGE and stained with Coomassie blue stain. The encapsulin band is near the 30 kDa marker of the PageRuler Unstained Protein Ladder (highlighted with a dashed black box). **iii**: Mass spectrum of Empty-Enc displaying one charge state distribution (highlighted with purple circles), which corresponds to the mass of the encapsulin monomer protein (28969.7 Da). **iv:** Histograms of the size distribution of Empty-Enc from negative stain TEM. A Gaussian curve was fitted to the data by nonlinear least squares regression, showing that individual Empty-Enc nanocompartments have a mean diameter of 21.13 nm with a standard

deviation of 1.80 nm. **B: i.** Size exclusion chromatogram trace of Loaded-Enc. Loaded-Enc elutes at 65 ml, the peak is stressed with an orange diamond (◆) Non-encapsulated EncFtn proteins are eluted at 80ml, exclusively in the Loaded-Enc SEC run. **ii.** SDS-PAGE of loaded-Enc at different stages of the purification. Proteins were resolved by 15% acrylamide SDS-PAGE and stained with Coomassie blue stain. Loaded-Enc encapsulin band is near 30 kDa (highlighted with a dashed black box) and the EncFtn bands are highlighted by dashed green boxes. EncFtn exists as both monomer and dimer bands, as is typical for these proteins. **iii**: Mass spectrum of proteins present in Loaded-Enc are highlighted by coloured circles; encapsulin nanocompartment protein in purple (28813.2 Da), EncFtn monomer in blue (14667.4 Da), and EncFtn dimer in green (29334.5 Da). **iv:** Histogram of the size distribution of Loaded-Enc as observed by negative stained TEM. Nanocompartments of Loaded-Enc were found to have an average diameter of 21.42 nm with a standard deviation of 1.71 nm.

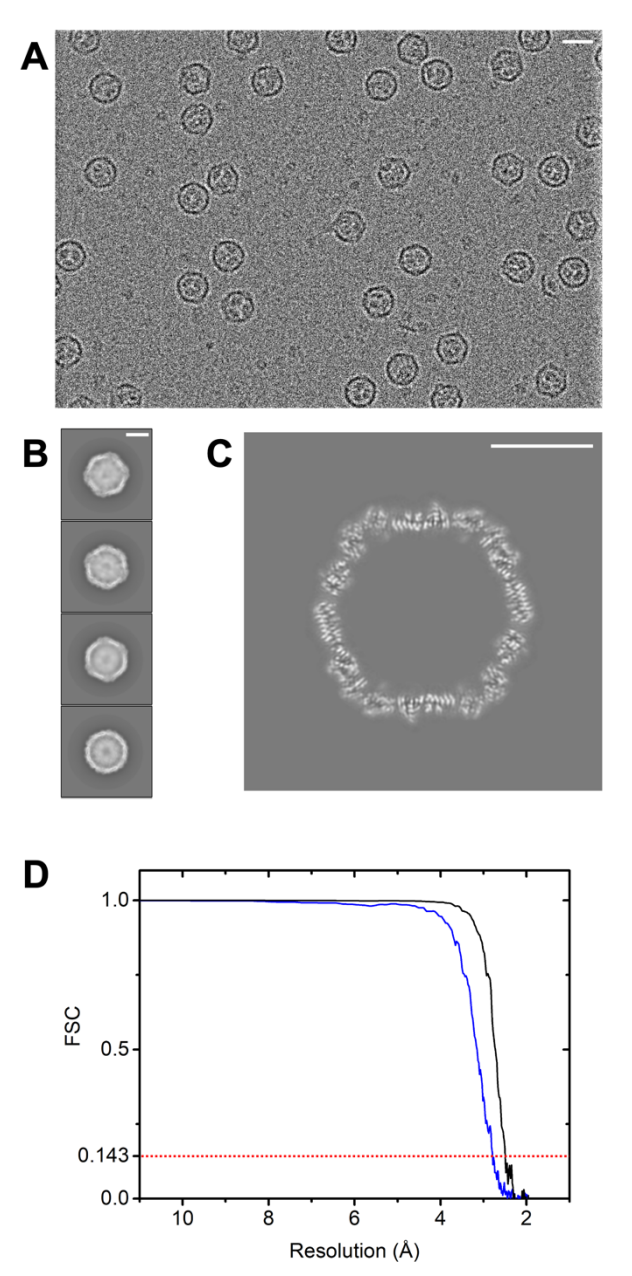

Figure 2 – figure supplement 1: Representative cryo-EM micrograph and additional data from the icosahedral reconstruction of the *H. ochraceum* encapsulin complex

A: Representative cryo-EM micrograph of Loaded-Enc. **B**: 2D classes of Loaded-Enc **C**: Central slice of Loaded-Enc from icosahedral processing. A white scale bar representing 10 nm is shown in the upper right corner of **A**, **B** and **C**. **D**: Gold standard FSC curve of the masked (black) and unmasked (blue) icosahedral reconstruction half maps. The FSC 0.143 threshold is highlighted with a dashed red line.

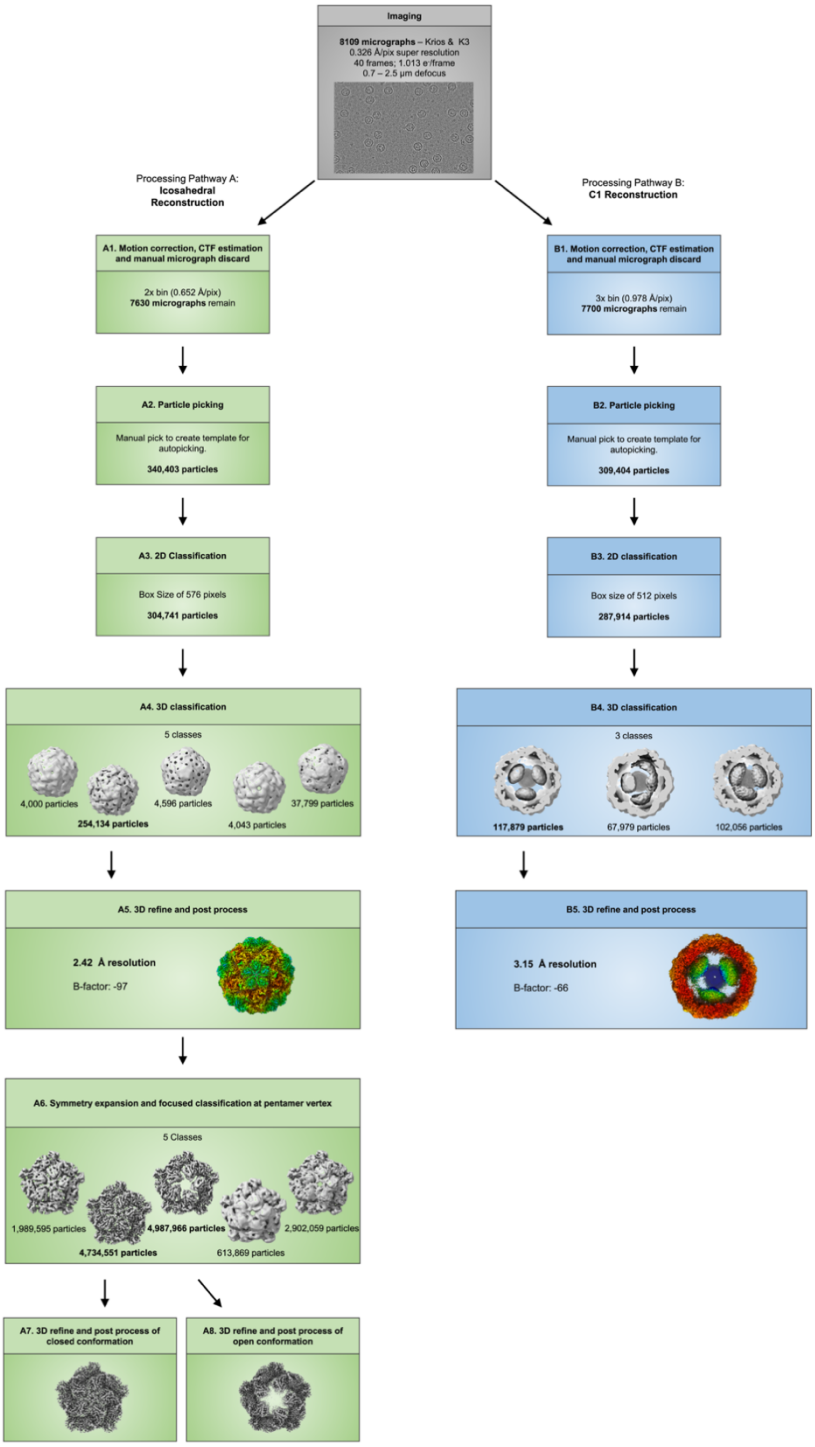

Figure 2 – figure supplement 2: Cryo-EM Processing Workflow
The processing pipeline within Relion 3.1 used to obtain the single particle reconstruction of Loaded-Enc. The green and blue boxes highlight divergence in the workflow where two approaches were executed simultaneously to obtain the highest possible resolution structures of exterior and interior of Loaded-Enc.

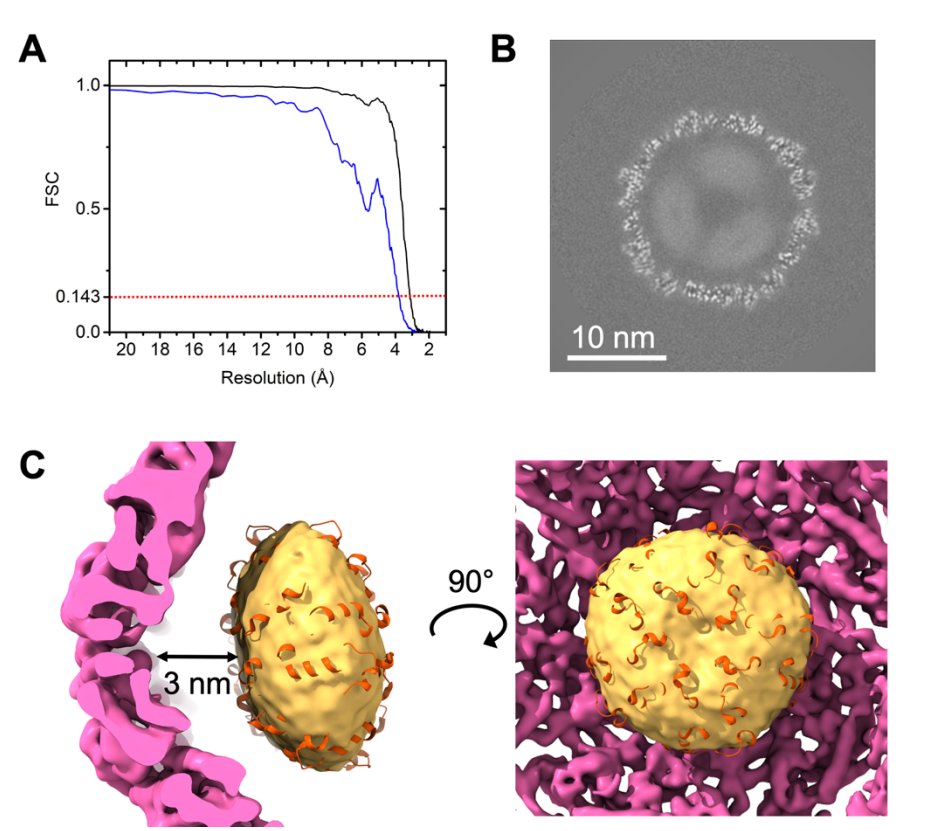

Figure 3 – figure supplement 1: Supplementary cryo-EM data of Loaded-Enc C1 reconstruction.
A: Gold standard FSC curve of the masked (black) and unmasked (blue) icosahedral reconstruction half maps. The FSC 0.143 threshold is highlighted with a dashed red line. **B**: Central slice of Loaded-Enc from C1 reconstruction after 3D refinement. A white scale bar representing 10 nm is shown in the lower left corner. **C**: The crystal structure of the EncFtn from *H. ochraceum* (orange, PDB 5N5F) docked into the interior density (yellow) of the C1 reconstruction. A 3 nm gap is observed between the Enc shell (pink) and the EncFtn (yellow and orange).

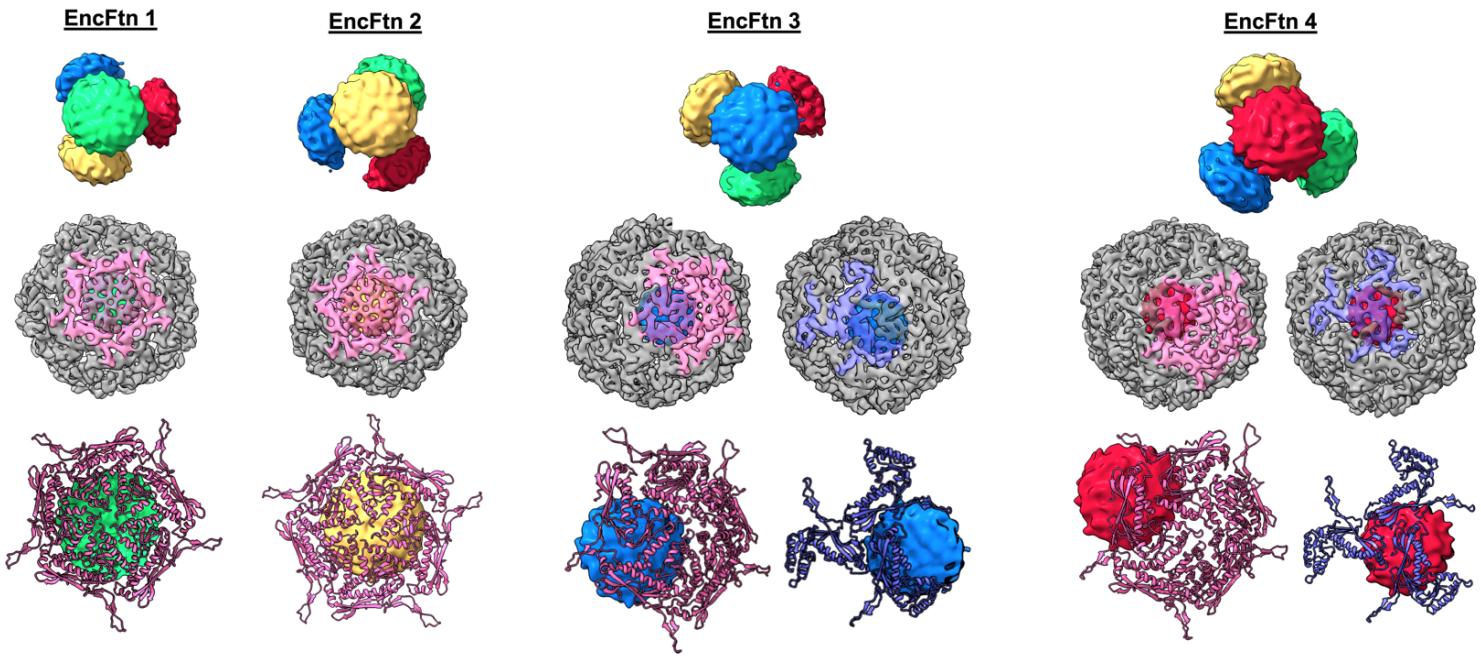

Figure 3 – figure supplement 2: Visualization of the distinct EncFtn environments within the icosahedral encapsulin nanocompartment.
Top panels: Each EncFtn within the encapsulin nanocompartment has been individually colored (as in **Figure 4 C**) and numbered. Four orientations of EncFtn tetrahedral are shown with a different EncFtn in the foreground. **Middle panels:** Each EncFtn of the Loaded-Enc complex as viewed from outside of the encapsulin nanocompartment. EncFtn complexes are in the same orientation as the top panels. A pentamer of the encapsulin nanocompartment has been colored pink to allow direct correlation between the EncFtn location and the 5-fold pore of encapsulin. EncFtn 3 and EncFtn 4 do not align with the 5-fold pore and so the encapsulin 3-fold pore has also been shown and colored purple. **Bottom panels:** The relationship between each EncFtn and the pores of encapsulin. The encapsulin 5-fold pore is shown by pink cartoons, the encapsulin 3-fold pore by purple cartoons and the EncFtn are colored as in the top panels. EncFtn 1 and EncFtn 2 are in broadly equivalent environments which are aligned with the 5-fold pore of the encapsulin nanocompartment.

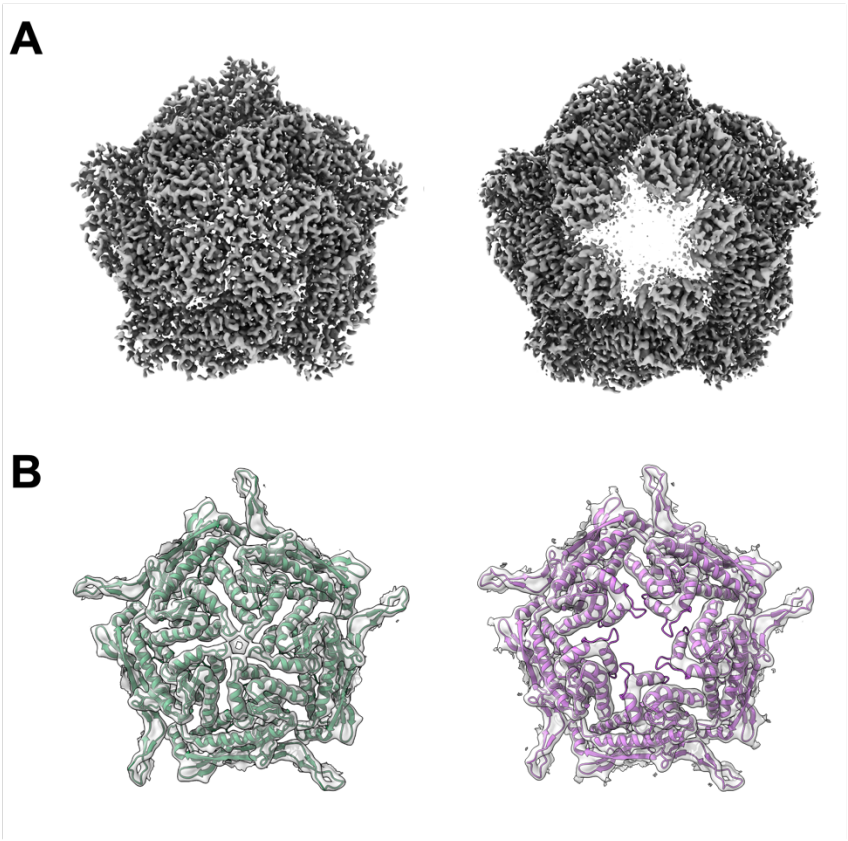

Figure 4 – figure supplement 1: Electronic potential maps and models from masked 3D-classification and refinement centered on the five-fold symmetry axis of icosahedral reconstruction.
A: CryoEM maps of the ‘closed’ (left) and ‘open’ (right) pentamer conformations from symmetry expansion of the icosahedral five-fold axis. **B**: The Gaussian smoothed cryoEM maps with docked models of the ‘closed’ (green, left) and ‘open’ (purple, right) conformations. Smoothed EM maps allow easy visualization of the docked secondary structure.

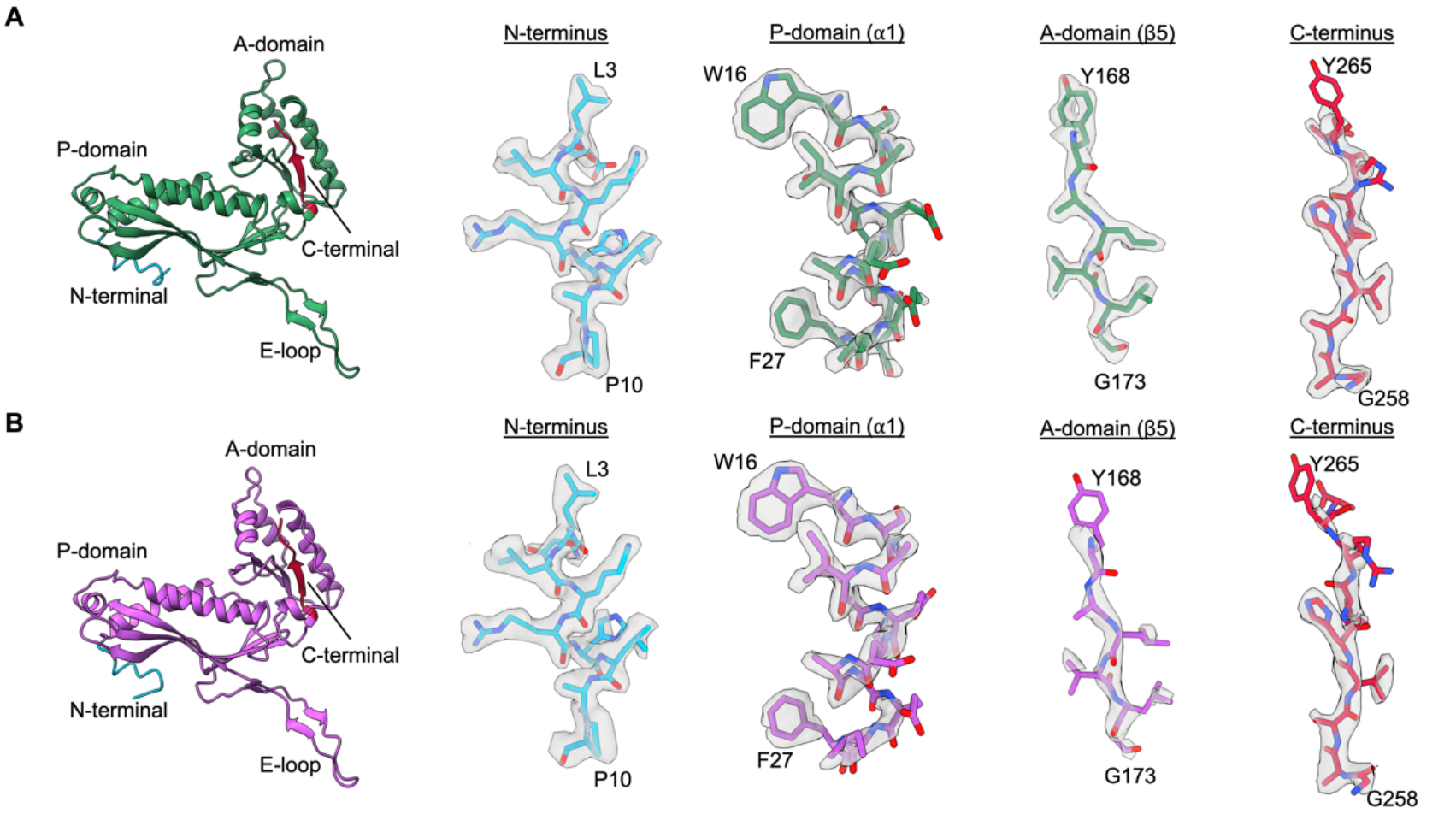

Figure 4 – figure supplement 2: Model fit of the 'open' and 'closed' atomic models produced from open and closed maps

The atomic models of the monomer subunit of the 'closed' (green, **A**) and 'open' (purple, **B**) pentamer conformations. Representative maps are shown across the monomer chain illustrating the fit of the atomic model and side chains into the cryoEM map.

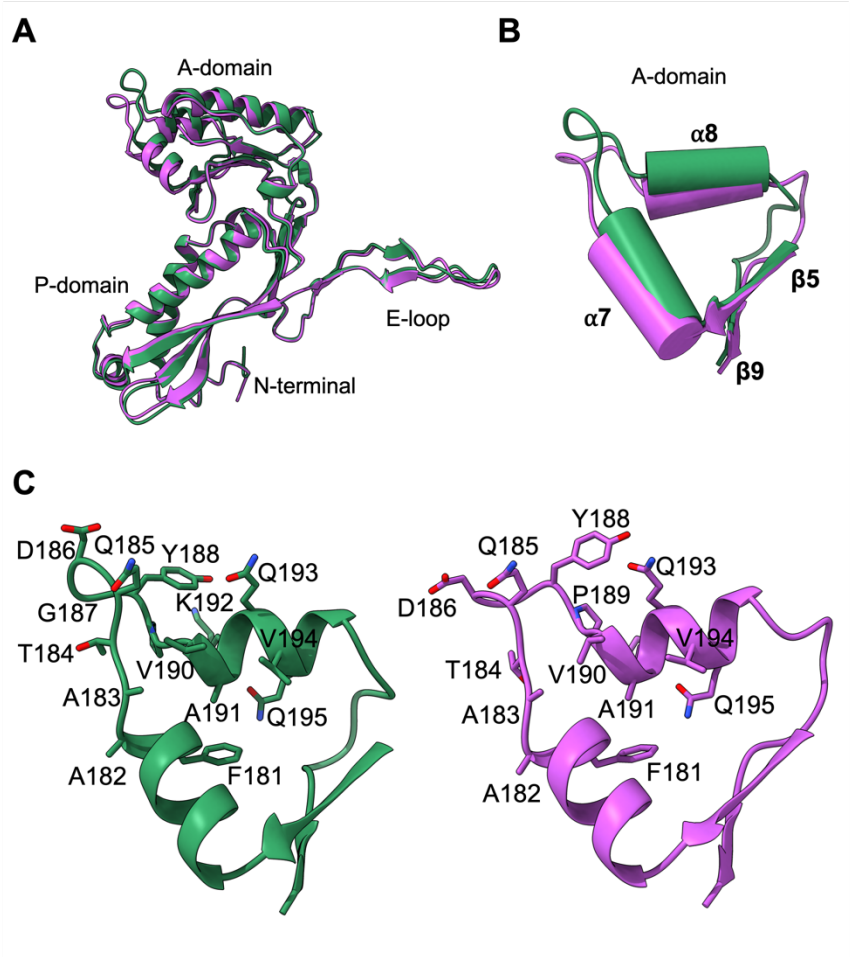

Figure 4 – figure supplement 3: Comparison of the A-domain of 'closed' and 'open' pentamer structures

A: Overlay of the monomers from the 'closed' (green) and 'open' (purple) pentamer conformations displaying shifted A-domains. **B**: Overlay of the A-domain of the 'closed' (green) and 'open' (purple) conformations highlighting the change in alpha helices 7 and 8. **C**: The residues present at the 5-fold pore and in alpha helices 7 and 8.

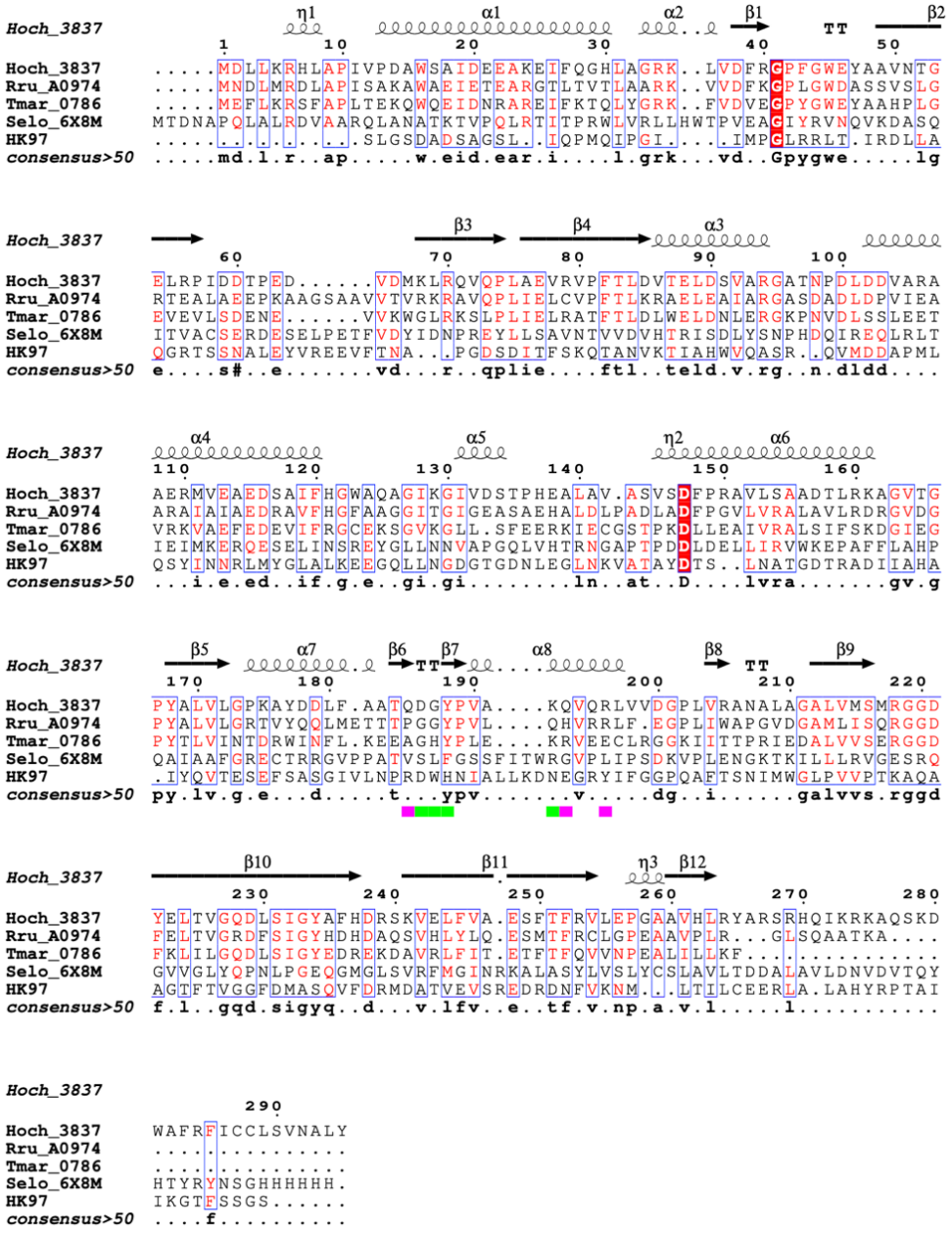

Figure 4 – figure supplement 4: Sequence alignment of encapsulins and HK97.

Protein sequence alignment of encapsins from the *H. ochraceum* (Hoch_3837), *R. rubrum* (Rru_A0974), *T. maritima* (Tmar_0786) and *S. elongatus* (Selo_6X8M). Encapsulins share the HK97-fold from the HK97 bacteriophage capsid which is also shown in the sequence alignment (HK97). The residues in the 5-fold pore in the 'closed' pentamer conformation are underlined in purple. The additional residues which become exposed and form the 5-fold pore in the 'open' conformation are underlined in green. Protein sequences were sourced from uniprot and the alignment was performed with Clustal Omega, and then formatted using ESPript.

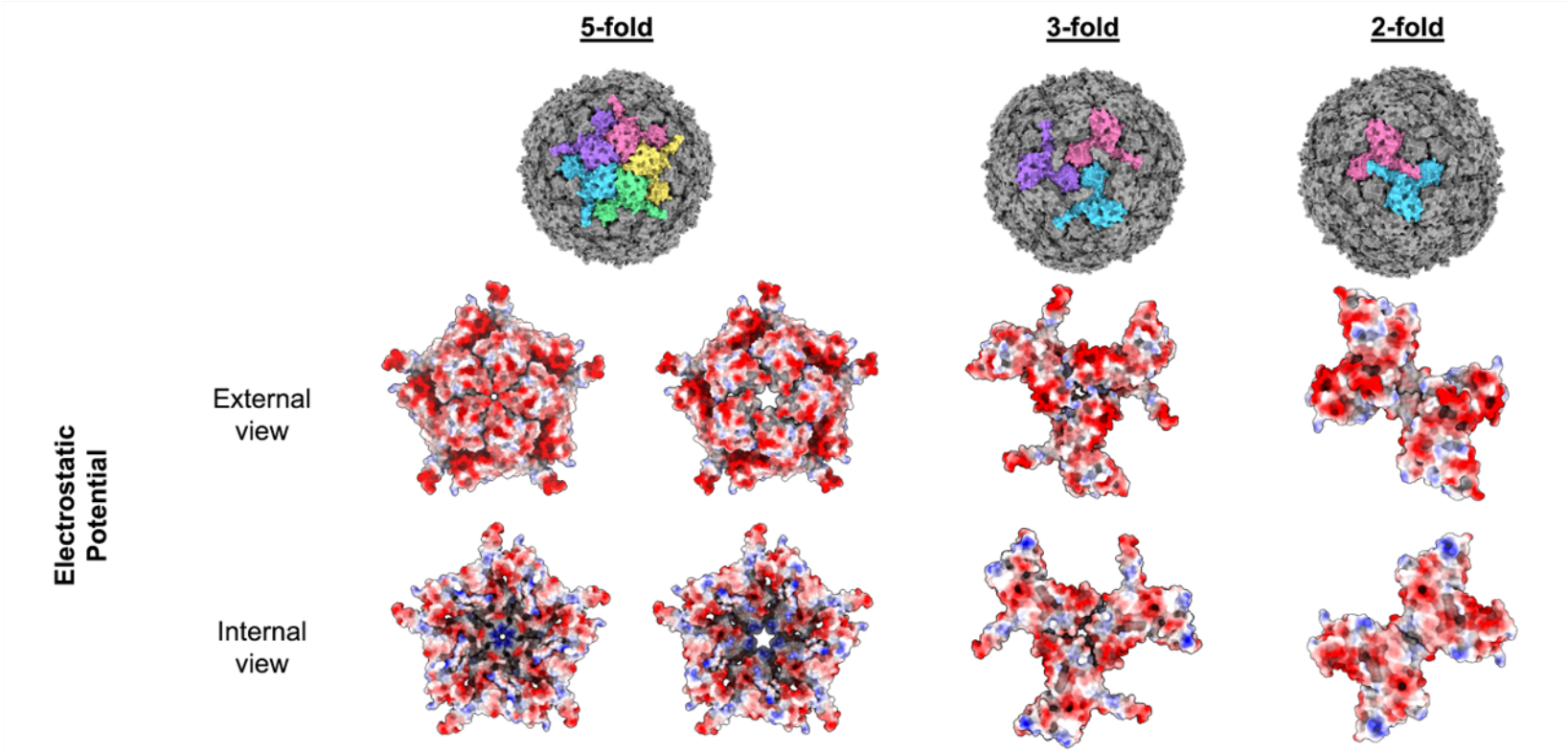

Figure 4 – figure supplement 5: Electrostatic properties of the encapsulin nanocompartment pores.
From top to bottom: The top row shows a T=1 encapsulin in three orientations to show the 5-fold, 3-fold and 2-fold symmetry axes. Monomers have been coloured individually (pink, yellow, green, blue and purple) to highlight the symmetry axes. The second and third row show the 5-fold (closed and open conformations), 3-fold and 2-fold Loaded-Enc symmetry axes coloured with electrostatic surfaces (positive charges shown in red and negative in blue).

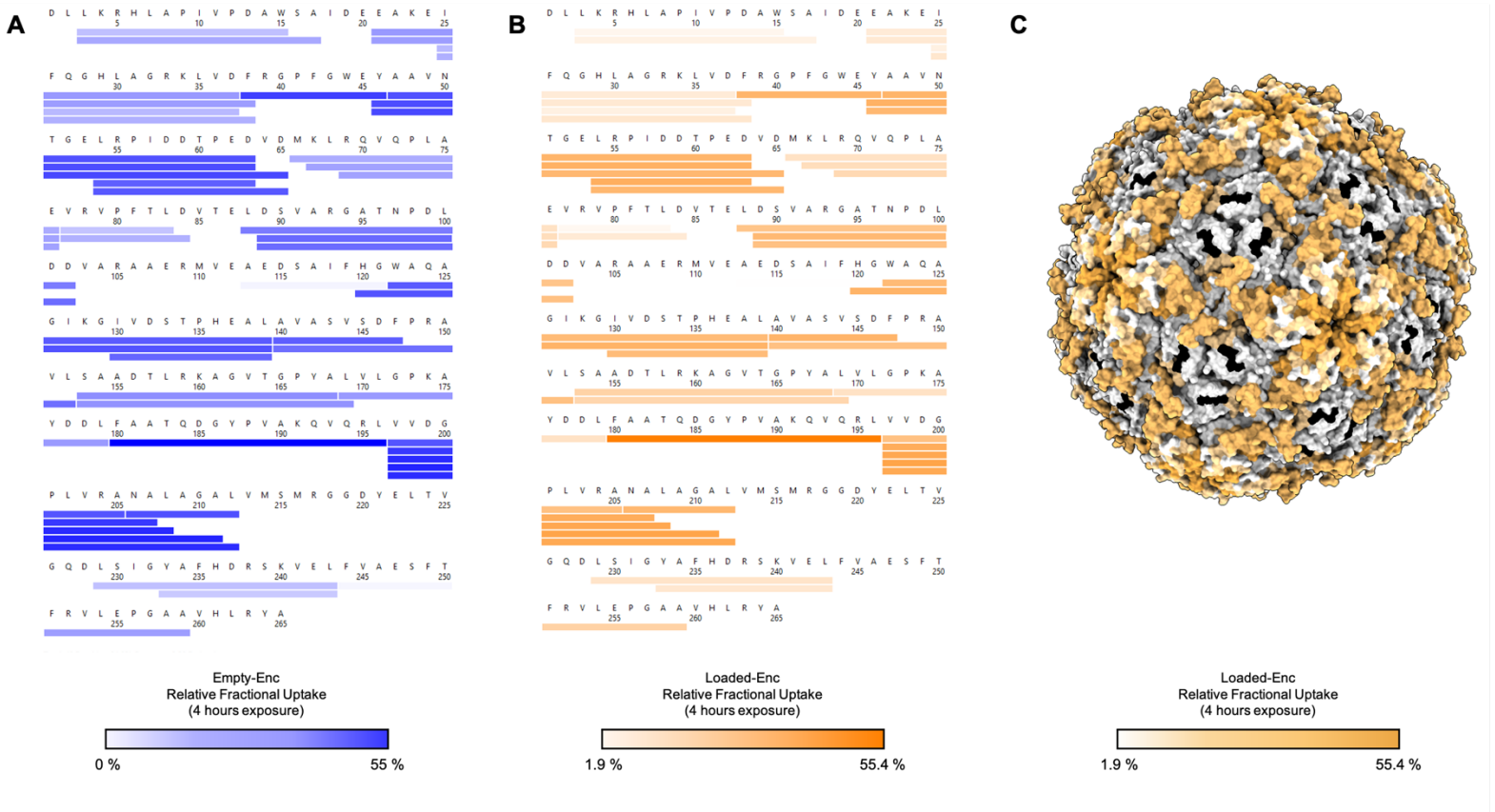

Figure 6 – figure supplement 1: Sequence coverage and deuterium uptake after HDX MS analysis of Empty-Enc and Loaded-Enc peptides. HDX coverage maps showing the relative fractional deuterium uptake for both Empty-Enc (**A**) and Loaded-Enc (**B**) after 4 hours of deuterium labelling. There are 40 observed peptides common to both states, providing 84 % protein sequence coverage with 2.35 redundancy. Color keys are shown under **A** and **B**. (**C**) The relative fractional uptake of Loaded-Enc displayed on the encapsulin nanocompartment. Areas colored black are representative of no sequence coverage.

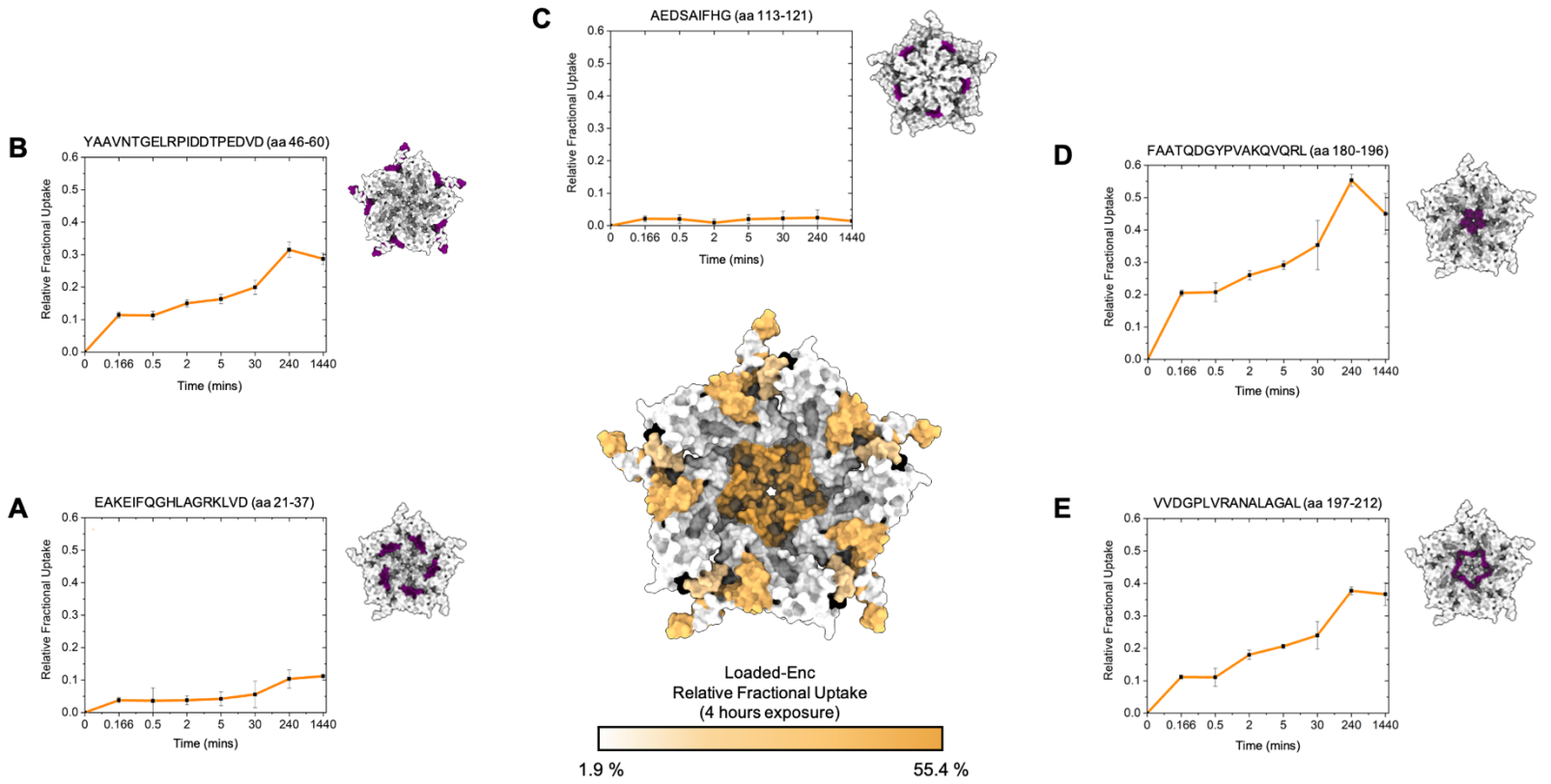

Figure 6 – figure supplement 2: Deuterium fractional uptake of Loaded-Enc peptides.
HDX-MS of the encapsulin 5-fold pore highlighting the amount of deuterium incorporation after 4 hours of exposure, colored according to uptake (a color key is shown at the bottom of the central figure). (**A**) to (**E**), Uptake plots for individual peptides showing the relative deuterium uptake over time. An encapsulin pentamer is shown with the corresponding peptide sequence colored purple. (**A**) residues 21-37, whose proximity is close to the proposed localization sequence for EncFtn (11.2 ± 0.2 % uptake at 4 hours). (**B**) residues 46-60, which overlays with the proposed 2-fold symmetry pore (31.7 ± 2.2 % uptake at 4 hours), (**C**) residues 113-121, a highly protected exterior peptide (0.2 ± 0 % at 4 hours), (**D**)

residues 180-196 the potential 5-fold symmetry pore (55.4 ± 2.2 % at 4 hours), (**E**) interior residue 197-212 in close proximity to the potential 5-fold symmetry pore site (37.7 ± 1.2 % at 4 hours).

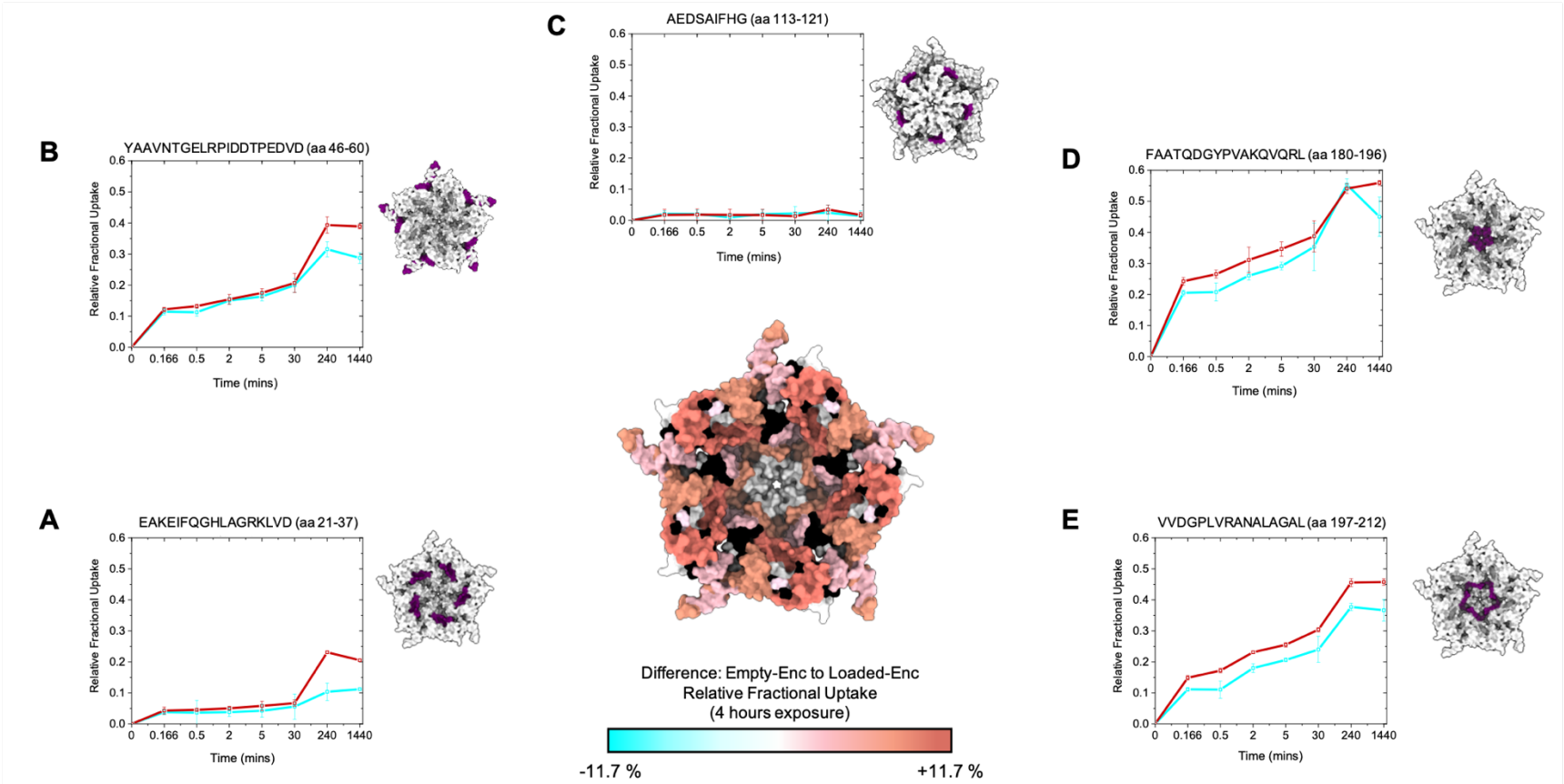

Figure 6 – figure supplement 3: Differential HDX-MS analysis of Empty-Enc and Loaded-Enc.
The amount of deuterium incorporation over time is shown for five representative peptides (**A** - **E**) for both Empty-Enc (r lines) and Loaded-Enc (cyan lines). An encapsulin pentamer is shown next to each uptake chart with the corresponding peptide colored in purple. The central figure shows the difference in exchange is displayed on the structure of the 5-fold pore of encapsulin, with a color key below and areas of no sequence coverage colored black. (**A**) Location of residues 21-37 difference of +11.7 ± 1.5 %, (**B**) residues 46-65 with a difference of +7.6 ± 1.8 % (**C**) Location of residues 113-121 with a difference of 0.1 ± 0 %, (**D**) residues 180-196 with a difference of 0 ± 0.3 %, (**E**) interior residue 197-212 with a difference of 7.9 ± 1.7 %.