The 2.7 Å resolution cryo-EM reconstruction of *Mycobacterium tuberculosis* encapsulin nanocompartment containing DyP peroxidase.

Rhys Willmore

University of Cape Town

Thesis submitted in fulfilment of the requirements for the degree of Master of Medical Science in Medical Biochemistry in the Department of Integrative Biomedical Sciences.

02/08/2023

Supervisor: Jeremy David Woodward

Table of contents

DECLARATION

I, Rhys Willmore, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 02/08/2023

Acknowledgements

The process of completing this thesis has been the most difficult undertaking of my life and many people have helped me so much during the course of it. First and foremost, I’d like to thank my late father who unfortunately did not get to see me finally submit this thesis but always encouraged me to keep going even though he didn’t always fully grasp my work. It is for him that I soldiered through and kept working. I’d then like to thank my mom as she was always ready to give the best advice and to give me a jumpstart when necessary. And most crucially of all, my partner who has lived through my highs and lows over the course of this thesis. I would also like to thank my supervisor, Jeremy Woodward, for his insights and for being incredibly understanding throughout this project. Even going so far as to travel to Stellenbosch with me to carry out crucial lab work.

I would like to thank everyone at the Electron Microscope Unit at U.C.T. who has helped me at some point in my work. And although she is no longer a part of the EMU, I would like to thank Kaylene Baron for being incredibly helpful during the first two years of my work.

For the native work, I would like to thank Digby Warner for allowing me to use the facilities of the MMRU and Veneshley Samuels for guiding me through this unfamiliar work despite having her own research to carry out.

We acknowledge Diamond for access and support of the Cryo-EM facilities at the UK national electron bio-imaging centre (eBIC), (proposal bi24039-10), funded by the Wellcome Trust, MRC, and BBSRC. Dr Alistair Siebert for assistance with data collection.

This work was funded by GCRF-START (UK Science and Technology Facilities Council grant ref. ST/R002754/1).

Abbreviations

ABTS – 2,2′-azinobis (3-ethylbenzo-6-thiazolinesulfonic acid).

A-domain – Axial domain.

AU – Absorbance units.

B. linens – Brevibacterium linens.

BSA – Bovine serum albumen.

CTF – Contrast transfer function.

DDS – Drug delivery system.

DyP – Dye-decolourising peroxidase.

E. coli – Escherichia coli.

E-loop – Elongated loop.

EM – Electron microscopy.

EMDB – Electron Microscopy Data Bank.

Enc – encapsulin

FLP – Ferritin-like proteins.

FolB - 7,8-dihydroneopterin aldolase.

H. ochraceum – Haliangium ochraceum.

IMAC – Immobilized metal affinity chromatography.

IMEF - Iron-mineralizing Enc-associated firmicute.

IPTG – Isopropyl β- d-1-thiogalactopyranoside.

LB – Luria-Bertani.

M. smegmatis – Mycobacterium smegmatis.

M. tuberculosis – Mycobacterium tuberculosis.

M. xanthus – Myxococcus xanthus.

MCS – Multiple cloning site.

MS – Mass spectrometry.

P. furiosus – Pyrococcus furiosus.

PDB – Protein Data Bank.

P-domain – Peripheral domain.

Q. thermotolerans – Quasibacillus thermotolerans.

R. jostii – Rhodococcus jostii.

R. rubrum – Rhodospirillum rubrum.

ROS – Reactive oxygen species.

S. elongatus – Synechococcus elongatus.

SDS-PAGE – Sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

SEC – Size exclusion chromatography.

T. maritima – Thermotoga maritima.

TP – Targeting peptide.

Abstract.

Mycobacterium tuberculosis has evolved many persistence factors in response to the host generated immune response as a means of survival. One such immune response generated by humans is the use of reactive oxygen species (ROS) such as H_2O_2 to cause damage to *M. tuberculosis*. Encapsulin (from here on referred to as Enc) nanocompartments and the cargo proteins within have been implicated as persistence factors and decreased viability of cells has been shown when they are knocked out. Previous research has found that dye-decolorizing peroxidase (DyP) is encapsulated by these Enc nanocompartments, as has been shown in *Mycobacterium smegmatis.* This is done by way of a C-terminal targeting peptide, with Enc and DyP also being part of the same operon in the genome of *M. tuberculosis*. However, not much is understood about the structure and function of this system. Both encapsulin and DyP were expressed and purified recombinantly in *E. coli*. Cryo- EM particle processing and EM map reconstruction was carried in an attempt to generate a density map of encapsulated DyP along with model building. Native expression and purification were carried out followed by negative-stain EM on this sample.Successful recombinant expression and purification of DyP and Enc was achieved, with a high-resolution 2.7 Å cryo-EM structure of the Enc nanocompartment obtained, but no encapsulated DyP was visualized. A model of both the Enc monomer and multimer was built, with comparisons to *M. smegmatis* in charge around the fivefold pore showing differences. The purification of Enc from *M. tuberculosis* was successful and a negative stain reconstruction of the nanocompartment was obtained. 2D classes showed what could have been DyP but it did not show up in any 3D models. Ferritin was shown to only be outside of the nanocompartment. The high-resolution map showed high similarity to other T=1 Enc nanocompartments. The multimer model built called into question the exact function of this nanocompartment as the charge distribution differed from closely related *M.* smegmatis. When these charge changes are correlated to findings of catalytic analysis done in the same lab, it indicates that more research is needed to understand the exact function of this system in *M. tuberculosis*. The 2D classes showed that DyP is the only cargo protein that appears to be present in *M. tuberculosis* nanocompartments, with ferritin being identified outside the nanocompartment.

Introduction

M. tuberculosis research importance

Mycobacterium tuberculosis is the infectious agent responsible for the disease, tuberculosis (TB). As per the WHO Global Tuberculosis Report in 2022, it was estimated that 10.6 million people were ill with TB in the previous year with 1.4 million deaths in HIV negative people (1). It has been found that the highest cause of HIV-related mortality is due to infection with TB (2). This is due the immunodeficiency caused by HIV infection increasing how quickly the TB infection progresses from latent to active disease (3). The occurrence of TB within South Africa is thought to be 852 cases per 100 000 population and puts South Africa on the list of the 30 countries with the highest burden of TB compiled by the WHO (4). This means that research into new methods of combating TB is vital.

Persistence factors and the reaction by the body

M. tuberculosis has many persistence factors that allow it to evade the human immune system and medical treatments. The first line of defence that the body utilises is the initial ingestion of the bacteria by alveolar macrophages (5). Alveolar dendritic cells as well as monocyte-derived macrophages then also join this process of phagocytosis (6,7). The phagosome and lysosome fuse within the macrophage, with the bacteria within being exposed to a number of different stresses such as a lowered pH and contact with reactive oxygen species (ROS) (8). The bacteria are able to evade autophagy by employing various strategies. The production of CpsA carries out inhibition of NADPH oxidase recruitment to the infected phagosomes, which favours growth of the bacteria by reducing ROS production (9). *M. tuberculosis* is also able to inhibit the maturation and acidification of the phagolysome (10). The first of those is accomplished by hindering maturation of both the phagosome (11) and lysosome (12). The bacteria can also prevent the two from fusing (13). The second of these is accomplished in a number of ways, but a well-studied one is mediated by the ptkA protein and leads to a less detrimental environment for the mycobacteria (14). Upon phagocytosis, macrophages carry out a respiratory burst and this generates reactive oxygen intermediates (10). Two examples of these intermediates are H_2O_2 and NO. A number of genes have increased expression when both those ROS are present such as KatG and TrxB2 (15).

CFP-29

CFP-29 (each monomeric unit has a molecular weight of 29 kDa) was found in exosomes infected by *M. tuberculosis* (16). CFP-29 was initially compared to the M18 Linocin protein found within *Brevibacterium linens,* and was found to have a 58% amino acid identity and a similar molecular mass (16). Linocin forms a high molecular mass polymeric structure (17) elutes near the void volume of a Superose 6 gel filtration column indicated that CFP29 forms into a complex that was at least 700 kDa in size (16). It is postulated that the genes for Enc expression are a remnant left from bacteriophage infection, in particular the order *Caudovirales* (18). This may have led to Enc systems being present in at least 15 bacteria that encapsulate a variety of cargo proteins (19). Many seem to play a role in the stress response of these bacteria (20) when they are attempting to adapt to stationary phase. Within *Myxococcus xanthus,* enhanced expression of the *EncA* gene is induced upon starvation conditions and monomers are assembled into iron storage vessels while mutants that had mutations of the *EncA* gene were more sensitive to oxidative stress (21). In the case of *M. tuberculosis*, it was identified as an antigen that induced a strong T-cell response in both mice and humans with TB (22). During a stress test coupled with transposon screening, the expression of this protein was crucial to the survival of *M. tuberculosis* when exposed to H_2O_2 at a low concentration and this led to the conclusion that CFP29 may be crucial for survival of *M. tuberculosis* inside the phagosome (23).

Enc has also been detected in different places: in the culture filtrate and membrane of *M. tuberculosis* cultures (16), as well as the cell membrane and the cytoplasm of *Mycobacterium smegmatis* (24,25). *M. tuberculosis* Enc has been reported to be membrane enriched with a small amount found in the culture filtrate (16). In the case of *R. jostii*, it was found in the bacterial culture filtrate supernatants (26) despite the lack of a signal peptide. The presence of Enc in the culture filtrate could be due to secretion systems present in many bacteria that utilise a secretion signal when a protein lacks a signal peptide (27).

Enc nanocompartment.

Enc nanocompartments are shell-like protein compartments that are icosahedral in shape with a virus capsid-like structure (28). They are thought to be an example of compartmentalization within the cell and serve a similar purpose to larger compartments such as mitochondria and lysosomes (20). The purpose of this compartmentalization is to allow detoxification, catalytic, and storage functions which would otherwise be incompatible to exist within the same cell (29). They are composed of a monomeric unit that self-assemble, with the structure of the monomers determining the overall architecture and triangulation number (T) (29).

Structure of Enc nanocompartment: monomer,

The field of research regarding cargo loaded Enc nanocompartments is a rapidly growing one but there are many Enc systems that have not been characterized fully (30), with an example being the one found in *M. tuberculosis*. This is where the widespread nature of Enc amongst different bacteria becomes important. Enc in *M. tuberculosis* is proposed to be similar to the Enc found within *T. maritima* (28) and *M. smegmatis* (24), with the Enc found within *M. smegmatis* having an amino acid sequence identity of 83% and thus making it an ideal research surrogate. There are numerous reasons for comparison between Enc of *M. tuberculosis* and *M. smegmatis.* The genetic similarity between the two make it the best candidate for comparison and inference of characteristics. The similarities continue with both *M. smegmatis* living in the soil (31) and *M. tuberculosis* being able to survive in soil for at least 12 months and maintain its pathogenicity (32). It has also been found that *M. smegmatis can* infect mouse macrophages such as the J774A.1 macrophage (33). These similarities make comparisons in structure and function between the Enc systems important as *M. smegmatis* Enc and DyP are better characterised at present.

While there may be variation in the sequence and secondary structures when comparing Enc's, there are always three conserved domains present. These are similar to the capsomere gp5 found within the HK97 bacteriophage (34). The first region is the peripheral domain (P), which forms most of the monomer and is usually made up of two long and two short α-helices as well as three antiparallel β-sheets that make up the N-terminal region (35). These regions of the P-domain house a hydrophobic region that is site of cargo protein binding (28). The second is the axial domain (A) which has three α-helical segments next to five parallel and antiparallel β-sheets which surround the C-terminus. This domain makes up a five-fold symmetry axis by interacting solely with the A domains of other monomers (35). The third region is an elongated loop (E) composed of two-antiparallel β-strands and a region that is unordered. The E-loop is where the two-fold symmetry pore is formed between two dimers. Due to the unordered region, the E loop is the most flexible with this domain having a prominent role in assembly as well as determining the triangulation number (T) and topology of Enc nanocompartment multimers (34,36).

Structure of Enc: multimer.

The flexible and unordered region of the E-loop has an important role in determining the triangulation number of any Enc nanocompartment by the affecting the number of structural units that make up a face of the icosahedral superstructure, otherwise known as the T number. Within Enc nanocmpartments, there are three classes of icosahedral Enc architecture with different T numbers. The first class (T=1) has a diameter of 20-24 nm and are made up of 60 monomers. These include *T. maritima* (28), *R. jostii* (37) and most recently discovered *M. smegmatis* (Fig. 1B) (24). The Enc nanocomaprtment found within *M. tuberculosis* will most likely be part of the T=1 due to the sequence similarity to *M. smegmatis* mentioned previously and due to the diameter of the complex being ~220 Å or 22 nm (38). The other classes include a 30-32 nm diameter, 180-mer with T=3 symmetry present in *Pyrococcus furiosus* (39) and *M. xanthus* (Fig. 1B) (21) as well as one found within *Q. thermotolerans* which forms a 42 nm, 240-mer with T=4 symmetry (Fig. 1B) (36). In all the abovementioned classes there is a persistent icosahedral structure which has 12 pentameric vertices that align to six five-fold symmetry axes (29).

Nanocompartment pores.

At each of these six five-fold symmetry axes, a pore of the nanocompartment is formed. This pore is termed the pentameric pore, with pores also forming at the three-fold axes of symmetry along with a third type of pore forming at the interaction between two subunits (Fig. 1A) (28) with the major exception being found in *Quasibacillus thermotolerans*, which does not possess a clear two-fold pore (Fig. 1B)

(36). In *R. jostii* and *M. smegmatis* Enc nanocompartments, these pores form at protomer junctions and have been postulated to selectively allow substrates into the nanocompartment (24,37).*R. jostii* Enc nanocompartments have a diameter of ~ 5 Å with the majority of substrates being too large to pass through the pore (37) (e.g. guaiacol has a diameter of 6.68 Å). *M. smegmatis* has a larger pore diameter of 6.8 Å which would in theory allow a larger range of substrates to enter the nanocompartment (24). However, there is no structural evidence to show if the five-fold pore is selective or if it opens in these two bacteria. But out of the three pore types, they are thought to be the most likely to play a role in the passage of substrates into the nanocompartment as they are the largest in diameter of the different pores (28,40). Along with the size of the five-fold pores, the hydrophobicity and charge of the amino acids that make up the five-fold pores are thought to allow molecules with a certain charge into the nanocompartment, such as the five-fold pore observed in *T. maritima* which seems to alter the activity of the ferritin-like protein and the permeability of the nanocompartment shell (28). There is only one example to date of structures showing the five-fold pore in both the open and the closed conformation. Ross *et al.* (41) showed that the five-fold pore diameter had variability within the same cryo-EM dataset whilst studying the Enc of *H. ochraceum*. When looking at work done on the Enc of *M.* tuberculosis, Contreras *et al.* (38) stated that the Enc nanocompartment within *M. tuberculosis* has positively charged pores that may play a role selectively allowing molecules into the Enc nanocompartment. These proposed mechanisms of selectivity regarding the five-fold pore have not been proven structurally with the only results obtained coming from catalytic tests of the encapsulated cargo proteins (28).

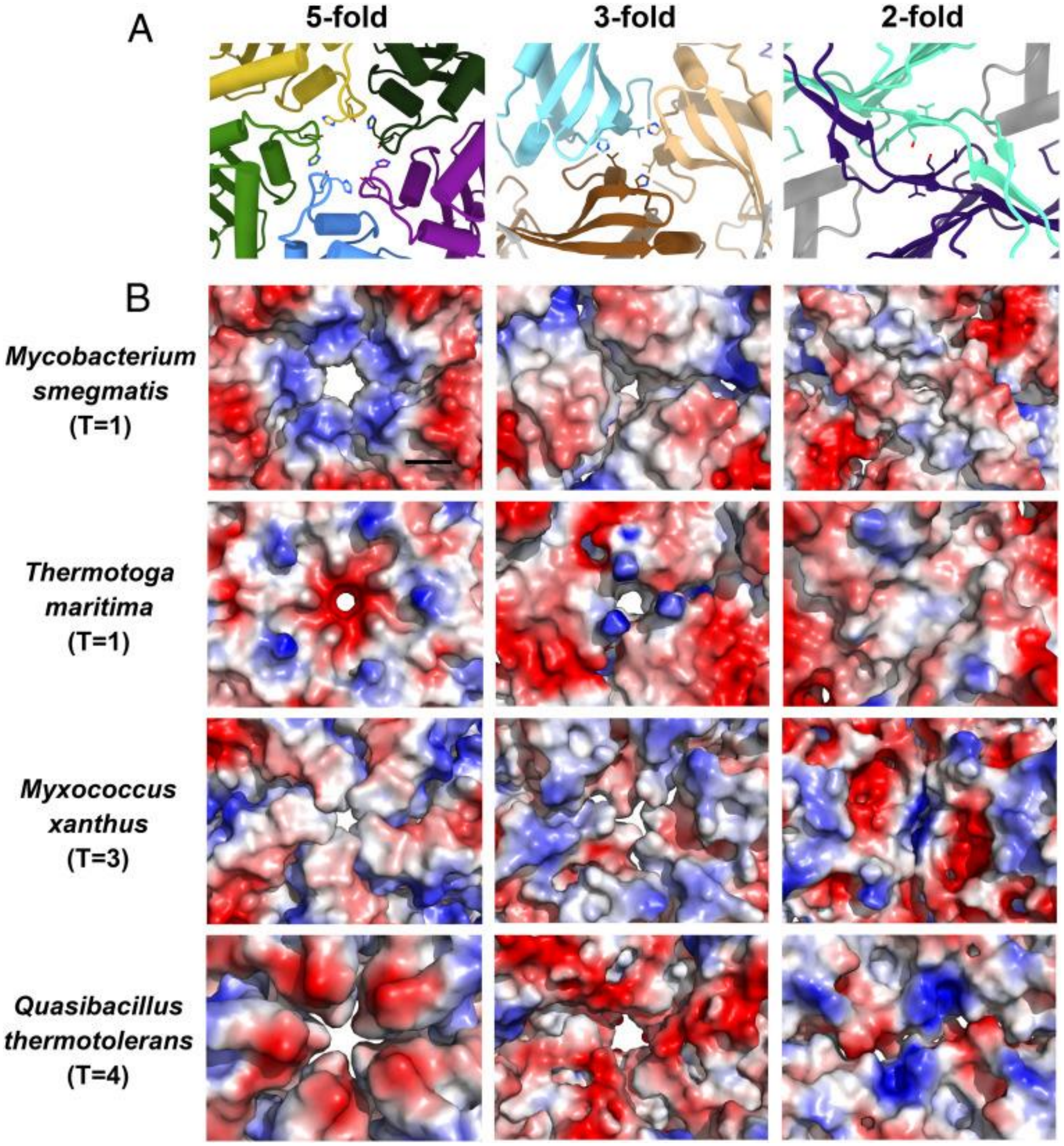


Figure 1: Figure reprinted from (24) **The pores on the encapsulin capsids. (*A*) Cartoon representation of the pores down the fivefold, threefold, and twofold symmetry axes on the encapsulin shell of *M. smegmatis*. (*B*) Electrostatic surface representation of the pores down the fivefold, threefold, and twofold symmetry axes on the encapsulin capsids from *M. smegmatis*, *T. maritima* (PDB 3DKT), *M. xanthus* (PDB 4PT2), and *Q. thermotolerans* (PDB 6NJ8). (The black scale bar is 10 Å.)**

Cargo proteins

The presence of enzymes within compartments is a form of cellular organization called compartmentalization (42). There are a number of Enc systems that have been studied at this point in time. There are encapsulated DyP systems in both *Rhodococcus jostii* and *Mycobacterium smegmatis* with both displaying different catalytic activity levels with certain substrates (24,37). There are the ferritin-like proteins present within both *Thermotoga maritima* and *M. xanthus* which are involved in iron mineralization and storage (21,28). The exact amount and arrangement of encapsulated enzymes on the inside is not clear in all cases, but examples have been shown *in vivo* that two hexameric DyP enzymes can be incorporated into one nanocompartment in *M. smegmatis* (24) while 12 BfrB enzymes can fit into the lumen of the nanocompartment found within *R. rubrum* (20). While there is genetic similarity between *M. smegmatis* and *M. tuberculosis,* it has not been proven that they share a similar amount or conformation of DyP inside of their respective nanocompartments. These cargo proteins have a variety of functions that range from catabolism of certain reactions (37) to antioxidant defence (43,44).

Potential cargo proteins: DyP-type peroxidase

The generally accepted function of peroxidases has been that they catalyse the oxidation of a number of organic and inorganic substrates by breaking down H_2O_2 and other peroxides (45). DyP-type peroxidases are heme peroxidases but differ from other heme peroxidases by way of a change at the heme distal region where a histidine residue has been changed to an aspartic acid residue which changes its reaction mechanism (46). There are three currently defined classes of DyP-type peroxidases, those being classes P (Primitive), I (Intermediate) and V (adVanced) (43). This classification scheme is based upon structure-based sequence alignments which resulted in the grouping of DyP-type peroxidases by tertiary structure (43).

In relation to Enc nanocompartments, a lignin peroxidase DyPB found within *R. jostii* (47) was shown to be encapsulated and to have increased catalytic activity whilst degrading lignin when compared to the unencapsulated form (37). This was a surprising finding as the pore size is 5 Å in diameter which should prevent lignin oxidation from occurring (37) and indicates that DyP could function differently whilst encapsulated. The presence of an encapsulated DyP within *M. tuberculosis* has been proposed as both Enc and DyP are part of a two gene operon (38), with DyP containing a C-terminal extension similar to the encapsulated DyP of *B. linens* (28). The presence of an encapsulated DyP to *M. tuberculosis* has been proposed as a response to ROS produced by macrophages (15). Stress tests were carried out where where *Escherichia coli* cultures recombinantly expressing *M. tuberculosis* encapsulated and unencapsulated DyP were exposed to H_2O_2, with the cultures that expressed the encapsulated DyP showing greater resistance to oxidative stress (23).

Reaction mechanism of DyP

In a study by Contreras (38), when DyP is purified, it has a pink colour which can be inferred as a bound heme molecule. Tests were then carried out for heme-dependent peroxidase activity, where a tetrameric holo-Mt-DyP formed a heme-dependent peroxidase intermediate. The study also tested peroxidase activity against guaiacol and ABTS as substrates. The study then showed that the apo-enzyme had no activity when compared to the holo-enzyme which shows that there is heme-dependent peroxidase activity. Not much has been confirmed *in vitro* about the reaction mechanism of DyP peroxidase but it is thought that there are three Compounds or States that the peroxidase is found in (Fig. 2) (48). The first is the Resting compound that has an Fe^{3+} at its core. A molecule of H_2O_2 is then cleaved and water is produced along with Compound I which now has an oxygen bound to the heme-iron with an unpaired electron present (49). This was also shown by monitoring the change in Soret absorption band, which broadened and had a decreased intensity once H_2O_2 was added and implies that it operates in a similar way to other peroxidases (46). This same research also showed that the Compound I form of that DyP was able to spontaneously return to the resting DyP after around 20 minutes (46). Compound I is responsible for the oxidation of a substrate and forms Compound II and an oxidized substrate which the bacterium can use. Compound II is then returned to the Resting compound by being reduced by a second substrate with a second oxidized substrate and a water molecule being produced (50). But there is some controversy around the presence of Compound II. A new reaction scheme was proposed where Compound I returns to resting by way of a two-electron reduction and that Compound II only forms when in alkaline and neutral pH whilst H_2O_2 is present (51). Compound II has not been shown at pH 3 yet, despite this being the optimal pH for dye decolourizing (52). A Compound II has also not been shown for either Class P or V DyP at low pH, of which the DyP found within *M. tuberculosis* is Class P. So, it is still unclear exactly how DyP peroxidase functions in terms of substrate use and optimal reaction conditions in all cases. The reaction mechanism of DyP-type peroxidases has been well studied and tested but the question of how it would function inside of a nanocompartment that has pore sizes that potentially limit the substrate availability remains unanswered. The previously mentioned size of the pore is an obvious limiting factor which would influence the availability of substrates for the progress of the reaction represented in Figure 1. This raises the question of whether the DyP would be able to carry out its reaction with only H_2O_2 present, in a manner similar to other catalase enzymes that can use a second molecule of H_2O_2 to return to the resting state (53). If this is the case, then the versatility of DyP makes it a vital target to interrupt.

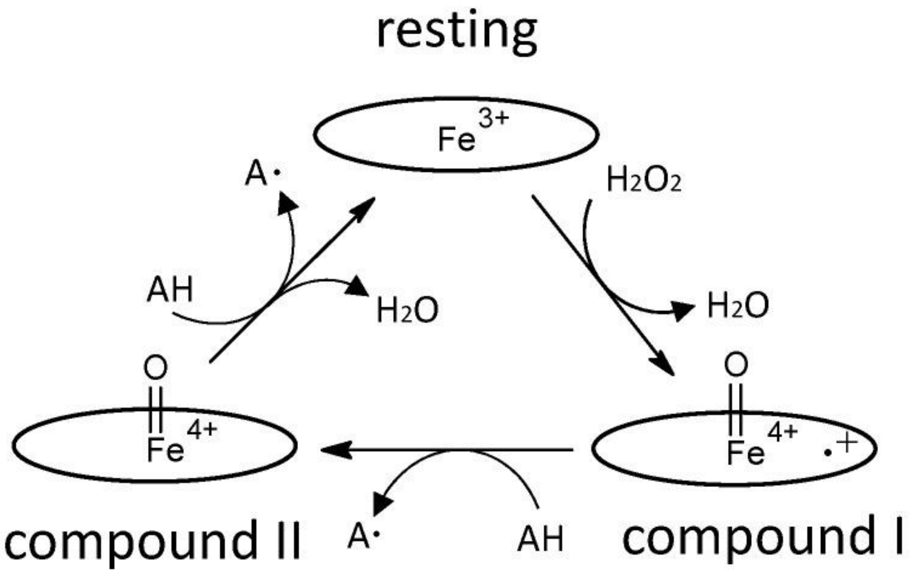


Figure 2. Reprinted from reference (48)**. Catalytic cycle of a typical heme peroxidase. Oval denotes the heme plane in the enzyme. AH is a substrate.**

The reaction mechanism in Figure 2 is the consensus of much of the literature regarding how DyP reacts with hydrogen peroxide. But in recent catalytic research done in Wiswedel (54), there was potential evidence that would indicate that a substrate is not necessarily needed for this reaction to occur. When the absorbance of the solution containing only DyP and hydrogen peroxide was monitored at 470 nm, an increase in the absorbance was observed. This is likely caused by the formation of bubbling and a cloudy precipitate within the reaction vessel. This coupled with the lower absorbance values seen when measuring the Soret peak shift (from 402 nm to 400 nm) could indicate that the amount of DyP absorbing light has been decreased.

Reaction mechanism of horseradish peroxidase.

When readily available, DyP will use a substrate to aid in breaking down harmful hydrogen peroxide as has been shown in many studies. But there is also thought that DyP could interact with H_2O_2 by itself with no substrate present because all it would need is an electron donor that serves a similar function to the previously mentioned substrate.When considering the possibility that DyP can catalyse the decomposition of H_2O_2 without the presence of a substrate, the reaction pathways of horseradish peroxidase become a likely alternative.

Horseradish peroxidase utilises three activity pathways described in Fig. 2 (55,56). The first is the catalase-like/pseudo catalase pathway that uses a second molecule of H_2O_2 as an electron donor to allow the reduction of Compound $I{\cdot}H_2O_2$ back to the native peroxidase enzyme while releasing O_2 and H_2O via the k_3 reaction step. The second pathway is the peroxidatic one and utilises substrates as electron donors to reduce peroxidase from Compound I back to the native peroxidase via the k_4 to k_7 reaction steps. The third pathway is the inactivation of the peroxidase and occurs when there is an excess amount of H_2O_2 and subsequent formation of oxygen free radicals and occurs via the k_i reaction step, with this inactivated version of the peroxidase increasing in concentration as the concentration of the H_2O_2 increases (56). These pathways all exist at the same time within solution with the first and second being protective pathways to prevent inactivation. The potential ability of DyP and particularly encapsulated DyP to be able to produce oxygen in the absence of an appropriate substrate in the hypoxic environment of a granuloma would be very beneficial to an aerobe. These results could give more information as to how DyP operates inside of Enc nanocompartments.

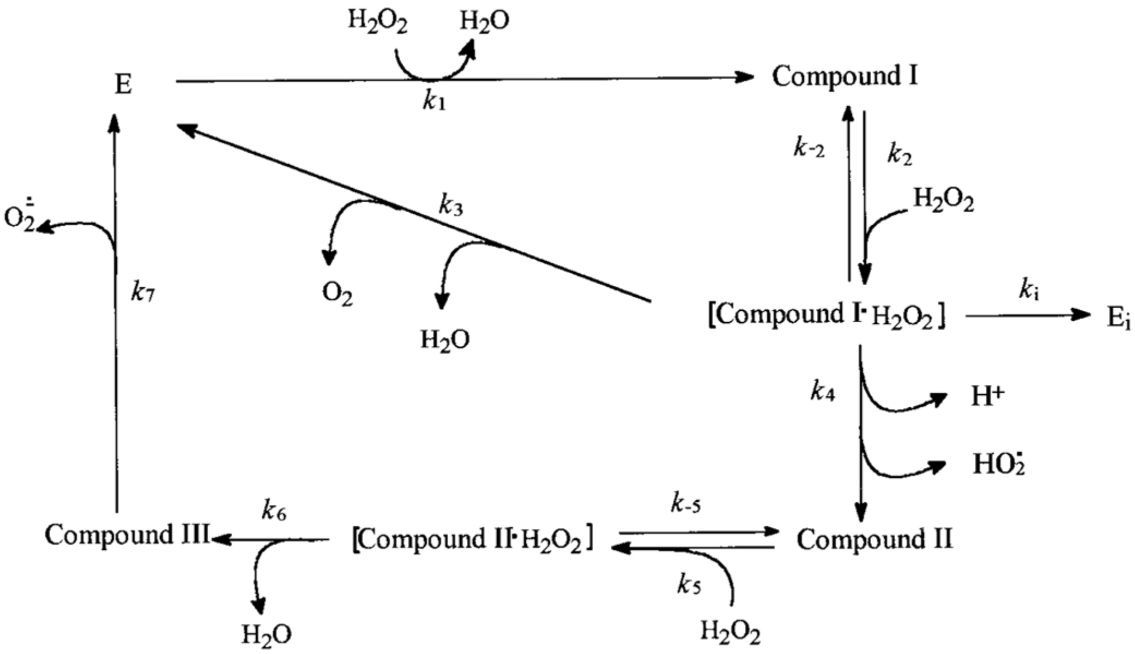


Figure 3. Reprinted from reference (56).

Model of the reaction of HRP-C with H_2O_2. E and Ei are the native and inactive enzyme, respectively; compound I, II, III are enzyme intermediates; compound I·H_2O_2 and compound II·H_2O_2 are complexes between the respective enzyme intermediates and H_2O_2. Principal features of the model are as follows. (1) The enzyme exhibits two catalytic cycles (catalase-like/peroxidase activities) with distinct rate constants and stoichiometries. (2) The complex compound I·H_2O_2 has a central role. The catalase-like (O_2 releasing) activity involves a direct two-electron reduction of H_2O_2. The peroxidatic (compound III) pathway proceeds stepwise in two single-electron reductions. The inactivation of HRP-C (to give Ei) is the third pathway from the complex. (3) The distribution of enzyme between the two catalytic pathways and one inactivating pathway results from two partition ratios: $r_c=k_3/k_i$ (catalase-like route/inactivation) and $r_{compound\ III}=k_4/k_i$ (compound III route/inactivation). The total value of the overall partition ratio (r) is: $r=r_c+r_{compound\ III}$ and corresponds to the number of turnovers performed by one mole of enzyme before its inactivation.

Potential cargo proteins: Ferritin-like protein (FLP).

The next potential cargo protein of Enc nanocompartments within *M.* tuberculosis are FLPs. Ferritins are utilised by a variety of organisms to ensure iron homeostasis by catalysing the conversion of Fe^{2+} into Fe^{3+} by utilizing O_2 or H_2O_2 and storing the resulting Fe^{3+} for controlled release with this process designed to prevent the reactive Fe^{2+} from interacting with any free radicals (57,58). The ferritins that are of particular interest are termed FLPs. Previously studied encapsulated FLP systems are found within *Rhodospirillum rubrum* (59), *T. maritima* (60), *M. xanthus* (21) and *Haliangium ochraceum* (41). The ferritin found within *R. rubrum* was shown to not have an enclosed compartment in its structure allowing for iron storage in the way of traditional ferritins which led to suggestions that this particular ferritin-like protein being encapsulated within a nanocompartment allows it to store significantly more iron (59). When the iron mineralization of the different forms of ferritin within *R. rubrum* were tested, the Enc and ferritin complex was able to sequester more iron ions than the apoferritin that was not in complex with Enc (59). Structural analysis of this ferritin showed that the ferritin was different to classical ferritin cages and could only store iron when in complex with an Enc nanocompartment (59) and was confirmed by McHugh *et al.* (21). The Enc found within *M. xanthus* actually contains a ferritin-like domain that allows the nanocompartment to mineralize iron without needing an encapsulated ferritin, whilst being able to store 30 000 (59) iron atoms. This is far larger when compare the 4 500 stored by ferritins (58).

The BfrB found within *M. tuberculosis* follows the conserved structure of the ferritin family which are made up of 24 subunits that form a complex with octahedral symmetry that enclose a hollow cavity which allows iron storage (61). This would indicate that unlike the ferritin-like protein found within *R. rubrum*, it seems to be able to store iron and would not need to be encapsulated within a nanocompartment to function. This BfrB is thought to mainly play a role in iron homeostasis by way of iron mineralization and storage (44) with a second function of this being the conversion of Fe (II) to Fe (III) to stop it from taking part in Fenton reactions. When there are low oxygen levels (62) and the cells are undergoing stationary phase (63), BfrB alongside BfrA expression is upregulated.

Potential cargo protein: aldolase.

Contreras *et al*. (38) demonstrated the cargo loading ability of 7,8-dihydroneopterin aldolase (FolB) from *M. tuberculosis* when expressed recombinantly in *E. coli*. This protein is part of the folate biosynthesis pathway found in *M. tuberculosis* and its specific function is to convert 7,8-dihydroneopterin to 6-hydroxymethyl-7,8-

dihydropterin and glycolaldehyde (64). However, FolB in *M. tuberculosis* is not part of a two gene operon with Enc and has not been observed as a cargo protein for Enc in any other bacteria so far, which makes it an unlikely candidate for encapsulation.

Antioxidant roles of potential cargo proteins.

Given the functions of both DyP and FLPs, it is possible that they both play antioxidant roles within *M.* tuberculosis. The ferritin-like proteins convert Fe (II) into stable Fe (III) (65) and DyP is responsible for detoxifying certain reactive oxygen species such as H_2O_2 as part of their proposed reaction mechanism. Chen *et al.* (52) describes how the resting version of the ferric DyP enzyme is oxidized by H_2O_2 into an active version. They then also went on to propose that this heme peroxidase family is responsible for catalysing H_2O_2 dependent oxidation of assorted molecules and is mostly present in bacteria with the DyP being studied in this thesis being present in *M. tuberculosis* (52). This reaction mechanism has, however, not been shown in vitro as far as DyP peroxidase in *M. tuberculosis* and thus the physiological role as well as the exact mechanism of action that is utilised by DyP peroxidase is unclear. Fe (II) catalyses the conversion of H_2O_2 into a hydroxyl free radical which is called Fenton's reaction and can lead to biological damage that is harmful to *M. tuberculosis* (66), so it would be beneficial to have both of the enzymes present to prevent cell damage. The presence of these free radicals is part of the defence mechanism employed by the human body whereby inflammatory cells generate ROS as a means of killing any infecting pathogen (67).

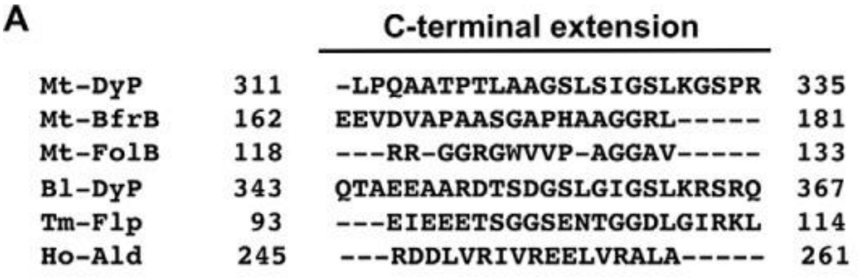
When *M. tuberculosis* is exposed to low doses of H_2O_2, a number of genes that are part of H_2O_2 scavenging and iron uptake are activated while much larger levels of H_2O_2 (greater than 50 mM) lead to majority cell death due to DNA damage with a few bacilli surviving (15). This could then supply a function or a way to determine the function of this protein in *M. tuberculosis*.

Interactions between the Enc and the cargo protein.

Cargo proteins are targeted to the inside of the Enc nanocompartment by a short C-terminal peptide sequence that is called a targeting peptide (TP) (68). This TP binds to a hydrophobic region of an interior conserved N-terminal helix present within the P-domain using hydrophobic and ionic interactions (29,68). This interaction between the TP and the hydrophobic region is difficult to visualise due to the flexibility of the TP as well as the variable occupancy of the encapsulated cargo proteins (30). The principle of the C-terminal TP has been demonstrated in *T. maritima* by Cassidy-Amstutz *et al.* (69) who added a C-terminal TP to green fluorescent protein and found that these proteins were encapsulated. The strength of the TP-binding pocket interaction can also be changed by mutating residues of the anchor motif with a complete change or removal of the TP causing a large reduction in encapsulation (68,69).

The main cargo protein of interest, DyP, has a C-terminal TP composed of 37 amino acids mainly made up of hydrophobic amino acids (28). Any DyP that is part of an operon with Enc are co-regulated to be transcribed at the same time (38). An example of one of these operons is the one found in *B. linens* where M18 linocin protein (which is an Enc) along with a DyP that has a C-terminal extension, is part of a two-gene operon (28). This Enc has a 58% amino acid identity to the Enc found within *M. tuberculosis* and was therefore proposed that these Encs are close homologs (28) The core TP anchor sequence motif of GSLXIGSLK (with X being any amino acid) is consistently present in all DyP that are part of two-gene operons with Enc, while the anchor motif is less conserved within bacterioferritin (28). The TP of BfrB does not contain this motif and the C-terminal extension is not very similar (Fig. 3). The amino acid sequence of the *M. tuberculosis* BfrB C-terminal TP is similar to the FLP found within *M. smegmatis* Bfr (61), and there are no recorded instances of encapsulated ferritin in *M. smegmatis*. While FLPs are present as part of a two-gene operon in *T. maritima* (28), this has not been shown in *M. tuberculosis*. The other way that BfrB could be encapsulated is as a secondary cargo protein, which are not close to the Enc operon (19).

To date, there is still no structural confirmation of any encapsulated proteins present within *M. tuberculosis.* Lien *et al.* (23) carried out mass spectrometry (MS) upon the nanocompartment fraction from a purified *M. tuberculosis* sample and confirmed the presence of DyP. There were no other proteins present in this MS analysis and the transmission electron microscope images showed occupied nanocompartments (23). To explore the ability of *M. tuberculosis* Enc to encapsulate various cargo enzymes, Contreras *et al.* (38) conducted experiments that tested the possibility of the three *M. tuberculosis* enzymes mentioned above. The sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gels in the study show the co-purification of these enzymes when this overexpression is carried out, with faint bands of each protein being present in the same sample as Enc. There are EM images showing examples of a small number of occupied nanocompartments but the overall number of filled nanocompartments for the BfrB and the FolB examples were not numerous while the DyP images showed more apparent occupancy. When the C-terminal extensions of the various enzymes are compared, they do not share much similarity in amino acid sequence (Fig. 3A), and it is unclear why these proteins would be encapsulated.



A

		C-terminal extension	
Mt-DyP	311	-LPQAATPTLAAGSLSIGSLKGSPR	335
Mt-BfrB	162	EEVDVAPAASGAPHAAGGRL-----	181
Mt-FolB	118	---RR-GGRGWVVP-AGGAV-----	133
Bl-DyP	343	QTAEEAARDTSDGSLGIGSLKRSRQ	367
Tm-Flp	93	---EIEEETSGGSENTGGDLGIRKL	114
Ho-Ald	245	---RDDLVRIVREELVRALA-----	261

Figure 4. Fig. 1A reprinted from reference (38)**. C-terminal extensions of potential Mt-Enc protein cargos. *A*, C-terminal extension residue sequences of Mt-DyP, Mt-BfrB, and Mt-FolB beyond what is seen in homologs that are not encapsulated as well as *B. linens* DyP (Bl-DyP), *T. maritima* ferritin-like protein (Tm-Flp), and a putative *H. ochraceum* aldolase (Ho-Ald) thought to be encapsulated by its C-terminal tail to BMCs.**

Pore dynamics.

The structure of Enc nanocompartments and the different cargo proteins have been well studied, but questions still remain about how the substrates of these cargo proteins enter nanocompartment. The overall charge and amino acid distribution of these pores seems to correspond with the substrates of the encapsulated proteins within. The fivefold pores of *T. maritima* have a mostly neutral charge with hydrophobic residues present (40) and a ferritin-like protein is contained within the nanocompartment (28). The entrance to this pore has a ring of histidine's which have an affinity for the primary substrate of the ferritin-like protein, Fe^{2+}, while the interior of the pore has negatively charged residues which form a pathway to the FLP (30). Another example of a charged five-fold pore is that of *Synechococcus elongatus*, which encapsulates a cysteine desulferase with both these proteins being upregulated in response to sulphur starvation (70). This five-fold pore has an overall positive charge and would most likely be allowing deprotonated cysteine into the lumen of the nanocompartment. In a study carried out by Rhamanpour and Bugg (37), it was shown that encapsulated DyPB present in *R. jostii* was able to use lignin as a substrate. According to this study, the pores are wide enough to handle the entrance of Fe (II) and hydrogen peroxide but not wide enough to allow dye 2,2′-azinobis (3-ethylbenzo-6-thiazolinesulfonic acid) (ABTS) into the nanoncompartment which is why it came as a surprise when they discovered that the activity of the unencapsulated DyPB using ABTS as its substrate was like that of encapsulated DyPB. This could indicate that opening of the pores to allow substrates into the complex is more dynamic than first thought and warrants more research. The positively charged residues around the fivefold pore of *M. smegmatis* has led to the theory that negatively charged aromatic substrates enter the nanocompartment and are oxidized by DyP (24). It has been postulated that the five-fold pore of the *M. tuberculosis* nanocompartment will be like the one found in *M. smegmatis* in terms of charge distribution. This has not been proven, but the high amino acid sequence similarity between the Enc pf *M. smegmatis* and *M. tuberculosis* (83%) makes it plausible that the structure of Enc could be conserved, however the charge distribution may not. This has not been confirmed and will be one of the focuses of this thesis.

Uses in biotechnology.

The importance of understanding the Enc systems present in many bacteria does not only include their importance as targets of therapeutics, but their uses in biotechnology. Drug Delivery Systems (DDS) have led to more efficient therapies and diagnostic tools to be used in the biomedical field (71). Van de Steen *et al.* (71) confirmed that it was possible to use Enc nanocompartments as a viable DDS when targeted to breast cancer cells. These DDS were recombinantly expressed *T. maritima* Enc with an antibody mimetic protein on the outside with a cytotoxic protein in the lumen of the Enc. When this was applied to breast cancer cells, there was an increase in apoptosis. Another potential use of Enc is the concept of delivering immunogenic cargos and also displaying the immunogenic epitope on the surface of the nanocompartment. Lagoutte *et al.* (72) showed that an ectodomain of a matrix protein from *Influenza A* virus H1N1 was needed to improve the correct assembly of nanocompartments loaded with a reporter protein so the cargo and shell engineering go hand in hand. The compartmentalization of catalytic processes can also be exploited to accelerate biochemical reaction kinetics (73). This is of particular interest when it comes to the production of antimicrobial peptides and silver nanoparticles which when produced in conjunction show better antimicrobial activity than commercial citrate-coated silver nanoparticles (74).

In this thesis, a high resolution cryo-Electron Microscopy (EM) structure of the *M. tuberculosis* Enc nanocompartment is obtained along with a negative stain reconstruction of native Enc and ferritin from *M. tuberculosis*. Purification pipelines for both recombinant expression in *E. coli* and native expression in *M. tuberculosis* will be obtained. The purification of native *M. tuberculosis* proteins will be of interest as the presence of a cell wall that contains mycolic acids, makes sample preparation for proteomic research difficult to carry out (75) which has led to fewer methods of purifying intact proteins (76) and the purification methods used will be informative for further studies. These findings will be contrasted to previous work and compared to the structure of the Enc found within other bacteria to gain more insight into the Enc within *M. tuberculosis*. This coupled with catalytic studies carried out by Wiswedel (54) will give a more complete understanding of the Enc and DyP present in *M. tuberculosis*.

Materials and methods

Plasmid construct and synthesis.

The genetic and plasmid constructs were carried out by Genscript.

Rv0798c was cloned into pET-28a(+) using the NcoI and XhoI cloning sites to produce Rv0798c_pET-28a(+). A mutagenesis was carried out on Rv0798c_pET-28a(+) where the stop codon at the end of Rv0798c was removed so that Enc would be produced with a C-terminal His-Tag attached to produce Rv0798c_pET-28a(+)-His tag.

Rv0799c was cloned into pET-28a(+) using the NcoI and XhoI cloning sites to produce Rv0799c_pET-28a(+). The Rv0799c_pET-28a(+) was used to subclone the Rv0799c into pETDuet-1 using the NdeI and XhoI cloning sites to produce pETDuet-1-Rv0799c-no His-tag.

The Rv0798c_pET-28a(+)-His tag was used to subclone the Rv0798c into pETDuet-1-Rv0799c-no his-tag using the NcoI and XhoI cloning sites to produce pETDuet-1-Rv0799c-no histag-Rv0798c-C-His-tag.

The inserts are as follows with the highlighted yellow sections being genomic sequence of each protein.

Gene: Rv0798c. Vector: pET-28a(+). Construct: Rv0798c_pET-28a(+)

Ccatgggcaacaatctctaccgcgatttggcaccggtcaccgaagccgcttgggcggaaatcgaattggaggcggcgcggacgttcaagcgacacatcgccgggcgccgggtggtcgatgtcagtgatcccgggggcccgtcaccgcggcggtcagcaccggccggctgatcgatgttaaggcaccaaccaacggcgtgatcgcccacctgcgggccaagcaaaccccttgtccggctacgggttccgtttaccctgtcgcgcaacgagatcgacgacgtggaacgtggctctaaggactccgattgggaaccggtaaaggaggcggccaagaagctggccttcgtcgaggaccgcacaatattcgaaggctacagcgccgcatcaatcgaagggatccgcagcgcgagttcgaacccggcgctgacgttgcccgaggatccccgtgaaatccctgatgtcatctcccaggcattgtccgaactgcggttggccggtgtggacggaccgtattcggtgttgctctctgctgacgtctacaccaaggttagcgagacttccgatcacggctatcccatccgtgagcatctgaaccggctggtggacggggacatcatttgggccccggccatcgacggcgcgttcgtgctgaccactcgaggcggcgacttcgacctacagctgggcaccgacgttgcaatcgggtacgccagccacgacacggacaccgtgcgcctctacctgcaggagacgctgacgttcctttgctacaccgccgaggcgtcggtcgcgctcagccactaactcgag

Gene: Rv0798c-no stop codon. Vector: pET-28a(+). Construct: Rv0798c_pET-28a(+)-His tag.

Ccatgggcaacaatctctaccgcgatttggcaccggtcaccgaagccgcttgggcggaaatcgaattggaggcggcgcggacgttcaagcgacacatcgccgggcgccgggtggtcgatgtcagtgatcccgggggcccgtcaccgcggcggtcagcaccggccggctgatcgatgttaaggcaccaaccaacggcgtgatcgcccacctgcgggccagcaaaccccttgtccggctacgggttccgtttaccctgtcgcgcaacgagatcgacgacgtggaacgtggctctaaggactccgattgggaaccggtaaaggaggcggccaagaagctggccttcgtcgaggaccgcacaatattcgaaggctacagcgccgcatcaatcgaagggatccgcagcgcgagttcgaacccggcgctgacgttgcccgaggatccccgtgaaatccctgatgtcatctcccaggcattgtccgaactgcggttggccggtgtggacggaccgtattcggtgttgctctctgctgacgtctacaccaaggttagcgagacttccgatcacggctatcccatccgtgagcatctgaaccggctggtggacggggacatcatttgggccccggccatcgacggcgcgttcgtgctgaccactcgaggcggcgacttcgacctacagctgggcaccgacgttgcaatcgggtacgccagccacgacacggacaccgtgcgcctctacctgcaggagacgctgacgttcctttgctacaccgccgaggcgtcggtcgcgctcagccacctcgag

Gene: Rv0799c. Vector: pET-28a(+). Construct: Rv0799c_pET-28a(+)

ccatggctgtgcctgctgtctctccgcagccgattctggcgccgttgactccggccgcgatcttcctggtcgccaccatcggcgccgacg
gcgaggcgacggtgcacgacgcactgtcgaagatctccggcctggtgcgtgcgatcggcttccgtgacccgaccaaacacttgtccgtgg
tcgtctcgatcgggtccgacgcctgggaccgcttgttcgccggacctcgacccaccgagctgcatccgttcgtcgagctgaccggaccac
ggcataccgccccggcaactcccggcgatctgttgttccacatccgggccgagaccatggacgtgtgcttcgagttggctggccgcattc
tcaagtcgatgggcgatgcggtgaccgttgtcgacgaagtgcacggcttccggtttttcgacaaccgtgatctgctgggctttgtcgatg
gcaccgaaaacccaagtggcccaatcgccatcaaggccaccacgatcggtgacgaggatcgaaatttcgccggctcgtgttatgtgcacg
tgcagaagtatgtgcacgatatggcctcgtgggaatccctctcggtcaccgagcaggagcgtgtgatcggccggaccaagctcgacgaca
tcgaactcgacgacaacgcgaagccagccaattcacacgtagcgctcaacgtcatcaccgacgacgacggcaccgagcgcaagatcgtgc
gacacaacatgccgttcggcgaagtcggcaagggcgagtacggcacttacttcatcggatattcgcgcacgccaacggtgaccgagcaga
tgctgcgcaacatgtttctcggcgatccggcaggcaacaccgatcgagtgcttgacttttccaccgcggtgaccggcggactgttcttct
cacccaccatcgactttctcgaccatccaccgcccctaccgcaggcggcgacgccaactctggcagccgggtcgctatcgatcggcagct
tgaaaggaagcccccgatgactcgag

Bacterial cell and culture preparation,

Competent cells were prepared in the following way. Luria-Bertani (LB)/agarose plates (1% (w/v) sodium chloride, 0.5% (w/v) yeast extract, 1%(w/v) tryptone, 1.5% (w/v) agarose) containing no antibiotics were prepared as well as Buffer A (100 mM MgCl2) and B (100 mM $CaCl_2$), a glycerol stock of DH5α *E. coli* was used to streak DH5α *E. coli* out onto the LB/agarose plates. The plates were sealed with parafilm (Bemis, Oshkosh, USA) and incubated at 37°C overnight. A single colony was isolated and used to inoculate a 5 ml LB media starter culture. The starter culture was incubated with shaking at 37°C overnight. 0.5 ml of the starter culture was used to inoculate a 50 ml LB culture. This culture was incubated with shaking at 37°C. Once the absorbance reading at 600 nm (OD600) reached 0.6 – 0.8, the culture was cooled on ice for 10 minutes, divided between two chilled 50 ml centrifugal tubes and pelleted by centrifugation at 5 000 rpm for 5 min at 4°C.

The two cell pellets were each resuspended with 50 ml of ice-cold buffer A prepared in milli-Q purified water, autoclaved, and cooled to 4°C and incubated on ice for an hour. The cells were pelleted by centrifugation (5 000 rpm, 5 min, 4°C) and each pellet was resuspended with 0.5 ml of ice-cold buffer B (100 mM CaCl2 (Sigma, Saint Louis, USA), 15% (w/v) glycerol, prepared in milli-Q purified water, autoclaved, and cooled to 4°C).

Transforming chemically competent cells.

Aliquots of competent cells were thawed whilst on ice. Once these were fully thawed, 5 µl of both ligation reactions and positive controls (25 ng of DNA) were mixed with the competent cells and were incubated for 15 minutes on ice. Heat shock at 42°C was performed for two minutes. The cells were then placed back on ice for two minutes. 1 ml of LB media was then added, and they were incubated for one hour at 37°C with shaking. Aliquots of 50 µl and 200 µl were plated out on LB/agar containing the correct antibiotic that each construct was resistant to (Kanamycin for the plasmids that had the DyP peroxidase gene and ampicillin for plasmids that had both together). These were incubated overnight at 37°C. A single colony was used to inoculate a 10 ml starter culture which was then grown overnight at 37°C with shaking. Glycerol stocks were then made using this.

Expressing Dyp-Peroxidase and Encapsulated DyP-peroxidase.

A 10 ml starter culture of Nutrient Broth media with the appropriate antibiotic and concentration (30 µg/ml of Kanamycin for DyP by itself and 100 µg/ml of Ampicillin for Enc and DyP together) were inoculated using the corresponding glycerol stock. These were then incubated at 37°C overnight with shaking. The starter culture with the most growth after visual inspection was then used to inoculate a 1 L culture with the appropriate antibiotic. This culture was then grown at 37°C for ~3 hours until an OD600 in the range of 0.4-0.6 AU was measured, indicating that log-phase had been reached. 1 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG) and 0.5 mM α-aminolevulinic acid were then added to the 1L culture to induce expression of the recombinant proteins. The culture was incubated at 20°C overnight with shaking. A sample of the induced culture was taken for SDS-PAGE analysis and the cells were centrifuged at 4 000 x g for 30 mins to harvest them. In some cases, these were stored at -20°C overnight.

Purification of recombinant proteins

The purification methods to purify recombinant Enc and DyP were adapted from Contreras *et al.* (38) with modifications where necessary.

Purifying Encapsulated DyP-Peroxidase.

Harvested cells containing expressed encapsulated DyP were then thawed on ice if they had been stored at -20ºC. The pellet was then weighed and for each gram of wet weight, 5 mL of lysis buffer was added (50 mM TRIS-HCl, 350 mM NaCl, 10 mM imidazole, 10% v/v glycerol, protease inhibitor at pH 7.4). The resuspended cell pellet was then sonicated for 5 minutes (15 seconds on, 15 seconds off with the temperature not going above 15ºC). The lysate was centrifuged at 20 000 x g for 30 min at 4°C. The supernatant was then syringe filtered using a 0.45 µm syringe filter. The supernatant was loaded onto a Histrap Ni^{2+} Immobilized Metal Affinity Chromatography (IMAC) column (GE Health) equilibrated with 50mM TRIS-HCl, 350mM NaCl, 10mM Imidazole, 10% v/v glycerol at pH 7.4. The proteins were eluted using the above buffer containing a linear gradient of 10 mM to 500 mM Imidazole. The relevant fractions were identified using SDS PAGE and the relevant fractions concentrated using a 30 kDa Amicon filter (Millipore). The protein concentration was then calculated using UV/vis spectrophotometer. The molar extinction coefficient at 280 nm (ε280nm ~ 28420 $m^{-1}cm^{-1}$) as predicted by Expasy Protparam was used.

Purifying Dyp-Peroxidase.

Harvested cells containing expressed DyP were thawed on ice if they had been stored at -20ºC. The pellet was then weighed and for each gram of wet weight, 5 mL of lysis buffer (50 mM TRIS-HCl, 350 mM NaCl, 10% v/v glycerol, protease inhibitor and pH 7.4) was added. The resuspended cell pellet was sonicated for 5 mins (15 seconds on, 15 seconds off with the temperature never going above 15ºC). The sonicate was then centrifuged at 20 000 x g for 30 minutes at 4ºC. 60 µL of both the pellet and supernatant were taken for SDS page analysis. The supernatant was carried forward. Ammonium sulphate precipitation was carried out. A 40% cut was made, and the ammonium sulphate was added gradually with stirring while the sample was on ice. Once all the ammonium sulphate had been added and dissolved, the sample was incubated on ice with stirring for one hour. This sample was then centrifuged at 10 000 x g for 10 minutes at 4ºC. The supernatant was discarded.

The DyP sample is then resuspended by adding 10 mL of anion equilibration buffer (50mM NaCl, 50mM TRIS-HCl and pH 7.4). This sample was filtered using a 0.45 µm syringe filter to clean up any remaining debris. This sample was loaded onto a desalting column equilibrated with anion equilibration buffer to remove excess salt from the ammonium sulphate precipitation. The fractions were run on an SDS PAGE gel and the fractions containing DyP were pooled and filtered using a syringe filter as previously described. This sample was then loaded onto a HisTrap Q FF column (GE Health) equilibrated with anion equilibration buffer. The proteins were eluted using a linear NaCl gradient of 50-100 mM. The corresponding eluted fractions were collected and run on an SDS PAGE gel. The fractions that contained DyP were concentrated down to 3 mL using a 30 KDa Amicon concentrator (Millipore). This sample was purified further using size exclusion chromatography (SEC) on a S200 Increase 10/300 GL SEC column (Cytiva) equilibrated using size exclusion buffer (100 mM NaCl, 50 mM TRIS-HCl and pH 7.4). Samples of fractions from the resulting curve were then run on an SDS PAGE gel. The concentration of the DyP in different regions of the curve were determined by using a UV/vis spectrophotometer. The molar extinction coefficient at 280 m (ε280nm ~ 18 575 m-1cm-1) as predicted by the Expasy Protparam tool was used. A calibration curve was used to determine which regions corresponded to which forms of the protein and what the size of each form was.

Cryo-EM sample preparation and imaging.

3 µL of the purified encapsulin sample was applied to a glow discharged carbon coated copper grid, allowed to incubate for 30 seconds followed by blotting away of excess liquid, 3 µL of distilled water was applied to wash the sample followed by blotting of the grid. The sample was the vitrified.

The dataset was collected using a Thermo Fisher Scientific 300 keV Titan Krios microscope equipped with a Gatan K3 direct detection camera. The images were acquired using Thermo Fisher Scientific Smart EPU automated data collection software, with a super resolution pixel size of 0.53 Å (1.06 Å physical pixel size) at a magnification of 81k×. Each movie was recorded with a total electron dose of 45 e–/Å^2. The defocus range was set at -0.5 to 3 µm.

Image Processing.

Image processing was done in RELION 3.1.0 (77,78). The movies were aligned, dose weighted, summed, distortion corrected, and binned by two-fold within Fourier space which gave a 1.06 Å pixel size. This was done using MotionCor2 (79). Contrast transfer function (CTF) estimation was then applied to these movies using CTFFIND-4.1 (80). These results were then subjected to two selection steps to rule out low quality micrographs. The first step ruled out any micrographs that had Thon rings that had a CtfFigureOfMerit value that was smaller than 0.13. These selected micrographs were then subjected to the second selection step which ruled out any micrographs with Thon rings that had a CtfMaxResolution value bigger than 3.5. Particles were manually picked and extracted. These were used in 2D classification. One representative class containing was used in template-based auto-picking which resulted in an initial subset of particles. These were then extracted with a pixel box size of 400-pixels. Ten rounds of 2D classification were carried out and the resulting 3 classes. These were used to auto pick particles. These were extracted at 400-pixel box size and were subjected to 2D classification. Seven representative classes were selected. These were extracted using a 400-pixel box size and the particles were rescaled to 80 pixels to give a pixel size of 5.3 Å/pixel and therefore were easier to work with. I1 symmetry was imposed to generate an initial model. This was then subjected to 3D auto-refinement. A mask was created from this model. Both the model and the mask were rescaled back to 1.06 Å/pixel using relion_image_handler. The particle stack of the seven selected classes and the rescaled map and model were used in a 3D auto refinement. A mask was created from this map and was used along with one of the unfiltered half-maps from the 3D auto refinement in a PostProcess to generate a final reconstruction.

Building of model and refining process.

To build the Enc shell structure, the *M. smegmatis* Enc structure (7BOJ) was docked as a rigid body into the 2.7 Å map. This was done by docking as a rigid body in UCSF Chimera 1.16 (81). The structure was then built in WinCoot 0.9.7 (82) by comparing amino acid sequences and manually mutating amino acid residues. This was then run through Phenix 1.20.1-4487 (83) real space refine and was refined iteratively until a final structure was generated. Secondary structure was imposed during the real space refine.

Symmetry expansion and DyP feature determination.

To generate a map of Enc and DyP that had at the very least, a localization signal of encapsulated DyP present, localized masking and symmetry expansion was attempted to highlight a region of the inside of Enc that could have the DyP or localization signal present. A map surrounding one of the sites where DyP could have been found was created in Chimera 1.15 and was imported into RELION 3.1.0 and converted into a mask. A particle stack was selected from the refine 3D job that was part of the processing above and was symmetry expanded. These symmetry expanded particles, the mask and the map surrounding the DyP site were submitted into two 3D classification jobs. The reference map was lowpass filtered to 40 Å and C1 symmetry was used. The number of classes was set to three, the regularisation parameter T was set to twenty and forty for the two different runs, there were 20 iterations, the mask diameter was set to 300 and no sampling was used.

Processing after symmetry expansion.

The next 2 sections were based off methods and structures from the Tang *et al.* (24).

Masking using mask and map generated by Tang *et al.* (24).

A subset of micrographs was initially selected, and a subset of particles selected from these. These were extracted and underwent 2D classification to generate templates for auto picking. Five classes were selected as representative classes. These were used as the template for auto picking with representative classes being selected. The particles from these selected classes were used to create an initial model. This was used as the reference in a 3D classification in conjunction with the Enc-DyP mask from Tang *et al.* (24) followed by using both the Enc-DyP mask and map from Tang *et al.* (24).

Utilising Tang *et al.* (24) reconstruction pipeline.

An initial pool of micrographs was used to manually pick particles as a subset to use in 2D classification. These were extracted and underwent 2D classification to generate templates for auto picking. Certain classes were selected as representative classes. These yielded an initial particle stack. These were then extracted at 4.24 Angstrom/pixel. 2D classes were made of this stack and the best classes were selected through a series of 2D classifications until a more refined stack was obtained. An initial model was generated using this stack and I1 symmetry was imposed. 3D classes were then generated. A bigger particle stack was used to create a model. This was used along with the resized and recentred particle stack (that potentially had DyP features) in a refine 3D job.

Working natively.

The next step to obtain a reconstruction of encapsulated DyP was to begin working with *M. tuberculosis* itself.

M. tuberculosis growth.

The bacterial strain used is the H37Ra strain which has had multiple mutations to decrease its virulence but to remain an ideal surrogate for research into the proteome. Two 10 ml starter cultures (Middlebrook 7H9 broth supplemented with 10% v/v OADC, 0.2% v/v glycerol and 0.05% w/v Tween-80) were inoculated and incubated at 37^{O}C until they had an absorbance of between 1.2 and 1.8 AU when measured at an OD of 600 nm. These starter cultures were then diluted down to 0.2 AU. Multiple dilution steps were carried out until there was ~1 L of culture split between multiple culture flasks. These were incubated at 37^{O}C until they reached an absorbance of ~2 AU. These were centrifuged at 4 000 x g for 30 minutes and the pellets were stored at -80^{o}C until needed.

M. tuberculosis purification.

The purification methods were adapted from Tang *et al.* (24). The culture pellets were resuspended in Buffer A (20 mM MOPS, 100 mM NaCl, protease inhibitor cocktail, pH 7.4) in a ratio of 10 ml per 1 g of wet pellet weight. This was lysed using a Qsonica Sonicator Q700. The sample was lysed in a pattern of 15 seconds on, 15 seconds off for 4 minutes with the temperature of the sample ranging from 5 to 8ºC. Once lysing had been carried out, the lysate was centrifuged at 20 000 x g for 15 minutes to collect any unlysed cells and separate the cytosol and membrane from the cell wall fraction. The supernatant from this contained the cytosolic and membrane fraction and was ultracentrifuged at 150 000 x g for 2 hours at 8ºC with the membrane pellet being collected at the end of this step. The following steps were carried out at 4ºC where possible. The membrane pellet was resuspended in Buffer A that contained 1% [wt/vol] Brij-35% and was sonicated on ice to resuspend the pellet. The pattern of sonication was 5 seconds on and 5 seconds off at 3 W output using a microtip until the pellet was resuspended. The Enc was extracted using slow stirring for 2 h. This was centrifuged at 20 000 x g for 1 hour to pellet the insoluble components. The supernatant obtained was loaded onto a Ni-NTA column (GE Health) equilibrated in Buffer B (20 mM MOPS, 100 mM NaCl, 0.05% [wt/vol] Brij-35, pH 7.4). After protein binding, the column was washed with Buffer B containing 50 mm imidazole. The Enc was eluted on a gradient with Buffer B containing 1 M imidazole. Fractions that showed absorbance at 280 nm were concentrated down to between 80 and 90 µl using ultrafiltration with a 100 kDa cutoff centrifugational filter (Millipore). This sample was then run on an HiPrep 16/60 Sephacryl S-300 HR column (Cytiva) pre-equilibrated with Buffer A. Fractions were pooled to be used for negative stain grid production.

Negative stain sample preparation.

Carbon-coated copper grids were glow discharged at 25 mA for 30 seconds. 3 µl of sample were placed on a grid and incubated for 30 seconds. It was then washed with a small amount of distilled water three times followed followed by two applications of 2% uranyl acetate, with blotting using clean filter paper between each. The grid was left to dry after the final stain application.

Negative stain Data acquisition.

The native negative stain dataset was collected at 3.02 Å/pixel using a Tecnai 200 kV F20 microscope with a DE-16 (Direct Electron) direct electron detector. The images were captured using SerialEM (84) as the automated data collection program. The microscope was operated at 200 kV at an electron dose of ~40 e/ Å^2 and at a defocus of ~1.5 µm.

Processing.

An initial set of particles were manually picked, and this was carried forward. All the processing work was carried out in RELION 3.1.0. Two initial rounds of 2D classification were carried out and a subset of particles that displayed the best resolution as well as DyP features were selected. An initial model was generated with I1 symmetry. This was then used in a 3D classification job which used C1 symmetry. The next steps were to be more selective in the process of picking 2D classes. Only classes with even better resolution and features that looked like the DyP hexamer or dodecamer from different views were included with these classes being narrowed down significantly. An I1 initial model was produced and run through 3D refine. This was then used in a 3D classification job followed which was applied to initial particle set. The classes with the most DyP-like features were then selected and used as the model in a 3D refine job.

A bacterioferritin complex was thought to be present in micrographs so a set of particles that appeared to be around the correct size were manually picked. 2D classification was carried out and any classes that looked to be ferritin were used to generate an initial 3D model. This was then run through a 3D auto refine job.

Results

1. Expression and purification.

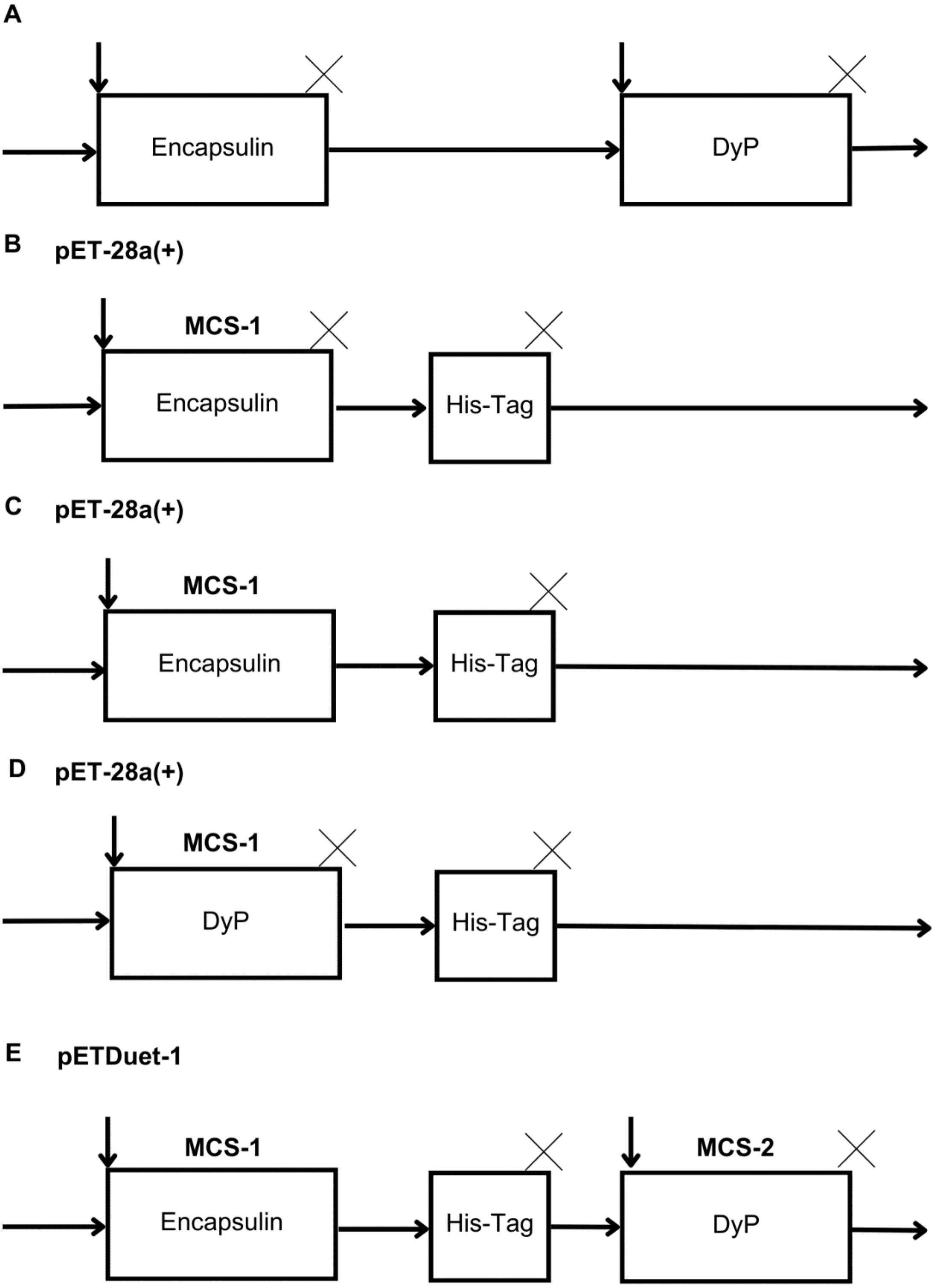


Figure 5. Simplified plasmid constructs used to express Enc and DyP together, and DyP alone. Downward arrows indicate start codons and crosses indicate

stop codons. MCS-1 and 2 refer to multiple cloning sites. (A) The native Enc and DyP operon in *M. tuberculosis*. These are controlled by the same promoter. (B) The pET-28a(+) plasmid that Enc was cloned into to form Rv0798c-pET-28a(+)-no His-tag. (C) The mutated Rv0798c-pET-28a(+)which removed the stop codon at the end of Enc so that the His-tag is added to the C-terminal end to produce Rv0798c-pET-28a(+) His-tag. (D) The pET-28a(+) plasmid that DyP was cloned into to form Rv0799c-pET-28a(+)-no His-tag. (D) The pETDuet-1 plasmid that the His-tagged Rv0798c and non-His-tagged Rv0799c were cloned into to produce pETDuet-1-Rv0799c-no Histag-Rv0798c-C-His-tag.

Enc and DyP occur in a two gene operon natively within *M. tuberculosis* (Fig. 5A). The gene for Enc (RV0798c) was cloned into pET-28a(+) and mutated to not end in a stop codon and allow the addition of a 6xHis-tag on the C-terminal end (Fig. 5B and C). The gene for DyP (Rv0799c) was then cloned into pET-28a(+) in a manner that led to no His-tag (Fig. 5D). DyP was cloned into the pETDuet-1 plasmid at MCS-2 while the gene for His-tagged Enc was cloned into the plasmid at MCS-1.

The expression of DyP alone used the Rv0799c-pET-28a(+)-no His-tag plasmid (Fig. 5D). It was not cloned in frame with the histidine tag as it was unclear how the presence of a Histidine tag would affect the function of the DyP.

pETDuet-1-Rv0799c-no Histag-Rv0798c-C-His-tag (Fig. 5E) was the plasmid used to express both Enc and DyP together. This was done so that the association of DyP and the Enc nanocompartment would be tested.

2.1. Enc and DyP expressed together.

In an attempt to co-purify and visualise encapsulated DyP, they were simultaneously expressed and purified. The proteins were purified using IMAC purification, which resulted in a broad peak consisting of one major peak in absorbance at 280 nm followed by a region where the absorbance was lower, with the entire peak encompassing 5 fractions, with the most prominent peak being present in fractions 28 and 29 (Fig. 6A and B). These fractions were picked based upon the absorbance at both 280 nm (the reading used to measure protein concentration) and at 402 nm, which is the wavelength at which heme absorbs light. DyP is classified as a heme peroxidase, and this was used as a way to determine the fractions which would have both Enc and DyP, since the Enc should bind to the column by way of the added His Tag. Fractions 28 and 29 were run on an SDS PAGE gel to determine the presence of both Enc and DyP (Fig. 6A). Both Enc and DyP were present in fraction 29 which correlates to the highest peak in the main region of interest. When comparing the relative band intensity of the Enc and DyP bands in Fig. 6A using the ImageJ Integrated Density measurement, there was ~4 times more Enc than DyP, which was different to what has been seen in *M. smegmatis*, where the expected ratio is 1 DyP to 10 Enc monomer for the hexameric form and 1 DyP to 5 Enc monomer for the dodecameric form. This shows that there was a different composition of DyP forms present in the sample. The large peak at 402 nm found within the flow through (FT) fractions at the start of graph in Fig. 6A was thought to be due to free DyP and heme that was not associated with Enc nanocompartments and was therefore not bound to the column. The presence of both Enc and DyP in the same IMAC fractions (Fig. 6A) despite not being expressed in a similar manner to the native operon structure in *M. tuberculosis* showed successful recombinant expression as well as an association between the two (Fig. 5A and E).

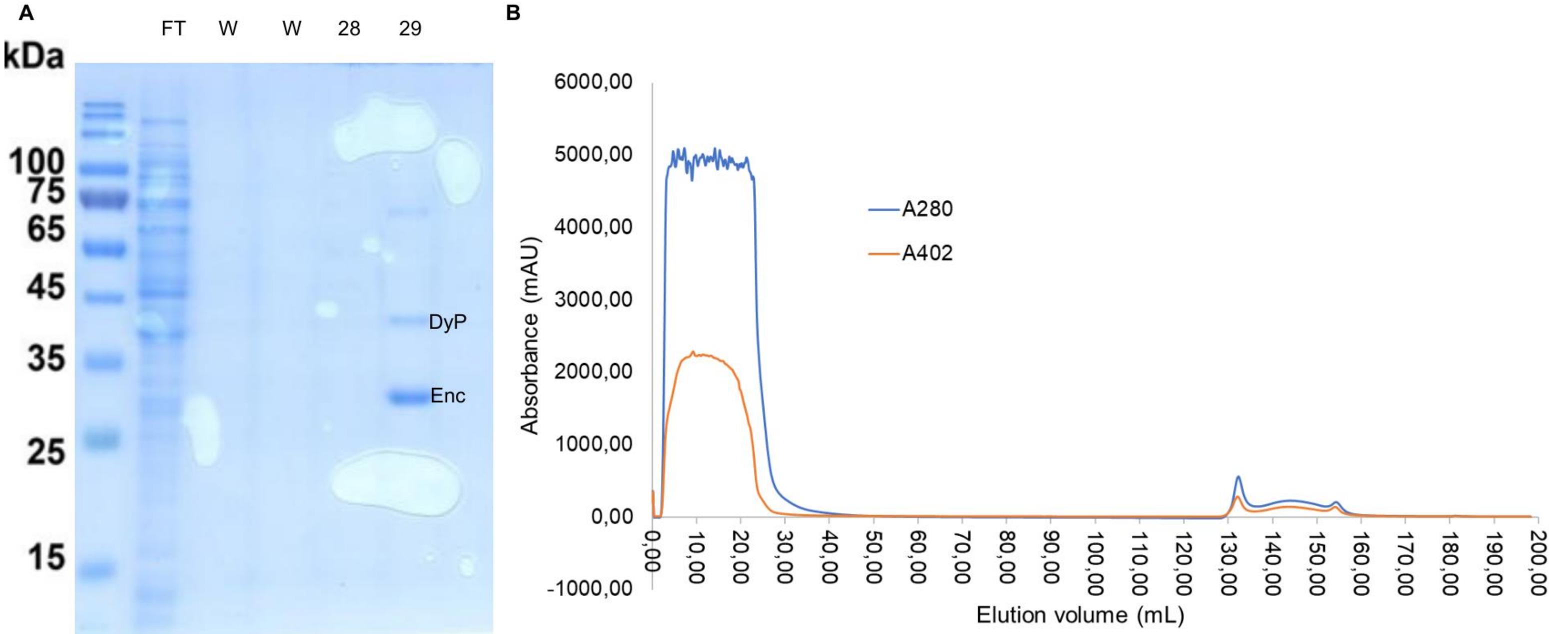


Figure 6. Purification of Enc and DyP peroxidase expressed together. (A) SDS PAGE gel image of the elution profile in A. The first lane was the FT, the second and third lanes are two regions of the wash stage (W), the third was fraction 28 and the last was fraction 29 with bands showing the presence of both Enc and DyP. Monomeric Enc is ~29 kDa in size while DyP is ~36 kDa. (B) Immobilized Metal Affinity Chromatography.

The elution curve measured at 280 nm is shown in blue. The elution curve measured at 402 nm is shown in orange.

2.2. DyP alone.

Following the purification of Enc and DyP within the same sample, the next target was the characterisation of DyP by itself. Ammonium sulphate precipitation was used as a crude initial purification step and DyP was found at 40% ammonium sulphate as per previous work (25). After anion exchange chromatography, fractions 21 and 22 had peaks present when measuring absorbance at 280 and 402 nm (Fig. 7C), with a sample of the fractions showing the presence of DyP when run on an SDS-PAGE gel (Fig. 7A). The next step was to use size exclusion chromatography to purify and fractionate the DyP. This was done to see whether different forms of DyP could be observed and to confirm which was more prevalent, with three peaks being observed in the resulting chromatogram (Fig. 7D) and run on an SDS-PAGE gel to confirm the presence of DyP. (Fig. 7B). These corresponded to the sizes of three forms of DyP. The centre fraction of each peak as well as one fraction before and after, were pooled. Each of the peaks had three lanes in the SDS PAGE gel (Fig. 7B). The molecular mass of each form was calculated using the calibration curve of the SEC column. The first and smallest peak was 144 kDa (I) and this would make it the tetrameric form. The next two peaks were the dimeric form at 72 kDa (II) and the monomeric form at 36 kDa (III). Neither the hexameric nor the dodecameric form found within *M. smegmatis* (24) were found in this sample. The most abundant form of DyP in the sample was thought to be the monomer as the absorbance of peak III was the highest (Fig. 7D) and the corresponding bands on the SDS PAGE gel had the darkest intensity (Fig. 7B). When measuring the integrated density values of the bands using ImageJ; the bands that contained the tetrameric form of DyP measured at ~10000, while the bands that contained the dimer all measured at ~12000, with the bands that contained the monomeric form measuring at ~14000.

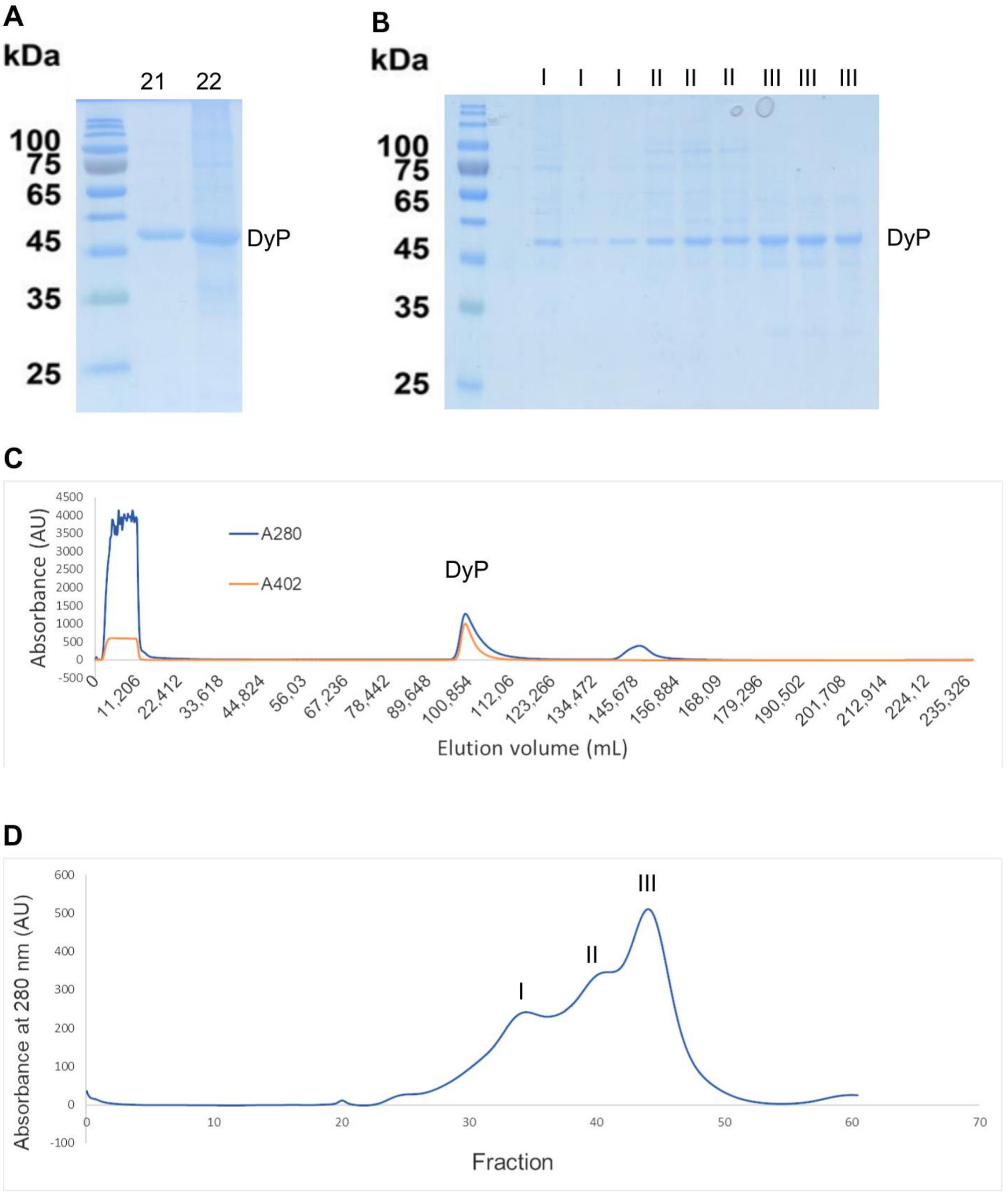


Figure 7. Purification of His-tagged DyP-peroxidase. (A) SDS PAGE gel image of fraction 21 and 22 which contained DyP, which has a molecular mass of ~37 kDa. (B) SDS PAGE gel image of SEC elution curve with the middle of the peak and one fraction either side indicated by the numerals. I was made up of fractions 33, 34 and 35. II was made up of fractions 39, 40 and 41. III was made up of fractions 43, 44 and 45. The highest absorbance peak corresponds to the most intense bands in the gel. (C) Anion exchange chromatogram of the sample with the blue line representing the absorbance at 280 nm and the orange line representing the absorbance at 402 nm. (D) Chromatogram of the

separation of DyP by SEC. (I) Tetramer, (II) dimer and (III) monomer. The monomeric form had the highest absorbance. (D)

Enc-DyP and DyP catalytic analysis were carried out in Wiswedel (54), which was centred on exploring the catalytic function of the DyP found in *M. tuberculosis*. The proteins and purification methods used in Wiswedel (54) are the same as the ones in this thesis and the outcome of the catalytic tests were deemed relevant. The work was done by an honours student in the same lab. Both the DyP by itself and the sample that contained both Enc and DyP were catalytically active which shows that the constructs and recombinant expression produced catalytically active proteins and complexes.

3. Reconstruction pipeline

3.1. Data collection.

All the reconstruction statistics are provided in Table 1. 6204 micrographs were captured at a pixel size of 1.06 Å and a magnification of 81 000x. The defocus range was between -0.5 and 3 μm.

	Enc shell
PDB entry	8PYS
EMDB entry	EMD-18031
Data collection	
Microscope	Titan Krios
Voltage	300 kV
Magnification	81 000x
Detector	K3
Data collection software	EPU
Electron exposure (e^{-} /Å^2)	45
Defocus range (μm)	-0.5 to 3
Pixel size (Å)	1.06
Data processing	
Number of micrographs:	882

Initial particles	150 444
Final particles	121 235
Symmetry	I1
Resolution (Å)	
FSC threshold 0.143	2.7
Refinement	
Initial model used (PDB code)	7BOJ
Model resolution (Å) (FSC threshold 0.143)	2.5
Map sharpening B factor (Å^2)	-115.585
Map correlation coefficient	
CC (mask)	0.86
CC (box)	0.42
CC (peaks)	0.22
CC (volume)	0.75
Model composition	
Non-hydrogen atoms	2 034
Protein residues	265
B factors (Å^2)	
Protein	26.39
R.M.S. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.497
Validation	
MolProbity score	1.15
Clashscore	2.97
Rotamer outliers (%)	0.00
Ramachandran plot	
Favoured (%)	97.72
Allowed (%)	2.28
Outliers (%)	0.00
Cβ outliers (%)	0.00

Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general.	0.0/0.0
CaBLAM outliers (%)	0.77

Table 1. Data collection and refinement statistics for recombinantly expressed DyP peroxidase-loaded Enc from *M. tuberculosis*.

3.2. Initial processing.

The two main aims of the reconstruction pipelines used were to generate a reconstruction of the *M. tuberculosis* Enc nanocompartment and a reconstruction where DyP was seen inside of the nanocompartment. The imaging of the sample was not part of this thesis as it was carried out prior to the start of the project. In a number of the micrographs captured, the Enc concentration was too high leading to overlapping particles. To generate a high resolution cryo-EM reconstruction of Enc, the selection criteria and picking used were stringent due to the nature of the micrographs, with only 882 micrographs being used from a total of 6204. The initial number of particles picked from these micrographs was 150 444 and the final particle stack comprised 121 235 particles. DyP was not apparent within any of the 2D classes obtained (Fig. 8A). An initial model with I1 symmetry imposed was generated from these particles. This was subjected to 3D auto-refinement and resulted in a map with a resolution of 10.87 Å. The particle stack of 121 235 particles and the rescaled map and model were used in a 3D auto refinement that yielded a 2.92 Å map. A mask was created from this map and was used along with one of the unfiltered half-maps from the 3D auto refinement in a PostProcess job. The reconstruction pipeline produced EMD-18031, a 2.7 Å cryo-EM reconstruction of the Enc nanocompartment found within *M. tuberculosis* (Fig. 8B) with icosahedral symmetry. Fig. 8C shows a cross section of the nanocompartment and the inside of the fivefold pore opposite to one pictured in Fig. 8B. Fig. 8D shows the fivefold pore, highlighting five monomeric units present in EMD-18031 once Segger had been run to separate out individual subunits.

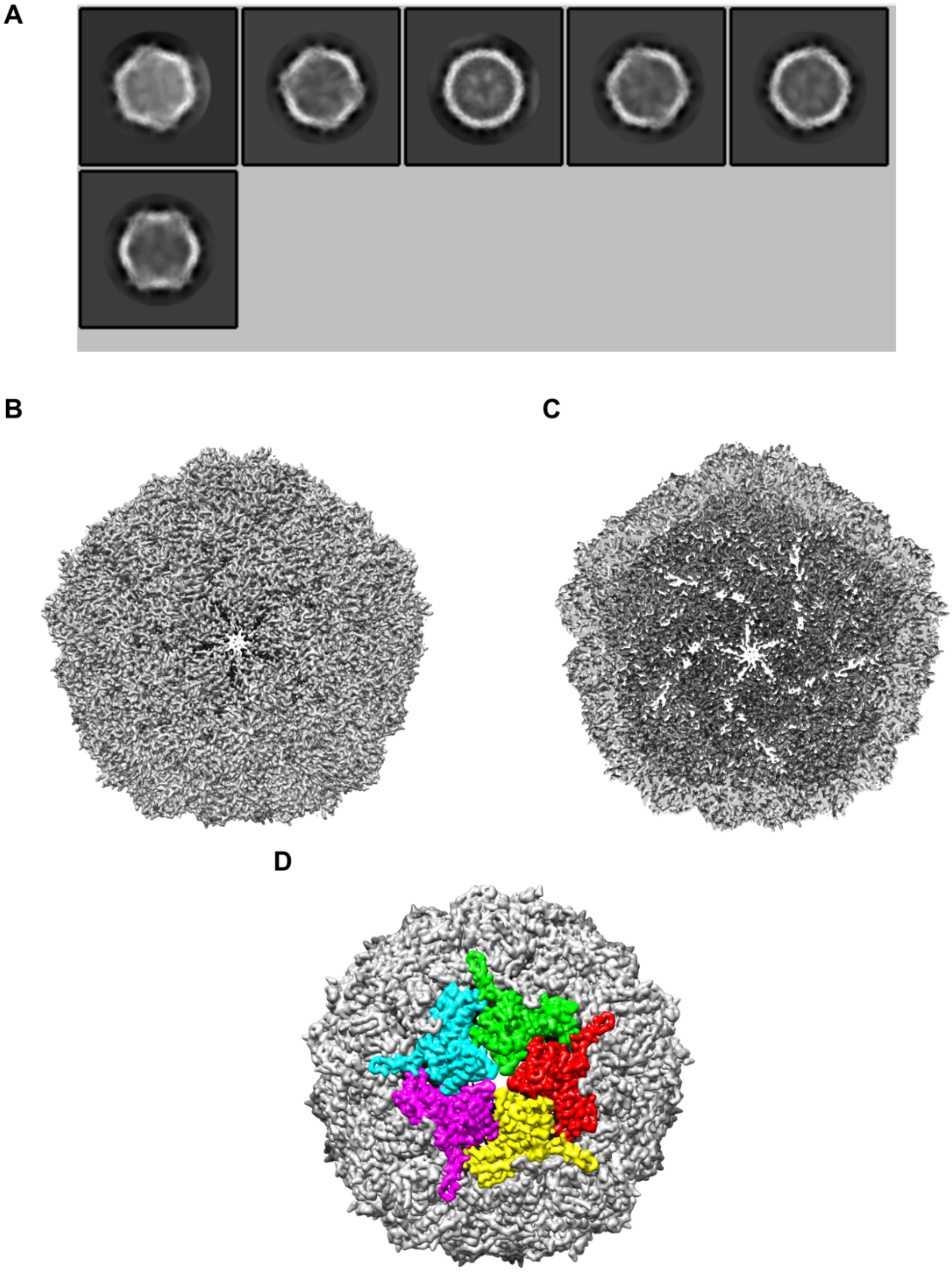


Figure 8. 2D class averages and shell architecture of density map EMD-18031. (A) The best 2D classes made up of 121235 particles which were used in further processing. (B) Exterior view when aligned with the five-fold axis of symmetry. (C) Cross section of nanocompartment down the five-fold axis. (D) The five monomers of a five-fold pore highlighted in colour. EMD-18031 was

volume filtered using Gaussian smoothing and run through Segger so that the monomeric units were easily selectable.

3.3. Analysis of EMD-18301.

The nanocompartment has a diameter of ~240 Å and a nanocompartment wall thickness of ~25 Å (Fig. 9A). The local resolution map of EMD-18031 (Fig. 9B) showed that the regions around the icosahedral five-fold pore displayed a lower resolution than the other regions of the structure, but the change in resolution was not large with the five-fold region being at ~3.0 Å compared to 2.7 Å resolution of an average for the entire map. This is in agreement with analysis of local resolution maps for the *H ochraceum* nanocompartment (41), although the range of resolutions are smaller for EMD-18031. The overall structure and shape of EMD-18031 (*M. tuberculosis* Enc) and EMD-30130 (*M. smegmatis* Enc) are very similar (Fig. 9C). When superimposed, the monomeric unit of EMD-18031 fits over the same monomeric unit of the EMD-30130 multimer with very little density falling outside (Fig. 9D). These two structures therefore have similar overall architecture. Any differences between them can’t be seen at this level of the structure and must arise when comparing the protein models of both the monomeric and multimeric unit.

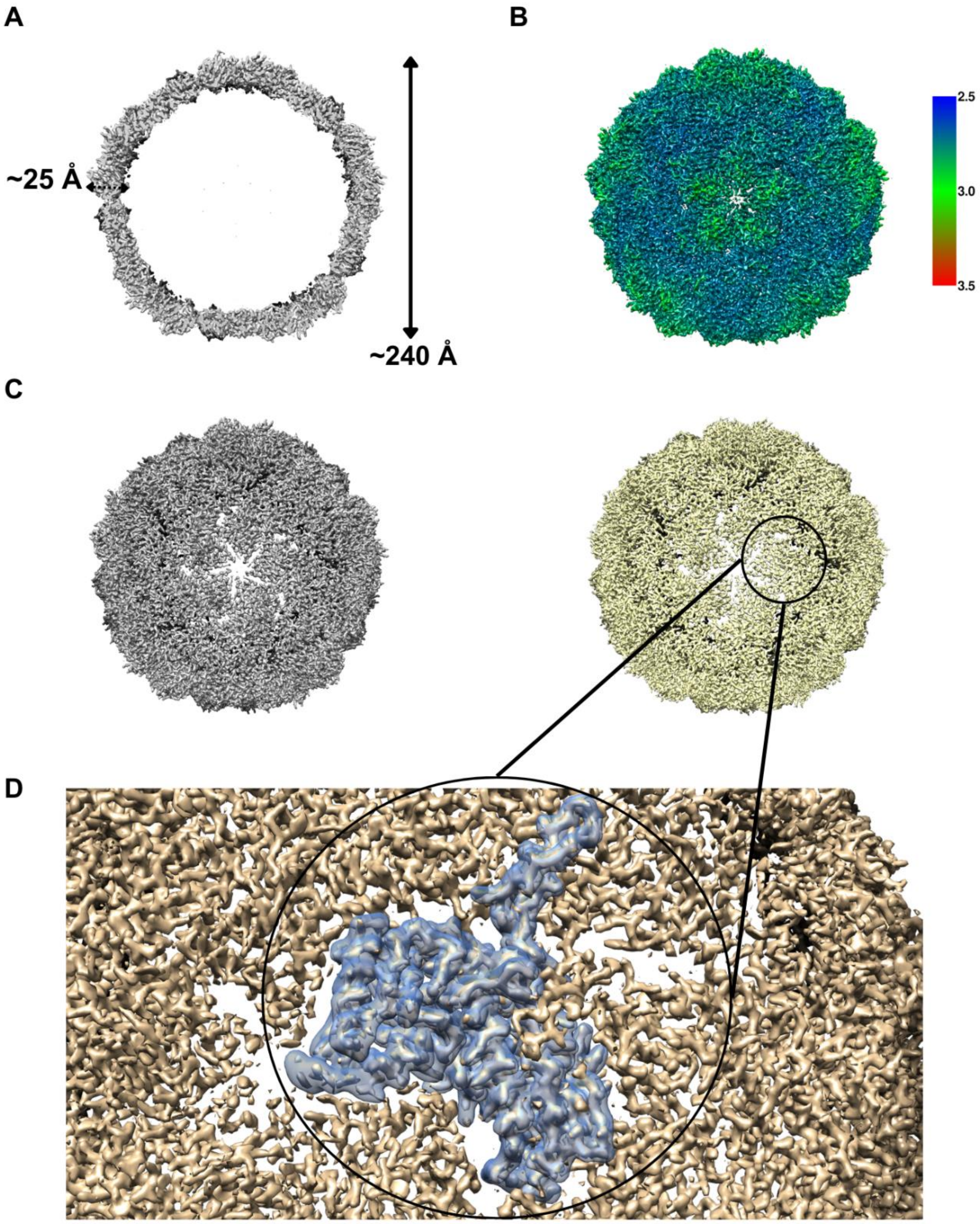


Figure 9. Properties of EMD-18031. (A) A cross section of the shell showing the diameter of the nanocompartment and the thickness of the shell. (B) Local resolution analysis of EMD-18031 where the regions around the five-fold pores deviated the most from the resolution. Colour key of the resolution mapping is shown to the righthand side. (C) Comparison of the general icosahedral shape and fivefold pore structure of EMD-18031 (Map on the left) and EMD-30130 (Map on the right). (D) A highlighted monomeric unit of EMD-30130 overlayed

with the corresponding monomeric density of EMD-18031 to show the similarity of both maps.

4. Model building.

4.1. Monomer structure.

A multimeric structure containing 60 8PYS monomeric units and displayed icosahedral T=1 symmetry was generated during the course of this work (Fig. 10A). The monomeric structure of Enc (8PYS) was built in Coot by manually mutating amino acid residues, using the Enc of *M. smegmatis* as a base model. This monomeric structure was refined in real space using Phenix with the cryo-EM 2.7 Å reconstruction as the reference map (Fig. 10B). The secondary structures from *M. smegmatis* Enc were imposed during Phenix real space refinement as some of the structures were missing after each amino acid was manually mutated, with the resulting statistics (Table 1) obtained being within the acceptable range. While there were clashes present within the monomer, these were unavoidable due to the placement of the clashes within the interior, with a total of 12 clashes. The monomeric model had a MolProbity score of 1.15 with no Geometry Restraint outliers present. The Clash score was 2.97. The overlaps were largely caused by hydrogens interacting with other hydrogens, with this being deemed acceptable as it isn't possible to predict the orientations of the hydrogens at 2.7 Å resolution. There were no rotamer outliers present and any allowed rotamers were kept that way as changes would have caused severe clashes. There were no poorly fitting residues present within the monomer. There were also no Ramachandran outliers present.

The majority of these *M. smegmatis* Enc secondary structures imposed during this real space refinement didn't raise any concerns in the Phenix output statistics, with any clashes being changed In Coot where possible. This structure was compared to the *M. smegmatis* Enc 7BOJ (24)(Fig. 10C). The amino acid sequence changes between 8PYS and 7BOJ with the changes spread throughout the structure (Fig. 10C). These changes sometimes resulted in secondary structures being shortened or no longer being present with β-sheets being most affected. There is an important change in a loop of A domain that is located directly at the five-fold pore that abolishes a beta sheet with a D present instead of an A at aa 186, right next to the histidine that forms part of the histidine ring of the fivefold pore. 8PYS was also compared to the Alphafold generated structure of Cfp29 (Uniprot entry: I6WZG6) (Fig. 10E), and was found to largely be similar in secondary structure placement and length. The *M. tuberculosis* Enc built for this thesis was similar in overall structure to the other two monomers it was compared to with a few differences when it came to length of secondary structures (Fig. 10F). Some of the secondary structures, such as the α-helix found at amino acids 124 to 126, were not present in the *M. smegmatis* monomer but were found in the other two. This is interesting to note as there is an amino acid change at the start of that α-helix from an A to an S.

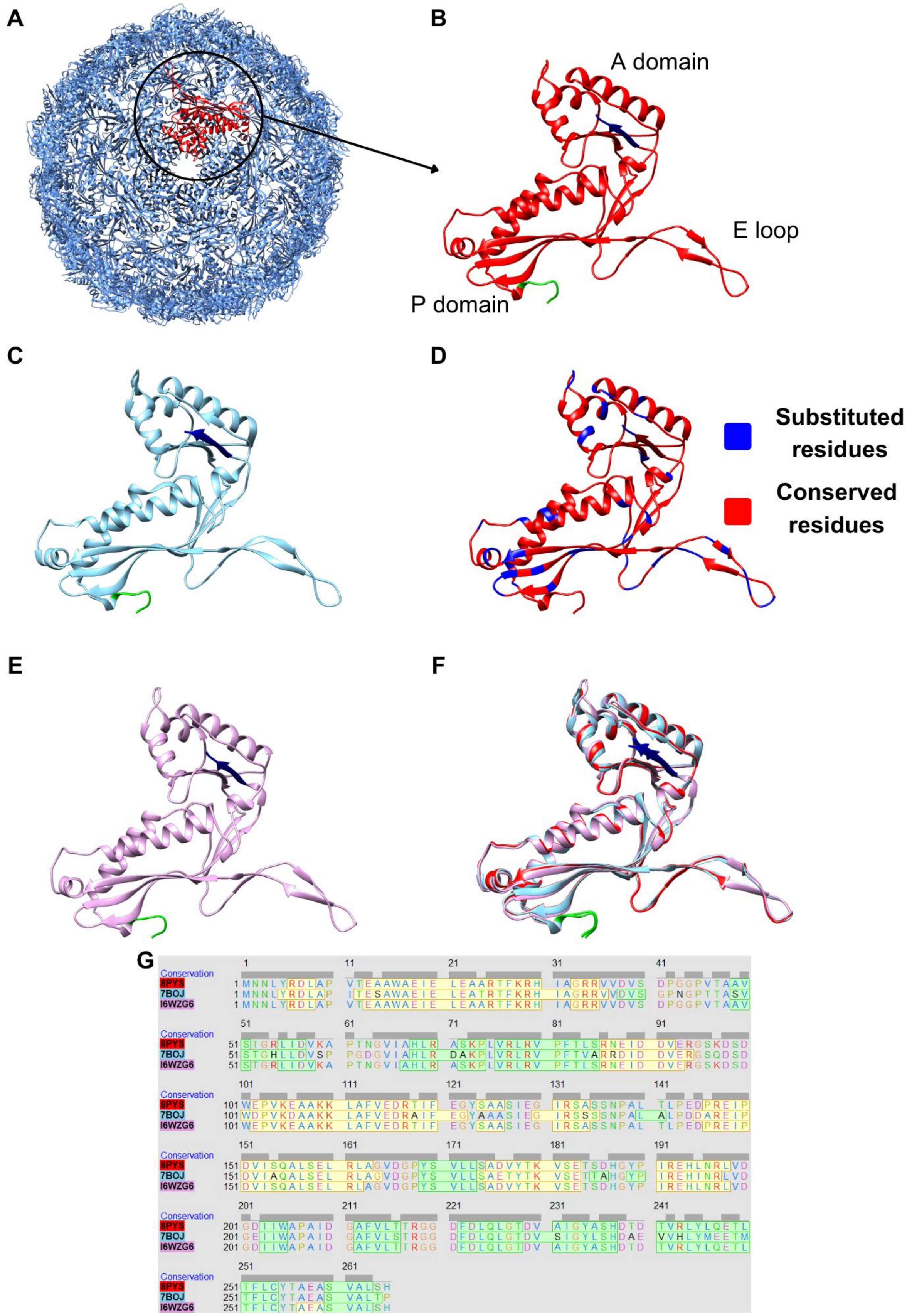


Figure 10. Structure of the Enc shell monomer. Where applicable, the N-terminus is highlighted in green while the C-terminus is highlighted in red.

(A) The multimeric form of encapsulin with a monomer highlighted in red. (B) The monomeric structure 8PYS of *M. tuberculosis* with the different structural folds labelled: P-domain, A-domain and E-loop. (C) *M. smegmatis* monomer structure (7BOJ)(24)**. (D) A map of the conserved amino acids when 8PYS is compared to the sequence of 7BOJ. Red means the presence of a conserved sequence. Blue means the presence of a substituted residue. (E) Alphafold generated structure of Cfp-29 (Uniprot entry: I6WZG6). (F) Overlay comparison of the three monomers. (G) Sequence comparisons of the aligned monomers with α-helices in yellow and β-sheets in green. The first sequence is the one from this thesis (red), the second is the one of 7BOJ (blue), and the third is of I6WZG6 (pink). Amino acids that are different when compared to the sequence of 8PYS are indicated by the grey rectangle above each residue.**

Each monomer of Enc has 3 regions. These are the P, A and E regions (the peripheral and axial domains as well as the elongated loop shown in Figure 10B). The proposed TP binding site, that is part of the P-domain and is said to be made up of mainly hydrophobic residues (35), was present within the structure built in this thesis (Figure. 11A) and contains a number of hydrophobic residues. In *H. ochraceum*, the residues within the Enc monomer that make contact with the TP of the encapsulated ferritin are L4, F27, L31, F39, D229 and I232 (41). When looking at the amino acids present at those points in the sequence in both *M. smegmatis* and *M. tuberculosis,* the amino acids are the same in both except for L31 and F39 which are still a hydrophobic Isoleucine and Valine, respectively (Fig. 10G). This indicates that this binding pocket is probably conserved. When looking at the charge distribution of the TP binding pocket (Fig. 11C), there appears to be a hollow in the corresponding region where a predominantly negative and neutral charge is seen. The presence of this hollow indicates that the TP binding pocket is most likely conserved within Enc proteins that encapsulate cargo proteins such as DyP and FLPs as it is the same composition in multiple bacteria.

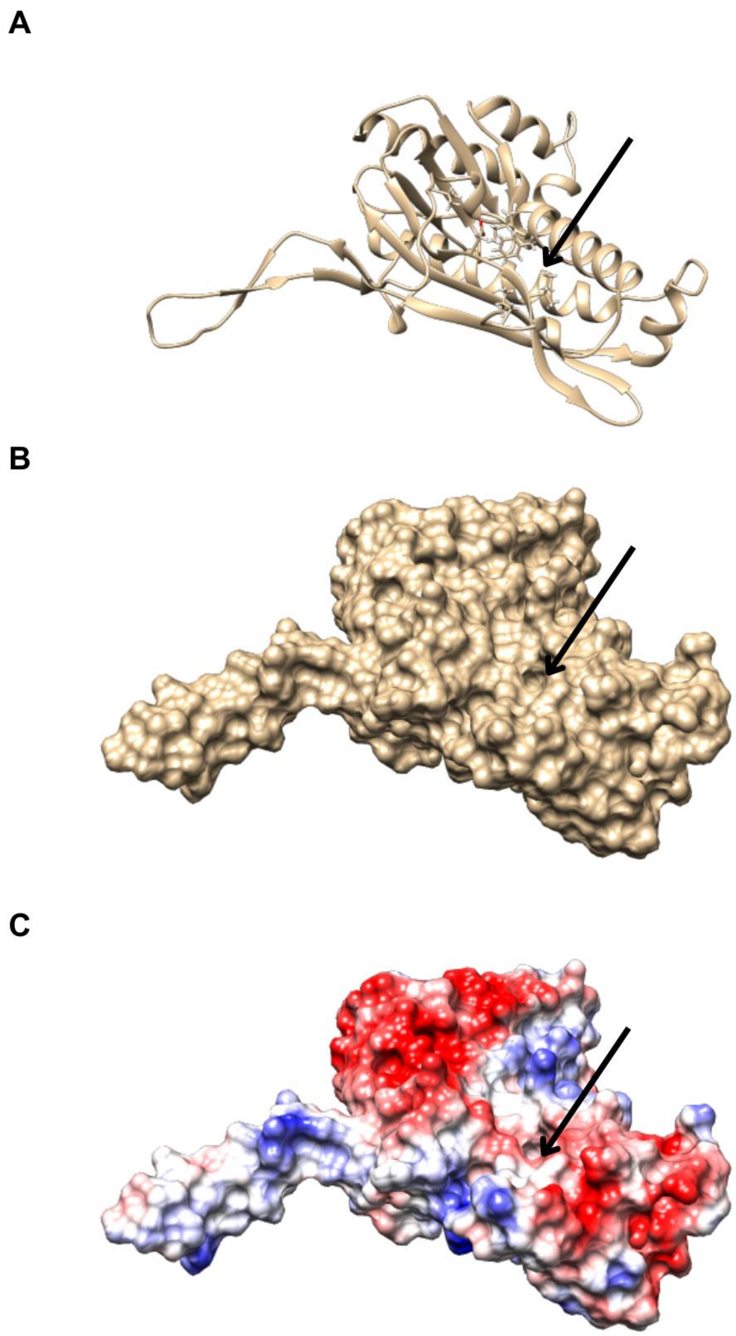


Figure 11. TP binding site with the arrow indicating the region of interest. (A) Cartoon representation of the Enc binding pocket with relevant amino acids displayed. (B) Surface representation of the binding pocket. (C) Surface representation of the binding pocket with amino acid charges indicated. Blue is positive and red is negative.

4.2.1. Overall quality of 8PYS.

The contour level chosen to test overall fit was chosen in comparison to similar structures such as 7BOJ. Overall, the fit of the multimeric (Fig. 12A) and the monomeric (Fig. 12B) versions without hydrogens was very good with very few poorly fitting residues. The glutamine at 144 was the worst fitting amino acid when analysing how well 8PYS fits into a monomer of the EMD-18031 density. This Glutamine is part of the A-domain and is one of the residues around the fivefold pore. This glutamine along with a few other residues in this region are among the worst fitting residues of 7BOJ (Fig. 11D). This corresponds to the regions of local resolution variance shown in Fig. 9B. This flexibility of the fivefold pore would explain some of the amino acid residues in this region not fitting into the density map as well as the rest of the model.

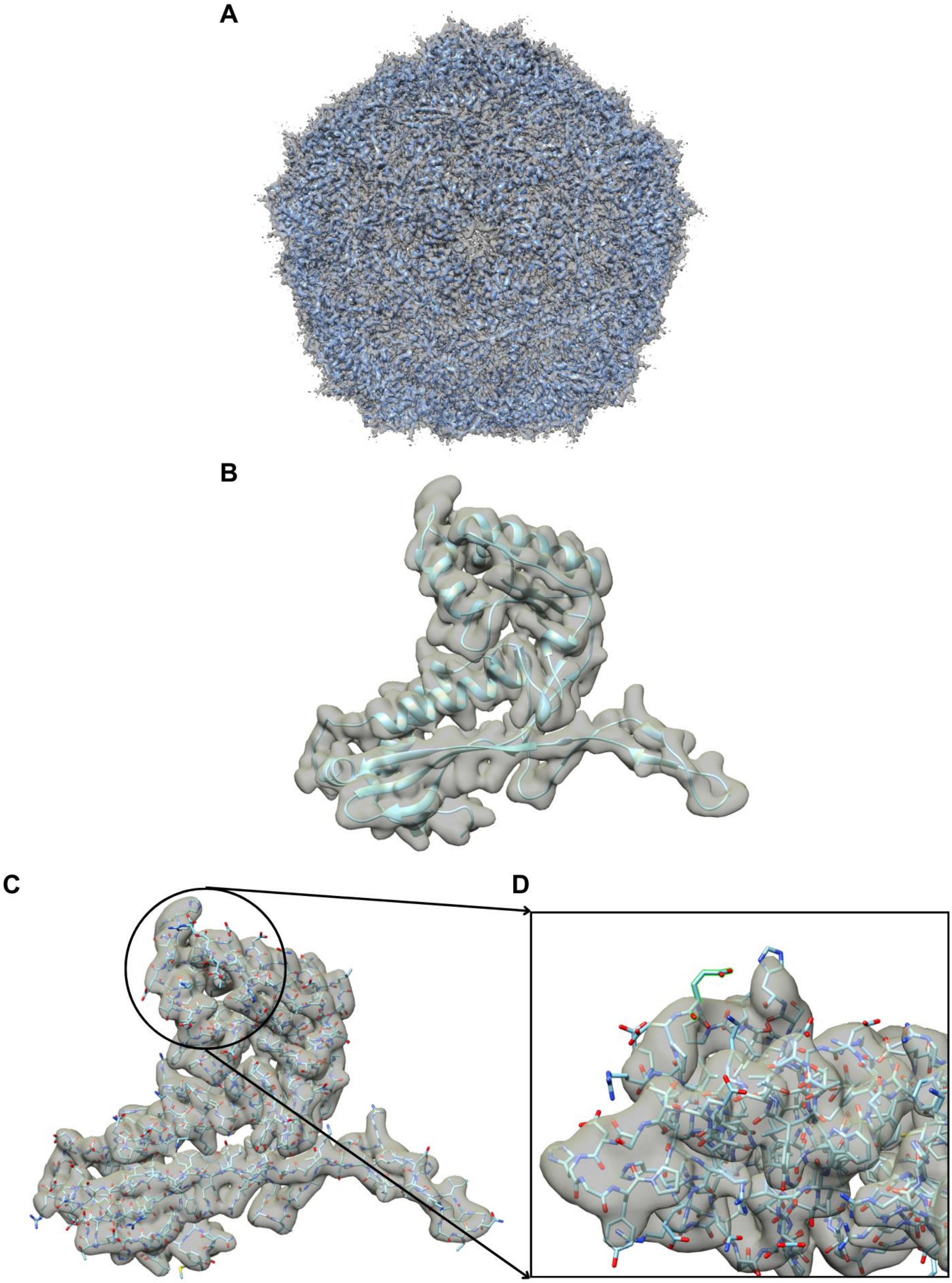


Figure 12. Multimeric and monomeric forms of 8PYS docked into EMD-18031. (A) The 8PYS multimer docked into EMD-18031. (B) The monomeric 8PYS

docked into a monomeric section of EMD-18031 obtained by Gaussian smoothing and Segger. (C) The monomeric 8PYS displaying atoms and bonds docked into a monomeric section of EMD-18031 obtained by Gaussian smoothing and Segger. (D) A zoomed in section of the P domain showing the section that contained the E residue (highlighted in green) and a few other residues that did not fit well.

4.2.2. Structure of the pores.

Each region has a different role to play in the formation of the three symmetry axes that are formed in the Enc shell. At each symmetry axis, an opening is formed, with these openings being referred to as pores. Two pores are found at the three-fold and five-fold axes of symmetry while a two-fold axis of symmetry is found where two monomeric units interact. The fivefold pore has a ring of five A domains (Fig. 13A). An alpha helix from each A domain as well as a loop of amino acids form the main part of the fivefold pore structure. This loop, which is the main region of the fivefold pore, has a slightly different amino acid makeup to the one in *M. smegmatis*. The loop is present from 184 to 190, with a largely hydrophilic makeup and two hydrophobic amino acid when compared to the three of *M. smegmatis* Enc. There are also no secondary structures present here as opposed to the two β-strands found in *M. smegmatis*. The diameter of the fivefold pore is ~8 Å, but this then decreased down to ~6.8 Å when the Van der Waals radii of the hydrogens (1.2 Å) were subtracted due to the difficulty in predicting the orientation of hydrogens given the resolution of the density map obtained and used during the fitting process. The threefold pore is made up of the P-domains of three monomeric units (Fig. 13B). The specific regions of the P-domain that make up the pore are the N-terminal helix and one of the strands of a β-sheet found in the P-domain. The final symmetry related axis, the twofold axis pore, is formed by the interaction of two β-strands from two E-loops that run parallel to each other to form a β-sheet composed of 4 β-strands (Fig. 13C).

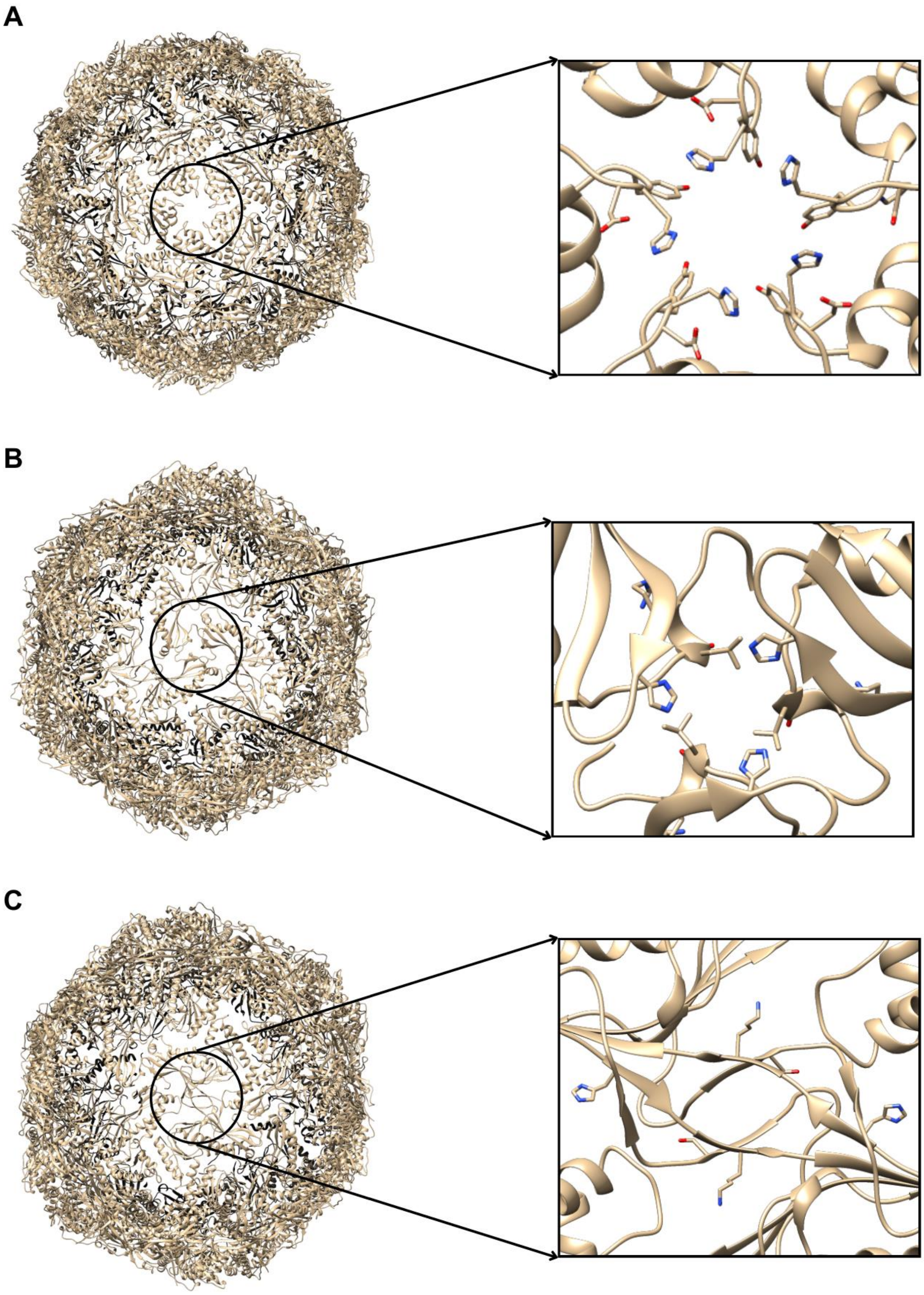


Figure 13. The pores of the Enc nanocompartment. Cartoon representation of the five-fold pore (A), the three-fold pore (B) and the two-fold pore (C). The

images on the fight are zoomed in images of the different pores to highlight the opening. Key amino acids that contribute to charge are shown as stick structures with charges present. Positively charged is blue and negatively charged is red.

4.2.3. Charge of the pores.

Fivefold pore.

The *M. tuberculosis* Enc amino acid sequence has 83% identity to that of *M. smegmatis* (Fig. 9D). The amino acid changes appear to have led to a change in the overall charge distribution of the pores. One notable change is present around the five-fold pore, where an aspartic acid is present in *M. tuberculosis* rather than an alanine in the case of *M. smegmatis* at amino acid 186 (Fig. 10G and Fig. 13A). This changed the overall charge of the exterior area surrounding the pore to a more negative one, but the histidine at amino acid 187 was left untouched although its charge is not easily visible (Fig. 14A). This led to a more negative and neutral charge directly around the five-fold pore in the *M. tuberculosis* Enc nanocompartment (Fig. 14A) when compared to the five-fold pore found in *M. smegmatis* Enc nanocompartments (Fig. 14C). Other than this change directly around the pore, there are no other major charge differences when comparing the five-fold pores of *M. tuberculosis* and *M. smegmatis*.

The interior side of the pentameric pore has groupings of negatively charged residues that form five lines radiating outwards from the pore (Fig. 14B). Each line is comprised of one alpha helix that stretches from 191 to 198 in the sequence (Fig. 10G) with the furthest section of the negative charge being comprised of two loops of amino acids. The reason for the negative charge, despite the presence of more positively charged residues in the helix, is due to the negatively charged amino acid side chains that face into the interior of the Enc nanocompartment. These form channels of negative charge separated by pockets of positively charged residues. The interior of the fivefold pore found within *M. smegmatis* Enc nanocompartments is similar in charge distribution (Fig. 14D).

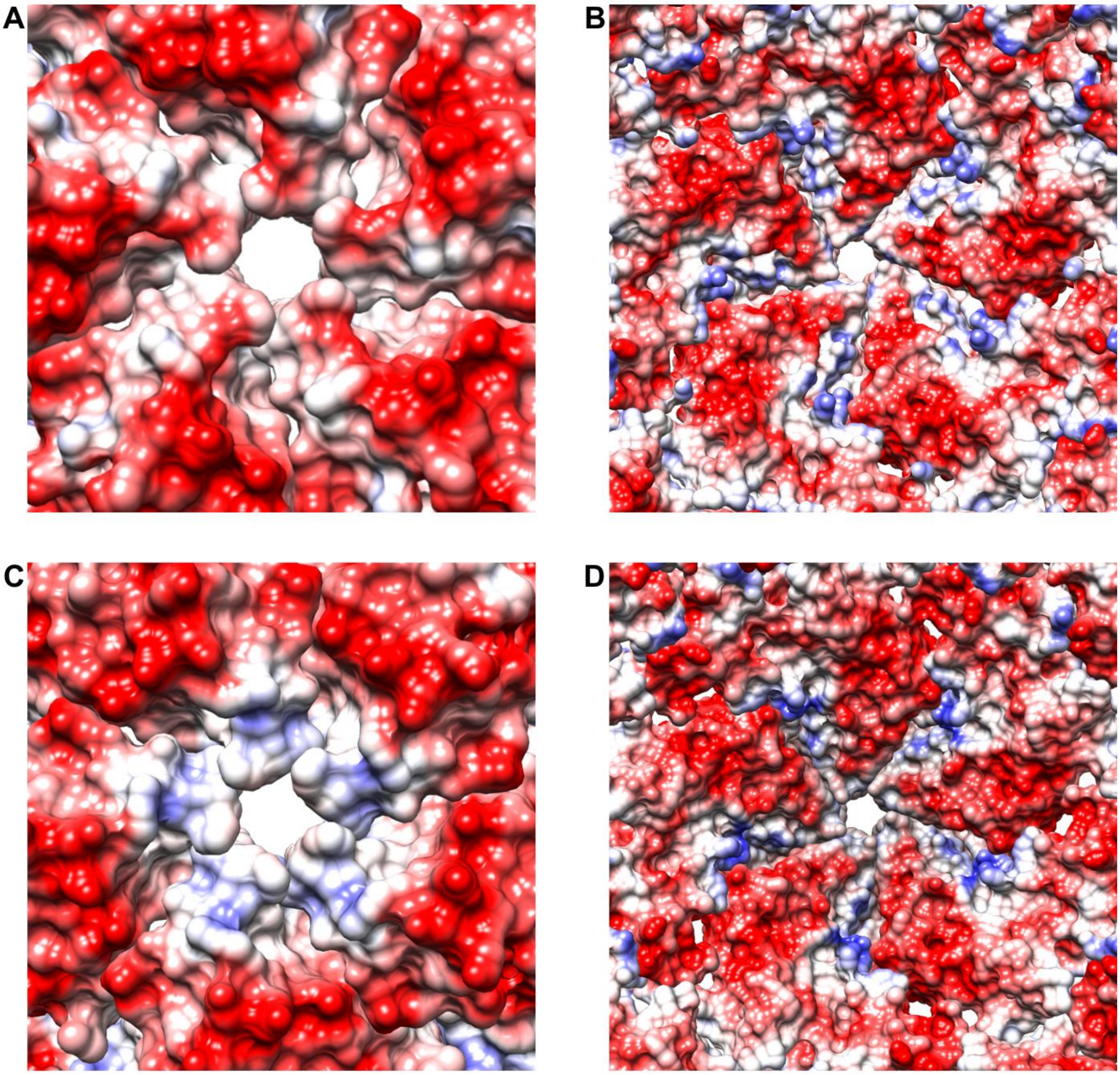


Figure 14. Electrostatic charge distribution of the five-fold pores from different Enc nanocompartments with red meaning a negative charge and blue a positive one. (A) The exterior of the *M. tuberculosis* Enc nanocompartment five-fold pore. (B) The interior of the *M. tuberculosis* Enc nanocompartment five-fold pore. (C) The exterior of the *M. smegmatis* Enc nanocompartment five-fold pore. (D) The interior of the *M. smegmatis* Enc nanocompartment five-fold pore.

Threefold and twofold pores.

The three-fold pore is also largely the same except for a few changes in charge in the areas around the pore. There are more positively charged amino acids present in the beta sheet that extends from amino acid 241 to 254 (Fig. 10G and 13B). This leads to more positive and neutral charges (Fig. 15A) when compared to the three-fold of *M. smegmatis* Enc. The two-fold pore follows this trend of charge changes (Fig. 13B). In the case of the two-fold pore, two regions around the pore have more positive charges compared to the *M. smegmatis* counterpart (Fig. 10G and 13C).

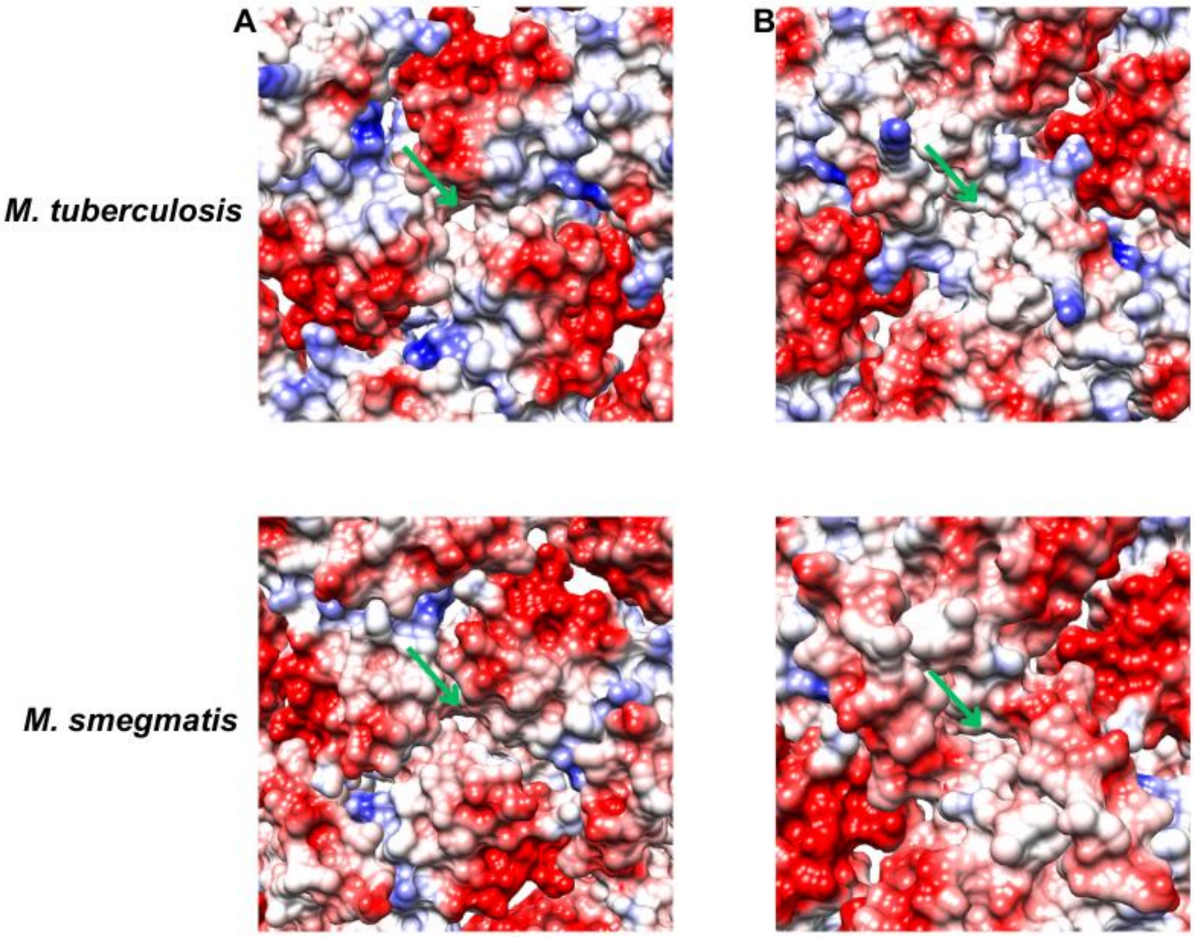


Figure 15. Comparison of the charges around the two other Enc nanocompartments pore types of both *M. tuberculosis* and *M. smegmatis*. Each pore opening is indicated by the green arrow. (A) 3-fold pore. (B) 2-fold pore.

5. Further attempts at obtaining a structure of DyP inside of the nanocompartment.

5.1. Masking the interior.

To enhance the signal from any potential DyP present in the dataset, two methods were used on this dataset and map. The first was to create a mask encompassing just the interior of the Enc nanocompartment and the nanocompartment itself (Fig. 16). This was done to decrease the noise from outside the nancompartment which would make sure the signal of the DyP was enhanced. None of the classes obtained had any density present within the nanocompartment except the yellow class but the density only formed at a lower contour threshold and did not resemble DyP.

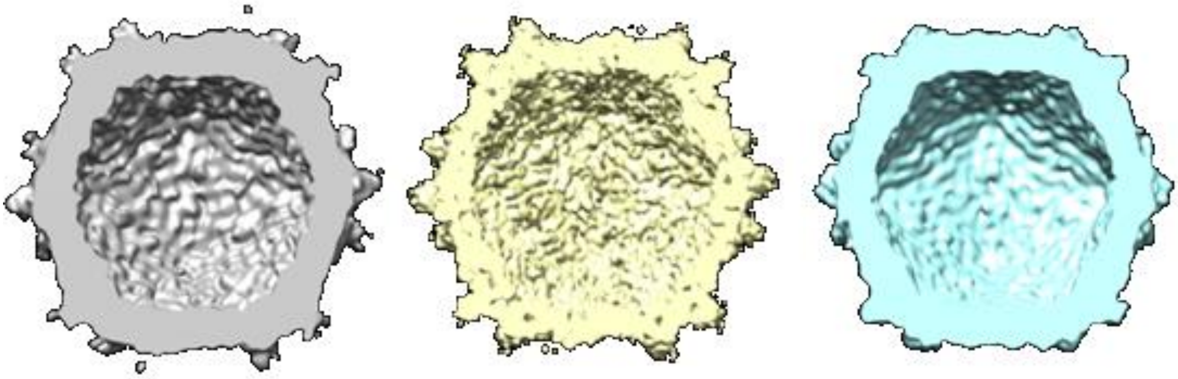


Figure 16. 3D classes of the masked interior of the Enc nanocompartment.

5.2. Symmetry expansion and focused classification.

The second was to symmetry expand the particle stack that was used to generate the 2.7 Å map followed by focused classification. This was done to increase the likelihood that any asymmetric feature such as DyP would be more equally represented in the dataset. This particle stack was then used in a localised focused classification which used a spherical mask (generated in Chimera) of an area that is part of the P-domain and is said to be where the TP of the DyP would bind (Fig. 17A). It is comprised of a hydrophobic pocket between the α-helix at the N-terminus and a three stranded β-sheet (Fig. 17A). Even at a lower threshold when using the volume viewer function in Chimera, the 3D classes obtained did not show convincing

TP density and showed no density other than the Enc at a higher threshold (Fig. 17B and C).

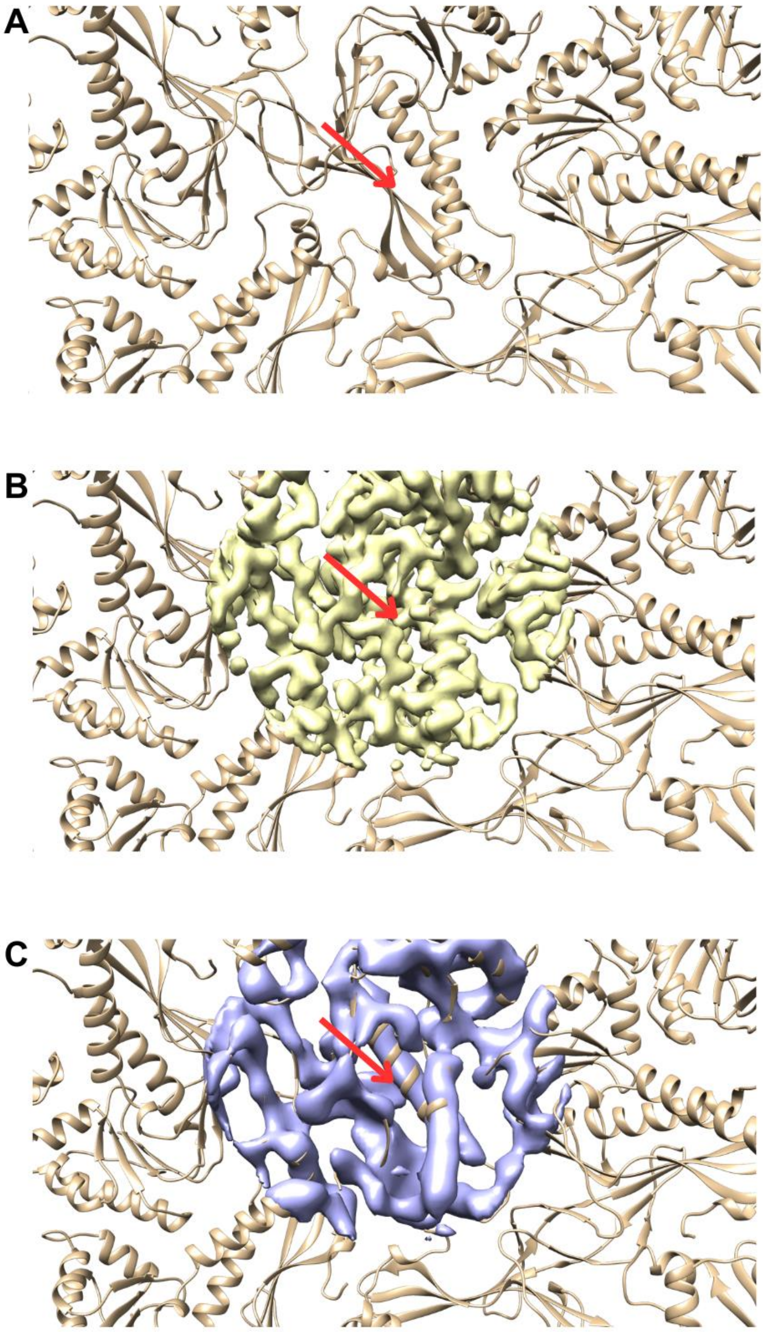


Figure 17. Symmetry expanded and focused classified 3D classes with the region of TP binding indicated by the blue arrow. (A) Cartoon representation of binding pocket and surrounding area. (B) 3D classes of masked region docked overlayed on the Enc structure.

5.3. Tang *et al* (24).

While attempting to reconstruct encapsulated DyP, a density map showing dodecameric DyP inside of an Enc nanocompartment for *M. smegmatis* (EMD-30132) was published (24). Due to the high genetic similarity between *M. tuberculosis* and *M. smegmatis*, it was decided that the reconstruction pipeline described in Tang *et al.* (24) should be used as the template for further reconstruction attempts using the cryo-EM dataset generated in this thesis.

5.3.1. Using the Tang *et al* (24) map and mask.

The exact pipeline started by using a larger number of micrographs and particles during processing as this would increase the likelihood of visualising DyP. 5 984 micrographs were used with 996 466 particles being auto picked. This initial particle stack was then refined down to 644 182 by way of 2D classification. These were used to make an initial model at 11.2 Å. Two different 3D Classifications were carried out using this data set. The first of which was to use this initial model in combination with the mask that Tang et al (24) used in their paper in an attempt to bias the 3D classes into showing DyP (Fig. 18A). The plan was to use the particles in those classes to generate a model and then to impose DyP density upon the 3D classification so there would be more classes that had interior density present, with these classes being used in subsequent processing steps. This first classification did not yield convincing DyP density inside the 3D classes, even at low thresholding. The second round of 3D Classifications carried out took the particle stack that was used to create the initial model and applied the density map and the mask from Tang et al (24) (Fig. 18B). There were two classes that showed density (the blue and the red density maps in Fig. 18B), but the resolution of the Enc shell was not high enough and the density was not convincing.

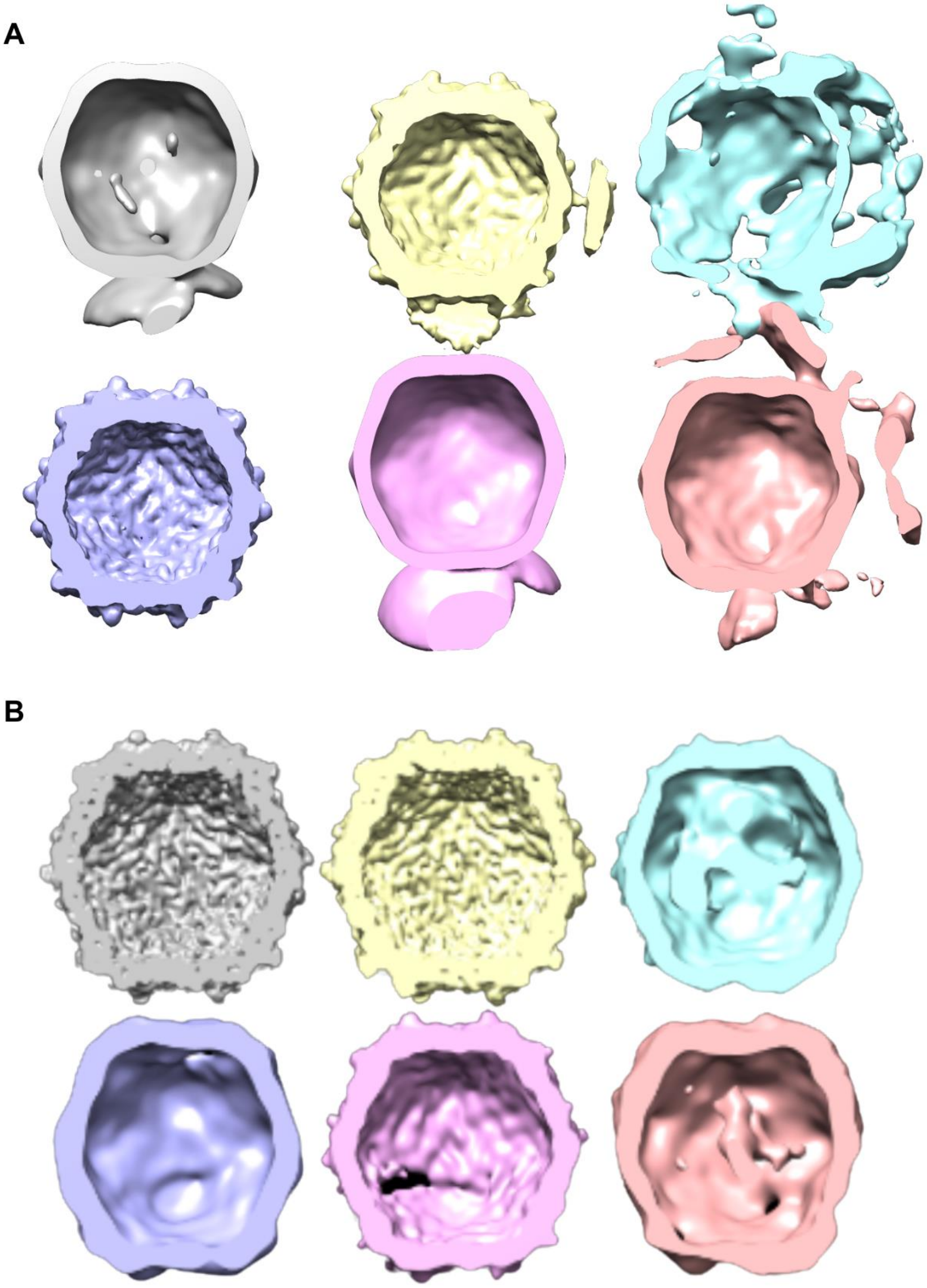


Figure 18. 3D classes reconstructed using different model and mask combinations. (A) 3D classes made using 11.2 Å initial model generated above and the mask made by Tang *et al.* (24) **as the template. (B) 3D classes made using the EMD-30132 density map and the mask made by Tang *et al.*** (24) **as the template.**

5.3.2. Following the Tang *et al.* (24) reconstruction pipeline.

The final attempt using the recombinant data set was to follow the exact same reconstruction pipeline as Tang *et al.* (24), as it proved successful in that paper. The pipeline used a far less stringent approach than the one initially used in this thesis. 5 984 micrographs were selected initially. 1 013 particles were manually picked as a subset to use in 2D classification. These were extracted and underwent 2D classification to generate templates for auto picking. Five classes were selected as representative classes. These yielded a particle stack of 996 466 (Fig. 19A). One notable issue with the data set was the presence of many overlapping particles due to a high concentration sample. This decreased the number of viable particles in the dataset. These particles were then extracted at 4.24 Angstrom/pixel. 2D classes were made of this stack and the best classes were selected through a series of 2D classifications until there was a stack of 138 454 (Fig. 19B).

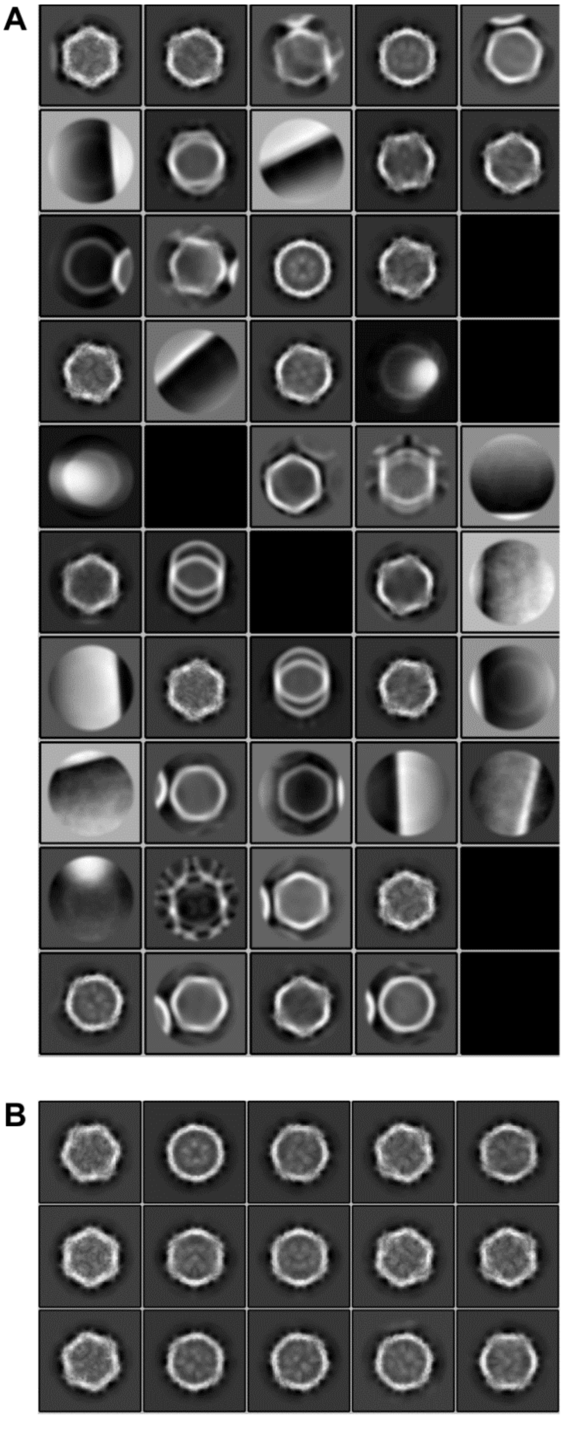


Figure 19. 2D classifications following the Tang *et al.* (24) **pipeline. (A) Initial 2D classes made up of 996 466 particles. (B) Final selected 2D classes made up of 138 454 particles.**

An initial model at 15 Å was generated using this stack with I1 symmetry imposed (Fig. 20A). 3D classes were generated from this which didn't yield DyP being within the nanocompartment (Fig. 20B). This differs from what Tang *et al.* (24) observed. At this step during the reconstruction pipeline, they saw a single class showing good DyP features. Since no DyP features were seen, it was decided that a bigger particle stack was required. A bigger particle stack of 644 182 particles was used to create an initial model of 11.223 Å (Fig. 20C). This model was used in a 3D classification job but did not yield any classes that displayed DyP features, even when the thresholding used when viewing the maps was adjusted (Fig. 20D). This illustrates that even when a far larger portion of the dataset was included in the reconstruction pipeline and a successful reconstruction pipeline of encapsulated DyP from the closely related *M. smegmatis* was followed, DyP was not able to be reconstructed out of this dataset. A plausible explanation was the lack of either a hexameric or dodecameric form in the samples, as shown by the purification data (Fig. 7D).

Figure 20 is a summative flow diagram showing the various stages of the reconstruction process and how each step transitions into the next.

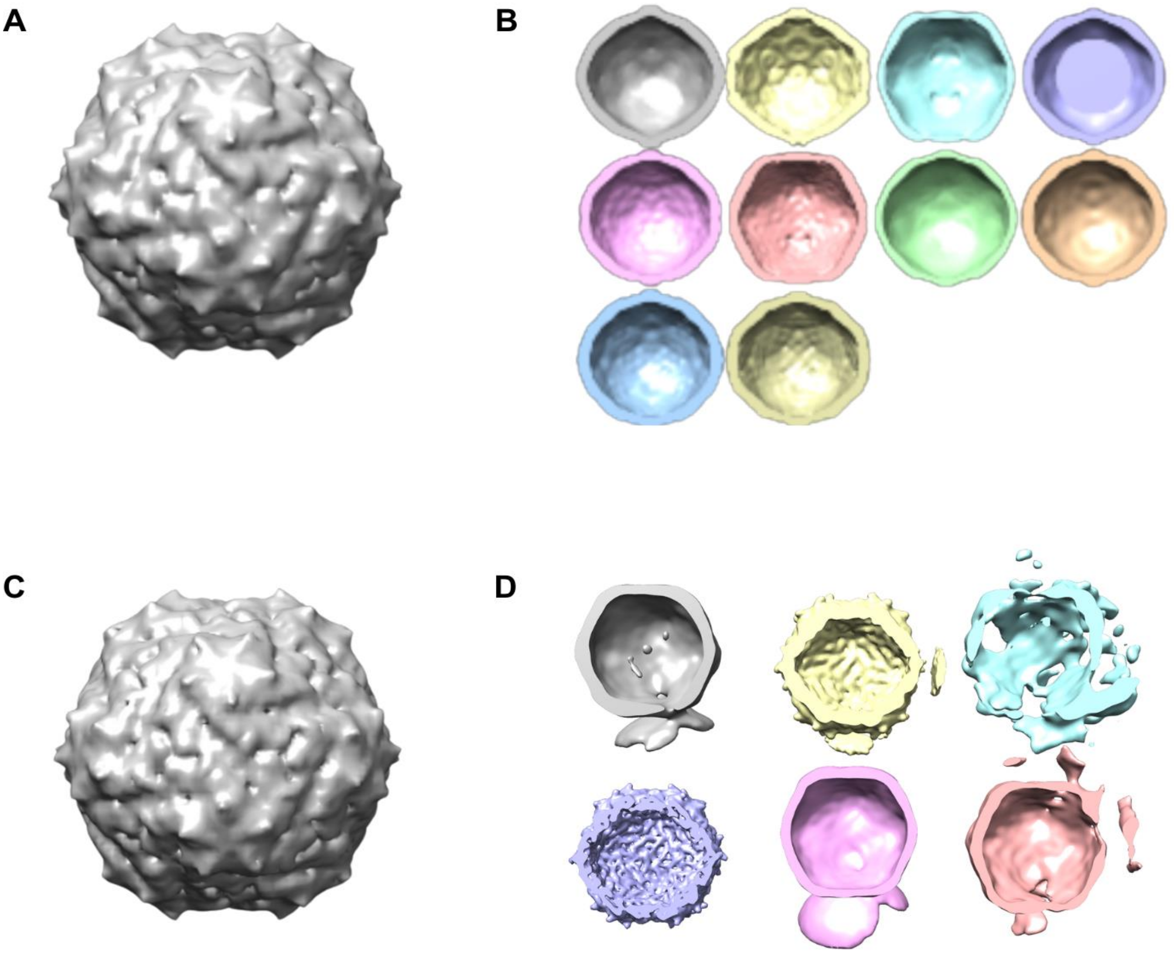


Figure 20. Reconstruction process using Tang *et al.* (24) **reconstruction pipeline. (A) 15 Å reconstruction. (B) 3D classes using the 15 Å model as a reference. (C) 11.223 Å model using a larger dataset. (D) 3D classes using the 11.223 Å model as a reference.**

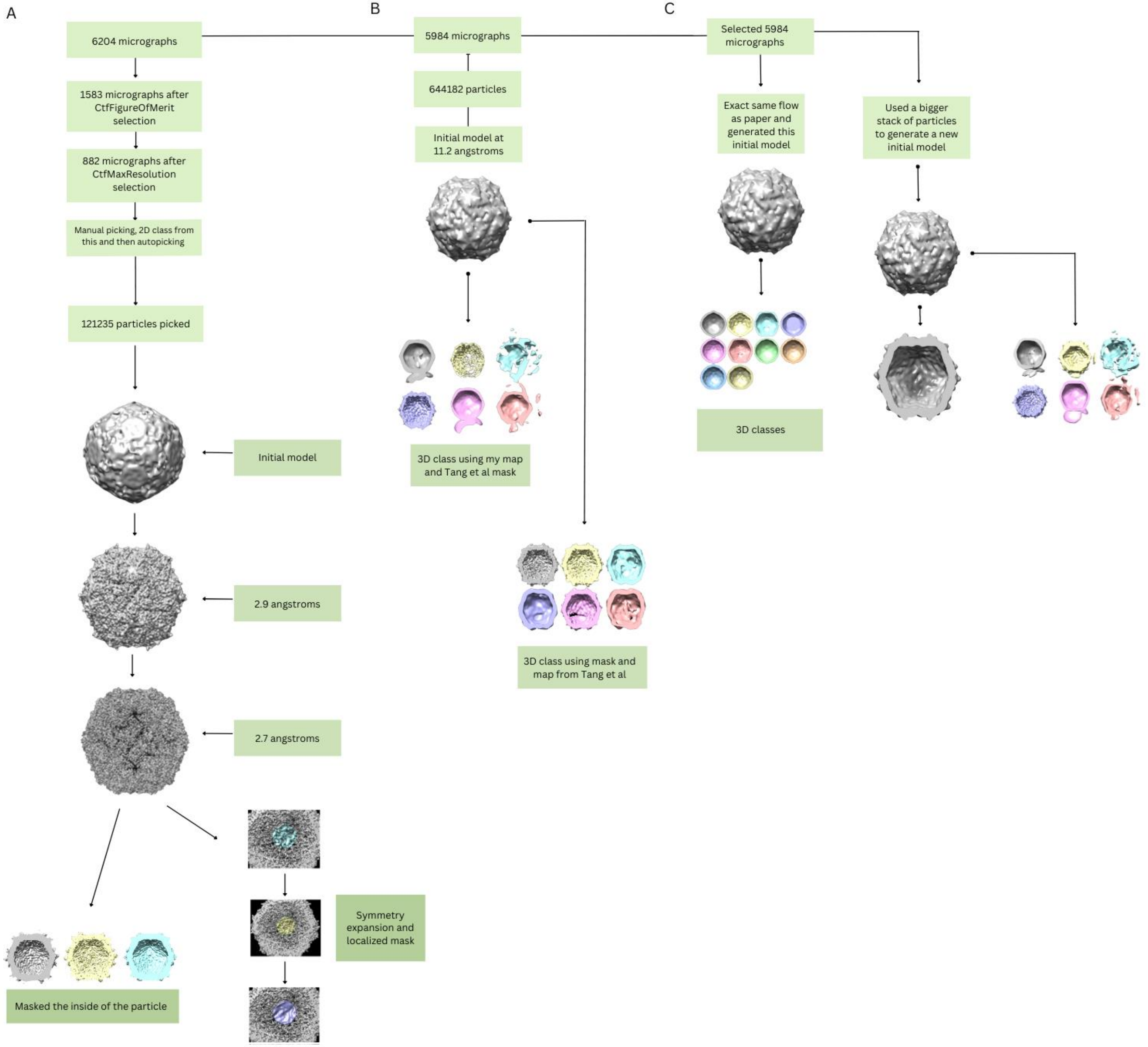


Figure 21. Various reconstruction pipelines. (A) 2.7 Å reconstruction pipeline with DyP reconstruction attempts. (B) Combination of masks and maps from *M. smegmatis*. (C) Exact same pipeline as the one that reconstructed *M. smegmatis* Enc-DyP.

6. Native *M. tuberculosis* Enc.

The SDS-PAGE gels of the purified sample where Enc and DyP were recombinantly expressed together in *E. coli* showed bands of both Enc and DyP (Fig. 6A). This result coupled with the catalytic studies done by Wiswedel (54) showed that both Enc and DyP were present and catalytically active within the same sample. This makes the difficulty in reconstructing a map of any forms of encapsulated DyP, unexpected. Other studies of encapsulated DyP expressed in *E. coli* showed that the recombinantly expressed DyP was still catalytically active (38). When DyP was expressed and purified by itself, three forms came out in the SEC data Fig. 7D). The 3 forms were thought to be the tetramer, dimer, and monomer. This finding is consistent the findings of Contreras *et al*. (38), who saw the same forms but in different ratios. These forms contrast with findings in *M. smegmatis* where the forms found are the monomer, hexamer and the dodecamer (24). This indicated that the next step should be to attempt to study Enc and DyP natively.

The next steps taken were to explore native expression in *M. tuberculosis* strain H37Ra in hopes that the complex forms more readily under native conditions.

6.1. Native *M. tuberculosis* expression, purification, and micrograph capture.

The *M. tuberculosis* H37Ra cell cultures were grown until ~2 AU when measured at 600 nm. This was thought to be an ideal point as a previous study had shown that stationary phase was reached at ~15 days with an absorbance of 1.5 AU when measured at 600 nm (85).The amount of culture able to be grown was severely limited due to biosafety restrictions. One litre was grown and yielded around 7.5 grams of pellet. Fractions 10 to 14 were pooled from nickel affinity IMAC as the chromatogram gradient at that section differed from imidazole curve present within the chromatogram (Fig. 22A). This yielded a sample which still required further purification upon negative stain viewing. Each grouping of fractions had to be analysed by negative stain microscopy to assess their suitability, with fractions 10 to 14 being the ones taken forward. The sample was run through the SEC column in an attempt to separate the Enc from the bulk of the other protein in the sample. Due to the size of the Enc nanocompartment, it was thought that it could come off the column near void volume due to the size constraints of the column used (1 kDa-1.5

MDa range) with the Enc nanocompartment having a total molecular mass of 1.74 MDa. This meant that the protein would supposedly come off the column around the 60 mL mark (Fig. 22B). There is a noticeable oscillating pattern present in the chromatogram with the region around the void volume being the noticeably different section as a peak emerged. Different fraction groups were pooled and concentrated down but fractions just on and after the void volume proved to be the fractions that would be taken forward for negative stain EM and image processing.

The Enc nanocompartments were not abundant in the image set that was captured. 634 images were taken of the sample and the average number of particles per micrograph was 4.7. When looking at the micrographs, there appeared to be a number of different proteins present along with contaminant. The lack of Enc nanocompartments and low purity of the sample meant that cryo-EM could not be carried out. The nanocompartments also proved to be variable in occupancy and orientation of density when looking at the micrographs captured with the majority being empty (Fig. 22C).

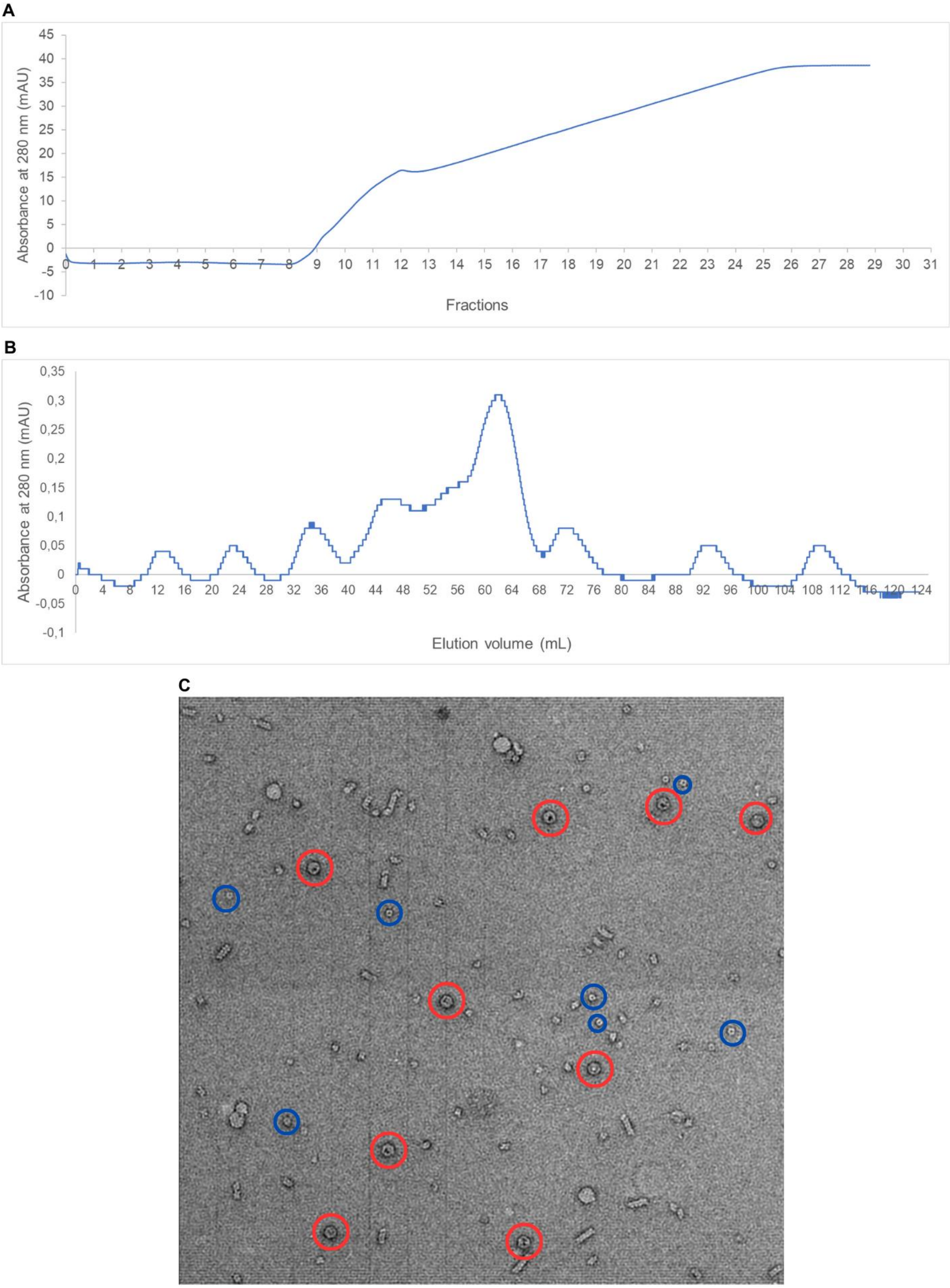


Figure 22. Native purification data and resulting micrographs. (A) Chromatogram showing the elution profile from the Ni-NTA purification. Each

fraction was 1 mL and the fractions pooled were 10-14. This part of the curve was deemed to not be part of the imidazole elution curve due to the difference in slope gradient. (B) Chromatogram of the elution profile from SEC purification with each fraction being 2 ml. Void volume is at around 60 mL elution volume (Indicated by the black arrow). The fractions just after 60 mL were taken forward. (C) Examples of particles selected for Enc nanocompartments (red circle) and ferritin (blue circle).

6.2. Image processing.

6.2.1. DyP

An initial set of 2 952 particles that were classified as Enc nanocompartments were picked manually, and 2D classification was carried out on these particles with varying apparent occupancy (Fig. 23A). To select for the best particles, 2D classification was carried out and a final subset of 2 475 particles were taken forward. There was heterogeneity in the occupancy and density seen on the interior of the nanocompartment within the 2D classes (Fig. 23B). There were examples of empty nanocompartments (yellow circle) as well as examples of what appeared to be the dodecameric form of DyP viewed along the twofold axis (red circle), where two black circles were surrounded by white density. There were also examples of what appeared to be one visible DyP hexamer density (green circle) and what looked like a slightly angled pair of DyP’s (pink circle). Interestingly, a few classes showed what appeared to be three DyP densities with threefold symmetry forming a triangle shape (blue circle), but it was not possible to confirm whether this was an actual conformation of DyP inside of the nanocompartment. Resulting initial model yielded was at 17.49 Å with I1 symmetry (Fig. 23C). In Fig. 23C, the cross-section view of the initial model shows the beginning of an interior density at a lower threshold, but it was unclear whether this density was artefactual or represents protein. When this initial model was 3D refined to 13.01 Å (Fig. 23E), the density was no longer present at any thresholding level, and it seemed that the density could have been lost during the application of symmetry during the 3D refinement step. When 3D classification was carried out using this map as the reference (Fig. 23E), two out of the four classes kept the shape of the nanocompartment. When looking at the grey and yellow classes, a structure began to form when the threshold was low. However, this structure was not visible when processing was take further.

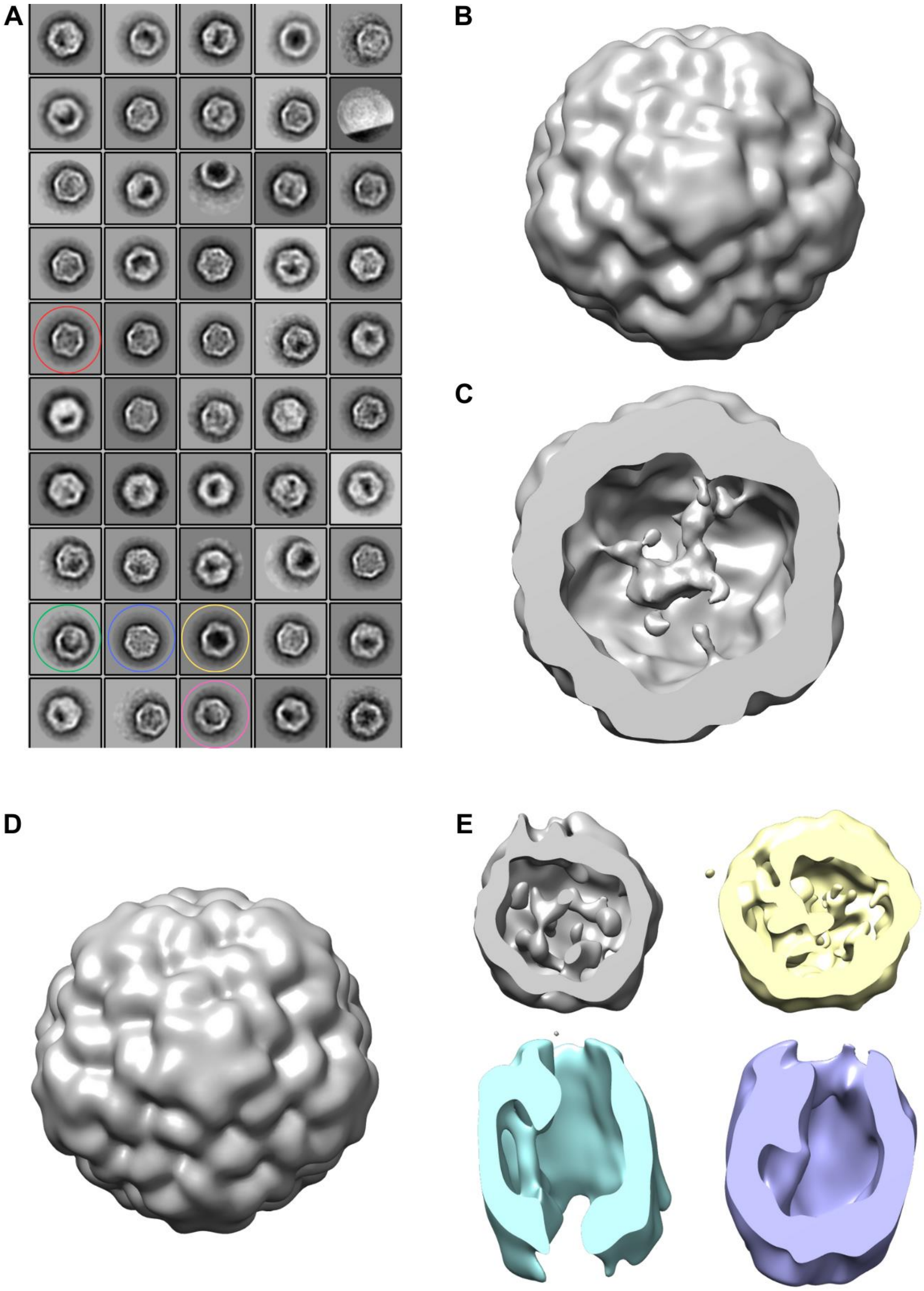


Figure 24. Representative 2D classes and reconstructions. (A) All the 2D classes from an initial round of 2D classification used in a 17.4952 Å initial model. The classes of interests are circled. Red indicates two DyP along the twofold axis. Green inidicates one DyP. Blue indicates three DyP. Yellow

indicates an empty nanocompartment. Pink indicates an angled view of two DyP. (B) 17.4952 Å initial model. (C) A cross section view of the 17.4952 Å initial model. (D) 3D refined model of the Enc shell at 13.01 Å. (E) Interior views of the 3D classes made using the 13.01 Å model as a template.

The reconstruction statistics of the maps generated are included in Table 2.

	Native Enc shell	Native bacterioferritin
Data collection		
Microscope	F20	F20
Voltage	200	200
Magnification	14 500 x	14 500 x
Detector	DE-16	DE-16
Data collection software	Serial EM	Serial EM
Defocus (μm)	-1.5	-1.5
Electron exposure (e^{-} /Å^2)	40	40
Pixel size (Å)	3.02	3.02
Map reconstruction		
Number of micrographs:	634	634
Initial particles	2 952	1 035
Final particles	2 475	506
Symmetry	I1	O
Resolution (Å)	13.01	18

Table 2. Data collection statistics for native *M. tuberculosis* Enc and bacterioferritin

The next step was to be more selective in the process of picking 2D classes. Classes that showed better resolution and DyP features were selected and yielded a set of 498 particles (Fig. 25A). This was used to make an initial model with I1 symmetry (Fig. 25B). This was refined, yielding an 18.8 Å map (Fig. 25C).

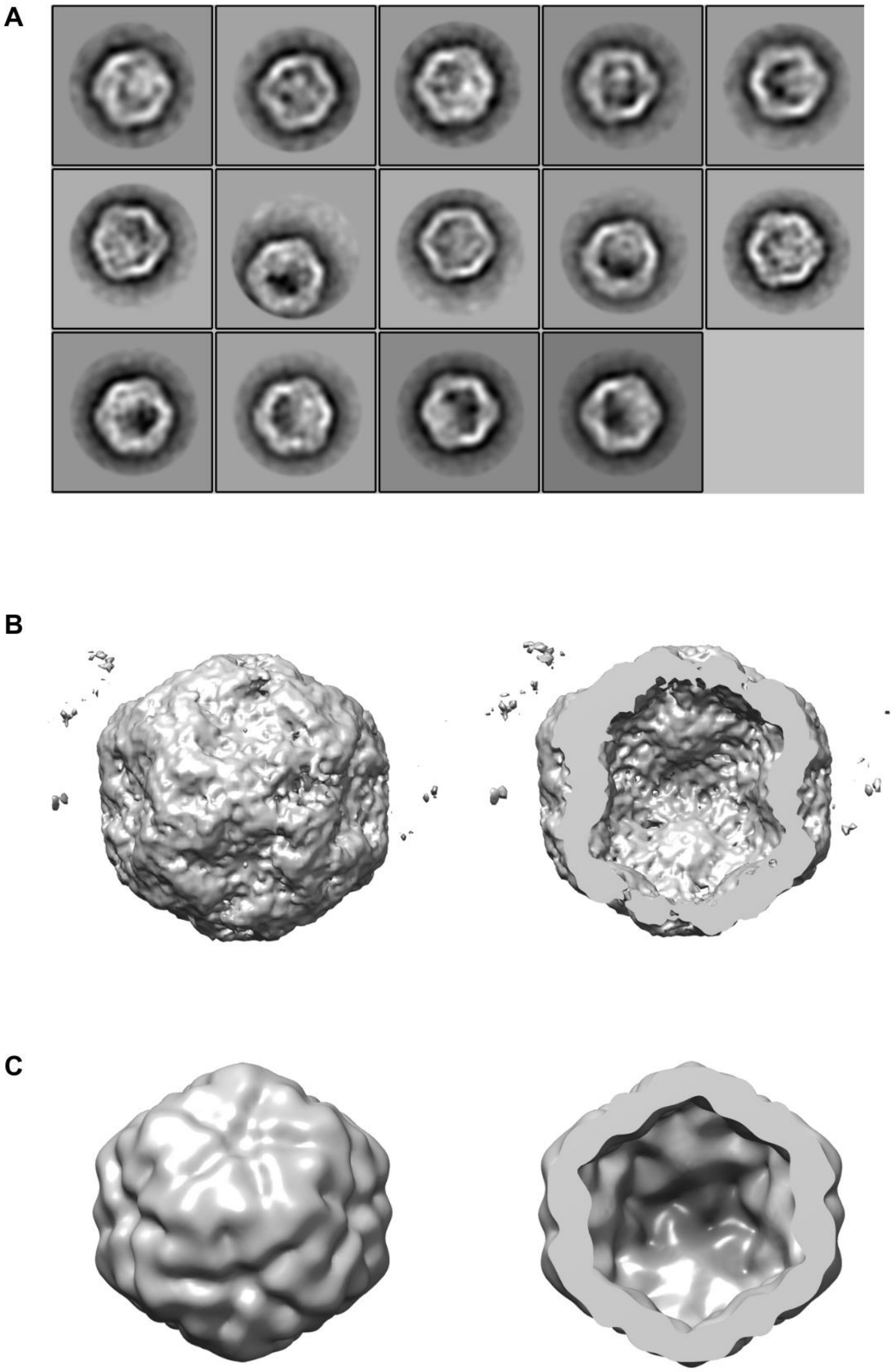


Figure 25. Initial reconstructions obtained when being more selective during 2D classification. (A) Representative 2D class averages of the 498-particle stack. (B) 9. 57 Å initial model and cross section with I1 symmetry imposed. (C) 18.8 Å refined map and cross section.

This 18.8 Å map was used for 3D classification with the total particle stack of 2 952. The initial and the refined maps had Enc shell features at low resolution but did not show any convincing DyP features (Fig. 26A and B). The 3D classes showed what could be DyP, especially the class averages when viewed in Relion, with blurry density that could be DyP present within 2 of the classes (Fig. 26A). The class that had the best resolution and DyP features, the first class from left to right in Fig. 26A and the grey class in Fig. 26B, was refined but no density was clear within the reconstruction (Fig. 26C). Due to the limit of resolution imposed by negative stain microscopy, there were no clear examples of DyP present in any of the reconstruction attempts.

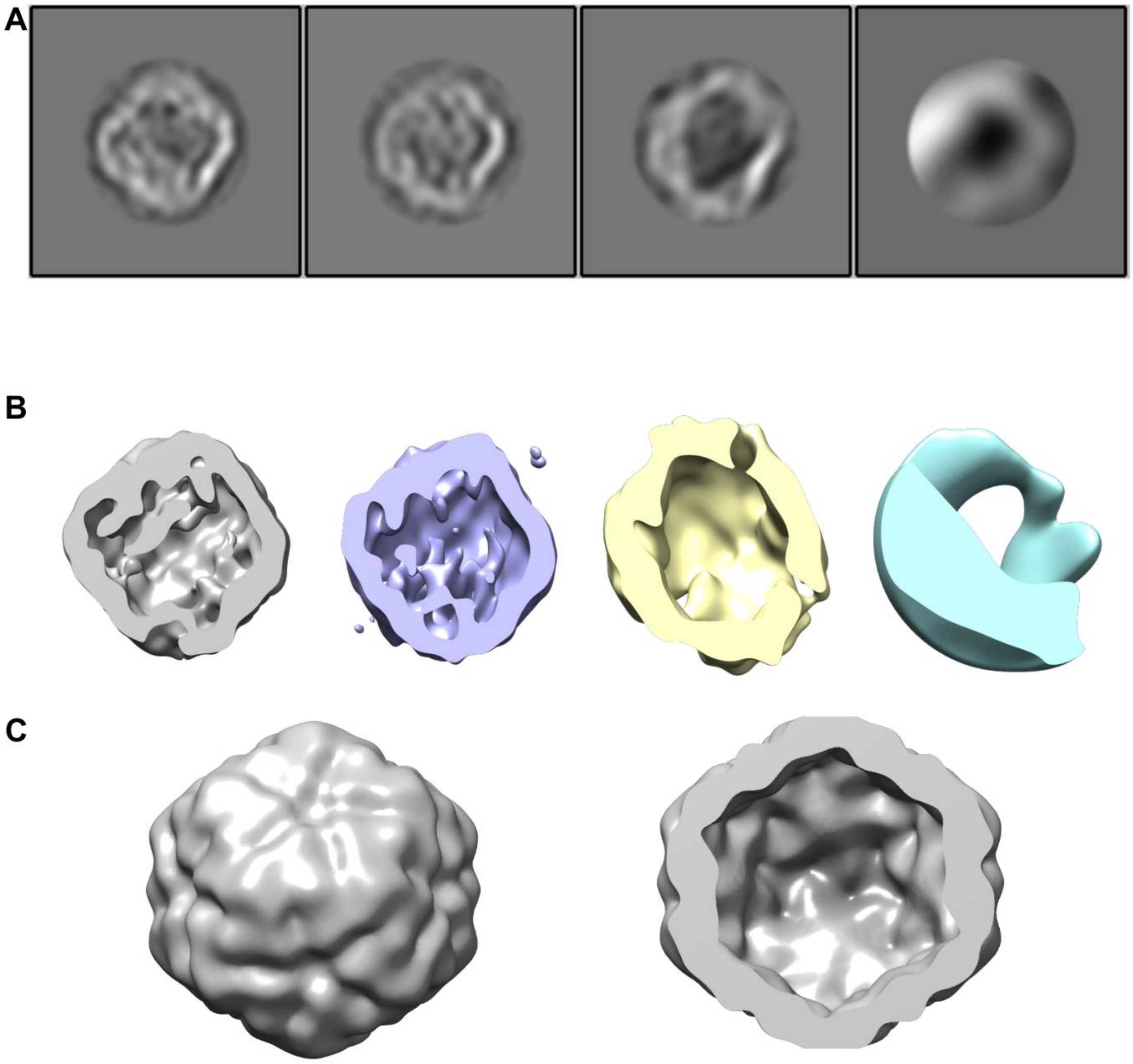


Figure 26. Further reconstructions obtained using the 18.8 Å reconstruction. (A) Class averages obtained from the 3D classification using the 18.8 Å map as a template with two out of 4 classes showing possible DyP features. (B) Cross

section view of the representative density maps of the class averages. (C) Refined map using the particles from the grey 3D class.

6.2.2. Native bacterioferritin.

A bacterioferritin complex was thought to be present in the micrographs so a set of 1 035 particles that appeared to be around the correct size were manually picked. A 2D classification was carried out on these particles with most of the 2D classes present having ferritin-like features (Fig. 27A). These 2D classes were compared to the classes obtained by Kirykowicz and Woodward (25) and this was used as the basis of their selection for use in a reconstruction that yielded an initial model with O symmetry that was 3D refined to 18 Å (Fig. 27B). This model had comparable low resolution features when compared to an existing recombinantly expressed ferritin model produced by Gijsbers *et al.* (86) (Fig. 27C). The selection of many proteins around the same size and shape as ferritin proved to be an adequate way to go about confirming the presence of a protein within a sample. It has been shown previously that *M. tuberculosis* ferritin co-purifies with Enc when the two are recombinantly expressed within *E. coli* (38), but the presence of ferritin inside of Enc nanocompartments has never been confirmed using structural analysis. There are no clear examples of ferritin present within any of the 2D classes shown in Fig. 24A and ferritin only appears to be present in the sample outside of the Enc nanocompartment.

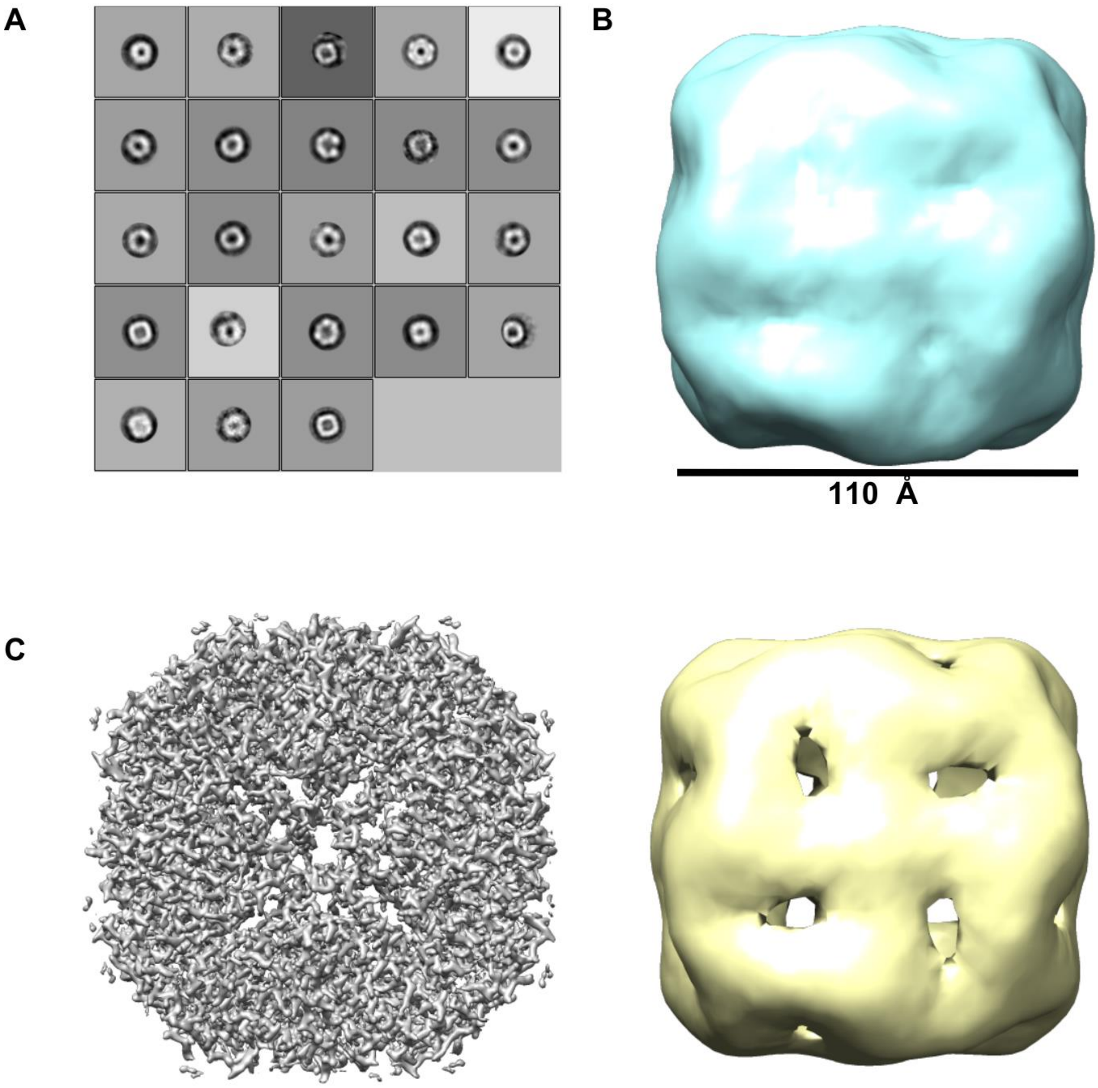


Figure 27. Particle processing of native *M. tuberculosis* ferritin. (A) 2D classes of all the particles picked thought to be ferritin. (B) External structure of the ferritin reconstruction with accompanying scale bar. (C) Comparison of external reconstruction at a threshold that shows the octahedral symmetry (Yellow) to an existing structure of ferritin (Grey) (7O6E)(86)**.**

Discussion

Expression and purification.

The plasmid constructs used in the recombinant expression of both proteins (Fig. 5) led to Enc and DyP being present in both the IMAC elution curve obtained (Fig. 6B) as well as bands on the subsequent SDS-PAGE gel analysis (Fig. 6A). The presence of both Enc and DyP in the same sample would strongly indicate an interaction between them as the plasmid construct used to express both proteins was constructed in such a way that only Enc would be Histidine-tagged and would therefore bind to the column (Fig. 5E). The genes for Enc and DyP are part of the same operon in the genome of *M. tuberculosis* and are therefore expressed together (28). When the SDS PAGE gel results and the plasmid construct are put together it shows that DyP is associated with the Enc nanocompartment. The ratio of DyP to Enc being 1:4 is similar to the ratio dodecameric DyP to Enc. This could possibly indicate that a range of DyP forms were associated to the Enc nanocompartments in the samples shown in Fig. 6A and B. This ratio varies depending upon which form of DyP is most present within the sample, which in the case of the *M. smegmatis* would be the dodecamer (1 DyP: 5 Enc) and the hexamer (1 DyP: 10 Enc). While these forms were not seen in the purification data of this thesis, it is unclear why those forms were not present. The most abundant form of DyP when expressed by itself was found to be the monomer followed by the dimer and then the tetramer. This would indicate that there is potentially a difference when *M. tuberculosis* DyP is recombinantly expressed in *E. coli*. The lack of hexamer and dodecamer in the sample did not mean that the three forms obtained were not catalytically active. The catalytic study done by Wiswedel (54), who was working with the same sample as this thesis, showed that the DyP was still active against a variety of substrates. The catalytic activity taken together with the purification of both Enc-DyP and DyP alone shows that whilst functional, the formation of recombinant DyP does not occur in the same way as in *M. smegmatis,* where the three forms of DyP obtained were the monomer, hexamer and dodecamer (24).

The suitability of recombinant expression in *E. coli* for structural studies is not ideal as the proteins tend to misfold and become insoluble with Goldstone *et al.* (87) reporting that under half of the *M. tuberculosis* proteins that are recombinantly expressed in *E. coli* are present in a soluble form. This could be due to the lack of a compatible molecular chaperone being present. The chaperones that could be responsible for correct folding in *M. tuberculosis* are not dependent on ATP with an example of this being Cpn60s (88) and this could indicate that the chaperone proteins in the *E. coli* and *M. tuberculosis* could be different. A potential way to solve this issue is to explore recombinant expression of *M. tuberculosis* DyP and Enc using *M. smegmatis* expression systems or to discover what chaperone protein is responsible for proper folding and recombinantly express it along with the DyP and Enc in an *E. coli* expression vector. The Hsp60 chaperone GroEL1 found within *M. smegmatis* was shown to co-elute and with proteins that had His-tags present and it was shown that mutating the Histidine-rich C-terminus could improve purification while still being a functional protein (89). This allowed recombinant expression and purification of His-tagged *M. tuberculosis* proteins in *M. smegmatis*, thus mimicking native conditions. Further genetic work would have to be done to prevent the naturally occurring Enc and DyP from being produced in the *M. smegmatis* expression host so that the *M. tuberculosis* versions of the proteins could be expressed.

As it stands, there is no structure for a DyP-type peroxidase in *M. tuberculosis*, let alone one for encapsulated DyP. There has been more success with expressing ferritin recombinantly in *E. coli* with Ross *et al.* (41) recombinantly expressing an encapsulated ferritin from *H. ochraceum* and obtaining a density map. A density map was also successfully achieved with recombinant expression of *M. tuberculosis* BrfB in *E. coli* (86).

Cryo-EM reconstruction of Enc naocompartment.

The initial approach to the processing was that the DyP would be visible inside at least a few of the initial 2D classes when picking the best micrographs and particles, with these being used in further reconstruction steps. The only high-resolution cryo-EM map generated was that of the Enc nanocompartment shell and was at 2.7 Å (Fig. 8B). When analysing the local resolution map of EMD-18031, the region around the five-fold pores had the lowest resolution. This has been postulated to be due to the flexible nature of this pore (41). Ross *et al*. (41) reconstructed the open and closed confirmations of the five-fold pore found within the Enc nanocompartment of

H. cohraceum (Fig. 28) and this could be the reason for the difference in local resolution around the five-fold pore of EMD-18031.

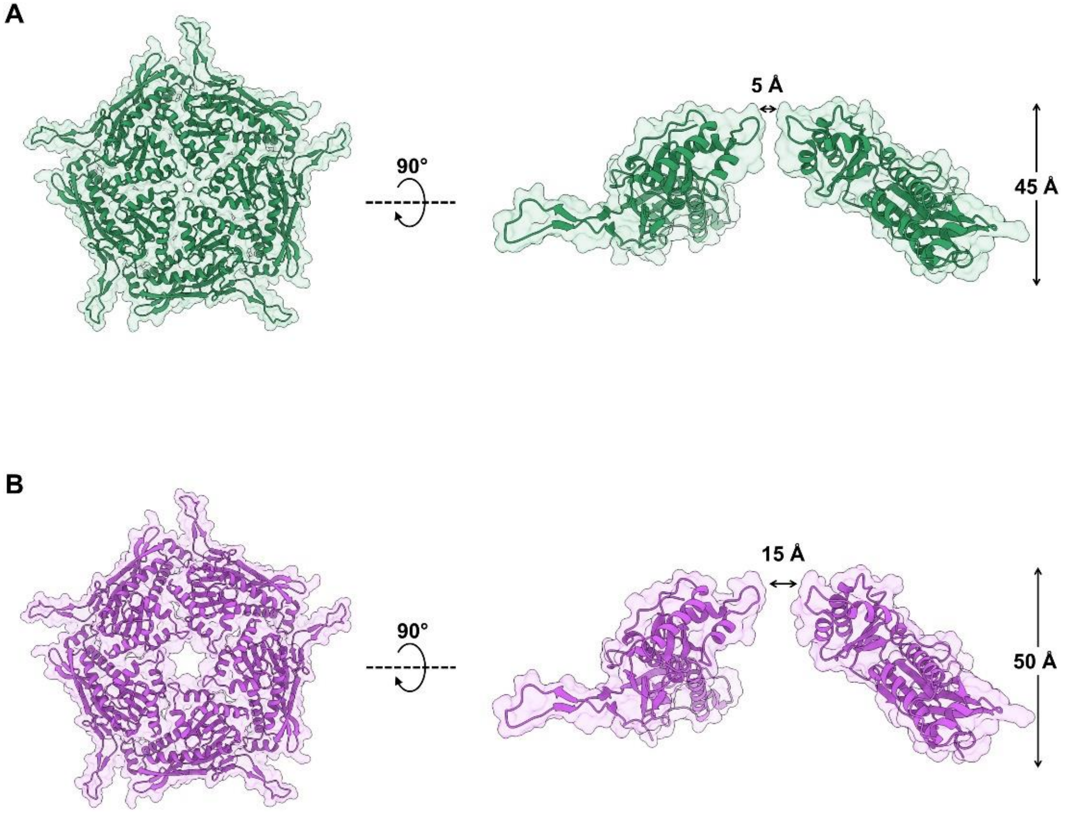


Fig. 28. Figure reprinted from (41)**. Open and closed conformations of the dynamic fivefold pore of the H. ochraceum encapsulin shell. Masked 3D refinements centered on the fivefold pore of the icosahedral reconstruction were performed after symmetry expansion of the asymmetric units. The closed [(A), green] and open [(B), purple] conformations of the Loaded-Enc shell pentamer are represented with transparent solvent-accessible surface rendered over cartoon secondary structure. Both closed and open pentamers are shown in top-down (left-hand side) and side-on views (right-hand side). The closed conformation has a fivefold pore diameter of 5 Å, while the equivalent diameter of the open conformation is 15 Å. To allow for the widening of the pore, the open pentamer protrudes from the nanocompartment by 5 Å.**

No clear trace of DyP was detected throughout the process of generating this map (Fig. 8B). This is most likely due to the stringency of the initial selection criteria which meant that the number of micrographs and particles used was potentially not large enough to highlight any asymmetric features present but was very good at generating a high-resolution Enc shell. A paper attempting to generate a structure of a TP tagged heterologous protein loaded *T. maritima* Enc used a similar approach in their attempt to generate a cryo-EM map containing the encapsulated cargo protein

(90). This paper used a similar particle stack of 112 241 when creating their initial reconstruction and produced a refined map of 2.87 Å (Fig. 29A). The difficulty of visualising cargo proteins within Enc nanocompartments is illustrated by their results, where they reconstructed the Enc shell but were unable to obtain the cargo protein within. However, they showed 2D classes where density could be seen on the interior of the nanocompartment (Fig. 29B). paper contained examples of density in their micrographs and 2D class which indicated the presence of cargo protein, unlike what was found during this thesis.

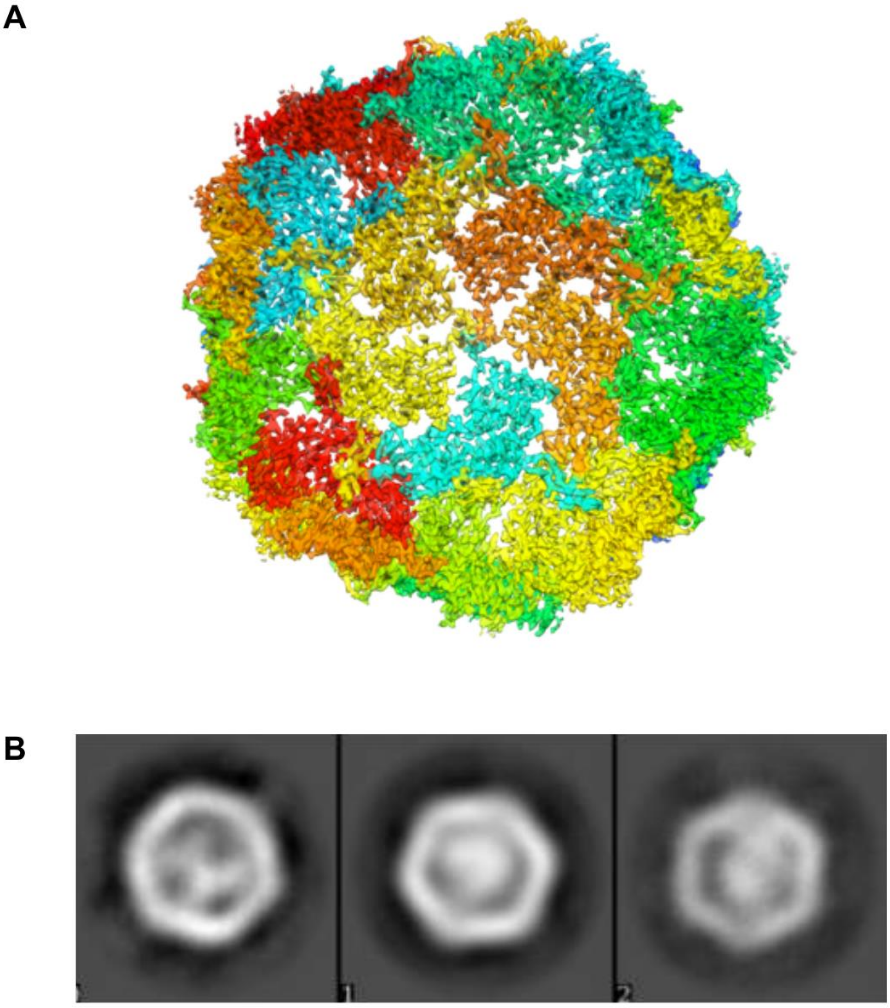


Figure 29. Comparative results from the cryo-EM processing of *T. maritima* Enc. Figures reprinted from (90)**. (A) The 2.87 Å density map of *T. maritima* Enc (EMD-22617). (B) Representative 2D class averages that were used to obtain EMD-22617**

This shared method has been successful in other cases. A density map showing ferritin inside the Enc nanocompartment of *H. ochraceum* was produced by applying no symmetry to reconstruct the nanocompartment followed by taking the 3D classes that showed four distinct densities in a tetrahedral fashion, with the best class from this 3D classification were then used to generate a refined reconstruction (Fig. 29) (41).

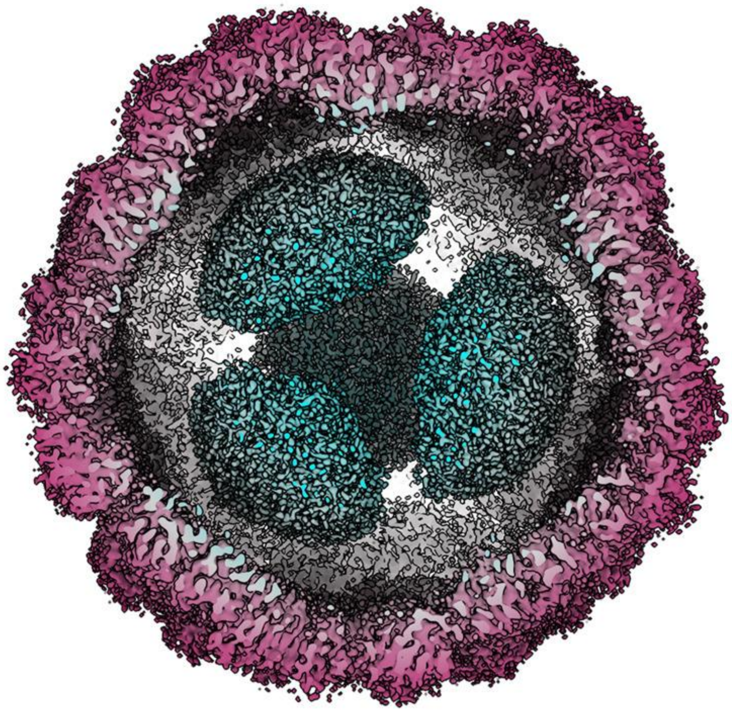


Figure 30. Figure reprinted from (41)**. Radially coloured cryo-EM–derived map of the Loaded-Enc nanocompartment displaying the interior EncFtn (cyan).**

Model building, refinement, and comparison to other models.

The next steps were to produce a monomeric model of Enc. This was accomplished by manually mutating the Enc from *M. smegmatis* residue by residue in Coot with and without fitting yielded an Enc monomer that was missing secondary structures found in *M. smegmatis*. The fitting in Coot upon amino acid mutation was not optimal and as a result, the secondary structures that were most obviously present needed to be imposed in Phenix real space refinement which led to more acceptable statistics and overall fit. This discrepancy when manually mutating each residue is most likely caused by overfitting carried out by Coot whereby the secondary structure is abolished when using the ‘Mutate and Auto Fit Rotamer’ function as Coot will automatically adjust the structure to fit the map. The clashes found within the monomer were largely due to the addition of Hydrogens to the model during the evaluation process and were not able to be adjusted as any changes caused more clashes. The density map created during this thesis and used during the building process, EMD-18031, was not at a high enough resolution to be able to predict hydrogens as the resolution needed is approaching 1.2 Å (91) and this means they were removed when carrying out final statistical calculations.

The manually mutated and real space refined structure (8PYS) was compared to the one found within *M. smegmatis* (7BOJ) and the Alphafold structure linked to the Enc Uniprot entry (I6WZG6). The secondary structures were similar when compared, with minor differences in the length of certain regions. The differences in amino acid sequence between the Enc of *M. tuberculosis* and *M. smegmatis* resulted in instances of secondary structures being shorter in 8PYS. The TP binding site of the P-domain remains largely conserved (Fig. 10G and 11A). This is important to note as the conservation of this region could mean that the Enc nanocomaprtment of *M. tuberculosis* encapsulates a cargo protein that possesses an appropriate TP, as is the case in the numerous examples described by Sutter *et al*. (28). One of the notable changes occurs in the A-domain. This change abolished a β-sheet and the amino acid changed to a negatively charged aspartic acid when compared to the more neutrally charged alanine. This region of the monomer possessed the highest number of residues not fitting the EMD-18031 density map, most likely due to the previously discussed lower resolution around the five-fold pore.

However, the overall structure of the A-domain, P-domain and E-loop did not change, and it can be inferred that the major differences between 8PYS and 7BOJ that affect how the Enc functions are at the primary structure level of each protein.

Multimeric structure

In general, the size of the Enc naocompartment five-fold pore found in many species has been shown to be around the same diameter, which would be considered a limiting factor for the functioning of any encapsulated enzyme. But Rhamanpour and Bugg (37) showed that there was peroxidase activity against lignin when they were testing the encapsulated DyPB present in *R. jostii*. This same study also showed that the Enc nanocompartment went from a diameter of around 220 Å to a diameter of around 310 Å when exposed to pH 3 conditions. This structural flexibility could indicate that the pore is indeed able to be enlarged to allow large substrates to pass through but no such examples of this were seen in any of the 3D classes generated by this current study. If the pores can be enlarged in this way, it could explain the activity of encapsulated DyP against various substrates when there is a high concentration of hydrogen peroxide bringing the pH down (24,37,38). In Contreras *et al.* (38), the activity of encapsulated DyP seemed to be decreased but still present when compared to the unencapsulated version. This could be explained by the decreased presence of an external stimulus to mediate the opening of the pore as was shown by Ross *et al.* (41) where the fivefold pore of *H. ochraceum* was shown in both closed and open conformations with the pore diameter going from 5 to 15 Å. This same Enc nanocompartment strongly interacts with Tb^{3+} ions on the outside during simulations of molecular dynamics and this could indicate the need for a stimulus to bind to the shell to enable opening. Another example of molecules binding to the exterior of the Enc nanocomompartment shell is the flavin binding pocket present in *T. maritima* (40,60), with the exact function of this interaction unclear. But it is hard to discern exactly what could cause this change in confirmation without knowing the exact conditions that the nanocompartment finds itself.

It is therefore thought that the charge and hydrophobicity of the amino acids that are part of the five-fold pore play an important role in allowing substrates into the nanocompartment. When comparing the 8PYS 60-mer and the 60-mer of 7BOJ, some differences emerged. Despite the size of the fivefold pore in the *M. tuberculosis* nanocompartment being similar to the one in *M. smegmatis*, there are differences when comparing the amino acid residues around the pore. The charge distribution around the pores in *M. tuberculosis* is more negative than in *M. smegmatis* due to a single change from an alanine to an aspartic acid with the rest of the amino acids being neutral (Fig. 13A and 14A). The conservation of the ring of histidines could be aiding the function of DyP inside the nancompartment as, in theory, DyP is the most likely cargo protein and would need the correct charge to be present at the fivefold pore to aid in moving substrates into the nanocompartment. However, the charge of the histidine residues is not clearly visible in the charge distribution of the fivefold pore (Fig. 14A), and this could mean that the histidine may not play a large role in the import of substrates into the lumen of the nanocompartment. Contreras *et al*. (38) postulated that the five-fold pore of the *M. tuberculosis* Enc nanocompartment may be predominantly positively charged based upon the sample containing Enc and DyP having minimal activity against neutral guaiacol and instead having more activity against negatively charged ABTS. With this structure, it is not certain that charge is the sole selective measure by which substrates are chosen. Another factor that is thought to aid the import of substrates into the nanocompartment is the interior arrangement of the amino acids of the five-fold pore. This arrangement of amino acids is similar when comparing *M. tuberculosis* and *M. smegmatis* as it has lines of negative charged amino acids which could be used to funnel the negatively charged substrates further into the lumen of the nanocompartment by means of opposite charges (Fig. 14B). *T. maritima* also has a similar arrangement of charges but with a more neutral and positive overall charge directly around the interior of the pore (due to the histidine's) with the negative charges being used to funnel Fe^{2+} ions towards the encapsulated ferritin (30).

Another such selective factor of the five-fold pore is the hydrophobicity of the amino acid residues. The fivefold pore in *M. tuberculosis* is comprised of mainly polar, hydrophilic amino acids, which could explain the selectivity for ABTS instead of guaiacol when testing encapsulated DyP activity. ABTS has a solubility of 50 mg/ml compared to guaiacol which has a solubility of 30 mg/ml. Pyrogallol, with a solubility of 625 mg/ml, was tested as a substrate for the DyP in the Enc nanocompartment of *M. smegmatis* and a negligible difference was seen when compared to the DyP alone with the main difference being a lagging reaction curve which initially required more H_2O_2 to be at the same point as the DyP by itself before overtaking it (24). This could mean that substrate selectivity could be influenced by how hydrophilic said substrate is and the encapsulation of DyP could help it to remain active at higher concentrations of H_2O_2. The potential encapsulation of BfrB could explain the negative charge surrounding the fivefold pore and its polar amino acids as it would enable the funnelling of positively charged iron into the nanocompartment to be mineralized. Contreras *et al*. (38) did prove that BfrB can be encapsulated when a C-terminal TP is added but it does not have a naturally occurring C-terminal TP and is not part of a two gene operon with Enc like DyP. However, there are certain cargo proteins that are not necessarily in a two gene operon with Enc, but are still encapsulated, which are called secondary cargo (19). Ferritin-like proteins could be such a protein in *M. tuberculosis*, but there are no concrete examples showing their encapsulation yet.

Examples of other Enc shells in *Q. thermotolerans* (36) and *M. xanthus* (92) have negatively charged fivefold pores which are key in attracting and allowing Fe^{2+} into the lumen of the nanocompartment to where the encapsulated ferritin is present. Despite encapsulating ferritin as well, *T. maritima* has a neutral and hydrophobic five-fold pore with the interior of the pore being negative (Fig. 31A) (28,40). The conserved histidine's have a high affinity towards Fe^{2+} (30), with these nanocompartments found to encapsulate ferritins that act upon iron free radicals which have net positive charges and are therefore shuttled into these nanocompartments. The hydrophobic pore should hinder the influx of Fe^{2+} into the lumen but is still hypothesized to be the most likely of its three types of pores for the Fe^{2+} to flow through. In contrast. *S elongatus* possesses positively charged residues that protrude outwards as spikes (Fig. 31B) (70).

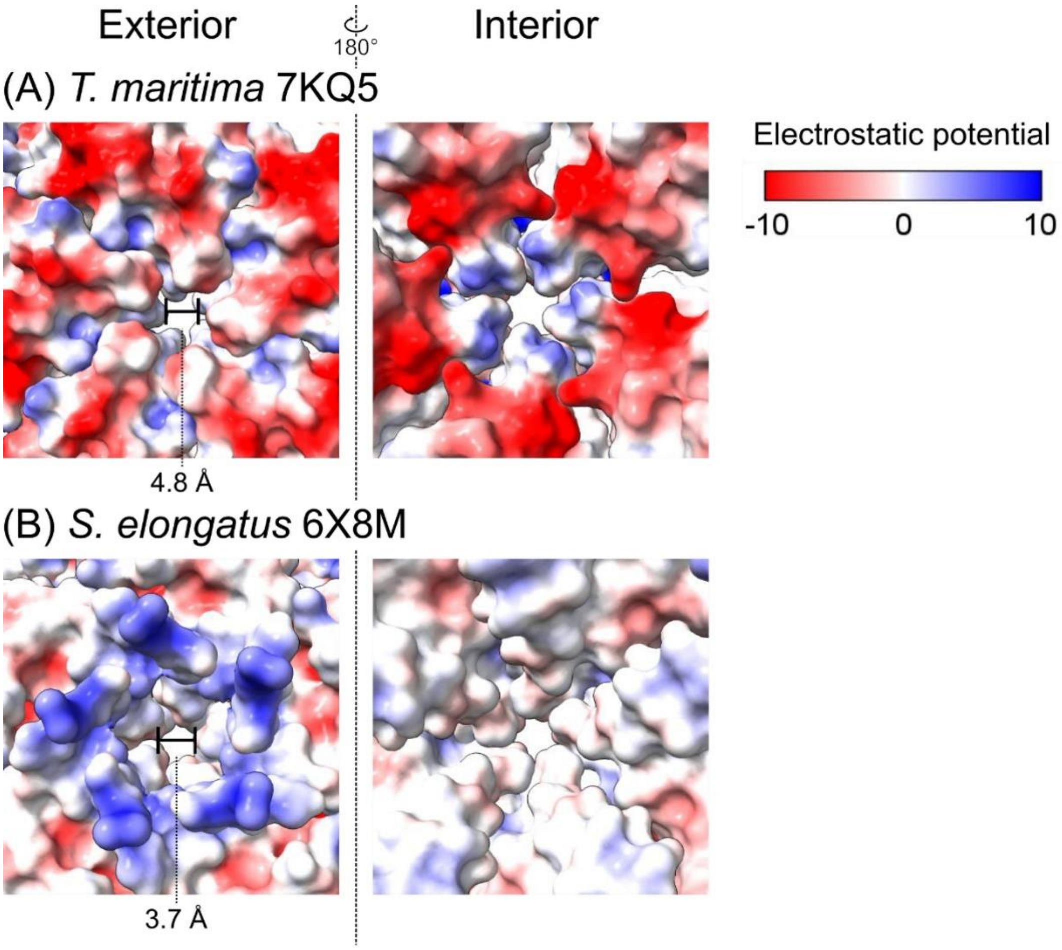


Figure 31. Reprinted from (30)**. Pores at the fivefold axes of (A) T. maritima, (B) S. elongatus, and (C) M. smegmatis encapsulins. The figures in the left column are the views from the exterior of the encapsulins and the ones in the right column are views from the interior of the encapsulins. The surface representation is colored by electrostatic potential where blue is positive, and red is negative. The measurements in the center of the figures indicate the diameter of the pores in Å.**

Correlation of primary structure to catalytic study carried out by Wiswedel (54)

The charge and hydrophobicity of the five-fold pores seem to impact the choice of substrates. When Wiswedel (54) tested the activity of DyP against guaiacol in the presence of H_2O_2, less tetra-guaiacol was formed than expected. There were also questions raised around the use of dyes as a substrate for DyP, with no trend emerging when a variety were tested with Enc and DyP together in the same sample. Interestingly, 2,6-dimethoxyphenol was shown to be catabolised by DyP with this catabolism completely abolished in the Enc and DyP sample. This substrate has a solubility of 20 mg/mL and is less soluble than ABTS and guaiacol, which could once again point to the selection of more hydrophilic substrates, as mentioned previously. The Contreras *et al*. (38) catalytic results using guaiacol obtained similar results. These differences in encapsulated vs unencapsulated function were also tested by Lien *et al.* (23), who saw that encapsulated DyP provided much greater protection against H_2O_2 treatment when compared to unencapsulated DyP which points to some sort of stabilising effect. The nanocompartment pores could also be imposing a limit on the flow of H_2O_2 into the nanocompartment which protects the DyP from being overwhelmed and degraded as has been seen when DyP is treated with higher levels of H_2O_2 (43).

These results taken together raise the question of whether the difference in guaiacol conversion is due to decreased DyP activity or due to a lack of dissolved oxygen gas, which is needed to interact with guaiacol to form tetra-guaiacol.

This catalytic research carried out by Wiswedel (54) concluded that it was unclear whether or not the recombinantly expressed DyP used in the catalytic tests was actually responsible for the decomposition of the various substrates tested. The common consensus in the literature is that DyP needs a substrate to oxidize so it can return to the resting state and without a substrate, it would catalyse the decomposition of H_2O_2 but would be stuck in the Compound I form (Fig. 2). DyP was able to catalyse the conversion of H_2O_2 without the presence of a substrate such as guaiacol or ABTS to act as an electron donor, which would indicate that the DyP could function like a catalase and use the electron from breaking down a second molecule of H_2O_2 to return it back to the resting form (53). This catalysis of the decomposition of H_2O_2 in the presence or absence of a substrate such as ABTS to act as the electron donor could indicate that DyP can function either way depending on the conditions. The pseudo catalase mechanism (Fig. 3) mentioned previously could account for the H_2O_2 decomposition in the absence of a substrate with the inactivated DyP being the white precipitate seen when excess H_2O_2 is added to the DyP solution. It could be favourable for the DyP to aid in production of an oxidized substrate as a source of food as well as catalysing the decomposition of H_2O_2 for the bacterium in the challenging conditions of the granuloma. It may also allow the bacterium to still decompose H_2O_2 in the absence of a substrate at the expense of some of the DyP while also producing O_2 as a by-product. The purpose of DyP could be to protect the bacterium from oxidative stress.

The physical constraints of the fivefold pore size means that some sort of stimulus would still be needed in order for the pores to open, regardless of the amino acid composition. This is should most likely be the next step in figuring out the functioning of this system.

Further attempts at reconstructing DyP inside of the nanocompartment.

When it became clear that the DyP would a reconstruction of encapsulated DyP was not forthcoming whilst reconstructing the Enc nanocompartment, multiple methods were used in attempts to highlight DyP. The first attempt at bringing out DyP was done using a mask of the inside of Enc however this didn't yield a density that resembled Enc. This was done to mask out any density from outside of the nanocompartment and to focus more upon the inside. The density within the nanocompartment only appeared at very low threshold values and was not convincing.

The second was an attempt at creating a small, localised, and spherical mask to focus on the signal specifically where it was thought that DyP and the TP would be found. The hallmark of a TP that targets a cargo protein into an encapasulin nanocompartment is one that is rich in alanine, proline and glycine residues and contain either one or two Glycine-Serine-Leucine motifs (28). The TP is highly flexible and difficult to visualize with a large amount of variety in binding strength and specificity (93). This is even more applicable when the cargo protein occupancy of the nanocompartments seems to be incredibly low or non-existent. But it was thought that the hydrophobic pocket of the P-domain that the DyP is thought to bind to with strong ionic and hydrophobic bonds (93) was an ideal target of focused classification. The amino acid residues that form part of the hydrophobic pocket and that have previously been shown to make contact with the TP of the cargo protein in *H. ochraceum* (41), are largely conserved when looking at *M. smegmatis* and *M. tuberculosis*, with the two changes being switches to other hydrophobic amino acids.

The exact contact point of the DyP within the nanocompartment of *M. smegmatis* is not clear as the dodecamer and hexamer were shown to be flexible (24), but Kirykowicz and Woodward (25) saw that DyP density was present at the 3-fold axis and could feasibly be binding to the hydrophobic pocket which is part of the 3-fold symmetry axis.

Focused classification of this region was coupled with symmetry expansion. Symmetry expansion is the process of assigning every symmetry-related orientation to each particle within a dataset (94). In a model with a certain symmetry such as the nanocompartment, symmetric and individual elements within the structure are sometimes heterogenous in conformation, which therefore causes the complex to be asymmetric. In such an example, every unit of density is separated from the overall structure and handled as individual particles that are connected by the complex's overall symmetry. Symmetry expansion of the dataset is accomplished by using the corresponding transformation on the experimental 2D images. The transformation is calculated using the complex's overall symmetry taken together with the subtraction. Any asymmetric element is then obtained by way of combining 3D masked classification and refinement using only local angular searching (95). This is to allow every asymmetric unit to contribute to the map equally (94). This technique is a widely used way of resolving asymmetric features, particularly when the overall symmetry of the reconstructed protein is icosahedral. Local refinement was also used to allow the particle to only rotate a small amount to increase the chances of seeing the asymmetric feature (in this case DyP). This is a widespread method and has shown to be successful in Ross *et al.* (41) where the pore dynamics of *H. ochraceum* were explored and a reconstruction of encapsulated ferritin was obtained using symmetry expansion and masked classification. However, in this thesis, this method did not yield even the TP of the supposed hexamer or dodecamer that DyP forms inside Enc. This is not an uncommon problem with cargo proteins or the TP of the cargo protein. The lack of a TP density can often be due to low occupancy or can be lost when symmetry is imposed during reconstruction (30). The issue with these methods most likely came about prior to masking, during the symmetry specifications throughout the reconstruction of EMD-18031. The application of I1 symmetry may have averaged out any hint of DyP present within the data.

Utilising Tang *et al.* (24) construction pipeline.

Once these attempts had proven unsuccessful, the successful pipeline that was used to reconstruct encapsulated DyP from *M. smegmatis* was attempted (24). When following this pipeline, the biggest problem encountered was the lack of clearly identifiable DyP within both the micrographs and in the resulting 2D class averages. At both of those stages of processing, DyP was visible within the *M. smegmatis* dataset (Fig. 23A) and one class average from a subsequent 3D classification step showed good DyP features. Even in the negative stain dataset of Kirykowicz and Woodward (25), DyP was visible within the 2D class averages (Fig. 23B) and was able to be reconstructed within the nanocompartment. The 2D classes and particles obtained during the recombinant expression portion of this thesis showed very little in the way of DyP features (Fig. 8A and 19) and is in line with the work done by Contreras *et al.* (38), from which the methods of purification were taken. The particles shown in that paper do not convincingly show the hexameric or dodecameric form of DyP seen in *M. smegmatis,* but the SEC and SDS-PAGE gel results mirror the results of this thesis. There are examples of successfully encapsulated proteins when using recombinant expression. H. ochraceum (41) and R. rubrum (59) Enc-ferritin systems have been recombinantly expressed within E. coli and both led to fully formed complexes. The recombinant expression of M. tuberculosis of BfrB has also led to a crystal structure (61), although the encapsulation of this BfrB has not been confirmed yet. The lack of a clear DyP in this recombinant dataset indicates that there is an issue with the formation of DyP when recombinantly expressed.

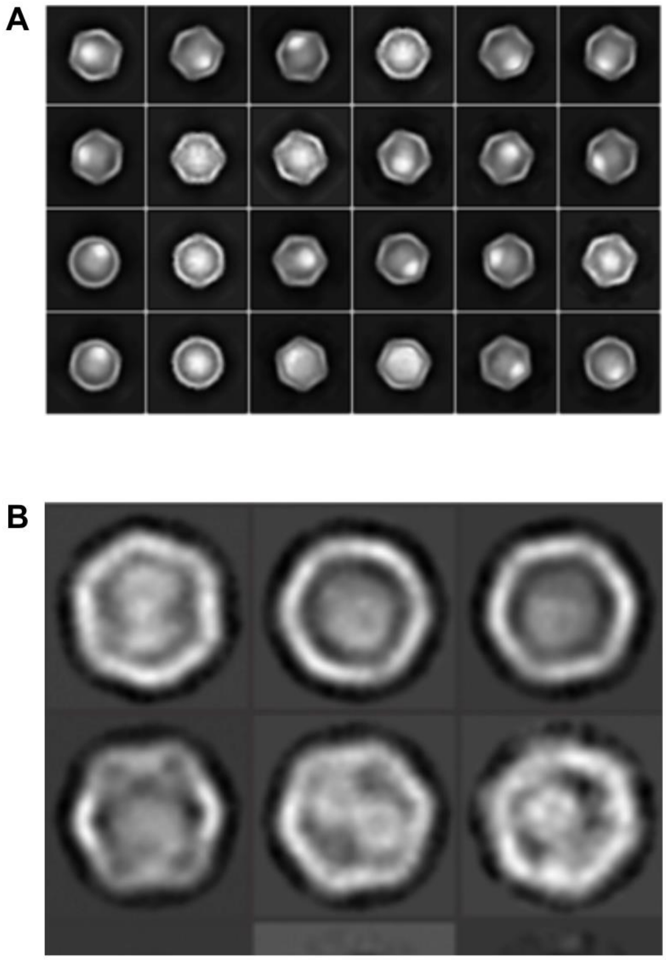


Figure 32. Examples of 2D classes from selected works involving Enc nancompartments. (A) Cryo-EM 2D class averages of M. smegmatis Enc taken from Tang et al. (24)**. (B) Negative stain 2D class averages of M. smegmatis Enc from Kirykowicz and Woodward** (25)**.**

M. tuberculosis native work.

Rationale.

All the attempts to generate a structure of encapsulated DyP had been unsuccessful. Other research that used recombinant expression of *M. tuberculosis* Enc and DyP had been carried out, but no maps have been obtained showing encapsulated DyP. The EM images obtained by Contreras *et al*. (38) of recombinantly expressed Enc and DyP do not clearly show the DyP hexamer or dodecamer inside of the nanocompartment, but the accompanying SDS PAGE gel shows a similar result to the ones obtained in this thesis. There appears to be density present, but it is not a similar size to the density found within *M. smegmatis*. Lien *et al.* (23) also carried out recombinant expression and more of the nanocompartments in their EM images showed what could have been DyP, but no further structural work took place. This shows that recombinant expression of these two proteins together leads to an association between them, but it is unclear in what form the DyP takes inside of the nanocompartment. The lack of a structure produced by recombinant expression led to the use of native expression and purification in *M. tuberculosis.*

Expression and purification.

There has been very little research that has been done into the structures of the native proteins found within *M. tuberculosis*. The difficulties primarily stem from the difficulty with culturing and the biosafety restrictions surrounding *M. tuberculosis* with many papers choosing to recombinantly express the proteins of interest in *E. coli* (23,38). The growing of the cultures took was one the least efficient portions of this process as the cultures needed to be in late stationary phase to ensure Enc had been produced. More research is needed on when exactly Enc is produced during the growth cycle of *M. tuberculosis* as this will streamline the process. Some methods employed to explore the proteome of *M.* tuberculosis heat inactivate the cells if they are looking to carry out Mass spectrometry (MS) (96) while others utilise bead beating to break apart the cells and to ensure there are few if any unlysed cells remaining (97). Neither of these was an option as the former would have affected the protein structure and function while the latter was not available, so sonication had to be used. The efficacy of this method could unfortunately not be assessed due to limited sample and safety concerns but proved to be an adequate method with the resulting sample yielding a membrane pellet after ultracentrifugation.

The purification pipeline that was obtained during the course of this paper was successful in obtaining examples of Enc nanocompartments at a relatively good purity given the limited sample and the questions around the exact method of purification required. More fine tuning is needed, but the groundwork for an efficient purification pipeline is present. While Enc did seem to bind to the IMAC column, the strength of this binding is unclear. There is a ring of histidine’s present around the five-fold pore, but it is unclear how strongly this would interact to the Nickel affinity column used. Lien *et al.* (23) utilised a sucrose gradient to purify native *M. tuberculosis* nanocompartments and had relatively good success but had to look at each fraction to determine exactly where the Enc was present. Despite this, Enc was purified and separated further by utilising an SEC column. The sample was not pure enough, and this led to the decision to utilise negative stain EM.

Negative stain reconstruction process.

Enc nanocompartment and DyP.

None of the 3D density maps generated from the native *M. tuberculosis* Enc sample had a clear example of the DyP hexamer or dodecamer that previous publications saw natively in *M. smegmatis*. There were 2D classes that showed what appeared to be different orientations of the dodecamer as well as a few that showed just the hexamer. There were also 2D classes that showed what appeared to be three DyP particles (Fig. 24A), but these did not show up clearly in any 3D maps that followed. When looking at the density map of the DyP within the nanocompartment present within *M. smegmatis*, a large density forms within the EMD-30132 density map before the DyP dodecamer forms when using the volume viewer threshold function. It is present right above the dodecamer and raises questions about the presence of a 3rd DyP.

When looking at all the 2D classes, there appeared to be a large amount of variability in occupancy of the nanocompartments and orientation of the DyP within (Fig. 24A). This is consistent with previous work carried out in *M. smegmatis* (24,25). Tang *et al.* (24) had a large number of 2D class averages when they carried out reference free 2D classification (Fig. 32A) while Kirykowicz and Woodward (25) also obtained different 2D class averages but saw less diversity in occupancy and orientation of DyP due to a low number of particles (Fig. 32B).

When compared to the 2D classes obtained by Tang *et al*. (24) and Kirykowicz and Woodward (25), the clarity with which the DyP can be seen is far less. The 2D class averages that were obtained by Kirykowicz and Woodward (25) are the most comparable as they are taken using negative stain and the presence of DyP in its dodecameric form is clear in a few of these classes.

Kirykowicz and Woodward (25) used different methods of purification followed by native PAGE and mass spectrometry in combination with molecular mass estimation from generated models as well as fitting using homologous structures in order to identify the proteins within their samples. When looking at the 2D class averages of the Enc particles in their research, there is a density present in the nanocompartment. This was summarily icosahedrally averaged in the reconstruction and produced an unknown mass which was then identified as DyP. The Enc was reconstructed and had C3 symmetry applied to it, and it was at this point that the density appeared more clearly. This finding by Kirykowicz and Woodward (25) deviates from what was obtained in this thesis as the 2D class averages show the presence of a density within the nanocompartment, but when these particles are used in a reconstruction, no such density persists, no matter which combinations of symmetry were attempted.

Upon seeing that the DyP was not going to be easily reconstructed inside the nanocompartment, a few different combinations of methods were used to enhance the DyP signal. Different combinations of symmetry were used to try and enhance the signal from the DyP. The first attempts used an initial model that had I1 symmetry and then the 3D classification runs used C1 symmetry to try and avoid the asymmetric DyP signal being lost, but there was no DyP clearly visible in any class obtained. Attempts using an initial model with C1 symmetry and 3D classification with C1 symmetry also didn’t yield any results. This would indicate that the occupancy of the DyP inside the Enc is highly variable and clearly the resolution of the data set was not adequate for showing DyP clearly.

A mask that encompassed the inside of the Enc nanocompartment with as little of the Enc included as possible was used and didn’t show a clear DyP particle. This was followed by a more selective 2D classification systemin an attempt to isolate only the particles that showed clearer DyP structures, with this also proving unsuccessful.

It appeared that the main problem holding back the processing of these particles was the fact that negative stain was used. The limit of resolution of negative stain is around 18 to 20 Å (98). It was thought the improved contrast provided by negative stain would help with visualising the interior of the nanocompartment and therefore the DyP (99). When the sample was initially viewed using negative stain to assess the quality, the overall number of Enc present in the micrographs examined showed that there was a low concentration of Enc nanocompartments and that the sample wasn't well suited to cryo-EM (100). All the work done to bring out the DyP signal does show that the presence of DyP inside of Enc is highly variable and a far larger, high quality data set was required to generate an adequate reconstruction.

Bacterioferritin.

It is known that BfrA and BfrB are part of the *M. tuberculosis* genome and are produced but it is unclear whether it is encapsulated in *M. tuberculosis* and has not been reconstructed. BfrB and Enc are part of a two-gene operon in some cases such as *T. maritima*, with BfrB having a C-terminal extension that is used as a targeting peptide (101), with Contreras *et al*. (38) proposing that it could be encapsulated in *M. tuberculosis*. This led to thoughts that BfrB could be present in the native dataset. A subset of proteins around the right size were used, with the particles being selected based upon the bacterioferritin 2D classes found within Kirykowicz and Woodward (25). This proved an efficient way of confirming the presence of bacterioferritin in the sample with the reconstruction resembling, at low resolution, a recently generated crystal structure of *M. tuberculosis* BfrB (7O6E) that was recombinantly expressed (Fig. 27C) (61). This rudimentary similarity means that the bacterioferritin within *M. tuberculosis* most likely functions in a similar way and therefore can store iron within the cavity present in the centre.

However, there were no examples of bacterioferritin present within the 2D classes or the 3D models of Enc nanocompartment (Fig. 23 and 24A). This shows that the findings of Contreras *et al.* (38), were potentially not physiologically relevant as far as

BfrB is concerned. The SDS-PAGE in that research showed that BfrB and DyP co-eluted with Enc but this was by way of the addition of a C-terminal TP which has been shown to cause encapsulation (69). When the C-terminus was truncated, the SDS-PAGE gel showed a dark band present in the flow through and none in the fractions with Enc. The natural C-terminal extension of BfrB is found on the inside of the fully formed oligomer and is thought to aid in heat stabilising the protein with truncated versions of the protein unfolding when exposed to temperatures as low as 30^{O}C (61).

There is a possibility that BfrB is a secondary cargo protein within *M. tuberculosis* with the TP being present at the N-terminal end of the protein. Giessen and Silver (19) discovered a variety of secondary cargo proteins whilst doing whole genome searches using three different TP consensus sequences with ferredoxins having an N-terminal TP. The presence of an N-terminal TP would not mean anything if the abolition of the C-terminal peptide led to the protein unfolding at a far lower temperature when compared to the wild type. This would perhaps indicate why Contreras *et al.* (38) saw a difference when comparing the truncated version to the normal one as the cultures that the proteins were expressed in were grown at 37^{o}C. Lien *et al.* (23) also did not identify any examples of encapsulated BfrB when the proteins were expressed natively which would indicate that the result obtained by Contreras *et al.* (38) was not physiologically relevant. Native Enc and cargo protein systems are generally present in a two-gene operon structure, with the cargo followed by the Enc protein with both under the control of a single promoter (68), of which the ferritin in *M. tuberculosis* is not an example. There is still a possibility that it is a secondary cargo protein, as BfrB was found to a secondary cargo protein in all cases observed (19). This means that the conditions for its encapsulation in *M. tuberculosis* are clearly different than what has been tested. When looking at the ferritin of *M. smegmatis*, this ferritin was not seen to be encapsulated natively and with BfrB sharing a 71.82% identity with *M. tuberculosis*, this strongly points to BfrB not being readily encapsulated natively within *M. tuberculosis* in the conditions tested. The fact that the C-terminal peptide tag is on the interior of the oligomer makes it unclear whether it is indeed used as a means of targeting the protein to the interior of the nanocompartment and would make sense as the targeting peptide if each monomer of the bacterioferritin is individually imported into the nanocomaprtment and assembled inside. The C-terminal peptide has been shown previously to interact with a mostly hydrophobic binding pocket but the position of the C-terminal peptide inside of BfrB contradicts this. This coupled with the lack of a clearly encapsulated 2D class with the findings by Lien *et al.* (23) show that BfrB might not be encapsulated. This is contrasted by Cassidy-Amstutz *et al.* (69) who showed that adding a C-terminal tag to heterologous proteins in *T. maritima* caused the proteins to be directed inside of the nanocompartment.

Limitations.

The recombinant sample preparation and micrograph capture took place prior to the start of this thesis. The protein concentration of the sample was too high and due to limited available slots for viewing the sample at the Electron Bio-Imaging Centre at Diamond Light Source, another sample was not sent.

When working with *M. tuberculosis* H37Ra, the main limitation of this study was that only one litre of culture was allowed to be grown in the Biosafety Level 2 laboratory where the work was conducted. Tang *et al.* (24) grew a much larger amount of *M. smegmatis* culture (6 L of liquid culture) and obtained a much larger pellet (40 g of pellet). This coupled with the slow growth of the cultures meant that the amount of Enc present within the pellet was not as high as was hoped for, which showed in the lack of particles present within the negative stain micrographs (Fig. 22A). There are not many well defined purification processes for purifying intact proteins for structural analysis in *M. tuberculosis* and as a result the purification of the Enc was adapted from the purification of Enc from *M. smegmatis* (24) and it is unclear how efficient this was. The nickel affinity column used to purify the Enc nanocompartments was chosen on the rationale that there is a naturally occurring ring of histidine residues around every five-fold pore, however the ring of histidine's may not have been able to interact strongly enough with the column when compared to proteins that have had a hexa-histidine tag added, such as the recombinantly expressed Enc. The sample obtained was not suitable for cryo-EM microscopy due to the low concentration of Enc and the lack of purity. This same reason is why no catalytic tests were carried out. Only negative stain was used when making grids due to this low concentration of Enc and as such the resolution of any reconstructions was limited. One main reason for carrying out negative stain was due to improved contrast within the micrographs captured. Kirykowicz and Woodward (25) used negative stain EM in the identification of protein complexes in *M. smegmatis* and had success in generating a map showing DyP inside of the naocompartment. It was hoped that this contrast would allow for identification of the DyP inside of the nanocompartment.

Conclusion.

This study did not result in a reconstruction of the encapsulated DyP that it initially set out to achieve but recombinant Enc and DyP were expressed with a high-resolution structure of the Enc nanocompartment being obtained. A monomeric and multimeric structure were built of Enc, with new insight gained into the charge and hydrophilicity of the amino acids around the fivefold pore. A rudimentary growth and purification pipeline was tested natively on *M. tuberculosis* strain H37Ra which showed promise and could be improved on in the future. Whether DyP or BfrB is present inside of the nanocompartment when natively expressed remains unclear, however there is some promise to the presence of DyP. The field of native work on *M. tuberculosis* proteins is relatively small due to the difficulty of growing *M. tuberculosis*, even the H37Ra strain. Many studies that have worked to explore the proteome have either used recombinant expression or have not been looking to conserve the structure of the proteins and have carried out heat inactivation as a means of making sure the sample is safe to work with. The methods described in this paper could provide a way to express and purify proteins natively. This could lead to a better understanding of structure and function when compared to recombinant expression with native conditions being important to understanding these proteins better. The next step would be to work on improving these methods to obtain a purer sample.

Future work.

Future work on this topic could be to figure out a better method of recombinantly expressing encapsulated DyP so that the DyP forms into a visible complex within the nanocompartment so that a high-resolution reconstruction can be obtained. Maintaining the two-gene operon structure of the Enc and DyP by using codon optimization as this has been shown to increase production of a number of enzymes in a variety of hosts (102). If the correct form of DyP present in *M. tuberculosis* is the hexameric and dodecameric forms, then the reason for their formation needs to be discovered. Expanding upon the methods used by Lien *et al.* (23) could potentially yield visibly encapsulated DyP as the images obtained in that research show what could be encapsulated DyP. Immunogenic protein expression of *M. tuberculosis* proteins within *E. coli* has been studied and improved in a few ways such as improving the design of the expression constructs and optimizing the growth and purification conditions used (103). Another route would be to explore the use of specifically optimised *M. smegmatis* strains such as the one produced in Noens *et al.* (89).

The native study of Enc and DyP should also be refined as this provides more relevant physiological information. The optimal growth time and exactly which stage of growth the Enc and DyP is produced would provide information that could potentially decrease the overall growth time of the cultures as this was a major limiting factor during the native part of this thesis. This would be done by taking a sample of the culture at difference stages of growth, lysing the cells, and running the sample on an SDS-PAGE gel. The purification process also needs to be explored to generate a much purer sample that could then be used in catalytic studies and to generate a high resolution cryo-EM reconstruction. Lien *et al.* (23) used a sucrose gradient to purify the native nanocompartment which could be a promising method to explore. Another option is to do more in-depth testing on which column and which combinations of purification methods are best suited to purify encapsulated DyP.

Data Availability

The 3D cryo-EM density map and co-ordinate data have been deposited into the PDB and EMDB database. The EMDB accession number of the 3D cryo-EM density map of the DyP cargo loaded Enc nanocompartment in this thesis is EMD-18031. The accession number of the co-ordinates for the Enc nanocompartment are 8PYS.

References

1. Global Tuberculosis Report 2022 [Internet]. Geneva: World Health Organization; 2022 p. 51. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022

2. Lawn SD, Kerkhoff AD, Vogt M, Wood R. HIV-associated tuberculosis: relationship between disease severity and the sensitivity of new sputum-based and urine-based diagnostic assays. BMC Med. 2013 Dec;11(1):231.

3. Kwan CK, Ernst JD. HIV and Tuberculosis: a Deadly Human Syndemic. Clin Microbiol Rev. 2011 Apr;24(2):351–76.

4. Ayles H, Mureithi L, Simwinga M. The state of tuberculosis in South Africa: what does the first national tuberculosis prevalence survey teach us? Lancet Infect Dis. 2022 Aug;22(8):1094–6.

5. Van Crevel R, Ottenhoff THM, Van Der Meer JWM. Innate Immunity to *Mycobacterium tuberculosis*. Clin Microbiol Rev. 2002 Apr;15(2):294–309.

6. Henderson RA, Watkins SC, Flynn JL. Activation of human dendritic cells following infection with Mycobacterium tuberculosis. J Immunol Baltim Md 1950. 1997 Jul 15;159(2):635–43.

7. Thurnher M, Ramoner R, Gastl G, Radmayr C, Böck G, Herold M, et al. Bacillus Calmette-Guérin mycobacteria stimulate human blood dendritic cells. Int J Cancer. 1997 Jan 6;70(1):128–34.

8. Maphasa RE, Meyer M, Dube A. The Macrophage Response to Mycobacterium tuberculosis and Opportunities for Autophagy Inducing Nanomedicines for Tuberculosis Therapy. Front Cell Infect Microbiol. 2020;10:618414.

9. Köster S, Upadhyay S, Chandra P, Papavinasasundaram K, Yang G, Hassan A, et al. Mycobacterium tuberculosis is protected from NADPH oxidase and LC3-associated phagocytosis by the LCP protein CpsA. Proc Natl Acad Sci U S A. 2017 Oct 10;114(41):E8711–20.

10. Zhai W, Wu F, Zhang Y, Fu Y, Liu Z. The Immune Escape Mechanisms of Mycobacterium Tuberculosis. Int J Mol Sci. 2019 Jan 15;20(2):340.

11. Schüller S, Neefjes J, Ottenhoff T, Thole J, Young D. Coronin is involved in uptake of Mycobacterium bovis BCG in human macrophages but not in phagosome maintenance. Cell Microbiol. 2001 Dec;3(12):785–93.

12. Walburger A, Koul A, Ferrari G, Nguyen L, Prescianotto-Baschong C, Huygen K, et al. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science. 2004 Jun 18;304(5678):1800–4.

13. Gutierrez MG, Mishra BB, Jordao L, Elliott E, Anes E, Griffiths G. NF-kappa B activation controls phagolysosome fusion-mediated killing of mycobacteria by macrophages. J Immunol Baltim Md 1950. 2008 Aug 15;181(4):2651–63.

14. Wong D, Li W, Chao JD, Zhou P, Narula G, Tsui C, et al. Protein tyrosine kinase, PtkA, is required for Mycobacterium tuberculosis growth in macrophages. Sci Rep. 2018 Jan 9;8(1):155.

15. Voskuil MI, Bartek IL, Visconti K, Schoolnik GK. The Response of Mycobacterium Tuberculosis to Reactive Oxygen and Nitrogen Species. Front Microbiol [Internet]. 2011 [cited 2023 Mar 6];2. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2011.00105/abstract

16. Rosenkrands I, Rasmussen PB, Carnio M, Jacobsen S, Theisen M, Andersen P. Identification and Characterization of a 29-Kilodalton Protein from *Mycobacterium tuberculosis* Culture Filtrate Recognized by Mouse Memory Effector Cells. Infect Immun. 1998 Jun;66(6):2728–35.

17. Valdés-Stauber N, Scherer S. Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. Appl Environ Microbiol. 1994 Oct;60(10):3809–14.

18. Radford DR. Understanding the Encapsulins: Prediction and Characterization of Phage Capsid-like Nanocompartments in Prokaryotes [Thesis]. [Toronto]: University of Toronto; 2014.

19. Giessen TW, Silver PA. Widespread distribution of encapsulin nanocompartments reveals functional diversity. Nat Microbiol. 2017 Mar 6;2(6):17029.

20. Nichols RJ, Cassidy-Amstutz C, Chaijarasphong T, Savage DF. Encapsulins: molecular biology of the shell. Crit Rev Biochem Mol Biol. 2017 Sep 3;52(5):583–94.

21. McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, et al. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. EMBO J. 2014 Sep 1;33(17):1896–911.

22. Weldingh K, Andersen P. Immunological evaluation of novel Mycobacterium tuberculosis culture filtrate proteins. FEMS Immunol Med Microbiol. 1999 Feb;23(2):159–64.

23. Lien KA, Dinshaw K, Nichols RJ, Cassidy-Amstutz C, Knight M, Singh R, et al. A nanocompartment system contributes to defense against oxidative stress in Mycobacterium tuberculosis. eLife. 2021 Nov 9;10:e74358.

24. Tang Y, Mu A, Zhang Y, Zhou S, Wang W, Lai Y, et al. Cryo-EM structure of *Mycobacterium smegmatis* DyP-loaded encapsulin. Proc Natl Acad Sci. 2021 Apr 20;118(16):e2025658118.

25. Kirykowicz AM, Woodward JD. Shotgun EM of mycobacterial protein complexes during stationary phase stress. Curr Res Struct Biol. 2020;2:204–12.

26. Gabashvili AN, Chmelyuk NS, Efremova MV, Malinovskaya JA, Semkina AS, Abakumov MA. Encapsulins—Bacterial Protein Nanocompartments: Structure, Properties, and Application. Biomolecules. 2020 Jun 26;10(6):966.

27. Daleke MH, Ummels R, Bawono P, Heringa J, Vandenbroucke-Grauls CMJE, Luirink J, et al. General secretion signal for the mycobacterial type VII secretion pathway. Proc Natl Acad Sci. 2012 Jul 10;109(28):11342–7.

28. Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, et al. Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat Struct Mol Biol. 2008 Sep;15(9):939–47.

29. Jones JA, Giessen TW. Advances in encapsulin nanocompartment biology and engineering. Biotechnol Bioeng. 2021 Jan;118(1):491–505.

30. Wiryaman T, Toor N. Recent advances in the structural biology of encapsulin bacterial nanocompartments. J Struct Biol X. 2022;6:100062.

31. T JAS, J R, Rajan A, Shankar V. Features of the biochemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis. J Infect Public Health. 2020 Sep;13(9):1255–64.

32. Ghodbane R, Medie FM, Lepidi H, Nappez C, Drancourt M. Long-term survival of tuberculosis complex mycobacteria in soil. Microbiology. 2014 Mar 1;160(3):496–501.

33. Anes E, Peyron P, Staali L, Jordao L, Gutierrez MG, Kress H, et al. Dynamic life and death interactions between Mycobacterium smegmatis and J774 macrophages. Cell Microbiol. 2006 Jun;8(6):939–60.

34. Chmelyuk NS, Oda VV, Gabashvili AN, Abakumov MA. Encapsulins: Structure, Properties, and Biotechnological Applications. Biochem Biokhimiia. 2023 Jan;88(1):35–49.

35. Almeida AV, Carvalho AJ, Pereira AS. Encapsulin nanocages: Protein encapsulation and iron sequestration. Coord Chem Rev. 2021 Dec;448:214188.

36. Giessen TW, Orlando BJ, Verdegaal AA, Chambers MG, Gardener J, Bell DC, et al. Large protein organelles form a new iron sequestration system with high storage capacity. eLife. 2019 Jul 8;8:e46070.

37. Rahmanpour R, Bugg TDH. Assembly *in vitro* of *Rhodococcus jostii* RHA1 encapsulin and peroxidase DypB to form a nanocompartment. FEBS J. 2013 May;280(9):2097–104.

38. Contreras H, Joens MS, McMath LM, Le VP, Tullius MV, Kimmey JM, et al. Characterization of a Mycobacterium tuberculosis Nanocompartment and Its Potential Cargo Proteins. J Biol Chem. 2014 Jun;289(26):18279–89.

39. Akita F, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, et al. The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. J Mol Biol. 2007 May;368(5):1469–83.

40. Wiryaman T, Toor N. Cryo-EM structure of a thermostable bacterial nanocompartment. IUCrJ. 2021 May 1;8(3):342–50.

41. Ross J, McIver Z, Lambert T, Piergentili C, Bird JE, Gallagher KJ, et al. Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment. Sci Adv. 2022 Jan 28;8(4):eabj4461.

42. Cornejo E, Abreu N, Komeili A. Compartmentalization and organelle formation in bacteria. Curr Opin Cell Biol. 2014 Feb;26:132–8.

43. Yoshida T, Sugano Y. A structural and functional perspective of DyP-type peroxidase family. Arch Biochem Biophys. 2015 May;574:49–55.

44. Pandey R, Rodriguez GM. A Ferritin Mutant of Mycobacterium tuberculosis Is Highly Susceptible to Killing by Antibiotics and Is Unable To Establish a Chronic Infection in Mice. Flynn JL, editor. Infect Immun. 2012 Oct;80(10):3650–9.

45. Patricia Twala P, Mitema A, Baburam C, Aliye Feto N, OMICS Research Group, Department of Biotechnology, Vaal University of Technology, Vanderbijlpark, South Africa. Breakthroughs in the

discovery and use of different peroxidase isoforms of microbial origin. AIMS Microbiol. 2020;6(3):330–49.

46. Sugano Y, Muramatsu R, Ichiyanagi A, Sato T, Shoda M. DyP, a Unique Dye-decolorizing Peroxidase, Represents a Novel Heme Peroxidase Family. J Biol Chem. 2007 Dec;282(50):36652–8.

47. Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TDH. Identification of DypB from *Rhodococcus jostii* RHA1 as a Lignin Peroxidase. Biochemistry. 2011 Jun 14;50(23):5096–107.

48. Sugano Y, Yoshida T. DyP-Type Peroxidases: Recent Advances and Perspectives. Int J Mol Sci. 2021 May 24;22(11):5556.

49. Lang G, Spartalian K, Yonetani T. Mössbauer spectroscopic study of compound es of cytochrome c peroxidase. Biochim Biophys Acta BBA - Gen Subj. 1976 Nov;451(1):250–8.

50. Hiner ANP, Ruiz JH, López JNR, Cánovas FG, Brisset NC, Smith AT, et al. Reactions of the Class II Peroxidases, Lignin Peroxidase andArthromyces ramosus Peroxidase, with Hydrogen Peroxide. J Biol Chem. 2002 Jul;277(30):26879–85.

51. Shrestha R, Huang G, Meekins DA, Geisbrecht BV, Li P. Mechanistic Insights into Dye-Decolorizing Peroxidase Revealed by Solvent Isotope and Viscosity Effects. ACS Catal. 2017 Sep 1;7(9):6352–64.

52. Chen C, Shrestha R, Jia K, Gao PF, Geisbrecht BV, Bossmann SH, et al. Characterization of Dye-decolorizing Peroxidase (DyP) from Thermomonospora curvata Reveals Unique Catalytic Properties of A-type DyPs. J Biol Chem. 2015 Sep;290(38):23447–63.

53. Yuzugullu Karakus Y. Typical Catalases: Function and Structure. In: Dulce Bagatini M, editor. Glutathione System and Oxidative Stress in Health and Disease [Internet]. IntechOpen; 2020 [cited 2023 Dec 8]. Available from: https://www.intechopen.com/books/glutathione-system-and-oxidative-stress-in-health-and-disease/typical-catalases-function-and-structure

54. Wiswedel R. Investigating the substrate catalysis of free and encapsulated dye-decolourising peroxidase in Mycobacterium tuberculosis. [Cape Town]: University of Cape Town; 2020.

55. Hiner ANP, Hernández-Ruiz J, Williams GA, Arnao MB, Garcı́a-Cánovas F, Acosta M. Catalase-like Oxygen Production by Horseradish Peroxidase Must Predominantly Be an Enzyme-Catalyzed Reaction. Arch Biochem Biophys. 2001 Aug;392(2):295–302.

56. Hernández-Ruiz J, Arnao MB, Hiner ANP, García-Cánovas F, Acosta M. Catalase-like activity of horseradish peroxidase: relationship to enzyme inactivation by H2O2. Biochem J. 2001 Feb 15;354(1):107–14.

57. Rivera M. Bacterioferritin: Structure, Dynamics, and Protein–Protein Interactions at Play in Iron Storage and Mobilization. Acc Chem Res. 2017 Feb 21;50(2):331–40.

58. Andrews SC. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. Biochim Biophys Acta BBA - Gen Subj. 2010 Aug;1800(8):691–705.

59. He D, Hughes S, Vanden-Hehir S, Georgiev A, Altenbach K, Tarrant E, et al. Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. eLife. 2016 Aug 16;5:e18972.

60. LaFrance BJ, Cassidy-Amstutz C, Nichols RJ, Oltrogge LM, Nogales E, Savage DF. The encapsulin from Thermotoga maritima is a flavoprotein with a symmetry matched ferritin-like cargo protein. Sci Rep. 2021 Nov 23;11(1):22810.

61. Khare G, Gupta V, Nangpal P, Gupta RK, Sauter NK, Tyagi AK. Ferritin Structure from Mycobacterium tuberculosis: Comparative Study with Homologues Identifies Extended C-Terminus Involved in Ferroxidase Activity. Agrewala JN, editor. PLoS ONE. 2011 Apr 8;6(4):e18570.

62. Tullius MV, Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, et al. Discovery and characterization of a unique mycobacterial heme acquisition system. Proc Natl Acad Sci. 2011 Mar 22;108(12):5051–6.

63. Voskuil MI, Visconti KC, Schoolnik GK. Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis. 2004 Jan;84(3–4):218–27.

64. Haußmann C, Rohdich F, Schmidt E, Bacher A, Richter G. Biosynthesis of Pteridines in Escherichia coli. J Biol Chem. 1998 Jul;273(28):17418–24.

65. Chiancone E, Ceci P, Ilari A, Ribacchi F, Stefanini S. Iron and proteins for iron storage and detoxification. BioMetals. 2004 Jun;17(3):197–202.

66. Winterbourn CC. Toxicity of iron and hydrogen peroxide: the Fenton reaction. Toxicol Lett. 1995 Dec;82–83:969–74.

67. Shastri MD, Shukla SD, Chong WC, Dua K, Peterson GM, Patel RP, et al. Role of Oxidative Stress in the Pathology and Management of Human Tuberculosis. Oxid Med Cell Longev. 2018 Oct 11;2018:1–10.

68. Altenburg WJ, Rollins N, Silver PA, Giessen TW. Exploring targeting peptide-shell interactions in encapsulin nanocompartments. Sci Rep. 2021 Mar 2;11(1):4951.

69. Cassidy-Amstutz C, Oltrogge L, Going CC, Lee A, Teng P, Quintanilla D, et al. Identification of a Minimal Peptide Tag for *in Vivo* and *in Vitro* Loading of Encapsulin. Biochemistry. 2016 Jun 21;55(24):3461–8.

70. Nichols RJ, LaFrance B, Phillips NR, Radford DR, Oltrogge LM, Valentin-Alvarado LE, et al. Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. eLife. 2021 Apr 6;10:e59288.

71. Van de Steen A, Khalife R, Colant N, Mustafa Khan H, Deveikis M, Charalambous S, et al. Bioengineering bacterial encapsulin nanocompartments as targeted drug delivery system. Synth Syst Biotechnol. 2021 Sep;6(3):231–41.

72. Lagoutte P, Mignon C, Stadthagen G, Potisopon S, Donnat S, Mast J, et al. Simultaneous surface display and cargo loading of encapsulin nanocompartments and their use for rational vaccine design. Vaccine. 2018 Jun;36(25):3622–8.

73. McDowell HB, Hoiczyk E. Bacterial Nanocompartments: Structures, Functions, and Applications. Galperin MY, editor. J Bacteriol. 2022 Mar 15;204(3):e00346-21.

74. Giessen TW, Silver PA. Encapsulation as a Strategy for the Design of Biological Compartmentalization. J Mol Biol. 2016 Feb;428(5):916–27.

75. Rabodoarivelo MS, Aerts M, Vandamme P, Palomino JC, Rasolofo V, Martin A. Optimizing of a protein extraction method for Mycobacterium tuberculosis proteome analysis using mass spectrometry. J Microbiol Methods. 2016 Dec;131:144–7.

76. Bespyatykh JA, Shitikov EA, Ilina EN. Proteomics for the Investigation of Mycobacteria. Acta Naturae. 2017;9(1):15–25.

77. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012 Dec;180(3):519–30.

78. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife. 2018 Nov 9;7:e42166.

79. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods. 2017 Apr;14(4):331–2.

80. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015 Nov;192(2):216–21.

81. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

82. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr. 2010 Apr 1;66(4):486–501.

83. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr Sect Struct Biol. 2019 Oct 1;75(10):861–77.

84. Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol. 2005 Oct;152(1):36–51.

85. Peñuelas-Urquides K, Villarreal-Treviño L, Silva-Ramírez B, Rivadeneyra-Espinoza L, Said-Fernández S, León MBD. Measuring of Mycobacterium tuberculosis growth: a correlation of the optical measurements with colony forming units. Braz J Microbiol. 2013;44(1):287–90.

86. Gijsbers A, Zhang Y, Gao Y, Peters PJ, Ravelli RBG. *Mycobacterium tuberculosis* ferritin: a suitable workhorse protein for cryo-EM development. Acta Crystallogr Sect Struct Biol. 2021 Aug 1;77(8):1077–83.

87. Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Shaun Lott J. A new Gateway® vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. Protein Expr Purif. 2008 Jan;57(1):81–7.

88. Qamra R, Mande SC, Coates ARM, Henderson B. The unusual chaperonins of Mycobacterium tuberculosis. Tuberculosis. 2005 Sep;85(5–6):385–94.

89. Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT, Wilmanns M. Improved mycobacterial protein production using a Mycobacterium smegmatis groEL1ΔCexpression strain. BMC Biotechnol. 2011 Dec;11(1):27.

90. Xiong X, Sun C, Vago F, Klose T, Zhu J, Jiang W. Cryo-EM Structure of Heterologous Protein Complex Loaded Thermotoga Maritima Encapsulin Capsid. Biomolecules. 2020 Sep 19;10(9):1342.

91. Wang HW. A commentary of "Cryo-EM achieves atomic resolution" in 10 remarkable discoveries from 2020 in Nature. Fundam Res. 2022 Mar;2(2):349–50.

92. Eren E, Wang B, Winkler DC, Watts NR, Steven AC, Wingfield PT. Structural characterization of the Myxococcus xanthus encapsulin and ferritin-like cargo system gives insight into its iron storage mechanism. Struct Lond Engl 1993. 2022 Apr 7;30(4):551-563.e4.

93. Jones JA, Benisch R, Giessen TW. Encapsulin cargo loading: progress and potential. J Mater Chem B. 2023;10.1039.D3TB00288H.

94. Goetschius DJ, Lee H, Hafenstein S. CryoEM reconstruction approaches to resolve asymmetric features. In: Advances in Virus Research [Internet]. Elsevier; 2019 [cited 2023 Mar 6]. p. 73–91. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0065352719300247

95. Serna M. Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes. Front Mol Biosci. 2019 May 15;6:33.

96. Yimer SA, Birhanu AG, Kalayou S, Riaz T, Zegeye ED, Beyene GT, et al. Comparative Proteomic Analysis of Mycobacterium tuberculosis Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence. Front Microbiol. 2017 May 9;8:795.

97. Gopinath V, Raghunandanan S, Gomez RL, Jose L, Surendran A, Ramachandran R, et al. Profiling the Proteome of Mycobacterium tuberculosis during Dormancy and Reactivation. Mol Cell Proteomics. 2015 Aug;14(8):2160–76.

98. Scarff CA, Fuller MJG, Thompson RF, Iadanza MG. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. J Vis Exp. 2018 Feb 6;(132):57199.

99. Ohi M, Li Y, Cheng Y, Walz T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online. 2004;6:23–34.

100. Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods. 2016 May;100:3–15.

101. Rurup WF, Snijder J, Koay MST, Heck AJR, Cornelissen JJLM. Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. J Am Chem Soc. 2014 Mar 12;136(10):3828–32.

102. Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol [Internet]. 2014 [cited 2023 Dec 6];5. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2014.00021/abstract

103. Piubelli L, Campa M, Temporini C, Binda E, Mangione F, Amicosante M, et al. Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins. Microb Cell Factories. 2013 Dec;12(1):115.