

Transmission electron microscopy study of iron oxide nanoparticles formed in *Myxococcus xanthus* encapsulins through bio-mineralisation



University of
Sheffield

Harry McDowell

Supervisor: Dr. Thomas Walther

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

in the

Department of Electronic and Electrical Engineering
Faculty of Engineering
The University of Sheffield

April 15, 2025

0.1 Acknowledgement

This project would not have been completed without the help of a number of individuals, not just through direct contributions to the research but also in terms of moral support throughout the past four years. I would like first and foremost thank my supervisors Dr. Thomas Walther and Dr. Egbert Hoiczky who have guided me through the process and been incredibly supportive and encouraging. I have benefited immensely from the knowledge you have both shared with me. Coming from a biological background I would like to thank Thomas especially for his patience in teaching me. The scientific discussions I shared with the both of you were both enjoyable and incredibly useful in helping me to develop my scientific outlook. I would also like to thank the Department of Electronic and Electrical Engineering for providing funding for my PhD, as without it the project would not have been possible. I am grateful for the Sorby centre and the Faculty of Science Electron Microscopy Centre, for their help and support in collecting the experimental data for this project.

I would like to thank Diamond Light Source for access and support in use of the electron Physical Science Imaging Centre (Instrument E01, proposal number MG35899-3) that contributed to the results presented here. I would like to thank Dr. Mohsen Danaie for his support both on the microscope and afterwards. Additionally, I am grateful to the Ernst Ruska Centre for access and support in completing experimental work for this project (proposal number E-023), especially to Dr. Jo Janghyun and Dr. Penghan Lu who operated the microscopes.

I would like to thank my lab colleges, Shuang Li, Jeffery So and Koen Semeijn for their companionship, support and scientific discourse on a daily basis. Finally, I would like to thank my friends, family and partner Hannah for their support during the project. Their enduring love and support helped me get through the hardest times.

0.2 Abstract

Encapsulins have been recently identified as iron storage nano-compartments whose protein shell have been the focus of study in a wide range of species. However, little work to date has sought to characterize the nanoparticles synthesised inside these nano-compartments. In this thesis, a comprehensive investigation is made into the iron mineralizing encapsulin from the bacterium *Myxococcus xanthus*, with the aim to characterize the nanoparticles inside the natively isolated encapsulin and identify the role played by the three cargo proteins in the iron mineralising function of the system. Of the three cargo proteins - EncB, EncC and EncD - two are labelled as ‘ferritin-like’ proteins, however it is unclear why there are two proteins with the same function packaged into a single encapsulin. To do this, a range of electron microscopical techniques were employed, including a number of imaging and spectroscopic techniques. It was identified that, the nanoparticles of the natively isolated *M. xanthus* were morphologically distinct from previous observations: significantly smaller, amorphous and with a higher Fe:P ratio. The function of the cargo proteins was explored by generating encapsulin structures containing reduced numbers of cargo proteins and synthesising nanoparticles inside them using two distinct approaches. No difference was identified between the total amount of Fe inside the encapsulins containing reduced numbers of cargo proteins or the size of the nanoparticles synthesised. However, a difference was identified in the Fe:P ratio between encapsulins containing either EncB or EncC using both electron dispersive x-ray spectroscopy and electron energy loss spectroscopy, which may be important for understanding their biological function. Furthermore, high resolution scanning transmission electron microscopy was used to identify the role phosphorous plays in the formation of crystalline material inside the encapsulin. Thus, this study has contributed to our understanding the function of the *M. xanthus* encapsulin system.

Contents

0.1	Acknowledgement	
0.2	Abstract	
0.3	List of abbreviations	1
1	Introduction	3
1.1	Compartmentalization	4
1.2	Encapsulins	4
1.2.1	On the genealogy of encapsulins	4
1.2.2	Family units	5
1.2.3	Shell structure	6
1.2.4	Structure of basal shell unit	7
1.2.5	Shell assembly	8
1.2.6	The pores of the encapsulin shell	8
1.2.7	Identification of cargo proteins	10
1.2.8	Cargo loading peptides	11
1.2.9	Cargo loading stoichiometry	13
1.2.10	Biological function of encapsulin	14
1.2.11	Relationship between the shell and cargo	14
1.3	Iron storage encapsulins	15
1.3.1	Biom mineralization	15

CONTENTS

1.3.2	Iron and oxidative stress	16
1.3.3	Sequence identity of the ferritin-like proteins of encapsulins	17
1.3.4	Structural identity of ferritin like proteins	17
1.3.5	Proposed mechanism for iron storage by encapsulin systems	18
1.3.6	The encapsulin system from <i>M. xanthus</i>	21
1.4	Electron microscopy	24
1.4.1	Interactions of electrons with matter	24
1.4.2	Elastic and inelastic scattering	25
1.4.3	The scanning transmission electron microscope	26
1.4.4	Spherical aberration	28
1.4.5	Bright field imaging	29
1.4.6	Annular dark field and high angle annular dark field imaging	30
1.4.7	Energy dispersive X-ray spectroscopy	31
1.4.8	Electron energy loss spectroscopy	32
1.4.9	Comparison of EDXS and EELS	35
1.5	Aims and Objectives	36
2	Isolation and generations of encapsulin protein complexes	39
2.1	Introduction	39
2.2	Experimental procedures	40
2.2.1	Strains and culturing conditions	40
2.2.2	Isolation of encapsulin from <i>M. xanthus</i>	41
2.2.3	PCR and plasmid construction	42
2.2.4	Transformation of <i>E.coli</i>	43
2.2.5	Protein expression	46
2.2.6	Purification of proteins from <i>E. coli</i>	46
2.2.7	Polyacrylamide gel electrophoresis	47

2.2.8	SDS gel quantification	48
2.2.9	Negative stain for transmission electron microscopy	48
2.2.10	Negative stained image analysis	48
2.2.11	Template matching	49
2.3	Results	50
2.3.1	Isolation of encapsulins from <i>M. xanthus</i>	50
2.3.2	Long term stability of encapsulins	50
2.3.3	Gene amplification	52
2.3.4	Protein expression conditions	52
2.3.5	Overexpression of 6xHis-tagged EncA	54
2.3.6	Simultaneous expression of shell protein EncA and single cargo proteins	54
2.3.7	Simultaneous expression of EncA and multiple cargo proteins	57
2.3.8	Cargo loading efficiency	59
2.3.9	Visualisation of structures in TEM	59
2.3.10	The presence of the cargo proteins during overexpression reduce the number of shells showing T=1 morphology	60
2.4	Discussion	64
3	Electron microscopical investigation of encapsulin derived nanoparticles	67
3.1	Experimental procedures	69
3.1.1	Iron loading <i>in vivo</i>	69
3.1.2	Iron loading <i>in vitro</i>	69
3.1.3	Encapsulins used in this study	70
3.1.4	Magnetic separation of Fe filled encapsulins	70
3.1.5	Electron microscopy	70
3.1.6	Grid preparation	70
3.1.7	Image preprocessing	73

CONTENTS

3.1.8	Image analysis	73
3.1.9	Encapsulin shell image analysis	73
3.1.10	Encapsulin core region analysis	75
3.1.11	Internal nanoparticle analysis	75
3.1.12	4D-STEM and data processing	75
3.1.13	Selected area electron diffraction	76
3.1.14	HR-STEM and data processing	77
3.2	Results	78
3.2.1	Encapsulin shell thickness	78
3.2.2	<i>M. xanthus</i> encapsulins show core size variability in different isolations .	80
3.2.3	Size of internal nanoparticles	80
3.2.4	Iron mineralizing properties of encapsulin proteins	85
3.2.5	Effect of different Fe concentration in the growth media on encapsulin loading	87
3.2.6	Attempted isolation of Fe loaded encapsulins using magnets	90
3.2.7	Comparison of core sizes at different Fe concentrations	92
3.2.8	Largest diameter cores achieved in <i>E. coli</i> expressed encapsulins	94
3.2.9	EncB and EncC synthesise nanoparticles with the same diameters . . .	95
3.2.10	Observation of individually distributed ~ 2 nm nanoparticles	95
3.2.11	Comparing the HAADF intensity and the size of the Fe rich region in samples isolated from <i>M. xanthus</i> with those isolated from <i>E. coli</i> . . .	97
3.2.12	Loading encapsulins <i>in vitro</i> with increasing concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$	99
3.2.13	Size measurements of nanoparticles expressed <i>in vitro</i>	101
3.2.14	4D-STEM investigation of the nanoparticles inside the <i>M. xanthus</i> en- capsulin	104
3.2.15	High Resolution STEM imaging of the <i>M. xanthus</i> encapsulin	106

3.2.16	Structural characterization of nanoparticles from encapsulins overexpressed in <i>E. coli</i>	108
3.2.17	HR-STEM investigation of EncA+EncB+EncC loaded <i>in vitro</i>	109
3.3	Discussion	111
3.4	Appendix	115
3.4.1	Identification of nanoparticles from TEM images	115
3.4.2	Identification of nanoparticles from HAADF images	116
3.4.3	Iron oxide lattice parameters	117
3.4.4	<i>D</i> -spacings measured from HR-STEM images of encapsulins isolated from <i>M. xanthus</i>	118
3.4.5	<i>D</i> -spacings measured from HR-STEM images of EncA+EncB+EncC	121
4	Characterization of encapsulins derived nanoparticles using EDXS	123
4.1	Experimental procedures	124
4.1.1	Electron microscopy	124
4.1.2	Encapsulin samples	124
4.1.3	Data processing: Digital Micrograph	124
4.1.4	Elemental map generation using HyperSpy	126
4.2	Results	127
4.2.1	Chemical mapping of the <i>M. xanthus</i> encapsulin using EDXS	127
4.2.2	Chemical mapping of the encapsulin of <i>M. xanthus</i> on Au grids	129
4.2.3	Chemical mapping of encapsulins overexpressed in <i>E. coli</i>	131
4.2.4	Chemical mapping of EncA+EncB+EnC loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$	135
4.2.5	Quantification of the atomic composition of the nanoparticles inside the encapsulin using EDXS	137
4.3	Discussion	139

4.4	Appendix	141
4.4.1	Chemical maps from empty encapsulins isolated from <i>E.coli</i>	141
5	Characterization of encapsulin-derived nanoparticles using EELS	143
5.1	Experimental methods	144
5.1.1	Sample preparation	144
5.1.2	Reference material preparations	145
5.1.3	Scanning transmission electron microscopy	145
5.1.4	Pre-processing of EELS SI datasets	146
5.1.5	Use of binary masks as spectrum image region of interest tools	146
5.1.6	HyperSpy quantification routine	146
5.1.7	Digital Micrograph quantification routine	147
5.1.8	Estimation of the number of iron atoms	148
5.1.9	Quantification of the Ca L-edge	148
5.1.10	Multiple Least Squares Fitting	149
5.1.11	Fe L _{2,3} ELNES fingerprinting	149
5.1.12	Quantification of the Fe L _{2,3} fine structure	150
5.2	Results	151
5.2.1	Chemical maps of encapsulin samples from EELS SIs	151
5.2.2	Separation of the P L _{2,3} and the S L _{2,3} edges using MLLS fitting	158
5.2.3	Observation of the presence of Ca in <i>E. coli</i> generated constructs	160
5.2.4	Calculations of the atomic composition of the encapsulin shell	163
5.2.5	Loss of protein shell material due to beam damage during scanning	165
5.2.6	Calculation of the atomic composition of the encapsulin derived nanoparticles	167
5.2.7	Nanoparticles within a single shell have the same atomic composition	168

5.2.8	Estimates of the number of Fe atoms inside <i>M. xanthus</i> encapsulins compared with HAADF intensity	171
5.2.9	Comparing estimates of the number of Fe atoms inside encapsulins over expressed in <i>E. coli</i>	172
5.2.10	Measurements of individually distributed particles	172
5.2.11	Fingerprinting of the Fe L _{2,3} edge near-edge structure	176
5.2.12	Change in Fe L _{2,3} edge shape with increasing electron dose	178
5.2.13	Quantification of Fe L _{2,3} edge	180
5.3	Discussion	181
5.4	Appendix	185
5.4.1	Calculation of convergence angles	185
5.4.2	Calculation of collection angles	185
6	Final Discussion	187
6.1	Novel model of nanoparticle arrangement inside encapsulins	191
7	Conclusion	195
8	Future work	199

CONTENTS

List of Figures

1.1 Encapsulin structures and their constituting monomeric proteins. (a)

Examples of different atomically solved encapsulin shells. The shells are scaled to size, and the HK97 bacteriophage capsid is included as a reference. The complete structures from: *T. thermotolerans* (PDB: 3DKT), *Syn. elongatus* (PDB: 6C9E), *M. xanthus* (PDB: 4PT2), *Q. thermotolerans* (PDB:6NJ8) and *HK97* bacteriophage capsid (PDB: 1OHG) are shown. The pentameric faces are shown in yellow and the hexameric faces are shown in blue. (b) Encapsulin shell monomer from *M. xanthus* the A-domain, P-domain and E-loop are labelled. (c) Overlay of three encapsulin shell proteins highlighting the conservation of the HK97 structure across multiple encapsulin monomers. The shell proteins for *T. maritima* (orange), *M. xanthus* (green) and *Q. thermotolerans* (purple) are shown. 9

1.2 Structure and organisation of encapsulated ferritin-like proteins. A)

Non-FOC-containing dimer and B) FOC-containing dimer from the FLP of *R. rubrum*. Iron atoms are shown in red. C) FLP decamers from a number of different species superimposed to highlight their structural similarity. The structures shown are *M. xanthus* EncB in red (PDB:7S5K), *R. rubrum* in green (PDB:5DA5), *P. furiosus* in purple (PDB:5N5E), *H. ochraceum* in orange (PDB:5N5F) and *M. xanthus* EncC in blue (PDB:7S8T). D) Multiple sequence alignment for encapsulated ferritin-like proteins using Clustal Omega [90]. Top to bottom: EncB from *M.xanthus*, EncC from *M. xanthus*, *T. maritima*, *H. Ochraceum*, *R. rubrum* [91]. 19

1.3 General model for iron oxidation by encapsulin FLP cargo proteins.

(a) Truncated EncC (PDB: 7S8T) atomic model superimposed over model of *M. xanthus* encapsulin shell (PDB: 4PT2). The cargo is modelled over the 5-fold pore; the suggested entry point for the ferrous substrate. (b) Schematic representation of ferrooxidation pathway from encapsulin FLP systems. Fe^{2+} is transported through the shell pore to the FOC containing α helical bundles of the FLP cargo. Here it is oxidised using O_2 and the resulting ferric substances leaves the cargo protein to be mineralized on its surface before being released into the lumen [23]. 21

1.4 Structure and sequence of *M. xanthus* encapsulin cargo proteins. A)

Partial atomic structure of EncB (PDB:7S5C) in grey. Overlaid with the structural prediction (AlphaFold) of the full length structure (light blue) the CLP region is highlighted (red) B) Partial atomic structure of EncB (PDB:7S5C) in grey. Overlaid with the structural prediction (AlphaFold) of the full length structure (light blue) the CLP region is highlighted (red). C) Sequence alignment of the three cargo proteins: EncB, EncC and EncD. Highly conserved regions, including the ExxH iron-binding motif and the CLP are highlighted. [98, 99] 23

1.5 The different secondary signals emitted from the interaction of electrons with matter. 25**1.6 Schematic diagram of a typical Scanning Transmission Electron Microscope.** The electron gun, condenser lens, scan coils, objective lens, probe forming apertures and detectors are labelled. 27

- 1.7 Comparative BF and HAADF dark field images of an encapsulin isolated from *M. xanthus*. (a) BF (b) HAADF.** Both images were captured at 8 Mx on a spherical aberration corrected JEOL ARM200F operating at 80keV. The scale bar represents 10 nm. The different contrast mechanisms are clearly visible. In the HAADF image the Fe rich core has a higher atomic number (Z value) and thus appears brighter than the predominantly C (lower Z value) protein shell. In the BF image, due to the general low mass and density of the sample the contrast appears poor across the image, despite the high spatial resolution. This is especially evident for the C containing protein shell that is too light to produce enough contrast to be visualised with this technique. 29
- 1.8 The optical transfer function (OTF) for HAADF imaging compared with the phase contrast transfer function (CTF) for BF imaging for the same defocus and spherical aberration.** This figure has been reproduced from [130]. 31
- 1.9 Example EDXS spectra taken from a sample of encapsulins isolated from *M.xanthus*.** The peaks representing C K α , N K α , O K α , Fe L α , Cu L α , Na K α , Mg K α , Si K α , P K α , Fe K α and Cu K α are labelled. 33
- 1.10 Example low loss and high loss electron energy loss spectra taken of encapsulins from the bacteria *M. xanthus*. (a) Low loss spectra.** Two spectral features have been highlighted with arrows, the zero loss peak (top) and the plasmon peak (bottom). **(b) High loss spectra.** Several core loss edges have been labelled with arrows. From left to right: C K-edge, N K-edge, O K-edge and Fe L-edge. The core loss edges are seen as features on the continuously decaying background. 35

2.1	Encapsulin isolation from <i>M. xanthus</i>. (a) Image of encapsulin isolation stained with 1% uranyl acetate, the insert shows four negatively unstained encapsulins. The image was captured at 13kx. (b) Unstained image of encapsulin isolation, the insert shows four unstained encapsulins. The image was captured at 11kx. Both images were taken using a Tecnai Spirit. For both images the scale bar represents 100 nm and the insert scale bar represents 20 nm. (c) Histogram representing the distribution of measured particle diameters. Above it, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are quoted.	51
2.2	Observation of long term stability of encapsulin proteins. (a) Negatively stained image of encapsulin shells taken in November 2020 (left) and in April 2024 (right). Both images were captured on a Tecnai Spirit at 11kx magnification and stained with 1% uranyl acetate. Scale bars represent 100 nm. (b) SDS-PAGE gels of isolated encapsulins. The <i>M. xanthus</i> encapsulin sample was run in the left lane in both images, and bands representing the proteins EncA (32kDa), EncB (17kDa), EncC (13kDa) and EncD (11kDa) can be seen highlighted with arrows. Additionally, a non-specific band of a co-purified protein can be seen at ~30kDa in both images, this band is also present in SDS-PAGE gels of <i>M. xanthus</i> encapsulins that have been published in the literature [18]. In both gels the bands of the See Blue Prestained 2 protein ladder in the right lane of each image is labelled with the size of each protein band in kDa.	53

- 2.3 Purification of C-terminal tagged 6xHis EncA from *E. coli* Rosetta(DE3) following expression with 0.1 mM IPTG for 18 hours at 24 °C. (a)** Alphafold prediction model of a C-terminal tagged 6xHis EncA, the tag is highlighted in yellow on a single subunit of the encapsulin shell. The complete encapsulin shell atomic structure is shown in blue (PDB:4PT2). **(b)** SDS gel showing the purification of 6xHis tagged EncA from *E. coli* Rosetta(DE3). Each lane in the SDS-PAGE gel is labelled, from left to right: Lane 1: whole cell lysate. Lanes 2-4: three wash steps, the concentration of the EncA protein exceeded the bed volume of the column therefore some of this protein is eluted during the wash steps. Lanes 5-7: three elution steps, here the band at ~36kDa representing the EncA protein can be clearly seen. Lane 8: See Blue Prestained protein ladder (Invitrogen), the bands representing 50, 36, 30 and 16kDa are labelled. 55
- 2.4 SDS-PAGE analysis of encapsulins expressed in *E. coli* and loaded with different cargo proteins.** Lanes are labelled 1-6 from left to right. Lane 1: See Blue Prestained protein laddered (Invitrogen), the bands representing 50, 36, 30, 16 and 4kDa are labelled. Lane 2: sample of empty encapsulin. The band representing the EncA shell protein can be seen at 36kDa. Lane 3: sample of encapsulin containing the cargo protein EncB. The bands representing both the EncA shell protein (36kDa) and the EncB cargo protein (17kDa) can be clearly seen. Lane 4: sample of encapsulin containing the cargo protein EncC. The bands representing both the EncA shell protein (36kDa) and the EncC cargo protein (13kDa) can be clearly seen. Lane 5: sample of encapsulin containing the cargo protein EncB. The bands representing both the EncA shell protein (36kDa) and the EncD cargo protein (11kDa) can be clearly seen. Lane 6: sample of encapsulin containing the cargo proteins EncB and EncC. The bands representing both the EncA shell protein (36kDa), EncB cargo protein (17kDa) and EncC cargo protein (13kDa) can be clearly seen. . . 58

2.5	Electron micrographs of generated encapsulin constructs negatively stained with 1% uranyl acetate. (a) EncA (b) EncA+EncB (c) EncA+EncC (d) EncA+EncD (e) EncA+EncB+EncC . In the top right of each image there is a insert showing a selection of negatively stained encapsulins. There is variation between the structures within an isolation, this is most likely due to the staining pooling in the middle of the encapsulin causing them to collapse inwards to varying degrees. The images were captured using a Phillips CM100 operating at 80 keV at a magnification of 15.5kx. Image scale bars represent 100 nm and insert scale bars 20 nm. Importantly, the electron micrographs show no discernible difference between these encapsulins overexpressed in <i>E. coli</i> and encapsulins isolated from <i>M. xanthus</i>	61
2.6	Presence of cargo proteins stabilizes the formation of the larger T=3 encapsulin shell. (a) templates generated of T=1 and T=3 shells. Top left: example of a negatively stained T=1 shell and top right: corresponding template of this structure. Bottom left: example of a negatively stained T=3 shell and bottom right: corresponding template of this structure. The templates consisted of two blurred concentric circles. (b) Bar chart showing the number of observed T=1 shells compared to T=3 shells for the different encapsulin cargo complexes expressed in <i>E. coli</i> . (c) The result of template matching of T=1 and T=3 templates to a negatively stained image of EncA proteins. Encapsulins identified as T=1 shells are marked with red dots, and those identified as T=3 shells are marked with green dots.	63
3.1	Flow chart showing steps taken to identify and measure encapsulin shells.	74
3.2	Flow chart showing steps taken to identify and measure encapsulin cores.	75
3.3	Flow chart showing steps taken to identify and measure nanoparticles properties.	76

3.4	Measuring the thickness of the encapsulin shell using different imaging techniques. (a) BF (left), HAADF (middle) and SE (right) images of <i>M. xanthus</i> encapsulins. Images captured simultaneously on a Hitachi HF5000 FE-S/TEM at a magnification of 60kx. Scale bars represents 50 nm. (b) Three magnified examples of shell recognition from (a). Top to bottom: BF image, HAADF image, SE image, HAADF image overlay with shell region from BF images (blue), HAADF overlaid with shell regions from HAADF images (green) and HAADF images with shell region from SE images (red). (c) Distribution of shell thickness calculated from HAADF (green) and SE images (red). Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown.	79
-----	--	----

3.5	Comparison of <i>M. xanthus</i> core size from different isolations.(a) HAADF image of encapsulin isolation 1, imaged on a JEOL F200 operated at 200kV and a magnification of 500kx (b) HAADF image of encapsulin isolation 2, imaged on a JEOL F200 operated at 200kV and a magnification of 600kx. (c) HAADF image of encapsulin isolation 3, imaged on a Hitachi HF5000 FE-S/TEM operating at 200keV at a magnification of 60kx. (d)-(f) Histograms presenting the average size of the Fe rich core measured for the three <i>M. xannthus</i> isolations: isolations 1 (red), isolation 2 (green), isolation 3 (green). Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown. (g) Scatter plot showing the major and minor diameters of the three isolations.	81
-----	---	----

- 3.6 Measurements of the nanoparticle diameters and inter-nanoparticle distances for the *M. xanthus* using two imaging methods.** (a) Unstained electron micrograph captured on a JEOL F200 at a magnification of 400kx. Scale bar represents 20 nm. Insert: two encapsulins, example nanoparticles highlighted with green arrows. (b) HAADF image of encapsulins captured on a JEOL ARM 200F at a magnification of 3milx. Scale bar represents 20 nm. Insert: A single encapsulin with example nanoparticles highlighted with green arrows. (c) - (d) Histograms representing the average nanoparticle diameters calculated from TEM and HAADF images respectively. (e) - (f) Histograms representing the average inter-nanoparticle distances calculated from BF TEM and HAADF STEM images respectively. Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown. 84
- 3.7 Unstained electron micrographs of encapsulins containing single cargo proteins after growth in media containing 4 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$.** (a) EncA, (b) EncA+EncB, (c) EncA+EncC (d) EncA+EncD. The Fe mineralized regions appear as the areas with darker contrast in the images, while the encapsulin shells are visible as the low contrast regions. Electron micrographs of unstained sample were taken at a magnification of 15.5kX on a Phillips CM100. Scale bars represent 100nm. 86
- 3.8 Encapsulin constructs grown in media containing different concentrations of $\text{Fe}(\text{NH}_4)_2\text{SO}_4$.** (a) HAADF images of unstained samples from each growth condition. Scale bar represents 50 nm. The images captured in the 2 mM and 4 mM conditions have lighter backgrounds because carrier grids with thicker C films were used during data collection. (b) Bar chart comparing the number of shells that showed iron mineralization for the different constructs generated under different concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ in the media. . . . 89

- 3.9 Magnetic separation of encapsulins loaded with Fe.** SDS gels showing the results of testing the magnetic properties of encapsulin derived nanoparticles using MACS cell separator columns. **(a)** *M. xanthus* encapsulins (left) and EncA+EncB (right). **(b)** EncA+EncC (left) and EncA+EncB+EncC (right). Both gels have 7 lanes that contain the following samples: Lane 1 and Lane 5: Flow through, Lane 2 and Lane 6: wash step, Lane 3 and Lane 7: elution step, Lane 4 contains the See Blue Prestained protein ladder. The bands representing 50kDa, 36kDa, 30kDa, 16kDa, 4kDa are labelled. 91
- 3.10 Histograms representing the average core size measured inside encapsulins containing different combinations of cargo proteins after growth in media supplemented with different concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$.** Particle diameter was calculated from binary images generated using a multi-Otsu threshold from HAADF images [165]. **(a) - (c)** Histograms from samples of EncA+EncB. **(d) - (f)** Histograms from samples of EncA+EncC. **(g) - (i)** Histograms from samples of EncA+EncB+EncC. For all samples, core diameter after growth in 2 mM, 4 mM and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ are shown (left to right). The number of cores measured (n), the sample mean ($\bar{x}$) and standard deviation (σ) for each sample are shown above each histogram. 93
- 3.11 Largest diameter cores observed from encapsulins expressed in *E. coli*.** Encapsulins were induced using 0.01 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ was added to the media after 2 hours. Particle diameter was calculated from binary images generated using a multi-Otsu threshold from HAADF images [165]. **(a)-(c)** Histograms representing the average core size measured for EncA+EncB, EncA+EncC and EncA+EncB+EncC respectively. The number of cores measured (n), the sample mean ($\bar{x}$) and standard deviation (σ) for each sample are shown above each histogram. 94

- 3.12 Nanoparticle diameters measured for encapsulins overexpressed in *E. coli*.** (a) HAADF reference image of EncA+EncB. An insert is included where a number of individual nanoparticles are highlight by green arrow. (b) Histogram of average particle size measured from samples of EncA+EncB. (c) Reference HAADF image of EncA+EncC. An insert is included where a number of individual nanoparticles are highlight by red arrow.(d) Histogram of average particle size measured from samples of EncA+EncC. All reference images were captured on a JEOL F200 at 250kx magnification. Scale bars represent 50 nm. Above each histogram the number of particles measured (n), the mean ($\bar{x}$) and standard deviation (σ) (both in nm) are given. 96
- 3.13 Observation and localisation of individual ~ 2 nm nanoparticles.** Images of a sample of EncA+EncB+EncC where individual nanoparticles can be observed. (a) Reference HAADF image, the contrast has been enhanced for easier observation of individual nanoparticles. The insert shows a high magnification region where the individual ~ 2 nm nanoparticles can be clearly seen. Scale bar represents 10 nm. (b) The same image with the individual nanoparticles labelled with red and the boundary of the encapsulin proteins in the image labelled with green. The boundary does not perfectly track the expected shape of the protein shell, this is because of the Gaussian filter (width 2 pixels) applied to the image to reduce noise. The localisation of both features to the same image region is clearly evident. (c) Histogram representing the average particle diameter. The number of particles measured (n), the mean ($\bar{x}$) and standard deviation (σ) (both in nm) are given above the histogram. 98
- 3.14 Plots showing the intensity of the core regions from HAADF images against the diameter of the core.** The constructs shown are: (a) *M. xanthus* encapsulins, (b) EncA+EncB, (c) EncA+EncC, (d) EncA+EncB+EncC. The three samples shown in (b)-(d) are from encapsulins expressed in *E. coli* and grown in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The HAADF image contrast was normalised by dividing the intensity value of each pixel in the image by the highest intensity pixel. 100

- 3.15 Loading encapsulins containing EncB and EncC with different concentrations of Fe *in vitro*.** The following loading conditions were used: **(a)** 0 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, this is a negative control showing the empty protein shells. **(b)** 0.25 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ **(c)** 1 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ **(d)** 2 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ **(e)** 4 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and **(f)** 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The insert in each image is a representative example of a single shell for each growth condition, the scale bars all represent 20 nm. The increase in contrast of the core region as the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ increases can be clearly seen in these BF TEM images. The micrographs **(a)** - **(d)** were captured on the Tecnai Spirit at 100 kx and micrographs **(e)** - **(f)** were captured on the JEOL F200 at 100 kx. The scale bars in all images represent 50 nm. **(g)** Graph of the average core size measured from BF TEM images against the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The data for each loading condition is represented by a colour: blue (0.25 mM), red (1 mM), green (2 mM), purple (4mM) and orange (8mM). For each growth condition, 100 core diameter measurements are represented. . 102
- 3.16 The diameter of nanoparticles inside encapsulin shells supplemented with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$.** **(a)** HAADF reference image of sample. The image was captured using a JEOL F200 at a magnification of 1Mx. The scale bar represents 20 nm. **(b)** Higher magnification HAADF image of single encapsulin captured on a FEI Titan G2 CHEMISTEM at 3.6Mx magnification. Individual 2 nm nanoparticles have been highlighted with green arrows. The scale bar represents 5 nm. **(c)** Histogram representing the distribution of measured particle diameters. Above it, the sample size (N), the mean ($\bar{x}$), and the standard deviation (σ) (both in nm) are quoted. 103
- 3.17 4D-STEM investigation of the encapsulins isolated from *M. xanthus*.** The images are paired, **(a)/(b)**, **(c)/(d)** and **(e)/(f)**. The first image in the pair is a reference BF image of the investigated region. Highlighted in red is the probe position from which the diffraction pattern in the second image is taken from. The scale bar represents 50 nm. In the diffraction pattern, partial diffraction spots are highlighted with red circles. Scale bar represents 0.02 mrad. 105

- 3.18 **HR-STEM imaging of the encapsulin from *M. xanthus*.** (a) Reference BF image of encapsulin. (b) Reference HAADF image of encapsulin. (c) HAADF image of encapsulin where three potentially crystalline regions are highlighted in boxes and labelled. (d) Higher magnification images of the three identified ROIs 1-3, top to bottom. The lattice plans have been highlighted with yellow lines. *D*-spacings from top to bottom: 203 ± 8 pm, 242 ± 12 pm and 168 ± 8 pm. Scale bar represents 5 Å. 107
- 3.19 **SAED from core regions of encapsulins containing either EncB or EncC.** Diffraction patterns from cores of (a) EncA+EncB and (c) EncA+EncC. Images were captured on a JEOL F200 at 200keV, with a camera length of 800 mm and the 50 μ m selected area aperture inserted. The diffraction patterns represents a densely loaded 800x800 nm area of the grid. In both images, only a diffuse rings representing the amorphous material of the carbon film can be seen. 108
- 3.20 **HR-STEM images of encapsulins containing the cargo proteins EncB and EncC loaded with 8 mM of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$.** (a) HAADF image of an encapsulin captured using a FEI Titan G2 80-200 ChemiSTEM operating at 200 keV. The image was captured at 7.2Mx magnification and the scale bar represents 2 nm. The three ROIs highlight regions that show a regular arrangement of atoms. The bright spots in the image are probably individual Fe atoms. As they appear across the whole image they are most likely not associated with the internal nanoparticles. Several image features appear to be horizontally elongated; these are most likely scan artefact together with sample drift during data collection. This scan artefact is a known issue with this microscope set up at higher magnifications. (b) Higher magnification images of the three ROIs labelled in (a). From top to bottom the images are ROIs 1-3. On each image the lattice planes are highlighted with yellow lines. The scale bars represent 5 Å. 110

3.21	Identification of nanoparticles from TEM images. (a) Reference TEM image of unstained encapsulins captured at 300kx magnification. Scale bar represents 50 nm. The insert shows a single unstained encapsulin. Scale bar represents 5nm. (b) Same as (a). Overlay of image regions defined as ‘nanoparticle’ after image processing. The outlines of the identified regions are shown in red.	115
3.22	Identification of nanoparticles from HAADF images. (a) Reference HAADF image of unstained encapsulins captured at 300kx magnification. Scale bar represents 20 nm. The insert shows a single unstained encapsulin. Scale bar represents 5 nm. (b) Same as (a). Overlay of image regions defined as ‘nanoparticle’ after image processing. The outlines of the identified regions are shown in red.	116
4.1	Chemical mapping of encapsulin isolated from <i>M. xanthus</i>. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The L α and K α lines of all elements present are labelled. (b) HAADF reference image of sampled region, the scale bar represents 50 nm. (c) Elemental maps showing the spatial distribution of C_K α , N_K α , O_K α , Na_L α , Mg_K α , Si_K α , P_K α , Cl_K α , Ca_K α , Mn_K α , Fe_K α . Cu is present as the sample has been imaged on a Cu carrier grid, however the chemical map has been omitted as the intensity of its signal causes the signal from the other elements to become lost. An intensity bar representing the contrast against counts can be seen on the right. The max counts for many elements are 0 which is an artefact caused by the binning of data and the subsequent rounding of the number of counts.	128
4.2	EDXS analysis of <i>M. xanthus</i> encapsulins on Au grids. (a) Summed EDXS spectrum of all probe positions. The spectrum is plotted on a log scale for easier visualisation. The L α and K α lines of all elements present are labelled. The insert is a magnification of the spectral region between 0-1.5 keV. (b) Elemental maps generated from the Fe_K α and Cu_K α lines. An intensity bar representing the contrast against counts can be seen on the right.	130

- 4.3 **Chemical mapping of elemental cores inside EncA+EncB shells. (a)**
Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. **(b)** HAADF reference image of spectral image. Scale bar represents 50 nm. **(c)** Chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cu_K α , Fe_K α , K_K α , Mg_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , S_K α , Si_K α and Ti_K α . An intensity bar representing the contrast against counts can be seen on the right. 132
- 4.4 **Chemical mapping of elemental cores inside EncA+EncC shells. (a)**
Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. **(b)** HAADF reference image of spectral image. Scale bar represents 20 nm. **(c)** chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cu_K α , Fe_K α , K_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , Si_K α . An intensity bar representing the contrast against counts can be seen on the right. 133
- 4.5 **Chemical mapping of elemental cores inside EncA+EncB+EncC shells.**
(a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. **(b)** HAADF reference image of spectral image. Scale bar represents 20 nm. **(c)** chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cu_K α , Fe_K α , K_K α , Mg_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , S_ α and Si_ α . An intensity bar representing the contrast against counts can be seen on the right. 134

- 4.6 **EDXS chemical maps of EncA+EncB+EncC loaded with Fe following the addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$.** (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. the L- and K α lines of all elements present are labelled. (b) Reference HAADF image of the SI. Scale bar represents 20 nm, the C contamination build up during scanning can be seen. (c) Chemical maps showing the spatial distributions of C_K α , Ca_K α , Cl_K α , Cu_K α , Fe_K α , K_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , Si_K α , Ti_K α . An intensity bar representing the contrast against counts can be seen on the right. 136
- 4.7 **Chemical mapping of empty EncA+EncC encapsulins.** (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. (b) HAADF reference image of spectral image. Scale bar represents 50 nm. (c) Chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cr_K α Cu_K α , Fe_K α , K_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , Si_K α . An intensity bar representing the contrast against counts can be seen on the right. 142
- 5.1 **Example binary masks of encapsulin shell.** (a) Reference ADF image, scale bar represents 20 nm. (b) Binary mask of encapsulin shell, generated from a single class defined using a multi-Otsu threshold [165]. The foreground in the binary mask (in (b)) matches the region of low contrast surrounding the bright Fe rich core in the ADF image (in (a)). 147
- 5.2 **Chemical maps generated from EELS spectra of encapsulins from *M. xanthus*.** (a) HAADF reference image of sampled region. The reference image is 528x528 pixels and the SI maps are 66x66 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction. 153

- 5.3 **Chemical maps generated from EELS spectra of the EncA+EncB encapsulins.** (a) HAADF reference image of sampled region. The reference image is 408x544 pixels and the SI maps are 56x63 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction. 154
- 5.4 **Chemical maps generated from EELS spectra of EncA+EncC encapsulins.** (a) HAADF reference image of sampled region. The reference image is 416x552 pixels and the SI maps are 52x69 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction. 155
- 5.5 **Chemical maps generated from EELS spectra of EncA+EncB+EncC encapsulins.** (a) HAADF reference image of sampled region. The reference image is 608x392 pixels and the SI maps are 76x49 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction. 156
- 5.6 **Chemical maps generated from EELS spectra of EncA+EncB+EncC encapsulins loaded with Fe *in vitro*** (a) HAADF reference image of sampled region. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.5 eV/ch. This SI was captured on the JEOL F200 operating at 200 keV. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction. . . . 157

- 5.7 Comparison of P and S localisation within a single *M. xanthus* encapsulin.** (a) HAADF reference image of single encapsulin. (b) P L-edge map of counts. (c) S L-edge map of counts. (d) Energy loss spectrum covering range of 125 - 190 eV from a single encapsulin. The overlapping P and S L-edges can be clearly seen. The edge onsets are labelled with green dotted lines. (e) Map of P distribution by MLLS fitting over 15 eV window (136-151 eV). (f) Map of S distribution by MLLS fitting over a range of 15 eV (165-180 eV), showing the anti-correlation to the P distribution. 159
- 5.8 The Ca L_{2,3} edge can be observed in the core regions of samples generated in *E. coli*.** Spectrum from the following constructs are shown: *M.xanthus*: (a) core, (b) shell. The following samples have all been expressed in *E.coli*. EncA+EncB: (c) core, (d) shell. EncA+EncC: (e) core,(f) shell. EncA+EncB+EncC (g) core, (h) shell. EncA+EncB+EncC loaded *in vitro* (i) core, (j) shell. The green dotted line marks the onset of the Ca L_{2,3} edge. Individual EELS spectra are shown instead of chemical maps, because of the difficulty in modelling the Ca edge due to its low signal and proximity to the C K-edge. 162
- 5.9 Chemical maps showing the loss of the encapsulin shell following exposure to increasing electron dose.** (a) HAADF reference image of sampled encapsulin. The encapsulin imaged here is composed of EncA+EncB loaded with 8 mM Fe in the presence of 0.01 mM of the inducer in *E. coli*. (b) Chemical maps generated from the S L_{2,3}, C K, N K and O K-edges. The electron dose increases from 34 $e^-/\text{\AA}^2$ in the first row and ends at 120 $e^-/\text{\AA}^2$ in the final row. 166
- 5.10 Chemical composition of individual nanoparticles within encapsulin shells from *M. xanthus*.** Three examples of individually resolved nanoparticles are shown. Left to right: HAADF reference image of sampled encapsulin, map generated from the Fe L_{2,3} edge with the sampled individual nanoparticles highlighted with green boxes and tabulated atomic composition of the highlighted nanoparticles. Sampled ROIs listed in the table moving down the image from top (ROI 1) to bottom (ROI 3). 170

- 5.11 **Normalised HAADF intensity against the estimated number of Fe atoms inside encapsulins from *M. xanthus*.** A linear relationship is expected between the HAADF intensity and no. of Fe atoms [202]. A linear fit to the experimental data yields an R^2 value of 0.504. The relatively poor fit and the observed offset of 0.3968 are most likely due to additional contrast contributions from both the protein shell and the carbon coating of the support film. 171
- 5.12 **Estimates of the number of Fe atoms per encapsulins shell for the different encapsulin constructs generated in this work.** For each dataset 30 encapsulins were measured. **MX** (blue) = Complete encapsulins isolated from *M. xanthus*. **AB** (red)= EncA+EncB, **AC** (green)= EncA+EncC **ABC** (yellow)= EncA+EncB+EncC. For AB, AC and ABC the encapsulins were overexpressed in *E. coli* grown in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and 0.05 mM of the inducer IPTG. **ABC *in vitro*** (purple)= EncA+EncB+EncC construct, empty encapsulin constructs were isolated and loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ *in vitro*. 173
- 5.13 **Chemical mapping of individually distributed nanoparticles after sequential scanning of the same area.** Left to Right: HAADF reference image. Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K, Fe L_{2,3}. 175
- 5.14 **Fe L_{2,3} EELS edge fingerprinting against reference standards.** The two vertical dotted black lines represent the maxima values of the two end reference materials, ilmenite (708.7 eV) and haematite (710.5 eV). MX represents the encapsulins isolated from *M. xanthus*. All spectra were collected on the JEOL F200 operating at 200 keV, using the 30 μm condenser aperature with a nominal spot size of 6 and a dwell time of 0.01s. The spectra have been deconvolved using Fourier-ratio deconvolution and background subtracted with an inverse power law between 600-700 eV. Dual EELS spectra were collected with an offset of 660 eV and a dispersion of 0.1 eV/ch. 177

5.15	The change in the Fe $L_{2,3}$ edge of the EncA+EncB+EncC core with increasing electron dose. The electron dose increases from (a) to (c) and from (d) to (f). For a given exposure time, the electron dose is higher when using probe size 4 compared to probe size 6. The experimental conditions used were: (a) probe size 6 at 0.01 s/pixel, (b) probe size 6 at 0.1 s/pixel, (c) probe size 6 at 1 s/pixel, (d) probe size 4 at 0.01s/pixel dwell time, (e) probe size 4 at 0.1 s/pixel, (f) probe size 4 at 1 s/pixel. All spectra were collected using a JEOL F200 operating at 200 keV. The spectra have been deconvolved using Fourier-ratio deconvolution and background subtracted with an inverse power law between 600-700 eV. Dual EELS spectra were collected with an offset of 660 eV and a dispersion of 0.1 eV/ch. The electron probe was centred on an encapsulin and the single reading for the given time was captured. For (c)-(f) the spectra represent 25 summed measurements. However, due to the poor signal intensity generated for the imaging conditions in (a) and (b) 25 measurements was too low to give a useful spectra, resulting in far higher numbers of measurements needing to be summed. This is why the signal to noise ratio is higher in these two spectra.	179
5.16	Plot of the $L_3/(L_3+L_2)$ against the L_3 maximum peak position. The reference material calibration curve (black), encapsulins from <i>M. xanthus</i> (red), EncA+EncB (blue), EncA+EncC (green) and EncA+EncB+EncC (purple) are shown on the plot. The $L_3/(L_3+L_2)$ ratio was calculated from two windows (708 eV to 710 eV and 719.7 eV to 721.7 eV).	181
5.17	Indexed diffraction patterns of polycrystalline gold used to calculate convergence angles. (a) 10 μm condenser aperture and (b) 40 μm condenser aperture.	186

6.1	Proposed model for iron nanoparticle arrangement inside the encapsulin based on the experimental results of this work.	
	The model contains various elements: the EncA protein shell (blue), the cargo proteins EncB (green) and EncC (purple), Fe^{2+} (orange) and Fe^{3+} (red). Free Fe^{2+} in the cytoplasm of the cell is transported through the pores in the encapsulin shell where it contacts in the ferritin like cargo proteins attached to the inner surface of the encapsulin shell. Here the Fe^{2+} is oxidised into Fe^{3+} and transported to the surface of the cargo protein for mineralization to occur. The resulting nanoparticle product remains bound to the surface of the cargo protein, where it is stored.	193

List of Tables

1.1	Table highlighting a selection of known encapsulin systems and their cargos.	12
2.1	Bacterial strains used in this study.	40
2.2	List of the primers sequences used in this study.	44
2.3	List of plasmids generated as part of this study. The parent plasmid strains are also included in the table.	45
2.4	The efficiency of cargo loading in encapsulin constructs calculated using SDS densitometry. The data from native <i>M. xanthus</i> encapsulins isolated by other groups is included as a reference [18].	60
3.1	Encapsulins generated for use in this study. For the encapsulins expressed in <i>E. coli</i> the concentration of IPTG used to induce the protein expression and the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ added to the growth media are shown.	71
3.2	Electron microscopes used during the course of this study. The microscopes, operating voltage in kV, imaging capabilities, and locations are given. The following acronyms are used: Transmission Electron Microscope (TEM), Bright Field (BF), Scanning/Transmission Electron Microscope (S/TEM), Scanning Transmission Electron Microscope (STEM), High Angle Annular Dark Field (HAADF), Selected Area Electron Diffraction (SAED), Annular Dark Field (ADF), Secondary Electron (SE) and 4 Dimensional Scanning Transmission Electron Microscopy (4D-STEM).	72

3.3	Calculation of the volume of internal cargo proteins. The volumes were calculated based on the published atomic models [28, 97]	83
3.4	Table comparing the HAADF intensity of the Fe rich cores of different encapsulin constructs from normalised HAADF images. The mean ($\bar{x}$), standard deviation (σ) and number of cores measured (N) are given.	99
3.5	The lattice parameters of iron oxides phases used in this study. The valence state, chemistry, name, space group, space group number (No.), a brief description of the space group (comment), the lattice parameters (in pm) and the source the lattice parameters were taken from.	117
3.6	Table of the measured d-spacings from HR-STEM images of the nanoparticles inside encapsulins isolated from <i>M. xanthus</i>. The errors represent the uncertainty when measurements were made from selected line profiles. the different iron oxides phases are colour coded: ferrihydrite (red), goethite (green), haematite (blue), iron(III) phosphate (purple), magnetite(orange), maghemite (yellow) and wuestite(black).	120
3.7	Table of the measured d-spacings from HR-STEM images of EncA+EncB+EncC. Empty EncA+EncB+EncC encapsulins were purified and loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The error represents the uncertainty when the measurements were made from selected line profiles. The different iron oxides phases are color coded: ferrihydrite (red), goethite (green), haematite (blue), magnetite (purple), maghemite (yellow) and wuestite (black).	122
4.1	Table of electron microscopes used to collected EDXS data during this study. The machine, operating voltage, capabilities and location are given.	124
4.2	Table of encapsulins samples used in this work. The host bacterium, proteins present in each encapsulin, and the experimental conditions under which nanoparticle synthesis occurred are given.	125
4.3	Atomic composition of the nanoparticles inside the different encapsulins. The stdev represents one standard deviation. Quantifications were carried out in Digital Micrograph using the k -factors from the software.	138

5.1	Table of encapsulins constructs used in this work. The host organism and the experimental conditions used to generate the samples are indicated. . .	144
5.2	Core edge fitting parameters used in both Digital Micrograph and HyperSpy. The numerical values are given in eV.	148
5.3	Atomic% of Ca in the core regions of encapsulin constructs. Pixels representing encapsulin cores were summed across all frames in a single data set. Atomic% was calculated using HyperSpy.	163
5.4	Atomic composition of EncA (Gene Identifier: MXAN_3556). The number of atoms, relative composition, and percentage composition of a single EncA monomer calculated using the Xpansy ProtParam tool [214]. The data for H has been omitted.	164
5.5	Atomic composition of the encapsulins protein shell calculated using routines in both Digital Micrograph and HyperSpy. The protein shells from samples isolated from <i>M.xanthus</i> were used for analysis.	164
5.6	Atomic composition of the iron containing nanoparticles of generated encapsulins. Stdev is one standard deviation.	169
5.7	Calculated convergence and collection angles used for a camera length of 120 mm.	186

0.3 List of abbreviations

6xHis = 6 Histidine residues

ADF = Annular Dark Field

A-domain = Axial Domain

Amp = Ampicillin

BF = Bright Field

CCD = Charge-Coupled Device

cNMP = cyclic Nucleotide

cryo-EM = Cryo Electron Microscopy

CsCl = Caesium Chloride

CTT = Casitone Tris Media

ddH₂O = Double Distilled Water

DMSO = Dimethyl Sulfoxide

Dps = DNA Binding Proteins from Starved Cells

DyP = Dye-Decolorizing Peroxidase

EDS = Energy-Dispersive X-ray Spectroscopy

EELS = Electron Energy-Loss Spectroscopy

ELNES = Electron Near Edge Structure

E-loop = Elongated Loop

FLP = Ferritin Like Protein

FOC = di-iron Ferroxidase Centre

FWHM = Full Width Half Maxima

GIF = Gatan Imaging Filter

GuHCl = Guanidine chloride

HAADF = High Angle Annular Dark Field

HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HR-STEM = High Resolution Scanning Transmission Electron Microscopy

IMEF = Iron Mineralizing activity Encapsulin-associated Firmicute

IPTG = Isopropyl β -D-1-thiogalactopyranoside

HL = High Loss

Kan = kanomycin

LIST OF TABLES

LB = Lysogeny Broth

LL = Low Loss

MCS = Multiple Cloning Site

MLLS = Multiple Linear Least Square

OD₅₉₅ = Optical Density at wavelength of 595nm

OD₆₀₀ = Optical Density at wavelength of 600nm

ROI = Region Of Interest

ROS = Reactive Oxygen Species

RPM = Revolutions Per Minute

PBS = Phosphate Buffered Saline

PCA = Principle Component Analysis

PCR = Polymerase Chain Reaction

P-domain = Peripheral Domain

SAED = Selected Area Electron Diffraction

SDD = Silicon Drift Detectors

SDS-PAGE = Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SE = Secondary Electron

STEM = Scanning Transmission Electron Microscopy

TEM = Transmission Electron Microscopy

TPM = Tris-Phospahte Magnesium starvation media

ZLP = Zero Loss Peak

Chapter 1

Introduction

This chapter is adapted from the following publications, published during the period of study:

McDowell H.B and Hoiczky, E. (2022) Bacterial nanocompartments: structures, functions and applications. *Journal of Bacteriology*, 204 (3). e00346-21. ISSN 0021-9193

A. R. Quinton, **H. B. McDowell**, and E. Hoiczky, Encapsulins: Nanotechnology's future in a shell, p. 1–48. Elsevier, 2023.

1.1 Compartmentalization

Compartmentalization can be defined as the process of dividing a larger space into smaller functional units. For cellular life, this involves the segregation of metabolic processes into defined intracellular compartments; creating specific micro-environments for enhanced efficiency and protection from toxic intermediates [1]. The most commonly cited examples of cellular compartmentalization are the chloroplast and mitochondria - large membrane bound organelles - which are responsible for photosynthetic and respiratory functions within eukaryotes [2, 3]. The absence of observable parallels to these structures in prokaryotes led to the long-standing belief that prokaryotic life was too simple to exhibit the level of structural complexity characteristic of eukaryotes.

However, this long-held belief began to shift with technological advancements in microscopy, which first led to the discovery of gas vesicles. Gas vesicle, are a large sub-cellular compartments constructed of protein rather than lipids and with a size which allowed it to be observed using a compound light microscope [4, 5]. Following this, improvements in electron microscopy led to the observation of the first protein based microcompartment: the carboxysome [6]. This is a 40-200 nm (quasi)isocahedral shell that contains carbonic anhydrase and RuBisCO and is essential in the carbon fixation step of photosynthesis [7]. Similar microcompartments, later deemed metabolosomes, were identified in *Salmonella* when the bacteria was grown on ethanolamine or 1,2-propanediol [8, 9, 10, 11]. It was however not until 2008 that the first bacterial nanocompartment was identified: the encapsulin [12].

1.2 Encapsulins

1.2.1 On the genealogy of encapsulins

Encapsulins are a class of recently discovered protein nanocompartments, so called because their diameters do not exceed 100 nm. The first known observation of these structures was made in 1994, when high molecular weight aggregates were observed in the supernatant of *Brevibacterium linens* [13]. Following this initial identification, observations were made of highly homologous proteins in *Mycobacterium leprae* [14], *Mycobacterium tuberculosis* [15] and *Thermotoga maritima* [16]. However, the nomenclature of ‘encapsulin’ only became common-

place after the atomic structures were solved for systems from the archaea *Pyrococcus furiosus* and the hyperthermophilic bacterium *T. maritima* [17, 12] in 2007 and 2008 respectively. Since the initial discovery, encapsulins have been identified and studied in detail in a wide range of different species including: *Myxococcus xanthus* [18], *Quasibacillus thermotolerans* [19], *Rhodococcus jostii* [20] and *Synechococcus elongatus* [21]. More recently still, large scale genetic studies have identified that these systems are widespread with over 6000 homologues in 31 bacterial and 4 archeal phyla [22].

1.2.2 Family units

Encapsulins have been classified into four families based on structural similarities between the different members. The families have been denoted as Family 1 (classical encapsulins), Family 2, Family 3 and Family 4 [22, 21]. Family 1 encapsulins are the best studied and their functions have been inferred from their cargo proteins. These encapsulins are involved in protecting the bacteria from oxidative stress, most commonly from reactive oxygen species (ROS), by either isolating ferrous iron from the greater cell volume [19, 23, 18] or through the activity of encapsulated dye-decolorizing peroxidase (DyP) enzymes [24, 25]. Family 2 encapsulins are predicted to be the most abundant family and are sub-divided into either 2A or 2B depending on the presence (2A) or absence (2B) of an internal cNMP-binding domain [22]. Currently, experimental work has focused on Family 2A encapsulins identifying that they are upregulated under sulphur starvation in the species *Synechococcus elongatus* and *Acinetobacter baumannii* [21], indicating a role in sulphur metabolism. No functions for Family 2B encapsulins have been experimentally verified, and all functions are inferred from bio-informatics. Family 3 encapsulins, also known as ‘natural product encapsulins’, have been linked with the synthesis of peptide based products. However, exact functions for these systems remain unclear. Finally, Family 4 encapsulins, or ‘A-domain encapsulins’, are highly truncated HK97-fold containing constructs, found predominantly in hyperthermophilic Archaea. However, as with Family 3 systems to date no experimental investigation has been undertaken, and thus their function remains speculative [22].

1.2.3 Shell structure

A number of encapsulin shell structures have been determined using both X-ray crystallography and cryo-electron microscopy (Cryo-EM): *P. furiosus*, *T. maritima*, *M. xanthus*, *Q. thermotolerans*, *Syn. elongatus* and *Haliangium ochraceum* [17, 12, 18, 19, 21, 26]. These structures revealed that encapsulin shells are icosahedral, self-assembling from a single repeating protomer that organizes into both hexameric and pentameric facets. To form the larger icosahedral structure, a minimum of 12 pentamers are required, however the scale of the structure can be increased through the addition of hexamers. To quantitatively describe the complex organisation of the icosahedral shell, the T value is used, which was originally introduced to characterize viral capsids [27]. To date, three different encapsulin shell sizes have been described.

The smallest encapsulins are T=1 structures, and have been identified in the species, *T. maritima* [12], *R. jostii* [20] and *M. tuberculosis* [24]. These structures are 20-24nm in diameter formed of 60 monomers arranged into 12 pentameric faces.

The second identified class of encapsulin structure are larger and denoted as T=3 structures. T=3 encapsulins are formed from 180 monomers arranged into 12 pentameric and 20 hexameric faces with a diameter of 32nm. Currently, encapsulins with these dimensions have been explored in *M. xanthus* [18] and *P. furiosus* [17]. Curiously, the encapsulin from *M. xanthus* shows a divergent morphology when overexpressed in *E. coli*. A minority of structures show T=1 topology, the percentage of which has been reported as 36% [28] and 20% [18]. Additionally, the T=3 encapsulin shell from *M. xanthus* can show heterogeneity when overloaded with cargo proteins: a percentage of shells show fusions, elongations or expansions [29]. Moreover, point mutations to key residues of the shell protein result in the formation of tetrahedral shells [30].

Finally, the largest encapsulin is classified as T=4 and is found in the species *Q. thermotolerans* [19]. This structure is formed from 240 monomers organised into 12 pentameric and 30 hexameric faces and has a diameter of 42 nm. To date, this is the largest known encapsulin structure, with an internal volume that is 530% larger than T=1 shells and 220% larger than T=3 shells.

The extent of our structural knowledge about Family 2 encapsulins comes from the

atomically resolved structure of the encapsulin from *Syn. elongatus*. This structure is 24.5 nm in diameter and is formed from 60 monomers arranged into 12 pentameric faces [21]. The structure of this encapsulin has a 'spikey' morphology, that marks it as different from known Family 1 structures. The morphology is caused by an extended C-terminus which protrudes from the capsid along the 5-fold axis of symmetry.

1.2.4 Structure of basal shell unit

The base unit of the encapsulin shell is a HK97 like-fold containing protein. So called because of its similarity to the major capsid protein of the T=7 HK97 lambdoid bacteriophage [31, 32], despite the lack of sequence similarity [17, 12]. Since its initial discovery, the HK97 fold has been identified a key architectural feature of many different viral families [33, 34]. Structurally, the HK97 fold is defined by its three domains, the axial (A-domain), peripheral (P-domain) and elongated domain (E-loop) (See Figure 1.1) [32].

The A-domain is found at the C-terminus and represents a major structural element of the HK97 fold. It is formed from the organisation of three α -helices and five β -sheets arranged in either parallel or anti-parallel orientation, and is essential for the interaction of different monomers in the formation of the hexameric and pentameric facets [32]. Interestingly, it is differences in this domain that cause the 'spikey' morphology of the Family 2A encapsulin [21] (See Figure 1.1a).

The P-domain is found at the N-terminus of the monomer and covers the majority of its sequence. It is comprised of a two long α -helices, two short α -helices and three anti-parallel β -sheets. Protected within this domain is the hydrophobic binding pocket which is essential for the interaction between the encapsulin shell and its internalized cargo [12, 28]. Differences in the arrangement of the N-terminus marks another key structural difference between the Family 1 and Family 2 structures. Instead of the extended helix present in the Family 1 structures, those from Family 2 possess a shorter helix and an extended N-terminal arm [21]. Of note, modifications to this N-terminal arm, leads the *Syn. elongatus* encapsulin to adopt a non-covalent chain mail topology, similar to many viral capsids. In the formation of this structural feature the N-terminal arm of one monomer links with the N-terminal arm of its dimer forming concatemeric rings which increase the thermal stability of the structure [21]. Furthermore, the encapsulin from *Q. thermotolerans* also shows this property, where the E-loop

and P-domain of two capsid monomers are arranged head to tail, forming the aforementioned concatemeric rings [19].

Finally, the E-loop represents the region of highest variability between the different solved encapsulin structures [12, 18, 35, 36]. The domain is comprised of a number of β -strands and an unordered region and functions as the place of interaction between dimers. For example, the *T. maritima* encapsulin E-loop has a 60° rotation compared with other structures, which aids in the formation of the smaller T=1 structures [12]. Of note, a similar 60° rotation can be observed in the *M. xanthus* encapsulins when the divergent T=1 shell is formed [28].

1.2.5 Shell assembly

The process of encapsulin shell assembly is poorly understood, in part because of the technical difficulties present with observing dynamic processes at the nanoscale. However, the most prominent hypothesis is that encapsulins self assemble through the repeated addition of dimers. This process is initiated by the interaction of the E-loop regions of two monomers across the 2-fold symmetry axis. Following this initial interaction, repeated pairs of dimers are added to the developing structure until the complete shell is achieved [37]. The proposed self assembly has been evidenced by the successful heterologous expression of encapsulin in a range of non-native hosts including: *E. coli* [12], *Saccharomyces cerevisiae* [38] and the HEK293T mammalian cell line [39] without the need of additional proteins. This marks the encapsulins as unique from potentially related viral capsids, which require additional scaffolding proteins to form their, albeit, larger T-value shells [40, 41, 42].

1.2.6 The pores of the encapsulin shell

Pores are a universal feature of protein based compartments, they are present in the structures of all known bacterial micro- and nano-compartments, from Dps to carboxysomes [43, 44]. Pores perform the vital role of facilitating the exchange of substrates between the encompassing cellular volume and the regulated internal environment through steric and chemical selection [12, 17, 18]. Encapsulins are no exception to this, with pores found at three locations on the shell surface; the sites of symmetry (three- and five-fold pores) and at the interfaces between protomers (two-fold pores) and across studied systems, these pores range

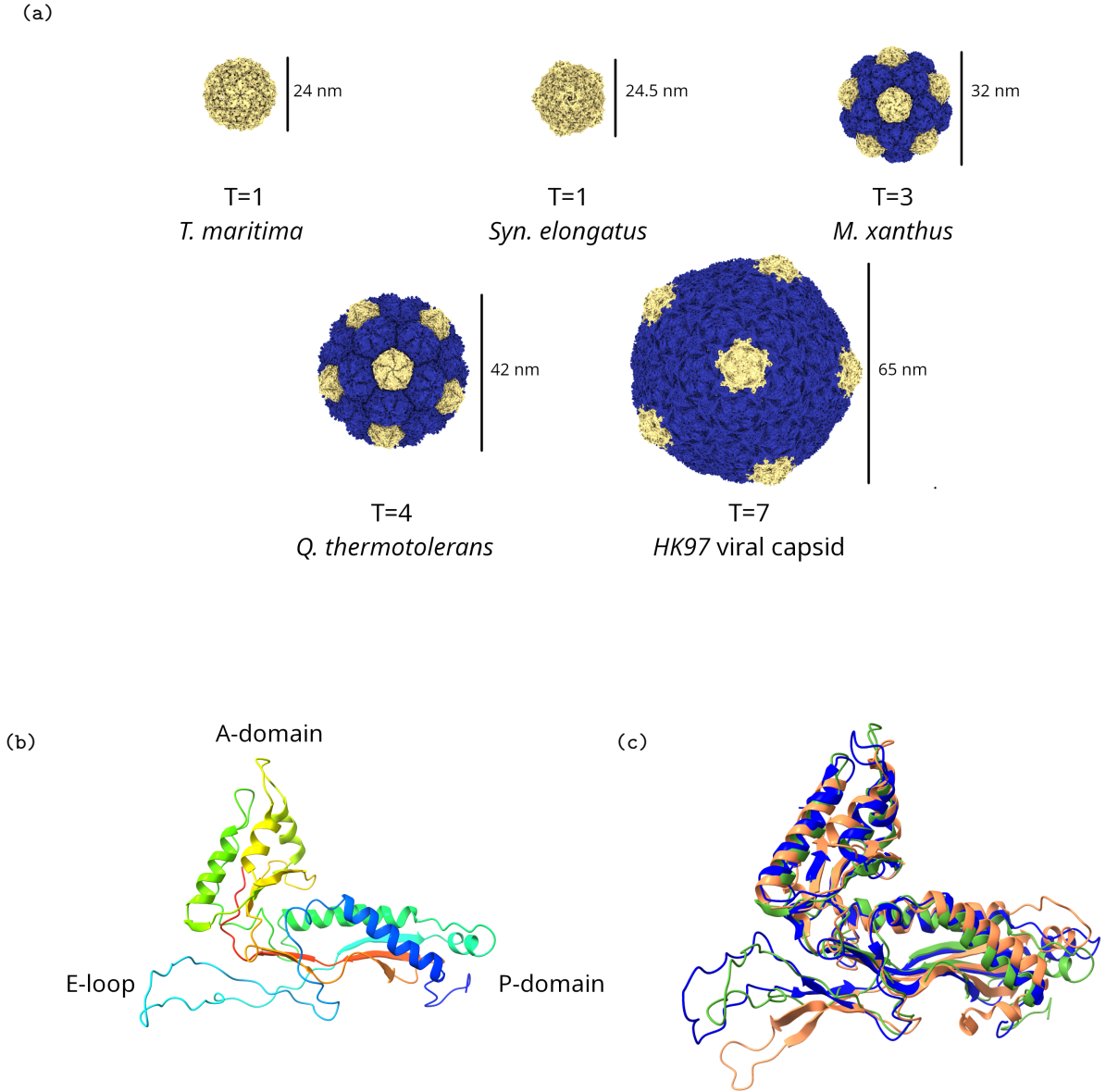


Figure 1.1: Encapsulin structures and their constituting monomeric proteins. (a) Examples of different atomically solved encapsulin shells. The shells are scaled to size, and the HK97 bacteriophage capsid is included as a reference. The complete structures from: *T. thermotolerans* (PDB: 3DKT), *Syn. elongatus* (PDB: 6C9E), *M. xanthus* (PDB: 4PT2), *Q. thermotolerans* (PDB: 6NJ8) and HK97 bacteriophage capsid (PDB: 1OHG) are shown. The pentameric faces are shown in yellow and the hexameric faces are shown in blue. (b) Encapsulin shell monomer from *M. xanthus* the A-domain, P-domain and E-loop are labelled. (c) Overlay of three encapsulin shell proteins highlighting the conservation of the HK97 structure across multiple encapsulin monomers. The shell proteins for *T. maritima* (orange), *M. xanthus* (green) and *Q. thermotolerans* (purple) are shown.

from 3Å to 7Å in diameter. [12, 18, 19, 45, 46].

The atomic structures of encapsulins has revealed the relationship between the structural arrangement of the pore and its electrostatic properties. The *T. maritima* five-fold pores are ringed by five histidine residues. Histidine residues arranged in this manner are known to coordinate metal ions like nickel or iron, a principle used in immobilized metal-affinity chromatography to purify proteins with surface-exposed histidine tags [47]. Therefore, their presence on the surface suggests specificity for the divalent iron substrate. The pores at the three-fold site of symmetry are uncharged, and thus are unlikely to be involved directly in the selective transport of the desired substrate across the protein shell [48]. Examination of the 240 monomer structure from *Q. thermotolerans* reveals that all of the pores are negatively charged - due to the presence of aspartate and glutamate residues - allowing for rapid channelling of its positively ferrous iron substrate [19] In the case of the *Syn. elongatus* and *M. smegmatis* encapsulins the pores are positively charged to facilitate the transport of the substrates, deprotonated L-cysteine and negatively charged aromatic compounds respectively [21, 49, 50].

Access through the encapsulin shell pore is controlled not only through electrostatic effects, but also through the conformational arrangement of the pore itself. The structure of the five-fold pore of the *H. ochraceum* encapsulin has been solved in both an open (15Å) and closed (5Å) state [26]. Suggesting there is dynamic control over the rate of flow of substrate into the core of the encapsulin. Whilst, the two-fold pores from the *Q. thermotolerans* encapsulin are blocked by two flexible arginine residues. It is believed that these arginine residues act as a gate, opening and closing the pore as required similar to pores found in some ferritins [19, 51].

1.2.7 Identification of cargo proteins

The cargo protein of a given encapsulin system is necessary and sufficient for defining the function of that system. Internalized proteins were first identified as areas of increased electron density in the hydrophobic pockets of the inner shell surface of the *T. maritima* crystal structure. An 8-amino-acid region of this density was resolved, identifying the protein as a ferritin-like protein (FLP), encoded upstream of the shell protein in the same operon [12]. *In silico* research led to the identification of a number of different cargo proteins located

upstream of the gene for their respective encapsulin shell protein [35, 52]. These include: FLP [12, 18], DyP (dye decoloring peroxidase) [20, 24, 53], bacterioferritin, 7,8-dihydroneopterin aldolase (FolB) [24], Iron mineralizing activity encapsulin-associated firmicute protein (IMEF) [19], nitrite reductase domain fused to a hydroxylamine oxidoreductase domain (NIR-HAO) [35], and cysteine desulphurase [21, 54].

1.2.8 Cargo loading peptides

The ‘cargo loading peptide’ (CLP) is essential for the inclusion of the cargo protein into the encapsulin shell. In family 1 encapsulins, the CLP is located at the C-terminus of the cargo protein, and consists of predominantly alanine, glycine and proline residues which allow for hydrophobic interactions and salt bridges with the aforementioned hydrophobic pocket on the inner face of the encapsulin shell [12]. The CLP is separated from the catalytic domain by a 10-150 residue flexible linker, primarily composed of glycine and proline. This flexible region likely reduces steric clashes, enabling optimal cargo loading [57]. Furthermore, the fusion of a CLP to a non native cargo results in its encapsulation into fully folded structures, whilst excising it from existing cargo prevents their inclusion [24, 58]. This result, led to the conclusion that the CLP is essential for the encapsulation process.

The simple model of C-terminally tagged cargo protein upstream of the shell in a co-regulated operon is not sufficient for describing the array of different encapsulin models since identified. For example, the encapsulins from *M. xanthus* and *M. tuberculosis* have additional cargo proteins ((EncC, EncD, in the case of *M. xanthus*, and DyP, FolB in *M. tuberculosis*) that are encoded at distal genomic loci from their shell protein. These additional proteins are deemed ‘secondary cargo proteins’ and were identified through their conserved C-terminal CLP regions [18, 24]. Although it has been suggested, in the *T. maritima* system, that the presence of the shell and cargo in the same operon supports a model of co-translational loading the existence of multi-cargo systems challenges this idea [12, 35]. It is currently unclear how cargos coded in distant genomic locations are loaded alongside the upstream cargo, but suggests that there may be a complex system controlling cargo loading than is currently known. The IMEF-Fd encapsulin system found in firmicutes may represent a possible answer; as the cargo proteins are tagged at the C- or N- terminus respectively [35]. There are additional variations that have been observed for Family 1 encapsulins. In the *P. furiosus* system as

Species	T-value (Family)	Cargo	No. of cargo per shell	Experime- ntal method	Reference
<i>T. maritima</i>	1 (1)	FLP	2-5	Cryo-EM	[55]
<i>R. jostii</i>	1 (1)	DyPB	6	UV visible assay	[20]
<i>B. lines</i>	1 (1)	DyP	6 or 12	Native mass spectrome- try	[56]
<i>M. smegma- tis</i>	1 (1)	DyP	6 or 12	Cryo-EM	[49]
<i>M. xanthus</i>	3 (1)	EncB, EncC, EncD	36, 92, 47	Cryo-EM	[18]
<i>M. tubercu- losis</i>	3 (1)	DyP, FolB, BfrB	6, 4 or 8, 24	EM	[24]
<i>P. furiosus</i>	3 (1)	FLP	180	X-ray crys- tallography	[17]
<i>Q. thermo- tolerans</i>	4 (1)	IMEF	84	Cryo-EM	[19]
<i>A. bauman- nii</i>	1 (2A)	CyD	12 or 24	Cryo-EM	[54]

Table 1.1: Table highlighting a selection of known encapsulin systems and their cargos.

the cargo protein and shell are fused into a single polypeptide chain, meaning there is no CLP present [17]. The encapsulin from *Kuenenia stuttgartiensis* also employs a non-CLP dependent cargo loading system. In these encapsulins the shell proteins have an additional dihaem cytochrome domain on their N-termi which is orientated towards the inside of the shell. The cargo is a NIR-HAO fusion protein that adheres to the inner face of the shell through its interaction with the additional dihaem cytochrome [35, 59].

Family 2 encapsulins lack the C-terminal CLP region that is common place amongst Family 1 systems. In its place, Family 2 systems employ long (20-260 residues) intrinsically disordered N-terminal targeting domains (TD) [22, 21, 54]. For the *Syn. elongatus* system, it was identified *in vitro* that the N-terminal 255 residue long domain is necessary and sufficient for cargo encapsulation. Within the TD, 20-30 highly conserved residues were identified, these residues were not contained to a single part of the TD but dispersed throughout it. The entirety of the TD was identified as essential for cargo binding, as the addition of any small part of the conserved residues to non-native protein did not result in encapsulation [22, 21, 57].

1.2.9 Cargo loading stoichiometry

As previously stated, in Family 1 systems the CLP interacts with a hydrophobic pocket on the inner surface of the encapsulin shell. As these binding pockets are found within the structure of each shell monomer - and not spanning several - the maximal cargo load of an encapsulin is equal to the number of monomers that constitute its shell: 60 for T=1, 180 for T=3 and 240 for T=4 [57]. However, the ‘actual’ number of cargos found per shell is far lower than this, as cargo loading is hindered by steric clashes and the need for shell closure [12, 37].

These stoichiometric hindrances underpin the observed preference in the localisation of CLP-shell interactions identified following the solution of atomic structures. Notably, in T=1 encapsulins, the CLP is found interacting with the monomers around the five-fold pore. Supporting the notion that the five-fold pores are the major access points between the lumen and the greater cell volume [12]. Meanwhile, in the larger T=3 systems CLP-shell interactions were observed as evenly distributed across the five-, three-, and two-fold axes with no substantial differences between CLP-shell interactions on pentameric and hexameric facets [28]. Finally, in the largest T=4 shells, CLP-shell interactions were observed at both the two- and

five-fold pores, however, the interactions were found to be tighter at the two-fold pores [19].

1.2.10 Biological function of encapsulin

The function of an encapsulin system is dependent on the biochemical activity of its cargo proteins. In Family 1 systems, a number of different functions have been identified that all centre around redox chemistry; oxidative stress response, iron storage and ammonium-oxidation. Oxidative stress response, in encapsulins is mediated through the activity of DyP cargo proteins. DyP enzymes bind haem and are able to catalyse the breakdown of substrates using hydrogen peroxide [60]. Despite the natural substrate being unknown, *in vitro* experiments have identified that DyP proteins are active against polyphenolic compounds such as azo dyes [53], and that they are able to break down lignin and other peroxidase substrates [61, 20]. A second function commonly observed in Family 1 encapsulins is iron storage. Iron is essential for cellular function, however, excess iron can cause stress through the production of free hydroxyl radicals. Hence it is hypothesised that these encapsulins are an additional iron storage system, alongside ferritins, that are activated during times of stress. For example, it has been shown that FLP-containing encapsulins are upregulated during times of starvation stress in *M. xanthus* [18] and in another work that the FLP loaded system was able to accumulate 9x more iron than apoferritin [23]. Moreover, a unique Family 1 system exists in anammox species: NIR-HAO fusion systems. The exact biological role of these systems is currently unknown, however, a role in detoxifying the harmful intermediates generated during the anammox process has been suggested. These intermediates include; nitric oxide, hydroxylamine and hydrazine [35, 52].

1.2.11 Relationship between the shell and cargo

The encapsulin system comprises two distinct yet interdependent components, both essential for its functional integrity. The encapsulin shell exhibits high structural stability, showcasing resistance against high temperature, extreme pH conditions, and the denaturing effects of guanidine hydrochloride [62]. Additionally, the encapsulin shell shows resistance to a number of proteases, a feature that is often not shared by its cargo [63, 64, 58, 62]. It is thus likely, that the encapsulin shell performs protective functions for its more fragile packaged cargo. Peculiarly, atomic force microscopy revealed that the enclosing of the cargo weakens

the mechanical stability of the encapsulin shell by introducing conformational changes in the shell monomers [37]. This evidence is interesting given the suggested protective relationship of the shell and cargo, and may indicate that the packaging of cargo proteins influences the potential disassembly of the structure.

Encapsulation of enzymes is a common approach for increasing their efficiency. For example, in the carboxysome the encapsulation of Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) increases its efficiency by creating an environment with a higher concentration of CO₂ [65, 10, 66]. A similar, effect has also been observed for encapsulin cargo proteins. *In vitro* studies with the DyP cargo of the *R. jostii* system, revealed that encapsulation increases the peroxidase activity of this protein 8-fold compared to an unencapsulated version [20].

In encapsulin systems where the cargo proteins catalyse the oxidation of ferrous species, the cargo and shell are both required for adequate function. The ferrooxidation domain is located on the cargo protein, and if the reaction proceed unencapsulated the ferric iron species generated is released into the greater cell volume where it may have toxic effects, thus, the shell is required to isolate this ferric species [23, 18, 19]. In *T. maritima* a shell associated flavin moiety has been identified that is bound to a specific pocket of the shell. Flavins are known redox cofactors and its presence may indicate a coupling between iron oxidation and flavin reduction during the catalytic activity of the FLP cargo protein [55, 48, 50]. However, when iron loading assays were performed *in vivo* under aerobic conditions no differences were observed between systems with flavins and ones without. Suggesting the functionality of the flavin may be active under anaerobic conditions [55]. Thus, the shell may perform additional roles in supporting the catalytic function of its cargo.

1.3 Iron storage encapsulins

1.3.1 Biomineralization

Bacterial iron storage proceeds through biomineralization, which is a process by which organic systems control the mineralization of inorganic substrates. The scale of this process is wide spread, ranging from the deposition of calcium in the bones of vertebrates to the storage of iron in small protein complexes [67, 68]. Generally, for biomineralization to take place within a cell several conditions must be met; ions must be super saturated, a size constrained

compartment is required for nucleation to take place and the environmental conditions must promote mineralization.

1.3.2 Iron and oxidative stress

Iron is the most abundant element on Earth and the 4th most abundant element in the Earth's crust [69]. Due to its electronic configuration, it exists in one of two stable states either as soluble ferrous (Fe^{2+}) or insoluble ferric (Fe^{3+}) ions. As iron regularly forms complexes with a number of different organic molecules its presence is vital for the functioning a wide variety of different biological processes. Such as its role as a co-factor in a number of electron transfer reactions [70].

Before the appearance of molecular oxygen around 2.1 and 2.4 billion years ago, the environment was highly reducing, meaning that ions with high reducing potentials such as Fe^{2+} were commonly found and incorporated into cellular life [71]. However, after the appearance of molecular oxygen, there was a large shift in the prevailing iron chemistry causing a decrease in the bio-availability of ferrous iron as it was oxidised into insoluble ferric complexes and precipitated [72, 73, 74]. Despite the decrease in availability, ferrous iron remained important in many biological reactions due to its redox states, large redox potential window (-600 to +500 mV), and its ability to adopt different spin states depending on the ligand environment [74, 75]. Importantly, if the cellular ferrous iron content is not carefully maintained its uncontrolled oxidation into ferric iron could lead to the formation of toxic reactive oxygen species through Fenton chemistry [76, 77]. This reaction proceeds via the interaction of Fe^{2+} and H_2O_2 :



Both H_2O_2 and OH^- are not reactive enough to oxidise cellular components. However, the generated hydroxyl radical ($\text{OH}\cdot$), which is short lived but highly reactive, freely oxidizes a number of basic cellular components; nucleic acids, proteins, phospholipids and carbohydrates. If left unmanaged, this process can cause high levels of damage and lead to cell death [78, 79, 73]. Thus, cells must maintain their iron homeostasis to prevent this.

Intracellular iron content is monitored by metalloregulatory proteins such as the ferric

uptake regulator (Fur) protein. The Fur protein can act as a transcriptional activator or repressor to increase or decrease the production of proteins involved in iron storage, sensing and efflux [80, 81]. For example, under iron limited conditions, de-repression of Fur-related operons results in increased expression of siderophores and upregulation of Fe^{3+} uptake pathways [82]. Whilst, in the presence of excess Fe^{2+} Fur induces the expression of efflux systems to decrease the intracellular Fe^{2+} concentration [83]. Iron storage is the other major method employed by cells to maintain iron homeostasis. When the concentration of intracellular ferrous iron exceeds what is required storage of the excess is facilitated by the proteins of the ferritin-like super family [84]. The major members of this family are: bacterial ferritins, bacterioferritins, DNA-binding proteins from starved cells, encapsulin ferritin-like proteins and IMEF [85, 86, 87, 88, 45].

1.3.3 Sequence identity of the ferritin-like proteins of encapsulins

To date, a number of different encapsulin systems have been identified to have ferroxidation performing cargo, with the systems from *T. maritima* and *P. furiosus* being the first identified [17, 12]. Since, several additional systems have been identified that all perform the same ‘iron mineralization’ function with similar ferroxidation performing cargos. This includes but is not limited to the systems from: *M. xanthus*, *Q. thermotolerans*, *H. ochraceum* and *R. rubrum* [18, 19, 23, 87]. Despite these systems possessing the same highly conserved iron-binding motif (ExxH) and the same tertiary structures, there is limited sequence similarity between the cargo proteins from the different systems. For example, the *R. rubrum* ferritin-like protein has a 53% sequence identity with *T. maritima* FLP, a 29% identity with the *P. furiosus* FLP and finally a 33% identity with the EncC protein from *M. xanthus* [23].

1.3.4 Structural identity of ferritin like proteins

The determination of atomic structures for a number of different FLPs has revealed a conserved tertiary and quaternary organisation across a range of different species [12, 18, 23, 87]. The tertiary structure consists of two longer antiparallel α -helices and a shorter α -helix. Ten of these monomers are arranged to give a ‘decameric ring’ quaternary structure, where each conserved C-terminal CLP region points outward from the centre of the ring, allowing for it to associate freely with the shell for efficient cargo loading (See Figure 1.2).

The decameric ring is formed through the association of five pairs of dimers, which are arranged such that there are two types of dimer interfaces. The first dimer interface contains the di-iron ferroxidase centre (FOC) and thus is the interface along which the ferroxidation reaction occurs. This interface is maintained through the co-ordination of iron ions by two glutamates and a histidine residue that are found on each monomer [23]. The second dimeric face, is the back-to-back interface between the FOC containing interfaces. These dimer interfaces are solvent facing and thus mediated by a mixture of hydrophobic interactions, hydrogen bonds and salt bridges, as support by the solved crystal structures [87, 89]. It is believed that the non-FOC containing dimer interfaces are the first to form, and then the formation of the catalytically active interface proceeds only when Fe^{2+} is present to give the complete decameric structure [23, 87, 89].

1.3.5 Proposed mechanism for iron storage by encapsulin systems

The process of iron mineralization is a complex multi-step process. The inherent complexity combined with the inability of current experimental methods - Cryo-EM and X-ray Crystallography - to provide high resolution data for dynamic processes severely limits our understanding of how this process occurs. In recent years, and as more FLP systems have been studied, models have been suggested to explain the mechanisms underpinning the process, and have highlighted the importance of the role of the shell in the function of the cargo.

Generally, free Fe^{2+} ions have to pass through the pores in the encapsulin shell and be translocated to the docked cargo protein. Of note, no protein has been identified experimentally or bioinformatically that could possess the function of delivering Fe^{2+} ions to encapsulin pores, thus it is assumed that this process occurs passively within the cell. Upon reaching the cargo protein, Fe^{2+} is oxidised into Fe^{3+} at the FOC centres using up O_2 in the reaction. The Fe^{3+} then travels to the surface of the cargo protein where mineralization occurs. The mineralized product is then released into the lumen of the encapsulin for storage [23, 28] (See Figure 1.3). Interestingly, it has been observed that the encapsulin shell is essential for the efficiency of this process. TEM observations of isolated FLP cargo proteins incubated *in vitro* with Fe^{2+} found large heterogeneous deposits of iron on the grid. The heterogeneous deposits did not share the highly regular morphology observed when the FLP cargo was ex-

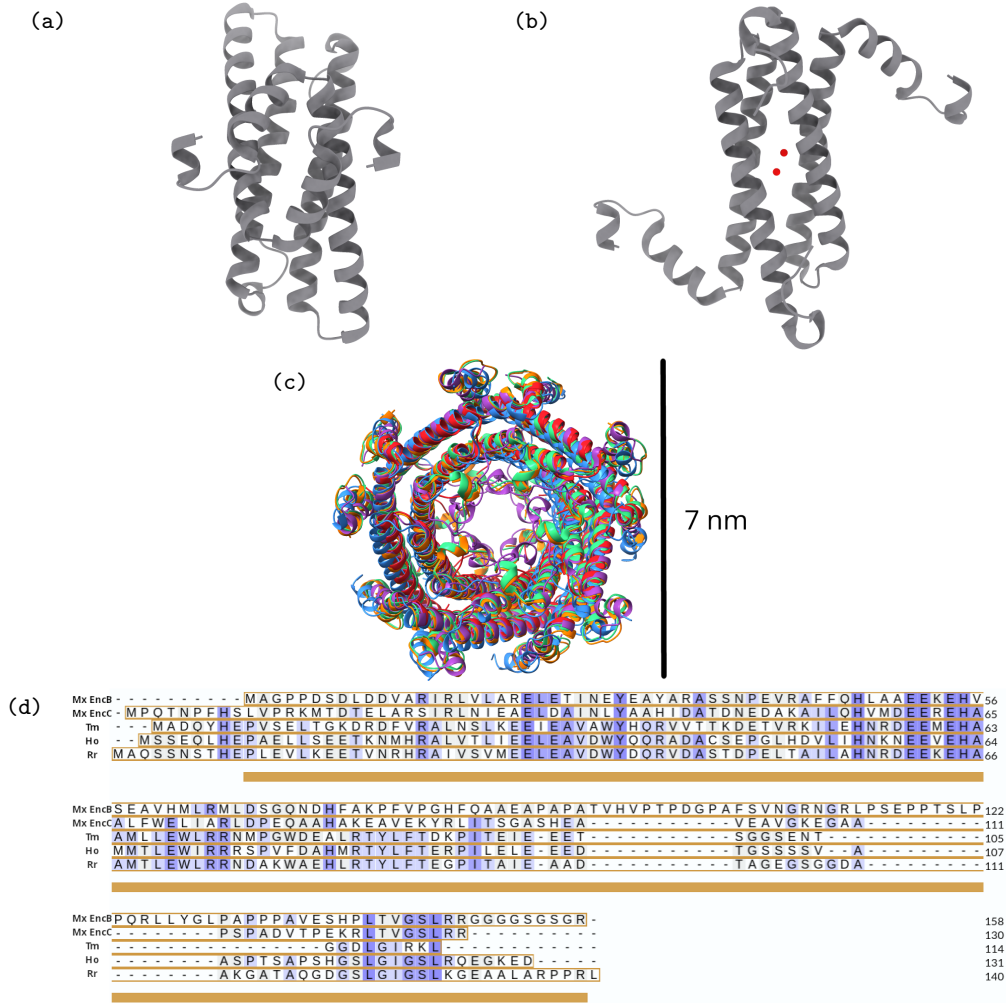


Figure 1.2: Structure and organisation of encapsulated ferritin-like proteins. A) Non-FOC-containing dimer and B) FOC-containing dimer from the FLP of *R. rubrum*. Iron atoms are shown in red. C) FLP decamers from a number of different species superimposed to highlight their structural similarity. The structures shown are *M. xanthus* EncB in red (PDB:7S5K), *R. rubrum* in green (PDB:5DA5), *P. furiosus* in purple (PDB:5N5E), *H. ochraceum* in orange (PDB:5N5F) and *M. xanthus* EncC in blue (PDB:7S8T). D) Multiple sequence alignment for encapsulated ferritin-like proteins using Clustal Omega [90]. Top to bottom: EncB from *M. xanthus*, EncC from *M. xanthus*, *T. maritima*, *H. Ochraceum*, *R. rubrum* [91].

pressed alongside its shell protein and incubated with Fe^{2+} in the same manner, thus it was concluded that the shell is essential for proper iron nanoparticle mineralization [23].

The FOC-containing cargo proteins contain a dimer interface that co-ordinates 2 Fe^{2+} ions in a 5-coordinate complex through 2 conserved glutamates and a histidine that are found on both of the monomers, similar to the coordination found in traditional ferritins [23, 28]. Additionally, there is a conserved tyrosine in the FOC region. In other ferritin systems tyrosine aids the conversion of O_2 to H_2O by forming a tyrosyl radical as an intermediate in the reaction [92, 93]. Additional to the iron co-coordinating sites in the FOC, metal-binding sites have been observed on the surface of the decameric ring that are consistent with the uptake of the Fe^{2+} before the FOC and the release of Fe^{3+} after ferroxidation in the FOC [23, 87, 28, 89].

In *M. xanthus*, the arrangement of iron-binding sites in EncC was both the same as described above and as the FLP proteins from *P. furiosus* and *R. rubrum* [23, 87, 28]. However, a novel arrangement of binding sites was observed in the atomic structure of the EncB protein. In EncB, one of the conserved glutamate residues is rotated away from the iron-binding site by 180° . Whilst the tyrosine residue is shifted closer to the iron-binding site, together these changes indicate the iron is not bound by a high affinity 5-coordinating arrangement but instead a lower affinity 4-coordinate state. Additionally, a glutamic acid-histidine ladder was observed along the peripheral region of the FOC-containing α -helices. A single iron ion was consistently found coordinated by 4 glutamates and 2 histidines at the centre of this ladder and thus it has been proposed that this region functions as homologue to the ‘site C’ found in other bacterial ferritins [28, 93, 88, 94, 95, 96]. The identified ladder region is negatively charged and thus is able to attract Fe^{2+} ions, therefore it is likely that this regions functions as a gateway for ferrous ions. Futhermore, the proteins were crystallized in the presence of Ca^{2+} , and were later observed bound to a site close to the FOC. It was assumed that this may indicate the Fe^{3+} transit site for the exit of the molecule from the FOC. As the Ca-binding site was observed between the FOC and the central channel of the decamer this would support a model where the Fe^{3+} leaves the FLP through the central channel.

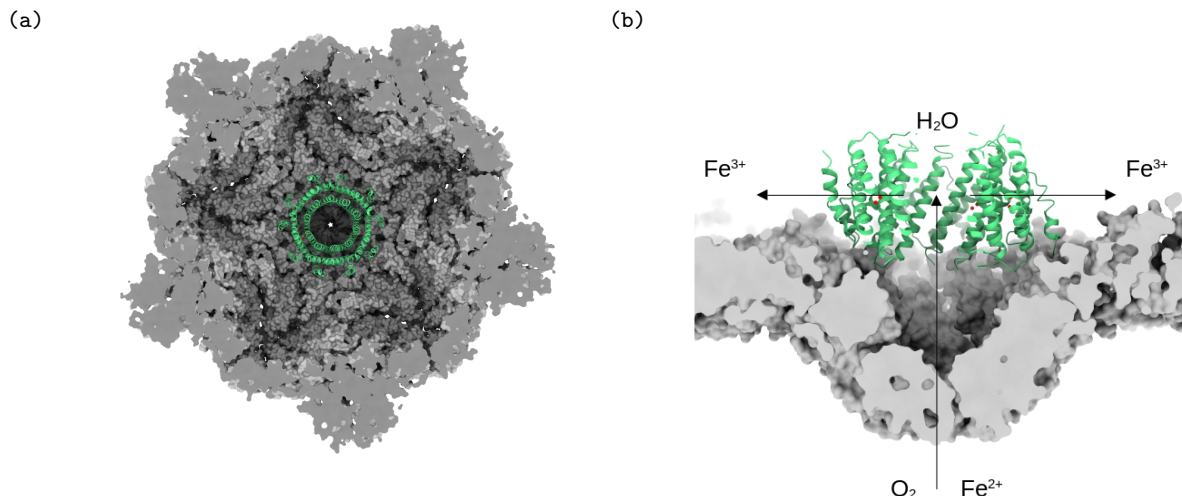


Figure 1.3: General model for iron oxidation by encapsulin FLP cargo proteins. (a) Truncated EncC (PDB: 7S8T) atomic model superimposed over model of *M. xanthus* encapsulin shell (PDB: 4PT2). The cargo is modelled over the 5-fold pore; the suggested entry point for the ferrous substrate. (b) Schematic representation of ferrooxidation pathway from encapsulin FLP systems. Fe²⁺ is transported through the shell pore to the FOC containing α helical bundles of the FLP cargo. Here it is oxidised using O₂ and the resulting ferric substances leaves the cargo protein to be mineralized on its surface before being released into the lumen [23].

1.3.6 The encapsulin system from *M. xanthus*

M. xanthus encodes a multi-cargo encapsulin system consisting of: EncA, EncB, EncC and EncD. EncA is the shell protein, and forms a 32 nm wide 180 monomer T=3 encapsulin shell. EncB and EncC are ferritin-like proteins with highly conserved iron-binding motifs (ExxH). EncD lacks this conserved iron-binding site and instead functions as a ferric reductase [97]. The three cargo proteins share a conserved C-terminal region (LTVGSLRR), which is used to load the proteins into the shell (Figure 1.4c). EncB and EncC are both decameric structures formed of repeating pairs of anti-parallel α -helical dimers. The EncB decamer is slightly smaller than the EncC decamer, being 7.0 nm in diameter compared to 8.5 nm for EncC. Both structures have the same thickness of 5.0 nm [28]. Although the N- and C- termini are not confidently resolved in the crystal structure, it is proposed that the N-termini are angled towards the centre of the complex and the C-termini are organised around the circumference; which is in agreement with AlphaFold predicted structural models (Figure 1.4) [98, 99]. The C-terminus being arranged in this manner is important, as it allows free interaction between the CLP and the binding pocket on the inner lumen of the encapsulin shell. The charge distribution on the EncD surface is even, however there are localized negative areas

concentrated around the dimerization interface, which in EncC is mostly negativity charged [28].

It is likely that the encapsulin is involved in the stress response of the bacterium. Firstly, it has been shown that deletion of the *encA* (MXAN_3556) gene from *M. xanthus* results in tan-coloured colonies that are unable to transform into yellow colonies. This phase variant strain was also deficient in the production of DKxanthene and myxovirescin and showed deficiencies in aggregation, swarming and sporulation [100]. As the production of DKxanthene and the environmental iron levels are linked, it was suggested that the deletion of the *encA* gene resulted in a mutant that was unable to sense its environmental iron levels, thus forcing it to become tan phase locked. This variant phenotype has only been observed following the deletion of the shell protein and not the cargo proteins [101, 102]. Additionally, it has been observed that the expression of encapsulin is increased upon starvation and when $\Delta encA$ mutants are exposed to H_2O_2 they show higher sensitivity compared to the wild type. Suggesting that the encapsulin system likely plays a role protecting the cell from oxidative stress through the removal of free Fe^{2+} from the greater cell volume during starvation or antibiotic induced stress [18].

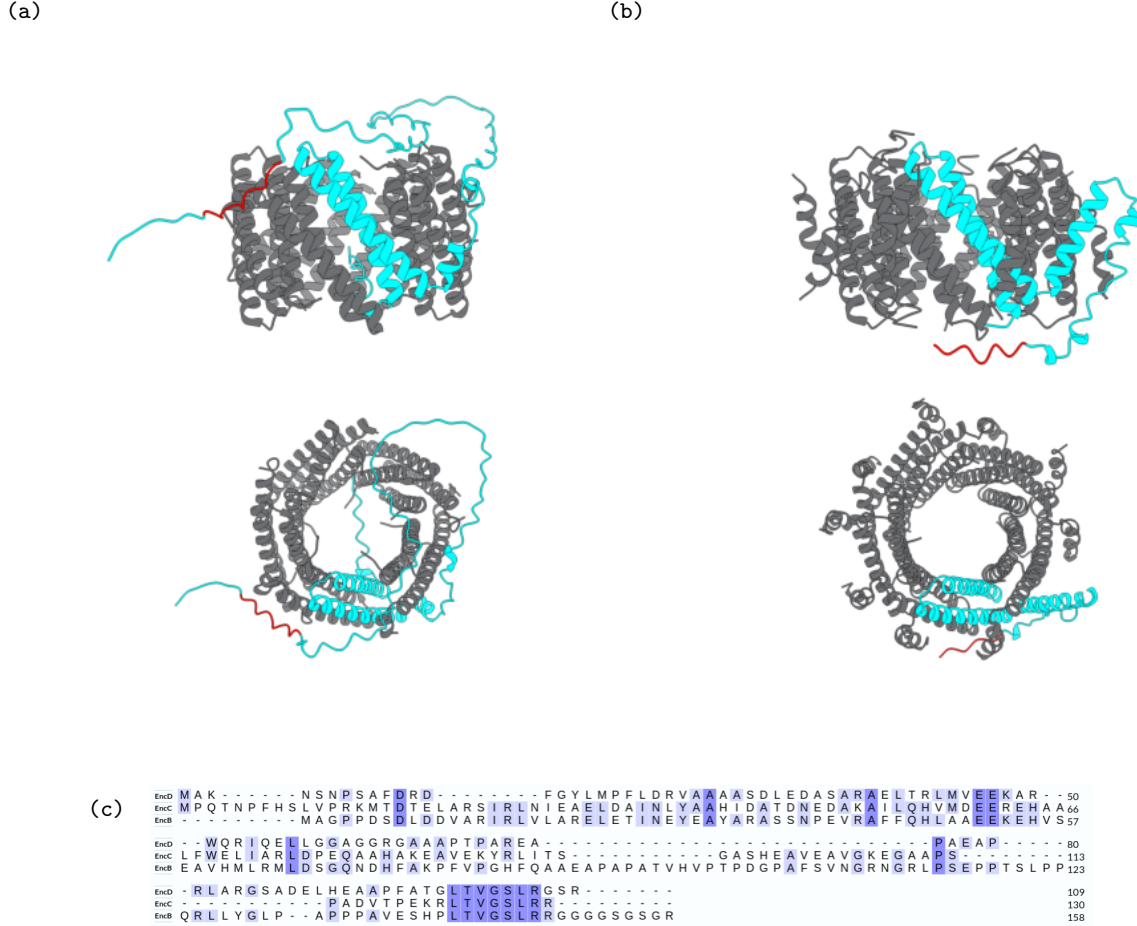


Figure 1.4: Structure and sequence of *M. xanthus* encapsulin cargo proteins. A) Partial atomic structure of EncB (PDB:7S5C) in grey. Overlaid with the structural prediction (AlphaFold) of the full length structure (light blue) the CLP region is highlighted (red) B) Partial atomic structure of EncB (PDB:7S5C) in grey. Overlaid with the structural prediction (AlphaFold) of the full length structure (light blue) the CLP region is highlighted (red). C) Sequence alignment of the three cargo proteins: EncB, EncC and EncD. Highly conserved regions, including the ExxH iron-binding motif and the CLP are highlighted. [98, 99]

1.4 Electron microscopy

1.4.1 Interactions of electrons with matter

In electron microscopy, an electron is used as the contrast generating source of illumination. Electrons are a form of ionizing radiation, which means they possess sufficient energy to displace inner-shell electrons from atoms. The ionization process can cause a variety of different secondary signals to be emitted which can be used for materials characterization (see Figure 1.5) [103].

During the illumination of a sample with electrons a number of electrons are transmitted without interacting with an atom, these form the direct beam. Others collide directly with an atom and are scattered through angles greater than 90° , causing them to leave the sample in the opposite direction to the incident beam; these are called backscattered electrons. The majority of electrons pass near to atomic nuclei and suffer some sideways deflection called elastic scattering, without significant loss of kinetic energy. If the electrons are deflected through small angles they can be used to form a Bright Field (BF) image. If the deflection is larger, the electron can be used for dark-field TEM using Bragg spots, or for even larger angles, for Annular Dark Field (ADF) STEM imaging using Rutherford-type and thermal-diffuse (phonon) scattering. Inelastic scattering by bound electrons will lead to considerable loss of kinetic energy, forming the basis of Electron Energy Loss Spectroscopy (EELS). The deceleration of incident electrons by charges in the material can lead to the emission of X-ray radiation called bremsstrahlung which forms a continuous spectrum of X-ray energies unlike the discrete values observed for characteristic X-rays; this contributes only to the background of Energy Dispersive X-ray (EDX) spectra [104]. Auger electrons [105] and characteristic X-rays are alternative ways to emit energy when an ionised electron returns to its ground state. Auger electrons are used to study the properties of surfaces [106], whilst characteristic X-rays provide information about the chemical composition of the sample that come from deeper in the sample [107]. Finally, cathodoluminescence is light emitted from the recombination of electron-hole pairs created by the illumination of semiconductor materials with the electron beam [108]. .

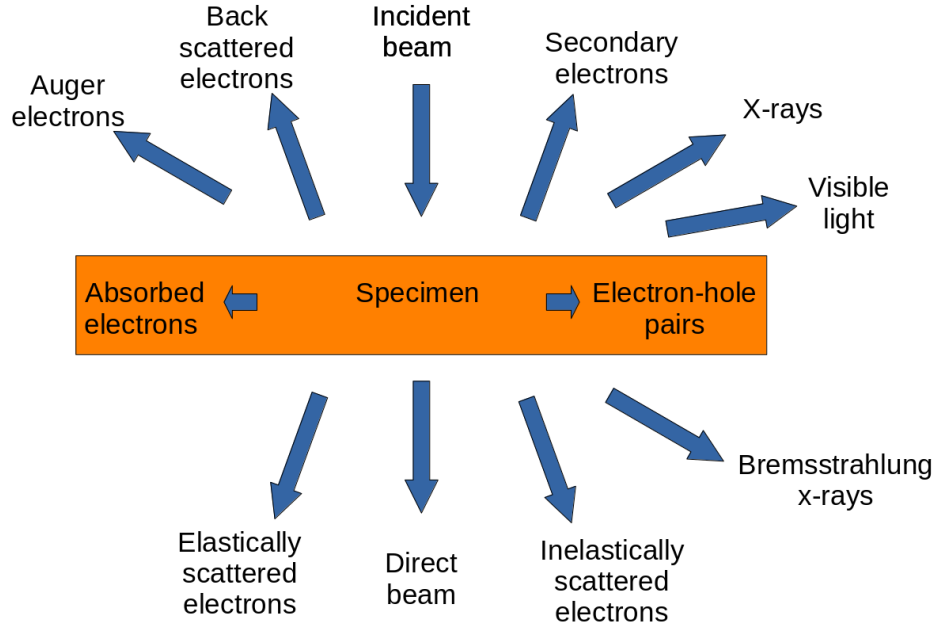


Figure 1.5: The different secondary signals emitted from the interaction of electrons with matter.

1.4.2 Elastic and inelastic scattering

The interaction of an electron with the specimen can be classified as either elastic or inelastic. Elastic scattering occurs when the incident electron interacts with the atomic nuclei which is a large charged mass compared with the electron. If the electron passes close enough it is deflected and scattered through large angles (0° to 180°). However, the majority of electrons pass through the sample further away from the atomic nuclei and therefore are scattered by less than a couple of degrees. If the material is amorphous, each scattering event happens independently but in the case of a crystalline material the interference between scattered electron waves gives rise to characteristic patterns which are referred to as diffraction. Inelastic scattering is defined as the interaction of incident electrons with the bound electrons. It is these inelastic interactions that give rise to the secondary emission effects discussed previously. Phonon scattering is another type of quasi-elastic scattering that can be observed. Secondary electrons are produced from the thermal vibrations of the bulk state of the material. These scattering events typically have very low energy loss associated with them ($<0.1\text{eV}$) and they cannot be recorded on conventional electron detection equipment [109].

1.4.3 The scanning transmission electron microscope

The first Scanning Transmission Electron Microscope (STEM) was developed by von Ardenne in the late 1930s, several years after the development of the Transmission Electron Microscope (TEM) [110, 111, 112]. The electron optics in the STEM differ slightly from those in the TEM because the electron beam is converged into a probe and used to illuminate only a small region of the sample rather than illuminating the whole specimen region with a parallel beam. Images are formed by rastering the probe over the sample and collecting both the electrons that are transmitted straight through the sample and those that are scattered through higher angles. This differs from the previously developed SEM techniques where backscattered and secondary electrons were used in reflection geometry to probe the surface properties of bulk materials.

Initially, a high voltage (typically between 1-4kV) is applied to the electron gun causing emission of electrons. Electron emission is achieved by either thermionic emission or through field emission from the tip, or a combination of both in a Schottky field emission gun (FEG). In thermionic emission a crystal with a low work function is heated to a temperature where electrons are emitted and in a field emission gun an electric field is applied to the tip causing it to emit electrons due to tunneling effects. Schottky emission is a combination of both of these, thus it can be considered thermally assisted field emission [113]. The electrons then pass through an aperture hole and are further accelerated to typically 80-300kV. The electron optics consists of a series of electromagnetic lenses that demagnify the electron source to create the probe that passes through the sample. The beam-forming lens arranged between the accelerator and the specimen in the microscope is known as the condenser lens. Also arranged before the specimen are a series of scan-coils required to raster the probe across the specimen. Finally, there is an objective lens directly before the specimen which performs the final demagnification step. Spherical aberration in this lens is dominant in limiting the resolution obtainable by the electron optics system, and thus a condenser aperture is inserted to limit the contribution of spherical aberration to the probe. However, the use of an aperture has two consequences: its use results in a diffraction limit to the probe size and a loss in signal intensity from the sample due to a decrease in beam current, as electrons that do not pass through the aperture do not contribute to the final image [103, 114].

As previously identified, the interactions of the electrons emitted from the electron

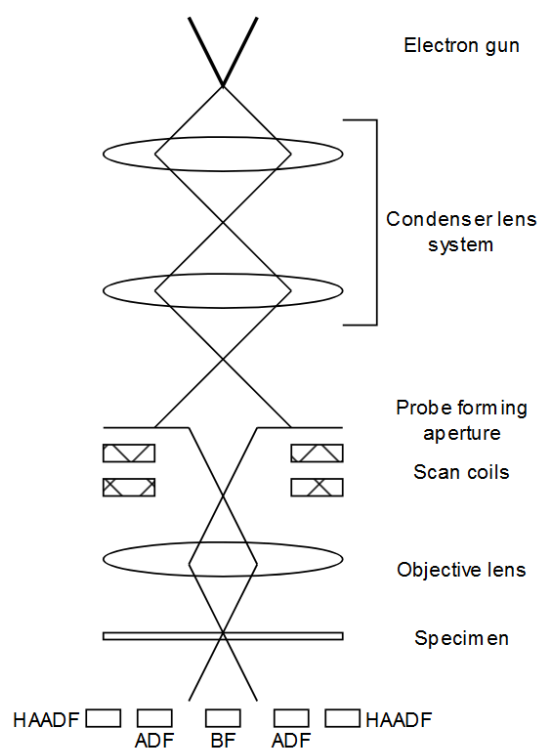


Figure 1.6: Schematic diagram of a typical Scanning Transmission Electron Microscope. The electron gun, condenser lens, scan coils, objective lens, probe forming apertures and detectors are labelled.

gun with the specimen can produce a variety of signals, the majority of which can be detected simultaneously using a STEM instrument. However, the most commonly collected and used signals are: electrons that do not deviate far from the optical axis and form BF images. Electrons that leave the specimen at high angles relative to the optical axis and form either ADF or High Angle Annular Dark Field (HAADF) images. X-rays emitted through the ionization of the sample by the electron beam are used in Energy Dispersive X-ray Spectroscopy (EDXS). Finally, electrons that lose energy due to inelastic interactions with the atoms in the specimen can be used in EELS. For optimum signal detection the sample should generally be thinner than 100 nm [114, 103].

1.4.4 Spherical aberration

STEM imaging is fundamentally different from conventional parallel beam TEM. In a STEM the electron probe is formed by the electromagnetic lenses focusing a parallel electron beam to a point and as with all electromagnetic lens it is subject to spherical aberrations. Spherical aberration in an electron lens causes electrons near the optical axis and electrons far from the optical axis to be focused differently [115, 114]. In a STEM this results in an unfocused and larger probe degrading the possible resolution. To tackle this problem, the spherical aberration corrector was developed [116, 117]. A spherical aberration corrector uses a series of multi-pole lenses to adjust the electron beam path such that all electrons are focused to the same point. The benefits of aberration correction in STEM instruments are drastic; firstly it facilitates the formation of a smaller probe, and as the maximum possible resolution is tied to the size of the probe it drastically improves the spatial resolution. The use of this technology allows atomic resolution to be achieved for many crystalline materials and has resulted in the characterization of a wide variety of materials; biologically this includes the characterization of the ferritin mineral core [118] and the 3D reconstructions of elements of the mammalian cell cytoskeleton [119]. Secondly, reductions to spherical aberration mean that larger condenser apertures can be inserted into the path of the electron beam to form the probe. A larger aperture gives a larger convergence angle, meaning that the electron probe will be much more intense, giving better signal-to-noise-ratios (SNR) in spectroscopic techniques. Over time the capacities of this technology have continuously improved and the highest resolution achieved to date is now sub=0.4Å, where a single atom defect was recorded

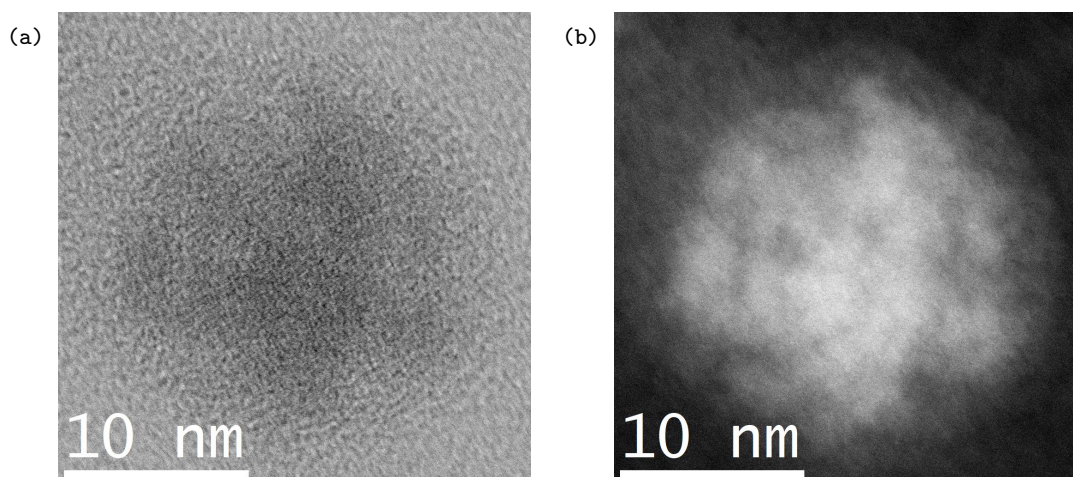


Figure 1.7: Comparative BF and HAADF dark field images of an encapsulin isolated from *M. xanthus*. (a) BF (b) HAADF. Both images were captured at 8 Mx on a spherical aberration corrected JEOL ARM200F operating at 80keV. The scale bar represents 10 nm. The different contrast mechanisms are clearly visible. In the HAADF image the Fe rich core has a higher atomic number (Z value) and thus appears brighter than the predominantly C (lower Z value) protein shell. In the BF image, due to the general low mass and density of the sample the contrast appears poor across the image, despite the high spatial resolution. This is especially evident for the C containing protein shell that is too light to produce enough contrast to be visualised with this technique.

in a sample of molybdenum disulphide [120, 121].

1.4.5 Bright field imaging

In BF STEM imaging the detector is placed on the optical axis and collects the signal from the electrons scattered through low angles and from the transmitted beam. As a result of this, there is always some signal picked up by the detector, which is what gives the technique its name. The contrast in BF is roughly proportional to the atomic number of the atoms being imaged [122]. Additionally, if there is more specimen present the BF intensity gets darker and in regions with minimal specimen the intensity appears brighter (as can be seen in Figure 1.7a). This is called mass-thickness contrast. Mass-thickness contrast is dominant when the sample material is amorphous, such as would be the case for proteins however if the material is crystalline diffraction contrast must also be taken into consideration.

1.4.6 Annular dark field and high angle annular dark field imaging

Annular Dark Field (ADF) and High Angle Annular Dark Field (HAADF) imaging are so called because of the geometry of the detectors used to capture them [123, 124]. The detector itself is a large annular detector that is placed around from the optical axis of the microscope. The only signal that reaches this detector is from interactions between the electrons and the atomic nuclei that result in the electrons being scattered through large angles. For ADF imaging this is electrons that are scattered to $<\sim 3^\circ$ and for HAADF imaging it is electrons scattered to angles greater than this up to the maximum acceptance angle of the instrument (typically 10-20°) [125]. Thus, the background in the image remains dark until there is strongly scattering specimen in the beam path, giving the second part of the nomenclature. As the scattering power of an atomic nucleus is related to its atomic number, Z , the larger the atomic number the greater the intensity of the scattering and the higher the contrast in both ADF and HAADF imaging. The relationship between the atomic number and the contrast can be described using the following equation:

$$C \propto NZ^{1.6-2} \quad (1.2)$$

Where C is the contrast and Z is the atomic number of the atoms in the specimen and N is their number. The exponent of the Z value is dependent on the inner collection angle of the detector used and atomic screening between the electron and atomic nuclei by the atomic electrons [126, 127]. The relationship between the atomic number and the contrast has resulted in this type of contrast being known as Z-contrast.

High resolution ADF and HAADF techniques are much more easily interpretable than their conventional TEM counterparts. This is because of the difference in the Optical Transfer Function (OTF) in HAADF imaging and the CTF for BF imaging (see Figure 1.8). Briefly, in BF imaging the CTF is due to the interference of electrons scattered through low angles. Thus, the CTF is highly sensitive to changes in defocus specimen thickness, and aberrations will cause oscillations and contrast inversions at different spatial frequencies. Contrastingly, in ADF imaging techniques images are formed by highly scattered electrons and thus are less sensitive to the effects of defocus and aberrations. Resulting much easier to interpret images, making HAADF much more preferable for high resolution imaging of heavy metal atoms in

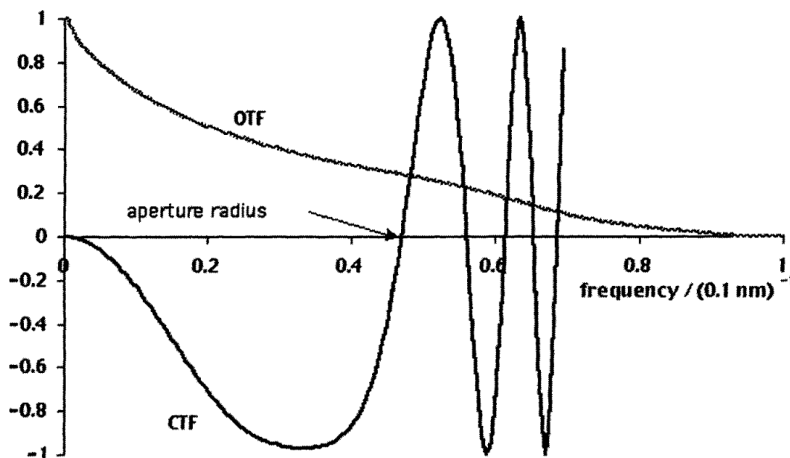


Figure 1.8: The optical transfer function (OTF) for HAADF imaging compared with the phase contrast transfer function (CTF) for BF imaging for the same defocus and spherical aberration. This figure has been reproduced from [130].

thin support films [114, 103].

A major limitation of ADF imaging techniques is the low sensitivity, because of their smaller scattering cross section, for light elements, such as those found in biological specimens. Despite the inability to provide high spatial resolution of biological specimens, HAADF has been successfully used to correctly measure the mass of proteins present in a number of different instances [128, 129]. Finally, as this imaging technique deals with scattered electrons the transmitted electrons that carry chemical information about the sample can be collected by an EELS spectrometer and additional EDXS systems can collect X-rays emitted simultaneously. Thus, ADF techniques are useful for combining with spectroscopy techniques to give images containing a combination of chemical and spectral information.

1.4.7 Energy dispersive X-ray spectroscopy

As previously stated X-rays are a type of secondary emission caused by the interaction of the electron beam with the sample. When an atom is ionised, and an electron transitioned away from the nucleus, there is a period where the hole in the electron shell can be filled by an electron from a higher energy shell. When this process occurs there is a chance that it results in the emission of a characteristic X-ray. The fluorescence yield of X-rays is heavily dependent on the atomic number of the atom; for atoms where $Z > 32$ the yield exceeds 50%, however for elements where $Z < 11$ the yield drops to $< 2\%$ [131, 132]. However, X-rays are not the

only type of emission that can occur, as sometimes an Auger electron is emitted instead. The probability of Auger electron and X-ray emissions for an ionised atom always totals 100%. Thus, as the X-ray emission increases with atomic number, Auger electron emission decreases. If a K-shell electron was to be removed via ionization and the hole filled by an L electron it would result in the emission of a $K\alpha$ X-ray, whilst if the hole was filled by a M electron it would cause the emission of a $K\beta$ X-ray. However, not all conceivable transitions between the existing electronic energy levels are allowed, those that are allowed are defined by quantum-mechanical selection rules [107, 133]. Due to this relationship, emitted X-rays can only possess a discrete set of energies. Thus, in an EDXS spectrum, there are characteristic peaks that represent the chemical elements present in the specimen. These peaks are super-imposed on a background of low intensity which is formed from stray radiation from the column and from bremsstrahlung X-ray production (see Figure 1.9). Although theoretically all elements can be identified based on their EDX signal, the physical parameters of the detector can hamper detection. Traditional EDX detectors employed thick beryllium windows to protect them from the hydrocarbon contamination in the microscope. These thick windows were strongly absorbing and prevented the analysis of any elements with an energy of 1000 eV or less ($Z=11$) [125]. However, the subsequent development of both ultra-thin windows and atmospheric pressure polymer windows allowed for the measurements of elements down to boron. Finally, the development of the windowless Silicon Drift Detectors (SDDs) allowed for measurements of elements down to lithium [133, 125]. The combination of this technique with a STEM instrument provides a method for generating Spectrum Images (SI) where at each probe position within the scanned x and y directions there exists a third axis, the energy axis. The energy axis contains spectra representing the elements present at each probe position [134]. Using the generated SI it is possible to resolve spatial information about the distribution of elements within the sample.

1.4.8 Electron energy loss spectroscopy

Electron Energy Loss Spectroscopy (EELS) was first developed in the 1940s and has evolved into a technique that is commonly used for quantifying the elements present and their local environment in a material down to atomic resolution [135, 136, 137]. EELS works by measuring the change in kinetic energy experienced by a forward scattering electron after it

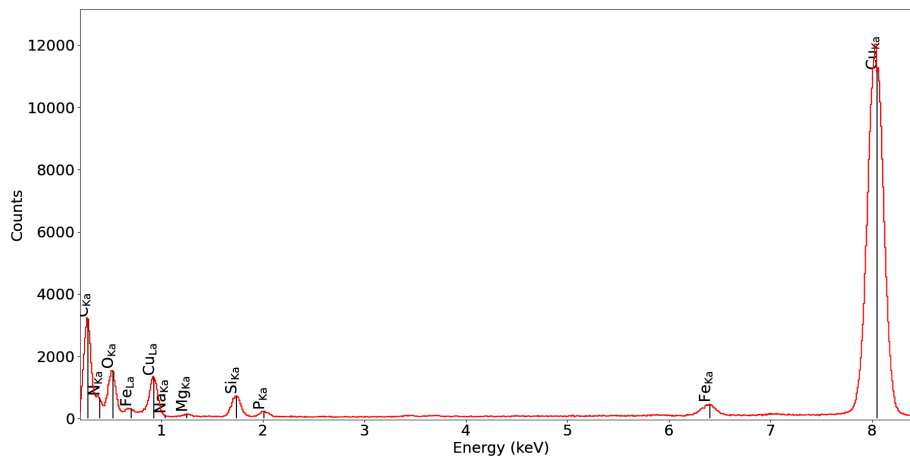


Figure 1.9: Example EDXS spectra taken from a sample of encapsulins isolated from *M.xathus*. The peaks representing C K α , N K α , O K α , Fe L α , Cu L α , Na K α , Mg K α , Si K α , P K α , Fe K α and Cu K α are labelled.

has interacted inelastically with an atom in the specimen material therefore it is a measure of the primary excitation of the material. Thus, a core loss edge represents the kinetic energy lost by the incident electron as it excites an electron inelastically from an inner atomic shell [125]. The core loss edge onset is therefore the lowest amount of energy required for this ionization to take place. The ionization can still be caused by incident electrons with higher kinetic energies than this minimum value, which causes the ‘edge’ shape of the higher energy core loss region. As unlike the clear peaks seen in an EDXS spectrum, the edges in an EELS spectrum have an initial onset followed by an exponential decay until they return to the background.

As the EELS spectrometer detects electrons that are scattered through low angles, the detector is placed at the bottom of the microscope column after the specimen. This requires that the BF detector be removed from the beam path, meaning that BF techniques cannot be used simultaneously with EELS acquisition. However, EELS does not require the electrons that are scattered through high angles, therefore the ADF and HAADF detectors can be used alongside it, providing both spatial and elemental information about the specimen region being imaged. The most commonly used spectrometer set-up is a post-column spectrometer such as the Gatan Imaging Filter (GIF) [103]. This spectrometer uses a magnetic prism to sort the entering electrons by their kinetic energies, and an aperture can be inserted between the specimen and the spectrometer entrance to limit the electrons and improve energy resolution.”

The energy loss spectrum is complex containing a number of different features that can provide information about the material being imaged. The first and most intense region is the Zero Loss Peak (ZLP), which is centred on zero loss and contains transmitted electrons which have not undergone any significant energy loss. This includes electrons that are elastically scattered and those that are inelastically scattered with energy loss lower than the detection limit of the detector. This region also contains the effects of phonon scattering due to the thermal vibrations of the material. Phonon scattering has an energy loss that is $<0.1\text{eV}$ and thus cannot be resolved with a standard electron spectrometer system [138, 139]. Following the ZLP there is the Low Loss (LL) region, whose major feature is normally the bulk plasmon peak. The bulk plasmon peak normally resides in the region of $1\text{-}30\text{eV}$ and is the result of the polarisation of the material as the electron is transmitted [103]. It is not limited to a single peak, as if the sampled region is thicker multiple plasmon peaks can occur. At higher energy losses, the core loss region of the EEL spectrum can be found. This region is characterized by a smooth decaying background caused by the rapid decrease in electron density as energy loss increases. Superimposed on the background are core loss edges that represent inner shell excitations. As the inner shell binding energies are dependent on the atomic number, the edges represent the elements that are present in the material. In thicker materials the shape of the edge is convoluted by multiple plasmon scattering events; the incident electron is scattered multiple times as it passes through the material. This effect can be removed from the core loss edge by deconvolving it with the low loss region to give a single scattering distribution [140].

At high energy resolutions, features of core loss edges can be resolved that provide additional chemical information about the atoms. These are known as the Energy-Loss Near Edge Structure (ELNES) and the EXtended Electron Energy-Loss Fine Structure (EXELFS). The ELNES covers a region of 50 eV following the onset of the core loss edge and contains information about the chemical bonding environment of the atoms. For example, modulations in the ELNES of transition metals can be used to identify the valence state of the atom and has led to standard-less identification of the valence state of the material [141, 142]. The EXELFS region is present at energies higher than the onset energy in the spectrum and provides information on the scattering of ejected electrons. This region can provide information on the inter atomic distances and on the coordination numbers of the atoms present [114].

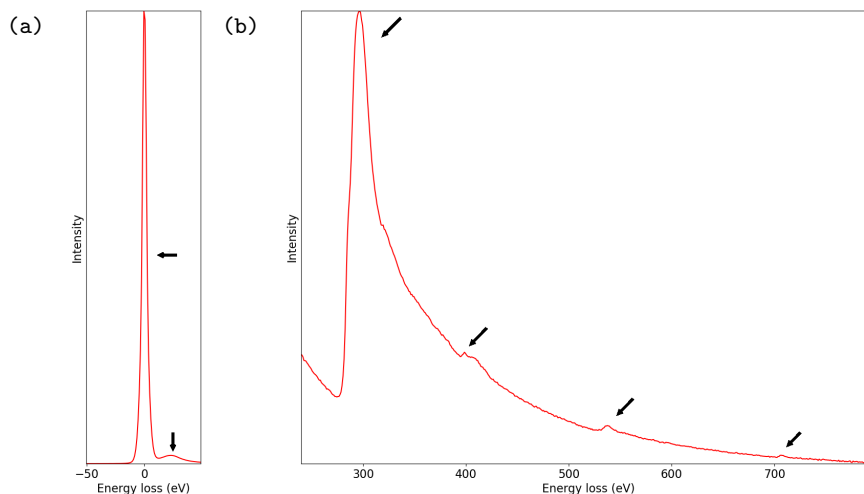


Figure 1.10: Example low loss and high loss electron energy loss spectra taken of encapsulins from the bacteria *M. xanthus*. (a) Low loss spectra. Two spectral features have been highlighted with arrows, the zero loss peak (top) and the plasmon peak (bottom). (b) High loss spectra. Several core loss edges have been labelled with arrows. From left to right: C K-edge, N K-edge, O K-edge and Fe L-edge. The core loss edges are seen as features on the continuously decaying background.

1.4.9 Comparison of EDXS and EELS

As discussed, both EXDS and EELS are useful spectroscopic methods for identifying and quantifying the elements present in a sample. However, there are some key differences between them that alter their usefulness under different experimental conditions. Firstly, the collection efficiency of EELS is much higher than that of EDXS. This is because the EELS spectrometer is located on the optical axis and collects the forward low angle scattered electrons so it has a collection efficiency of between 20% and 50%. Because of the isotropic nature of X-ray production, and the limited total collection angle of the EDXS detectors, that need to be located sideways above or below the specimen, only a small fraction (typically 1-10%) of the X-rays are detected. Furthermore, because the K-line fluorescence is directly linked to the atomic number of the nuclei EDXS is better suited for imaging heavier elements, while conversely EELS has a higher sensitivity to lighter elements [133, 125]. Secondly, the spatial resolution in EELS is higher than in EDX, this is because of the spectrometer entrance aperture that can be used to limit the angular range of electrons that are detected. In EDX there is no limiting aperture and therefore electrons scattered through large angles in thicker specimens also contribute to the signal. The two techniques also differ in how easily they can be quantified. In EELS, measured edge intensities can be converted to atomic ratios using

cross-sections and known experimental parameters (the incident beam energy, the convergence angle and the collection angle). The cross-sections used have been identified as accurate within typically 5% for K-edges and within 15% for L-edges [143, 125]. The quantification in EDXS is dependent on the physical parameters of the spectrometer that was used and on the thickness of the sample, where thicker samples can re-absorb generated X-rays. Under conditions where there is no absorption or secondary fluorescence the relative ratios of elements can be calculated using the k -factor. However, the k -factors cannot be accurately calculated and must be measured experimentally using samples of known composition under the desired imaging conditions in the same instrument. Without doing this the accuracy cannot exceed 15% [125]. Sample thickness plays a huge role in the usefulness of the techniques. In EELS increased sample thickness results in plural scattering. Although various deconvolution methods exist to remove the effect of this plural scattering the sample thickness is limited to typically 100 nm if using a STEM with a 200 kV acceleration voltage after which plural scattering increases the background to the edges with energy loss lower than 1000 eV to the point where the edges become blurred. EDXS is less sensitive to sample thickness, but it is still limited to samples with a thickness of typically no more than a few hundred nm. However, quantification of light elements becomes very difficult when sample thickness exceeds 100 nm [125].

1.5 Aims and Objectives

The initial aim of this research project was based on reports in the literature describing uniform spherical, 5 nm mono dispersed nanoparticles within the encapsulin shells of *M. xanthus* [18]. The proposed regular morphology suggested an ordered atomic structure, which could yield materials with desirable characteristics - particularly magnetic properties. Such characteristics would be significant given the existing challenges in manufacturing nanoparticles with these properties [144, 145].

However, preliminary low-resolution characterization experiments (see Sections 3.2.6 and 3.2.14) revealed that the encapsulated nanoparticles inside encapsulins isolated from *M. xanthus* did not display the desired crystallinity or magnetic properties. As a result, the research focus shifted toward a more detailed characterization of the iron-containing mineral stored within the encapsulin nanocompartments—an area not well investigated in the literature prior to this study. With the aim of answering the question: what is the iron containing

mineral found inside the encapsulin?

To achieve this, a combination of electron microscopy techniques was employed. TEM, HAADF-STEM and HR-STEM were used to investigate nanoparticle size, morphology, and atomic structure. EDXS and EELS were used to assess elemental composition, relative elemental ratios, number of iron atoms per encapsulin, and the valence state of the iron containing mineral.

Additionally, the individual contribution of the three cargo proteins - EncB, EncC and EncD - in the formation of the iron containing nanoparticles had not been well studied. This study aims to identify and characterize the minerals generated by two of these cargo proteins - EncB and EncC - to further classify their functional roles in the system. Both proteins have previously been identified through bio informatic and structural analysis as ferritin-like, suggesting they may contribute to iron oxidation and mineralization within the encapsulin compartment [18, 28]. EncD is not investigated in this study, as its function was unknown at the outset of the project. However, recent work has established that EncD functions as a ferric reductase [97]. This question would be answered by investigating encapsulin constructs containing either single or double cargo proteins in an *E. coli* over expression system subsequently loaded with iron. The resulting materials would then be investigated using the electron microscopy techniques described above. As an addition, the role of phosphorous in the formation of material inside the encapsulin would also be explored, by the generation of material inside encapsulins in a phosphorous free environment.

The first results chapter details the experimental work carried out to obtain encapsulin samples for electron microscopy investigation in subsequent chapters. This chapter describes the isolation of native encapsulins from *M. xanthus*, as well as the expression and purification of encapsulins with varying cargo content within an *E. coli* expression system. The work involved generating a series of petDUET-1 plasmids containing a C-terminally 6xHis tagged encapsulin shell protein (EncA) and one of the cargo proteins in the second multiple cloning site. Additionally, a pCOLADuet-1 plasmid containing EncC was generated and successfully co-expressed with the petDUET-1 plasmid in an *E. coli* expression system, resulting in the production of encapsulin constructs with controllable cargo protein composition.

The second results chapter describes a series of experiments using various electron microscopy imaging techniques to investigate the nanoparticles inside encapsulins isolated

from *M. xanthus* and the encapsulin constructs described in Chapter 1. First, methods for loading encapsulins generated in *E. coli* with iron are discussed, including both *in vivo* and *in vitro* techniques. The thickness of the encapsulin shell, the size, and the distribution of the nanoparticles inside the encapsulins are explored using BF, HAADF, and SE imaging. Finally, a combination of Selected Area Electron Diffraction (SAED) and HR-STEM is used to examine the crystallinity of the nanoparticles inside the native encapsulin shell, constructs expressed in *E. coli* and inside constructs loaded with iron *in vitro* with no phosphorous present.

In the third results chapter, the encapsulins generated in Chapter 1 are analyzed using low resolution EDXS. The data presented identifies elemental signals from the encapsulins, including trace elements and maps them to the core of the encapsulin. A quantification of the chemical composition of the material within the encapsulin is also presented. These experiments were also instrumental in defining the experimental parameters (specifically the EELS detector offset and dispersion) used for the higher spatial resolution EELS experiments discussed in Chapter 4.

Chapter 4, the final results chapter presented in this thesis, investigates encapsulins using EELS. This chapter provides quantification of the chemical composition of the encapsulin protein shell and the nanoparticles stored within. The high spatial resolution achievable in EELS experiments enables discussion of the chemical composition of individual nanoparticles within the same shell. Additionally, experiments centred on the analysis of the Fe L_{2,3} Energy Loss Near Edge Structure (ELNES) allow for characterization of the valence state of the iron that is found inside the encapsulin.

Chapter 2

Isolation and generations of encapsulin protein complexes

2.1 Introduction

In this chapter, the vector design, protein expression and purification of encapsulins containing single cargo proteins - EncA+EncB, EncA+EncC, EncA+EncD - will be presented. Alongside the generation of a more complex system containing a combination of cargo proteins: EncA+EncB+EncC. To achieve this, the EncA protein was C-terminal tagged with a 6xHis tag and cloned into the petDUET-1 plasmid alongside one of the cargo proteins. The expression of multi-cargo complexes was achieved by cloning EncC onto the pCOLADuet plasmid and expressing the proteins from both plasmids simultaneously. Furthermore, The effect of the cargo protein on the proper formation of the encapsulin shell in *E. coli* will also be presented. All protocols used for the generation of these constructs followed standard molecular biological protocols and the successful formation of the encapsulin shells loaded with the desired cargo proteins was confirmed using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Transmission Electron Microscopy (TEM) of negatively stained samples. The reduced cargo encapsulins were generated so that the iron mineralizing function of the different cargo proteins could be compared in future analysis.

The isolation of the native encapsulin from *M. xanthus* will also be presented in this chapter. As well an experiment into its long term structural stability.

CHAPTER 2. ISOLATION AND GENERATIONS OF ENCAPSULIN PROTEIN COMPLEXES

Name	Description	Genotype	Source
<i>M. xanthus</i> DK1622	Wild type strain	WT	[146]
<i>E. coli</i> DH5 α	For molecular cloning	<i>fhuA2</i> Δ (<i>argF-lacZ</i>) <i>U169</i> <i>phoA</i> <i>glnV44</i> Φ 80 Δ (<i>lacZ</i>) <i>M15</i> <i>gyrA96</i> <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>thi-1</i> <i>hsdR17</i>	ThermoFisher
Rosetta(DE3)	For heterogeneous protein expression	<i>fhuA2</i> [<i>lon</i>] <i>ompT</i> <i>gal</i> (λ <i>DE3</i>) [<i>dcm</i>] Δ <i>hsdS</i> λ <i>DE3</i> = λ <i>sBamHI</i> <i>o</i> Δ <i>EcoRI-B</i> <i>int::</i> (<i>lacI::PlacUV5::T7</i> <i>gene1</i>) <i>i21</i> Δ <i>nin5</i>	Novagen

Table 2.1: Bacterial strains used in this study.

2.2 Experimental procedures

2.2.1 Strains and culturing conditions

Bacterial strains used are referenced in Table 2.1. *M. xanthus* cultures were grown in CTT (1% Tryptone, 10 mM Tris-HCl pH 8.0, 8 mM MgSO₄ 1 mM KH₂PO₄) liquid medium on a shaking platform at 200 Revolutions Per Minute (RPM), or on CTT 1.5% agar (Sigma) plates at 32 °C. *E. coli* strains were grown in Lysogeny Broth (LB) (Sigma; 10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract) liquid medium on a shaking platform at 200 RPM, or on LB 1.5% agar plates, supplemented with appropriate antibiotics (Ampicillin 10 μ g/ml, Chloramphenicol 25 μ g/ml, Kanamycin 50 μ g/ml) at 37 °C. For protein over expression conditions, filter-sterilised IPTG (0.01 mM, 0.05 mM or 0.1 mM) was added to the liquid growth media.

2.2.2 Isolation of encapsulin from *M. xanthus*

The method used here was modified from [18]. A *M. xanthus* DK1622 starter culture (1 x 400 ml) was grown shaking at 250 RPM and 32 °C in CTT media (1% tryptone, 10 mM Tris-HCl pH 8.0, 8 mM MgSO₄ 1 mM KH₂PO₄) until an optical density at 600 nm (OD₆₀₀) of 0.4 was reached. The starter culture was then used to induce 8 x 800 ml Fernbach flasks. The flasks were returned to the incubator shaking 250 RPM at 32 °C for 24 hours. The cells were harvested by centrifugation (10 mins, 10,000 RPM, SLA3000 rotor) and re-suspended in 3 litres of TPM media (10 mM tris-HCl pH 8.0, 8 mM MgSO₄ 1 mM KH₂PO₄) before being returned to the Fernbach flasks and starved overnight for 18 hours at 250 RPM and 32 °C. Starved cells were harvested via centrifugation (10 mins, 10,000 RPM, SLA3000 rotor). Cell pellets were re-suspended in ddH₂O and cooled on ice for 1 hour. Cells were broken by ultrasonification using a Soniprep 150 plus ultrasonicator for 10 mins on ice (alternating on and off for 30 seconds at 60% amplitude), the cell lysate was clarified by spinning for 10 mins at 10,000 RPM using the SLA3000 rotor. 15% (NH₄)SO₄ was slowly added to the cell lysate and agitated using a magnetic stirring rod to. The suspension was centrifuged for 10 mins at 10,000 RPM using the SLA3000 rotor and the pellet was discarded. 2% PolyEthylene Glycol (PEG) 8k was added and slowly dissolved using a magnetic stirrer, then the solution was centrifuged for 15 mins at 20,000 RPM using a JA25.50 rotor. The supernatant was then discarded and the remaining orange pellet dissolved in ddH₂O before 0.5% N-dodecyl- β -D-maltoside was added. 0.5% N-dodecyl- β -D-maltoside was added as it is a mild detergent, and therefore increases membrane solubility resulting in a cleaner final product. 0.3 g/ml CsCl were added to the solution and it was pipetted on top of a 5.7 M CsCl cushion in an ultracentrifuge tube. The gradient was then centrifuged for 18 hours at 38,000 RPM in a SW41 rotor. The encapsulins were found enriched in a brown pellet at the bottom of the ultracentrifuge tube, the CsCl gradient was discarded and the pellet re-suspended in ddH₂O. The sample was dialysed against ddH₂O for 24 hours to remove the CsCl before being layered atop a 10-40% sucrose gradient and centrifuged for 2.5 hours at 20,000 RPM using a SW41 rotor. The pellet was discarded and the remaining supernatant spun at 55,000 RPM for 90 mins. Pelleted encapsulins were then re-suspended in ddH₂O and stored at 4°C. Before examination in a STEM instrument, the purified encapsulin sample was dialysed against ddH₂O at a concentration of 1:1,000,000 to remove any remaining sucrose from the sample, as this is a source of additional carbon that

could cause carbon contamination to build up during data collection.

2.2.3 PCR and plasmid construction

The primers used in this study are listed in Table 2.2. The primers were designed based on the genes stored in the Ensembl Bacteria database [147] and ordered from Sigma-Aldrich. Genes encoding for *encB* (MXAN_3557) and *encD* (MXAN_2410) were codon optimized for expression in *E. coli* and synthesised by TwistBioscience. The two genes were synthesized and codon optimized because attempted expression using the genes from the *M. xanthus* genome resulted in minimal to no expression; most likely because of the large GC rich regions present in both of these genes, and it is known that genes with highly enriched GC regions are known to have a poor protein yield when expressed in *E. coli* [148].

Polymerase Chain reaction (PCR) were carried out using either Q5 Hot Start High Fidelity Master Mix (New England Biolabs) or DreamTaq DNA Polymerase (ThermoFisher) using the protocol described by the manufacturer. For the PCR, an initial 5 minute denaturation step at 95 °C, followed by 32 cycles of a 1 minute denaturation step at 95 °C, a 30 second annealing step at 55-65 °C depending on primer combination used, a 1 minute elongation step at 72 °C (all genes were <1 kb in length so 1 minute for sufficient for elongation) after 32 cycles there was a final 7 minute elongation step at 72 °C. PCR reactions were supplemented with up to 2% DMSO where appropriate, as many of the genes being amplified had high % GC content. DMSO is a known and commonly used additive for improving the yield of PCR product for genes with high % GC content [149]. PCR products were run on a 1% agarose gel containing CYBER SAFE after visualisation using a UV box the bands containing the desired amplification product were cut from the gel and purified with a GeneJet PCR Purification Kit (ThermoFisher) using modified instructions from the manufacturer. The modification made was using 21 μ l instead of 50 μ l of elution buffer in the final elution step to increase the final DNA concentration.

Following purification, a double restriction digest was undertaken on both the amplified gene and the desired plasmid in the restriction buffer optimal for both restriction enzymes and the reaction was carried out for 1 hour at 37 °C. All restriction enzymes were purchased from New England Biolabs. After 1 hour, 0.5 μ l of Shrimp Alkaline Phosphatase (New England Biolabs) was added to the restriction digest reaction containing the plasmid, and incubated

for a further 15 minutes at 37 °C. Shrimp Alkaline Phosphatase removes the 5' phosphate group from the end of the digested DNA preventing the recircularization of the plasmid. The digestion reaction mixtures were then heat inactivated by heating to 65 °C for 20 mins before the DNA product was purified using a GeneJet PCR Purification Kit (ThermoFisher).

The ligation reactions were carried out using T4 DNA ligase purchased from New England Biolabs. A reaction mixture was generated based on the manufacturers recommendations: 10 μ l T4 DNA Ligase Buffer (New England Biolabs), 50 ng of vector, required ng of insert to give 1:3 ratio of vector to insert, 1 μ l T4 DNA ligase and finally ddH₂O to give total reaction volume of 20 μ l. After one hour at room temperature the reaction mixture was used in a transformation reaction with *E. coli* DH5 α (see Section 2.2.4).

Successful ligation was confirmed using PCR. A PCR reaction mixture was generated using 10 ng of purified plasmid and primers that bound to the plasmid directly. Briefly, if the gene was ligated into the Multiple Cloning Site (MCS) 1 of the petDUET-1 (Novagen) plasmid the standard 'pET Upstream' and 'DuetDOWN1' primers were used and if the gene was ligated into the MCS2 the standard 'DuetUP2' and the 'T7 Terminator' primers were used. For ligations into the pCOLADuet-1 plasmid, the standard 'ACYCDuet1 Primer' and 'DuetDOWN1' were used. Successful ligation of the gene into the MCS was confirmed after running the PCR product on a 1% agarose gel containing CYBER SAFE and comparing the resulting band size to a negative control. A complete list of the plasmids generated as part of this study can be found in Table 2.3.

2.2.4 Transformation of *E.coli*

For the transformation of *E. coli*, 10 ng of purified vector or whole ligation product were used to transform chemically competent Rosetta(DE3) or DH5 α strains of *E. coli*, respectively. DH5 α is an *E. coli* strain optimized for the replication of recombinant DNA, and Rosetta(DE3) is a strain optimized for the expression of proteins. Importantly, the Rosetta(DE3) strain contains plasmids that encode tRNAs for codons rarely used in *E. coli*. Thus, the Rosetta(DE3) strain was chosen for overexpression as it should lead to higher protein yields if used for the expression of genes from *M. xanthus* as they contain codons that are not normally used in *E. coli*. 50 μ l of chemically competent cells were mixed with the appropriate construct and heat shocked at 43 °C for 30 seconds. Cells were then placed on ice

CHAPTER 2. ISOLATION AND GENERATIONS OF ENCAPSULIN
PROTEIN COMPLEXES

Name	Sequence	Source
MXAN_3556 F NdeI	ATACATATGCCTCTGG AGCCACATTTTCATGC	This study
MXAN_3556 R XhoI 6xHis	TATCTCGAGCTAGTGG TGATGGTGAATGATGAG ATCCTCCTCCTCCAGA TCCTCCTCCTCCGCGA CGTTCCGTGGCGCC	This study
MXAN_3557 F NcoI	ATACCATGGCCGGCCC TCCCGACAGTGAAT	This study
MXAN_3557 R HindIII	TATAAGCTTTCAGCGA CCGCTGCCGGA	This study
MXAN_4464 F NcoI	ATACCATGGGAACGGA CACCGAGCTGGCC	This study
MXAN_4464 R HindIII	TATAAGCTTCTACCGG CGCAGGCTGC	This study
MXAN_2410 F NcoI	ATACCATGGCGAAGAA CTCGAATCCCC	This study
MXAN_2410 R HindIII	TATAAGCTTTCAGCGG CTGCCCCCTGAG	This study
pET Upstream	ATGCGTCCGGCGTAGA	Novagen
DuetDOWN1	AACATGTGCCGGCGTA TTAG	Novagen
DuetUP2	TTGTACACGGCCGCAT AATC	Novagen
T7 Terminator	GGCGACTCGTTATTGA TCG	Novagen
ACYCDuet1	CGATCTCGACGCTCTC CCT	Novagen

Table 2.2: List of the primers sequences used in this study.

CHAPTER 2. ISOLATION AND GENERATIONS OF ENCAPSULIN PROTEIN COMPLEXES

Name	Parent	Resistance Cassette	Description	Source
petDuet-1	-	Amp	For the expres- sion of two pro- teins in <i>E. coli</i> using the T7 promoters.	Novagen
pCOLADuet-1	-	Kan	For the expres- sion of two pro- teins in <i>E. coli</i> using the T7 promoters.	Novagen
petDuet- EncA6xHis	petDuet-1	Amp	For the ex- pression of EncA with a C-terminal 6xHis tag.	This study
petDuet- EncA+EncB	petDuet-1	Amp	For the ex- pression of EncA-6xHis and EncB.	This study
petDuet- EncA+EncC	petDuet-1	Amp	For the ex- pression of EncA-6xHis and EncC.	This study
petDuet- EncA+EncD	petDuet-1	Amp	For the ex- pression of EncA-6xHis and EncD.	This study
pCOLADuet- EncC	pCOLADuet-1	Kan	For the expres- sion of EncC.	This study

Table 2.3: List of plasmids generated as part of this study. The parent plasmid strains are also included in the table.

for 4:30 mins before being recovered for 1 hour at 37 °C in 1 ml liquid LB media. After the recovery step, cells were spread onto LB plates supplemented with the appropriate antibiotics.

Following the successful transformation of DH5 α with a ligation mixture, colonies were checked to see if they contained plasmids with an inserted gene. Single colonies were picked from the transformation plate and grown for 18-24 hours in 10 ml of LB broth supplemented with the appropriate antibiotics. The plasmids were then isolated using a GeneJET Plasmid Miniprep Kit (Thermofisher) with a modification to the final step of the manufacturer's instructions: using 21 μ l instead of 50 μ l of elution buffer in the final elution step to increase the final DNA concentration. The presence of the desired gene was checked using PCR (see Section 2.2.3). Following, the successful transformation of Rosetta(DE3), single colonies were picked from the transformation plate and used for protein expression.

2.2.5 Protein expression

For heterogeneous protein expression, Rosetta(DE3) transformed with the desired plasmids were used. A starter culture containing the appropriate antibiotics was prepared and grown overnight for 16 - 20 hours at 37 °C. The starter culture was then used to inoculate a larger culture at a dilution of 1:100. Cells were grown again at 37 °C to an OD₆₀₀ of 0.4-0.6 before being induced with IPTG. Initially, a concentration of 1 mM IPTG was used, however, this resulted in the formation of insoluble inclusion bodies visible as large white pellets following cell lysis. The IPTG concentration was lowered iteratively until no inclusion bodies were observed. Similarly, the temperature and length of induction were iteratively decreased from 37 °C and 18 hours until the optimal overexpression conditions was identified. This condition was found to be a maximum of 0.1 mM of the inducer IPTG, and 18 hours at 24 °C shaking at 200 RPM. This condition was then used for overexpression of encapsulins in *E. coli* regardless of the cargo proteins being expressed alongside the shell. After 18 hours, cells were chilled on ice and harvested by centrifugation at 5000 x *g* for 10 min at 4 °C.

2.2.6 Purification of proteins from *E. coli*

Cells were thawed on ice for 20 mins then re-suspended in lysis buffer (50 mM Tris, 500 mM NaCl at pH 8.0). Cells were then broken by ultrasonication using a Soniprep 150 plus ultrasonicator in rounds of 10 seconds on and 20 seconds off for a total of 5 mins at an

amplitude of 30% on ice. The sonicated mixture was then clarified by centrifuging at 8000 x *g* for 10 mins. In this step the remaining large cell fragments are pellets and were discarded. The 6xHis tag added to the C-terminus of the EncA protein was used for purification via His-tag:Ni-NTA affinity chromatography.

A His-tag:Ni-NTA affinity chromatography column (Qiagen) was equilibrated with equilibration buffer (50 mM Tris, 500 mM NaCl, pH 8.0, 10 mM imidazole), before the clarified lysate was run through the column. The column was then washed three times in washing buffer (50 mM Tris, 500 mM NaCl, pH 8.0, 20 mM imidazole) to ensure all of the remaining cellular debris was washed out of the column. The washing buffer contained 20 mM imidazole to remove any proteins that had bound non-specifically to the column. The protein of interest was eluted using an elution buffer (50 mM Tris, 500 mM NaCl, pH 8.0, 250 mM imidazole) which contained higher concentration of imidazole (250 mM) and out-competed the His-tagged protein for binding to the Ni resulting in the protein being eluted from the column. The elution step was carried out three times to ensure all of the bound protein was eluted from the column. Presence of protein in the final elution fraction was checked using a Bradford Protein assay (Biorad) and measuring the absorbance at 595 nm.

2.2.7 Polyacrylamide gel electrophoresis

Gel electrophoresis was undertaken in denaturing conditions. Protein sample was mixed with equal volume of 1x SDS sample buffer and heated in a 100 °C waterbath for 10 mins. Samples were then loaded onto a 4-20% Tris-Glycine gel (Novex) set into a XCell Surelock Electrophoresis cell (Invitrogen). All gels were run with the See Blue Prestained (Invitrogen) protein marker, to allow for estimation of protein size. The gel was run at 100 V until the sample had migrated successfully into the gel, voltage was then increased to 200 V and the sample run until the dye front reached the end of the gel.

For visualization, gels were first removed from the gel tank and after opening the plastic cassette gels were incubated with Quick Coomassie Stain (NeoBioTech) overnight, before being de-stained with ddH₂O several times. Gels were then visualized with the G:BOX Chemi-G9 system (Syngene) using the Syngene acquisition software (Syngene).

2.2.8 SDS gel quantification

Purified protein samples were loaded into each well of a 4-20% Tris-Glycine gel (Novex) and run until the dye front reached the end of the gel. Gels were stained with Quick Coomassie Stain (NeoBioTech) for 18 hours before being de stained with several washes of ddH₂O. Gels were then visualized with the G:BOX Chemi-G9 system (Syngene) using the Syngen acquisition software (Syngene). Following visualization, background intensity was removed using a rolling ball method. This is a commonly used background subtraction technique that works by averaging the intensity of pixels in a defined neighbourhood [150]. Measurements of density were then calculated using the Gel Analysis package in ImageJ [151]. Estimates of the number of proteins present in the sample could then be made based on the ratio of the measured densities.

2.2.9 Negative stain for transmission electron microscopy

Continuous carbon grids purchased from either Electron Microscopy (CF400-Cu) or EMResolutions (CLC400Cu25-UT), were glow discharged for 1 minute. Glow discharging is a common method used for increasing the hydrophilic properties of the grid surface increasing the adhesion of proteins and thus the sample density on the grid. 3 μ l of protein sample was applied to the grid and excess liquid blotted off using filter paper (Whatman). Grids were then washed three times using ddH₂O. Samples to be visualized with negative stain were stained by placing the grid in 1% uranyl acetate for 1 minute. Excess uranyl acetate was then blotted with filter paper (Whatman) and the grids were left to air dry for several minutes. Negative stained samples were imaged using either a Phillips CM100 or a Tecnai T12 Spirit in the Faculty of Science Electron Microscopy Facility at the University of Sheffield.

2.2.10 Negative stained image analysis

Due to the complex structure, negatively stained images were segmented using the Segment Anything Model (SAM) from Meta [152]. The model was set up and run on a Google Collaberations document to make use of GPU acceleration. The resulting binary images were downloaded and analysed using the scikit-image and openCV packages [153, 154].

2.2.11 Template matching

Template matching was undertaken using a combination of functions from the OpenCV and Scikit-image python libraries [154, 153]. Template matching is a method for identifying the occurrence of an image patch or ‘template’ in a larger image using cross-correlation [155]. Templates representing encapsulins with 24 nm (T=1) and 32 nm (T=3) diameters were generated using the `circle()` function from the OpenCV library [154]. The generated templates consisted of two concentric circles, which were blurred using a Gaussian function (with a width of 2 pixels). The `match_template()` function from Scikit-image was then used to identify the occurrence of the differently sized encapsulin templates within a negatively stained electron micrograph. The number of matches was then counted using the `measure.label()` function from the Scikit-image library [153].

2.3 Results

2.3.1 Isolation of encapsulins from *M. xanthus*

Encapsulins were successfully isolated from *M. xanthus*, using a modified method from the established protocol in the literature [18]. The modifications made were, firstly, the incubation of the sample with 0.5% N-dodecyl- β -D-maltoside before the CsCl gradient step, this increased the membrane solubilization resulting in a cleaner final product isolation. Secondly, following the CsCl gradient the encapsulins were found enriched in a brown pellet at the bottom of the ultracentrifuge tube instead of in the expected grey/white fractions in the CsCl cushion. Thus, the CsCl gradient could be discarded and the pellet redissolved in ddH₂O. The isolated product was highly enriched in encapsulins, with very little contamination of either membranes or other cellular debris. The encapsulins were highly homogeneous and presented the expected 32 ± 1 nm ($N=1217$) size in stained images (see Figure 2.1). Furthermore, imaging of unstained samples revealed the expected dense core.

2.3.2 Long term stability of encapsulins

At the start of the research project, encapsulins from *M. xanthus* were isolated and imaged using negatively stained TEM. The material was then stored in ddH₂O at 4 °C and imaged again at the end of the research period. As can be seen in from the negatively stained electron micrographs (Figure 2.2a) the encapsulin structure is stable over this period of time; as the encapsulins appear the same in both images and there is no evidence of shell breakage. The lower density of encapsulins on the grid in the image from 2024 is most likely due to the sample being re-suspended in a ddH₂O before the grids were made, resulting in lower encapsulin density in the solution. This step was taken because of loss of water due to evaporation that occurred whilst the sample was in storage.

When visualized using SDS-PAGE analysis the band representing the EncA protein does not run at its expected height in the gel from 2020. It is expected that the protein should run at ~ 32 kDa, however, it runs at ~ 40 kDa when first isolated in 2020 and at ~ 32 kDa in 2024 (Figure 2.2). The additional height of the EncA band in 2020 is unexpected. It may be due to an unknown post translational modification to the protein or down to incomplete

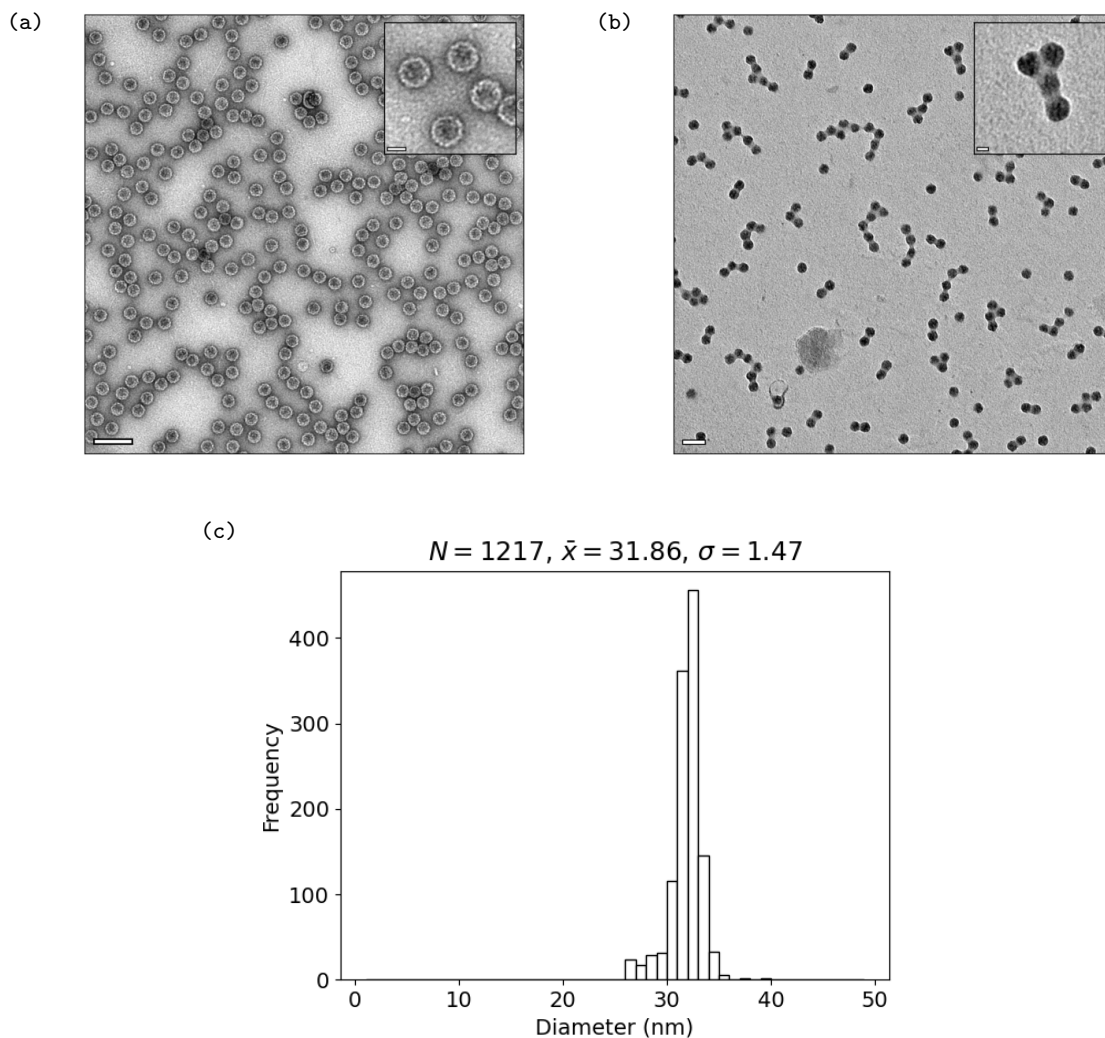


Figure 2.1: Encapsulin isolation from *M. xanthus*. (a) Image of encapsulin isolation stained with 1% uranyl acetate, the insert shows four negatively unstained encapsulins. The image was captured at 13kx. (b) Unstained image of encapsulin isolation, the insert shows four unstained encapsulins. The image was captured at 11kx. Both images were taken using a Tecnai Spirit. For both images the scale bar represents 100 nm and the insert scale bar represents 20 nm. (c) Histogram representing the distribution of measured particle diameters. Above it, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are quoted.

denaturation of the EncA protien before the gel was run. The bands for the three cargo proteins EncB (17kDa), EncC (13kDa) and EncD (11kDa) can be seen in both gels (Figure 2.2), and they run at the expected heights based in what has been previously observed in the literature [18, 28].

Thus, the encapsulin shell from *M. xanthus* is stable for a period of at least 4 years under the storage conditions used here. Futhermore, as bands representing the cargo proteins can be seen in the SDS gel, it is likely that they are also stable over this time period.

2.3.3 Gene amplification

Initially, all individual genes from the encapsulin system were amplified from *M. xanthus* genomic DNA and cloned into the petDUET-1 expression vector. EncA with an additional C-terminal 6xHis was successfully cloned into MSC1 and the cargo proteins were cloned MSC2 to create three plasmids. However, after transforming these plasmids into the *E. coli* Rosetta(DE3) and attempting an overexpression, only the EncA and EncC proteins were identified as successfully expressed, no expression was observed for either EncB or EncD. Initial attempts made to correct this included; designing different primers, changing the order of genes on the plasmid, and using different plasmids. However, none of these changes resulted in the expression of either EncB or EncD. Therefore, it was assumed that the lack of expression was due to the EncB and EncD sequences having large regions with high GC content which often cause issues when performing overexpressions in *E. coli* [156]. Therefore, the genes for the EncB and EncD proteins were codon optimized and synthesized (Twist Bioscience). The resulting fragments were then used as templates for PCR reactions to generate EncB and EncD fragments that could be ligated into the MSC2 site on the petDuet-EncA6xHis plasmid. In the end, five plasmids were generated using both petDUET-1 and pCOLADuet-1 to allow a combination of different encapsulins shell (EncA) and cargo proteins to be expressed: EncA+EncB, EncA+EncC, EncA+EncD and EncA+EncB+EncC.

2.3.4 Protein expression conditions

A series of different expression conditions were tested; IPTG concentrations, temperatures, and induction times. For IPTG concentrations greater than 0.1 mM the protein product was found in insoluble inclusion bodies following sonification. Similar results were observed

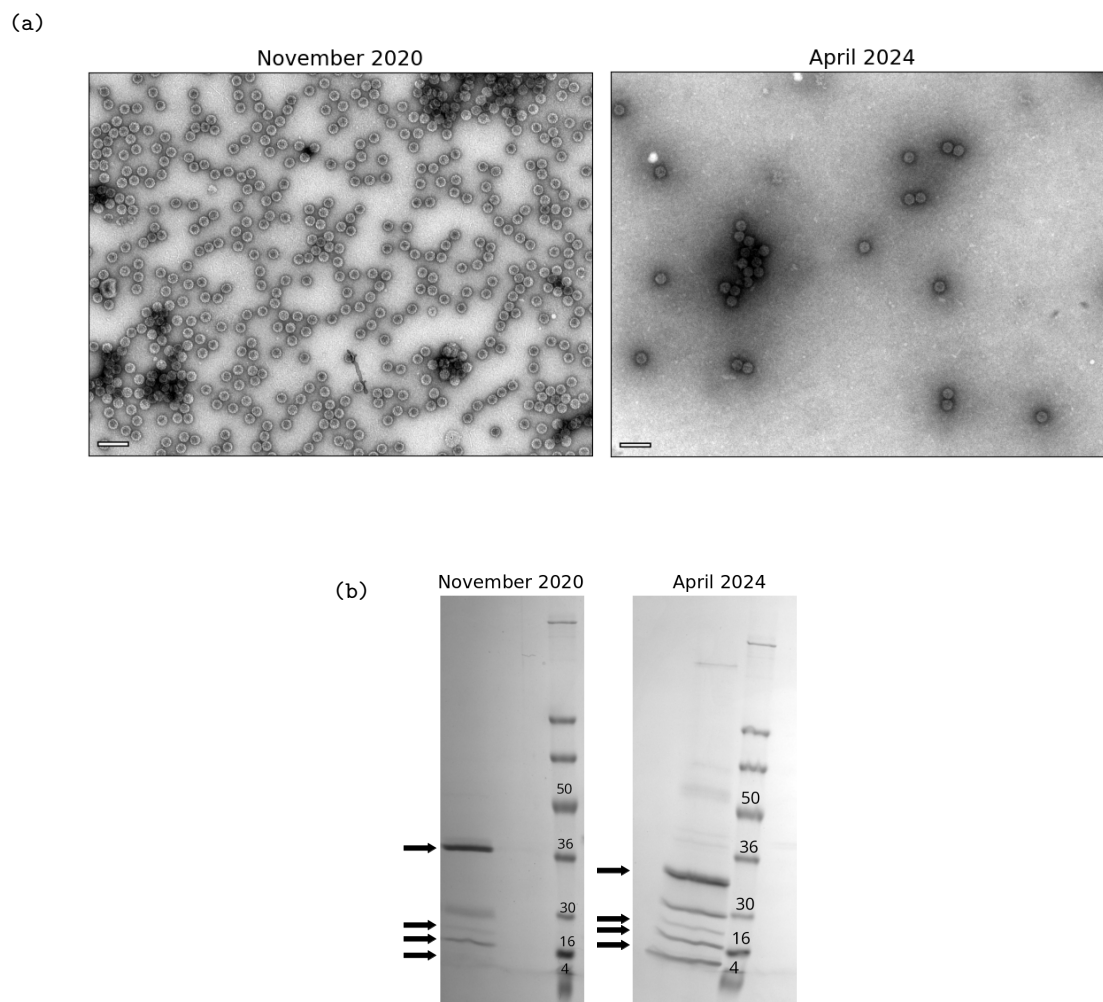


Figure 2.2: Observation of long term stability of encapsulin proteins. (a) Negatively stained image of encapsulin shells taken in November 2020 (left) and in April 2024 (right). Both images were captured on a Tecnai Spirit at 11kx magnification and stained with 1% uranyl acetate. Scale bars represent 100 nm. (b) SDS-PAGE gels of isolated encapsulins. The *M. xanthus* encapsulin sample was run in the left lane in both images, and bands representing the proteins EncA (32kDa), EncB (17kDa), EncC (13kDa) and EncD (11kDa) can be seen highlighted with arrows. Additionally, a non-specific band of a co-purified protein can be seen at ~30kDa in both images, this band is also present in SDS-PAGE gels of *M. xanthus* encapsulins that have been published in the literature [18]. In both gels the bands of the See Blue Prestained 2 protein ladder in the right lane of each image is labelled with the size of each protein band in kDa.

when using temperatures higher than 24 °C for time periods of longer than 3 hours. Therefore, for large scale purifications cultures were induced with a maximum of 0.1 mM IPTG for 18 hours at 24 °C. Expression under this condition resulted in high yield of soluble product (see Figure 2.4).

2.3.5 Overexpression of 6xHis-tagged EncA

It was decided that a 6xHis-tag would be used to purify the encapsulins from *E. coli* instead of the reported size exclusion chromatography method because of the increased ease of purification granted by the 6xHis tag and the large number of different encapsulin complexes that would be purified over the course of the research project [157]. Both the C- and N-termini appeared to be potential candidates for the 6xHis tag, however after modelling both options using Alphafold and comparing to the published atomic model [98, 18] it became clear that only the addition of the 6xHis tag to the C-terminus would result in an outward facing tag that could be used for purification (see Figure 2.3a). The tag was appended to the C-terminus through into inclusion in the reverse primer using during the PCR amplification. The generated EncA6xHis fragment was inserted into MCS1 on the petDUET-1 plasmid using the NdeI and XhoI restriction enzymes. The addition of the tag allowed for the successful purification of the expressed EncA protein using His-tag:Ni-NTA affinity chromatography. The presence of a band at 36kDa in the elution fractions on the SDS gel is evidence of the successful overexpression and purification of the EncA protein. The band at 36kDa is higher than the weight of the EncA protein that has previously been calculated bio-informationally and observed in earlier purifications [18]. The extra weight may be from the addition of the 6xHis tag and the longer (GGSGGS) linker region between the tag and the EncA protein. Under these conditions, the yield of protein exceeded the bed volume in the column resulting in protein being eluted in the wash steps (Figure 2.3b).

2.3.6 Simultaneous expression of shell protein EncA and single cargo proteins

The gene coding for the cargo protein EncC was successfully amplified from isolated genomic DNA using the primers listed in Table 2.2. The amplified fragment was then inserted into MCS2 on the petDUET-1-EncA6xHis plasmid using the NcoI and HindIII restriction

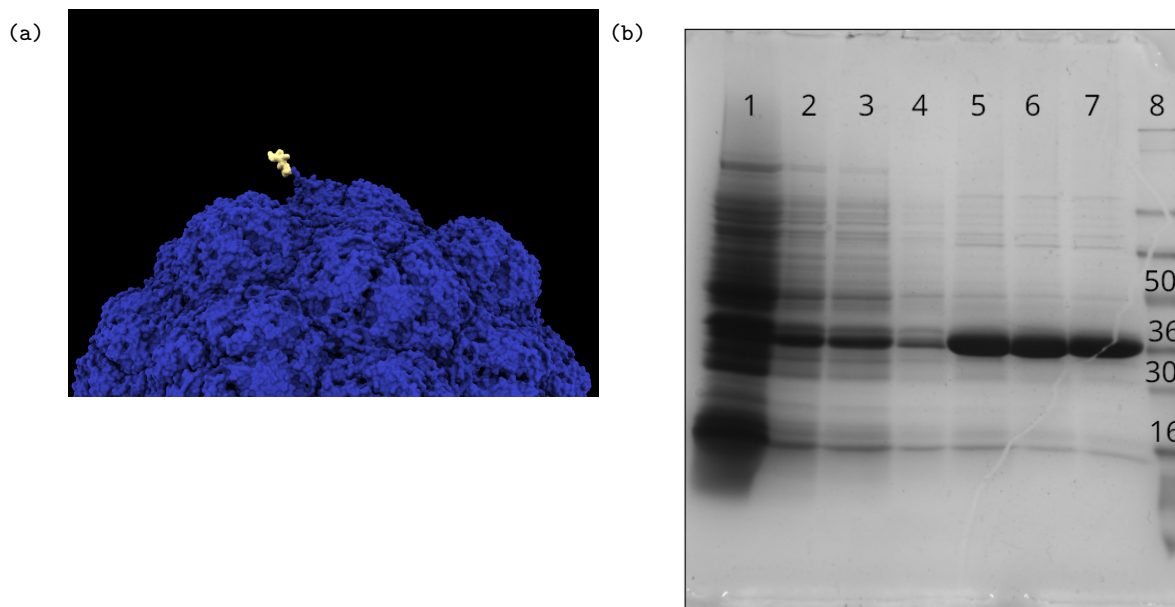


Figure 2.3: Purification of C-terminal tagged 6xHis EncA from *E. coli* Rosetta(DE3) following expression with 0.1 mM IPTG for 18 hours at 24 °C. (a) AlphaFold prediction model of a C-terminal tagged 6xHis EncA, the tag is highlighted in yellow on a single subunit of the encapsulin shell. The complete encapsulin shell atomic structure is shown in blue (PDB:4PT2). (b) SDS gel showing the purification of 6xHis tagged EncA from *E. coli* Rosetta(DE3). Each lane in the SDS-PAGE gel is labelled, from left to right: Lane 1: whole cell lysate. Lanes 2-4: three wash steps, the concentration of the EncA protein exceeded the bed volume of the column therefore some of this protein is eluted during the wash steps. Lanes 5-7: three elution steps, here the band at ~36kDa representing the EncA protein can be clearly seen. Lane 8: See Blue Prestained protein ladder (Invitrogen), the bands representing 50, 36, 30 and 16kDa are labelled.

enzymes. The generated plasmid, petDuet-1-EncA-EncC, was sufficient for the expression and purification of the two proteins simultaneously and bands representing both proteins can be seen in the elution fractions in the SDS-PAGE gel (see Figure 2.4). Based on hypothesis in the literature [12, 18, 19], cargo loading occurs co-translationally, without any additional proteins or co-factors. Thus, it is most likely that the two bands seen in the SDS-PAGE gel represent a formed encapsulin shell loaded with EncC cargo protein. Especially, as the cargo is untagged and therefore should not remain in the column after the wash steps.

Amplifying *encB* from the genome of *M. xanthus* and inserting it into the MSC2 of petDUET-1-EncA6xHis did not result in expression of the protein. The lack of expression was most likely caused by high GC (70%) content of the gene as previously discussed. Therefore, the gene for this protein was synthesised (TwistBioscience) and codon optimised for expression in *E. coli*. The synthesised product was amplified and inserted into the petDUET-1-EncA6xHis plasmid, using the XbaI and HindIII restriction enzymes. Evidence of the successful expression can be observed in the SDS gel where the two bands at 36kDa and 17kDa representing EncA and EncB respectively can be seen (Figure 2.4).

As with *encB*, the amplification of *encD* and its insertion into petDUET-1-EncA did not result in expression. Thus, *encD* was also synthesised (TwistBioscience) and codon optimized for expression in *E. coli*. The optimized gene was then inserted into petDuet-1-EncA6xHis plasmid using the XbaI and HindIII restriction sites. The successful expression of this construct can be seen by the presence of two bands in the elution fraction of the SDS gel, the two bands can be seen at 36kDa and 11.5kDa (Figure 2.4). Interestingly, there is another *E. coli* protein with a molecular weight of 11.5kDa that can be seen in all of the lanes in the SDS gels. Initially, there was some confusion as to whether this protein was EncD or a non-specific protein that was eluted under the same conditions as the tagged protein. Mass spectrometry was carried out to confirm the presence of EncD in the SDS-PAGE gel. Finding that the other protein present at the same height in the gel were the *E. coli* stress response protein ElaB (ECBD_139) which is linked to protecting the cell from oxidative and heat shock stress [158], and the DUF883-domain containing protein YqjD which is an inner membrane protein that associates with the 30S ribosome (ECBD_064) [159].

2.3.7 Simultaneous expression of EncA and multiple cargo proteins

During the next phase of the experiment, the multi-cargo encapsulin containing EncB and EncC was generated through the insertion of *encC* into the MSC1 of the petCOLADuet-1 plasmid. Co-transformation of this plasmid with petDUET-1-EncA-EncB resulted in an *E. coli* Rosetta(DE3) strain that is able to express the encapsulin shell protein EncA and two cargo proteins EncB and EncC simultaneously. Induction under the same conditions as the single cargo constructs, results in a high yield of multi-cargo complexes that could be purified using a His-tag:Ni-NTA affinity chromatography column. As with the single cargo complexes, the loading of the cargo occurs co-translationally, and no further steps were required to induce cargo loading. The multi-cargo encapsulins were confirmed through the observation of two bands representing EncB (17kDa) and EncC (13kDa) alongside the 36kDa band that represents the encapsulin shell protein EncA (see Figure 2.4).

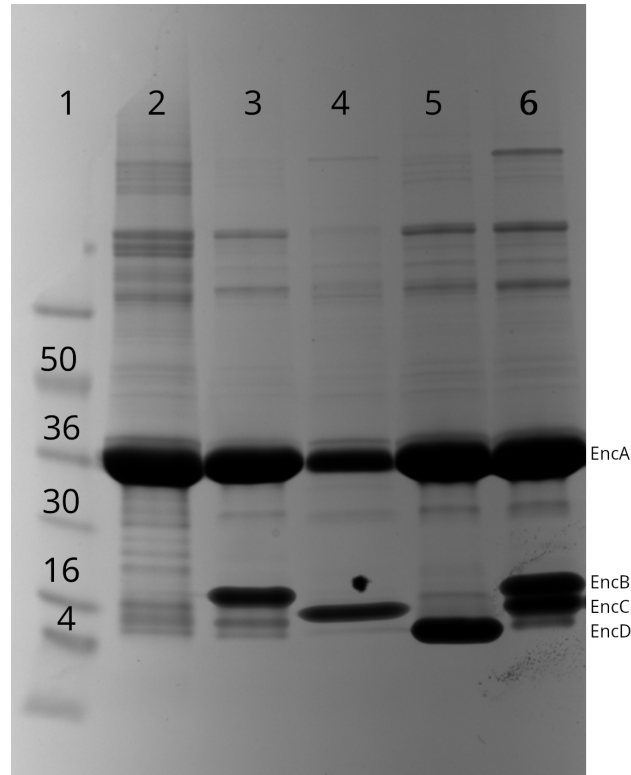


Figure 2.4: SDS-PAGE analysis of encapsulins expressed in *E. coli* and loaded with different cargo proteins. Lanes are labelled 1-6 from left to right. Lane 1: See Blue Prestained protein ladder (Invitrogen), the bands representing 50, 36, 30, 16 and 4kDa are labelled. Lane 2: sample of empty encapsulin. The band representing the EncA shell protein can be seen at 36kDa. Lane 3: sample of encapsulin containing the cargo protein EncB. The bands representing both the EncA shell protein (36kDa) and the EncB cargo protein (17kDa) can be clearly seen. Lane 4: sample of encapsulin containing the cargo protein EncC. The bands representing both the EncA shell protein (36kDa) and the EncC cargo protein (13kDa) can be clearly seen. Lane 5: sample of encapsulin containing the cargo protein EncD. The bands representing both the EncA shell protein (36kDa) and the EncD cargo protein (11kDa) can be clearly seen. Lane 6: sample of encapsulin containing the cargo proteins EncB and EncC. The bands representing both the EncA shell protein (36kDa), EncB cargo protein (17kDa) and EncC cargo protein (13kDa) can be clearly seen.

2.3.8 Cargo loading efficiency

SDS-PAGE gel densitometry is a commonly used method for estimating the relative amount of protein that is present in a gel, it is thus useful for estimating the numbers of cargo proteins that are present in the different encapsulins generated in this work [160, 161]. Firstly, it should be noted that there is a difference in the cargo protein content of the native *M. xanthus* encapsulins isolated here and those isolated in previous studies. In this work, the number of cargo proteins per shell is lower than the value published in the literature [18] (See Table 2.4). In the samples of encapsulins investigated here the shells contained a total of ~ 96 cargo proteins, with ~ 18 copies of EncB, ~ 61 copies of EncC and ~ 17 copies of EncD. The difference is most likely due to the fact that the EncB and EncD bands are relatively faint on the gel, due to the low concentration of encapsulin initially loaded into the gel. However, the relative ratios of cargo proteins in the encapsulins studied here are consistent with those reported in the literature, with EncC being the dominant cargo protein in both cases.

Calculations made from encapsulins expressed in *E. coli* reveal a loading of ~ 63 copies of EncB, ~ 96 copies of EncC and ~ 78 copies of EncD within the shell for the generated single cargo constructs. For the constructs loaded with EncB and EncC, there are ~ 61 copies of both proteins inside the encapsulin. When compared with the natively isolated structures from *M. xanthus*, it becomes evident that when a certain cargo protein is lacking higher numbers of those that are still available are loaded. However, across all single and multi-cargo encapsulins generated in *E. coli* the total number of cargo proteins was lower than the highest reported values in the *M. xanthus* encapsulins [18]. This is most likely due to the differences between native and heterologous expression in the two hosts. In *E. coli*, expression and cargo loading are not regulated in the same manner as in the native host, *M. xanthus* resulting in lower amounts of cargo protein loading being loaded into the encapsulin shells.

2.3.9 Visualisation of structures in TEM

Purified encapsulin constructs were negatively stained with 1% uranyl acetate and visualized using a Phillips CM100 (See Figure 2.5). The isolated structures were highly homogeneous, generally showing the expected ~ 32 nm diameter, and the structures appear very similar morphologically to those isolated from *M. xanthus* (see Section 2.3.1). This is

Construct	EncB	EncC	EncD	Total number of cargo proteins	Source
<i>M. xanthus</i> encapsulins	~36	~92	~47	~175	[18]
<i>M. xanthus</i> encapsulins	~18	~61	~17	~96	This study
EncA+EncB	~63	-	-	~63	This study
EncA+EncC	-	~96	-	~96	This study
EncA+EncD	-	-	~78	~78	This study
EncA+EncB +EncC	~61	~61	-	~122	This study

Table 2.4: The efficiency of cargo loading in encapsulin constructs calculated using SDS densitometry. The data from native *M. xanthus* encapsulins isolated by other groups is included as a reference [18].

a good indicator that the presence of the 6xHis tag on the C-terminus does not impede the ability of EncA to form complete encapsulin shells. The morphological similarity also provides further support for the suggestions made based on the SDS-PAGE analysis, that the cargo proteins are loaded inside the encapsulin shells. This is because no other proteins appear either associated with the outside of the EncA shells or in the background of the negatively stained electron micrographs. Although the vast majority of the encapsulin shells are similar to those isolated from *M. xanthus* a small percentage still appear to show the previously identified T=1 morphology [28, 18].

2.3.10 The presence of the cargo proteins during overexpression reduce the number of shells showing T=1 morphology

When the encapsulin system from *M. xanthus* is overexpressed in *E. coli* it is common to observe a subset of the population with a smaller diameter shell. These are structures which have formed the smaller T=1 icosahedron instead of the expected T=3 icosahedron. To examine the relationship between this phenomenon and the presence of cargo proteins, the

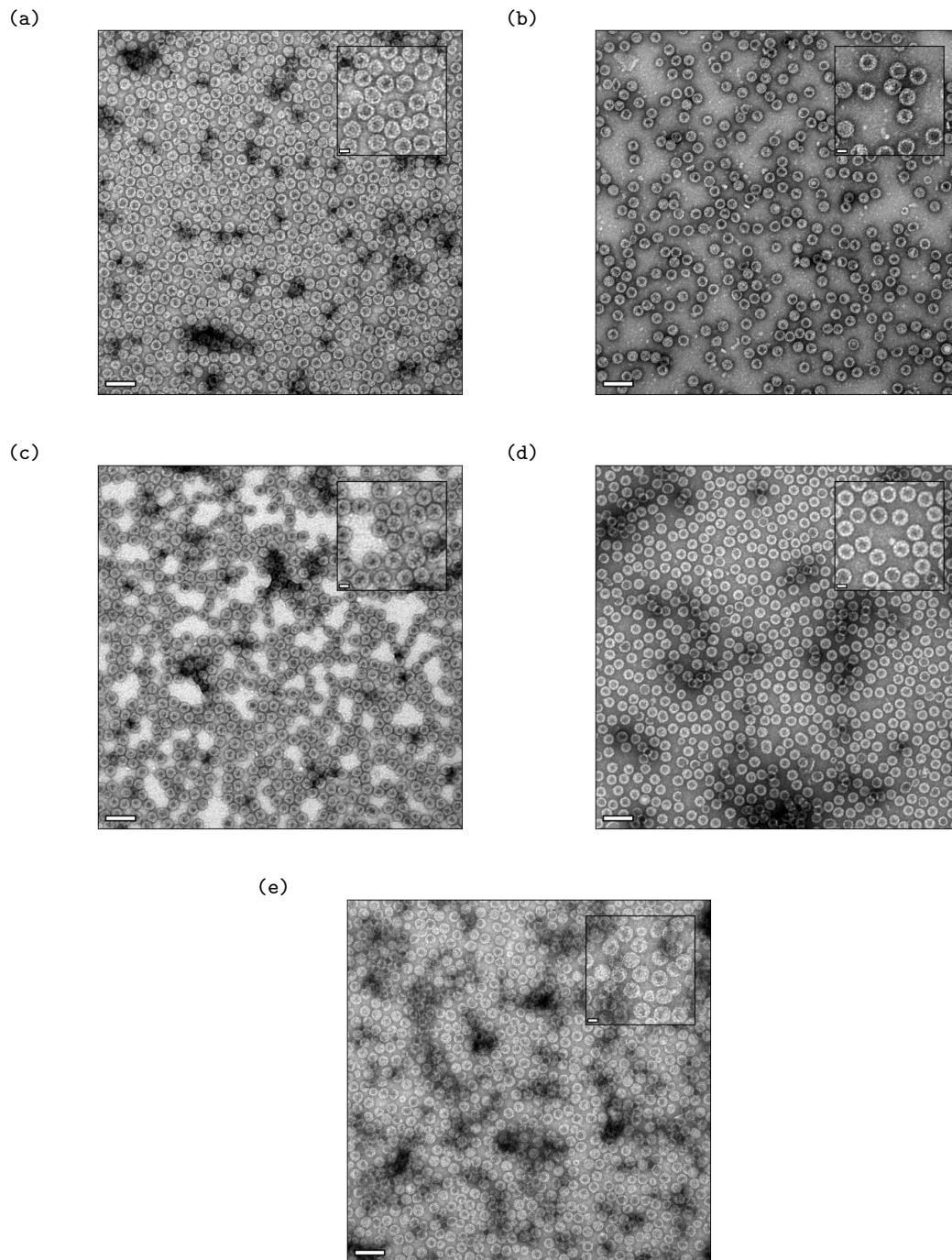


Figure 2.5: Electron micrographs of generated encapsulin constructs negatively stained with 1% uranyl acetate. (a) EncA (b) EncA+EncB (c) EncA+EncC (d) EncA+EncD (e) EncA+EncB+EncC . In the top right of each image there is a insert showing a selection of negatively stained encapsulins. There is variation between the structures within an isolation, this is most likely due to the staining pooling in the middle of the encapsulin causing them to collapse inwards to varying degrees. The images were captured using a Phillips CM100 operating at 80 keV at a magnification of 15.5kx. Image scale bars represent 100 nm and insert scale bars 20 nm. Importantly, the electron micrographs show no discernible difference between these encapsulins overexpressed in *E. coli* and encapsulins isolated from *M. xanthus*.

numbers of encapsulins showing both morphologies were quantified from negatively stained micrographs using a template matching approach (see Figure 2.6a).

When EncA is expressed in isolation, the percentage of the population which shows the T=1 morphology was recorded at 21.9% ($N=2379$). This value is in agreement with previous observations made in the literature: 36% [28] and 20% [18]. When a cargo protein is co-expressed alongside EncA there is a decrease in the number of shells showing the T=1 morphology. This decrease is to 5.5% ($N=1064$) for EncA+EncB, 1.2% ($N=2085$) for EncA+EncC and 0.8% ($N=2216$) for EncA+EncD. Thus, it is likely that the formation of the 32 nm encapsulin shell is stabilized by the presence of the cargo proteins under normal overexpression conditions. The observations made here are in agreement with observations reported in the literature, where highly loaded encapsulin shells expressed in *E. coli* did not show the T=1 morphology [38, 29]. Additionally, due to their relatively simple appearance in negatively stained TEM images, template matching is a viable and quick method for quantifying different sizes of encapsulin structures.

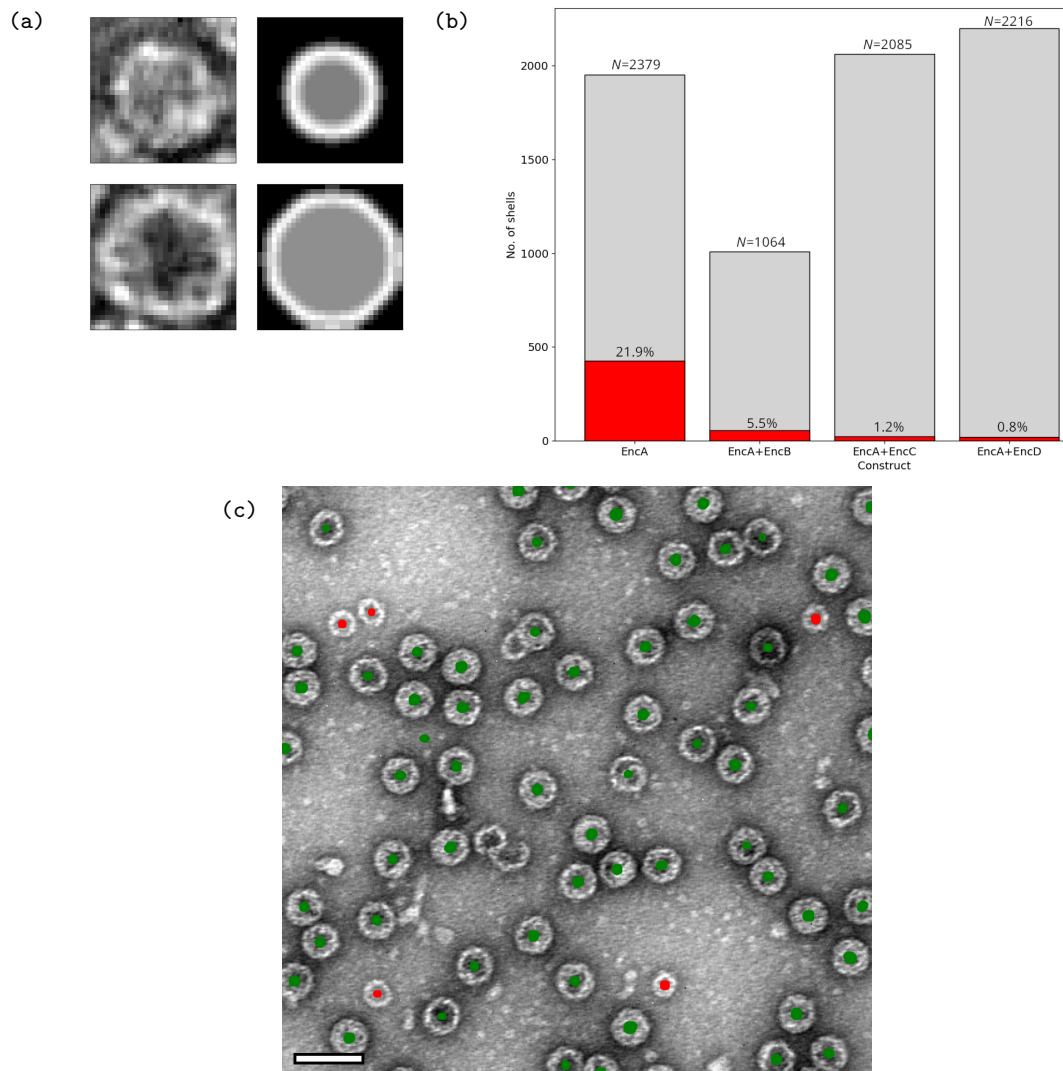


Figure 2.6: Presence of cargo proteins stabilizes the formation of the larger T=3 encapsulin shell. (a) templates generated of T=1 and T=3 shells. Top left: example of a negatively stained T=1 shell and top right: corresponding template of this structure. Bottom left: example of a negatively stained T=3 shell and bottom right: corresponding template of this structure. The templates consisted of two blurred concentric circles. (b) Bar chart showing the number of observed T=1 shells compared to T=3 shells for the different encapsulin cargo complexes expressed in *E. coli*. (c) The result of template matching of T=1 and T=3 templates to a negatively stained image of EncA proteins. Encapsulins identified as T=1 shells are marked with red dots, and those identified as T=3 shells are marked with green dots.

2.4 Discussion

The experiments presented here show that the encapsulin from *M. xanthus* can be isolated successfully using the protocol previously described in the literature [18], and that encapsulins containing either single or multiple cargoes can be overexpressed and purified from *E. coli* using a C-terminal 6xHis tag. The encapsulins isolated from *M. xanthus* were highly enriched and had the same 32 nm shell and dense core expected based on previous studies. However, there were some key differences; most notably, SDS densitometry revealed a difference in the total number of cargo proteins. The encapsulins isolated here contained a total of ~ 96 cargo proteins compared to the ~ 175 reported in the literature [18]. Observations of different but viable numbers of cargo proteins is interesting and indicates that controlling the amount of cargo proteins may be important for providing higher levels of control over the function of the encapsulin. Similar perhaps, to how different combinations of H- and L-chains effect the efficiency of iron uptake by mammalian ferritin. Where a combination of both results in more efficient iron uptake than having only either one present [162]. To provide higher levels of control over cargo loading it is most likely that additional proteins would be required, similar to the scaffold proteins that are required for successful viral capsid formation [163], however no such system has been identified either experimentally or bio-informatically in *M. xanthus*. Furthermore, the observation made here and previously in the literature [12, 38, 39], that encapsulins can be expressed and loaded with cargo in non-native hosts suggests no such system exists. Thus it is unclear why this difference in the cargo loading was observed, however, it may be due to the low levels of protein loaded onto the SDS gel used for densitometry resulting in an artefact during analysis. Additionally, the *M. xanthus* encapsulin is highly resistant to degradation over time. Evidenced, by the same structure being observed in negatively stained images after a period of 4 years, this result further supports previously made observations about the stability of these system [62]. High levels of structural stability may be important for the function of the encapsulin system, especially if it is the cells long term iron storage. Coupled with previous observations of encapsulins in cultural supernatant [20, 13], the stability of the protein shell may be required for ensuring the iron is available for other cells in the colony after the original cell dies.

Expression of the encapsulin proteins in *E. coli* results in the production of high yields

of pure and homogeneous protein constructs. The insertion of EncA and a single cargo protein into the two multiple cloning sites of the petDUET-1 plasmid was sufficient for the generation and purification of encapsulins consisting of the shell protein, EncA and a single cargo protein. Transformation of pCOLADUET-1 containing additional cargo protein into the overexpression strain was sufficient for the production and purification of EncA and multiple cargo proteins. The combination of the results from both SDS-PAGE and negatively stained TEM indicate that the cargo proteins were successfully loaded into the encapsulin shell. As the cargo proteins were eluted alongside the 6xHis tagged EncA protein in the final step from the His-tag:Ni-NTA affinity chromatography column despite not having their own 6xHis tags. Furthermore, in electron micrographs only complete encapsulins were visible on the grid indicating that the cargo proteins are not free in the solution. Quantification of the SDS gels indicates that the *E. coli* expressed encapsulins contained higher numbers of the individual cargo proteins than the native *M. xanthus* encapsulins. However, the total number of cargo proteins was always found to be lower even in conditions where two cargo proteins were expressed simultaneously. This might be due to the preferential binding sites on the inner face of the encapsulin shell that were identified previously from the atomic structure [28]. This indicates that despite the abundance of single cargo proteins in the *E. coli* overexpressions the preferential binding the lumen are most likely maintained, despite the high levels of similarity between the cargo loading peptide of the three *M. xanthus* cargo proteins. Alternatively, the lower levels of cargo loading may result from expressing the system in a non-native host. In the *E. coli* overexpression strain, encapsulin production is not as tightly regulated as in its native host, which could explain the reduced cargo loading observed in these constructs. Finally, it was observed that the presence of cargo proteins when EncA is over expressed in *E. coli* reduces the number of aberrant T=1 capsids that are formed, which is in agreement with similar observations that have been made in the literature [29]. The reason for the aberrant shell formation is currently unknown, however, it may be that under the expression conditions used in *E. coli* the encapsulin shell is closed before any cargo loading can occur.

The genetic and protein constructs generated in this chapter are used in experimental work described in the following chapters.

Chapter 3

Electron microscopical investigation of encapsulin derived nanoparticles

In this chapter, a number of different electron microscopical imaging techniques are used to characterize isolated *M. xanthus* encapsulins as well as encapsulins expressed in *E. coli* containing reduced numbers of cargo proteins. Firstly, following the successful expression of the encapsulin proteins in *E. coli*, methods for loading iron into the empty encapsulins are discussed, including both *in vivo* and *in vitro* approaches. These experiments were undertaken so that the relative iron mineralizing function of the cargo proteins could be studied and compared. Furthermore, nanoparticles were synthesised *in vitro* to assess the importance of the cellular context on this process.

The thickness and shape of the *M. xanthus* encapsulin shell and the size and morphology of the iron rich core are explored using Secondary Electron (SE) and High Angle Annular Dark Field (HAADF) imaging. A combination of Bright Field (BF) TEM and HAADF field techniques are used to quantify the size and relative distribution of nanoparticles both inside the *M. xanthus* and *E. coli* expressed encapsulins. These experiments have been undertaken to allow the iron mineralizing function of the *M. xanthus* encapsulin to be characterized and compared to previous isolation made in the literature, as well as allowing the iron mineralizing function of the cargo proteins to be compared within the context of the encapsulin protein shell. Finally, a combination of Selected Area Electron Diffraction (SAED) and High Resolution Scanning Transmission Electron Microscopy (HR-STEM) are used to explore the

atomic arrangement of the nanoparticles inside the encapsulins shell. Providing insight into the possible relationship between the crystallinity of the nanoparticles and the wider biological function of the encapsulin iron storage system.

3.1 Experimental procedures

3.1.1 Iron loading *in vivo*

Encapsulins were loaded with iron *in vivo* to better mimic the conditions found in the *M. xanthus* cell. A starter culture of *E. coli* Rosetta(DE3) containing the desired plasmids was grown overnight in LB (Sigma; 10 g/L tryptone, 5 g/L NaCl, 5g/L yeast extract) media containing the appropriate antibiotics. This preparation was then used to inoculate a larger culture at a dilution of 1:100. Cells were grown to an OD₆₀₀ of 0.4-0.6 before being induced with 0.05 mM IPTG and grown for two hours at 30 °C, shaking at 200 RPM. Cells were harvested at 5,000 x *g* for 10 mins before being re suspended in LB media containing the following: 50 mM HEPES (Fisher Scientific), 4 mM trisodium citrate (Fison Analytical Reagents) and the desired concentration of Fe(NH₄)₂(SO₄)₂ (VWR chemicals, 99.8 % purity). Cultures were then left for 18 hours shaking at 200 RPM in a 30 °C incubator. Following this incubation, cells were harvested at 5000 x *g* for 10 mins and stored at 4 °C.

3.1.2 Iron loading *in vitro*

In vitro methods for loading Fe into encapsulins were used to create a P free environment, to examine the effect of P on the formation of crystalline material inside the encapsulin. Stock solutions of 2.5 mM and 20 mM Fe(NH₄)₂(SO₄)₂ were prepared in 0.1% (v/v) HCl. Encapsulins containing EncB and EncC were overexpressed and purified from *E. coli*, using the previously described method (see Section 2.2.5) and stored in a storage buffer (50 mM Tris, 500 mM NaCl at pH 8). The encapsulins were confirmed to be empty through observation in TEM. The prepared stock solutions of Fe(NH₄)₂(SO₄)₂ were diluted ten-fold in 30 μM of protein solution. The resulting solution was agitated and left exposed under aerobic conditions for 1 hour. This step was repeated with the 20 mM solution until a final concentration of 8 mM Fe was reached. The resulting solutions were dialysed against the storage buffer for 48 hours to remove precipitated Fe.

3.1.3 Encapsulins used in this study

The encapsulins and the experimental conditions can be found in Table 3.1. Briefly, the encapsulins used in this study consisted of native structures isolated from *M. xanthus* and a number of different structures overexpressed in *E. coli* and loaded through the addition of various concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the growth media or through the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to purified empty encapsulins. For detailed protocols on how the overexpressed encapsulins were loaded with iron see Sections 3.1.1 and 3.1.2.

3.1.4 Magnetic separation of Fe filled encapsulins

Constructs were tested for magnetic properties using a MACS magnetic cell separator (Miltenyi Biotec). Initially, the column was washed with sample storage buffer (10 mM Tris, or 50 mM HEPES 500 mM NaCl at pH 8). Encapsulin samples from either *M. xanthus* or *E. coli* were run through the column and the flow through collected, the column was then washed with 100 μl of storage buffer. Material remaining in the column was eluted by removing the column from the magnetic field and plunging after the addition of 50 μl of storage buffer. The elution was then checked for its protein content using SDS-PAGE (see Section 2.2.7).

3.1.5 Electron microscopy

A number of different microscopes and imaging techniques were used over the course of this study. A complete list can be found in Table 3.2.

3.1.6 Grid preparation

Before grid preparation, the preparation area and tweezers were cleaned with 70% ethanol. For TEM imaging of unstained samples, thick continuous carbon film grids were purchased from Electron Microscopy (CF400-Cu), for STEM imaging of unstained samples, ultra-thin continuous carbon on lacy carbon grids were purchased from EMResolutions (CLC400Cu25-UT). Regardless of carrier type, grids were glow discharged for 1 minute, a 3 μl drop of the sample was pipetted onto the grid and left for a sufficient time for the sample to settle on the grid. Excess sample was then blotted off using filter paper (Whatman), washed

CHAPTER 3. ELECTRON MICROSCOPICAL INVESTIGATION OF
ENCAPSULIN DERIVED NANOPARTICLES

Host	Construct	IPTG Conc. (mM)	Fe(NH ₄) ₂ (SO ₄) ₂ (mM)
<i>M. xanthus</i>	EncA, EncB, EncC, EncD	-	-
<i>E. coli</i>	EncA	0.05	8
<i>E. coli</i>	EncA+EncB	0.01	8
		0.05	2
		0.05	4
		0.05	8
		0.1	8
<i>E. coli</i>	EncA+Enc	0.01	8
		0.05	2
		0.05	4
		0.05	8
		0.1	8
<i>E. coli</i>	EncA+EncD	0.05	8
<i>E. coli</i>	EnA+EncB+EncC	0.01	8
		0.05	2
		0.05	4
		0.05	8
		0.1	8
Loaded <i>in vitro</i>	EnA+EncB+EncC	0.05	0.25
			1
			2
			4
			8

Table 3.1: Encapsulins generated for use in this study. For the encapsulins expressed in *E. coli* the concentration of IPTG used to induce the protein expression and the concentration of Fe(NH₄)₂(SO₄)₂ added to the growth media are shown.

Machine	Operating voltage (kV)	Type	Imaging capability	Location
Phillips CM100	100	TEM	BF TEM	Faculty of Science Electron Microscopy Facility, Sheffield
Tecnai T12 Spirit	100	TEM	BF TEM	Faculty of Science Electron Microscopy Facility, Sheffield
JEOL F200	200	S/TEM	BF TEM, BF STEM, HAADF STEM, SAED	Sorby Centre, Sheffield
FEI Titan G2 CHEMISTEM	200	S/TEM	HAADF STEM	Ernst Ruska Centre, Juelich, Germany
Hitachi HF5000	200	S/TEM	BF STEM, ADF STEM, SE STEM	Ernst Ruska Centre, Juelich, Germany
Tescan Tensor	100	STEM	4D-STEM	Ernst Ruska Centre, Juelich, Germany
JEOL ARM200F	80	S/TEM	BF STEM, ADF STEM, HAADF STEM	ePSIC, Diamond Labs, Harwell

Table 3.2: Electron microscopes used during the course of this study. The microscopes, operating voltage in kV, imaging capabilities, and locations are given. The following acronyms are used: Transmission Electron Microscope (TEM), Bright Field (BF), Scanning/Transmission Electron Microscope (S/TEM), Scanning Transmission Electron Microscope (STEM), High Angle Annular Dark Field (HAADF), Selected Area Electron Diffraction (SAED), Annular Dark Field (ADF), Secondary Electron (SE) and 4 Dimensional Scanning Transmission Electron Microscopy (4D-STEM).

three times using ddH₂O and left to air dry.

3.1.7 Image preprocessing

Before image data processing in Python3 steps were taken to enhance the detection of features in the images. These steps were applied to all images regardless of the imaging technique or the microscope they were captured on. The steps taken are as follows:

- A 3x3 median filter was applied to the image to reduce shot noise.
- The images were normalised, so that all intensity values fell between 0-1, making thresholding easier.
- Large images (4096x4096 pixels) were tiled before processing (into 1024x1024 pixels) and stitched back together afterwards. This step was taken to decrease the script runtime.
- Finally, a Gaussian filter (width of 2 pixels) was applied to the image to further reduce noise and allow for easier feature

3.1.8 Image analysis

All images were analysed using Python3 as it allowed easy batch processing of large numbers of images. The majority of analysis was carried out using the scikit-image and openCV packages [153, 154]. Scripts were written that identified desired features using thresholding (manual, Otsu and multi-Otsu were used [164, 165]) or through the identification of pixels with the highest contrast value. Desired region properties could then be measured, including: diameter, maximum axis length, minimum axis length, total intensity and inter-region distance.

3.1.9 Encapsulin shell image analysis

Analysis of encapsulin shells were undertaken from BF, DF and SE images. A general overview of steps taken in the Python scripts can be seen in the workflow diagram in Figure 3.1.

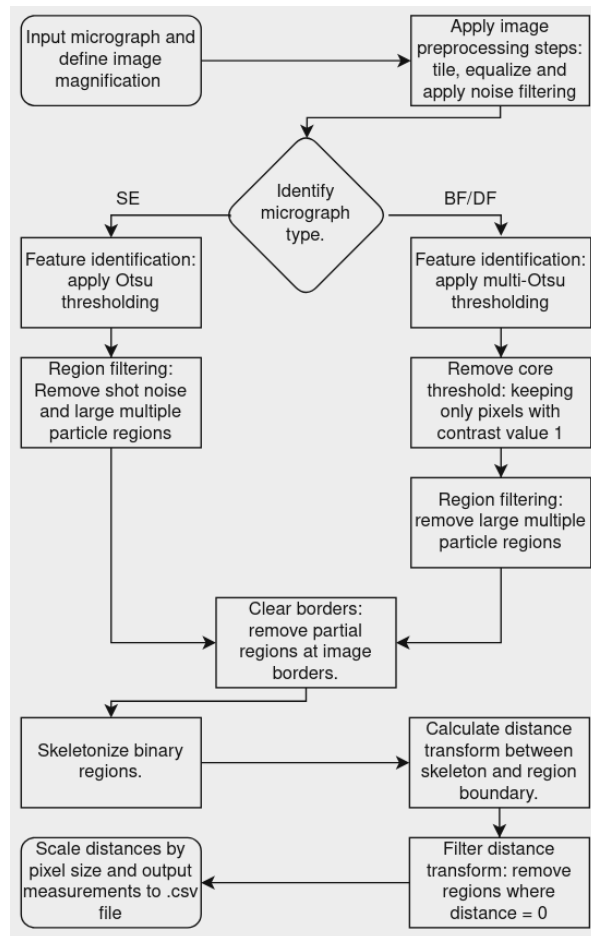


Figure 3.1: Flow chart showing steps taken to identify and measure encapsulin shells.

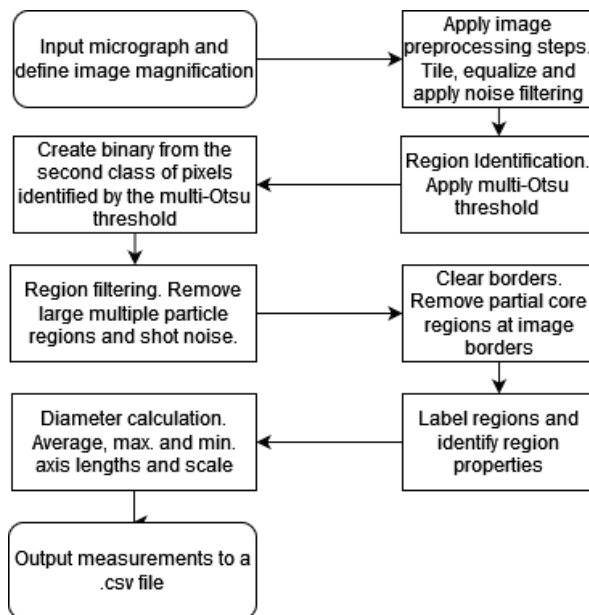


Figure 3.2: Flow chart showing steps taken to identify and measure encapsulin cores.

3.1.10 Encapsulin core region analysis

Encapsulin core regions were analysed from DF images. A general overview of the steps taken can be seen in Figure 3.2.

3.1.11 Internal nanoparticle analysis

Internal nanoparticles were analysed from BF TEM, BF and ADF images. Before analysis BF images were inverted such that the same script could be used for all images. A general overview of the steps taken can be seen in Figure 3.3.

3.1.12 4D-STEM and data processing

4D-STEM was carried out using a Tescan Tensor at Ernst-Ruska Centre in Juelich, Germany. The microscope was operated by Dr. Penghan Lu at 100kV with a convergence angle of 22 mrad and a probe size of ~ 1 nm. For data processing, the 4D-STEM dataset was first opened using lazy loading, part of the Python package [166]. Lazy loading makes use of out-of-core computation algorithms to load datasets larger than the available memory by chunking. A binary mask was generated from the spatial image based on the `flood_fill()` scripts used previously for analysing the nanoparticles inside the *M. xanthus* encapsulin (see Section

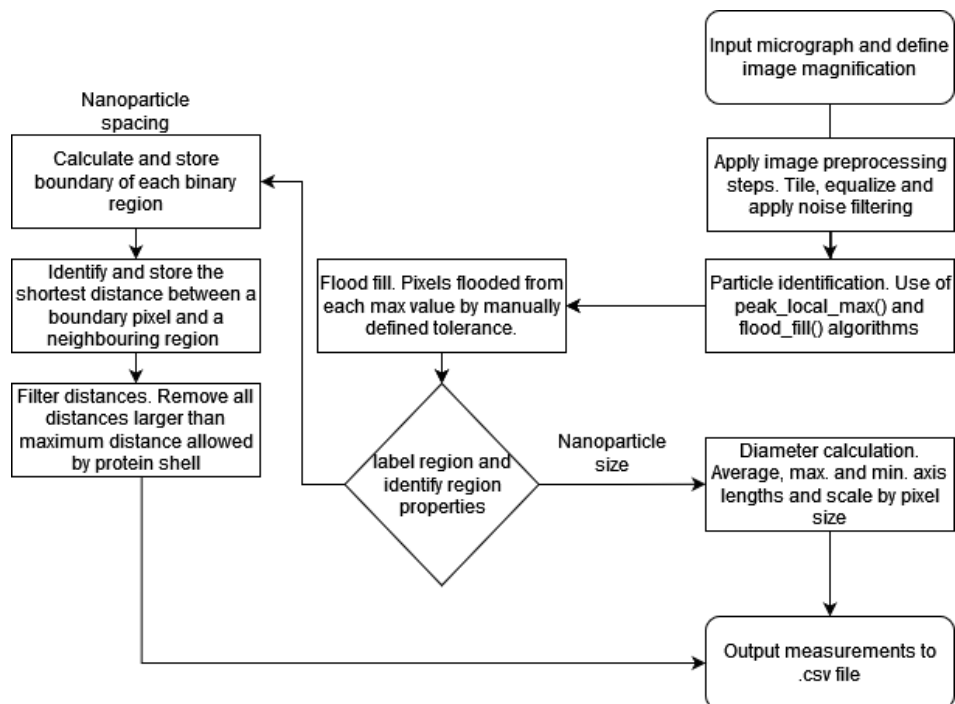


Figure 3.3: Flow chart showing steps taken to identify and measure nanoparticles properties.

3.1.11). Masking and removal of the background in this way greatly increased the speed of data processing. The spectral images from masked pixels were extracted from the dataset, visualized on a log scale and then saved. Images were visualized on a log scale because the intensity from the direct beam far exceed the brightness of any possible diffraction spots, so the only contrast would otherwise come from the direct beam. After saving, the images were investigated manually to identify regions that showed diffraction contrast.

3.1.13 Selected area electron diffraction

Before imaging, ultra-thin C coated grids (EM-Resolutions, CLC400Cu25-UT) were heavily loaded with encapsulins to ensure a high density of material and increased likelihood of signal generation. The JEOL F200 was first aligned in TEM mode and a selected area aperture (50 μm) was inserted and centred over a suitable sample region. Diffraction patterns were collected at a camera length of 800 mm using the OneView camera (Gatan).

3.1.14 HR-STEM and data processing

HR-STEM experiments were conducted at both the Diamond Light Source and the Ernst Ruska Centre. In the Diamond's ePSIC lab in Harwell the microscope was operated together with Dr. Mohsen Danaie. The JEOL ARM200F was operated at 80 kV, with the 30 μm probe forming aperture inserted and the spherical aberration corrector activated. These conditions gave a convergence angle of 22.6 mrad and a probe size of $\sim 0.1\text{nm}$. Both BF and HAADF images were collected. At the Ernst Ruska Centre, the FEI Titan G2 ChemiSTEM was operated by Dr. Jo Janghyun at 200 kV, the microscope was set up to give a convergence angle of 22.4 mrad and the spherical aberration corrector was activated. Only HAADF images were collected.

For data analysis, Digital Micrograph 3.6 (Gatan) was used. Images from the JEOL ARM200F were loaded and the magnification value in the meta data corrected after it had been noticed the values were incorrect due to an earlier service visit. Images from the FEI Titan G2 ChemiSTEM were loaded and then scaled based on the calibrated scale bar present on the image. Lattice planes were identified and measured using the line profile tool in Digital Micrograph integrated over a 50 pixel wide area.

3.2 Results

3.2.1 Encapsulin shell thickness

Various imaging techniques, including Bright Field (BF), High Angle Annular Dark Field (HAADF), and Secondary Electron (SE) imaging, were employed to measure the thickness of the encapsulin shell (see Figure 3.4). HAADF imaging provides visualization of the shell because contrast is dominated by the atomic number of the elements present. Consequently, the protein shell appears as a low-contrast region encircling the bright core. In contrast, SE imaging is sensitive to topological variations, enabling clear distinction between the shell and the background.

SE images of encapsulins displayed bright ring-like features identified as the protein shell. Some of these ring structures appeared incomplete, as if parts of the shell were missing. This phenomenon is particularly notable in areas where multiple shells are in contact with each other. In such regions, low topological change in between touching encapsulin shells results in low SE signal. Additionally, similar partial ring structures were observed in some mono-dispersed shells. In this case, the phenomenon is attributed to electrostatic forces acting on the shells during sample loading, which may distort the protein shell without causing shell breakage.

No measurements of the shell were made using BF images due to insufficient contrast between the protein shell and the background from the carbon support film (Figure 3.4a). However, both HAADF and SE images contained identifiable protein shells. The average shell thickness measured from HAADF images was 3 ± 2 nm ($N=44871$) and the average thickness measured from SE images was 3 ± 2 nm ($N=40058$). The consistent thickness values measured between the two imaging techniques is a good indication that the same structure was measured in both instances. Additionally, the average thickness values obtained from these measurements align with previously published structural data [18, 28], confirming the accuracy of the measurements made here. Additionally, the agreement between the techniques indicates that the structures examined in the SE images were indeed the encapsulin shells, and that SE and HAADF imaging techniques are valid approaches for measuring protein sizes under these conditions.

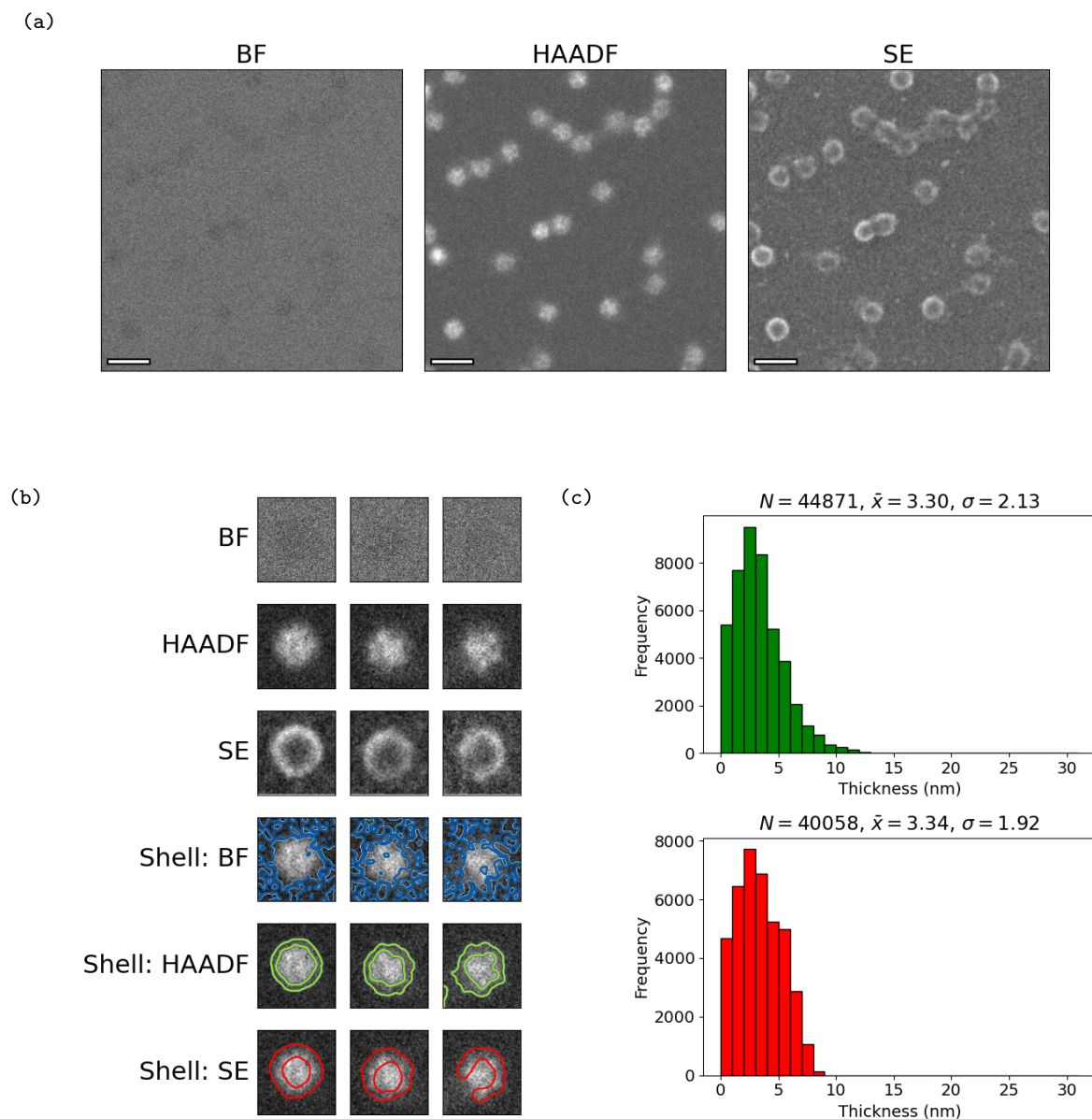


Figure 3.4: Measuring the thickness of the encapsulin shell using different imaging techniques. (a) BF (left), HAADF (middle) and SE (right) images of *M. xanthus* encapsulins. Images captured simultaneously on a Hitachi HF5000 FE-S/TEM at a magnification of 60kx. Scale bars represents 50 nm. (b) Three magnified examples of shell recognition from (a). Top to bottom: BF image, HAADF image, SE image, HAADF image overlay with shell region from BF images (blue), HAADF overlaid with shell regions from HAADF images (green) and HAADF images with shell region from SE images (red). (c) Distribution of shell thickness calculated from HAADF (green) and SE images (red). Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown.

3.2.2 *M. xanthus* encapsulins show core size variability in different isolations

Images of unstained encapsulins were captured with ADF/HAADF STEM and BF STEM imaging techniques. In DF imaging the cores are easily distinguishable as regions of high contrast. In the BF images however, the contrast between the core and background is too low to allow for easy identification, and thus ADF/HAADF images were used for calculations of core size (see Figure 3.5). The average core diameter differed between the three isolations. In the first isolation the average diameter of the core region was measured as 22 ± 5 nm ($N=542$), in the second 18 ± 3 nm ($N=1453$) and the third 24 ± 3 nm ($N=239$). However, the difference in the average core size between isolations was not found to be significant.

The core of the encapsulin often appears non-symmetrical, indicating that the entire lumen is not filled with nanoparticles (Figure 3.5g). This is evident when investigating the maximum and minimum core diameters for each core region. In isolation 3, the average maximum diameter measured was 26 ± 3 nm ($N=239$) and the average minimum diameter measured was 21 ± 3 nm ($N=239$). As previous experiments have identified that the internal lumen of the protein shell had a diameter of 26 nm (see section 3.2.1), a maximum measurement of 26 nm for the core region is a good indicator that the entire diameter within the encapsulin shell can be used for nanoparticle storage.

3.2.3 Size of internal nanoparticles

The images of encapsulins where the internal nanoparticles could be resolved were combined from all three isolations and separated by the imaging technique used, either BF TEM or HAADF STEM imaging. Given the complex nature of how the nanoparticles are arranged in the image - stacked on top of each other - thresholding of the image was insufficient for identifying the individual nanoparticles. Instead nanoparticles were identified using a `peak_local_max()` algorithm in the Sci-kit image package [153]. This algorithm allowed the individual nanoparticles to be identified as they were the brightest features in the image. A flooding algorithm (also from the Sci-kit image package [153]) was then used to find the edges of the identified nanoparticles, the flooding algorithm works by grouping particles that have contrast values within a defined range of each other moving outwards from an initial

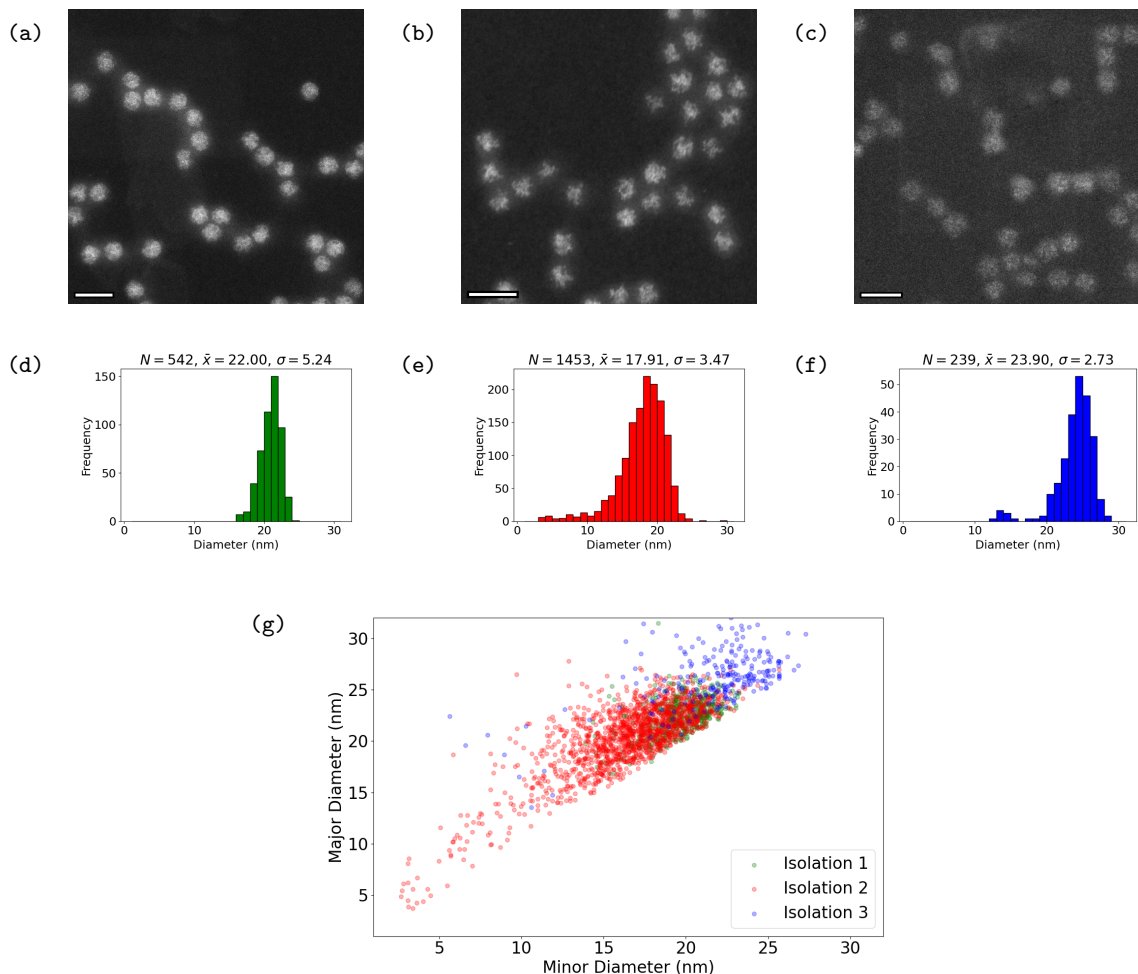


Figure 3.5: Comparison of *M. xanthus* core size from different isolations.(a) HAADF image of encapsulin isolation 1, imaged on a JEOL F200 operated at 200kV and a magnification of 500kx (b) HAADF image of encapsulin isolation 2, imaged on a JEOL F200 operated at 200kV and a magnification of 600kx. (c) HAADF image of encapsulin isolation 3, imaged on a Hitachi HF5000 FE-S/TEM operating at 200keV at a magnification of 60kx. (d)-(f) Histograms presenting the average size of the Fe rich core measured for the three *M. xanthus* isolations: isolations 1 (red), isolation 2 (green), isolation 3 (green). Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown. (g) Scatter plot showing the major and minor diameters of the three isolations.

seed point (for examples of images where nanoparticles have been identified see Appendix 3.4.1 and Appendix 3.4.2 for examples of TEM and HAADF images respectively). Using this technique, the same number of nanoparticles were measured per shell between the two imaging techniques: 40 ± 7 ($N=13$) from the encapsulins imaged with HAADF STEM and 42 ± 13 ($N=13$) for the encapsulins imaged with BF TEM. Identifying ~ 40 nanoparticles per shell indicates that the number of nanoparticles per shell is far higher than the previously made estimates of 15 [18]. Given the difficulty in identifying individual nanoparticles, and the overlapping arrangement, it is likely that the actual number of nanoparticles per shell is higher still than the value identified here.

The average nanoparticle diameter measured from HAADF STEM images was 1.8 ± 1.0 nm ($N=321$) whilst the average diameter measured from BF TEM images was 1.8 ± 0.6 nm ($N=824$). The similarity between the measurements made using the two imaging techniques is a good indicator that the size measured is reliable. Combined and accounting for systematic errors in measurements due to specimen drift and resolution limits, this gives an average nanoparticle size of ~ 2 nm. Additionally, the distance between the boundary of each nanoparticle and its closest neighbour was measured to assess the distribution of nanoparticles inside the shell. An average inter-nanoparticle distance of 2 ± 1 nm ($N=288$) was measured from HAADF STEM images, whilst an average distance of 1 ± 1 nm ($N=296$) was calculated from BF TEM images. Again, there is strong agreement between these two numbers, however, both sets of data are highly variable due to the low sampling of nanoparticles in many of the shells.

Using the previously identified internal diameter of the encapsulin core of 26 nm (see Section 3.2.2) and by calculating the volume of space filled by the internal cargo proteins an upper estimate the number of nanoparticles per shell can be made. The volume of the cargo proteins is based on dimensions reported in the literature from atomic models. However, this estimation does not account for unordered regions that could not be resolved at the atomic level. Consequently, the values presented here likely represent an underestimation of the true volume of the cargo proteins. The cargo proteins form decameric rings with dimensions of 7 nm x 5 nm (EncB) and 8.5 nm x 5 nm (EncC) [28]. EncD instead forms dimers, that are localized in pairs [97] and can be modelled as a rectangular prism with dimensions 4.6 nm x 5.8 nm x 2 nm giving a volume of 56 nm³. Combined, with the number of cargo proteins

CHAPTER 3. ELECTRON MICROSCOPICAL INVESTIGATION OF
ENCAPSULIN DERIVED NANOPARTICLES

Cargo Protein	No. per shell	No. of proteins per quaternary structure	Volume per quaternary structure (nm ³)	Total volume (nm ³)
EncB	~18	10	192	384
EncC	~61	10	283	1698
EncD	~17	4	56	784

Table 3.3: Calculation of the volume of internal cargo proteins. The volumes were calculated based on the published atomic models [28, 97]

per shell calculated from SDS quantification (Table 2.4) the total volume of the encapsulin filled by cargo proteins is estimated at $\sim 2866 \text{ nm}^3$. The total internal volume is $\sim 9203 \text{ nm}^3$ meaning that the volume remaining for nanoparticles to be loaded into is $\sim 6337 \text{ nm}^3$. The nanoparticles have an average diameter of 1.8 nm and therefore an average volume of 3 nm^3 and thus the maximum number of nanoparticles that could be loaded per shell is ~ 2000 .

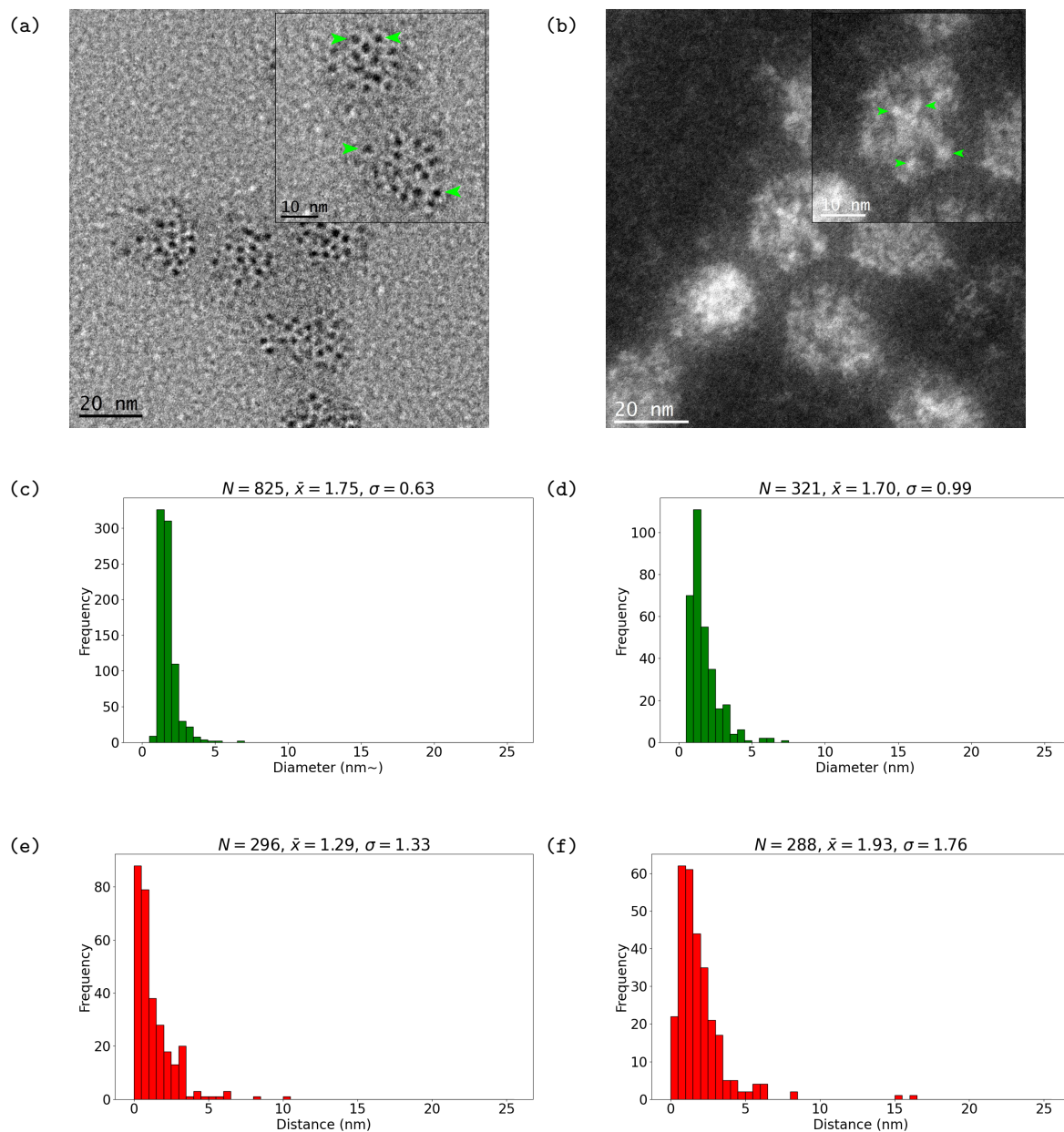


Figure 3.6: Measurements of the nanoparticle diameters and inter-nanoparticle distances for the *M. xanthus* using two imaging methods. (a) Unstained electron micrograph captured on a JEOL F200 at a magnification of 400kx. Scale bar represents 20 nm. Insert: two encapsulins, example nanoparticles highlighted with green arrows. (b) HAADF image of encapsulins captured on a JEOL ARM 200F at a magnification of 3milx. Scale bar represents 20 nm. Insert: A single encapsulin with example nanoparticles highlighted with green arrows. (c) - (d) Histograms representing the average nanoparticle diameters calculated from TEM and HAADF images respectively. (e) - (f) Histograms representing the average inter-nanoparticle distances calculated from BF TEM and HAADF STEM images respectively. Above each histogram, the sample size (N), the mean ($\bar{x}$) and the standard deviation (σ) (both in nm) are shown.

3.2.4 Iron mineralizing properties of encapsulin proteins

It has already been shown that the function of the *M. xanthus* encapsulin lies in Fe storage and the two cargo proteins EncB and EncC are labelled as the ‘ferritin-like’ proteins responsible for the Fe mineralization [18]. However, it is unclear whether all of the proteins are required for the mineralization process to occur. To test this, *E. coli* strains transformed with the overexpression plasmids were induced using 0.05mM IPTG and then 4 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ was added to the culture after a two hour expression window. The resulting purified structures were imaged without stain in a TEM to confirm the presence of mineralized Fe.

In all images, the low-contrast, poorly resolved spherical regions correspond to the encapsulin shell. Due to the absence of negative staining, these proteins generate minimal contrast, making them difficult to detect. Conversely, the Fe-rich regions are easily identified because of their higher density and atomic number results in significantly greater contrast (see Figure 3.7). Within the micrographs of EncA+EncB and EncA+EncC we can see evidence of Fe mineralization. Furthermore, it is clear that the process is occurring within the encapsulins as each high contrast region has a low contrast halo, which is indicative of the protein shell. All of the encapsulins in samples of EncA and EncA+EncD lack this increased contrast at their centre, indicating that no Fe mineralization has taken place. Interestingly, in a small number of encapsulins in both these images there appears to be a small dark core. However, due to the infrequency of this observation and the lack of consistent localization with protein rich regions, it is most likely these are dirt particles collected either in the sample or on the grid. In the micrograph of EncA, large cubic structures can also be seen. As the elution buffer has a high salt content (500 mM NaCl) these structures are most likely salt that has precipitated out of solution.

These findings indicate that EncA alone cannot perform the iron mineralizing function of the encapsulin system and the presence of either EncB or EncC is essential for this process to occur. Additionally, the inclusion of EncD inside the EncA shell does not result in Fe mineralization occurring, this is most likely because of its function as an Fe reductase which has recently been identified [97].

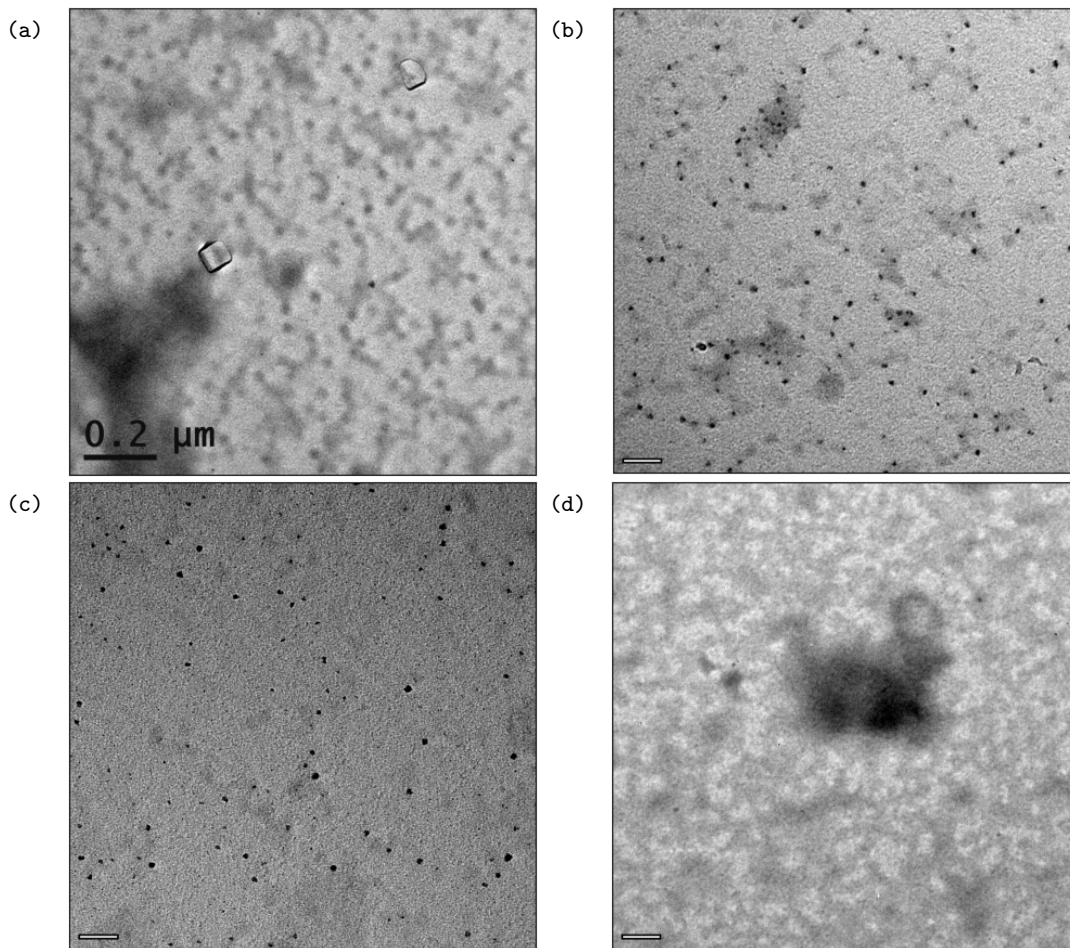


Figure 3.7: Unstained electron micrographs of encapsulins containing single cargo proteins after growth in media containing 4 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. (a) EncA, (b) EncA+EncB, (c) EncA+EncC (d) EncA+EncD. The Fe mineralized regions appear as the areas with darker contrast in the images, while the encapsulin shells are visible as the low contrast regions. Electron micrographs of unstained sample were taken at a magnification of 15.5kX on a Phillips CM100. Scale bars represent 100nm.

3.2.5 Effect of different Fe concentration in the growth media on encapsulin loading

When expressed and purified from *E.coli*, encapsulins were found to lack stored material (see Figure 3.7). This is most likely due to the fact that *E. coli* maintains its own iron homeostasis system meaning that it does not need to use the additional encapsulins for iron storage [182, 183]. Two major methods have been identified for loading Fe into over expressed encapsulins: either through the addition of Fe to the growth media during over expression [19], or adding Fe to the purified encapsulins [23]. The former was selected, because it was felt that it would more closely mimic the environment inside the *M. xanthus* cell, and thus the iron containing mineral product would more closely reflect what would be expected inside the native system. However, it is important to state that the stress on the cell during over expression and the starvation conditions required for encapsulin production in *M. xanthus* are different and this difference may factor into the observed results.

Thus, to induce the mineralization of Fe atoms into encapsulins, additional $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ was added after protein expression had been induced for 2 hours at 30 °C. It was identified that the number of shells which showed signs of iron mineralization increased as the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ in the media increased (see Figure 3.8). The largest increase was a 10-fold increase in the number of EncA+EncB shells that showed signs of Fe mineralisation, from 2% after the addition of 2 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to 21% after the addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (Figure 3.8). Additional experiments were carried out concentrations of Fe exceeding 8 mM, however, under those conditions the overexpression cultures died before any encapsulins could be purified. This is most likely because at those concentrations the additional Fe became toxic for the cells. Although the number of shells showing mineralisation increased with Fe concentration in the media, the percentage of mineralized shells remained low, most likely due to the aforementioned *E. coli* homeostasis mechanisms that would be working to remove the excess Fe from the cell limiting the amount available for the overexpressed encapsulins to interact with. Even at the highest loading condition (8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ in EcA+EncB+EncC), only a maximum of 31% of the shells imaged showed signs of mineralization. In contrast, encapsulins isolated from *M. xanthus* consistently showed 100% loading. This may be due to the expression being more tightly controlled in *M. xanthus* however, it may also be because the encapsulins were purified from a pellet at the bottom of

a CsCl gradient. As a CsCl gradient separates by density, this may have artificially selected for only highly loaded structures.

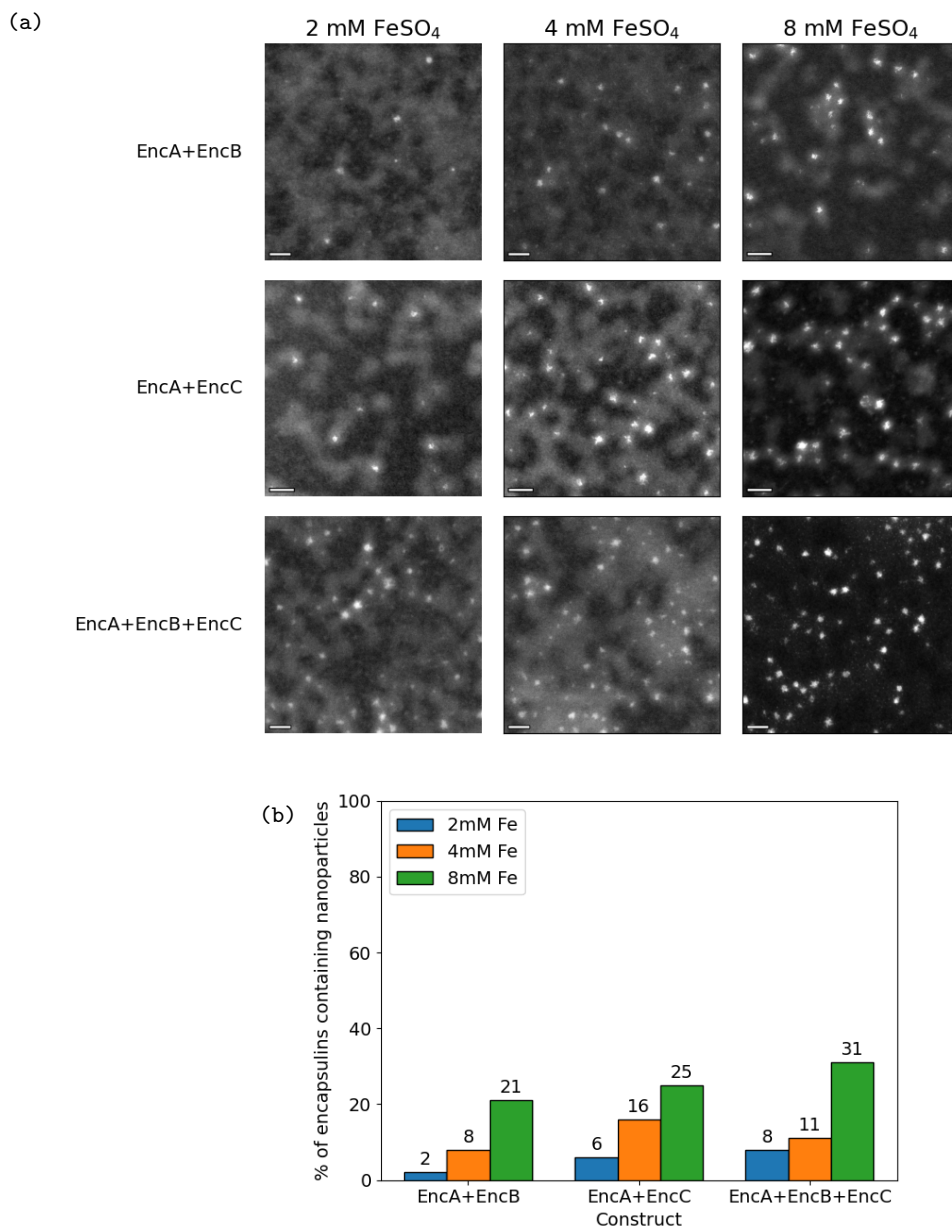


Figure 3.8: Encapsulin constructs grown in media containing different concentrations of $\text{Fe}(\text{NH}_4)\text{SO}_4$. (a) HAADF images of unstained samples from each growth condition. Scale bar represents 50 nm. The images captured in the 2 mM and 4 mM conditions have lighter backgrounds because carrier grids with thicker C films were used during data collection. (b) Bar chart comparing the number of shells that showed iron mineralization for the different constructs generated under different concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ in the media.

3.2.6 Attempted isolation of Fe loaded encapsulins using magnets

The encapsulin is an Fe storage nanocompartments and therefore depending on the material that is present in the core the encapsulins may have magnetic properties [167]. To test this, previously isolated Fe loaded encapsulins were run through a MACS magnetic separator column (Miltenyi Biotec). If the material was magnetic it would remain in the column whilst in the presence of a magnetic field, and could be eluted following its removal from the magnetic field. However, for all material tested, both the structures isolated from *M. xanthus* and those generated in *E. coli*, the encapsulins were eluted from the column in either the initial flow through and in the first wash (see Figure 3.9). In the three encapsulins EncA+EncB, EncA+EncC and EncA+EncB+EncC, there is a small band visible in the elution step running at the height of EncA (~36kDa). Although this may represent a small number of shells containing magnetic material, it is most likely due to the protein sticking to the column and not being removed completely during the wash steps. This indicates that the material inside the encapsulin is most likely non-magnetic, and that it cannot be easily isolated using the MACS magnetic separator system.

This result significantly impacted the overarching scope of the project, as it was among the initial characterization experiments conducted to test the original hypothesis. We initially hypothesized that the nanoparticles inside the *M. xanthus* encapsulin might possess magnetic properties, based on previous observations in the literature [18]. This possibility was particularly interesting since magnetic nanoparticles of this size possess a myriad potential applications in electronic engineering [44]. However, the negative result from this experiment prompted a shift in the thesis aims: moving away from identifying potential engineering applications for the encapsulin derived nanoparticles and toward characterizing the material naturally present in these structures and understanding the contribution of the different cargo proteins to the generation of this material.

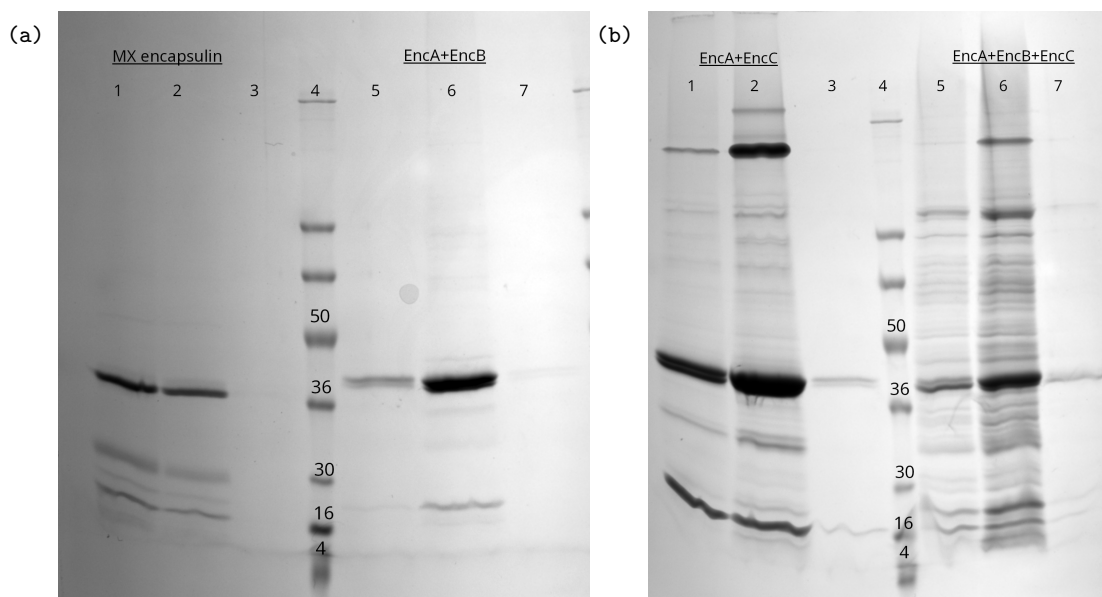


Figure 3.9: Magnetic separation of encapsulins loaded with Fe. SDS gels showing the results of testing the magnetic properties of encapsulin derived nanoparticles using MACS cell separator columns. **(a)** *M. xanthus* encapsulins (left) and EncA+EncB (right). **(b)** EncA+EncC (left) and EncA+EncB+EncC (right). Both gels have 7 lanes that contain the following samples: Lane 1 and Lane 5: Flow through, Lane 2 and Lane 6: wash step, Lane 3 and Lane 7: elution step, Lane 5 contains the See Blue Prestained protein ladder. The bands representing 50kDa, 36kDa, 30kDa, 16kDa, 4kDa are labelled.

3.2.7 Comparison of core sizes at different Fe concentrations

Both the EncB and EncC cargo proteins are ferritin-like proteins and both are essential for mineralizing Fe inside the encapsulin shell (see Figure 3.7). However, it remains unclear why *M. xanthus* has two proteins with the same proposed function distally located on the genome [18, 28, 23, 87]. One possibility is that the cargo proteins mineralize Fe at different rates, and thus can be implemented differently depending on the needs of the cell. To test this, the diameters of the Fe containing cores were measured from HAADF images following expression in media containing $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (2 mM, 4 mM and 8 mM).

In all three tested samples — encapsulins containing only EncB, encapsulins containing only EncC, and encapsulins containing both EncB and EncC — the average size of the Fe containing core increased as the concentration of Fe added to the media increased, indicating, that this is a valid method for controlling Fe loading into encapsulins whilst they are being overexpressed in *E. coli*. The diameter of the cores of EncB containing encapsulins were measured at 9.6 ± 5.0 nm ($N=115$), 11.3 ± 2.5 nm ($N=372$) and 14.8 ± 2.4 nm ($N=413$) for the 2 mM, 4 mM and 8 mM conditions respectively (Figure 3.10). In comparison, the core regions of the EncC containing encapsulin were measured as 8.7 ± 3.2 nm ($N=228$), 11.1 ± 3.9 nm ($N=128$) and 13.8 ± 2.9 nm ($N=228$) for the 2 mM, 4 mM and 8 mM conditions respectively. No significant difference was observed in core sizes between encapsulins containing either EncB or EncC under any of the tested Fe concentration growth conditions. Furthermore, when both cargo proteins are present the core size measured was 10.2 ± 2.3 nm ($N=1109$), 10.5 ± 3.8 nm ($N=250$) and 14.8 ± 2.4 nm ($N=414$) for the 2 mM, 4 mM and 8 mM growth conditions respectively (Figure 3.10). The differences in core size between encapsulins containing either EncB or EncC and encapsulins containing both EncB and EncC was also found to not be statistically significant.

Given the lack of a significance in size differences across the different concentrations of Fe added, it is likely that the differences between individual conditions are not sufficiently representative of a real difference between the iron mineralizing properties of these proteins. Therefore, it is likely that there is no real difference between the total amount of Fe mineralized by either the EncB or EncC proteins under these conditions. Additionally, there is no increase in core size observed by having both cargo proteins present in the sample.

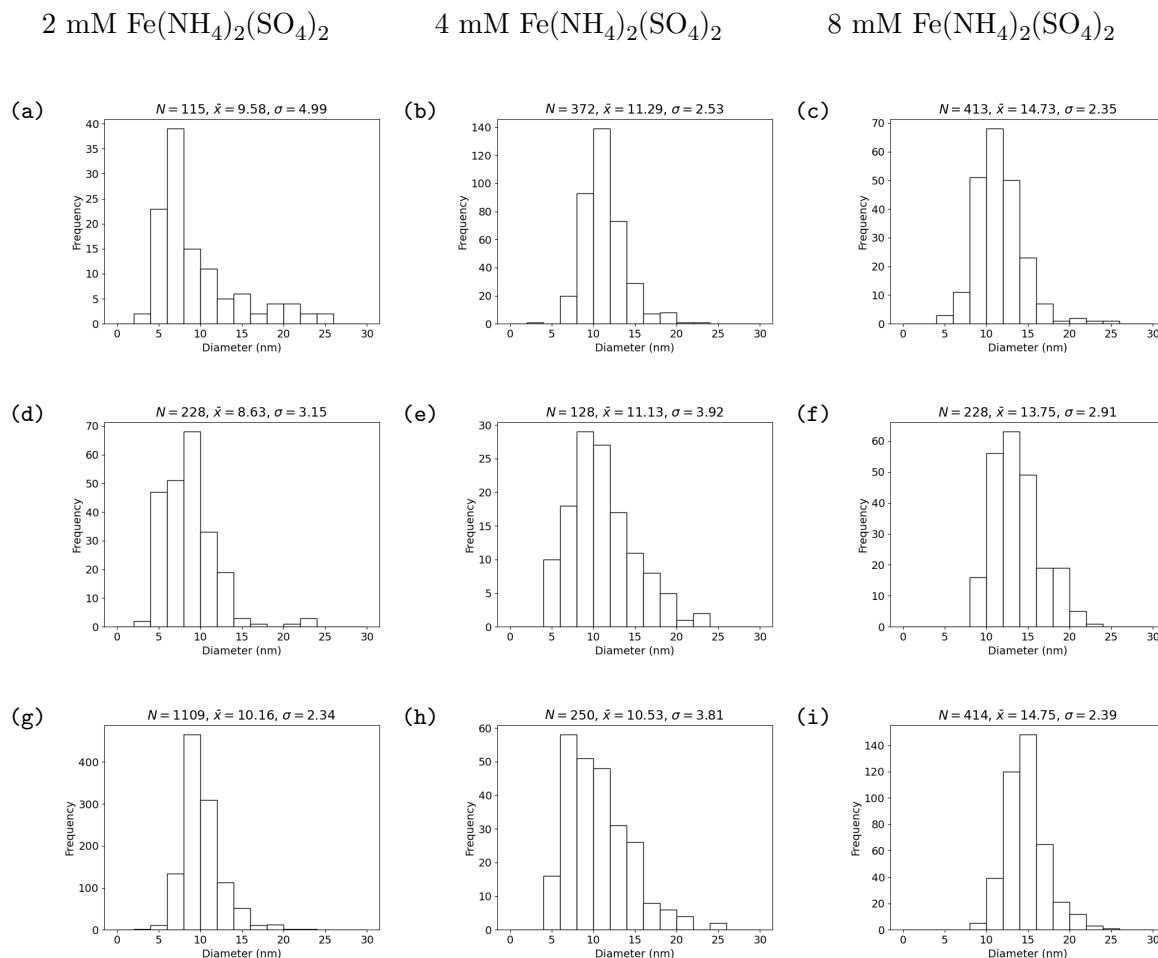


Figure 3.10: Histograms representing the average core size measured inside encapsulins containing different combinations of cargo proteins after growth in media supplemented with different concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. Particle diameter was calculated from binary images generated using a multi-Otsu threshold from HAADF images [165]. (a) - (c) Histograms from samples of EncA+EncB. (d) - (f) Histograms from samples of EncA+EncC. (g) - (i) Histograms from samples of EncA+EncB+EncC. For all samples, core diameter after growth in 2 mM, 4 mM and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ are shown (left to right). The number of cores measured (n), the sample mean ($\bar{x}$) and standard deviation (σ) for each sample are shown above each histogram.

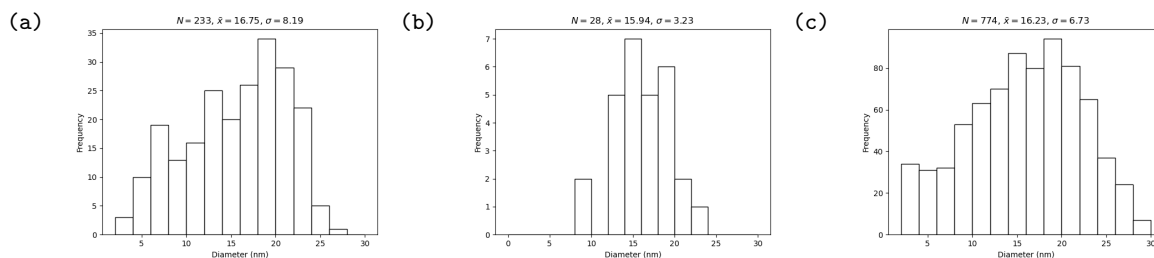


Figure 3.11: Largest diameter cores observed from encapsulins expressed in *E. coli*. Encapsulins were induced using 0.01 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ was added to the media after 2 hours. Particle diameter was calculated from binary images generated using a multi-Otsu threshold from HAADF images [165]. (a)-(c) Histograms representing the average core size measured for EncA+EncB, EncA+EncC and EncA+EncB+EncC respectively. The number of cores measured (n), the sample mean ($\bar{x}$) and standard deviation (σ) for each sample are shown above each histogram.

3.2.8 Largest diameter cores achieved in *E. coli* expressed encapsulins

After growth in media containing the highest concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (8 mM), the Fe containing cores in *E. coli* were ~ 10 nm smaller than those observed in the complete *M. xanthus* encapsulin (see Sections 3.2.2 and 3.2.7). This difference is indicative that the growth conditions used may not have been optimal for Fe loading. One reason for this may be that protein overexpression from the pETDuet-1 plasmids is controlled by the viral T7 promoter, which is several times more productive than standard *E. coli* promoters, resulting in large numbers of encapsulins within a single *E. coli* cell [168]. This is especially important given the low number of encapsulins observed in thin sections of *M. xanthus* cells [18]. In an attempt to achieve core sizes more similar to those observed in the native host, encapsulin expression was induced with low levels of IPTG (0.01 mM) to reduce the number of structures per *E. coli* cell. The result of this was an increase in core size compared with previous growth conditions in *E. coli*; 16.8 ± 8.2 nm ($N=233$) for EncA+EncB, 15.9 ± 3.2 nm ($N=28$) for EncA+EncC and 16.2 ± 6.7 nm ($N=774$) for EncA+EncB+EncC (Figure 3.11). However, under these conditions the cores were still significantly smaller than those measured inside the *M. xanthus* encapsulins. Indicating that this growth condition is also insufficient for restoring the core size observed in the native encapsulins. Notably, the number of encapsulins sampled for EncA+EncC was substantially lower than all other samples because of the low sample density on the grid.

3.2.9 EncB and EncC synthesise nanoparticles with the same diameters

In all electron micrographs of unstained encapsulin samples that had been overexpressed in *E. coli* captured on a TEM, the cores appear to consist of a single irregularly shaped nanoparticle (for an example see Figure 3.7). This observation led to the initial idea that the encapsulins containing single cargo proteins were unable to produce the nanoparticles observed in structures isolated from *M. xanthus*. However, when the samples were imaged with HAADF STEM at higher spatial resolutions, it was revealed that the protein shells were filled with individual nanoparticles. The nanoparticles were measured and found to have an average size of ~ 2 nm, which is consistent with the size of the nanoparticles observed inside the cores of the encapsulins isolated from *M. xanthus* (see Section 3.2.10). The nanoparticles from encapsulins containing only EncB had a diameter of 1.9 ± 1.0 nm ($N=813$) and the nanoparticles from the encapsulins containing only EncC had a diameter of 1.9 ± 1.0 nm ($N=548$) (Figure 3.12). The difference between these two was found to not be significant at the $p=0.05$ level using an unpaired t-test. This result suggests two important facts: firstly, that EncB and EncC synthesis nanoparticles that are the same size, and that the presence of a single cargo protein inside the shell is necessary and sufficient for nanoparticle formation and is not reliant on the presence of all the cargo proteins.

3.2.10 Observation of individually distributed ~ 2 nm nanoparticles

In a small number of isolations very small bright features were observed distributed across the sample in HAADF STEM imaging (see Figure 3.13a). The bright features were spherical and appear to be similar to the nanoparticles observed inside the densely filled encapsulins. When measured, the features had an average diameter of 2.0 ± 1.0 nm ($N=9775$). This diameter was not significantly different from those measured for nanoparticles synthesized by either EncB or EncC (see Section 3.2.9). Furthermore, greater than 98% of the features were found associated with the protein rich region of the sample suggesting they are encapsulin associated (see Figure 3.13b). Samples that showed these nanoparticles were dialysed against the elution buffer for 48 hours. If the nanoparticles were freely dispersed in the solution, it is expected that they would be removed during this process. Conversely, if the nanoparticles were encapsulated within an encapsulin shell, they would not be lost during the dialysis process. This has been shown to work, because several dialysis steps were used

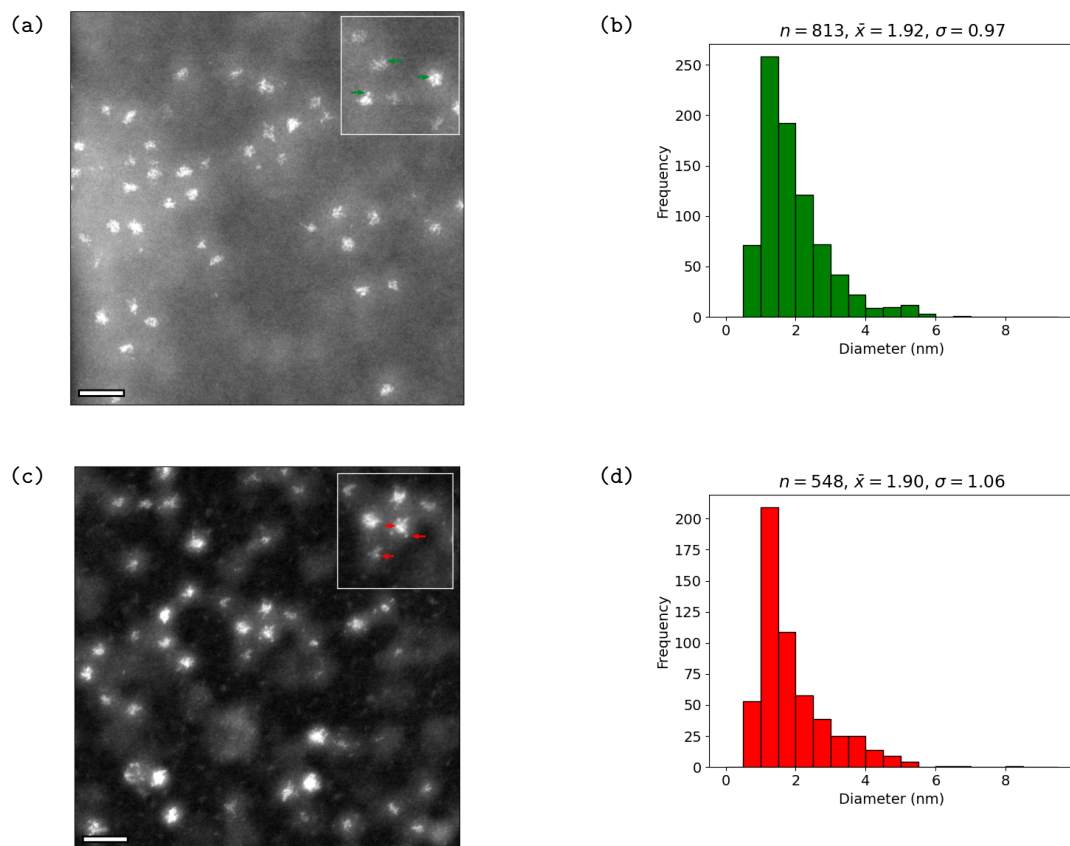


Figure 3.12: Nanoparticle diameters measured for encapsulins overexpressed in *E. coli*. (a) HAADF reference image of EncA+EncB. An insert is included where a number of individual nanoparticles are highlight by green arrow. (b) Histogram of average particle size measured from samples of EncA+EncB. (c) Reference HAADF image of EncA+EncC. An insert is included where a number of individual nanoparticles are highlight by red arrow. (d) Histogram of average particle size measured from samples of EncA+EncC. All reference images were captured on a JEOL F200 at 250kx magnification. Scale bars represent 50 nm. Above each histogram the number of particles measured (n), the mean ($\bar{x}$) and standard deviation (σ) (both in nm) are given.

during the isolation of encapsulins from *M. xanthus* without loss of the internal nanoparticles. Examination of the sample following the dialysis did not result in an observable loss of the distributed nanoparticles, providing more evidence that they are associated with the inside of the encapsulin. Taken together, these pieces of evidence are good indicators that the bright features are nanoparticle associated with the encapsulin proteins in the sample. Thus, there may be encapsulins within this sample that contain either a single or a couple of nanoparticles, most likely due to low levels of iron inside the *E. coli* cell. However, this phenomena could not be repeated consistently.

3.2.11 Comparing the HAADF intensity and the size of the Fe rich region in samples isolated from *M. xanthus* with those isolated from *E. coli*

Due to the Z-contrast mechanism of HAADF STEM imaging, the brightness of each pixel is strongly dependent on both the atomic number (Z) of the atoms that are present and their density [125]. In these samples, Fe has the highest atomic number of all atoms that are present and thus is responsible for the majority of the image contrast. Therefore, the intensity of HAADF contrast in the Fe rich regions can be correlated with the number of Fe atoms present (see Section 5.2.8). The intensity of the Fe rich regions of encapsulins isolated from *M. xanthus* and those isolated from *E. coli* following growth in media containing 8 mM of Fe were compared. All data was collected in the same session and the HAADF detector set up was not altered between the imaging of the different samples.

The resulting plots of average intensity against core diameter reveal several key findings. Firstly, in the sample isolated from *M. xanthus* the distribution in diameter and intensity appears to be tightly regulated, whilst in the samples containing either EncB, EncC or both of these proteins, there is much greater variation in both the average intensity and diameter, with the widest variation in intensity being observed in EncA+EncB+EncC (see Figure 3.14). Secondly, the HAADF intensity of EncA+EncB, EncA+EncC is not significantly different from the average intensity of encapsulins from *M. xanthus* ($p=0.05$) (see Table 3.4). Similar contrast in a smaller region is indicative of a higher density of Fe atoms in these regions. There are multiple explanations for this: the nanoparticle spacing that is present in isolates from *M. xanthus* is absent in these structures, resulting in more nanoparticles being

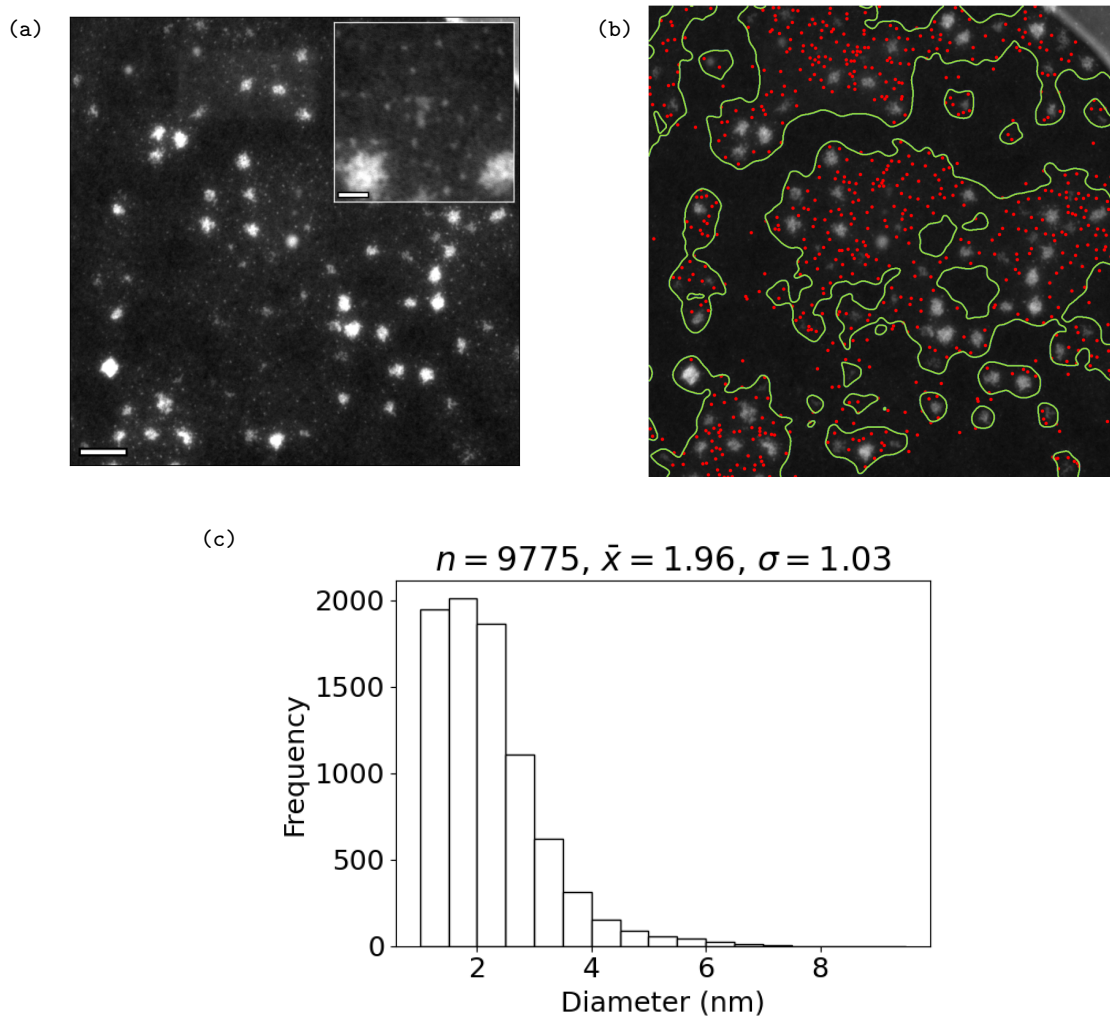


Figure 3.13: Observation and localisation of individual ~ 2 nm nanoparticles. Images of a sample of EncA+EncB+EncC where individual nanoparticles can be observed. **(a)** Reference HAADF image, the contrast has been enhanced for easier observation of individual nanoparticles. The insert shows a high magnification region where the individual ~ 2 nm nanoparticles can be clearly seen. Scale bar represents 10 nm. **(b)** The same image with the individual nanoparticles labelled with red and the boundary of the encapsulin proteins in the image labelled with green. The boundary does not perfectly track the expected shape of the protein shell, this is because of the Gaussian filter (width 2 pixels) applied to the image to reduce noise. The localisation of both features to the same image region is clearly evident. **(c)** Histogram representing the average particle diameter. The number of particles measured (n), the mean ($\bar{x}$) and standard deviation (σ) (both in nm) are given above the histogram.

CHAPTER 3. ELECTRON MICROSCOPICAL INVESTIGATION OF ENCAPSULIN DERIVED NANOPARTICLES

Sample	$\bar{x}$	σ	N
<i>M. xanthus</i> encapsulin	0.7	0.1	860
EncA+EncB	0.7	0.2	218
EncA+EncC	0.7	0.2	228
EncA+EncB+EncC	0.5	0.2	341

Table 3.4: Table comparing the HAADF intensity of the Fe rich cores of different encapsulin constructs from normalised HAADF images. The mean ($\bar{x}$), standard deviation (σ) and number of cores measured (N) are given.

arranged into a smaller area, or that the material could be significantly structurally different resulting in higher numbers of Fe atoms per unit area. Finally, the average intensity values from EncA+EncB+EncC were found to be statistically significantly different from all other constructs imaged ($p=0.05$). The reason for this is unclear, but as there are two cargo proteins being expressed to the singular shell protein this may lead to improper loading of cargo into shells, and increased iron mineralization by free cargo proteins inside the *E. coli*. This may occur because the encapsulin cargo proteins are known to still perform their catalytic function without the shell being present [23]. Leading to a reduction in the amount of free ferrous iron that can be loaded into the encapsulin and thus smaller less intense cores on average. This result has implications for our understanding of the encapsulin system. Namely, that when encapsulins are expressed without the full set of cargo proteins the resulting iron filled core regions are not as tightly regulated as in the native structures where all of the cargo proteins are present. As the cores from encapsulins with fewer cargo proteins showed much higher variability in terms of core size and intensity in HAADF images. Meaning that the cargo proteins may play an axillary role in maintaining an arrangement of nanoparticles inside the encapsulin.

3.2.12 Loading encapsulins *in vitro* with increasing concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$

The addition of Fe to the growth media of *E. coli* cells expressing encapsulins is important for assessing the Fe loading characteristics whilst retaining the cellular context. However, loading encapsulins with metal outside of the cellular context provides more con-

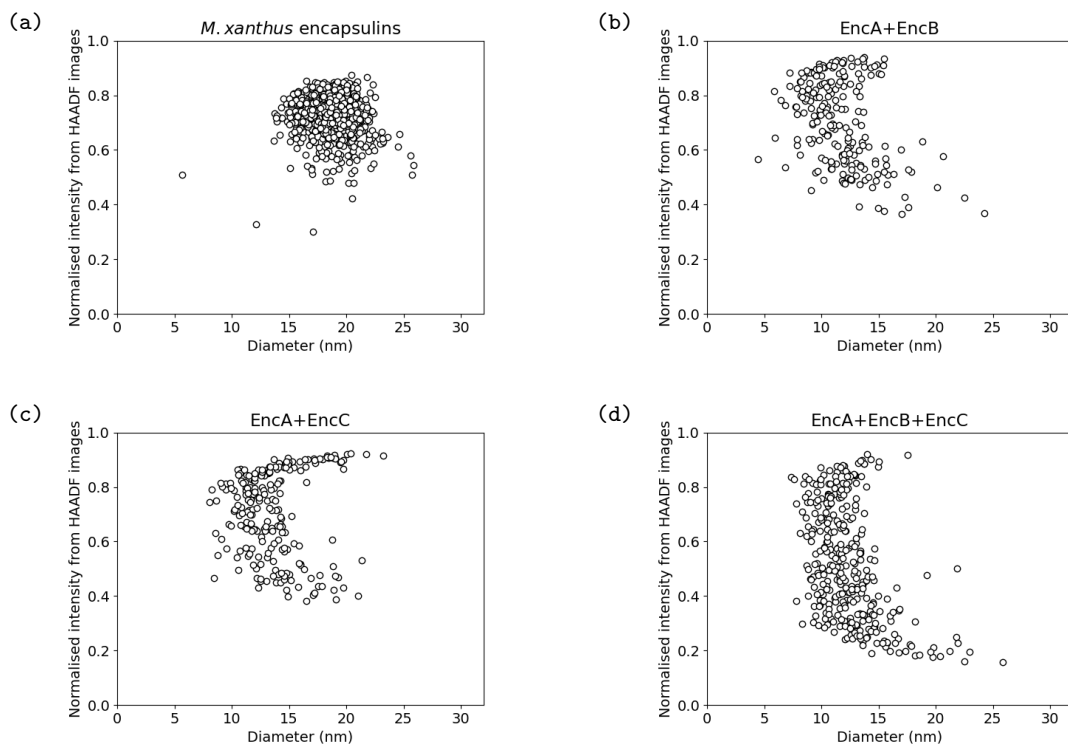


Figure 3.14: Plots showing the intensity of the core regions from HAADF images against the diameter of the core. The constructs shown are: (a) *M. xanthus* encapsulins, (b) EncA+EncB, (c) EncA+EncC, (d) EncA+EncB+EncC. The three samples shown in (b)-(d) are from encapsulins expressed in *E. coli* and grown in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The HAADF image contrast was normalised by dividing the intensity value of each pixel in the image by the highest intensity pixel.

trol and can result in the formation of materials with distinct structures [169, 170]. Empty EncA+EncB+EncC encapsulins were purified (the lack of mineralization was confirmed using TEM) and combined with various concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (see Figure 3.15). Initially, an Fe concentration of 0.25 mM was used based on similar experiments carried out by other groups [23, 97], however, when visualized in TEM the cores had drastically lower contrast than had been observed inside encapsulins loaded *in vivo*. Thus, higher concentrations of Fe were utilized, up to 8 mM of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The clear increase in contrast in the encapsulin cores compared with the surrounding protein suggests that mineralization can be induced in the EncA+EncB+EncC encapsulin by the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, and that the amount of mineralization increases with the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ used (see Figure 3.15). The low contrast of the mineralized Fe after the addition of 0.25 mM of Fe compared with the protein shell in TEM images resulted in increased difficulty in separating these two image regions. This difficulty may explain why the average size of the Fe rich regions were larger in the 0.25 mM condition compared with the 1 mM condition.

3.2.13 Size measurements of nanoparticles expressed *in vitro*

As with the other samples generated in this work, the resolution from the basic TEM instrumentation available was insufficient for identifying if the core region consisted of a singular or multiple nanoparticles, thus, the encapsulins were imaged using HAADF-STEM. The use of this method, revealed that, as with the encapsulins from *M. xanthus*, the protein shell was filled with nanoparticles (see Figure 3.16). Furthermore, the diameters of these nanoparticles were measured at 2.1 ± 0.6 nm ($N=102$). Accounting for measurement errors when segmenting the nanoparticles from the background, this is the same as the nanoparticle sizes measured under all other conditions. Indicating that the nanoparticle formation is not affected by the conditions under which Fe is introduced to the encapsulins.

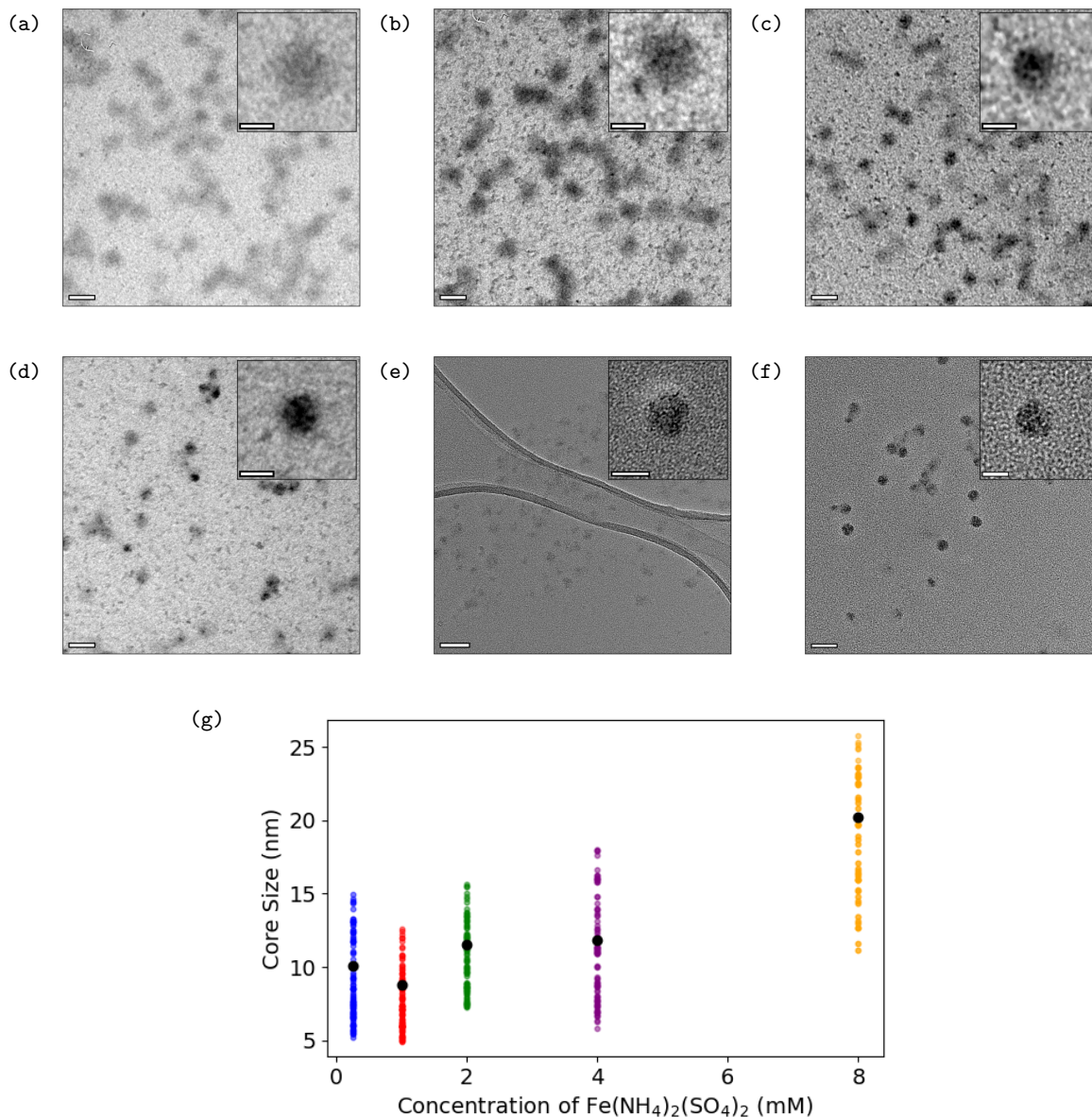


Figure 3.15: Loading encapsulins containing EncB and EncC with different concentrations of Fe *in vitro*. The following loading conditions were used: (a) 0 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, this is a negative control showing the empty protein shells. (b) 0.25 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (c) 1 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (d) 2 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ (e) 4 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and (f) 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The insert in each image is a representative example of a single shell for each growth condition, the scale bars all represent 20 nm. The increase in contrast of the core region as the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ increases can be clearly seen in these BF TEM images. The micrographs (a) - (d) were captured on the Tecnai Spirit at 100 kx and micrographs (e) - (f) were captured on the JEOL F200 at 100 kx. The scale bars in all images represent 50 nm. (g) Graph of the average core size measured from BF TEM images against the concentration of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. The data for each loading condition is represented by a colour: blue (0.25 mM), red (1 mM), green (2 mM), purple (4mM) and orange (8mM). For each growth condition, 100 core diameter measurements are represented.

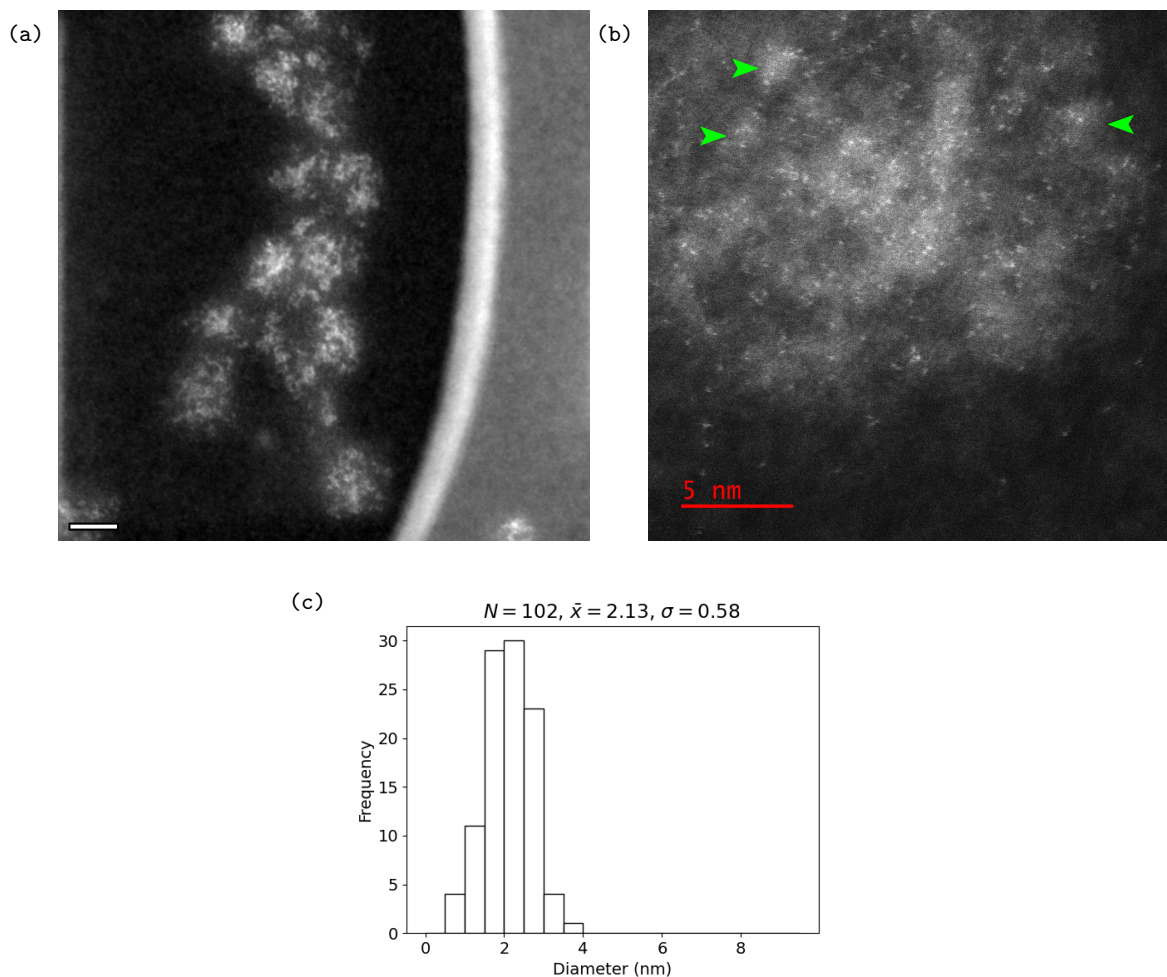


Figure 3.16: The diameter of nanoparticles inside encapsulin shells supplemented with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. (a) HAADF reference image of sample. The image was captured using a JEOL F200 at a magnification of 1Mx. The scale bar represents 20 nm. (b) Higher magnification HAADF image of single encapsulin captured on a FEI Titan G2 CHEMISTEM at 3.6Mx magnification. Individual 2 nm nanoparticles have been highlighted with green arrows. The scale bar represents 5 nm. (c) Histogram representing the distribution of measured particle diameters. Above it, the sample size (N), the mean ($\bar{x}$), and the standard deviation (σ) (both in nm) are quoted.

3.2.14 4D-STEM investigation of the nanoparticles inside the *M. xanthus* encapsulin

Despite it being known that the nanoparticles inside the encapsulin contain Fe, to date no investigation has been made into the atomic structure of these particles [18]. Thus, a sample of encapsulins isolated from *M. xanthus* was investigated using the new Tescan Tensor STEM at the ER-C in Juelich, Germany. In the vast majority of images captured, no diffraction was observed. All that could be seen were the diffuse rings from the C film on the grid. However, in some of the probe positions, faint partial spots could be seen; they were infrequent (only around 3 probe positions in the dataset of 40,000 probe positions) and only marginally above the background (see Figure 3.17) meaning that it is not abundantly clear if they are proper diffraction spots or noise. This result indicates that the majority of the nanoparticles inside the *M. xanthus* encapsulin are most likely amorphous and, at most, have some near-field order but no crystalline microstructure.

Due to the machine setup, the spatial resolution is limited to approximately 1 nm. Since the nanoparticles inside the encapsulin are less than 2 nm in diameter, this resolution limit likely prevents proper sampling of the nanoparticles, possibly explaining the absence of observable diffraction. However, since diffraction was observed in a few probe positions, it is possible that the nanoparticles contain crystalline regions smaller than 1 nm, which could explain the limited detection of diffraction spots.

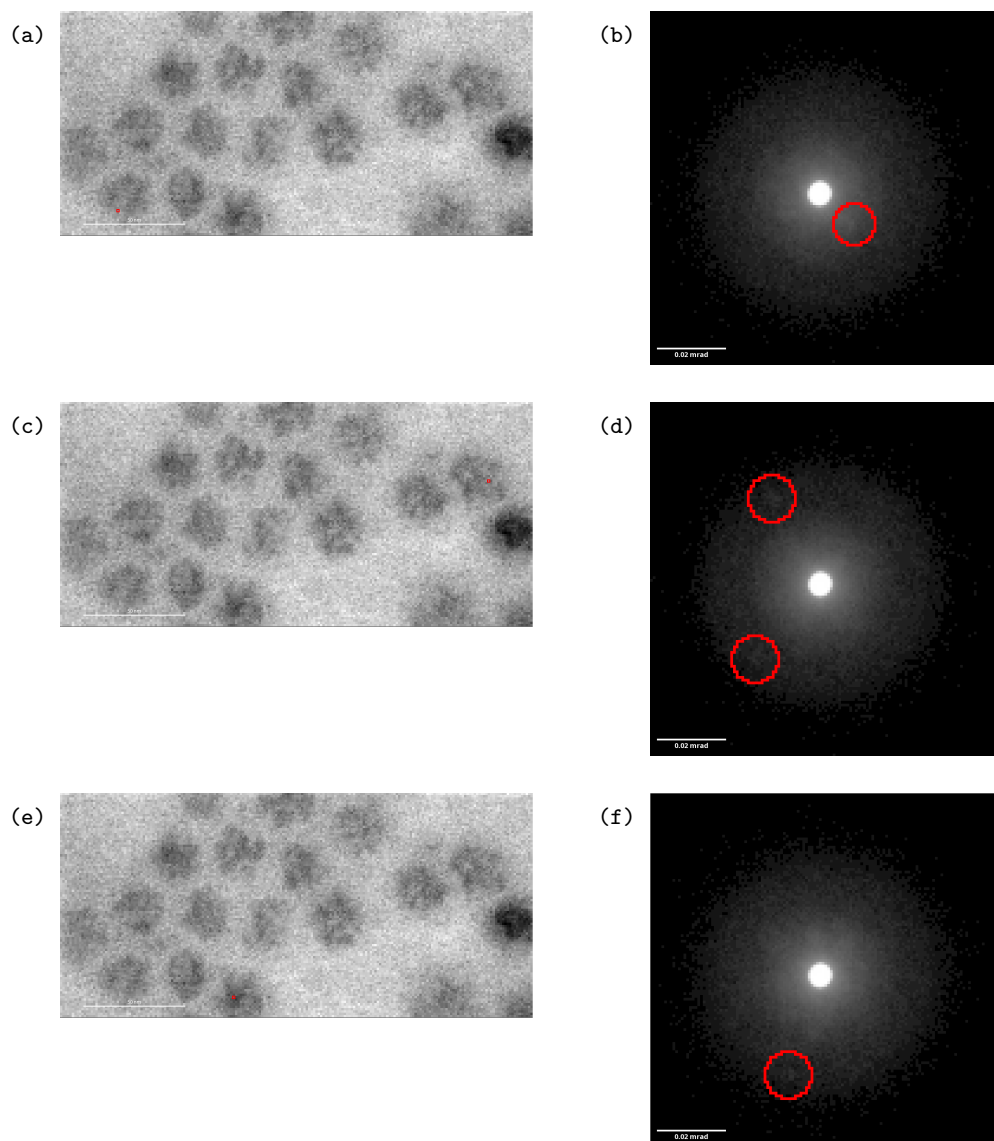


Figure 3.17: 4D-STEM investigation of the encapsulins isolated from *M. xanthus*. The images are paired, (a)/(b), (c)/(d) and (e)/(f). The first image in the pair is a reference BF image of the investigated region. Highlighted in red is the probe position from which the diffraction pattern in the second image is taken from. The scale bar represents 50 nm. In the diffraction pattern, partial diffraction spots are highlighted with red circles. Scale bar represents 0.02 mrad.

3.2.15 High Resolution STEM imaging of the *M. xanthus* encapsulin

To check if the nanoparticles inside the encapsulin of *M. xanthus* showed sub-nm sized regions with crystalline arrangement, the material was imaged using an aberration corrected STEM (JEOL ARM200F) as this gave a spatial resolution of ~ 0.1 nm which is sufficient to see potential crystalline regions along many different viewing directions. Both BF and HAADF images were captured, however, due to the relatively low sample thickness it was difficult to distinguish between the background and nanoparticles in the BF images (see Figure 3.18a). Additionally, it appears that there are regions with apparent regular arrangements in the BF that are not present in the HAADF image making it difficult to distinguish crystalline regions from artefacts, due to the complexity of the contrast transfer function in BF imaging. As a result of these concerns, the HAADF image exclusively was used for further analysis.

The nanoparticles inside the encapsulin are super imposed on each other. The ramification of this arrangement, is enhanced difficulty in separating columns of atoms in one nanoparticle in the top layer from the atoms in the nanoparticle beneath it. However, it was still possible to identify sample regions containing regular structures (see Figure 3.18c). These regions contained low numbers of lattice fringes, 3-7, and did not exceed 1.5 nm in size. A number of different inter-planar spaces were measured from the nanoparticles inside the *M. xanthus* encapsulins: 160 ± 8 , 195 ± 9 , 215 ± 11 and 242 ± 12 . A single iron oxide phase could not be identified to explain these measurements which could be explained as ferrihydrite ($\text{Fe}_2\text{O}_3 \cdot 0.5\text{H}_2\text{O}$), haematite ($\alpha\text{Fe}_2\text{O}_3$), iron(III) phosphate (αFePO_4) and magnetite (Fe_3O_4) (see Appendix 3.4.4).

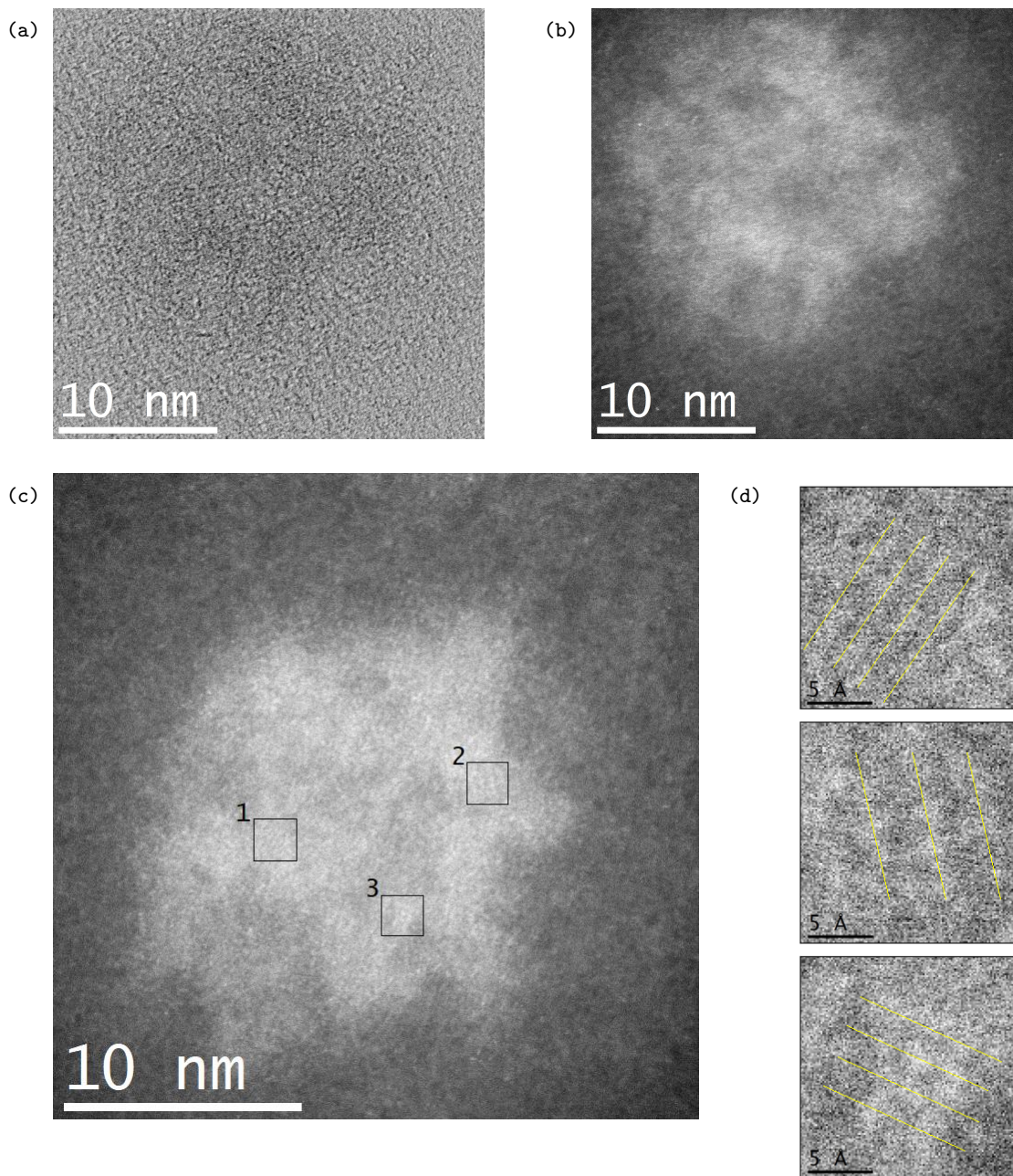


Figure 3.18: HR-STEM imaging of the encapsulin from *M. xanthus*. (a) Reference BF image of encapsulin. (b) Reference HAADF image of encapsulin. (c) HAADF image of encapsulin where three potentially crystalline regions are highlighted in boxes and labelled. (d) Higher magnification images of the three identified ROIs 1-3, top to bottom. The lattice plans have been highlighted with yellow lines. D -spacings from top to bottom: 203 ± 8 pm, 242 ± 12 pm and 168 ± 8 pm. Scale bar represents 5 Å.

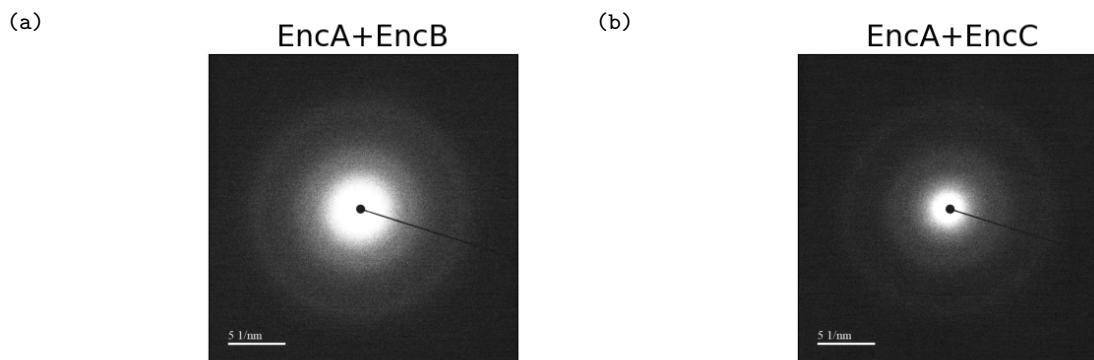


Figure 3.19: SAED from core regions of encapsulins containing either EncB or EncC. Diffraction patterns from cores of (a) EncA+EncB and (c) EncA+EncC. Images were captured on a JEOL F200 at 200keV, with a camera length of 800 mm and the 50 μm selected area aperture inserted. The diffraction patterns represents a densely loaded 800x800 nm area of the grid. In both images, only a diffuse rings representing the amorphous material of the carbon film can be seen.

3.2.16 Structural characterization of nanoparticles from encapsulins over-expressed in *E. coli*

Following investigations into the crystallinity of nanoparticles synthesized within encapsulins from *M. xanthus*, the crystallinity of nanoparticles from overexpressed encapsulins from *E. coli* were investigated using Selected Area Electron Diffraction (SAED). A selected area aperture was inserted and centred over grid regions dense in nanoparticles. For these experiments, encapsulins isolated from media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ were used as they had the largest cores and thus were more likely to produce a signal. However, no diffraction was observed from any of the materials imaged, only diffuse rings (see Figure 3.19). The presence of these diffuse rings strongly suggests the nanoparticles are amorphous in nature. However, as with the material inside the encapsulins isolated from *M. xanthus*, there may be small crystalline regions in the material that cannot be identified using this method.

3.2.17 HR-STEM investigation of EncA+EncB+EncC loaded *in vitro*

To identify if the samples of empty encapsulins isolated from *E. coli* and loaded with 8 mM of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ *in vitro* contained crystalline material, samples were examined using HR-STEM HAADF imaging on the spherically aberration corrected FEI Titan G2 ChemiS-TEM at the Ernst Ruska Centre in Juelich, Germany. The nanoparticles synthesized using this method exhibited some crystalline characteristics. The largest continuous crystalline region observed had a length of 2.6 nm (see Figure 3.20). Measurements taken from nanoparticles inside ten different encapsulin shells revealed a range of inter-planar spacings.: 176 ± 8 pm, 204 ± 8 pm, 224 ± 11 pm, 278 ± 13 pm and 300 ± 15 pm. The measured inter-planar spacings with error could be interpreted as only a single iron oxide phase: Goethite ($\alpha\text{-FeO}(\text{OH})$) (see Appendix 3.4.5).

Due to the complex arrangement of the nanoparticles within the shell, it was difficult to identify the boundaries of individual nanoparticles. Additionally, scan artefacts caused some features to appear horizontally elongated. As a result, it was not possible to determine whether the individual nanoparticles were mono-crystalline or poly-crystalline.

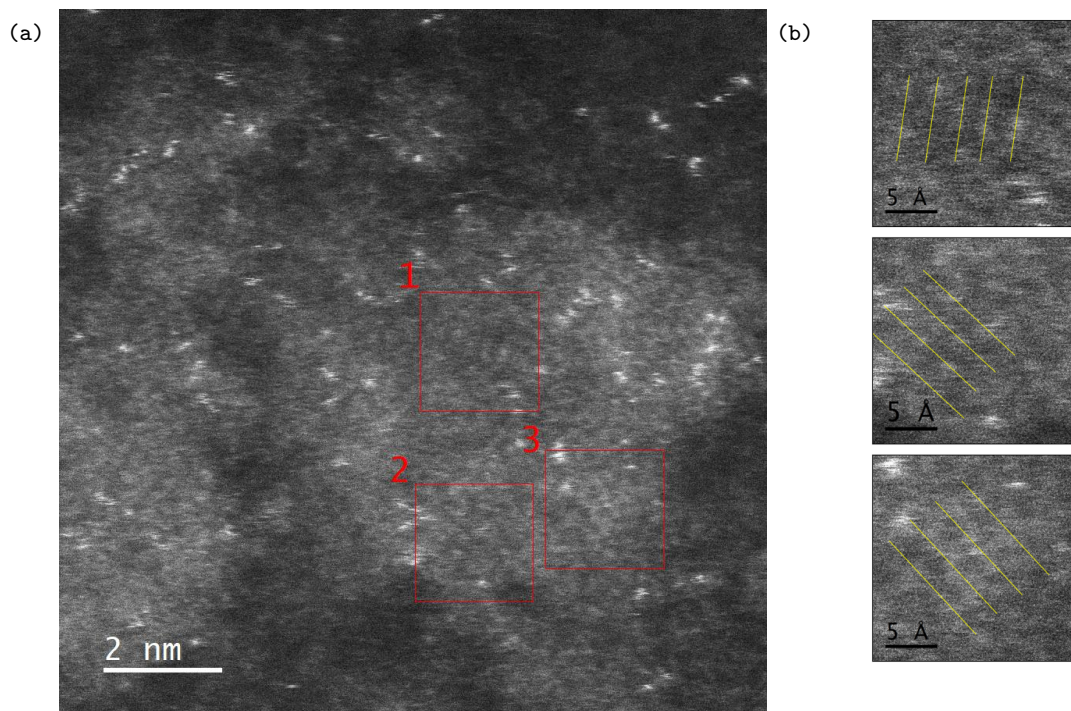


Figure 3.20: HR-STEM images of encapsulins containing the cargo proteins EncB and EncC loaded with 8 mM of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. (a) HAADF image of an encapsulin captured using a FEI Titan G2 80-200 ChemiSTEM operating at 200 keV. The image was captured at 7.2Mx magnification and the scale bar represents 2 nm. The three ROIs highlight regions that show a regular arrangement of atoms. The bright spots in the image are probably individual Fe atoms. As they appear across the whole image they are most likely not associated with the internal nanoparticles. Several image features appear to be horizontally elongated; these are most likely scan artefact together with sample drift during data collection. This scan artefact is a known issue with this microscope set up at higher magnifications. (b) Higher magnification images of the three ROIs labelled in (a). From top to bottom the images are ROIs 1-3. On each image the lattice planes are highlighted with yellow lines. The scale bars represent 5 Å.

3.3 Discussion

This work analyses the structure, size, and distribution of nanoparticles within the encapsulin of *M. xanthus* using various imaging techniques. Additionally, the structure, size, and distribution of nanoparticles formed by encapsulins expressed in *E. coli* were investigated. These encapsulins contained reduced numbers of cargo proteins and were loaded with Fe through various methods. Using a combination of SE and HAADF imaging the dimensions of the encapsulin protein shell were calculated to compare the results from these images with previous investigations. It was identified that the protein shell has a thickness of ~ 3 nm and that the Fe rich core can have a maximum diameter of ~ 26 nm. These measurements agree with the measurements made previously from the atomic model of the encapsulin shell [18]. Despite not regularly being used due to the high beam damage suffered by biological material when exposed to a STEM probe [171, 126], both of these methods proved useful for measuring protein dimensions, despite the lack of high spatial resolution. Especially SE techniques, as the inner edge of the encapsulin protein shell could be identified in the images allowing for much easier measurements of shell thickness, adding a useful technique to the investigation of encapsulins as well as other protein complexes.

Secondly, the encapsulins isolated from *M. xanthus* as part of this study were filled with nanoparticles as was expected based on observations that had previously been made [18]. However, the nanoparticles inside the encapsulins isolated here had significantly smaller diameters than those measured by other groups: ~ 2 nm measured using both BF TEM and HAADF STEM compared with ~ 5 nm measured using cryo-electron tomography [18]. Due to the disparity between these two observations, measured nanoparticle diameters from *M. xanthus* encapsulins isolated in this study were confirmed by using HAADF STEM alongside BF TEM. Contrast in HAADF imaging is roughly proportional to the atomic number (Z) of the atom present squared and as Fe (major component of the nanoparticles) has a higher atomic number than C (major component in the background) it provides more easily interpretable images [172, 173]. This has been shown to result in easier detectability of nanoparticles with diameters < 3 nm when using ADF compared with BF, thus providing a possible explanation for the difference in nanoparticle size observed here and in previous publications [174]. The higher detectability of small nanoparticles in ADF techniques may also explain the large dif-

ference in the average number of nanoparticles identified per shell between the isolations made here and those report in the literature As in this study an average of ~ 650 nanoparticles were identified per shell compared with the ~ 15 identified in previous works [18]. The observed 2 nm nanoparticle diameters were further supported by the observation of similar size nanoparticles inside all of the encapsulins overexpressed in *E. coli*. Regardless of the cargo protein that was present (either EncB, EncC or both of these proteins) or the conditions under which the Fe was loaded into the encapsulins (either by the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the growth media or the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to isolated empty encapsulin proteins). Combined, these results indicate that the nanoparticles inside the encapsulin from *M. xanthus* are most likely ~ 2 nm in diameter rather than the 5 nm that was previously measured. loaded

Thirdly, diffraction experiments identified that the nanoparticles isolated from the *M. xanthus* encapsulins and those isolated from encapsulins expressed in *E. coli* after the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the growth media were amorphous. However, HR-STEM experiments on the *M. xanthus* encapsulin revealed that small regions (< 1.5 nm in size) of some nanoparticles exhibit regular atomic arrangement, suggesting that some crystallization occurs within the encapsulin shell. However, a single iron oxide phase was insufficient for explaining all of the measured d -spacings. The observation of multiple phases is not uncommon in ferritin-derived materials [175, 176]. Thus, it makes sense that the nanoparticles inside the encapsulin would also show this characteristic as they are synthesised by ‘ferritin-like’ proteins. The amorphous nature of the nanoparticles may be important for the eventual release of Fe from the encapsulin, as it has been hypothesized that amorphous materials can be more readily mobilized under Fe-limited conditions [177, 178]. This is important biologically given the encapsulin is upregulated in response to starvation stress [18]. By storing Fe as an amorphous material that that is more easily mobilized, the cell can more quickly return to vegetative growth under more favourable conditions. HR-STEM analysis of the encapsulin containing EncB and EncC loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ *in vitro* contained regions with regular arrangements and the measured d -spacings that could be described as the mineral Goethite. Previous work has identified that the cores of encapsulins isolated from bacteria contain P [18, 19], and P is know to affect the crystallization of the iron oxide nanoparticles [179, 180]. The *in vitro* loading conditions were designed to generate nanoparticles in a P free environment to identify its role in nanoparticle formation. The results indicate that the P most likely prevents the nanoparticles inside the encapsulin forming crystalline minerals, which is important for

explaining how amorphous material is stored and maintained inside the encapsulins.

As has been discussed previously, no differences were observed between the sizes of the Fe rich cores in the encapsulins containing only EncB or EncC or encapsulins containing both of these proteins. Suggesting, that there is no difference in the relative Fe uptake between encapsulins containing these two cargo proteins, and that both of them being present does not increase Fe uptake. Thus, it is still unclear as to why there would be two highly similar cargo proteins (in terms of both structure and function) encoded at distal genomic locations and what role they play in the Fe storage function of the encapsulin system [22, 181]. However, the encapsulins expressed in *E. coli* produced Fe-containing cores that were smaller than those from the isolated *M. xanthus* encapsulins across all growth conditions. When comparing the HAADF intensity between the two groups, the intensities were similar despite the difference in core size. This indicates, that the spatial arrangement of nanoparticles that has in part come to define the *M. xanthus* encapsulin is absent in encapsulins containing single or double cargo proteins, resulting in a more tightly packed arrangement. This could signify that the combination of all the cargo proteins is required to ensure the proper arrangement and storage of nanoparticles inside the encapsulin shell, and this proposed auxiliary function may be the answer to the question of why there are multiple cargo proteins with seemingly similar Fe uptake characteristics. However, the differences in the internal nanoparticle arrangement may also in part be due to the difference the host bacteria. Iron is inherently insoluble at neutral pH and in oxic environments and thus *M. xanthus* as an obligate aerobe is most likely more sensitive to maintaining proper Fe homeostasis than *E. coli* which can tolerate lower oxygen levels. Especially as the expression of proteins related to Fe homeostasis is tightly linked to the phase variation exhibited by the bacterium [101]. Whilst, *E. coli* has its own iron regulatory system including its own Fe storage proteins Bfr, FtnA and FtnB [182, 183]. It is likely therefore, that the Fe content of individual cells differs between the two species, and thus there would be a difference in the availability free Fe^{2+} ions for the encapsulin inside the cell. This may partially explain the differences in core sizes observed between encapsulins expressed in the two species.

Finally, encapsulins have been implicated in stress response. Specifically in *M. xanthus*, the encapsulin production is upregulated as a response to starvation [18]. Thus, the observed amorphous structure and the high surface area to volume ratio of the ~ 2 nm nanoparticles

potentially allow for the stored Fe to be liberated rapidly from the shell following the a return of the cell to an optimal growth environment. Suggesting a relationship between the stress response function and stucture and arrangement of nanoparticles inside the encapsulin. However, similar observations of complex nanoparticle arrangements inside the shell have not been made for other studied encapsulins systems where instead a single large Fe containing nanoparticle has been observed [23, 19]. In the study of other systems, primarily low magnification TEM was used to identify the Fe nanoparticles inside the encapsulins. Thus, the seeming difference between the Fe storage in *M. xanthus* and the other studied systems may be down to the type of imaging used, and high spatially resolved techniques (such as HAADF STEM) may be required to visualise a similar nanoparticle arrangement in the other systems. However, it should also be noted that in the study of other systems *E. coli* overexpression strains were used and as has been suggested here that may result in the loss of internal nanoparticle arrangement. Futher supporting the notion that highly spatial resolution microscopical techniques are important for characterizing the nanoparticles inside of encapsulin shells.

3.4 Appendix

3.4.1 Identification of nanoparticles from TEM images

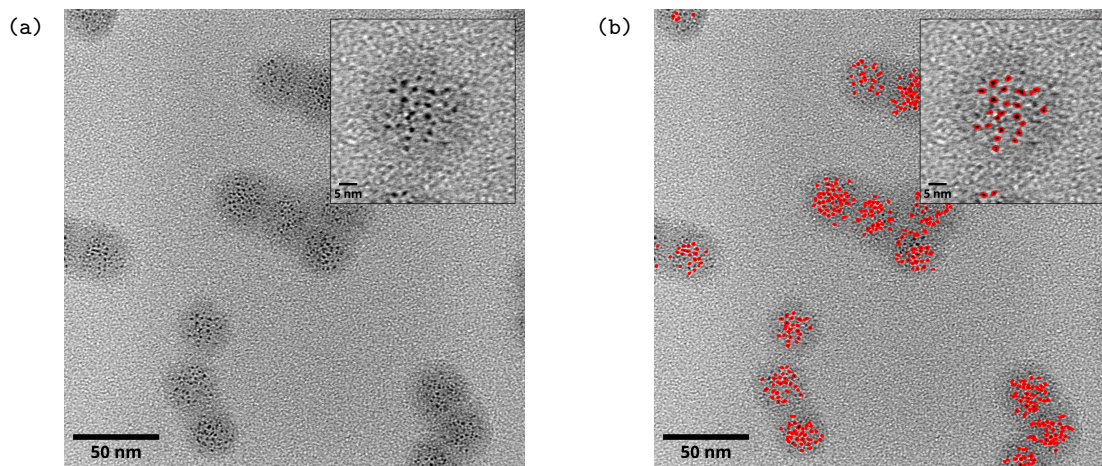


Figure 3.21: Identification of nanoparticles from TEM images. (a) Reference TEM image of unstained encapsulins captured at 300kx magnification. Scale bar represents 50 nm. The insert shows a single unstained encapsulin. Scale bar represents 5nm. (b) Same as (a). Overlay of image regions defined as 'nanoparticle' after image processing. The outlines of the identified regions are shown in red.

3.4.2 Identification of nanoparticles from HAADF images

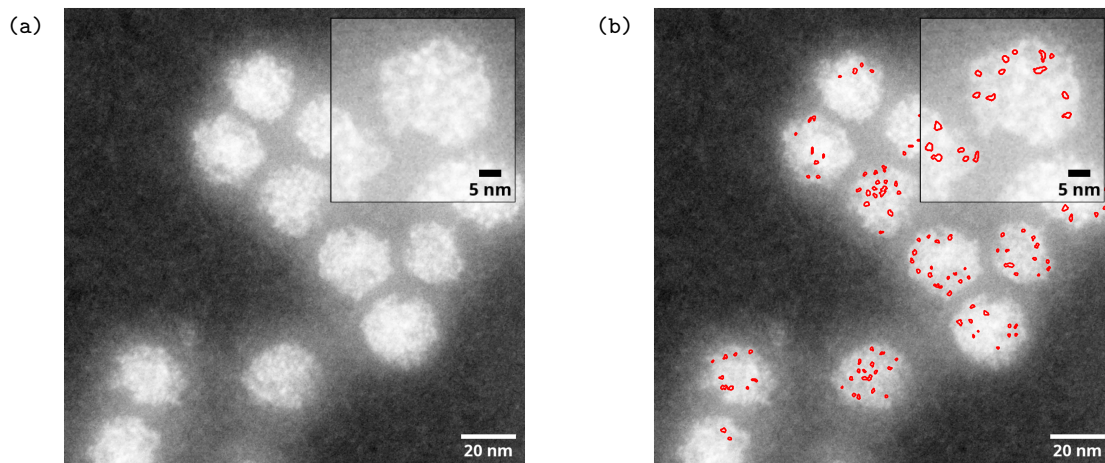


Figure 3.22: Identification of nanoparticles from HAADF images. (a) Reference HAADF image of unstained encapsulins captured at 300kx magnification. Scale bar represents 20 nm. The insert shows a single unstained encapsulin. Scale bar represents 5 nm. (b) Same as (a). Overlay of image regions defined as ‘nanoparticle’ after image processing. The outlines of the identified regions are shown in red.

3.4.3 Iron oxide lattice parameters

Valence	Chemis-try	Name	Space group	No.	Comment	Lattice param-eters (pm)	Source
II	FeO	Wuestite	Fm $\bar{3}$ m	225	Cubic	$a=430$	[184]
II/III	Fe ₃ O ₄	Magnetite	Fd $\bar{3}$ m	227	Cubic, inverse spinel	$a=840$	[185]
III	α -Fe ₂ O ₃	Haematite	R $\bar{3}c$	167	Hexagonal	$a=504$, $c=1375$	[186]
	γ -Fe ₂ O ₃	Maghemite	Fd $\bar{3}$ m	227	Cubic, spinal	$a=833$	[187]
	α -FeO(OH)	Goethite	Pbnm	62	Orthor- hombic	$a=460$, $b=995$, $c=302$	[188]
	Fe ₂ O ₃ - ·0.5H ₂ O	6-line Ferrihy- drite	P63mc	186	Hexagonal	$a=296$, $c=940$	[189]
	α -FePO ₄	Iron phos- phate	P3 ₁ 21	152	Trigonal	$a=500$, $c=1125$	[190]

Table 3.5: The lattice parameters of iron oxides phases used in this study. The valence state, chemistry, name, space group, space group number (No.), a brief description of the space group (comment), the lattice parameters (in pm) and the source the lattice parameters were taken from.

3.4.4 *D*-spacings measured from HR-STEM images of encapsulins isolated from *M. xanthus*

<i>D</i> -spacing (pm)	± (pm)	Matching phases	Phase <i>d</i> -spacing (pm)
160	8	Ferrihydrite(-233) Ferrihydrite(123) Ferrihydrite(-323) Ferrihydrite(-2-13) Goethite(-300) Haematite(-2-13) Haematite(-2-10) Haematite(123) Haematite(-312) Iron(III) Phosphate(-311) Iron(III) Phosphate(-2-12) Iron(III) Phosphate(-2-11) Iron(III) Phosphate(122) Iron(III) Phosphate(-2-1-1) Iron(III) Phosphate(-322) Magnetite(-3-3-3) Magnetite(-3-33) Magnetite(-333) Magnetite(333) maghemite(333) Wuestite(-202) Wuestite(-2-20) Wuestite(022)	163 163 163 163 153 155 165 155 160 160 155 160 155 160 155 162 162 162 162 160 152 152 152
195	9	Ferrihydrite(013) Ferrihydrite(-103) Haematite(-214) Haematite(-223) Haematite(114) Iron(III) Phosphate(-222) Iron(III) Phosphate(022) Magnetite(-402) Magnetite(-3-13) Magnetite(133) Magnetite(-3-3-1) Magnetite(-3-31) Magnetite(024) Magnetite(-133) Magnetite(-4-20) Magnetite(-204) Magnetite(-313) Maghemite(-3-13) Maghemite(133) Maghemite(-3-3-1) Maghemite(-3-31) Maghemite(-133) Maghemite(-313)	198 198 203 197 203 199 199 188 193 193 193 193 188 193 188 188 198 191 191 191 191 191 191
215	11	Ferrihydrite(-112) Goethite(-2-20) Goethite(-2-10) Haematite(-202) Iron(III) Phosphate(-122) Iron(III) Phosphate(112) Iron(III) Phosphate(-200) Iron(III) Phosphate(-1-12) Iron(III) Phosphate(-213)	225 209 224 208 225 225 214 225 205

CHAPTER 3. ELECTRON MICROSCOPICAL INVESTIGATION OF ENCAPSULIN DERIVED NANOPARTICLES

<i>D</i> -spacing (pm)	Measurement error (pm)	Matching phases and Miller Indices (in brackets)	Predicted <i>d</i> -spacing (pm)
		Iron(III) Phosphate(-123)	205
		Iron(III) Phosphate(-201)	210
		Iron(III) Phosphate(-1-13)	205
		Iron(III) Phosphate(113)	205
		Iron(III) Phosphate(-212)	225
		Magnetite(004)	210
		Magnetite(-400)	210
		Wuestite(-200)	215
		Wuestite(002)	215
242	12	Ferrihydrite(011)	248
		Ferrihydrite(004)	248
		Ferrihydrite(-101)	248
		Goethite(-200)	230
		Goethite(-111)	245
		Goethite(-1-11)	245
		Goethite(-101)	252
		Goethite(111)	245
		Goethite(-1-1-1)	245
		Haematite(-1-12)	236
		Haematite(-211)	247
		Haematite(111)	247
		Haematite(-122)	236
		Haematite(-1-1-1)	247
		Iron(III) Phosphate(-114)	232
		Iron(III) Phosphate(-1-10)	247
		Iron(III) Phosphate(-211)	241
		Iron(III) Phosphate(-1-11)	241
		Iron(III) Phosphate(111)	241
		Iron(III) Phosphate(-1-1-1)	241
		Iron(III) Phosphate(014)	232
		Magnetite(-2-22)	242
		Magnetite(222)	242
		Magnetite(-222)	242
		Magnetite(113)	253
		Magnetite(-2-2-2)	242
		Magnetite(-311)	253
		Magnetite(-113)	253
		Magnetite(-1-13)	253
		Magnetite(-3-1-1)	253
		Magnetite(-3-11)	253
		Maghemite(-2-22)	240
		Maghemite(222)	240
		Maghemite(-222)	240
		Maghemite(113)	251
		Maghemite(-2-2-2)	240
		Maghemite(-311)	251
		Maghemite(-113)	251
		Maghemite(-1-13)	251
		Maghemite(-3-1-1)	251
		Maghemite(-3-11)	251
		Wuestite(-111)	248
		Wuestite(-1-1-1)	248
		Wuestite(-1-11)	248
		Wuestite(111)	248

Table 3.6: Table of the measured d -spacings from HR-STEM images of the nanoparticles inside encapsulins isolated from *M. xanthus*. The errors represent the uncertainty when measurements were made from selected line profiles. the different iron oxides phases are colour coded: ferrihydrite (red), goethite (green), haematite (blue), iron(III) phosphate (purple), magnetite(orange), maghemite (yellow) and wuestite(black).

3.4.5 *D*-spacings measured from HR-STEM images of EncA+EncB+EncC

CHAPTER 3. ELECTRON MICROSCOPICAL INVESTIGATION OF ENCAPSULIN DERIVED NANOPARTICLES

<i>D</i> -spacing (pm)	Measurement error (pm)	Matching phases and Miller Indices (in brackets)	Predicted <i>d</i> -spacing (pm)
176	8	Ferrihydrite(-114) Goethite(-211) Goethite(-2-21) Goethite(-2-11) Goethite(-201) Goethite(-2-2-1) Goethite(-2-1-1) Haematite(024) Magnetite(-422) Magnetite(-4-2-2) Magnetite(-2-24) Magnetite(224) Magnetite(-224) Magnetite(-4-22)	173 180 172 180 183 172 180 184 170 171 171 171 171 171
204	8	Ferrihydrite(013) Ferrihydrite(-103) Goethite(-2-20) Haematite(-214) Haematite(-202) Haematite(-223) Haematite(114) Magnetite(004) Magnetite(-400)	198 198 209 203 208 197 203 210 210
224	11	Ferrihydrite(004) Ferrihydrite(-112) Goethite(-2-10) Goethite(-200) Wuestite(-200) Wuestite(002)	235 225 224 230 215 215
278	13	Goethite(011) Haematite(-104)	289 270
300	15	Ferrihydrite(-1-10) Goethite(001) Magnetite(-202) Magnetite(-2-20) Magnetite(022) Maghemite(-202) Maghemite(-2-20) Maghemite(022)	298 302 297 297 297 295 295 295

Table 3.7: Table of the measured *d*-spacings from HR-STEM images of **EncA+EncB+EncC**. Empty EncA+EncB+EncC encapsulins were purified and loaded with 8 mM Fe(NH₄)₂(SO₄)₂. The error represents the uncertainty when the measurements were made from selected line profiles. The different iron oxides phases are color coded: ferrihydrite (red), goethite (green), haematite (blue), magnetite (purple), maghemite (yellow) and wuestite (black).

Chapter 4

Characterization of encapsulins derived nanoparticles using EDXS

In this chapter, the internal nanoparticles from both the encapsulins isolated from *M. xanthus* and encapsulins with reduced numbers of cargo proteins expressed in *E. coli* are characterized using Energy Dispersive X-ray Spectroscopy (EDXS). The data collected provides insights into the elemental composition of the nanoparticles inside the encapsulins and provides a quantification of the major elements present. The identification of the major elements present was important not only for the characterization of the encapsulin synthesised nanoparticles, but also for the identification of the experimental parameters required for future electron energy loss experiments. Furthermore, these experiments identify the trace elements that are present in the samples and suggest a reason for their presence.

Microscope	Operating voltage (kV)	Type	Location
JEOL F200	200	S/TEM	Sorby Centre, Sheffield
Tescan Tensor	100	4D-STEM	Ernst Ruska Centre, Juelich

Table 4.1: Table of electron microscopes used to collected EDXS data during this study. The machine, operating voltage, capabilities and location are given.

4.1 Experimental procedures

4.1.1 Electron microscopy

A list of the electron microscopes used can be found in Table 4.1. EDXS data was collected from both machines in STEM mode using double Silicon Drift Detectors (SDD). The EDXS spectra were acquired simultaneous with HAADF images.

4.1.2 Encapsulin samples

A list of the encapsulin samples used in this study can be found in Table 4.2.

4.1.3 Data processing: Digital Micrograph

The EDXS data collected using the JEOL F200 was processed using the Digital Micrograph 3.6 (Gatan) software suite. Briefly, the spectra were loaded into the software and spectral peaks were identified using the ‘auto-identify’ feature within the software. After using this feature, the spectra were manually checked to ensure all peaks had been correctly identified as sometimes the software struggled to recognise peaks close to the noise level. Before chemical maps were generated, the signal-to-noise ratio was improved by using Principle Component Analysis (PCA). This is a commonly used method and can greatly improve the identification of peaks, especially in noisy EDXS datasets [191]. For quantification, the ‘quantify’ function within Digital Micrograph was used with the k -factors provided by the software.

Bacterium	Proteins	Growth conditions
<i>M. xanthus</i>	EncA, EncB, EncC, EncD	Isolated from native host
<i>E. coli</i>	EncA, EncB	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)(\text{SO}_4)_2$
<i>E. coli</i>	EncA, EncC	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)(\text{SO}_4)_2$
<i>E. coli</i>	EncA, EncB, EncC	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)(\text{SO}_4)_2$
<i>E. coli in vitro</i>	EncA, EncB, EncC	Isolated empty encapsulins loaded with Fe following the addition of 8 mM $\text{Fe}(\text{NH}_4)(\text{SO}_4)_2$

Table 4.2: Table of encapsulins samples used in this work. The host bacterium, proteins present in each encapsulin, and the experimental conditions under which nanoparticle synthesis occurred are given.

4.1.4 Elemental map generation using HyperSpy

The EDXS data collected using the Tescan Tensor was processed using the HyperSpy python package [166]. Due to the large size of the SI dataset, the file had to be opened using the lazy loading function from the HyperSpy package [166]. Lazy loading uses out-of-core algorithms to chunk the data, allowing for datasets larger than the available computer memory to be opened. Due to the large size of the data the SI was binned 10x10 in the spatial axes. Binning the data in this way, greatly reduced the size of the dataset allowing it to be loaded entirely into memory for analysis. Binning has an additional benefit of greatly improving the signal to noise ratio improving the usefulness of the dataset. The peaks in the dataset were manually identified and set using the `.add_elements()` function. Elemental maps were generated using the `.get_line_intensities()` and `.plot()` functions.

4.2 Results

4.2.1 Chemical mapping of the *M. xanthus* encapsulin using EDXS

The EDXS maps of encapsulins isolated from *M. xanthus* provide novel information about the chemical composition of the nanoparticles synthesised inside the encapsulin. Most importantly, the nanoparticles are enriched for O, Fe and P, evidenced by the high count number and shared localization in the chemical maps (see Figure 4.1c). The presence of O in the elemental composition of the nanoparticles is a novel observation, as previous work only identified the presence of Fe and P using inductively coupled plasma mass spectrometry [18]. Several other elements were also identified in the sample, namely: C, N, Na, Mg, Si, Cl, Ca, Mn and Co. As the sample carrier grid is made of Cu, this is the source of the large Cu signal in the spectra. The C signal is derived from two sources, firstly the C film that covers the Cu carrier grid, and secondly C is a major component of the protein encapsulin shell. The Si signal comes from the Silicon Drift Detector that was used to collect the X-ray signals and the N is a key component of the protein shell. The remaining elements - Na, Mg, Cl, Ca, Co and Mn- are found distributed across the sampled region. These ions are all present in small amounts inside the bacterial protoplasm (inorganic molecules amount to only 1% of the total dry mass of the cell [192]) which may explain their presence. It should also be noted that Mg - in the form of MgSO_4 - is a major component of the *M. xanthus* growth media which may also explain its presence in the isolated encapsulins.

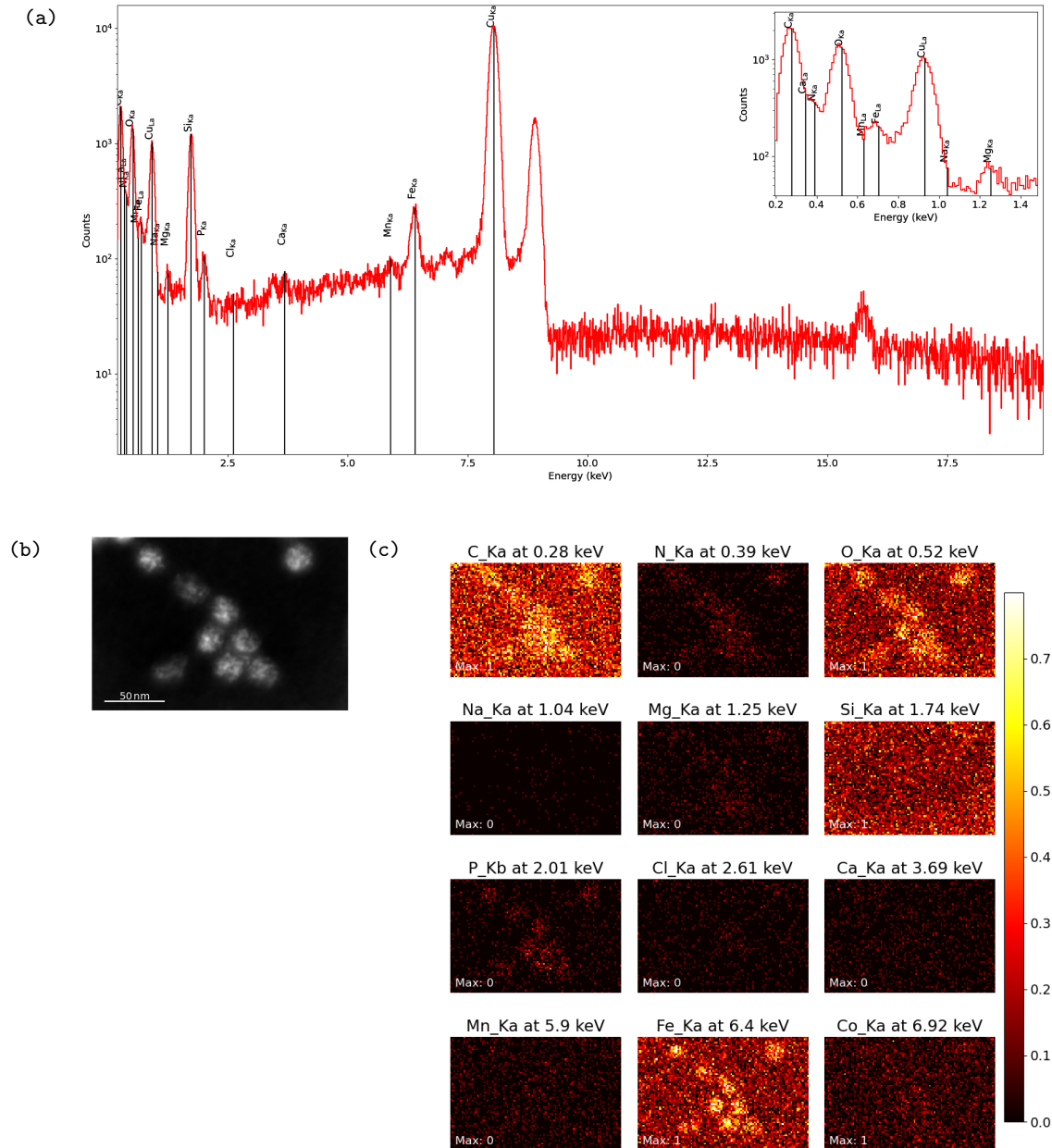


Figure 4.1: Chemical mapping of encapsulin isolated from *M. xanthus*. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The L_α and K_α lines of all elements present are labelled. (b) HAADF reference image of sampled region, the scale bar represents 50 nm. (c) Elemental maps showing the spatial distribution of C_{Kα}, N_{Kα}, O_{Kα}, Na_{Lα}, Mg_{Kα}, Si_{Kα}, P_{Kα}, Cl_{Kα}, Ca_{Kα}, Mn_{Kα}, Fe_{Kα}. Cu is present as the sample has been imaged on a Cu carrier grid, however the chemical map has been omitted as the intensity of its signal causes the signal from the other elements to become lost. An intensity bar representing the contrast against counts can be seen on the right. The max counts for many elements are 0 which is an artefact caused by the binning of data and the subsequent rounding of the number of counts.

4.2.2 Chemical mapping of the encapsulin of *M. xanthus* on Au grids

As previously discussed, the sample carrier grid produces its own x-ray signal and if the element the grid is made of is also present in the sample, the signal from the sample can become lost. As Cu is a micro-nutrient that is found in low levels in *M. xanthus*, it was important to identify if Cu was also present in the isolated encapsulin samples [193]. Thus, further EDXS experiments were carried out using carrier grids made of Au instead of Cu. Chemical mapping of the EDXS data from the *M. xanthus* encapsulin on Au grids reveals that Cu is present in low amounts. Furthermore, it was found localised to the same image region as the Fe signal (see Figure 4.2). This indicates that there is a small Cu signal associated with the encapsulin, however, due to the low spatial resolution of the dataset, it is unclear if it is localised with the internal nanoparticles or the protein shell. The association of Cu with the encapsulin may be explained by the Irwin-Williams series [194]. The Irwin-Williams series states that the stability of transition metal complexes across the first row of the periodic table increases up to Cu. Meaning, that the presence of Cu in association with the encapsulin might be due to its naturally higher stability in complex than Fe.

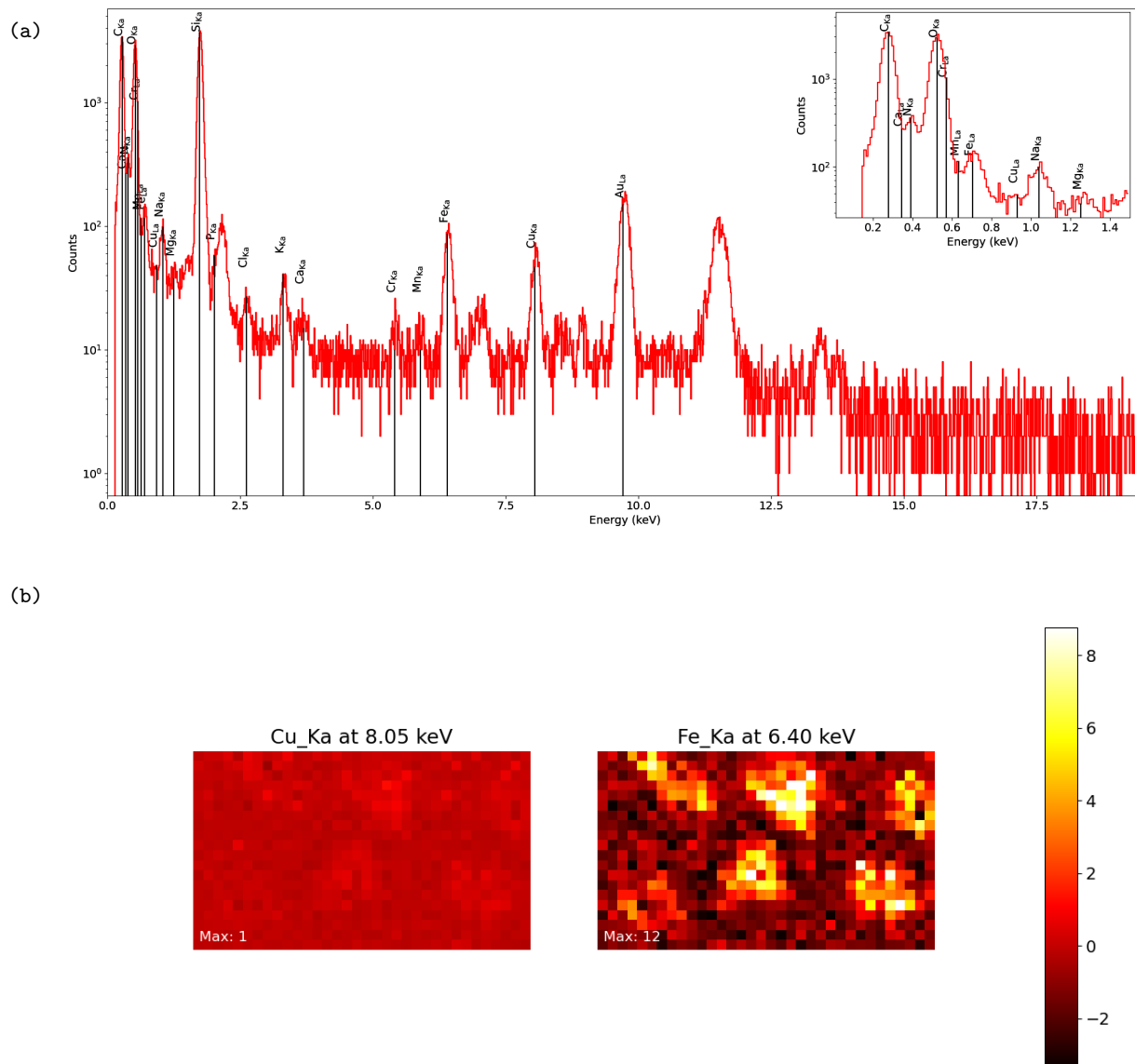


Figure 4.2: EDXS analysis of *M. xanthus* encapsulins on Au grids. (a) Summed EDXS spectrum of all probe positions. The spectrum is plotted on a log scale for easier visualisation. The $L\alpha$ and $K\alpha$ lines of all elements present are labelled. The insert is a magnification of the spectral region between 0-1.5 keV. (b) Elemental maps generated from the $Fe_{K\alpha}$ and $Cu_{K\alpha}$ lines. An intensity bar representing the contrast against counts can be seen on the right.

4.2.3 Chemical mapping of encapsulins overexpressed in *E. coli*

Chemical mapping of encapsulins expressed in *E. coli* reveals that the major elemental components of the internal nanoparticles are O, P and Fe as with the native structures isolated from *M. xanthus*. Also, as the samples were analysed on C coated Cu grids, a large C and Cu signal can be seen in the spectra. A number of trace elements were found associated with the encapsulin nanoparticles, these were: Ca, Cl, N, Co, K, Mg, Mn, and Na. The Na and Cl signal most likely comes from the salt used in the elution buffer. NaCl is often a component of protein buffers as it aids in protein solubility. The proposed function of NaCl in protein stabilization may explain why it is found localised to the same image region as the Fe, P and O signals. The N signal comes from the protein shell, and thus is also localized to the same image region as the elements representing the internal nanoparticles. Finally, trace amounts of the metals Ca, Co, Mg and Mn were observed (see Figures 4.3, 4.4 and 4.6). As previously stated, these are all inorganic ions that are found in the bacterial cytoplasm. Despite, these elements being found in trace amounts in the $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ that was added to the growth media, they are found in the EDXS spectra for encapsulins isolated from cultures where $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ was not added (see Appendix 4.4.1). Further supporting the idea that these trace elements were most likely accumulated from *E. coli* cytoplasm. One final note, the only difference observed between the three samples was the presence of Cr in EncA+EncC. The signal for this element is found dispersed over the entire image, thus the possible source of the signal is unclear, but the signal is most likely from either a contaminant in the electron microscope or from a component of the electron microscope itself⁴³.

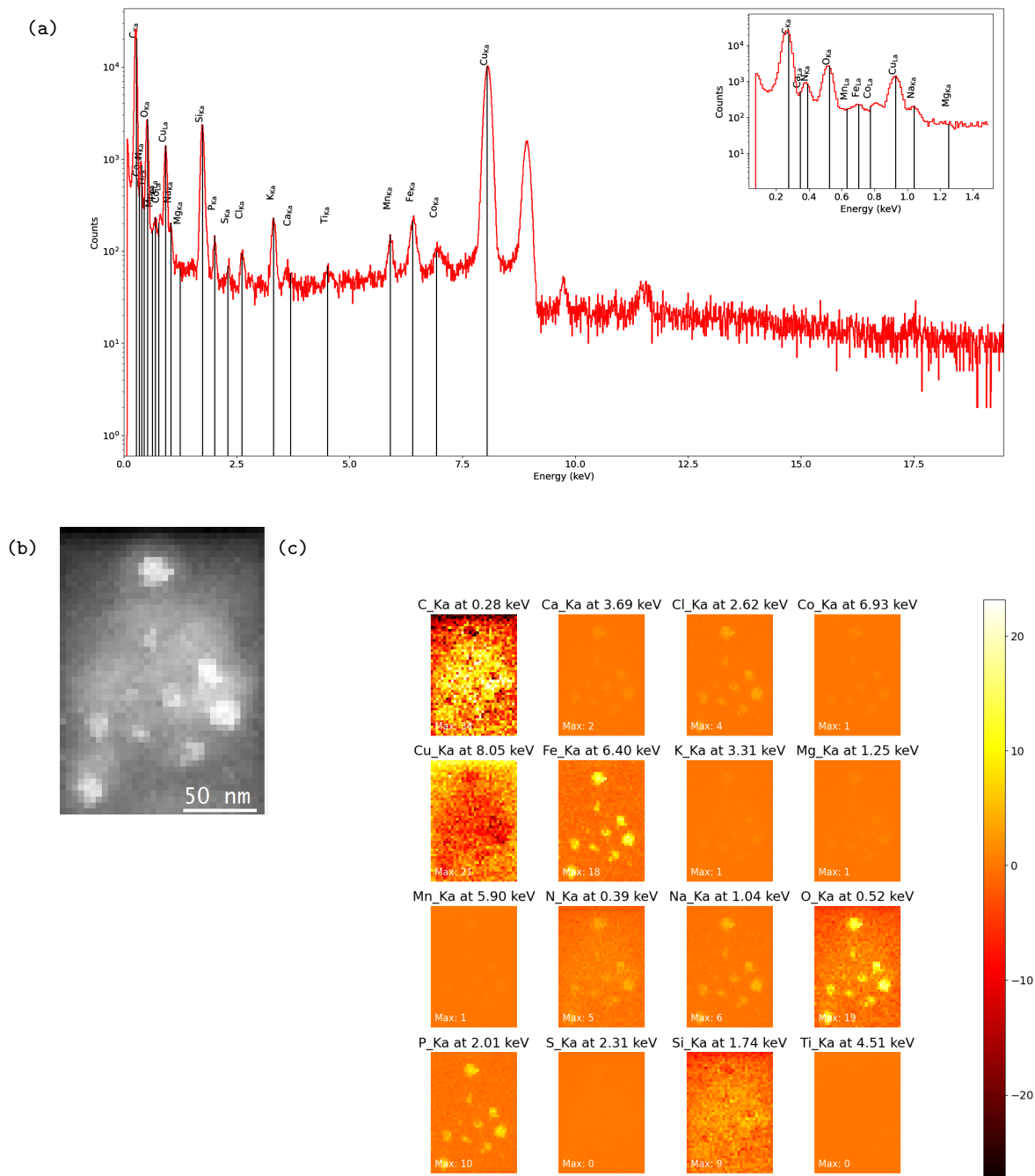


Figure 4.3: Chemical mapping of elemental cores inside EncA+EncB shells. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. (b) HAADF reference image of spectral image. Scale bar represents 50 nm. (c) Chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cu_K α , Fe_K α , K_K α , Mg_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , S_K α , Si_K α and Ti_K α . An intensity bar representing the contrast against counts can be seen on the right.

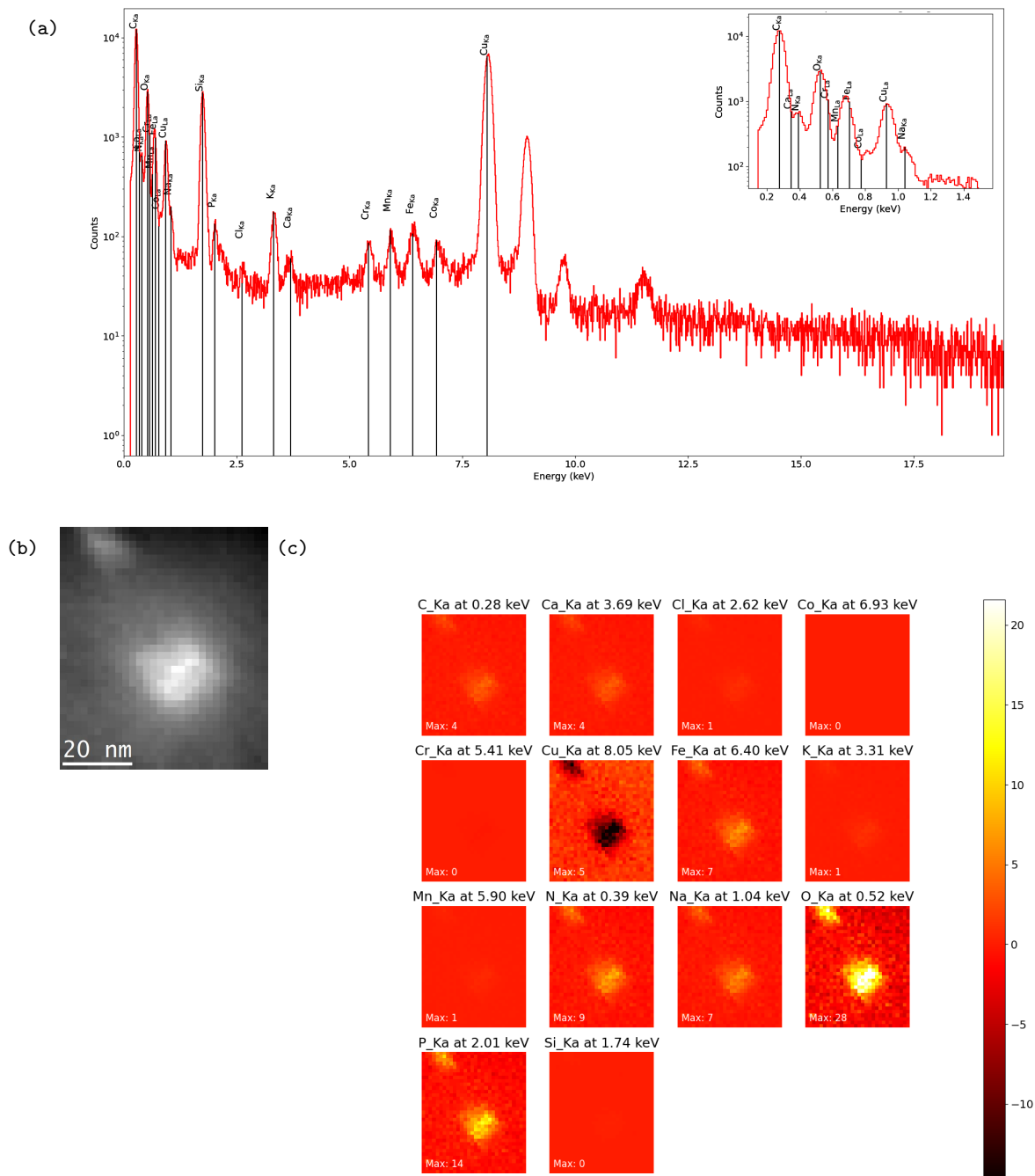


Figure 4.4: Chemical mapping of elemental cores inside EncA+EncC shells. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. (b) HAADF reference image of spectral image. Scale bar represents 20 nm. (c) chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_K α , Co_K α , Cu_K α , Fe_K α , K_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , Si_K α . An intensity bar representing the contrast against counts can be seen on the right.

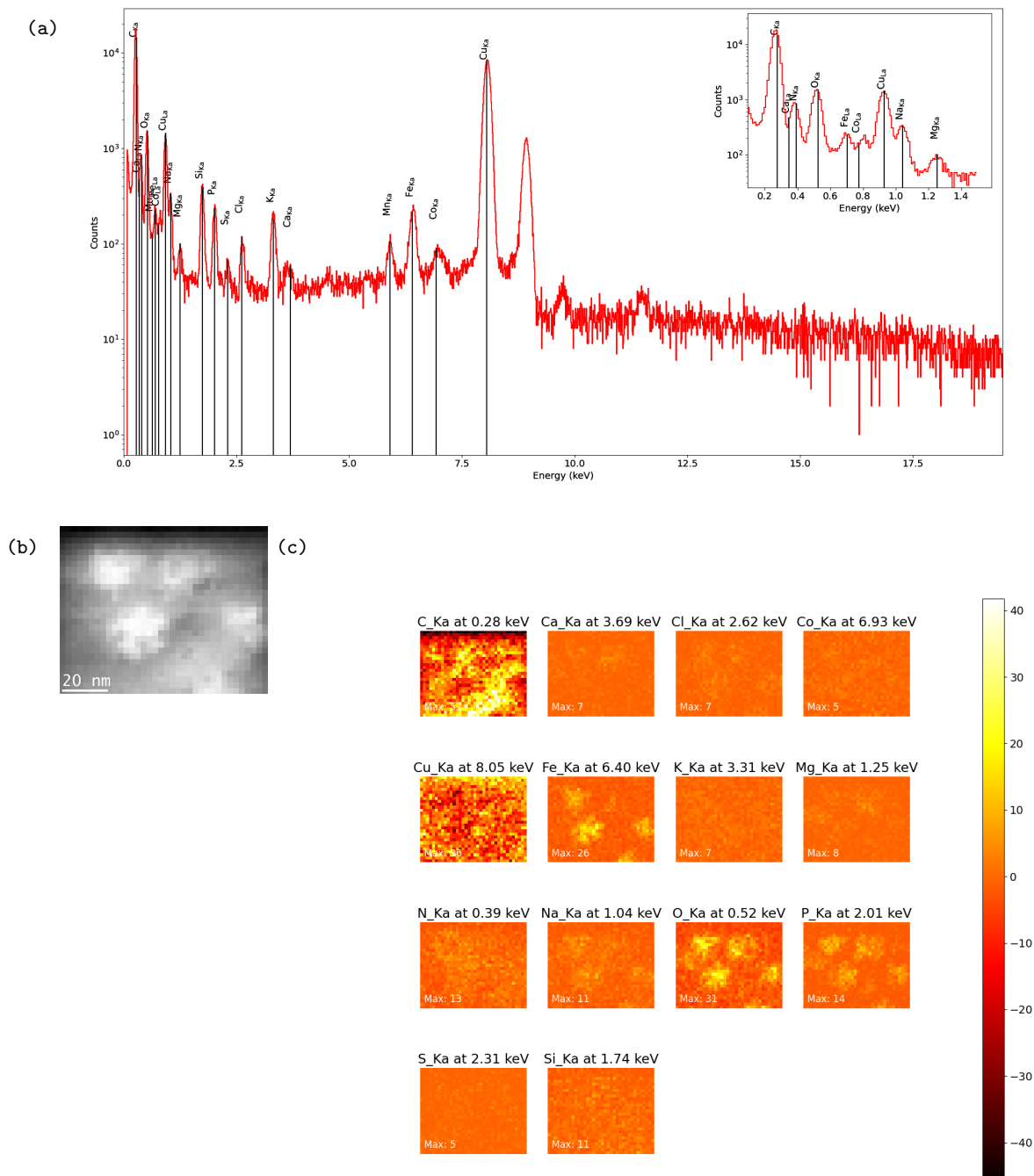


Figure 4.5: Chemical mapping of elemental cores inside EncA+EncB+EncC shells. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. **(b)** HAADF reference image of spectral image. Scale bar represents 20 nm. **(c)** chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_L α , Co_K α , Cu_K α , Fe_K α , K_K α , Mg_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , S_K α and Si_K α . An intensity bar representing the contrast against counts can be seen on the right.

4.2.4 Chemical mapping of EncA+EncB+EnC loaded with 8 mM

$\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$

The samples of EncA+EncB+EnC loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ *in vitro* were analysed using EDXS. Unlike previous samples, the nanoparticles did not have P associated with the Fe, and thus the major signal from the nanoparticles is from only Fe and O. Indicating that the cores are free of P and that this method of loading Fe into the encapsulin is sufficient for generating nanoparticles that are free of P. The other elements present are the same as has been discussed previously (see Sections 4.2.1 and 4.2.3).

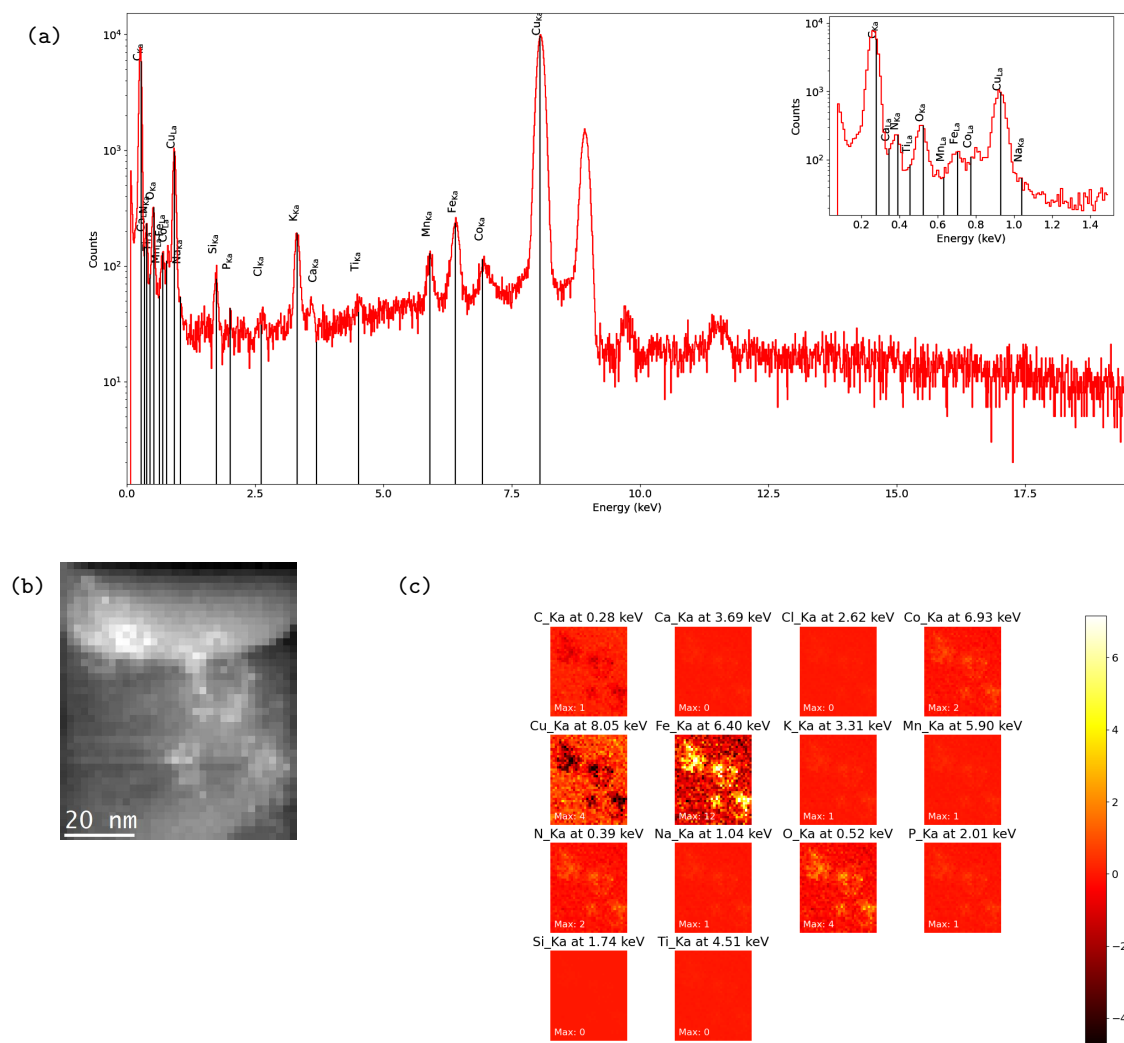


Figure 4.6: EDXS chemical maps of EncA+EncB+EncC loaded with Fe following the addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. the L_{α} and K_{α} lines of all elements present are labelled. (b) Reference HAADF image of the SI. Scale bar represents 20 nm, the C contamination build up during scanning can be seen. (c) Chemical maps showing the spatial distributions of C_Kα, Ca_Kα, Cl_Kα, Cu_Kα, Fe_Kα, K_Kα, Mn_Kα, N_Kα, Na_Kα, O_Kα, P_Kα, Si_Kα, Ti_Kα. An intensity bar representing the contrast against counts can be seen on the right.

4.2.5 Quantification of the atomic composition of the nanoparticles inside the encapsulin using EDXS

The atomic composition of the nanoparticles inside all encapsulins generated were quantified from EDXS spectra using the routines present inside Digital Micrograph. In samples of encapsulins from *M. xanthus* that were used for quantification, an O containing contamination layer formed during data collection, the presence of this layer drastically increased the O signal and resulted in a loss of spatial resolution in the O signal maps. The cause of the contamination is unknown and was not present in any other sample used in this work. However, it was still possible to identify that the Fe:P ratio of the nanoparticles was 1.2 ± 0.1 .

Quantification of the EDXS reveals key differences between the encapsulins containing EncB and those containing EncC. The encapsulins expressed in *E. coli* in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ had higher Fe:P ratios than the encapsulin isolated from *M. xanthus*. The Fe:P ratios measured were 2.4 ± 0.1 , 2.9 ± 0.2 and 2.7 ± 0.1 for EncA+EncB, EncA+EncC and EncA+EncB+EncC respectively. Thus, the Fe:P ratio was higher for encapsulin containing only EncC than in encapsulins containing only EncB and when both cargo proteins were present the Fe:P ratio was between these two extremes. Finally, as was previously identified, the encapsulins loaded with Fe *in vitro* contained no P and only O and Fe. The atomic ratio of these two elements in the material was 1:1.

Host	Pro- teins	At% O	$\pm$ stdev	At% P	$\pm$ stdev	At% Fe	$\pm$ stdev	Fe:P	$\pm$ stdev
<i>M. xanthus</i>	EncA, EncB, EncC, EncD	92	5	4	1	5	1	1.2	0.1
<i>E. coli</i>	EncA, EncB	52	3	14	2	34	3	2.4	0.1
<i>E. coli</i>	EncA, EncC	57	3	11		32	3	2.9	0.2
<i>E. coli</i>	EncA, EncB, EncC	54	3	14	2	30	3	2.7	0.1
<i>E. coli</i> + <i>in vitro</i>	EncA, EncB, EncC	52	3	-	-	48	3	-	-

Table 4.3: Atomic composition of the nanoparticles inside the different encapsulins. The stdev represents one standard deviation. Quantifications were carried out in Digital Micrograph using the k -factors from the software.

4.3 Discussion

The study presented here uses EDXS to identify the elements present inside the encapsulins isolated from *M. xanthus* and those overexpressed in *E. coli*. Firstly, the observation of O as a major component in the encapsulin core is a novel observation, as previously published work identified that Fe and P were the only significant elemental components [18]. This presence of O in the nanoparticles means that the *M. xanthus* encapsulins synthesized material is more chemically similar to both the material synthesised to by other encapsulins and bacterial ferritin systems and most likely consists of a oxy-hydroxide-phosphate [195, 196, 197].

A number of different trace elements were found localised to the same image region as the major constituting elements of the nanoparticles. These elements - Ca, Co, Cu, Mg, Mn - may be localised to this region for several reasons. Firstly, all of these elements exist as 2+ ions, which is the same as the Fe^{2+} which is mineralized into the encapsulin [192]. Thus, these ions may be able to traverse the shell of the encapsulin using the same mechanism as the Fe^{2+} , resulting in them being accumulated inside the encapsulin and into the nanoparticles. However, due to the low spatial resolution in the datasets it is impossible to tell if these ions are indeed localized to nanoparticles inside the shell or to the surrounding protein. Secondly, the encapsulin cargo proteins have additional metal-binding sites, to which Ca is bound in the solved crystal structure [28]. However, as the proteins were crystallized in the presence of Ca it is unclear if the Ca was already present or bound to the metal-binding site during crystallization. Finally, it is known that all of these elements are able to bind directly to the active site of ferritin proteins and act as proxy ions, thus, the ions may have traversed the shell and be interacting directly with the ‘ferritin-like’ cargo proteins explaining their localization in the EDXS chemical maps [88, 198].

The spatial resolution in EDXS datasets is limited by the trade-off between the exposure time required for signal generation at each probe position and the buildup of carbon contamination during scanning. X-ray emission requires longer exposure times, especially for light elements such as O and P, due to their low K-line fluorescence yield [131, 125]. However, these longer exposure times cannot be achieved for these samples without causing significant carbon contamination, which severely reduces spatial resolution. Moreover, the poor X-ray yield from light elements results in the elemental maps lacking both localization and com-

positional information about the encapsulin protein shell which is composed entirely of light elements: C, N, O and S.

The quantification of the EDXS data is discussed alongside the Electron Energy Loss Spectroscopy (EELS) quantification in the following chapter (see Section 5.3).

4.4 Appendix

4.4.1 Chemical maps from empty encapsulins isolated from *E.coli*

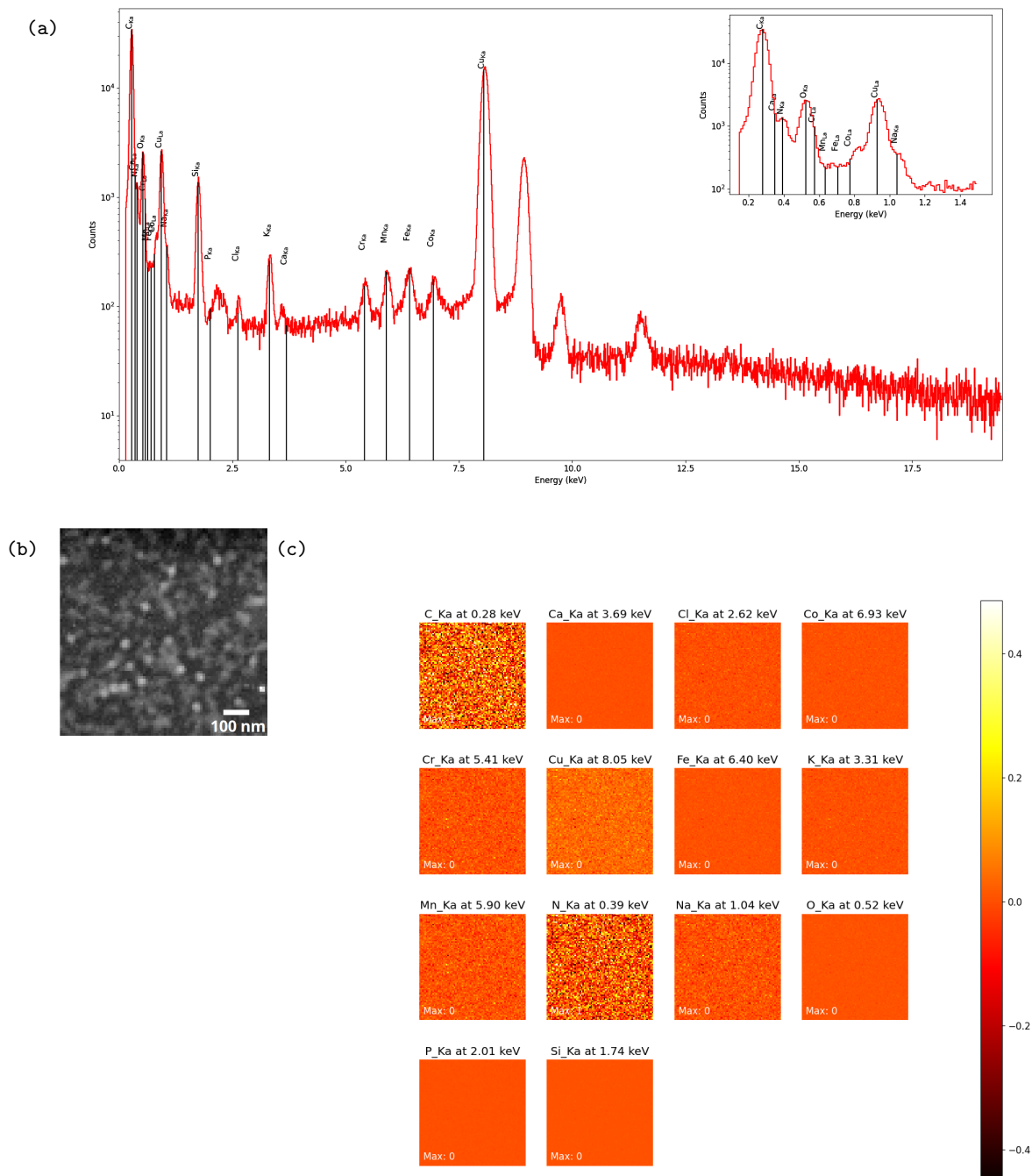


Figure 4.7: Chemical mapping of empty EncA+EncC encapsulins. (a) Summed EDXS spectrum of all probe positions in SI. The spectrum is plotted on a log scale for easier visualisation. The peaks representing the L- and K α lines of all elements present are labelled. The insert is a magnification of 0 - 1.5 keV. (b) HAADF reference image of spectral image. Scale bar represents 50 nm. (c) Chemical maps showing the spatial distribution of C_K α , Ca_K α , Cl_K α , Co_K α , Cr_K α , Cu_K α , Fe_K α , K_K α , Mn_K α , N_K α , Na_K α , O_K α , P_K α , Si_K α . An intensity bar representing the contrast against counts can be seen on the right.

Chapter 5

Characterization of encapsulin-derived nanoparticles using EELS

In this chapter, the core material from both encapsulins isolated from *M. xanthus* and encapsulins containing reduced numbers of cargo proteins expressed in *E. coli* are characterized using Electron Energy Loss Spectroscopy (EELS). Several experiments are undertaken: quantification of the elemental content of the cores, quantifications of the protein shell, estimations of the number of Fe atoms stored and the identification of the Fe valence state using the Fe L_{2,3} Energy Loss Near Edge Structure (ELNES) region. Importantly, due to its higher sensitivity to light elements compared with Energy Dispersive X-ray Spectroscopy (EDXS), shorter exposure times could be used reducing the build up of carbon contamination during scanning and increasing spatial resolution. The higher spatial resolution allowed for individual nanoparticles be quantified and thus it was possible to compare nanoparticle inside a single encapsulin shell.

Bacterium	Proteins	Growth conditions
<i>M. xanthus</i>	EncA, EncB, EncC and EncD	Isolated from native host
<i>E. coli</i>	EncA, EncB	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$
<i>E. coli</i>	EncA, EncC	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$
<i>E. coli</i>	EncA, EncB, EncC	Isolated following growth in media containing 0.05 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$
<i>E. coli</i>	EncA, EncB, EncC	Empty encapsulins were isolated and loaded with Fe <i>in vitro</i> through the addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$

Table 5.1: Table of encapsulins constructs used in this work. The host organism and the experimental conditions used to generate the samples are indicated.

5.1 Experimental methods

5.1.1 Sample preparation

The samples used in this chapter Table 5.1. Briefly, the samples imaged consisted of: encapsulin structures isolated from *M. xanthus*, a number of structures overexpressed in *E. coli* and loaded with iron by the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the growth media, and a sample of empty EncA+EncB+EncC purified following overexpression in *E. coli* before being loaded with iron *in vitro*. For detailed protocols on how the overexpressed encapsulins were loaded with iron see Section 3.1.1 and 3.1.2.

5.1.2 Reference material preparations

Samples of ilmenite and haematite dispersed on lacy carbon grids were provided by Dr. T. Walther. Magnetite reference material was purchased (Inoxia) and suspended in 99.9% acetone before being loaded onto a holey carbon Cu grid (AGS147-H7, Agar scientific). The excess suspension was wicked off with filter paper (Whatman). Magnetite grids were screened using a conventional light microscope to ensure adequate material distribution on the grid before being loaded into the JEOL F200 for analysis.

5.1.3 Scanning transmission electron microscopy

Electron microscopy analysis was performed using either a JEOL ARM200F at ePSIC - Diamond Light Source, Harwell - or a JEOL F200 - Sorby Centre, University of Sheffield. For the experiments carried out at the Diamond Light Source, the JEOL ARM200F was operated by Dr. Danaie Mohsen. On the JEOL ARM200F, Annular Dark Field (ADF) imaging and EELS spectroscopy was carried out at 80 keV with the probe forming spherical aberration corrector activated. A 30 μm condenser aperture was used which gave a convergence angle (α) of 22.6 mrad and a ≈ 1 Å probe. A 5 mm spectrometer entrance aperture was used, giving a collection angle (β) of 103 mrad. The convergence angle is the angle by which the electron beam is converged by the electron optics in a STEM, and the collection angle is the angular range over which scattered electrons are collected by the spectrometer. Dual EELS spectra were collected, with a detector offset of 132 eV and an energy dispersion of 0.25 eV/ch. The microscope had a K3 Continuum GIF (Gatan) allowing for *in situ* data to be collected. In an *in situ* dataset, each Spectrum Image (SI) generated is comprised of several passes of the electron beam. The *in situ* data collection was stopped when obvious carbon contamination build up could be observed in the ADF image.

On the JEOL F200, ADF imaging and EELS spectroscopy were carried out at 200 keV. A 40 μm condenser aperture was used which gave a convergence angle of 12 mrad, the 5 mm spectrometer entrance aperture was used which gave a collection angle of 17 mrad. For the calculation of these angles see Appendix 5.4.1 and Appendix 5.4.1. Dual EELS spectra were collected over a range of offsets and dispersions.

5.1.4 Pre-processing of EELS SI datasets

Pre-processing of EELS data was carried out with either Digital Micrography (GMS 3.6 Suite, Gatan) or HyperSpy 2.0[166]. Initially, the Low Loss (LL) spectra was shifted such that the Zero Loss Peak (ZLP) was centred over 0, with the same shift being applied to the High Loss (HL) spectra. *In situ* data from the JEOL ARM200F was then aligned to remove any spatial drift that occurred during data collection. Finally, to improve the signal-to-noise ratio, Principle Component Analysis (PCA) decomposition was applied to the HL spectra. This is a commonly used noise reduction method that has been shown to be effective when applied to EELS datasets [199, 200].

5.1.5 Use of binary masks as spectrum image region of interest tools

The SI picker tool in Digital Micrograph 3.6, which is used to select Regions Of Interests (ROI) and integrate the signal from groups of pixels, is limited to a square, therefore it is inefficient for sampling EELS SI regions that have an irregular shape, such as the encapsulin protein shell (see Figure 5.1) for example binary masks). To allow for better sampling of this region, the python library Sci-kit image was combined with the HyperSpy 2.0library [153, 166]. Using the multi-Otsu function within Sci-kit image a binary image of the ROI (the protein shell) was generated [165]. The multi-Otsu function identified a specified number of classes in an image by identifying the minimum point between peaks in an intensity histogram of the image. The generated binary mask was then multiplied with the original HL and LL dataset, to give masked spectra. The foreground pixels were then summed and the metadata copied from the original SIs. The resulting spectra could then be used for quantification.

5.1.6 HyperSpy quantification routine

The resulting spectra from binary masking were quantified using the HyperSpy package [166]. Before quantification the microscope parameters were applied (accelerating voltage, convergence angle (α), collection angle (β)) and when required the HL spectra were deconvolved with the LL spectra using the Fourier-ratio method (generally $t/\lambda < 0.3$, indicating single scattering conditions). The background was modelled using an inverse power law and the `smart_fit()` method used to fit the background and core loss edges. The edges were iden-

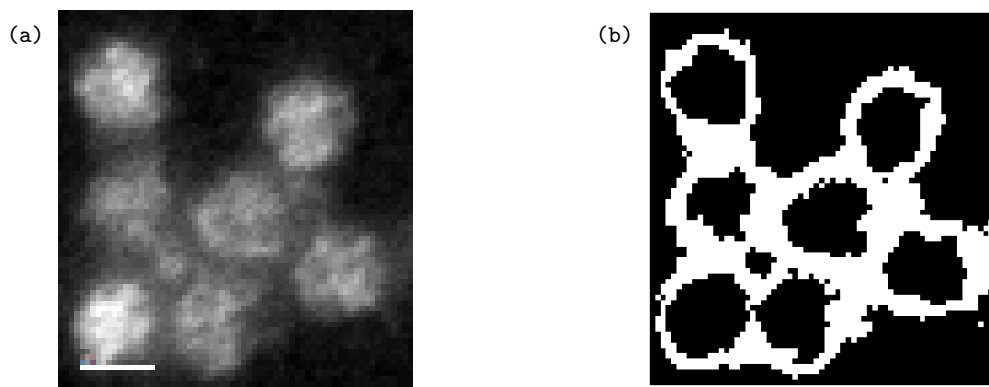


Figure 5.1: Example binary masks of encapsulin shell. (a) Reference ADF image, scale bar represents 20 nm. (b) Binary mask of encapsulin shell, generated from a single class defined using a multi-Otsu threshold [165]. The foreground in the binary mask (in (b)) matches the region of low contrast surrounding the bright Fe rich core in the ADF image (in (a)).

tified automatically using the package and the fitting parameters used can be seen in Table 5.2. The `smart_fit()` function is specific for fitting EELS models; it iterates along the energy-loss-axis, fitting first the background and then the edge sequentially until all edges are fitted simultaneously. Visual observation of the model and experimental data was used to confirm the quality of the fit before quantification was carried out.

5.1.7 Digital Micrograph quantification routine

For the majority of cases, quantification was carried out in the Digital Micrograph GMS 3.6 Suite (Gatan). Firstly, a region of interest (ROI) marker was centred over the core to generate a summed HL and LL spectra. Details of the edge and background fitting can be found in Table 5.2. Notably, the P $L_{2,3}$ and the S $L_{2,3}$ edges over-lap, therefore the S $L_{2,3}$ ELNES region is included in the quantification of P. For all other edges, a 40 eV gap is used after the edge onset to remove the ELNES region from the quantification. Finally, despite the optimal relative sample thickness (generally $t/\lambda < 0.3$ indicating single scattering conditions) the effects of multiple scattering were removed by low loss deconvolution using the Fourier-ratio method.

Edge	Total fit start	Total fit end	Edge on- set	Width, Δ	Back- ground model	ELNES exclusion gap
P L	88	254	132	122	Power law	-
S L	88	254	164	49	Power law	40
C K	215	397	283	72	Power law	40
N K	334	527	401	85	Power law	40
O K	452	703	532	131	Power law	40
Fe L	582	949	708	201	Power law	40

Table 5.2: Core edge fitting parameters used in both Digital Micrograph and HyperSpy. The numerical values are given in eV.

5.1.8 Estimation of the number of iron atoms

The number of Fe atoms per encapsulin was estimated based on a previously published method [201, 202]. The number of Fe atoms N was estimated using equation 5.1:

$$N = \frac{I_{Fe}(\beta, \Delta)}{I_0(\beta)\sigma_{Fe}(\beta, \Delta)} \quad (5.1)$$

where N is the areal density of atoms (atoms/nm²) contributing to the Fe L_{2,3} ionisation edge, $I_{Fe}(\beta, \Delta)$ represents the integral intensity of the Fe L-edge, I_0 is the zero loss integral, β is the collection angle, Δ is the signal integration width and $\sigma_{Fe}(\beta, \Delta)$ is the partial inelastic scattering cross-section for Fe. The resulting areal density was then multiplied by the area in nm to give an estimate of the number of atoms per encapsulin shell.

5.1.9 Quantification of the Ca L-edge

The Ca L-edge has an edge onset at 346 eV and sits on the high energy shoulder of the C K-edge (284 eV), thus its visibility is highly dependent on the concentration of C present in the sample. Thus to image the Ca L-edge, data was collected from samples on ultra-thin C film carrier grids (Agar Scientific) at low electron doses (typically $< 100 e^-/\text{\AA}^2$). After data collection, the spectra from two ROIs (only protein shells or core regions) were summed and

quantified using the method described in Section 5.1.5.

5.1.10 Multiple Least Squares Fitting

Multiple Linear Least Squares (MLLS) fitting was carried out in Digital Micrograph 3.6 (Gatan). MLLS is a useful method for both separating overlapping edges and Energy-Loss Near Edge Structure (ELNES) distribution mapping [203, 204]. In this work it was used to improve the spatial resolution in maps of the overlapping P and S L-edges. Before MLLS fitting was used, HL spectra were deconvolved with their LL using the Fourier-ratio method and the background was removed by modelling an inverse power law to the region preceding the edge onset: 90 eV - 130 eV for the P and S L-edges.

To separate the P and S L-edges, reference spectra were obtained either from the nanoparticles (P L-edge) or from the surrounding encapsulin shell proteins (S L-edge). A 15 eV window between 136 eV - 151 eV was used for the P L-edge, as this encompasses the first two peaks of the P L-edge. A 15 eV window of 165 eV - 180 eV was used to map the distribution of the S L-edge as this is the window where $L_{2,3}$ edge onset occurs.

5.1.11 Fe $L_{2,3}$ ELNES fingerprinting

To measure changes in the Fe $L_{2,3}$ ELNES, spectra were collected from three reference materials with known Fe valence states (ilmenite, magnetite and haematite) and from the experimentally generated constructs. Spectra were collected with a detector offset of 660, 680 and 700 eV (to check for detector stability) and at a dispersion of 0.1 eV/ch. This energy dispersion was selected as it accounts for the natural line width of the Fe L-edge which is ~ 0.36 eV [131]. Spectra from experimental samples were collected at a range of electron doses to check for beam induced reduction of the Fe in the sample. Following data acquisition, spectra were deconvolved with the simultaneously collected LL spectra, and the background was removed by fitting an inverse power law to the region immediately preceding the edge onset. Experimental spectra were then aligned with the spectra from the reference materials to compare the shape of the Fe $L_{2,3}$ edge.

5.1.12 Quantification of the Fe $L_{2,3}$ fine structure

The valence state of the Fe $L_{2,3}$ edge was quantified using the ratio of L_3 to L_2 method outlined in the literature [142, 205]. This method involves integrating Fe $L_{2,3}$ net intensity under two 2 eV wide windows, 708 - 710 eV for L_3 and 719.7 eV - 721.7 eV for L_2 . Firstly, a universal curve was generated by plotting $L_3 / (L_3 + L_2)$ against the position of the maxima of the L_3 edge for the three reference materials: ilmenite (Fe^{2+}), magnetite (Fe^{2+}/Fe^{3+}) and haematite (Fe^{3+}). The data from experimental samples could then be plotted against this curve to estimate the ratio of Fe valence states in the encapsulin cores.

5.2 Results

5.2.1 Chemical maps of encapsulin samples from EELS SIs

As was identified using Energy Dispersive X-ray (EDX) analysis, the three elements P, O and Fe are enriched in the nanoparticles inside encapsulins. However, using EDX no chemical information could be collected about the proteinaceous shell, due to a combination of the techniques low sensitivity to light elements and the short dwell times used during data collection. EELS is not limited by these factors and thus can be used to map the encapsulin shell at high spatial resolutions [132, 125].

The encapsulin shell is composed of C, N, O and S (see Table 5.4). The S atoms are located in the cysteine residues at positions 121 and 276 of the EncA protein sequence [18]. These residues form disulphide bonds, which contribute to the structural stability of the protein. All of the elements have core loss edges within 600 eV of each other (in the energy loss range of 132 - 703 eV), thus the necessary signal could be captured in a single HL spectra using an offset of 132 and a dispersion of 0.25 eV/ch.

Chemical maps of encapsulins from *M. xanthus* (Figure 5.2) and constructs generated in *E. coli* (Figures 5.3, 5.4 and 5.5) reveal that the elements P, O and Fe are localized to the nanoparticles. This finding is consistent with previous EDX results (see Chapter 4). Additionally, due to EELS having a higher sensitivity for light elements, the intensity maps for C, N and S reveal a ring structure encircling each of the Fe-rich nanoparticles. Notably, the total signal in the C map is much greater than that of the other elements because the continuous C film on the carrier grid creates a large background signal. Despite the high background signal the characteristic ring can still be seen from the protein C signal. Additionally, in the EncA+EncC C map the ring of the protein shell appears more diffuse than in the other samples, because of the increased levels of contamination that built up during data collection. The presence of this structure is a good indicator that under the imaging conditions used on the JEOL ARM200F the protein shell is not destroyed during data collection.

Finally, when a sample of empty EncA+EncB+EncC encapsulin were incubated with $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and agitated, mineralization occurred within the lumen of the encapsulin.

The chemical maps of the resulting structures demonstrate that the mineralization has occurred within the EncA protein shell. As there is increased signal for C, N and S in the same region as the Fe rich core. However, the spatial resolution is not as high as for the other samples, meaning the characteristic ring structure can not be resolved. Furthermore, the maps also demonstrate that these conditions effectively produce nanoparticles devoid of P indicating that nanoparticle formation is not dependant on the presence of P inside the encapsulin shell.

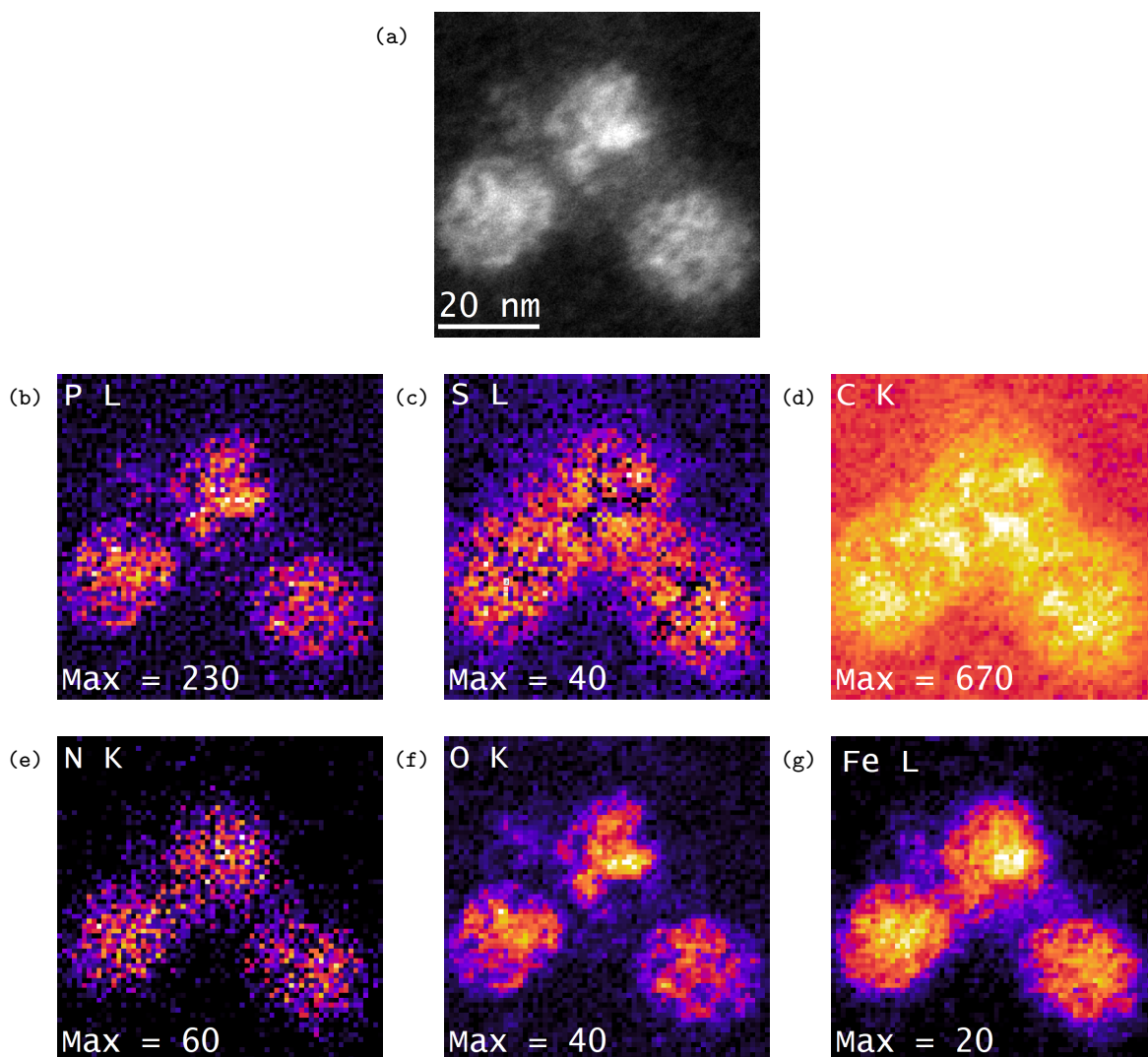


Figure 5.2: Chemical maps generated from EELS spectra of encapsulins from *M. xanthus*. (a) HAADF reference image of sampled region. The reference image is 528x528 pixels and the SI maps are 66x66 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction.

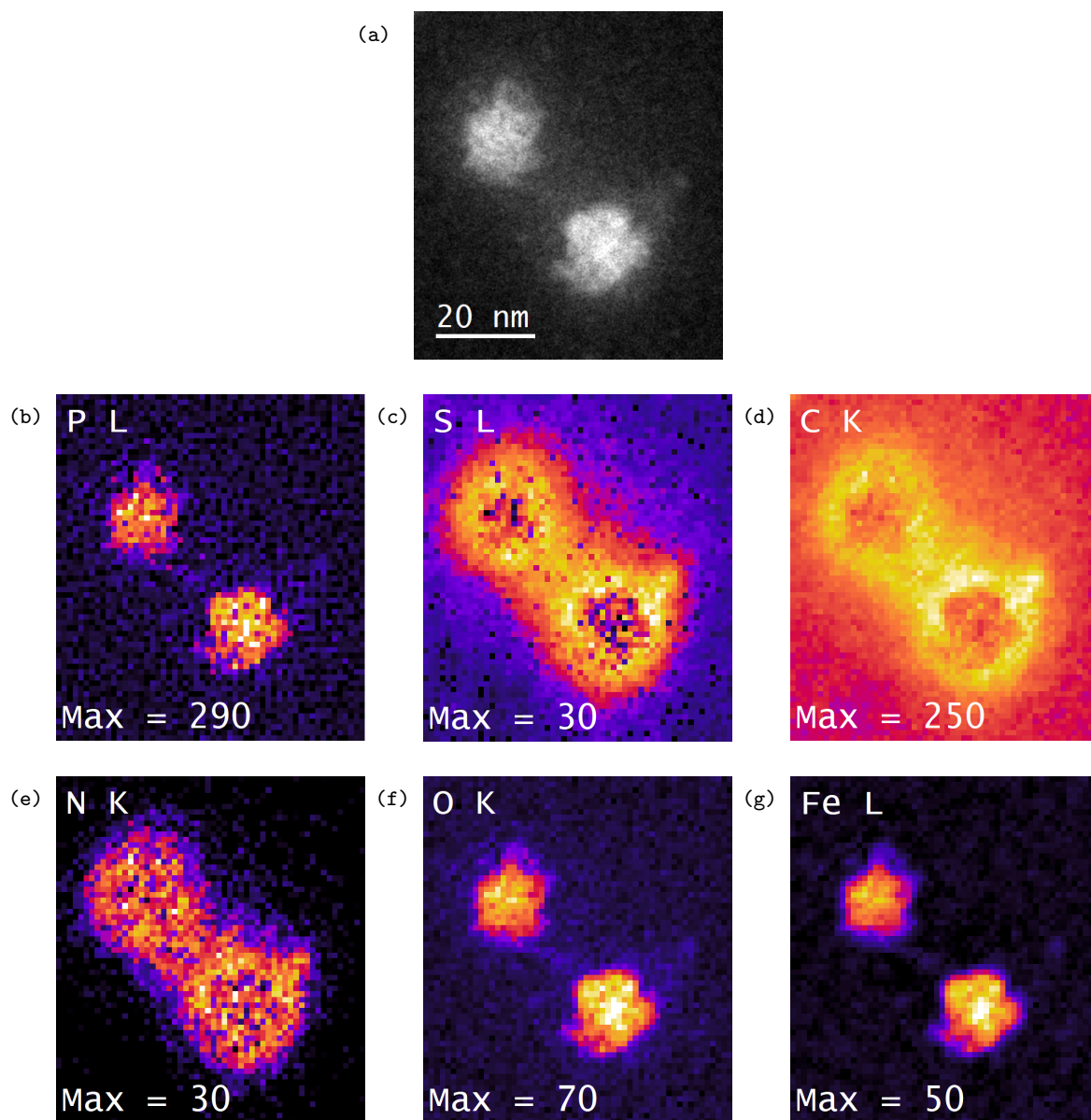


Figure 5.3: Chemical maps generated from EELS spectra of the EncA+EncB encapsulins. (a) HAADF reference image of sampled region. The reference image is 408x544 pixels and the SI maps are 56x63 pixels. (b) - (g) Elemental maps showing the spatial distribution of P $L_{2,3}$, S $L_{2,3}$, C K, N K, O K and Fe $L_{2,3}$. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction.

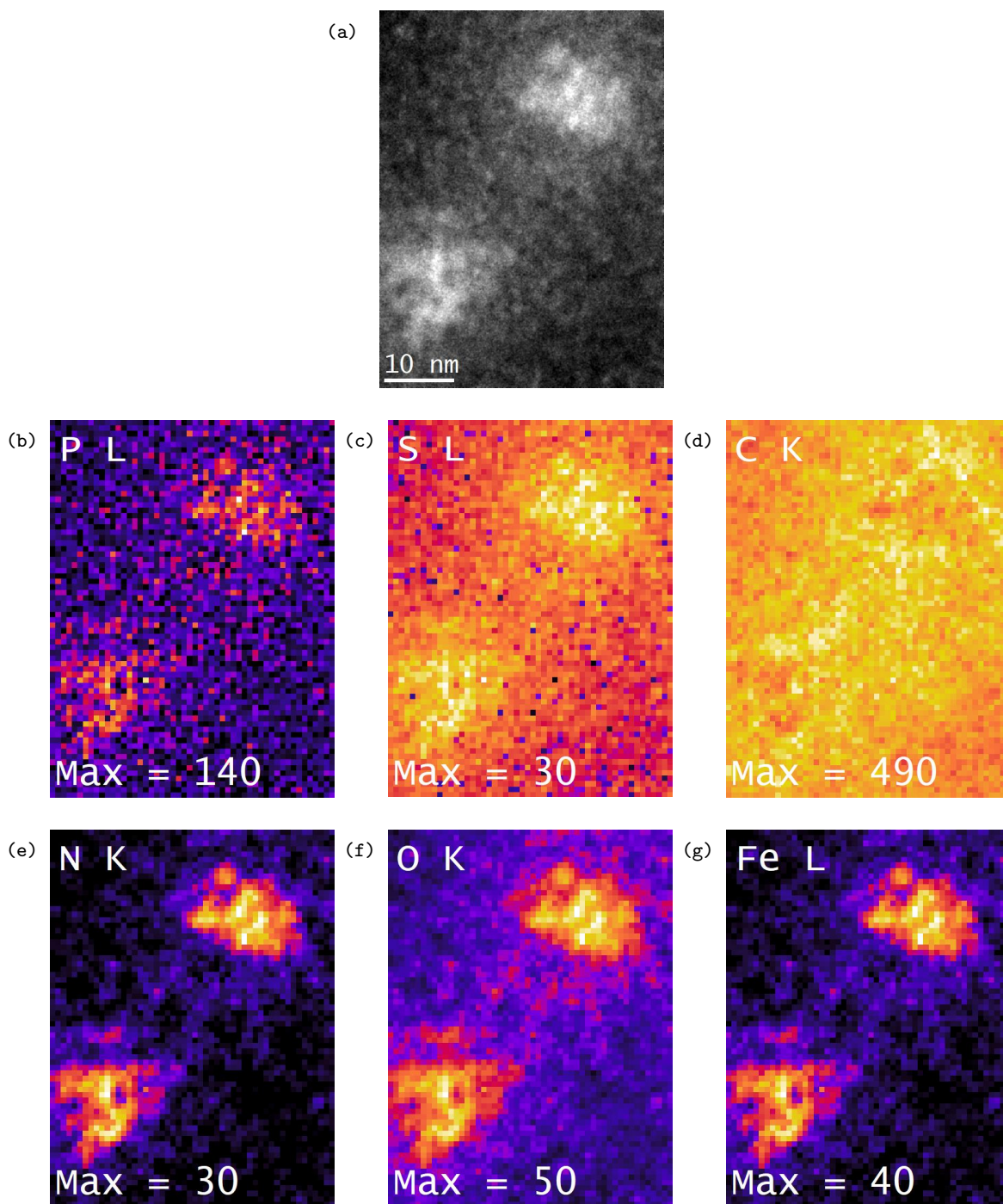


Figure 5.4: Chemical maps generated from EELS spectra of EncA+EncC encapsulins. (a) HAADF reference image of sampled region. The reference image is 416x552 pixels and the SI maps are 52x69 pixels. (b) - (g) Elemental maps showing the spatial distribution of P $L_{2,3}$, S $L_{2,3}$, C K, N K, O K and Fe $L_{2,3}$. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction.

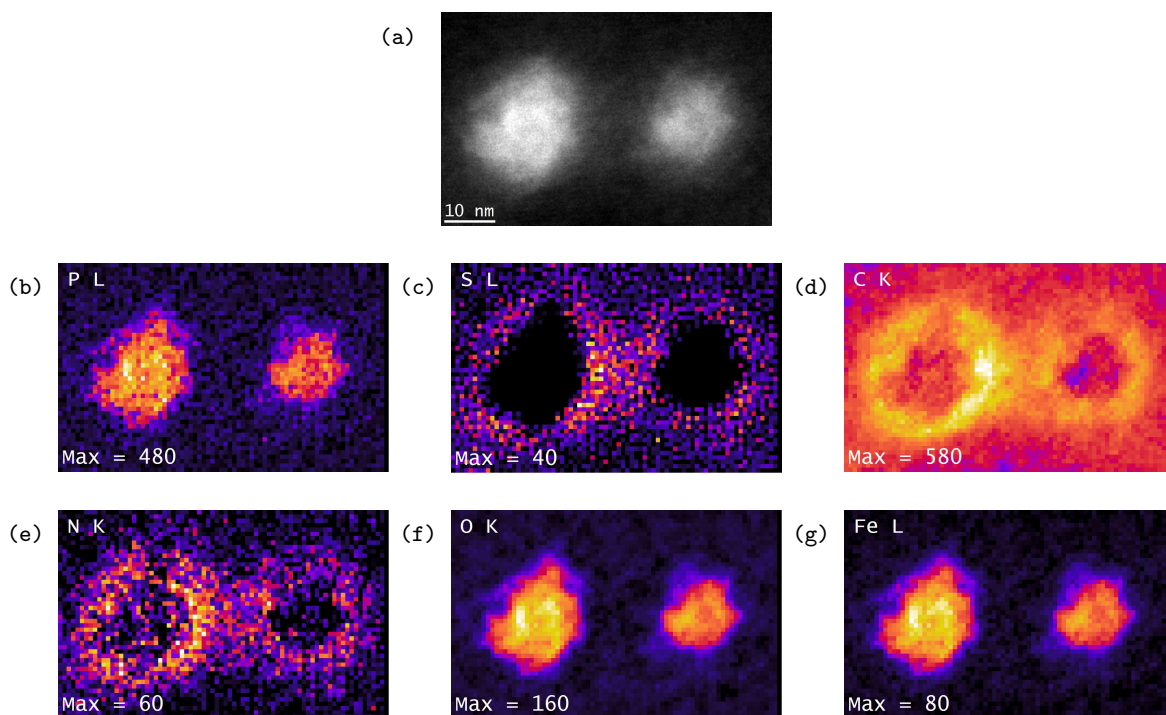


Figure 5.5: Chemical maps generated from EELS spectra of EncA+EncB+EncC encapsulins. (a) HAADF reference image of sampled region. The reference image is 608x392 pixels and the SI maps are 76x49 pixels. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.25 eV/ch. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction.

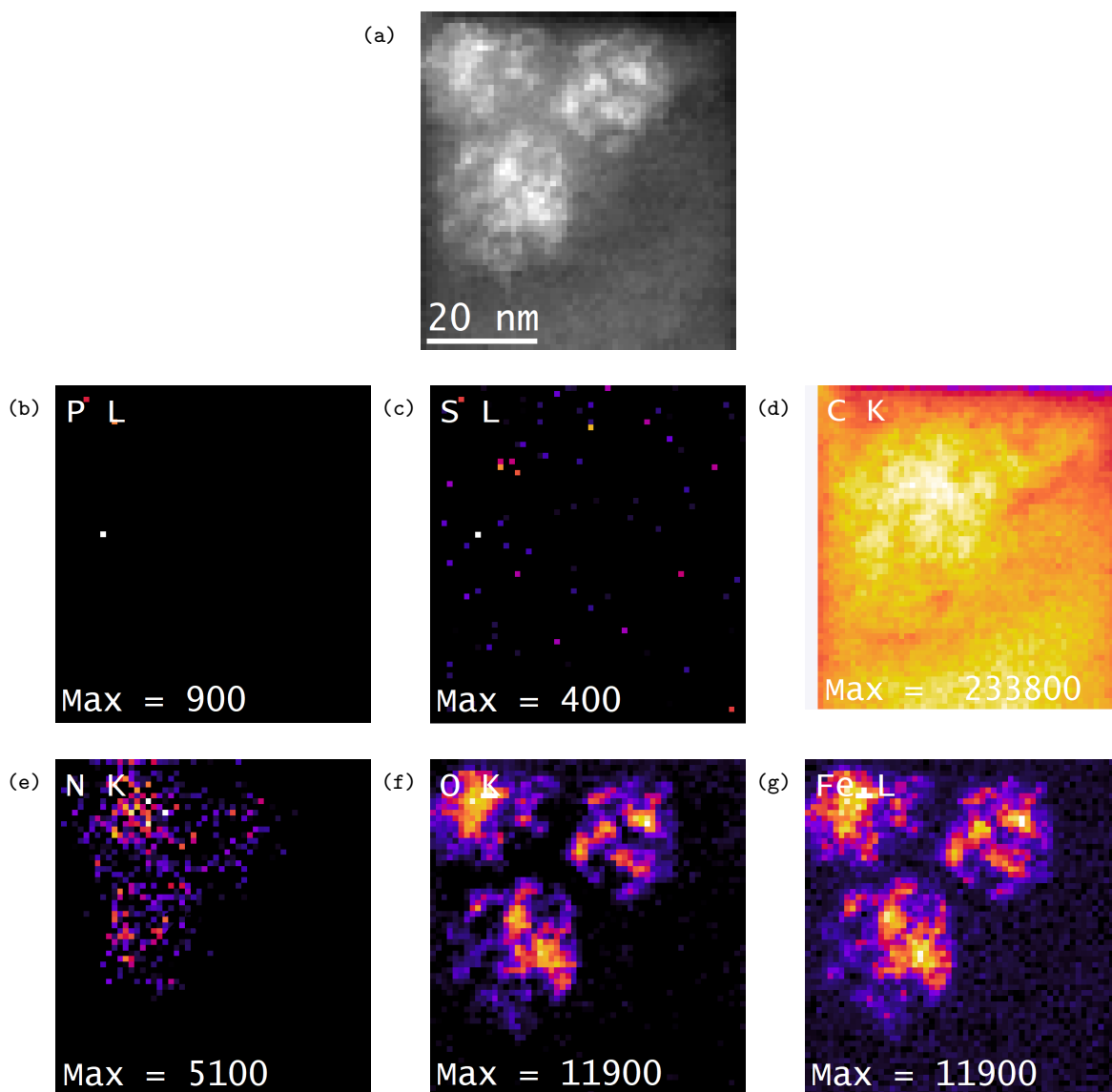


Figure 5.6: Chemical maps generated from EELS spectra of EncA+EncB+EncC encapsulins loaded with Fe *in vitro* (a) HAADF reference image of sampled region. (b) - (g) Elemental maps showing the spatial distribution of P L_{2,3}, S L_{2,3}, C K, N K, O K and Fe L_{2,3}. The SI was captured with an offset of 132 eV and a dispersion of 0.5 eV/ch. This SI was captured on the JEOL F200 operating at 200 keV. The max count value represents the pixel with the highest counts value following deconvolution and background subtraction.

5.2.2 Separation of the P $L_{2,3}$ and the S $L_{2,3}$ edges using MLLS fitting

In the majority of the chemical maps shown previously, the location of the P and S signals can be easily separated. However, in the maps of the *M. xanthus* encapsulins they both appear to be localized across the nanoparticles and shell (see Figure 5.2). This may be in part due to the overlapping nature of the P $L_{2,3}$ and S $L_{2,3}$ core loss edges (see Figure 5.7b) making it difficult to separate them when using the integration method [125]. To address this problem, and improve the spatial resolution of the P and S chemical maps, Multiple Least Squares (MLLS) fitting was used. MLLS is a commonly used method for separating overlapping edges and works by mapping known internal reference spectra to unknown regions [203, 206]. It has been applied successfully to a number of cases including that of the P $L_{2,3}$ and S $L_{2,3}$ edges [204]. In this instance, a reference spectra for the S $L_{2,3}$ edge was generated from the surrounding ring of protein shell and the P reference from the nanoparticle with the largest number of P counts.

The result of MLLS fitting significantly improved the spatial resolution compared to chemical maps generated using the integration method of the P and S signals (see Figure 5.7). The enhanced spatial resolution reveals that the two elements are localized to different regions, the P to the nanoparticles and the S to the spaces in between. This distinction confirms that P is directly associated with the nanoparticles, rather than being confined to the protein surfaces inside the nanoparticle shell. Finally, the localisation of S to the spaces in between the nanoparticles indicates that these are positions where the beam passed directly through the encapsulins without interfering with any of the nanoparticles, an interesting observation that highlight the low occupancy of the shell by nanoparticles.

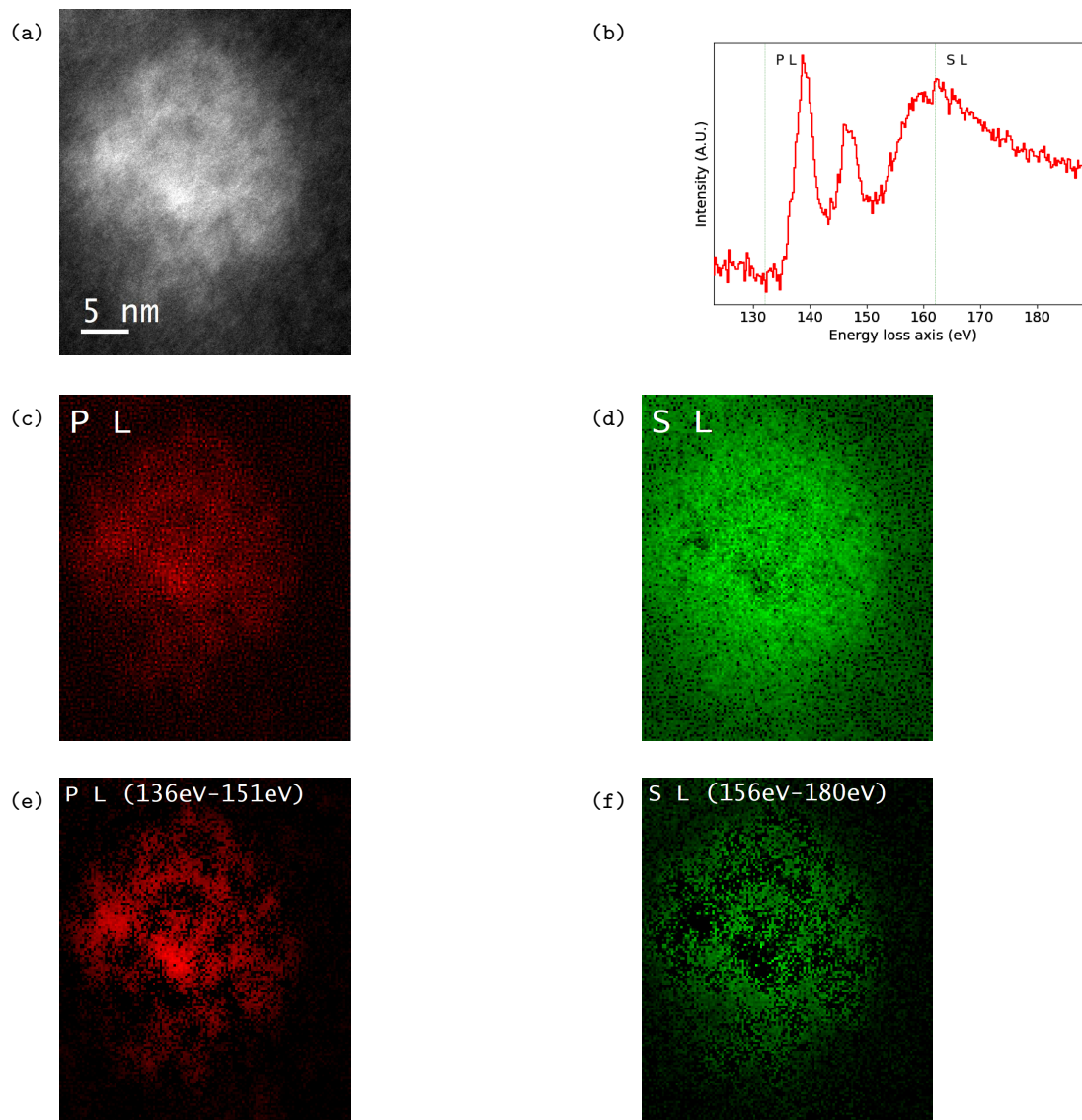


Figure 5.7: Comparison of P and S localisation within a single *M. xanthus* encapsulin. (a) HAADF reference image of single encapsulin. (b) P L-edge map of counts. (c) S L-edge map of counts. (d) Energy loss spectrum covering range of 125 - 190 eV from a single encapsulin. The overlapping P and S L-edges can be clearly seen. The edge onsets are labelled with green dotted lines. (e) Map of P distribution by MLLS fitting over 15 eV window (136-151 eV). (f) Map of S distribution by MLLS fitting over a range of 15 eV (165-180 eV), showing the anti-correlation to the P distribution.

5.2.3 Observation of the presence of Ca in *E. coli* generated constructs

During EELS data collection fast acquisition times were used to limit the build up of C contamination thus retaining spatial resolution. However, the effect of this was low signal to noise ratio. Thus, to identify small spectral features pixels representing larger regions had to be summed.

The result of this, is the identification of a Ca signal from the cores of the encapsulin complexes (see Figure 5.8). Ca was identified only in the encapsulins that had been expressed and loaded with Fe *in vivo*, it was not found associated with either the encapsulins isolated from *M. xanthus* or with the empty shells that were later loaded with Fe *in vitro*. The observation of Ca in encapsulins purified from *E. coli* and not in encapsulin isolated from *M. xanthus* may be due to the differences in the amount of Ca present in the cytoplasm of either bacteria. It is currently believed that *E. coli* cells contain 90 ± 10 nM of Ca [207], however, similar measurements have not been made for *M. xanthus*. Thus it is impossible to identify if the concentration of Ca in the cytoplasm can explain this observation. Furthermore, Ca was not found associated with the protein shell in any of the spectra, indicating that it is only associated with inorganic material in the core. The absence of Ca association with the protein shell is interesting, given that Ca is frequently associated with proteins and has previously been detected bound to the encapsulin cargo proteins following overexpression and crystallization [208, 209, 28]. The Ca content of the core was quantified and it was identified that the atomic% of Ca ranged between 0.6% in EncA+EncB+EncC to 2.6% in EncA+EncB of the total core material (see Table 5.3).

Initially, the presence of Ca was believed to have derived from the Phosphate Buffered Saline (PBS) that was being used in the elution buffer. However, after changing the elution buffer to a Tris buffer (which is Ca free) the presence of Ca in the core persisted. Meaning that the Ca accumulation occurs during the Fe mineralization inside the *E. coli* cell.

Ca is commonly found in bacterial cytoplasm and has been implicated in several important cellular processes including maintaining the cell cycle and motility [210, 211, 212]. Furthermore, it has been found associated with the *M. xanthus* cargo proteins following their overexpression in *E. coli* and crystallization [28]. However, its presence in the core is unexpected, especially as no Ca is found in the isolated empty encapsulins. Intracellular Ca and Fe

both have valence states of $2+$, and have similar atomic radii of 140 and 180 pm respectively [213]. Thus, it is likely that they can both traverse the pores in the encapsulin shell. As the encapsulins expressed in *E. coli* were incomplete - missing one or two cargo proteins - this may confer a lack of control over what ions have access through the pores.

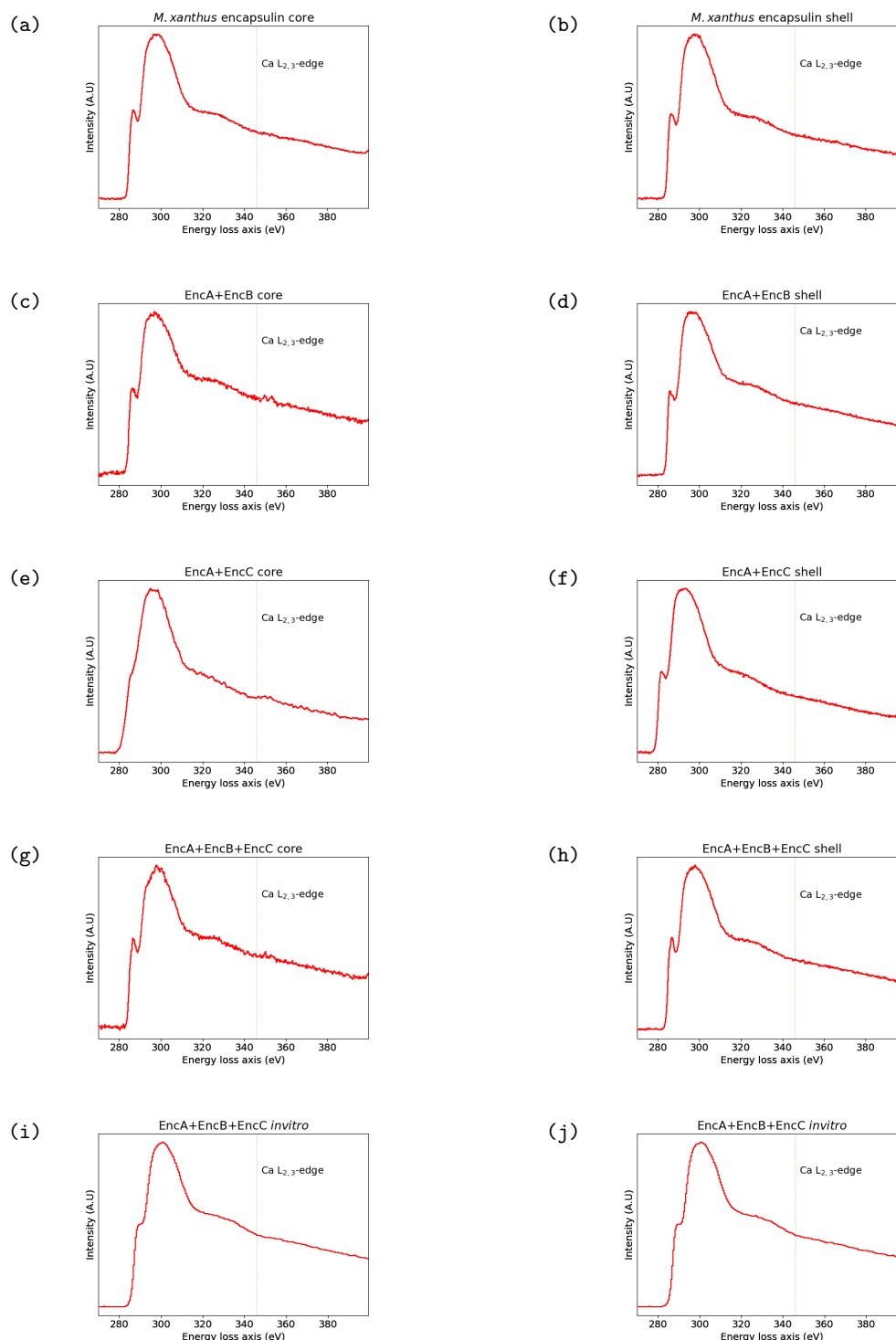


Figure 5.8: The Ca $L_{2,3}$ edge can be observed in the core regions of samples generated in *E. coli*. Spectrum from the following constructs are shown: *M.xanthus*: (a) core, (b) shell. The following samples have all been expressed in *E.coli*. EncA+EncB: (c) core, (d) shell. EncA+EncC: (e) core, (f) shell. EncA+EncB+EncC (g) core, (h) shell. EncA+EncB+EncC loaded *in vitro* (i) core, (j) shell. The green dotted line marks the onset of the Ca $L_{2,3}$ edge. Individual EELS spectra are shown instead of chemical maps, because of the difficulty in modelling the Ca edge due to its low signal and proximity to the C K-edge.

Construct	Ca at.%	$\pm$ (standard deviation)
<i>M. xanthus</i>	0.0	0.0
EncA+EncB	0.6	0.1
EncA+EncC	2.6	0.3
EncA+EncB+EncC	0.3	0.1
EncA+EncB+EncC + <i>in vitro</i>	-	-

Table 5.3: Atomic% of Ca in the core regions of encapsulin constructs. Pixels representing encapsulin cores were summed across all frames in a single data set. Atomic% was calculated using HyperSpy.

5.2.4 Calculations of the atomic composition of the encapsulin shell

The HyperSpy 2.0python package provides the ability to generate custom shaped ROIs based on image thresholding meaning that for complex shaped regions, such as the ring shape of the protein shell, the whole region can be summed easily into one spectrum. This increases the total signal being used for analysis when compared with the limited square ROI in Digital Micrograph. However, despite the difference in ROI both routines agree on the atomic composition of the shell, within error bars. In both instances, the C signal is dominant, equalling $\sim 94\%$ of the total atomic composition (see Table 5.5). This elevated percentage compared with the calculated percentages for the EncA monomer (Table 5.4) is the result of the additional signal from the C film on the carrier grid. Also in the atomic percentages calculated from the protein sequence there are equal amounts of O and N, however both routines give twice the percentage composition of O compared to N. This difference may be accounted for by either the inclusion of a number of pixels from the nanoparticles inflating the O signal, or the enhanced difficulty in modelling the power law background for the N K-edge (edge onset: 401 eV) from the trailing end of the C K-edge (edge onset: 284 eV) upon which it sits. Finally, despite the low number of S atoms present, their signal from the S L-edge could be identified, measured and mapped over the correct region. Thus, both of these routines are useful tools for quantifying the atomic composition of the encapsulin shell.

Element	No. of atoms	Relative comp.	At. %
C	1452	1	64
N	422	0.29	18
O	432	0.3	19
S	10	0.006	0.5

Table 5.4: Atomic composition of EncA (Gene Identifier: MXAN_3556). The number of atoms, relative composition, and percentage composition of a single EncA monomer calculated using the Xpansy ProtParam tool [214]. The data for H has been omitted.

-	HyperSpy		Digital Micrograph	
Element	Atomic%	$\pm$ (standard deviation)	Atomic%	$\pm$ (standard deviation)
C	93	5.8	95	1.6
N	2	0.78	2	1
O	4	1	4	1.2
S	1	0.2	0.4	0.2

Table 5.5: Atomic composition of the encapsulins protein shell calculated using routines in both Digital Micrograph and HyperSpy. The protein shells from samples isolated from *M.xanthus* were used for analysis.

5.2.5 Loss of protein shell material due to beam damage during scanning

The calculations of the atomic composition of the encapsulin shell made above were from the characteristic ring structure surrounding the Fe rich core, as this is the thickest protein region and provides the strongest signal. However, the shell encapsulates the nanoparticles, thus there is a 3 nm layer of protein above and below the nanoparticles that must be considered during quantification. As proteins are highly likely to suffer ionization damage altering their structure when imaging using a STEM instrument [215], the stability of this 3 nm layer was measured over increasing electron dose (see Figure 5.9).

In situ data collected using the JEOL ARM200F reveals the instability of the encapsulin shell under increasing electron dose. The first pass of the electron beam ($34\text{ e}^-/\text{\AA}^2$) produced maps where signal from N and S is distributed across the entire encapsulin, indicating that the shell is present. After the second data frame was collected ($75\text{ e}^-/\text{\AA}^2$), in both the S and N maps an empty space has formed across part of the central region of the encapsulin, indicating a loss of the protein shell in this region. In the final data frame ($120\text{ e}^-/\text{\AA}^2$), there is further degradation in the central region clearly visible in the S and N maps. Additional changes can be seen to the of ring of protein that surrounds the core, indicating the loss of this part of the structure as well. Notably, in the last frame, distinct changes are visible on the surface of the O containing core, which may be the start of beam damage to the core.

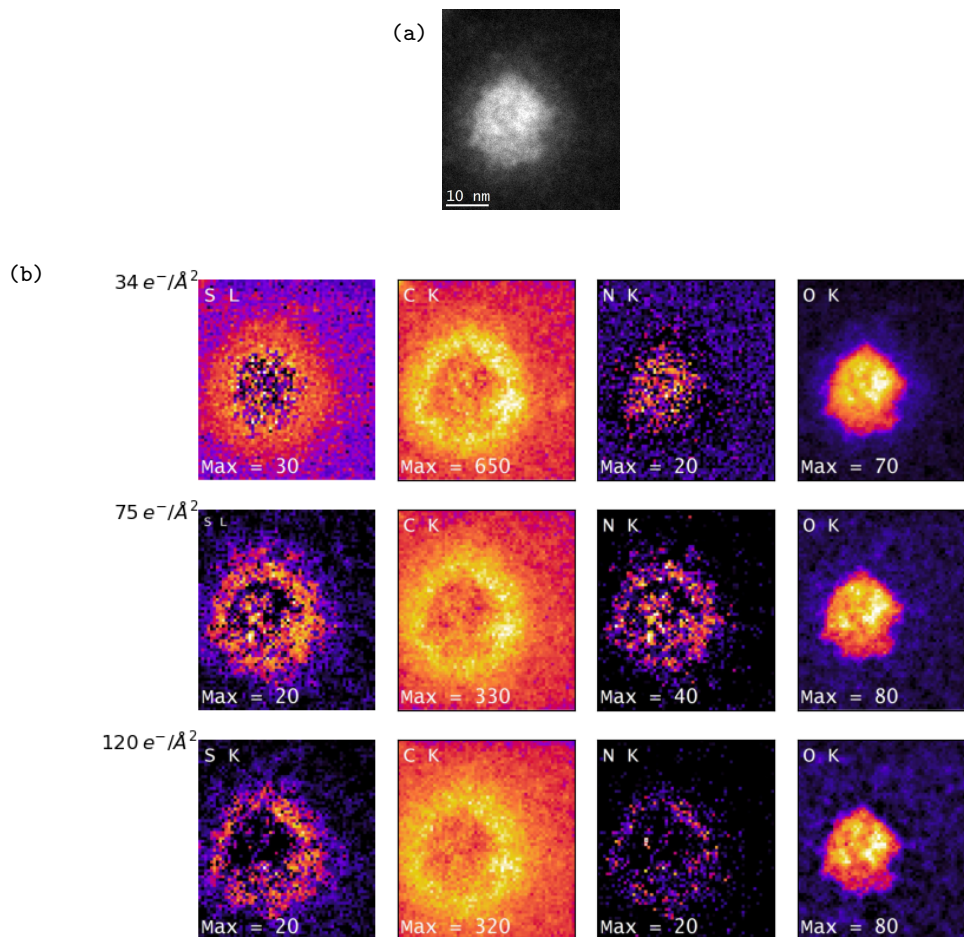


Figure 5.9: Chemical maps showing the loss of the encapsulin shell following exposure to increasing electron dose. (a) HAADF reference image of sampled encapsulin. The encapsulin imaged here is composed of EncA+EncB loaded with 8 mM Fe in the presence of 0.01 mM of the inducer in *E. coli*. (b) Chemical maps generated from the S L_{2,3}, C K, N K and O K-edges. The electron dose increases from $34 e^-/\text{\AA}^2$ in the first row and ends at $120 e^-/\text{\AA}^2$ in the final row.

5.2.6 Calculation of the atomic composition of the encapsulin derived nanoparticles

The nanoparticles from all constructs generated were quantified from EELS spectra using the routines present in Digital Micrograph. It was identified that the nanoparticles inside encapsulins isolated from *M. xanthus* have a general chemical formula of FePO_4 , and an Fe:P ratio of $1.2 \pm 0.2:1$ (see Table 5.6). This Fe:P ratio differs significantly from previously made measurements where a Fe:P ratio of 4:1 was identified using inductively coupled plasma mass spectroscopy [18]. The difference in the Fe:P ratio could explain the smaller nanoparticles observed in these structures compared with those published in the literature (see Section 3.2.3), as it is known that the presence of P impacts the size and shape of iron oxide nanoparticles synthesised both chemically and inside ferritin protein cages [179, 180].

The encapsulin constructs generated in *E. coli* were expressed in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and 0.05 mM of the inducer molecule IPTG. We found a difference in atomic composition, both with the nanoparticles imaged from *M. xanthus* as well as between the individual constructs. The nanoparticles from EncA+EncB and EncA+EncB+EncC have general formulas of FeO_5P_2 whilst those from EncA+EncC have a general formula of FeO_5P_4 (see Table 5.6).

In all encapsulins generated inside the *E. coli* expression system the nanoparticles have higher concentrations of O and lower concentrations of P when compared with the native assembled encapsulin from *M. xanthus*. This lower amount of P measured may be due to a number of different reasons: variations in the concentration of cytosolic phosphate anions between the two different hosts, fluctuations caused by the status of the host cell as protein overexpression is considered a highly stressful state for a cell or from differences in the amounts of phosphate in the growth media used to growth the two strains. *E. coli* is grown in LB media, which contains both Tryptone and Yeast Extract, whilst *M. xanthus* is grown in CTT media that contains both Casitone and KH_2PO_4 , all of these can be considered sources of P. However, as Tryptone, Casitone and Yeast Extract are natural products the exact contribution of these compounds to the total P content of the growth media is unknown.

There are also differences in the Fe:P ratio between EncA+EncB/EncA+EncB+EncC and EncA+EncC, with a Fe:P ratio of 2:1 in the former and 4:1 in the latter (see Table 5.6).

Whilst scanning the beam across the sample a higher concentration of C building up with scan time can be observed in the C chemical map of EncA+EncC compared with those of the other constructs (see Figure 5.4). The increase in this material causes a broadening of the plasmon peak in the P L-edge pre-edge region. The ramification of this broadening is a more difficult background fit using an inverse power law for the P L-edge and therefore increased error in its calculation, which may explain part of the observed variance between the Fe:P ratios. Especially, as the Fe:P ratio measured for EncA+EncB and EncA+EncB+EncC are the same. SDS quantification of the number of cargo proteins in the encapsulin containing both cargos suggests that shells contain roughly equal numbers of EncB and EncC (see Section 2.3.8). Thus, it is expected that the Fe:P ratio of EncA+EncB+EncC would be 3:1 however, this is not what was measured.

5.2.7 Nanoparticles within a single shell have the same atomic composition

The use of spherical aberration correction allows higher probe currents to be maintained at smaller probe sizes in STEM, enhancing spectroscopy at higher spatial resolutions [216, 127] whilst the lower electron dose required for signal generation in EELS compared to EDXS results in lower C contamination build-up during data collection, thus retaining spatial resolution. The combination of these two factors has allowed for individual single nanoparticles to be resolved and characterized within a single encapsulin from *M. xanthus*.

We identified that within a single encapsulin shell of *M. xanthus* all of the sampled particles had the same atomic composition of FePO_4 (see Figure 5.10). However, there was some variation between different shells. Additionally, the presence of nanoparticles with the same atomic composition in a single *M. xanthus* shell provides further evidence that the products of the two cargo proteins, EncB and EncC are the same.

Host	Pro- teins	At% P	$\pm$ stdev (er- ror bar)	At% O	$\pm$ stdev	At% Fe	$\pm$ stdev	Fe:P	$\pm$ stdev
<i>M. xanthus</i>	EncA, EncB, EncC, EncD	12	4	72	3	15	2	1.2	0.2
<i>E. coli</i>	EncA, EncB	8	1	74	4	16	4	2	0.3
	EncA, EncC	4	1	79	2	16	2	4	1.8
	EncA, EncB, EncC	7	1	78	2	14	2	2	0.3
<i>E. coli</i> + <i>in vitro</i>	EncA, EncB, EncC	0	0	64	3	36	3	-	-

Table 5.6: Atomic composition of the iron containing nanoparticles of generated encapsulins. Stdev is one standard deviation.

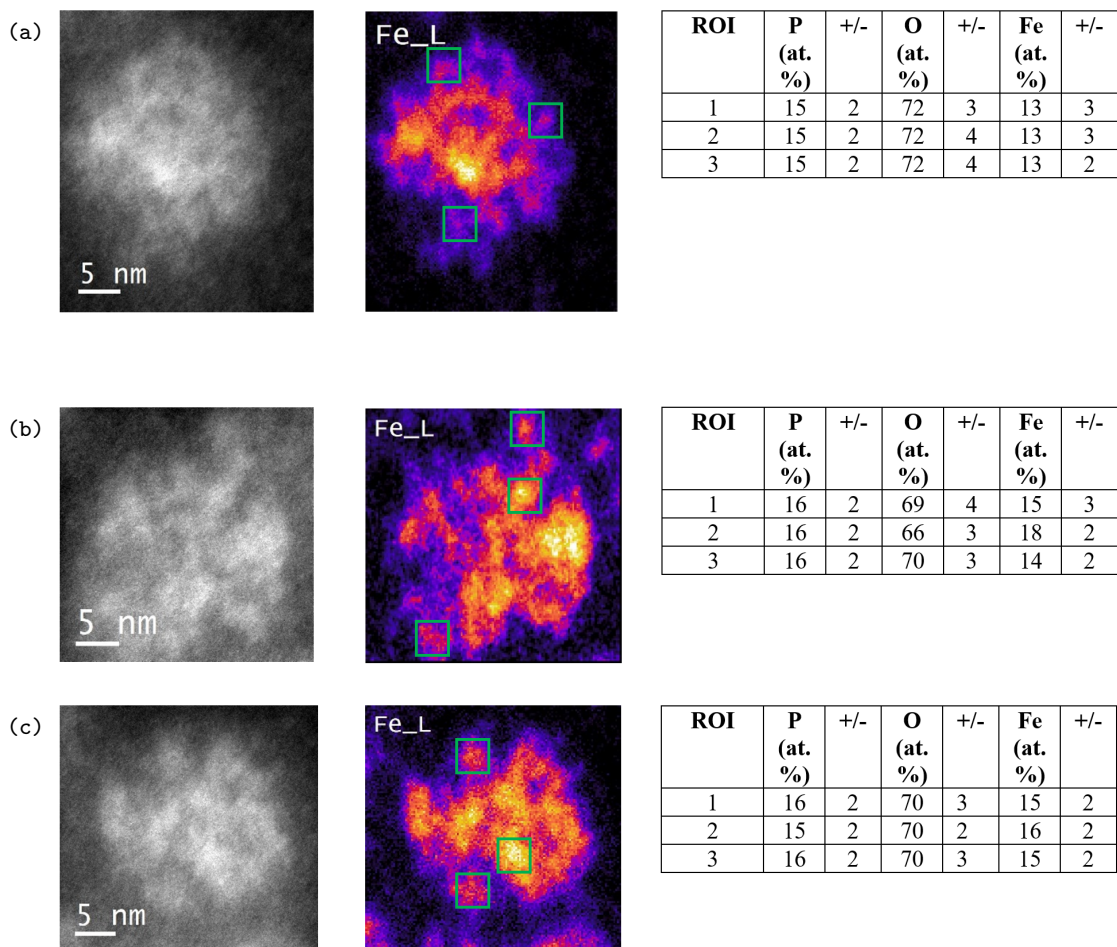


Figure 5.10: Chemical composition of individual nanoparticles within encapsulin shells from *M. xanthus*. Three examples of individually resolved nanoparticles are shown. Left to right: HAADF reference image of sampled encapsulin, map generated from the Fe $L_{2,3}$ edge with the sampled individual nanoparticles highlighted with green boxes and tabulated atomic composition of the highlighted nanoparticles. Sampled ROIs listed in the table moving down the image from top (ROI 1) to bottom (ROI 3).

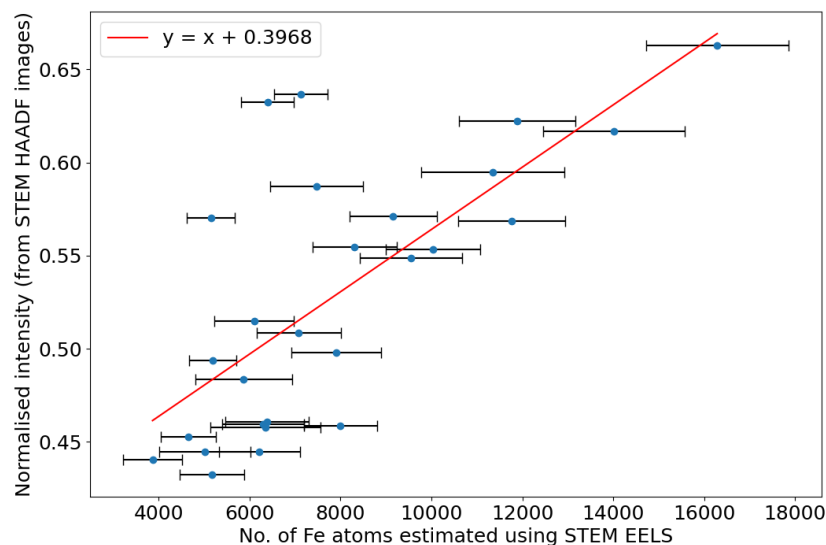


Figure 5.11: Normalised HAADF intensity against the estimated number of Fe atoms inside encapsulins from *M. xanthus*. A linear relationship is expected between the HAADF intensity and no. of Fe atoms [202]. A linear fit to the experimental data yields an R^2 value of 0.504. The relatively poor fit and the observed offset of 0.3968 are most likely due to additional contrast contributions from both the protein shell and the carbon coating of the support film.

5.2.8 Estimates of the number of Fe atoms inside *M. xanthus* encapsulins compared with HAADF intensity

The number of Fe atoms per encapsulin was estimated for 30 *M. xanthus* encapsulins. Encapsulins were found to contain between ~ 4000 and ~ 17000 Fe atoms with an average of $\sim 8000 \pm 1000$ (see Figure 5.11). The largest volumetric density estimated for the number of Fe atoms was 4.7 atoms/nm^3 . Thus an upper estimate for the number of Fe atoms per 1.8 nm nanoparticle is ~ 15 atoms (see Section 3.2.3). Using the previous calculation for the maximum number of nanoparticles that can be stored: 2000. The maximum number of Fe atoms that can be stored in a single shell taking into account the internal volume filled by the cargo proteins is ~ 26000 . However, a more realistic estimate, that accounts for spacing between the nanoparticles would be ~ 9000 . The estimates made here are lower than the previous estimate of ~ 30000 made in the literature using STEM mass measurements [18] as they account for loss of space due to the cargo proteins and the inter-particle spaces.

5.2.9 Comparing estimates of the number of Fe atoms inside encapsulins over expressed in *E. coli*

The number of Fe atoms per shell was estimated for 30 of each encapsulin construct generated in *E. coli*. The average number of Fe atoms estimated per shell for the EncA+EncB construct was 16000 ± 11000 , for the EncA+EncC construct an average of 13000 ± 7000 Fe atoms was estimated and the EncA+EncB+EncC construct was estimated to have an average of 14000 ± 8000 (see Figure 5.12). Despite the large difference in spread between the data sets, no statistical significance was identified. Additionally, the estimated average number of Fe atoms for the encapsulins loaded *in vitro* was 15000 ± 8000 . The difference between the number of Fe atoms per shell in this samples and those loaded *in vivo* was found to not be statistically significant, suggesting that the method by which iron is supplied to the encapsulin is not the most important factor in defining the number of Fe atoms loaded per shell. However, a statistically significant difference was identified between the number of Fe atoms in the constructs and in the native structure. Suggesting that addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the growth media resulted in an over loading of the encapsulins compared to what is found in *M. xanthus*.

Finally, the encapsulin containing the highest number of Fe atoms measured contained only the EncB protein. Here the shell was estimated to contain ~ 38000 Fe atoms. This estimate exceeds previous estimates for the maximum number of Fe atoms that can be stored per shell of ~ 30000 suggesting that the maximum number is higher than this estimate [18]. However, it should be noted that this structure was lacking two of the cargo proteins and thus there is more space inside the lumen than in naturally occurring structures so it does not reflect what would happen inside a normal *M. xanthus* cell.

5.2.10 Measurements of individually distributed particles

As previously discussed, in some of the isolations of encapsulins generated in *E. coli* encapsulin shells containing single nanoparticles were observed (see Section 3.2.10). Although they appear morphologically similar to nanoparticles observed in more densely packed cores elucidation of their chemical composition is important for confirming they are the same. However their size provided several challenging for analysis with a STEM instrument. To

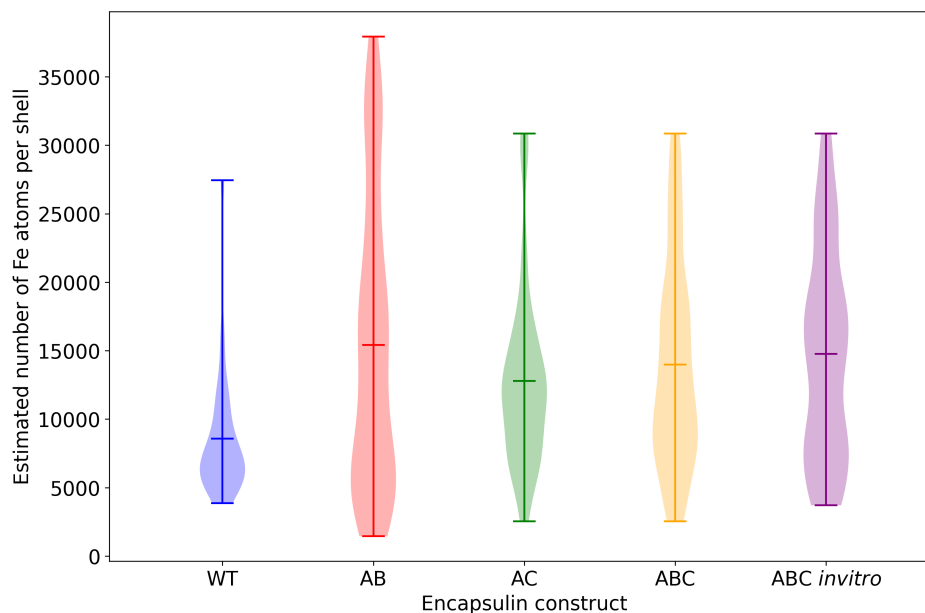


Figure 5.12: Estimates of the number of Fe atoms per encapsulins shell for the different encapsulin constructs generated in this work. For each dataset 30 encapsulins were measured. **MX** (blue) = Complete encapsulins isolated from *M. xanthus*. **AB** (red) = EncA+EncB, **AC** (green) = EncA+EncC **ABC** (yellow) = EncA+EncB+EncC. For AB, AC and ABC the encapsulins were over-expressed in *E. coli* grown in media containing 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ and 0.05 mM of the inducer IPTG. **ABC *invitro*** (purple) = EncA+EncB+EncC construct, empty encapsulin constructs were isolated and loaded with 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ *in vitro*.

address this, firstly, a spherical aberration corrected machine was used, as it allows for smaller probe sizes and thus a higher density measurements per particle. Secondly, low pixel dwell times - and thus low electron doses - were required to prevent the accumulation of damage and loss of the material.

As can be seen in Figure 5.13, the individual nanoparticles consisted of P, O and Fe. However, the nanoparticles were destroyed by the electron beam after an accumulated dose of $599\text{ e}^-/\text{\AA}^2$. Before this occurred, atomic quantification of the particles revealed they had a chemical composition of $\sim\text{FeO}_5\text{P}_{5.5}$. This composition is within error of the compositions calculated for the nanoparticles inside encapsulins cores and is thus a good indicator that these are the same structures. Following the collection of the first data frame, the average density of Fe atoms calculated was 2 atoms/nm^3 meaning that each 2 nm particle contained an average of ~ 9 iron atoms. Following the collection of the second data frame the density of Fe atoms is 1 atom/nm^3 , indicating that between these first two frames a significant amount of the particle has been destroyed. Following this frame, no signal could be collected from the nanoparticles.

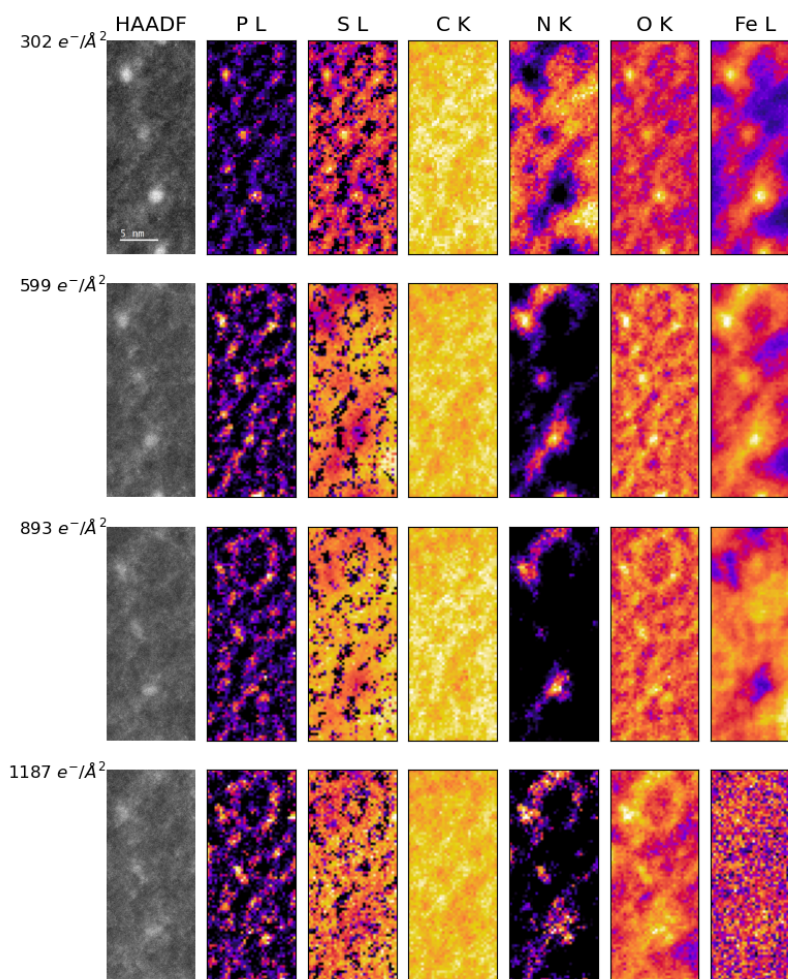


Figure 5.13: Chemical mapping of individually distributed nanoparticles after sequential scanning of the same area. Left to Right: HAADF reference image. Elemental maps showing the spatial distribution of P $L_{2,3}$, S $L_{2,3}$, C K, N K, O K, Fe $L_{2,3}$.

5.2.11 Fingerprinting of the Fe $L_{2,3}$ edge near-edge structure

It has been shown previously that the $L_{2,3}$ edge ELNES of transition metals are sensitive to the oxidation state of the metal atoms [217, 142]. In the case of Fe, there is a shift in the position of the L_3 maximum depending of the valence state. If the material is Fe^{3+} dominant, such as in haematite, the peak maximum can be found at 710.5 eV whilst if the material is Fe^{2+} dominant, such as ilmenite, the maxima is at 708.7 eV. In minerals that exhibit mixed valence, such as magnetite, the maximum falls somewhere between these two extremes.

The encapsulin cargo proteins are ferritin-like proteins, meaning they are involved in catalysing the oxidation of Fe^{2+} into Fe^{3+} . It is therefore hypothesised that the encapsulin core contains only Fe atoms in the 3+ oxidation state. To examine this, the Fe $L_{2,3}$ edge is compared with those of the known reference materials: haematite, magnetite and ilmenite. These reference materials were chosen because the shape of the Fe $L_{2,3}$ edge depends solely on the valence state of the Fe atoms, and not on the other elements present in the material. The Fe $L_{2,3}$ region of the encapsulins isolated from *M. xanthus*, EncA+EncC and EncA+EncB+EncC all appear to match the shape of the haematite $L_{2,3}$, i.e, a peak at ~ 710.5 eV and a characteristic shoulder on the lower energy side of the edge (see Figure 5.14). This similarity indicates that the nanoparticles inside the encapsulin are most likely Fe^{3+} dominant. Interestingly, the Fe $L_{2,3}$ for the EncA+EncB construct matches more closely to the edge of the magnetite reference sample indicating that the core could be of mixed valence between Fe^{2+} and Fe^{3+} . However due to the sensitivity of Fe atoms to reduction by the electron beam, it is possible that the observed mixed valence is an artefact [218, 201, 219].

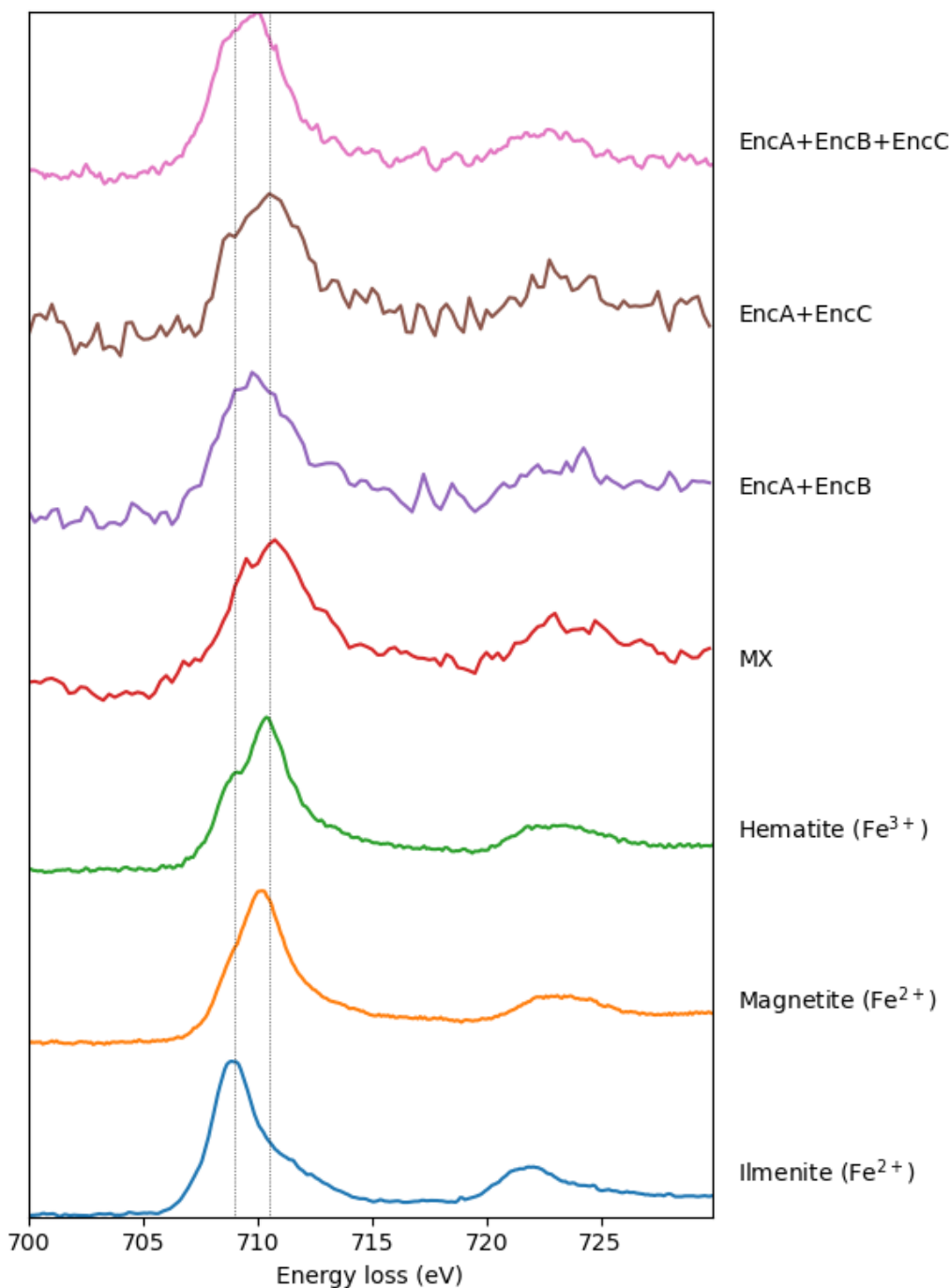


Figure 5.14: Fe $L_{2,3}$ EELS edge fingerprinting against reference standards. The two vertical dotted black lines represent the maxima values of the two end reference materials, ilmenite (708.7 eV) and haematite (710.5 eV). MX represents the encapsulins isolated from *M. xanthus*. All spectra were collected on the JEOL F200 operating at 200 keV, using the 30 μm condenser aperture with a nominal spot size of 6 and a dwell time of 0.01s. The spectra have been deconvolved using Fourier-ratio deconvolution and background subtracted with an inverse power law between 600-700 eV. Dual EELS spectra were collected with an offset of 660 eV and a dispersion of 0.1 eV/ch.

5.2.12 Change in Fe $L_{2,3}$ edge shape with increasing electron dose

Initially, all experiments examining the Fe $L_{2,3}$ ELNES structure, returned edges whose shape matched that of ilmenite indicating a Fe^{2+} dominant material (see Section 5.2.11). However, this did not seem to match what was hypothesised based on the function of the cargo proteins.

It has been well explored in the literature that ferritin cores are easily damaged by the electron beam, suffering reduction that increases as a function of electron dose [218, 219]. The damage can be visualized as a shift in the Fe L_3 peak maxima from ~ 710.5 eV (for Fe^{3+}) to ~ 708.7 eV (for Fe^{2+}). As the iron mineralizing proteins of the encapsulin system are 'ferritin-like' it is expected that a similar damage mechanism might affect the samples generated in this work. To examine this, samples were imaged at sequentially decreasing electron doses using a dispersion of 0.1 eV/ch. These experiments were carried out on the JEOL F200 using an acceleration voltage of 200 keV. On this machine the beam current has not been calibrated for STEM mode, therefore, no measurement of the electron dose can be provided. Instead the electron dose was controlled by changing the spot size (6 to 4 is an increase in intensity) and the pixel dwell time during data collection. It was identified that the Fe $L_{2,3}$ edge undergoes the previously mentioned shift in peak maxima (see Figure 5.15), indicating that when collecting spectra centred around the Fe $L_{2,3}$ on this microscope the optimal conditions for preventing reduction of the material by the electron beam is a probe size of 6 and a pixel dwell time of up to 0.1s/pixel.

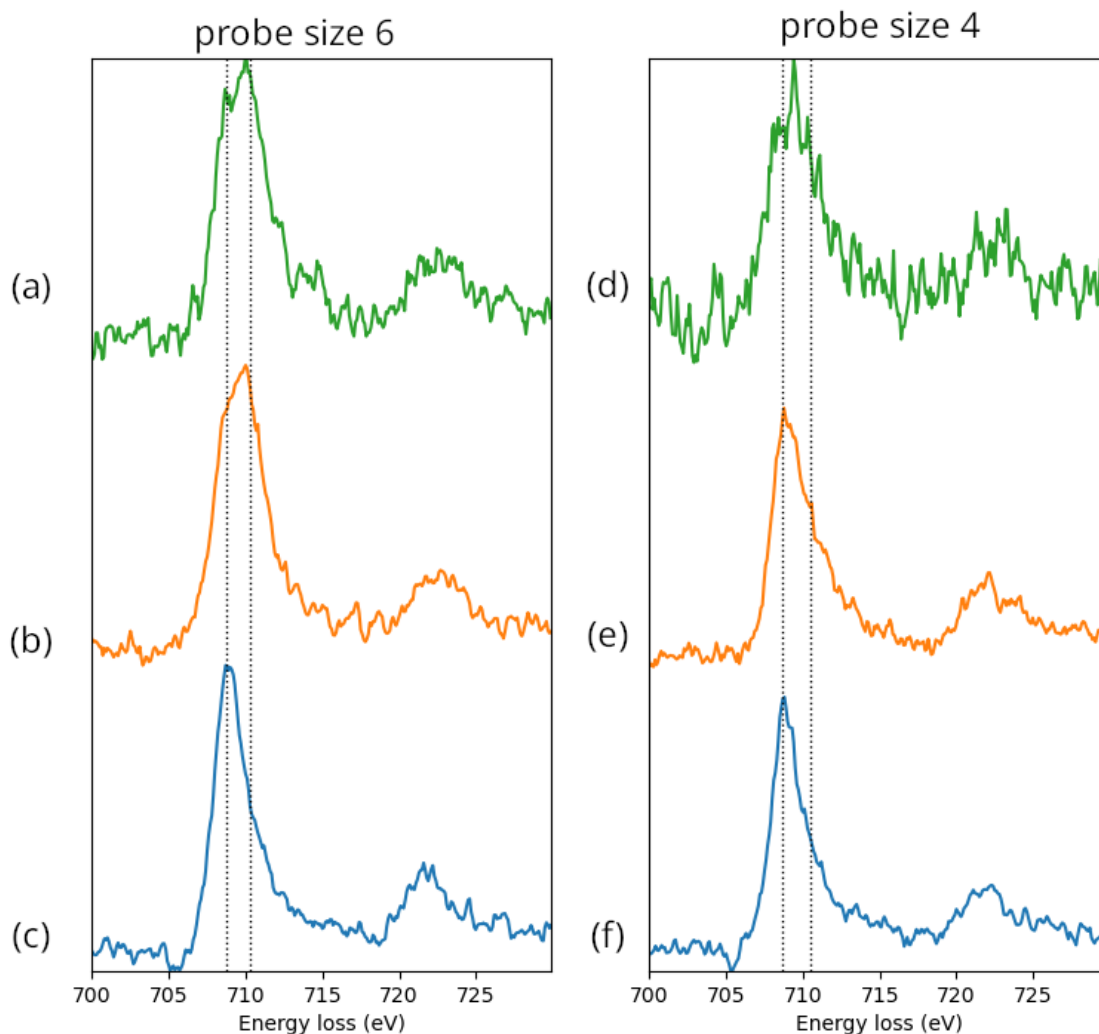


Figure 5.15: The change in the Fe $L_{2,3}$ edge of the EncA+EncB+EncC core with increasing electron dose. The electron dose increases from (a) to (c) and from (d) to (f). For a given exposure time, the electron dose is higher when using probe size 4 compared to probe size 6. The experimental conditions used were: (a) probe size 6 at 0.01 s/pixel, (b) probe size 6 at 0.1 s/pixel, (c) probe size 6 at 1 s/pixel, (d) probe size 4 at 0.01s/pixel dwell time, (e) probe size 4 at 0.1 s/pixel, (f) probe size 4 at 1 s/pixel. All spectra were collected using a JEOL F200 operating at 200 keV. The spectra have been deconvolved using Fourier-ratio deconvolution and background subtracted with an inverse power law between 600-700 eV. Dual EELS spectra were collected with an offset of 660 eV and a dispersion of 0.1 eV/ch. The electron probe was centred on an encapsulin and the single reading for the given time was captured. For (c)-(f) the spectra represent 25 summed measurements. However, due to the poor signal intensity generated for the imaging conditions in (a) and (b) 25 measurements was too low to give a useful spectra, resulting in far higher numbers of measurements needing to be summed. This is why the signal to noise ratio is higher in these two spectra.

5.2.13 Quantification of Fe $L_{2,3}$ edge

The differences in the shape of the Fe L-edge between Fe atoms in different valence states can be quantified using a method that calculates the ratio of the L_3 to L_2 edges [142, 205, 220]. In this method, intensities are integrated over two windows (708 - 710 eV and 719.7 - 721.7 eV) from the deconvolved and background subtracted spectra and compared with a universal curve made from reference materials with known valences.

Comparison of materials generated as part of this work with the calibration curve confirms the observation of the ELNES fingerprinting experiment (see Figure 5.16). The nanoparticles inside the *M. xanthus* encapsulin are dominated by ferric species. However, this was not observed for constructs expressed in *E. coli* where the average $L_3/(L_3+L_2)$ and spread of data varied greatly from measurements made from the reference materials. The similarity in the $L_3/(L_3+L_2)$ ratios from these materials compared with the ratio calculated for magnetite suggests that these materials are mixed valence. Some of this variance can be explained by the reduction of the Fe due to the electron beam (see Section 5.2.12) if the sampled encapsulin was damaged during region surveying before data collection, despite care being taken during the data acquisition. Addition error in the L_3 peak maxima position may have been introduced by a calibration error of the detector offset, as the data from samples and from reference materials were captured on different days. However, additional care was taken to set the offset using the drift tube to prevent this error.

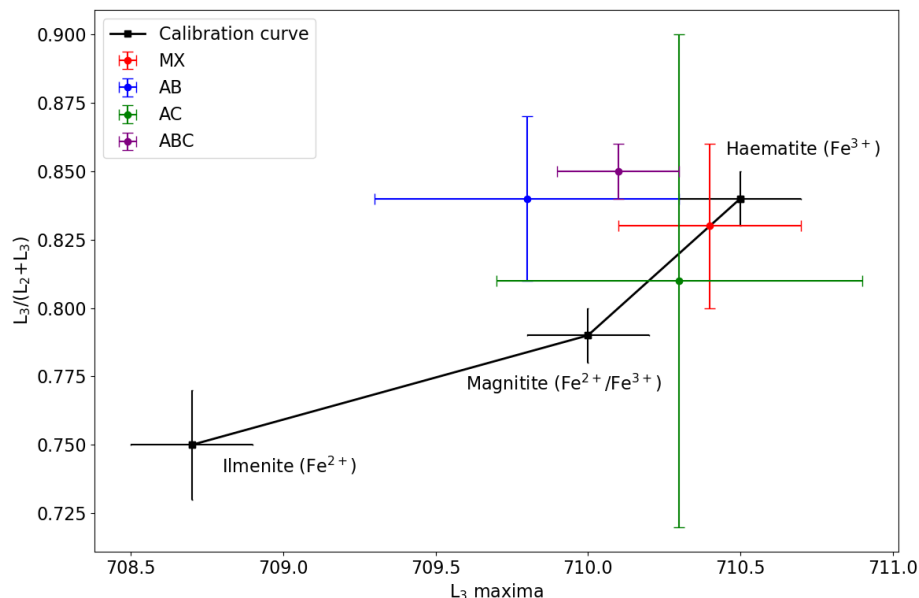


Figure 5.16: Plot of the $L_3/(L_3+L_2)$ against the L_3 maximum peak position. The reference material calibration curve (black), encapsulins from *M. xanthus* (red), EncA+EncB (blue), EncA+EncC (green) and EncA+EncB+EncC (purple) are shown on the plot. The $L_3/(L_3+L_2)$ ratio was calculated from two windows (708 eV to 710 eV and 719.7 eV to 721.7 eV).

5.3 Discussion

This study presents an EELS analysis of the chemical composition of nanoparticles encapsulated within the encapsulins isolated from *M. xanthus*. We also examine the nanoparticles formed within encapsulins containing either single or multiple cargo proteins. It was identified through investigations into the Fe $L_{2,3}$ edge fine structure that the nanoparticles are dominated by Fe^{3+} . This is in agreement with what was hypothesised based on the predicted function of the cargo proteins, and suggests that the Fe^{2+} ions that traverse the pores in the shell are oxidized during the mineralization process [18, 23]. In the encapsulins containing fewer numbers of cargo proteins it was identified that the Fe atoms in the nanoparticles are mixed valence, containing both Fe^{3+} and Fe^{2+} . This result could indicate that the combination of all the cargo proteins is required to prevent the inclusion Fe^{2+} in the nanoparticles inside the encapsulin. The result of removing cargo proteins would therefore be an increase in the Fe^{2+} content of the nanoparticles, as there is no cargo protein present to oxidise the Fe^{2+} atom after it has entered the encapsulin through the pore, the Fe^{2+} ion is incorporated directly into the nanoparticle.. However, as previously discussed, the observed difference may

also be due to the sensitivity of Fe to reduction under the electron beam (see Section 5.2.11).

The atomic composition of the nanoparticles was also quantified. In the structures isolated from *M. xanthus* a general atomic composition of $\text{Fe}^{3+}\text{PO}_4$ was estimated. This quantification was carried out assuming no contribution in O K-edge signal from the encapsulin shell as under the conditions used for data acquisition the shell was destroyed during analysis. This general formula would indicate the presence of a Fe oxy-hydroxide-phosphate, which is similar to those observed in other bacterial ferritin systems [221, 222, 223, 224]. Furthermore, the average Fe:P ratio calculated for the nanoparticles was, 1.2:1 which strongly agrees with the Fe:P ratio calculated using EDXS (see Section 4.2.5). However, it differs dramatically from measurements made previously from structures isolated in *M. xanthus*, where a Fe:P ratio of 4:1 was recorded using inductively coupled plasma mass spectrometry [18]. The Fe:P ratio measured here does however align with the Fe:P ratio identified in the *Q. thermotolerans* encapsulin, where a Fe:P ratio of 1:1 was measured using STEM-EELS [19] and is more similar to measurements made in other bacterioferritin systems where a ratio of 1-2:1 is often observed [195, 196, 197]. Interestingly, SDS quantification in this study revealed that the isolated encapsulins predominantly contained EncC, similar to the other structures investigated in the literature [18]. Thus, it is unclear if the differences in the Fe:P ratio observed is related to the cargo protein content of the encapsulin shell.

The encapsulins expressed in *E. coli* containing either EncB or EncC produced nanoparticles with different Fe:P ratios, 2:1 in the former and 4:1 in the latter. Although these ratios are not the same as the ratios observed from the EDXS data, they are similar in that Fe:P ratio was found to be higher in encapsulin containing EncB than in encapsulins containing EncC. In the EDXS experiment, Fe:P ratios of 2.4:1 and 2.9:1 were observed for encapsulins containing EncB and EncC respectively (see Section 4.2.5). Although these ratios differ from the *M. xanthus* encapsulin - possibly due to the P availability within the *E. coli* cell compared with the *M. xanthus* cell [180] - the difference between them may be important for understanding the function of the encapsulin. If the observed difference in the chemical composition of nanoparticles synthesised by EncB and EncC is representative of real difference in the native structure it would suggest that the P content of the encapsulin can be controlled by altering the number of cargo proteins present. Which may be important for the functioning of the system, as the presence of P has been shown to play an important role in formation of

nanoparticles inside ferritin cages [179, 180]. However, this would require high levels of control over cargo loading, and currently the prevailing model is one of co-translational loading [12, 56, 37], where the cargoes are loaded based on their proximity to the forming encapsulin shell. This hypothesis is also unsupported by the experimental result that all of the nanoparticles inside the encapsulin have the same chemical composition as measured with high spatial resolution STEM EELS (Section 5.2.7). It is also important to note that the nanoparticles of encapsulins containing both EncB and EncC expressed in *E. coli* had the same Fe:P ratio as the encapsulin containing only the EncB protein, whilst in the EDXS data this encapsulin had an Fe:P ratio of 2.7:1 (see Section 4.2.5), which is in between the Fe:P ratio identified for EncB and EncC containing encapsulins. The EELS result cannot be explained by a higher quantity of EncB compared to EncC as the SDS quantification identified the same number of both cargo proteins were present (see Table 2.3.8). However, the EDXS result being in between the two extremes makes sense if the core is composed of nanoparticles made by both cargo proteins. Additionally, quantification of single nanoparticles from highly spatially resolved *M. xanthus* encapsulins revealed that all of the sampled nanoparticles had the same Fe:P ratio. The results of these two experiments indicate that if the P content is controlled the system may be more complicated than simply the cargo proteins that are present.

Estimates of the number of Fe atoms per shell reveal that the actual amount of Fe stored within the encapsulin is significantly lower than previous estimates of the maximum storage capacity. Using STEM mass measurements, it was previously estimated that encapsulin could store up to ~ 30000 Fe atoms [18]. Since this initial estimation, the atomic structures of the three cargo proteins have been resolved [28, 97], thus allowing a more refined estimate to be made that accounts for the space taken up by the cargo proteins leading to a revised maximum capacity of approximately ~ 26000 Fe atoms in the *M. xanthus* encapsulin about 15% smaller than the previous estimate. On average, encapsulins isolated from *M. xanthus* in this study contained ~ 8000 Fe atoms per shell. This estimate makes sense in the context of the number of Fe atoms estimated for other encapsulins of different sizes. An estimate of 4100 Fe atoms was made for the encapsulins of *R. rubrum* which has a diameter of 24 nm [23], and an estimate of 23000 was made for the 42 nm encapsulin of *Q. thermotolerans* [19].

No significant difference was found between the estimations of the number of Fe atoms

made in encapsulins containing either EncB or EncC or both of these cargo proteins. The implication of this is that all of the nanoparticles in encapsulins of *M. xanthus* have the same Fe content and it is only the P content that changes between nanoparticles synthesised by the different cargo proteins. However, due to low spatial resolution in the EELS SI maps no individual nanoparticles could be measured from either of these constructs to confirm this directly. Furthermore, having both of the cargo proteins present did not result in a higher Fe content of the cores. These results indicate that there is also no difference in the rate of Fe mineralization between EncB and EncC.

A significant difference was observed in the number of Fe atoms between the overexpressed encapsulins and those isolated from *M. xanthus*. The observed difference may be due to several reasons. Firstly, the amount of Fe atoms present in *M. xanthus* is unknown. Therefore, it is impossible to predict if the 8 mM of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ added to the media far exceeded what would be available normally. *E. coli* has its own iron homeostasis system that would be actively reducing the intracellular Fe^{2+} content [74]. Secondly, as previously discussed, the cores of the encapsulins generated in *E. coli* do not show the same internal nanoparticle distribution as the structures isolated from *M. xanthus* leading to a higher density of particles and therefore a higher number of Fe atoms stored. Finally, it has been previously shown that there is a large variability in the number of nanoparticles inside the cores of the overexpressed encapsulins (see Section 3.2.5). The implication of this is it may have caused sampling bias when conducting the STEM spectroscopy experiments. If the Fe rich region was larger it meant more data points could be collected per region, giving better signal to noise for that region. This was especially apparent on the JEOL F200 which is not spherically aberration corrected, and thus larger probe sizes were required to give the sufficient beam current for data collection. Therefore, the significantly higher estimations of the number of Fe atoms may be an artefact caused by preferentially selecting encapsulins with larger cores during data collection.

5.4 Appendix

5.4.1 Calculation of convergence angles

A sample of poly crystalline gold on lacy carbon was purchased (Agar Scientific) and loaded into the JEOL F200 in STEM mode. The 40 μm condenser aperture was inserted and aligned over the centre of the Ronchigram. With the viewing screen down and the scan set to spot mode the beam was scanned over the sample until a clear diffraction pattern could be seen. The viewing screen was lifted and an image captured on the OneView Camera (Gatan). Immediately after image capturing the beam was blanked to prevent damage to the camera. This process was repeated for the 10 μm aperture. Convergence angles were then calculated using:

$$\alpha = \frac{r}{R} \cdot \frac{\lambda}{a} \sqrt{h^2 + l^2 + k^2} \quad (5.2)$$

α = convergence angle

r = radius of the $[0,0,0]$ spot measured in $1/\text{nm}$

R = distance between the same point in the $[0,0,0]$ spot and a diffracted spot in $1/\text{nm}$

λ = wavelength of the electron beam at 200 kV = 2.51pm

a = lattice constant for gold = 407.8pm

hkl = Miller indices representing the diffraction spot

Indexed diffraction patterns of polycrystalline gold can be seen in Figure 5.17.

5.4.2 Calculation of collection angles

JEOL F200 was aligned in STEM mode and the 40 μm condenser aperture was inserted. All detectors and the OneView camera (Gatan) were removed, leaving an unblocked path to the Gatan Imaging Filter (GIF). Imaging mode on the GIF camera was selected and the view started. Images were collected sequentially with both the HiRes (2.5 mm) and HiSNR (5mm) spectrometer entrance apertures inserted. The convergence angle of the transmitted beam is known, therefore, the collection angle of each entrance aperture was calculated as the ratio of

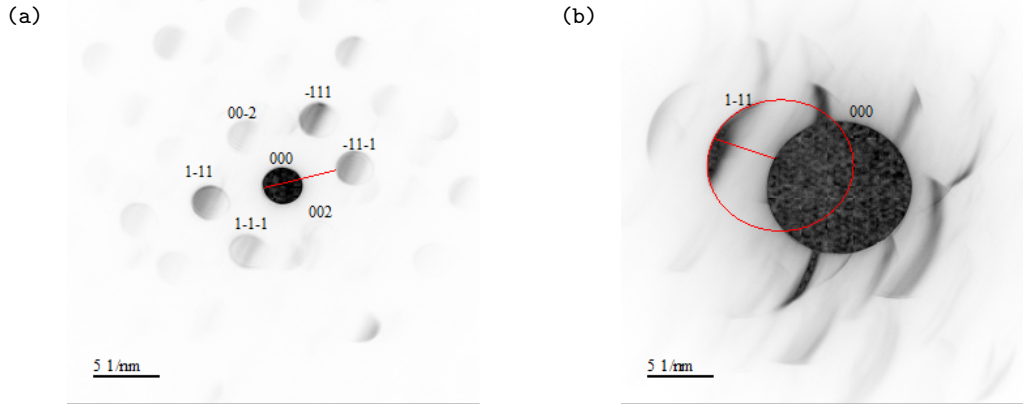


Figure 5.17: Indexed diffraction patterns of polycrystalline gold used to calculate convergence angles. (a) 10 μm condenser aperture and (b) 40 μm condenser aperture.

the radius between that of the transmitted beam and the shadow of the entrance aperture.

Aperture	Convergence angle (α)	Collection angle (β)
10 μm condenser aperture	2.96 mrad	-
40 μm condenser aperture	12.34 mrad	-
HiSNR entrance aperture	-	17.14 mrad
HiRes entrance aperture	-	8.54 mrad

Table 5.7: Calculated convergence and collection angles used for a camera length of 120 mm.

Chapter 6

Final Discussion

The research presented over the course of this work provides new insight into the nanoparticles synthesised by the *M. xanthus* encapsulin and into the function of the two ‘ferritin-like’ cargo proteins: EncB and EncC. A combination of different imaging techniques were used to image and then measure the size of the nanoparticles inside encapsulins isolated from *M. xanthus*, identifying that these nanoparticles have a diameter of ~ 2 nm. This is far smaller than the diameter of ~ 5 nm that was previously measured using cryo-electron tomography [18]. The observation that the nanoparticles inside the *M. xanthus* encapsulin are ~ 2 nm is supported by the fact that the same diameter nanoparticles were observed inside all encapsulins generated in *E. coli*, regardless of whether the encapsulins were supplied with Fe *in vivo* or *in vitro*. Structural characterization of the nanoparticles inside *M. xanthus* using SAED and HR-STEM revealed that the majority of material inside the encapsulin is amorphous, which is different from suggestions of a crystalline material hypothesised in the literature [28]. The same observation was made for all of the encapsulins overexpressed in *E. coli* and loaded with Fe through the addition of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ to the media. This has important implications biologically, as amorphous material is believed to be more easily mobilized under conditions where the cell is Fe-limited thus amorphous material may be more useful to the cell when recovering from starvation [177, 178]. However, the HR-STEM experiments revealed small areas (~ 1.5 nm) of nanoparticles inside the *M. xanthus* encapsulins that possessed crystalline arrangement, and no single iron oxide phase could explain these measurements. A generally amorphous structure with small crystalline regions is often observed inside ferritin cages and thus in that context the atomic arrangement of the nanoparticles inside the M.

xanthus encapsulins aligns with established ferritin mineralization patterns [175, 176]. The presence of amorphous material is most likely the result of the high P concentration in the nanoparticles, as encapsulins loaded iron in a P free environment mineralized material with a crystalline atomic arrangement. Quantification from EELS identified a general formula of FePO_4 and both EDXS and EELS quantification measured a Fe:P ratio of 1.2:1. This Fe:P is far higher than the Fe:P ratio previously measured inside encapsulins from *M. xanthus* using inductively coupled plasma mass spectrometry [18]. However, it is much more similar to the minerals found inside both other studied encapsulin systems and to bacterial ferritin systems [195, 221, 222, 197, 196, 223, 224]. The differences in the nanoparticles sizes and between the P content of the nanoparticles measured here and those measured previously may also be linked, as the presence of P can impact the size of the nanoparticles formed inside of ferritin cages and in *in vitro* nanoparticle synthesis [179, 180, 225]. Thus combining all the experimental results together, the isolated encapsulins from *M. xanthus* were highly loaded with amorphous 2 nm nanoparticles, that had a general formula of $\text{Fe}^{3+}\text{PO}_4$ and an Fe:P ratio of 1.2:1.

To examine if the presence of P impacts the crystallinity of the nanoparticles inside the encapsulin a sample of encapsulins containing the cargo proteins EncB and EncC were isolated from *E. coli* and then loaded with Fe through the addition of 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. These experimental conditions were designed such that P was not present in any buffer or reagent and this was confirmed with EDXS and EELS as minimal P signal was found associated with the nanoparticles. HR-STEM analysis identified that the nanoparticles inside these structures showed far higher crystallinity than the nanoparticles isolated from *M. xanthus*, and the *d*-spacing measured indicated that the mineral present was Goethite ($\alpha\text{-FeO}(\text{OH})$). Quantification of the EDXS and EELS data from these structures, gave a general atomic formula of FeO and FeO_2 respectively. Combining these general formula with the results of the HR-STEM experiment, it is most likely that the crystalline material represents Goethite. However, it is unclear how much of the observed crystallinity is the result of action of the cargo proteins, as Goethite is the most prevalent hydrous iron oxide precipitated from aqueous solutions [226, 227]. Despite this, it is clear that the presence of P is important for the maintenance of amorphous nanoparticles inside the encapsulin.

Experiments were also carried out that focused on the function of the two ‘ferritin-like’

cargo proteins; EncB and EncC. Encapsulins containing either of these proteins were overexpressed and isolated from *E. coli* that had been grown in media containing $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$. An encapsulin was also generated that contained both of these proteins. EncD was omitted as bio-informatics and structural data had previously identified it lacked an Fe-binding domain [18, 97] and initial experiments carried out here identified that encapsulin containing only this cargo showed no Fe mineralization (Section 3.2.4). No clear differences were measured between the size of Fe rich cores synthesised inside encapsulin containing either cargo proteins over a range of concentrations of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ added to the media. Additionally, there was no effect from having both of these cargo proteins present. However, quantifications using both EDXS and EELS identified that the encapsulins containing EncB had a higher Fe:P ratio than those containing EncC. In EDXS Fe:P ratios of 2:1 and 2.7:1 were measured and in EELS Fe:P ratios of 2:1 and 4:1 were measured respectively. This difference in quantification results between these two methods may be in part due to the lower sensitivity of EDXS to light elements ($Z < 11$) and the different signal collection efficiencies between the two methods at the low pixel dwell times used for all experiments in this work [125]. These results indicate that the nanoparticles synthesised by the EncB protein may have a higher P content than those synthesised by the EncC protein, especially under the experimental conditions used.

Furthermore, the nanoparticles inside the encapsulin containing both of these cargo proteins had Fe:P ratio of 2:1 and 2.7:1 in EELS and EDXS respectively. As previously identified the presence of P may impact the size of nanoparticles synthesised inside ferritins [225]. Thus providing an argument for the differences observed in the nanoparticle sizes between the isolations from *M. xanthus* made here and those made in previous studies. However, encapsulins containing either EncB or EncC measured here both contained nanoparticles with the same nominal diameter of ~ 2 nm, indicating that this difference in P concentration does not impact the nanoparticle size. Combining these results together, it was identified that there is no difference between either the size of the Fe-rich core or the size of the nanoparticles synthesised inside encapsulins containing either or both cargo proteins, however, spectroscopic analysis identified that there is a difference in the chemical composition of the nanoparticles synthesised by either cargo protein, with those synthesised by EncB having a higher P content than those synthesised by EncC.

The difference in the P content between the nanoparticles synthesised by EncB and

EncC has some interesting biological implications. As experiments carried out here have linked a lack of P to a crystalline material forming inside the encapsulin shell, and furthermore crystalline materials have been identified as having lower bio-availability than amorphous materials [177, 178]. The differences in the nanoparticles synthesised by EncB and EncC may represent an opportunity for higher levels of control over the mineral that is stored inside the encapsulin. This may be important in allowing finer tuning of the response of the bacteria to the starvation stress that resulted in the production of the encapsulin. Under low stress conditions it may be important to store Fe in a more bio-available form for quicker recovery. Alternatively, it has been shown that encapsulins are also expressed under normal cell growth conditions [18], and Fe storage under these conditions may have a different requirement than under stressful conditions, thus resulting in a different combination of cargo proteins being used. However, this would require the identification of a more complex system for loading cargo proteins into the shell - such as the scaffold proteins used during viral capsid formation [163] - and the current model is one where the cargo is loaded co-translationally in a passive manner [56].

One final point, the largest differences observed between different encapsulins were between those isolated from *M. xanthus* and those overexpressed in *E. coli*. In the former an average of ~ 8000 Fe atoms were measured per shell and in the latter an average of ~ 13000 - ~ 16000 was measured per shell, and under all experimental conditions the overexpressed encapsulins had smaller cores than the encapsulins isolated from *M. xanthus* despite having the same average nanoparticle size. Furthermore, the cores from overexpressed encapsulins showed equal average contrast in HAADF imaging compared with the encapsulins isolated from *M. xanthus* despite their smaller size, and the estimates of the number of Fe atoms present made using EELS indicate that these encapsulins have roughly twice as many Fe atoms in them as those isolated from *M. xanthus*. The most probable explanation for this result is that if only either EncB or EncC (or both of these proteins) are present inside the encapsulin the nanoparticles are packed more closely. Indicating that all of the cargo proteins are required to achieve the spatial arrangement of nanoparticles seen inside the native encapsulin shell. However, it is also important to point out that the two bacterial strains have different Fe homeostasis mechanisms and different Fe requirements for their metabolism [182, 183, 101], thus the availability of free Fe^{2+} differs between the two cells. Furthermore, overexpression and growth in media containing high concentrations of Fe do not represent the normal growth

conditions for an *E. coli* cell. Overexpression alone puts substantial pressure on the cell, as the T7 promoter used for the petDUET-1 overexpression plasmid is twice as productive as a normal *E. coli* promoter [168]. Whilst, a large influx of Fe^{2+} would activate the *E. coli* Fur system to produce efflux pumps putting further pressure on the cell [83]. Thus, these conditions may partially explain some of the differences seen between the encapsulins isolated from *M. xanthus* and those overexpressed in *E. coli*, pointing to the inherent limitations host systems for investigating physiological processes.

6.1 Novel model of nanoparticle arrangement inside encapsulins

Comparison of native encapsulin structures isolated from *M. xanthus* and recombinant encapsulin structures purified from *E. coli* provides the foundation for a novel model of iron storage within the encapsulin nanocompartment (see Figure 6.1). Key differences have been observed between the arrangement of nanoparticles inside both groups of structures. In the encapsulins isolated from *M. xanthus*, the diameter of the nanoparticle containing ‘core’ structure was measured at ~ 24 nm (Section 3.2.2) and found to contain on average 8000 Fe atoms (Section 5.2.8). Whilst in the recombinant encapsulins isolated from *E. coli*, the largest diameter of a ‘core’ region was measured at ~ 15 nm (EncA+EncB+EncC - Section 3.2.5) and across the constructed investigated, the number of Fe atoms was recorded at between 13000 and 16000 Fe atoms (Section 5.2.9). Furthermore, even under the experimental conditions aimed at producing the largest core inside a recombinant encapsulin (0.01 mM IPTG and 8 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$) the largest diameter did not exceed ~ 17 nm (Section 3.2.8). Thus, recombinant encapsulins isolated from *E. coli* contain smaller, more tightly packed cores with a higher number of Fe atoms compared to encapsulins isolated from the native host, *M. xanthus*.

Given this information, a new model for the storage of iron inside the encapsulin can be suggested. Free Fe^{2+} is delivered passively to the encapsulin, where it traverses the open pores located at the 5-fold-axis. Upon entry, the Fe^{2+} diffuses to the cargo proteins - EncB and EncC - where it is transported to the FOC and oxidised to Fe^{3+} . Due to the lower affinity of the Fe^{3+} for the the binding sites in the FOC, it is replaced by another Fe^{2+} and moved through the central channel fo the cargo protein complex to the outer surface of this protein complex. In previous models, it has been proposed that the surface of the cargo protein complex is the site of nucleation of the iron containing mineral which then leaves

the surface of the protein to be stored in the lumen to be stored as a crystalline mineral ($[\text{FeO}(\text{OH})]_8[\text{FeO}(\text{H}_2\text{PO}_4)]$) [28]. However, the experimental observations made about the *M. xanthus* encapsulin in this work (low Fe atom density, consistent inter-nanoparticle spacing, and large core region), suggest that the nanoparticles remain bound to the surface of the cargo proteins rather than being released into the encapsulin lumen for storage (see Figure 6.1). Furthermore, the results described here suggest that the nanoparticles are structurally amorphous with general formula of $\text{Fe}^{3+}\text{PO}_4$ rather than the crystalline ferrihydrite mineral $[\text{FeO}(\text{OH})]_8[\text{FeO}(\text{H}_2\text{PO}_4)]$. Remaining bound to the surface of the cargo proteins confers potential advantages when the liberation of Fe from these structures is considered. As has been previously identified, the third encapsulin cargo protein - EncD - is a ferric reductase and has been implemented in the release of Fe from these structures [97]. Under this assumption, that EncD is essential for iron liberation, a model where the nanoparticles are bound to the surface of the cargo proteins would infer greater levels of control and access to these nanoparticles when Fe release is required, compared to models proposing that nanoparticles are freely stored within the encapsulin lumen. Thus, this model provides a better understanding of an internal arrangement of nanoparticle that allows the encapsulin to function as both a system for oxidative stress resistance and more generalized iron homeostasis. Thus, in this updated model, the nanoparticles remain bound to the surface of the cargo protein complexes, near the inner surface of the encapsulin shell, where they are stored as an amorphous material ($\text{Fe}^{3+}\text{PO}_4$).

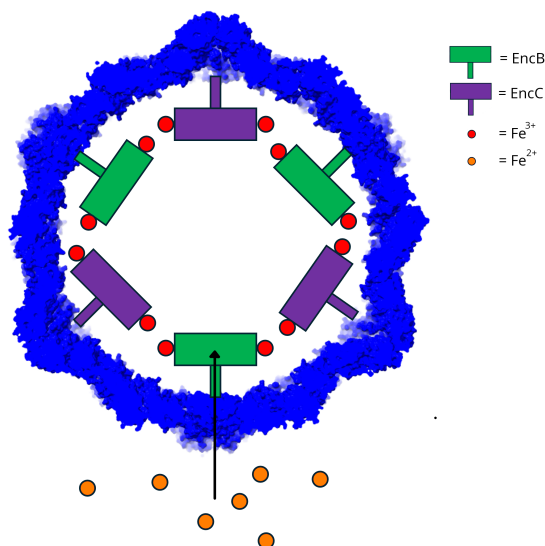


Figure 6.1: Proposed model for iron nanoparticle arrangement inside the encapsulin based on the experimental results of this work. The model contains various elements: the EncA protein shell (blue), the cargo proteins EncB (green) and EncC (purple), Fe^{2+} (orange) and Fe^{3+} (red). Free Fe^{2+} in the cytoplasm of the cell is transported through the pores in the encapsulin shell where it contacts in the ferritin like cargo proteins attached to the inner surface of the encapsulin shell. Here the Fe^{2+} is oxidised into Fe^{3+} and transported to the surface of the cargo protein for mineralization to occur. The resulting nanoparticle product remains bound to the surface of the cargo protein, where it is stored.

Chapter 7

Conclusion

The research presented in this thesis offers new insights into the encapsulin system of *M. xanthus*, through a novel characterization of the material mineralized and stored by the system. The findings in this work have identified that, the structures contain large numbers of 2 nm spherical nanoparticles that are generally amorphous in their atomic structure. Despite showing no long range order, the nanoparticles do show some small regions with regular crystalline arrangement when investigated with HRSTEM, however, no single iron oxide phase was sufficient for explaining the observed crystallinity. The novel characterization of the material inside the *M. xanthus* encapsulin presented here provides novel support for the function of the encapsulin in oxidative stress response. The functional significance is especially evident when the release of Fe from the system is considered. The small size of the encapsulated nanoparticles, smaller than the previously reported 5 nm [18] results in a high surface area-to-volume ratio, enhancing their chemical reactivity. This property is likely to facilitate the rapid restoration of cellular iron levels when cells return to optimal growth conditions following starvation. In addition, the amorphous nature of the stored material further contributes to this process, as iron is more readily mobilized from amorphous minerals than from crystalline forms [177, 178]. Thus the small amorphous nanoparticles inside the *M. xanthus* encapsulin are advantageous for the cells recovery following oxidative stress. Additionally, the findings in this work identify that the nanoparticles inside the encapsulin contain O as well as the previously identified Fe and P, meaning that they are more similar chemically to both the material inside ferritins and other encapsulin systems [19, 227]. It was identified that the dominant valence state of the Fe inside the encapsulin is 3+, with quantification of the L_{2,3}

edge indicating that there is a small amount of the 2+ Fe present. It could not however, be identified if the small amount of Fe^{2+} measured was an artefact resulting from the sensitivity of Fe to reduction by the electron beam [218, 201, 219]. As a result, it can be concluded that the vast majority of the Fe atoms that enter the encapsulin are oxidised and stored in nanoparticles with a general formula of $\text{Fe}^{3+}\text{PO}_4$. Highly spatially resolved measurements of individual nanoparticles within the same encapsulin shell reveal that this general formula is correct at the individual nanoparticle level, and thus all of the nanoparticles inside a single encapsulin shell have the same chemical composition. Finally, it was identified that encapsulin on average contain 8000 Fe atoms, a number far low than the previously estimated maximum of ~ 30000 [18] or the maximum of ~ 26000 atoms estimated here. The storage of relatively low numbers of Fe atoms compared with the proposed maximum may be indicative of the expression conditions of the encapsulin system. As the system is associated with the starvation stress response, it is likely that encapsulin proteins are produced rapidly by the cell in response to the stressor. Resulting in large numbers of encapsulins relative to the number of free ferrous ions in the cell, which would function to increase the rate of Fe uptake across the cell. Furthermore, low iron density would also serve to increase the rate of Fe release back into the cell upon its return to normal growth conditions, thus allowing the cell to respond more quickly to the nutrient content of its environment.

The generation of encapsulin constructs with reduced numbers of cargo proteins enables conclusions to be drawn regarding the functional roles of the two iron mineralizing cargo proteins, EncB and EncC. Several functional similarities were observed between the encapsulins containing only one of these two proteins. Specifically, it was found that they both produce Fe containing core regions with no significant difference in size, they both produce nanoparticles with the same ~ 2 nm diameter, they both produce amorphous nanoparticles and when loaded with Fe both contain the same number of Fe atoms on average. Additionally, no significant difference was found between these two encapsulin constructs and encapsulins containing both EncB and EncC across any of these features. Indicating that there is no difference between the amount of Fe mineralized by these two proteins nor between the physical properties of the nanoparticles they produce. However, the EDXS and EELS results presented in this work, suggest that there is a difference in the chemical composition of the nanoparticles produced by both proteins. It was identified that nanoparticles synthesised by EncB have a higher Fe:P ratio than the nanoparticles synthesised by EncC. These results suggest that

there may be some influence on the chemical composition of the mineral core from the cargo protein present inside the encapsulin, potentially allowing for higher levels of control over the mineralized product than has previously been thought. However, a potential mechanism underpinning controlled cargo protein loading into encapsulin is currently unknown.

Additionally, experimental work carried out suggests that the presence of P is important for the production of the desired mineral inside the encapsulin. It was found that loading encapsulins with Fe in a P free environment results in the formation of crystalline Goethite nanoparticles instead of amorphous nanoparticles. This result may have important biological implications, as there is a relationship between the crystallinity of a material and its bio-availability [177, 178]. Therefore, the presence of P within the encapsulin during nanoparticle formation may play a key role in promoting the formation of an amorphous mineral, facilitating the efficient release of Fe when cells return to favourable growth conditions.

Finally, a novel model for the arrangement of nanoparticles inside the encapsulin can be suggested based on the comparison between the native structures isolated from *M. xanthus* and the recombinant structures generated in *E. coli* (see Section 6.1). The results of this work indicate that the core size is larger but the number of Fe atoms is lower in the native structures isolated from *M. xanthus* when compared with the recombinant structures produced in *E. coli*, despite both containing the same size nanoparticles. Therefore it can be concluded that, the nanoparticles in the recombinant structures are more densely packed and do not have the inter nanoparticle spacing found in the native structures. It is therefore proposed that the nanoparticles are not released from the surface of the cargo proteins to be stored in the lumen of the encapsulin as has been suggested previously [23, 28]. Instead they remain bound and are stored on the surface of the cargo proteins. This model for iron storage inside the encapsulin confers the advantage of higher levels of control and access to the nanoparticles for the ferric reductase EncD. The model suggested here provides support for the idea that the encapsulin functions as a more generalized iron homeostasis system alongside its already defined function in oxidative stress response.

Chapter 8

Future work

Looking forward, there are several key experimental area that could be focused on to extend the study undertaken here. Firstly, due the uncertainty in the observed phenotypes caused by the use of the *E. coli* host for overexpression, it may be useful to generate the reduced cargo encapsulins in a *M. xanthus* background. This would help to eliminate the uncertainty and provide additional proof that the observed tighter arrangement of nanoparticles is correct. Futhermore, it has been suggested that the differences observed between the native isolated encapsulins here and those isolated previously could be the result of the environment under which the isolation was undertaken. Therefore, it would be interesting to repeat the experiment in the city the initial isolation was undertaken in to observed in the differences in the ratio of Fe:P is maintained.

Bibliography

- [1] E. Cornejo, N. Abreu, and A. Komeili, "Compartmentalization and organelle formation in bacteria," *Current Opinion in Cell Biology*, vol. 26, p. 132–138, Feb 2014.
- [2] M. T. Waters and J. A. Langdale, "The making of a chloroplast," *The EMBO Journal*, vol. 28, p. 2861–2873, Sep 2009.
- [3] L. Ernster and G. Schatz, "Mitochondria: a historical review.," *The Journal of cell biology*, vol. 91, p. 227s–255s, Feb 1981.
- [4] H. Klebahn, "Gasvakuolen, ein bestandteil der zellen der wasserblü- tebildenden phycochromaceen," *Flora (Jena)*, vol. 80, no. 241–283, 1895.
- [5] A. E. Walsby, "Gas vesicles," *Microbiological Reviews*, vol. 58, p. 94–144, Mar 1994.
- [6] G. Drews and W. Niklowitz, "Beiträge zur cytologie der blualgen," *Archiv für Mikrobiologie*, vol. 24, no. 2, p. 147–162, 1956.
- [7] J. M. Shively, F. Ball, D. H. Brown, and R. E. Saunders, "Functional organelles in prokaryotes: Polyhedral inclusions (carboxysomes) of *thiobacillus neapolitanus*," *Science*, vol. 182, p. 584–586, Nov 1973.
- [8] E. Kofoid, C. Rappleye, I. Stojiljkovic, and J. Roth, "The 17-gene ethanolamine (*eut*) operon of *salmonella typhimurium* encodes five homologues of carboxysome shell proteins," *Journal of Bacteriology*, vol. 181, p. 5317–5329, Sep 1999.
- [9] T. A. Bobik, G. D. Havemann, R. J. Busch, D. S. Williams, and H. C. Aldrich, "The propanediol utilization (*pdu*) operon of *salmonella enterica* serovar typhimurium lt2 includes genes necessary for formation of

- polyhedral organelles involved in coenzyme b₁₂-dependent 1,2-propanediol degradation,” *Journal of Bacteriology*, vol. 181, p. 5967–5975, Oct 1999.
- [10] G. D. Havemann, E. M. Sampson, and T. A. Bobik, “Pdua is a shell protein of polyhedral organelles involved in coenzyme b₁₂-dependent degradation of 1,2-propanediol in salmonella enterica serovar typhimurium lt2,” *Journal of Bacteriology*, vol. 184, p. 1253–1261, Mar 2002.
- [11] A. Pokhrel, S.-y. Kang, and C. Schmidt-Dannert, “Ethanolamine bacterial microcompartments: from structure, function studies to bioengineering applications,” *Current Opinion in Microbiology*, vol. 62, p. 28–37, Aug 2021.
- [12] M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, and N. Ban, “Structural basis of enzyme encapsulation into a bacterial nanocompartment,” *Nature Structural and Molecular Biology*, vol. 15, no. 9, pp. 939–947, 2008.
- [13] N. Valdés-Stauber and S. Scherer, “Isolation and characterization of linocin m18, a bacteriocin produced by *brevibacterium linens*,” *Applied and Environmental Microbiology*, vol. 60, p. 3809–3814, Oct 1994.
- [14] N. Winter, J. A. Triccas, B. Rivoire, M. C. V. Pessolani, K. Eiglmeier, E.-M. Lim, S. W. Hunter, P. J. Brennan, and W. J. Britton, “Characterization of the gene encoding the immunodominant 35 kda protein of *mycobacterium leprae*,” *Molecular Microbiology*, vol. 16, p. 865–876, Jun 1995.
- [15] I. Rosenkrands, P. B. Rasmussen, M. Carnio, S. Jacobsen, M. Theisen, and P. Andersen, “Identification and characterization of a 29-kilodalton protein from *mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells,” *Infection and Immunity*, vol. 66, p. 2728–2735, Jun 1998.
- [16] P. M. Hicks, K. D. Rinker, J. R. Baker, and R. M. Kelly, “Homomultimeric protease in the hyperthermophilic bacterium *thermotoga maritima* has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria,” *FEBS Letters*, vol. 440, p. 393–398, Dec 1998.

-
- [17] F. Akita, K. T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara, and A. Nakagawa, “The crystal structure of a virus-like particle from the hyperthermophilic archaeon *pyrococcus furiosus* provides insight into the evolution of viruses,” *Journal of Molecular Biology*, vol. 368, no. 5, pp. 1469–1483, 2007.
- [18] C. A. McHugh, J. Fontana, D. Nemecek, N. Cheng, A. A. Aksyuk, J. B. Heymann, D. C. Winkler, A. S. Lam, J. S. Wall, A. C. Steven, and E. Hoiczyk, “A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress,” *The EMBO Journal*, vol. 33, pp. 1896–1911, July 2014.
- [19] T. W. Giessen, B. J. Orlando, A. A. Verdegaaal, M. G. Chambers, J. Gardener, D. C. Bell, G. Birrane, M. Liao, and P. A. Silver, “Large protein organelles form a new iron sequestration system with high storage capacity,” *eLife*, vol. 8, p. e46070, July 2019.
- [20] R. Rahmanpour and T. D. H. Bugg, “Assembly in vitro of *rhodococcus jostii* rha1 encapsulin and peroxidase dypb to form a nanocompartment,” *FEBS Journal*, vol. 280, p. 2097–2104, Apr 2013.
- [21] R. J. Nichols, B. LaFrance, N. R. Phillips, D. R. Radford, L. M. Oltrogge, L. E. Valentin-Alvarado, A. J. Bischoff, E. Nogales, and D. F. Savage, “Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism,” *eLife*, vol. 10, p. e59288, apr 2021.
- [22] M. P. Andreas and T. W. Giessen, “Large-scale computational discovery and analysis of virus-derived microbial nanocompartments,” *Nature Communications*, vol. 12, Aug 2021.
- [23] D. He, S. Hughes, S. Vanden-Hehir, A. Georgiev, K. Altenbach, E. Tarrant, C. L. Mackay, K. J. Waldron, D. J. Clarke, and J. Marles-Wright, “Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments,” *eLife*, vol. 5, Aug 2016.
- [24] H. Contreras, M. S. Joens, L. M. McMath, V. P. Le, M. V. Tullius, J. M. Kimmey, N. Bionghi, M. A. Horwitz, J. A. Fitzpatrick, and C. W. Goulding, “Characterization

- of a mycobacterium tuberculosis nanocompartment and its potential cargo proteins,” *Journal of Biological Chemistry*, vol. 289, p. 18279–18289, Jun 2014.
- [25] K. A. Lien, K. Dinshaw, R. J. Nichols, C. Cassidy-Amstutz, M. Knight, R. Singh, L. D. Eltis, D. F. Savage, and S. A. Stanley, “A nanocompartment system contributes to defense against oxidative stress in mycobacterium tuberculosis,” *eLife*, vol. 10, Nov 2021.
- [26] J. Ross, Z. McIver, T. Lambert, C. Piergentili, J. E. Bird, K. J. Gallagher, F. L. Cruickshank, P. James, E. Zarazúa-Arvizu, L. E. Horsfall, K. J. Waldron, M. D. Wilson, C. L. Mackay, A. Baslé, D. J. Clarke, and J. Marles-Wright, “Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment,” *Science Advances*, vol. 8, Jan 2022.
- [27] D. L. D. Caspar and A. Klug, “Physical principles in the construction of regular viruses,” *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 27, p. 1–24, Jan 1962.
- [28] E. Eren, B. Wang, D. C. Winkler, N. R. Watts, A. C. Steven, and P. T. Wingfield, “Structural characterization of the myxococcus xanthus encapsulin and ferritin-like cargo system gives insight into its iron storage mechanism,” *Structure*, vol. 30, pp. 551–563.e4, Apr 2022.
- [29] S. Kwon, M. P. Andreas, and T. W. Giessen, “Structure and heterogeneity of a highly cargo-loaded encapsulin shell,” *Journal of Structural Biology*, vol. 215, p. 108022, Dec 2023.
- [30] T. N. Szyszkka, M. P. Andreas, F. Lie, L. M. Miller, L. S. R. Adamson, F. Fatehi, R. Twarock, B. E. Draper, M. F. Jarrold, T. W. Giessen, and Y. H. Lau, “Point mutation in a virus-like capsid drives symmetry reduction to form tetrahedral cages,” *Proceedings of the National Academy of Sciences*, vol. 121, May 2024.
- [31] R. Duda, “Structural transitions during bacteriophage hk97 head assembly,” *Journal of Molecular Biology*, vol. 247, p. 618–635, Apr 1995.
- [32] W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, and J. E. Johnson, “Topologically linked protein rings in the bacteriophage hk97 capsid,” *Science*, vol. 289, p. 2129–2133, Sep 2000.

-
- [33] X. Agirrezabala, J. A. Velázquez-Muriel, P. Gómez-Puertas, S. H. Scheres, J. M. Carazo, and J. L. Carrascosa, “Quasi-atomic model of bacteriophage t7 procapsid shell: Insights into the structure and evolution of a basic fold,” *Structure*, vol. 15, p. 461–472, Apr 2007.
- [34] E. V. Koonin, V. V. Dolja, M. Krupovic, A. Varsani, Y. I. Wolf, N. Yutin, F. M. Zerbini, and J. H. Kuhn, “Global organization and proposed megataxonomy of the virus world,” *Microbiology and Molecular Biology Reviews*, vol. 84, May 2020.
- [35] T. W. Giessen and P. A. Silver, “Widespread distribution of encapsulin nanocompartments reveals functional diversity,” *Nature Microbiology*, vol. 2, Mar 2017.
- [36] H. Moon, J. Lee, J. Min, and S. Kang, “Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform,” *Biomacromolecules*, vol. 15, p. 3794–3801, Sep 2014.
- [37] J. Snijder, O. Kononova, I. M. Barbu, C. Uetrecht, W. F. Rurup, R. J. Burnley, M. S. T. Koay, J. J. L. M. Cornelissen, W. H. Roos, V. Barsegov, G. J. L. Wuite, and A. J. R. Heck, “Assembly and mechanical properties of the cargo-free and cargo-loaded bacterial nanocompartment encapsulin,” *Biomacromolecules*, vol. 17, p. 2522–2529, Jul 2016.
- [38] Y. H. Lau, T. W. Giessen, W. J. Altenburg, and P. A. Silver, “Prokaryotic nanocompartments form synthetic organelles in a eukaryote,” *Nature Communications*, vol. 9, Apr 2018.
- [39] F. Sigmund, C. Massner, P. Erdmann, A. Stelzl, H. Rolbieski, M. Desai, S. Bricault, T. P. Wörner, J. Snijder, A. Geerlof, H. Fuchs, M. Hrabě de Angelis, A. J. R. Heck, A. Jasanoff, V. Ntziachristos, J. Plitzko, and G. G. Westmeyer, “Bacterial encapsulins as orthogonal compartments for mammalian cell engineering,” *Nature Communications*, vol. 9, May 2018.
- [40] M. E. Cerritelli and W. F. Studier, “Assembly of t7 capsids from independently expressed and purified head protein and scaffolding protein,” *Journal of Molecular Biology*, vol. 258, p. 286–298, May 1996.
- [41] J. Prevelige, P.E., D. Thomas, and J. King, “Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells,” *Biophysical Journal*, vol. 64, p. 824–835, Mar 1993.

- [42] S. R. Casjens and R. W. Hendrix, “Bacteriophage lambda: Early pioneer and still relevant,” *Virology*, vol. 479–480, p. 310–330, May 2015.
- [43] P. Mahinthichaichan, D. M. Morris, Y. Wang, G. J. Jensen, and E. Tajkhorshid, “Selective permeability of carboxysome shell pores to anionic molecules,” *The Journal of Physical Chemistry B*, vol. 122, p. 9110–9118, Sep 2018.
- [44] K. Zeth, “Dps biomineralizing proteins: multifunctional architects of nature,” *Biochemical Journal*, vol. 445, p. 297–311, Jul 2012.
- [45] E. M. Williams, S. M. Jung, J. L. Coffman, and S. Lutz, “Pore engineering for enhanced mass transport in encapsulin nanocompartments,” *ACS Synthetic Biology*, vol. 7, p. 2514–2517, Oct 2018.
- [46] L. S. R. Adamson, N. Tasneem, M. P. Andreas, W. Close, E. N. Jenner, T. N. Szyszka, R. Young, L. C. Cheah, A. Norman, H. I. MacDermott-Opeskin, M. L. O’Mara, F. Sainsbury, T. W. Giessen, and Y. H. Lau, “Pore structure controls stability and molecular flux in engineered protein cages,” *Science Advances*, vol. 8, Feb 2022.
- [47] H. Block, B. Maertens, A. Spriestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn, and F. Schäfer, *Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC)*, p. 439–473. Elsevier, 2009.
- [48] T. Wiryaman and N. Toor, “Cryo-em structure of a thermostable bacterial nanocompartment,” *IUCrJ*, vol. 8, p. 342–350, Apr 2021.
- [49] Y. Tang, A. Mu, Y. Zhang, S. Zhou, W. Wang, Y. Lai, X. Zhou, F. Liu, X. Yang, H. Gong, Q. Wang, and Z. Rao, “Cryo-em structure of *Mycobacterium smegmatis* dyp-loaded encapsulin,” *Proceedings of the National Academy of Sciences*, vol. 118, Apr 2021.
- [50] T. Wiryaman and N. Toor, “Recent advances in the structural biology of encapsulin bacterial nanocompartments,” *Journal of Structural Biology: X*, vol. 6, p. 100062, 2022.
- [51] E. C. Theil, X. S. Liu, and T. Tosha, “Gated pores in the ferritin protein nanocage,” *Inorganica Chimica Acta*, vol. 361, p. 868–874, Mar 2008.

-
- [52] J. C. Tracey, M. Coronado, T. W. Giessen, M. C. Y. Lau, P. A. Silver, and B. B. Ward, “The discovery of twenty-eight new encapsulin sequences, including three in anammox bacteria,” *Scientific Reports*, vol. 9, Dec 2019.
- [53] S. J. Kim and M. Shoda, “Purification and characterization of a novel peroxidase from *Geotrichum candidum* dec 1 involved in decolorization of dyes,” *Applied and Environmental Microbiology*, vol. 65, p. 1029–1035, Mar 1999.
- [54] R. Benisch, M. P. Andreas, and T. W. Giessen, “A widespread bacterial protein compartment sequesters and stores elemental sulfur,” *Science Advances*, vol. 10, Feb 2024.
- [55] B. J. LaFrance, C. Cassidy-Amstutz, R. J. Nichols, L. M. Oltrogge, E. Nogales, and D. F. Savage, “The encapsulin from *Thermotoga maritima* is a flavoprotein with a symmetry matched ferritin-like cargo protein,” *Scientific Reports*, vol. 11, Nov 2021.
- [56] J. Snijder, M. van de Waterbeemd, E. Damoc, E. Denisov, D. Grinfeld, A. Bennett, M. Agbandje-McKenna, A. Makarov, and A. J. R. Heck, “Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry,” *Journal of the American Chemical Society*, vol. 136, p. 7295–7299, May 2014.
- [57] J. A. Jones, R. Benisch, and T. W. Giessen, “Encapsulin cargo loading: progress and potential,” *Journal of Materials Chemistry B*, vol. 11, no. 20, p. 4377–4388, 2023.
- [58] C. Cassidy-Amstutz, L. Oltrogge, C. C. Going, A. Lee, P. Teng, D. Quintanilla, A. East-Seletsky, E. R. Williams, and D. F. Savage, “Identification of a minimal peptide tag for in vivo and in vitro loading of encapsulin,” *Biochemistry*, vol. 55, p. 3461–3468, Jun 2016.
- [59] P. Javidpour, S. Deutsch, V. K. Mutalik, N. J. Hillson, C. J. Petzold, J. D. Keasling, and H. R. Beller, “Investigation of proposed ladderane biosynthetic genes from anammox bacteria by heterologous expression in *E. coli*,” *PLOS ONE*, vol. 11, p. e0151087, Mar 2016.
- [60] K. Frade, C. M. Silveira, B. A. Salgueiro, S. Mendes, L. O. Martins, C. Frazão, S. Todorovic, and E. Moe, “Biochemical, biophysical, and structural analysis of an unusual dyp from the extremophile *Deinococcus radiodurans*,” *Molecules*, vol. 29, p. 358, Jan 2024.

- [61] M. Ahmad, J. N. Roberts, E. M. Hardiman, R. Singh, L. D. Eltis, and T. D. H. Bugg, “Identification of dypb from rhodococcus jostii rha1 as a lignin peroxidase,” *Biochemistry*, vol. 50, p. 5096–5107, May 2011.
- [62] I. Boyton, S. C. Goodchild, D. , A. Elbourne, L. E. Collins-Praino, and A. Care, “Characterizing the dynamic disassembly/reassembly mechanisms of encapsulin protein nanocages,” *ACS Omega*, vol. 7, p. 823–836, Dec 2021.
- [63] W. F. Rurup, J. Snijder, M. S. T. Koay, A. J. R. Heck, and J. J. L. M. Cornelissen, “Self-sorting of foreign proteins in a bacterial nanocompartment,” *Journal of the American Chemical Society*, vol. 136, p. 3828–3832, Feb 2014.
- [64] A. Tamura, Y. Fukutani, T. Takami, M. Fujii, Y. Nakaguchi, Y. Murakami, K. Noguchi, M. Yohda, and M. Odaka, “Packaging guest proteins into the encapsulin nanocompartment from rhodococcus erythropolis n771,” *Biotechnology and Bioengineering*, vol. 112, p. 13–20, Sep 2014.
- [65] G. C. Cannon, R. S. English, and J. M. Shively, “In situ assay of ribulose-1,5-bisphosphate carboxylase/oxygenase in thiobacillus neapolitanus,” *Journal of Bacteriology*, vol. 173, p. 1565–1568, Feb 1991.
- [66] S. Tanaka, C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, and T. O. Yeates, “Atomic-level models of the bacterial carboxysome shell,” *Science*, vol. 319, p. 1083–1086, Feb 2008.
- [67] A. L. Boskey, “Biom mineralization: An overview,” *Connective Tissue Research*, vol. 44, p. 5–9, Jan 2003.
- [68] Y. Chen, Y. Feng, J. G. Deveau, M. A. Masoud, F. S. Chandra, H. Chen, D. Zhang, and L. Feng, “Biom mineralization forming process and bio-inspired nanomaterials for biomedical application: A review,” *Minerals*, vol. 9, p. 68, Jan 2019.
- [69] P. A. Frey and G. H. Reed, “The ubiquity of iron,” *ACS Chemical Biology*, vol. 7, p. 1477–1481, Aug 2012.
- [70] P. J. Kiley and H. Beinert, “The role of fe-s proteins in sensing and regulation in bacteria,” *Current Opinion in Microbiology*, vol. 6, p. 181–185, Apr 2003.

- [71] R. P. Hong Enriquez and T. N. Do, “Bioavailability of metal ions and evolutionary adaptation,” *Life*, vol. 2, p. 274–285, Oct 2012.
- [72] T. W. Lyons, C. T. Reinhard, and N. J. Planavsky, “The rise of oxygen in earth’s early ocean and atmosphere,” *Nature*, vol. 506, p. 307–315, Feb 2014.
- [73] D. Galaris, A. Barbouti, and K. Pantopoulos, “Iron homeostasis and oxidative stress: An intimate relationship,” *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1866, p. 118535, Dec 2019.
- [74] S. C. Andrews, A. K. Robinson, and F. Rodríguez-Quñones, “Bacterial iron homeostasis,” *FEMS Microbiology Reviews*, vol. 27, p. 215–237, Jun 2003.
- [75] E. Eren, N. R. Watts, F. Montecinos, and P. T. Wingfield, “Encapsulated ferritin-like proteins: A structural perspective,” *Biomolecules*, vol. 14, p. 624, May 2024.
- [76] M. Merkofer, R. Kissner, R. C. Hider, U. T. Brunk, and W. H. Koppenol, “Fenton chemistry and iron chelation under physiologically relevant conditions: electrochemistry and kinetics,” *Chemical Research in Toxicology*, vol. 19, p. 1263–1269, Sep 2006.
- [77] P. Cornelis, Q. Wei, S. C. Andrews, and T. Vinckx, “Iron homeostasis and management of oxidative stress response in bacteria,” *Metallomics*, vol. 3, no. 6, p. 540, 2011.
- [78] Y. Luo, E. S. Henle, and S. Linn, “Oxidative damage to dna constituents by iron-mediated fenton reactions,” *Journal of Biological Chemistry*, vol. 271, p. 21167–21176, Aug 1996.
- [79] M. Sánchez, L. Sabio, N. Gálvez, M. Capdevila, and J. M. Dominguez-Vera, “Iron chemistry at the service of life,” *IUBMB Life*, vol. 69, p. 382–388, Feb 2017.
- [80] I. Delany, R. Rappuoli, and V. Scarlato, “Fur functions as an activator and as a repressor of putative virulence genes in neisseria meningitidis,” *Molecular Microbiology*, vol. 52, p. 1081–1090, Apr 2004.
- [81] S. W. Seo, D. Kim, H. Latif, E. J. O’Brien, R. Szubin, and B. O. Palsson, “Deciphering fur transcriptional regulatory network highlights its complex role beyond iron metabolism in escherichia coli,” *Nature Communications*, vol. 5, Sep 2014.

- [82] P. Chandrangsou, C. Rensing, and J. D. Helmann, “Metal homeostasis and resistance in bacteria,” *Nature Reviews Microbiology*, vol. 15, p. 338–350, Mar 2017.
- [83] G. Guan, A. Pinochet-Barros, A. Gaballa, S. J. Patel, J. M. Argüello, and J. D. Helmann, “Pfet, a p1b4-type atpase, effluxes ferrous iron and protects bacillus subtilis against iron intoxication,” *Molecular Microbiology*, vol. 98, p. 787–803, Sep 2015.
- [84] S. C. Andrews, “The ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1800, p. 691–705, Aug 2010.
- [85] M. A. Carrondo, “Ferritins, iron uptake and storage from the bacterioferritin viewpoint,” *The EMBO Journal*, vol. 22, p. 1959–1968, May 2003.
- [86] M. Rivera, “Bacterioferritin: Structure, dynamics, and protein–protein interactions at play in iron storage and mobilization,” *Accounts of Chemical Research*, vol. 50, p. 331–340, Feb 2017.
- [87] D. He, C. Piergentili, J. Ross, E. Tarrant, L. R. Tuck, C. L. Mackay, Z. McIver, K. J. Waldron, D. J. Clarke, and J. Marles-Wright, “Conservation of the structural and functional architecture of encapsulated ferritins in bacteria and archaea,” *Biochemical Journal*, vol. 476, p. 975–989, Mar 2019.
- [88] K. H. Ebrahimi, E. Bill, P.-L. Hagedoorn, and W. R. Hagen, “Spectroscopic evidence for the role of a site of the di-iron catalytic center of ferritins in tuning the kinetics of fe(ii) oxidation,” *Molecular BioSystems*, vol. 12, no. 12, p. 3576–3588, 2016.
- [89] J. Ross, T. Lambert, C. Piergentili, D. He, K. J. Waldron, C. L. Mackay, J. Marles-Wright, and D. J. Clarke, “Mass spectrometry reveals the assembly pathway of encapsulated ferritins and highlights a dynamic ferroxidase interface,” *Chemical Communications*, vol. 56, no. 23, p. 3417–3420, 2020.
- [90] F. Madeira, N. Madhusoodanan, J. Lee, A. Eusebi, A. Niewielska, A. R. N. Tivey, R. Lopez, and S. Butcher, “The embl-ebi job dispatcher sequence analysis tools framework in 2024,” *Nucleic Acids Research*, vol. 52, p. W521–W525, Apr. 2024.
- [91] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, “Fast,

- scalable generation of high-quality protein multiple sequence alignments using clustal omega,” *Molecular Systems Biology*, vol. 7, Jan 2011.
- [92] D. R. Hall, J. M. Hadden, G. A. Leonard, S. Bailey, M. Neu, M. Winn, and P. F. Lindley, “The crystal and molecular structures of diferric porcine and rabbit serum transferrins at resolutions of 2.15 and 2.60 Å, respectively,” *Acta Crystallographica Section D Biological Crystallography*, vol. 58, p. 70–80, Dec 2001.
- [93] K. Honarmand Ebrahimi, P.-L. Hagedoorn, and W. R. Hagen, “Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin,” *Chemical Reviews*, vol. 115, p. 295–326, Nov 2014.
- [94] P. D. Hempstead, A. J. Hudson, P. J. Artymiuk, S. C. Andrews, M. J. Banfield, J. R. Guest, and P. M. Harrison, “Direct observation of the iron binding sites in a ferritin,” *FEBS Letters*, vol. 350, p. 258–262, Aug 1994.
- [95] S. Pfaffen, R. Abdulqadir, N. E. Le Brun, and M. E. Murphy, “Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom,” *Journal of Biological Chemistry*, vol. 288, p. 14917–14925, May 2013.
- [96] J. Tatur, W. R. Hagen, and P. M. Matias, “Crystal structure of the ferritin from the hyperthermophilic archaeal anaerobe pyrococcus furiosus,” *JBIC Journal of Biological Inorganic Chemistry*, vol. 12, p. 615–630, Feb 2007.
- [97] E. Eren, N. R. Watts, J. F. Conway, and P. T. Wingfield, “Myxococcus xanthus encapsulin cargo protein encd is a flavin-binding protein with ferric reductase activity,” *Proceedings of the National Academy of Sciences*, vol. 121, May 2024.
- [98] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, “Highly accurate protein structure prediction with alphafold,” *Nature*, vol. 596, p. 583–589, Jul 2021.

- [99] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Židek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar, “Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models,” *Nucleic Acids Research*, vol. 50, p. D439–D444, Nov 2021.
- [100] P. Meiser, H. B. Bode, and R. Müller, “The unique dkxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation,” *Proceedings of the National Academy of Sciences*, vol. 103, p. 19128–19133, Dec. 2006.
- [101] K. Dziewanowska, M. Settles, S. Hunter, I. Linquist, F. Schilkey, and P. L. Hartzell, “Phase variation in myxococcus xanthus yields cells specialized for iron sequestration,” *PLoS ONE*, vol. 9, p. e95189, Apr 2014.
- [102] D. Kim, J. Choi, S. Lee, H. Hyun, K. Lee, and K. Cho, “Mutants defective in the production of encapsulin show a tan-phase-locked phenotype in myxococcus xanthus,” *Journal of Microbiology*, vol. 57, p. 795–802, Jun 2019.
- [103] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*. Springer Science 38; Business Media, 2nd ed., Aug 2009.
- [104] G. R. Blumenthal and R. J. Gould, “Bremsstrahlung, synchrotron radiation, and Compton scattering of high-energy electrons traversing dilute gases,” *Reviews of Modern Physics*, vol. 42, p. 237–270, Apr 1970.
- [105] A. Jablonski, “Effects of Auger electron elastic scattering in quantitative AES,” *Surface Science*, vol. 188, p. 164–180, Sep 1987.
- [106] L. A. Harris, “Analysis of materials by electron-excited Auger electrons,” *Journal of Applied Physics*, vol. 39, no. 3, p. 1419–1427, 1968.
- [107] B. K. Agarwal, *X-Ray Spectroscopy: An Introduction*. Springer, June 2013.
- [108] E. Y. Xie, Z. Z. Chen, P. R. Edwards, Z. Gong, N. Y. Liu, Y. B. Tao, Y. F. Zhang, Y. J. Chen, I. M. Watson, E. Gu, R. W. Martin, G. Y. Zhang, and M. D. Dawson, “Strain

- relaxation in ingan/gan micro-pillars evidenced by high resolution cathodoluminescence hyperspectral imaging,” *Journal of Applied Physics*, vol. 112, Jul 2012.
- [109] O. Krivanek, T. Lovejoy, and N. Dellby, “Aberration-corrected stem for atomic-resolution imaging and analysis,” *Journal of Microscopy*, vol. 259, p. 165–172, May 2015.
- [110] M. Knoll and E. Ruska, “Das elektronenmikroskop,” *Zeitschrift für Physik*, vol. 78, p. 318–339, May 1932.
- [111] M. von Ardenne, “Das elektronen-rastermikroskop,” *Zeitschrift für Physik*, vol. 109, p. 553–572, Sep 1938.
- [112] M. Von Ardenne, P. Hawkes, and T. Mulvey, *On the history of scanning electron microscopy, of the electron microprobe, and of early contributions to transmission electron microscopy*, p. 25–50. Elsevier, 2021.
- [113] M. T. Otten and W. M. Coene, “High-resolution imaging on a field emission tem,” *Ultramicroscopy*, vol. 48, p. 77–91, Jan. 1993.
- [114] S. J. Pennycook and P. D. Nellist, *Scanning Transmission Electron Microscopy: Imaging and Analysis*. Springer Science and Business Media, Mar 2011.
- [115] N. Tanaka, “Present status and future prospects of spherical aberration corrected tem/stem for study of nanomaterials,” *Science and Technology of Advanced Materials*, vol. 9, p. 014111, Jan. 2008.
- [116] M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius, and K. Urban, “A spherical-aberration-corrected 200kv transmission electron microscope,” *Ultramicroscopy*, vol. 75, p. 53–60, Oct. 1998.
- [117] O. Krivanek, N. Dellby, and A. Lupini, “Towards sub-Å electron beams,” *Ultramicroscopy*, vol. 78, p. 1–11, Jun 1999.
- [118] N. Jian, M. Dowle, R. D. Horniblow, C. Tselepis, and R. E. Palmer, “Morphology of the ferritin iron core by aberration corrected scanning transmission electron microscopy,” *Nanotechnology*, vol. 27, p. 46LT02, Oct 2016.

- [119] N. de Jonge, R. Ramachandra, B. Northan, N. Poirier-Demers, and D. Drouin, “Three-dimensional aberration-corrected scanning transmission electron microscopy of biological specimens,” *Microscopy and Microanalysis*, vol. 16, p. 848–849, Jul 2010.
- [120] R. Erni, M. D. Rossell, C. Kisielowski, and U. Dahmen, “Atomic-resolution imaging with a sub-50-pm electron probe,” *Physical Review Letters*, vol. 102, Mar 2009.
- [121] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, “Electron ptychography of 2d materials to deep sub-ångström resolution,” *Nature*, vol. 559, p. 343–349, Jul 2018.
- [122] E. Okunishi, I. Ishikawa, H. Sawada, F. Hosokawa, M. Hori, and Y. Kondo, “Visualization of light elements at ultrahigh resolution by stem annular bright field microscopy,” *Microscopy and Microanalysis*, vol. 15, p. 164–165, Jul 2009.
- [123] A. V. Crewe, J. Wall, and L. M. Welter, “A high-resolution scanning transmission electron microscope,” *Journal of Applied Physics*, vol. 39, p. 5861–5868, Dec 1968.
- [124] A. V. Crewe, J. Wall, and J. Langmore, “Visibility of single atoms,” *Science*, vol. 168, p. 1338–1340, Jun 1970.
- [125] R. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*. New York: Springer Science 38; Business Media, 3rd ed., Jul 2011.
- [126] K. Sohlberg, T. J. Pennycook, W. Zhou, and S. J. Pennycook, “Insights into the physical chemistry of materials from advances in haadf-stem,” *Physical Chemistry Chemical Physics*, vol. 17, no. 6, p. 3982–4006, 2015.
- [127] T. Walther, “A new experimental procedure to quantify annular dark field images in scanning transmission electron microscopy,” *Journal of Microscopy*, vol. 221, p. 137–144, Feb. 2006.
- [128] D. Thomas, “Mass analysis of biological macromolecular complexes by stem,” *Biology of the Cell*, vol. 80, no. 2–3, p. 181–192, 1994.
- [129] C. Goldsbury, U. Baxa, M. N. Simon, A. C. Steven, A. Engel, J. S. Wall, U. Aebi, and S. A. Müller, “Amyloid structure and assembly: Insights from scanning transmission electron microscopy,” *Journal of Structural Biology*, vol. 173, p. 1–13, Jan 2011.

-
- [130] P. Nellist and S. Pennycook, *The principles and interpretation of annular dark-field Z-contrast imaging*, p. 147–203. Elsevier, 2000.
- [131] M. O. Krause and J. H. Oliver, “Natural widths of atomic k and l levels, $k\alpha$ x-ray lines and several kll auger lines,” *Journal of Physical and Chemical Reference Data*, vol. 8, pp. 329–338, 1979.
- [132] R. D. Leapman and J. A. Hunt, “Comparison of detection limits for eels and edxs,” *Microscopy Microanalysis Microstructures*, vol. 2, no. 2–3, p. 231–244, 1991.
- [133] R. Schneider, “Energy-dispersive x-ray spectroscopy (edxs),” *Surface and Thin Film Analysis*, p. 293–310, Apr. 2011.
- [134] C. Jeanguillaume and C. Colliex, “Spectrum-image: The next step in eels digital acquisition and processing,” *Ultramicroscopy*, vol. 28, p. 252–257, Apr 1989.
- [135] G. Ruthemann, “Diskrete energieverluste schneller elektronen in festkoerpern,” *Die Naturwissenschaften*, vol. 29, p. 648–648, Oct 1941.
- [136] J. Hillier and R. F. Baker, “Microanalysis by means of electrons,” *Journal of Applied Physics*, vol. 15, p. 663–675, Sep 1944.
- [137] N. D. Browning, M. F. Chisholm, and S. J. Pennycook, “Atomic-resolution chemical analysis using a scanning transmission electron microscope,” *Nature*, vol. 366, p. 143–146, Nov. 1993.
- [138] A. D. Baden, P. A. Cox, R. G. Egdell, A. F. Orchard, and R. J. D. Willmer, “Observation of surface optical phonons on $\text{SrTiO}_3(100)$,” *Journal of Physics C: Solid State Physics*, vol. 14, p. L1081–L1084, Dec 1981.
- [139] O. L. Krivanek, T. C. Lovejoy, N. Dellby, and R. Carpenter, “Monochromated stem with a 30 mev-wide, atom-sized electron probe,” *Microscopy*, vol. 62, p. 3–21, Jan 2013.
- [140] P. Thomas and P. Midgley, “Image-spectroscopy – ii. the removal of plural scattering from extended energy-filtered series by fourier deconvolution,” *Ultramicroscopy*, vol. 88, p. 187–194, Aug 2001.

- [141] T. Walther and H. Stegmann, *Comparison of monochromated electron energy-loss with X-ray absorption near-edge spectra: ELNES vs. XANES*, p. 65–66. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [142] P. A. van Aken, B. Liebscher, and V. J. Styrsa, “Quantitative determination of iron oxidation states in minerals using fe l 2,3 -edge electron energy-loss near-edge structure spectroscopy,” *Physics and Chemistry of Minerals*, vol. 25, p. 323–327, May 1998.
- [143] R. Egerton, Y.-Y. Yang, and S. Cheng, “Characterization and use of the gatan 666 parallel-recording electron energy-loss spectrometer,” *Ultramicroscopy*, vol. 48, p. 239–250, Mar. 1993.
- [144] A. Schröfel, G. Kratošová, I. Šafařík, M. Šafaříková, I. Raška, and L. M. Shor, “Applications of biosynthesized metallic nanoparticles – a review,” *Acta Biomaterialia*, vol. 10, p. 4023–4042, Oct. 2014.
- [145] Y. Li, X. Duan, Y. Qian, L. Yang, and H. Liao, “Nanocrystalline silver particles: Synthesis, agglomeration, and sputtering induced by electron beam,” *Journal of Colloid and Interface Science*, vol. 209, p. 347–349, Jan. 1999.
- [146] D. Kaiser, “Social gliding is correlated with the presence of pili in myxococcus xanthus.,” *Proceedings of the National Academy of Sciences*, vol. 76, no. 11, pp. 5952–5956, 1979.
- [147] P. W. Harrison, M. R. Amode, O. Austine-Orimoloye, A. Azov, M. Barba, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S. K. Bhurji, S. Boddu, P. R. Branco Lins, L. Brooks, S. Ramaraju, L. Campbell, M. C. Martinez, M. Charkhchi, K. Chougule, A. Cockburn, C. Davidson, N. De Silva, K. Dodiya, S. Donaldson, B. El Houdaigui, T. Naboulsi, R. Fatima, C. G. Giron, T. Genez, D. Grigoriadis, G. Ghattaoraya, J. G. Martinez, T. Gurbich, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Lodha, D. Marques-Coelho, G. Maslen, G. Merino, L. Mirabueno, A. Mushtaq, S. Hossain, D. Ogeh, M. P. Sakthivel, A. Parker, M. Perry, I. Piližota, D. Poppleton, I. Prosovetkaia, S. Raj, J. Pérez-Silva, A. Salam, S. Saraf, N. Saraiva-Agostinho, D. Sheppard, S. Sinha, B. Sipos, V. Sitnik, W. Stark, E. Steed, M.-M. Suner, L. Surapaneni, K. Sutinen, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, D. Ware, E. Wass, N. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakraborty, B. Flint, S. Giorgetti, L. Haggerty, G. Ilsley, J. Keatley, J. Loveland, B. Moore,

- J. Mudge, G. Naamati, J. Tate, S. Trevanion, A. Winterbottom, A. Frankish, S. E. Hunt, F. Cunningham, S. Dyer, R. Finn, F. Martin, and A. Yates, “Ensembl 2024,” *Nucleic Acids Research*, vol. 52, p. D891–D899, Nov. 2023.
- [148] E. Redwan, “The optimal gene sequence for optimal protein expression in escherichia coli: principle requirements,” *Arab. J. Biotechnol*, vol. 11, no. 1, pp. 493–510, 2006.
- [149] S. Frackman, G. Kobs, D. Simpson, D. Storts, *et al.*, “Betaine and dmso: enhancing agents for pcr,” *Promega notes*, vol. 65, no. 27-29, pp. 27–29, 1998.
- [150] Sternberg, “Biomedical image processing,” *Computer*, vol. 16, p. 22–34, Jan 1983.
- [151] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, “Fiji: an open-source platform for biological-image analysis,” *Nature Methods*, vol. 9, p. 676–682, Jun 2012.
- [152] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” *arXiv.org*, Apr 2023.
- [153] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, “scikit-image: image processing in Python,” *PeerJ*, vol. 2, p. e453, 6 2014.
- [154] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [155] J. Lewis, “Fast normalized cross-correlation,” *Ind. Light Magic*, vol. 10, 10 2001.
- [156] K. S. Dewi and A. M. Fuad, “Improving the expression of human granulocyte colony stimulating factor in escherichia coli by reducing the gc-content and increasing mrna folding free energy at 5’-terminal end,” *Advanced Pharmaceutical Bulletin*, vol. 10, p. 610–616, Aug. 2020.
- [157] M. P. Andreas and T. W. Giessen, “Heterologous expression and purification of encapsulins in streptomyces coelicolor,” *MethodsX*, vol. 9, p. 101787, 2022.

- [158] Y. Guo, X. Liu, B. Li, J. Yao, T. K. Wood, and X. Wang, “Tail-anchored inner membrane protein elab increases resistance to stress while reducing persistence in escherichia coli,” *Journal of Bacteriology*, vol. 199, May 2017.
- [159] H. Yoshida, Y. Maki, S. Furuike, A. Sakai, M. Ueta, and A. Wada, “Yqjd is an inner membrane protein associated with stationary-phase ribosomes in escherichia coli,” *Journal of Bacteriology*, vol. 194, p. 4178–4183, Aug. 2012.
- [160] M. Gassmann, B. Grenacher, B. Rohde, and J. Vogel, “Quantifying western blots: Pitfalls of densitometry,” *ELECTROPHORESIS*, vol. 30, p. 1845–1855, June 2009.
- [161] K. Ohgane and H. Yoshioka, *Quantification of Gel Bands by an Image J Macro, Band/Peak Quantification Tool v1*. Oct. 2019.
- [162] J. L. Smith, “The physiological role of ferritin-like compounds in bacteria,” *Critical Reviews in Microbiology*, vol. 30, p. 173–185, Jan. 2004.
- [163] M. C. Good, J. G. Zalatan, and W. A. Lim, “Scaffold proteins: Hubs for controlling the flow of cellular information,” *Science*, vol. 332, p. 680–686, May 2011.
- [164] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, p. 62–66, Jan 1979.
- [165] P.-S. Liao, T.-S. Chen, and P. C. Chung, “A fast algorithm for multilevel thresholding,” *J. Inf. Sci. Eng.*, vol. 17, pp. 713–727, 2001.
- [166] F. d. l. Peña, E. Prestat, V. T. Fauske, J. Lähnemann, P. Burdet, P. Jokubauskas, T. Furnival, C. Francis, M. Nord, T. Ostasevicius, K. E. MacArthur, D. N. Johnstone, M. Sarahan, J. Taillon, T. Aarholt, pquinn dls, V. Migunov, A. Eljarrat, J. Caron, T. Nemoto, T. Poon, S. Mazzucco, actions user, N. Tappy, N. Cautaerts, S. Somnath, T. Slater, M. Walls, pietsjoh, and H. Ramsden, “/hyperspy: v2.1.1,” Jul 2024.
- [167] R. M. Cornell and U. Schwertmann, *The Iron Oxides: Structure, Properties, Reactions, Occurrences and Uses*. John Wiley and Sons, Dec 2006.
- [168] S. Tabor, “Expression using the t7 rna polymerase/promoter system,” *Current Protocols in Molecular Biology*, vol. 11, Jul 1990.

- [169] F. C. Meldrum, B. R. Heywood, and S. Mann, “Magnetoferritin: in vitro synthesis of a novel magnetic protein,” *Science*, vol. 257, p. 522–523, Jul 1992.
- [170] Y. Zhang, X. Wang, C. Chu, Z. Zhou, B. Chen, X. Pang, G. Lin, H. Lin, Y. Guo, E. Ren, P. Lv, Y. Shi, Q. Zheng, X. Yan, X. Chen, and G. Liu, “Genetically engineered magnetic nanocages for cancer magneto-catalytic theranostics,” *Nature Communications*, vol. 11, Oct 2020.
- [171] W. H. Massover, “Radiation damage to protein specimens from electron beam imaging and diffraction: a mini-review of anti-damage approaches, with special reference to synchrotron x-ray crystallography,” *Journal of Synchrotron Radiation*, vol. 14, p. 116–127, Dec. 2006.
- [172] W. D. Pyrz and D. J. Buttrey, “Particle size determination using tem: A discussion of image acquisition and analysis for the novice microscopist,” *Langmuir*, vol. 24, p. 11350–11360, Aug. 2008.
- [173] S. Sueda, K. Yoshida, and N. Tanaka, “Quantification of metallic nanoparticle morphology on tio₂ using haadf-stem tomography,” *Ultramicroscopy*, vol. 110, p. 1120–1127, Aug. 2010.
- [174] Walther T, “The limits of lattice imaging compared to annular dark-field imaging for determining size distributions of nano-particles,” vol. 2 (Materials Sciences), pp. 119–120, Belgian Society for Microscopy, Liege, 1 2004. Accessed on 2024/09/22.
- [175] N. Gálvez, B. Fernández, P. Sánchez, R. Cuesta, M. Ceolín, M. Clemente-León, S. Trasobares, M. López-Haro, J. J. Calvino, O. Stéphan, and J. M. Domínguez-Vera, “Comparative structural and chemical studies of ferritin cores with gradual removal of their iron contents,” *Journal of the American Chemical Society*, vol. 130, p. 8062–8068, May 2008.
- [176] N. E. Le Brun, A. Crow, M. E. Murphy, A. G. Mauk, and G. R. Moore, “Iron core mineralisation in prokaryotic ferritins,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1800, p. 732–744, Aug 2010.

- [177] G. D. Watt, R. B. Frankel, D. Jacobs, H. Huang, and G. C. Papaefthymiou, “Iron(2+) and phosphate interactions in bacterial ferritin from *azotobacter vinelandii*,” *Biochemistry*, vol. 31, p. 5672–5679, June 1992.
- [178] R. K. Watt, R. J. Hilton, and D. M. Graff, “Oxido-reduction is not the only mechanism allowing ions to traverse the ferritin protein shell,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1800, p. 745–759, Aug. 2010.
- [179] G. Niraula, J. A. Coaquira, F. H. Aragon, B. M. Galeano Villar, A. Mello, F. Garcia, D. Muraca, G. Zoppellaro, J. M. Vargas, and S. K. Sharma, “Tuning the shape, size, phase composition and stoichiometry of iron oxide nanoparticles: The role of phosphate anions,” *Journal of Alloys and Compounds*, vol. 856, p. 156940, Mar 2021.
- [180] A. A. Reutovich, A. K. Srivastava, G. L. Smith, A. Foucher, D. M. Yates, E. A. Stach, G. C. Papaefthymiou, P. Arosio, and F. Bou-Abdallah, “Effect of phosphate and ferritin subunit composition on the kinetics, structure, and reactivity of the iron core in human homo- and heteropolymer ferritins,” *Biochemistry*, vol. 61, p. 2106–2117, Sep 2022.
- [181] A. V. Almeida, A. J. Carvalho, and A. S. Pereira, “Encapsulin nanocages: Protein encapsulation and iron sequestration,” *Coordination Chemistry Reviews*, vol. 448, p. 214188, Dec. 2021.
- [182] H. Abdul-Tehrani, A. J. Hudson, Y.-S. Chang, A. R. Timms, C. Hawkins, J. M. Williams, P. M. Harrison, J. R. Guest, and S. C. Andrews, “Ferritin mutants of *escherichia coli* are iron deficient and growth impaired, and *fur* mutants are iron deficient,” *Journal of Bacteriology*, vol. 181, p. 1415–1428, Mar. 1999.
- [183] J. P. McHugh, F. Rodríguez-Quñones, H. Abdul-Tehrani, D. A. Svistunenko, R. K. Poole, C. E. Cooper, and S. C. Andrews, “Global iron-dependent gene regulation in *escherichia coli*,” *Journal of Biological Chemistry*, vol. 278, p. 29478–29486, Aug. 2003.
- [184] R. W. G. Wyckoff and E. D. Crittenden, “Vi. herstellung und kristallstruktur von ferrooxyd (feo),” *Zeitschrift für Kristallographie - Crystalline Materials*, vol. 63, p. 144–147, Nov. 1926.
- [185] A. A. Claassen, “The scattering power of oxygen and iron for x-rays,” *Proceedings of the Physical Society of London*, vol. 38, p. 482–487, Jan. 1925.

- [186] R. L. Blake, R. E. Hessevick, T. Zoltai, and L. W. Finger, "Refinement of the hematite structure," *American Mineralogist*, vol. 51, pp. 123–129, 02 1966.
- [187] T. Pecharromán, C. González-Carreño and J. E. Iglesias, "The infrared dielectric properties of maghemite, γ -Fe₂O₃, from reflectance measurement on pressed powders," *Phys Chem Minerals*, vol. 22, p. 21–29, 1995.
- [188] J.-L. Hazemann, J. Bézar, and A. Manceau, "Rietveld studies of the aluminium-iron substitution in synthetic goethite," *Materials Science Forum*, vol. 79–82, p. 821–826, Jan. 1991.
- [189] V. A. Drits, B. A. Sakharov, A. L. Salyn, and A. Manceau, "Structural model for ferrihydrite," *Clay Minerals*, vol. 28, no. 2, p. 185–207, 1993.
- [190] G. J. Long, A. K. Cheetham, and P. D. Battle, "Study of the iron-phosphorus-oxygen system by moessbauer effect, neutron diffraction, magnetic susceptibility, and analytical electron microscopy: some pitfalls and solutions in the analysis of a complex mixture," *Inorganic Chemistry*, vol. 22, p. 3012–3016, Oct. 1983.
- [191] M. S. I. Khan, S.-W. Oh, and Y.-J. Kim, "Power of scanning electron microscopy and energy dispersive x-ray analysis in rapid microbial detection and identification at the single cell level," *Scientific Reports*, vol. 10, Feb 2020.
- [192] F. C. Neidhardt, J. L. Ingraham, and M. Schaechter, *Physiology of the Bacterial Cell: A Molecular Approach*. Sinauer Associates, Incorporated, 1990.
- [193] J. Pérez, J. Muñoz-Dorado, and A. Moraleda-Muñoz, "The complex global response to copper in the multicellular bacterium *Myxococcus xanthus*," *Metallomics*, vol. 10, no. 7, p. 876–886, 2018.
- [194] H. Irving and R. J. P. Williams, "637. the stability of transition-metal complexes," *Journal of the Chemical Society (Resumed)*, p. 3192, 1953.
- [195] S. Mann, J. V. Bannister, and R. J. Williams, "Structure and composition of ferritin cores isolated from human spleen, limpet (*Patella vulgata*) hemolymph and bacterial (*Pseudomonas aeruginosa*) cells," *Journal of Molecular Biology*, vol. 188, p. 225–232, Mar 1986.

- [196] A. Treffry, P. M. Harrison, M. I. Cleton, W. C. de Bruijn, and S. Mann, “A note on the composition and properties of ferritin iron cores,” *Journal of Inorganic Biochemistry*, vol. 31, p. 1–6, Sep 1987.
- [197] V. J. Wade, A. Treffry, J.-P. Laulhère, E. R. Bauminger, M. I. Cleton, S. Mann, J.-F. Briat, and P. M. Harrison, “Structure and composition of ferritin cores from pea seed (*pisum sativum*),” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1161, p. 91–96, Jan 1993.
- [198] E. C. Theil, “Ferritin: The protein nanocage and iron biomineral in health and in disease,” *Inorganic Chemistry*, vol. 52, p. 12223–12233, Oct. 2013.
- [199] F. de la Peña, M.-H. Berger, J.-F. Hocheplied, F. Dynys, O. Stephan, and M. Walls, “Mapping titanium and tin oxide phases using eels: An application of independent component analysis,” *Ultramicroscopy*, vol. 111, p. 169–176, Jan 2011.
- [200] G. Haberfehlner, S. F. Hoefler, T. Rath, G. Trimmel, G. Kothleitner, and F. Hofer, “Benefits of direct electron detection and pca for eels investigation of organic photo-voltaics materials,” *Micron*, vol. 140, p. 102981, Jan 2021.
- [201] Y.-H. Pan, A. Brown, K. Sader, R. Brydson, M. Gass, and A. Bleloch, “Quantification of absolute iron content in mineral cores of cytosolic ferritin molecules in human liver,” *Materials Science and Technology*, vol. 24, p. 689–694, Jun 2008.
- [202] Y.-H. Pan, K. Sader, J. J. Powell, A. Bleloch, M. Gass, J. Trinick, A. Warley, A. Li, R. Brydson, and A. Brown, “3d morphology of the human hepatic ferritin mineral core: New evidence for a subunit structure revealed by single particle analysis of haadf-stem images,” *Journal of Structural Biology*, vol. 166, p. 22–31, Apr 2009.
- [203] H. Shuman and A. Somlyo, “Electron energy loss analysis of near-trace-element concentrations of calcium,” *Ultramicroscopy*, vol. 21, p. 23–32, Jan 1987.
- [204] R. Leapman and C. Swyt, “Separation of overlapping core edges in electron energy loss spectra by multiple-least-squares fitting,” *Ultramicroscopy*, vol. 26, p. 393–403, Jan 1988.

- [205] P. A. van Aken and B. Liebscher, “Quantification of ferrous/ferric ratios in minerals: new evaluation schemes of Fe L₂₃ electron energy-loss near-edge spectra,” *Physics and Chemistry of Minerals*, vol. 29, p. 188–200, Apr 2002.
- [206] R. Twesten, “Phase identification and mapping based on valence loss eels and elnes,” *Microscopy Today*, vol. 17, p. 16–19, May 2009.
- [207] P. Gangola and B. P. Rosen, “Maintenance of intracellular calcium in escherichia coli,” *Journal of Biological Chemistry*, vol. 262, p. 12570–12574, Sept. 1987.
- [208] C. A. McPhalen, N. C. Strynadka, and M. N. James, *Calcium-Binding Sites in Proteins: A Structural Perspective*, p. 77–144. Elsevier, 1991.
- [209] L. F. Milles, E. M. Unterauer, T. Nicolaus, and H. E. Gaub, “Calcium stabilizes the strongest protein fold,” *Nature Communications*, vol. 9, Nov 2018.
- [210] V. Norris, S. Seror, S. Casaregola, and I. Holland, “A single calcium flux triggers chromosome replication, segregation and septation in bacteria: a model,” *Journal of Theoretical Biology*, vol. 134, p. 341–350, Oct 1988.
- [211] T. P. Pitta, E. E. Sherwood, A. M. Kobel, and H. C. Berg, “Calcium is required for swimming by the nonflagellated cyanobacterium synechococcus strain wh8113,” *Journal of Bacteriology*, vol. 179, p. 2524–2528, Apr 1997.
- [212] H. Jones, I. Holland, and A. Campbell, “Direct measurement of free Ca²⁺ shows different regulation of Ca²⁺ between the periplasm and the cytosol of escherichia coli,” *Cell Calcium*, vol. 32, p. 183–192, Oct 2002.
- [213] J. C. Slater, “Atomic radii in crystals,” *The Journal of Chemical Physics*, vol. 41, p. 3199–3204, Nov 1964.
- [214] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, *Protein Identification and Analysis Tools on the ExPASy Server*, p. 571–607. Totowa, NJ: Humana Press, 2005.
- [215] R. Leapman and M. Aronova, “Advances and applications of eels in biology,” *Microscopy and Microanalysis*, vol. 26, p. 482–485, Jul 2020.

- [216] R. Erni and N. D. Browning, “Valence electron energy-loss spectroscopy in monochromated scanning transmission electron microscopy,” *Ultramicroscopy*, vol. 104, p. 176–192, Oct 2005.
- [217] L. Garvie and A. Craven, “High-resolution parallel electron energy-loss spectroscopy of mn l2,3-edges in inorganic manganese compounds,” *Physics and Chemistry of Minerals*, vol. 21, Aug 1994.
- [218] Y. Pan, A. Brown, R. Brydson, A. Warley, A. Li, and J. Powell, “Electron beam damage studies of synthetic 6-line ferrihydrite and ferritin molecule cores within a human liver biopsy,” *Micron*, vol. 37, p. 403–411, Jul 2006.
- [219] Y.-H. Pan, G. Vaughan, R. Brydson, A. Bleloch, M. Gass, K. Sader, and A. Brown, “Electron-beam-induced reduction of fe3+ in iron phosphate dihydrate, ferrihydrite, haemosiderin and ferritin as revealed by electron energy-loss spectroscopy,” *Ultramicroscopy*, vol. 110, p. 1020–1032, Jul 2010.
- [220] A. P. Brown, S. Hillier, and R. M. D. Brydson, “Quantification of fe-oxidation state in mixed valence minerals: a geochemical application of eels revisited,” *Journal of Physics: Conference Series*, vol. 902, p. 012016, Sep 2017.
- [221] J. S. Rohrer, Q. T. Islam, G. D. Watt, D. E. Sayers, and E. C. Theil, “Iron environment in ferritin with large amounts of phosphate, from azotobacter vinelandii and horse spleen, analyzed using extended x-ray absorption fine structure (exafs),” *Biochemistry*, vol. 29, p. 259–264, Jan 1990.
- [222] P. Doig, J. W. Austin, and T. J. Trust, “The helicobacter pylori 19.6-kilodalton protein is an iron-containing protein resembling ferritin,” *Journal of Bacteriology*, vol. 175, p. 557–560, Jan 1993.
- [223] H. Aitken-Rogers, C. Singleton, A. Lewin, A. Taylor-Gee, G. R. Moore, and N. E. Le Brun, “Effect of phosphate on bacterioferritin-catalysed iron(ii) oxidation,” *JBIC Journal of Biological Inorganic Chemistry*, vol. 9, p. 161–170, Dec 2003.
- [224] A. García-Prieto, J. Alonso, D. Muñoz, L. Marcano, A. Abad Díaz de Cerio, R. Fernández de Luis, I. Orue, O. Mathon, A. Muela, and M. L. Fdez-Gubieda, “On

- the mineral core of ferritin-like proteins: structural and magnetic characterization,” *Nanoscale*, vol. 8, no. 2, p. 1088–1099, 2016.
- [225] T. Kuwata, D. Sato, Y. Yanagida, E. Aoki, K. Fujiwara, H. Yoshimura, and M. Ikeguchi, “Morphological difference of escherichia coli non-heme ferritin iron cores reconstituted in the presence and absence of inorganic phosphate,” *JBIC Journal of Biological Inorganic Chemistry*, vol. 27, p. 583–594, Aug. 2022.
- [226] I. G. Macara, T. G. Hoy, and P. M. Harrison, “The formation of ferritin from apoferritin. kinetics and mechanism of iron uptake,” *Biochemical Journal*, vol. 126, p. 151–162, Jan. 1972.
- [227] N. Chasteen and P. M. Harrison, “Mineralization in ferritin: An efficient means of iron storage,” *Journal of Structural Biology*, vol. 126, p. 182–194, Jun 1999.