NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTAMENT OF
CHEMISTRY

FUNCTIONAL CHARACTERIZATION AND DESIGN OF PROTEIN NANOCAGES

ANA VIANA DE ALMEIDA

Master in Biochemistry

DOCTORATE IN RADIATION BIOLOGY AND BIOPHYSICS

NOVA University Lisbon

September 2022

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTAMENT OF
CHEMISTRY

FUNCTIONAL CHARACTERIZATION AND DESIGN OF PROTEIN NANOCAGES

ANA VIANA DE ALMEIDA

Master in Biochemistry

DOCTORATE IN RADIATION BIOLOGY AND BIOPHYSICS

NOVA University Lisbon

September 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

FUNCTIONAL CHARACTERIZATION AND DESIGN OF PROTEIN NANOCAGES

ANA VIANA DE ALMEIDA

Master in Biochemistry

Adviser: Maria Alice dos Santos Pereira
Associate Professor, NOVA University of Lisbon

Co-advisers: Pedro António de Brito Tavares
Assistant Professor, NOVA University of Lisbon

Examination Committee:

Chair: Orlando Manuel Neves Duarte Teodoro,
Full Professor, NOVA University of Lisbon

Rapporteurs: Christian Betzel,
Full Professor, Universität Hamburg
Patrick England,
Senior Technical Officer and Head of the Molecular Biophysics Platform, Institute Pasteur

Adviser: Maria Alice dos Santos Pereira,
Associate Professor, NOVA University of Lisbon

Members: Søren Vrønning Hoffmann,
Senior Researcher, Aarhus University
Bruno José Cardoso Vieira,
Researcher, Instituto Superior Técnico da Universidade de Lisboa
Paulo Manuel Assis Loureiro Limão-Vieira
Full Professor, NOVA University of Lisbon

DOCTORATE IN RADIATION BIOLOGY AND BIOPHYSICS

NOVA University Lisbon
September 2022

Functional Characterization and Design of Protein Nanocages

To my family
To João

ACKNOWLEDGMENTS

This thesis was only possible with the invaluable collaborations and contribution of several people/institutions who supported me during these years. Thus, I would like to show my deepest gratitude:

To my supervisors Alice Pereira and Pedro Tavares. Thank you for accepting in your laboratory and for everything you have done for me over these years. For everything you taught me, for your guidance and for always motivating me both professionally and personally with your example. Thank you for giving me all the conditions to develop this work, for giving me new challenges and also the opportunity to try different things.

I will always be grateful for your companionship and opportunity to grow under your guidance, I deeply hope our memories will grow over the year with scientific collaboration and friendship.

To Paula da Fonseca and Edward Morris for accepting me in your laboratory and for always being available to teach me so much in such a short time. Thanks to you, I accomplished one of my personal and professional goals during my PhD.

To Paulo Limão-Vieira for everything you have taught me both in the classroom and outside the laboratory. I will always remember your guidance and kindness you have gave me. Thank you for the opportunity (together with Alice Pereira and Pedro Tavares) to be part of the RaBBiT doctoral program.

To Søren Hoffmann and Nykola Jones for their collaboration, scientific discussion and for providing us with everything during our trips to ASTRID. To Clément Blanchet for the scientific collaboration and samples measurement. To João Carlos Waerenborgh and Bruno Vieira for the Mössbauer spectra measurements and for always being available. It was a pleasure to work with you all and I hope we can continue these productive collaborations.

To the Institutions that supported this work and myself: the RaBBiT Doctoral Programme (PD/00193/2012), UCIBIO and CEFITEC. To the COST Action CA15126 (ARBRE-MOBIEU). To Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Ensino Superior for my funding during this work (Ph.D. fellowships PD/BD/135477/2017 and COVID/BD/152498/2022).

To my past and present colleagues in the Molecular Biophysics Group. Daniela Penas, Ana Carvalho, João Jacinto, João Guerra, Nuno Coelho, Raquel Pacheco, André Dias, Carlos Rodrigues, Carolina Buga, Carolina Luciano, Diogo Silva, Gonçalo Santos, Jéssica Soares, João Cruz and Miriam Colaço. Specially to all the student that I had the pleasure to co-supervise, I have learned valuable lessons from you.

To Shirin Akbar, for taught me so much and for all the company during my stay in Glasgow.

To Idalina Martins, Conceição Lino, Elisabete Ferreira, Inês Santos, Maria José Carapinha and Teresa Sequeira Carlos. Thank you for your irreplaceable help to the Molecular Biophysics group and to myself during these years.

To my closest friends, Sarvesh Kumar, João Jacinto, Ana Spencer, Catarina Madeira, Catarina Pires, David Cruz, João Paulo Santos. Thank you for all your love and friendship.

To Lina, Belarmino and Miguel for having me as a member of your family.

To my mother Maria, my father António for being everything I would ever ask for. This thesis and my own success are also the result from your own work and love. To my brother Gabriel.

To João, you are my everything, this is also yours.

ABSTRACT

Compartmentalization is an essential cellular mechanism that allows cells to create and organize controlled microenvironments for specific metabolic pathways, increase their reaction rate and/or protect the cell from the harmful effect of substrates or products. Due to the lack of organelles, prokaryotes produce protein-based compartments by protomer self-assembly.

Encapsulins, one of the bacterial nanocompartments most recently described, are protein nanocages with the ability of sequestering other smaller proteins (cargo proteins) within their cavity. The physiological function of encapsulins seems to be determined by the type of cargo proteins encapsulated. *Myxococcus xanthus* encapsulin (EncA) is constituted by 180 protomers assembled into a 32 nm wide cage protein. This protein naturally encapsulates two ferritin-like proteins (EncB and EncC) and a third protein with no predicted activity. The encapsulation of cargo proteins and the characterization of the complex are described in this thesis. No major structural changes in EncA were detected upon cargo encapsulation but the assembly of EncC is shown to be iron-dependent. The function of the shell protein in stabilizing and protecting cargo proteins from thermal denaturation is also demonstrated. Additionally, a novel function of binding and protection of circular double stranded plasmid DNA (pUC19) by EncA was discovered and characterized. Finally, the intrinsic ability of EncA to mineralize iron was observed and described as similar to L-chain ferritins. The encapsulation of EncC or EncB within EncA renders the complexes into a ferritin-like catalytic active system. Differences in the chemical nature of the mineral core formed in the presence of molecular oxygen and hydrogen peroxide were probed, as well as the detection of ferric intermediates in EncC during the ferroxidation and iron mineralization reactions using Mössbauer spectroscopy.

Dps, or DNA-binding protein from starved cells, is another bacterial nanocompartment composed of 12 monomers assembled into a cube-like cage protein ~ 9 nm wide. These proteins are known to bind and protect DNA and to accumulate ferric iron in their cavity, protecting the cell from reactive oxygen species. However, the mechanism by which labile iron returns to the cellular medium is still poorly understood. As such, the second part of this thesis describes the iron release mechanism from the Dps of *Marinobacter hydrocarbonoclasticus* while using WrbA, a flavoprotein, as an electron-transfer partner, NADH as an electron donor and two types of iron acceptors: an inorganic compound (1,10-phenanthroline) and a metalloprotein (rubredoxin). Although in aerobic conditions the iron is only released when an iron chelator is present, labile iron is released into the solvent in anaerobic condition. However, the presence of a chelator increases the rate of release. Rubredoxin was proven to be suitable as a putative biological partner for the iron release from Dps with faster iron release than in the presence of 1,10-phenanthroline.

Keywords: Encapsulin; Cargo protein; Ferritin-like protein; DNA-binding protein from starved cells (Dps); Iron homeostasis; Structural dynamics; DNA binding and condensation; Iron oxidation and mineralization; Iron release from ferritins; Kinetics; Dynamic Light Scattering; Synchrotron Radiation Circular Dichroism; Small Angle X-ray Scattering; Mössbauer spectroscopy

RESUMO

A compartimentalização é um mecanismo celular essencial que permite que as células se organizem internamente e criem microambientes controlados permitindo vias metabólicas específicas, o aumento da taxa reacional e/ou a proteção celular dos efeitos nocivos de substratos ou produtos. Devido à ausência de organelos, os organismos procariontes produzem compartimentos através da produção e auto-organização de protómeros.

As encapsulinas, um tipo de nanocompartimento proteico descoberto recentemente, são *nanocages* com capacidade de encapsular outras proteínas de menores dimensões ("proteínas-carga") dentro da sua cavidade. A função biológica das encapsulinas ainda é relativamente desconhecida, mas parece estar associada à função das proteínas-carga encapsuladas. A encapsulina de *Myxococcus xanthus* (EncA) é composta por 180 monómeros organizados de modo a formar uma proteína esférica oca com 32 nm de diâmetro. Neste organismo, a EncA encapsula duas proteínas semelhantes a Ferritinas (EncB e EncC) e uma terceira sem atividade prevista. O trabalho nesta tese descreve o estudo deste sistema, a ausência de alterações estruturais notórias na EncA aquando do encapsulamento das proteínas-carga e a necessidade da presença de ferro para garantir a oligomerização completa da EncC. Observou-se, ainda, a função protetora da encapsulina das proteínas-carga durante a desnaturação térmica. Adicionalmente, foi descoberta como nova função da EncA a capacidade de se ligar e proteger DNA plasmídico circular (pUC19). Finalmente, demonstrou-se a atividade da EncA em realizar a mineralização de ferro de forma semelhante às ferritinas do tipo L e que a encapsulação das proteínas-carga EncB e EncC converte o complexo num sistema catalítico. As diferenças na natureza do mineral férrico quando este se forma na presença de oxigénio molecular ou de peróxido de hidrogénio e a deteção de intermediários férricos durante a reação de ferroxidação pela EncC foram analisadas através da espectroscopia de Mössbauer.

A Dps, *DNA-binding protein from starved cells*, é outro nanocompartimento bacteriano composto por 12 monómeros que formam uma proteína oca, quase cúbica, com 9 nm de largura. Sabe-se que essas proteínas se ligam ao DNA bacteriano, protegendo-o, e armazenam ferro na sua cavidade protegendo o organismo de danos por espécies reativas de oxigénio. No entanto, o mecanismo pelo qual o ferro é libertado para o meio celular de forma biodisponível é um fenómeno pouco estudado. Assim, na segunda parte desta tese é descrito o mecanismo de libertação de ferro da Dps de *Marinobacter hydrocarbonoclasticus*, acoplando a reação à ação da flavo-proteína WrbA como parceira de transferência eletrónica, NADH como doador de eletrões e dois tipos de aceitadores de ferro: um composto inorgânico (1,10-fenantrolina) e uma metaloproteína (rubredoxina). Apesar de em condições aeróbicas o ferro apenas ser libertado na presença de um quelante, este é libertado para a solução em condições anaeróbicas. Contudo, a presença do quelante aumenta a taxa de libertação. A rubredoxina mostrou ser um bom candidato como parceiro biológico para a libertação de ferro sendo que sua presença aumenta a taxa de libertação de ferro comparativamente à molécula inorgânica.

Keywords: Encapsulina; Proteínas-carga; *Ferritin-like protein*; Dps; Homeostasia de ferro; Dinâmica estrutural; Ligação e condensação de DNA; Oxidação e mineralização de ferro; Libertação de ferro; Cinética; Dispersão Dinâmica de Luz; Dicroísmo Circular com Radiação Sincrotrão; Dispersão de Radiação raio-X a Ângulo Baixo; Espectroscopia de Mössbauer.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

δ	Isomer Shift
ε	Molar extinction coefficient
ΔE_Q	Quadrupole splitting
Δε	Molar circular dichroism
ΔH_m	Enthalpy change of the unfolding transition
A-domain	Axial domain
Abs	Absorbance
AFM	Atomic Force Microscopy
Bfr	Bacterioferritin
CLP	Cargo Loading Peptide
CM	Carboxymethyl-cellulose
Cryo-EM	Cryogenic Electron Microscopy
CyD	Cysteine Desulfurase
Cyt	Cytochrome
DEAE	Diethylaminoethyl Cellulose
DLS	Dynamic Light Scattering
Dmax	Maximum protein diameter
Dps	DNA-binding Proteins from Starved cells
DTT	Dithiothreitol
DyP	Dye-decolorizing Peroxidase
EDTA	Ethylenediaminetetraacetic acid
E-loop	Elongated/Extended Loop
EMSA	Electrophoretic Mobility Shift Assays
EPR	Electron Paramagnetic Resonance

Fd	Ferredoxin
FF	Fast Flow
Flp	Ferritin-like protein
FMN	Flavin Mononucleotide
FOC	Ferroxidation Centers
FolB	Folate Biosynthesis
Ft	Canonical Maxiferritins
Ftn	Bacterial Ferritins
H-chain	Heavy Chain
HD	Hydrodynamic Diameter
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HK97	Hong Kong 97 *Escherichia coli* virus
IEX	Ion Exchange chromatography
IMEF	Iron-Mineralizing activity Encapsulin-associated Firmicute protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
L-chain	Light Chain
LB	Luria-Bertani Medium
MOPS	(3-(N-morpholino)propanesulfonic acid)
MW	Molecular Weigh
MWCO	Molecular Weight Cut-Off
NADH	Nicotinamide Adenine Dinucleotide
NAP	Neutrophil-Activating Protein
NIR-HAO	Nitrite Reductase domain fused to a Hydroxylamine Oxidoreductase domain
OD	Optical Density
P-domain	Peripheral Domain
PDB	Protein Data Bank
pI	Isoelectric Point
PMSF	Phenylmethylsulfonyl Fluoride
Rd	Rubredoxin
R*g*	Radius of Gyration
ROS	Reactive Oxygen Species
R_S	Stokes Radius
SAXS	Small Angle X-ray Scattering

SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SRCD	Synchrotron Radiation Circular Dichroism
rSRCD	Rapid Synchrotron Radiation Circular Dichroism
T	Triangulation number
t	Time
T*agg*	Aggregation Temperature
TCA	Trichloroacetic Acid
TEM	Transmission Electron Microscopy
TEMED	Tetramethylethylenediamine
T*m*	Melting Temperature
Tmp	Temperature
Tris	Tris(Hydroxymethyl)Aminomethane
UV	Ultraviolet Radiation
V_e	Elution Volume
WrbA	NAD(P)H Dehydrogenase (Quinone) Flavodoxin
x-mer	Protein oligomer with *x* protein subunits, *x* is the number of subunits
Z-average	Intensity weighted mean hydrodynamic diameter

1

INTRODUCTION

Sections 1.1 and 1.2 were adapted from the following manuscript, published during this thesis:
Ana V. Almeida, Ana J. Carvalho, Alice S. Pereira, "Encapsulin nanocages: Protein encapsulation and iron sequestration" Coord. Chem. Rev., 448: 214188, (2021). DOI: 10.1016/j.ccr.2021.214188.

Section 1.3.2.3.1 was adapted from the following manuscript, published during this thesis:
Ana V. Almeida, João P. Jacinto, João P. L. Guerra, Bruno J. C. Vieira, João C. Waerenborgh, Nykola C. Jones, Søren V. Hoffmann, Alice S. Pereira, Pedro Tavares., "Structural features and stability of apo- and holo-forms of a simple iron–sulfur protein" Eur. Biophys. J., 50(3-4): 561-570 (2021). DOI: 10.1007/s00249-021-01546-0.

1.1 Cell Compartmentalization

Cells spatially control and constrain metabolically related enzymes through a process called compartmentalization. It was initially suggested that cell compartmentalization was limited to eukaryotic organelles (such as mitochondria, lysosomes, or peroxisomes) due to the lack of intracellular lipid-based membranes in eubacteria and archaea. However, with the discovery of protein-based compartments, prokaryotic cells were found to have the capacity to maintain distinct controlled environments within the cell [1]. These bacterial micro- or nanocompartments are generated by the assembly of a specific number of identical subunits forming highly symmetrical large-scale structures (Figure 1.1) [2]. These grant a localized concentration of enzymes, facilitate the transfer of substrates and products between reactions, sequester toxic substances or block reactive intermediates from reaching the cytosol, where they might cause cellular damage [1,2].

Bacterial compartments have a wide array of sizes and roles within the cell. The carboxysome is an 80 – 140 nm polyhedral proteinaceous bacterial microcompartment that enhances CO_2 fixation in cyanobacteria and chemoautotrophic bacteria due to the presence of key enzymes in the inner cavity [3,4]. Ethanolamine utilization (Eut) microcompartments are present in Enterobacteriaceae and in some Firmicutes and are 100 – 200 nm wide. They prevent the accumulation of the harmful intermediate acetaldehyde in the cytosol while reducing the loss of carbon source by compartmentalization [5,6]. The lumazine synthase (LS) nanocompartment (16 – 29 nm) forms an icosahedral shell that contains an enzymatic complex (riboflavin synthase) involved in riboflavin biosynthesis [7]. The smallest type of nanocages corresponds to the proteins from the Ferritin Family, octahedral protein cages formed by the assembly of 24 subunits with 12 nm of diameter (or 12 in the case of the mini-ferritin DNA-binding protein from starved cells, Dps, with 9 nm) with ferroxidase activity. These proteins accumulate and store a mineral iron core inside their cavity, controlling its bioavailability [8].

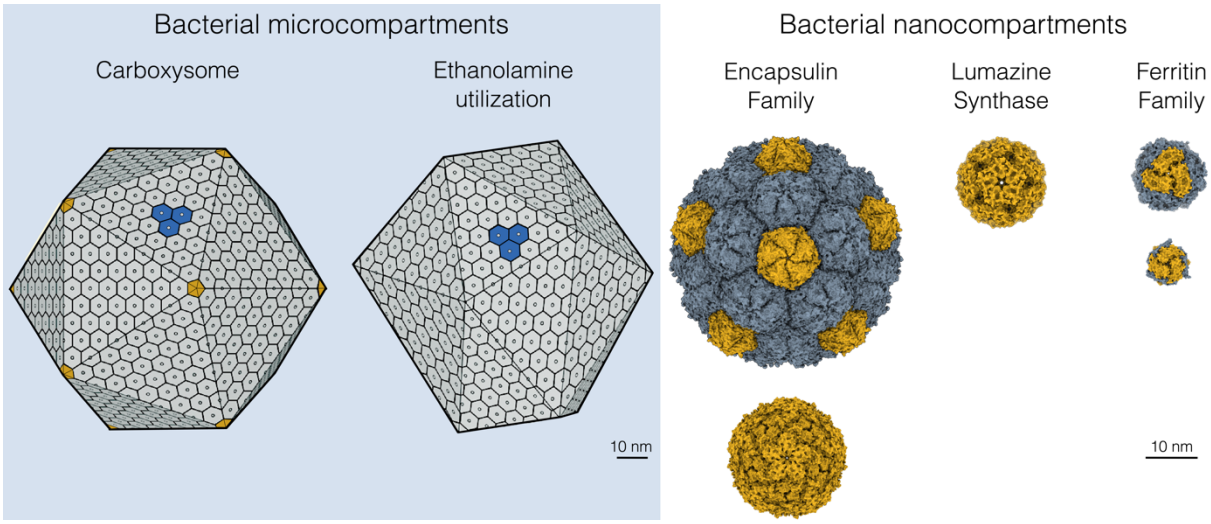

Figure 1.1: Quaternary structure of micro- and nanocompartment shells. Left: Simplified models of the carboxysome and ethanolamine utilization microcompartment assemblies. Right: Atomic structures of bacterial

nanocompartments: two types of encapsulins, lumazine synthase and two members of the Ferritin Family. The carboxysome shell is constituted by pentameric and hexameric vertices, rendering an icosahedral shape. The Eut shell is solely composed of hexamer faces resulting in a less geometrically regular shell. The pentameric vertices are shown in yellow and the hexameric capsomer is identified as blue for carboxysome, Eut, encapsulins and LS. In the Ferritin family the 3-fold channel formed by adjacent N-terminal regions is represented in yellow. Adapted from [9]. Encapsulin from *Quasibacillus thermotolerans* (top structure, PDB: 6NJ8) and *Thermotoga maritima* (bottom structure, PDB: 7KQ5); LS from *Aquifex aeolicus* (PDB: 1HQK); Ferritin from *Mycobacterium tuberculosi*s (top structure, PDB: 3OJ5); and Dps from *Escherichia coli* (bottom structure, PDB: 1DPS).

1.2 Encapsulins

1.2.1 Discovery

A new type of protein nanocompartment was recently identified and classified as the Encapsulin Family (Figure 1.1). Encapsulin systems comprise the interplay between at least two distinct types of proteins: one that self-assemblies into an icosahedral shell-like structure 24 to 42 nm wide (the encapsulin shell) and functional cargo proteins that are encapsulated in the cavity of the former (the cargo proteins). There was some initial disparity regarding the nomenclature of the encapsulin genes upon their identification. In 1994, encapsulins were originally identified in *Brevibacterium (Bre.) linens* as a high molecular weight aggregate classified as linocin M18 [10]. Homologs of this high molecular weight aggregate were later identified in the supernatants of *Mycobacterium (Myc.) tuberculosis*, *Mycobacterium (Myc.) leprae* and *Thermotoga (The.) maritima* cultures [11–13]. In *Streptomyces griseus*, a multimeric protein was identified in the cytoplasm (27 nm of diameter) with a positive effect on DNA replication at late growth phases [14]. These discrepancies were clarified upon the publication of the crystallographic structures of *Pyrococcus (Pyr.) furiosus* and *The. maritima* encapsulins, which showed that the previously described high molecular weight aggregates were protein nanocages loaded with functional cargo proteins [15,16]. So far, the activity of the encapsulin shell is believed to be limited to a structural scaffolding function, while the catalytic activity of the system is ensured by the cargo proteins.

1.2.2 Family of Proteins

Encapsulin monomers share high structural homology with viral capsid protomers (such as the Hong Kong 97 (HK97)-like from Caudovirales), when compared with other bacterial nanocompartments [15–17]. Encapsulins are widely distributed in bacterial and archaeal phyla, representing the larger distribution than any other large-scale compartments in Prokarya. Encapsulins can be found in anaerobic thermophilic microbes, aerobic mesophilic soil bacteria, cyanobacteria and host-associated microbes, both common and pathogenic, which demonstrates the relevance of researching these proteins [18].

The encapsulins proteins are generally classified as Family 1 (classical), Family 2, Family 3 and Family 4, depending on the type of cargo proteins they encapsulate [19]. According to the cargo proteins of Family 1 encapsulins, the function of these systems is closely related with the protection against oxidative stress by preventing the formation of reactive oxygen/nitrogen species (ROS/RNS) through the removal of ferrous iron from the solution or by consuming toxic intermediates of oxygen and nitrogen reactions [18,20–22]. Family 2 encapsulins are evolutionarily distant from their Family 1 counterparts and are expected to be the most numerous Encapsulin sub-family [19–23]. It is also proposed that Family 2 can be further subdivided into Family 2a and Family 2b according to the absence or presence of an internal cyclic nucleotide-monophosphate-binding domain in the E-loop (see section 1.2.3.3). Based on a single study of a member of the Family 2a, the encapsulin shell and its cargo protein seem to be associated with the rescue of sulfur upon starvation. However, there is no experimental evidences currently reported for members of the Family 2b, and all the information about this subclass is based on bioinformatic analysis [19,23,24]. Two additional families have been newly described: Family 3 and Family 4. Family 3 (or natural product encapsulin) is associated to biosynthetic gene clusters resulting in the production of peptide-based natural products. However, the function of their cargo proteins is still unclear. Family 4 (or A-domain encapsulin) only contains the A-domain of the HK97-fold due to a major sequence deletion. This family is mostly found in anaerobic thermophiles found in submarine hydrothermal vents. It is hypothesized that Family 4 might self-assemble as pentamers facets or even larger complexes, yet their function is still speculated [19].

1.2.3 Shell Protein

1.2.3.1 Structure of the Encapsulin Shell

Recently, the high-resolution structures of *Pyr. furiosus*, *The. maritima*, *Myxococcus (Myx.) xanthus*, *Quasibacillus (Qua.) thermotolerans* and *Synechococcus (Syn.) elongatus* encapsulins determined by X-ray crystallography and Cryogenic Electron Microscopy (Cryo-EM) revealed that encapsulins and Caudovirales capsid proteins share high structural similarity. The resemblance lies on the architecture of the monomers and their self-assembly into an icosahedral protein shell [15–17,21,23,25]. Such as viral capsids, encapsulin shells are classified by their size and triangulation number (T). T is a nomenclature used in virology that describes the complexity of a capsid. It is based on the number of equilateral triangles on its surface [26]. The gp5 phage capsid protein of HK97 bacteriophage is composed of 420 monomers. These protomers assemble into a 66 nm wide icosahedral shell with 12 pentameric and 60 hexameric faces (T = 7) (Figure 1.2A) [17]. All encapsulins described so far are smaller than gp5. Likewise, encapsulins are nanocompartments formed by the self-assembly of identical subunits into an icosahedral shell-like structure. Additionally, the encapsulin shell of Families 1 and 2 exhibit differences in protomer organization and overall structure (Figure 1.2).

The first Family is further subdivided into three distinct architectures. The smallest type of encapsulin was found in *The. maritima*. It is composed of 60 identical monomers (~ 31 kDa/monomer) that form a 23 – 24 nm wide cage structure with an inner diameter of 22 nm. It presents an icosahedral T = 1 symmetry with 12 pentamers and no hexamers (Figure 1.2A) [15]. Other examples of 60-mer encapsulins are present in *Rhodospirillum (Rhs.) rubrum*, *Rhodococcus (Rhc.) jostii*, *Myc. turberculosis*, *Bre. linens* and *Mycobacterium (Myc.) smegmatis* [20,22,27–29].

The second encapsulin Family 1 sub-group comprises the *Pyr. furiosus* and *Myx. xanthus* (38.8 and 32.5 kDa per monomer) homologs. Both are built from 180 copies of the same protomer organized in 12 pentamers and 20 hexamers occupying icosahedral vertices and faces. The 180-mer architecture forms a T = 3 symmetry, creating a nearly spherical structure. These encapsulins have an outer diameter of 32 – 36 nm and a 33 Å-thick shell (Figure 1.2A) [16,21]. Notably, the encapsulin from the *Kuenenia (Kue.) stuttgartiensis* Anommox bacterium has an additional diheme cytochrome (cyt) domain fused to the N-terminus of each monomer. It belongs to the Family 1, is a T = 3 encapsulin and assembles into a 180-mer nanocompartment regardless of the truncation of the N-terminal cyt domain. However, the cage of this variant is smaller (29 nm) than the wild-type (diameter of 33 nm) [18]. Strikingly, the expression of the *Myx. xanthus* encapsulin in *Escherichia (Esc.) coli* resulted in two distinct types of shell organization. Besides the expected T = 3 encapsulin, empty protein shells with T = 1 symmetry were also obtained (36 % of the total protein produced) (Figure 1.2A) [30]. This behavior had also been described in a previous work [21].

The third and largest encapsulin Family 1 sub-group includes the protein from *Qua. thermotolerans*, characterized by Cryo-EM. It revealed an icosahedral T = 4 shell topology resulting from the self-assembly of 240 identical subunits, rendering a cage 42 nm wide [25] (Figure 1.2A). The organization of the shell consists of 12 pentameric and 30 hexameric capsomeres. The *Qua. thermotolerans* encapsulin displays a non-covalent chainmail topology established by the interaction of E-loops with two adjacent P-domains through electrostatic, hydrophobic, and potential anion-π interactions. HK97 also presents this topology albeit with distinct physico-chemical properties due to the covalent nature of the interactions [25,31].

The Encapsulin Family 2 is currently solely comprised by the *Syn. elongatus* encapsulin, which is 24.5 nm wide and results from the self-assembly of 60 identical monomers creating an icosahedral T = 1 symmetry [23] (Figure 1.2A). The most notorious structural difference between Family 1 and 2 is observed at the protruding vertices that form the 5-fold symmetry axis leading to a ‘‘spike” morphology in Family 2. Similar to the *Qua. thermotolerans* protein, the encapsulin from *Syn. elongatus* has a chainmail-like topology. This organization is achieved by the interconnection of the extended N-terminal region with the adjacent subunit [23].

1.2.3.2 Assembly Mechanism of the Encapsulin Shell

A report on the *Bre. linens* encapsulin using native mass spectrometry (native MS) suggests a two-step assembly pathway. Initially, there is the self-association into dimers promoted by the E-loop of each subunit across the 2-fold icosahedral symmetry axis. In a second step, a new set of dimers repeatedly assemble into the structure until the complete shell is formed. The cargo protein is encapsulated during this process [28].

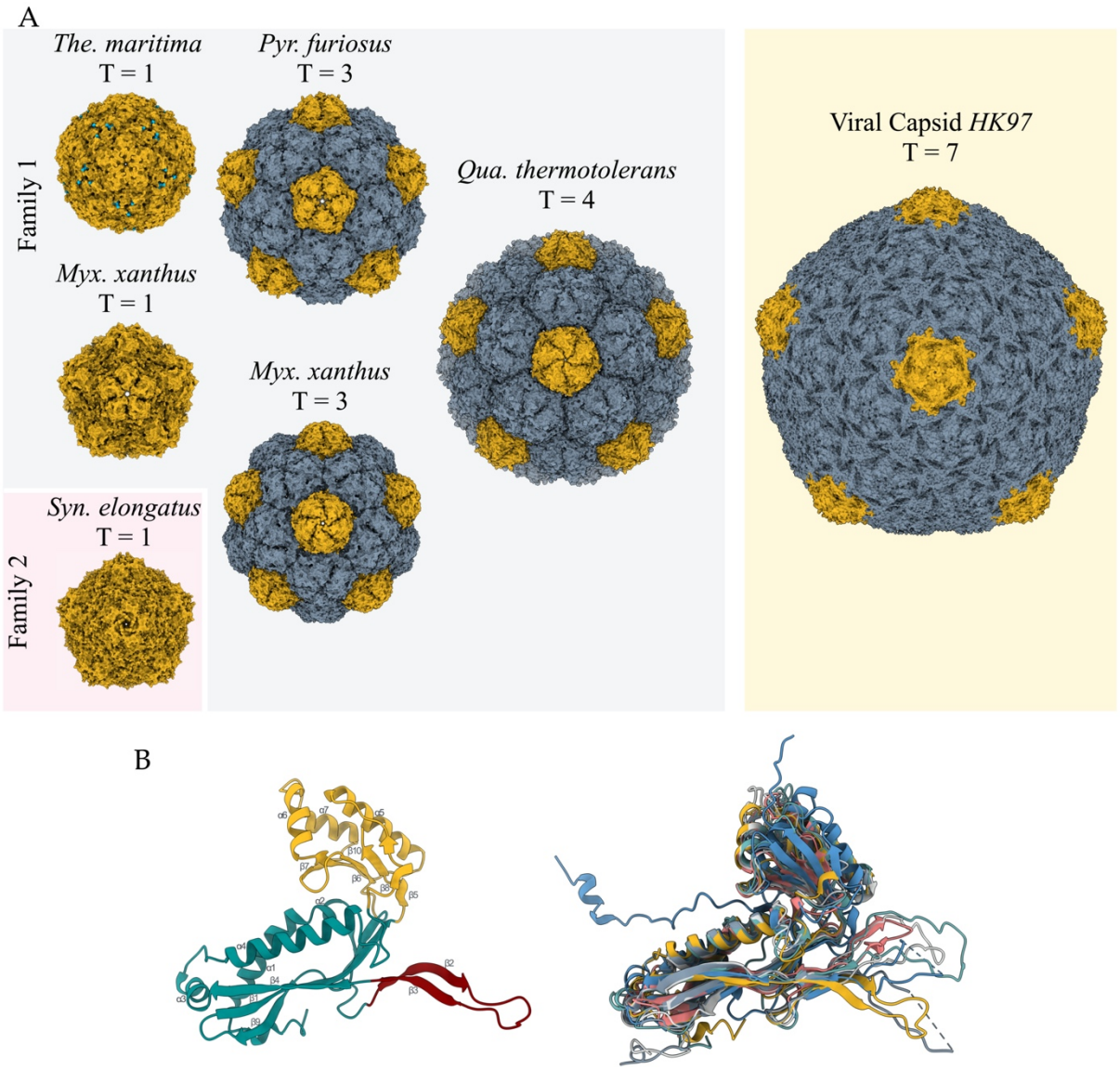

Figure 1.2: Structural overview of the Family 1 and 2 encapsulin proteins. (A) Size comparison between assembled encapsulins. Hexameric facet (in gray) and pentameric vertices (in yellow) for the *The. maritima*, *Syn. elongatus*, *Pyr. furiosus*, *Myx. xanthus* (T = 1 and T = 3), *Qua. thermotolerans* proteins and HK97 viral capsid (proteins in scale). The FMN co-factor of *The. maritima* is colored as blue atoms. **(B)** Tertiary structure of the encapsulin monomers. Left: *The. maritima* monomer colored according to its domains: P-domain in cyan, E-loop in red and A-domain in yellow. The α and β structures are numbered. Right: Superposition of Family 1 and 2 monomers. *The. maritima* in yellow (PDB: 7KQ5), *Syn. elongatus* in blue (PBD: 6X8M), *Pyr. furiosus* in green (PDB: 2E0Z), *Myx. xanthus* T = 1 in

dark gray (PBD: 7S21) and T = 3 in light gray (PBD: 4PT2), *Qua. thermotolerans* in salmon (PDB: 6NJ8). The gaps in the structures are displayed as dashed lines.

1.2.3.3 Structure of the Protomer

Although the sequence identity between encapsulins and viral capsids is relatively low, there are major structural similarities between their monomers. Encapsulin monomers have a HK97-like fold, first reported for the capsid protein gp5 of HK97 lambdoid bacteriophage [15–17]. The HK97-like fold renders a monomer with three different domains, in a α+β structure comprising seven α-helices (α1–α7) and ten β-strands (β1–β10) (Figure 1.2B) [15,16]. Due to their conservation, these domains are termed as the P-domain, A-domain, and E-loop (for peripheral, axial, and elongated/extended, respectively).

The P-domain entails most of the monomer (Figure 1.2B in cyan). It possesses two long (α1 and α4) and two short (α2 and α3) α-helices and three antiparallel β-sheets (β1, β4 and β9) that cover the N-terminal and other non-consecutive regions along the primary sequence. The α-helix and β-sheet regions of this domain protect a hydrophobic pocket that establishes the interaction with the cargo protein [15]. Surprisingly, *Syn. elongatus* exhibits an extended unordered N-terminal region, more characteristic of bacteriophage protein structures [17,23,32]. In the case of the *Pyr. furiosus* encapsulin, the high flexibility of the initial 109 residues prevented atomic resolution [16].

The A-domain creates a compact motif formed by three α-helical (α5, α6 and α7) segments accommodated against five parallel and antiparallel β-sheets (β5, β6, β7, β8 and β10) that enclose the C-terminus (Figure 1.2B in yellow). This domain has almost no contact points with the other two domains from the same monomer, but it is essential in the pentameric and hexameric assemblies that form the encapsulin pores through contacts with the A-domains from other monomers using the highly-flexible loop between helices α6 and α7 [14]. The *Syn. elongatus* encapsulin A-domain presents an elongated C-terminal region, which leads to the distinct Family 2 ''spike" morphology [23].

The E-loop is composed of two-antiparallel β-strands (β2 and β3) and an unordered region (Figure 1.2B in dark red). This domain establishes the contacts between dimers, forming the two-fold symmetry pores. The E-loop is the most variable of all domains between encapsulin homologs due to its flexible nature. The *The. maritima* encapsulin presents a shorter loop with a 60° rotation. This rotation provides a tighter contact with the neighboring protomers. Additionally, the *Myx. xanthus* variant that self-associate into a T = 1 symmetry also display a similar 60° rotation. Differences in the length and angle of this region may be the reason for the distinct assemblies found within this encapsulin sub-group [15,16]. In the *Pyr. furiosus* protein, the subunits forming either pentameric or hexameric capsomeres present noticeable structural variations in both the A-domain loop and the E-loop [16].

1.2.3.4 Structure of the Pore

All known protein micro and nanocompartments, including Dps, ferritins, encapsulins and even carboxysomes, present more than one type of pores. The unique properties of each pore type provide the selective passage of substrates to the cavity trough chemical or steric selection [15,23,25,33–35].

Encapsulins present three types of openings: the 5-, 3- and 2-fold pores. These pores can be 3 to 7 Å wide and differ in their electrostatic properties. In Family 1 the 2-fold channels are constituted by negatively charged residues (Figure 1.3). In the *The. maritima* encapsulin, the 5-fold pores are uncharged and mainly formed by histidine residues. The 3-fold pores are not evolutionarily conserved. In the *The. maritima* shell the 3-fold pores are positively charged [15] (Figure 1.3). Besides having a greater volume, *Qua. thermotolerans* encapsulin also displays the widest pore with 7.2 Å. The pores in this protein are all negatively charged, allowing the channeling of positively charged substrates. However, the 2-fold pores located in 60-mer and 180-mer encapsulins are missing in *Qua. thermotolerans* due to the presence of two asparagine side chains that hinder the opening (Figure 1.3) [25].

Both *The. maritima* and *Qua. thermotolerans* encapsulins present extra electron density in their atomic structures due to the presence of metal ions. In *The. maritima* the histidine residues in the 5-fold channels are coordinating a metal ion [15]. The *Qua. thermotolerans* protein presents extra density in both the 5- and 3-fold pores [25]. A tightly bound flavin-based molecule was found near the 3-fold pore in the *The. maritima* encapsulin (Figure 1.3A) [36]. More recently, this flavin molecule was also found in the interface between two subunits interacting with a tryptophan (W90), suggesting that the *The. maritima* encapsulin might be a flavoprotein, binding preferentially to FMN (flavin mononucleotide) and riboflavin [36,37]. It is predicted that encapsulins from other organisms may also contain this flavin moiety. However, its function is still unknown.

The negatively charged pores in encapsulin Family 1 compensate the positive charges usually present in their substrates. Contrarily, the pores in *Syn. elongatus* encapsulin (from Family 2) are positively charged (Figure 1.3), which is attuned with the channeling of L-cysteine as substrate [23]. Interestingly, a recent study probing the flexibility of the *The. maritima* encapsulin pentameric vertices revealed the dynamic conformation of the pore, with the detection of an ‘‘open” form with a 24 Å aperture and a ‘‘closed” pore, 9 Å wide [38]. This observation suggests a dynamic feature of the encapsulin pore.

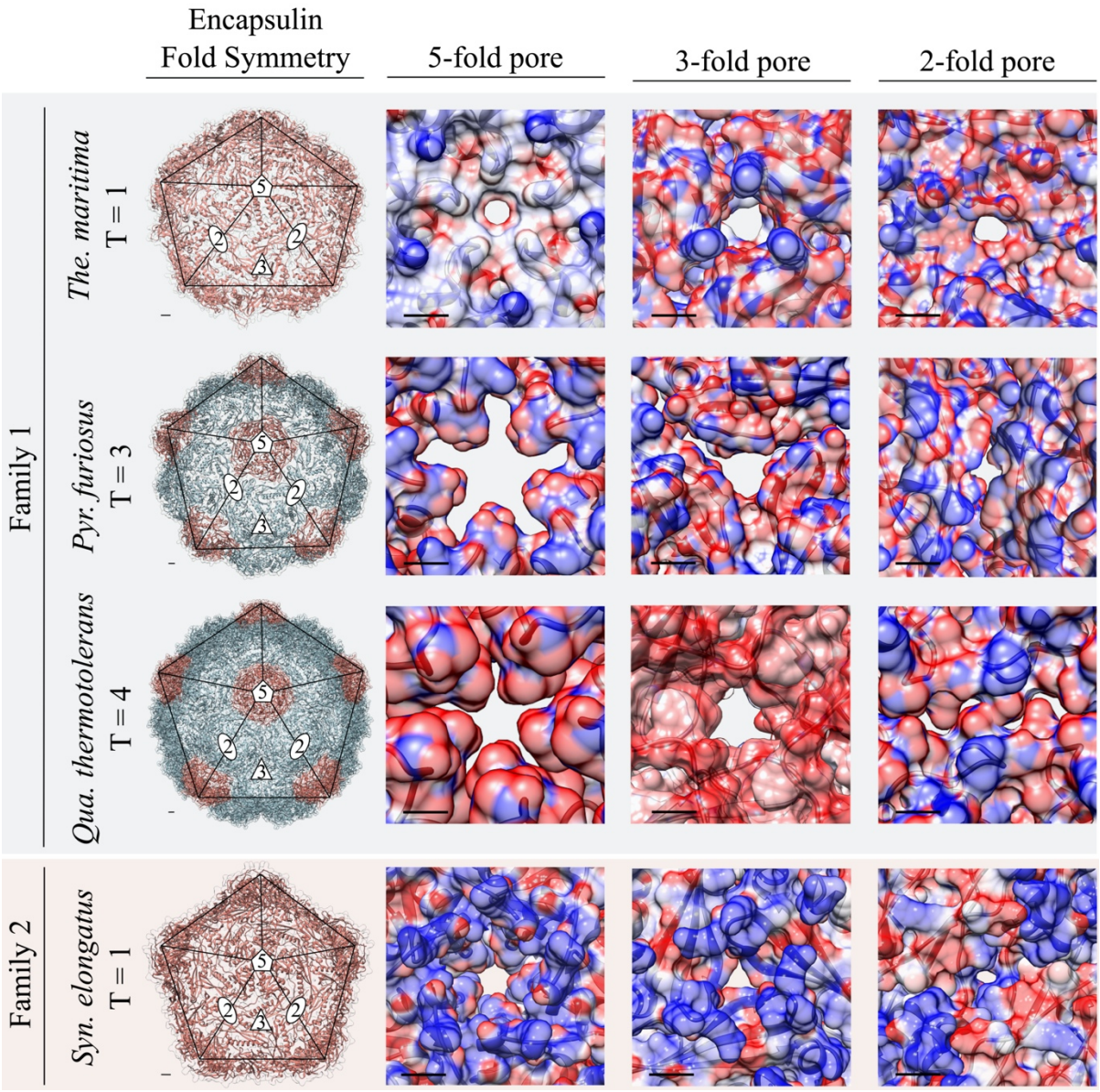

Figure 1.3: Electrostatic surface representation of the encapsulin pores. On the left the three pores are represented with the icosahedral symmetry position. The pentagon corresponds to the 5-fold pore, the triangle to the 3-fold and the ellipse to the 2-fold pore. On the right, the 5-fold, 3-fold and 2-fold pores are colored with blue to red according to their electrostatic potential, from $+10\ kTe^{-1}$ to $-10\ kTe^{-1}$. The scale bar corresponds to 10 Å. The electrostatic surface was calculated using the APBS software [39]. *The. maritima* (PDB:3DKT), *Syn. elongatus* (PBD:6X8M), *Pyr. furiosus* (PDB: 2E0Z), *Myx. xanthus* (PBD:4PT2) and *Qua. thermotolerans* (PDB: 6NJ8). The electrostatic surface was calculated using APBS software.

1.2.4 Cargo Proteins

So far, the function of an encapsulin system is always correlated with the proteins found within the cavity. The ability of the encapsulin shell to harbor cargo proteins was first identified in the *The. maritima* system. Additional electron density was found in its atomic

structure near the internal hydrophobic pockets of the protein shell, in which 8 extra residues (Figure 1.4) were identified, not corresponding to any region of the primary sequence of the encapsulin protomer [15]. These residues were found to be part of the C-terminal sequence of a ferritin-like protein (Flp), encoded by a gene found in the encapsulin operon. The structure of the whole encapsulated Flp was impossible to determine due to the flexibility of its C-terminal region [15]. Subsequent bioinformatics studies identified this C-terminal coding sequence conserved in several other genes usually present immediately upstream the *enc* gene, depending on the organism: Flp, DyP (dye-decolorizing peroxidase), folate biosynthesis (FolB), bacterioferritin (Bfr), ferredoxin (Fd), iron-mineralizing activity encapsulin-associated Firmicute protein (IMEF), nitrite reductase domain fused to a hydroxylamine oxidoreductase domain (NIR-HAO), hemerythrin-like, rubrerythrin-like proteins and biphenyl dehydrogenase were found associated with members of encapsulins Family 1, while cysteine desulfurase (CyD) and polyprenyl diphosphate synthase were found adjacent to members of the Family 2 [15,18,22,23,25].

1.2.4.1 Cargo Loading Peptide

The deletion of the conserved amino acid sequence described above prevents the encapsulation of endogenous cargo proteins and its artificial fusion to the C-terminus of exogenous proteins allows their encapsulation. Therefore, this sequence was designated as ''cargo loading peptide'' (CLP) and deemed necessary for the incorporation of proteins by encapsulins [15,18,22,23,25,40–42]. CLP are predominantly composed of alanine, proline and glycine residues that interact with the hydrophobic pocket of the P-domain and the N-terminal of the encapsulin protomer (Figure 1.4). In fact, the interaction between the encapsulin and the cargo protein molecules is predominantly established by hydrophobic bonds and salt bridges [15].

Besides the CLP found in the *The. maritima* encapsulin, other variations to the encapsulation process have been found. For example, in the *Pyr. furiosus* system the first 109 residues (located inside the encapsulin shell) correspond to the cargo protein that is fused to the encapsulin, resulting in a single polypeptide chain. As such, there is no CLP motif in this case [16]. Furthermore, in the IMEF-encapsulin-Fd system, the IMEF cargo protein presents the CLP at the C-terminus, whereas Fd is tagged at the N-terminus [18]. There is currently only one encapsulin system described containing a single cargo protein with a N-terminal CLP, shared between the cargo proteins of Encapsulin Family 2. Strikingly, the CLP domain of the cargo protein from *Kuenenia* (*Kue.) stuttgartiensis* encapsulin–cyt system is absent. It is suggested that the interaction among the shell and the cargo is established between the cyt domain of the encapsulin shell and the NIR domain of the NIR-HAO fusion protein [18,43].

In *The. maritima* the gene coding for the cargo protein was identified in close proximity to the encapsulin gene, as a core operon [15]. However, cargo proteins can also be found throughout the genome of the organism, and are termed as secondary cargo proteins, such as the EncC and EncD cargo proteins of *Myx. xanthus* or BfrB and FolB of *Myc. tuberculosis*

[18,21,22]. The protein environment inside an encapsulin shell is not necessarily homogeneous. The encapsulins from *Myx. xanthus*, *Myc. tuberculosis* and from Firmicutes can simultaneously encapsulate different types of cargo proteins, such as the EncB-EncC-EncD, DyP-BfrB-FolB and IMEF-Fd cargo proteins, respectively [18,21,22].

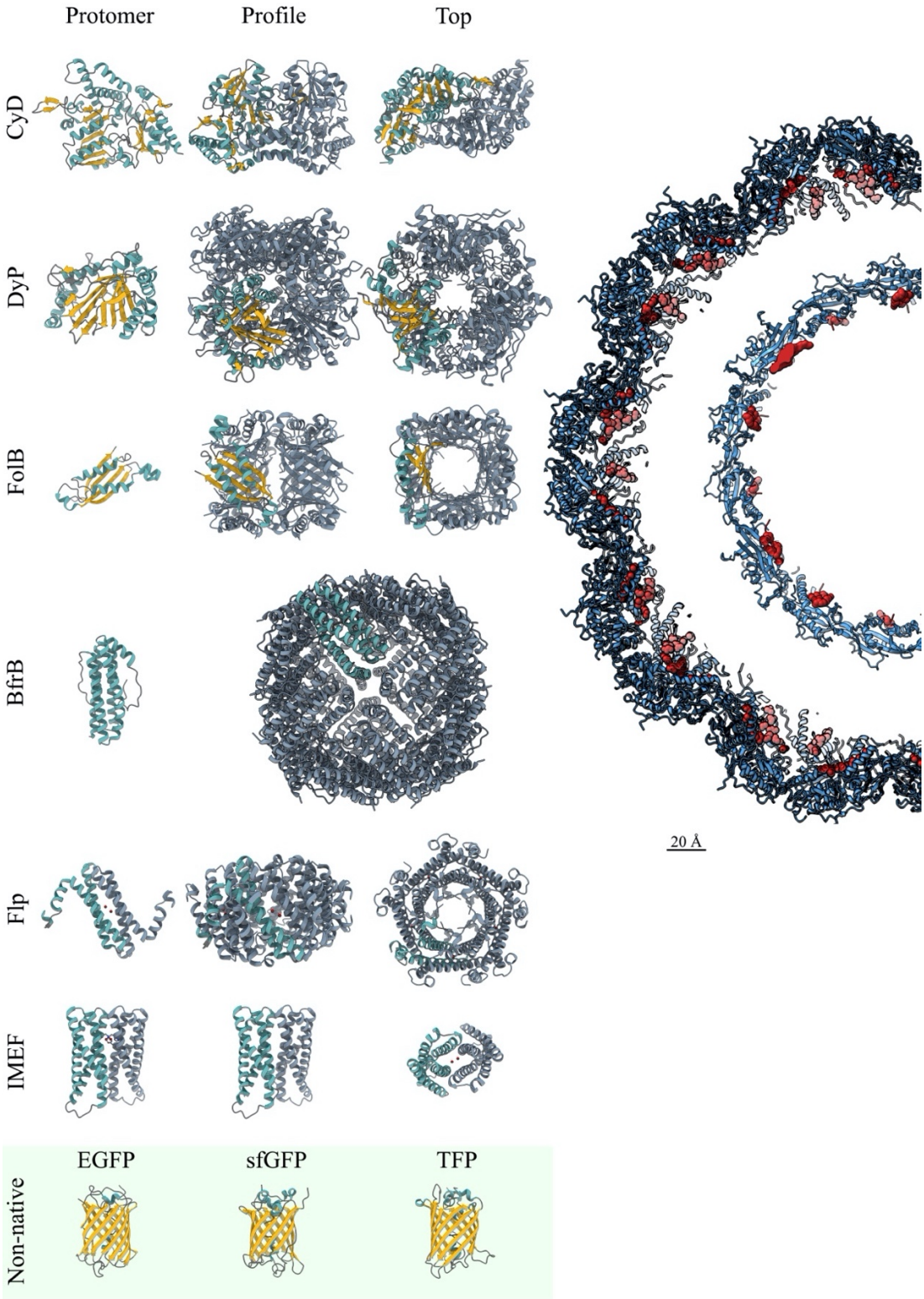

Figure 1.4: High resolution structures of native and non-native cargo proteins of encapsulin systems. Left: Top and profile views of cargo proteins and their protomer/functional units. The secondary structure elements of a

monomer of each cargo protein is colored with α-helices in green, β-sheets in yellow and coiled coils in dark gray, with the remaining monomers represented in gray; the iron atoms are represented as orange spheres with their binding residues shown as atoms. Right: The outer monomers of the encapsulin shell are represented in steel blue and the cargo loading peptides are shown as atoms through red spheres. All structures are represented in a 20 Å scale. Native cargo proteins of encapsulins from Family 2: Cysteine desulfurase (CyD) (PDB: 6C9E) from *Legionella pneumophila*; native cargo proteins of encapsulins from the Family 1: Dye-decolorizing peroxidase B (DypB) (PDB: 3QNR) from *Rhc. jostii*, folate biosynthesis enzyme (FolB) (PDB: 1NBU) from *Myc. tuberculosis*, bacterioferritin B (BfrB) (PDB: 3UNO) from *Myc. tuberculosis*, ferritin-like protein (Flp) (PDB: 5DA5) from *Rhs. rubrum*, iron-mineralizing encapsulin-associated Firmicute (IMEF) (PDB: 6N63) from *Qua. thermotolerans*; non-native cargo proteins: Cyan fluorescent protein derived (TFP) (PDB: 2HQK) from *Clavularia*, enhanced green fluorescent protein (EGFP) (PDB: 4EUL) from *Aequorea victoria*, superfolder green fluorescent protein (sfGFP) (PDB: 2B3P) from *Aequorea victoria*.

The CLP that tether to the encapsulin protomers were identified in the three sub-families from Classical Encapsulins. In 60-mer encapsulins the CLP (and thus the cargo protein) were found directly below the 5-fold pore in *The. maritima* (Figure 1.4), which suggests that these represent the putative pathway for substrate entry and binding [15]. Contrarily, the CLP was located near both 5- and 2-fold symmetry pores in 240-mer encapsulins, but with tighter interaction with the 2-fold pores [25]. The location of the CLP in 180-mer systems was recently identified at the 2-, 3- and 5-fold axes, resulting in similar distributions of the cargo proteins next to the pentameric and hexameric facets [30].

1.2.4.2 Cargo Function

So far there is no evidence that the encapsulin shell presents any function besides the role of cargo stabilization and protection. Encapsulins are known to be resistant to chemical stresses such as pH variation or denaturation by chaotropic agents, mechanical forces and high temperatures [15,27,28,40]. They are also highly resistant to unspecific proteolysis, while their cargo proteins are usually not [40,42]. Thus, the encapsulation process may have evolved to increase the lifetime of the cargo proteins, protecting them from different external factors and maintaining suitable conditions for their enzymatic activity.

In Family 1 encapsulin systems the activity of the cargo proteins is intrinsically related with redox chemistry. DyP-type proteins are members of the Heme Peroxidase Family that use hydrogen peroxide to catalyze the oxidation of a wide variety of substrates, such as lignin or anthraquinone dyes [44,45]. While, FolB cargo proteins are involved in folate (or vitamin B9) biosynthesis. Folate derivatives are essential cofactors for amino acid and nucleic acid synthesis. NIR-HAO cargo protein is suggested to oxidize hydrazine resulting in the formation of N_2. Moreover, distinct cargo proteins such as bacterioferritin, Flp and IMEF perform ferroxidation of ferrous species by the consumption of molecular oxygen.

1.2.4.3 Encapsulin Shell and Cargo Protein System

The encapsulation process limits the range and number of substrate molecules that pass through the shell and reach the cargo protein, as in the case of guaiacol, a substrate of *Myc. tuberculosis* DyP [22]. In some cases, the encapsulin cage also enhances the activity of the cargo protein as shown for the encapsulated DypB of *Rhc. jostii*, with a ~ 8-fold increase in the rate of lignin degradation [27]. The coupling of the NIR and HAO reactions in a fusion protein encapsulated inside the 180-mer encapsulin shell of Anammox bacteria may be considered the first depiction of sequential steps of a metabolic pathway controlled by spatial restriction of a Encapsulin protein cage [18]. Moreover, the Flp and IMEF cargo proteins exhibit ferroxidase activity but no capacity to harbor the ferric species generated, which leads to their toxic precipitation in solution. As such, the partnership with an encapsulin cage enables them to function akin to a large-scale ferritin-like system [20,21].

1.2.5 Iron Sequestration and Oxidative Stress Protection

As previously mentioned, encapsulins are thought to be involved in iron mineralization and storage as well as oxidative stress protection, among other functions. Although the catalytic mechanisms are still unknown, they are thought to act by removing the toxic species, Fe^{2+} ions and H_2O_2, from the intracellular environment.

Iron is often considered the most vital transition metal in Biology. It is present in the active sites of multiple proteins responsible for carrying out a wide variety of functions, from electron-transfer to oxygen transport and nitrogen fixation, among others. It is also the fourth most abundant element in the Earth's crust [46]. In water and at physiological pH, Fe^{2+} ions are relatively soluble (10^{-2} M) whereas the ferric form Fe^{3+} is considered insoluble (10^{-18} M). Though soluble, in the presence of H_2O_2 free Fe^{2+} ions rapidly react to form ROS by the Fenton reaction:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{\bullet} + OH^{-} \qquad \text{Equation 1}$$

Cycling between the Fe^{3+} and Fe^{2+} ionic forms can proceed via the superoxide radical anion:

$$Fe^{3+} + O_2^{\bullet -} \rightarrow Fe^{2+} + O_2 \qquad \text{Equation 2}$$

Combination of the two equations represents the so-called Haber-Weiss reaction:

$$O_2^{\bullet -} + H_2O_2 \rightarrow O_2 + OH^{\bullet} + OH^{-} \qquad \text{Equation 3}$$

Hydroxyl radicals ($OH^{\bullet}$) can react with almost all types of biomacromolecules, nucleic acids, proteins, carbohydrates and lipids, impairing essential functions that can lead to cell death [47]. Hence, living organism have developed several mechanisms to cope with this toxic drawback, including the production of proteins from the Ferritin Family that control the intracellular concentration of free iron. They sequester free Fe^{2+} ions in solution and convert them to a ferric mineral (through ferroxidation and mineralization activities) that is stored inside the protein's cavity and released on demand.

1.2.6 Ferritin-like Proteins

The encapsulins of *The. maritima* and *Pyr. furiosus* were the first reported encapsulins with an Flp identified as cargo protein [15,16,48]. Later, the *Myx. xanthus* homologue was characterized with two Flp proteins: EncB and EncC. Additionally, *Rhs. rubrum* and *Haliangium (Hal.) ochraceum* encapsulins were also reported to have this type of cargo protein [20,48]. The primary sequence (and tertiary structure) of Flp cargo proteins from different organisms is similar. The encapsulated Flp of *Rhs. rubrum* shows 53 % sequence identity with its *The. maritima* counterpart, 33 % with *Myx. xanthus*, and 29 % with *Pyr. furiosus* homologues and the iron binding motif, EXXH (X represents any amino acid residue), is conserved in these proteins [20] (Figure 1.5A and B)

1.2.6.1 Structural Dynamics of Ferritin-like Proteins

Crystallographic structures of the free form of Flp from *Nitrosomonas (Nit.) europaea*, *The. maritima*, *Rhs. rubrum*, *Hal. ochraceum*, *Pyr. furiosus* and both Flp from *Myx. xanthus* encapsulins revealed a new type of assembly, distinct from the previously described ferritins and Dps, displaying a toroidal arrangement composed of 10 identical subunits (Figure 1.5B). Additionally, the crystal structure of a truncated domain of the Flp present in the N-terminal of *Pyr. furiosus* encapsulin shows a similar organization [20,48]. Each monomer contains two long antiparallel α-helices followed by a shorter α-helix (Figure 1.5B, left). The CLP-containing C-terminus is pointed towards the outside of the ring, suggesting that these residues are available to interact with the encapsulin shell.

The decameric structure results from the assembly of five dimers, with two types of dimer interfaces. One of these interfaces contains a di-nuclear FOC, where two iron atoms are coordinated by two glutamates and one histidine residue of each monomer (Figure 1.5B, right). In the other type of dimer interface (formed by a back-to-back dimer conformation) these residues are exposed to the solvent and as such no FOC is formed (Figure 1.5B, middle). Although typical ferritins are 24-mer proteins and contain their FOCs within the four-helix bundle of each monomer, the architectures of the di-iron ferroxidase centers of Flps and ferritins are superimposable [20,48]. The dimerization of some Flp is dependent on the presence of Fe^{2+} ions. Contrarily, Fe^{3+} does not affect the oligomer state of the protein. Additionally, the oligomerization may be induced through the protonation of a conserved histidine involved in FOC formation. The substitution of the histidine residue prevented the decameric assembly, leading to the accumulation of monomers and dimers in the *Rhs. rubrum* Flp or only dimers in the case of the *Hal. ochraceum* protein. Other results further suggest that Flp oligomerization is achieved by self-assembly of non-FOC dimers followed by the association of five of these dimers through the formation of FOC-containing interfaces upon iron binding [20,48,49].

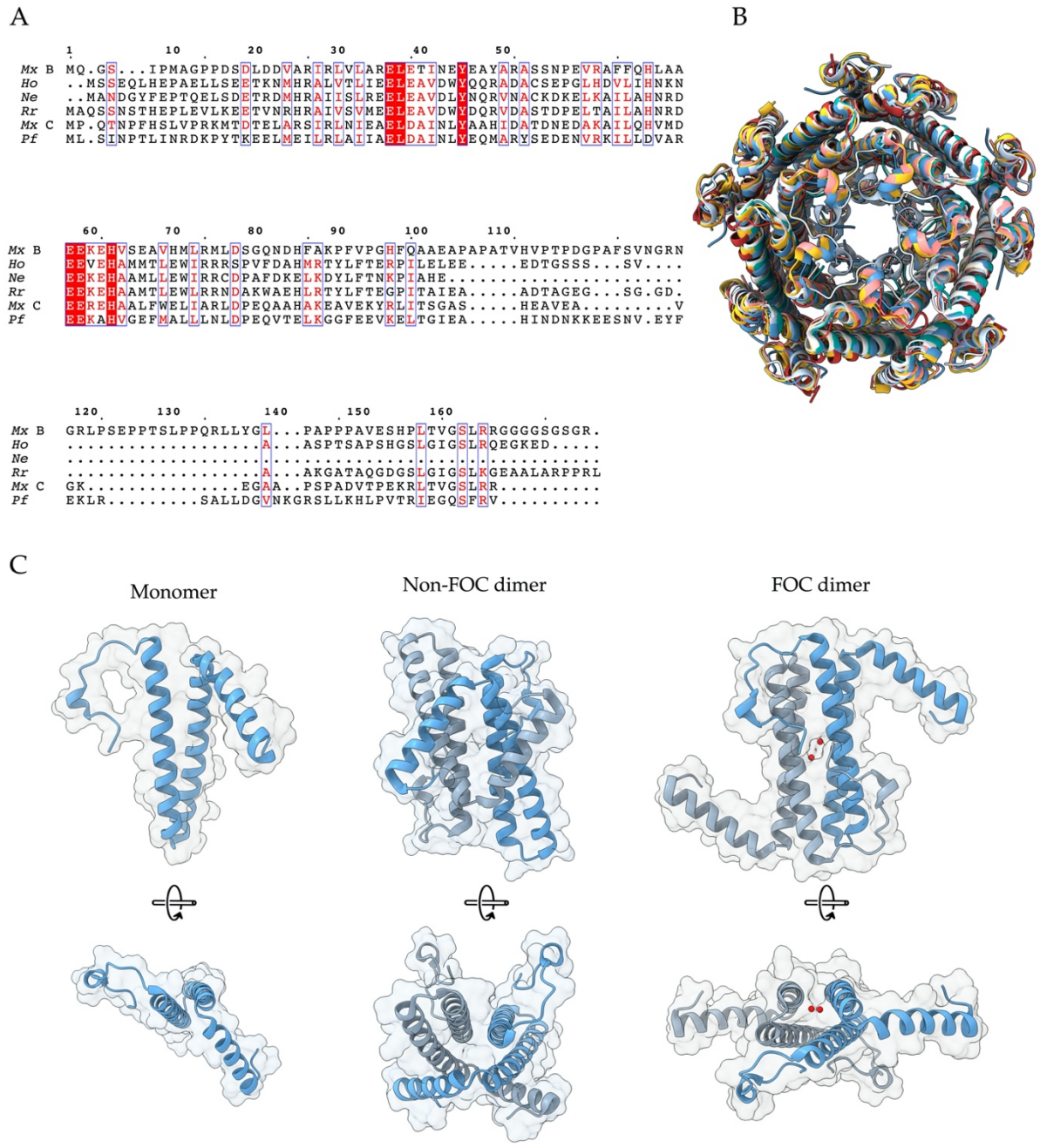

Figure 1.5: Comparison between sequence and structural organization of different Flp cargo proteins. (A) Primary sequence alignment (Clustal Omega and ESPriot) [50,51] of the EncB from *Myx. xanthus (Mx B)*, *Hal. ochraceum (Ho)*, *Nit. europaea (Ne)* and *Rhs. rubrum (Rr)* as well as the EncC from *Myx. xanthus (Mx C) and Pyr. furiosus (Pf).* **(B)** The toroidal decameric structure of the Flp from different organisms are superimposed as cartoons. EncB from *Myx. xanthus* in light gray (PDB: 7S5K), *Hal. ochraceum* in green (PDB: 5N5F), *Nit. europaea* in blue (PDB: 3K6C), *Rhs. rubrum* in salmon (PDB: 5DA5), EncC from *Myx. xanthus* in red (PDB: 7S8T), *Pyr. furiosus* in dark gray (PDB: 5N5E). **(C)** Structure of the monomer, non-FOC dimer and FOC dimer of Flp cargo proteins: Left: a monomer; Middle: a non-FOC dimer interface; Right: A FOC dimer interface of *Pyr. furiosus* Flp. Iron atoms are shown as orange spheres (PDB: 5N5E).

1.2.6.2 Iron Ferroxidation

To accomplish their iron oxidation and mineralization function, the encapsulin-Flp cargo system has to rapidly attract Fe^{2+} ions in solution, translocate them to the active sites in the encapsulated Flp cargo proteins, catalyze their ferroxidation, and direct the ferric species to nucleation/growth sites at the inner surface of the encapsulin shell (Figure 1.6D). While a wealth of structural information on the encapsulin-cargo system, unloaded encapsulin shell and free Flp cargo proteins have been reported in the last years, the aforementioned process is far from being elucidated. Recently, the Flp-encapsulin system has been described as a member from the Ferritin Family.

The presence of dense granules (~ 24 nm in diameter) inside the *Myx. xanthus* encapsulin shell, in close proximity to the Flp cargo proteins was detected using Cryo-EM, scanning transmission electron microscopy (STEM), energy-dispersive X-ray (EDX) spectroscopy and inductively coupled plasma mass spectrometry (ICP-MS). On average, each shell displayed 14 granules composed of iron and phosphate in an approximate molar ratio of 4:1, with 5 – 6 nm of diameter, accounting to a total of ~ 30,000 Fe atoms. The encapsulin-Flp complex from *Rhs. rubrum* was obtained by recombinant co-expression in *Esc. coli* [20]. TEM data revealed the presence of the Flp cargo protein within the encapsulin shell. The complex displayed ferroxidase activity, solubilizing all ferric iron. As before, TEM micrographs revealed highly regular dense particles of ~ 5 nm in diameter proving the formation of the ferric mineral within the encapsulin shell.

The atomic structure of the Flp cargo proteins revealed a di-nuclear FOC in the dimer interface with differential occupancy of the iron binding site that is most probably dependent on the experimental conditions or can be explained by the occurrence of a transient catalytic site [15,20,48,49,52]. Besides the FOCs, additional metal binding sites on the outer surface and near the center of the open ring were detected, consistent with translocation of hydrated Fe^{2+} ions from the outside to the FOC and exit of Fe^{3+} species from the FOC to the putative nucleation sites at the encapsulin inner surface, respectively.

The FOC formed by the Flp dimer interface is occupied by two iron atoms coordinated by three glutamates and one histidine residue from each monomer as a 5-coordinate Fe^{2+} complex (state 1) (Figure 1.6A) similar to the di-nuclear coordination geometry of the FOC found in ferritins [20,30,48]. Additionally, a conserved tyrosine residue is present in the vicinity of this center. In bacterial ferritins this same tyrosine is suggested to form an intermediate tyrosyl radical associated with the conversion of O_2 to H_2O_2 [53–55].

A new type of iron coordination was revealed by the crystal structure of *Myx. xanthus* EncB (state 2). Here, residue E28 suffered a 180° rotation from the iron atom increasing the distance from 2.7 Å to 4.1 – 4.4 Å (Figure 1.6B). This rotation leads to a distinct iron coordination center from the previous high-affinity 5-coordinated state resulting in a lower affinity 4-coordinated state [30]. Additionally, Eren *et al.* detected a glutamate-histidine (three glutamate and two histidine residues from each dimer) ladder along the α-helix near the FOC (Figure 1.6C) functioning as a secondary iron binding center. At the center of this ladder an iron atom is coordinated by four glutamate and two histidine residues similar to the "site C"

of bacterial ferritins placed ~ 7.0 Å away from the FOC. The overall ferroxidation mechanism is still unclear, yet two mechanisms are currently proposed. In the first hypothesis "site C" is not occupied by an additional Fe^{2+} and both Fe^{2+} bound to the FOC are oxidized forming a peroxodiferric intermediate resulting in the formation of $Fe^{3+}$and H_2O_2. In the second hypothesis "site C" is occupied by Fe^{2+} and the conserved tyrosine acts as an electron and H^+ donor to the peroxodiferric intermediate resulting in Fe^{3+} and Fe^{4+} ions [30]. Additional studies are still needed to understand the ferroxidation mechanism of Flp cargo proteins.

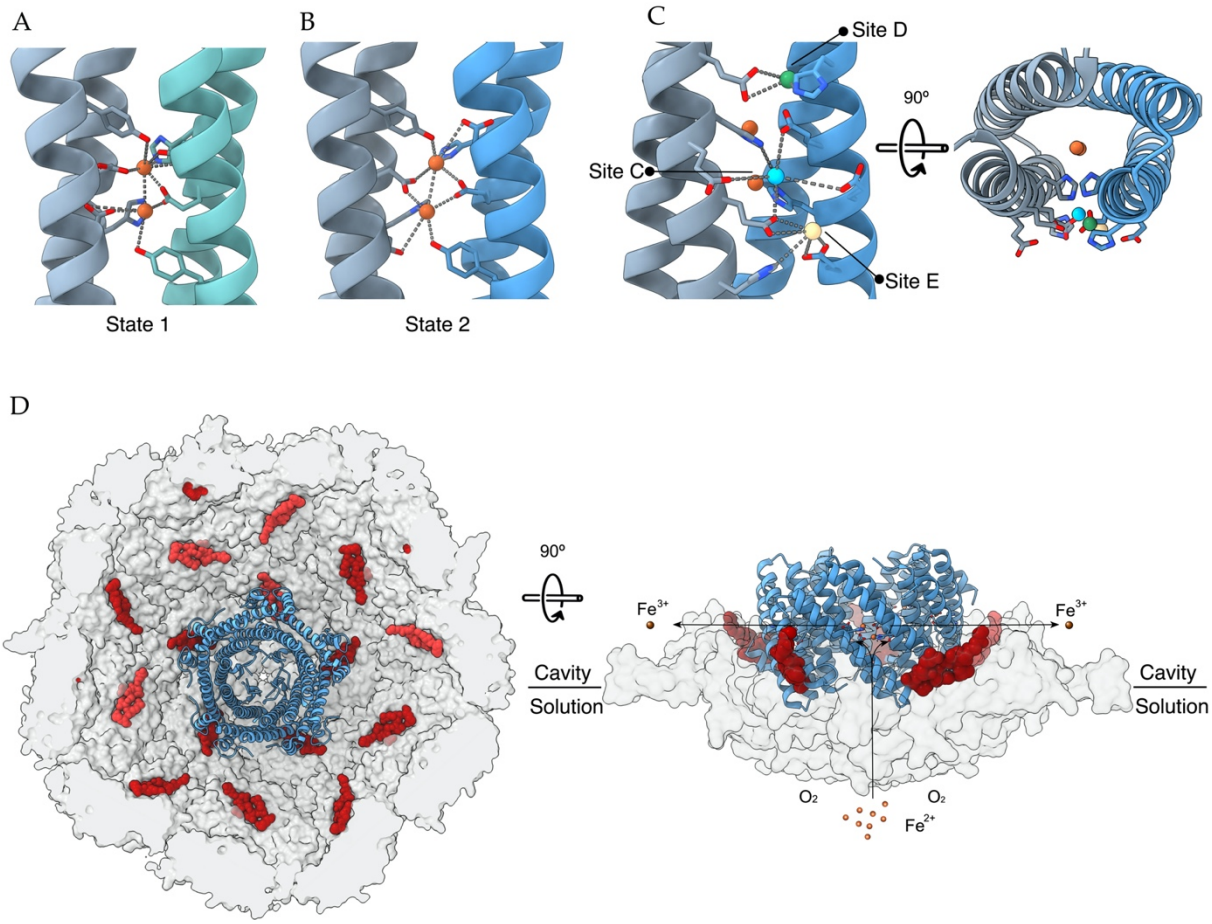

Figure 1.6: The ferroxidation center and predicted location of Ferritin-like proteins within the shell protein. (A) "State 1" di-iron binding FOC of *Myx. xanthus* EncC as previously detected for other Flp. **(B)** "State 2" di-iron binding FOC of *Myx. xanthus* EncB. **(C)** Additional iron binding sites formed by the glutamate-histidine ladder in EncB. The residues from each α-helix that form the FOC are displayed in gray and in green for EncC and gray and blue for EncB. Iron ions are shown as orange spheres. Interactions between the residues and iron are shown as dashed lines. The iron atom is shown as a blue sphere in site C, in green in site D and in yellow in site E (adapted from [30]). **(D)** Left: Truncated encapsulated Flp domain from *Pyr. furiosus* (PDB: 5N5E, in blue) docked to the *The. maritima* encapsulin shell, in light gray. The position of the Flp C-terminus agrees with the location of the CLP domains (in dark red) found in the structure of the *The. maritima* encapsulin shell. The Flp toroidal decameric structure (in blue) faces the 5-fold pore of the shell, suggesting a possible iron entry route. Right: Schematic illustration of the iron oxidation process in encapsulin-Flp systems. Fe^{2+} ions are channeled through the encapsulin pores reaching the di-iron FOCs at the Flp dimer interface. At the FOC the iron is coordinated by 2 glutamate and 1 histidine residues from each subunit and is oxidized using O_2 as co-substrate. The resulting ferric species are

subsequently translocated from the FOCs to nucleation/growth sites at the inner surface of the encapsulin cavity to form the ferric mineral. Adapted from [48].

Additionally, two extra iron binding sites were detected above and below the ladder (Figure 1.6C, sites D and E). All residues from site C and four out of five of the ladder residues are conserved in *Myx. xanthus* EncC, *Rhs. rubrum*, and *Hal. ochraceum*, while three out of five are conserved in *Pyr. furiosus*. Similar to some bacterial ferritins, this electronegative ladder is suggested to be the "gateway" for Fe^{2+} entry. A distinct additional iron binding site was detected in *Rhs. rubrum* [56] suggested by the authors as the iron entry site (E31 and E34). However, these residues are not conserved between Flp homologues (*Myx. xanthus* EncB and EncC and *Pyr. furiosus*) (Figure 1.5A). Therefore, different Flp may have distinct iron entry and translocation routes.

The ferroxidase activity of the *Rhs. rubrum* Flp cargo protein was measured using UV-Visible spectroscopy, following the absorbance at 315 nm in 50 mM Tris-HCl, pH 8.0 with 150 mM NaCl at room temperature. Fe^{2+} ions (in either a 2- or 10-fold molar excess relative to the FOC) were added to a truncated His-tagged recombinant form of Flp in the presence of atmospheric O_2 [20]. The iron oxidation rate was dependent of the amount of Fe^{2+} added, being faster for higher concentrations. Precipitation of ferric oxide aggregates was detected above 10 – 15 Fe/Flp decamer, demonstrating the inability of this protein to mineralize and store the ferric mineral. Site-directed mutagenesis of the iron ligands at the FOC (E32A, E62A, and H65A) reduced or abolished (in the case of E62A mutant) the rate of iron oxidation. A similar protocol was used to assess the ferroxidase activity of the *Hal. ochraceum*, and *Pyr. furiosus* Flp. An excess of 10 Fe^{2+}/FOC was added in the presence of O_2 in 10 mM HEPES, pH 8.0 containing 150 mM NaCl, at room temperature [48]. Although not identical, the progress curves obtained showed that all proteins were active and able to oxidize Fe^{2+} ions. Notably, the iron content of these proteins after purification (when produced fused to a StrepII-tag) was consistent with 40 – 80 % occupancy of the FOCs, which could explain their different kinetic behavior. Inhibition by Zn^{2+} was assayed since the protein also contained considerable amounts of this metal. Akin to classical maxi-ferritins, Zn^{2+} inhibited the ferroxidation reaction.

1.2.6.3 Arrangement of the Flp within the Encapsulin Shell

As previously mentioned, the toroidal structure of the encapsulated Flp is not compatible with its predicted iron management function, and thus the ferric mineral core has to be formed inside the encapsulin shell [15,48,49]. The inability of Flp to store the ferric mineral adds another level of complexity and requires a high level of functional coordination with the encapsulin shell. For mineralization, Fe^{2+} ions must be translocated from the solvent through the encapsulin shell, redirected to the FOCs in the Flp and finally released into the shell cavity for nucleation and mineral growth (Figure 1.6D).

The Flp density detected inside the *The. maritima* encapsulin was recently solved confirming that the position of the Flp decamer with 5-fold symmetry is practically aligned

with the 5-fold axis interface of the shell protein with a cargo protein loading capacity varying between 3 to 5 Flp per encapsulin [36]. However, in *Hal. ochraceum*, 4 Flp decamers were incorporated inside the encapsulin shell with a symmetry mismatch. They were spaced inside the encapsulin as the vertices of a tetrahedron, with two Flp decamers in line with the 5-fold pore and the remaining two decamers located between the 3- and 5-fold pores [38]. In *Pyr. furiosus*, the Flp is fused to the N- terminal of the encapsulin shell, in a stoichiometry of 1:1 cargo to encapsulin protomer [16]. Despite the conservation of the structural and biochemical properties of the decameric Flp proteins, one can question if the loading capacity predicted for the 60-mer encapsulin (solely composed of pentamers) would be the same as the 180-mer ones (formed by both pentamers and hexamers). How the Flp rearrangement into decamers occurs inside the encapsulin is not yet understood, because it would require enough linker flexibility to allow the interaction of Flp domains between hexamers and adjacent pentamers. The employment of icosahedral symmetry in the density reconstruction of EncB and EncC loaded into the encapsulin shell resulted in the loss of internal density indicating that the position of the cargo proteins did not follow the same symmetry as the shell. Notwithstanding, the EncC density was recovered near the 5-fold-axis, similar to the Flp of the *The. maritima* 60-mer encapsulin [30].

Overall, the Encapsulin-Flp complexes accommodate up to ~ 4,100 iron atoms in the 60-mer encapsulin of *Rhs. rubrum*, similar to the capacity of ~ 2,000 – 4,000 iron atoms in classical ferritins, while *Myx. xanthus* encapsulin is able to store ~ 30,000 iron ions in its 180-mer shell, an order of magnitude higher than ferritins [20,21,57,58].

1.2.7 Biotechnological Applications

Over the years, encapsulins have been extensively characterized providing highly engineerable and modular protein templates. These studies created new opportunities to assign natural or artificial nanocage proteins with novel biotechnological applications [59].

Protein nanocages have a great potential as catalytic nanoreactors allowing the production of valuable pharmaceutical compounds by promoting the activity of non-native cargo proteins [60] or the controlled production of antimicrobial peptides [61,62]. Also, the confinement of larger products such as the melanin polymer within the encapsulin shell demonstrates the attractiveness of this system as a nanocontainer for a reporter in multispectral optoacoustic tomography [63]. Similarly, the natural capacity to form large mineral core in the presence of Flp cargo proteins reveals its potential use as a negative contrast agent for magnetic resonance imaging [64].

The encapsulin system has also been used for medical applications. The encapsulation of a fluorescent flavoprotein mini-Singlet Oxygen Generator led to the controlled production of ROS resulting in a phototoxic effect on lung cancer cells cultured *in vitro* [65]. Additionally, the external surface of the encapsulin shell can be functionalized and targeted to specific tissues [66] while its cavity is loaded with anticancer prodrugs, increasing both their solubility

and target specificity [67]. Moreover, encapsulin shells are highly resistant to protease activity enhancing its applicability for oral administration as a drug-delivery system. This property was further enhanced upon functionalization with an mEETI-II knottin mini-protein [68].

The use of encapsulins as a template for the design and development of new materials has also been explored. Gold nanoparticles with CLP-tags were successfully loaded inside encapsulins. This new material has shown high potential to be used in photothermal tumor therapy [69]. In another approach, silver precipitating AG4 peptides were fused to the N-terminus of an encapsulin shell orientating this peptide towards the cavity. This protein engineering worked as a template for the formation of homogeneous and size-controlled silver nanoparticles [70].

These examples illustrate the versatility and potential of the encapsulin system in different biotechnology fields as a novel type of nanomaterial.

1.2.8 Current Challenges and Perspectives

Encapsulin research is still relatively recent and Encapsulin Family 1 is the most extensively described so far. However, Family 2 is predicted to be the most widely distributed and to represent the largest share of Encapsulin proteins [18]. Yet only one member of Family 2 was characterized so far [23]. Moreover, besides bioinformatics studies there is no formal knowledge regarding Families 3 and 4. Even if one would focus solely on Family 1, there are many types of predicted cargo proteins that are still uncharacterized, such as the encapsulated Dps or double Flp fused proteins [18].

The cargo proteins better described so far are the Flp. Many structural studies have clarified the role of the residues conserved in the FOC, which revealed some structural resemblance to the FOCs found in ferritins [20,30,48]. However, the Flp ferroxidation mechanism remains unclear. The iron entry pathway towards the FOC, the ferroxidation intermediates, the ferric iron release and particularly the iron accumulation and mineralization processes still need clarification.

Furthermore, some encapsulins systems can simultaneously harbor different types of cargo proteins, such as the *Myc. tuberculosis* system, which encapsulates three distinct cargo proteins: DyP-BfrB-FolB [22]. Is there any evolutionary advantage for this multi-cargo composition? Likewise, in Firmicutes the encapsulation of IMEF-Fd is suggested to expand the iron storage capacity similarly to bacterioferritin-associated Fd [18]. Furthermore, the *Myx. xanthus* encapsulin encapsulates two types of Flp (EncB and EncC) [21]. What are the advantages? Is this system similar to the predicted double Flp fused cargo proteins?

Moreover, the intrinsic function of shell proteins is still unclear since some of their cargo proteins also naturally exist in their free form in solution, such as ferritins or DyP. Is there any advantage for the encapsulation of free functional proteins? Is there any additional function of the shell protein not currently known?

Although encapsulin systems were firstly described in 1994 their vigorous research is much more recent. In addition to fundamental research, many biotechnological applications have been described for this system. A better basic understanding of encapsulins might further promote new and outstanding applications.

1.2.9 *Myxococcus xanthus* organism

Although Bacteria and Archaea are unicellular organisms in some cases, they show incipient multicellularity with the formation of simple clusters or filaments. A more restrict types of microorganisms (such as can *Caulobacter*, *Pseudomonas* and myxobacteria) are capable to self-organize into a multicellular structure [71]. Myxobacteria have made the transition from a single cell to multicellular life with cooperative behaviors and multicellular development. Under depleted condition, they form fruiting bodies that are multicellular biofilms varying from simple mounds to three dimensional structures [72] showing the path to multicellularity. However, the multicellularity in Myxobacteria is transitory and not obligatory since in favorable environments the organism do not initiate the multicellular state [73] and can even be evolutionarily lost if not required [74].

Myx. xanthus life cycle is composed by two phases: the cooperative predation and multicellular development. In both cases the organism used two motility systems: individual and group motility. In the presence of nutrients, the cells move in coordination forming multicellular biofilms [75,76]. And upon starvation, the cells start to exchange extracellular chemical signals and physical contacts to form a millimeter-long fruiting body.

Myx. xanthus produces an encapsulin system composed of an encapsulin shell, EncA, an 180-mer 32 nm wide shell (GenBank ABF87797.1) and 3 cargo proteins, EncB and EncC (Figure 1.7) (both containing ferritin-like domains with a highly conserved iron binding motif) and a third cargo protein, EncD, with unknown function [21].

Kim *et al.* have shown that the deletion of the Δ*encA* (and Δ*encF*) gene in *Myx. xanthus* impaired the formation of the fruiting body as the cells became unable to produce DKxanthene and myxoviresin, which are essential for this cell adaptation. The ability of the cells to agglutinate was also impaired [77]. The authors correlated this Δ*encA* phenotype with the cellular inability to sense iron availability. However, the Δ*encC* and Δ*encD* mutants (deletion of the putative ferritin-like cargo proteins) maintained the proper transition to fruiting body formation. Additionally, McHugh *et al.* revealed that under starvation conditions the encapsulin nanocompartments were overexpressed and their assembly augmented. The exposure of a *Myx. xanthus* encapsulin deletion mutant to H_2O_2 resulted in higher levels of cell death, suggesting a detoxifying role against oxidative stress by the encapsulin system, likely through the removal of free Fe^{2+} ions from the intracellular medium [21]. As such, the *Myx. xanthus* encapsulin system is suggested to function as a major iron storage container with a capacity to accommodate up to ~ 30,000 irons.

Although EncB (GenBank ABF88760.1) and EncC (GenBank ABF92698.1) proteins share only 30 % sequence identity, their C-terminal CLP is highly conserved with the

consensus sequence LTVGSLRR [30]. The most significant sequence disparity between EncB and EncC is due to a longer C-terminal tail in EncB, containing 36 additional residues. Even though the C-terminal regions of Flp proteins such as EncB and EncC are predicted to be unordered, this region is abnormally lengthy in EncB when compared with its homologues from other organisms (Figure 1.7).

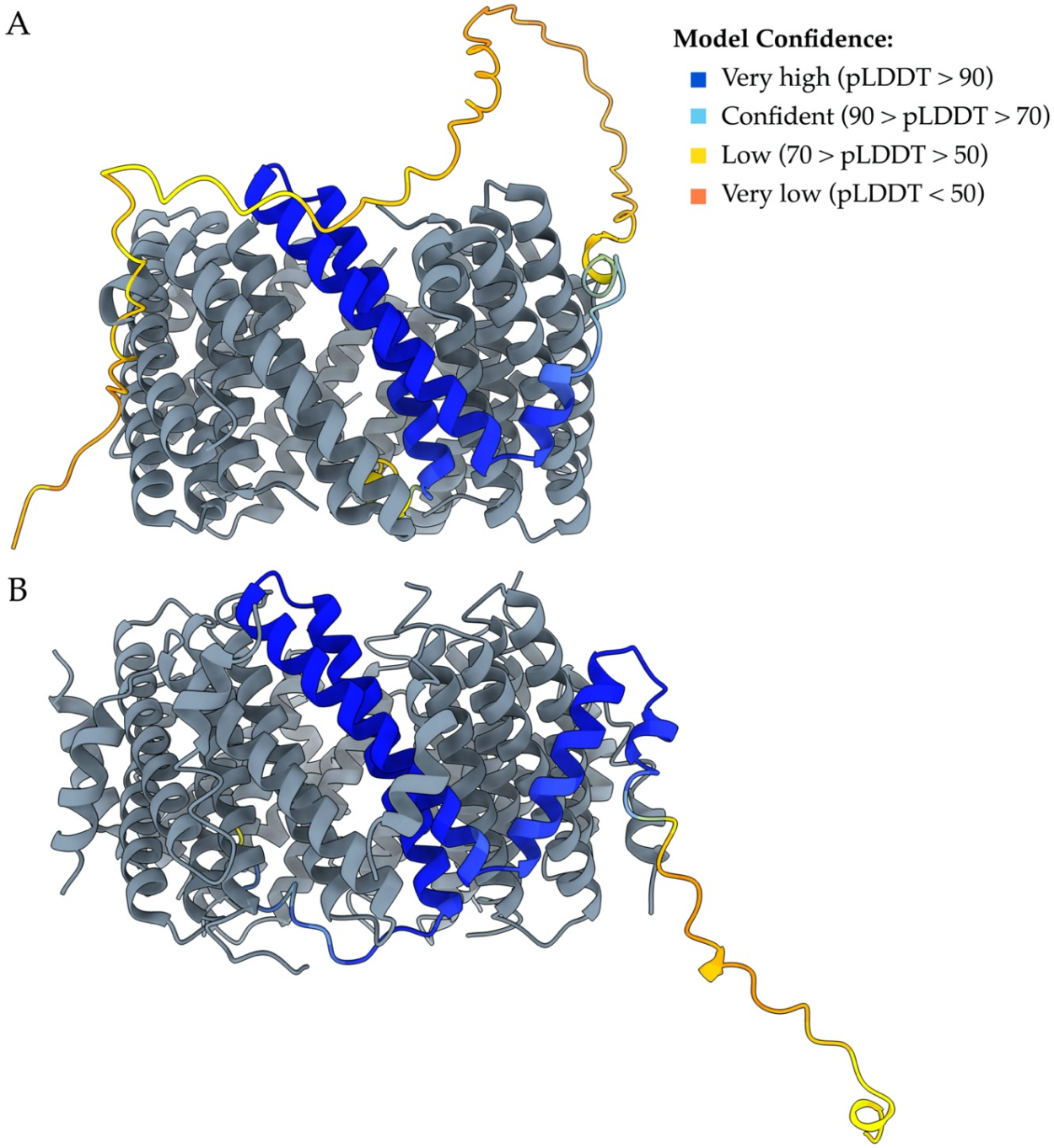

Figure 1.7: Structure prediction of EncB and EncC by AlphaFold. (A) Cartoon representation of the atomic structure of residues 15 to 76 EncB decamer (PDB: 7S5K, in gray) and structure prediction of the complete monomer sequence (residues 1 – 164) in a color gradient cartoon. **(B)** Cartoon representation of the atomic structure of residues 12 to 93 of the EncC decamer (PDB: 7S8T, in gray) and structure prediction of the complete monomer sequence (residues 1 – 130) in a color gradient cartoon. The gradient represents the per-residue confidence metrics (pLDDT) in the prediction from blue (very high) to orange (very low) [78,79].

1.3 Dps

The Ferritin family of proteins belongs to the Ferritin-like superfamily, which comprises different proteins that share a four-helix bundle structural ancestor [80].

The members of the Ferritin family are generally subdivided according to their size: The maxi-ferritins (from the assembly of 24 subunits, ~ 500 kDa in total) and the mini-ferritins (with 12 monomers, ~ 250 kDa in total). This family of proteins is widespread along the three kingdoms of life and are the main cellular mechanism responsible for maintaining iron homeostasis within the cell [81,82].

Maxi-ferritins are further subdivided into canonical maxiferritins (Ft), bacterial ferritins (Ftn) and bacterioferritins (Bfr). In eukaryotes only Ft are present and are composed of the assembly of different ratios of the H-chain (heavy) and L-chain (light) depending on the organism and tissue [83]. The H-chain has a conserved FOC within the four-helical bundle responsible for iron catalysis while in the L-chain a structure stabilized by salt bridges replaces the FOC. The L-subunits are associated with iron nucleation in the Ft cavity [84,85]. In bacteria, the Ftn are composed of homopolymers identical to the eukaryotic H-chain [86]. Bfr are similar to Ftn but have 12 additional heme groups, non-covalently bound to methionine residues in-between the dimers. These heme cofactors are suggested to be involved in ferric iron reduction and ferrous iron release from the cavity [87,88].

Mini-ferritins are currently solely composed of one subdivision: the DNA-binding protein from starved cells (Dps). This nomenclature is related with the discovery of the protein. In 1992, Dps was found overexpressed and tightly bound to the chromosomic DNA of *Esc. coli* cells in the starvation-induced late stationary phase, without an apparent sequence specificity. Dps proteins were found to be essential for the organisms ability to control oxidative stress [89]. In the absence of the *dps* gene the cells were highly susceptible to DNA damage upon acute hydrogen peroxide stress, revealing its importance for DNA protection. Different nomenclatures have been used for the term "mini-ferritins" depending on background of their discovery. Therefore, in addition to Dps, these proteins can also be termed Dpr (peroxide resistance), NAP (neutrophil-activating protein), or MrgA (metalloregulated genes) [90–92].

1.3.1 Structural Characterization

The first atomic structure of a Dps was published in 1998 and revealed a protein with a cubic structure with rounded corners, approximately 9 nm wide [93]. As the other members from the Ferritin family, the Dps monomer has a four-helix bundle motif where two pairs of helices connected by a small loop (A and B, C and D) that interact with a smaller helix in between (BC helix) through a longer loop (BC loop) (Figure 1.8A). The difference in the quaternary structure of the maxi- and mini-ferritins is due to the absence of a C-terminal helix (in mini-ferritins) responsible for the 24-mer oligomerization (Figure 1.8). The Dps FOC (with

two monomeric centers) lies in the interface between opposing subunits that form an anti-symmetric dimer. This interface is stabilized by hydrophobic interactions. The FOC is formed by residues (usually aspartate, glutamate and histidine) from both sides of the interface, resulting in a total of 12 FOCs per protein. The Dps FOC spatial organization is also distinct from their maxi-ferritin counterparts since the latter are located in the center of each monomer as a di-nuclear center (Figure 1.8) [58,86,94–96].

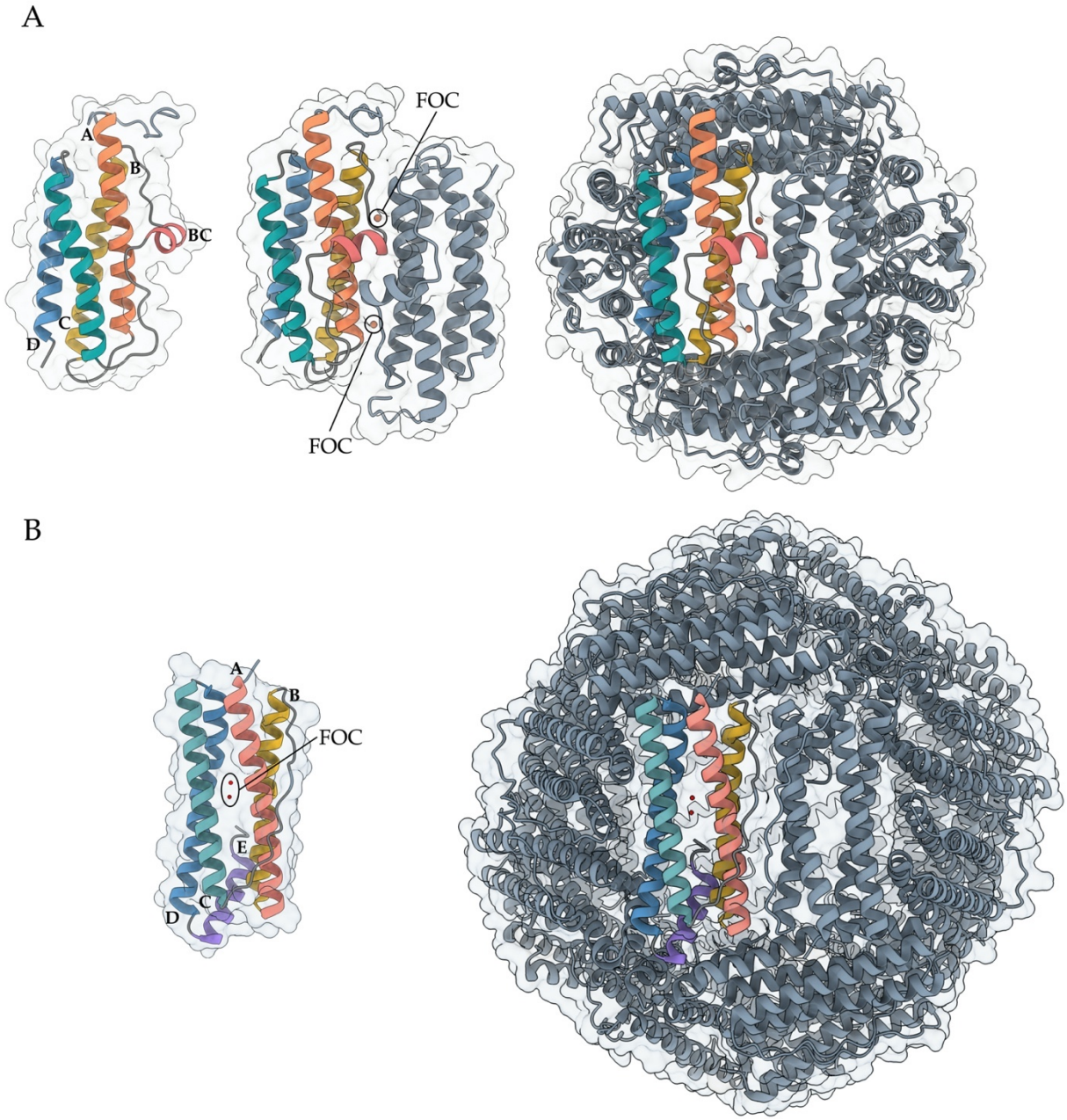

Figure 1.8: Structure of a mini- and maxi-ferritin. (A) *Esc. coli* Mini-ferritin (PDB: 1DPS). Left: Cartoon representation of the monomer (helix A in salmon, B in yellow, C in green, D in blue and BC in pink). Middle: The structural and functional unit (dimer) of the protein, with the ferroxidation centers (FOCs) identified with iron atoms shown as orange spheres. Right: Fully assembled Dps dodecamer. **(B)** *Esc. coli* Maxi-ferritin (PDB: 4XGS). Left: Cartoon representation of the monomer (helix A in salmon, B in yellow, C in green, D in blue and E in purple).

The ferroxidation center is identified with iron atoms shown as orange spheres. Right: Fully assembled 24-mer maxi-ferritin.

The dodecameric structure of Dps results from the self-assembly of six dimers into a 23-symmetry shell protein with a neutral-to-negatively charged surface and more accentuated negative pores formed by the interface between adjacent N-terminal tails (Figure 1.9A). Dps assembles into a hollow cage with an internal diameter of approximately 4.0 – 4.5 nm. Its internal surface is mainly negatively charged due to the presence of carboxylate groups of the amino acid side chains, ensuring a suitable environment for ferric iron nucleation and mineralization [97]. The negatively charged hydrophilic cavity can accommodate up to 500 iron atoms as a ferric mineral core.

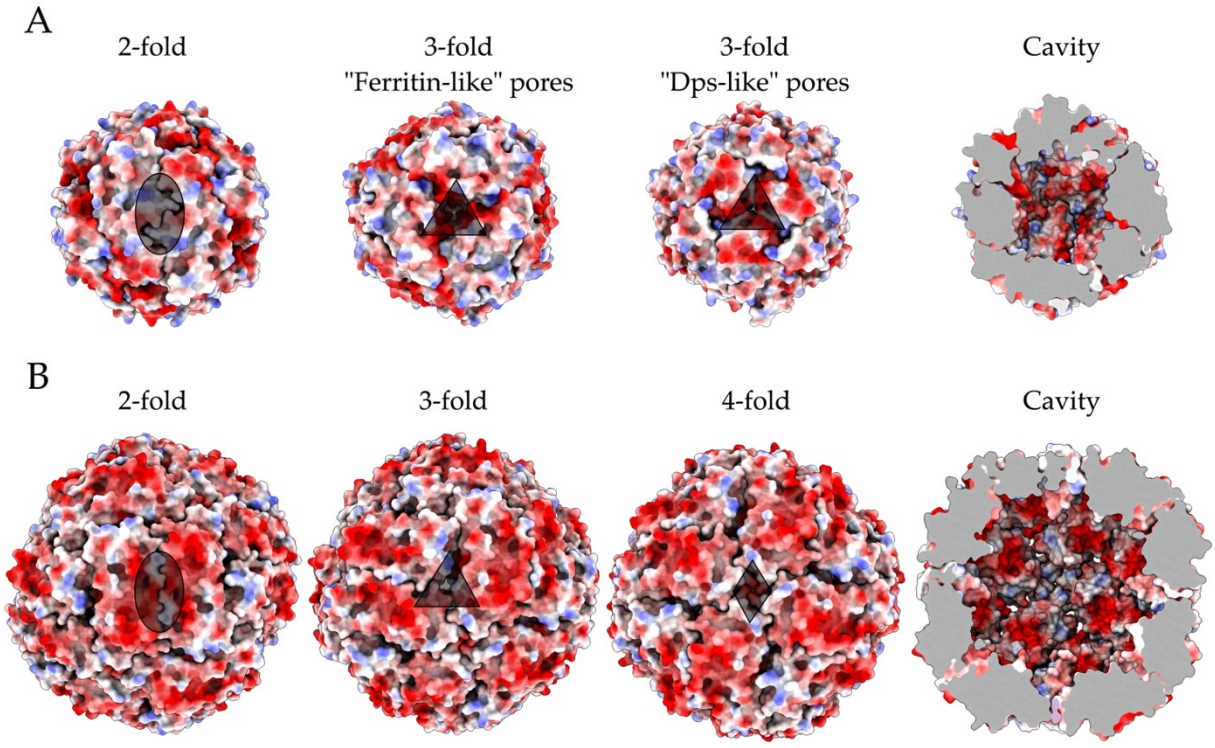

Figure 1.9: Representation of the external and internal electrostatic surface potentials of a mini- and a maxi-ferritin. (A) *Esc. coli* mini-ferritin (PDB: 1DPS), highlighting the 2-fold (dimer), 3-fold (Ferritin-like pores) and 3-fold (Dps-like) interfaces and the cavity. **(B)** *Esc. coli* maxi-ferritin (PDB: 4XGS), depicting the 2-fold, 3-fold and 4-fold interfaces and the cavity.

The translocation of ferric iron species towards the cavity occurs through pores and channels [98]. The interaction between the N-terminal regions of three monomers forms the 3-fold named "ferritin-like" pores (due to their similarity with the Ftn and Bfr 3-fold pores (Figure 1.9B) that are proposed to be responsible for the entry and exit of iron ions [99]. The C-terminal region of the monomers forms a unique type of 3-fold pores named "Dps-like" pores, whose function is still unclear [100]. The difference between this type of pore and the 4-fold pores found in the Ftn and Bfr is due to the absence of a fifth helix (helix E, Figure 1.8B).

The chemical environment of the two types of pores found in Dps are distinct, since the "ferritin-like" pores are highly hydrophilic (with acidic residues) and funnel-shaped, generating a charge gradient that facilitates the entry of Fe^{2+} ions [54]. The "Dps-like" pores are generally less hydrophilic and less conserved between Dps homologues.

Contrary to what was observed in the atomic structures solved using X-ray crystallography, in solution the N-terminal tails are not compressed against the protein. Using Small Angle X-ray Scattering these were found extended towards the solvent and responsive to changes in the ionic strength of the buffer [101,102].

1.3.2 Protein Activity

Dps has two major functions: one is directly correlated with its discovery, since the proteins were found interacting with DNA. The second is related with the general function of the proteins from the Ferritin Family, its capacity to perform ferroxidation, nucleation and mineralization of iron inside its cavity. Additionally, given the importance of iron in biological systems, labile iron must return to the cytosol, depending on the cellular needs. As such, an additional function of the protein is believed to be to guarantee the controlled release of labile iron and balance its bioavailability with its toxicity.

1.3.2.1 DNA binding

The external surface of the Dps is predominantly negatively charged, which would theoretically hinder its ability to bind to the negatively charged DNA backbone (Figure 1.9A). However, this interaction is established through electrostatic interactions between specific positively charged residues in the N-terminal extensions of the protein (mainly lysines and histidines) with phosphate groups in DNA. The electrostatic nature of this interaction makes the concentration of ions (Na^+, Cl^- and Mg^{2+}) an important variable capable of disturbing the binding [103,104]. In some Dps the extended C-terminal region is also accountable for the interaction with DNA.

The *in vitro* characterization of the DNA-binding properties using Electrophoretic Mobility Assays (EMSA) and Atomic Force Microscopy (AFM) showed that Dps preferentially binds to the ends of linear DNA. With genomic DNA, Dps interacted with sites that are known to be specific for nucleoid proteins. These results suggest that the interaction is not completely unspecific [105,106].

The interaction between Dps and plasmid DNA formed large complexes of protein-DNA condensates in physiologic conditions (Figure 1.10). The condensation activity seems to be related with the number of lysine residues in the N-terminal regions and higher number of lysines and their protonation increases the affinity to DNA [107]. Contrary to what was initially believed, a recent work revealed that in the Dps–DNA complex, DNA is not wrapped around the Dps molecule, contradicting the "histone-like" DNA binding model [108].

The length flexibility of the N-terminal tails are important for the interaction. The deletion of the first (8 or 18) residues of the N-terminal (eliminating 2 or 3 lysine residues, respectively) negatively affected the formation of DNA condensates [107]. Also, the N-terminal region of *Lactococcus lactis* DpsA and DpsB have a α-helix instead of the typical extended conformation which affects their DNA-binding properties [109]. In *Marinobacter (Mar.) hydrocarbonoclasticus* the replacement of a neutral residue by a negative one (lowering the pI of the tail from 8.2 to 5.9) impaired the interaction with DNA, revealing that a correct protonation of the tail is essential for the interaction [110].

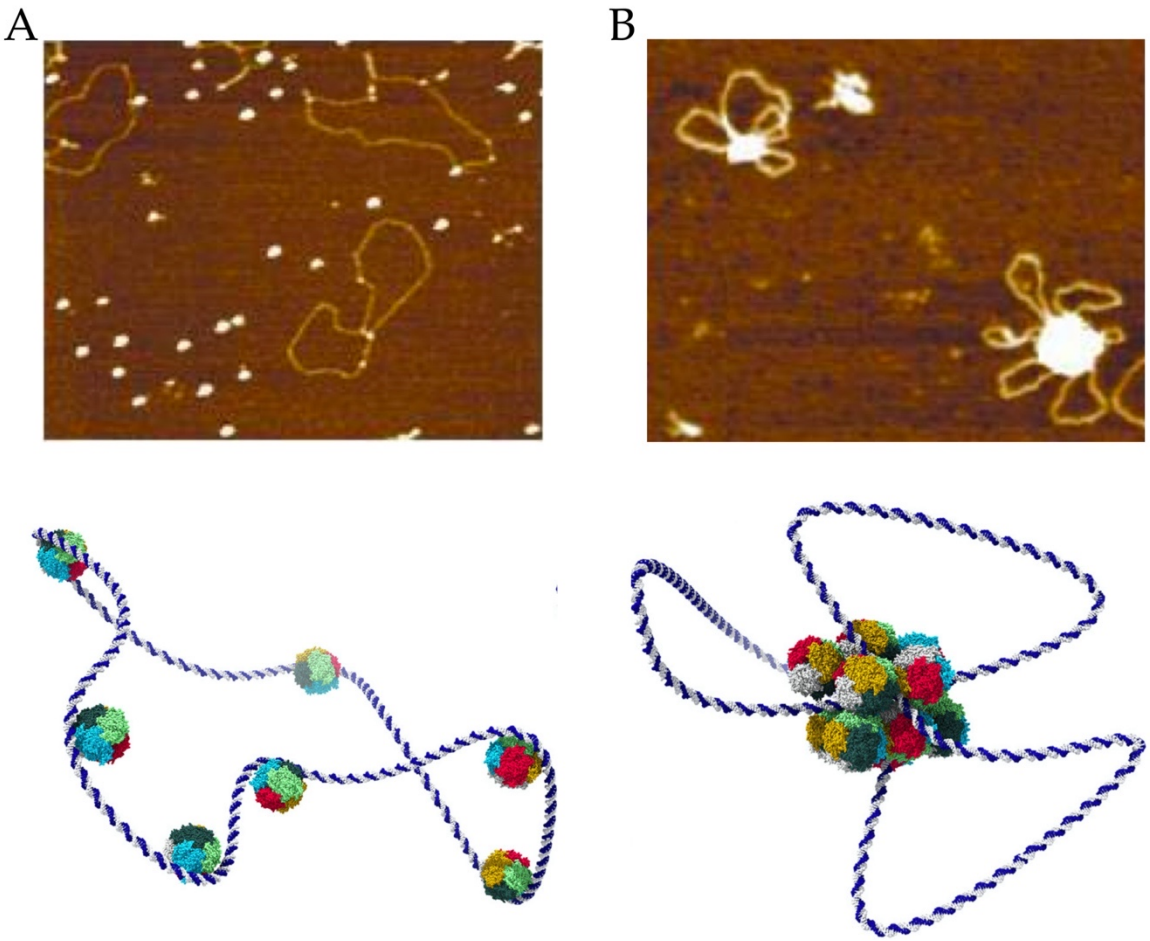

Figure 1.10: Representation of Dps–DNA interactions. (A) Individual mini-ferritin molecules bind to a single plasmid DNA molecule converting it from its supercoiled state to a more relaxed conformation. **(B)** Due to protein self-aggregation the DNA is further condensed into high-order Dps–DNA structures. Adapted from [80,111].

In the *Helicobacter (Hel.) pylori* NAP no extended N- or C-terminal exist and the DNA interaction occurs through its unique positively charged surface [91]. Additionally, two types of morphology were detected upon interaction with DNA. In physiologic conditions, the NAP surface is positively charged favoring DNA condensation. However, at higher pH (pH 8.0), the interaction still occurs but no longer forms a condensate. Instead, each protein binds to

DNA separately, converting it from its supercoiled conformation to a more relaxed "beads-on-a-string" morphology (Figure 1.10A) [112].

Not all Dps have the ability to interact with DNA, such as the Dps from *Listeria (Lis.) innocua*, *Bacillus anthracis*, *Agrobacterium tumefaciens* or *Streptococci*. However, their capacity to indirectly protect DNA is still granted by their ferroxidase activity, limiting the formation of ROS [113–116].

1.3.2.2 Ferroxidation

The main catalytic function of Dps proteins is their ferroxidase activity. For the iron to reach the ferroxidation center it first passes through the negatively charged 3-fold "ferritin-like" pores [80]. The funnel-shape environment is created by the presence of three conserved aspartate and/or glutamate residues that favors the entry of ferrous irons. The pore then constricts from ~ 10 to ~ 3 Å wide reaching the FOC, starting with hexa-aquo coordination that gradually turns into an interaction with carboxylates from the protein (Type I channel) [99]. In *Microbacterium arborescens* two iron atoms were identified in the channel in a hexa-aquo coordination sphere coordinated by aspartate residues.

As mentioned, the location of the FOCs in Dps is one of the most distinct features from this protein sub-family, located in the dimer interface. At the FOCs, Fe^{2+} ions are oxidized to Fe^{3+} using either hydrogen peroxide or molecular oxygen as co-substrate, the former being the preferential oxidant (named fast ferroxidation reaction, in the millisecond range), 100-fold faster than the ferroxidation with oxygen (slow ferroxidation reaction, in the seconds-to-minutes range) [117,118]. In addition to removing (and storing) iron atoms from the solution, the consumption of hydrogen peroxide as co-substate prevents the formation of hydroxyl radicals, limiting ROS production and protecting the integrity of the cell (see section 1.2.5).

The ferroxidation center is highly conserved between mini-ferritins. The FOCs are usually mono-nuclear and coordinated by one aspartate and one glutamate residues from one subunit and one histidine from the opposing monomer, with indirect binding by one histidine and one aspartate residues through water contacts (Figure 1.11A, left) [97,119]. A di-nuclear site was also detected in some Dps, with an additional iron ion coordinated by one glutamate and one aspartate residues of the same subunit. The function of this site is still unclear, and its occupancy is much lower than the mono-nuclear FOC (Figure 1.11A, middle) [98]. Furthermore, a distinct FOC was identified in cyanobacteria composed of three histidine and one glutamate residues ("His-type" FOC) (Figure 1.11A, right) [120–122].

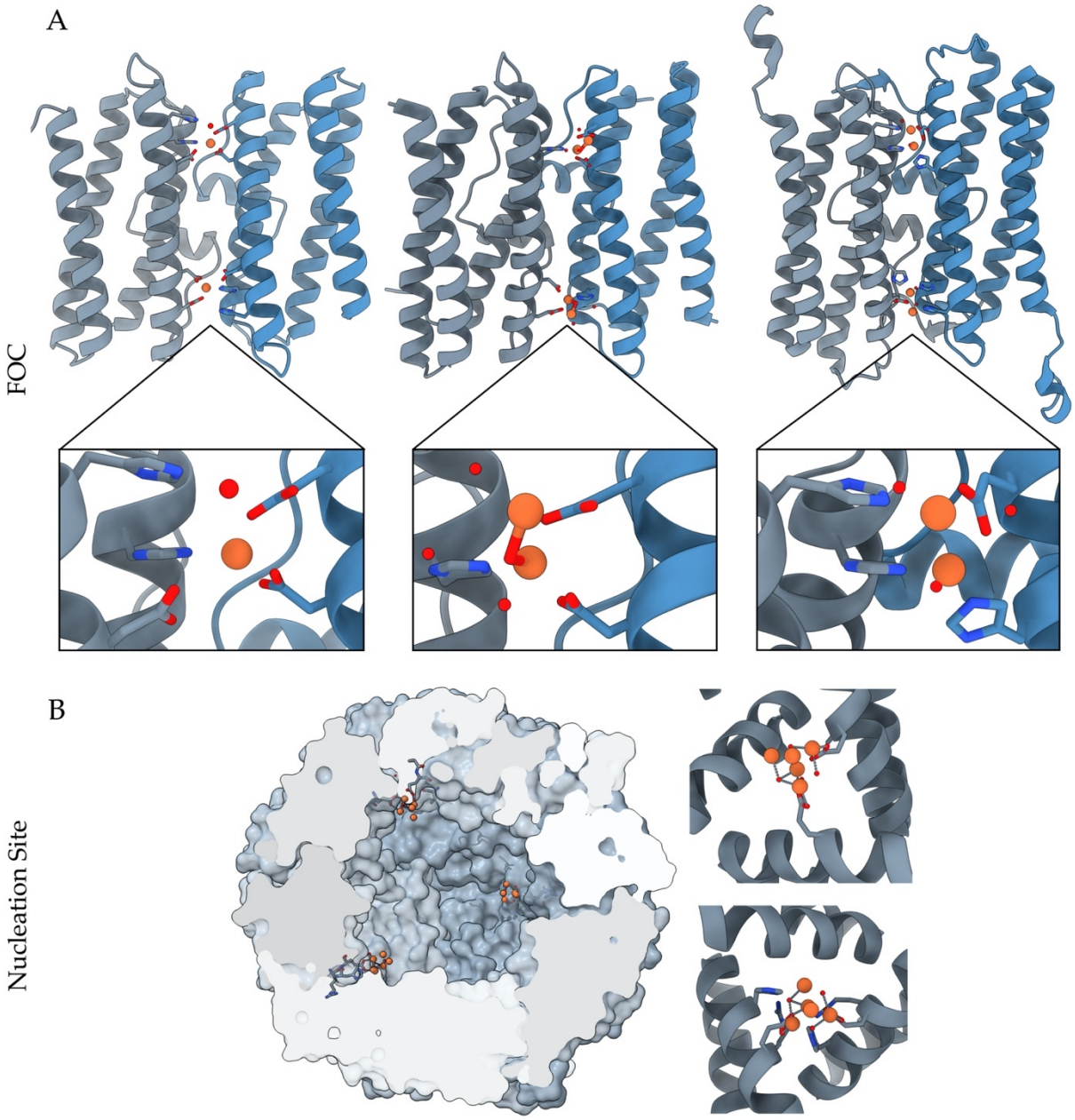

Figure 1.11: Structural representation of the types of ferroxidase and nucleation centers in mini-ferritins. (A) Types of FOC. Left: mono-nuclear FOC (with two Asp, one Glu and two His residues, PDB: 1QGH), middle: di-nuclear FOC (with one Asp, one Glu and one His residues, PDB: 1N1Q). Right: "His-type" FOC (with one Glu and three His residues, PDB: 1HJH). **(B)** The mineralization mechanism starts with two types of Nucleation center. Top: nucleation center I with Glu residues. Bottom: Nucleation center II closely stabilized by Glu and His residues (PDB: 1TKP).

After the ferroxidation reaction occurs, ferric ions leave the FOC and move towards the internal lining of the protein, rich in carboxylate residues, where the nucleation process occurs. Distinct nucleation sites have been detected using X-ray crystallography. The structure of the *Hal. salinarum* Dps revealed two nucleation sites, one in the 2-fold interface and another in the 3-fold interface placed ~ 13 and ~ 12 Å apart from the FOC, respectively) (Figure 1.11B).

In *Lis. innocua* Dps, two internal residues (glutamate and aspartate residues) near the 3-fold channel ("Dps-like" pore) adjacent to the FOCs were found to act as a nucleation center [97] (Figure 1.11B). The accumulation of ferric species in the nucleation sites leads to the formation of the ferrihydrite mineral core [123]. Cryo-EM data of an iron-loaded Dps showed a 10 – 15 Å wide mineral core near the "Dps-like" pore [124]. Upon formation of a small mineral core, alternative mineralization reactions occur even in the absence of oxygen or hydrogen peroxide as oxidants.

In anaerobic conditions and without hydrogen peroxide a previously existent small mineral core revealed the autocatalytic ability of further oxidizing and mineralizing ferrous ions [125].

1.3.2.3 Iron Release

The iron release mechanism from the proteins of Ferritin family is still poorly understood but it is believed to occur through distinct processes: 1) by degradation of the protein by lysosomal (eukaryotes) or 2) proteasomal degradation leading to an iron burst in solution and/or 3) by release promoted by reducing agents, the latter being the process most studied so far [126].

Strong reducing agents such as dithionite, thiols, dihydroflavin, or large molecules such as flavoproteins have been used for the reduction of the Classical Ft mineral core. Afterwards, ferrous iron is released and its presence in solution is usually monitored by complexation with chromophores (such as 2,2'-bipyridyl) which allows its spectroscopic detection [99,115,127–131]. However, is still unclear if the iron is released from the "Dps-like" or by the "Ferritin-like" pores [99,132–134]. Nevertheless, by monitoring the release mechanism from Ftn and Dps, a biphasic behavior was noticed. First, iron release is described with a faster kinetics behavior justified by the surface dissolution of the mineral core (following a first-order kinetics) followed by a slower process due to bulk mineral dissolution [115,123,135]. The complete iron release was achieved after 35 – 50 min in the presence of a 750- to 1250-fold molar excess of reducing agent and 1250- to 5800-fold molar excess of 2,2'-bipyridyl relative to a mini-ferritin sample harboring 240 – 280 iron ions.

1.3.2.3.1 WrbA

Flavodoxins are small acidic proteins (14 – 23 kDa) noncovalently bound to a flavin mononucleotide (FMN) co-factor functioning as an electron-shuttling protein (Figure 1.12A, left) [136]. The flavin reduction/oxidation in solution (pH 7.0) has a midpoint potential of ~ 200 mV. While incorporated in the protein the redox potential spans from ~ 400 to ~ 60 mV, spanning the range of the redox of the system [137].

These proteins are widely distributed in bacteria, functioning as an electron donor to several redox enzymes: ribonucleotide reductase systems, nitric oxide synthase, biotin synthase, pyruvate formate-lyase, and cobalamin-dependent methionine synthase, among others [138–142].

WrbA was first reported after being co-purified with a tryptophan repressor (TrpR) and therefore termed tryptophan repressor binding protein (WrbA) [143]. Further studies identified the presence of an FMN co-factor, grouping this protein as an novel member from the flavodoxin proteins (Figure 1.12A, right) [144]. However, its binding affinity to FMN is much lower compared with flavodoxin.

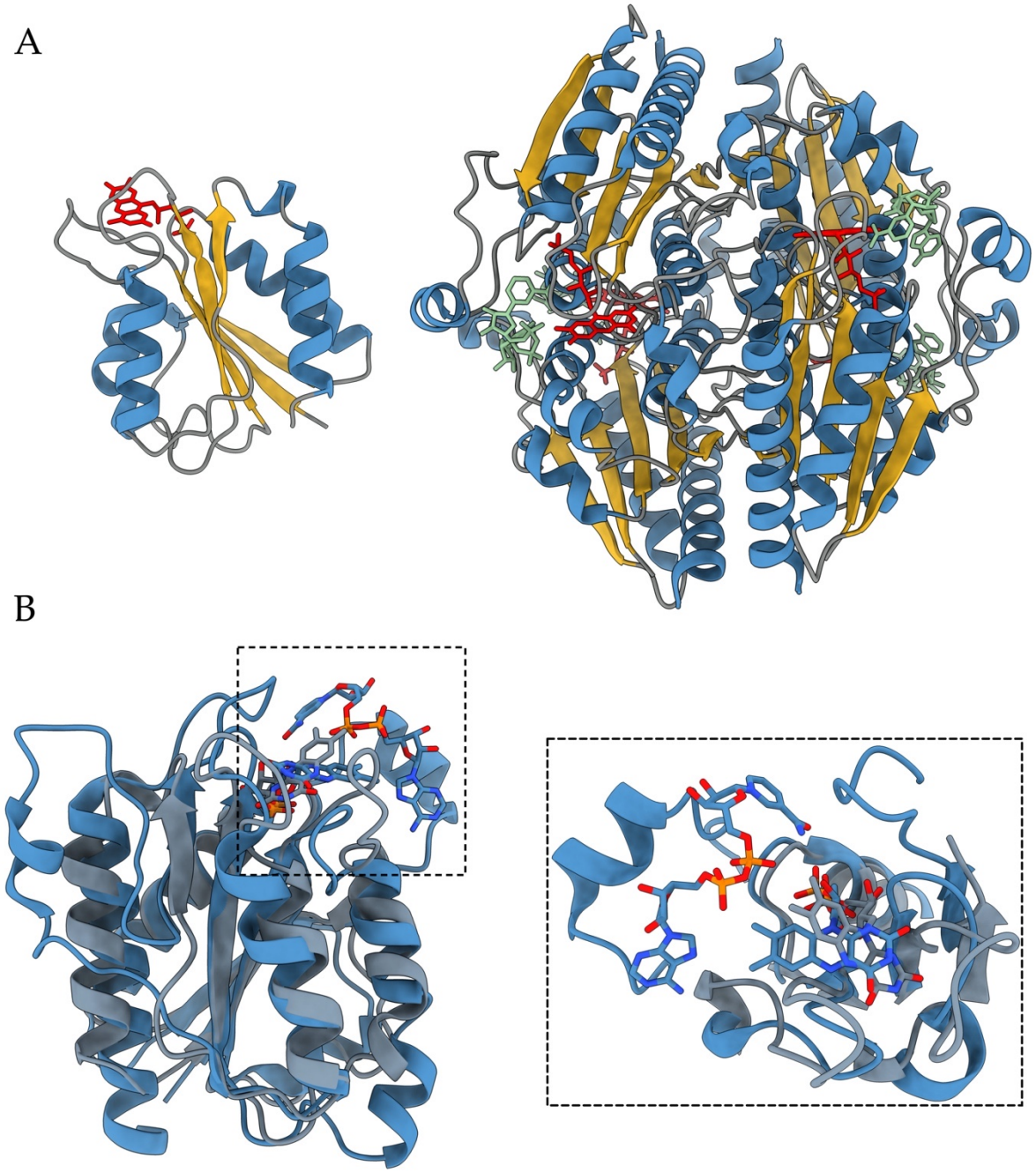

Figure 1.12: Structural comparison between flavodoxin and WrbA. (A) Left: Cartoon representation of monomeric flavodoxin with an FMN co-factor. Right: Tetrameric WrbA with FMN and NADH co-factors. α-helices are shown in blue, β-sheet in yellow and loops in gray. FMN and NADH are represented as sticks in red and green, respectively. **(B)** Left: Superposition of the flavodoxin monomer (in gray) and a WrbA monomer (in blue); Right: Close-up of the FMN and NADH binding sites. (Flavodoxin PDB: 1AKU and WrbA PDB: 3B6J).

Through structural homology and sequence analysis tools this protein was predicted to have the typical β-α-β flavodoxin with an FMN present at the end of a β-sheet, similar to flavodoxins (Figure 1.12B). However, this analysis predicted the existence of an additional conserved αβ-unit [145] uncharacteristic of classical flavodoxins and thus being further classified as a long chain flavodoxin [146]. Additionally, WrbA is found as a tetramer [147–150] while typical flavodoxins are monomers/dimers [144,151–153] (Figure 1.12A). As a tetrameric protein, WrbA has four active sites, each one with a flavin cofactor. While the exact function of WrbA is still not clear, sequence analysis predicted that WrbA proteins are also homologous to a NAD(P)H:quinone oxidoreductase (NQO) (Figure 1.12A, right). This prediction was than confirmed in *Esc. coli* and *Archaeoglobus fulgidus* and therefore WrbA was also classified as a new class of type IV NQO [147]. Unlike flavoproteins, WrbA transfers a pair of electrons with no detectable formation of a stable semiquinone, analogous to NQO [154].

1.3.2.3.1 Rubredoxin

Iron-sulfur center containing proteins are found in all Domains of life and play key roles in fundamental metabolic processes such as cellular respiration (either aerobic or anaerobic), nitrogen fixation, photosynthesis, nucleic acid modification, DNA replication and repair or gene expression, among others [155–157]. The role of iron-sulfur centers may vary, being part of inter/intra-molecular electron transfer chains or enzyme active sites. Additionally, iron-sulfur centers can alter the protein structure through redox state or center interconversion, and by doing so, regulate protein activity. From a structural point of view, iron-sulfur centers can be quite complex multinuclear centers including mixed-coordination environments, of which nitrogenase P-clusters and FeMoCo cofactors, novel tetranuclear centers, hydrogenases or carbon monoxide dehydrogenase are good examples [158–162]. However, iron-sulfur centers can also be very simple as is the case of Rubredoxin (Rd) type centers. Rd are small, monomeric proteins made of a single polypeptide chain of ca. 50 amino acid residues and a single iron atom that can vary between +2 and +3 oxidation states (Figure 1.13A). Contrary to all other iron-sulfur centers, the Rd-type center does not contain inorganic sulfide and can be more accurately described as an iron atom tetrahedrally coordinated by four cysteinyl sulfurs. Other metals can substitute iron, either *in vivo* or *in vitro*. For example, using heterologous expression in *Esc. coli*, it was shown that zinc can replace iron in *Desulfovibrio (Des.) vulgaris* desulfoferrodoxin and *Clostridium (Clo.) pasteurianum* Rd [163–165]. It was also possible to replace iron with zinc, indium, gallium, mercury, cobalt, cadmium, nickel, and copper using *in vitro* procedures [108,166–175].

Initially found in sulfate and/or nitrate reducing bacteria, Rd and Rd-type centers have also been found in other organisms [157,176]. In some cases, Rd-type centers are part of a larger protein, such in the case of rubrerythrin and desulfoferrodoxin (a class I superoxide reductase). Notably, to the best of our knowledge, Rd-type centers have only been implicated in electron transfer functions, either in electron transfer chains or in intramolecular electron transfer to the active site. The deposited structures are quite similar and show a structure

dominated by the tetrahedrally bound metal, containing α-helix, turns and triple-stranded anti-parallel β-sheet structural segments. Figure 1.13A shows the superposition of 3D-structures obtained for the different organisms.

The structure similarity is quite striking, even considering that the protein amino acid sequences only have six totally conserved amino acid residues. From the sequence alignment (Figure 1.13B) a motif $C(X)_2C(X)_3Y(X)_{18\text{-}25}CPXC$ is recognizable and seems essential for metal binding and folding. The metal-bound, holo-Rd structure has been quite well described. However, no structural features have been attributed to the apo, metal-free, form. This is in part due to the lack of stability, with the possibility of cysteinyl residue oxidation and protein aggregation.

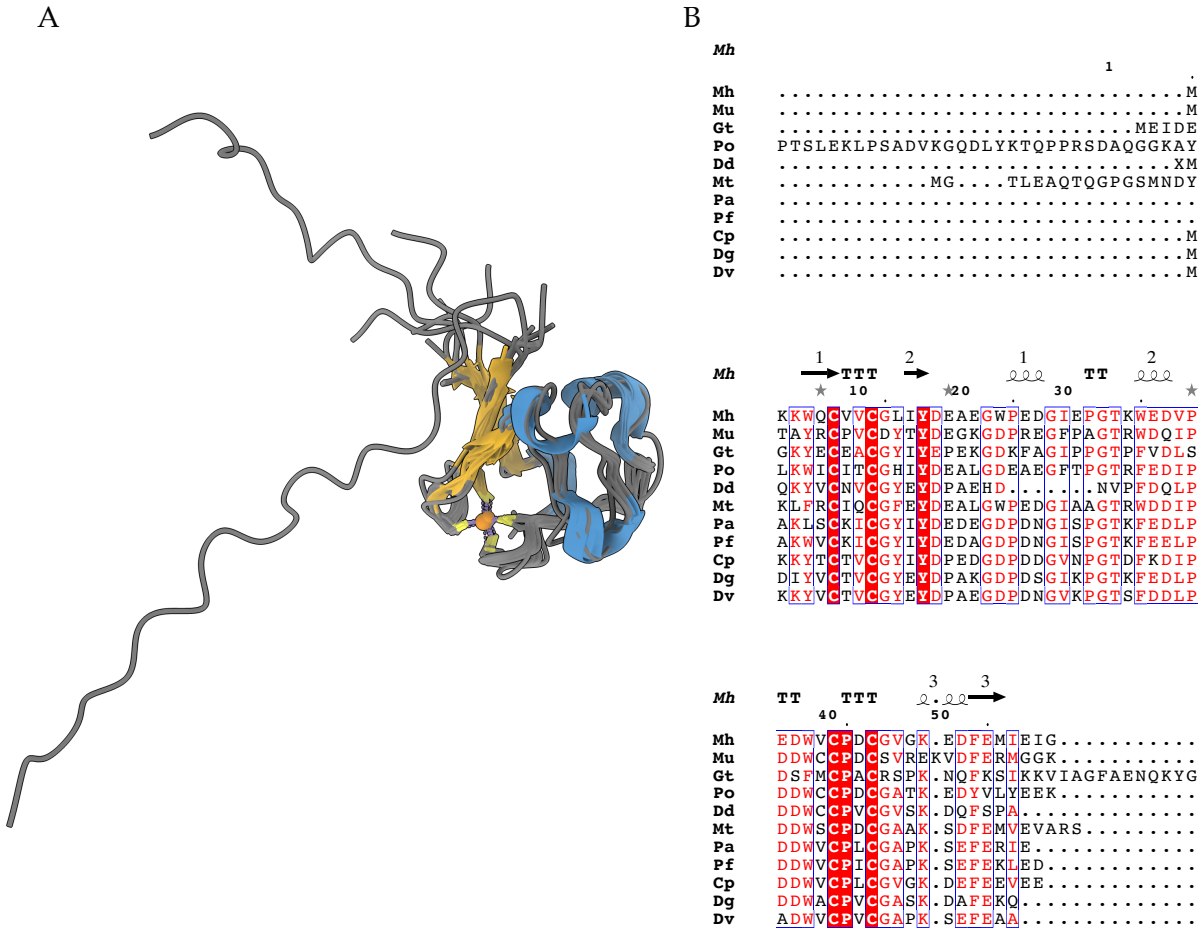

Figure 1.13: Structure and sequence comparison between rubredoxin proteins from different organisms. **(A)** Structural comparison between *Clo. pasteurianum* (4RXN), *Des. desulfuricans* (6RXN), *Des. gigas* (2DSX), *Des. vulgaris* (8RXN), *Pseudomonas (Pse.) oleovorans* (1S24), *Pyr. furiosus* (1CAA), *Pyr. abyssi* (2PYA), *Myc. tuberculosis* (2KN9), *Myc. ulcerans* Agy99 (2M4Y), and *Guillardia (Gui.) theta* (1DX8). α-helices are displayed in blue, β-sheets in yellow and coiled coils in dark gray, **(B)** Sequence alignment, prepared using Clustal Omega and ESPript [50,51], of *Mar. hydrocarbonoclasticus (Mh)* Rd with other known rubredoxins from *Myc. ulcerans* Agy99 (Mu), *Gui. theta (Gt)*, *Pse. oleovorans (Po)*, *Des. desulfuricans (Dd)*, *Myc. tuberculosis (Mt)*, *Pyr. abyssi (Pa)*, *Pyr. furiosus (Pf)*, *Clo. pasteurianum (Cp)*, *Des. gigas (Dg)*, and *Des. vulgaris (Dv)*. Secondary structure elements are displayed on top, according to PDB secondary structure assignments (α-helices are displayed as squiggles, β-strands as arrows, strict β-turns as TT

letters and strict α-turns as TTT). Strict identity is displayed as a red box with white characters, and a red character is used to display sequence identity.

1.3.3 *Marinobacter hydrocarbonoclasticus*

Marinobacter genus is a heterotrophic, and in some specific cases mixotrophic, [177,178] and it is are present in a wide variety of marine and saline terrestrial settings. Only a limited number of these organisms are strict aerobes while its majority are facultative anaerobes [179].

Mar. hydrocarbonoclasticus was isolated from hydrocarbon-polluted sediment, specifically from an oil refinery along the French Mediterranean cost [180]. It is most noticeably known for degrading different types of hydrocarbons (aliphatic and aromatic) as its carbon and energy source under aerobic conditions, showing this bacteria potential for bioremediation in environments contaminated with crude oil. Furthermore, under anaerobic conditions *Mar. hydrocarbonoclasticus* can consume citrate, acetate or succinate as carbon and energy source acting as a denitrifier under anaerobic environment [180].

2

OBJECTIVES AND OVERVIEW

2.1 Objectives

The research projects reported in this Thesis focused on two distinct but thematically connected themes: the structural and functional properties of the newly discovered Encapsulin family of proteins as well as the iron release mechanism from Dps.

Due to their relatively recent discovery, there are several questions still to be answered concerning Encapsulins. Specifically, one of the aims of this work was to clarify the quaternary structure of *Myxococcus xanthus* encapsulin cargo proteins both free in solution and upon encapsulation and to understand if the encapsulation mechanism grants any protection to the cargo proteins.

Additionally, up to this moment the function of the encapsulin system has been solely related with the type of cargo protein encapsulated. With this work we intended to evaluate potential intrinsic functions of the shell protein in the absence of any cargo protein, such as its hypothetical ability to interact and protect DNA molecules. Moreover, the previously characterized ferroxidase function of the Flp-Encapsulin system was further characterized through kinetic studies and spectroscopic analysis of iron mineral cores formed within EncA in the presence of molecular oxygen or hydrogen peroxide as co-substrates.

Although there is currently a relatively extensive knowledge regarding Dps proteins, especially regarding its iron mineralization and DNA-binding properties, the iron release mechanism from within Dps is still poorly understood. Therefore, Dps of *Marinobacter hydrocarbonoclasticus* was selected to study the iron release mechanism due to the facultative anaerobic nature of this organism. Is the iron release disturbed by the presence or absence of molecular oxygen? How is the release mechanism affected by the presence of an iron acceptor in solution? Does the chemical nature of the iron binding molecule change the behavior and kinetics of the release mechanism?

To answer all these questions different biochemical, structural and spectroscopic techniques such as Size Exclusion Chromatography, Electrophoresis Mobility Shift Assay, Small Angle X-ray Scattering, Dynamic Light Scattering, Atomic Force Microscopy,

Synchrotron-Radiation Circular Dichroism, UV-Visible, Electron Paramagnetic Resonance and Mössbauer spectroscopy were applied in a multidisciplinary, integrative manner.

2.2 Thesis Overview

Chapter 1 described the most recent scientific developments in the study of the bacterial mechanisms that mediate cellular compartmentalization, focusing on the Encapsulin Family of proteins harboring Ferritin-like cargo proteins and on Dps proteins, as a ubiquitous biological defense mechanism against oxidative stress by mediating the controlled storage of iron from solution, among other biochemical functions.

Briefly, Chapters 3, 4, 5 and 6 describe the efforts made to clarify the structural organization and function of the *Myxococcus xanthus* Encapsulin system, while Chapter 7 delves into the iron release mechanism from *Marinobacter hydrocarbonoclasticus* Dps.

Chapter 3 describes the development of production and purification protocols for recombinant expression of the Encapsulin system (EncA, EncAB, EncAC, EncB and EncC) as iron-free and soluble proteins. The characterization of the cargo encapsulation process and its loading efficiency is also shown.

Chapter 4 clarifies the structural changes in the shell protein envelope upon cargo protein encapsulation, the oligomeric state of the cargo proteins and the impact of iron addition on their quaternary structure. The protective role of the shell protein after encapsulation of the cargo proteins was also investigated upon temperature stress.

Chapter 5 unveils the iron oxidation and mineralization function of the shell protein and the effect of the encapsulated cargo proteins. The mineral core formed within EncA in the presence of either molecular oxygen or hydrogen peroxide was analyzed by Mössbauer spectroscopy, as well as the ferric species formed by free EncC during ferroxidation.

Chapter 6 focuses on a newly proposed function for the encapsulin as a DNA binding protein and its ability to protect DNA against enzymatic digestion using EMSAs and AFM.

Chapter 7 explores the iron release from a Dps in the presence of an electron transfer protein partner (WrbA protein) and two types of iron acceptor: an inorganic iron chelator (1,10-phenanthroline) and an iron-sulfur center containing protein partner with high affinity for ferrous iron (Rubredoxin).

The final concluding chapter integrates the results presented in this Thesis and discusses the main conclusions and the impact of this research work while also presenting future perspectives and potential follow-up experiments.

3

PRODUCTION AND PURIFICATION OF THE ENCAPSULIN SYSTEM

This chapter details the vector construction, protein expression tests and large-scale production and purification of the proteins from the *Myxococcus xanthus* encapsulin system. The protocols and results obtained using standard expression and purification techniques for the production of the free forms of EncA and its cargo proteins EncB and EncC, as well as the encapsulated EncAB and EncAC complexes are described in this section. Moreover, the encapsulation of EncB and EncC by EncA was confirmed using biochemical characterization techniques (size exclusion chromatography (SEC) and gel electrophoresis under denaturation conditions (SDS-PAGE)) and the encapsulation efficiency determined using densitometric analysis of SDS-PAGE gels.

3.1 Experimental Procedure

3.1.1 Production of the Proteins from the Encapsulin System

The genes encoding the encapsulin shell protein (EncA) (GenBank ABF87797.1) and its cargo proteins EncB (GenBank ABF88760.1) and EncC (GenBank ABF92698.1) from *Myx. xanthus* were optimized for recombinant expression in *Esc. coli*. The plasmids harboring the genes encoding each protein sequence were obtained by chemical synthesis (Invitrogen GeneArt Gene Synthesis, ThermoFischer Scientific). The *encA* gene was subcloned into a pET21-c expression vector (Novagen) using EcoRI and NdeI enzymes (Figure A.1A, appendix), while the *encC* and *encB* genes were inserted into pET28-c (Novagen) using NcoI, NdeI, through recombinant DNA technology (Figure A.1B and C, appendix). NZY5α competent cells (NZYTech) were transformed with the ligation reactions. Positive clones were isolated from a Luria-Bertani (LB)-agar plate containing 100 mg/L ampicillin for cells pET-21c-EncA transformants and with 50 mg/L kanamycin plate for EncB pET-28c-EncB and EncC pET-28c-EncC and grown in liquid LB medium supplemented with the appropriate antibiotic for plasmid DNA isolation.

The pET-21c-EncA, pET-28c-EncB and pET-28c-EncC expression vectors were used to transform *Esc. coli* BL21(DE3) competent cells (NZYTech) for expression of each protein and for co-expression of the EncAB and EncAC complexes.

Small-scale expression tests were performed by inoculating 50 mL of liquid media (supplemented its appropriate antibiotic) with a 500 μL inoculum grown overnight at 37 °C, in LB medium. The overexpression was optimized by testing parameters such as the concentration of Isopropyl β-D-1-thiogalactopyranoside (IPTG at 0.1, 0.5 or 1 mM) and the duration and temperature of induction (3 h at 37 °C or overnight at 22 °C). The expression levels were normalized based on the cellular density ($OD_{600\,nm}$) and evaluated through polyacrylamide SDS-PAGE gels.

For large-scale expression all proteins and complexed were obtained under the same conditions: the competent cells containing the plasmid of interest were grown in 0.5 L of LB medium (25 g/L, NZYTech) (on a 2 L flask) supplemented with the appropriate antibiotics at 37 °C, 220 rpm, up to an $OD_{600\,nm}$ of ~ 0.8 and induced with 0.5 mM IPTG overnight at 22 °C, with orbital shaking at 220 rpm. The cells were harvested by centrifugation at 11,000× *g* during 15 min at room temperature (Z 36 HK, HERLME LaborTechnik) and the pellet was resuspended in 10 mM Tris-HCl buffer pH 7.6 and lysed using an ultrasonic homogenizer (LabsonicM, Sartorius) in the presence of protease inhibitors (1 mM phenylmethylsulfonyl fluoride (PMSF) and 10 mM Benzamidine) and DNase I (Merck).

EncA, EncAB and EncAC were obtained as soluble proteins and purified following the same protocol. Protein fractionation and purity assessment was carried out using 12.5 % polyacrylamide SDS-PAGE gels.

3.1.2 Purification of the Proteins from the Encapsulin System

3.1.2.1 Purification of EncB

Since EncB was expressed in the form of inclusion bodies, the pellet obtained after centrifugation of the cell extract was incubated in 80 mL of 10 mM Tris-HCl pH 7.6 buffer, 250 mM NaCl, 1 % Triton-X100 and 1 M urea during 1 h at room temperature, with a gentle rocking. This was followed by a centrifugation at 11,000× *g* for 30 min (Z 36 HK, HERLME LaborTechnik). The pellet was resuspended in 80 mL of 10 mM Tris-HCl pH 7.6 buffer, 250 mM NaCl, 1 % Triton-X100 and 8 M urea during 1 h at room temperature to solubilize the inclusion bodies.

The protein solubilized with 1 M urea was dialyzed overnight against 4 L of 10 mM Tris-HCl buffer, pH 7.6, 1 M urea at 4 °C and purified to homogeneity through ion exchange chromatography (IEX) and gel filtration for polishing. All purification steps were performed at 4 °C using an ÄKTA prime Plus system (Cytiva). Solubilized EncB was loaded into a DEAE-Sepharose fast flow (FF) chromatography (XK 26/40 cm, Cytiva) column pre-equilibrated with 10 mM Tris-HCl buffer, pH 7.6 at a flow rate of 5 mL/min. After a washing step, a linear NaCl (0 – 500 mM) in the same buffer was applied to elute adsorbed proteins. Fractions containing recombinant EncB were pooled, concentrated in a Vivaspin 20 (MWCO 10 kDa, Sartorius) and applied into a Superdex 200 prep grade Size Exclusion Chromatography (SEC) (XK 16/60 cm, Cytiva) equilibrated with 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. The fractions containing EncB were pooled, concentrated and stored at −80 °C until further use.

3.1.2.2 Purification of EncC

The EncC soluble cell extract was further ultracentrifuged at 180,000× *g* for 1 h, at 4 °C (Beckman Coulter 70 Ti type rotor) to remove additional cellular components. The resulting supernatant was dialyzed overnight against 4 L of 10 mM Tris-HCl buffer, pH 7.6 and then purified to homogeneity through IEX and SEC polishing steps, similar to soluble EncB (Section 3.1.2.1).

3.1.2.3 Purification of EncA Shell, EncAB and EncAC Complexes

Cellular extracts of either EncA or EncAB and EncAC complexes were dialyzed overnight at 4 °C against 4 L of 10 mM Tris-HCl buffer, pH 7.6 and loaded into a cation exchange CM Sepharose FF (XK 26/10, Cytiva) column pre-equilibrated with 10 mM Tris-HCl buffer, pH 7.6 at a flow rate of 5 mL/min. All three proteins were collected in the flow-through fractions. After a second dialysis against 10 mM Tris-HCl buffer, pH 7.6, the flow-through fraction was then loaded (flow rate of 5 mL/min) into a DEAE-Sepharose FF column (XK 26/40, Cytiva) pre-equilibrated with 10 mM Tris-HCl buffer, pH 7.6 and once again eluted as the flow-through. The fractions containing the proteins of interest were then pooled, concentrated using a Vivacell (MWCO 100 kDa, Sartorius) and loaded into a HiPrep Sephacryl

S-500 HR (XK 16/60 cm, Cytiva) SEC column equilibrated with 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. The fractions containing either EncA, EncAB and EncAC were pooled, concentrated and stored at −80 °C until further use.

3.1.3 General procedures

3.1.3.1 Protein Purity Assessment

Protein purity was evaluated throughout the purification process using polyacrylamide gel electrophoresis and catalase contamination test.

3.1.3.1.1 Polyacrylamide Gel Electrophoresis

Protein electrophoresis in denaturating conditions, SDS-PAGE, were performed by preparing a stacking gel with 5 % acrylamide and a separating gel with 12.5 – 17 % acrylamide. Accordingly, a solution of 30:0.8 of acrylamide/bisacrylamide was mixed with the separating buffer solution (375 mM Tris-HCl, pH 8.8, 0.1 % SDS) or with stacking buffer solution (125 mM Tris-HCl, pH 6.8, 0.1 % SDS). Gel polymerization was achieved with 0.1 – 0.2 % of ammonium persulfate and TEMED.

The protein samples were mixed with sample buffer (62.5 mM Tris-HCl, pH 6.8, 2.5 % SDS, 10 % glycerol, 5 % β-mercaptoethanol and 0.1 % bromophenol blue) and boiled for 2 min. The samples were then loaded into a gel set in an electrophoresis system (MiniGel Tank, Invitrogen) that was filled with electrophoresis running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS pH 8.3). A commercial protein marker was used in every gel for estimation of the molecular weight (Low Molecular Weight (LMW) II or NZYColour I Protein Marker, NZYTech). The electrophoresis system was set at 35 mA (per gel) for 60 – 80 min (depending on the acrylamide content). The protein bands were stained with BlueSafe (NZYTech) and the background was distained with water. The gels were imaged using a white light transilluminator and a Gel Logic 100™ Imaging System (Kodak).

3.1.3.1.1 Catalase Contamination Test

The presence of catalase, a possible contaminant, was tested by diluting 10 μL of the protein fractions under evaluation with 100 μL of milli-Q H_2O and adding 10 μL of 30 % H_2O_2 (w/v). After 5 min, the fractions were pooled based on a comparative qualitative assessment of the formation of O_2 bubbles, a product of catalase contamination.

3.1.3.2 Protein concentration determination

Protein concentration was estimated by measuring the absorbance at 280 nm and using the molar extinction coefficient determined by analysis of the primary sequence of each protein (ExPASy ProtParam online tool) (Table 3.1) [181]. Additionally, the concentration of EncAB and EncAC was determined based on the loading efficiency of each complex. All

measurements were performed on a Thermo Scientific Evolution 210 or 300 UV-Visible spectrophotometer. Standard quartz cells with 1.0 cm pathlength were used. Due to its relatively low absorptivity the quantification of EncB was performed using a 0.5 cm quartz cuvette.

Table 3.1: Theoretical molar extinction coefficients and molecular weight predicted from the protein sequence from the *Myx. xanthus* encapsulin system.

	Molar Extinction Coefficients ($M^{-1} cm^{-1}$)		Molecular weight (kDa)
	280 nm	205 nm	
EncA	31,400	1120,640	32.5
EncB	4,470	586,900	17.5
EncC	8,480	406,100	12.8

Molar extinction coefficient and molecular weight of the monomer of each protein using Protparam and Anthis method [181,182].

3.1.3.3 Encapsulation Loading Efficiency

Serial dilutions of calibrated EncA, EncB and EncC samples were prepared between 3.5 and 1.5 μg. In parallel, a serial dilution of each EncAB and EncAC of unknown concentration was prepared. The samples were loaded into a 12.5 % polyacrylamide SDS-PAGE gel and stained with Blue Safe. The densitometric analysis of the bands was performed using Image J [183] and used to determine the loading efficiency.

3.1.3.4 Size-Exclusion Chromatography

Samples containing 3 mg of either EncA, EncAB, EncAC, EncB and EncC were injected into a SEC Sephacryl S-500 HR (XK 16/60 cm, Cytiva) column at a flow rate of 1.5 mL/min and the elution profile was monitoring by measuring the absorbance at 280 nm.

3.2 Results and Discussion

3.2.1 Production and Purification of Encapsulin System Proteins

Different overexpression conditions (IPTG concentration, temperature and time of induction) were tested using *Esc. coli* BL21(DE3) competent cells and the different expression vectors. Expression of pET21-EncA, pET28-EncB and pET28-EncC (Figure A.1, appendix) resulted in the production of EncA (32.5 kDa monomer), EncB (17.5 kDa) and EncC (12.8 kDa), respectively (Figure 3.1A, B and D). Their complexes (EncAB and EncAC) were produced by co-expressing each vector pair (Figure 3.1C and E). The production of the EncABC complex was attempted without success. With the exception of EncB, all proteins were expressed in the soluble form. As such, EncB was solubilized using a urea renaturation protocol.

3.2.1.1 EncB

EncB was successfully expressed in all conditions tested using the pET28-EncB vector and *Esc. coli* BL21(DE3) competent cells (Figure 3.1B). For large-scale production the overnight at 22 °C expression with 0.5 mM IPTG induction was the selected conditions. As previously mentioned, EncB was obtained in the insoluble fraction and as such a solubilization protocol was used (in 10 mM Tris-HCl pH 7.6 buffer, 250 mM NaCl, 1 % Triton-X100 and 1 M urea). Protein recovery at 1 M urea was deemed complete since a very low amount of protein was detected in the 8 M urea solubilization step performed afterwards (Figure 3.2B).

After overnight dialysis against 10 mM Tris-HCl pH 7.6 buffer, EncB was loaded into a DEAE FF anion exchange chromatography column and eluted between 340 and 415 mM NaCl (Figure 3.2A). The presence of two bands in the EncB fraction is noticeable in the corresponding SDS-PAGE electrophoresis gel (Figure 3.2B). One band exhibits a faster migration pattern than the 20 kDa marker, and the second band a slower migration than the 15 kDa marker. However, when loaded into a SEC Sephacryl S500 column the protein is eluted as a single peak (Figure 3.3). These migration abnormalities suggest protein hydrolysis even in the presence of protease inhibitors. The hydrolysis pattern observed may correspond to up to 40 residues. EncB has a predicted unordered flexible C-terminus region (Figure A.2A, appendix) when compared with its Flp homologues, with at least 60 additional residues, vulnerable to protease activity. Although no EncB proteolysis was reported in previous works it was also impossible to solve the atomic structure of this region [30]. This solubilization and purification protocol typically yielded 40 mg of pure protein per liter of culture.

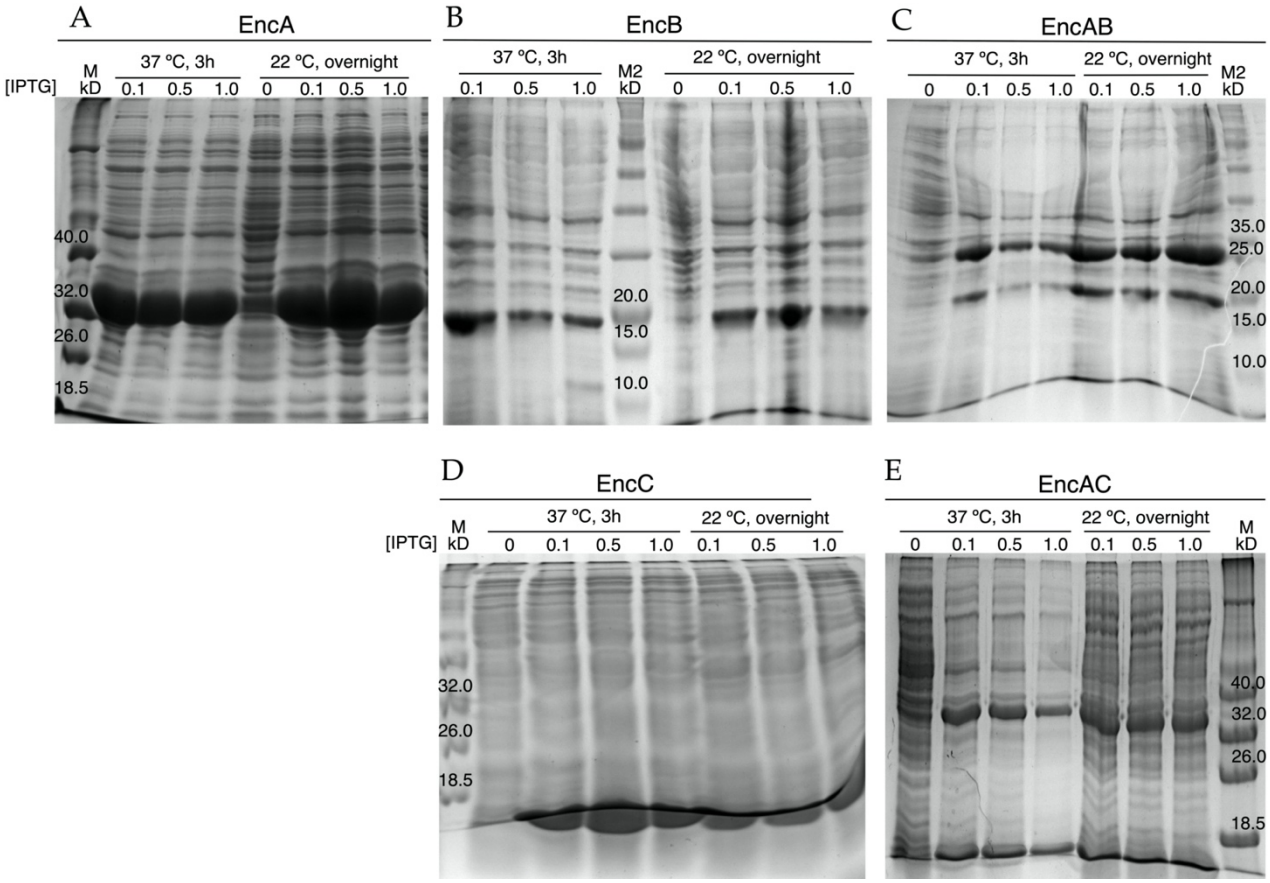

Figure 3.1: SDS-PAGE analysis of the *Myx. xanthus* encapsulin system protein expression tests. Expression tests of **(A)** EncA, **(B)** EncB, **(C)** EncAB, **(D)** EncC and **(E)** EncAC in *Esc. coli* BL21(DE3) competent cells in LB medium. Protein expression was induced at $OD_{600\,nm}$ ~ 0.8 with different IPTG concentrations (0, 0.1, 0.5, 1.0 mM) for 3 h at 37 °C or overnight at 22 °C. M – LMW II, M2 – NZYColour I protein marker. Sample loading was normalized based on the $OD_{600\,nm}$ (12.5 % gels).

3.2.1.2 EncC

EncC was successfully expressed in all conditions tested as evidenced by an overexpressed protein with a faster migration than the 18.5 kDa marker (the protein sequence predicts a monomer with a molecular weight of approximately 12.8 kDa) (Figure 3.1D). Expression of EncC in its soluble form was achieved in *Esc. coli* BL21(DE3) competent cells containing the pET28-EncC vector following an overnight induction with 0.5 mM IPTG at 22 °C. The protein was loaded into a DEAE FF column and eluted between 340 and 390 mM NaCl (Figure 3.2C). After elution from the DEAE column the fractions were pooled according to the purity assessment by SDS-PAGE (Figure 3.2D). The protein was loaded into a Superdex 200 Prep grade SEC column resulting in a pure and homogeneous protein. This procedure yields a total of 60 mg of pure protein per liter of culture.

3.2.1.3 EncA, EncAB and EncAC

EncA was successfully expressed under all conditions tested resulting in an intense band with a migration similar to the 32 kDa protein marker (Figure 3.1A). Similarly, EncAB and EncAC were expressed in all conditions and exhibited two bands after co-expression with each vector pair (Figure 3.1C and E).

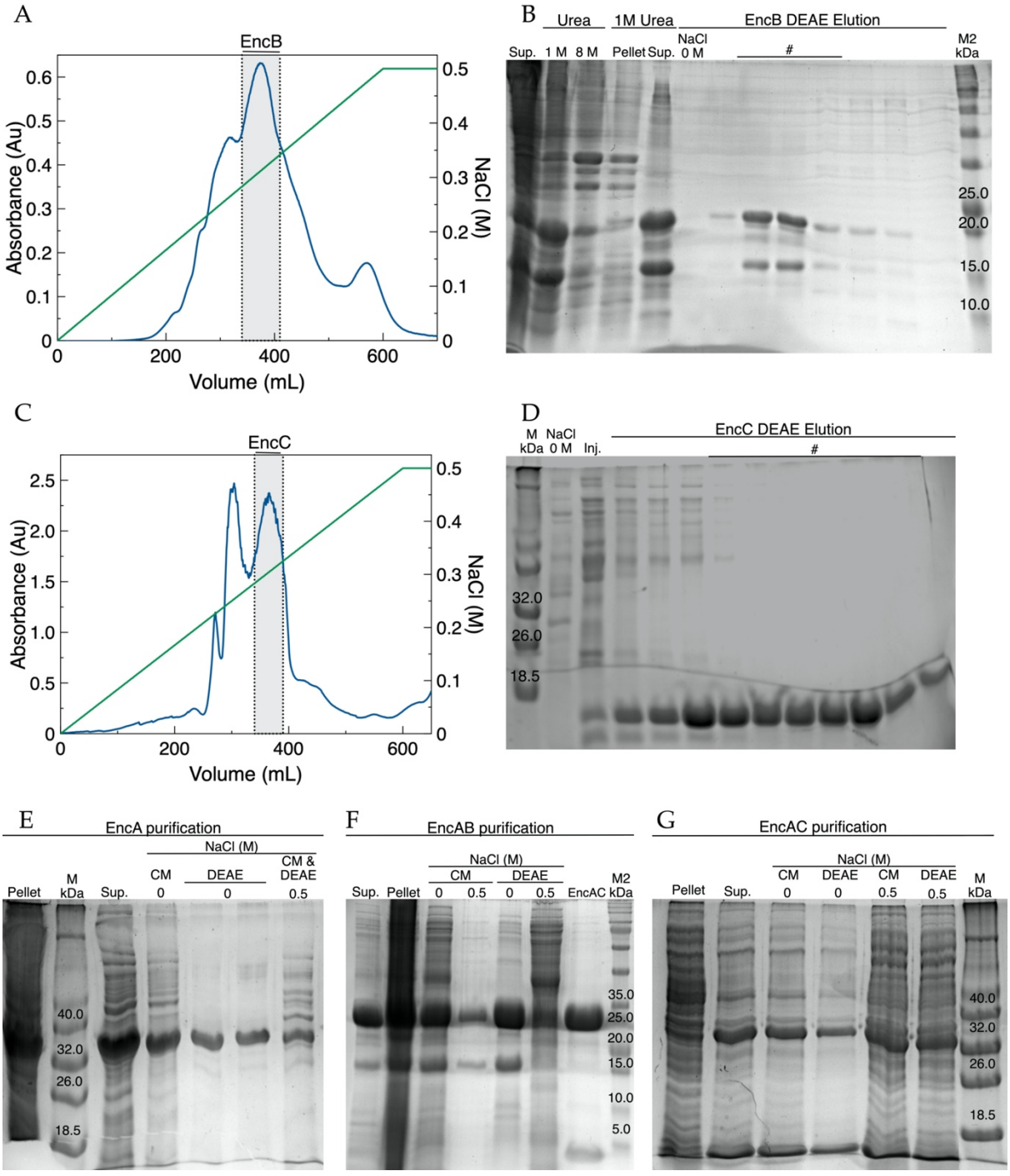

Figure 3.2: Purification of the proteins from the *Myx. xanthus* encapsulin system. (A) Elution profile of EncB purification on a DEAE Sepharose FF column using a continuous NaCl gradient. The fractions selected are represented in gray. **(B)** SDS-PAGE analysis of the solubilization of EncB' inclusion bodies (with 1 M and 8 M urea)

and analysis of fractions from the IEX column. **(C)** Elution profile of EncC from the DEAE Sepharose FF column using a continuous NaCl gradient. The fractions selected are represented in gray. **(D)** SDS-PAGE analysis of EncC purification step. **(E)** SDS-PAGE analysis of the purification process of EncA, **(F)** EncAB and **(G)** EncAC. # indicates the number of the fractions that were later pooled, Pellet – insoluble fractions from the centrifugation step, Sup. – Supernatant from the centrifugation step, Inj. – Fraction injected into the column, 0 – 0 M of NaCl to remove the non-interacting protein from the column, 0.5 – 0.5 M NaCl to remove the absorbed proteins, M – LMW II and M2 – NZYColour I protein markers (12.5 % gels).

The EncAB expression test gel shows a band with a similar migration to the EncA band and a second band similar to free EncB (Figure 3.1C). An analogous result is observed for the EncAC co-expression with the presence of two bands with similar migration to free EncA and EncC (Figure 3.1E). For large-scale production all proteins were expressed using 0.5 mM IPTG overnight at 22 °C. Although the pI of the EncA is relatively low (pI of 5.43), the outer surface of the shell seems to have a mixed charge, with positive, negative and neutral regions at pH 7.0 (Figure A.3, appendix). Therefore, the proteins were incapable of binding to either CM (cationic) or DEAE (anionic) resins (Figure 3.2E, F and G). However, both CM and DEAE chromatographic steps were successful by removing the positively and negatively charged contaminants, allowing the purification of EncA, EncAB and EncAC complexes as the flow-through. When encapsulated, EncB was expressed in a soluble form and protected from the enzymatic hydrolysis previously described for its free form. Still, a small proteolytic fraction is noticeable at the lane with the highest load in the SDS-PAGE electrophoresis, suggesting some level of hydrolysis occurring before encapsulation (Figure 3.3A, lane 4). This purification protocol yields pure homogenous recombinant EncA shell and EncAB or EncAC complexes with a typical yield of ~ 100 mg of pure protein per liter of culture.

3.2.2 Loading Efficiency

After purification using anion exchange chromatography, the resulting free EncB and EncC fractions exhibit a single peak when eluted from a SEC Sephacryl S-500 column (Figure 3.3C). Contrary to results previous published [21,30,64], EncA recombinantly expressed in *Esc. coli* cells appear as a single peak corresponding to a homogeneous T = 3 cage protein with no detectable T = 1 conformation. Both the EncAB and EncAC complexes eluted with an elution volume similar to EncA. Furthermore, SDS-PAGE gels of the EncAB and EncAC complexes confirm the correct encapsulation of the cargo proteins (Figure 3.3A and B).

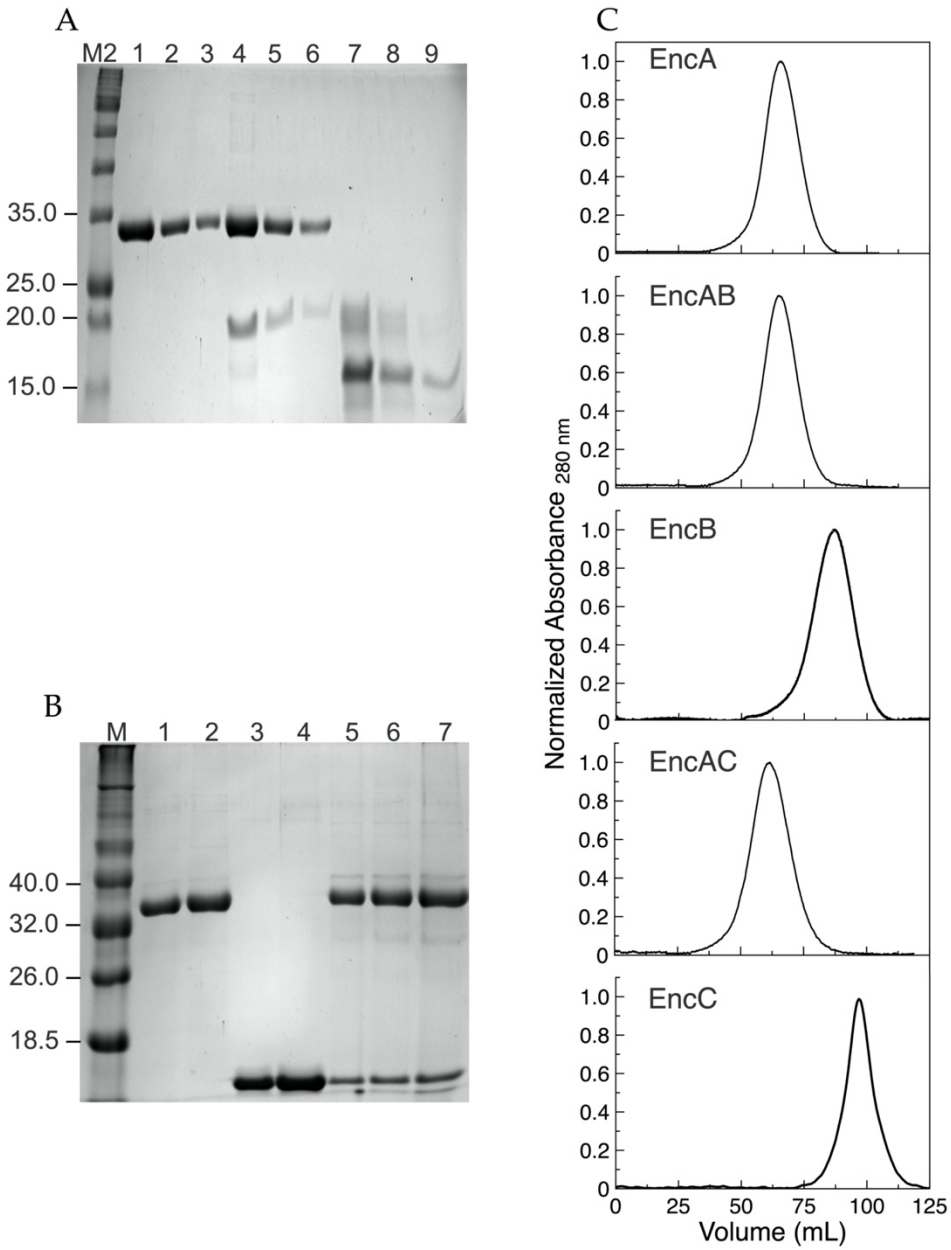

Figure 3.3: Assessment of cargo protein encapsulation by EncA and quantification of loading efficiency. (A) EncA and EncB samples were loaded alongside a serial dilution of EncAB samples of unknown concentration on an SDS-PAGE gel (12.5 %). 1 – 3: 3.5, 1.5 and 0.5 μg of EncA, 4 – 6: Serial dilution of EncAB, 7 – 9: 3.5, 1.5 and 0.5 μg of EncB. **(B)** EncA and EncC samples were loaded alongside a serial dilution of EncAC samples of unknown concentration on an SDS-PAGE gel (12.5 %). 1 – 2: 1.5 and 3.5 μg of EncA, 3 – 4: 1.5 and 3.5 μg of EncC, 5 – 7: Serial dilution of EncAC. **(C)** EncA, EncAB, EncB, EncAC and EncC elution profiles from a Sephacryl S-500 SEC column in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. M – LMW II and M2 – NZYColour I Protein Marker (12.5 % gels).

The loading capacity of the *Myx. xanthus* encapsulin was estimated by SDS-PAGE densitometry, in similarity with other works (Figure 3.3A and B). In this case, densitometric analysis of the protein samples resulted in a loading capacity of 52 ± 4 copies of EncB and 150 ± 7 copies of EncC inside each EncA 180-mer cage in the EncAB or EncAC complexes,

respectively. McHugh *et al.* determined that the cargo protein accounted for 40 % of the total mass of the complex with ~ 36 copies of EncB, ~ 92 copies of EncC and ~ 47 copies of EncD [21]. A more recent work with the co-expression of EncAB and EncAC only yielded a 30 – 40 % cargo protein content, suggesting a lower loading efficiency [30]. These data suggest that the loading efficiency depends on the experimental conditions tested.

EncA, EncAB and EncAC are eluted with similar volumes indicating a similar particle hydrodynamic radius (the column exclusion limits of the Sephacryl S500 GL column are 40 kDa to 20 MDa) and their elution profiles exhibit no detectable elution bands characteristic of EncB or EncC, indicating total encapsulation of each cargo protein. Additionally, free EncB has a larger apparent hydrodynamic radius when compared with EncC. These differences will be further analyzed in the following section.

In conclusion, high yields of pure and homogeneous EncA, EncB and EncC were obtained using the expression and purification protocols described in this section. The co-expression of the shell and each cargo encoding genes ensure the expression of the EncAB and EncAC complexes. Free EncB protein was obtained in the insoluble fraction and solubilization with 1 M urea and renaturation protocol led to the recovery of a homogeneous soluble protein fraction. Upon its co-expression with EncA the resulting complex was soluble. The loading efficiency of cargo encapsulation by EncA was determined using SDS-PAGE densitometry showing higher loading of cargo protein inside of EncA when comparing with previous studies.

The proteins herein described will be used in studies described in sections 4, 5 and 6.

4

STRUCTURAL CHARACTERIZATION OF THE *MYXOCOCCUS XANTHUS* ENCAPSULIN SYSTEM

This chapter presents the structural characterization of the encapsulin shell (EncA) from *Myxococcus xanthus* with and without its cargo proteins (EncB or EncC) by Small Angle X-ray Scattering (SAXS), Size Exclusion Chromatography (SEC), Dynamic Light scattering (DLS) and Synchrotron Radiation Circular Dichroism (SRCD).

The envelope of the encapsulin shell was compared while empty and harboring the cargo proteins. EncB and EncC oligomerization and the effect of metal binding were analyzed.

The thermostability of the proteins were also evaluated to verify if cargo protein encapsulation leads to changes in complex stability.

4.1 Experimental procedure

4.1.1 Small-angle X-ray Scattering

Small-angle X-ray Scattering (SAXS) data were measured at the EMBL beamline P12 at the PETRA III storage ring (DESY, Hamburg) [184,185] in collaboration with Dr. Clément Blanchet.

Scattered X-ray photos wavelength of 1.24 Å (10 keV) and a sample detector distance of 3.1 m and I (s) was recorded for $0.02 < s < 4.5$ nm^{-1} with $s = (4\pi\sin 2\theta)/\lambda$.

Independently, 50 – 100 μL of each protein at 10 mg/mL (8.0 mg/mL for EncB) were centrifuged at 11,000× g and loaded into a SEC column at a flow rate of 0.5 mL/min for SEC-SAXS analysis [185]. EncC and EncB were loaded into a Superdex 200 increase 5/150 GL (Cytiva), while EncA, EncAC and EncAB were loaded into a Superose 6 increase 10/300 GL (Cytiva), both pre-equilibrated with 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. Between each injection the column was equilibrated with 1.5 columns volumes of buffer. Approximately 900 frames (1 frame/s) were collected for each sample. Data were reduced using the SASflow pipeline [186]. The 2D images were radially averaged, frames with protein signal and no trace of radiation damage were averaged and used for further processing. The buffer signal was subtracted from the protein signal.

The one-dimensional data were analyzed using ATSAS program suite (PRIMUS, CHROMIX, GNOM) for analysis of the scattering profiles, Guinier plots, Kratky plots and pair distance distribution ($P(r)$) calculation [187]. Each protein envelope was obtained by *ab initio* shape reconstructions with DAMMIN [188] for EncA, EncAB and EncAC first applying P1 and icosahedral symmetry and with GASBOR [189] for EncB and EncC first with P1 and then applying a P2 restriction for EncC and P52 for EncB based on the oligomer state data. The flexible regions of EncB and EncC were predicted with CORAL [190] using the P1 symmetry and the atomic structure of each Flp. The decameric structure was used for EncB while the dimeric FOC and non-FOC monomers were used for the EncC modeling. At least 10 interactions of each protein were obtained (in DAMMIN, GASBOR and CRYSOL) to validate the models. The final model for each protein was obtained using DAMAVER [191] to align and average the models obtained by DAMMIN. The envelope was visualized using ChimeraX [192].

4.1.2 Size Exclusion Chromatography

The quaternary structures of EncB and EncC and their iron-loaded forms were evaluated using a calibrated Superdex GL 200 10/300 (Cytiva) with exclusion limits between 10 and 600 kDa pre-equilibrated in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. The column was calibrated with commercial protein markers: Catalase (232 kDa), Bovine Serum Albumin (66 kDa), Ovalbumin (44.3 kDa), Carbonic Anhydrase (29.4 kDa), and Cytochrome *c*

(13.2 kDa). The void volume was determined with individual injections of blue dextran < 1.0 mg (NZYTech). The apparent molecular weight and the Stokes radii were estimated using the calibration curve (Figure A.4, appendix), according with Equation 4 and Equation 5:

$$Log_{10}(MW) = -1.546\frac{V_e}{V_o} + 7.677 \qquad \text{Equation 4}$$

$$Log_{10}(HD) = -1.028\,\frac{V_e - V_o}{V_t - V_o} + 0.99 \qquad \text{Equation 5}$$

where MW is the apparent molecular weight in kDa, HD is the hydrodynamic diameter in nm, V_e is the elution volume of the protein, V_o is the void volume and V_t is the total volume of the column.

Approximately 1 mg of EncB and EncC incubated with different ferrous iron concentrations (0, 1, 2, 4 and 8 Fe^{2+}/dimer) in aerobiosis were injected into a calibrated Superdex 200 10/300 GL (Cytiva) at a flow rate of 0.5 mL/min. $FeSO_4{\cdot}7H_2O$ was dissolved in acidic milli-Q water at pH 2.0, under an inert atmosphere and quantified by the 1,10-phenanthroline method (calibration curve in Figure A.5, appendix) [193].

4.1.3 Synchrotron Radiation Circular Dichroism

Synchrotron Radiation Circular Dichroism (SRCD) spectra of EncA, EncAB, EncAC, EncB and EncC were acquired at the AU-CD beam line at the ASTRID2 synchrotron radiation source (ISA, Aarhus University, Denmark), using the periscope chamber setup. Samples containing ~ 1.0 mg/mL of protein freshly dialyzed against 10 mM MOPS buffer, pH 7.0, and 240 mM NaF were recorded with 1 nm steps and a dwell time of 2.0 s per step, in triplicates, using a 0.01008 cm pathlength quartz cell in the wavelength range of 170 – 280 nm, at 25 °C. The correct pathlength of the cell was determined by an interference technique [194].

A mixture of previously dialyzed proteins was prepared mimicking the cargo loading capacity ratio previously determined (see section 3.2.2). Therefore 0.86 mg/mL of EncA was incubated with 0.14 mg/mL of EncB and 0.75 mg/mL of EncA was added to 0.25 mg/mL of EncC and measured as previously described.

Freshly dialyzed EncC (1.0 mg/mL) was incubated with 4 Fe^{2+}/EncC dimer for 10 min at room temperature and in aerobic conditions and measured at 25 °C.

Measurement at 25 °C and with temperature scans between ~ 5 and ~ 83 °C, with either 5 or 10 °C increment steps, were collected for all proteins. When the protein scan reached ~ 83 °C the temperature was decreased to 25 °C following the same increment steps. The molar circular dichroism ($\Delta\varepsilon$) for each spectrum was calculated using the absorbance and the protein molar extinction coefficient at 205 nm [182]. The secondary structure content for each protein was determined using DichroWeb, with CDSSTR as analysis program and SP175 dataset [195].

The melting temperature was obtained by monitoring the CD signal intensity at 192, 209 and 220 nm. The data was fitted with a two-state thermal denaturation equation:

$$\mathrm{SI} = \frac{(\alpha_N + \beta_N \times Tmp) + (\alpha_D + \beta_D \times Tmp)e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_m}\right)}{RT}}}{1 + e^{-\frac{\Delta \mathrm{H}_m\left(1-\frac{Tmp}{T_m}\right)}{R \times Tmp}}} \qquad \text{Equation 6}$$

Where ΔH_{m} is the enthalpy change at the unfolding transition midpoint (kcal/mol), Tm the melting temperature, Tmp is the temperature in Kelvin, R is the universal gas constant (1.987 cal/mol/K) and α_{N}, β_{N}, α_{D} and β_{D} are normalization factors, as well as the slope and intercept of the baselines before and after the transition.

4.1.4 Dynamic Light Scattering

Dynamic Light Scattering (DLS) measurements were acquired with a scattering angle of 90°, at 25 °C, during 30 sec, in triplicates using a HORIBA SZ100 equipped with a 10 mW 532 nm laser. Temperature scans were acquired in a 5 °C steps with an equilibration time of 1 min at each temperature and obtained triplicates or until stabilization between measurements. DLS data were analyzed using the equipment built-in software assuming a polydisperse sample, a particle refractive index of 1.6 (organic sample) and water settings as dispersion medium (index of 1.333). Samples with 0.5 mg/mL of EncA, EncAB and EncAC and samples with 1.0 mg/mL of EncB and EncC were centrifuged at 11,000× *g* for 15 min in 200 mM MOPS buffer pH 7.0, 200 mM NaCl before each measurement.

4.2 Results

4.2.1 Structural Characterization and Conformational Dynamics

All datasets show good signal-to-noise ratio up to $s = 2\ nm^{-1}$ (Figure 4.1A) and good linearity with a small deviation at low s^2 value (bellow $0.015\ nm^{-2}$) in the Guinier analysis from the scattering data, suggesting low radiation damage and/or aggregation (Figure 4.1B).

Free EncA and EncAB or EncAC complexes exhibit similar SAXS scattering curves, typical of large spherical particles with a hollow core and correctly folded as shown in the Kratky plot (Figure 4.1D) [196]. The values of radius of gyration (R_g) for these protein samples are similar, varying between 14.4 and 14.9 nm, and D_{max} values of 35.2 − 35.5 nm (Figure 4.1C). The values obtained using the Guinier or the P *(r)* methods are similar (Table 4.1) [197]. The dimensions are consistent with the atomic structure of EncA, both with and without cargo proteins [21,30], corresponding to a sphere with a diameter of ~ 32 nm. The *P(r)* curve of EncA has a shoulder at ~ 15.2 nm corresponding to the internal cavity diameter (according to the atomic structure its hollow cavity is 13 nm wide) (Figure 4.1C). The EncAC *P(r)* curve has a shoulder at ~ 9.0 nm, while the EncAB *P(r)* curve shows no detectable shoulder. This difference in the internal cavity size might be related with the presence of the cargo protein, their loading efficiency and cargo dimensions/organization.

The molecular envelopes of EncA, EncAC and EncAB were obtained by *ab initio* modeling using DAMMIN [188]. P1 and icosahedral symmetries generated similar models. However, P1 symmetry resulted in a better scattering curve fitting. A total of 10 models were averaged and filtered to obtain the final model for each sample using DAMAVER [191] (Figure 4.1E). The atomic structure of EncA (PDB: 4PT2) docks well into the envelopes of EncA, EncAB and EncAC and the theoretical scattering curve of the atomic structure (by CRYSOL [198]) is shown as a blue line in Figure 4.1A.

Regarding EncB, the R_g is 4.66 ± 0.01 nm and a D_{max} of 16.9 nm (Table 4.1). The atomic structure of EncB shows a toroidal structure with D5 symmetry, 7.0 nm of diameter and a thickness of 5.0 nm [30]. The thickness of the EncB envelope calculated from SAXS data is similar (~ 5.6 nm), while the diameter considerably differs. This difference may be explained by the presence of the additional 84 residues in the C-terminal extensions of each EncB monomer that are absent in the atomic structure. Contrary to the core of the protein, these 84 residues are predicted to be unordered (Figure A.2, appendix). This may be the reason why the Kratky plot of EncB exhibits a partially flexible behavior (Figure 4.1D). The EncB envelope was generated by *ab initio* modeling using GASBOR [189] and applying a P52 (D5) symmetry. The crystal structure docks well into the generated SAXS model except for the extra protrusions that extend from the center, supporting the conclusion that the protruding N-terminal extensions may represent the residues absent in the atomic structure. The position of the flexible terminal region of EncB was predicted using CORAL confirming its position towards the solution (Figure A.6A, appendix) [190].

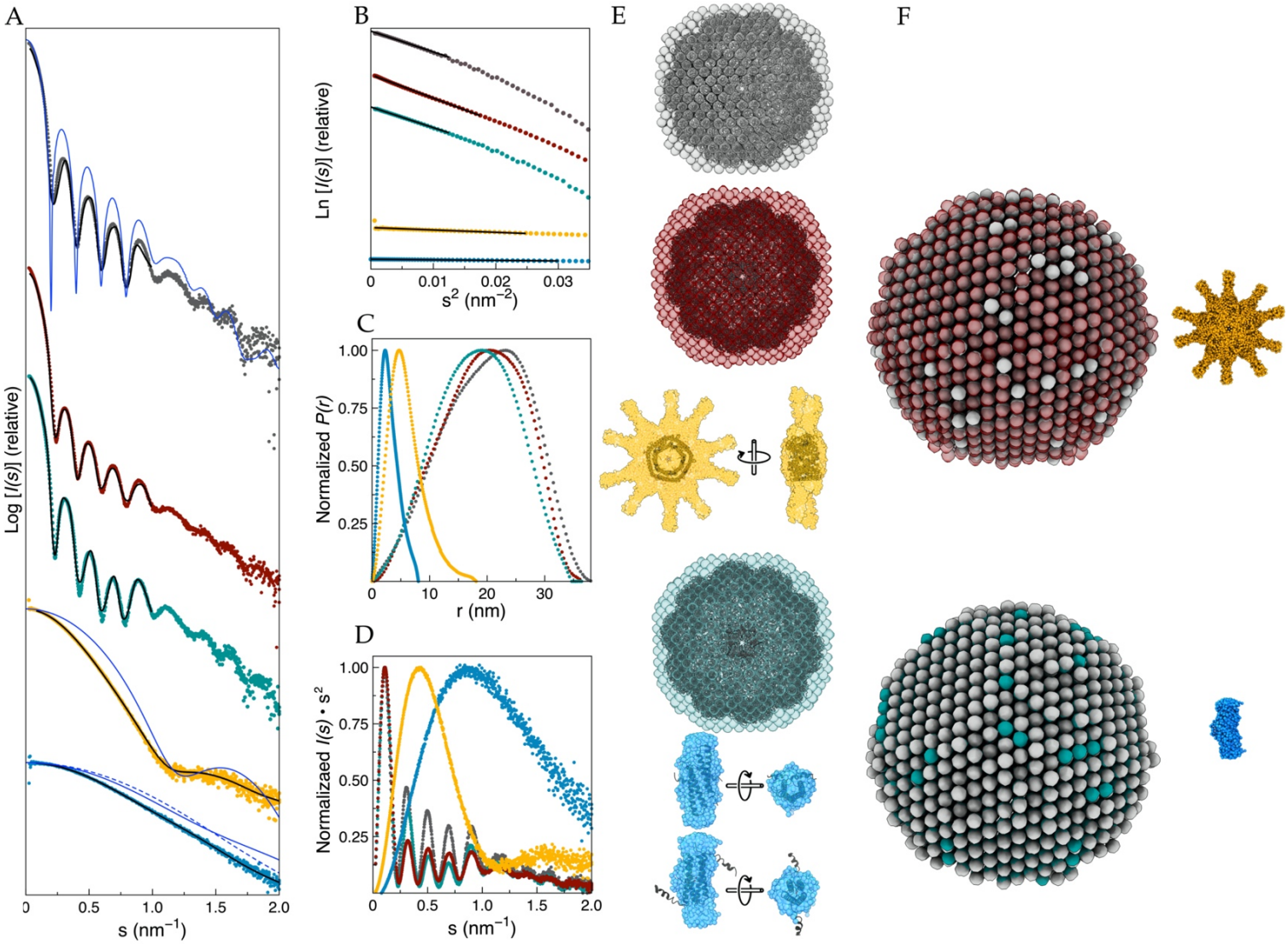

Figure 4.1: SAXS profiles and envelope models of the proteins from the *Myx. xanthus* Encapsulin system. **(A)** Experimental scattering curves for EncA (in gray), EncAB (in red), EncAC (in green), EncB (in yellow) and EncC (in blue). DAMMIN fits for EncA, EncAB and EncAC and GASBOR fits for EncB and EncC (as black lines). CRYSOL calculated scattering curves from the atomic structures of EncA (PDB: 4PT2), EncB (PDB: 7S5C) and EncC (PDB: 7S8T) (in blue lines). For EncC, full blue line is the FOC dimer and dotted blue line is the non-FOC dimer. **(B)** Guinier plot (dots) and linear fit (lines) of the scattering profiles. **(C)** Pair distance distribution *P(r)* curves. **(D)** Kratky plots. **(E)** *Ab initio* models obtained from each dataset: EncA, EncAB, EncAC were generated by DAMMIN while EncB and EncC were generated by GASBOR. The ribbon structures of each protein are superimposed to the envelopes. The EncC envelope is superimposed to a non-FOC dimer (Top) and a FOC dimer (Bottom). **(F)** Size comparison between the EncA, EncAB, EncB, EncAC and EncC envelopes (in scale).

Table 4.1: SEC-SAXS data collection and analysis parameters for the *Myx. xanthus* Encapsulin system.

	EncA	EncAB	EncAC	EncB	EncC
Data Collection					
SEC column	Superose 6 increase 10/300 GL			Superdex 200 increase 10/300 GL	
Concentration (mg/mL)	10.0	10.0	10.0	8.0	10.0
Buffer composition	200 mM MOPS buffer, pH 7.0 and 200 mM NaCl				
Structural Parameters					
I (0) from Guinier	663,783 ± 974	102,728 ± 114	1,060,810 ± 1412	2,265 ± 1.7	4,230 ± 5.1
R_g from Guinier (nm)	14.83 ± 0.05	14.41 ± 0.04	14.94 ± 0.04	4.66 ± 0.01	2.43 ± 0.01
I (0) from *P(r)*	673,700.00	92,350.00	1,045,000.00	2,259.00	4,230.00
R_g from *P(r)*	14.9	14.4	14.5	4.7	2.5
D_{max} from *P(r)*	35.3	35.2	35.5	16.9	7.7
Porod Volume estimate (nm^3)	16,683.5	22,945.6	18,523.0	394.7	40.1
Molecular Weight (kDa)					
Estimation from data	nd	nd	nd	208.0	25.9
Theoretical molecular weight				17.5	12.7
Ab initio		DAMMIN		GASBOR	
Symmetry	P1	P1	P1	P52	P2
$\chi 2$	144	209	231	1.86 – 2.71	1.11 – 1.44
Rigid body modelling	–	–	–	CORAL	
Symmetry				P1	P1
$\chi 2$				4.88 – 7.46	1.96 – 3.20
Structured residues				14–79 [PDB: 7S5C]	1–80 [PDB: 7S8T]
Flexible residues				1–13 and 80–164	81–116

nd – not determined

EncC has a scattering profile typical of an elongated protein (Figure 4.1). The protein has a R_g of 2.43 ± 0.01 nm and a D_{max} of 7.7 nm. The data is consistent with a dimer assembly (Table 4.1). According to the atomic structure of other Flp proteins, the structural building block can be either the non-FOC dimer (6.2 nm wide) or the FOC dimer (6.5 nm wide). The differences between the scattering curves predicted from the atomic structure of the non-FOC and FOC dimers (by CRYSOL) and the experimental dataset may be due to the presence of additional 37 residues found in the C-terminal region of the EncC protein sequence (similar to EncB), responsible for tethering to the EncA shell (Figure A.2, appendix). The EncC envelope was obtained with both P1 and P2 symmetry using GASBOR. The P2 model is shown to reflect the putative protein quaternary structure. The model consists of an elongated envelope to which both the FOC and the non-FOC dimers can be superimposed (with a better superposition with the non-FOC dimer). Similarly, the C-terminal of the protein was predicted with CORAL and this region is shown to be solvent exposed (Figure A.6B, appendix). Therefore, while EncB was found in a decameric form, EncC seems to assemble into a dimer. However, with the resolution limit of the SAXS data it is impossible to ascertain whether the dimer is formed by the FOC or by the non-FOC building block.

Didi *et al.* showed that the presence of Fe^{2+} ions induced the oligomerization of *Rhs. rubrum* Flp [20]. To test if the addition of ferrous iron promoted a higher oligomerization state of the *Myx. xanthus* EncC, the protein was titrated with Fe^{2+} and each iron-loaded sample was loaded into a calibrated Superdex 200 10/300 GL (Cytiva) column (Figure 4.2B).

The elution volume of EncC while incubated with different amounts of iron was compared with the elution volume of decameric EncB (Figure 4.2A, V_e = 11.0 mL). In the absence of Fe^{2+} ions, EncC eluted as a single peak with an elution volume of 15.7 mL (Figure 4.2B) corresponding to an apparent hydrodynamic diameter of 8.1 nm, similar to the D_{max} of 7.7 nm determined by SAXS. The addition of Fe^{2+} (1, 2, 4, and 8 Fe^{2+}/EncC dimer) gradually shifted the peak to smaller elution volumes, from 15.7 mL to 12.1 mL, consistent with a hydrodynamic diameter of 12 nm (Table 4.2). Additions of iron above 10 Fe^{2+} atoms per EncC dimer led to protein precipitation. Similarly, the addition of Fe^{2+} ions to EncB also induced protein precipitation. According to the SAXS data, the peak of the apo-form at 15.7 mL corresponds to the dimeric EncC, whereas the EncB peak at 11.0 mL (hydrodynamic diameter of 13.7 nm) is consistent with the decameric EncB.

The results from the SEC experiment suggest that the addition of Fe^{2+} ions to EncC converts the dimer state into a decamer structure similar to EncB. The difference between the decameric EncB (13.7 nm) and EncC (12.0 nm) elution volumes may be explained by the difference in the length of their flexible protruding C-terminal extensions. The EncC and EncB extensions comprise around 20 residues and 80 residues, respectively. Noteworthy, recently obtained decameric toroidal atomic structure of EncC was determined in the presence of $FeCl_2$, revealing irons atoms in the FOC [30].

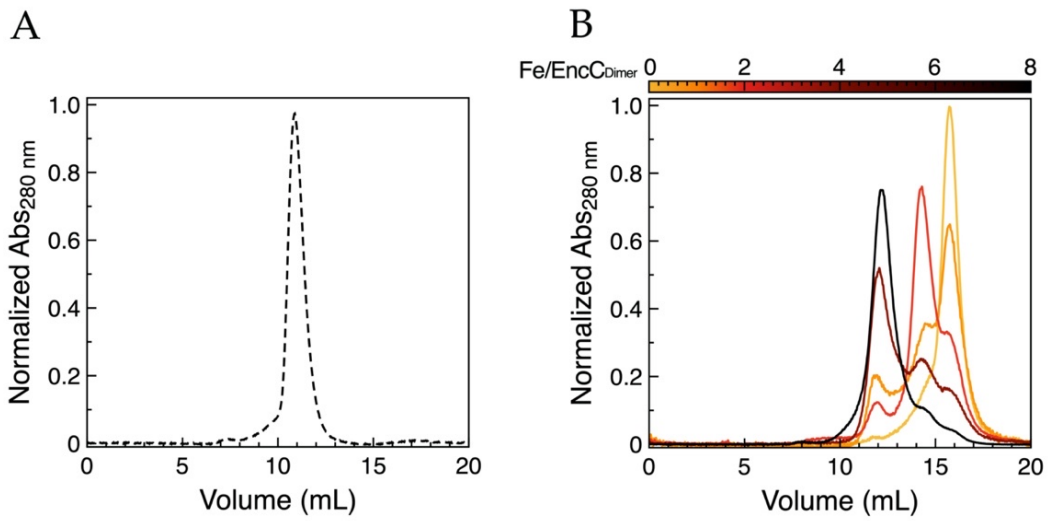

Figure 4.2: Effect of iron loading on the quaternary structure of EncB and EncC. (A) SEC elution profile of the apo-form of EncB (dotted line). **(B)** EncC in its apo-form (in yellow) and upon incubation with 1 to 8 Fe^{2+} per EncC dimer (from orange to black curves). Superdex 200 10/300 GL (Cytiva) in 200 mM MOPS buffer pH 7.0, 200 mM NaCl.

Table 4.2: Effect of incubation with Fe^{2+} ions on the quaternary structure of EncC, analyzed by SEC.

Fe/EncC dimer	Ve (mL)	HD (nm)
0	15.7	8.1
1	15.7	8.1
2	14.3	9.3
4	12.0	12.0
8	12.1	12.0

V_e – Elution volume; HD – Hydrodynamic diameter

4.2.2 Secondary Structure Assessment and Thermal Protection

Besides probing the quaternary structure and conformational dynamics of the *Myx. xanthus* Encapsulin system, the secondary structure of these proteins was also evaluated.

All SRCD spectra exhibit negative peaks at 209 and 220 nm and a positive peak at 192 nm, coherent with folded proteins with high helical content (Figure 4.3A). A quantitative assessment of the percentage of each secondary structure element for each protein was obtained by deconvoluting the SRCD spectra using DichroWeb [195] and comparing with the secondary structure content determined from their respective atomic structures (PDB structures: 4PT2 and 7S20 for EncA, 7S5C and 7S5K for EncB and 7S8T for EncC) using 2Struc [199] (Table 4.3). The atomic structures of EncB and EncC are incomplete and lack the final 84 or 37 residues, respectively, while the percentages obtained in the SRCD experiment reflect the entire protein sequence. In solution, EncA displays a secondary structure content similar

to its atomic structures with 26.5 % α-helix, 26.5 % β-strand and 47.0 % of others (turns and unordered) (Table 4.3). Regarding the cargo proteins, EncB has 52.5 % α-helix, 8.0 % β-strand and 39.5 % of others while EncC exhibits 76.4 % α-helix, 0.0 % β-strand and 23.6 % of other structures. The global secondary structure percentages for the encapsulated forms EncAB and EncAC are similar, reflecting 31.3 – 34.2 % α-helix, 22.1 – 22.2 % β-sheet and 43.7 – 46.5 % others.

Comparing the secondary structure content of each cargo protein, as determined by SCRD, with their atomic structure, for EncB the 32.4 – 33.5 % decrease in helical structures is compensated by a 8.0 % increase in β-sheet and 24.4 – 25.5 % of others. For EncC the differences are less significant (Table 4.3). As mentioned, these differences are most likely due to the incomplete nature of their atomic structures since the unsolved region amounts to 51 % of the total sequence of EncB C-terminal region and 32 % in EncC. Taking this into account, together with the SRCD data, one can conclude that the C-terminal extensions of EncB and EncC are mostly unordered as predicted by bioinformatic studies (Figure A.2, appendix).

Additionally, considering the effect of iron on the Flp protein structures, Didi *et al.* also described that in the absence of iron the protein displays a low amount of secondary (and tertiary) ordered structures [20]. Yet, despite the EncC dependence of iron to assemble as a decamer, its secondary structure is not affected upon iron addition, as determined from the SRCD data (Figure 4.3A, last row), suggesting that the overall secondary structure of the dimer state is similar to the fully assembled decamer.

Moreover, the thermal stability of the proteins was investigated and compared to assess the impact of protein cargo encapsulation (spectra presented on Figure A.7, appendix). The denaturation curves were plotted at three wavelengths (192, 209 and 220 nm, Figure 4.3B). To understand the relevance of cargo protein encapsulation, artificial mixtures of EncA and either EncB or EncC (mimicking the loading stoichiometry of each complex) were also analyzed (Figure 4.3C). By following the denaturation of empty EncA and fitting the experimental curve with a two-state thermal denaturation equation (Equation 6) it is noticeable that the shell protein is a highly thermostable protein with a melting temperature (T_m) of 76 ± 0.7 °C), albeit lower than the T_m value of 86.6 °C reported for the *Qua. thermotolerans* encapsulin [25].

No significant differences were detected between the melting curves of EncA, EncB, EncAB and the EncA+EncB mixture, suggesting that EncB is equally thermostable, exhibiting similar T_m and enthalpy energy change (ΔH_m) upon thermal denaturation (Table 4.4). EncC has a considerably lower T_m (41 ± 2.6 °C). In this case, the encapsulation of EncC (EncAC complex) increased the T_m of the complex to 81 ± 0.3 °C, 3.9 °C higher than EncA and 40 °C higher than free EncC. The artificial mixture of EncA and EncC showed no protective effect, with each protein being independently denaturated. These results confirm that encapsulation shields the cargo protein from thermal denaturation and additionally suggests that cargo binding to the shell tends to increase the overall thermostability of the complex.

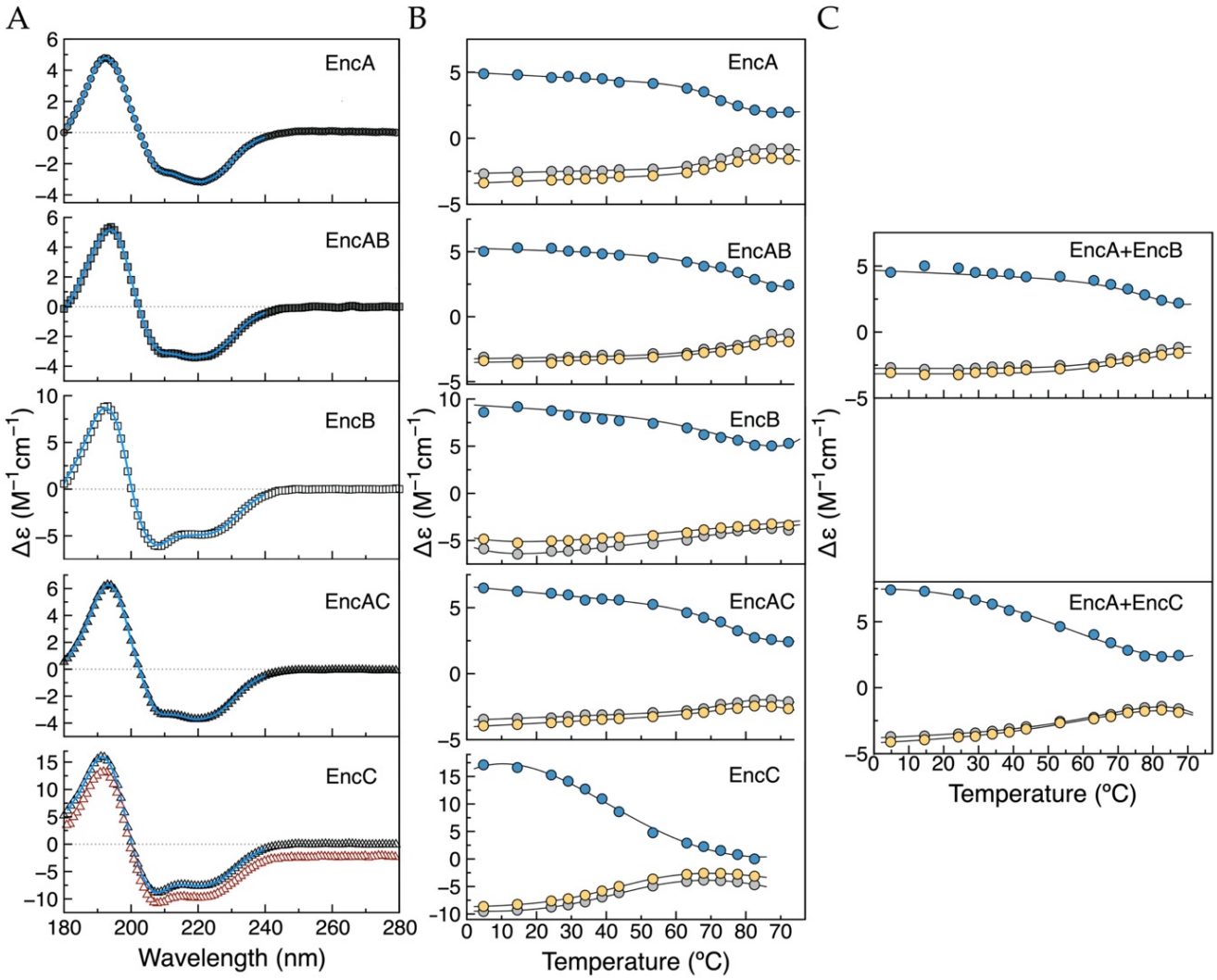

Figure 4.3: Secondary structure and thermostability assessment of the encapsulin system. (A) SRCD spectra in Molar circular dichroism ($\Delta\varepsilon$) of EncA (full circles), EncAB (full squares), EncB (empty squares), EncAC (full triangles), EncC (empty triangles) and EncC incubated with 4 Fe^{2+} per EncC dimer during 10 min at atmospheric conditions (empty red triangles with an offset of −2.5 $M^{-1}cm^{-1}$) at 25 °C. Fit from the deconvolution of each spectrum using Dichroweb in blue lines. **(C)** Molar circular dichroism ($\Delta\varepsilon$) denaturation curves of EncA, EncAB, EncB, EncAC and EncC. **(D)** Molar circular dichroism ($\Delta\varepsilon$) denaturation curves of mixture of EncA with EncB (Top) and mixture of EncA with EncC (Bottom). The melting curves were plotted by monitoring the signal at 192 (in blue), 209 (in yellow) and 220 nm (in gray) as a function of temperature and fitted with a two-state denaturation curve (Equation 6) with a solid line.

Table 4.3: Secondary structure composition of the free proteins and complexes of the Encapsulin system by deconvolution of the SRCD spectra obtained at 25 °C or analysis of atomic structures of each protein.

			Secondary structure content (%)		
Protein	**Residues**	**Analysis**	**α-Helix**	**β-Sheet**	**Others**
EncA	1 – 294	DichroWeb	26.5 ± 0.5	26.5 ± 0.2	47.0 ± 0.1
	9 – 285	PDB: 4PT2	27.8 ± 0.3	27.6 ± 2.4	44.6 ± 2.7
	9 – 285	PDB: 7S20	29.0 ± 1.4	28.4 ± 3.6	42.6 ± 2.1
EncAB		DichroWeb	31.3 ± 0.1	22.2 ± 0.0	46.5 ± 0.1
EncB	1 – 164	DichroWeb	52.5 ± 0.2	8.0 ± 0.1	39.5 ± 0.4
		AlphaFold	39.0	0.0	61.0
	14 – 80	PDB: 7S5C	86.0 ± 2.3	0.0 ± 0.0	14.0 ± 2.3
		PDB: 7S5K	84.9 ± 2.7	0.0 ± 0.0	15.1 ± 2.7
EncAC		DichroWeb	34.2 ± 0.2	22.1 ± 0.1	43.7 ± 0.2
EncC	1 – 116	DichroWeb	76.4 ± 0.1	0.0 ± 0.0	23.6 ± 0.1
		AlphaFold	61.6	0.0	38.5
	1 – 79	PDB: 7S8T	70.7 ± 3.1	0.0 ± 0.0	29.3 ± 3.1

The secondary structure content determined from the atomic structures was achieved using 2Struc

Nevertheless, thermal denaturation not only impacted in the secondary structure of the proteins but also their monodispersity and aggregation. DLS temperature scans were used to evaluate the propensity of these proteins to self-aggregate due to thermal denaturation (Figure 4.4).

At 25 °C, EncA, EncAB and EncAC show hydrodynamic diameter (Z-average) values similar to those obtained by SAXS with a hydrodynamic diameter of 35 ± 0.7 nm for EncA, 37 ± 0.6 nm for EncAB and 40 ± 2.7 nm for EncAC (Figure 4.4A). Samples of each free cargo protein also exhibit results consistent with the SAXS datasets, since EncB has a hydrodynamic diameter of 14 ± 0.7 nm and EncC of 6.2 ± 0.5 nm (Figure 4.4A). All proteins showed polydispersity indexes below 0.3 at 25 °C.

During the temperature scan, the particle size of EncA is stable until an abrupt increase around 70 °C, with similar results obtained for EncAB and EncAC at ~ 65 °C and ~ 70 °C (Figure 4.4B). As such, the T_m obtained by SRCD and the aggregation temperature (T_{agg}) determined by DLS were identical despite monitoring different phenomena. However, EncB started to form large particles around 45 °C. This T_{agg} is quite different from the T_m determined by SRCD (Table 4.4). Thus, this suggests that as the temperature increases EncB starts to self-aggregate at lower temperatures while maintaining its secondary structure until it eventually further denatures. The encapsulation of EncB by EncA protects it from this aggregation event probably due to its tethering to the shell, minimizing self-association contacts. EncC shows a

similar T_m and T_{agg} values (Table 4.4). Therefore, in this case the encapsulation not only protects the cargo protein from secondary structure changes upon temperature increase but also from self-aggregation.

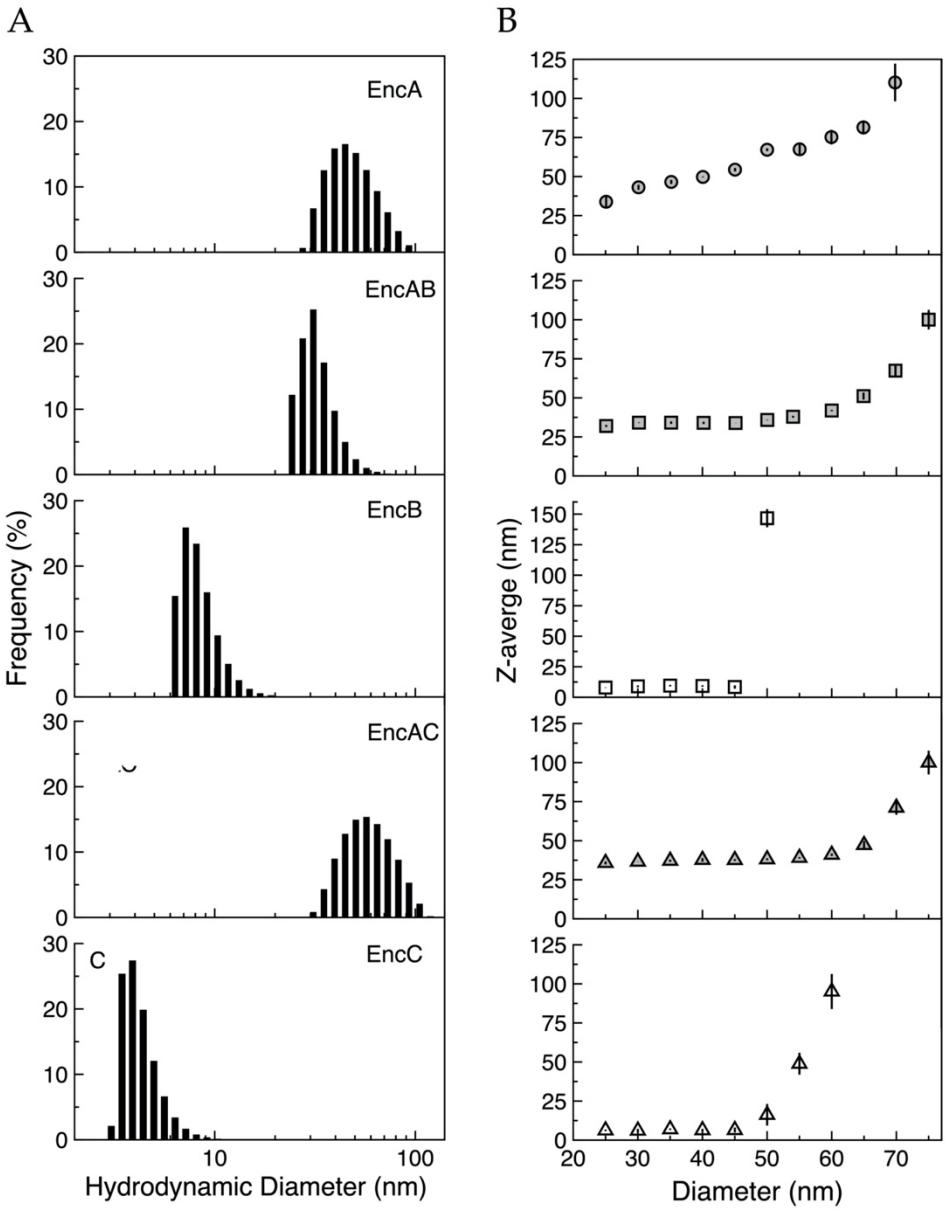

Figure 4.4: Particle size distribution of the *Myx. xanthus* Encapsulin system during thermal denaturation. (A) Hydrodynamic diameter (Z-average) of EncA, EncAB, EncB, EncAC and EncC at 25 °C. **(B)** Temperature scans of EncA (full circles), EncAB (full squares), EncB (empty squares), EncAC (full triangles) and EncC (empty triangles). The data was obtained in triplicates or until stabilization between measurements.

Table 4.4: Melting temperature and enthalpy at thermal denaturation midpoint of encapsulin proteins, complexes and artificial mixtures from the SRCD temperature scans and aggregation trigger temperature by DLS.

Protein	T_m (°C)*	ΔH_m (kcal/mol)*	T_{agg} (°C)**
EncA	76.9 ± 2.8	42.3 ± 2.1	≈ 70
EncAB	78.9 ± 0.6	67.3 ± 3.0	≈ 65
EncB	78.1 ± 2.6	45.1 ± 3.2	≈ 45
EncA+EncB	77.4 ± 0.7	42.8 ± 7.5	
EncAC	80.8 ± 0.3	38.9 ± 5.2	≈ 70
EncC	41.0 ± 2.6	23.0 ± 4.3	≈ 50
EncA+EncC	41.2 ± 1.7 69.1 ± 0.8	23.7 ± 1.6 39.2 ± 2.6	

* Parameters determined from SRCD experimental data fitted using a two-state melting model (Equation 6)

** Trigger temperature for aggregate formation as measured by DLS

4.3 Conclusions and Discussion

This chapter described the structural characterization of the proteins of the *Myx. xanthus* Encapsulin system. EncA can harbor three types of cargo proteins. Two of them, EncB and EncC, share high similarity with each other and with Flp proteins found in other encapsulin systems.

Here, SAXS data confirmed that the encapsulation of cargo proteins does not affect the overall dimensions of the shell protein in solution, showing similar D_{max} (35.2 – 35.5 nm) and R_g (14.41 – 14.94 nm) values, as well as superimposable envelope models. The presence of the cargo proteins in EncAB and EncAC seems to decrease the internal diameter from ~ 15.2 to ~ 9.0 nm for EncAC and to a non-detectable cavity for EncAB. These differences should be related the organization of the cargo proteins placed inside the cavity.

The SAXS model obtained for EncB in its apo-form reveals that it assembles as a decamer with toroidal morphology, similar to its atomic structure. Additionally, the *ab initio* envelope unveiled the presence of extensions corresponding to the 84 residues long C-terminal ends, which are absent from the atomic structure currently available. Although the EncAB has a lower cargo load, due to the presence of these long extensions the EncB cargo protein can be further away from the encapsulin shell resulting in an absent detection of the internal cavity in EncAB complex as suggested by the *P(r)* profile.

Interestingly, the apo form of EncC was found in the dimeric form. However, it is not possible to identify the type of dimers present in solution in the experimental conditions tested, since both non-FOC and FOC dimers are similar to the *ab initio* model and have an analogous reconstituted scattering profile when calculated using CRYSOL. Moreover, the addition of iron to EncC induces the assembly into a decamer, with the decrease of the elution volume in SEC analysis upon the addition of 8 Fe per dimer. There appears to be an intermediate state consistent with a tetrameric structure in the sample loaded with 4 Fe per dimer. The differences in the elution volume of EncB (13.7 nm) and EncC (12.0 nm) may be related with the presence of 84 residues in the C-terminal extension of EncB while EncC only presents 21 residues. Ross *et al*, showed that the Flp dimer is formed by the non-FOC interface and the addition of Fe^{2+} allows the self-association of five of these dimers into the final toroidal structure by contacts between the FOC interfaces [49]. This observation is attuned with the results presented in this work.

Although EncC requires ferrous iron to assemble as decamers, its secondary structure is already fully folded in its dimeric apo-form, with similar SRCD results obtained in the absence and presence of ferrous iron, contrary to other Flp that were previously suggested to be unstructured in the absence of Fe^{2+} ions [20].

The secondary structure analysis of EncA by SRCD renders compositions of each element similar to its atomic structure [21,30]. Both EncAB and EncAC exhibit a higher α-helix content than EncA due to the presence of either cargo proteins (both α-helix rich proteins) inside the cavity. EncB showed 7.9 % of β-sheet content which is not present in its atomic

structure. This suggests that the additional residues in the C-terminal extension may be composed of unordered regions and some β-sheet structures.

EncA exhibited high thermostability in both SRCD and DLS temperature scans, with a melting temperature (T_m) of 75.5 ± 0.74 °C obtained from SRCD and an aggregation temperature (T_{agg}) around 70 °C from DLS. The presence of EncB or EncC slightly increases the thermostability of the complex in 2 °C for EncAB and 3.9 °C for EncAC, which may be explained by stabilization of the shell protein through the tethering of the CLP sequences of the cargo protein. Regarding the cargo proteins, EncB secondary structure seems to be highly thermostable (T_m of 78.1 ± 2.62 °C) but starts to self-aggregate at ~ 45 °C (DLS data). Upon encapsulation EncB decamers are probably spaced apart, which hinders the self-aggregation, preventing EncB denaturation at lower temperatures. Similarly, EncC is a less thermostable protein with a T_m of 41.0 ± 2.6 °C and a T_{agg} of ~ 50 °C. Upon encapsulation this protein is significantly stabilized, with a T_m of 80.8 ± 0.3 °C and a T_{agg} of ~ 70 °C for the EncAC complex.

With this work we confirmed that cargo proteins EncB and EncC assemble into a decameric quaternary structure although EncC only achieves this oligomer state in the presence of iron. The position and organization of the C-terminal regions were characterized revealing a predominant unstructured tail as predicted by bioinformatics (Figure A.2, appendix). Additionally, this region likely extends towards the solvent as protrusions, possibly increasing the availability to interact with the cage protein. The encapsulation of these proteins is essential to increase their thermostability, either by preventing the loss of secondary structure or by avoiding self-aggregation.

5

UNVEILING THE IRON OXIDATION MECHANISM OF THE ENCAPSULIN SYSTEM

This chapter explores the iron oxidation activity of the encapsulin shell (EncA) from *Myx. xanthus*, both with and without its cargo proteins (EncB or EncC) using UV-Visible and Mössbauer spectroscopies.

The physiological relevance for the encapsulation of cargo proteins inside the cavity of EncA was evaluated by monitoring the iron mineralization kinetics of the protein complex under oxygen saturated conditions. Moreover, detection and quantification of iron species in each sample were achieved by Mössbauer spectroscopy. The iron oxidation and storage capacities of EncA and EncAC were compared, either in the presence of molecular oxygen or hydrogen peroxide.

The iron oxidation properties of free EncC were also evaluated and compared with EncAC, its encapsulated counterpart.

5.1 Experimental Procedure

5.1.1 Iron Uptake

The formation of ferric species was analyzed using UV-Visible spectroscopy as previously described (section 3.1.3.2). Spectra of either EncA, EncAB, EncAC, EncB and EncC were collected between 230 and 800 nm. Two oxidants were used as co-substrates for iron oxidation (hydrogen peroxide and molecular oxygen).

5.1.1.1 Iron Loading Capacity with Hydrogen Peroxide

EncA, EncAB, EncAC, EncB and EncC were incubated with consecutive additions of $FeSO_4$ in the presence of a 2-fold molar excess of H_2O_2 in 200 mM MOPS buffer pH 7.0 and 200 mM NaCl. 16 nM of EncA, EncAB and EncAC were incubated with 29 μM of $^{56}Fe^{2+}$ ions (1,800 Fe/protein per addition) and 58 μM of hydrogen peroxide per addition. Otherwise, successive additions of 29 μM of $^{56}Fe^{2+}$ and 58 μM of hydrogen peroxide were added to 29 μM of dimeric EncC and EncB. Spectra were acquired 5 min after incubation of each addition until no spectral changes could be detected or when evident protein precipitation occurred. The maximum iron storage capacity was determined by plotting the absorbance at 350 nm corrected for the baseline ($A_{350\,nm} - A_{800\,nm}$) and considering the dilution factor upon each addition.

The H_2O_2 stock solution was quantified by UV spectroscopy at 230 and 240 nm using the corresponding molar extinction coefficients ($\varepsilon_{230\,nm} = 72.4\ M^{-1}cm^{-1}$ and $\varepsilon_{240\,nm} = 43.6\ M^{-1}cm^{-1}$).

5.1.1.2 Iron Loading and Kinetics with Oxygen

The sample buffer (200 mM MOPS buffer, pH 7.0 and 200 mM NaCl) was saturated with oxygen for 15 min. The protein was transferred to a cuvette coupled to a magnetic stirring setup for sample homogenization. Spectra were acquired 4 h after each iron addition until protein precipitation or when no changes in the absorbance were detected.

Moreover, to study the kinetics of the iron oxidation reaction, the absorbance at 350 nm was measured after iron addition in oxygen saturated conditions. For the "high load" samples 16 nM of EncA, EncAB and EncAC were incubated with either 1,800, 3,600, 7,200, 10,800, 18,000 and 21,600 Fe^{2+}/protein. To study the iron oxidation reaction at "low load", 160 nM of EncA, EncAB and EncAC were incubated with 90, 180, 360, 720 and 1,800 Fe^{2+}/protein. 28 μM of dimeric EncC were incubated with 1, 2, 4, 7 and 10 Fe^{2+}/dimer.

5.1.2 Mössbauer Spectroscopy

Mössbauer spectra were acquired at 4 and 80 K in transmission mode using a 25-mCi ^{57}Co source in a Rh matrix and a conventional constant acceleration spectrometer. The velocity scale was calibrated with α-Fe foil at room temperature and the isomer shift values (δ) are correlated to this standard. The $^{57}FeSO_4$ solution was prepared by acidic dissolution of a ^{57}Fe metal foil (> 95 % enrichment) with H_2SO_4 in an anaerobic chamber as described by Ravi *et al.* [199] followed by iron quantification using the 1,10-phenanthroline method [193]. The analysis was carried out using the WMOSS software [200]. The spectra were acquired at C²TN, Instituto Superior Técnico, Universidade de Lisboa, in collaboration with Dr. João Carlos Waerenborgh and Dr. Bruno Vieira.

EncA, EncAC and EncC were incubated with $^{57}Fe^{2+}$ ions and reacted with either molecular oxygen or hydrogen peroxide as co-substrate. In the "high iron loads" samples 0.56 μM of EncA and EncAC in oxygen saturated buffer (200 mM MOPS buffer pH 7.0 and 200 mM NaCl) were incubated with $^{57}FeSO_4$ at a molar ratio of 1,800 (1 mM Fe) and 10,800 Fe per protein (6 mM Fe) for either 10 min or 24 h before being transferred to a Mössbauer cup and frozen in liquid nitrogen. The spectra were measured at either 55 K or 4 K without an external magnetic field.

For "low iron loads", 11 μM of EncA and EncAC in oxygen saturated buffer (200 mM MOPS buffer pH 7.0 and 200 mM NaCl) were incubated with different ^{57}Fe/protein ratios (90, 180 and 360, corresponding to 1, 2 and 4 mM ferrous ions) for 15 min under molecular oxygen saturation conditions. The spectra were measured at 80 K without an external magnetic field.

EncC samples were prepared under anaerobic conditions inside an anaerobic glovebox and in oxygen saturation. In the anaerobic samples the protein stocks and other reagents were degassed by performing three cycles of atmosphere substitution with argon using a Schlenk line as follows: frozen at −80 °C for at least 10 min followed by 3 min vacuum (while still frozen), 15 min of degassing and flushing with argon and a final 5 min argon saturation. The reagents were then transferred into an anaerobic chamber (< 4 ppm O_2, MBLab, MBraun) where the additions were prepared. 1 mM of dimeric EncC was incubated with 2 and 7 mM of $^{57}Fe^{2+}$ for 5 min and transferred to a Mössbauer cup and frozen in liquid nitrogen. Under oxygen saturation 1 mM of dimeric EncC in oxygen saturated buffer was incubated with 1, 2, 7 and 10 mM of $^{57}Fe^{2+}$ during 5 min. The spectra were measured at either 80 K or 4 K without an external magnetic field.

5.2 Results

5.2.1 Iron Loading Capacity

The iron oxidation activity of the proteins and complexes from the *Myx. xanthus* encapsulin system was assessed in the presence of molecular oxygen and hydrogen peroxide as co-substrates of the reaction (Figure 5.1). The proteins were successfully expressed in their apo-form since no absorbance was detected at 350 nm. The addition of ferrous iron led to a broad absorption band between 300 and 500 nm, characteristic of the formation of ferric species (Figure A.8, appendix).

Strikingly, EncA seems to be able to mineralize iron inside in its cavity with either molecular oxygen or hydrogen peroxide (Figure 5.1A and B, first row). A sigmoidal curve is observed in the presence of O_2, probably associated with two events. In an initial phase, and up to 7,200 irons per EncA, the amount of ferric species produced may be considered very small. At higher iron loads, above 7,200 irons per protein, the concentration of ferric species increases linearly (Figure 5.1A, first row). Above 24,000 irons per protein precipitation of ferric species start to occur, suggesting protein saturation. In the presence of H_2O_2, the amount of iron species detected per amount of iron added is linear until saturation at around 19,500 irons per protein (Figure 5.1B, first row).

The ferroxidation behavior of EncAC in the presence of O_2 is similar to EncA (with a slightly sharper slope after 7,200 irons per protein, Figure 5.1A, third row) but its loading capacity with H_2O_2 is slightly diminished to 18,000 irons per protein when comparing with the empty shell protein (Figure 5.1B, third row). This decrease is likely due to the spatial hindrance caused by the presence of EncC within the cage's cavity. For EncAB, the reaction in the presence of O_2, and contrary to the EncA or EncAC, is monophasic, with a steeper slope of product formation per substrate amount compared to EncA alone, which suggests a higher product formation in the presence of this cargo protein (Figure 5.1A, second row). The EncAB reaction with hydrogen peroxide leads to an iron loading similar to EncAC (Figure 5.1B, second row).

Although encapsulin shell may not be an enzyme in the conventional sense, the data presented shows that the iron is oxidized inside leading to a storage of large quantities of iron, with or without the ferritin-like cargo proteins encapsulated.

Regarding the free cargo proteins in solution, the addition of iron to EncB unfortunately leads to protein precipitation (as such, no data is shown). Notwithstanding, its rate-enhancing activity can be inferred by comparing the activity of EncAB with EncA. Moreover, and as previously mentioned, the addition of ferrous iron to EncC favors its self-association. Thus, for better readability, all EncC data are shown considering the dimeric state of the protein as the functional unit (one FOC per dimer). The addition of ferrous iron to EncC leads to the formation of ferric species and saturation is reached at approximately 10 irons per FOC in the presence of H_2O_2 and 7 iron with O_2, with ferric iron and protein precipitation occurring above this ratio (Figure 5.1). This suggests that EncC is unable to store a mineral

core and might provide the catalytic active sites for faster iron oxidation within the EncA cage. Therefore, all proteins and complexes studied here have shown to be functionally active, and the encapsulation of cargo proteins within EncA enhances the efficiency and iron loading capacity of the complex.

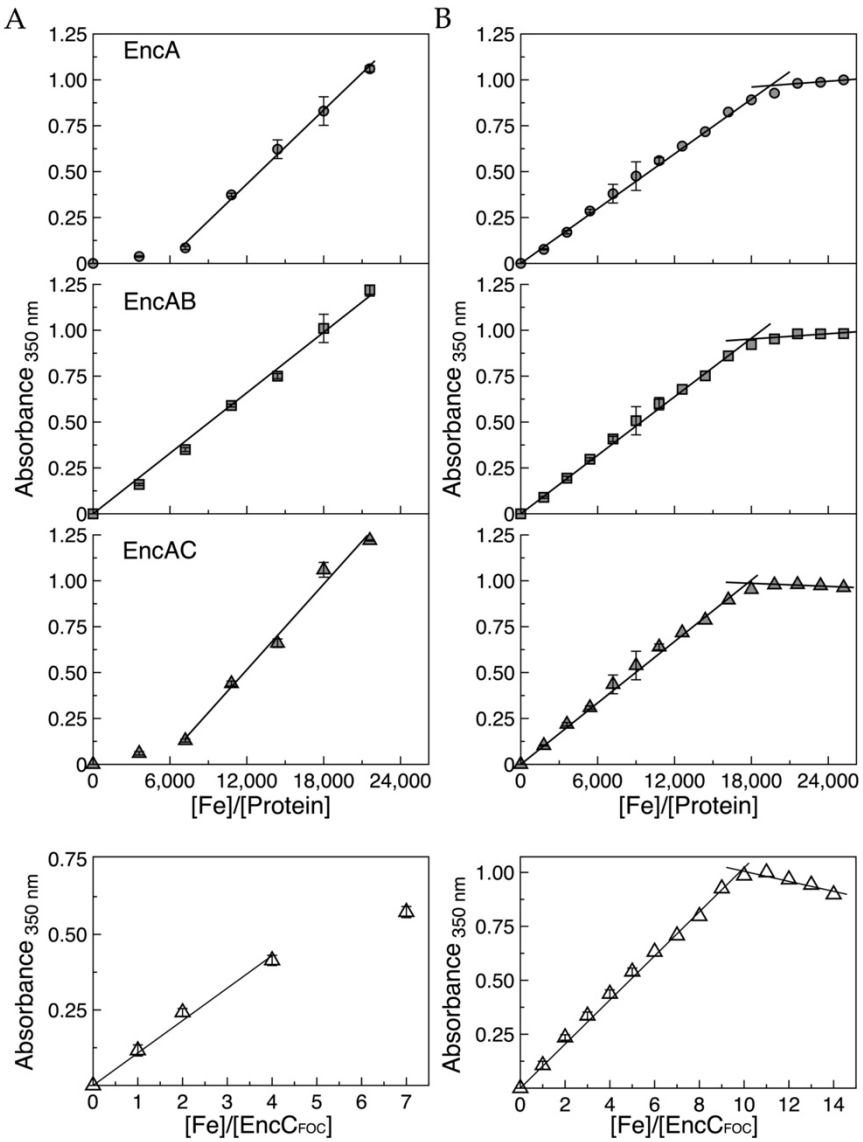

Figure 5.1: Spectroscopic evidence of the iron oxidation activity of the *Myx. xanthus* encapsulin system. Iron loading in the presence of **(A)** O_2 and **(B)** H_2O_2. EncA, EncAB and EncAC samples (16 nM) were incubated with different concentrations of Fe^{2+} ions. EncC (29 μM dimer) was incubated with 1 to 7 Fe^{2+}/EncC per FOC, in O_2 saturation conditions. The spectra were acquired 4 h after each addition and the absorbance at 350 nm was used to follow product formation. For the H_2O_2 experiments, successive additions of 29 μM Fe^{2+} with a 2-fold molar excess of H_2O_2 to EncA, EncAB and EncAC samples (16 nM) were performed. EncC (29 μM FOC) was incubated with a molar ratio of 1 to 14 Fe^{2+} per FOC) and a 2-fold molar excess of H_2O_2. The spectra were acquired 5 min after absorbance stabilization. The addition of iron substrate to EncB resulted in instant sample precipitation. The solid lines overlaying the experimental data were included to facilitate the comparison between the different proteins.

5.2.2 Iron Oxidation Kinetics

To further explore the iron oxidation capacity of these proteins, kinetic assays of EncA, EncAB, EncAC were performed in oxygen saturation conditions by addition of "high" (over 1,800 Fe/protein) and "low" (under 1,800 Fe/protein) loads of ferrous iron substrate and 1, 2, 4, 7 and 10 Fe^{2+} per EncC FOC.

The time courses of the reactions of EncA with 1,800 or 3,600 Fe^{2+}/protein are characterized by a nearly linear oxidation behavior (Figure 5.2A, first row). For higher iron loads, the reaction rate is enhanced, and a second phase appears with magnitude increasing with the substrate concentration. This sigmoidal kinetic profile is typical of an autocatalytic mineral surface reaction, previously observed in L-chain ferritins that, contrary to the H-type counterpart, lack the FOC centers [201]. Most likely, in the initial stages of each reaction (first phase), the metal is translocated to the encapsulin cavity through its negative pores and small ferric iron clusters are formed by spontaneous oxidation at specific nucleation sites on the surface of the inner cavity [202,203]. Once a mineral core with a reasonable size is attained, it becomes autocatalytic. When 1,800 and 3,600 (Figure 5.2A, first row, yellow and orange data points) ferrous irons are added per protein, the size of the initial mineral core is probably not large enough to promote mineral growth at its surface, as opposed to the samples reacted with 7,200, 10,800, 18,800 and 21,600 irons per protein (Figure 5.2A, first row, red to black data points).

A similar kinetic behavior is observed for the EncAB and EncAC complexes that show a sigmoidal progressive curve, with some subtle differences (Figure 5.2A, second and third row). Those differences are more noticeable for the highest iron load assay, in which the proteins were incubated with 21,600 Fe^{2+}/protein (Figure 5.2A, second and third row, black data points). A shorter lag-phase is observed for the EncAB and EncAC complexes compared to EncA where the plateau is reached later.

To further understand the difference between the EncA, EncAB and EncAC iron oxidation and mineralization kinetics, the proteins were also reacted with lower iron loads (Figure 5.2B). As previously observed, EncA exhibits a longer lag-phase when incubated with 1,800 irons per protein (Figure 5.2B, first row). However, the presence of EncB or EncC within EncA changes the progress curve from a sigmoidal to an exponential curve (Figure 5.2B, second and third row). This type of kinetics clarifies the catalytic activity of the complex and is associated with the presence of FOCs in the cargo proteins, responsible for the fast oxidation of ferrous iron (ferroxidation reaction), as reported for the *Rhs. rubrum* Encapsulin-Flp systems [56]. Although EncAC is more efficient than EncA, presenting shorter time range to complete the iron oxidation, the EncAB complex is even faster than EncAC, reaching a plateau after ~ 15 min, whereas EncAC achieves the same level of product formation after more than 90 min (Figure 5.2B, second and third row, the black data points (1,800 Fe per protein)). The difference between the apparent ferroxidation and mineralization kinetics of EncAC and EncAB might be related with the ferroxidation capacity of the cargo protein due to its superficial charge

and/or its position relatively to the shell pores. Additional experiments are needed to elucidate these differences.

The kinetic profiles of free EncC protein in solution (Figure 5.2B, fourth row) display a considerably higher ferroxidation rate while comparing with its encapsulated form. This might suggest that the rate-limiting step in the EncAC ferroxidase reaction is iron translocation from solution into the EncA cavity and the EncC active site.

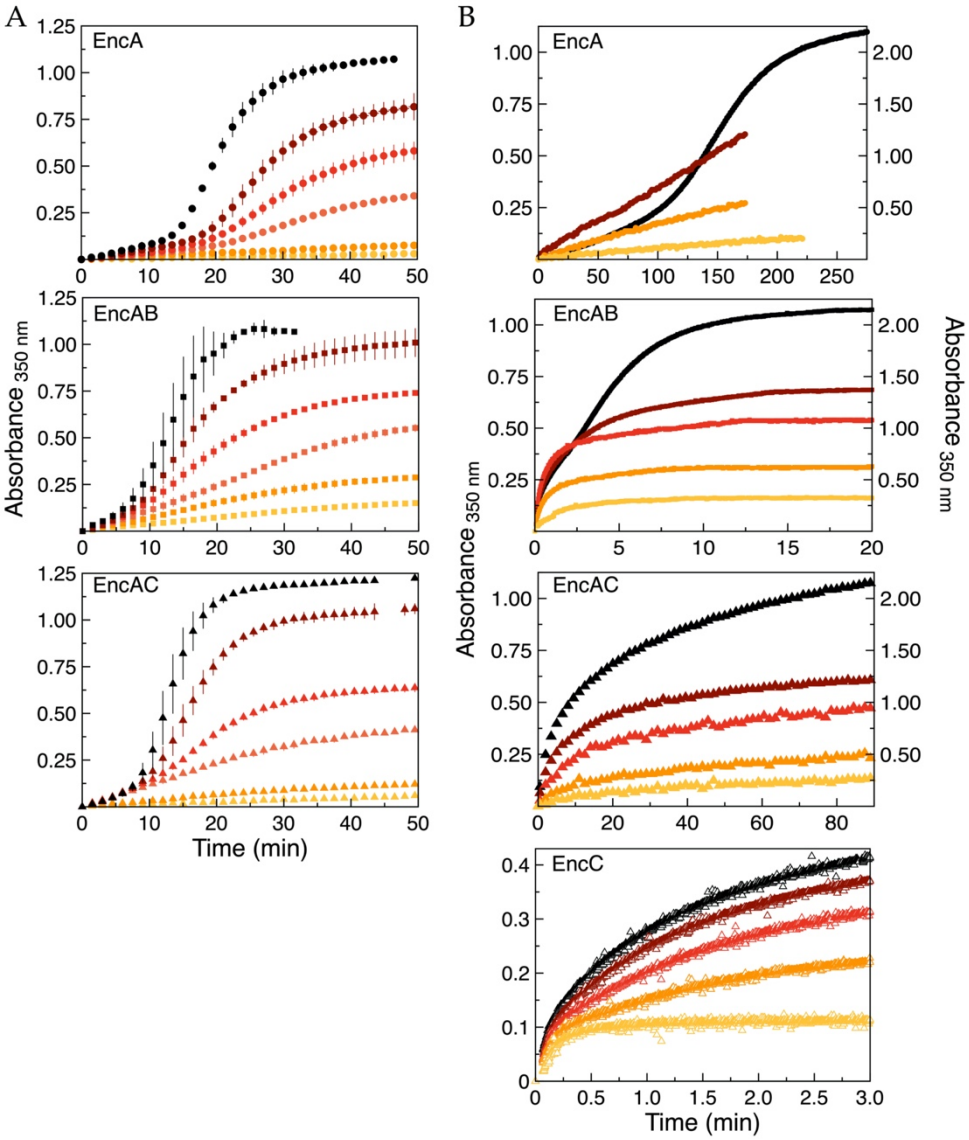

Figure 5.2: Progress curves of the iron oxidation by the *Myx. xanthus* encapsulin system with high and low iron loads in the presence of O_2 saturation conditions, using UV-Visible spectroscopy. (A) Kinetic profiles of iron oxidation with high iron loads (1,800 (in yellow), 3,600 (in orange), 7,200 (in coral), 10,800 (in red), 18,000 (in dark red) and 21,600 Fe^{2+}/protein (in black)) of EncA (circles), EncAB (squares) and EncAC (triangles). **(B)** Kinetic traces of the iron oxidation reaction with low loads of iron by EncA (circles), EncAB (squares), EncAC (triangles) (90 (in yellow), 180 (in orange), 360 (in red), 720 (in dark red) and 1,800 Fe^{2+}/protein (in black, is displayed as secondary axis)) and EncC (empty triangles) (1 (in yellow), 2 (in orange), 4 (in red), 7 (in deep red) and 10 Fe^{2+}/FOC (in black)).

5.2.3 Monitoring of Mineral Core Formation

The product of the iron oxidation in the presence of EncA and EncAC with either oxygen or hydrogen peroxide were characterized using Mössbauer spectroscopy at either 55 K (Figure 5.3) or 4 K (Figure 5.4) in the absence of an external magnetic field.

The Mössbauer spectra of EncA and EncAC reacted with $^{57}Fe^{2+}$ ions in the presence of either O_2 or H_2O_2 as co-substrate show no detectable ferrous species, suggesting total oxidation even after only 10 min (Figure 5.3). Considering the apparent iron kinetics using UV-Visible spectroscopy (Figure 5.2A), the presence of unreacted ferrous substrate was expected in these conditions. However, the complete iron oxidation observed may be explained due to an increased concentration of both protein and iron substrate (35-times higher), as required for Mössbauer spectroscopy.

The reactions of both EncA and EncAC with either oxidants resulted in a the sharp spectra observed in Figure 5.3 that can be explained by the contribution of two quadrupole doublets with $\delta = 0.48 \pm 0.01$ mm/s, and $\Delta E_Q = 0.58 \pm 0.02$ mm/s (Species A.Ia, red line) and $\Delta E_Q = 1.00 \pm 0.02$ mm/s (Species A.Ib, orange line) for EncA and $\delta = 0.48 \pm 0.01$ mm/s, and $\Delta E_Q = 0.57 \pm 0.02$ mm/s (Species AC.Ia, red line) and $\Delta E_Q = 1.00 \pm 0.02$ mm/s (Species AC.Ib, orange line) for EncAC (Table 5.1). These parameters are characteristic of high-spin Fe^{3+} species coordinated with oxygen/nitrogen ligands, as previously described in other protein nanocages [204].

The relative contribution of these iron species seems to indicate that in each data set (for EncA and EncAC), species A/AC.Ib decays into species A/AC.Ia along the reaction (Table 5.1). As such, the mineral core formed at the initial stages of the reaction with H_2O_2 (10 min) eventually evolves into another type of mineral core obtained after a longer period of incubation (24 h) (Figure 5.3B, D).

Moreover, the incubation of EncA with 10,800 iron atoms in the presence of O_2 for 24 h revealed a paramagnetic ferric species, with a sextet appearing in the spectrum at 55 K. This sextet signal suggests the presence of a larger mineral core as the mineralization reaction progresses due to the mineral reorganization along the reaction time (Figure 5.3A). The decreased contribution of the Species A.Ib with the concomitant appearance of the paramagnetic species indicates that this species correspond to the typical ferric mineral core.

Interestingly, the ferroxidation of 1,800 Fe by EncAC in the presence of O_2 (Figure 5.3C) resulted in a species distribution similar to the product of the reaction with H_2O_2 in both EncA and EncAC (Table 5.1). This suggest that a similar mineral core is being formed in these conditions possibly due to the characteristic ferroxidase activity of EncC.

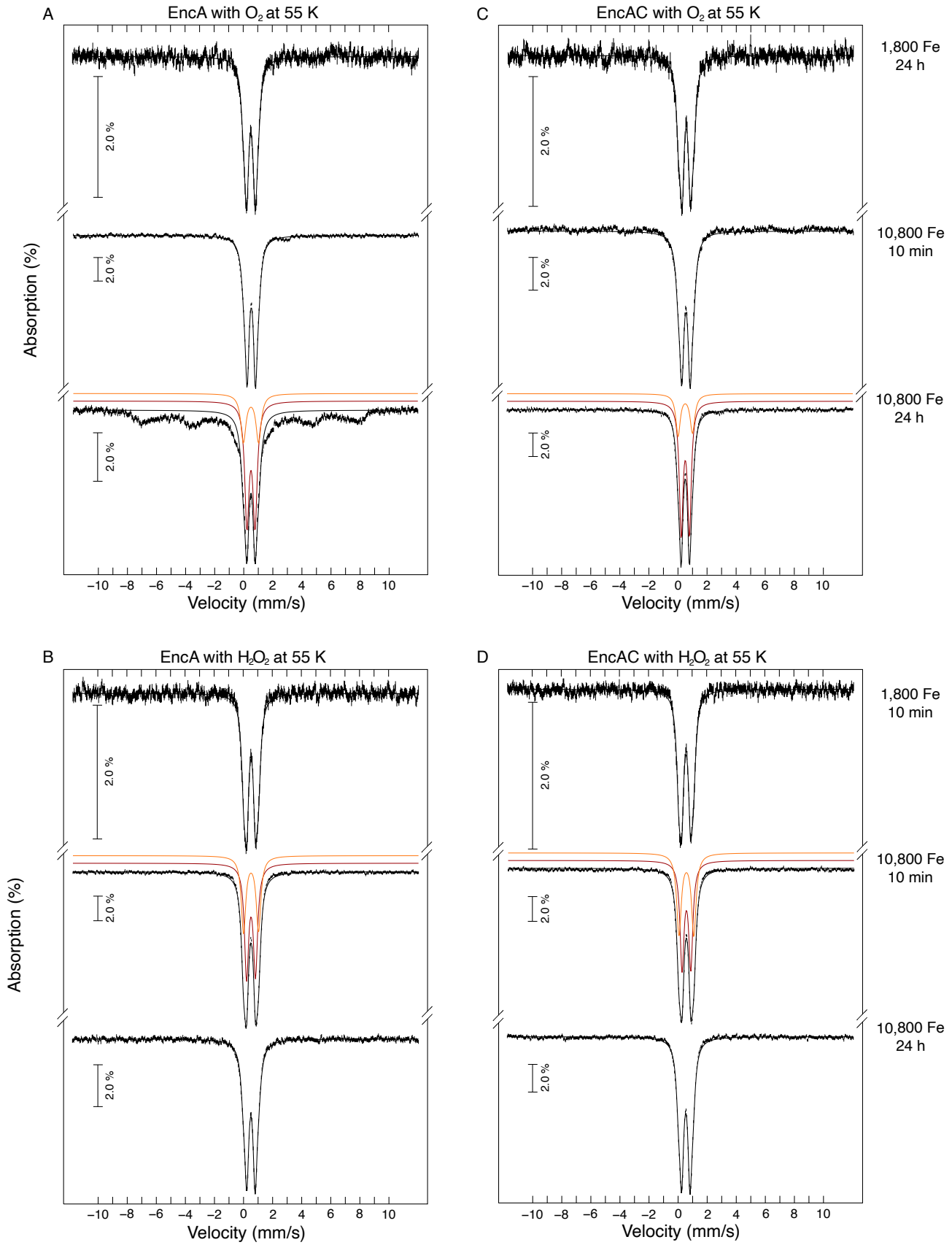

Figure 5.3: Mössbauer spectra of the EncA and EncAC reacted with ferrous iron in the presence of molecular oxygen and hydrogen peroxide. 1,800 and 10,800 $^{57}Fe^{2+}$ / protein and incubated with either molecular oxygen in the

presence of **(A)** EncA or **(C)** EncAC or hydrogen peroxide with **(B)** EncA and **(D)** EncAC for 10 min and 24 h, as displayed in the labels on the top and right side. The solid black lines represent the theoretical spectra reconstucted by fitting using the parameters from Table 5.1. In each data set, species A.Ia is represented as a red line (δ = 0.48 ± 0.01 mm/s and ΔE_Q = 0.58 ± 0.02 mm/s for EncA and species AC.Ia ΔE_Q = 0.57 ± 0.02 mm/s for EncAC) and represented as a orange line for species A.Ib and AC.Ib (δ = 0.48 ± 0.01 mm/s and ΔE_Q = 1.00 ± 0.02 mm/s for EncA and EncAC.

Table 5.1: Mössbauer parameters resulting from the deconvolution of the spectra from the EncA and EncAC iron oxidation experiments with "high iron loads" using either molecular oxygen and hydrogen peroxide as co-substrates.

	Mössbauer parameters of iron species			
	EncA		**EncAC**	
	A.Ia	**A.Ib**	**AC.Ia**	**AC.Ib**
δ (mm/s)	0.48(1)		0.48(1)	
ΔE_Q (mm/s)	0.58(2)	1.0(2)	0.57(2)	1.0(2)
Linewidth (mm/s)	0.34(2)	0.39(3)	0.33(2)	0.39(2)
	Quantification of iron species (%)			
Time (min)	**A.Ia**	**A.Ib**	**AC.Ia**	**AC.Ib**
1,800 Fe, O_2, 24 h	77.3	22.7	58.4	51.6
10,800 Fe, O_2, 10 min	73.7	27.3	72.5	27.5
10,800 Fe, O_2, 24 h	38.9*	15.3*	71.6	28.4
1,800 Fe, H_2O_2, 10 min	51.7	48.3	50.6	49.4
10,800 Fe, H_2O_2, 10 min	57.6	42.4	53.4	46.6
10,800 Fe, H_2O_2, 24 h	74.7	25.3	71.1	28.9

The values in parentheses are the uncertainties of each parameter

* The sextet signal corresponds to 45.7 % of the overall area

To further study the effect of both oxidants and/or if the presence of EncC encapsulated by EncA has an impact in the chemical nature of the mineral core formed, spectra were acquired at 4 K (Figure 5.4). At temperatures higher than the blocking temperature, an intrinsic characteristic of superparamagnetic ferric species (normally higher than 30 K [205]), the sextet signal collapses into quadrupole doublets due to the spin relaxation rate being faster than the resonance frequency of the ^{57}Fe nucleus, behaving as a paramagnet above its blocking temperature [204].

The spectra at 4 K (Figure 5.4) primarily show 6-line magnetic hyperfine patterns resulting from hyperfine splitting, typical of superparamagnetic ferric species. Holo-Ferritins have been characterized as ideal superparamagnets due to their organic shell, which keeps the ferrihydrite mineral cores apart from each other, guaranteeing that dipole-dipole interactions

can be ignored, at least above 1 K. As such, the mineral cores behave as isolated magnetic particles either to thermal excitation or application of an external magnetic field [206].

According to Figure 5.4, a similar property is observable in encapsulins, although the predicted size of its mineral core should be significantly larger than the Ferritin core. Notwithstanding, the formation of a large iron mineral core inside EncA still results in a superparamagnetic species.

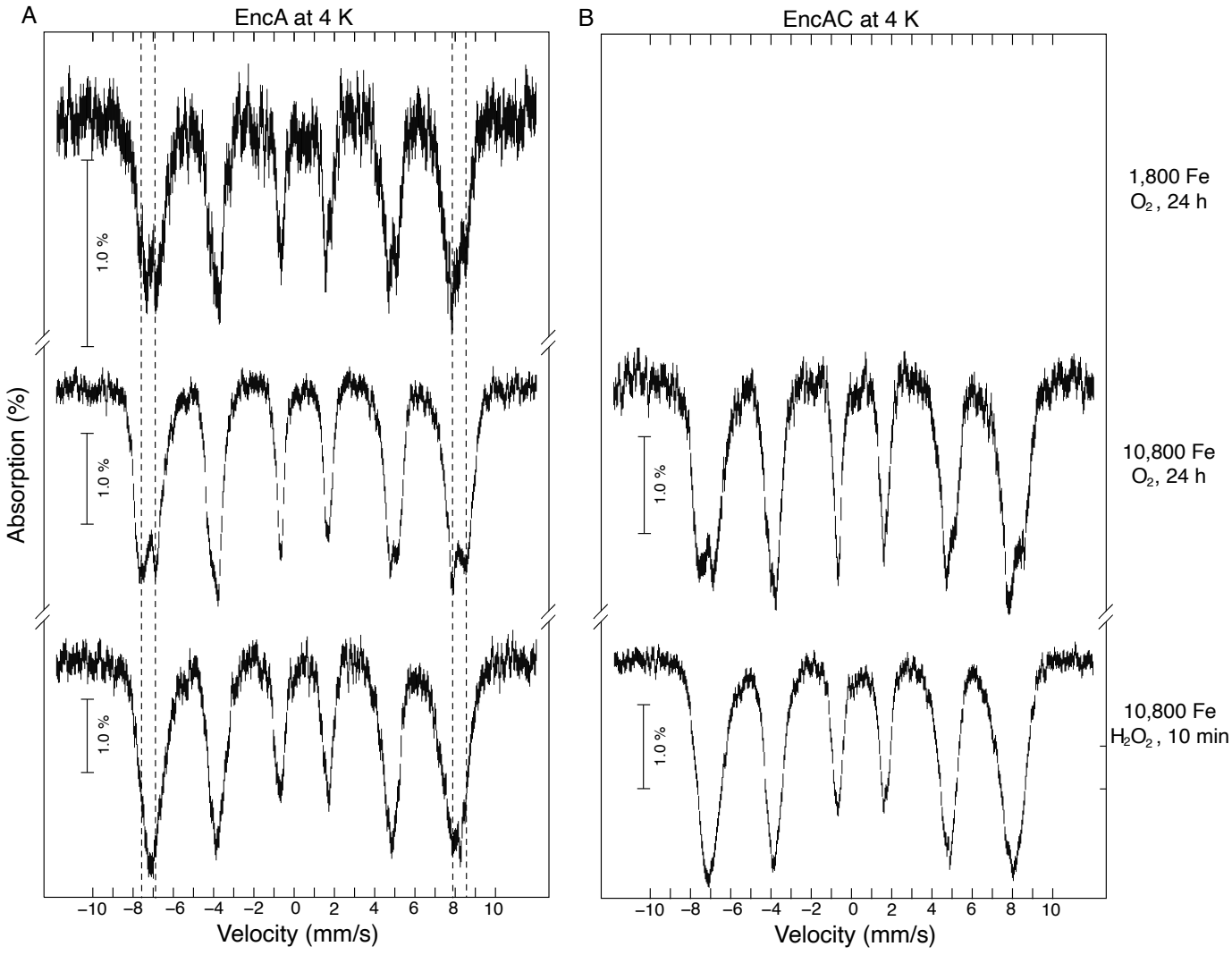

Figure 5.4: Evaluation of the superparamagnetic properties of the iron mineral core formed inside EncA and EncAC using molecular oxygen and hydrogen peroxide. (A) EncA was incubated with either 1,800 or 10,800 $^{57}Fe^{2+}$/protein and **(B)** EncAC was incubated with 10,800 $^{57}Fe^{2+}$/protein in the presence of molecular oxygen (24 h) and hydrogen peroxide (10 min), as displayed in the labels on the top and right side. The vertical dashed lines highlight the hyperfine splitting of the major superparamagnetic species from the oxidation of 10,800 Fe by EncA. Spectra were acquired at 4 K with no external magnetic field.

The spectra in Figure 5.4A with low loads (1,800 Fe/protein, top spectra) in the presence of oxygen show the heterogeneity of the ferrihydrite core with the contribution of sextets from two major species, resulting from different hyperfine splitting phenomena. The

addition of higher amounts of iron (10,800 Fe/protein, Figure 5.4A, middle row spectra) results in a spectrum with a splitting to lower energies consistent with the presence of a larger mineral core (in agreement with the spectra obtained in 55 K (Figure 5.3A).

Additionally, there are some noticeable differences between the mineral cores formed with molecular oxygen and hydrogen peroxide. In the presence of hydrogen peroxide the mineral core is more monodisperse, with a more well-defined sextet (Figure 5.4A, bottom spectra).

Altogether, these results suggest that the addition of higher amounts of ferrous iron substrate to EncA or EncAC leads to the formation of a larger mineral core. Based on these results, the iron oxidation reaction with molecular oxygen should favor the formation of larger heterogeneous mineral cores while the hydrogen peroxide seems to form smaller mineral cores centers distributed along the internal lining of the encapsulin shell with a more homogeneous environment.

The presence of EncC inside EncA does not seem to change the nature of the ferric mineral (Figure 5.4B), nor the superparamagnetic characteristic of the ferrihydrite mineral. In this case, the Fe^{2+} ions are rapidly oxidized at the FOCs and then transferred to a nucleation site at the EncA inner surface where the mineral is formed independently of the presence of the cargo protein.

5.2.4 Iron oxidation species

The iron oxidation kinetics assay monitored using UV-Visible spectroscopy revealed noticeably distinct oxidation rates between EncA and EncAC upon addition of lower amounts of iron substrate (lower than 1,800 Fe/protein). In these conditions, the encapsulation of EncC led to a significant enhancement of the rate of ferroxidation (Figure 5.2B). Mössbauer spectra were acquired at 80 K in the absence of an external magnetic field to detect and characterize all iron species involved in the reactions catalyzed by EncA and EncAC in the presence of molecular oxygen (Figure 5.5 and Table 5.2).

Contrary to the samples reacted with higher iron loads (over 1,800 Fe/protein), the spectra of the samples incubated with 90, 180 and 360 $^{57}Fe^{2+}$/protein (with either EncA or EncAC) and O_2 for 15 min show the presence of unreacted ferrous iron in solution (Figure 5.5). The high spin ferrous iron species appears as a quadrupole doublet that is better fitted with three subspectra with the same δ and different ΔE_Q. These parameters are quite similar between EncA and EncAC (see Table 5.2). The resemblance between the EncA and EncAC parameters may indicate that the ferrous iron is mainly on the surface of the shell protein or that Fe^{2+} binding to EncC occurs in a very similar chemical environment.

The conversion of ferrous into ferric iron in the presence of molecular oxygen by both proteins is confirmed with the appearance of a quadrupole doublet with parameters typical of high-spin Fe^{3+} species with oxygen/nitrogen ligands, as previously mentioned. Upon the addition of 90 or 180 Fe^{2+}/EncA only one ferric species was observed (Figure 5.5A, first and

second row). The spectra of EncA with 360 Fe^{2+}/EncA present a new set of ferric iron parameters (species A.Id) was included due to the broadening of the signal (Figure 5.5A, third row). The EncAC with the lower ratio of 90 Fe^{2+}/EncAC both species were needed for a correct fitting. This is justified due to the formation of more product during the same time comparing with the EncA.

The two sets of ferric iron parameters (A.Ic vs AC.Ic and A.Id vs AC.Id) were also similar between EncA and the EncAC complex (see Table 5.2). Although species A.Ia and A.Ic seems to be quite similar (as well as AC.Ia and AC.Ic) (see Table 5.1 and Table 5.2), species A.Ib and A.Id (as well as AC.Id and AC.Ib) are reasonably dissimilar. This difference may indicate the presence of an intermediate species (A.Id and AC.Id) only detected during the low loading experiments.

The iron oxidation reaction kinetics of empty EncA is remarkably different to the EncAC complex, as shown by UV-Visible spectroscopy (Figure 5.2B). Once again, it is possible to detect an increase in reaction rate when the cargo protein is present (Table 5.2), showing a higher ferric iron fraction for the same reaction time of when compared with EncA revealing the functional impact of this cargo protein in the overall encapsulin system.

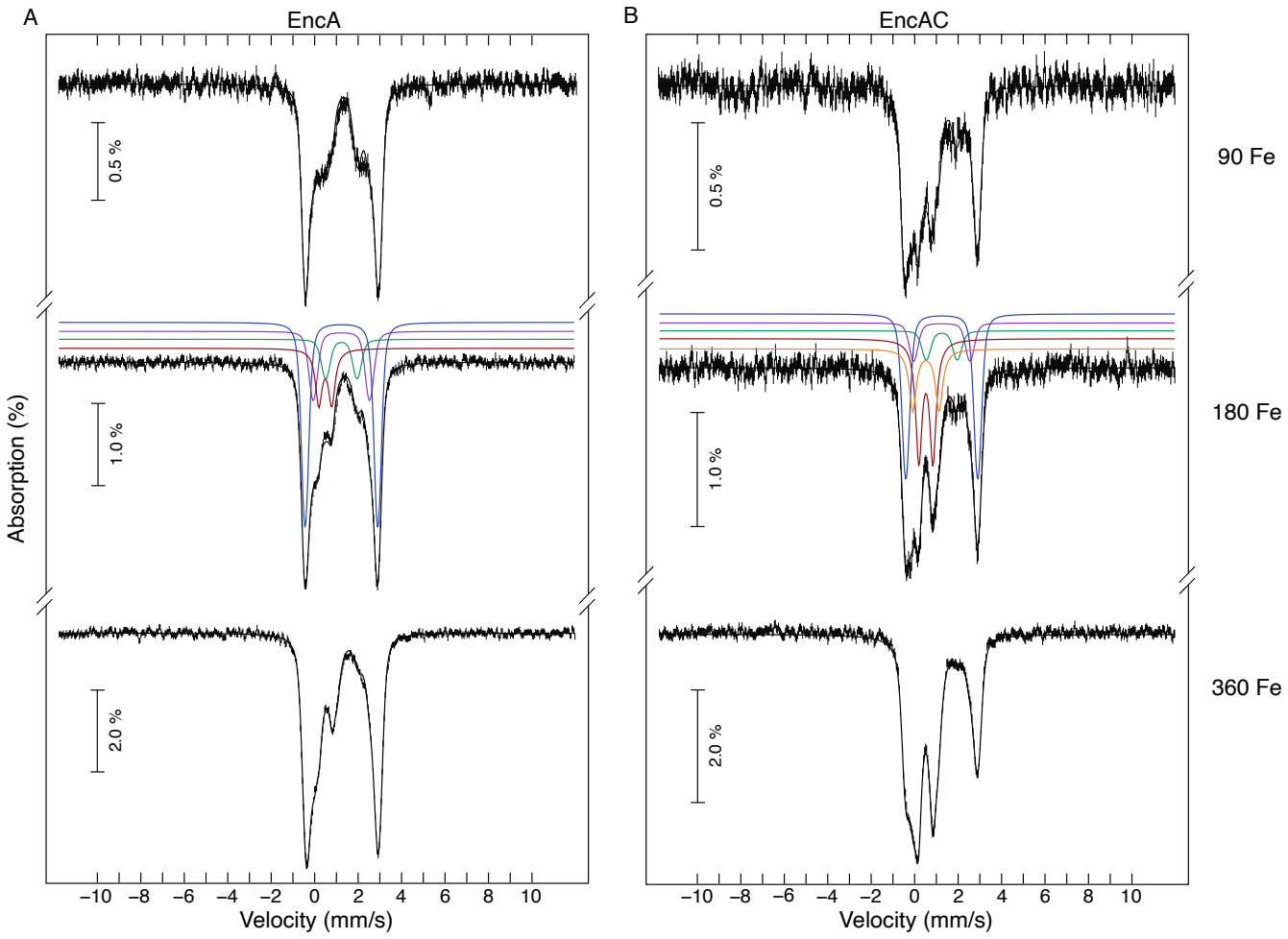

Figure 5.5: Mössbauer spectra of low iron loads of EncA and EncAC in the presence of molecular oxygen for 15 min. (A) EncA, **(B)** EncAC were incubated with either 90, 180 or 360 $^{57}Fe^{2+}$/protein for 15 min in the presence of

molecular oxygen as displayed in the labels on the top and right side. The solid black lines represent the theoretical spectra reconstucted using the parameters from Table 5.2: Species A.IIa and AC.IIa in purple line, Species A.IIb and AC.IIb in green line, Species A.IIc and AC.IIc in blue line, Species A.Ic and AC.Ic in red line and Species A.Id and AC.Id in orange line. The spectra were recorded at 80 K with no external magnetic field.

Table 5.2: Mössbauer parameters resulting from the deconvolution of the spectra of EncA and EncAC iron oxidation samples with low iron loads (90, 180 and 360 Fe^{2+}/ protein) and molecular oxygen, for 15 min.

	EncA Mössbauer parameters of iron species				
	Ferrous species			**Ferric species**	
	A.IIa	**A.IIb**	**A.IIc**	**A.Ic**	**A.Id**
δ (mm/s)		1.26(3)		0.54(3)	
ΔE_Q (mm/s)	2.58(3)	1.46(3)	3.35(2)	0.67(1)	1.20
Linewidth (mm/s)	0.45(2)	0.45(3)	0.37(2)	0.39(2)	0.40
	EncA Quantification of iron species (%)				
Sample					
90 Fe, O_2, 15 min	19.6	21.9	51.4	7.1	
180 Fe, O_2, 15 min	19.2	12.3	52.6	15.9	
360 Fe, O_2, 15 min	15.9	6.8	51.0	19.4	6.9

	EncAC Mössbauer parameters of iron species				
	Ferrous species			**Ferric species**	
	AC.IIa	**AC.IIb**	**AC.IIc**	**AC.Ic**	**AC.Ic**
δ (mm/s)		1.25(2)		0.51(1)	
ΔE_Q (mm/s)	2.58(3)	1.43(2)	3.3(2)	0.65(2)	1.2(3)
Linewidth (mm/s)	0.45(2)	0.45(2)	0.38(1)	0.38(3)	0.40(2)
	EncAC Quantification of iron species (%)				
Sample					
90 Fe, O_2, 15 min	7.5	12.3	36.9	26.9	16.4
180 Fe, O_2, 15 min	9.8	7.4	38.6	28.0	16.2
360 Fe, O_2, 15 min	9.4	4.6	29.0	38.7	18.3

The values in parentheses are the uncertainties of each parameter

5.2.5 EncC Iron Oxidation Species

Mössbauer spectra of EncC were measured to further characterize the ferroxidation mechanism (Figure 5.6). The protein was incubated with different equivalents of $^{57}FeSO_4$ under anaerobic conditions and in oxygen saturation conditions for 5 min.

EncC was incubated with the iron substrate inside an anaerobic chamber to characterize the binding sites of ferrous iron species. As expected, the spectra (Figure 5.6A) show a relatively symmetric quadrupole doublet that can be fitted with three sets of parameters characteristic of high-spin ferrous species as listed in Table 5.3. Two doublets have a similar δ parameters and distinct ΔE_Q (Species C.Ia, blue line and species C.Ib: green line, see Table 5.3) showing a similar environment to ferrous iron in 200 mM MOPS pH 7.0 buffer containing 200 mM NaCl (Figure A.9, appendix). The third species (C.Ic, purple line) with an $\delta = 1.23 \pm 0.01$ mm/s and a $\Delta E_Q = 3.19 \pm 0.03$ mm/s probably represents a specific binding site in the protein (absent in EncA and EncAC spectra), and are similar to Fe^{2+} atoms coordinated by oxygen/nitrogen ligands from glutamate and histidine residues found in the FOC of Ferritin protein [207]. In the sample incubated with 2 Fe^{2+}/FOC under anaerobic condition, this species accounts for 50 % of the total iron absorption, representing 1 Fe^{2+} per FOC. Reaction of the protein with an excess of Fe^{2+} ions, 7 Fe^{2+}/FOC, saturates the FOC (26.3 % of the total absorption, equivalent to 1.9 Fe^{2+} per FOC) (Table 5.3).

Under saturation of molecular oxygen (Figure 5.6B), the signal characteristic of the ferrous protein-bound species is converted into a subspectrum typical of ferric species, fitted with two quadrupole doublets sharing the same isomer shift value (Species C.IIa, dark red line and species C.IIb: light red line, see Table 5.3).

In the spectra of EncC samples reacted with 2 or more Fe^{2+}/FOC, the appearance of a broader spectrum (at 80 K) can only be fitted with an additional parameter set, with $\delta = 0.49 \pm 0.02$ mm/s and $\Delta E_Q = 1.61 \pm 0.02$ mm/s (species C.IIc, orange line). This species is similar to the a μ-1,2-peroxo di-ferric intermediate species identified in H-type ferritins [208].

No superparamagnetic species were detected at 4 K upon ferroxidation of 2 Fe^{2+} per FOC (Figure 5.6C). However, a sextet is detected with 7 Fe^{2+} per FOC confirming the presence of a ferric iron species with superparamagnetic properties that is no longer in fast relaxation (quadrupole doublet with $\delta = 0.49 \pm 0.02$ mm/s and $\Delta E_Q = 0.66 \pm 0.01$ mm/s, species C.IIa, dark red line in Figure 5.6), and is due to the formation of a small mineral core (Figure 5.6C).

At the stoichiometry of 10 Fe^{2+} per dimer the μ-1,2-peroxo di-ferric intermediate is no longer detected, most probably due to its low contribution to the overall spectra. Additionally, this species was absent in the spectra of the EncAC samples (Figure 5.5B).

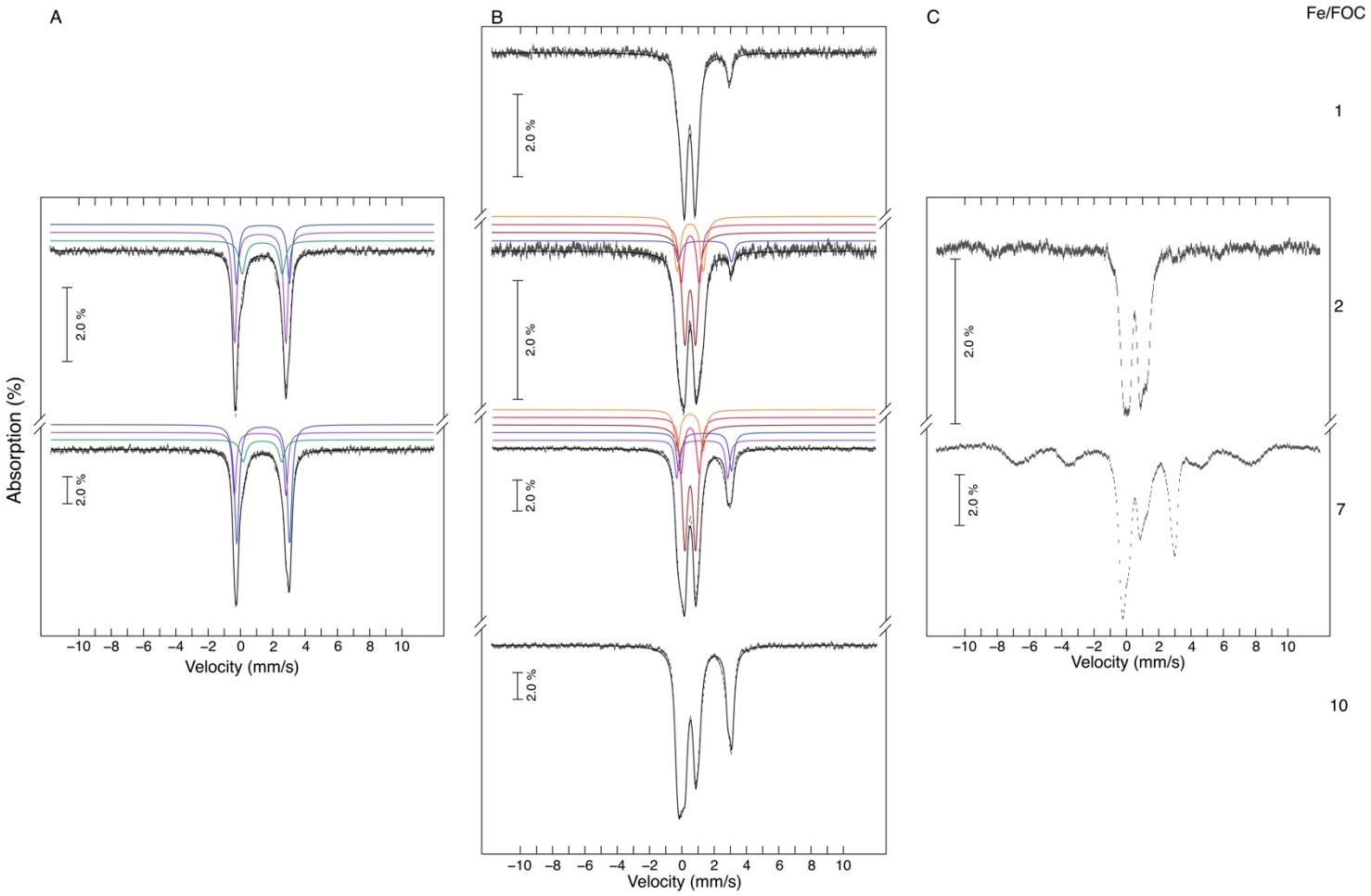

Figure 5.6: Mössbauer spectra of EncC reacted with Fe^{2+} ions in anaerobic conditions and in the presence of molecular oxygen, for 5 min. (A) Spectra of the EncC anaerobic iron loading (2 and 7 Fe^{2+}/FOC) in anaerobic conditions. **(B and C)** EncC incubated with 1, 2, 7 and 10 Fe^{2+}/FOC in the presence of molecular oxygen during 5 min and acquired at **(B)** 80 K and **(C)** 4 K, as indicated in the labels on the top and right side. The solid black lines represent the theoretical spectra reconstructed using the parameters from Table 5.3: Species C.Ia in blue line, Species C.Ib in green line, Species C.Ic in purple line, Species C.IIa in dark red line, Species C.IIb in red line and Species C.IIc in orange line. The spectra were recorded at 80 K and 4 K, with no external magnetic field.

Table 5.3: Mössbauer parameters of the iron species detected in EncC samples reacted with Fe^{2+} ions in anaerobic conditions (2 and 7 Fe^{2+}/dimer) and with oxygen saturation (1, 2, 7 and 10 Fe^{2+}/dimer), at 80 K.

	EncC Mössbauer parameters of iron species					
	Ferrous species			**Ferric species**		
	C.Ia	**C.Ib**	**C.Ic**	**C.IIa**	**C.IIb**	**C.IIc**
δ (mm/s)	1.41(1)		1.23(1)		0.49(2)	
ΔE_Q (mm/s)	3.28(2)	2.44(4)	3.19(3)	0.66(1)	1.10(2)	1.61(2)
Linewidth (mm/s)	0.34(3)	0.45(3)	0.30(2)	0.40(1)	0.36(2)	0.42(3)
	EncC Quantification of iron species (%)					
Sample						
2 Fe/FOC anaerobic conditions	27.3	22.2	50.5 [1.0]			
7 Fe/FOC anaerobic conditions	59.1	14.6	26.3 [1.9]			
1 Fe/FOC oxygen saturation	6.9		6.8 [0.07]	71.3	15.0	
2 Fe/FOC oxygen saturation	8.1			44.8	22.3	24.8 [0.5]
7 Fe/FOC oxygen saturation	12.6		11.0 [0.77]	42.9	18.7	14.8 [1.0]
10 Fe/FOC oxygen saturation	28.8		14.6 [1.46]	38.1	18.5	

The values in parentheses are the uncertainties of each parameter.

5.3 Discussion and Conclusion

The proteins used in this study were purified in their apo forms and thus considered ideal to investigate the iron oxidation mechanism of the encapsulin system.

To the best of our knowledge, we demonstrate for the first time the ability of an empty encapsulin nanocage (without any cargo protein) to incorporate Fe^{2+} ions. The ferrous ions are oxidized to form a ferric iron mineral, both in the presence of molecular oxygen or hydrogen peroxide. Using hydrogen peroxide the recombinant *Myx. xanthus* EncA can uptake a total of 19,500 irons per protein. In the presence of molecular oxygen, the maximum capacity of EncA increases up to 24,000 iron atoms per shell. These loading capacities are relatively close to those reported in the literature for the encapsulin-cargo protein complexes (EncABCD), ranging on average from 22,500 iron atoms to 30,000 iron atoms for the highest load possible per protein shell [21].

The progress curves of the iron oxidation reaction by EncA follows a sigmoidal curve similar to L-chain Ferritins (in which the FOC site is absent, and the auto-catalytic properties of the mineral core mediate the process). At lower iron ratios (1,800 and 3,600 Fe/EncA) the lag-phase is more extensive and is justified by the size of the mineral core being below the threshold for initiating the auto-catalytic reaction.

The iron loading assays using the EncAC and EncAB complexes with O_2 displayed loading capacities similar to EncA, but lower capacities when H_2O_2 was used as co-substrate, being able to uptake 18,000 iron atoms per complex versus 19,500 for EncA.

The kinetics of the reactions catalyzed by EncAB and EncAC with larger amounts of Fe^{2+} ions (higher than 1,800 Fe/protein) presented similar profiles, with shorter lag-phases and the consumption of the substrate is achieved sooner relatively to the free EncA.

For lower loads (below 1,800 Fe/protein), the progress curves were no longer sigmoidal and exponential curves were observed for the EncAB or EncAC complexes. This behavior is similar to the H-type ferritins and encapsulated Flp from *Rhs. rubrum* [56]. The EncAB complex exhibited a faster iron oxidation rate, reaching a plateau after ~ 15 min, while EncAC absorbance stabilization was observed after ~ 90 min. Although it was not possible to study the iron oxidation reaction of free EncB due to protein precipitation, it is reasonable to infer its catalytic properties by evaluating the activity of the EncAB complex in comparison with EncA and/or EncAC. A possible explanation for the faster oxidation activity of the EncAB complex (when compared to EncAC) may be the electrostatic charge of the surface of both cargo proteins. Looking at the atomic structure of EncB and EncC (PDB: 7S5K and 7S8T), the surface of EncC appears to be mostly negatively charged (Figure A.10A, appendix) while EncB appears to be more neutral. However, a considerable number of residues in both the N- and C- terminal regions of the EncB are absent in the structure. The residues in the N-terminal region are responsible for the formation of the Flp central pore and thus may affect its functional properties. The structure prediction of the missing residues using AlphaFold (Figure A.10B, appendix) suggests that the EncB internal pore is noticeably more negative, and thus may be more efficient in attracting ferrous irons towards to the FOC sites.

Free EncC was able to uptake up to 10 iron atoms per FOC (using hydrogen peroxide as co-substrate) until subsequent protein precipitation and accumulation of ferric iron in solution. This lower iron oxidation capacity is in agreement with other Flp from other organisms for which the formation of iron oxide aggregates were described after the addition of 10 – 15 Fe^{2+} per FOC [20,48]. The ferroxidation kinetics of EncC showed that the free protein in solution is more efficient (faster oxidation activity) when compared to its encapsulated form, indicating that the rate limiting step of reaction is the diffusion of iron towards the negative pores of EncA and translocation to the EncC FOCs.

The Mössbauer spectra of EncA and EncAC loaded with higher amounts of ^{57}Fe (over 1,800 Fe/protein) showed only ferric species, indicating an increased rate of iron oxidation and mineralization with higher protein and ferrous iron substrate concentrations, when compared with the UV-Visible assays. The ferric species formed inside EncA and EncAC have similar parameters characteristic of a high spin Fe^{3+} species coordinated by oxygen/nitrogen ligands. The spectra at 4 K showed a sextet confirming the superparamagnetic properties of the mineral core formed within the shell. Iron oxidation with molecular oxygen favors the formation of larger heterogeneous mineral cores while the reaction with hydrogen peroxide forms smaller, more homogenous mineral cores, most likely distributed over the internal lining of the protein nanocage, with similar chemical environments. However, no differences were detected between the mineral core formed by EncA and EncAC, suggesting that the mineralization step is mainly performed by the shell protein.

On the contrary, the EncA and EncAC samples with lower iron loads reacted for 10 min in the presence of O_2 contained unreacted ferrous iron species with similar Mössbauer parameters between EncA and EncAC, suggesting that these ferrous ions are coordinated with residues from the shell protein. The oxidation reaction generates high-spin Fe^{3+} species with similar parameters in both proteins. Despite the similarities in the parameters of the ferric iron species, the relative percentages of ferric species formed during the same reaction time and the same iron substrate concentration are distinct. It is clear that the presence of EncC inside the protein shell increases the ferroxidation rate and therefore the formation of ferric mineral core in agreement with the kinetic data obtained by UV-Visible spectroscopy.

In anerobic conditions, the ferrous species bound to the free EncC present a distinct Mössbauer parameters from the ones detected for the ferrous iron substrate in buffer or the ones detected in the surface of the shell protein. In fact, the parameters are similar to other ferrous species described in Ferritin proteins. However, these species were not detected when EncC is encapsulated within EncA, suggesting that in this case the ferrous ions are preferentially bound to the shell protein.

Moreover, an intermediate was detected in EncC in the presence of molecular oxygen, with parameters similar to the μ-1,2-peroxo di-ferric species also described in Ferritins [208]. The atomic structure of EncC shows that the iron binding centers are close enough to allow a bridge between two irons through oxygen atoms. Eren *et al*. hypothesized that the formation of a peroxo di-ferric intermediate would be followed by the formation of two Fe^{4+} intermediates due to the presence of two tyrosine residues in the vicinity which would act as

electron and H^+ donors [30]. However, all spectra were fitted without the need to include any subspectrum with parameters typical of Fe^{4+} species, either due to its absence or due to a putative low lifetime of the intermediate, preventing its accumulation and detection.

The μ-1,2-peroxo di-ferric intermediate was also not detected in the EncAC complex. Either this intermediate was not detectable in the spectra due to the presence the other species in the sample or the encapsulation of EncC inside EncA decreases the lifetime of this intermediate species, hindering its detection.

6

CONDENSATION AND PROTECTION OF DNA BY *MYXOCOCCUS XANTHUS* ENCAPSULIN: A NOVEL FUNCTION

The function of encapsulin systems is solely related to their cargo proteins. In this section we describe the binding and protection of circular double stranded DNA (pUC19) by *Myxococcus xanthus* EncA using Electrophoretic Mobility Shift Essays (EMSA), Atomic Force Microscopy (AFM) and DNase protection assays.

The interaction between EncA, EncC and EncAC with pUC19 was evaluated and the binding parameters determined by EMSA. The morphology of the complex formed between EncA and pUC19 was unveiled using AFM. Additionally, the protective effect of EncA binding to plasmid DNA was tested in the presence of DNase I. Moreover, the secondary structure and thermal stability of EncA was monitored using SRCD and the changes upon DNA interaction were also evaluated.

This chapter was adapted from the following manuscript, published during this thesis: Ana V. Almeida, Ana J. Carvalho, Tomás Calmeiro, Nykola C. Jones, Søren V. Hoffmann, Elvira Fortunato, Alice S. Pereira, Pedro Tavares. "Condensation and protection of DNA by the *Myxococcus xanthus* encapsulin: a novel function" Int. J. Mol. Sci. 2022, 23,7829. DOI: 10.3390/ijms23147829

6.1 Experimental Procedure

6.1.1 pUC19 Production and Purification

The pUC19 plasmid (Figure A.11, appendix) was produced using *Esc. coli* NZY5α (NZYTech) cells and purified with the NZYMidiprep kit (NZYTech). Plasmid isolation was performed according to the manufacturer's instructions [209]. After purification the concentration of pUC19 was determined using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific) and the absorbance at 260 nm. The average molar extinction coefficient for double-stranded DNA was considered to be 0.020 ng/μL^{-1} cm^{-1}.

6.1.2 Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSAs) were performed using supercoiled pUC19 plasmid in 50 mM MOPS buffer, pH 7.0, 50 mM NaCl (designated as low ionic strength buffer) and in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl (termed as high ionic strength buffer). Proteins and pUC19 were dialyzed against each binding buffer for 30 min using a Slide-A-Lyzer mini-device (ThermoFisher Scientific) immediately before each binding reaction.

A serial dilution set of EncA and EncAC samples, between 1.86 and 0.02 μM, and monomeric EncC between 336 and 3.88 μM were incubated with 5 nM of pUC19 in a total volume of 20 μL, using both high and low ionic strength buffers, for 30 min at room temperature. 1 μL of GreenSafe stain (NZYTech) was added to each sample and 10 μL were loaded into a 1 % agarose gel in TAE buffer (40 mM Tris-acetate buffer, pH 8.0, 0.1 mM EDTA), run at 80 V for about 1 h at room temperature. The data was obtained at least in triplicates. The gel was imaged using a Safe Imager™ transilluminator (Invitrogen) and a Gel Logic 100™ Imaging System (Kodak). The relative amounts of free and protein–DNA complexes were quantified by measuring the disappearance of the supercoiled conformation using Fiji/ImageJ [183]. The formation of the protein–DNA complexes was plotted against protein concentration and fitted with a Hill equation:

$$f = \frac{f_{max}\ [Protein]^n}{K_D + [Protein]^n} \qquad \text{Equation 7}$$

where f is the fraction of saturation, f_{max} corresponds to 100 % complex formation, [Protein] is the concentration of protein available to bind, n is the Hill coefficient and K_D is the macroscopic apparent dissociation constant and corresponds to a measure of the affinity of the protein to the DNA.

6.1.3 DNase I Protection Assay

Controlled digestion assays were performed using EncA (1.24 μM) pre-incubated with supercoiled pUC19 (5 nM) in both low and high ionic strength buffers. After the 30 min pre-incubation with DNA, 1.2 μg/μL of DNase I (Sigma) and 1 mM $MgCl_2$ were added to the samples. The reaction was stopped at different reaction times (0.5, 1, 15, and 30 min) by adding 5 mM EDTA. The same procedure was performed for free pUC19 samples. The samples were stained with GreenSafe and analyzed by electrophoresis using 1 % agarose gels in TAE buffer.

6.1.4 Atomic Force Microscopy Imaging

EncA and EncAC (both free and incubated with pUC19 at a stoichiometry of 1:1 (1.5 nM) at room temperature during 30 min in 50 mM MOPS buffer, pH 7.0 and 50 mM NaCl containing 5 mM $NiCl_2$) were imaged using Atomic Force Microscopy (AFM) imaging. The samples were deposited onto a fresh cleaved grade V1 muscovite mica (Ted Pella, Inc), incubated for 10 min, washed with deionized water, and dried with a weak stream of nitrogen gas. AFM images were collected in air, at room temperature and atmospheric pressure, using an Asylum Research MFP-3D Standalone (Oxford Instruments) operated in alternate contact mode with commercial silicon cantilevers (Olympus AC160TS, Olympus Corporation, f0 = 300 kHz; k = 26 N/m). Images were processed using the Gwyddion modular program.

6.1.5 Synchrotron Radiation Circular Dichroism

Synchrotron Radiation Circular Dichroism (SRCD) spectra were recorded on the AU-CD beam line at the ASTRID2 synchrotron radiation source (ISA, Aarhus University, Denmark) as described in section 4.1.3. Samples contained EncA at 1.0 mg/mL both with and without pUC19 in a 1:1 molar equivalent stoichiometry, prepared as described before in 10 mM MOPS buffer, pH 7.0 and 240 mM NaF.

6.2 Results

6.2.1 Characterization of the DNA Binding Properties of Encapsulins

The ability of encapsulins to bind circular double stranded plasmid DNA was evaluated using Electrophoretic Mobility Shift Assays (EMSA) (Figure 6.1).

The EMSAs were performed using the supercoiled form of the pUC19 plasmid (the electrophoretic pattern of the pUC19 in different conformations is presented in Figure A.11, appendix in low (50 mM MOPS buffer, pH 7.0 with 50 mM NaCl) ionic strength conditions.

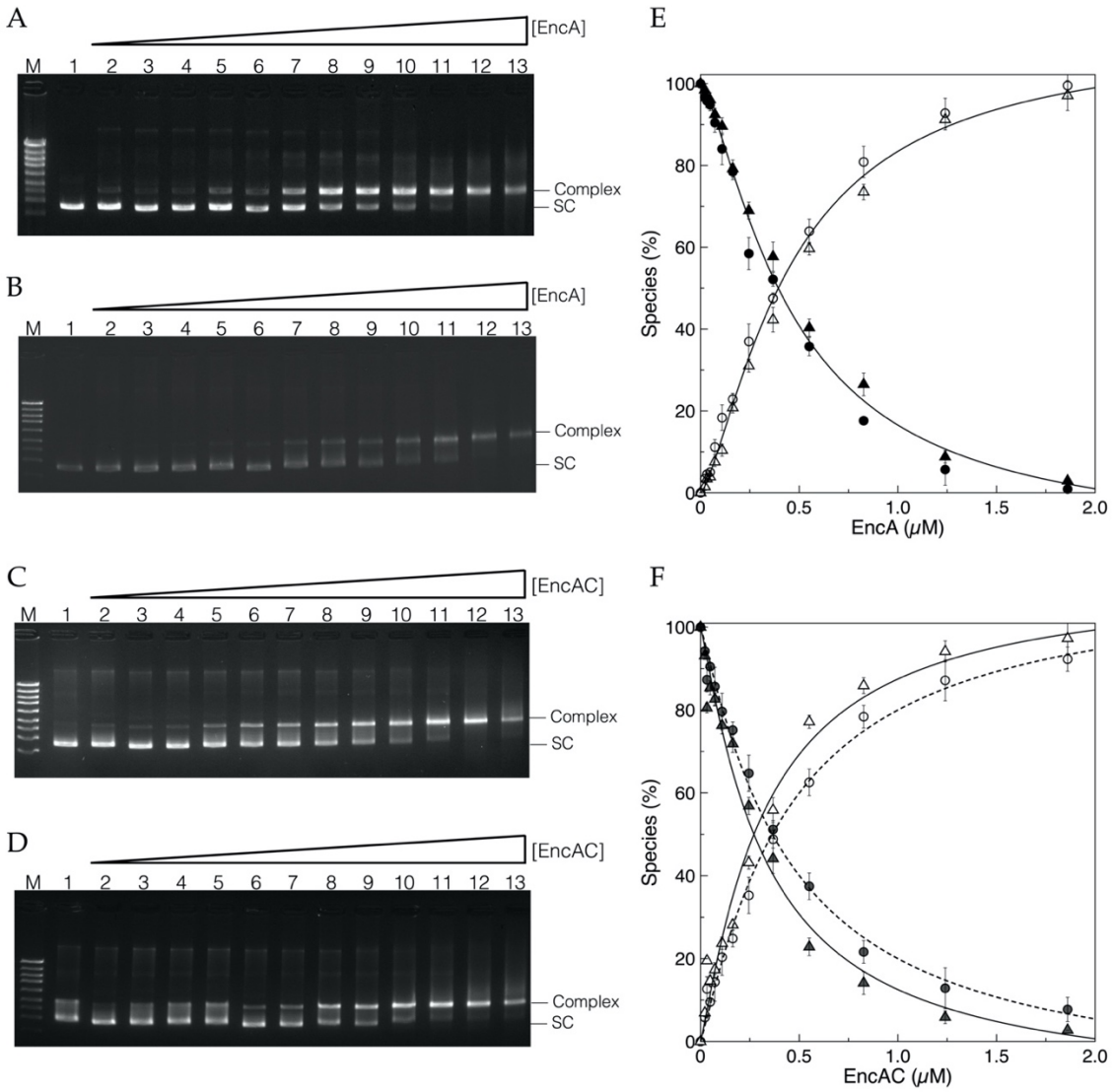

Figure 6.1: Binding of encapsulin to supercoiled pUC19 (5 nM) in high and low ionic strength conditions. (A) EMSA of EncA in 50 mM MOPS buffer, pH 7.0, 50 mM NaCl and **(B)** 200 mM MOPS buffer pH 7.0, 200 mM NaCl. **(C)** EMSA of EncAC complex in 50 mM MOPS buffer, pH 7.0, 50 mM NaCl and **(D)** 200 mM MOPS buffer pH 7.0, 200 mM NaCl. 1 to 13 – Binding of EncA and EncAC to the plasmid DNA (5 nM) with increasing protein concentrations: 0, 0.02, 0.03, 0.05, 0.07, 0.11, 0.16, 0.24, 0.37, 0.55, 0.83, 1.24 and 1.86 μM. The free form of the supercoiled plasmid pUC19 (SC) and the protein–pUC19 complexes bands (Complex) are labelled. M – NZYLadder II (Figure A.11C, appendix). **(E)** Hill plots of DNA binding to EncA and **(F)** EncAC from densitometric

analysis of three sets of experiments in either 50 mM MOPS buffer pH 7.0, 50 mM NaCl (circles) and 200 mM MOPS buffer, pH 7.0, 200 mM NaCl (triangles). Free pUC19 SC is plotted as full (and the complex as empty) circles or triangles. The lines on top of the experimental data are theoretical fits using Equation 7. For the binding of EncA to pUC19 a single set of parameters, K_D and n, were used for the theoretical fit, whereas for the EncAC complex two sets of parameters were employed for the binding reaction in low (dashed lines) and high (full lines) ionic strength conditions.

The results indicate that EncA is capable of binding to the supercoiled form of pUC19. Binding of EncA to plasmid DNA leads to the formation of EncA–pUC19 complexes that migrate slower when comparing with the free supercoiled form (Figure 6.1A and B). Increasing the protein concentration upshift the supercoiled pUC19 band to the relaxed form of the plasmid, with an apparent saturation at around 2 μM of EncA. Furthermore, at protein concentrations higher than 0.83 μM, a smear with smaller electrophoretic mobility is noticed. Additionally, the formation of EncA–pUC19 complexes was not affected by increasing the ionic strength in the binding reaction (Figure 6.1B).

At both ionic strengths, the formation of the complex was plotted and fitted with the Hill equation (Figure 6.1E). Under these conditions, the apparent dissociation constant (K_D) was 0.3 ± 0.1 μM for the low and high ionic strength conditions with a Hill coefficient (n) of 1.4 $\pm$ 0.1, indicating a moderate positive cooperativity on DNA binding, similar to the ones reported for proteins from the Ferritin family, such as *Des. vulgaris* Dps-like bacterioferritin ($n = 1.3 \pm 0.2$), Dps proteins from *Mar. hydrocarbonoclasticus* ($n = 1.2 \pm 0.1$), or *Deinococcus radiodurans* ($n = 1.3 \pm 0.2$) binding to supercoiled plasmid DNA [110,207,210].

Contrary to EncA, the cargo protein alone (EncC) was not able to bind to supercoiled pUC19 (Figure A.12, appendix). Yet, the DNA binding capacity of the EncAC complex was comparable to EncA, leading to the formation of EncAC–pUC19 complexes with similar electrophoretic migration profiles (Figure 6.1C and D). As for EncA, the interaction between the EncAC complex and supercoiled pUC19 was not affected by the increase of the ionic strength (Figure 6.1D), resulting in a K_D of 0.4 ± 0.1 μM and 0.3 ± 0.1 μM in both low and high ionic strength buffers and a n value of 1.1 ± 0.1 for both buffer conditions (Figure 6.1F). As such, the interaction with DNA seems specific to EncA and not affected by the presence of EncC cargo protein inside its cavity. The pUC19 band shift observed in the EMSA assays from the supercoiled conformation to a form that migrates slower in the gel when incubated with the encapsulin proteins can either be explained by the relaxation of the plasmid DNA topology and/or to a mechanism of DNA condensation associated with protein self-aggregation. Both processes have been previously described for Dps protein nanocages binding to DNA (see section 1.3.2.1) [107].

The topology of the protein–DNA complexes was characterized by AFM and compared with the free forms of the proteins and plasmid DNA (Figure 6.2). The micrographs in Figure 6.2A show that pUC19 is predominantly in the supercoiled form, with strands crossing (bright spots on the DNA molecule). Free EncA (Figure 6.2B) displays a uniform spherical shape, with a diameter around 35 nm, in agreement with the diameter determined using SAXS (section 4.2.1) and with its atomic structure (PDB: 4PT2 and 7S20) [11,21]. The images of the EncAC complex show larger particles, probably owing to a broadening effect

due to AFM tip-sample convolution effects (Figure 6.2C). Binding of both EncA and EncAC proteins to the plasmid DNA caused the formation of larger protein–DNA condensates (Figure 6.2D and E). In the presence of DNA most protein molecules appear as very large aggregates surrounded by DNA loops. In some images, mostly for the EncAC–pUC19 complexes, the spherical protein particles are randomly bound to the DNA molecules (Figure 6.2E, left image) with a "beads-on-a string" morphology.

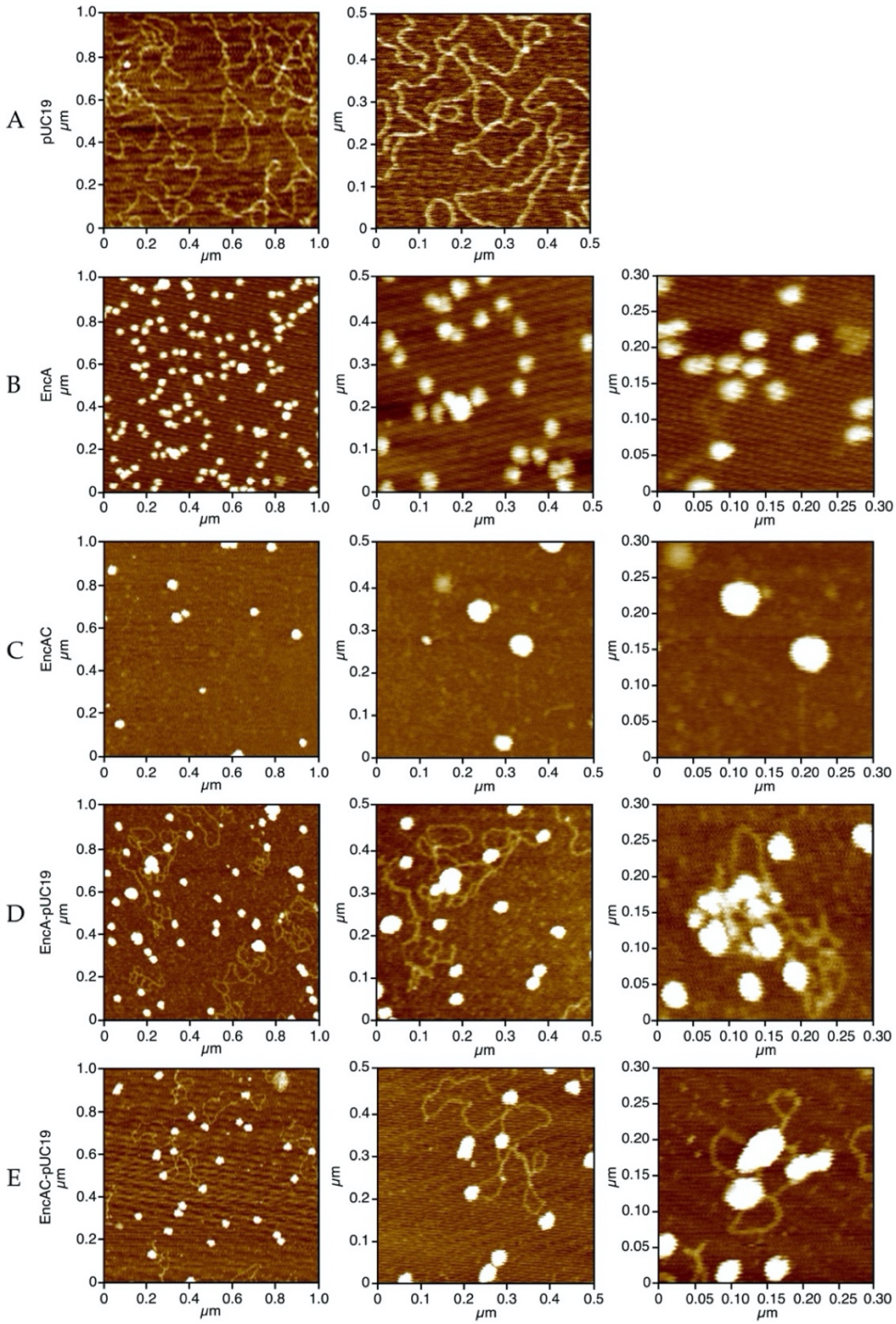

Figure 6.2: AFM imaging of encapsulin binding to supercoiled DNA. (A) pUC19; **(B)** EncA; **(C)** EncAC; **(D)** EncA–pUC19 complexes; **(E)** EncAC–pUC19 complexes. All samples were buffered in 50 mM MOPS, pH 7.0 containing 50 mM NaCl and 5 mM $NiCl_2$.

6.2.2 DNA protection assays

The controlled digestion of pUC19 with DNase I was tested in different conditions to probe the protective effect of encapsulins (Figure 6.3). The incubation of free pUC19 with DNase I leads to its complete hydrolysis within less than 1 min, regardless of the ionic strength of the buffer used in the reaction. When pUC19 was pre-incubated with 1.24 μM of EncA, and subsequently digested with DNase I, approximately 15 % of the supercoiled DNA was hydrolyzed after 1 min in low ionic strength (Figure 6.3A) and 35 – 40 % at high ionic strength (Figure 6.3B). For longer reaction times (30 min) the intensity of pUC19 band decreased by around 55 % at low ionic strength, indicating a considerable degree of digestion, whereas no further hydrolysis was observed at high ionic strength. These results demonstrate that the binding of encapsulin to the pUC19 molecules physically shields the DNA from DNase I enzymatic digestion. The different results obtained for both buffer conditions may be explained the level of looping and compaction of the protein–DNA condensates. The present data suggests that in the experimental conditions tested, higher concentrations of salts favor more compact aggregates, which might be physiologically relevant as part of a response mechanism in salt stressful conditions.

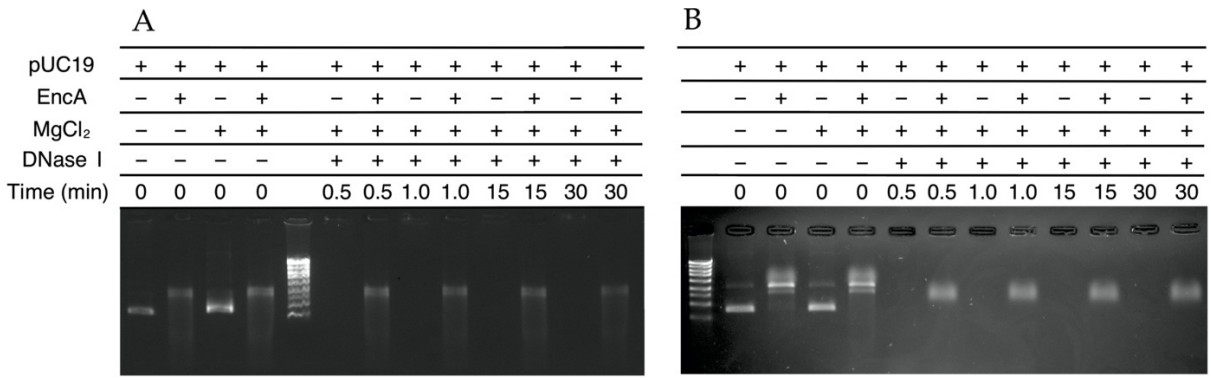

Figure 6.3: Protective effect of the EncA–pUC19 complex against DNase I digestion. Controlled digestion **(A)** in 50 mM MOPS, pH 7.0, 50 mM NaCl and **(B)** in 200 mM MOPS, pH 7.0, 200 mM NaCl. pUC19 (5 nM), either free or pre-incubated with EncA (1.24 μM), in the presence of 5 mM $MgCl_2$ was reacted with 1.2 μg/μL of DNase I at room temperature. The digestion was monitored after 0.5, 1.0, 15 and 30 min of hydrolysis reaction. The composition of each reaction is described above the gel. Marker: NZYLadder II (Figure A.11C, appendix).

6.2.3 Secondary structure assessment and thermal stability

The AFM dataset presented above demonstrate the changes in the morphology of pUC19 in the presence of encapsulin proteins (both empty or with its native cargo protein). In parallel, the secondary structure and thermal stability of encapsulin shell, both free and bound to plasmid DNA, were assessed by Synchrotron Radiation Circular Dichroism (SRCD). The previous SRCD data of EncA (section 4.2.2) are now directly compared with the EncA–pUC19

complex. Both SRCD spectra are nearly identical (Figure 6.4A and C), with negative peaks at 209 and 220 and a positive peak at 192 nm, consistent with a folded protein. The small spectral differences between EncA and EncA–pUC19 are justified by the spectral contribution of pUC19 (Figure A.13A, appendix). A more detailed analysis of the secondary structure of EncA was accomplished by spectral deconvolution using DichroWeb online tools [195] already described in section 4.1.3.

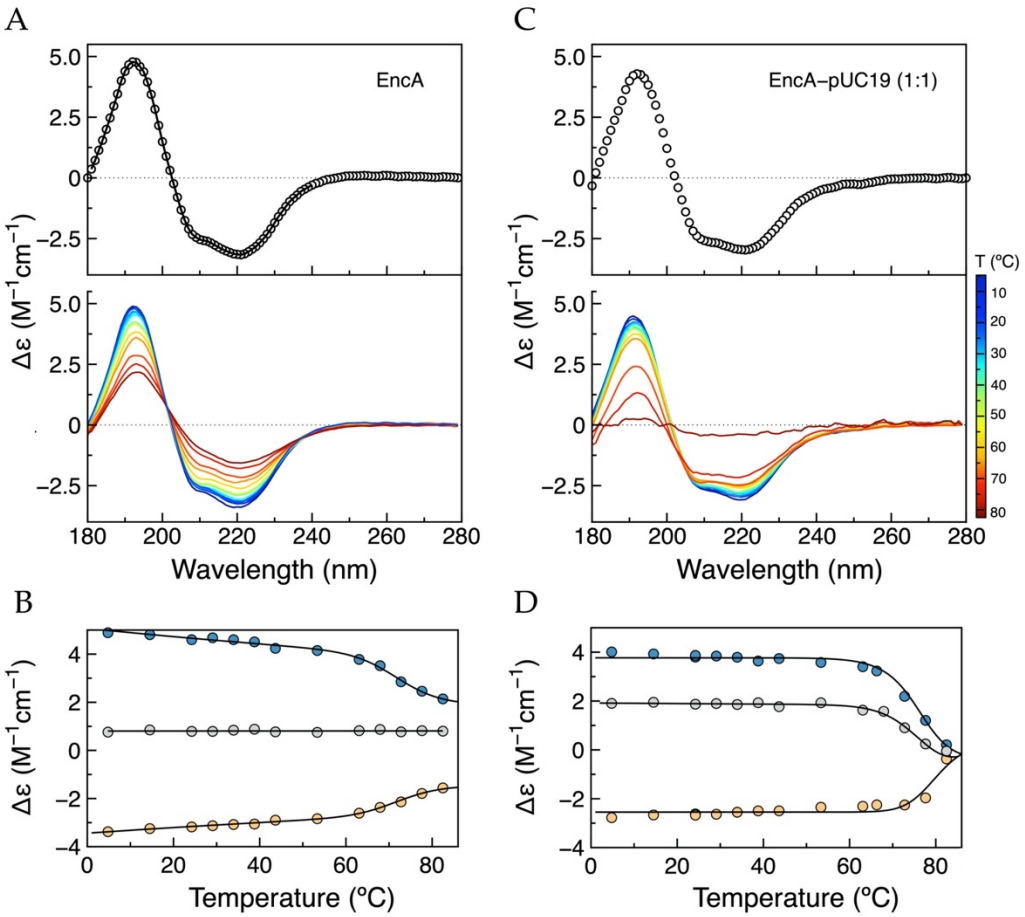

Figure 6.4: Secondary structure and thermostability assessment of free EncA and EncA–pUC19 complex by SRCD. (A) Spectra of EncA (top panel, open circles) at 25 °C with spectral reconstructed data from DichroWeb analysis (solid line) and temperature scans (bottom panel). **(B)** EncA thermal denaturation melting curves. **(C)** Spectra of EncA–pUC19 (top panel, open circles) and temperature scans (bottom panel). **(D)** EncA–pUC19 thermal denaturation melting curves. EncA and EncA–pUC19 thermal denaturation was monitored at 192 (in blue), 220 (in yellow), 202 nm (in gray) and the solid lines overlaying the experimental data result from a non-linear least-squares fit to the data based on a two-state denaturation model (Equation 6).

Temperature scans from ~ 5 to 83 °C were performed to compare protein folding and the thermal stability of EncA and the EncA–pUC19 complex (Figure 6.4A and C). In both cases, increasing the temperature does not affect the spectral shape but leads to a decrease in CD signal intensity, more extensive in the EncA–pUC19 sample, evidencing the loss of secondary

structure. Following the denaturation process at three distinct wavelengths (192, 220 and 202 nm) and fitting the experimental melting curves with a two-state thermal denaturation equation (Equation 6) show that EncA presents high thermal stability with a melting temperature (T_m) of 76 ± 1 °C (Figure 6.4B). In the presence of pUC19 the T_m increases to 79 ± 1 °C showing a slight increase in the thermostability of the protein when complexed to DNA (Figure 6.4D). Furthermore, a 4-fold increase in the enthalpy at the unfolding transition (ΔH_m) was determined for the EncA−pUC19 complex when comparing with free encapsulin (184 ± 5 vs 45 ± 2 kcal/mol). This difference may be explained by a higher number of non-covalent molecular interactions between EncA and the pUC19 molecule and between aggregated protein molecules in the condensates, requiring more energy to disrupt the complexes. Furthermore, the final denatured form of the EncA−pUC19 complex may have a less favorable hydration than the EncA one. The latter hypothesis is supported by the differences in the spectra of the final denatured states obtained for EncA and EncA−pUC19. Contrary to what is observed for the denatured form of EncA, there is a drastic loss of ellipticity from 180 to 280 nm pointing to the complete loss of secondary structure elements for the EncA−pUC19 complex. It is thus probable that the denatured state will be less prone to hydration.

6.3 Discussion and Conclusion

To our knowledge, we are reporting for the first time the capacity of an encapsulin cage protein to bind and protect circular double stranded DNA. The EMSA assays revealed that EncA protein was able to bind supercoiled plasmid DNA pUC19 with an apparent dissociation constant of 0.3 ± 0.1 μM and a Hill coefficient of 1.4 ± 0.1, similar to other protein nanocages [110,207,210]. The native cargo protein of this encapsulin system, EncC, did not bind to pUC19 under the conditions tested. However, when encapsulated within EncA shell, the EncAC complex displayed a DNA−binding capacity similar to the empty EncA ($K_D = 0.4 \pm 0.1$ μM and 0.3 ± 0.1 μM for low and high ionic strength, respectively, and $n = 1.1 \pm 0.1$), suggesting that the overall electrostatic charge and binding properties of the outer surface were not affected by the encapsulation process. Increasing the NaCl concentration of the binding buffer did not significantly affect the electrostatic interaction between the protein and DNA.

The EMSA data presented here show that in a first stage binding of EncA converts the plasmid DNA to a more relaxed form, with the protein appearing along the pUC19 molecules as "beads-on-a-string" as revealed by the AFM images. At higher protein concentrations the protein molecules bound to the DNA self-aggregate, looping and compacting the DNA. This mechanism has been described for the *Esc. coli* Dps, a nucleoid-associated protein, that binds DNA non-specifically, promoting its condensation through a protein self-aggregation process [107]. Furthermore, encapsulin binding protects the DNA from enzymatic digestion by DNase I, which suggests a protective effect caused by physical shielding.

The interaction between DNA and the outer surface of protein cages is relatively well characterized in Dps proteins and, in most cases, is mediated by positively charged amino acid residues, lysine and arginine, present in the protruding N- or C- terminal flexible extensions [80,110,211,212]. However, in encapsulin proteins the N-terminal regions of each monomer are positioned towards the cavity, interacting with the cargo protein, while the C- termini are part of the pore architecture and thus inaccessible for interaction. The encapsulin outer surface, like in the Dps proteins, is predominantly neutral to negatively charged, yet two distinct positively charged clusters are present. The first one, located within the monomer, is composed of 3 lysines and 3 arginines, and the second comprises 4 lysine and 4 arginine residues at each dimer interface (Figure A.14, appendix). Although these regions are located at the external surface of the encapsulin shell they are less solvent-accessible than the N- and C- terminal tails of Dps proteins.

Additionally, the overall secondary structure content of EncA is not greatly affected by DNA-binding. The EncA protein is highly thermostable with a T_m of 76 ± 1 °C, similar to the encapsulin from *Qua. thermotolerans* [25]. The interaction with pUC19 slightly increased the melting temperature by up to 3 °C. A significant difference was found for the value of the enthalpy at the unfolding transition, ΔH_m, which increased approximately 4-fold in the presence of the DNA (from 45 to 184 kcal/mol). Two phenomena can contribute to raise the ΔH_m. It is expected that the number of non-covalent bonds increases due to the formation of new protein–protein and protein–DNA molecular interactions, as illustrated by AFM results.

Moreover, in the presence of DNA, the SRCD spectrum of the fully denatured state is typical of unordered polypeptide chains which will lead to a less favorable hydration of this state. Taken together, during unfolding the DNA–protein system goes from a state with a higher number of non-covalent bonds to very few when compared to the EncA by itself, which can explain the 4-fold increased obtained for ΔH_m.

Myx. xanthus is used as a model organism to study the cooperative mobility and cell-to-cell signaling during multicellular development [213]. Under stress conditions, *Myx. xanthus* triggers a multicellular development cycle in which the individual rod-shape cell aggregate to form the fruiting bodies followed by spore development. Kim *et al.* have shown that Δ*encA* (and Δ*encF*) gene deletion in *Myx. xanthus* impaired the formation of the fruiting body as the cells became unable to produce DKxanthene and myxoviresin, which are essential for this cellular adaptation. The cells also became agglutination-impaired [77]. The authors correlated this Δ*encA* phenotype with a cellular inability to sense the availability of iron. However, the Δ*encC* and Δ*encD* mutants, with deleted genes encoding putative ferritin-like cargo proteins, maintained the proper transition to fruiting body formation. Following the observations by Kim *et al.*, an additional layer of understanding of this phenotype might be the hitherto uncharacterized ability of EncA to bind and protect DNA against external stresses. Besides being a particularly large iron storage protein that prevents the formation of reactive oxygen species through Fenton chemistry, EncA may also be part of the cellular defense response by DNA protection upon unfavorable conditions, possibly enabling the progress of the mycobacterial life cycle. Thus, EncA represents a new member of the nucleoid-associated protein family, inducing changes in the DNA topology that allow its protection under stress [211].

Further work is needed to understand if this DNA−binding and protection capacity is limited to the *Myx. xanthus* encapsulin and confirm the physiological relevance of these findings through *in vivo* studies.

7

IRON RELEASE FROM A MINI-FERRITIN

This chapter details the production and purification of the proteins used in the study of the iron release from the *Marinobacter (Mar.) hydrocarbonoclasticus* mini-ferritin. The production and purification of Dps and flavoprotein WrbA was optimized by Guilherme, M. [214] and Almeida, N. R. [131], while the Rubredoxin (Rd) expression vector had previously been prepared by Folgosa, F.

These proteins were used to characterize the reduction and release of the iron of the mineral core stored in Dps using a Flavodoxin NADH/NAD(P)H oxireductase system (WrbA) as a redox partner and NADH as an electron source. The reduction and release reactions were performed with and without a chemical chelator (1,10-phenanthroline) or a non-heme protein receptor (Rubredoxin (Rd)) as final iron receptors.

After the reduction of the mineral core it is important to determine whether the ferrous ions are inside or outside Dps. Initially, a chemical chelator (1,10-phenanthroline) was used to chelate the iron released to the extracellular medium and thus quantify the final product. Due to the anaerobic facultative nature of *Mar. hydrocarbonoclasticus* the iron release mechanism was explored both under anaerobic and aerobic conditions. The release mechanism was additionally investigated using a protein as iron receptor (apo-rubredoxin) to sequester the iron released from Dps. In this work, Rd was also identified as a suitable protein partner. Mössbauer spectroscopy allowed the identification, characterization, and quantification of the different iron species formed throughout the reaction. As such, the reduction of the ferric mineral core, the release of reduced ferrous ions and their chelation by Rd were monitored.

Section 7.2.3 was adapted from the following publication, published during this thesis: Ana V. Almeida, João P. Jacinto, João P. L. Guerra, Bruno J. C. Vieira, João C. Waerenborgh, Nykola C. Jones, Søren V. Hoffmann, Alice S. Pereira, Pedro Tavares., "Structural features and stability of apo- and holo-forms of a simple iron–sulfur protein" European Biophysics Journal, vol. 50(3-4):561-570 (2021) DOI: 10.1007/s00249-021-01546-0.

7.1 Experimental Procedure

7.1.1 Production of the Proteins from the Iron Release System

Dps (FQ312129.1), WrbA (WP_014422496.1) and Rubredoxin (Rd) (ABM20661.1) genes have been inserted in pET21 expression vectors prior to this work, using conventional recombinant DNA technology. The expression vectors (pET-21c-Dps, pET-21c-WrbA and pET-21c-Rd) were used to transform *Esc. coli* BL21(DE3) competent cells (NZYTech) following the manufacturer instructions and plated on LB-agar containing 100 mg/L of ampicillin.

For Rd small-scale expression tests the cells harboring the vector were inoculated in 50 mL of LB (25 g/L, NZYTech) containing 100 mg/L of ampicillin with a 500 μL inoculum grown overnight at 37 °C. The expression tests were performed by optimizing the concentration of IPTG (0.1, 0.5 or 1 mM) and the duration and temperature of induction (3 h at 37 °C or overnight at 22 °C). The cellular density between condition was compared ($OD_{600\ nm}$) and normalized to evaluate the protein expression through polyacrylamide SDS-PAGE gels.

For Dps, WrbA and Rd large-scale, cells harboring each of the expression vector were grown in LB medium with 100 mg/L of ampicillin at 37 °C, 220 rpm until an $OD_{600\ nm}$ around 0.6 – 0.8. The expression was induced with 0.5 mM IPTG at 37 °C during 3 h at 220 rpm. For Rd, additionally to the IPTG induction, the culture was supplemented with 0.2 mM of $FeSO_4$ (freshly prepared in milli-Q H_2O, pH 2.0) and grown overnight at 20 °C with orbital shaking at 200 rpm.

The cells were harvested by centrifugation at 11,000× *g* for 15 min at 8 °C (Z 36 HK, HERLME LaborTechnik) and resuspended in 10 mM Tris-HCl buffer, pH 7.6. Cell lysis was performed as previously described in section 3.1.1 followed by centrifugation at 10,000× *g* for 30 min at 8 °C and ultracentrifugation at 180,000× *g* for 90 min at 4 °C (Beckman Coulter type 40 Ti rotor). The cellular extracts were dialyzed overnight at 4 °C against 10 mM Tris-HCl buffer, pH 7.6. The production and purification of each protein was monitored using 12.5 % (15 % for Rd) SDS-PAGE gels.

7.1.2 Purification of the Proteins from the Iron Release System

All proteins purification procedures followed a two-step protocol. In the first step a DEAE-Sepharose FF preparative ionic exchange (IEX) chromatography in a XK 26/40 cm column (Cytiva) pre-equilibrated with 10 mM Tris-HCl buffer, pH 7.6 at a flow rate of 5 mL/min was used. After loading the cellular extract and elution of non-adsorbed proteins, the adsorbed ones were eluted with a linear gradient of 0 – 0.5 M NaCl in 10 mM Tris-HCl buffer, pH 7.6. The fractions containing the protein of interest were pooled and concentrated using a Vivaspin 20 (Sartorius) with a MWCO 100 kDa for Dps, 10 kDa for WrbA and 5 kDa for Rd.

7.1.2.1 Purification of Dps

After the preparative IEX chromatography step, Dps fractions were dialyzed against 10 mM Tris-HCl buffer, pH 7.6 and loaded into a Resource Q column (6 mL, Cytiva) pre-equilibrated with 10 mM Tris-HCl buffer, pH 7.6 at a flow rate of 1 mL/min. Elution was performed by applying a linear gradient of NaCl as described in the previous section. Protein purity was evaluated by SDS-PAGE electrophoresis and the catalase test was performed to assess potential catalase contamination as described in section 3.1.3.1.1.

7.1.2.2 Purification of WrbA and Rd

After the first chromatographic step, WrbA and Rd fractions were loaded into a Superdex 200 Prep Grade SEC column (XK 26/100 cm, Cytiva) pre-equilibrated with 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl at a flow rate of 1.5 mL/min. The fractions containing pure protein were pooled together based on SDS-PAGE purity assessment and catalase test.

7.1.2.2.1 Preparation of apo-Rd

As the Rd-producing *Esc. coli* cultures have been supplemented with $FeSO_4$ to improve the overexpression yield Rd was purified in its iron-loaded holo-form. In order to use Rd as an iron acceptor partner in the Dps iron release assay, apo-Rd was prepared as described in [215] with the following modifications: 10 mg of pure holo-Rd was precipitated using 5 % of trichloroacetic acid (TCA) in the presence of 0.5 M dithiothreitol (DTT) and incubated at 45 °C during 30 min. The resulting white precipitate was then collected by centrifugation at 7,700× *g* for 30 min and resuspended in 500 mM Tris-HCl buffer, pH 7.6 with 60 mM DTT. The precipitation process was repeated to ensure full iron demetallation. The buffer was then exchanged to 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl using a desalting column (PD SpinTrap G25, GE Healthcare Life Sciences) first in anaerobic conditions (as described in section 5.1.2) and then under atmospheric conditions.

7.1.3 Biochemical Characterization

7.1.3.1 Size-Exclusion Chromatography

The homogeneity of pure Dps, WrbA, Rd proteins was evaluated using the pre-packed Superdex 200 10/300 GL SEC column, previously calibrated as described in section 4.1.2.

Samples containing ~ 1 mg of each protein were loaded into the column at 0.5 mL/min. Additionally, 4 equivalents of FMN were added to the apo-form of WrbA to evaluate the oligomeric state of the holo-form. Similarly, 1.2 equivalents of Fe^{2+} were added to apo-Rd and loaded into the column.

7.1.3.2 UV-Visible Spectroscopy

The activity of each protein was confirmed using UV-Visible spectroscopy as previously described (section 3.1.3.2). The spectra of Dps, WrbA and Rd were acquired in 200 mM MOPS buffer pH 7.0, 200 mM NaCl.

Dps (250 μM apo-form) was incubated with 144 Fe^{2+}/protein under oxygen saturation conditions to promote mineral core formation. The spectrum was obtained 48 h after ferrous iron incubation. 8 μM of FMN-WrbA and apo-WrbA (tetrameric form) was incubated with a 4-fold excess of NADH to verify the reduction of the FMN co-factor. 250 μM of apo-Rd was incubated with 1.2 equivalent of $FeSO_4$ to confirm the iron-sulfur center reconstitution capacity. Furthermore, Apo-Rd was saturated with oxygen for 10 min and then loaded with iron by sequential additions of 25 μM (0.1 eqs.) of acidic ferrous iron sulphate. Each addition was only performed after stabilization of the absorbance spectrum, as well as in kinetic mode following the absorbance at 494 nm.

7.1.3.3 CD/SRCD

SRCD were acquired as mentioned in section 4.1.3. Temperature scans were obtained from 25 °C to ca. 90 °C, in 5 or 10 °C increments steps. Conventional Circular Dichroism (CD) spectra were acquired at the BioLab, Biological and Chemical Analysis Facility, UCIBIO-LAQV/Requimte, FCT NOVA, Portugal, using an Applied Photophysics Chirascan™ qCD spectrometer. Spectra were recorded with 1 nm steps and a dwell time of 2 sec per step in triplicates, using the same calibrated 0.01008 cm pathlength quartz cell (SUPRASIL, Hellma GmbH, Germany), for the wavelength range of 190 – 260 nm.

For SRCD samples, recombinant Rd and apo-Rd protein samples were exchanged to 10 mM MOPS buffer, pH 7.0 with 240 mM NaF using a desalting column (PD SpinTrap G25, GE Healthcare Life Sciences) inside an anaerobic chamber (MBraun). Holo-form of the protein form were prepared by incubating 1 mg/mL of apo-Rd with different equivalents (0.25, 0.5, 0.75, 1.0) of $FeSO_4$. For conventional CD analysis, apo-Rd samples at 1 mg/mL in the same buffer was incubated with different metal solutions such as $MnSO_4$, $FeSO_4$, $CoSO_4$, $NiCl_2$, $ZnSO_4$, $Cd(NO_3)_2$ or $Ce(SO_4)_2$ in a molar excess of 1.3. Excess of chloride ions were removed through dialysis.

The melting temperature (T_m) was calculated by fitting the SRCD signal intensity data at appropriate wavelengths as a function of temperature (K). A system with two consecutive transitions and three species was used. For this, a reversible equilibrium between an initial native form (N) and an intermediary form (I), followed by a second reversible equilibrium between the later and a denatured form (D) was postulated. In such cases the population percentages of each species are given by the following equations:

$$\chi_N = \frac{1}{1+e^{-\frac{\Delta H_{m,1}\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}+e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}\times e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,2}}\right)}{R\times Tmp}}}$$ Equation 8

$$\chi_i = \frac{e^{-\frac{\Delta H_{m,1}\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}}{1+e^{-\frac{\Delta H_{m,1}\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}+e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}\times e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,2}}\right)}{R\times Tmp}}}$$ Equation 9

$$\chi_D = \frac{e^{-\frac{\Delta H_{m,1}\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}\times e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,2}}\right)}{R\times Tmp}}}{1+e^{-\frac{\Delta H_{m,1}\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}+e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,1}}\right)}{R\times Tmp}}\times e^{-\frac{\Delta H_m\left(1-\frac{Tmp}{T_{m,2}}\right)}{R\times Tmp}}}$$ Equation 10

where ΔH_m is the enthalpy change of the unfolding transition, T_m is the melting temperature, Tmp is temperature in Kelvin, R is the universal gas constant [216,217].

7.1.3.4 EPR

Electron Paramagnetic Resonance (EPR) spectra were acquired on a MiniScope MS 400 spectrometer (Magnettech GmbH) setup operating at X-band (9.448 GHz), at 77 K by immersion in liquid nitrogen. Two sets of acquisition parameters were used: either a 200 mT field sweep with 10 mW microwave power, 0.4 mT modulation amplitude, 50x gain or a 20 mT field sweep with 19.9 mW microwave power, 0.5 mT modulation amplitude, 900x gain.

The g-values observed can be interpreted using a spin Hamiltonian as described before for a high-spin ferric ions, $S = 5/2$ [218].

$$\widehat{H} = D\left[S_Z^2 - \frac{1}{3S(S+1)} + E/D(S_Y^2 - S_X^2)\right] + \beta\vec{S}\cdot\tilde{g}\cdot\vec{H}$$ Equation 11

Samples with 1.25 mM of apo-Rd (in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl) were incubated with 1.5 mM ferrous iron (1.2 eqs.) for 15 min at room temperature before being flash-frozen in liquid nitrogen. A sample containing 1.5 mM recombinant Rd (in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl) was prepared in parallel as a control.

7.1.3.5 Mössbauer Spectroscopy

Mössbauer data were recorded as described in section 5.1.2. Rd samples were prepared inside an anaerobic chamber (MBraun). 1.2 mM of apo-Rd (in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl) was incubated with 1.0 mM ferrous iron for 15 min at room temperature, transferred to a Mössbauer cup and flash-frozen in liquid nitrogen.

7.1.3.6 DLS

DLS measurements were performed as described in section 4.1.4. Apo-Rd under anaerobic conditions, apo-Rd saturated with O_2 for 10 min, and zinc reconstituted apo-Rd samples were prepared. Final protein concentration was 2 mg/mL in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl. Samples were centrifuged for 15 min at 14000× *g* before each measurement.

7.1.4 Iron Release from a Mini-ferritin

The holo-Dps (harboring a mineral core with 144 Fe atoms), FMN-WrbA and apo-Rd proteins forms were used in the iron release assays.

7.1.4.1 Dps Iron Loading

To study the iron release mechanism in *Mar. hydrocarbonoclasticus* Dps, the first step was to load the protein with ferric iron to form a mineral core in its inner cavity, by the ferroxidation and mineralization reactions. A mineral core comprising 144 iron atoms per protein was prepared by addition of either $^{56}FeSO_4$ or $^{57}FeSO_4$ (for Mössbauer spectroscopy) in 200 mM MOPS buffer, pH 7.0 with 200 mM NaCl under oxygen saturation, and incubated for 3 days. The $^{56}FeSO_4.7H_2O$ solution was prepared in milli-Q H_2O at pH 2.0 under an inert atmosphere. The $^{57}FeSO_4$ was prepared as described at section 5.1.2.

7.1.4.2 Preparation of the Release Reaction

All reagents were quantified using their molar extinction coefficients (with the exception of 1,10-phenanthroline) (Table 7.1). When working under anaerobic conditions, the samples were prepared as described in section 5.1.2. When working in aerobic conditions, all reagents were kept in atmospheric conditions.

Table 7.1: Molar extinction coefficients of the reagents involved in the iron reduction and release reactions.

Reagent	λ (nm)	Molar Extinction Coefficients ($M^{-1}cm^{-1}$)
Dps	280	2.9×10^5
WrbA	280	7.6×10^4
Rd	280	2.3×10^4
NADH	340	6.2×10^3

The values refer to dodecameric Dps, tetrameric WrbA and monomeric Rd. Dps and Rd molar extinction coefficient were obtained using Protparam method [181,182] while WrbA was determined in previous work through BCA method [131].

7.1.4.3 Iron Release using 1,10-phenanthroline as Iron Chelator

The iron release assays performed using 1,10-phenanthroline as iron acceptor were prepared in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. 1 μM of Dps:144Fe was incubated with 12 μM of FMN-WrbA, 720 μM of NADH and 720 μM of 1,10-phenanthroline in anaerobic and atmospheric conditions. UV-Visible spectra of each component of the reaction were acquired using an Evolution 201 or Evolution 300 spectrophotometer (Thermo Scientific). To monitor the kinetics of the iron release reaction, spectra were recorded with a bandwidth of 1 cm, between 380 to 600 nm, with a 600 nm/min scan speed in 15 min intervals.

The product was quantified using the Lambert-Beer Law at 510 nm and the molar extinction coefficients of the ferroin complex (11.40 $mM^{-1}cm^{-1}$) (Figure A.5, appendix). The iron release progress curves were analyzed by non-linear least-square fits with a zero-order kinetics (Equation 12).

$$\%\ Iron\ Released\ (t) = k \times t \qquad \text{Equation 12}$$

where, t is the time in seconds, and k is the rate constant.

7.1.4.4 Iron Release using Rubredoxin as Iron Acceptor

7.1.4.4.1 Mössbauer Spectroscopy

Mössbauer data was acquired as described in section 5.1.2. As previously mentioned, the Dps:144Fe mineral core was prepared using $^{57}FeSO_4$ under oxygen saturation for 3 days. The samples were then degassed and transferred to the anaerobic chamber as previously described in section 5.1.2. 10.4 μM of Dps with 1.5 mM of ^{57}Fe mineral core (Dps:144Fe) was mixed with 125 μM of FMN-WrbA, 2.5 mM of apo-Rd and 7.5 mM of NADH in 200 mM MOPS buffer, pH 7.0, 200 mM NaCl (here termed master mix), inside the anaerobic chamber. 400 μL samples from the master mix were removed at different reaction times (5, 30, 70, 100, 135 and 400 min) and immediately flash-frozen in liquid nitrogen. The progression curves of the release of iron were represented by plotting the percentage of each iron species along the reaction. The curves were then analyzed by least-squares fitting using a zero-order equation (Equation 12).

7.2 Results

7.2.1 Production and Purification of Iron Release Proteins

7.2.1.1 Dps

Dps was successfully expressed using *Esc. coli* BL21(DE3) cells culture in LB medium (with 100 mg/L ampicillin) after induction at OD_{600} ~ 0.6 – 0.8 with 0.5 mM IPTG, during 3 h at 37 °C, 220 rpm. The protein was expressed in the soluble fraction as evidenced by the presence of a single band with an apparent molecular weight consistent with the monomer size (18.5 kDa). After fractionation of the cellular extract, Dps was purified using IEX chromatography. Based on the SDS-PAGE gels, Dps eluted from the anion column at a NaCl concentration between 220 to 260 mM. The fractions containing Dps were then loaded into a Resource Q column and eluted from this column between 240 to 335 mM NaCl. Protein purity was evaluated by SDS-PAGE and through a catalase test and only contaminant free Dps was used in the following experiments (Figure 7.1A). This expression and purification protocol typically yield ~ 100 mg of pure protein per liter of cell culture.

7.2.1.2 WrbA

WrbA was expressed using *Esc. coli* BL21(DE3) cells as described for Dps, exhibiting an SDS-PAGE band consistent with the apparent molecular weight of WrbA monomer (20.9 kDa) in the soluble fraction. WrbA was loaded into a preparative anion exchange column and eluted between 240 and 300 mM NaCl. This step allowed not only the purification, as well as the separation of both apo- (without the FMN co-factor) and holo-WrbA forms, the latter presenting a yellow coloration. From this point forward, apo-WrbA and FMN-WrbA were independently applied into a SEC column for final polishing. The selected pure fractions showed no catalase contamination (Figure 7.1B). Typically, 40 mg of FMN-WrbA and 25 mg of apo-WrbA per liter of cell culture were obtained.

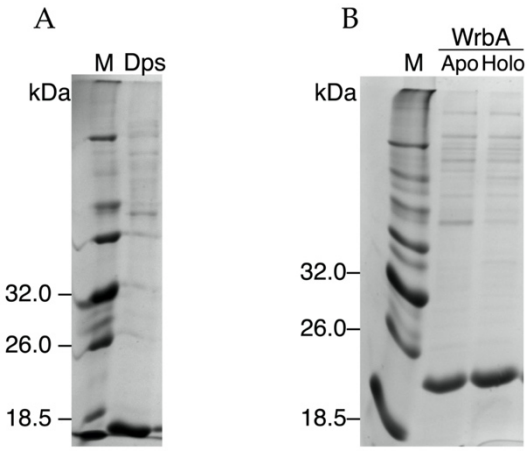

Figure 7.1: Purity assessment of Dps and WrbA. Pure stocks of **(A)** Dps and **(B)** WrbA in apo and holo (with FMN) form purity evaluation through SDS-PAGE electrophoresis (12.5 %). M– LMW II.

7.2.1.3 Rubredoxin

Rd overexpression in *Esc. coli* BL21(DE3) cells transformed with the pET21-Rd expression vector was tested in different conditions (IPTG concentration, induction temperature and duration) (Figure 7.2A). A protein with an apparent molecular weight below the 18.5 kDa protein marker band was expressed in all conditions tested (Figure 7.2A). Rd has a molecular weight of 6.26 kDa, suggesting that the band corresponds to the protein of interest. Therefore, the overnight expression at 22 °C, 200 rpm with 0.5 mM IPTG and 0.2 mM $FeSO_4$ conditions were selected as a standard overexpression protocol.

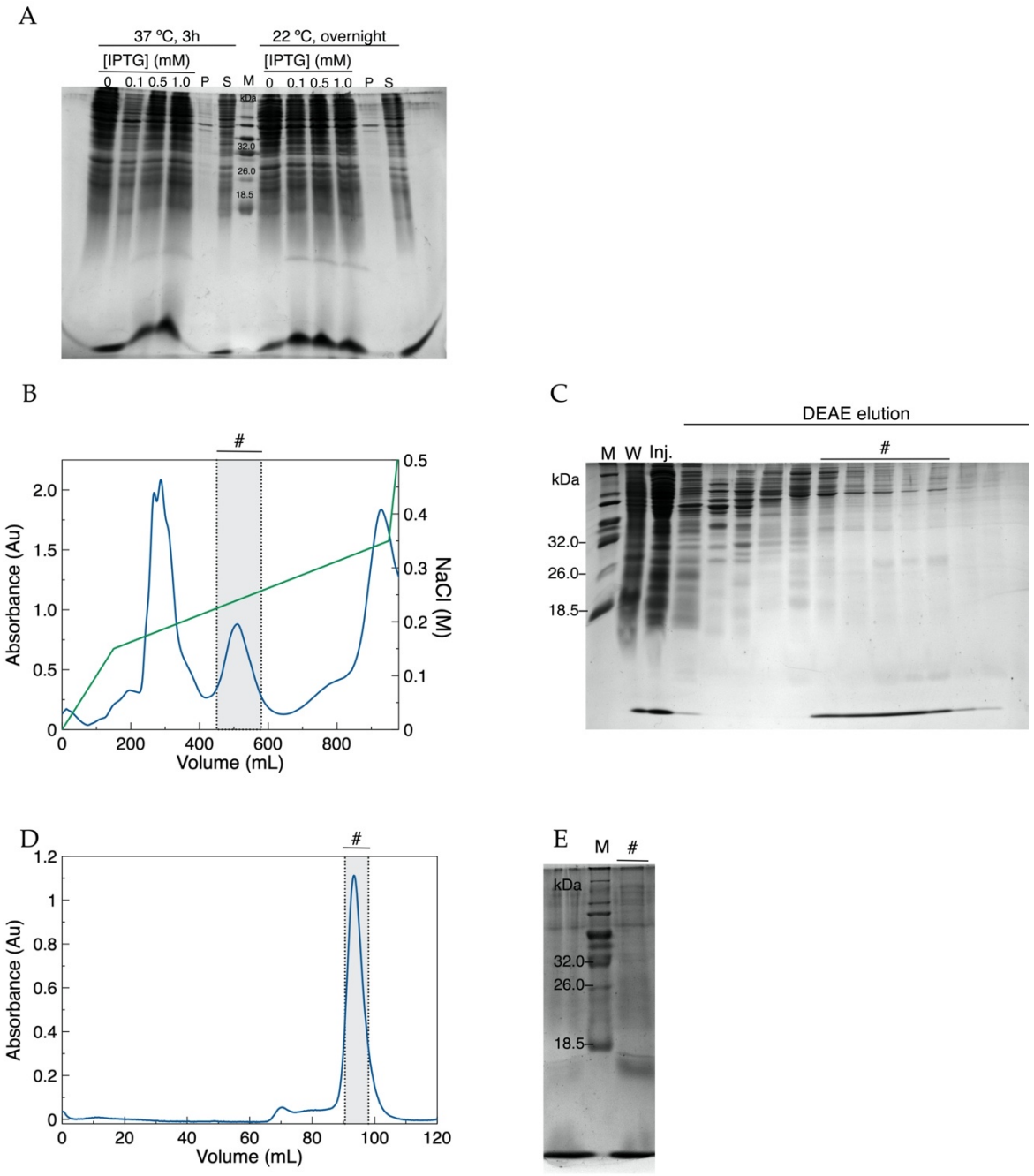

Figure 7.2: Expression tests and large-scale production and purification of rubredoxin. (A) Assessment of Rd expression levels by SDS-PAGE (15 %) after induction with IPTG (at a concentration of 0, 0.1, 0.5 or 1.0 mM) for 3 h

at 37 °C or at 22 °C overnight. P – pellet, S – soluble (both fractions obtained after centrifugation of lysed cells). **(B)** Rd elution profile (absorbance at 280 nm) from the DEAE Sepharose FF column using a discontinuous NaCl gradient. **(C)** SDS-PAGE (15 %) analysis of the chromatographic step presented in (B). **(D)** Rd elution profile from the Superdex 200 Prep grade column in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. **(E)** SDS-PAGE (15 %) of the Rd fractions from C. In each chromatographic step, the fractions selected are identified in gray. M – LMW II.

Large-scale production of the protein resulted in a soluble protein fraction. The purification process was started using a preparative DEAE anionic exchange column, with the protein eluting between 220 and 260 mM NaCl after applying a discontinuous gradient (Figure 7.2B). The presence of Rd in the eluted fractions was confirmed by SDS-PAGE and by a red coloration due to the presence of the iron-sulfur center (Figure 7.2C). Rd fractions were injected into a Superdex 200 Prep Grade size exclusion chromatography for a final protein polishing, assessed through SDS-PAGE (Figure 7.2D and E). The typical yield of a Rd production protocol was 215 mg per liter of cell culture.

7.2.2 Biochemical Characterization of the Iron Release Protein

The homogeneity, quaternary structure and activity of the proteins involved in this work were confirmed by UV-Visible spectroscopy and SEC (Figure 7.3).

The Dps was expressed and purified as the apo-protein form (Figure 7.3A, black line). The addition of ferrous iron substrate under oxygen saturation conditions led to the appearance of an absorption band between 310 and 390 nm, characteristic of the ferric species formed after ferroxidation and mineralization, confirming its catalytic activity (red line). Additionally, the apo-Dps is homogeneous, displaying a single peak in SEC chromatogram, with an elution volume of 12.8 mL (Superdex 200 10/300 GL column (Cytiva)) corresponding to a 200 kDa globular protein in its dodecameric state (Figure 7.3B).

The UV-Visible spectrum of pure FMN-WrbA exhibits three absorption bands (centered at 378, 446 and 470 nm) characteristic FMN (Figure 7.3C, black line). Accordingly, these bands are absent in the apo-WrbA spectrum (blue line). Upon NADH addition, the FMN co-factor is converted to the $FMNH_2$ reduced form, which lacks the spectral features, thus confirming the oxidoreductase activity of the protein (green line). Apo-WrbA and FMN-WrbA display a distinct elution profile in a SEC (Figure 7.3D). Homologues of *Mar. hydrocarbonoclasticus* WrbA were described as homotetramers in the presence of FMN. Similarly, the FMN-WrbA (black line) was eluted at 15.0 mL, which corresponds to a tetramer, while the apo-form (blue line) was eluted in two peaks at 15.0 and 16.8 mL, the first corresponding to a tetramer and the second to a dimer. The addition of FMN to the apo-WrbA completely converted it to the tetrameric state (red line).

The UV-Visible spectrum of Rd (Figure 7.3E) clearly resembles an oxidized rubredoxin visible spectrum with features at 760, 578, 494, 386 and 356 nm due to ligand-to-metal charge-transfer transitions (black line). An additional 279 nm absorbance peak is observed primarily due to the four tryptophans, one tyrosine and one phenylalanine sidechain absorption.

Precipitation with TCA and DTT resulted in the loss of the iron center (apo-Rd form, blue line) with a colorless apo-Rd showed only UV absorption peaks with a maximum at 279 nm with $\varepsilon_{279\,nm} = 23.5\ mM^{-1}cm^{-1}$ as per the absence of any spectral features. The addition of ferrous ions (under atmospheric conditions) to apo-Rd led to the reconstitution of its iron-sulfur center (red line). The reconstituted sample presented a more intense absorption spectrum, with a molar absorption coefficient is ca. 20 % higher than the observed for the as-purified recombinant protein, which can be explained by the possible presence of a small fraction of Rd containing zinc as the metal cofactor. Regarding the SEC (Figure 7.3F), recombinant Rd was eluted at 18.4 mL both with and without iron, in a single peak with an elution volume characteristic of monomeric protein species, and the SDS-PAGE gel migration pattern revealed that no significant protein degradation occurred during the precipitation step (Figure 7.3G).

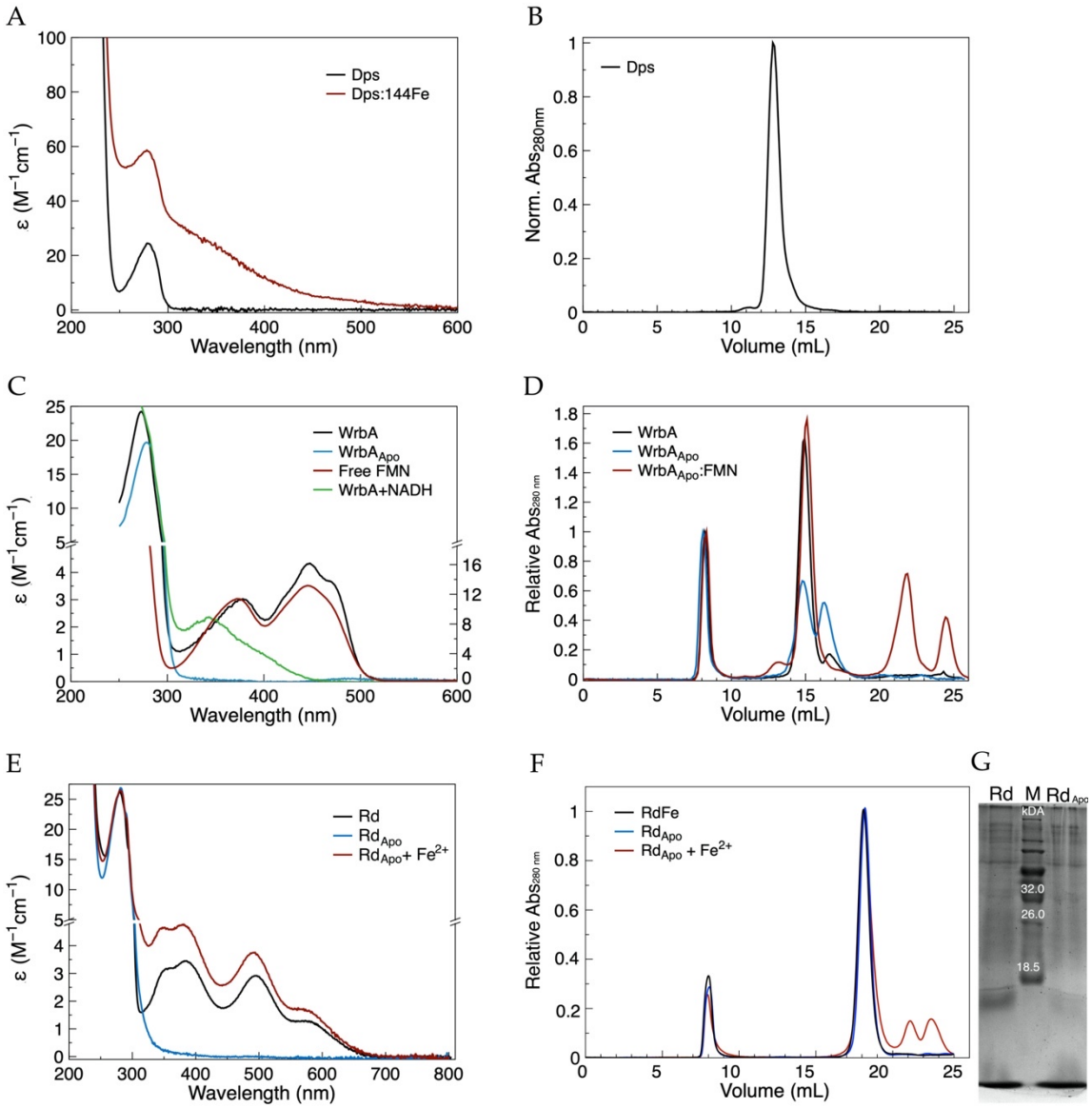

Figure 7.3: Biochemical characterization of the proteins used in the iron-release assays. (A) UV-Visible spectra of Dps as-purified (in black) and Dps loaded with 144 Fe/protein (in red). **(B)** SEC elution profile of Dps as-purified. **(C)** UV-Visible spectra of FMN-WrbA (in black), apo-WrbA (in blue), free FMN co-factor (in red) and FMN-WrbA

with NADH incubation (in green). **(D)** SEC elution profile of FMN-WrbA (in black), apo-WrbA (in blue) and apo-WrbA incubated with 4 equivalents of FMN (in red). **(E)** UV-Visible spectra of Rd as-purified (in black), apo-Rd form (in blue) and iron reconstituted holo-Rd (1.2:1 Fe/protein) (in red). **(F)** Elution profile of Rd expressed with iron (in black) and apo-Rd (in blue), iron reconstituted Rd (in red). **(G)** SDS-PAGE (17.5 %) gel of Rd expressed with iron and apo-Rd after precipitation with TCA and DTT. M – LMW II. SEC analysis was performed using a Superdex 200 10/300 GL column with 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. All UV-Visible spectra were obtained in milli-Q H_2O.

7.2.3 Manuscript: Structural features and stability of apo and holo forms of a simple iron sulfur protein

Ana V. Almeida[1], João P. Jacinto[1], João P. L. Guerra[1], Bruno J. C. Vieira[3], João Carlos Waerenborgh[3], Nykola C. Jones[2], Søren V. Hoffmann[2], Alice S. Pereira[1], Pedro Tavares[1]

[1] Molecular Biophysics Laboratory, UCIBIO/Requimte, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.
[2] ISA, Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark
[3] Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, 2695-066 Bobadela LRS, Portugal

The present work has been published on the 19th of May 2021 by the European Biophysical Journal DOI: 10.1007/s00249-021-01546-0

7.2.3.1 Purified recombinant Rd

The solid line in Figure 7.4A shows the EPR spectrum of the overexpressed protein at 77 K. This spectrum is characteristic of Rd-type centers in the ferric state, with principal features around *g*-values of 4.3, 4.7, 9.3, and 9.8 [219]. These data can be explained assuming a high-spin ferric $S = 5/2$ center with a unique rhombicity parameter $E/D \backsim 1/3$. For such a system, resonances around $g = 4.3$ and 4.7 arises from the $S_Z = \pm 3/2$ excited doublet. The resonances at g values equal to 9.3 and 9.8 originate from the $S_Z = \pm 1/2$ and $S_Z = \pm 5/2$ doublets. To gain information on the recombinant Rd secondary structure and thermal stability, Synchrotron Radiation Circular Dichroism (SRCD) spectroscopy was used (AU-CD beam line, ASTRID2, Aarhus University). SRCD data is particularly valuable since, by using the synchrotron beam, the spectrum can be recorded for higher flux of photons and energies (e.g. in the range of 175 to 195 nm). This extra spectral range provides better insight of structural information on the proteins, making it possible to detect subtle changes in protein secondary structure, thus monitoring their dynamic conformational behavior.

The solid line in Figure 7.4B shows the obtained Rd SRCD spectrum for the recombinant protein ferric state. This is highly similar to the only other Rd SRCD spectrum known [220], with two negative bands at 223 and 203 nm and a positive band at 189 nm. A detailed analysis was carried out with the DichroWeb server [221,222] using the CDSSTR method and the SP175 reference data set, optimized to 175 – 240 nm [220,223]. By applying a three state data model, this analysis results in a modelled 24 % of α-helices, 17 % of β-sheet and 59 % of other structures, which compares well to the average data of PDB structures values of 17 % of α-helices, 15 % of β-sheet and 68 % of other structures, as predicted by the DSSP method using the 2struc server [224]. From the above data, one can infer that the recombinant Rd protein, product of GenBank sequence ABM20661.1 [225], is a functional Rd protein.

7.2.3.2 Apo-Rd characterization

The apo-Rd protein was obtained as described before [170] with appropriate modifications and in aerobic conditions (see section 7.1.2.2.1). To further ascertain the solution structure of apo-Rd, the hydrodynamic diameter was assessed by dynamic light scattering (DLS) and a value of 4.7 ± 0.5 nm was obtained. While such a value is higher than expected, it is in agreement with the value obtained for the zinc substituted protein (4.1 ± 0.7 nm) under the same experimental conditions and can possibly be explained by effects due to the buffer's high ionic strength. Both SEC and DLS results show that apo-Rd protein is in a homogeneous monomeric state in solution. Figure 7.4B (dashed line) shows the SRCD spectrum obtained for the apo-Rd. Notably, the spectrum displays $\Delta\varepsilon$ maximum at 191 nm and minima at 204 and 231 nm, being consistent with a structure that can be modelled to 8 % of α-helices, 32 % of β-sheet and 60 % of other structures.

If compared to the data obtained for the overexpressed, recombinant Rd protein, one can see that only two fifths of the α-helix structures are retained with a concomitant conversion to β-sheet structure. It should be pointed out that this analysis relies heavily on the reference datasets used. Usually, a wide-ranging reference dataset covering the spectral range at which data was acquired should be expected to give best results and it is common to use the NRMSD parameter as a guide to find the most suitable reference dataset for the spectral analysis of the protein. In the case of apo-Rd a caveat concerning its spectral characteristics is worthwhile noting. The pronounced negative peak centered at 231 nm has little to no parallel in the reference datasets and could cause a bias in the secondary structure percentage analysis. Finally, the solution structure is maintained under aerobic conditions in the absence of any chemical reductant, and even in the presence of saturating molecular oxygen conditions. This fact, taken together with the SEC (Figure 7.3F) and DLS data, clearly shows that apo-Rd can maintain the same defined, monomeric, solution structure either in anaerobic or aerobic environments.

7.2.3.3 Apo-Rd metal-binding capacity

Metal binding capacity was tested with ferrous ions, and to a lesser extent with zinc, cobalt, cadmium, and nickel ions, since iron is the native metal center constituent. Ferrous ion binding assays were performed by simply incubating an appropriate amount of ferrous sulfate solution with apo-Rd protein. Samples were incubated either in the presence or absence of molecular oxygen. The result was monitored by UV-Visible, EPR, Mössbauer and SRCD spectroscopies. While the anaerobic addition of ferrous iron to apo-Rd results in no changes to the UV-Visible spectrum, ferrous binding and iron center reconstitution occurs as shown by Mössbauer spectroscopy results (Figure 7.4C).

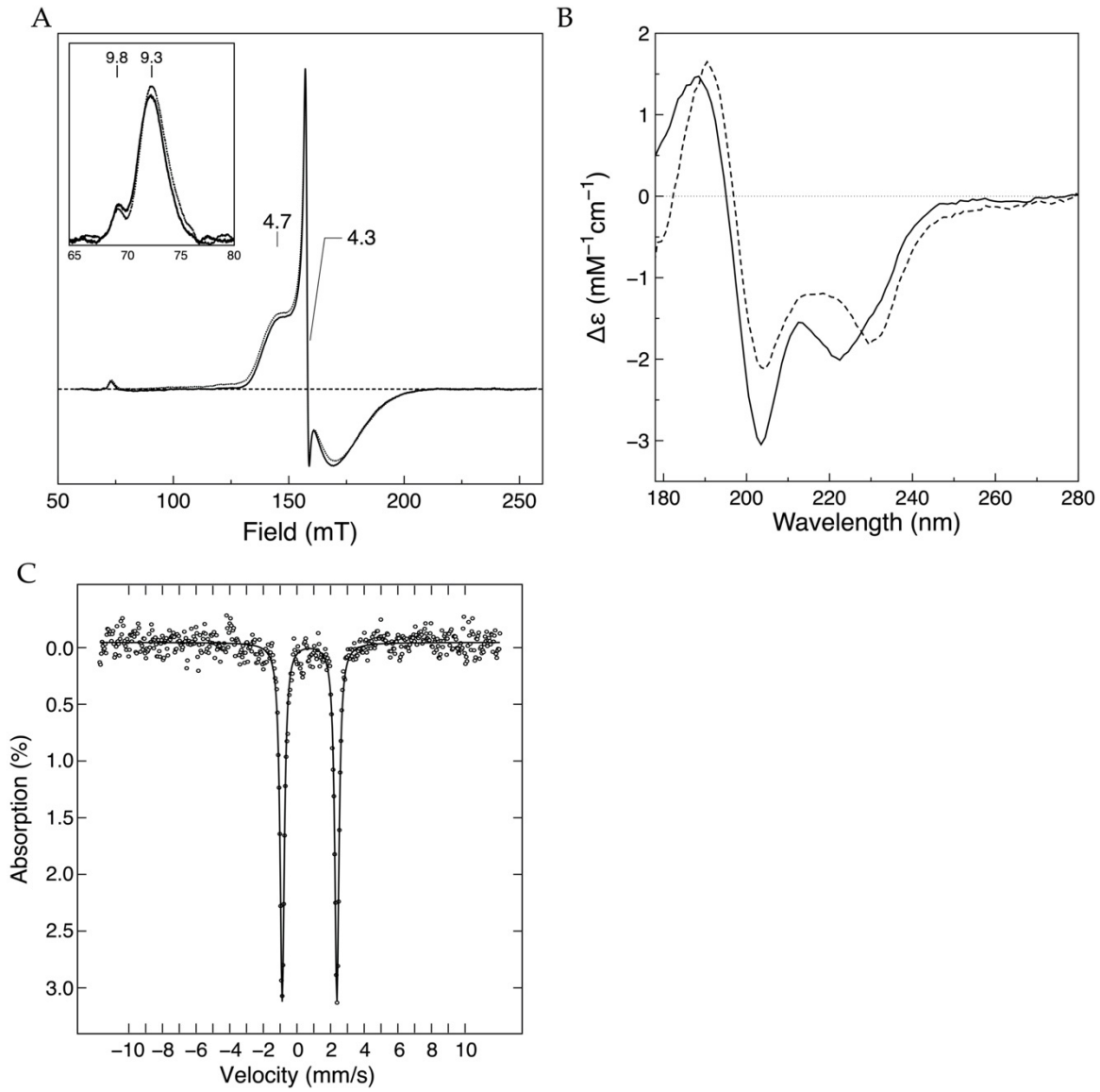

Figure 7.4: Spectroscopic characterization of the Rd, apo-Rd and reconstituted Rd. (A) EPR spectra of recombinant Rd (solid line), and iron-reconstituted Rd (dotted line) proteins recorded at 77 K. Other experimental conditions are: microwave frequency, 9.45 GHz; microwave power, 10 mW; modulation amplitude, 0.4 mT; receiver gain, 50. The inset shows an expansion of the recorded signal in the range of 65 to 80 mT. **(B)** SRCD spectra of recombinant Rd (solid line) and apo-Rd (dashed line) proteins acquired at 25 °C. (**C)** Mössbauer spectrum of the ^{57}Fe-reconstituted holo-Rd recorded at 80 K in the absence of an external magnetic field. The solid line is the result of a least squares fit to the observed quadrupole doublet. Sample prepared under anaerobic conditions.

Also, if ferrous ion addition is done in the presence of molecular oxygen a typical Rd-type visible spectrum develops (Figure 7.5A, top panel) at 494 nm rates of between 1.5 to 3.8 $M^{-1}cm^{-1}s^{-1}$ (depending on the ratio of the number ferrous ions to protein) and reaching a plateau at the point that all iron is chelated by the protein. Figure 7.5A, shows the results of twelve consecutive additions of ferrous iron between 0.1 and 1.2 ferrous ions per apo-Rd protein. Spectral changes are observed at 494 nm until a ratio of 0.96 ± 0.1 is attained, which is in agreement with the expected one iron per protein ratio. The EPR spectra obtained in aerobic

conditions for a ferrous ions per apo-Rd ratio of 1.2 (Figure 7.4A, dotted lines) shows the same spectral features observed in the as-purified recombinant protein, further confirming the reconstitution of the Rd-type center in the ferric state. It should be noted that no visible or EPR spectra are detectable for the ferrous state, prompting for the use of Mössbauer spectroscopy for the characterization of the iron species in this redox state.

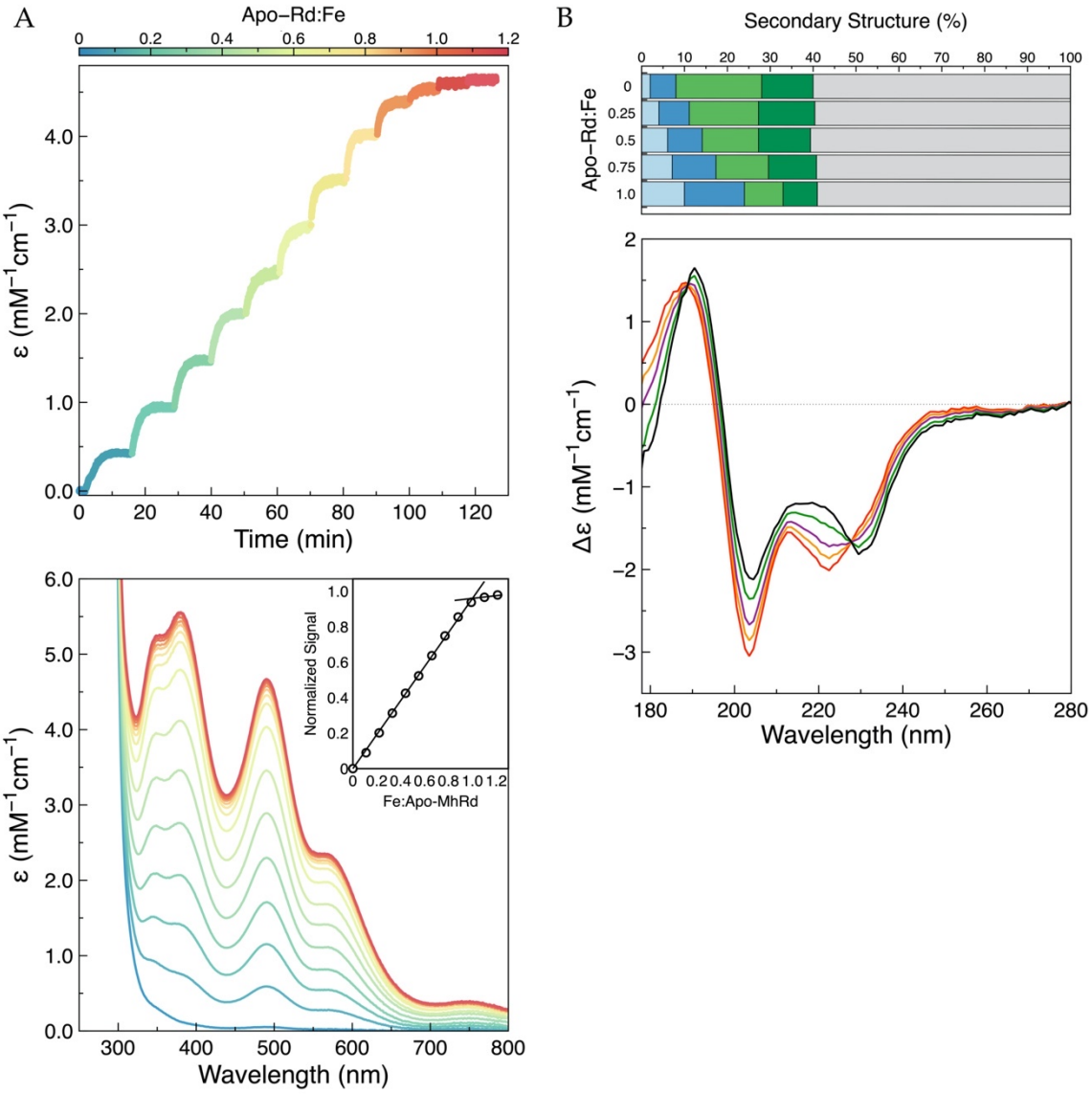

Figure 7.5: Spectral characterization of sequential addition of iron to apo-Rd. (A) Reconstitution of apo-Rd with iron ions. (Top panel) Kinetics of metal center reconstitution and oxidation. Ferrous ions were added sequentially under aerobic conditions with molar ratios between 0.1 and 1.2 Fe^{2+}/protein. Absorbance was recorded at 494 nm and converted into the protein molar absorption coefficient. (Bottom panel) UV-Visible spectra taken at the plateau after each iron addition. The inset shows plateau normalized absorbance as a function of Fe^{2+}/protein molar ratios. **(B)** SRCD spectra obtained for the reconstitution of apo-Rd with iron ions recorded at 25 °C. Five samples with molar ratios between 0.25 and 1.0 Fe^{2+}/protein were used: 0 (black line), 0.25 (green line), 0.5 (purple line), 0.75 (orange line) and 1 (red line). In the top panel the result of the DichroWeb based data analysis for each individual

spectrum is shown. Secondary structure elements are depicted in light blue (Regular helix), dark blue (Distorted helix), light green (Regular strand), dark green (Distorted strand), and gray (Others).

Indeed, ^{57}Fe Mössbauer spectroscopy is the technique of choice to study iron-containing proteins/enzymes since it is possible to obtain relevant information about the iron atoms in our sample [204]. Usually oxidation state, spin state and coordination environment can be inferred from spectral analysis. Figure 7.4C shows the Mössbauer spectrum of a sample resulting from the anaerobic incubation of apo-Rd with a ^{57}FeSO4 solution. The spectrum is constituted by a single sharp quadrupole doublet with parameters of δ equal to 0.69 ± 0.02 mm/s and ΔE_Q equal to 3.25 ± 0.02 mm/s. These are typical for high-spin ferrous ions with tetrahedral sulfur coordination and are identical to that observed in reduced Rd protein isolated from *Clo. pasteurianum* [226,227]. The fact that no other iron species is detected is evidence that all added iron was chelated by the protein in a homogeneous environment.

To probe secondary structure changes, SRCD spectra were acquired for various iron ratios Figure 7.5B. SRCD data shows that addition of ferrous ions leads to the change of the CD peaks around 191 and 231 nm with the concomitant appearance of peaks at 189 and 223 nm. Taking into account the characterizations described above for the recombinant Rd and apo-Rd, it is clear that ferrous addition promotes the conversion between the apo- and holo-forms of the protein. Notably, it is possible to find two true isosbestic points at 189 and 228 nm, which clearly indicates that only two interconvertible species are present in solution, since it would be extremely unlikely that a third species would contribute at two different energies with the same molar ellipticity. Also, the spectra obtained for under-stoichiometric iron ratio additions can be well described by a linear combination of the apo- and holo-forms spectra. Table 7.2 shows the good agreement between added ferrous ions to protein ratios and holo to apo calculated ratios. SRCD spectroscopic data quality allows the detailed data analysis with the DichroWeb server, using an optimized to 175 – 240 nm reference dataset. Figure 7.5B top panel shows the results of this analysis, with a noticeable increase in both helix components (from 2 % and 6 % to 10 % and 14 % for regular helix and distorted helix type structures) and corresponding decrease in strand components (from 20 % and 12 % to 9 % and 8 % for regular strand and distorted strand type structures), slight decrease in unordered regions and a conservation of turn regions. Preliminary experiments were performed with zinc, cobalt, cadmium, nickel for which it was possible to observe changes in the CD spectra consistent with metal binding (Figure A.15, appendix).

Table 7.2: Secondary structure element percentages obtained for the DichroWeb server analysis, using an optimized to 175 – 240 nm reference dataset.

		Secondary structure (%)					
		Regular helix	**Distorted helix**	**Regular strand**	**Distorted strand**	**Turns**	**Unordered**
Apo-Rd		2	6	20	12	15	45
Holo-Rd		10	14	9	8	18	41
Apo-Rd	+ 0.25 Fe/protein	4.0	7.1	16.2	13.1	16.2	43.4
	+ 0.50 Fe/protein	6.1	8.1	13.1	12.1	17.2	43.4
	+ 0.75 Fe/protein	7.1	10.2	12.3	11.2	18.4	40.8

7.2.3.4 Rd thermal stability

SRCD thermal denaturation studies showed profiles that cannot be explained by a single step process. Figure 7.6A and B bottom panel, show the SRCD spectra collected from 24 to 87 °C for both apo- and holo-Rd forms.

Although more perceptible in the former than in the latter form, it is noticeable that the initial spectrum does not convert to a different one, nor that it simply decreases in SRCD signal intensity. Figure 7.6A, top panel, shows the change in apo-SRCD signal intensity at four different wavelengths and the result of a simultaneous non-linear least-squares fit based on equations describing a two-step process with an intermediary species:

$$N \overset{T_{m,1}}{\longleftrightarrow} I \overset{T_{m,2}}{\longleftrightarrow} D$$

Where N, I and D, are native, intermediate and thermally denatured states of either apo- or holo-Rd (Figure A.16, appendix).

By applying this model to the apo-Rd data, we were able to obtain two melting temperature values of 49 ± 2 °C and 75 ± 3 °C, for the first and second transitions. The top panel in Figure 7.6A shows the fitted temperature dependence at specific wavelengths (Figure A.16, appendix, for correspondent species distributions). It is possible to observe that the large difference in melting temperatures allows for the intermediate species to accumulate, accounting for approximately 88 % of the total population around 61 °C. Thus, the spectrum acquired at this temperature is representative of the intermediate state structure. The prominent peak at 231 nm is lost, with the two other peaks being shifted to higher energies at 201 and 186 nm. Upon increasing the temperature this structure is lost and the protein becomes mainly structurally unordered. In the case of the holo-Rd, (Figure 7.6B) one observes a decrease in the absolute value of $\Delta\varepsilon$ over the monitored range of temperatures. If a three species model is used to rationalize spectral changes data 63 ± 2 °C and 79 ± 2 °C for the first and second transitions are obtained. The smaller difference in melting temperatures makes for

a less abrupt change CD signal as well as a lower percentage buildup of the intermediate species. It must be noted that, while the data prompts us to discard a single step process, more data will be needed to fully characterize thermal denaturation in both apo and holo proteins. For example, it will be important to probe different temperature ranges and possible reversibility of the denaturation process.

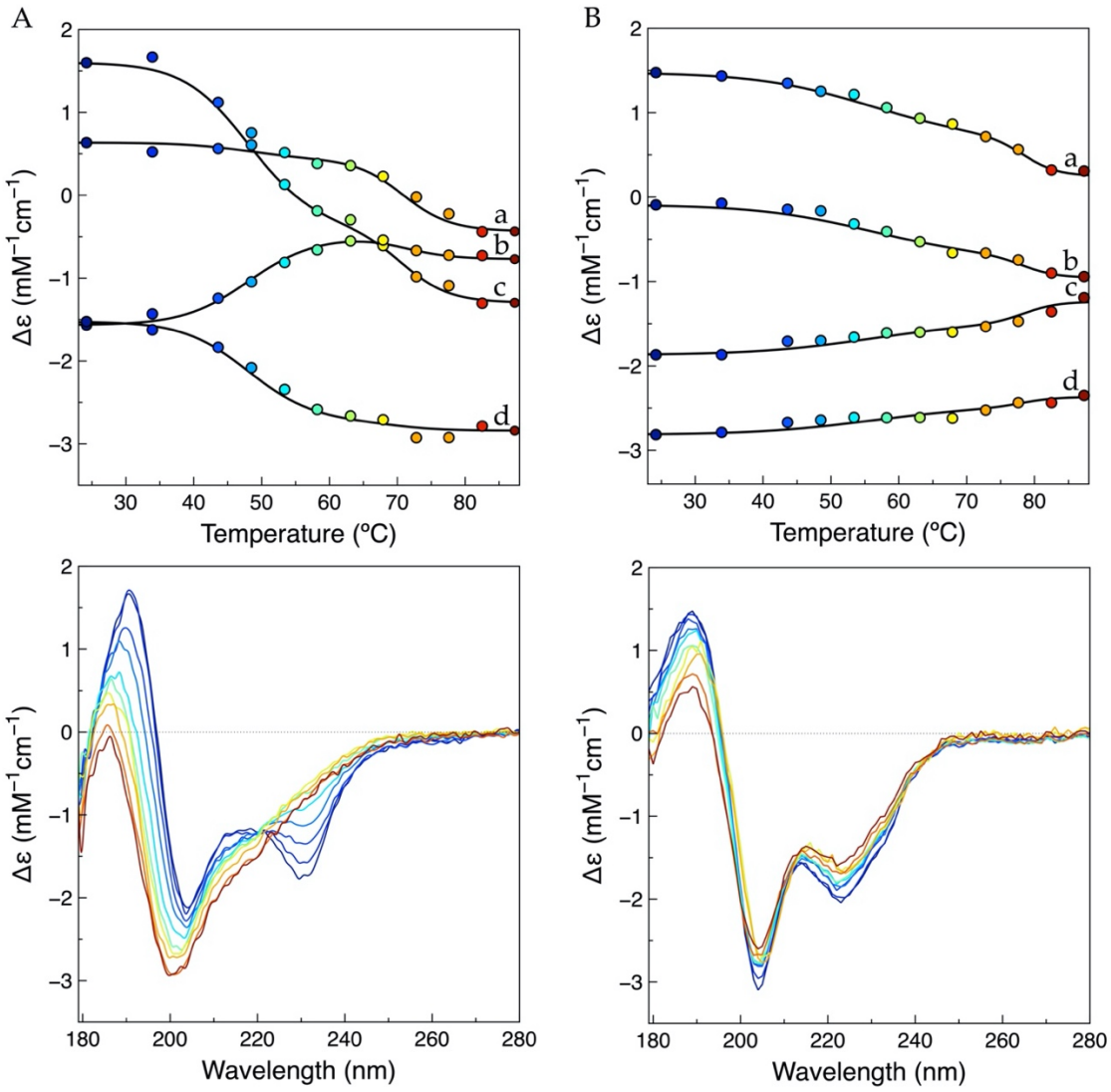

Figure 7.6: Secondary structure analysis upon thermal denaturation of apo-Rd and reconstituted Rd. (A) SRCD spectra of recombinant apo-Rd (Bottom panel) protein collected from 24 to 87 °C (spectra colors in accordance with temperature top scale). SRCD temperature profiles (Top panel). Δε at 184 (a), 233 (b), 191 (c), and 200 nm (d) as function of temperature. The solid lines are the results from a simultaneous non-linear least-squares fit to the data based on a two-step process. **(B)** SRCD spectra of recombinant Rd (Bottom panel) protein collected from 24 to 87 °C (spectra colors in accordance with temperature top scale). SRCD temperature profiles (Top panel). Δε at 189 (a), 196 (b), 226 (c), and 206 nm (d) as function of temperature. The solid lines are the result from a simultaneous non-linear least-squares fit to the data based on a two-step process.

7.2.3.5 Conclusion

The results presented here establish that the gene product of sequence ABM20661.1 from *Mar. hydrocarbonoclasticus* belongs to the Rd protein class of iron-sulfur proteins. The recombinant Rd protein is expressed in a holo metal-containing form. The apo form can be obtained without need of a reducing agent and/or anaerobic environment, with the apo-Rd being stable even in the presence of molecular oxygen partial pressures above the atmospheric one. SRCD data clearly shows that apo-Rd has a defined solution structure that can be stoichiometrically converted to the holo-Rd structure upon metal incubation. This reconstitution reaction can be accomplished under a molecular oxygen atmosphere. This differentiates Rd from other characterized Rd proteins for which thiol-reducing agent (usually dithiothreitol or 2-mercaptoetanol) is used in order to avoid cystine formation, with metal reconstitution performed under inert atmosphere incubation (usually argon). Apo- and holo-Rd structure temperature stability was tested, and it was found that the process is best described as a non-two-state denaturation. In this case, an intermediate species is formed during the denaturation process. From comparison of the results obtained for apo- and holo-Rd it is possible to conclude that the holo form is more stable than the apo form, with a calculated melting temperature for the first transition being ca. 14 °C higher. This is not surprising since the cysteinyl coordinated iron should further stabilize the protein native structure. However, the melting temperature for the second transition is the same within experimental error, meaning that temperature stability of both intermediate species is similar. A possible interpretation for this fact is that, above 70 °C, both intermediate structures will promote aggregation of polypeptide chains leading to a similar denaturation process. While the loss in absorbance at 205 nm (Figure A.17, appendix) seems to agree with this interpretation, further data is needed to support such conclusion. It is noteworthy that it is possible in the apo form for the intermediate species to accumulate to almost 90 % due to the thermal denaturation characteristics described above. The existence of an intermediate form has been observed for holo-Rubredoxins in a Single-Molecule Force Spectroscopy study [228]. In this work the authors used atomic force microscopy to probe the unfolding–refolding process of a cys-Rubredoxin-GB1-cys protein chimera and were able to show that such a process was reversible and proceeded via a Fe-$(SCys)_2$ intermediate species. In the case of this mechanical-induced unfolding, the iron ion was strictly required for the refolding process. A study conducted under anaerobic conditions on the holo-Rd from the mesophile *Clo. pasteurianum* also pointed to the existence of an intermediate species [229], showing that thermal stability was dependent on the metal binding (iron, zinc or cadmium) and that major structural changes would occur between 50 and 70 °C. In the case of Rd of the hyperthermophile *Pyr. furiosus*, different studies found that neither its folding nor the presence of iron justified by itself the high thermal stability of the protein. The tight packing of aromatic residues forming a hydrophobic core seems to contribute significantly to thermal stability and to induce similar solution structures for apo and holo forms [230]. Also, a complex unfolding process was described with the possible formation of at least three kinetic intermediates [231]. Contrary to this finding, for the Rd of the thermophile *Methanococcus*

jannaschii, it was reported a two-state denaturation process [232]. The aforementioned studies showed that structural stability was dependent in several factors, one being the non-covalent molecular interactions between key amino acid residues. Taken together with our results, it is possible that Trp4, Trp30 and Trp37 are key residues that support apo-Rd folding, not allowing for cystine formation, thus upholding the observed stable solution structure. The apo form can, from a physiological point of view, point to Rd functions not yet described. *Mar. hydrocarbonoclasticus* is considered to be potentially important for biotechnology and/or environmental applications since it is a hydrocarbon-degrading bacterial species. This species of eubacteria can be found in a wide variety of marine environments [233,234] and has the ability to grow under either anaerobic or aerobic conditions. However, aerobic conditions are required for hydrocarbon degradation. Under anaerobic growth it is able to reduce nitrate to dinitrogen, thus behaving as a denitrifier organism. In aerobic conditions molecular oxygen is used as the terminal electron acceptor. *Mar. hydrocarbonoclasticus* Rd can thus play an important role in both anaerobic and aerobic metabolisms. Besides the well-known electron transfer function, the possibility of having a stable apo form capable of metal binding even in the presence of molecular oxygen makes apo-Rd a suitable transition metal acceptor/scavenger. One could certainly foresee Rd as a cytoplasmatic protein capable of being an iron acceptor, keeping it in a soluble form even in aerobic conditions. We are currently pursuing experimental work to further understand the structural reasons for apo form stability and its possible physiological role.

7.2.4 Iron Release from a Mini-ferritin

The organism *Mar. hydrocarbonoclasticus* is a facultative anaerobe. Since molecular oxygen is a major enhancer of iron toxicity due to Fenton chemistry, the iron release mechanism was carried out both in the presence and absence of oxygen. Additionally, two different iron acceptors were used, an organic compound (1,10-phenanthroline) and a putative protein partner (apo-Rd).

7.2.4.1 Iron Release using 1,10-phenanthroline as Iron Chelator

1,10-phenanthroline was selected as iron chelator due to its high affinity to Fe^{2+} ions forming ferroin and its spectral properties that allow the monitorization of the release of Fe^{2+} ions from the ferric mineral in the Dps nanocage by UV-Visible spectroscopy. Although the size of 1,10-phenanthroline is similar to the width of the ferritin-like pores in Dps (7 – 10 Å), the latter are negatively charged, which limits its entrance of the neutral 1,10-phenanthroline into the Dps cavity.

7.2.4.1.1 Components of the Iron Release Assay

The individual spectral contribution of each component was evaluated by UV-Visible spectroscopy. At 510 nm, the maximum absorption peak of ferroin complex (the reaction product), no significant contribution of the other reagents was detected (Figure 7.7A). The spectrum characteristic of oxidized FMN-WrbA is lost upon addition of NADH due to FMN reduction. The only other component with absorption at 510 nm is Dps:144Fe for which the absorbance intensity corresponds to 2 % of the final ferroin signal. Therefore, this contribution was considered negligible.

The iron release assay is displayed in Figure 7.7B. As previously mentioned, due to the release of Fe^{2+} ions from Dps the ferroin complex gradually accumulates with time. Additionally, NADH was consumed during the reaction while reducing the ferric mineral core, as observed by the decay of the absorbance below 375 nm. The concentration and relative percentage of the iron released along the reaction were determined using the absorbance at 510 nm and the molar extinction coefficient of the ferroin complex.

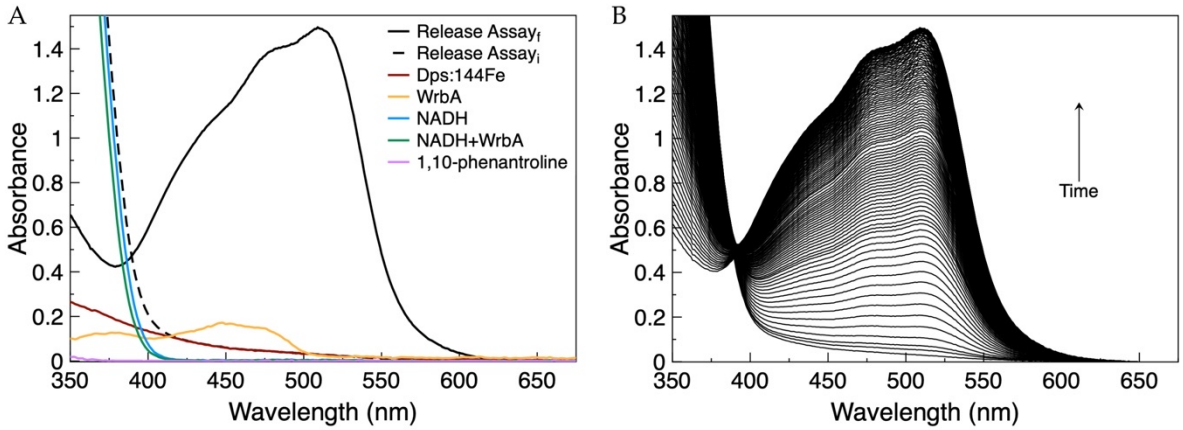

Figure 7.7: Reference UV-Visible spectra of the components of the iron release reaction and a typical example of an iron release assay. (A) Spectra of the reaction mixture at the starting point, $t = 0$ h (dashed line) and end point, $t = 24$ h (black solid line) of the iron release assay. Reference spectra of the component at the final concentration in the reaction mixture: 1 μM Dps:144 μM Fe (red line); 12 μM FMN-WrbA (yellow line); 720 μM NADH (blue line); 12 μM FMN-WrbA mixed with 720 μM NADH (green line); and 720 μM 1,10-phenanthroline (purple line). **(B)** A typical iron release assay was performed in anaerobic conditions. Reaction started after the addition of NADH. Measurements were performed every 15 min, for 24 h under anaerobic condition.

7.2.4.1.1 Iron Release under Aerobic Conditions

Two sequential assays were performed to study the impact of atmospheric oxygen in iron reduction and release from Dps. First, 1,10-phenanthroline was added at the reaction starting point and in the second assay the metal chelator was added 120 min after starting the reaction (Figure 7.8).

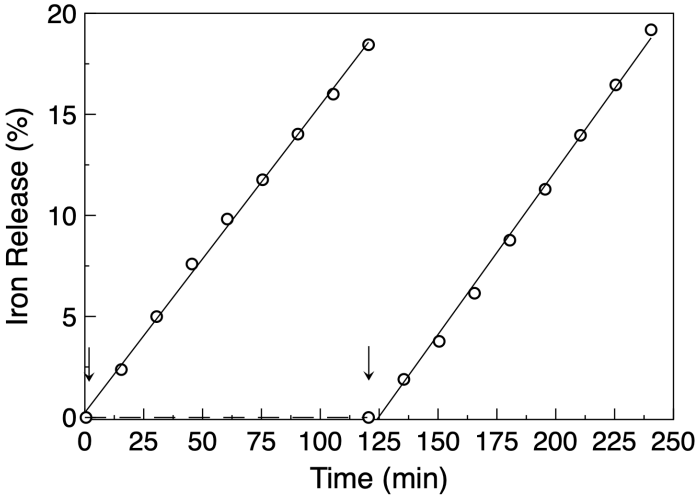

Figure 7.8: Iron release kinetics using 1,10-phenanthroline under atmospheric oxygen conditions. A reaction mixture containing 1 μM Dps:144 μM Fe, 12 μM WrbA, 720 μM NADH was prepared and then 720 μM 1,10-phenanthroline was added either at the starting of the reaction (left arrow) or after 120 min (right arrow) and the formation of the ferroin complex was monitored at 510 nm. The assays were performed in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. The spectra were acquired in 15 min intervals under atmospheric conditions.

It is noticeable that the iron release follows a zero-order kinetics behavior. In biological systems zero-order kinetics are observed in enzyme saturation conditions or found in diffusional exchanges controlled by pores [235,236]. In the experimental conditions tested, the iron release rate from Dps was determined to be 0.152 min^{-1}. Additionally, as previously reported using Mössbauer spectroscopy, the reduction of the mineral core follows a first-order kinetics with a rate of 0.011 ± 0.0003 min^{-1} [131]. No accumulation of iron outside the protein was detected when 1,10-phenanthroline was added ~ 120 min after the reduction starting point. If reduced ferrous ions had been released and accumulated outside the protein, a significant ferroin absorption signal increase should have been detected immediately after addition of 1,10-phenanthroline chelating agent. However, the rate of iron release occurred similarly to the assay when 1,10-phenanthroline was added at the initial of the reaction, following zero-order kinetics, at a rate of 0.162 min^{-1}. Therefore, under atmospheric conditions and in the absence of a metal chelator, ferric iron is either reduced but not released or it is instantly re-incorporated and re-oxidized at the FOCs. However, in the presence of a chelator such as 1,10-phenanthroline, the iron is immediately chelated upon release, preventing it re-oxidation by the protein, thus favoring the further iron release process into solution.

7.2.4.1.1 Iron Release under Anaerobic Conditions

The presence of a metal chelator was shown above to be essential to guarantee the removal of Fe^{2+} ions from the mini-ferritin in aerobic conditions. However, the effect of 1,10-phenanthroline (or another iron acceptor) in the iron release reaction in anaerobic conditions is unknown. Therefore, 1,10-phenanthroline was added to the kinetic mixture at different incubation times after starting the reaction. The results of the addition of 1,10-phenanthroline after 0, 155, 274, 414 and 610 min of reaction are shown in Figure 7.9.

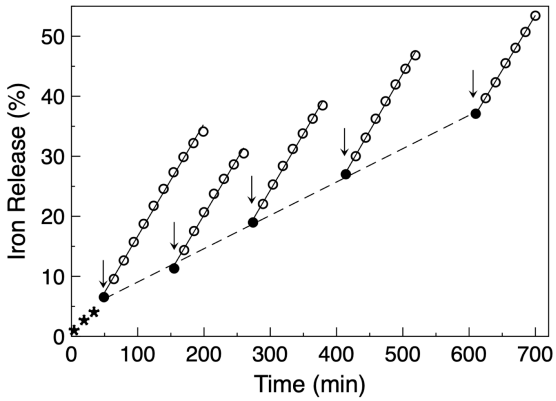

Figure 7.9: Iron release kinetics in anaerobic conditions upon addition of 1,10-phenanthroline at different times. A reaction mixture containing 1 μM Dps:144 μM Fe, 12 μM WrbA, 720 μM NADH and 720 μM 1,10-phenanthroline was prepared under anaerobic conditions. 1,10-phenanthroline was added 0, 155, 274, 414 and 610 min after addition of NADH to start the reaction. The assays were performed in 200 mM MOPS buffer, pH 7.0 and 200 mM NaCl. The spectra were acquired every 15 min under anaerobic conditions. The absorbance was immediately measured after 1,10-phenanthroline addition (full circle) and every 15 min afterwards (empty circles). The three initial time-points (black stars) were not used in data fitting. The dotted line represents the linear fit of the release

reaction without iron acceptor and each solid line represents the linear fit of the reaction after addition of 1,10-phenanthroline at each time-point.

Contrary to the release reaction in the presence of oxygen, the addition of 1,10-phenanthroline at different stages of the reaction under anaerobic conditions led to the instantaneous formation of a considerable amount of ferroin product due to the release of Fe^{2+} ions from the Dps protein into the buffer. As before, the release reaction follows a zero-order kinetics both with and without 1,10-phenanthroline (full lines vs dotted lines). However, the release rate in the absence of acceptor is 3.4-fold slower (0.056 min^{-1}) when compared with the release rate in the presence of the metal chelator. After the addition of the metal chelator, the release occurs at a rate similar to the one previously determined when 1,10-phenanthroline was added at the initial of the reaction (0.187 ± 0.003 min^{-1}).

7.2.4.2 Iron Release in the Presence of Rubredoxin as a Protein Partner

Rapid-SRCD was used to evaluate the binding kinetics of apo-Rd to ferrous iron in solution. The purpose of this study was to assess if apo-Rd was a good iron acceptor protein. Although preliminary, the results in Figure A.18, appendix section showed that the apo-Rd undergoes conformational changes after iron addition, adopting the secondary structure of the holo-form within 1.0 s. These kinetic data indicate that apo-Rd is perfectly suitable to be used as an iron acceptor protein partner in Dps iron release reaction since the iron release is still the rate limiting step from the kinetics perspective.

7.2.4.2.1 Iron release under anaerobic conditions

Iron release from Dps:144Fe in the presence of apo-Rd and under anaerobic conditions was analyzed using Mössbauer spectroscopy (Figure 7.10). At the first reaction time-point (5 min), the spectrum at 80 K shows a quadrupole doublet with parameters typical of a high-spin ferric species coordinated with oxygen/nitrogen ligands from the mini-ferritin, as previously described for *Mar. hydrocarbonoclasticus* Dps [125], isomer shift (δ) of 0.47 ± 0.02 mm/s and ΔE_Q of 0.58 ± 0.02 and 0.99 ± 0.02 mm/s (species Ia and Ib, see Table 7.3). Therefore, in the starting point of the reaction mixture all ^{57}Fe atoms are found in the form of the ferric mineral core. Along the reduction and release reactions (from 30 to 400 min) a distinct quadrupole doublet signal appears with parameters $\delta = 0.69$ mm/s and $\Delta E_Q = 3.25$ mm/s (Table 7.3). These parameters are characteristic of high-spin ferrous ions with tetrahedral sulfur coordination as previously reported for ^{57}Fe-reconstituted *Mar. hydrocarbonoclasticus* Rd under anaerobic conditions [237]. Thus, the ferric iron from the mineral core was reduced, released and bound to the Rd sulfur-center in its ferrous form. Although the iron was reduced before binding to Rd, no free ferrous iron was detected along the assay. The parameters of ferrous iron would be distinctive enough to allow its detection ($\delta = 1.35$ mm/s and $\Delta E_Q = 3.19$ mm/s [125]) indicating the rapid translocation of Fe^{2+} ions

through the channels, their release and binding to Rd, preventing its accumulation as labile iron.

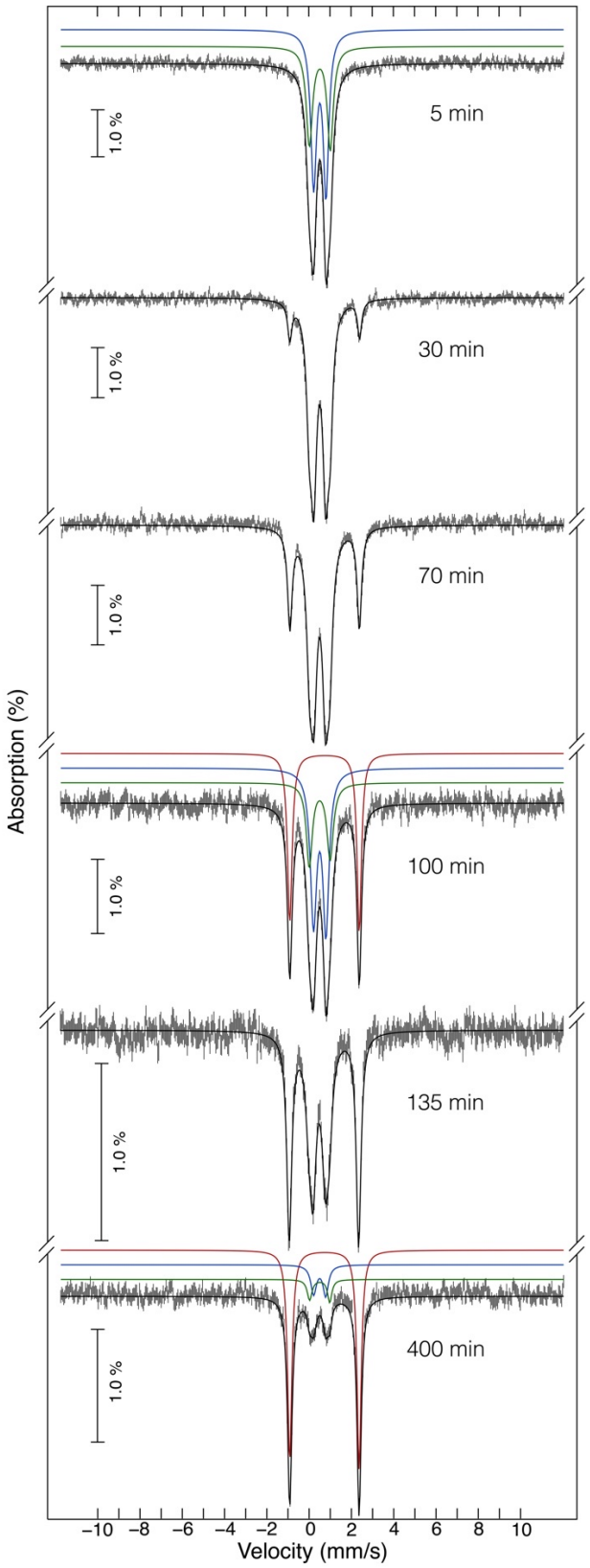

Figure 7.10: Mössbauer spectra of the iron release reaction from Dps in the presence of apo-Rd under anaerobic conditions. Dps harboring a mineral core (144 ^{57}Fe atoms per protein) was incubated with WrbA, NADH and apo-

Rd under anaerobic conditions. The iron release reaction was monitored at different reaction times after addition of NADH (5, 30, 70, 100, 135 and 400 min, top to bottom) using Mössbauer spectroscopy. The spectra were recorded at 80 K without an external magnetic field. The black solid lines are the result of a least squares fits using the parameters listed in Table 7.3. The blue and green lines represent the deconvolution of the Dps ferric mineral core species Ia and Ib, respectively. The red line represents the Rd ferrous iron-sulfur center spectral contributions.

Table 7.3: Mössbauer parameters of the iron species detected on the iron release reaction from *Mar. hydrocarbonoclasticus* Dps in the presence of apo-Rd, under anaerobic conditions.

	Mössbauer parameters of iron species		
	Ferric species		**Ferrous Iron-Sulfur Center**
	Ia	**Ib**	
δ (mm/s)	0.47(2)		0.69(1)
ΔE_Q (mm/s)	0.57(2)	0.97(2)	3.29(3)
Linewidth (mm/s)	0.35(2)	0.36(3)	0.27(2)
	Quantification of iron species (%)		
Time (min)			
5	58.5	40.9	0.60
30	58.7	32.4	8.90
70	45.0	31.3	23.7
100	42.9	21.8	35.3
135	34.2	20.8	45.0
400	13.7	8.30	78.0

The values in parentheses are the uncertainties of the species parameters.

The deconvolution of each spectrum into two ferric species (green and blue line) and one ferrous sulfur-center species (red line) (Table 7.3) allowed the determination of the relative contribution of each species throughout the reaction (Figure 7.11). As previously reported, the release reaction in the presence of apo-Rd follows zero-order kinetics (up to 135 min, see Figure 7.11) with a kinetic rate of 0.347 ± 0.009 min^{-1}. This is a ~ 2-fold increase in overall rate comparing to the release in the presence of iron chelator 1,10-phenanthroline, suggesting that apo-Rd facilitates ferrous iron release. Comparing the reduction kinetics [131] in the absence of a metal chelator and the release kinetics in the presence of Rd, an accumulation of ferrous iron was expected (Figure A.19, appendix). These results suggest that in the presence of an iron acceptor protein, the reduction of ferric iron and dissolution of the mineral is slowed down, representing the rate-limiting step. The presence of a large excess of apo-Rd (2.5 mM) compared to Dps (10.4 μM) and WrbA (125 μM) could physically impair electron transport between WrbA and Dps affecting electron transfer and the mineral core reduction rate.

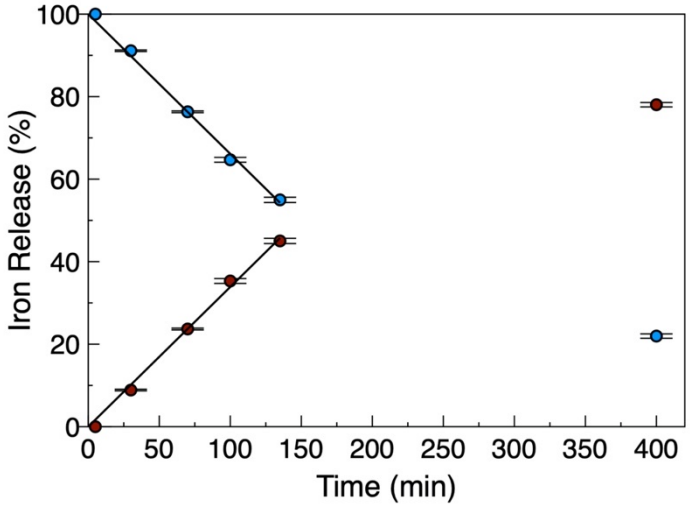

Figure 7.11: Deconvolution of the Mössbauer spectra monitoring iron release from Dps in the presence of apo-Rd under anaerobic conditions. The deconvolution of each spectrum into ferric (blue circles) and ferrous-sulfur center (red circles) species allows the determination of the percentage of iron released along the reaction. The initial data points were fitted using a zero-order reaction model (Equation 12) (solid lines).

7.3 Discussion and Conclusions

In this work we successfully produced Dps and WrbA following the protocols previously established and Rd expression was successfully achieved as a soluble protein. Protein activity was confirmed by UV-Visible spectroscopy: The iron oxidation activity of Dps was confirmed under oxygen saturation conditions; WrbA was expressed bound to FMN in its oxidized form, and the addition of NADH led to the reduction to $FMNH_2$ and subsequent loss of spectral features; monomeric apo-Rd was prepared by precipitation and iron chelation and the iron sulfur-center was reconstituted through iron addition in atmospheric conditions, recovering the spectral properties characteristic of the holo-form.

These proteins were used to characterize the iron reduction and release mechanism of *Mar. hydrocarbonoclasticus* Dps in the absence of any iron chelator and in the presence of either an organic compound (1,10-phenanthroline) or a protein partner (rubredoxin) as iron acceptor. The effect of the presence of oxygen was also explored due to the facultative anaerobe nature of this bacterium.

While the previously described iron reduction mechanism demonstrated an exponential behavior, the results monitoring iron release from the protein reveal a zero-order reaction [131]. As such, it is proposed that the Dps protein pores play a major role in controlling the bioavailability of iron, as well as the presence and absence of molecular oxygen and the type of iron acceptor present in solution. Under atmospheric conditions and in the absence of a metal chelator, the iron is not accumulated outside the mini-ferritin even after core reduction and dissolution, preventing the formation of ROS by labile iron. Upon addition of 1,10-phenanthroline, the release mechanism exhibited a similar rate (0.152 and 0.161 min^{-1}) regardless of when the iron chelator was added to the reaction mixture. However, under anaerobic conditions the release of reduced iron ions occurred even in the absence of a metal chelator. This suggests that the in the presence of oxygen the iron is probably also released but is immediately re-incorporated by the mini-ferritin and re-oxidized at the FOCs. In anaerobic conditions, the release rate is enhanced 3.4-fold upon 1,10-phenanthroline addition (0.056 min^{-1} and 0.187 ± 0.003 min^{-1}, in the absence and presence of the iron chelator). Moreover, the release reaction kinetics in the presence of 1,10-phenanthroline is not affected by the presence of oxygen. The use of 1,10-phenanthroline as iron acceptor can mimic the biologic behavior of other organic/inorganic compounds that bind iron in bacteria such as citrate or glutathione [238].

Rubredoxin was suggested as a putative physiologic iron-accepting protein partner due to its electron transfer function in different metabolic pathways. Although preliminary the apo-Rd seems able to bind all iron in solution within 1 sec of incubation.

Iron release from Dps the presence of Rd was studied using Mössbauer spectroscopy. At the initial phase of the reaction all iron atoms exhibited a quadrupole doublet characteristic of the mineral core with typical parameters of ferric iron bound to of nitrogen/oxygen ligands ($\Delta E_Q = 0.57 \pm 0.02$ and 0.97 ± 0.02 mm/s and $\delta = 0.47 \pm 0.02$ mm/s). After 30 min of reaction ~ 20 % the iron was reduced and transferred from the ferric mineral core to the Rd iron

binding site ($\delta = 0.69 \pm 0.01$ mm/s and $\Delta E_Q = 3.29 \pm 0.03$ mm/s, typical of ferrous iron with tetrahedral sulfur coordination). Based on the reduction rate previously described [131], an accumulation of ferric species would have been expected, yet no free ferrous species was detected. This suggests that the presence of apo-Rd may interfere with the interaction between Dps and its redox partner WrbA, decreasing the reduction rate. In agreement, the release reaction using 1,10-phenanthroline follows a zero-order kinetics (up to 135 min) with a rate constant of 0.347 ± 0.009 min^{-1}. The ~ 2-fold increase in the rate of release in the presence of apo-Rd comparing with 1,10-phenanthroline may indicate a more effective protein-protein interaction between the mini-ferritin and Rd, allowing a faster iron transfer.

This work clarified the effect of molecular oxygen in the iron release mechanism from Dps, as well as the interaction of the mini-ferritin with iron chelators. However, it is still necessary to understand why Rd apparently slows down the reduction of the mineral core. Does Rd block the interaction between Dps and WrbA? Does this less efficient reduction affect the overall release mechanism?

Additionally, the oxygen effect was tested only in atmospheric conditions. However, facultative anaerobes may live in environments with different oxygen levels. Therefore, it would be interesting to understand if and how the iron release mechanism in the absence of a metal chelator would be affected by different concentrations of oxygen. As such, additional studies are needed to answer these questions and have a deeper understanding of this complex system.

8

CONCLUSIONS AND FUTURE WORK

The research work developed during this thesis focused on the characterization of two types of protein nanocompartments: The functional and structural characterization of a member of the Encapsulin Family 1 and the iron release mechanism by Dps, a mini-ferritin.

The main goal for the first part of this work was to contribute to the biochemical and biophysical characterization of the *Myxococcus xanthus* Encapsulin system. The encapsulin shell (EncA) and its known cargo proteins (EncB and EncC), as well as their encapsulated forms (EncAB and EncAC) were successfully produced and purified. Even though EncB was obtained in the insoluble fraction, the protein was eventually obtained in a soluble one through the application of a renaturation protocol. Therefore, all proteins and complexes were obtained at high yields of pure and homogeneous fractions in the apo form. The encapsulation of the cargo proteins, the EncAB and EncAC complexes, obtained by co-expression in *Esc. coli* was determined to be 52 ± 4 copies of EncB monomers and 150 ± 7 copies of EncC monomers inside each EncA shell.

The results presented in chapter 4 confirmed that the encapsulation of the cargo proteins did not induce any major structural effect in the EncA envelope. The quaternary structure of EncB was determined as a decamer organized into a toroidal structure akin to other known Flp proteins. The presence of a large C-terminal region extending towards the solvent was also detected, matching the predicted unstructured tail that mediates the interaction with the shell protein, currently absent from the atomic structures available. Additionally, EncC was found in solution as an elongated dimer and the addition of ferrous iron (at least 4 Fe^{2+}/FOC) induced the conversion into higher oligomer states, ultimately forming a decamer. This result indicates the need for iron to be present to complete the stabilization of the FOC dimer interface in EncC.

Furthermore, the encapsulation of the cargo proteins by EncA had a protective effect against the thermal denaturation of EncB and EncC, either by preventing the loss of their secondary structure or by avoiding self-aggregation. The EncB aggregation event, which starts at ~ 45 °C in solution was stabilized at least up to ~ 65 °C in the EncAB complex, while the onset for EncC denaturation shifted from ~ 40 °C to ~ 70 °C upon encapsulation.

The iron oxidation kinetics and chemical nature of the iron species formed by the encapsulin proteins were evaluated in chapter 5. The intrinsic ability of EncA to mineralize iron was confirmed in the presence of either molecular oxygen and hydrogen peroxide, being able to uptake up to 24,000 and 19,500 iron atoms with each co-substrate. The iron mineralization activity of EncA was shown to be similar to the one observed in L-type ferritins, with the iron translocation and spontaneous oxidation steps occurring inside the cavity following a sigmoidal kinetic behavior due to the auto-catalytic property of the ferrihydrite mineral. The encapsulation of either EncB or EncC decreases the iron loading capacity to 18,000 ferrous iron atoms when hydrogen peroxide is used as a co-substrate. With molecular oxygen as substrate the iron oxidation kinetic curve shifts to a hyperbolic curve, revealing the rate enhancing activity of the encapsulin system when harboring these Flp cargo proteins. Also, the type of mineral core formed inside EncA and EncAC seems to be similar, described in both cases as the characteristic superparamagnetic ferric mineral in the presence of either co-substrate, indicating that the shell protein likely mediates the nucleation and mineralization process in both cases. A smaller, more homogeneous mineral core was detected in the reactions with hydrogen peroxide, suggesting different nucleation centers when compared with the product formed in the presence of molecular oxygen, in which the mineral core is organized into a larger granulate, as previously detected [21].

Free EncC in solution exhibited faster oxidation rates when compared to the encapsulated form. A ferroxidation intermediate species was identified as a μ-1,2-peroxo di-ferric species (with Mössbauer parameters $\delta = 0.49 \pm 0.02$ mm/s and $\Delta E_Q = 1.61 \pm 0.02$ mm/s), which might be justified by the proximity of di-nuclear iron binding sites (FOCs) at the interface of each EncC dimer [30]. However, this intermediate was not detected in the EncAC complex, probably due to its low half-life, which prevents its accumulation.

Chapter 6 described the interaction between the encapsulin shell and the encapsulated complexes with plasmid DNA. Electrophoresis Mobility Shift Assays revealed that the incubation of pUC19 with EncA changes the conformation of the plasmid from its supercoiled conformation to a new conformation with decreased electrophoretic mobility, evidencing a binding event. The apparent dissociation constant was determined to be $K_D = 0.3 \pm 0.1$ μM with a Hill coefficient of 1.4 ± 0.1, similar to other known protein nanocages [106,197,200]. Interestingly, EncC showed no ability to interact with DNA. However, upon encapsulation, the DNA-binding properties of the EncAC complex revealed to have similar parameters to EncA ($K_D = 0.4 \pm 0.1$ μM and 0.3 ± 0.1 μM at lower and higher ionic strengths, respectively, with $n = 1.1 \pm 0.1$). Additionally, the increase in ionic strength did not have a major effect in protein-DNA interaction. The DNA conformation in the novel DNA-encapsulin complex detected in the EMSAs was confirmed by Atomic Force Microscopy to be in a more relaxed conformation with EncA molecules bound to pUC19 in a beads-on-a-string morphology. The increase in protein concentration promoted DNA condensation through protein-protein self-aggregation. Moreover, this interaction was shown to protect the DNA against DNase I degradation, with EncA functioning as a physical shield. Regarding the protein, the interaction with pUC19 did not induce a major effect in its secondary structure nor in its thermostability

since the *Tm* only increased by 3 °C. Conversely, the EncA secondary structure denaturation profile revealed some notable differences after incubation with DNA, with a ~ 4-fold increase in the enthalpy change of denaturation. This increase can be due to a higher number of non-covalent bonds formed between the protein and the DNA molecules and/or the further denaturation of the protein secondary structure when the EncA–DNA complex had previously been formed.

Regarding the overall results obtained in the encapsulin system work, this system was shown to control the oxidative stress within the cell by harvesting ferrous irons in solution through iron oxidation and mineralization inside the cavity. Additionally, the Encapsulin shell is capable to bind and condensate circular DNA protecting this macromolecule as a physical shield. Therefore, this protein system has showed to act with a dual function by preventing the formation of reactive oxygen species capable of DNA degradation and by directly protecting the DNA similarly to Dps proteins. Therefore, these similarities between the encapsulin system and Dps proteins would be worth of research. Such as testing this dual function of encapsulin to interact with DNA while harboring a mineral core or the ability to perform iron oxidation while interacting with DNA.

In the future it would be of utmost interest to further characterize the changes in quaternary structure of EncC upon incubation of ferrous iron by Small-Angle X-ray Scattering. It would also be interesting to evaluate how EncB and EncC spatially organize within EncA cavity upon co-expression.

The next step regarding the iron oxidation properties of the encapsulin system would be to characterize the intermediate species formed by EncC (and EncB) and their encapsulated forms EncAB and EncAC using Rapid Freeze Quench coupled to Mössbauer spectroscopy, as well as to understand if the simultaneous presence of both cargo proteins would result in a different functional properties.

The DNA-binding ability of EncA must be further evaluated using different types of DNA such as genomic bacterial DNA and to evaluate if the interaction the DNA occurs in a site-specific manner, both in terms of protein interface and DNA sequence. These clarifications would increase this protein attractiveness for different DNA-related biotechnological applications.

In the second part of this thesis, Chapter 7, the iron release mechanism from *Mar. hydrocarbonoclasticus* Dps was probed using NADH as an electron donor, WrbA as an oxidoreductase protein partner and two types of iron chelators: 1,10-phenanthroline and a rubredoxin protein, in the presence and absence of molecular oxygen. In the absence of an iron chelator and under aerobic conditions, no iron ferrous is accumulated outside the protein. If any iron is released from the protein, the metal is most likely re-oxidized by the Dps protein. However, after addition of a 1,10-phenanthroline, the iron release follows a zero-order kinetic profile with an apparent rate of 0.157 ± 0.006 min^{-1}. In the absence of 1,10-phehanthroline and under anaerobic conditions, the iron release mechanism also follows a zero-order kinetics with

an apparent rate of 0.056 min^{-1}. Strikingly, the addition of 1,10-phenanthroline further enhances the kinetic rate of release (0.187 ± 0.003 min^{-1}).

Rubredoxin was also used to evaluate the iron release from the mini-ferritin. The holo form of rubredoxin has distinct Mössbauer parameters due to the presence of an iron-sulfur center (δ = 0.69 mm/s and ΔE_Q = 3.29 mm/s) and was confirmed using stopped-flow coupled to Synchrotron Radiation Circular Dichroism to bind iron in a time scale, significantly faster than the iron release. Under anaerobic conditions and in the presence of rubredoxin, iron release from Dps follows a zero-order kinetics (up to 135 min) with a rate of 0.347 ± 0.009 min^{-1}, approximately 2-fold faster compared to the reaction in the presence of 1,10-phenanthroline. Based on previous studies [129], an accumulation of reduced ferrous iron inside the protein was expected. Yet no free ferrous ions were detected, suggesting that the rubredoxin affected the electron transfer between WrbA and the Dps mineral core.

In the future it is important to further clarify how the presence of the iron acceptor protein affects the electron transfer reaction. Does the protein block the interaction between Dps and WrbA? Does rubredoxin somehow intersects the electron shuttle? Additionally, iron release from Dps and its rerouting to other iron harboring proteins such as Ferritins or Encapsulin-Flps (with a higher ferroxidation rate) would be interesting to study.

REFERENCE:

[1] S. Kang, T. Douglas, Some enzymes just need a space of their own, Science. 327 (2010) 42–43. https://doi.org/10.1126/science.1184318.

[2] C.A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments, Nat. Rev. Microbiol. 16 (2018) 277–290. https://doi.org/10.1038/nrmicro.2018.10.

[3] S. Tanaka, C.A. Kerfeld, M.R. Sawaya, F. Cai, S. Heinhorst, G.C. Cannon, T.O. Yeates, Atomic-level models of the bacterial carboxysome shell, Science. 319 (2008) 1083–1086. https://doi.org/10.1126/science.1151458.

[4] T.O. Yeates, C.A. Kerfeld, S. Heinhorst, G.C. Cannon, J.M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments, Nat. Rev. Microbiol. 6 (2008) 681–691. https://doi.org/10.1038/nrmicro1913.

[5] M. Held, M.B. Quin, C. Schmidt-Dannert, Eut bacterial microcompartments: Insights into their function, structure, and bioengineering applications, Microb. Physiol. 23 (2013) 308–320. https://doi.org/10.1159/000351343.

[6] S. Choudhary, M.B. Quin, M.A. Sanders, E.T. Johnson, C. Schmidt-Dannert, Engineered protein nano-compartments for targeted enzyme localization, PLoS One. 7 (2012) e33342. https://doi.org/10.1371/journal.pone.0033342.

[7] Y. Azuma, T.G.W. Edwardson, D. Hilvert, Tailoring lumazine synthase assemblies for bionanotechnology, Chem. Soc. Rev. 47 (2018) 3543–3557. https://doi.org/10.1039/c8cs00154e.

[8] D. He, J. Marles-Wright, Ferritin family proteins and their use in bionanotechnology, N. Biotechnol. 32 (2015) 651–657. https://doi.org/10.1016/j.nbt.2014.12.006.

[9] T.O. Yeates, M.C. Thompson, T.A. Bobik, The protein shells of bacterial microcompartment organelles, Curr. Opin. Struct. Biol. 21 (2011) 223–231. https://doi.org/10.1016/j.sbi.2011.01.006.

[10] N. Valdés-Stauber, S. Scherer, Isolation and characterization of Linocin M18, a bacteriocin produced by *Brevibacterium linens*, Appl. Environ. Microbiol. 60 (1994) 3809–3814. https://doi.org/10.1128/aem.60.10.3809-3814.1994.

[11] I. Rosenkrands, P.B. Rasmussen, M. Carnio, S. Jacobsen, M. Theisen, P. Andersen, Identification and characterization of a 29-kilodalton protein from *Mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells, Infect. Immun. 66 (1998) 2728–2735. https://doi.org/10.1128/IAI.66.6.2728-2735.1998.

[12] P.M. Hicks, K.D. Rinker, J.R. Baker, R.M. Kelly, Homomultimeric protease in the hyperthermophilic bacterium *Thermotoga maritima* has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria, FEBS Lett. 440 (1998) 393–398. https://doi.org/10.1016/S0014-5793(98)01451-3.

[13] N. Winter, J.A. Triccas, B. Rivoire, M.C. V. Pessolani, K. Eiglmeier, E.-M. Lim, S.W. Hunter, P.J. Brennan, W.J. Britton, Characterization of the gene encoding the immunodominant 35 kDa protein of *Mycobacterium leprae*, Mol. Microbiol. 16 (1995) 865–876. https://doi.org/10.1111/j.1365-2958.1995.tb02314.x.

[14] N. Saito, K. Matsubara, M. Watanabe, F. Kato, K. Ochi, Genetic and biochemical characterization of EshA, a protein that forms large multimers and affects developmental processes in *Streptomyces griseus*, J. Biol. Chem. 278 (2003) 5902–5911. https://doi.org/10.1074/jbc.M208564200.

[15] M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M.J. Loessner, K.O. Stetter, E. Weber-Ban, N. Ban, Structural basis of enzyme encapsulation into a bacterial nanocompartment, Nat. Struct. Mol. Biol. 15 (2008) 939–947. https://doi.org/10.1038/nsmb.1473.

[16] F. Akita, K.T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara, A. Nakagawa, The crystal structure of a virus-like particle from the Hyperthermophilic Archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses, J. Mol. Biol. 368 (2007) 1469–1483. https://doi.org/10.1016/j.jmb.2007.02.075.

[17] W.R. Wikoff, L. Liljas, R.L. Duda, H. Tsuruta, R.W. Hendrix, J.E. Johnson, Topologically linked protein rings in the bacteriophage HK97 capsid, Science. 289 (2000) 2129–2133. https://doi.org/10.1126/science.289.5487.2129.

[18] T.W. Giessen, P.A. Silver, Widespread distribution of encapsulin nanocompartments reveals functional diversity, Nat. Microbiol. 2 (2017) 17029. https://doi.org/10.1038/nmicrobiol.2017.29.

[19] M.P. Andreas, T.W. Giessen, Large-scale computational discovery and analysis of virus-derived microbial nanocompartments, Nat. Commun. 12 (2021) 4748. https://doi.org/10.1038/s41467-021-25071-y.

[20] D. He, S. Hughes, S. Vanden-Hehir, A. Georgiev, K. Altenbach, E. Tarrant, C.L. Mackay, K.J. Waldron, D.J. Clarke, J. Marles-Wright, Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments, Elife. 5 (2016) 1–31. https://doi.org/10.7554/eLife.18972.

[21] 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, E. Hoiczyk, A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress, EMBO J. 33 (2014) 1896–1911. https://doi.org/10.15252/embj.201488566.

[22] H. Contreras, M.S. Joens, L.M. McMath, V.P. Le, M. V. Tullius, J.M. Kimmey, N. Bionghi, M.A. Horwitz, J.A.J. Fitzpatrick, C.W. Goulding, Characterization of a *Mycobacterium tuberculosis* nanocompartment and its potential cargo proteins, J. Biol. Chem. 289 (2014) 18279–18289. https://doi.org/10.1074/jbc.M114.570119.

[23] R.J. Nichols, B. LaFrance, N.R. Phillips, D.R. Radford, L.M. Oltrogge, L.E. Valentin-Alvarado, A.J. Bischoff, E. Nogales, D.F. Savage, Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism, Elife. 10 (2021). https://doi.org/10.7554/eLife.59288.

[24] D. Radford, Understanding the encapsulins: Prediction and characterization of phage capsid-like nanocompartments in prokaryotes, 2015 (online publication).

[25] T.W. Giessen, B.J. Orlando, A.A. Verdegaal, M.G. Chambers, J. Gardener, D.C. Bell, G. Birrane, M. Liao, P.A. Silver, Large protein organelles form a new iron sequestration system with high storage capacity, Elife. 8 (2019). https://doi.org/10.7554/eLife.46070.

[26] D.L.D. Caspar, A. Klug, Physical principles in the construction of regular viruses, Cold Spring Harb. Symp. Quant. Biol. 27 (1962) 1–24. https://doi.org/10.1101/SQB.1962.027.001.005.

[27] R. Rahmanpour, T.D.H. Bugg, Assembly in vitro of *Rhodococcus jostii* RHA1 encapsulin and peroxidase DypB to form a nanocompartment, FEBS J. 280 (2013) 2097–2104. https://doi.org/10.1111/febs.12234.

[28] 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, A.J.R. Heck, Assembly and mechanical properties of the cargo-free and cargo-loaded bacterial nanocompartment encapsulin, Biomacromolecules. 17 (2016) 2522–2529. https://doi.org/10.1021/acs.biomac.6b00469.

[29] R.M. Putri, C. Allende-Ballestero, D. Luque, R. Klem, K.-A. Rousou, A. Liu, C.H.H. Traulsen, W.F. Rurup, M.S.T. Koay, J.R. Castón, J.J.L.M. Cornelissen, Structural characterization of native and modified encapsulins as nanoplatforms for *in vitro* catalysis and cellular uptake, ACS Nano. 11 (2017) 12796–12804. https://doi.org/10.1021/acsnano.7b07669.

[30] E. Eren, B. Wang, D.C. Winkler, N.R. Watts, A.C. Steven, P.T. Wingfield, Structural characterization of the *Myxococcus xanthus* encapsulin and ferritin-like cargo system gives insight into its iron storage mechanism, Structure. 30 (2022) 551-563.e4. https://doi.org/10.1016/j.str.2022.01.008.

[31] R.L. Duda, Protein chainmail, Cell. 94 (1998) 55–60. https://doi.org/10.1016/S0092-8674(00)81221-0.

[32] A. Fokine, P.G. Leiman, M.M. Shneider, B. Ahvazi, K.M. Boeshans, A.C. Steven, L.W. Black, V. V. Mesyanzhinov, M.G. Rossmann, Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry, Proc. Natl. Acad. Sci. 102 (2005) 7163–7168. https://doi.org/10.1073/pnas.0502164102.

[33] P. Mahinthichaichan, D.M. Morris, Y. Wang, G.J. Jensen, E. Tajkhorshid, Selective permeability of carboxysome shell pores to anionic molecules, J. Phys. Chem. B. 122

(2018) 9110–9118. https://doi.org/10.1021/acs.jpcb.8b06822.

[34] T. Tosha, R.K. Behera, H.-L. Ng, O. Bhattasali, T. Alber, E.C. Theil, Ferritin protein nanocage ion channels, J. Biol. Chem. 287 (2012) 13016–13025. https://doi.org/10.1074/jbc.M111.332734.

[35] K. Zeth, Dps biomineralizing proteins: multifunctional architects of nature, Biochem. J. 445 (2012) 297–311. https://doi.org/10.1042/BJ20120514.

[36] 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, Y.H. Lau, Pore structure controls stability and molecular flux in engineered protein cages, Sci. Adv. 8 (2022) 2021.04.26.441214. https://doi.org/10.1126/sciadv.abl7346.

[37] T. Wiryaman, N. Toor, Cryo-EM structure of a thermostable bacterial nanocompartment, IUCrJ. 8 (2021) 342–350. https://doi.org/10.1107/S2052252521001949.

[38] 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, J. Marles-Wright, Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment, Sci. Adv. 8 (2022) 2021.04.15.439977. https://doi.org/10.1126/sciadv.abj4461.

[39] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L.E. Felberg, D.H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D.W. Gohara, T. Dolinsky, R. Konecny, D.R. Koes, J.E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M.J. Holst, J.A. McCammon, N.A. Baker, Improvements to the APBS biomolecular solvation software suite, Protein Sci. 27 (2018) 112–128. https://doi.org/10.1002/pro.3280.

[40] C. Cassidy-Amstutz, L. Oltrogge, C.C. Going, A. Lee, P. Teng, D. Quintanilla, A. East-Seletsky, E.R. Williams, D.F. Savage, Identification of a minimal peptide tag for *in vivo* and *in vitro* loading of encapsulin, Biochemistry. 55 (2016) 3461–3468. https://doi.org/10.1021/acs.biochem.6b00294.

[41] W.F. Rurup, J. Snijder, M.S.T. Koay, A.J.R. Heck, J.J.L.M. Cornelissen, Self-sorting of foreign proteins in a bacterial nanocompartment, J. Am. Chem. Soc. 136 (2014) 3828–3832. https://doi.org/10.1021/ja410891c.

[42] A. Tamura, Y. Fukutani, T. Takami, M. Fujii, Y. Nakaguchi, Y. Murakami, K. Noguchi, M. Yohda, M. Odaka, Packaging guest proteins into the encapsulin nanocompartment from *Rhodococcus erythropolis* N771, Biotechnol. Bioeng. 112 (2015) 13–20. https://doi.org/10.1002/bit.25322.

[43] A. Tsuda, R. Ishikawa, H. Koteishi, K. Tange, Y. Fukuda, K. Kobayashi, T. Inoue, M. Nojiri, Structural and mechanistic insights into the electron flow through protein for cytochrome c-tethering copper nitrite reductase, J. Biochem. 154 (2013) 51–60. https://doi.org/10.1093/jb/mvt023.

[44] J.N. Roberts, R. Singh, J.C. Grigg, M.E.P. Murphy, T.D.H. Bugg, L.D. Eltis, Characterization of dye-decolorizing peroxidases from *Rhodococcus jostii* RHA1, Biochemistry. 50 (2011) 5108–5119. https://doi.org/10.1021/bi200427h.

[45] S.J. Kim, M. Shoda, Purification and characterization of a novel peroxidase from *Geotrichum candidum* Dec 1 involved in decolorization of dyes, Appl. Environ. Microbiol. 65 (1999) 1029–1035. https://doi.org/10.1128/AEM.65.3.1029-1035.1999.

[46] R. Crichton, Inorganic biochemistry of iron metabolism, John Wiley & Sons, Ltd, Chichester, UK, 2001. https://doi.org/10.1002/0470845791.

[47] M. Sánchez, L. Sabio, N. Gálvez, M. Capdevila, J.M. Dominguez-Vera, Iron chemistry at the service of life, IUBMB Life. 69 (2017) 382–388. https://doi.org/10.1002/iub.1602.

[48] D. He, C. Piergentili, J. Ross, E. Tarrant, L.R. Tuck, C.L. Mackay, Z. McIver, K.J. Waldron, D.J. Clarke, J. Marles-Wright, Conservation of the structural and functional architecture of encapsulated ferritins in bacteria and archaea, Biochem. J. 476 (2019) 975–989. https://doi.org/10.1042/BCJ20180922.

[49] J. Ross, T. Lambert, C. Piergentili, D. He, K.J. Waldron, C.L. Mackay, J. Marles-Wright, D.J. Clarke, Mass spectrometry reveals the assembly pathway of encapsulated ferritins and highlights a dynamic ferroxidase interface, Chem. Commun. 56 (2020) 3417–3420. https://doi.org/10.1039/C9CC08130E.

[50] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539. https://doi.org/10.1038/msb.2011.75.

[51] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) W320–W324. https://doi.org/10.1093/nar/gku316.

[52] C. Chang, E. Evdokimova, A. Savchenko, A.M. Edwards, A. Joachimiak, Crystal structure of protein Ne0167 from Nitrosomonas Europaea, To Be Publ. (n.d.). https://doi.org/10.2210/pdb3K6C/pdb.

[53] D.R. Hall, J.M. Hadden, G.A. Leonard, S. Bailey, M. Neu, M. Winn, 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 Crystallogr. Sect. D Biol. Crystallogr. 58 (2002) 70–80. https://doi.org/10.1107/S0907444901017309.

[54] K. Honarmand Ebrahimi, P.-L. Hagedoorn, W.R. Hagen, Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin, Chem. Rev. 115 (2015) 295–326. https://doi.org/10.1021/cr5004908.

[55] P. Nordlund, H. Eklund, Di-iron—carboxylate proteins, Curr. Opin. Struct. Biol. 5 (1995) 758–766. https://doi.org/10.1016/0959-440X(95)80008-5.

[56] C. Piergentili, J. Ross, D. He, K.J. Gallagher, W.A. Stanley, L. Adam, C.L. Mackay, A. Baslé, K.J. Waldron, D.J. Clarke, J. Marles-Wright, Dissecting the structural and functional roles of a putative metal entry site in encapsulated ferritins, J. Biol. Chem. 295 (2020) 15511–15526. https://doi.org/10.1074/jbc.RA120.014502.

[57] S.C. Andrews, Iron storage in bacteria, in: Adv. Microb. Physiol., 1998: pp. 281–351. https://doi.org/10.1016/S0065-2911(08)60134-4.

[58] P.M. Harrison, P. Arosio, The ferritins: molecular properties, iron storage function and cellular regulation, Biochim. Biophys. Acta - Bioenerg. 1275 (1996) 161–203. https://doi.org/10.1016/0005-2728(96)00022-9.

[59] A.M. Demchuk, T.R. Patel, The biomedical and bioengineering potential of protein nanocompartments, Biotechnol. Adv. 41 (2020) 107547. https://doi.org/10.1016/j.biotechadv.2020.107547.

[60] Y.H. Lau, T.W. Giessen, W.J. Altenburg, P.A. Silver, Prokaryotic nanocompartments form synthetic organelles in a eukaryote, Nat. Commun. 9 (2018) 1311. https://doi.org/10.1038/s41467-018-03768-x.

[61] T. Lee, T.S. Carpenter, P. D'haeseleer, D.F. Savage, M.C. Yung, Encapsulin carrier proteins for enhanced expression of antimicrobial peptides, Biotechnol. Bioeng. 117 (2020) 603–613. https://doi.org/10.1002/bit.27222.

[62] M.G. Scott, H. Yan, R.E.W. Hancock, Biological properties of structurally related α-helical cationic antimicrobial peptides, Infect. Immun. 67 (1999) 2005–2009. https://doi.org/10.1128/.67.4.2005-2009.1999.

[63] A.P. Jathoul, J. Laufer, O. Ogunlade, B. Treeby, B. Cox, E. Zhang, P. Johnson, A.R. Pizzey, B. Philip, T. Marafioti, M.F. Lythgoe, R.B. Pedley, M.A. Pule, P. Beard, Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter, Nat. Photonics. 9 (2015) 239–246. https://doi.org/10.1038/nphoton.2015.22.

[64] 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, G.G. Westmeyer, Bacterial encapsulins as orthogonal compartments for mammalian cell engineering, Nat. Commun. 9 (2018) 1990. https://doi.org/10.1038/s41467-018-04227-3.

[65] D. Diaz, X. Vidal, A. Sunna, A. Care, Bioengineering a light-responsive encapsulin nanoreactor: A potential tool for *in vitro* photodynamic therapy, ACS Appl. Mater. Interfaces. 13 (2021) 7977–7986. https://doi.org/10.1021/acsami.0c21141.

[66] Y. Bae, G.J. Kim, H. Kim, S.G. Park, H.S. Jung, S. Kang, Engineering tunable dual functional protein cage nanoparticles using bacterial superglue, Biomacromolecules. 19 (2018) 2896–2904. https://doi.org/10.1021/acs.biomac.8b00457.

[67] H. Moon, J. Lee, J. Min, S. Kang, Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform, Biomacromolecules. 15 (2014) 3794–3801. https://doi.org/10.1021/bm501066m.

[68] R. Klem, M. V. de Ruiter, J.J.L.M. Cornelissen, Protecting encapsulin nanoparticles with cysteine-knot miniproteins, Mol. Pharm. 15 (2018) 2991–2996. https://doi.org/10.1021/acs.molpharmaceut.8b00630.

[69] J. Kim, S. Park, J.E. Lee, S.M. Jin, J.H. Lee, I.S. Lee, I. Yang, J.-S. Kim, S.K. Kim, M.-H. Cho, T. Hyeon, Designed fabrication of multifunctional magnetic gold nanoshells and their application to magnetic resonance imaging and photothermal therapy, Angew. Chemie Int. Ed. 45 (2006) 7754–7758. https://doi.org/10.1002/anie.200602471.

[70] T.W. Giessen, P.A. Silver, Converting a natural protein compartment into a nanofactory

for the size-constrained synthesis of antimicrobial silver nanoparticles, ACS Synth. Biol. 5 (2016) 1497–1504. https://doi.org/10.1021/acssynbio.6b00117.

[71] J. Muñoz-Dorado, F.J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz, J. Pérez, Myxobacteria: Moving, killing, feeding, and surviving together, Front. Microbiol. 7 (2016) 1–18. https://doi.org/10.3389/fmicb.2016.00781.

[72] L.J. Shimkets, Intercellular Signaling During Fruiting-Body Development of *Myxococcus xanthus*, Annu. Rev. Microbiol. 53 (1999) 525–549. https://doi.org/10.1146/annurev.micro.53.1.525.

[73] M. Antonio, D. Schulze-Makuch, Toward a new understanding of multicellularity, Hypotheses Life Sci. 2 (2012) 4–14.

[74] G.J. Velicer, L. Kroos, R.E. Lenski, Loss of social behaviors by *Myxococcus xanthus* during evolution in an unstructured habitat, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12376–12380. https://doi.org/10.1073/pnas.95.21.12376.

[75] J.E. Berleman, J.R. Kirby, Deciphering the hunting strategy of a bacterial wolfpack, FEMS Microbiol. Rev. 33 (2009) 942–957. https://doi.org/10.1111/j.1574-6976.2009.00185.x.

[76] J. Pérez, A. Moraleda-Muñoz, F.J. Marcos-Torres, J. Muñoz-Dorado, Bacterial predation: 75 years and counting!, Environ. Microbiol. 18 (2016) 766–779. https://doi.org/10.1111/1462-2920.13171.

[77] D. Kim, J. Choi, S. Lee, H. Hyun, K. Lee, K. Cho, Mutants defective in the production of encapsulin show a tan-phase-locked phenotype in *Myxococcus xanthus*, J. Microbiol. 57 (2019) 795–802. https://doi.org/10.1007/s12275-019-8683-9.

[78] 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, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[79] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, 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, S. Velankar, AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. 50 (2022) D439–D444. https://doi.org/10.1093/nar/gkab1061.

[80] J.P.L. Guerra, J.P. Jacinto, P. Tavares, Miniferritins: Small multifunctional protein cages, Coord. Chem. Rev. 449 (2021) 214187. https://doi.org/10.1016/j.ccr.2021.214187.

[81] R.R. Crichton, J.-P. Declercq, X-ray structures of ferritins and related proteins, Biochim. Biophys. Acta - Gen. Subj. 1800 (2010) 706–718. https://doi.org/10.1016/j.bbagen.2010.03.019.

[82] Y. Zhang, B.P. Orner, Self-Assembly in the Ferritin Nano-Cage Protein Superfamily, Int. J. Mol. Sci. 12 (2011) 5406–5421. https://doi.org/10.3390/ijms12085406.

[83] L.F. Dickey, S. Sreedharan, E.C. Theil, J.R. Didsbury, Y.H. Wang, R.E. Kaufman, Differences in the regulation of messenger RNA for housekeeping and specialized-cell ferritin. A comparison of three distinct ferritin complementary DNAs, the corresponding subunits, and identification of the first processed in amphibia., J. Biol. Chem. 262 (1987) 7901–7907. https://doi.org/10.1016/S0021-9258(18)47653-3.

[84] A. Lewin, G.R. Moore, N.E. Le Brun, Formation of protein-coated iron minerals, Dalt. Trans. (2005) 3597. https://doi.org/10.1039/b506071k.

[85] A. Friedman, P. Arosio, D. Finazzi, D. Koziorowski, J. Galazka-Friedman, Ferritin as an important player in neurodegeneration, Parkinsonism Relat. Disord. 17 (2011) 423–430. https://doi.org/10.1016/j.parkreldis.2011.03.016.

[86] F. Bou-Abdallah, The iron redox and hydrolysis chemistry of the ferritins, Biochim. Biophys. Acta - Gen. Subj. 1800 (2010) 719–731. https://doi.org/10.1016/j.bbagen.2010.03.021.

[87] S.K. Weeratunga, C.E. Gee, S. Lovell, Y. Zeng, C.L. Woodin, M. Rivera, Binding of *Pseudomonas aeruginosa* apobacterioferritin-associated ferredoxin to bacterioferritin B promotes heme mediation of electron delivery and mobilization of core mineral iron, Biochemistry. 48 (2009) 7420–7431. https://doi.org/10.1021/bi900561a.

[88] H. Yao, Y. Wang, S. Lovell, R. Kumar, A.M. Ruvinsky, K.P. Battaile, I.A. Vakser, M. Rivera, The structure of the BfrB–Bfd complex reveals protein–protein interactions enabling iron release from bacterioferritin, J. Am. Chem. Soc. 134 (2012) 13470–13481. https://doi.org/10.1021/ja305180n.

[89] M. Almiron, A.J. Link, D. Furlong, R. Kolter, A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*., Genes Dev. 6 (1992) 2646–2654. https://doi.org/10.1101/gad.6.12b.2646.

[90] Y. Yamamoto, L.B. Poole, R.R. Hantgan, Y. Kamio, An iron-binding protein, Dpr, from Streptococcus mutans prevents iron-dependent hydroxyl radical formation *in vitro*, J. Bacteriol. 184 (2002) 2931–2939. https://doi.org/10.1128/JB.184.11.2931-2939.2002.

[91] G. Zanotti, E. Papinutto, W.G. Dundon, R. Battistutta, M. Seveso, G. Del Giudice, R. Rappuoli, C. Montecucco, Structure of the neutrophil-activating protein from *Helicobacter pylori*, J. Mol. Biol. 323 (2002) 125–130. https://doi.org/10.1016/S0022-2836(02)00879-3.

[92] L. Chen, J.D. Helmann, Bacillus subtilis MrgA is a Dps (PexB) homologue: evidence for metalloregulation of an oxidative-stress gene, Mol. Microbiol. 18 (1995) 295–300. https://doi.org/10.1111/j.1365-2958.1995.mmi_18020295.x.

[93] R.A. Grant, D.J. Filman, S.E. Finkel, R. Kolter, J.M. Hogle, The crystal structure of Dps, a ferritin homolog that binds and protects DNA, Nat. Struct. Biol. 5 (1998) 294–303. https://doi.org/10.1038/nsb0498-294.

[94] K.J. Hintze, E.C. Theil, Cellular regulation and molecular interactions of the ferritins, Cell. Mol. Life Sci. 63 (2006) 591–600. https://doi.org/10.1007/s00018-005-5285-y.

[95] P. Arosio, R. Ingrassia, P. Cavadini, Ferritins: A family of molecules for iron storage, antioxidation and more, Biochim. Biophys. Acta - Gen. Subj. 1790 (2009) 589–599. https://doi.org/10.1016/j.bbagen.2008.09.004.

[96] E.C. Theil, R.K. Behera, T. Tosha, Ferritins for chemistry and for life, Coord. Chem. Rev. 257 (2013) 579–586. https://doi.org/10.1016/j.ccr.2012.05.013.

[97] A. Ilari, S. Stefanini, E. Chiancone, D. Tsernoglou, The dodecameric ferritin from *Listeria innocua* contains a novel intersubunit iron-binding site, Nat. Struct. Biol. 7 (2000) 38–43. https://doi.org/10.1038/71236.

[98] B. Ren, G. Tibbelin, T. Kajino, O. Asami, R. Ladenstein, The multi-layered structure of Dps with a novel di-nuclear ferroxidase center, J. Mol. Biol. 329 (2003) 467–477. https://doi.org/10.1016/S0022-2836(03)00466-2.

[99] G. Bellapadrona, S. Stefanini, C. Zamparelli, E.C. Theil, E. Chiancone, Iron translocation into and out of *Listeria innocua* Dps and size distribution of the protein-enclosed nanomineral are modulated by the electrostatic gradient at the 3-fold "Ferritin-like" pores, J. Biol. Chem. 284 (2009) 19101–19109. https://doi.org/10.1074/jbc.M109.014670.

[100] S. Franceschini, P. Ceci, F. Alaleona, E. Chiancone, A. Ilari, Antioxidant Dps protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus*, FEBS J. 273 (2006) 4913–4928. https://doi.org/110.1111/j.1742-4658.2006.05490.x.

[101] J.P.L. Guerra, C.E. Blanchet, B.J.C. Vieira, A. V Almeida, J.C. Waerenborgh, N.C. Jones, S. V Hoffmann, P. Tavares, A.S. Pereira, The conformation of the N-terminal tails of *Deinococcus grandis* Dps is modulated by the ionic strength, Int. J. Mol. Sci. 23 (2022) 4871. https://doi.org/10.3390/ijms23094871.

[102] S.P. Santos, M.G. Cuypers, A. Round, S. Finet, T. Narayanan, E.P. Mitchell, C. V. Romão, SAXS structural studies of Dps from *Deinococcus radiodurans* highlights the conformation of the mobile N-terminal extensions, J. Mol. Biol. 429 (2017) 667–687. https://doi.org/10.1016/j.jmb.2017.01.008.

[103] L.F. Huergo, H. Rahman, A. Ibrahimovic, C.J. Day, V. Korolik, *Campylobacter jejuni* Dps protein binds DNA in the presence of iron or hydrogen peroxide, J. Bacteriol. 195 (2013) 1970–1978. https://doi.org/10.1128/JB.00059-13.

[104] S.Y. Lee, C.J. Lim, P. Dröge, J. Yan, Regulation of bacterial DNA packaging in early stationary phase by competitive DNA binding of Dps and IHF, Sci. Rep. 5 (2016) 18146. https://doi.org/10.1038/srep18146.

[105] V. V Melekhov, U.S. Shvyreva, A.A. Timchenko, M.N. Tutukina, E. V Preobrazhenskaya, D. V Burkova, V.G. Artiukhov, O.N. Ozoline, S.S. Antipov, Modes of *Escherichia coli* Dps interaction with DNA as revealed by atomic force microscopy, PLoS One. 10 (2015) e0126504. https://doi.org/10.1371/journal.pone.0126504.

[106] S.S. Antipov, M.N. Tutukina, E. V Preobrazhenskaya, F.A. Kondrashov, M. V Patrushev, S. V Toshchakov, I. Dominova, U.S. Shvyreva, V. V Vrublevskaya, O.S. Morenkov, N.A. Sukharicheva, V. V Panyukov, O.N. Ozoline, The nucleoid protein Dps binds genomic DNA of *Escherichia coli* in a non-random manner, PLoS One. 12 (2017) e0182800. https://doi.org/10.1371/journal.pone.0182800.

[107] P. Ceci, DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus, Nucleic Acids Res. 32 (2004) 5935–5944. https://doi.org/10.1093/nar/gkh915.

[108] E. V. Dubrovin, L.A. Dadinova, M. V. Petoukhov, E.Y. Soshinskaya, A.A. Mozhaev, D. V. Klinov, T.E. Schäffer, E. V. Shtykova, O. V. Batishchev, Spatial organization of Dps and DNA–Dps complexes, J. Mol. Biol. 433 (2021). https://doi.org/10.1016/j.jmb.2021.166930.

[109] T.J. Stillman, M. Upadhyay, V.A. Norte, S.E. Sedelnikova, M. Carradus, S. Tzokov, P.A. Bullough, C.A. Shearman, M.J. Gasson, C.H. Williams, P.J. Artymiuk, J. Green, The crystal structures of *Lactococcus lactis* MG1363 Dps proteins reveal the presence of an N-terminal helix that is required for DNA binding, Mol. Microbiol. 57 (2005) 1101–1112. https://doi.org/10.1111/j.1365-2958.2005.04757.x.

[110] J.P. Jacinto, D. Penas, J.P.L. Guerra, A. V. Almeida, N.C. Jones, S. V. Hoffmann, P. Tavares, A.S. Pereira, Dps–DNA interaction in *Marinobacter hydrocarbonoclasticus* protein: effect of a single-charge alteration, Eur. Biophys. J. 50 (2021) 513–521. https://doi.org/10.1007/s00249-021-01538-0.

[111] E. Chiancone, Role of Dps (DNA-binding proteins from starved cells) aggregation on DNA, Front. Biosci. 15 (2010) 122. https://doi.org/10.2741/3610.

[112] P. Ceci, L. Mangiarotti, C. Rivetti, E. Chiancone, The neutrophil-activating Dps protein of *Helicobacter pylori*, HP-NAP, adopts a mechanism different from *Escherichia coli* Dps to bind and condense DNA, Nucleic Acids Res. 35 (2007) 2247–2256. https://doi.org/10.1093/nar/gkm077.

[113] M. Bozzi, G. Mignogna, S. Stefanini, D. Barra, C. Longhi, P. Valenti, E. Chiancone, A novel non-heme Iron-binding ferritin related to the DNA-binding proteins of the Dps family in *Listeria innocua*, J. Biol. Chem. 272 (1997) 3259–3265. https://doi.org/10.1074/jbc.272.6.3259.

[114] E. Papinutto, W.G. Dundon, N. Pitulis, R. Battistutta, C. Montecucco, G. Zanotti, Structure of two iron-binding proteins from *Bacillus anthracis*, J. Biol. Chem. 277 (2002) 15093–15098. https://doi.org/10.1074/jbc.M112378200.

[115] P. Ceci, A. Ilari, E. Falvo, E. Chiancone, The Dps protein of *Agrobacterium tumefaciens* does not bind to DNA but protects it toward oxidative cleavage, J. Biol. Chem. 278 (2003) 20319–20326. https://doi.org/10.1074/jbc.M302114200.

[116] T. Haikarainen, A.C. Papageorgiou, Dps-like proteins: structural and functional insights into a versatile protein family, Cell. Mol. Life Sci. 67 (2010) 341–351. https://doi.org/10.1007/s00018-009-0168-2.

[117] E. Chiancone, P. Ceci, The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding, Biochim. Biophys. Acta - Gen. Subj. 1800 (2010) 798–805. https://doi.org/10.1016/j.bbagen.2010.01.013.

[118] M. Su, S. Cavallo, S. Stefanini, E. Chiancone, N.D. Chasteen, The so-called *Listeria innocua* ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties, Biochemistry. 44 (2005) 5572–5578. https://doi.org/10.1021/bi0472705.

[119] A. Ilari, G. Bellapadrona, D. Carbonera, M. Di Valentin, Disclosing the molecular mechanism of iron incorporation in *Listeria innocua* Dps by EPR spectroscopy, Appl. Magn. Reson. 51 (2020) 1543–1557. https://doi.org/10.1007/s00723-020-01287-x.

[120] C. Howe, V.K. Moparthi, F.M. Ho, K. Persson, K. Stensjö, The Dps4 from Nostoc punctiforme ATCC 29133 is a member of His-type FOC containing Dps protein class that can be broadly found among cyanobacteria, PLoS One. 14 (2019) e0218300. https://doi.org/10.1371/journal.pone.0218300.

[121] T. Minato, T. Teramoto, Y. Kakuta, S. Ogo, K. Yoon, Biochemical and structural characterization of a thermostable Dps protein with His-type ferroxidase centers and outer metal-binding sites, FEBS Open Bio. 10 (2020) 1219–1229. https://doi.org/10.1002/2211-5463.12837.

[122] F. Alaleona, S. Franceschini, P. Ceci, A. Ilari, E. Chiancone, *Thermosynechoccus elongatus* DpsA binds Zn(II) at a unique three histidine-containing ferroxidase center and utilizes O_2 as iron oxidant with very high efficiency, unlike the typical Dps proteins, FEBS J. 277 (2010) 903–917. https://doi.org/10.1111/j.1742-4658.2009.07532.x.

[123] A. Ilari, P. Ceci, D. Ferrari, G.L. Rossi, E. Chiancone, Iron incorporation into *Escherichia coli* Dps gives rise to a Ferritin-like microcrystalline core, J. Biol. Chem. 277 (2002) 37619–37623. https://doi.org/10.1074/jbc.M206186200.

[124] L. Dadinova, R. Kamyshinsky, Y. Chesnokov, A. Mozhaev, V. Matveev, A. Gruzinov, A. Vasiliev, E. Shtykova, Structural rearrangement of Dps-DNA complex caused by divalent Mg and Fe cations, Int. J. Mol. Sci. 22 (2021) 6056. https://doi.org/10.3390/ijms22116056.

[125] D. Penas, A.S. Pereira, P. Tavares, Direct evidence for ferrous ion oxidation and incorporation in the absence of oxidants by Dps from *Marinobacter hydrocarbonoclasticus*, Angew. Chemie Int. Ed. 58 (2019) 1013–1018. https://doi.org/10.1002/anie.201809584.

[126] B.B. Muhoberac, R. Vidal, Iron, ferritin, hereditary ferritinopathy, and neurodegeneration, Front. Neurosci. 13 (2019). https://doi.org/10.3389/fnins.2019.01195.

[127] F. Funk, J.-P. Lenders, R.R. Crichton, W. Schneider, Reductive mobilisation of ferritin iron, Eur. J. Biochem. 152 (1985) 167–172. https://doi.org/10.1111/j.1432-1033.1985.tb09177.x.

[128] S. Sirivech, E. Frieden, S. Osaki, The release of iron from horse spleen ferritin by reduced flavins, Biochem. J. 143 (1974) 311–315. https://doi.org/10.1042/bj1430311.

[129] G.D. Watt, D. Jacobs, R.B. Frankel, Redox reactivity of bacterial and mammalian ferritin: is reductant entry into the ferritin interior a necessary step for iron release?, Proc. Natl. Acad. Sci. 85 (1988) 7457–7461. https://doi.org/10.1073/pnas.85.20.7457.

[130] M. Castruita, L.A. Elmegreen, Y. Shaked, E.I. Stiefel, F.M.M. Morel, Comparison of the kinetics of iron release from a marine (*Trichodesmium erythraeum*) Dps protein and mammalian ferritin in the presence and absence of ligands, J. Inorg. Biochem. 101 (2007) 1686–1691. https://doi.org/10.1016/j.jinorgbio.2007.07.022.

[131] N.A.R. Almeida, Estudo da Libertação de Ferro em Miniferritinas, 2014 (online

publication).

[132] S.M. Williams, A. V. Chandran, M.S. Vijayabaskar, S. Roy, H. Balaram, S. Vishveshwara, M. Vijayan, D. Chatterji, A histidine aspartate ionic lock gates the iron passage in miniferritins from *Mycobacterium smegmatis*, J. Biol. Chem. 289 (2014) 11042–11058. https://doi.org/10.1074/jbc.M113.524421.

[133] S.-G. Kim, G. Bhattacharyya, A. Grove, Y.-H. Lee, Crystal structure of Dps-1, a functionally distinct Dps protein from *Deinococcus radiodurans*, J. Mol. Biol. 361 (2006) 105–114. https://doi.org/10.1016/j.jmb.2006.06.010.

[134] C. V. Romão, E.P. Mitchell, S. McSweeney, The crystal structure of *Deinococcus radiodurans* Dps protein (DR2263) reveals the presence of a novel metal centre in the N terminus, JBIC J. Biol. Inorg. Chem. 11 (2006) 891–902. https://doi.org/10.1007/s00775-006-0142-5.

[135] H.Q. Huang, Q.M. Lin, B. Kong, R.Y. Zeng, Y.H. Qiao, C.H. Chen, F.Z. Zhang, L.S. Xu, Role of phosphate and kinetic characteristics of complete iron release from native pig spleen ferritin-Fe, J. Protein Chem. 18 (1999) 497–504. https://doi.org/10.1023/A:1020653028685.

[136] R.G. Mothersole, M. Macdonald, M. Kolesnikov, M.E.P. Murphy, K.R. Wolthers, Structural insight into the high reduction potentials observed for *Fusobacterium nucleatum* flavodoxin, Protein Sci. 28 (2019) 1460–1472. https://doi.org/10.1002/pro.3661.

[137] M.W. Fraaije, A. Mattevi, Flavoenzymes: diverse catalysts with recurrent features, Trends Biochem. Sci. 25 (2000) 126–132. https://doi.org/10.1016/S0968-0004(99)01533-9.

[138] Z.-Q. Wang, R.J. Lawson, M.R. Buddha, C.-C. Wei, B.R. Crane, A.W. Munro, D.J. Stuehr, Bacterial flavodoxins support nitric oxide production by *Bacillus subtilis* nitric-oxide synthase, J. Biol. Chem. 282 (2007) 2196–2202. https://doi.org/10.1074/jbc.M608206200.

[139] O.M. Birch, M. Fuhrmann, N.M. Shaw, Biotin synthase from *Escherichia coli*, an investigation of the low molecular weight and protein components required for activity *in vitro*, J. Biol. Chem. 270 (1995) 19158–19165. https://doi.org/10.1074/jbc.270.32.19158.

[140] O. Ifuku, N. Koga, S. Haze, J. Kishimoto, Y. Wachi, Flavodoxin is required for conversion of dethiobiotin to biotin in *Escherichia coli*, Eur. J. Biochem. 224 (1994) 173–178. https://doi.org/10.1111/j.1432-1033.1994.tb20009.x.

[141] H.P. Blaschkowski, J. Knappe, M. Ludwig-Festl, G. Neuer, Routes of Flavodoxin and Ferredoxin Reduction in *Escherichia coli*, Eur. J. Biochem. 123 (2005) 563–569. https://doi.org/10.1111/j.1432-1033.1982.tb06569.x.

[142] K. Fujii, F.M. Huennekens, Activation of methionine synthetase by a reduced triphosphopyridine nucleotide-dependent flavoprotein system, J. Biol. Chem. 249 (1974) 6745–6753. https://doi.org/10.1016/S0021-9258(19)42122-4.

[143] W. Yang, L. Ni, R.L. Somerville, A stationary-phase protein of *Escherichia coli* that affects

the mode of association between the trp repressor protein and operator-bearing DNA., Proc. Natl. Acad. Sci. 90 (1993) 5796–5800. https://doi.org/10.1073/pnas.90.12.5796.

[144] R. Grandori, P. Khalifah, J.A. Boice, R. Fairman, K. Giovanielli, J. Carey, Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins, J. Biol. Chem. 273 (1998) 20960–20966. https://doi.org/10.1074/jbc.273.33.20960.

[145] R. Grandori, J. Carey, Six new candidate members of the α/β twisted open-sheet family detected by sequence similarity to flavodoxin, Protein Sci. 3 (1994) 2185–2193. https://doi.org/10.1002/pro.5560031204.

[146] S.G. Mayhew, M.L. Ludwig, 2 flavodoxins and electron-transferring flavoproteins, in: Enzymes, 1975: pp. 57–118. https://doi.org/10.1016/S1874-6047(08)60225-5.

[147] E. V. Patridge, J.G. Ferry, WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:Quinone oxidoreductase, J. Bacteriol. 188 (2006) 3498–3506. https://doi.org/10.1128/JB.188.10.3498-3506.2006.

[148] S.L.A. Andrade, E. V Patridge, J.G. Ferry, O. Einsle, Crystal structure of the NADH:Quinone oxidoreductase WrbA from *Escherichia coli*, J. Bacteriol. 189 (2007) 9101–9107. https://doi.org/10.1128/JB.01336-07.

[149] I. Kishko, J. Carey, D. Reha, J. Brynda, R. Winkler, B. Harish, R. Guerra, O. Ettrichova, Z. Kukacka, O. Sheryemyetyeva, P. Novak, M. Kuty, I. Kuta Smatanova, R. Ettrich, M. Lapkouski, 1.2 Å resolution crystal structure of *Escherichia coli* WrbA holoprotein, Acta Crystallogr. Sect. D Biol. Crystallogr. 69 (2013) 1748–1757. https://doi.org/10.1107/S0907444913017162.

[150] J. Wolfova, I.K. Smatanova, J. Brynda, J.R. Mesters, M. Lapkouski, M. Kuty, A. Natalello, N. Chatterjee, S.-Y. Chern, E. Ebbel, A. Ricci, R. Grandori, R. Ettrich, J. Carey, Structural organization of WrbA in apo- and holoprotein crystals, Biochim. Biophys. Acta - Proteins Proteomics. 1794 (2009) 1288–1298. https://doi.org/10.1016/j.bbapap.2009.08.001.

[151] J. Freigang, K. Diederichs, K.P. Schäfer, W. Welte, R. Paul, Crystal structure of oxidized flavodoxin, an essential protein in *Helicobacter pylori*, Protein Sci. 11 (2009) 253–261. https://doi.org/10.1110/ps.28602.

[152] Y.-C. Hsieh, T. Chia, H.-K. Fun, C.-J. Chen, Crystal structure of dimeric flavodoxin from *Desulfovibrio gigas* suggests a potential binding region for the electron-transferring partner, Int. J. Mol. Sci. 14 (2013) 1667–1683. https://doi.org/10.3390/ijms14011667.

[153] E. Ullmann, T.C. Tan, T. Gundinger, C. Herwig, C. Divne, O. Spadiut, A novel cytosolic NADH:quinone oxidoreductase from *Methanothermobacter marburgensis*, Biosci. Rep. 34 (2014) 893–904. https://doi.org/10.1042/BSR20140143.

[154] G. Nöll, E. Kozma, R. Grandori, J. Carey, T. Schödl, G. Hauska, J. Daub, Spectroelectrochemical investigation of a flavoprotein with a flavin-modified gold electrode, Langmuir. 22 (2006) 2378–2383. https://doi.org/10.1021/la051423n.

[155] H. Beinert, R.H. Holm, E. Münck, Iron-sulfur clusters: nature's modular, multipurpose structures, Science. 277 (1997) 653–659. https://doi.org/10.1126/science.277.5326.653.

[156] D.C. Johnson, D.R. Dean, A.D. Smith, M.K. Johnson, Structure, function, and formation of biological iron-sulfur clusters, Annu. Rev. Biochem. 74 (2005) 247–281. https://doi.org/10.1146/annurev.biochem.74.082803.133518.

[157] Y. Li, P. pan Liu, X. Ni, Molecular evolution and functional analysis of rubredoxin-like proteins in plants, Biomed Res. Int. 2019 (2019) 1–14. https://doi.org/10.1155/2019/2932585.

[158] P. Tavares, A.S. Pereira, C. Krebs, N. Ravi, J.J.G. Moura, I. Moura, B.H. Huynh, Spectroscopic characterization of a novel tetranuclear Fe cluster in an iron−sulfur protein isolated from *Desulfovibrio desulfuricans*, Biochemistry. 37 (1998) 2830–2842. https://doi.org/10.1021/bi9723008.

[159] S.J. Cooper, C.D. Garner, W.R. Hagen, P.F. Lindley, S. Bailey, Hybrid-cluster protein (HCP) from *Desulfovibrio vulgaris* (Hildenborough) at 1.6 Å resolution, Biochemistry. 39 (2000) 15044–15054. https://doi.org/10.1021/bi001483m.

[160] D.C. Rees, F. Akif Tezcan, C.A. Haynes, M.Y. Walton, S. Andrade, O. Einsle, J.B. Howard, Structural basis of biological nitrogen fixation, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 363 (2005) 971–984. https://doi.org/10.1098/rsta.2004.1539.

[161] W. Lubitz, H. Ogata, O. Rüdiger, E. Reijerse, Hydrogenases, Chem. Rev. 114 (2014) 4081–4148. https://doi.org/10.1021/cr4005814.

[162] M. Merrouch, M. Benvenuti, M. Lorenzi, C. Léger, V. Fourmond, S. Dementin, Maturation of the [Ni–4Fe–4S] active site of carbon monoxide dehydrogenases, JBIC J. Biol. Inorg. Chem. 23 (2018) 613–620. https://doi.org/10.1007/s00775-018-1541-0.

[163] M.K. Eidsness, S.E. O'Dell, D.M. Kurtz, R.L. Robson, R.A. Scott, Expression of a synthetic gene coding for the amino acid sequence of *Clostridium pasteurianum* rubredoxin, "Protein Eng. Des. Sel. 5 (1992) 367–371. https://doi.org/10.1093/protein/5.4.367.

[164] Y. Petillot, E. Forest, I. Mathieu, J. Meyer, J.M. Moulis, Analysis, by electrospray ionization mass spectrometry, of several forms of *Clostridium pasteurianum* rubredoxin, Biochem. J. 296 (1993) 657–661. https://doi.org/10.1042/bj2960657.

[165] F. Folgosa, C.M. Cordas, J.A. Santos, A.S. Pereira, J.J.G. Moura, P. Tavares, I. Moura, New spectroscopic and electrochemical insights on a class I superoxide reductase: evidence for an intramolecular electron-transfer pathway, Biochem. J. 438 (2011) 485–494. https://doi.org/10.1042/BJ20110836.

[166] E.T. Lode, M.J. Coon, Enzymatic ω-Oxidation, J. Biol. Chem. 246 (1971) 791–802. https://doi.org/10.1016/S0021-9258(18)62479-2.

[167] A. Thapper, A.C. Rizzi, C.D. Brondino, A.G. Wedd, R.J. Pais, B.K. Maiti, I. Moura, S.R. Pauleta, J.J.G. Moura, Copper-substituted forms of the wild type and C42A variant of rubredoxin, J. Inorg. Biochem. 127 (2013) 232–237. https://doi.org/10.1016/j.jinorgbio.2013.06.003.

[168] S.W. May, J. Kuo, Preparation and properties of cobalt (II) rubredoxin, Biochemistry. 17 (1978) 3333–3338. https://doi.org/10.1021/bi00609a025.

[169] A.T. Kowal, I.C. Zambrano, I. Moura, J.J.G. Moura, J. LeGall, M.K. Johnson, Electronic and magnetic properties of nickel-substituted rubredoxin: a variable-temperature magnetic circular dichroism study, Inorg. Chem. 27 (1988) 1162–1166. https://doi.org/10.1021/ic00280a015.

[170] I. Moura, M. Teixeira, J.J.G. Moura, J. LeGall, Spectroscopic studies of cobalt and nickel substituted rubredoxin and desulforedoxin, J. Inorg. Biochem. 44 (1991) 127–139. https://doi.org/10.1016/0162-0134(91)84025-5.

[171] P.R. Blake, J.B. Park, M.W.W. Adams, M.F. Summers, Novel observation of NH-S(Cys) hydrogen-bond-mediated scalar coupling in cadmium-113 substituted rubredoxin from *Pyrococcus furiosus*, J. Am. Chem. Soc. 114 (1992) 4931–4933. https://doi.org/10.1021/ja00038a084.

[172] C.J. Henehan, D.L. Pountney, M. Vašák, O. Zerbe, Identification of cysteine ligands in metalloproteins using optical and NMR spectroscopy: Cadmium-substituted rubredoxin as a model [Cd(CysS) 4] 2- center, Protein Sci. 2 (1993) 1756–1764. https://doi.org/10.1002/pro.5560021019.

[173] Z. Xiao, M.J. Lavery, M. Ayhan, S.D.B. Scrofani, M.C.J. Wilce, J.M. Guss, P.A. Tregloan, G.N. George, A.G. Wedd, The rubredoxin from *Clostridium pasteurianum*: Mutation of the iron cysteinyl ligands to serine. Crystal and molecular structures of oxidized and dithionite-treated forms of the Cys42Ser mutant, J. Am. Chem. Soc. 120 (1998) 4135–4150. https://doi.org/10.1021/ja973162c.

[174] M. Archer, A.L. Carvalho, S. Teixeira, I. Moura, J.J.G. Moura, F. Rusnak, M.J. Romão, Structural studies by X-ray diffraction on metal substituted desulforedoxin, a rubredoxin-type protein, Protein Sci. 8 (1999) 1536–1545. https://doi.org/10.1110/ps.8.7.1536.

[175] B.D. Fairbanks, M.P. Schwartz, C.N. Bowman, K.S. Anseth, Photoinitiated polymerization of PEG-diacrylate with lithium phenyl-2,4,6-trimethylbenzoylphosphinate: polymerization rate and cytocompatibility, Biomaterials. 30 (2009) 6702–6707. https://doi.org/10.1016/j.biomaterials.2009.08.055.

[176] K. Schweimer, S. Hoffmann, P. Rösch, H. Sticht, J. Wastl, U.G. Maier, Solution structure of a zinc substituted eukaryotic rubredoxin from the cryptomonad alga *Guillardia theta*, Protein Sci. 9 (2000) 1474–1486. https://doi.org/10.1110/ps.9.8.1474.

[177] K.M. Handley, M. Hery, J.R. Lloyd, *Marinobacter santoriniensis* sp. nov., an arsenate-respiring and arsenite-oxidizing bacterium isolated from hydrothermal sediment, Int. J. Syst. Evol. Microbiol. 59 (2009) 886–892. https://doi.org/10.1099/ijs.0.003145-0.

[178] K.M. Handley, M. Héry, J.R. Lloyd, Redox cycling of arsenic by the hydrothermal marine bacterium *Marinobacter santoriniensis*, Environ. Microbiol. 11 (2009) 1601–1611. https://doi.org/10.1111/j.1462-2920.2009.01890.x.

[179] K.M. Handley, J.R. Lloyd, Biogeochemical implications of the ubiquitous colonization of marine habitats and redox gradients by Marinobacter species, Front. Microbiol. 4 (2013) 1–10. https://doi.org/10.3389/fmicb.2013.00136.

[180] M.J. Gauthier, B. Lafay, R. Christen, L. Fernandez, M. Acquaviva, P. Bonin, J.-C. Bertrand, *Marinobacter hydrocarbonoclasticus* gen. nov., sp. nov., a New, Extremely

Halotolerant, Hydrocarbon-Degrading Marine Bacterium, Int. J. Syst. Bacteriol. 42 (1992) 568–576. https://doi.org/10.1099/00207713-42-4-568.

[181] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: J.M. Walker (Ed.), Proteomics Protoc. Handb., Humana Press, Totowa, NJ, 2005: pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571.

[182] N.J. Anthis, G.M. Clore, Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, Protein Sci. 22 (2013) 851–858. https://doi.org/10.1002/pro.2253.

[183] 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, A. Cardona, Fiji: an open-source platform for biological-image analysis, Nat. Methods. 9 (2012) 676–682. https://doi.org/10.1038/nmeth.2019.

[184] C.E. Blanchet, A. Spilotros, F. Schwemmer, M.A. Graewert, A. Kikhney, C.M. Jeffries, D. Franke, D. Mark, R. Zengerle, F. Cipriani, S. Fiedler, M. Roessle, D.I. Svergun, Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY), J. Appl. Crystallogr. 48 (2015) 431–443. https://doi.org/10.1107/S160057671500254X.

[185] M.A. Graewert, D. Franke, C.M. Jeffries, C.E. Blanchet, D. Ruskule, K. Kuhle, A. Flieger, B. Schäfer, B. Tartsch, R. Meijers, D.I. Svergun, Automated pipeline for purification, biophysical and X-ray analysis of biomacromolecular solutions, Sci. Rep. 5 (2015) 10734. https://doi.org/10.1038/srep10734.

[186] D. Franke, A.G. Kikhney, D.I. Svergun, Automated acquisition and analysis of small angle X-ray scattering data, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 689 (2012) 52–59. https://doi.org/10.1016/j.nima.2012.06.008.

[187] A. Panjkovich, D.I. Svergun, CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data, Bioinformatics. 34 (2018) 1944–1946. https://doi.org/10.1093/bioinformatics/btx846.

[188] D. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886. https://doi.org/10.1016/S0006-3495(99)77443-6.

[189] D.I. Svergun, M. V. Petoukhov, M.H.J. Koch, Determination of domain structure of proteins from X-ray solution scattering, Biophys. J. 80 (2001) 2946–2953. https://doi.org/10.1016/S0006-3495(01)76260-1.

[190] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P. V. Konarev, D.I. Svergun, New developments in the ATSAS program package for small-angle scattering data analysis, J. Appl. Crystallogr. 45 (2012) 342–350. https://doi.org/10.1107/S0021889812007662.

[191] V. V Volkov, D.I. Svergun, Uniqueness of *ab initio* shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864. https://doi.org/10.1107/S0021889803000268.

[192] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, T.E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers, Protein Sci. 30 (2021) 70–82. https://doi.org/10.1002/pro.3943.

[193] A. Besada, A facile and sensitive spectrophotometric determination of ascorbic acid, Talanta. 34 (1987) 731–732. https://doi.org/10.1016/0039-9140(87)80229-1.

[194] S.V. Hoffmann, M. Fano, M. van de Weert, Circular dichroism spectroscopy for structural characterization of proteins, in: A. Müllertz, Y. Perrie, T. Rades (Eds.), Anal. Tech. Pharm. Sci., Springer New York, New York, NY, 2016: pp. 223–251. https://doi.org/10.1007/978-1-4939-4029-5_6.

[195] A.J. Miles, S.G. Ramalli, B.A. Wallace, DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data, Protein Sci. 31 (2022) 37–46. https://doi.org/10.1002/pro.4153.

[196] D.I. Svergun, M.H.J. Koch, Small-angle scattering studies of biological macromolecules in solution, Reports Prog. Phys. 66 (2003) 1735–1782. https://doi.org/10.1088/0034-4885/66/10/R05.

[197] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503. https://doi.org/10.1107/S0021889892001663.

[198] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, D.I. Svergun, ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J. Appl. Crystallogr. 50 (2017) 1212–1225. https://doi.org/10.1107/S1600576717007786.

[199] D.P. Klose, B.A. Wallace, R.W. Janes, 2Struc: the secondary structure server, Bioinformatics. 26 (2010) 2624–2625. https://doi.org/10.1093/bioinformatics/btq480.

[200] I. Prisecaru, Mössbauer Spectral Analysis Software, wmoss.org, n.d.

[201] S. Levi, S.J. Yewdall, P.M. Harrison, P. Santambrogio, A. Cozzi, E. Rovida, A. Albertini, P. Arosio, Evidence of H- and L-chains have co-operative roles in the iron-uptake mechanism of human ferritin, Biochem. J. 288 (1992) 591–596. https://doi.org/10.1042/bj2880591.

[202] C. Pozzi, S. Ciambellotti, C. Bernacchioni, F. Di Pisa, S. Mangani, P. Turano, Chemistry at the protein–mineral interface in L-ferritin assists the assembly of a functional (μ 3 -oxo)Tris[(μ 2 -peroxo)] triiron(III) cluster, Proc. Natl. Acad. Sci. 114 (2017) 2580–2585. https://doi.org/10.1073/pnas.1614302114.

[203] S. Ciambellotti, C. Pozzi, S. Mangani, P. Turano, Iron biomineral growth from the initial nucleation seed in L-ferritin, Chem. – A Eur. J. 26 (2020) 5770–5773. https://doi.org/10.1002/chem.202000064.

[204] J.C. Waerenborgh, P. Tavares, A.S. Pereira, Mössbauer spectroscopy, in: A.S. Pereira, P. Tavares, P. Limão-Vieira (Eds.), Radiat. Bioanal. Spectrosc. Tech. Theor. Methods, Springer International Publishing, Cham, 2019: pp. 213–244. https://doi.org/10.1007/978-3-030-28247-9_8.

[205] N.D. Chasteen, P.M. Harrison, Mineralization in ferritin: An efficient means of iron storage, J. Struct. Biol. 126 (1999) 182–194. https://doi.org/10.1006/jsbi.1999.4118.

[206] G.C. Papaefthymiou, The Mössbauer and magnetic properties of ferritin cores, Biochim. Biophys. Acta - Gen. Subj. 1800 (2010) 886–897. https://doi.org/10.1016/j.bbagen.2010.03.018.

[207] C.G. Timóteo, M. Guilherme, D. Penas, F. Folgosa, P. Tavares, A.S. Pereira, Desulfovibrio vulgaris bacterioferritin uses H_2O_2 as a co-substrate for iron oxidation and reveals DPS-like DNA protection and binding activities, Biochem. J. 446 (2012) 125–133. https://doi.org/10.1042/BJ20111439.

[208] K.H. Ebrahimi, E. Bill, P.-L. Hagedoorn, W.R. Hagen, Spectroscopic evidence for the role of a site of the di-iron catalytic center of ferritins in tuning the kinetics of FeII oxidation, Mol. Biosyst. 12 (2016) 3576–3588. https://doi.org/10.1039/C6MB00235H.

[209] NZYTech, NZYMidiprep, Ed. Genes e Enzym. Lisbon. (2011) 2–3.

[210] K.H. Nguyen, L.T. Smith, L. Xiao, G. Bhattacharyya, A. Grove, On the stoichiometry of *Deinococcus radiodurans* Dps-1 binding to duplex DNA, Proteins Struct. Funct. Bioinforma. 80 (2012) 713–721. https://doi.org/10.1002/prot.23228.

[211] J. Hołówka, J. Zakrzewska-Czerwińska, Nucleoid associated proteins: The small organizers that help to cope with stress, Front. Microbiol. 11 (2020) 1–7. https://doi.org/10.3389/fmicb.2020.00590.

[212] S.M. Williams, D. Chatterji, An overview of Dps: Dual acting nanovehicles in prokaryotes with DNA binding and ferroxidation properties, in: Subcell. Biochem., United States, 2021: pp. 177–216. https://doi.org/10.1007/978-3-030-58971-4_3.

[213] D.R. Zusman, A.E. Scott, Z. Yang, J.R. Kirby, Chemosensory pathways, motility and development in *Myxococcus xanthus*, Nat. Rev. Microbiol. 5 (2007) 862–872. https://doi.org/10.1038/nrmicro1770.

[214] M.C.M. Guilherme, Estudos Mecanísticos e Estruturais da Oxidação e Armazenamento de Ferro por Ferritinas Rápidas, Univ. Nov. LISBOA Fac. Ciências e Tecnol. (2009) (online publication).

[215] F. Folgosa, Structural and Mechanistic Studies of Iron Containing Proteins, (2008) (online publication).

[216] E.K. Koepf, H.M. Petrassi, M. Sudol, J.W. Kelly, WW: An isolated three-stranded antiparallel β-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state, Protein Sci. 8 (2008) 841–853. https://doi.org/10.1110/ps.8.4.841.

[217] R. Varhač, D. Sedláková, M. Stupák, E. Sedlák, Non-two-state thermal denaturation of ferricytochrome c at neutral and slightly acidic pH values, Biophys. Chem. 203–204 (2015) 41–50. https://doi.org/10.1016/j.bpc.2015.05.002.

[218] I. Moura, B.H. Huynh, R.P. Hausinger, J. Le Gall, A.V. Xavier, E. Münck, Mössbauer and EPR studies of desulforedoxin from *Desulfovibrio gigas*., J. Biol. Chem. 255 (1980) 2493–2498. https://doi.org/10.1016/S0021-9258(19)85920-3.

[219] J. Meyer, J.-M. Moulis, Rubredoxin, in: Handb. Met., John Wiley & Sons, Ltd, Chichester, 2006: pp. 1–13. https://doi.org/10.1002/0470028637.met135.

[220] J.G. Lees, A.J. Miles, F. Wien, B.A. Wallace, A reference database for circular dichroism spectroscopy covering fold and secondary structure space, Bioinformatics. 22 (2006) 1955–1962. https://doi.org/10.1093/bioinformatics/btl327.

[221] L. Whitmore, B.A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases, Biopolymers. 89 (2008) 392–400. https://doi.org/10.1002/bip.20853.

[222] L. Whitmore, B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, Nucleic Acids Res. 32 (2004) W668–W673. https://doi.org/10.1093/nar/gkh371.

[223] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, Anal. Biochem. 287 (2000) 252–260. https://doi.org/10.1006/abio.2000.4880.

[224] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, Biopolymers. 22 (1983) 2577–2637. https://doi.org/10.1002/bip.360221211.

[225] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, Nucleic Acids Res. 41 (2012) D36–D42. https://doi.org/10.1093/nar/gks1195.

[226] P. Debrunner, C. Schulz, Mössbauer parameters of rubredoxin, a one-iron-sulfur protein, in: Mössbauer Eff. Methodol., Springer US, Boston, MA, 1976: pp. 155–167. https://doi.org/10.1007/978-1-4684-8073-3_7.

[227] V. V. Vrajmasu, E.L. Bominaar, J. Meyer, E. Münck, Mössbauer study of reduced rubredoxin as purified and in whole cells. Structural correlation analysis of spin hamiltonian parameters, Inorg. Chem. 41 (2002) 6358–6371. https://doi.org/10.1021/ic020508y.

[228] P. Zheng, Y. Wang, H. Li, Reversible unfolding-refolding of rubredoxin: A single-molecule force spectroscopy study, Angew. Chemie Int. Ed. 53 (2014) 14060–14063. https://doi.org/10.1002/anie.201408105.

[229] F. Bonomi, S. Iametti, S. Mazzini, D. Fessas, D.M. Kurtz, Thermal stability of *Clostridium pasteurianum* rubredoxin: Deconvoluting the contributions of the metal site and the protein, Protein Sci. 9 (2000) 2413–2426. https://doi.org/10.1110/ps.9.12.2413.

[230] S. Prakash, M. Sundd, P. Guptasarma, The key to the extraordinary thermal stability of P. furiosus holo-rubredoxin: Iron binding-guided packing of a core aromatic cluster responsible for high kinetic stability of the native structure, PLoS One. 9 (2014) e89703. https://doi.org/10.1371/journal.pone.0089703.

[231] S. Cavagnero, Z.H. Zhou, M.W.W. Adams, S.I. Chan, Unfolding mechanism of rubredoxin from *Pyrococcus furiosus*, Biochemistry. 37 (1998) 3377–3385. https://doi.org/10.1021/bi9721804.

[232] B.J. Henriques, L.M. Saraiva, C.M. Gomes, Probing the mechanism of rubredoxin thermal unfolding in the absence of salt bridges by temperature jump experiments, Biochem. Biophys. Res. Commun. 333 (2005) 839–844. https://doi.org/10.1016/j.bbrc.2005.06.004.

[233] P. Bonin, J.N. Barbotin, P. Dhulster, J.-C. Bertrand, Nitrate reduction in simulated microniches by a denitrifying marine bacterium, Can. J. Microbiol. 33 (1987) 276–279. https://doi.org/10.1139/m87-047.

[234] P. Bonin, M. Gilewicz, J.C. Bertrand, Denitrification by a marine bacterium *Pseudomonas nautica* strain 617, Ann. l'Institut Pasteur / Microbiol. 138 (1987) 371–383. https://doi.org/10.1016/0769-2609(87)90125-6.

[235] G.L. Smith, A.K. Srivastava, A.A. Reutovich, N.J. Hunter, P. Arosio, A. Melman, F. Bou-Abdallah, Iron mobilization from ferritin in yeast cell lysate and physiological implications, Int. J. Mol. Sci. 23 (2022) 6100. https://doi.org/10.3390/ijms23116100.

[236] D. Qureshi, A.K. Nayak, D. Kim, S. Maji, A. Anis, B. Mohanty, K. Pal, Polysaccharide-based polymeric gels as drug delivery vehicles, in: Adv. Challenges Pharm. Technol., Elsevier, 2021: pp. 283–325. https://doi.org/10.1016/B978-0-12-820043-8.00013-X.

[237] A. V. Almeida, J.P. Jacinto, J.P.L. Guerra, B.J.C. Vieira, J.C. Waerenborgh, N.C. Jones, S. V. Hoffmann, A.S. Pereira, P. Tavares, Structural features and stability of apo- and holo-forms of a simple iron–sulfur protein, Eur. Biophys. J. 50 (2021) 561–570. https://doi.org/10.1007/s00249-021-01546-0.

[238] R. Hider, M.V. Aviles, Y.-L. Chen, G.O. Latunde-Dada, The role of GSH in intracellular iron trafficking, Int. J. Mol. Sci. 22 (2021) 1278. https://doi.org/10.3390/ijms22031278.

A

APPENDIX

A.1 Production and Purification of the Encapsulin System

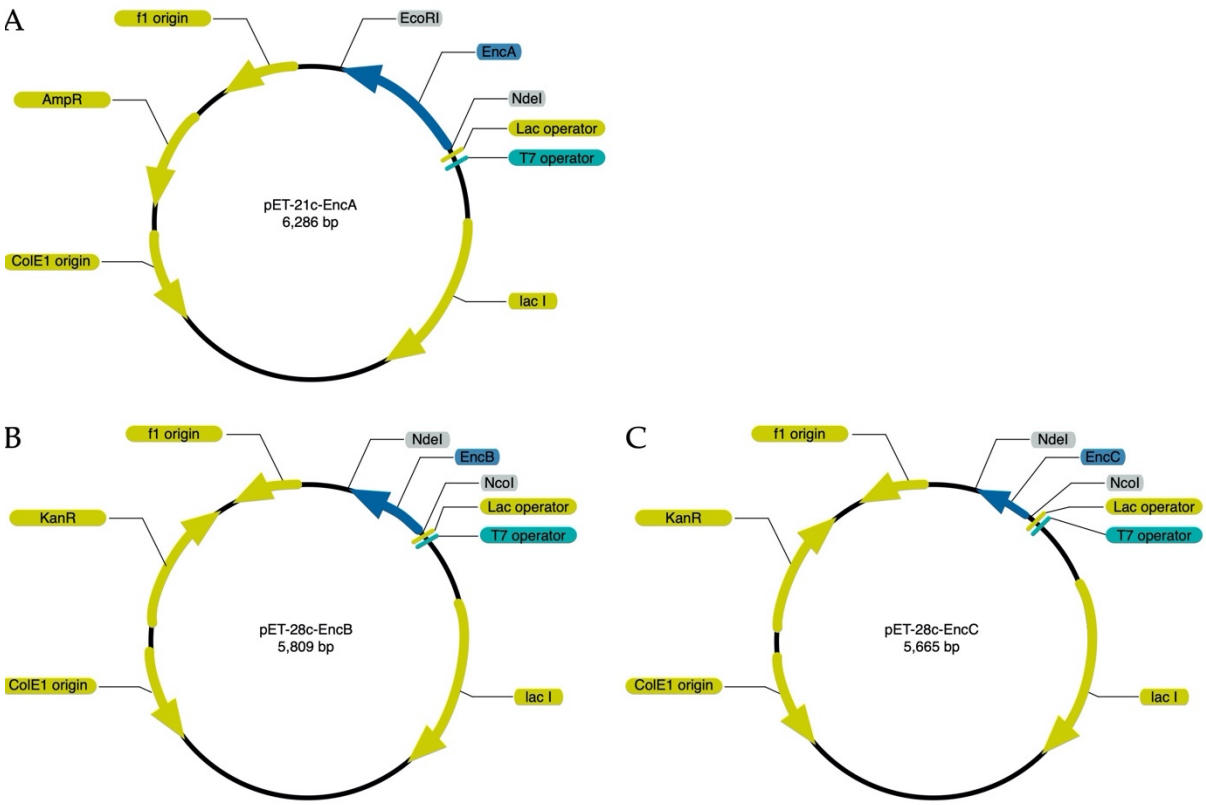

Figure A.1: Expression vector maps used to produce the proteins of the encapsulin system. (A) pET-21c-EncA, **(B)** pET-28c-EncB and **(C)** pET-28c-EncC.

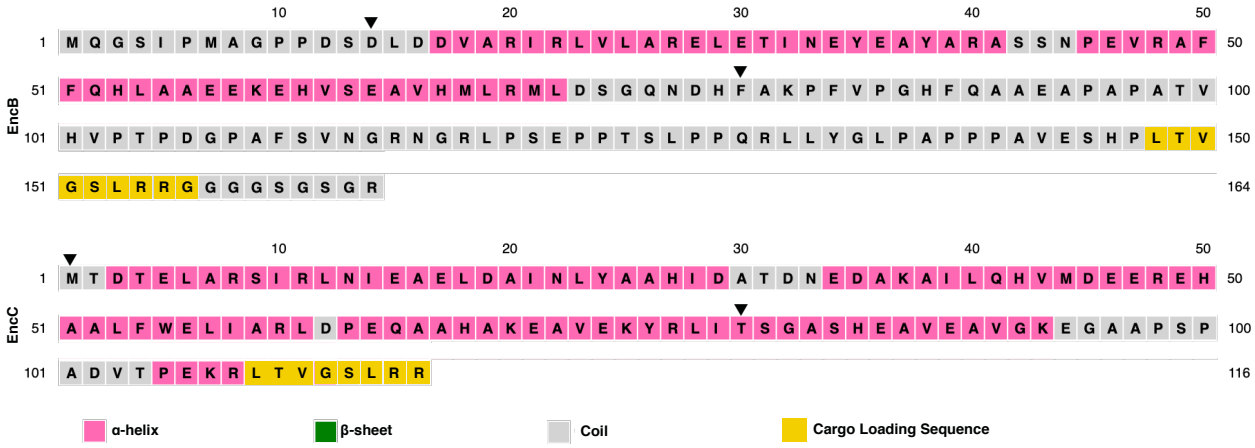

Figure A.2: Secondary structure prediction of the *Myx. xanthus* cargo proteins based on their primary sequences. The triangles represent the initial and final residues with atomic structure currently solved for EncB (PDB: 7S5C) and EncC (PDB:7S8T) [30].

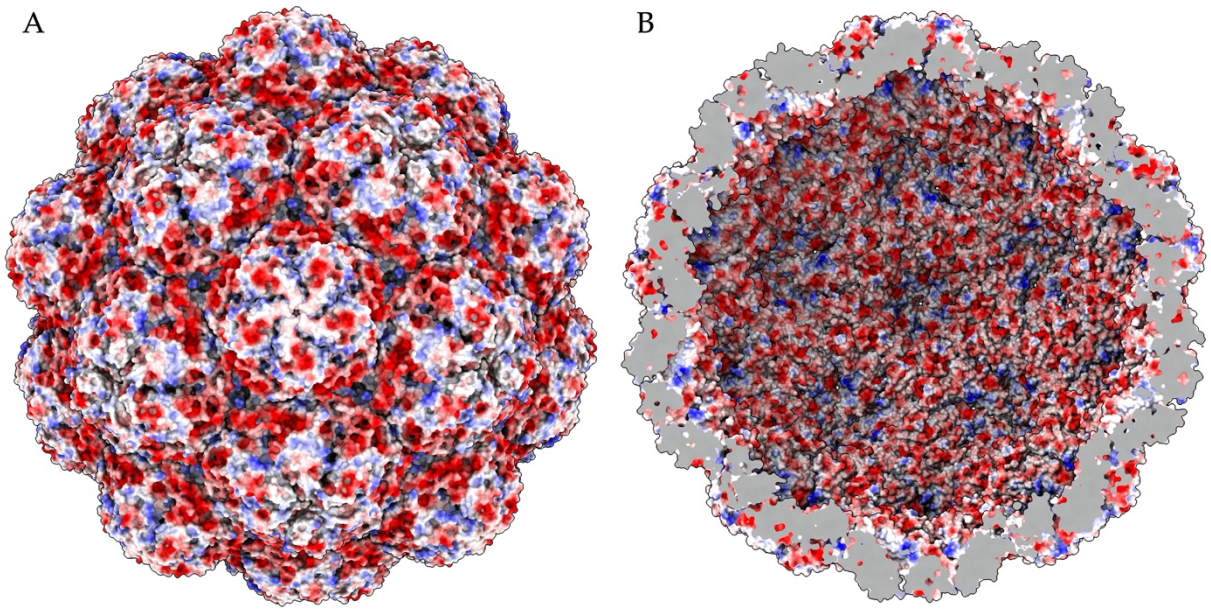

Figure A.3: Electrostatic surface potential of the *Myx. xanthus* encapsulin external and internal surfaces. (A) External surface and **(B)** internal surface. The electrostatic surface potentials are colored with a range from red (negative) to white (neutral) to blue (positive) corresponding to a +10 kTe^{-1} to 0 to −10 kTe^{-1}

A.2 Structural and Dynamic characterization of the *Myxococcus xanthus* Encapsulin system

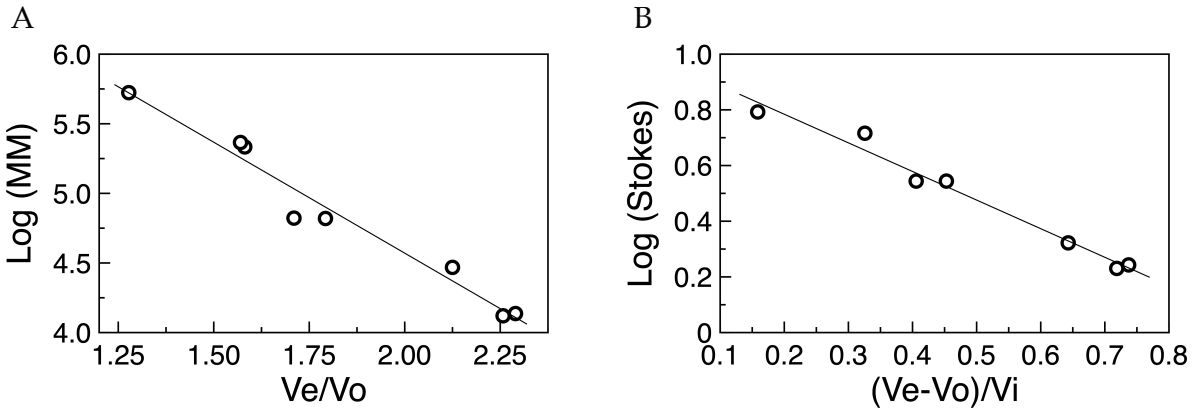

Figure A.4: Calibration of the Superdex 200 10/300 GL SEC column using protein standards. Several protein standards (Catalase, Bovine Serum Albumin, Ovalbumin, Carbonic Anhydrase, and Cytochrome *c*) were used to calibrate the column for the **(A)** apparent molecular weight and the **(B)** hydrodynamic diameter estimations using Equation 4 and Equation 5. The following equations were obtained: Log (MW) = −1.594 × Ve/Vo + 7.758 (R^2 = 0.966) and Log (Stokes) = −1.0279 × (Ve − Vo)/(Vi) + 1.00 (R^2 = 0.978). 1.0 mg/mL of each protein was injected in the column at a flow rate of 0.5 mL/min in 200 mM MOPS buffer, pH 7.0 with 200 mM NaCl.

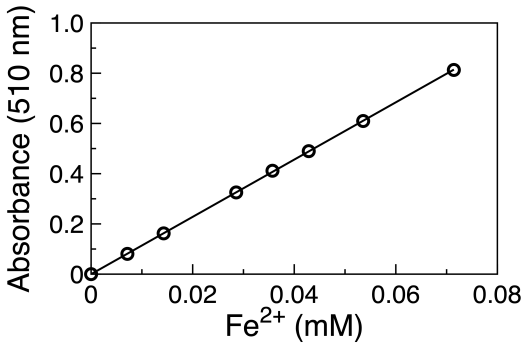

Figure A.5: Iron calibration curve for iron quantification by the 1,10-phenanthroline method. The equation obtained using linear regression at absorbance 510 nm = 11.40 × [Fe^{2+}] (R^2 = 1.00). Each experimental point reflects the mean of triplicates samples.

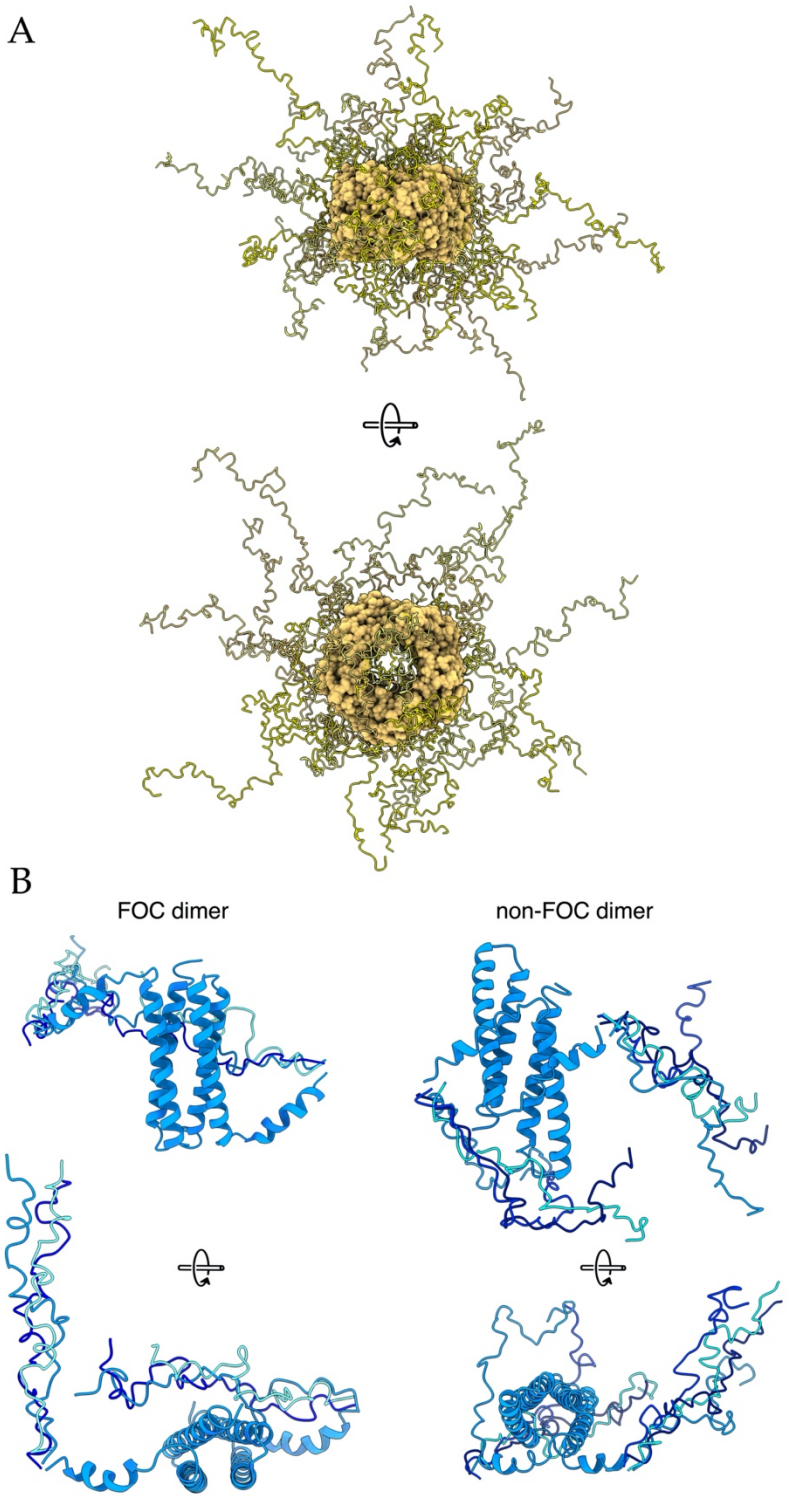

Figure A.6: Cartoon representation of EncB and EncC with atomic model determined by CORAL. Cartoon of **(A)** decameric EncB and **(B)** EncC FOC dimer (left) and EncC non-FOC dimer (right) from its atomic structure (PDB: 7S5K for EncB and 7S8T for EncC) and atomic model determined by CORAL. Views are rotated by 90° according to the axis for each model. The structure of EncB and EncC are colored as gray and the determined flexible region in blue for EncB and in green for EncC.

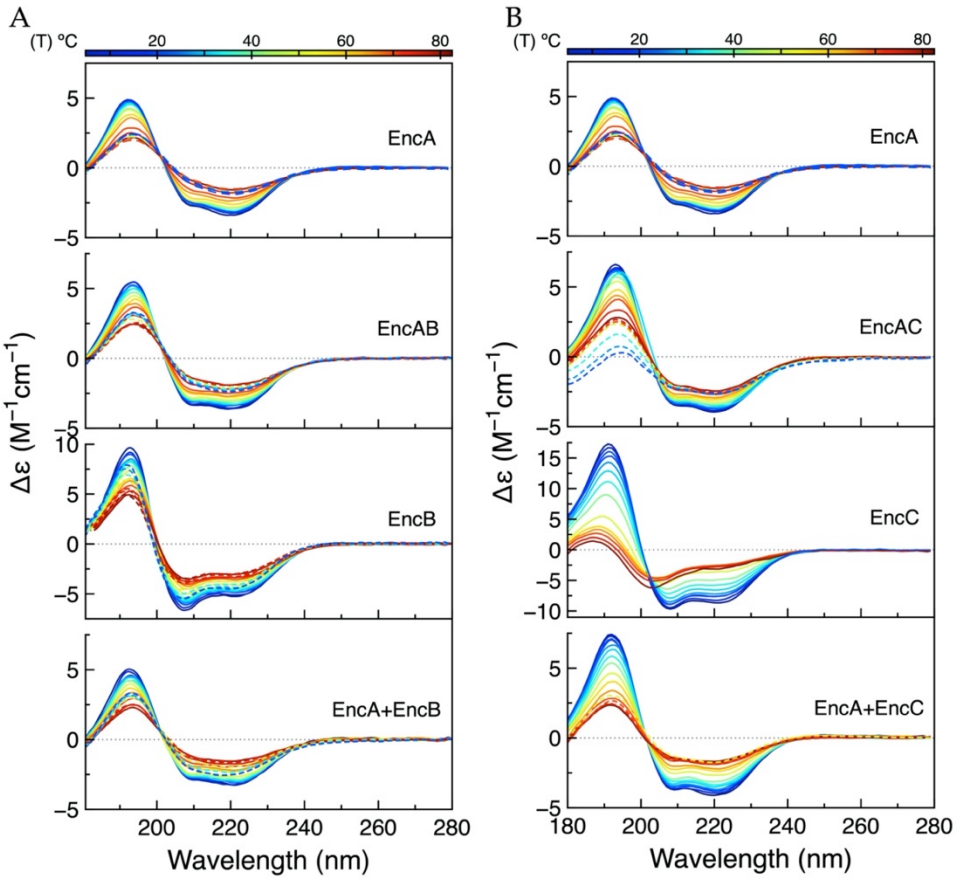

Figure A.7: Assessment of the thermostability of the encapsulin system by SRCD. (A) Molar circular dichroism ($\Delta\varepsilon$) denaturation curves of EncA, EncAB, EncB, EncA+EncAB and **(B)** EncA, EncAC, EncC and EncA+EncC. Full lines represent the spectral evolution as the temperature increases from 4.8 (blue) to 82.5 °C (red) and the dashed lines represents the subsequent temperature decrease.

A.3 Unveiling the Iron Oxidation Mechanism of the Encapsulin System

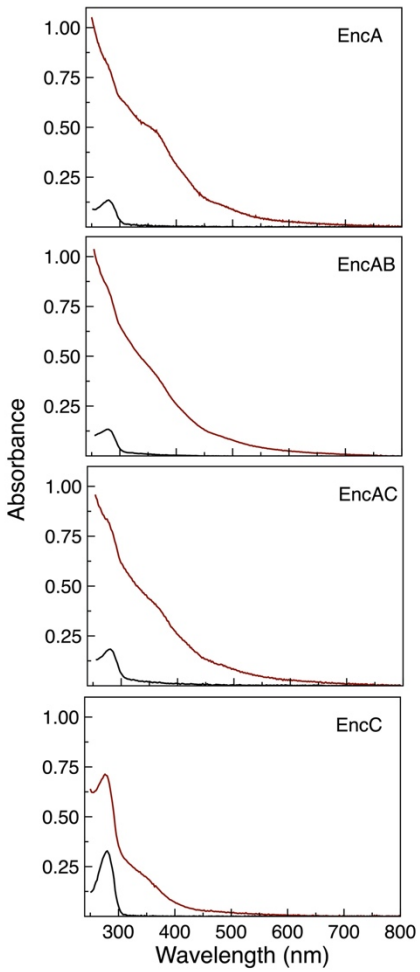

Figure A.8: UV-Visible spectra of the proteins from the *Myx. xanthus* encapsulin system as apo-proteins and after incubation with iron and molecular oxygen. The UV-Visible spectra of EncA, EncAB, EncAC and EncC are shown as black lines and after the addition of ferric iron in the presence of molecular oxygen as red lines.

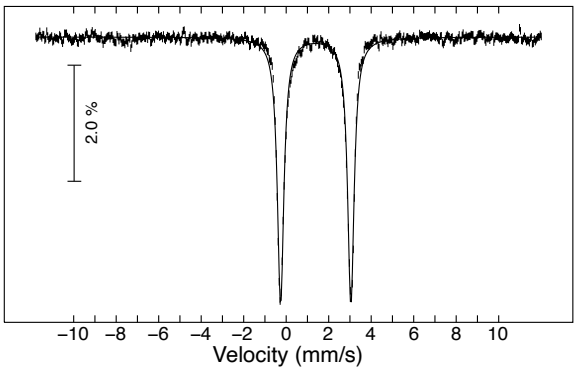

Figure A.9: Mössbauer spectra of 2 mM $^{57}Fe^{2+}$ in anerobic condition. Ferrous iron sulphate was added to 200 mM MOPS buffer pH 7.0 and 200 mM NaCl in an anaerobic chamber, transferred to a Mössbauer cup and frozen at −80 ºC. The spectrum was adquired at 80 K with no external magnetic field.

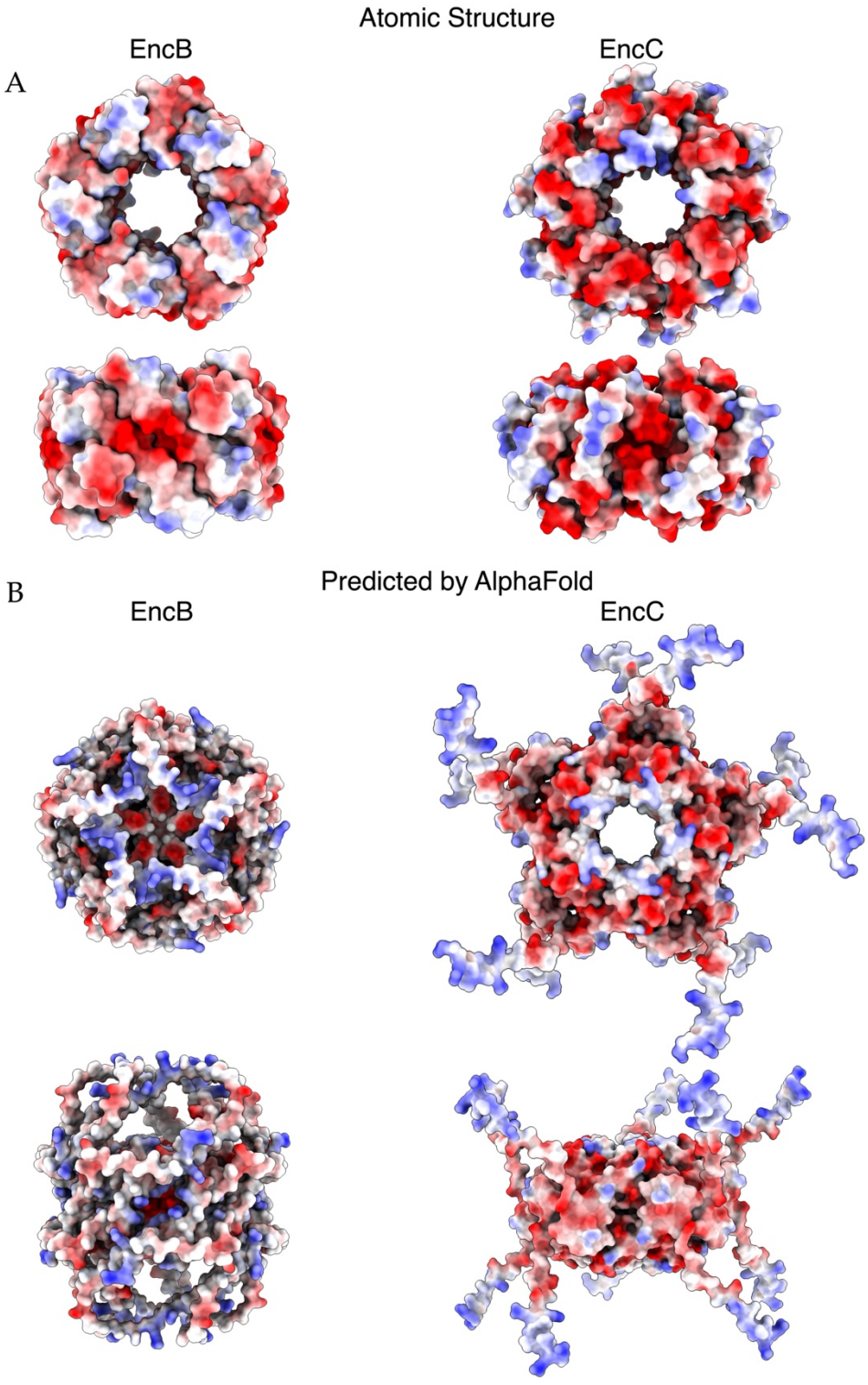

Figure A.10: Comparison of the EncB and EncC electrostatic surface potential from predicted structure. (A) Electrostatic surface potential of EncB (left) and EncC (right) in the atomic structures obtained by X-ray crystallography (7S5K and 7S8T); **(B)** Electrostatic surface potential of EncB (left) and EncC (right) from the structure predicted by the AlphaFold software. The electrostatic surface potentials are colored with a range from red (negative) to white (neutral) to blue (positive) corresponding to a +10 kTe^{-1} to 0 to −10 kTe^{-1}

A.4 Condensation and protection of DNA by *Myx. xanthus* encapsulin: a novel function

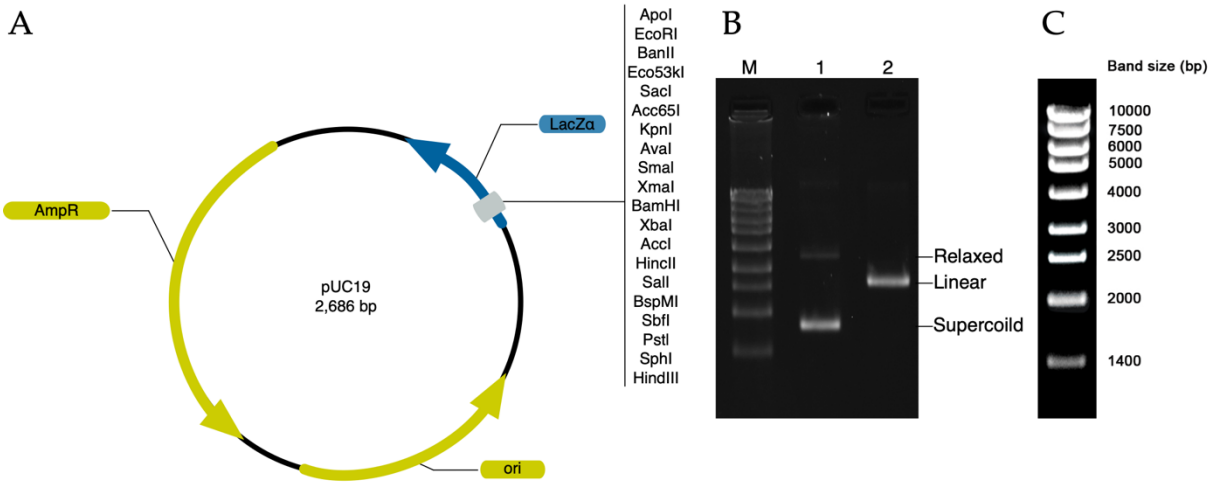

Figure A.11: pUC19 plasmid. (A) vector map of the pUC19 plasmid. **(B)** pUC19 migration profile in 50 mM MOPS pH 7.0 buffer containing 50 mM NaCl: 1 – As isolated pUC19 and 2 – Linear pUC19 after digestion with 10 U of EcoRI (NZYTech) for 1 h, at 37 °C; M – NZYDNA Ladder II. The supercoiled, linear and relaxed forms of the plasmid are identified on the right. **(C)** Band separation of the NZYDNA Ladder II in 1 % agarose gel obtained from the manufacturer's protocol.

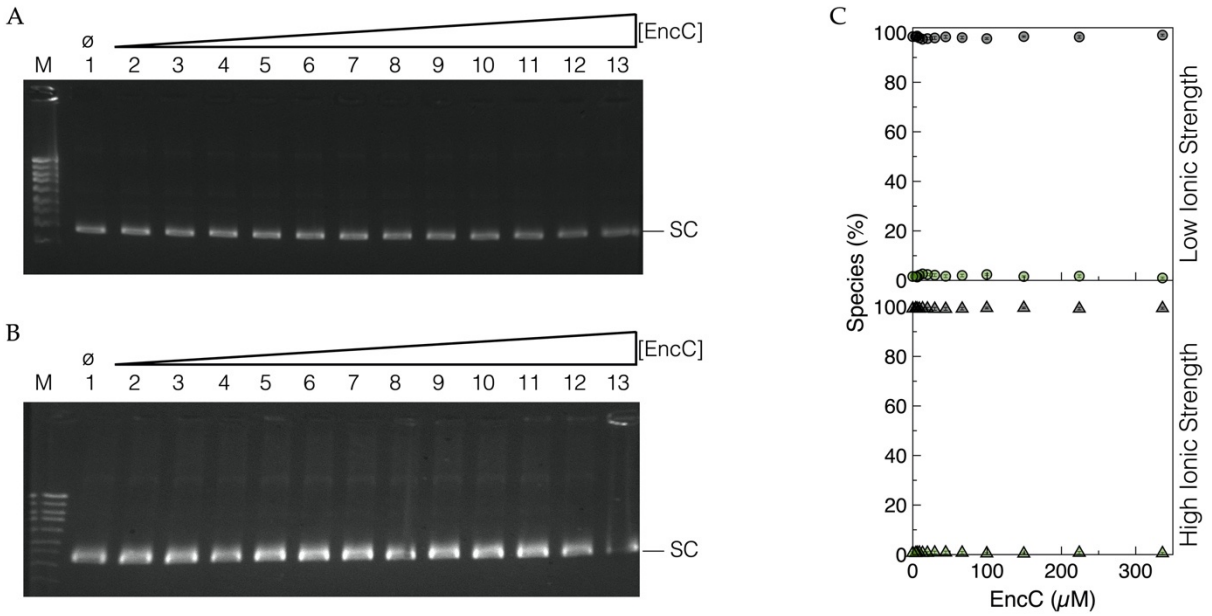

Figure A.12: Binding assay of EncC to supercoiled pUC19 (5 nM) at high and low ionic strength conditions. (A) EMSA in 50 mM MOPS buffer, pH 7.0, 50 mM NaCl and **(B)** 200 mM MOPS buffer pH 7.0, 200 mM NaCl. M – NZYLadder II ; 1 to 13 – Binding of EncC to plasmid DNA with increasing protein concentrations: 0, 3.88, 5.8, 8.7, 13.1, 19.7, 29.5, 44.2, 66.4, 99.6, 149, 224, 336 μM. The free form of the supercoiled plasmid pUC19 band (SC) is labelled. **(C)** Plots from the densitometric analysis of three sets of experiments in either 50 mM MOPS buffer pH 7.0, 50 mM NaCl (circles) and 200 mM MOPS buffer, pH 7.0, 200 mM NaCl (triangles). Free DNA is plotted as full markers, and the protein–DNA complex as empty circles or triangles.

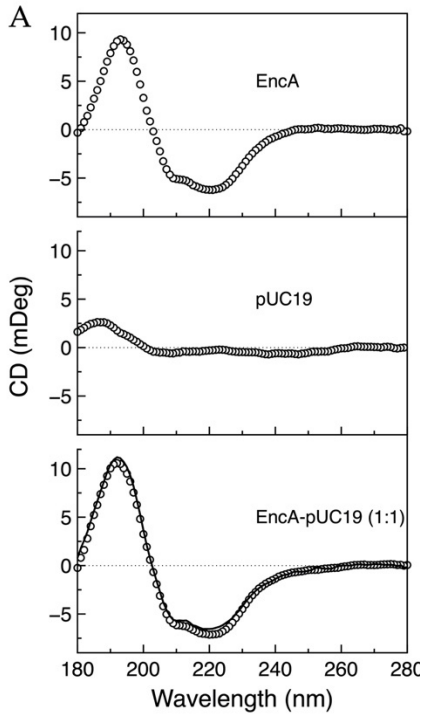

Figure A.13: Circular dichroism spectra of free EncA, pUC19 and the EncA−pUC19 complex. SRCD spectra of EncA (top panel), pUC19 (middle panel) and the EncA–pUC19 complex (top panel) in mDeg at 25 °C.

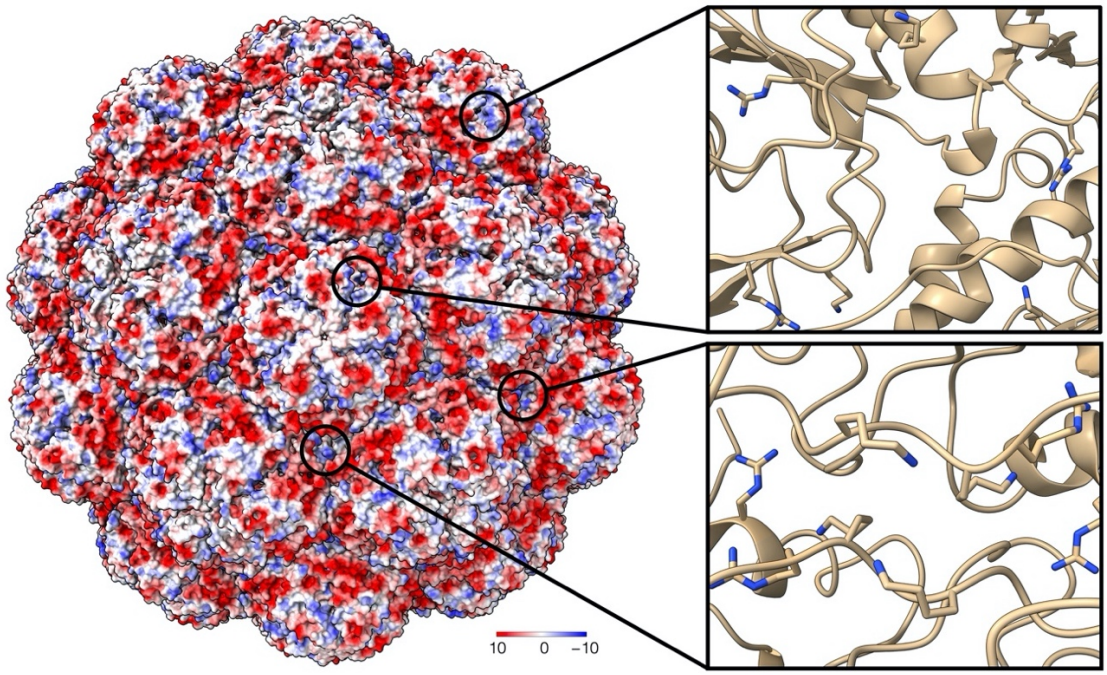

Figure A.14: Electrostatic surface potential representation of *Myx. xanthus* encapsulin EncA. The positively charged pockets (arginine and lysine residues) are magnified and displayed as sticks on the right side. The electrostatic surface potentials are colored with a range from red (negative) to white (neutral) to blue (positive) corresponding to a +10 kTe^{-1} to 0 to −10 kTe^{-1}.

A.5 Iron Release from a Mini-ferritin

A.5.1 Structural features and stability of apo and holo forms of a simple iron sulfur protein

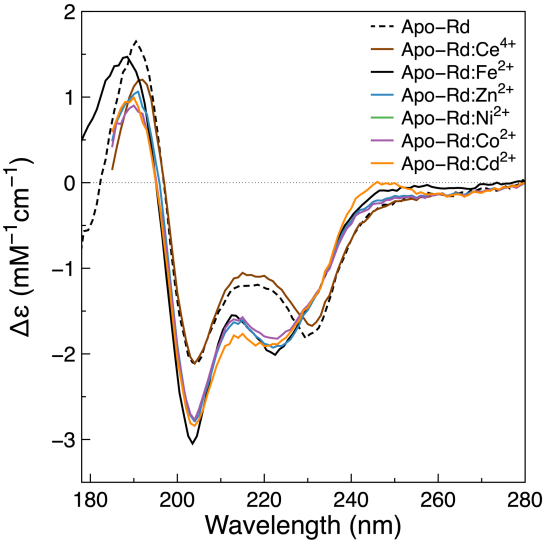

Figure A.15: Circular Dichroism spectra of Rd reconstituted with different transition metals at 25 °C. Buffer: 10 mM MOPS buffer, pH 7.0 with 240 mM NaF.

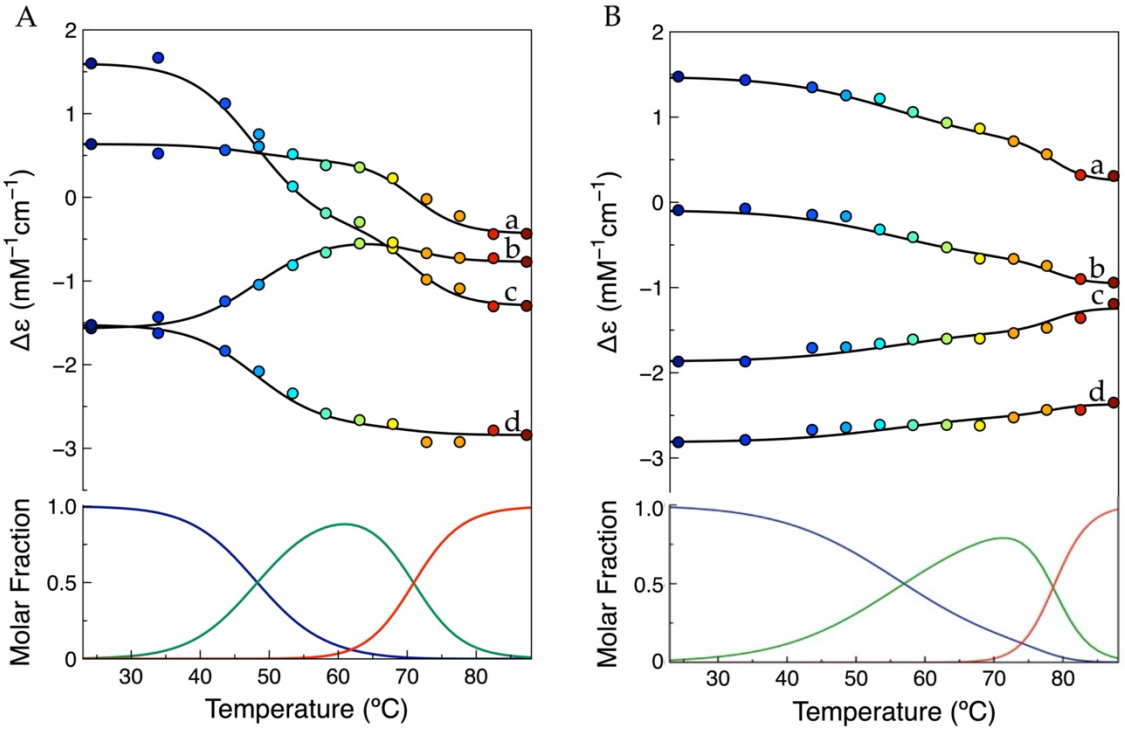

Figure A.16: SRCD temperature profiles for apo- and iron reconstituted holo-Rd. (A) SRCD temperature profiles for apo-Rd. (Top panel) Δε at (a) 184, (b) 233, (c) 191 and (d) 200 nm as a function of temperature. The solid lines

are the result from a global non-linear least-squares fit to the data based on a two-step process. (Bottom panel) Species distribution based on the used model. **(B)** SRCD temperature profiles for reconstituted Rd. (Top panel) $\Delta\varepsilon$ at (a) 189, (b) 196, (c) 226 and (d) 206 nm as function of temperature. The solid lines are the result from a global non-linear least-squares fit to the data based on a two-step process. (Bottom panel) Species distribution based on the used model. Blue, green and red lines represent molar fractions of native, intermediate and thermally denatured states, respectively.

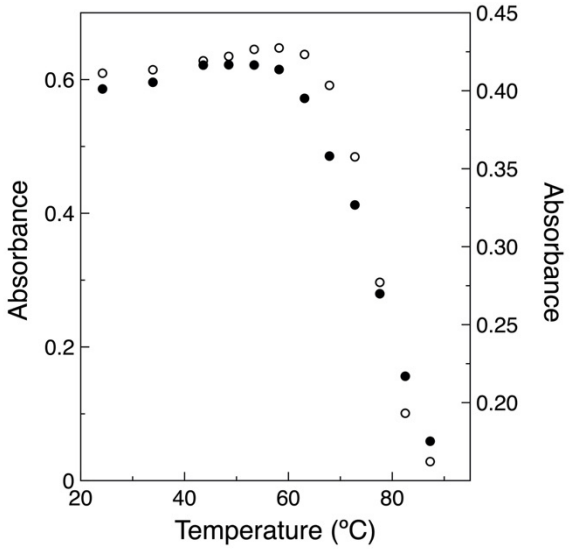

Figure A.17: Absorbance measured at 205 nm for apo-Rd (white circles) and iron reconstituted Rd (black circles) proteins, along a temperature scan from 24 to 87 °C.

A.5.2 Stopped-flow coupled to SRCD

Stopped-flow coupled to SRCD, also termed rapid-SRCD (rSRCD), measurements were obtained in the AMO beam line at ASTRID2 synchrotron radiation source (ISA, Aarhus University, Denmark). The stopped-flow instrument (SX20 Applied Photophysics Limited, Surry, UK) was used as a full automated single mixing mode. The wavelength for the experiment was set via the ASTRID2 control system by controlling the monochromator and the undulator. Both the apo-Rd and holo-Rd (apo-form reconstituted with iron) were mixed with a ferrous sulfate solution (prepared in acidic milli-Q water, pH 2.0, fixed with H_2SO_4) or buffer (10 mM potassium phosphate buffer, pH 7.0, 230 mM NaF) using two independent asymmetric syringes. 0.11 mg/mL of apo-Rd (or holo-Rd) were loaded into a 2.5 mL syringe while the Fe^{2+} solution or buffer was loaded into a 0.25 mL syringe providing a 1:10 ratio mixing. The protein and the Fe^{2+} were mixed in a T-mixer and measured in a 10-mm pathlength quartz cell at room temperature. Measurements at each wavelength (280, 232, 229, 221, and 202 nm) were done at least 10 times during 10 sec with 4000 points (400 points/sec) at high intensity (15 nm slit setting). The total mixing time (or dead time) was considered to be 0.04 sec. 10 separate measurements at each wavelength were averaged and data smoothing was applied by a 4 points moving average.

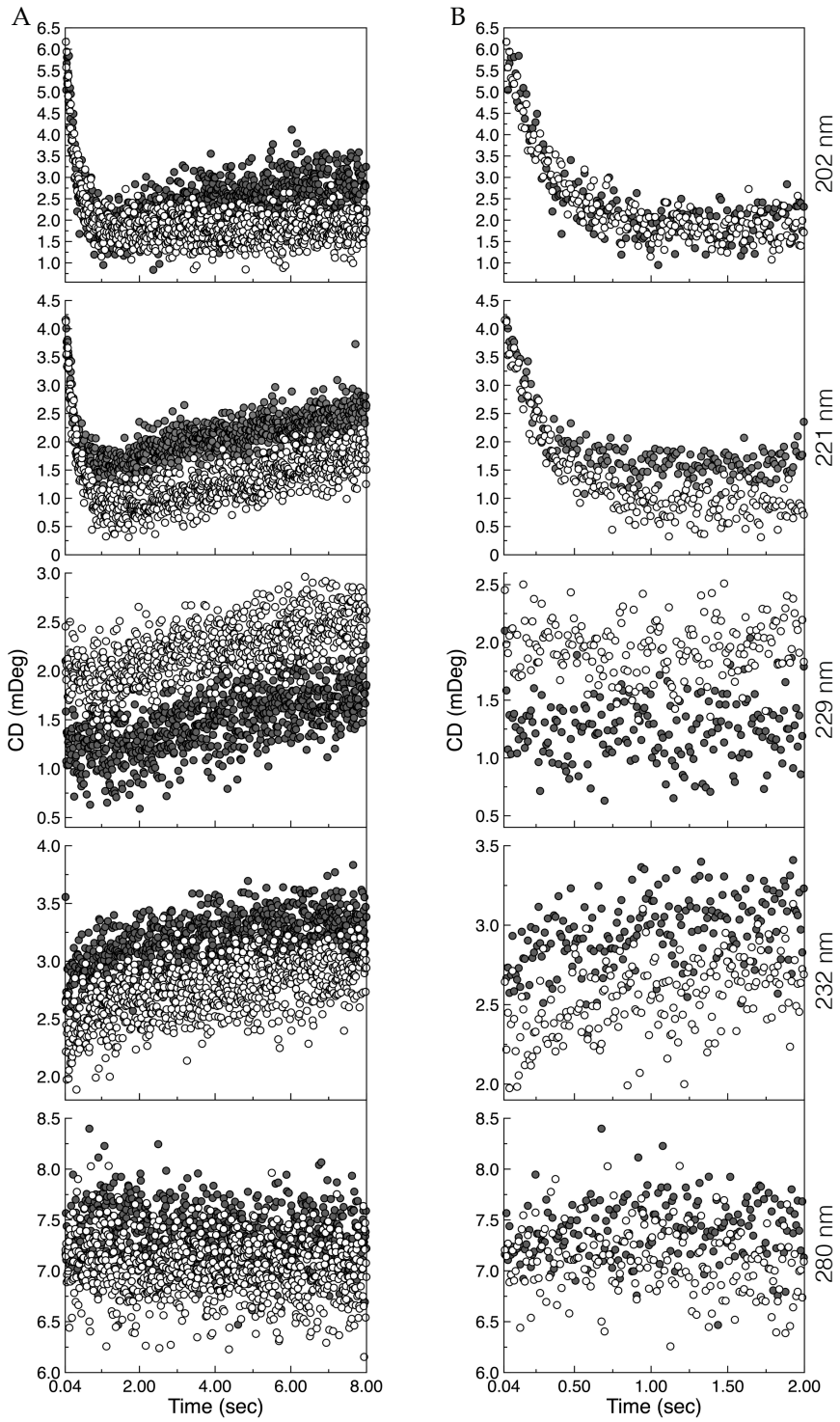

Figure A.18: Ferrous iron binding kinetics of apo-Rd by rSRCD. The reaction was monitored by measuring the secondary structure changes associated with the conversion of the apo-Rd into its iron loaded form at 280, 232, 229, 221, 202 nm (top to bottom). Reaction of apo-Rd with 1.8 Fe^{2+}/protein (empty circles) and 2.7 Fe^{2+}/protein (full circles). The reaction was monitored for **(A)** 8 s **(B)** and 2 s.

A.5.3 Iron Release from a Mini-ferritin

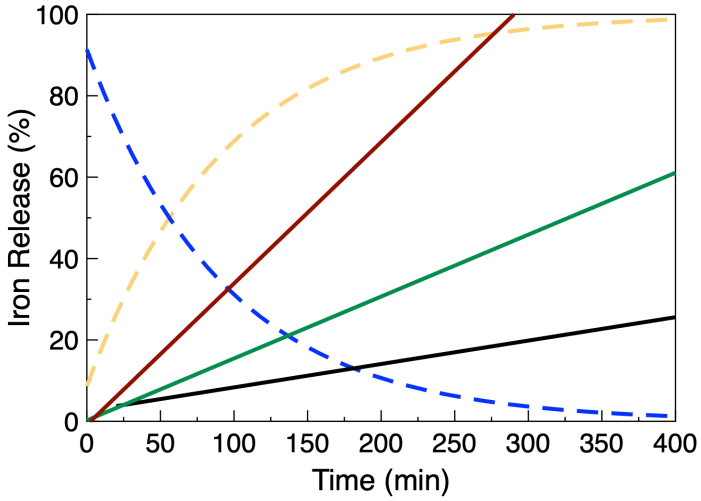

Figure A.19: Theoretical curves of the iron reduction and release from *Mar. hydrocarbonoclasticus* Dps. In a previously characterized reduction mechanism, the mineral core (blue dashed line) is reduced to ferrous iron (yellow line) either inside or outside the mini-ferritin. Under anaerobic conditions the iron is released (black line) following a zero-order equation and its rate of release is increased upon addition of 1,10-phenantroline (green line). In the presence of apo–Rd (red line) the rate of release still follows a zero-order equation with an increased rate.

2022

ANA VIANA DE ALMEIDA

FUNCTIONAL CHARACTERIZATION AND DESIGN OF PROTEIN NANOCAGES