Structural Basis of Enzyme Encapsulation into a Bacterial Nanocompartment

A dissertation submitted to the

ETH ZÜRICH

for the degree of

Doctor of Sciences (Dr.sc.nat.)

presented by

Markus Sutter

Dipl.-Biol. ETH Zürich

born December 16, 1976

citizen of Wald (ZH)

accepted on the recommendation of

Prof. Dr. Nenad Ban, Examiner

Prof. Dr. Eilika Weber-Ban, Co-examiner

Prof. Dr. Kaspar Locher, Co-examiner

2008

SUMMARY

Compartmentalization of enzymes confines metabolically related reactions and occurs in all domains of life. Large eukaryotic compartments are organelles such as mitochondria, lysosomes or the peroxisomes which are enclosed by lipid bilayers. On a much smaller scale, microcompartments built entirely of protein subunits have been described. A prominent example of a bacterial microcompartment is the carboxysome, which packages the enzyme ribulose bis-phosphate carboxylase oxygenase (RuBisCO) by forming polyhedral bodies. Similar microcompartments built from related shell-forming proteins, *eut* and *pdu* organelles, are also likely to package enzymes involved in metabolic reactions. Formation of cellular or molecular compartments increases the local concentration of enzymes, facilitates substrate transfer between a series of connected reactions, and sequesters potentially toxic substrates or products. Ferritin, for example, is an octahedral protein cage that oxidizes the highly toxic $Fe^{(II)}$ ions and deposits them into its cavity as an insoluble ferrihydrite form. Some multienzymes and enzyme complexes, such as the pyruvate dehydrogenase complex, lumazine synthase or the fungal fatty acid synthase, form reaction chambers into which substrates can enter but other macromolecules are excluded.

A large, yet little characterized family of conserved bacterial proteins forms large molecular weight assemblies with a diameter of about 20 nm. In the literature they are referred to as linocin-like proteins based on the observation that they are abundant in the culture supernatant of *Brevibacterium linens* which exhibits bacteriostatic activity towards various strains of *Listeria, Corynebacterium, Brevibacterium, Arthrobacter* and *Bacillus.* A close homologue of the *B. linens* protein (58 % identity) was found in *Mycobacterium tuberculosis.* However, neither the native nor recombinant form of this protein, CFP29 (culture filtrate protein with a mass of 29 kDa), displayed bacteriostatic activity against various *Listeria* strains. In *Thermotoga maritima* the homologous protein was characterized as a homomultimeric protease. It was shown to have proteolytic activity against various small synthetic peptides. Bactericidal activity against closely related species was tested but no growth inhibition was found.

In light of the varied biochemical data regarding the function of this poorly characterized protein family we decided to study it structurally. The protein from the thermophilic organism *T. maritima* was purified, crystallized and the structure was determined by X-ray crystallography. Based on the structural information we were able to design biochemical and electron microscopic experiments to establish that these proteins, which we refer to as

encapsulins, form nanocompartments that contain ferritin-like proteins or peroxidases, both enzymes involved in oxidative stress response. We demonstrate that these enzymes are specifically targeted to the interior of encapsulins via unique C-terminal extensions. We also show that the previously proposed function as a bacteriocin is unlikely since highly purified encapsulin shows no bactericidal activity. Encapsulation occurs in the native organism, as shown by the presence of the ferritin-like protein in *T. maritima* encapsulin crystals as well as when encapsulin and its cargo protein are expressed recombinantely, as shown in the case of *B. linens* DyP peroxidase/encapsulin.

ZUSAMMENFASSUNG

Die Kompartimentierung von Enzymen gruppiert zusammengehörige Stoffwechselprozesse und kommt in allen Lebewesen vor. In eukaryotischen Organismen gibt es grosse Organellen wir Mitochondrien, Lysosomen und Peroxisomen, die von einer Lipid-Doppelschicht umgeben sind. Es gibt aber auch viel kleinere Kompartimente, die aus Proteinen aufgebaut sind, sogenannte Mikrokompartimente. Das bekannteste Beispiel ist das Carboxysom, das polyedrische Körper bildet und in seinem Inneren das Enzym Ribulose-1,5-bisphosphat-carboxylase-oxygenase (RuBisCO) verpackt. Es gibt andere ähnlich aufgebaute Mikrokompartimente, die auch für bestimmte Stoffwechselprozesse verantwortlich sind.

Der Vorteil der Kompartimentierung ist dass die lokalen Enzymkonzentrationen erhöht werden und die Substrate einfacher und schneller zwischen den Enzymen transportiert werden können. Es werden auch eventuell toxische Reaktionen von Substraten oder Produkten vermieden. Ferritin ist ein oktaedrischer Proteinkomplex der die höchst schädlichen $Fe^{(II)}$ Ionen oxidiert und in seinem Inneren in einer unlöslichen Ferrihydrit-Form ablagert. Viele Multienzymkomplexe, wie z.B. Pyruvatdehydrogenase, Lumazinesynthase oder die Fettsäuresynthase aus Pilzen, bilden grosse Reaktionskammern, in denen die Reaktion von Substraten katalysiert wird, aber andere grosse Moleküle ausgeschlossen werden.

Eine grosse, wenig erforschte Familie von Proteinen bildet Strukturen mit einem Durchmesser von etwa 20 nm. Sie werden in der Literatur "Linocin" genannt, weil das Protein im Kulturüberstand von *Brevibacterium linens* vorkommt, der einen bakteriostatischen Effekt auf verschiedene Stämme der Gattungen der *Listeria, Corynebacterium, Brevibacterium, Arthrobacter* und *Bacillus* hat. Ein naher Verwandter des *B. linens* Proteins wurde in *Mycobacterium tuberculosis* gefunden (58 % Identität). Dieses Protein (genannt CFP29, von Kulturfiltratprotein mit einer Masse von 29 kDa) zeigte aber keine bakteriostatische Aktivität

iv

gegen Listerien, weder die nativ aufgereinigte noch die rekombinant produzierte Variante. In *Thermotoga maritima* wurde das homologe Protein als Protease charakterisiert da es proteolytische Aktivität gegen verschiedene kleine Tripeptid-Fragmente zeigte. Eine bakteriostatische Aktivität gegen nah verwandte Stämme konnte nicht nachgewiesen werden. Angesichts der unterschiedlichen biochemischen Daten bezüglich der Funktion dieser Proteine haben wir beschlossen das Protein strukturell zu untersuchen. Das Protein vom thermophilen Organismus *T. maritima* wurde nativ aufgereinigt, kristallisiert und die hochauflösende 3D Struktur bestimmt. Basierend auf dieser Struktur wurden biochemische und elektronenmikroskopische Studien durchgeführt, die beweisen dass diese Proteine, die wir Enkapsuline nennen, Nanokompartimente bilden, in deren Innerem Ferritin-ähnliche Proteine oder Peroxidasen verpackt werden, beides Enzyme, die etwas mit oxidativem Stress zu tun haben. Es wird gezeigt, dass diese Enzyme durch ihre C-terminalen Enden spezifisch an die Innenseite der Enkapsuline binden und so verpackt werden. Die früher angenommene Aktivität als Bakterizid ist sehr unwahrscheinlich ist da hoch aufgereinigtes Enkapsulin keine Aktivität zeigt. Die Enkapsulation erfolgt sowohl im nativen Organismus, wie gezeigt durch die Anwesenheit des Ferritin-ähnlichen Proteins in *T. maritima* Enkapsulin Kristallen, als auch wenn Enkapsulin und das verpackte Protein rekombinant exprimiert werden, wie gezeigt am Beispiel von *B. linens* Peroxidase/Enkapsulin.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Compartmentalization and Bacterial Microcompartments

Compartmentalization was a very important step in the emergence of cellular life since without a separating barrier from the environment it is not possible to keep the vital components from diffusing away. Examples of compartmentalization are found on all scale levels, from the grouping of cells in higher eukaryotes (organs) to intracellular subcompartments responsible for certain metabolic processes (organelles like mitochondria, lysosomes and chloroplasts) to bacterial microcompartments in prokaryotes.

1.1.1 Carboxysomes and related microcompartments

Bacterial microcompartments are intracellular structures formed by proteins that contain enzymes for a specific metabolic process. The best characterized of those is the carboxysome. They were first discovered as polyhedral bodies in electron micrograph thin sections of various cyanobacteria (Gantt and Conti 1969) and *Halothiobacilli* species (Shively, Decker et al. 1970). They were initially thought to be viruses but purification of these particles showed that they contain large amounts of the enzyme ribulose diphosphate carboxylase (RuBisCO, which is involved in fixing carbon dioxide) and these inclusions were therefore termed carboxysomes (Figure 1A) (Shively, Ball et al. 1973). Two major protein components of the surrounding shell have only recently been structurally characterized (Kerfeld, Sawaya et al. 2005; Tanaka, Kerfeld et al. 2008). The shell consists of two types of proteins forming either hexamers or pentamers which then arrange in an approximate T=75 icosahedral symmetry to form large particles with a diameter of about 1200 Å (Figure 1B). The shell seems to contain selective pores through which substrates or products can diffuse but larger molecules are excluded.
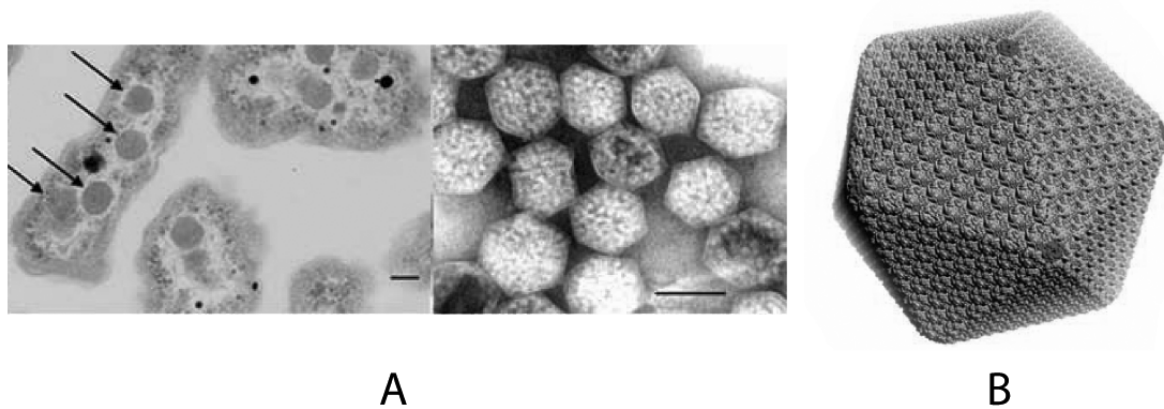
Figure 1: Carboxysomes in *Halothiobacillus neapolitanus* and a structural model. A: (left) Thin section of a *H. neapolitanus* cell, carboxysomes are indicated with arrows, (right) purified carboxysomes (scale bars 100 nm). B: Model of the carboxysome with hexamer subunits in light gray and pentamers at vertices in dark gray. From (Tanaka, Kerfeld et al. 2008), reprinted with permission from AAAS.

Similar bacterial microcompartments are also found in other species where they are responsible for metabolic processes. In the nonphotosynthetic bacterium *Salmonella enterica* a similar microcompartment was found to package diol dehydratase (Havemann, Sampson et al. 2002). This enzyme containing adenosyl-cobalamin (coenzyme B_{12}) is responsible for degradation of 1,2-propanediol. The presence of this microcompartment is only observed when the cells are grown on 1,2-propanediol as the only energy and carbon source. The operons for these large intracellular complexes contain at least 14 different proteins, some responsible for forming the shell and others coding for catalytic proteins.

In *Salmonella typhimurium* a similar operon structure was discovered containing 17 genes responsible for carboxysome-like shell proteins and enzymes involved in the use of ethanolamine as a carbon source (Kofoid, Rappleye et al. 1999).

1.1.2 Lumazine synthase

Lumazine synthase catalyzes the penultimate step of riboflavin synthesis in plants, fungi and microorganisms. The fungal enzyme forms to a pentamer while in plants the pentamers further assemble into a spherical shell with icosahedral symmetry (Persson, Schneider et al. 1999). This shell contains three molecules of the enzyme catalyzing the following step (the reaction to riboflavin) forming an $\alpha_3\beta_{60}$ complex (Ladenstein, Ritsert et al. 1994). This arrangement makes the reaction much more efficient by preventing the intermediate product from diffusing away. It is unclear though how the three α subunits are targeted to the interior of the shell.

2

Large enzymatic complexes are often encoded by single polypeptides fulfilling both structural and enzymatic roles. An example is the fungal fatty acid synthase where six alpha and beta subunits assemble to a large functional complex (Jenni, Leibundgut et al. 2007). The enzymatic domains are connected by structural parts and the whole complex forms two reaction chambers where the elongation cycle of the fatty acid synthesis occurs.
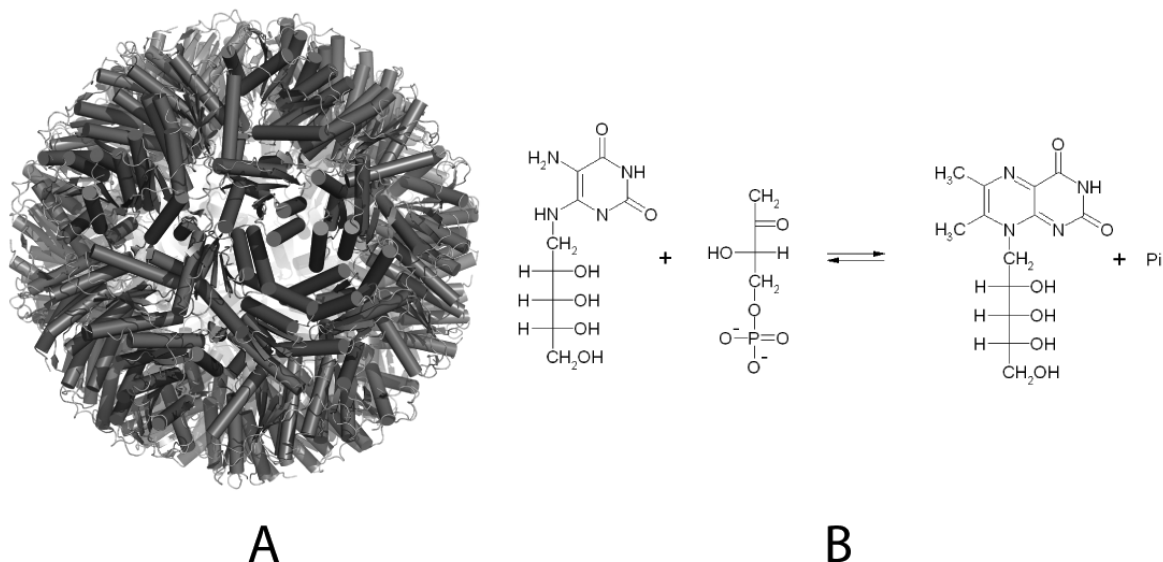
Figure 2: Structure and reaction of lumazine synthase. A: Structure of the spinach protein which forms an icosahedral shell of 60 subunits with an outer diameter of 150 Å and an inner diameter of about 70 Å. B: Reaction catalyzed by lumazine synthase (2,5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione and 3,4-dihydroxy-2-butanone 4-phosphate to 6,7-dimethyl-8-ribityllumazine).

1.2 Encapsulins / linocin-type Proteins

The protein which is described here is a member of a large, yet little characterized family (Figure 4) of conserved bacterial proteins (recently also detected in the archae *Candidatus Methanoregula*). In the literature they are referred to as linocin-like proteins based on the observation that they are abundant in the culture supernatant of *Brevibacterium linens* which exhibits bacteriostatic activity towards various strains of *Listeria, Corynebacterium, Brevibacterium, Arthrobacter* and *Bacillus*. These proteins were characterized as large molecular weight assemblies with a diameter of about 20 nm shown by gel filtration and negative stain electron microscopy experiments of the *Brevibacterium linens* (Figure 3) and *Thermotoga maritima* homologues (Valdes-Stauber and Scherer 1994; Hicks, Rinker et al. 1998). PCR analysis showed the presence of the *lin* gene in a majority of the tested strains of *Brevibacterium, Arthrobacter* and *Corynebacterium*. The homologous member of this protein

in *Mycobacterium tuberculosis* is called CFP29, from culture filtrate protein with a mass of 29 kDa (Rosenkrands, Rasmussen et al. 1998). It was found to be strongly immunogenic in mouse models. Even though the identity to the *B. linens* protein is 58 % it did not share its activity against various *Listeria* strains, neither the native not the recombinantely produced sample. Immunoblotting procedures localized the protein in culture supernatant and aqueous membrane fractions.
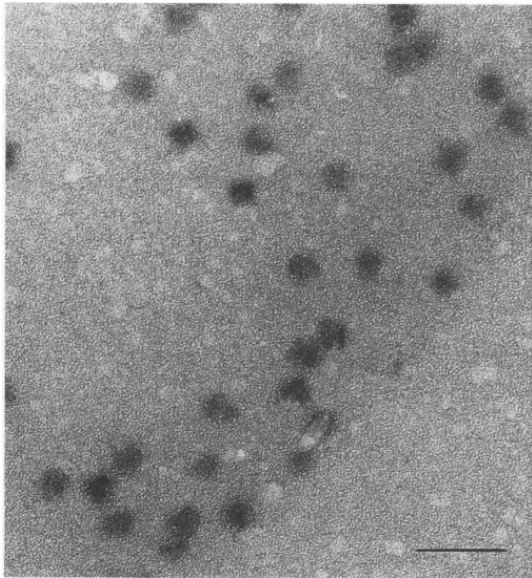
Figure 3: Transmission electron micrograph of globular structures as observed by Valdes-Stauber (Valdes-Stauber 1995).

In *Thermotoga maritima* the homologous protein was characterized as a homomultimeric protease (Hicks, Rinker et al. 1998). It was shown to have activity against various small synthetic peptides. As opposed to the other two characterized members this one was not purified from the culture supernatant but from the cytosolic fraction.

Figure 4: Alignment of selected encapsulin proteins. Red indicates identity and orange homologous substitution. Indicated below the alignment are the helical (cylinders) and beta sheet (arrows) regions deduced from the *T. maritima* crystal structure. The helix and beta sheet depictions are colored according to which domain they belong to, orange for P, blue for A and cyan for E. The numbers on top of the alignment refer to the *T. maritima* sequence. The residues within 4 Å of the cargo binding site are marked with an asterisk.

1.3 *Brevibacterium linens* and *Thermotoga maritima*

B. linens is a gram positive bacterium which is important for the ripening of cheese surfaces. It grows aerobically at temperatures between 8 to 37 °C with an optimum between 21 to 25 °C and a pH range of 6.0 to 9.8. One of the commonly used criteria to separate it from other bacteria is the relatively high salt tolerance of up to 15 %. The red pigment formation (carotenoids) is dependent on light. *B. linens* is used in the dairy industry in the red cheese smear (Figure 5A) where its presence prevents the growth of pathogenic *Listeria* strains (Eppert, Valdes-Stauber et al. 1997).
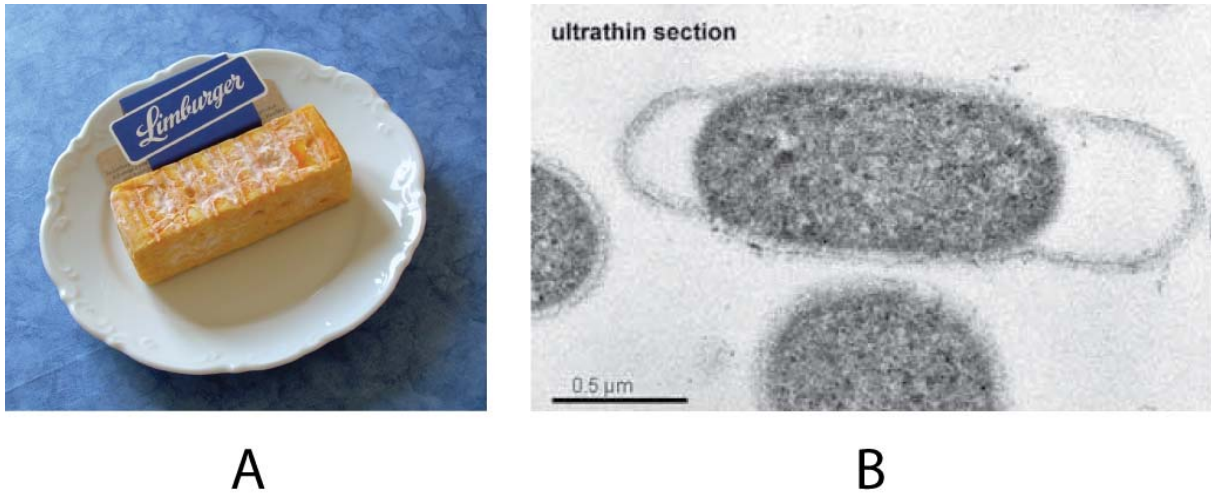
Figure 5: *B. linens* and *T. maritima*. A: Limburger cheese. The red part on the cheese is partly due to *B. linens* which is added to the smear and contributes to the ripening by its proteolytic activity and production of aroma and flavor constituents (Lenoir 1984). B: Thin section of a *T. maritima* cell (Image Source: K.O. Stetter)

T. maritima was first isolated from marine geothermal sediment near Vulcano, a small volcanic island close to Sicily. The organism grows at an optimal temperature of 80 °C and is a non-spore forming, rod-shaped bacterium (Figure 5B). The genome sequence of this organism was completed in 1999 and contains 1.8 Mbp. Analysis has revealed a high degree of similarity between *T. maritima* and archaea with nearly a quarter of the genome being archaeal in nature, indicating a lateral gene transfer between thermophilic eubacteria and archaea (Nelson, Clayton et al. 1999).

1.4 Iron Metabolism and Oxidative Stress Response

Iron is one of the most abundant metals in the earth's crust and is found to be essential for survival of almost all life forms. An important property of iron is its ability to exist in two different valence states, Fe(II) and Fe(III). It is found in active sites of enzymes involved in electron transport and oxygen carrier proteins. Even though iron is ubiquitous, it is a challenge for the cell to acquire it since under oxidising conditions and at neutral pH it is extremely insoluble, the concentration of soluble Fe(III) is only 1.4×10^{-9} M at pH 7. This is due to the reaction of water and oxygen with iron which produces insoluble hydrated oxides. Free iron is toxic to the cells when it reacts with hydrogen peroxide in the Fenton reaction:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + OH^{\cdot}$$
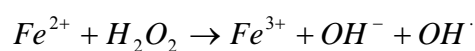
The resulting hydroxyl radicals can cause peroxidation of lipids and DNA strand breaks (Imlay, Chin et al. 1988). In a living cell most of the iron is bound to proteins or other iron

6

chelating agents like siderophores. Different kinds of proteins are needed for either acquisition of precious iron, transport, uptake into the cell or storage.

1.4.1 Ferritin

Ferritins are present in nearly all species and play a role in iron storage. This protein consists of 24 subunits with octahedral 432 symmetry of a small four helix bundle protein. The active site (called ferroxidase center) lies between two helices and is responsible for oxidization of iron from the Fe(II) form to an Fe(III) ferrhydrite form which is deposited in the center of the ferritin shell (Macara, Hoy et al. 1972). One ferritin molecule can store up to 4500 Fe(III) atoms in the center of the shell which measures 125 Å outer and 80 Å inner diameter (Figure 6). The structural conservation is very high but the primary sequences share low homology. In mammals, ferritins are composed of H and L chains. Depending on the tissue, a different composition of each chain is observed.

In bacteria there is an additional type of ferritin called bacterioferritin, which differs from animal and plant homologues by the presence of a heme group but it fulfills the same function (Andrews 1998).
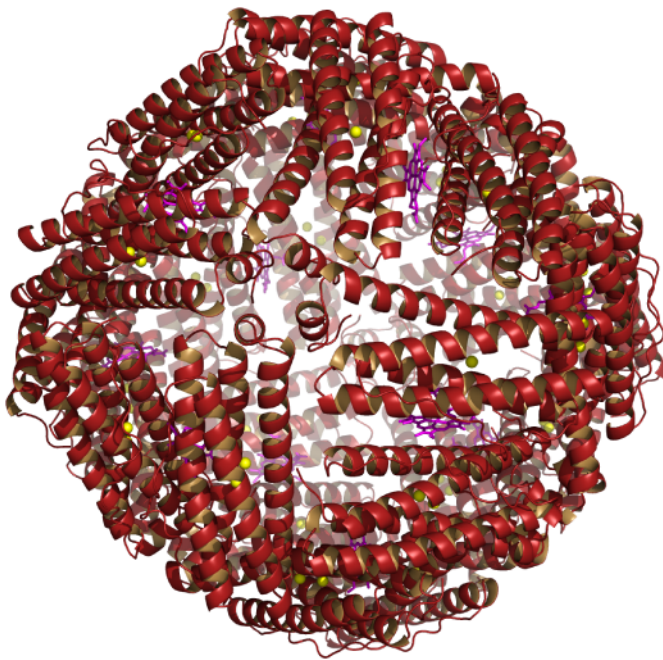
Figure 6: *E. coli* Bacterioferritin. The *E. coli* bacterioferritin consists of a 432 symmetrical arrangement of 4-helix bundle monomers. The 4-fold axis is visible in the center and the bound ions of the di-iron ferroxidase center (manganese in this structure) are indicated as yellow spheres. In between the subunits are heme groups which are drawn in stick form in magenta.

1.4.2 Oxidative Stress

All aerobic cells produce reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), superoxide ($O_2^{\cdot-}$) and the hydroxyl radical ($OH^{\cdot}$) which are normally in balance with antioxidants such as catalase and superoxide dismutase (Figure 7). Disruption of this balance due to either increase of ROS or depletion of antioxidants leads to oxidative stress (Finkel and Holbrook 2000).

While ROS are in general harmful to the cell causing random damage to proteins, DNA and lipids, they fulfill functions in signaling or host defense, especially in eukaryotes. Superoxide plays an important role in macrophages as a chemical defense mechanism against invading bacterial pathogens (Babior 1984). Cells have evolved an intricate network to sense the presence of oxidative stress which leads to induction of defensive enzymes (Storz and Imlay 1999).
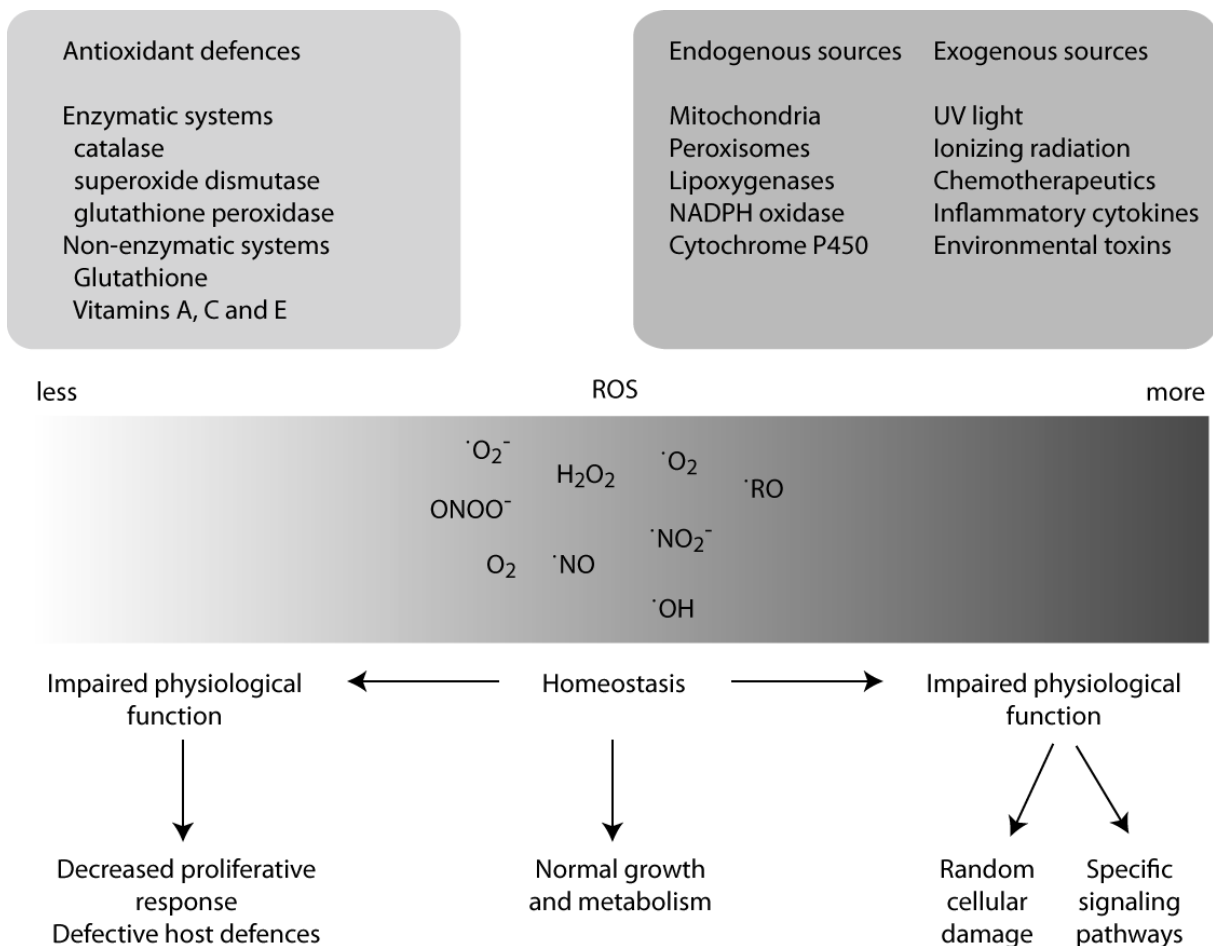
Figure 7: Schematic of the sources and effects of reactive oxygen species (ROS). ROS are harmful to the organism yet some of them are essential for metabolic processes. This requires a sophisticated regulation of the cellular antioxidant molecules. Adapted from (Finkel and Holbrook 2000).

1.5 Aims of this PhD project

Although encapsulins are found in bacteria from various different habitats, they have been poorly characterized so far which is in part due to the lack of structural information. The number of organisms where encapsulins are found increases steadily as new genome sequences become available. All the new sequences are annotated according to the work done on the *B. linens* protein ("Linocin").

The aim of this work is to elucidate the structure and function of this family of proteins. Experiments were designed to pursue structure determination at high resolution. This includes protein purification from native source material, crystallization, data collection and solving the phase problem using non-crystallographic symmetry averaging. From the clues provided by the crystal structure, the interaction with DyP or ferritin-like proteins was discovered and biochemical and electron microscopic experiments were performed to determine the structural basis of the packaging of these enzymes both in the native organism as well as when they are expressed recombinantely. This work provides the basis for further work on this family of proteins and provides insights into the general packaging principles that might be employed by various bacterial microcompartments.

2 MATERIALS AND METHODS

2.1 DNA Methods

2.1.1 Extraction of genomic DNA

Genomic DNA for cloning purposes was extracted from *Brevibacterium linens* and *Thermotoga maritima* using the following protocol (Bickley and Owen 1995). 1.5 ml of a bacterial culture was pelleted and then resuspended in 100 µl TE buffer (10 mM Tris·HCl pH 8.0, 1 mM EDTA pH 8.0). 600 µl GET buffer (5 M Guanidine hydrochloride, 100 mM EDTA pH 8.0, 0.5 % (v/v) Triton X-100) was added and the solution was mixed by slowly pipetting up and down. After addition of 250 µl 7.5 M NH_4OAc the solution was mixed by inverting the tube several times. DNA was then extracted by adding 500 µl chloroform reagent (chloroform:isoamyl alcohol, 24:1), mixing and incubating on ice for 10 min. After centrifugation for 10 min at 13 krpm in a microfuge the upper aqueous phase was recovered and then precipitated with 430 µl isopropanol. After a centrifugation step (13 krpm, 10 min) the supernatant was removed and the pellet washed three times with 1 ml 70 % ethanol. The pellet was then resuspended by addition of 100 µl TE buffer and incubation at 4 °C overnight on a shaker. DNA concentration was then determined using absorbance measurement at 260 nm.

2.1.2 PCR from genomic DNA

Primers for PCR (Polymerase chain reaction) were ordered from Microsynth and resuspended in a volume of ddH_2O corresponding to a final concentration of 100 µM. Pipetting scheme for PCR:

Pwo proofreading polymerase (2.5 U/µl)	1.0 µl
10 x Pwo complete buffer	5.0 µl
dNTP mix (10 mM)	1.0 µl
primer 1 (100 µM)	0.5 µl
primer 2 (100 µM)	0.5 µl
template DNA (100 ng/µl)	1.0 µl
ddH_2O	41.0 µl

All components were mixed on ice before starting the thermocycling reaction:

1.   3 min 97 °C
2.   **30 s   97 °C**
3.   **30 s   55-65 °C**
4.   **3 min 72 °C, back to step 2 for 29 times**
5.   5 min 72 °C

PCR products were analyzed on a 1 % agarose gel and visualized using ethidium bromide stain. Correct fragments were cut out of the gel and extracted using Qiagen Gel Extraction Kit. This DNA was then subjected to restriction digest or subcloned into TOPO vector (Shuman 1994) for sequencing.

2.1.3 Restriction enzyme digest and ligation

Restriction enzymes from Fermentas and New England Biolabs were used in the manufacturer's recommended buffers. In tests for the presence of restriction sites, 10 µl total volume assays were used, containing 0.2 µl of restriction enzyme 1 and 2, 1 µl of 10x buffer and 8.6 µl DNA. This mixture was incubated at 37 °C for 1 h and then analyzed on a 1 % agarose gel. For preparative digests volumes of 30 to 50 µl were used with appropriate enzyme amounts and digestion times from 2 to 12 hours. Plasmids were additionally treated with CIP (Calf Intestine Phosphatase) to prevent self-ligation.

Ligation was performed with T4 DNA Ligase from Fermentas. In a typical ligation reaction, 1 µl of plasmid DNA, 1 µl of insert and 6.5 µl ddH$_2$O were incubated at 45 °C for 5 min, then transferred on ice and 1 µl of 10 x ligase buffer and 0.5 µl of ligase were added. This was then incubated either for one hour at RT or 16 °C overnight.

2.1.4 Transformation of plasmid DNA

Ligation mix or plasmid DNA used for cloning was introduced into bacterial cells using electroporation. The cloning strain used was DH5α (F⁻ *endA1 hsdR17(rk⁻mk⁺) supE44 thi-1 λ⁻ recA1 gyrA96 relA1 φ80ΔlacAm15*). Electrocompetent cells were prepared according to the standard lab protocol (Appendix). 1-2 µl of DNA was added to 50 µl of cells on ice, transferred to the electroporation cuvette and subjected to a 2.4 kV pulse. 800 µl of LB medium were immediately added and the cells were regenerated for 45 min in a shaker at 37 °C. 200 µl of the cell suspension was then plated on a selective agar and incubated overnight at 37 °C.

2.1.5 DNA Sequencing

Sequencing of DNA was performed by Synergene biotech. 15 µl of a DNA miniprep were sent by mail. The following sequencing primers were used:

for pET derived vectors:

T7 (TAATACGACTCACTATAGGG)

T7term (GCTAGTTATTGCTCAGCGGT)

for TOPO cloning:

M13(-20)f (GTAAAACGACGGCCAGT)

M13rev (GGAAACAGCTATGACCATG)

for BlM18DyPEncapsulin:

self designed internal primer (GGGGGCAAGCTCGCGATAGAGG)

2.1.6 Electrocompetent cells preparation protocol

1. Inoculate 20 ml LB-Medium(low salt) o/n
2. Inoculate 1 L LB-Medium(low salt) in 5 L Erlenmeyer flasks with max. 10 ml of a fresh o/n culture (1/100)
3. Grow cells at 37 °C with vigorous shaking (150 rpm) to an OD600 of 0.6-0.8
4. To harvest the cells, chill the flask on ice for 15-30 min, and centrifuge in a cold rotor at 3000 g/12 min/4 °C
5. Remove as much of the supernatant as possible. Resuspend pellets in a total of 1 L of cold water. Centrifuge in a cold rotor at 3000 g/12 min/4 °C
6. Resuspend in 0.5 L of cold water. Centrifuge in a cold rotor at 3.000 g/12 min/4 °C
7. Resuspend in 20 ml of cold 10 % glycerol, and transfer into a 50 ml Greiner tube. Centrifuge in a tabletop at 3000 g/12 min/4 °C
8. Resuspend in a final volume of 2 ml cold 10 % glycerol. The cell concentration should be about 10^{10} cells/ml.
9. Aliquot the suspension in 50 µl Eppendorf tubes on dry ice and store them at -80 °C. The cells are good for at least 6 month under these conditions.

2.2 Protein methods

2.2.1 Protein analysis

Denaturing SDS-PAGE

Protein solution to be analyzed was treated with 6xSDS sample buffer (0.28 M Tris, 30 % glycerol, 1 % SDS (w/v), 0.5 M DTT, 0.0012 % bromphenol blue), boiled at 95 °C for 2-5 min and immediately loaded on a gel. Proteins were typically analyzed on 15 % SDS-PAGE gels (Laemmli 1970), running for 50 min with limiting voltage (210 V) and current (35 mA) in SDS electrophoresis buffer (0.25 M Tris base, 0.19 M glycine, 0.1 % SDS (w/v), pH 8.3).

15 % SDS-PAGE, recipe for 8 gels:

stacking gel:

      5 ml UGB

      12 ml ddH$_2$O

      3.2 ml 30 % Acrylamide

      40 µl 25 % APS

      40 µl TEMED

running gel:

      10.7 ml LGB

      8 ml ddH$_2$O

      20 ml 30 % Acrylamide

      80 µl 25 % APS

      80 µl TEMED

Gel staining – Coomassie Blue

Staining was usually done by incubation in Coomassie Blue staining solution (50 % ethanol, 0.05 % Coomassie brilliant blue R-250, 10 % acetic acid) for 1 hour, followed by destaining (50 % ethanol, 10 % acetic acid).

Gel Staining – Silver stain

To achieve higher sensitivity gels were stained with Biorad silver stain (Switzer, Merril et al. 1979). After fixation in 40 % methanol / 10 % acetic acid for 30 min the gel was immersed in oxidizer solution for 5 min. The gel was then washed intensively for up to 15 min until the

yellow color disappeared. After 20 min in the silver reagent the gel was rinsed with water and then developer solution was added. As soon as the solution turned brown it was replaced with fresh developer solution. When the staining of the bands was finished the gel was stored in 5 % acetic acid which stopped the reaction.

Molecular weight markers

Where not mentioned otherwise, the following molecular weight marker was used for denaturing SDS-PAGE gels (Figure 8)

MW, kDa	Protein	Source
116.0	β–galactosidase	E.coli
66.2	Bovine serum albumin	bovine plasma
45.0	Ovalbumin	chicken egg white
35.0	Lactate dehydrogenase	porcine muscle
25.0	REase Bsp98I	E.coli
18.4	β–lactoglobulin	bovine milk
14.4	Lysozyme	chicken egg white
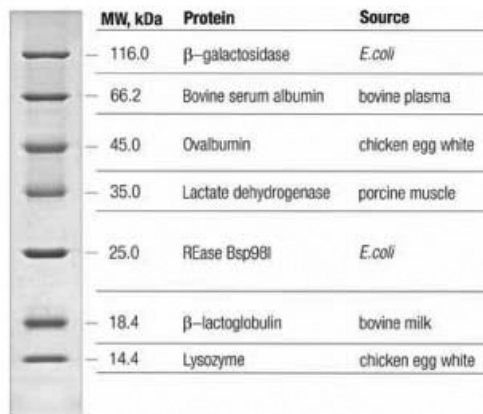
Figure 8: Fermentas Unstained Molecular Weight Marker (#SM0431)

2.2.2 Mass spectrometric analysis of proteins

Sample to be analyzed for mass spectrometry was given to the Protein Service Laboratory at Irchel where it was crystallized in a sinapinic acid matrix and analyzed in a MALDI-TOF spectrometer.

2.2.3 Protein expression

Expression in was done in Rosetta (DE3) cells (F$^-$ *ompT hsdS$_B$ (r$_B$$^-m_B$$^-$) gal dcm lacY1* (DE3) pRARE(Cm^R)). Genes from other organisms may have a slightly different usage of codons and a series of codons which are rare in *E. coli* could stall translation. This problem can be circumvented by using Rosetta cells, which contain the pRARE plasmid which codes for several tRNAs which are represented in lower amounts in *E. coli*.

2.2.4 Purification of *T. maritima* encapsulin from native source

Thermotoga maritima cells were obtained from Prof. K. O. Stetter. 20 g of cell pellet were resuspended in 100 ml buffer (10 mM Tris pH 7.4, 100 mM NH$_4$Cl, 10 mM MgCl$_2$, 6 mM β-me). DNaseI was added to a final concentration of 50 µg/ml before cracking the cells with an

Avestin EmulsiFlex-C5. The cell lysate was then cleared by centrifugation (2x25 min at 25 krpm in a Ti 70 rotor).

RNaseA was added (50 µg/ml) and the sample was incubated at RT for 2 h. 15 ml of cleared lysate were layered on top of a 16 ml sucrose cushion solution (38 % sucrose in 20 mM Tris pH 7.4, 500 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-me). The protein was then pelleted by centrifugation through the cushion (21 h at 31 krpm / 4 °C in a Ti 70 rotor). This resulted in firm lower and a soft upper pellet which was removed. After resuspension of the pellet in 700 µl buffer (50 mM Tris pH 7.4, 100 mM NH_4Cl, 0.3 mM $MgCl_2$, 6 mM β-me) on a shaker, the insoluble parts were removed with a short centrifugation step.

The resuspended solution was then loaded on a 10-50 % sucrose gradient (50 mM Tris pH 7.4, 500 mM NH_4Cl, 2 mM $MgCl_2$, 6 mM β-me) in SW-32 tubes and centrifuged for 17 h at 24 krpm. The gradient fractions were then collected and analyzed using SDS-PAGE. The fractions containing encapsulin were pooled and concentrated / buffer exchanged to buffer A (20 mM Tris pH 7.4, 150 mM NH_4Cl, 20 mM $MgCl_2$, 6 mM β-me) using Vivaspin 20 ml centrifugal concentrators. The resulting protein was then used for crystallization assays.

2.2.5 Purification of *B. linens* encapsulin from culture supernatant

Brevibacterium linens M18 was obtained from the „Deutsche Sammlung von Mikro-organismen und Zellkulturen GmbH" (DSM No. 20425). The cells were grown on PCB agar plates at room temperature and stored at 4 °C for up to one month. The culture was distinguishable from possible contaminations by its reddish color.

A 100 ml preculture was inoculated with several colonies from plate and grown O/N at 30 °C. The preculture was used to inoculate 9 L of culture medium in 6x5 L flasks and the cells were grown on a shaker at 30 °C for five days. The cells were then removed by centrifugation (SLC-6000 rotor, 7 krpm, 20 min) and the supernatant cooled to 4 °C. The supernatant was concentrated to a volume which was manageable for high speed centrifugation with a combination of a high flow peristaltic pump and a tangential flow membrane concentrator (Figure 9) with a cutoff of 30 kDa. This step removed small molecules and water from the supernatant and it could thereby be concentrated from 9-18 L to 100 ml.
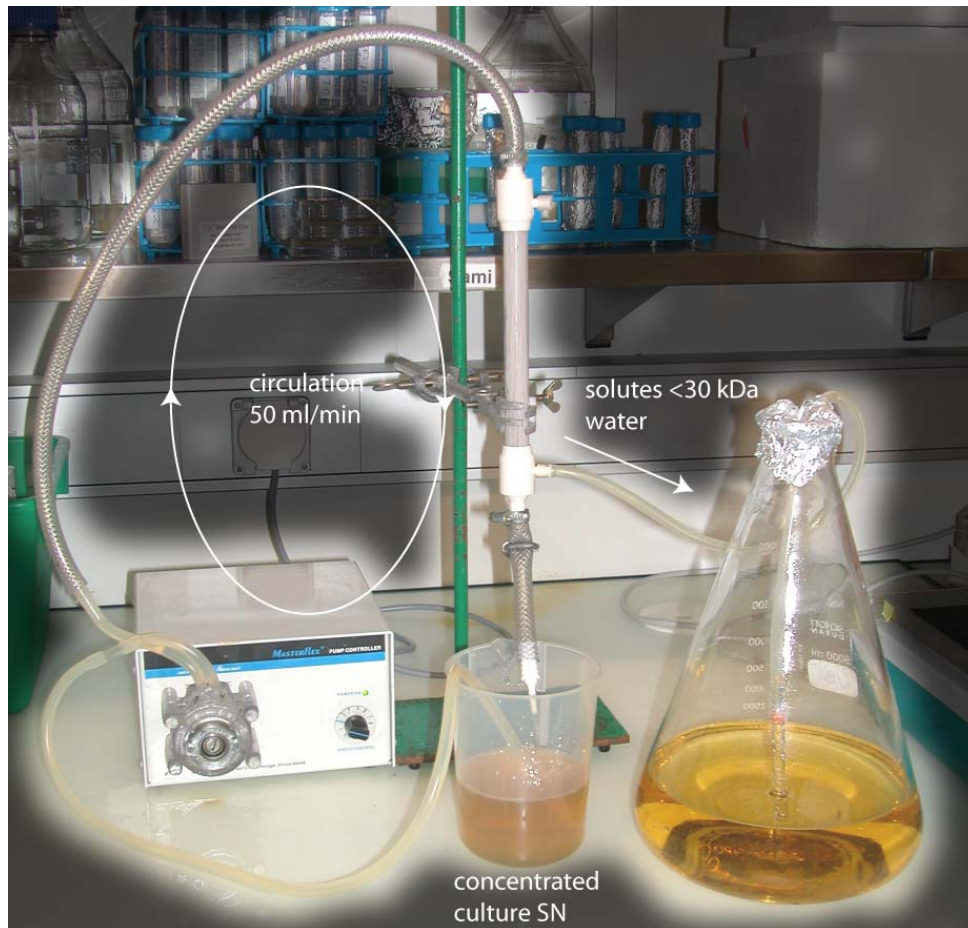
Figure 9: Concentration of culture supernatant. A high flow peristaltic pump is circulating the culture supernatant through the concentrator and the low molecular weight permeate is collected in the 5 L flask. Depicted is the final stage of the concentration step where only about 200 ml retentate is left.

In order to remove unwanted DNA and RNA the sample was then treated with 50 μg/ml DNAseI and 50 μg/ml RNAseA at RT for 3-4 h. Any remaining insoluble parts were removed with a 20 min Ti-70 spin (25 krpm, 4 °C). The cleared supernatant was then layered (16 ml per tube) on top of a 15 ml sucrose cushion (38 % sucrose in buffer A) and centrifuged for 21 h at 50 krpm in a Ti-70. The supernatant was discarded, the pellet washed with 500 μl 10 mM Tris pH 8.0 and then resuspended on a shaker in 700 μl buffer A. 0.5-1.0 ml of sample (per tube) was then applied on a 10-50 % sucrose gradient (in buffer A) and spun for 17 h at 28 krpm in a SW-28 rotor. Gradient fractions were then collected and the sucrose removed by concentration / buffer exchange. This is a modified protocol from the thesis of N. Valdés-Stauber (Stauber 1995).

2.2.6 Purification of recombinant *B. linens* encapsulin

Plasmid expressing the desired construct was transformed by electroporation into *E. coli* Rosetta (DE3) cells. After 45 min recovery 200 µl of the transformed cells were plated on a LB$_{AmpCam}$ plate and incubated O/N at 37 °C. The colonies from 3 plates were washed with 15 ml LB$_{AmpCam}$ which were then diluted to 100 ml and incubated for 1 h at 37 °C. 15 ml of this preculture was diluted into 1.5 L LB$_{AmpCam}$ and then grown at 37 °C until an OD$_{600}$ of 0.6 to 0.8. At this point the cells were cooled to 22 °C and then IPTG was added to a final concentration of 0.1 mM.

The cells were incubated O/N, harvested (SLC-6000 rotor, 7 krpm, 15 min) and resuspended in 100-300 ml buffer A. After addition of 50 µg/ml DNAseI the cells were passed through the cell cracker 2 times and the cell debris was removed with a SLA-1500 spin (12 krpm, 30 min). The remaining supernatant was treated with 50 µg/ml RNAseA and incubated at room temperature for 2 h. Any remaining insoluble parts were removed with a 20 min Ti-70 spin (25 krpm, 4 °C). The cleared supernatant was then layered (16ml per tube) on top of a 15 ml sucrose cushion (38 % sucrose in buffer A) and centrifuged for 17 h at 48 krpm in a Ti-70 rotor. The supernatant was discarded, the pellet washed with 500 µl FB and then resuspended on a shaker in 700 µl FB. Alternatively, the sample was centrifuged for 17 h at 40 krpm in a Ti-70 rotor and then the last 2 ml above the pellet were taken and buffer exchanged/concentrated to lower sucrose concentration. This step removes the ribosomes which are found in the pellet. 0.5-1.0 ml (per tube) of sample was then applied on a 10-50 % sucrose gradient (in 1x buffer A) and spun for 17 h at 28 krpm in a SW-28 rotor. Gradient fractions were then collected and the sucrose removed by concentration buffer exchange. Alternatively, the sample was applied on a MonoQ column and eluted in a gradient of 150 mM to 1000 mM NH$_4$Cl.

2.2.7 Purification of *B. linens* DyP

The soluble non-pelleting fraction from a co-expression of DyP with encapsulin was applied on a 5 ml FastFlow Q column and eluted with 300 mM NH$_4$Cl. The peak fractions were pooled, concentrated with Vivaspin concentrators and applied on a S-200 gelfiltration column. Peak fractions were pooled and concentrated which yielded >95 % pure protein.

2.2.8 Peroxidase activity of DyP

Peroxidase activity of DyP and DyP packaged in encapsulin were measured with guaiacol as substrate. The reaction was performed at 22 °C and contained 100 mM citrate buffer pH 4, 20 mM H_2O_2, 20 mM guaiacol and 3-4 µM DyP in a volume of 500 µl and was followed by absorption measurement at 470 nm.

2.2.9 *In vivo* assay of bactericidal activity against *Listeria* strains

100 µl of O/N preculture (0.5x BHI) of *Listeria ivanovii* (WSLC 3061) were added to 8 ml 1xBHI agar. Holes were cut out with a blue pipette tip and 10 µl of test solution were added. Alternatively 5-10 µl of the test solution was pipetted on top of the agar. The plates were then incubated at 30 °C for 10-12 h and then visually inspected for plaque formation. Activity was quantified by measuring the maximum possible dilution where inhibition was still observed.

2.3 Crystallization methods

2.3.1 Screening conditions

TM encapsulin sample used for crystallization was at concentrations of 5-12 mg/ml in buffer A (20 mM Tris pH 7.4, 20.0 mM $MgCl_2$, 150.0 mM NH_4Cl). Vapor diffusion technique was used in Chryschem 24-well sitting drop plates from Hampton Research. 750 μl of reservoir solution was added at room temperature, then 1-4 μl of the reservoir solution was pipetted to an equal amount of sample solution. Plates were sealed with "Manco Duck Crystal Clear" tape (Henkel) and incubated 19 °C. The additive screen used was designed by Marc Leibundgut (Leibundgut 2005). 70 μl of additive solution was added to 700 μl of reservoir prior to mixing with the protein solution.

Screen name	Description
Hampton Crystal Screen 1 & 2	sparse matrix screen, 96 conditions total (Jancarik, Scott et al. 1991)
pHvsMPD003	0.2 M NH_4OAc 0.1 M citrate pH 5.2-5.8 (0.2) 25-35 % MPD (2 %)
pHvsMPD004	0.0, 0.5 M NH_4OAc 0.1 M citrate pH 5.2-5.8 (0.2) 27, 31, 35 % MPD
pHvsMPD005	0.2 M NH_4OAc 0.1 M MES pH 5.2-5.8 (0.2) 25-35 % MPD (2 %)
pHvsMPD006	0.2 M NH_4OAc 0.1 M citrate pH 5.4-6.0 (0.2) 18-28 % MPD (2 %)
pHvsMPD009	0.2 M NH_4OAc 0.1 M citrate pH 4.6-5.8 (0.4) or 0.1 M HEPES pH 6.9 or 0.1 M Tris pH 7.4, 8.0, 8.5 7-40 % MPD (4 %)
pHvsMPD012	0.2 M NH_4OAc 0.1 M citrate pH 5.4-5.7 (0.1) 17-22 % MPD (1 %)
pHvsMPD013	0.2 M NH_4OAc 0.1 M citrate pH 5.3-6.0 (0.1) 23-34 % MPD (1 %)
pHvsMPD014	0.2 M NH_4OAc 0.1 M citrate pH 4.9-6.0 (0.1) 21.5-25 % MPD (0.5 %)
pHvsMPD015	0.05-0.35 M NH_4OAc (0.1 M) 0.1 M citrate pH 4.9-5.2 (0.1) 21-23.5 % MPD (0.5 %)

pHvsPEG001	0.2 M NH$_4$OAc 0.1 M citrate pH 5.2, 5.4 or 0.1 M MES pH 6.2 or 0.1 M Tris pH 7.4 2,5,9,14,20,27 % PEG3350
pHvs2prop001	0.2 M Na-citrate 0.1 M HEPES pH 7.2-7.8 (0.2) 16-26 % 2-propanol (2 %)

Table 1: Crystallization screens used in this study

2.3.2 Crystal stabilization

Various cryoprotecting agents were tested for stabilization (MPD, glycerol, ethylene glycol, PEG 550MME, PEG 3350). The following protocol was used to increase MPD concentration in the drop stepwise.

Reservoir solution (R): 21 % MPD, 0.2 M NH$_4$OAc, 0.1 M citrate pH 5.2

Stabilization solution 1 (S1): 35 % MPD, 0.2 M NH$_4$OAc, 0.1 M citrate pH 5.2

Stabilization solution 2 (S2): 50 % MPD, 0.2 M NH$_4$OAc, 0.1 M citrate pH 5.2

Time	drop size	Change	MPD concentration
0 min	4 µl		21 %
	12 µl	+8 µl R	
	16 µl	+4 µl S1	26 %
30 min	10 µl	-6 µl	
	16 µl	+6 µl S1	31 %
75 min	10 µl	-6 µl	
	16 µl	+6 µl S2	40 %
135 min	10 µl	-6 µl	
	16 µl	+6 µl S2	45 %

Table 2: Stabilization protocol of TM encapsulin crystals

Crystals were fished with a nylon loop of 0.1-0.2 µm diameter attached to the tip of a Hampton pin, flash frozen in liquid propane and stored in liquid nitrogen until use at synchrotron.

2.4 Electron microscopy methods

2.4.1 EM sample preparation

EM samples were prepared by conventional negative staining (Ohi, Li et al. 2004). Prior to sample application the grids were exposed to oxygen plasma to remove organic molecules. Sample for electron microscopic analysis was adjusted to suitable concentration (approx. 50 nM) by dilution with either ddH$_2$O or low concentrations of buffer (10 mM). 5 µl of sample was applied to a grid and removed after 1-2 min with filter paper. The grid was then applied on a 15 µl drop of a 2 % (w/v) solution of uranyl acetate. This procedure was repeated three times with filter paper drying steps in between.

2.4.2 EM Imaging

EM grids of *T. maritima* encapsulin were analyzed on a FEI Morgagni 268 transmission electron microscope. The voltage used was 100/200 kV and usual magnifications were between 32'000 and 100'000. Images were taken with the CCD camera with a resolution of 2048x2048 pixels and stored as 24 bit TIFF files. Size calibration of the resolution was done with a TMV (tobacco mosaic virus) sample.

EM samples of the *B. linens* DyP/encapsulin and DyP were prepared by conventional negative staining (Ohi, Li et al. 2004). Briefly, a solution of *B. linens* DyP/encapsulin or DyP was adsorbed to a thin carbon film and stained with 2 % (w/v) uranyl acetate solution for 2 minutes. For the general views of the *B. linens* DyP/encapsulin and *B. linens* DyP images were obtained at 40,000x magnification on a Fei Morgagni 268 electron microscope (FEI, Hillsboro, OR) and 60,000x magnification on a Fei F20 electron microscope (FEI, Hillsboro, OR), respectively, with CCD cameras. For 3D reconstruction of the *B. linens* DyP images were taken at 100,000x magnification under low dose conditions at 0.7-1.5 µm defoci with a Fei F20 electron microscope (FEI, Hillsboro, OR) equipped with a Tridiem Gatan energy filter at 200 kV on a Gatan UltraScan 1000 FT CCD camera (Gatan, Warrendale, PA) (14 µm pixel size). Image series were obtained using the TOM software toolbox(Nickell, Forster et al. 2005). A total of 2878 single particle images were CTF corrected (Sander, Golas et al. 2003), computationally coarsed to a final pixel size of 2.8 Å and used for image processing with the Imagic-5 software (van Heel, Harauz et al. 1996). After a "reference-free" alignment procedure (Dube, Tavares et al. 1993), images were subjected to a multivariate statistical analysis (van Heel and Frank 1981) and classification. The class averages were used for 3D

structure determination by the angular reconstitution approach (van Heel 1987). In the refinement cycles D3 point group symmetry was applied to the three-dimensional reconstruction. The resolution of the structure calculated by the Fourier shell correlation (FSC) function using the 3 σ criterion and 0.5 criterion was found to be 16 Å and 18 Å, respectively. The 3 σ FSC threshold curve was multiplied by square root 6 to account for the six fold redundancy of the D3 point group symmetry (Orlova, Dube et al. 1997). A homology model of the *B. linens* DyP was obtained using the Structure of *B. thetaiotaomicron* DyP (pdb: 2gvk) (Zubieta, Krishna et al. 2007) in Modeller (Sali and Blundell 1993). The homology model of the monomer was automatically placed into the EM density with Situs (Wriggers, Milligan et al. 1999) (cross correlation coefficient: 0.58). A model of the hexamer was obtained applying the D3 symmetry of the EM density to the monomer.

2.5 Computational methods

The program packages and publicly available scripts that have been used in this work for x-ray data processing, structure calculation, refinement of the structure, electron microscopy calculations and graphical display of the molecules are listed below. Scripts which were customized to a high degree are found in the appendix.

CCP4 package:

(Collaborative Computational Project 1994)

xray data handling	import_scaled, scaleit, cad, sfall
xray calculations	patterson, phistats, sigmaa
pdb file modifications	pdbset, superpose (Krissinel and Henrick 2004)
map calculations	fft
density modification	dm (Cowtan and Main 1993)
map / mask generation	ncsmask, maputils
refinement	refmac5 (Murshudov, Vagin et al. 1997)
pdb analysis	check

Uppsala software factory package:

pdb superposition	lsqman (Kleywegt and Jones 1995)
real space averaging	ave (Kleywegt and Read 1997)
map / mask modification	mama, mapman (Kleywegt and Jones 1999)

HKL package:

(Otwinowski and Minor 1997)

image display	xdisp
header analysis	readheader
integration	denzo_3d
scaling	scalepack

EM programs:

image acquisition	TOM (Nickell, Forster et al. 2005)
image processing	Imagic-5 (van Heel, Harauz et al. 1996)

modelling   Situs (Wriggers, Milligan et al. 1999)

                                        Modeller (Sali and Blundell 1993)

Standalone programs:

xray data integration / scaling   XDS (Kabsch 1993)

secondary structure calculation   dssp (Kabsch and Sander 1983)

self rotation function calculation   glrf (Tong and Rossmann 1997)

surface charge calculation   delphi (Rocchia, Sridharan et al. 2002)

model building   ONO (Jones, Zou et al. 1991)

model building   coot (Emsley and Cowtan 2004)

electron microscopy   EMAN (Tang, Peng et al. 2007)

image generation   povray (http://www.povray.org/)

pdb visualization   pymol (DeLano 2002)

sequence alignment   ClustalX 1.81 (Thompson, Gibson et al. 1997)

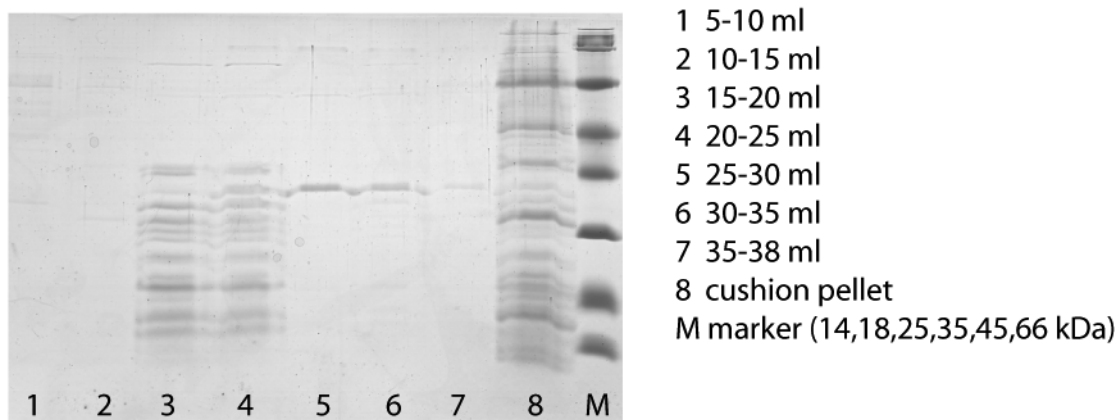refinement   phenix.refine (Afonine, Grosse-Kunstleve et al. 2005)

3 RESULTS

3.1 Purification and structural analysis of *T. maritima* encapsulin

3.1.1 Purification of native *T. maritima* encapsulin

Thermotoga maritima encapsulin was purified from native source using only centrifugation steps. After cell lysis centrifugation through a sucrose cushion removes the small cellular proteins (all proteins <200-300 kDa do not penetrate the sucrose cushion under the conditions used). In the following sucrose gradient centrifugation the large cellular complexes are further separated by size. The resulting protein shows a high purity in view of simple methods used for this preparation (Figure 10).

1 5-10 ml
2 10-15 ml
3 15-20 ml
4 20-25 ml
5 25-30 ml
6 30-35 ml
7 35-38 ml
8 cushion pellet
M marker (14,18,25,35,45,66 kDa)

Figure 10: SDS-PAGE of sucrose gradient fractions of native TM encapsulin purification. Fractions containing TM encapsulin (5+6) were pooled and the remaining sucrose was removed by buffer exchange. Ribosome is still present in the sample despite the RNase digest as seen by the ribosomal protein bands in lanes 3 and 4. Compared to the pellet shown in lane 8 the sucrose gradient fraction shows a high degree of purification.

3.1.2 EM analysis of *T. maritima* encapsulin

A purified sample of encapsulin was negative stained with uranyl acetate and analyzed in a transmission electron microscope (Figure 11A).
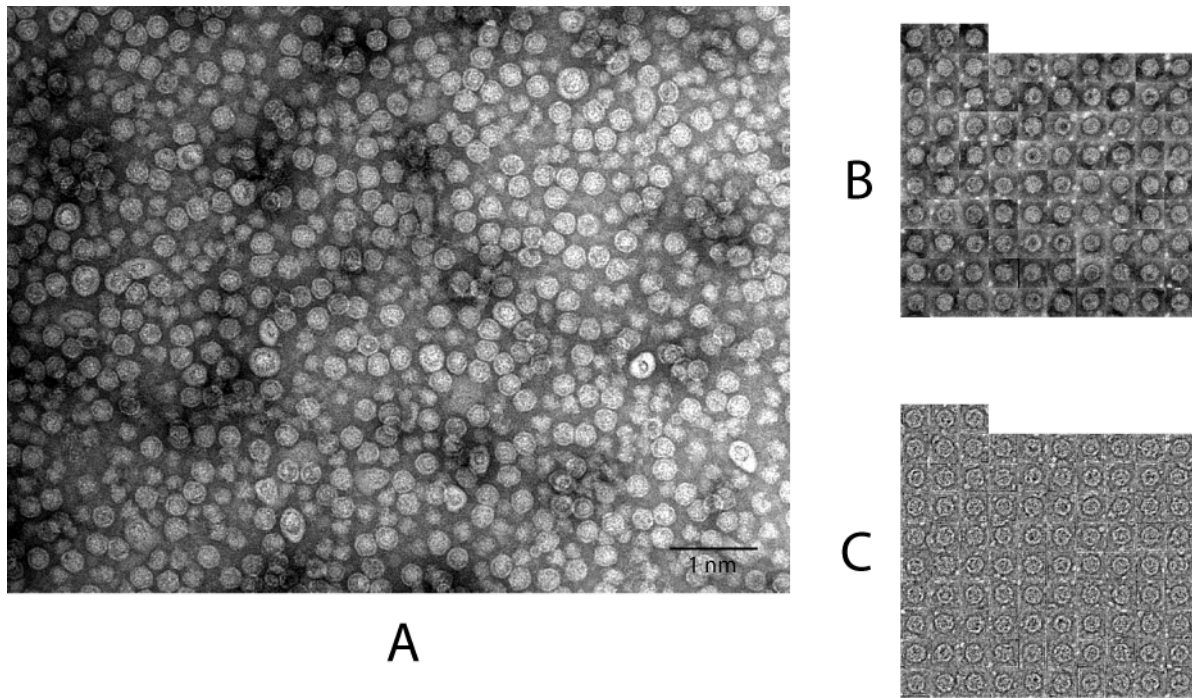
Figure 11: Negative stain image of native TM encapsulin. A: 32'000x magnified uranyl acetate stained sample of TM encapsulin. The majority of particles have the same size and the icosahedral symmetry is clearly visible. B: picked particles C: CTF corrected particles from B

From samples with a higher dilution it was possible to pick single particles for closer analysis (Figure 11B). The selected particles were corrected for resolution dependent distortion using the contrast transfer function (CTF) (Figure 11C) and an automated procedure was applied to classify the images with imposed icosahedral symmetry. An initial model was generated and the particles were assigned to the projections of this model. The reconstructed class averages were then used to generate a new model (Figure 12). The resolution was estimated to be around 50 Å and the information about the diameter and thickness of the shell proved to be very useful for experimental phasing of encapsulin crystals.

26

A   B
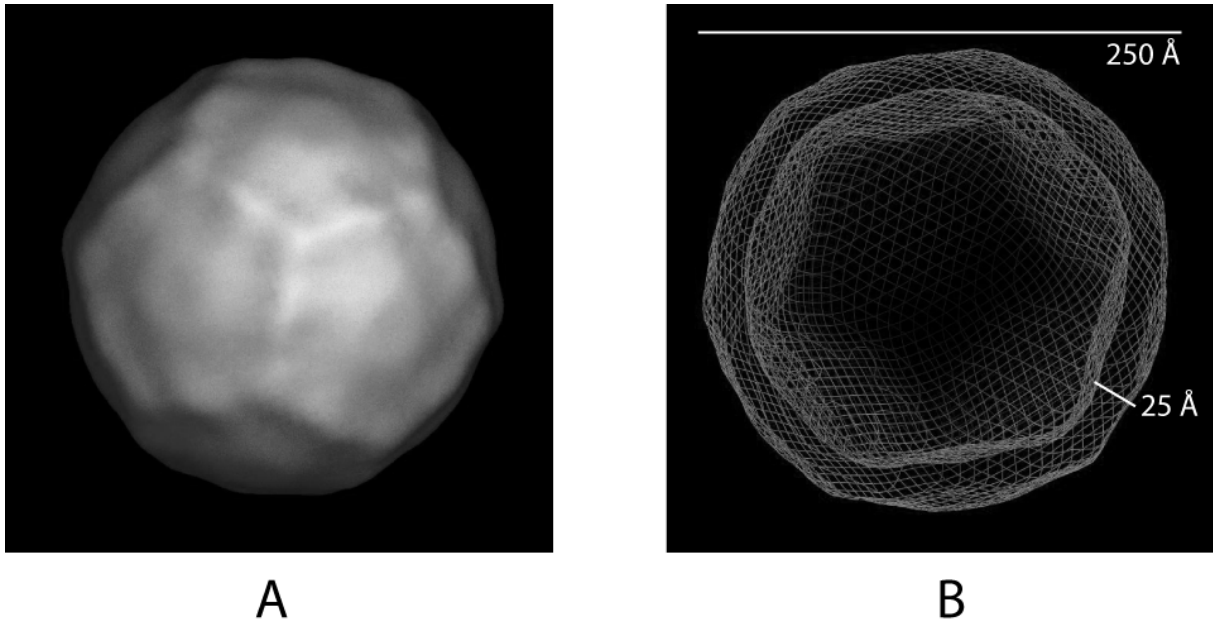
Figure 12: Calculated EM model. A: Surface rendering of the reconstructed model. The shell is quite smooth with only an elevation around the five-fold and along the pentamer. B: Mesh view cross-section with indicated approximate diameter and thickness.

3.1.3 Crystallization of *Thermotoga maritima* encapsulin

Initial encapsulin crystals were observed when a preparation of *Thermotoga maritima* 50S ribosomes were subjected to crystallization experiments. The ribosomes were purified with a sequence of sucrose cushion and gradient followed by dialysis. Hexagonal crystals were found in a reservoir condition of 30 % MPD 0.1 M sodium citrate pH 5.6 and 0.2 M ammonium acetate (Gutmann 2001).
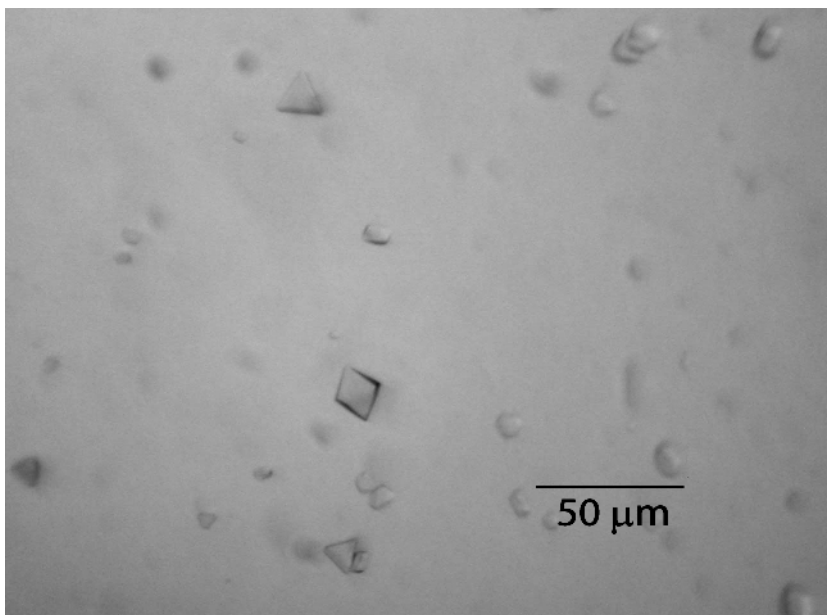
Figure 13: Initial crystals of native TM encapsulin. These crystals were obtained after optimizing the purification protocol to remove ribosomes by RNaseA digest. The conditions

are 0.1 M citrate pH 5.2, 25 % MPD and 200 mM ammonium acetate at a protein concentration of about 2 mg/ml.

Optimization of the crystals could proceed when the purification protocol was optimized to remove the ribosome contamination (Figure 13). Small crystals appeared after only one day but in the drops which yielded the largest crystals it took up to a week until they were visible. The crystals were stable for several months as judged by visual appearance. They often grew on the bottom of the well, making it harder to fish them out but they were quite resistant to mechanical stress. Remarkably, these crystals could grow at very low sample concentrations (as low as 0.1 mg/ml).

Screen name	Comments
Hampton Crystal Screen 1 & 2	Needles in various conditions (1/34, 2/9, 1/30, 2/40) and small crystals in 1/27
pHvsMPD001	Needles to small crystals
pHvsMPD002	Needles to small crystals
pHvsMPD003	Clear drops at lower MPD conc.
pHvsMPD004	Precipitation at 0.5 M NH$_4$OAc conditions
pHvsMPD005	Precipitation at MPD > 30 %
pHvsMPD006	Clear drops at lower MPD conc.
pHvsMPD009	Precipitation at higher pH, Small crystals and needles
pHvsMPD012	Clear drops at lower MPD conc (<20 %)
pHvsMPD013	Crystals of up to 220 μm size (26 % MPD / pH 5.4)
pHvsMPD014	Large crystals at lower pH
pHvsMPD015	Best diffracting crystals from conditions with 0.25 M NH$_4$OAc and pH 5.0
pHvsPEG001	Small round crystals with precipitate in conditions with 9 % PEG3350
pHvs2prop001	Bipyramidal small crystals at 24 % / 7.4
Additive screen	Many drops with medium size crystals

Table 3: Analysis of crystallization screens

In the additive screen many drops with crystals were observed which was presumably due to the original condition being already good. With some additives the crystals were layered (mostly with alcohol additives) or showed a rough surface (some salt additives) but the

general crystal morphology was still the same. The best crystals were observed in a condition in screen pHvsMPD015 (0.1 M citrate pH 5.0, 21 % MPD, 0.25 M NH$_4$OAc) at a protein concentration of 12 mg/ml (Figure 14). The range in which large crystals grow is quite narrow and not exactly the same for every prep which is why a fine screen with many conditions is necessary. A 1 % difference in MPD concentration usually makes a difference between getting 5-10 large crystals or a hundreds of small crystals.
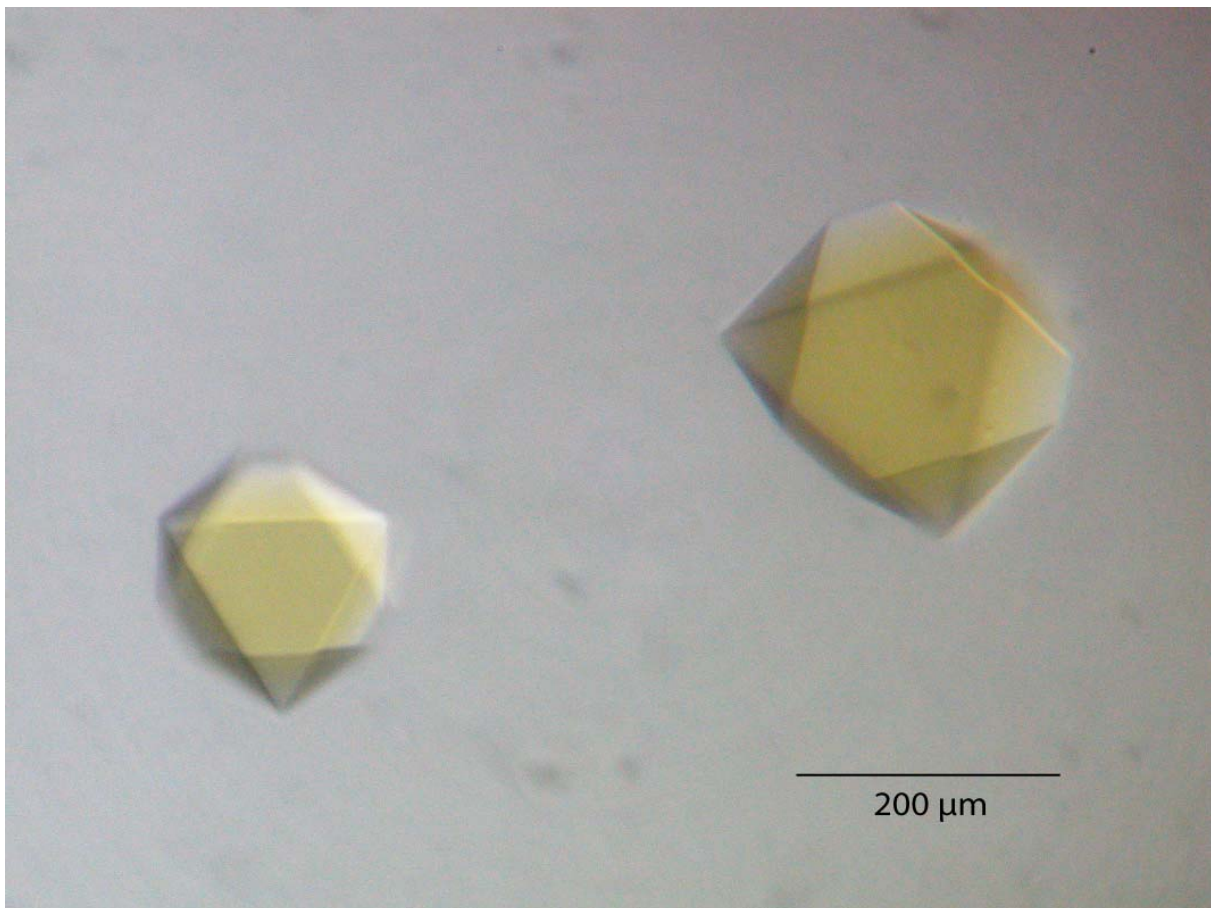
Figure 14: Improved crystals of native TM encapsulin. These bipyramidal crystals have dimensions of up to 250x250x100 µm.

MPD proved to be most suitable for cryoprotecting the crystals. Glycerol and ethylene glycol dissolved the crystals despite increasing MPD concentration at the same time while with PEG 550MME and PEG 3350 the crystals showed cracks or softened edges. Slow dehydration (punching a hole in the sealing tape) had a negative effect on crystal morphology (crack formation). Once stabilized (for the protocol see Table 2) the crystals were fairly rigid and could be easily fished from the drop.

3.1.4 Diffraction analysis

The frozen crystals were analyzed at the X06SA PX beamline of the SLS (Swiss Light Source) at the Paul Scherrer Institut (PSI) in Villigen. Diffraction data were recorded with a 165 mm marCCD detector. Most crystals diffracted to 4 Å or better but some were losing diffraction power more rapidly due to the harsh exposure conditions which were necessary to see high resolution diffraction.
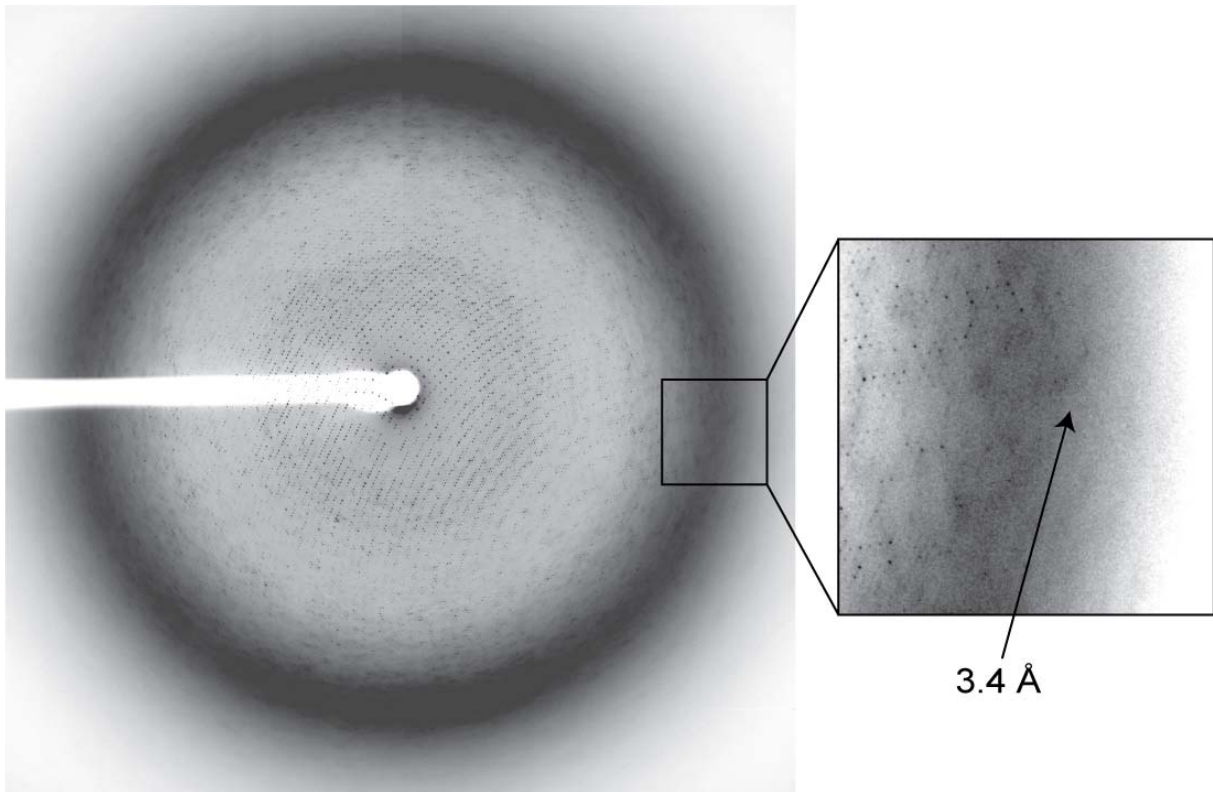
Figure 15: Diffraction pattern of TM encapsulin crystals (ms041_2). These crystals belong to the cubic space group $F4_132$ and have cell dimensions of 669.04 Å in each direction. While initial smaller crystals diffracted only to about 5 Å, these larger ones show visible diffraction up to 3.4 Å and have reasonable data statistics up to 3.1 Å.

Diffraction data were integrated and scaled with XDS (Kabsch 1993). A critical step was the determination of the correct beam center position as only a small offset could mean integrating the wrong lattice. During scaling a zero-dose correction was applied to account for radiation damage.

Space Group	F4₁32 (cubic face centered)
Unit cell [Å]	a = b = c = 669.04
	α = β = γ = 90 °
Resolution [Å]	50 – 3.1 (3.3-3.1)
Total observations	1'896'940
Unique reflections	220'157
Completeness	96.5 % (96.9 %)
Mosaicity	0.327
R_{merge}	20.7 % (62.7 %)
I/σ	7.82 (2.09)

Table 4: Data collection statistics of the crystal ms041_2. Data were recorded at SLS X06SA-PX, 90 images collected with 0.3 ° oscillation at a wavelength of 1 Å, processing done with XDS/XSCALE. Statistics for the highest resolution shell are in brackets.

From this data the Matthews coefficient (Matthews 1968; Kantardjieff and Rupp 2003) could be calculated to estimate the number of proteins in the asymmetric unit (asu). The Matthews coefficient V_M [Å³/Da] is calculated as follows:

$$V_M = \frac{V}{M_W \cdot Z \cdot N}$$

V = volume of the unit cell [Å³]

M_W = molecular weight of the monomer protein [Da]

N = number of molecules per asymmetric unit

Z = number of asu per unit cell / symmetry operators of the space group

Assuming a partial specific volume of 0.74 cm³/g for protein the solvent content can be derived from this. The most frequent value for protein crystals is 2.34 which equals about 47 % solvent content and is based on the analysis of 11'000 non-redundant protein structures (Kantardjieff and Rupp 2003). The normal range of this value is between 1.6 and 3.5 Å³/Da.

molecules/asu	V_M [Å³/Da]	Solvent content [%]
10	10.08	87.8
20	5.04	75.6
30	3.36	63.4
60	1.68	26.8

Table 5: Calculation of Matthews coefficient for different number of molecules/asu (assuming M_W = 30'800 Da, a = b = c = 669.04 Å, α = β = γ = 90 °, 96 asu/unit cell)

If the standard values would be valid for this case, a solvent content of 63 % would be reasonable but since this is presumably a large hollow spherical particle, solutions withhigher solvent content / lower number of molecules per asymmetric unit should still be considered.

3.1.5 Calculation and interpretation of the selfrotation function

In a self rotation function calculation the native Patterson is compared with itself, giving the rotational component of any non-crystallographic symmetry present in the crystal. In a calculation the rotation angle κ is fixed and the rotational space defined by Euler angles φ and φ is scanned for peaks. Self rotation functions were calculated for all angles likely to appear and the ones for κ=180,120 and 72 ° are shown in Figure 16.
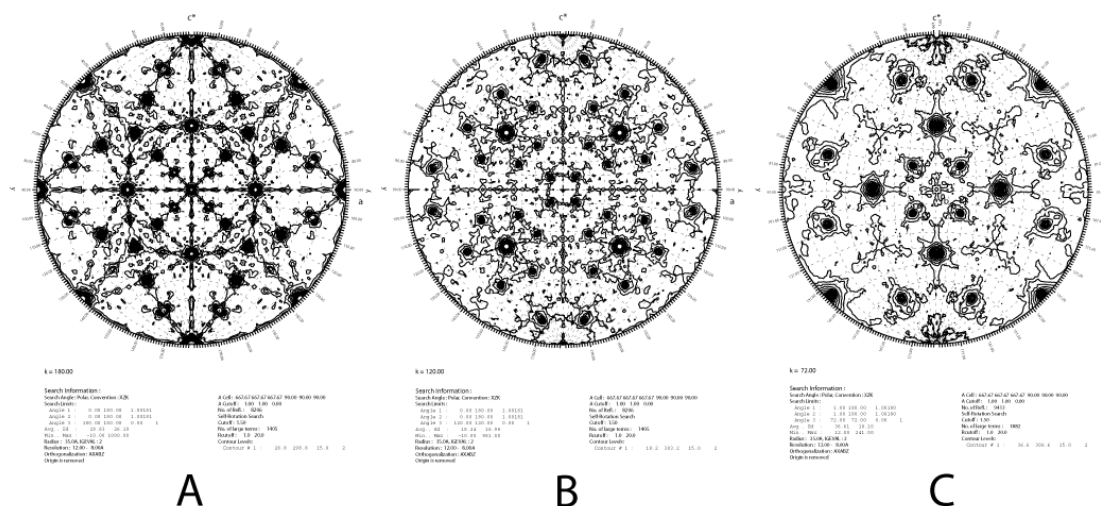
Figure 16: Selfrotation function of native Patterson. A, B, C: Selfrotation functions calculated with GLRF (Tong and Rossmann 1997), corresponding to 2,3 and 5-fold non-crystallographic symmetry. The data was used from 12-8.0 Å with a radius of 35 Å and unit cell parameters of a=b=c=669.04 Å, α=β=γ=90 °. The fourfold in the center is generated by the space group (F4₁32) as well as a threefold at φ = 45 °, φ=55 °.

The presence of 2, 3 and 5-fold non-crystallographic symmetry indicates a 532 icosahedral symmetry which is common in structural for virus particles. The smallest possible number of molecules to form such a particle is 60 corresponding to a triangulation number of T = 1. Geometrical arrangement shows that it is only possible to form icosahedrons which follow this formula (Caspar and Klug 1962):

$$T = H^2 + HK + K^2$$

where H and K are integers (including 0). In a T = 1 icosahedral structure all subunits are part of a pentamer, whereas in T = 3 and higher some subunits form hexamers.
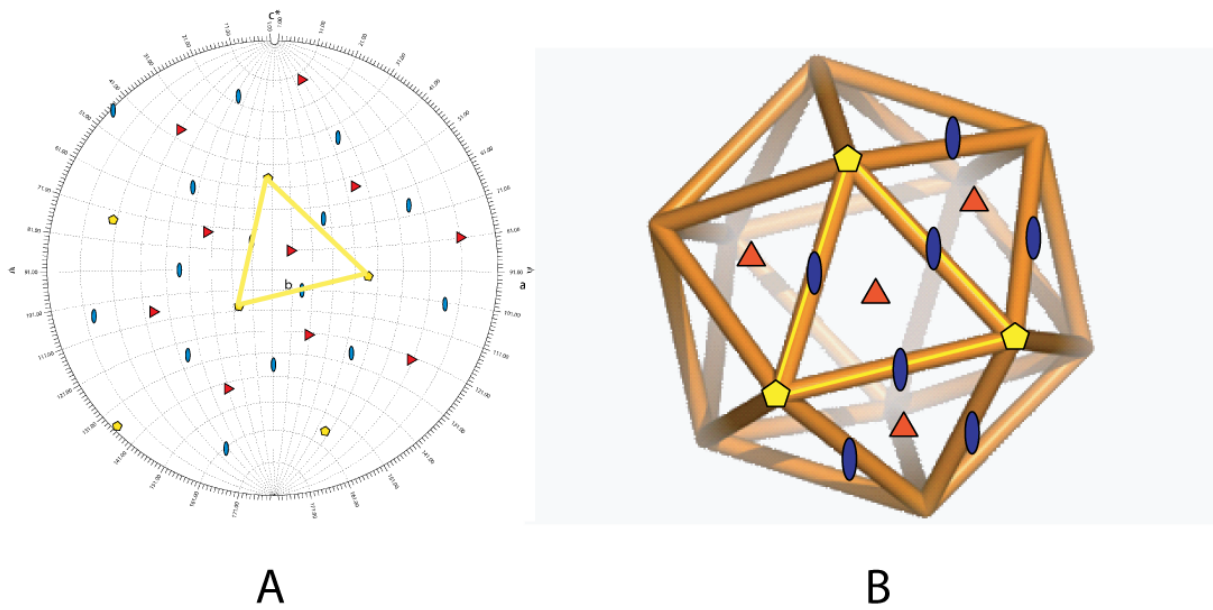
Figure 17: Assignment of unique selfrotation peaks to icosahedral symmetry. A: The peaks substituted by symbols (blue = 2-fold, red = 3-fold, yellow = 5-fold) and all the peaks generated by crystallographic symmetry were removed. B: The yellow triangle in A corresponds to the front face of this icosahedron with a few symmetry axis nearby indicated.

The peaks from the 2,3 and 5-fold sections of the selfrotation function can be mapped to a T=1 icosahedral symmetry (Figure 17). This means that 60 protein subunits of 30.8 kDa form one spherical particle. This is in agreement with the dimensions obtained from electron microscopy (approximate diameter 250 Å).

Usually it is not possible to derive the position of molecules from self rotation data alone since only the direction of the symmetry axis is known. The space group $F4_132$ is of a very high symmetry, containing 96 asymmetric units per cell. If the whole particle was present in one asymmetric unit the whole unit cell would have very large dimensions so it is clear that some of the particle symmetry axis overlap with space group symmetry. Most of the symmetry operators are screw axis (Figure 18B) which cannot be intramolecular. Of the few axis which are left only one threefold and one twofold have the right orientation and a point of intersection (Figure 18). Packing of the particles is possible if one shell is placed with its center at (1/8,1/8,1/8), assuming dimensions known from EM (Figure 12). The coincidence of particle and space group symmetry reduces the number of monomers in the asymmetric unit to 10. The positioning of these two axis narrows the positioning of the particle to two solutions which are related to each other by a rotation around the crystallographic threefold by 180 °. All the following calculations were done considering both possibilities.
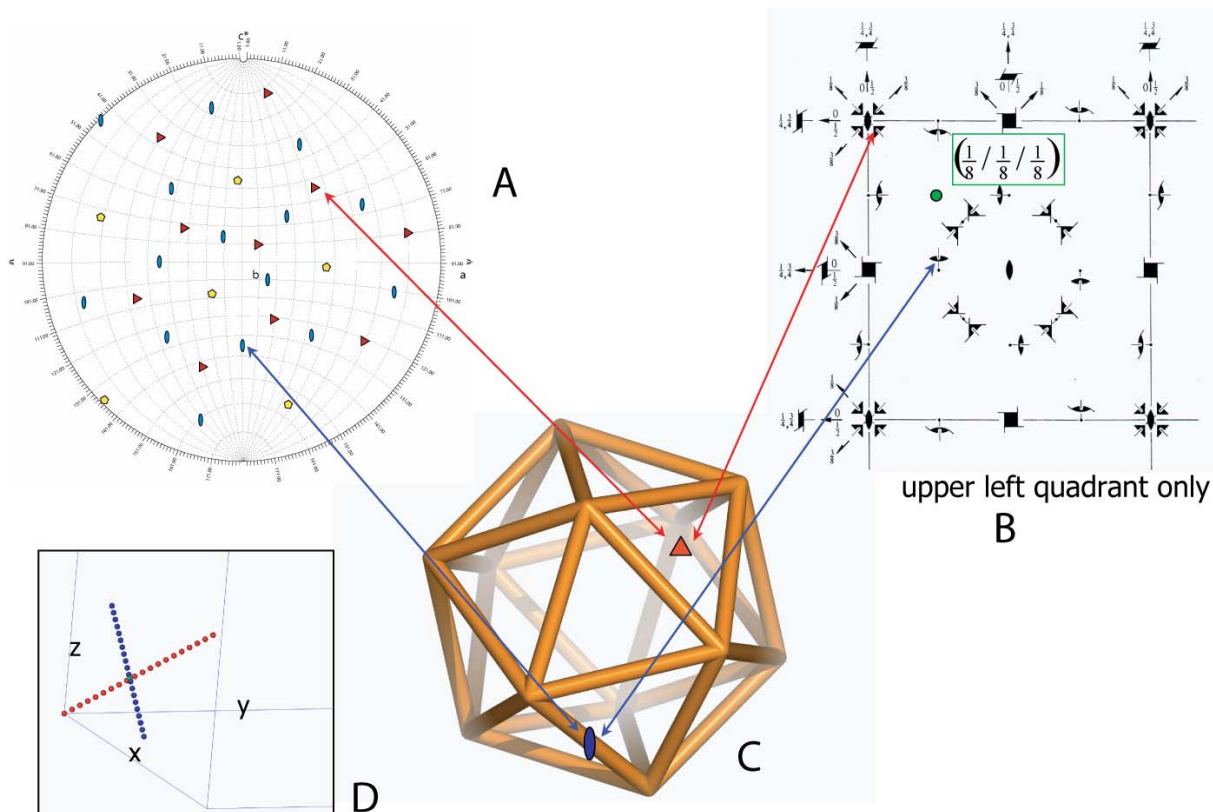
Figure 18: Determination of particle origin from selfrotation peaks. A: Icosahedral self-rotation peaks. B: F4$_1$32 space group symmetries (Hahn 1983), only upper left quadrant shown. The center of the particle is shown as a green dot at (1/8,1/8,1/8). C: Standard icosahedron; the perpendicular crystallographic axis are indicated. D: Depiction of the two- and three-fold axis in 3-dimensional xyz space.

3.1.6 Mask creation and symmetry averaging

The critical points when creating a mask for averaging are that there is neither overlap nor space in between the masks. The shape is actually not so important since it does not need to cover exactly one molecule. Therefore a slice is computationally the simplest solution. The mask was created with the following procedure:

- creating a grid of points in the standard orientation of an icosahedron with defined grid spacing and inner/outer diameter (see Appendix: 6.3.1 Calculation of monomer grid)
- applying fivefold and twofold operators to generate the 10-mer
- rotation and translation to the correct location; applying the crystallographic operators to the 10-mer creates the whole shell
- creating a mask around the grid; masks are harder to manipulate which is why this is done in the last step

Running all those calculations with scripts enabled the simple testing of various parameters, especially of the diameter of the shell.
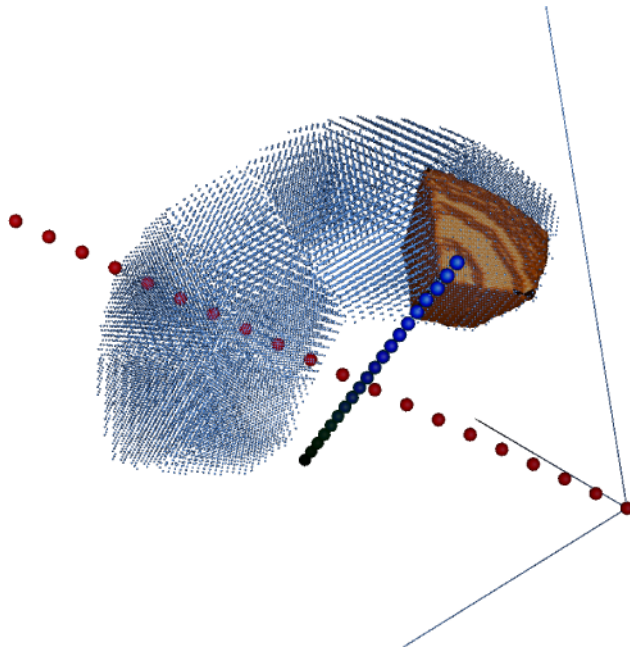
Figure 19: Calculation of 10-mer and monomer mask.
The whole 10-mer in the asymmetric unit is shown with a grid of points and one monomer is enclosed in a brown mask. The crystallographic symmetry axes are indicated by red (3-fold) and blue (2-fold) dots.

Starting phases for the averaging calculations were from the EM model or from a sphere of randomly distributed atoms. Averaging was then performed starting at 10 Å and extending in 100 steps to 5.2 Å. The first hint that the procedure worked was obtained when a clear solvent boundary was visible inside the mask, whereas for random phases the mask was just filled with noise. The first such density is shown in Figure 20A, which at first also looks like noise but when it is inverted to display at negative contour level, some helical features are clearly visible (Figure 20B). The first obtained phases were inverted (-φ) which is something which is likely to happen when such simple starting models are used.
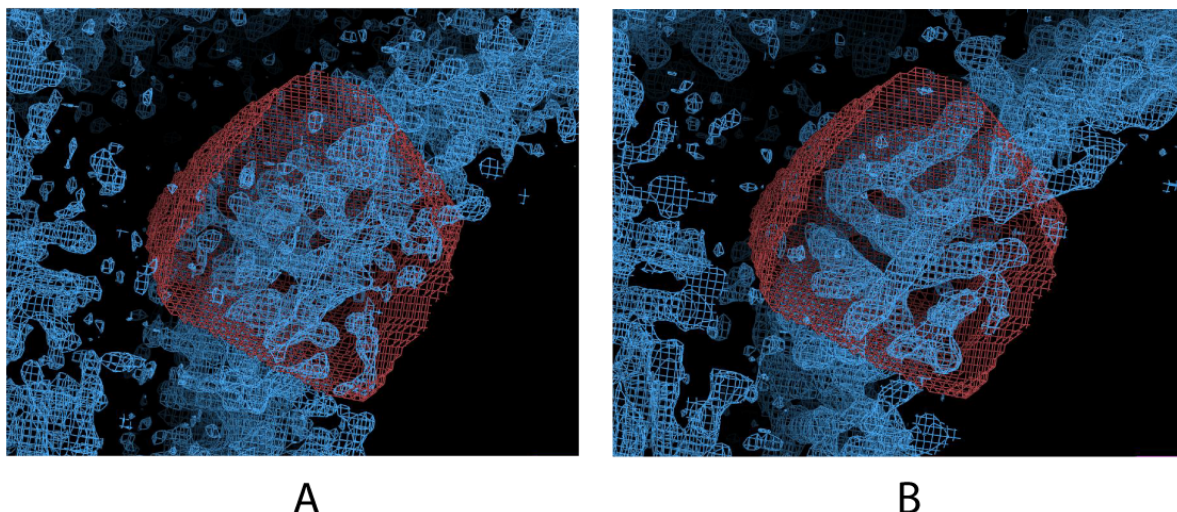
Figure 20: Calculated electron density from AVE calculations. A: F_0PhiCalc map from AVE, 100 steps of averaging from 10 to 5.2 Å, contoured at 2σ, with the monomer mask shown in red. B: same as a, but contoured at -2σ.

3.1.7 Improvement of the phases

The phases obtained from the AVE runs were then further improved with the DM program which has better options to integrate solvent modification. At the same time the best diffracting crystals were available so it was possible to calculate maps to a resolution of 3.1 Å. This first map (Figure 21A) was of quite remarkable quality for an experimentally phased map. It was soon obvious though that the phase is shifted by 180 °, resulting in an inversion of hand. This is best seen by the handedness of α-helices which is observable at this resolution (Figure 21A).

The normal approach in such a case is to invert the coordinates (-x / -y / -z) but this is not possible in this space group. Instead, the transformation is (x / y / z) to (¾-x / ¼-y / ¾-z). For atomic coordinates this transformation is straightforward yet for maps this is not implemented in the standard programs. The fastest way was to build a approximate C-α trace into the inverted density, apply the transformation and then use the model generated in the right hand as starting phases. This way a map of correct hand was obtained after averaging (Figure 21B). Further improvement was gained by increasing the number of steps of averaging. A very good experimental map was obtained which was suitable for building the model (Figure 21C).
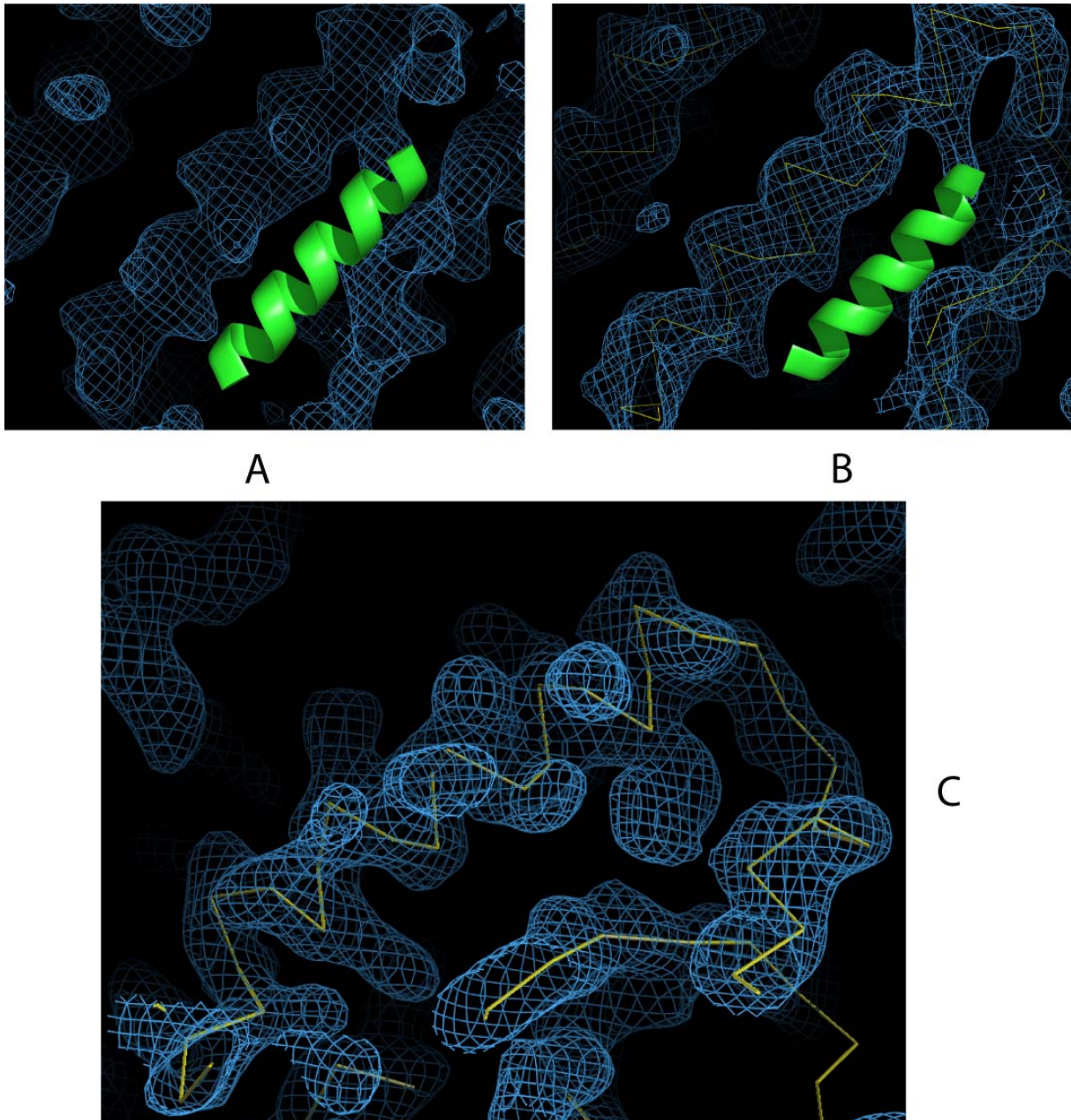
Figure 21: First high resolution electron density. A: 3.1 Å experimental density (F_0PhiCalc map from DM, 300 steps of averaging from 5.0 to 3.1 Å, contoured at 2σ) of a helical part with visible side chains, in green a left handed α-helix. B: 3.1 Å experimental density (F_0PhiCalc map from DM, 300 steps of averaging from 5.0 to 3.1 Å, contoured at 2σ) of a helix, in green a correct right-handed α-helix. C: 3.1 Å experimental density (F_0PhiCalc map from DM, 800 steps of averaging from 5.0 to 3.1 Å, contoured at 2σ) with C-α trace.

3.1.8 Model building and Refinement

A first model was built into the density with the program O (Jones, Zou et al. 1991). At the beginning only a poly-alanine trace was built and when the sequence was obvious from the bulky residues the sidechains were added. In some loop regions a clear assignment was not

possible from looking at the first experimental map while in other regions, especially in alpha helical regions, the sidechains were clearly identifiable.

During model building only one monomer was built and for refinement the symmetry was expanded to the whole 10-mer in the asymmetric unit. Refinement was performed with Refmac (Murshudov, Vagin et al. 1997) with tight 10-fold non-crystallographic symmetry restraints applied to all residues. Coordinate parameters were minimized in 20 to 30 cycles according to maximum likelihood residuals. Isotropic temperature factors were refined and a low xray data weighting term of 0.03 was chosen so as not to distort geometry. A random set of 5 % of reflections were left out during refinement in order to calculate the R_{free} values. This is potentially a problem since with high NCS the randomly chosen set of test reflections set aside for the R_{free} calculation are not really independent but taking a test set of free reflections from thin resolution shells, which are independent of reflections in other resolution ranges did not influence the ratio of R_{free}/R_{work}.

The structure from Refmac was then refined with Phenix (Afonine, Grosse-Kunstleve et al. 2005) which has better refinement strategies and the R factors decreased again significantly. Final R factors were 21.9 % for R_{work} and 23.9 % for R_{free}. Refinement statistics are summarized in Table 6.

Refinement	
Resolution (Å)	50–3.1
No. reflections	219,258
Rwork/Rfree	21.9/23.9
No. atoms	22070
Protein	21970
Water	100
B-factors	
Protein	128.3
Water	108.3
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.330

Table 6: Refinement statistics
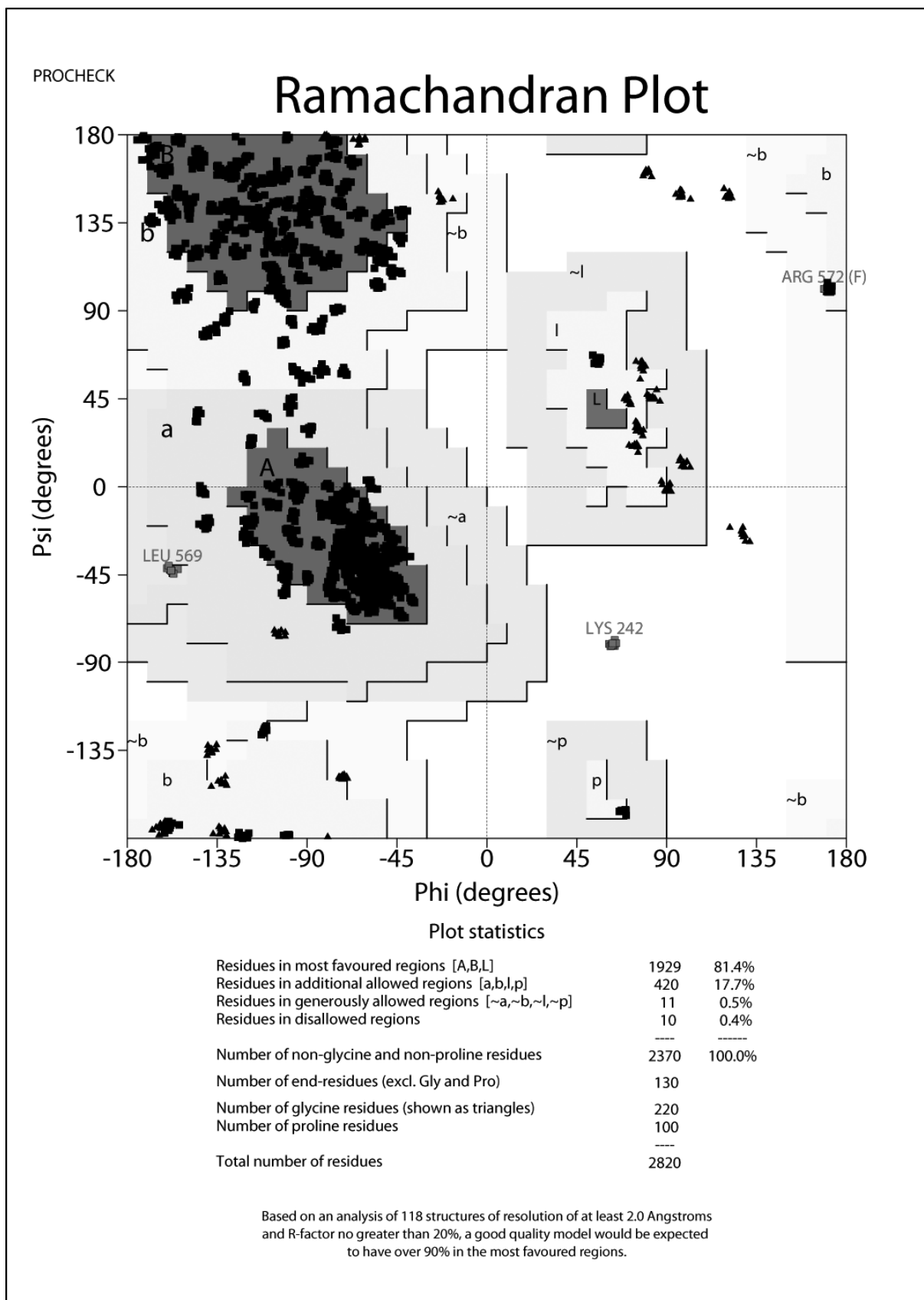
Figure 22: Ramachandran Plot. The Ramachandran diagram shows the distribution of the φ versus ψ angles of the amino acids. 81.4 % of the residues lie in the most favored region which is a reasonable value for a 3.1 Å structure. The only outlier is lysine 242 which is the middle residue of a gamma turn situated at the end of a β-hairpin structure. The average

values of a classic gamma turn are 75 ° for φ and -64 ° for ψ (Rose, Gierasch et al. 1985; Milner-White, Ross et al. 1988). The plot was calculated with PROCHECK.

3.1.9 Description of the Structure

The 31 kDa monomer of *Thermotoga maritima* encapsulin consists of two compact domains and a protruding loop (Figure 23). The P-domain forms the main body of the structure. It consists of a mixed alpha/beta structure, contains the N-terminus and is very fragmented with regard to primary sequence. A conserved hydrophobic core is located between the helical and beta sheet regions. The E-loop is responsible for the formation of contacts between the two-fold symmetry related subunits by providing a strand that completes a beta sheet formed by both subunits. The A-domain forms a compact structure consisting of three helical segments and a five stranded beta sheet which also contains the C-terminus. This domain has little connections to the rest of the monomer and mediates all contacts of the five-fold symmetry interface. The loop region close to the five-fold axis between helices α6 and α7 is particularly flexible judged based on electron density features and high temperature factors.
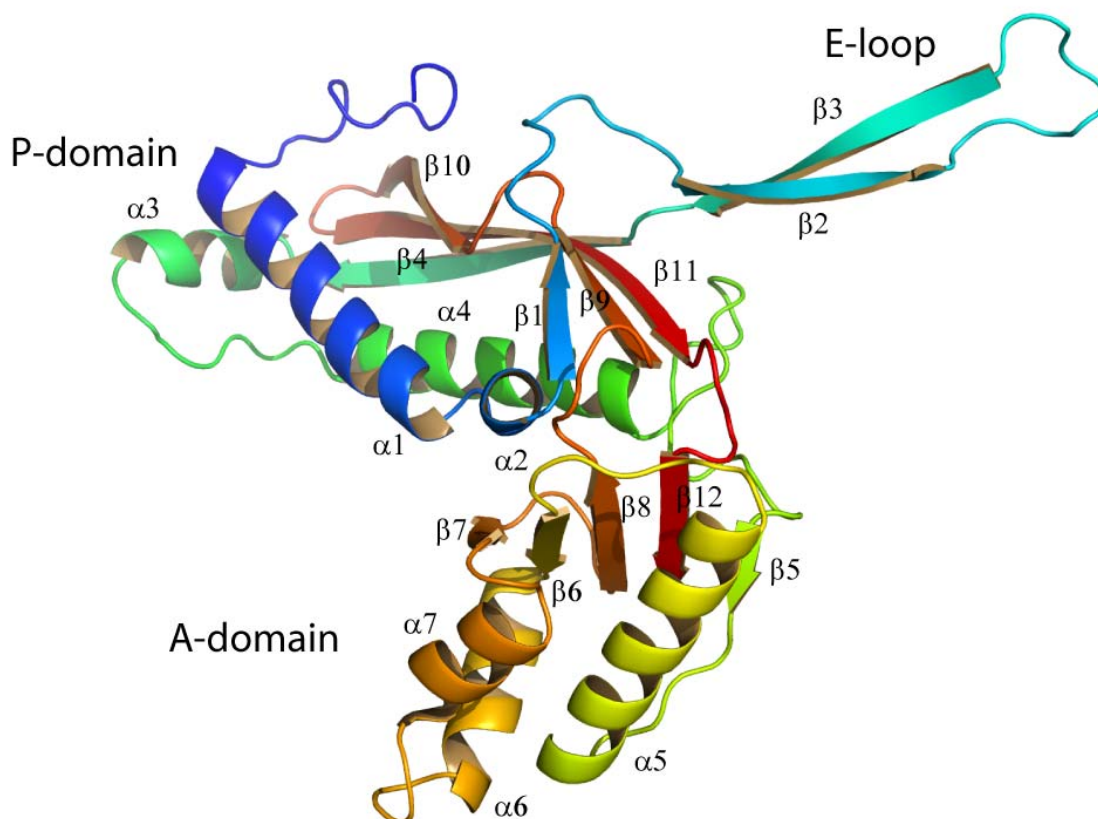
Figure 23: Monomer structure of *T. maritima* encapsulin. The monomer consists of the two domains A and P and the protruding E-loop. Coloring is in rainbow from N-terminus (blue) to C-terminus (red) and naming of the domains is according to the homologous gp5 major capsid protein of the HK97 virus (pdb: 1ohg) (Wikoff, Liljas et al. 2000; Helgstrand, Wikoff et al. 2003).

40

60 monomers of encapsulin assemble to a thin spherical shell with a diameter of 230-240 Å and a thickness of 20-25 Å (Figure 24A,B).
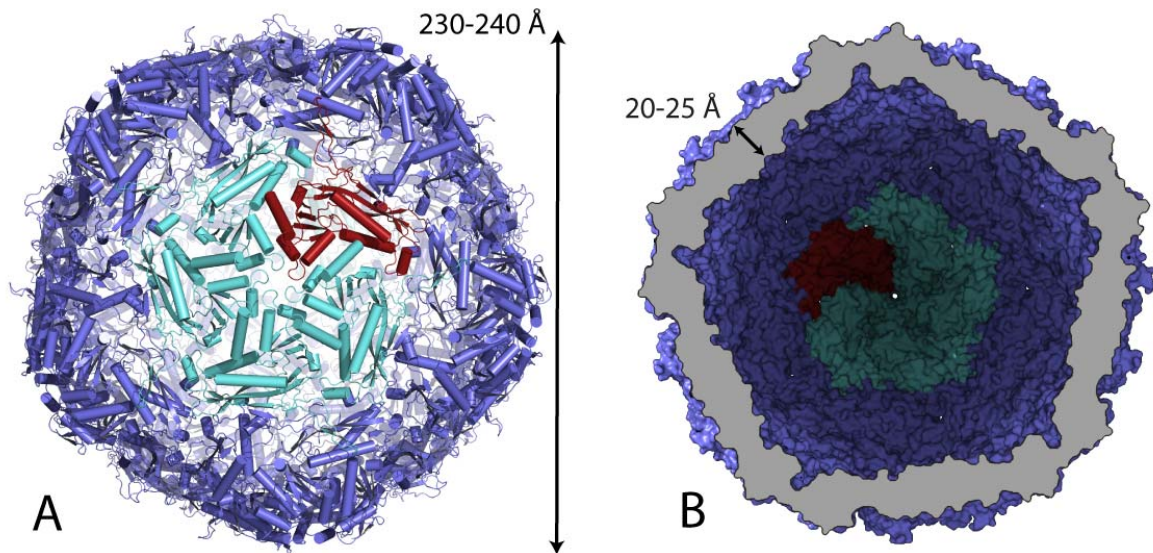
Figure 24. The spherical encapsulin shell. A: View from the outside on the five-fold symmetry axis, one pentamer is highlighted in cyan with one monomer in red. B: View to the inside of the shell which is cut open in the middle and displayed in a surface representation.

It is interesting to see how the protein forms such a large structure with only one monomer. In Figure 25A you can see how one monomer is surrounded by its symmetry related neighbors. A prominent feature is the long E-loop which overlaps with the 2-fold related molecule and forms a beta sheet structure where both proteins contribute strands (Figure 25C). In the space-filling view (Figure 25B) it is even more evident that the inter-subunit contact surface is quite large which explains the extraordinary stability of this particle.
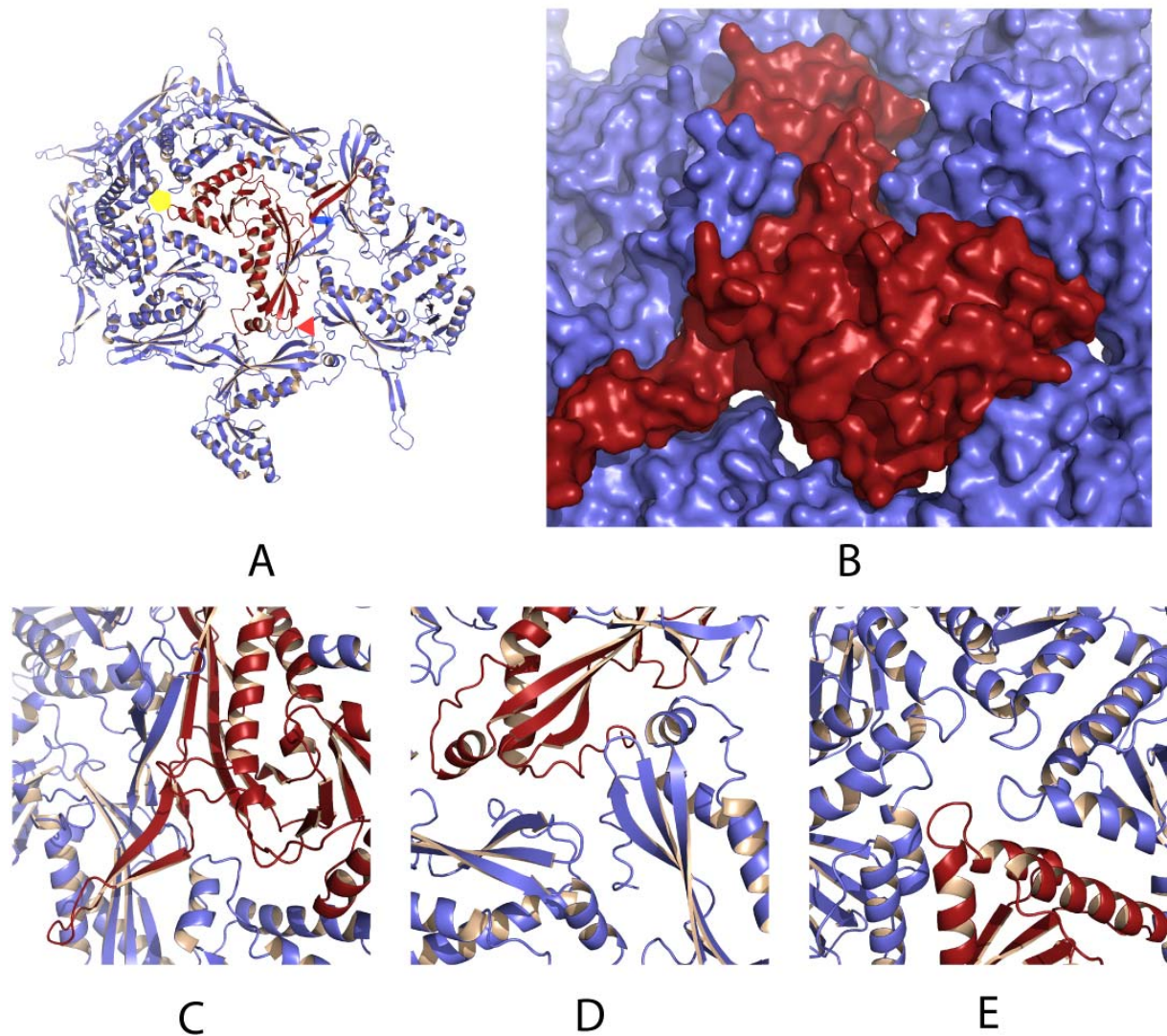
Figure 25: Arrangement of monomers in icosahedral symmetry. A: The protein monomer in blue and its symmetry related molecules, 5-fold axis indicated as a yellow pentagon, 3-fold as a red triangle and 2-fold as a blue ellipse. B: Space-filling view of a monomer and its surrounding symmetry related proteins. C,D,E: Close-up of the 2-,3- and 5-fold symmetry interactions.

From the closer look at the interactions between symmetry related molecules (Figure 25C-E) it could be reasoned that the 2-fold interaction is the first and strongest interaction due to the large overlap of the proteins. The interaction at the fivefold seems weakest and the loop there also has very poor density. This makes sense as structurally homologous proteins can form either a 5-fold or 6-fold interaction at this position. At the fivefold the proteins make a crater-like structure, a feature which was already visible in the very low resolution EM maps (Figure 12A). This part also forms crystal contacts between the shells.

encapsulin forms a thin shell with several openings. Two holes are located at the points of particle symmetry (Figure 26B,C) and one of them at the interface between two subunits

(Figure 26A). The latter forms a strongly negatively charged channel while the threefold hole contains two positively charged residues and the residues forming the fivefold hole are mainly uncharged.
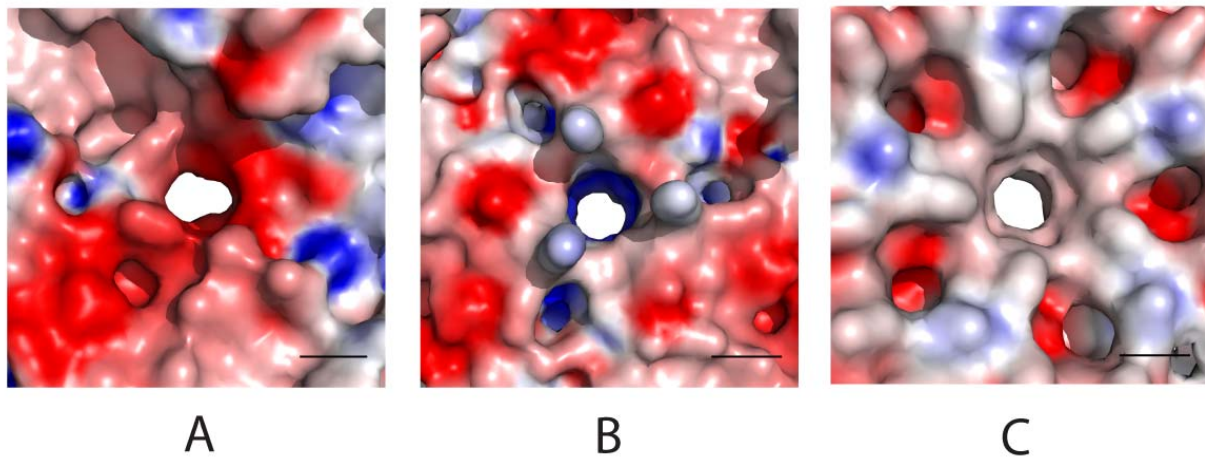
Figure 26. Electrostatic surface representation of the three pores in the shell of *T. maritima* encapsulin. The pore at the interface of two neighboring subunits (A) is lined with conserved negative charges while the threefold interface (B) is largely negative and the fivefold mainly uncharged (C). The surface is colored with a blue to red gradient from +10 to -10 kT/e. The bar length corresponds to 5 Å.

During the refinement of the model, additional electron density in the vicinity of the N-terminus became evident, which could not be explained by the amino acid sequence of the encapsulin protein alone (Figure 27). This short, continuous stretch of the Fo-Fc difference density shows features consistent with a bound peptide and can be observed in all ten monomers of the asymmetric unit. The three disordered residues of the N-terminus cannot explain this additional feature as it is disconnected from the main chain electron density and it is longer than could be accounted for by only three amino acids.
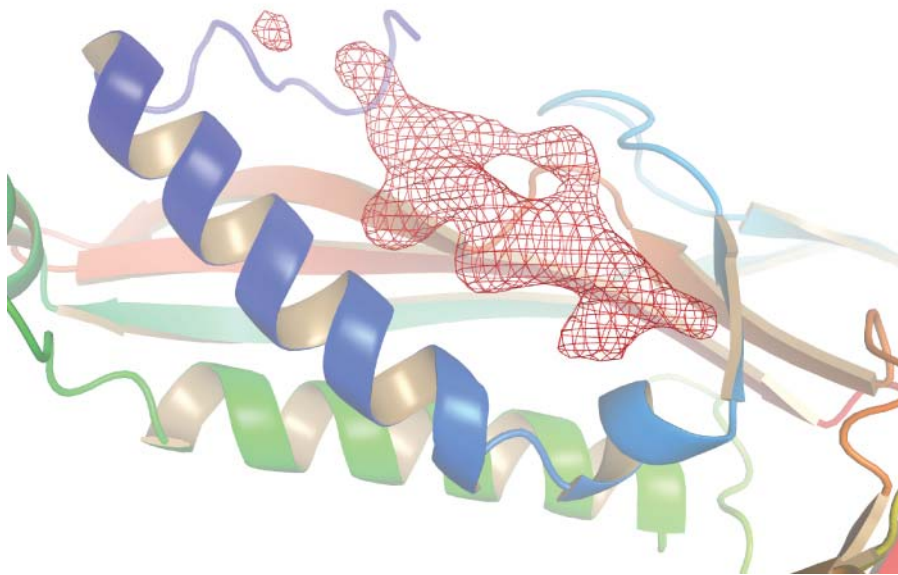
Figure 27. Extra density on the inside of the particle. Density is shown in red from a 10-fold averaged Fo-Fc omit map contoured at 6.5 σ.

3.1.10 Structural Homology Search (DALI)

A DALI search (Holm and Sander 1993) with the structure of the *Thermotoga maritima* encapsulin produced the following hits (search with poly-Ala model):

pdb	name	Z-score	rmsd (Å)	% identity	alignment length (aa)
1ohg	Structure of the ds-DNA bacteriophage HK97 mature empty capsid	13.4	3.4	11	199
1yue	Bacteriophage T4 capsid vertex protein gp24	10.5	3.8	12	198
2pk8	Crystal structure of an uncharacterized protein PF0899 from Pyrococcus furiosus	7.6	2.6	5	81

Table 7: DALI search results

The best hit is the protein gp5 from the bacteriophage HK97 capsid which forms the structural capsid pentamers and hexamers. The second hit is a bacteriophage T4 structural protein and the third from a protein of unknown function. The three-dimensional structure of encapsulin is apparently highly homologous to viral shell proteins. The homology is only based on

44

structure; there is no significant similarity of the primary amino acid sequences (lower than 15 % identity).

Remnants of phages are often found in bacteria but normally not just single genes are found but the whole assortment of genes necessary for the phage life cycle (Figure 28). If only one single structural gene is found then this would indicate that the inclusion happened early in the evolutionary timescale and that this gene was useful for the organism. Other phage proteins which maybe once were present there too would then have dropped out of the genome because they were not necessary.
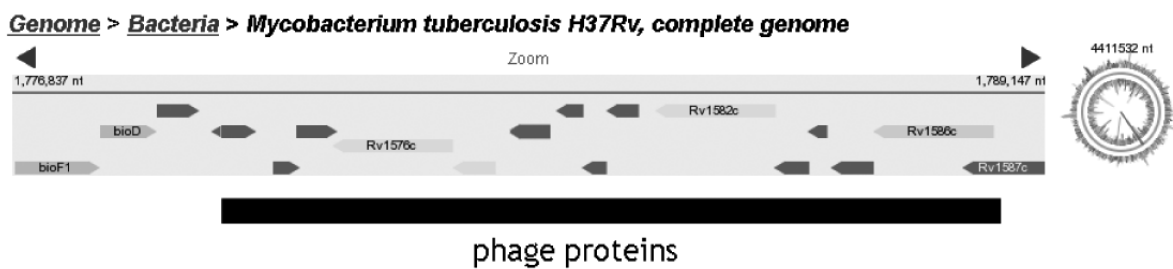
Figure 28: Remnant phage proteins in *Mycobacterium tuberculosis.* (adapted from http://www.ncbi.nlm.nih.gov)

3.2 The genetic context of encapsulins

A position-specific iterative BLAST search (Altschul, Gish et al. 1990; Altschul, Madden et al. 1997) with encapsulins reveals 49 homologous proteins (December 2007). A closer look at their genomic context shows that almost all of them are part of an operon that codes for two proteins where the gene for encapsulin lies downstream. In 28 cases the preceding protein is a specific type of peroxidase (referred to as dye-decolorizing peroxidase, DyP (Kim and Shoda 1999)) and in 17 cases it is a protein related to ferritin (ferritin-like protein, Flp). The space between the two genes is in most cases very small or they are even overlapping by a few bases, suggesting tight translational coupling (Figure 29B). We analyzed the DyP proteins and found that they are in general highly homologous, but when they appear in an operon with encapsulin they ubiquitously display a C-terminal extension of 30-40 amino acids (Figure 29A); the same is also true for Flps. The extension consists of a region of variable length with a high content of alanine, glycine and proline followed by a conserved anchor sequence Figure 29C).
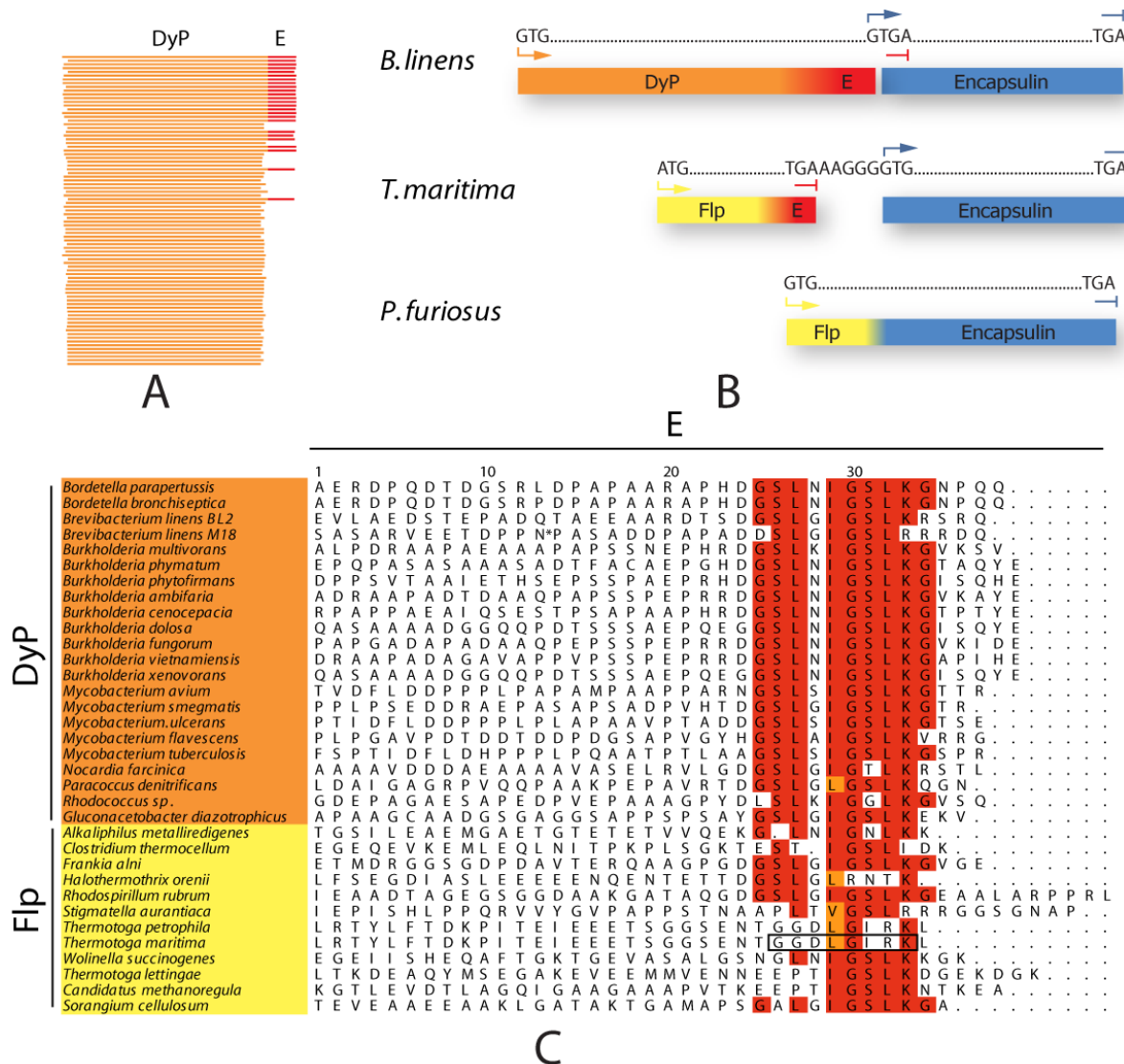
Figure 29: Packaged proteins possess a C-terminal extension. A: Schematic representation of the size of proteins resulting from an iterative BLAST search with DyP proteins against the protein data base, shown are the first 83 hits. The conserved DyP sequence is colored in orange and the extension sequence, which is only present if DyP is in an operon with encapsulin, is colored in red. B: Schematic illustrating the genetic context of encapsulins. GTG or ATG start codons are marked with an arrow and the TGA stop codons with a horizontal T. The translated proteins are shown as bars. E refers to the extension sequence. C: Multiple alignment of the extension sequences of DyP proteins (underlaid in orange) and ferritin-like proteins (yellow) from various species. Residues are colored according to conservation with red indicating identity and orange homologous substitution. The sequence built into the additional density of the *T. maritima* encapsulin is highlighted with a black box. The position of the *B. linens* M18 sequence where a stop codon was inserted to express a peroxidase without binding motif is marked with an asterisk.

3.2.1 DyP proteins

DyP proteins were first characterized in 1999 when their ability to decolorize various artificial dyes was found (Kim and Shoda 1999). They belong to a large family of heme containing peroxidases consisting of more than 150 bacterial and 6 eukaryotic (fungal) members. There

are several structures available of DyP proteins, two from bacterial members (pdb: 2hag/2iiz and 2gvk) and one from a fungus (pdb: 2d3q), but so far their function *in vivo* is still unclear. Figure 30 shows the structural alignment of the three known structures. Even though the bacterial and fungal members are from very different parts of the evolutionary tree their core structure aligns very well. The only significant differences are observable in several loop regions, with the fungal protein (shown in green) generally having longer extensions.
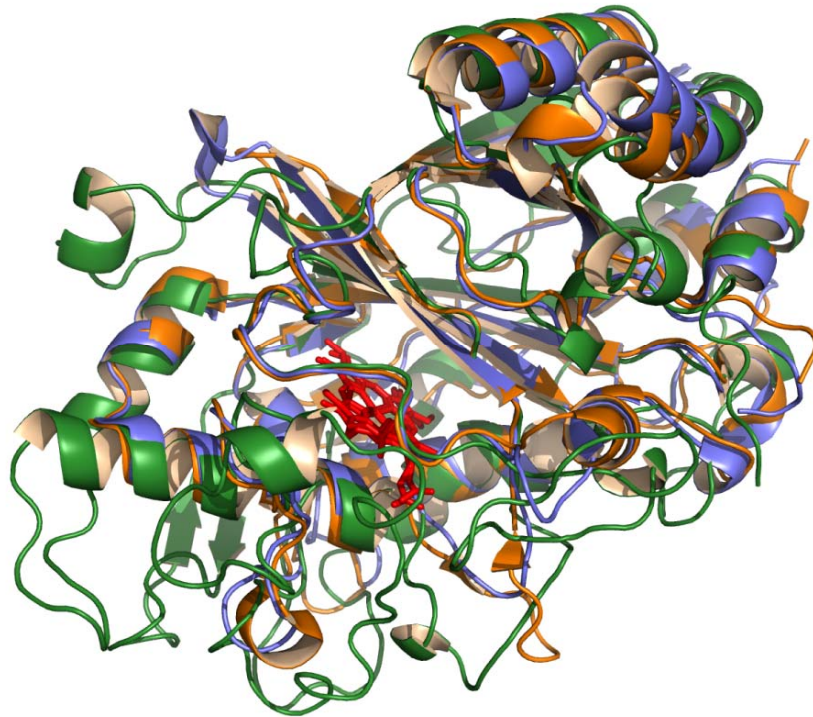
Figure 30: Alignment of the known structures of DyPs. The following structures are depicted: 2iiz from *Shewanella Oneidensis* (orange), 2gvk from *Bacteroides Thetaiotamicron* (lightblue) and 2d3q from *Thanatephorus Cucumeris* (green). The overall structure seems to be quite well conserved among the three solved members. The bacterial proteins have an overall rmsd of 1.7 Å to each other, and 2.9 Å to the fungal protein. If only the well aligned parts are evaluated, the rmsd values are 0.90 Å for 2gvk/2iiz, 1.46 Å for 2iiz/2d3q and 1.94 Å for 2gvk/2d3q. The heme prosthetic group present in the 2iiz and 2d3q structures is shown as red sticks.

3.2.2 Ferritin-like proteins

In some organisms, mostly anaerobes and/or thermophiles, the gene for encapsulin is preceded by putative ferritin-like proteins. The structure of such a protein from *Nitrosomonas europaea atcc 19718*A (56 % sequence identity to the *T. maritima* ferritin-like protein) was solved as part of a structure genomics project (Chang 2005) but no detailed functional analysis was carried out. It forms a pentamer of dimers arranged in a ring-like structure

(Figure 31A). The monomer consists of two long anti-parallel alpha helices of about 30 residues followed by a shorter 10-12 amino acid long helix. This fold is very similar to the one observed for ferritins (Figure 31B). Although ferritins have a different quaternary structure (it forms a 24-mer and a single subunit consists of four helices), the residues responsible for the diiron binding in the ferroxidase center are at the same in this ferritin-like protein. The active site residues superimpose very well (Figure 31C), therefore making a role in iron metabolism very likely.
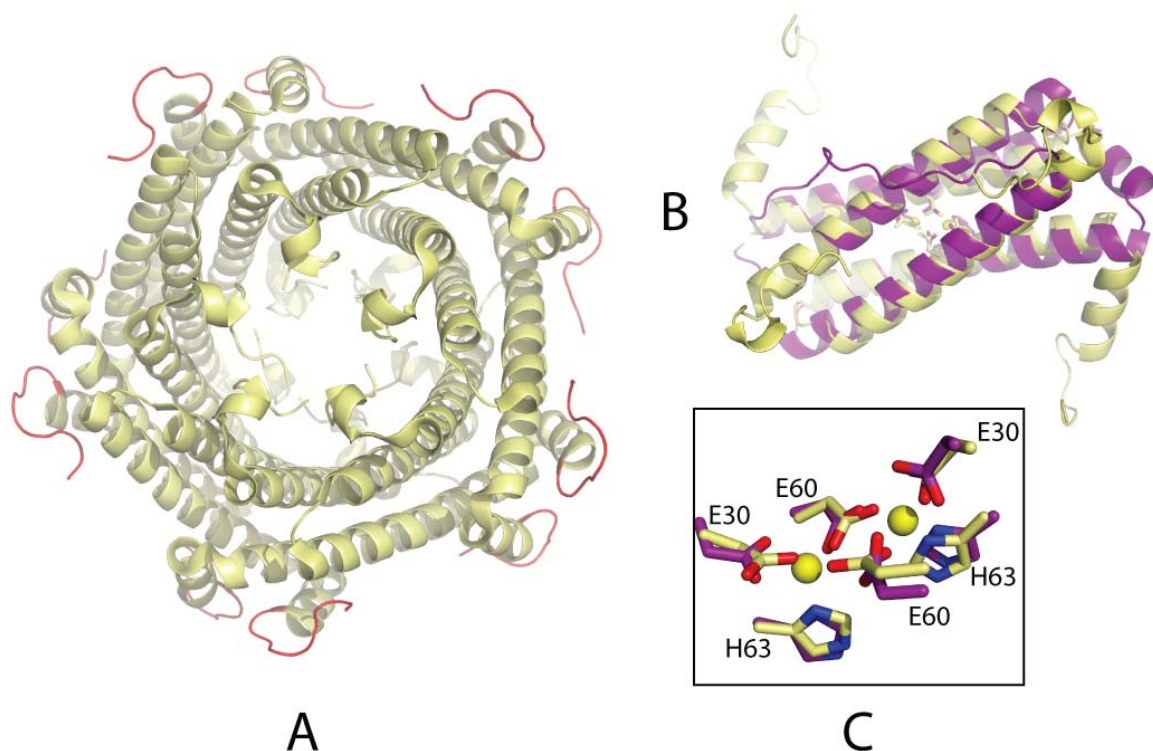
Figure 31: The protein preceding *T. maritima* encapsulin is related to ferritin. A: Proposed biological unit structure of the protein NE0167 which is formed of a pentamer of dimers. The C-terminal nine residues are marked in red. B: Alignment with *E. coli* bacterioferritin (pdb: 1bfr), the monomer is shown in purple with the metal ions (manganese) as yellow spheres and the coordinating residues as sticks. The superimposed dimer structure of NE0167 from *Nitrosomonas europaea* is shown in yellow. C: Close-up of B. The conserved residues in the ferroxidase active site align perfectly, even though there are no bound metal ions in the *N. europaea* structure. Residue numbering corresponds to the *N. europaea* ferritin-like protein.

3.2.3 *Pyrococcus furiosus* virus-like particle

Combining the crystallographic evidence for the additional density on the inside of the encapsulin with the analysis of the genomic environment suggests that ferritins and peroxidases with unique C-terminal extensions are interacting with encapsulins. According to the Rosetta Stone method for predicting protein interactions (Marcotte, Pellegrini et al. 1999),

two proteins are likely to interact if they have a homolog in a different organism where both are fused on one single polypeptide chain. Indeed, it is possible to identify larger proteins (343-373 amino acids) that in part align to encapsulins and are present in archae. One of these, a virus-like particle of *P. furiosus* (PfV), was structurally characterized in 2007 and likewise found to form a shell-like assembly (Akita, Chong et al. 2007). The fold of the monomer is very similar to encapsulin, but this protein assembles into a T=3 icosahedral particle. Interestingly, the first 109 amino acids of PfV, which cannot be aligned to encapsulins, are disordered in the structure. This region displays significant similarity with the ferritin-like protein from *T. maritima* both in the sequence of the conserved ferroxidase active site residues (Figure 32) and predicted secondary structure. The two regions of the sequence are not separated by an alanine, glycine, proline rich sequence similar to the extension sequences found in above mentioned peroxidases and ferritin-like proteins.
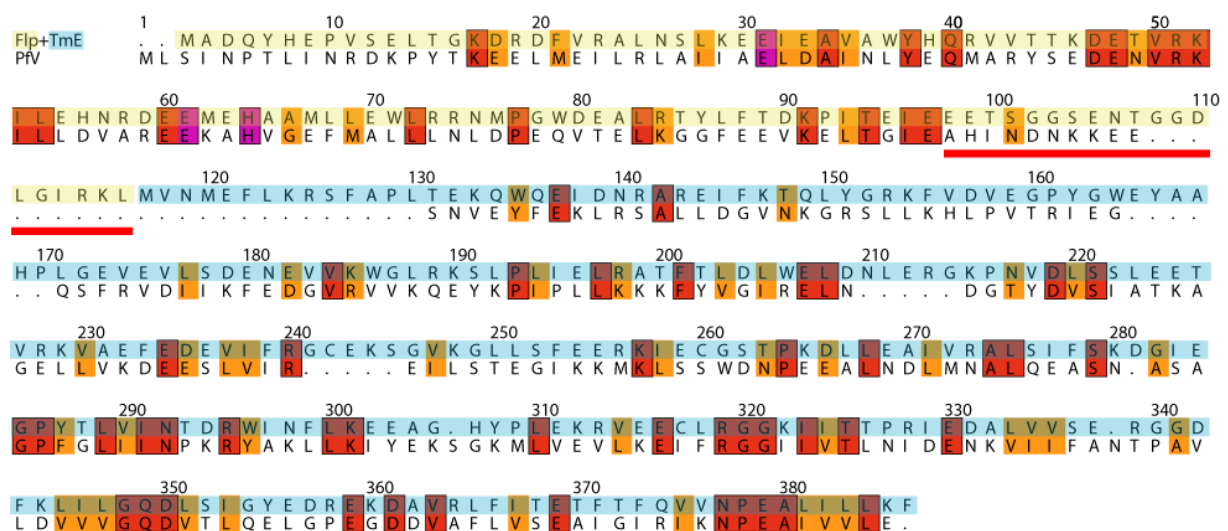
Figure 32: The protein preceding *T. maritima* encapsulin and the C-terminal region of the *P. furiosus* virus like particle are related. Sequence alignment of PfV and *T. maritima* Flp/encapsulin (underlaid in yellow/light blue). Residue coloring is according to conservation with red representing identity and orange homology. The ferredoxin active site residues are colored in magenta and the extension sequence is marked with a red bar.

3.2.4 Structural comparison with PfV and HK97 gp5

Comparing encapsulin to the two related shell proteins of PfV (Akita, Chong et al. 2007) from archaea and of gp5 from the Head II structure of the HK97 bacteriophage (Helgstrand, Wikoff et al. 2003), most structural elements are very conserved (Figure 33A) and the packing interactions of the monomers in the particle are very similar (Figure 33B-D). The rmsd values for the structural alignments of encapsulin to the two proteins are is 2.39 Å for 199 residues of PfV and 2.65 Å for178 residues of gp5, the smaller number of residues in the latter case is due

to the absence of the N-terminal helix (α1) in the phage protein Figure 33A, bottom left inset). The most striking difference is the 60° rotation of the E-loop in encapsulin compared to the related structures (Figure 33A, bottom right inset). The E loops are responsible for tight twofold interactions in encapsulin (Figure 33B-D), which are not observed in the larger PfV and HK97 particles. Large differences are also observed in the loops at the tip of the A domain which is responsible for the interactions at the five- or pseudo-sixfold axis (Figure 33A, top left inset). Interestingly, although at the sequence level encapsulin and PfV are more closely related than either is to gp5, encapsulin and the phage capsid protein both have an insertion loop at the end of helix α4 which is not present in PfV (Figure 33A, top right inset).
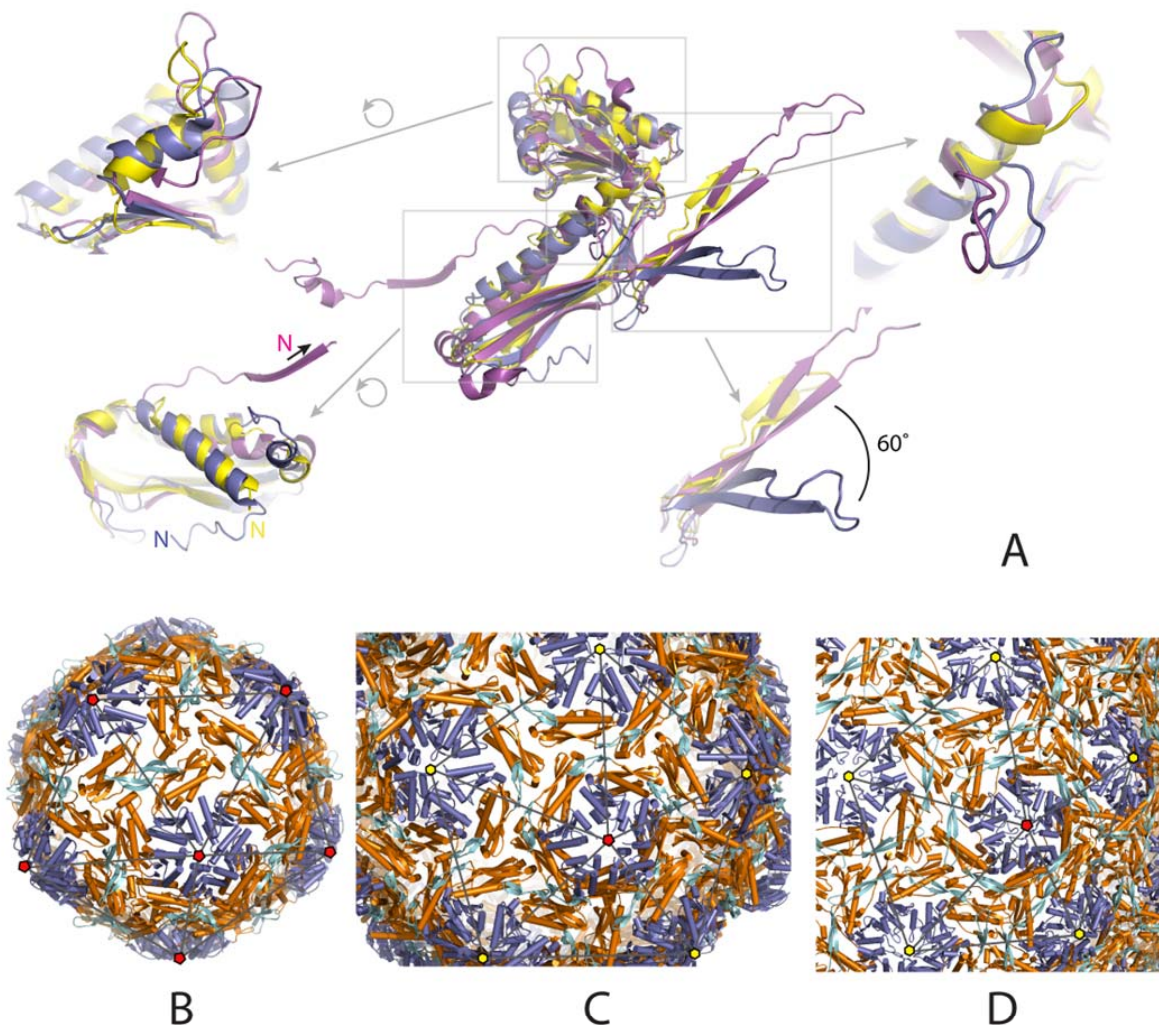
Figure 33: Structural comparison of *T. maritima* encapsulin with *P. furiosus* PfV and gp5 of the HK97 virus. A: Structural alignment of encapsulin (colored in blue), PfV (yellow) and gp5 (magenta). Four selected differences are shown as enlarged close-ups. The insets on the left are views from the back of the molecule shown in the center. B-D: Packing of the shell proteins of encapsulin (T=1 icosahedral symmetry, 240 Å diameter), PfV (T=3, 360 Å diameter) and gp5 (T=7, 660 Å diameter). The A, P and E domains are colored blue, orange and cyan, respectively and the icosahedral pentamers are traced with gray lines. Fivefold axis are indicated with red pentagons and pseudo-sixfold axis with yellow hexagons.

The residues responsible for the chain mail type crosslinking in the gp5 protein (Wikoff, Liljas et al. 2000), lysine 169 and asparagine 356, do not exist at the corresponding position in the *T. maritima* encapsulin structure and consequently no crosslinked residues were observed.

3.2.4 Additional density constitutes the binding interface for a packaged protein

Judging from the features of the additional segment of electron density on the interior of the encapsulin we concluded, as outlined above, that it might arise from a bound peptide. We could obtain a good fit to the density by interpreting it with a sequence of the anchor region of the C-terminal extension of the ferritin-like protein expressed in the same operon as the *T. maritima* encapsulin (Figure 34). The GGDLGIRK peptide built into the additional density refines well to produce a model with good geometry and results in a better overall free R-factor. The interaction between the peptide and the binding pocket on the interior of the spherical shell makes chemical sense since leucine and isoleucine are buried in a hydrophobic pocket formed by the conserved amino acids Phe30, Leu34, Leu233 and Ile249. The electron density also provides evidence for a salt bridge between the aspartate and the arginine in the sequence and a kink induced by the central glycine residue. The two central hydrophobic residues separated by one glycine provide the register while the directionality is defined by the two N-terminal glycine residues which would not fill enough density if the direction of the trace was reversed. Further tracing of the ferritin-like particle is not possible due to the flexibility of the extension. The protein would then disobey the crystallographic and non-crystallographic symmetry and would, therefore, not be visible in electron density maps. Interestingly, the structure of the *N. europaea* ferritin-like protein forms a pentamer of dimers in which five C-termini are separated by distances that would agree with the location of the additional density inside the encapsulin.
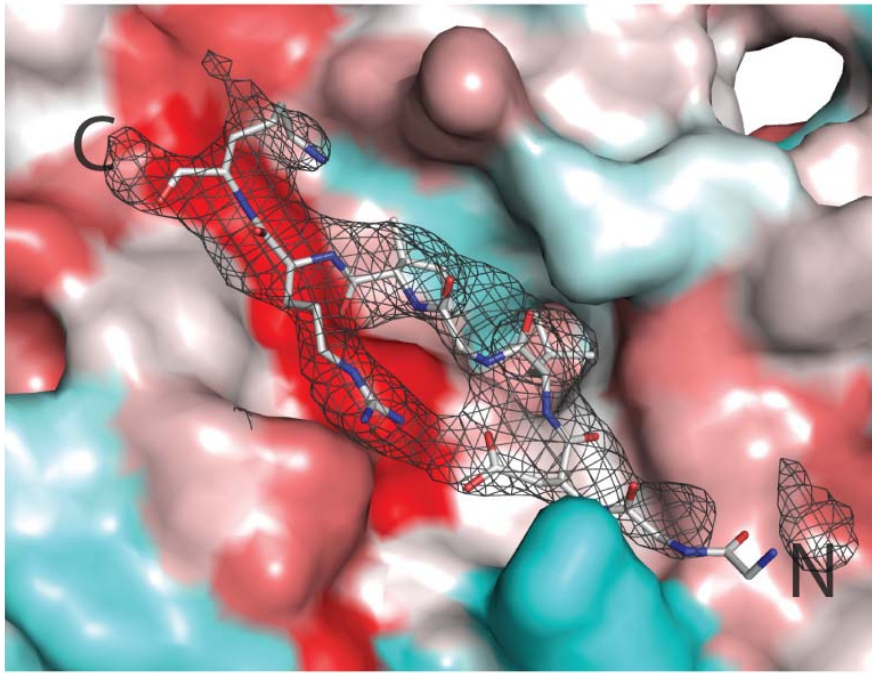
Figure 34: Peptide model built into the additional density. 10-fold averaged omit electron density map shown at a sigma cutoff of 5 with the peptide model in stick representation with labeled N/C termini and the *T. maritima* encapsulin surface colored according to conservation from cyan (low) over white (medium) to red (high).

The residues surrounding the binding site have a high degree of conservation (Figure 34) and, furthermore, the inside of the particle is in general much more conserved than the outside. Additional experimental support for the packaging of the putative ferritin inside the *T. maritima* encapsulin particles, purified from native source, is provided by the presence of the ferritin-like protein in washed and dissolved crystals which were subjected to trypsin-digest/MS analysis (Figure 35) and the observation that the crystals are dark yellow in color (Figure 14), indicating accumulation of iron, which was also confirmed with an X-ray fluorescence scan (data not shown).
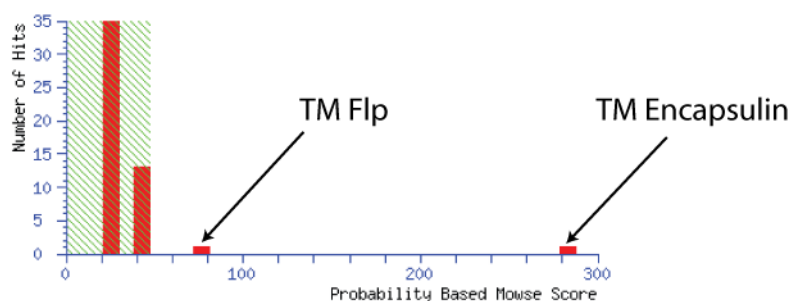
Figure 35: Probability based score of the two significant MS hits of dissolved crystals. Encapsulin crystals were washed twice and then dissolved in water. This sample was subjected to Trypsin digest/MALDI MS analysis. Peptides originating from tryptic digest fragments of both encapsulin and the *T. maritima* ferritin-like protein were identified.

3.3 Purification and functional analysis of *B. linens* encapsulin

3.3.1 Purification of *B. linens* M18 encapsulin from native source

A *Brevibacterium linens* M18 culture was grown and the culture supernatant was collected and concentrated according to protocol. The retentate had a dark brown color and a total protein concentration of approximately 1 mg/ml. At this stage it is not possible to visualize the encapsulin protein on a SDS-PAGE gel and coomassie staining. After pelleting the large molecular weight fraction of the culture supernatant and loading this on a 10-50 % sucrose cushion encapsulin is visible and can be identified due to its migration on an SDS-PAGE gel and also by MALDI-MS (Figure 36).
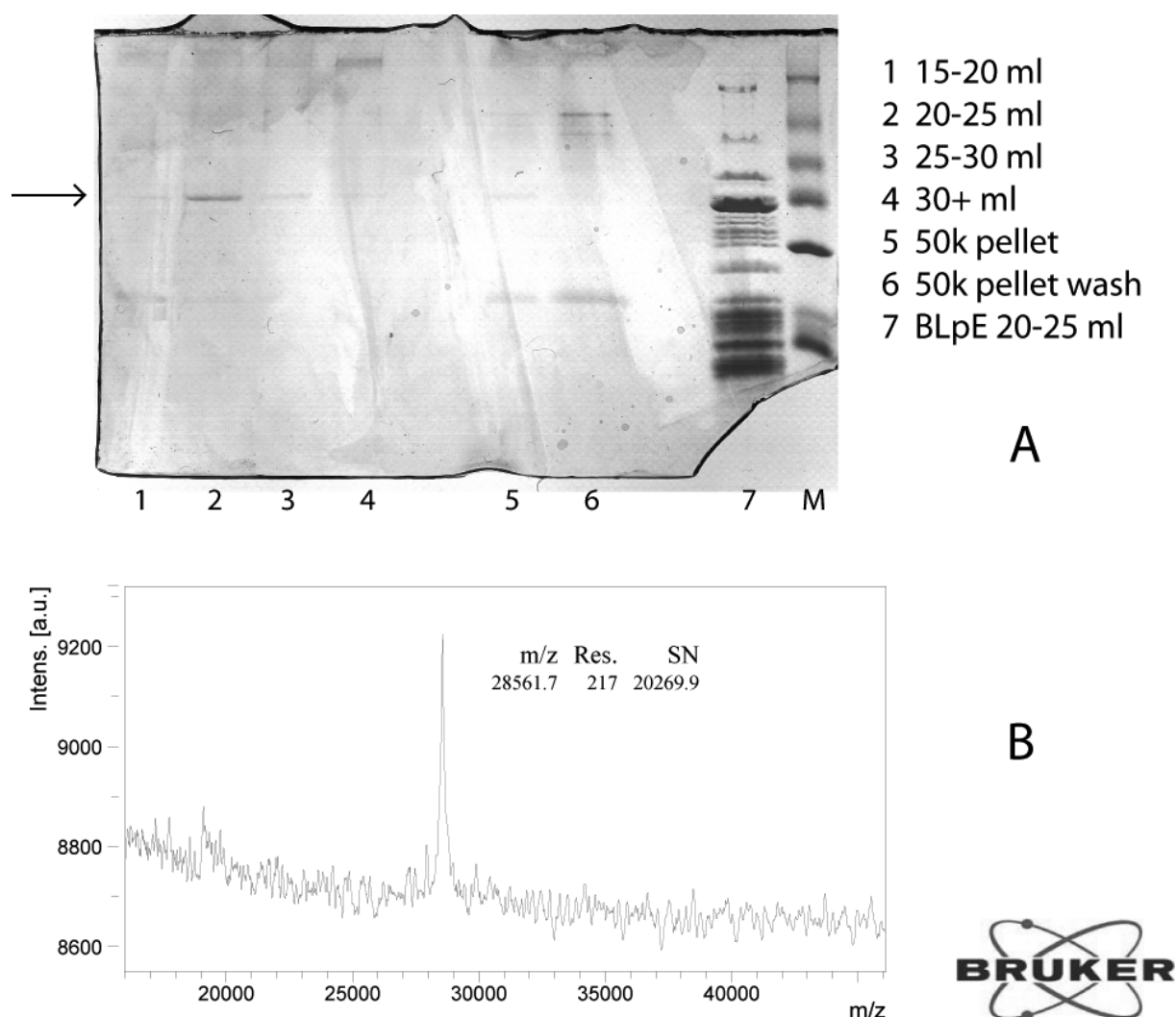
Figure 36: Analysis of purification of native *Brevibacterium linens* encapsulin/linocin M18. A: 15 % SDS-PAGE gel of purification of native encapsulin from culture supernatant. 1-4: fractions from final sucrose gradient (10-50 %) 5+6: fractions after O/N spin through sucrose cushion. 7: size comparison with sucrose gradient fraction of recombinant encapsulin. The arrow marks the position where encapsulin is expected to migrate. B: MALDI-MS of the

20-25 ml sucrose cushion fraction yields one peak of the correct mass (predicted: 28565 Da, measured: 28561.7 Da)

3.3.2 Purification of recombinant *B. linens* encapsulin

Recombinant encapsulin could be expressed from various different constructs, wild-type and with a his tag at the C-terminus with or without linker between protein and tag. The highest yield of soluble protein was achieved when the culture was grown at 37 °C and then cooled to 22 °C for induction with 0.1 mM IPTG overnight (Figure 37).
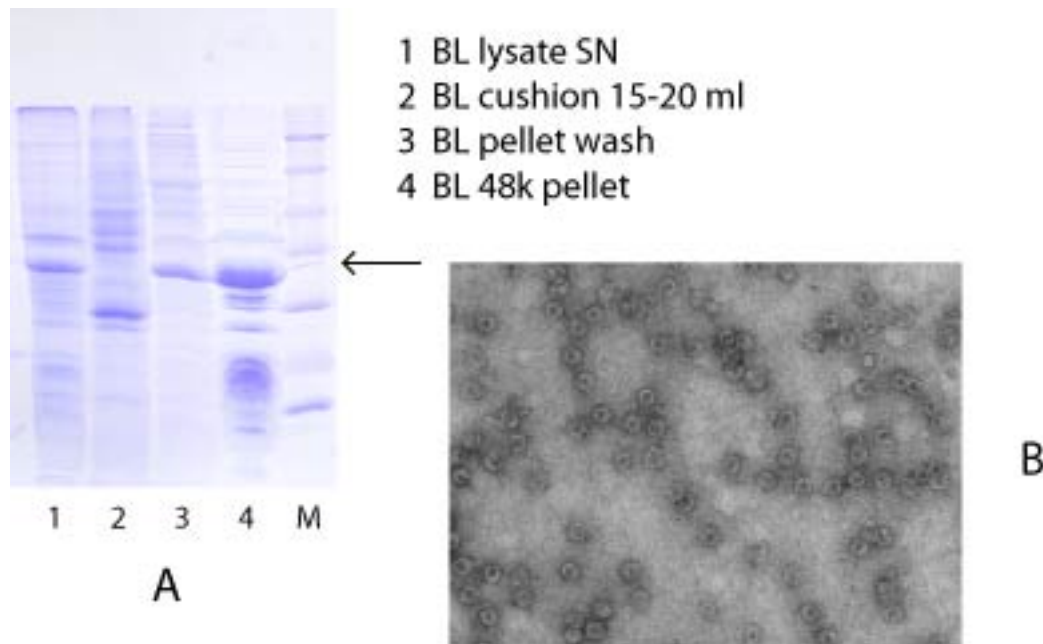
Figure 37: Preparation of encapsulin expressed in *E. coli*. A: SDS-PAGE gel of recombinant encapsulin preparation from lysate to pelleting through a sucrose cushion (1-4). Most of the protein is found in the pellet after 17 h at 48 krpm in a Ti-70 spin as shown in lane 4. The arrow indicates the expected migration of the encapsulin band. There is still some contamination with ribosomes as indicated by the various bands below. B: Negative stain electron microscopy picture of recombinant encapsulin at a magnification of 32'000x.

3.3.3 Coexpression of *B. linens* encapsulin with DyP

In order to demonstrate experimentally that the extension sequence targets the ferritins or the peroxidases to the interior of the spherical particle we cloned and expressed the two-gene-operon from *B. linens* M18 in *E. coli*. The part of the *Brevibacterium linens* genome coding for DyP followed by encapsulin was amplified and the fragment inserted into a pET21a vector. The ribosome binding site for encapsulin is contained within the region coding for the C-terminus of DyP. Expression levels are higher for DyP judging from the SDS-PAGE gel of the soluble lysate (Figure 38, lane 5).

When the sample is purified according to protocol (sucrose cushion / gradient) most of the DyP is found above the sucrose cushion but some part specifically co-migrates with the encapsulin particle to the same position in the 10-50 % sucrose gradient (Figure 38A, lane 8).
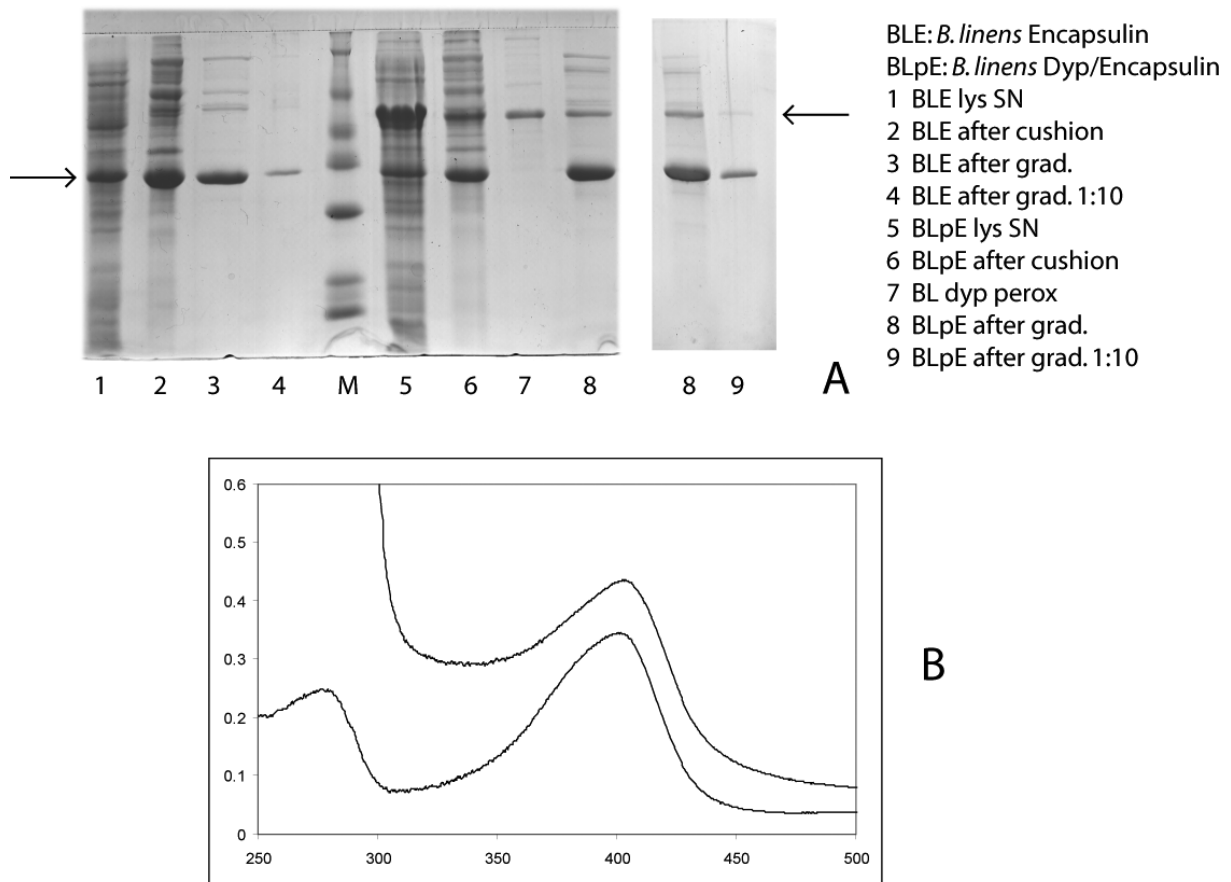
Figure 38: Purification of recombinant encapsulin without (1-4) and with coexpression of DyP (5-9). A: The left arrow indicates the position of encapsulin while the right arrow shows the migration of DyP in the 15 % SDS-PAGE gel. The peroxidase is co-migrating with the encapsulin as it is only found in fractions where both proteins are present. The ratio of encapsulin to peroxidase is about 1:10 as estimated from the intensity of the bands (lanes 8/9). B: UV absorption spectrum of samples corresponding to lanes 7 and 8, baseline corrected with 1x FB. The upper trace shows the copurified DyP and encapsulin with a higher ratio of A280/407 nm than the DyP alone.

DyP alone could be purified from the part of the preparation which did not enter the sucrose cushion. The fractions containing the peroxidase were brown in color, indicating the presence of a heme prosthetic group. This was confirmed by measuring an absorption spectrum which showed a peak at 407 nm (Figure 38B). Protein lacking a heme cofactor could be purified away using an anion exchange step. The migration of the protein in an S-200 gel filtration run indicated the formation of a multimeric structure.

3.3.4 EM analysis of DyP and co-expression with encapsulin

Electron microscopic analysis of the particles from the co-expression reveals additional protein inside the shells (Figure 39B) which is not present when encapsulin is expressed alone (Figure 39A). The protein packaged into the encapsulin shell has similar size and shape as the isolated BlDyP, which forms multimers according to electron microscopic experiments (Figure 39C).
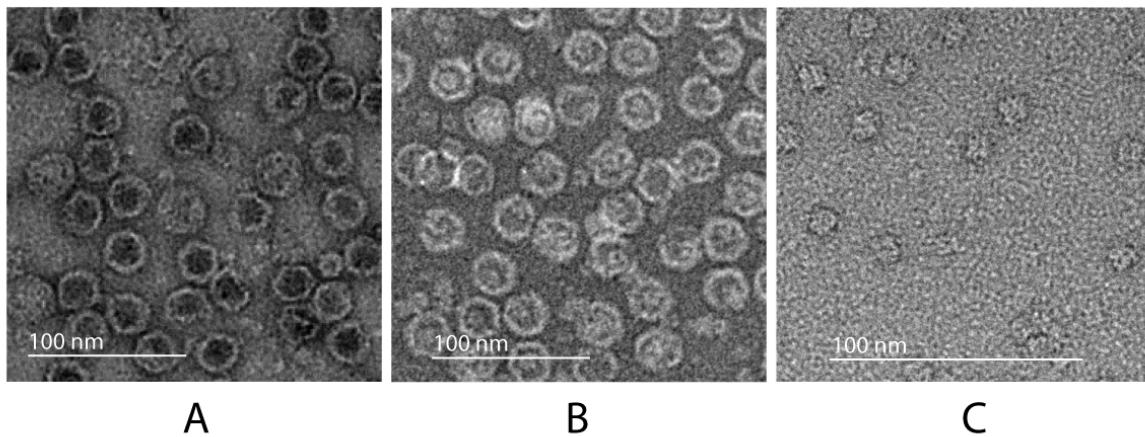
Figure 39: EM of uranyl acetate stained DyP/encapsulin samples. A: Negative stained sucrose gradient fraction containing encapsulin B: Negative stained sucrose gradient fraction containing encapsulin and co-migrated DyP. The diameter of the encapsulin particles is identical to those of the *T. maritima* particles. Inside the particles small particles are visible in most of the shells which are not present in the recombinant encapsulin alone. C: purified DyP sample from gel filtration showing particles with 3-fold and apparent 4-fold symmetry axes.

3.3.5 3D Reconstruction of *B. linens* DyP

We then proceeded to calculate a three-dimensional reconstruction of the BlDyP using negative stain electron microscopy. A total of 2878 single particle images of were picked, CTF corrected and used for a "reference-free" alignment procedure. The images were then subjected to a multivariate statistical analysis and classification. The class averages were used for 3D structure determination by the angular reconstitution approach. In the refinement cycles D3 point group symmetry was applied to the three-dimensional reconstruction. The resolution of the structure calculated by the Fourier shell correlation (FSC) function using the 0.5 criterion was found to be 18 Å. The monomers of a related peroxidase from *Bacteroides thetaiotaomicron* could be fitted into the reconstructed density revealing a trimer of dimers with D3 symmetry (Figure 40), which is the same oligomeric arrangement as observed for this peroxidase in the crystal (Zubieta, Krishna et al. 2007). The only difference is some

additional density at the C-terminus which can be accounted for by the extension of BlDyP (Figure 40, red/gray dotted arrows).
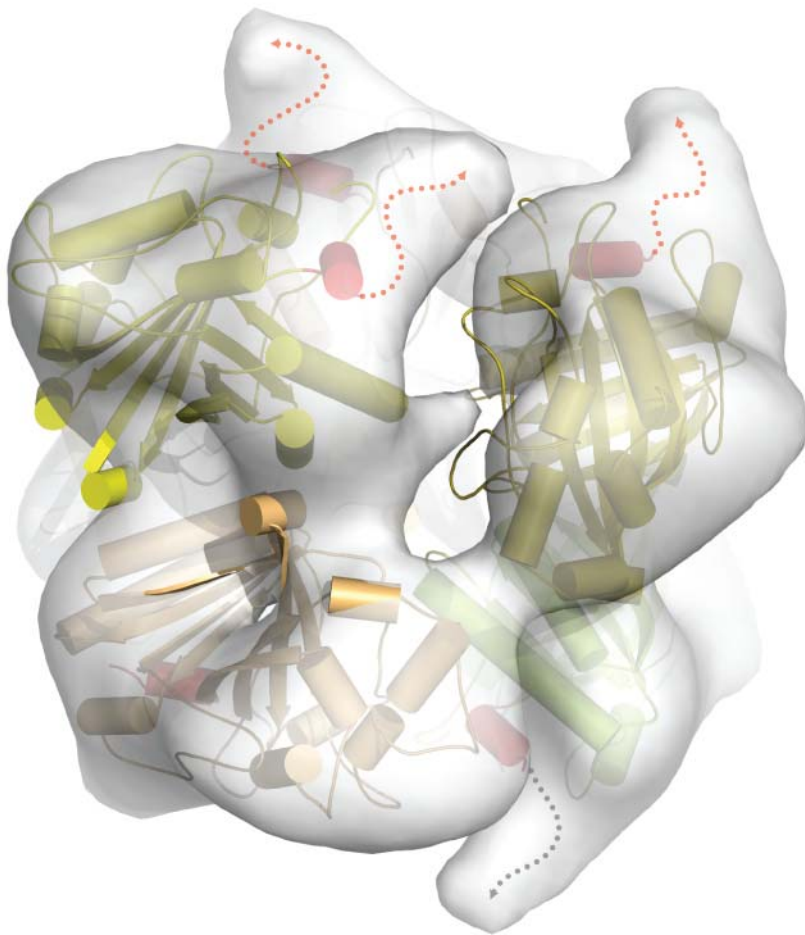
Figure 40: The EM structure of *B. linens* DyP. Fit of *B. thetaiotaomicron* monomers into the reconstructed BlDyP density. The C-terminal 10 residues are marked in red. A dotted line indicates the predicted continuation of the protein in the BlDyP homologue.

3.3.6 Coexpression of encapsulin with GFP / GFP_Linker

The construct which contains the whole DyP / encapsulin operon was modified by Quickchange mutagenesis to include an XmaI site in the linker region of DyP to replace it by hisGFP with the linker region in frame or by hisGFP followed by two stop codons (Figure 41A).
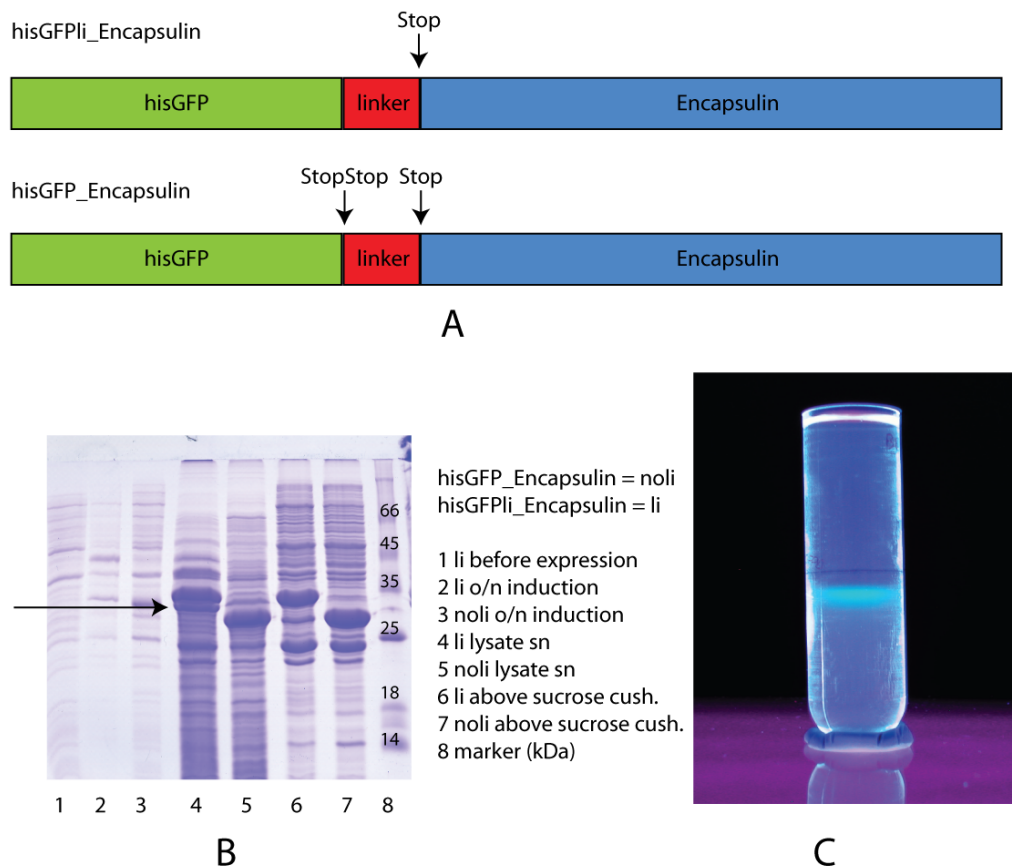
Figure 41: Expression of GFP/encapsulin constructs. A: hisGFP-encapsulin constructs B: 15 % SDS-PAGE of recombinantely expressed GFP/encapsulin constructs. The arrow indicates the position of encapsulin which is much higher expressed in the construct with the GFP-linker fusion (lane 4 compared to lane 5). Molecular weights: hisGFPli: 34.9 kDa, hisGFP: 29.4 kDa, encapsulin: 28.6 kDa. C: Co-migration of GFP within the encapsulin shell as shown in a 10-50 % sucrose gradient.

Expression levels for hisGFP with or without linker fusion were very high (large bands visible in Figure 41C, lanes 6 and 7) but the hisGFP_encapsulin construct produced much lower levels of encapsulin (lane 5, Figure 41B). After purification through a sucrose cushion and gradient the amount of GFP which co-migrated with encapsulin could be assessed. The fluorescence of GFP was measured both in a fluorimeter and qualitatively by excitation with UV light while the sample was still in the sucrose gradient tube (Figure 41C). Comparison of the fluorescence between hisGFPli_encapsulin and hisGFP_encapsulin showed no difference in co-purification with encapsulin when equal amounts of protein were applied.

3.3.7 Binding motif mutants of *B. linens* DyP

We constructed two mutants of the DyP peroxidase which are either lacking the last 25 amino acids or have a mutation in the binding motif (Figure 42). The mutants were expressed in *E.*

coli and subjected to the standard encapsulin preparation. Additionally, the DyPs were also purified from the fraction which did not enter the sucrose cushion.

DyP peroxidase which lacks the binding motif does not specifically co-migrate with encapsulin (Figure 42A, lane 2) despite of the huge excess of the peroxidase in the cellular lysate (Figure 42A lane 6). The particles were also analyzed under EM and they looked like the empty particles. The introduction of a stop codon in the extension seems to have an influence on expression of encapsulin which is very low (Figure 42A lane 6). The higher expression of encapsulin in the context of the non-truncated DyP could be due to ribosomes restarting translation instead of dissociating since the stop codon of DyP and start codon of encapsulin are so close together (see also Figure 29B).

The most conserved amino acid of the binding motif, isoleucine 257 was mutated to a large and hydrophilic arginine in order to disrupt the binding. The mutated peroxidase runs slightly higher in SDS-PAGE (Figure 42B lane 3) and is substantially less incorporated into encapsulin (Figure 42B lane 3).
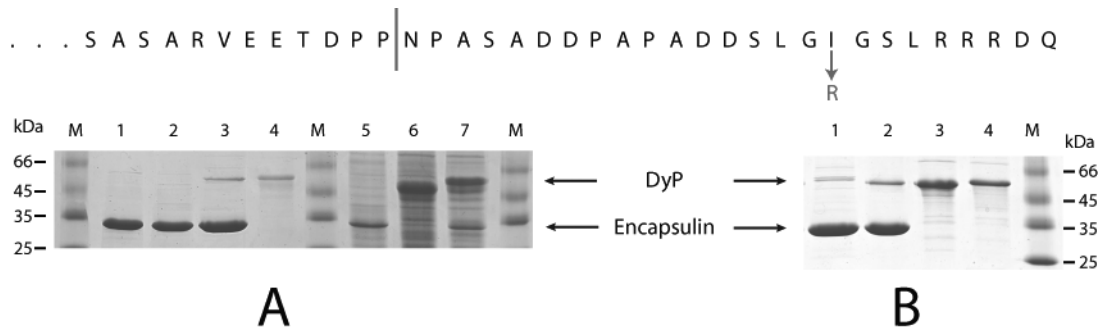
Figure 42: Polyacrylamide gels of *B. linens* DyP/encapsulin mutants. Above the gels is the sequence are the last 37 residues of BlDyP with the mutations indicated with a bar and an arrow (stop codon and I357R). A: Lanes 1-4 correspond to encapsulin alone, encapsulin/DyP lacking C-terminal extension, encapsulin/DyP and DyP alone. Lanes 5-7 are the cell lysates from which 1-3 are purified. The expected migration of DyP and encapsulin are marked with arrows. B: Lanes 2 and 4 correspond to the same as in A. Lane 1 is the co-purification of I357R mutant of DyP peroxidase with encapsulin, Lane 3 the DyP I357R mutant alone.

3.3.8 Activity assays of DyP peroxidase

Peroxidase activity can be measured with typical substrates such as guaiacol (Figure 43A) which reacts to a larger product with a distinct absorption at 470 nm (Tonami, Uyama et al. 2004). The reaction of DyP was fastest at a pH of 4 but overall the activity was quite low. DyP alone and DyP packaged in encapsulin showed comparable activities of 3.85 and 3.03 per minute per DyP, respectively (Figure 43B). Apparently guaiacol is not an optimal substrate for these peroxidases.
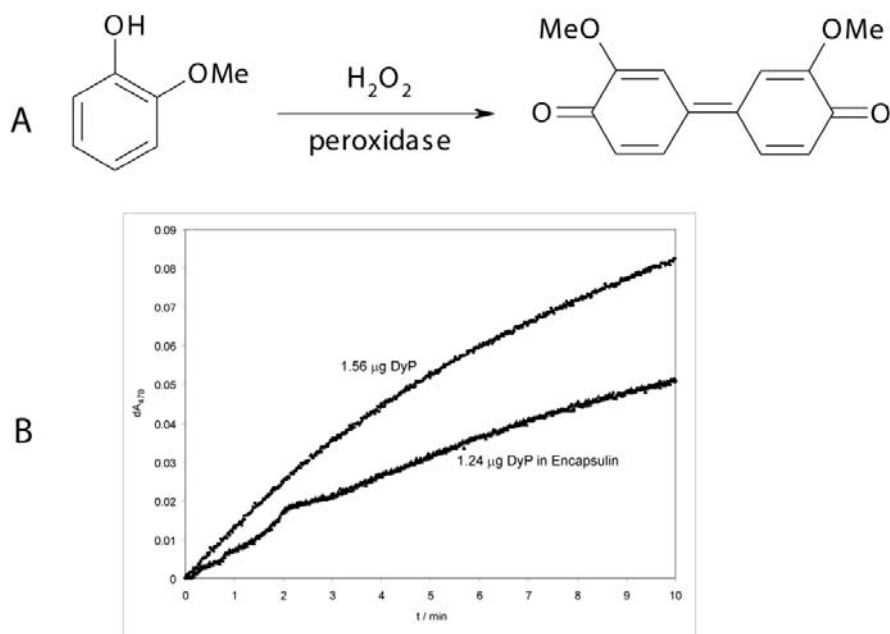
Figure 43: Peroxidase Assay with DyP. A: The reaction of guaiacol to the main product 3,3'-dimethoxy-4,4'-biphenoquinone. B: The reaction is followed at the absorption maximum of the product at 470 nm. DyP and DyP packaged in encapsulin both catalyze the reaction. The amounts of DyP present in the reaction are estimated from the characteristic absorption band of the heme at 407 nm.

3.3.9 Bactericidal activity of *B. linens* M18 does not coincide with the encapsulin particle

Anti-listerial activity of both recombinant and native encapsulin from *B. linens* M18 (Linocin M18) was tested on the most susceptible strain *Listeria ivanovii*. Its growth was inhibited by culture supernatant extracts of *B. linens* but showed no inhibition by recombinantely purified encapsulin or encapsulin/DyP complex (data not shown). A slightly modified centrifugation protocol could separate the anti-listerial activity from the main occurrence of the encapsulin shell protein (Figure 44). Bactericidal activity was assayed by the spot-on the lawn method (Barefoot and Klaenhammer 1983) using dilution series. The activity is given in activity units per ml (au/ml) which represents the highest dilution at which the sample is still active (Figure 44C). After ultrafiltration of the culture supernatant the small proteins were removed by ultracentrifugation through a 38 % sucrose cushion. Most of the activity went into the sucrose cushion but did not pellet (Figure 44A). Further separation on a 10-50 % sucrose gradient could localize the most active fraction between 21-27 % sucrose (Figure 44A). Encapsulin/linocin M18 would pellet in the ultracentrifugation step and migrate at 34-38 % sucrose (Figure 44B). The fact that purification of encapsulin does not coincide with enrichment of anti-listerial activity would also explain the very low yield of total activity of 0.25 % compared to the starting material which was reported previously (Valdes-Stauber and

61

Scherer 1994) and the existence of organisms which show anti-listerial activity without the presence of the *lin* gene (Valdes-Stauber and Scherer 1996).
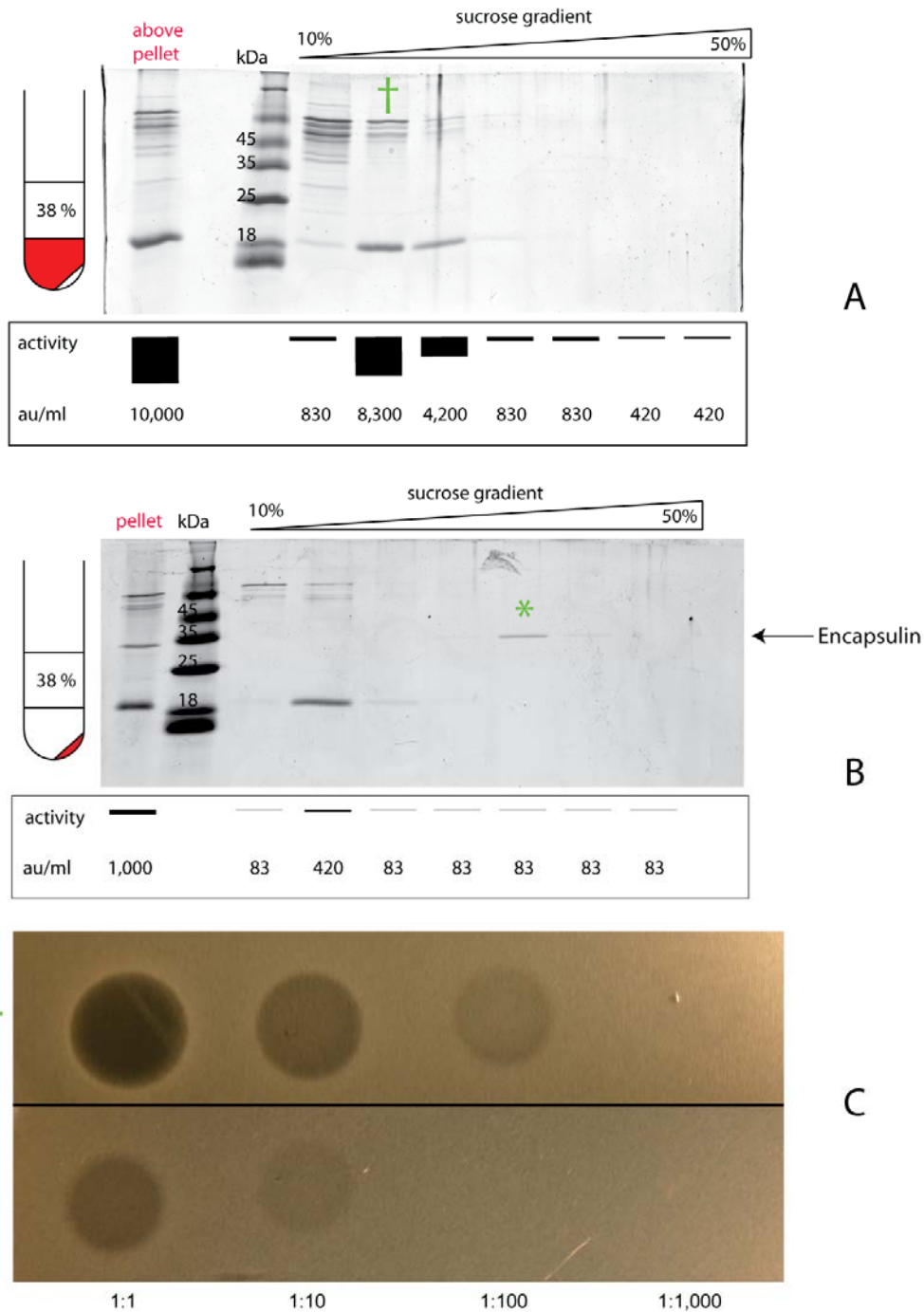
Figure 44: Native purification of encapsulin (linocin M18) from *B. linens* M18 culture supernatant and activity assays. A: The fractions above the pellet from the first ultracentrifugation step were further purified on a sucrose gradient and analyzed for activity and containing proteins (the strong band at low molecular weight is not responsible for bactericidal activity when the sample is further purified (data not shown)). B: Same as A, but the washed pellet after ultracentrifugation was taken instead. The marked encapsulin/linocin M18 band was identified by MALDI-MS. C: Serial dilution assays with the sucrose gradient fractions. The fraction with contains encapsulin/linocin M18 (*) only showed background activity while the peak of activity was located in a non-pelleting fraction (†).

4 DISCUSSION

Here we present the structure of *T. maritima* encapsulin together with biochemical and electron microscopic evidence for a cellular compartment that encapsulates enzymes. Encapsulins are newly characterized bacterial nanocompartments with functional similarities to bacterial microcompartments. The packaged enzymes are targeted to the interior of encapsulins via a specific peptide tag located on their C-terminus (Figure 45).
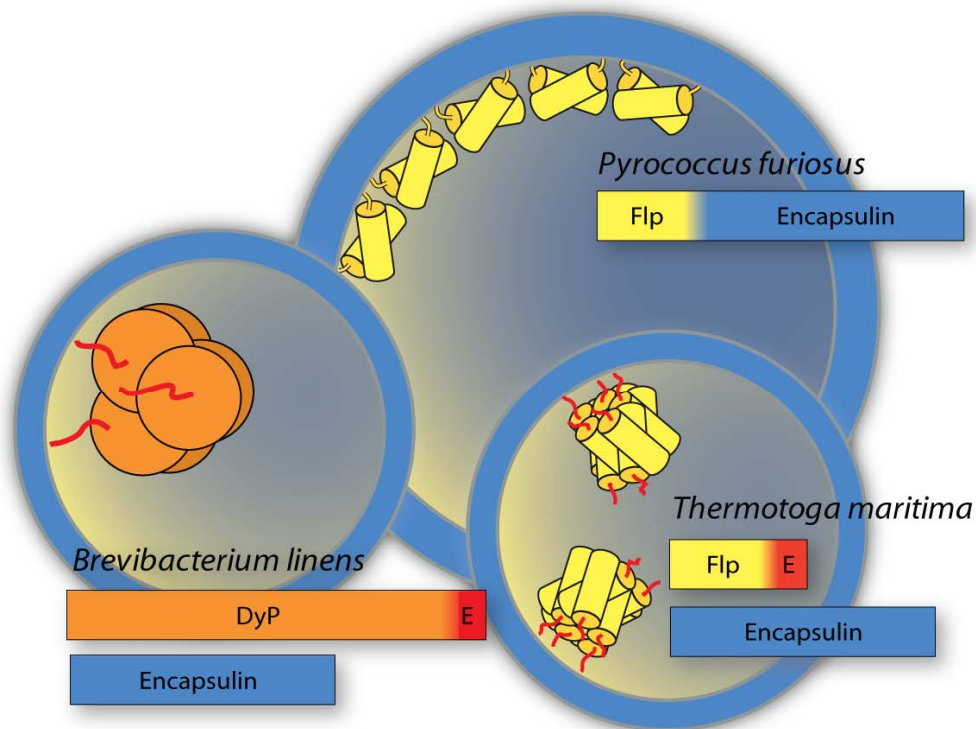
Figure 45: Schematic architectural and primary structure organization of the different types of encapsulins. Depicted as blue spheres are the encapsulins from *P. furiosus*, *B. linens* and *T. maritima* with their respective cargo proteins; their primary structures are depicted as bars. The extension sequence responsible for binding to the inside of encapsulin is marked in red.

These conclusions are supported by several lines of biochemical and structural evidence. I) A conserved binding site on the interior of the *T. maritima* encapsulin was identified. II) Additional electron density in the groove of the binding site in the interior of the particle is observed. III) The sequence of the C-terminal extension of the putative ferritin-like protein which is in the same operon as *T. maritima* encapsulin explains the additional electron density well and it can be identified in dissolved crystals of encapsulin by mass spectrometry. IV) Iron accumulates within the *T. maritima* encapsulin particles, as would be the case for

ferritins. This is indicated by the brownish color of the protein sample as well as an X-ray fluorescence scan on the crystal. V) Co-expression of the *B. linens* encapsulin with DyP containing the conserved C-terminal anchor sequence results in packaging of the peroxidase inside the encapsulin shell as shown by co-migration in a sucrose gradient and by negative stain electron microscopy. No encapsulation is observed when peroxidase without the packaging sequence is co-expressed. VI) Finally, in the archaeon *P. furiosus*, the genes for encapsulin and the ferritin-like protein are fused to produce one single polypeptide chain.

The *B. linens* M18 encapsulin was previously reported to act as a bacteriocin and therefore referred to as Linocin M18. This proposed function is unlikely, since the recombinantly expressed shell does not show bactericidal activity. Increasing the purity of the native *B. linens* M18 encapsulin does not correlate with an increase of bactericidal activity. It remains to be clarified which molecule in the culture supernatant of *Brevibacterium* species is responsible for the bactericidal activity. The *T. maritima* encapsulin was classified as a protease according to the measurement of peptidolytic activity towards chymotrypsin and trypsin peptide substrates(Hicks, Rinker et al. 1998). However, no structural features that would potentially constitute a protease active site could be identified in the structure.

Since both the packaged peroxidase and the packaged ferritin form oligomers they are likely to form multiple interactions with the interior of the encapsulin based on the symmetry match between the C-terminal extensions and the distribution of binding sites in the interior of the encapsulin (Figure 46B/C). The multiple binding contacts will increase the avidity of the interaction and the specificity of packaging. Consequently, efficient packaging would occur even if the affinity of individual interactions is not very high. Binding of the multimeric enzymes to the encapsulin subunits may also help in assembling the icosahedral shell but this interaction can not be critical, since empty encapsulins can be observed in electron micrographs.
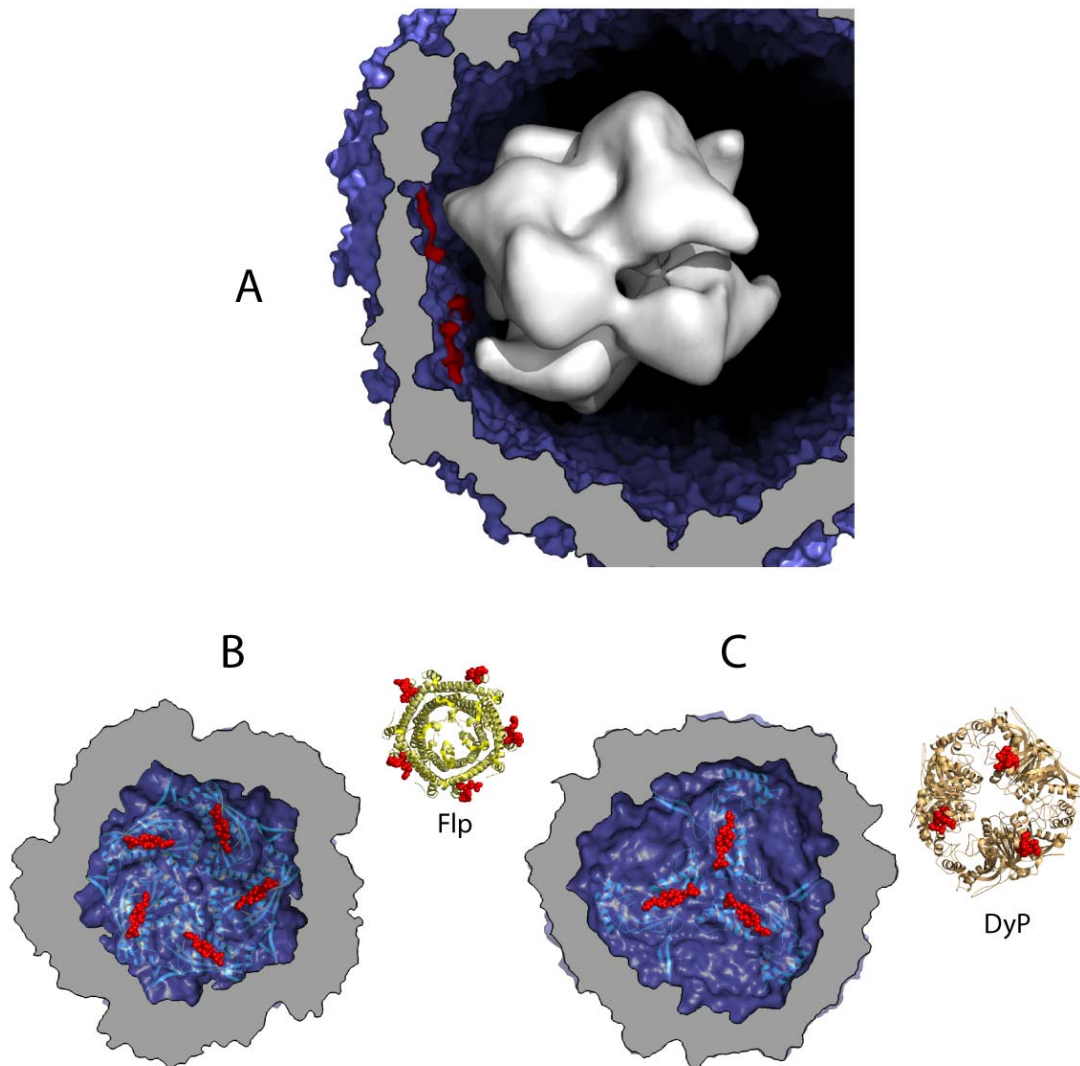
Figure 46: Packaging of oligomeric cargo proteins. A: Model of the BlDyP hexamer fitted into the *T. maritima* encapsulin shell with the bound peptide extensions shown in red. B/C: The binding sites on the inside of *T. maritima* encapsulin match the distribution of the C-terminal extensions of Flp and DyP proteins (the last 10 amino acids of the encapsulated protein and the bound peptide extensions are shown as red spheres).

Considering the size of the D3 hexamer of DyP, 240 kDa and dimensions of about 100 Å in each direction, it is unlikely that more than one DyP complex could be accommodated within the 220 Å diameter encapsulin shell (Figure 46A). Consequently, most filled encapsulin particles visualized by negative stain electron microscopy suggest that one peroxidase complex is encapsulated resulting in a 10:1 stoichiometry of the encapsulin to the packaged peroxidase (60 encapsulin monomers build the shell packaging one peroxidase hexamer). The ferritin-like pentamers of dimers packaged through interactions with the pentamers of *T. maritima* encapsulin are smaller and could in principle fully occupy the interior of the encapsulin by binding to all 12 pentamers. High occupancy of interactions is also likely based

on our crystallographic data, which revealed clear density for the anchor sequence of the encapsulated ferritin-like protein. The stoichiometry of the fused archaeal protein is obviously 1:1 but it is possible that a functional ferritin-like molecule is formed through dimeric interactions of the ferritin-like domains in the interior of the T=3 symmetric archaeal particle built out of 180 subunits (Figure 45).

The two different packaged proteins are a peroxidase and a protein which has the same fold as ferritins and contains the ferroxidase active site. The reactive centers of the ferritin-like protein packaged within the particle could oxidize the reactive ferrous ions, which are then retained in the huge interior of the particle in a mineral ferrihydrate form. The size of the particle (220 Å inner diameter compared to 80 Å for ferritin) makes it possible to store vast amounts of iron. The pores in the shell of encapsulin resemble the ones in ferritin, where also conserved charged and hydrophobic channels are observed (Takahashi and Kuyucak 2003). Interestingly, since ferrous ions can react with peroxide to generate very toxic hydroxyl radicals (the Fenton reaction), removing excess ferrous ions also protects the organism from adverse effects of metabolically produced peroxide (Harrison and Arosio 1996). The presence of these ferritin-like proteins in anaerobic bacteria, where due to the absence of oxygen iron toxicity is usually not a problem, can be explained as a protective mechanism in case of oxygen exposure (Rocha, Andrews et al. 1992). This might be especially important for these organisms as they generally have low levels of superoxide dismutase and catalase (Finegold and George 1989). Similarly, in pathogenic bacteria like *Mycobacteria* a peroxidase packaged within the encapsulin shell could provide resistance to the oxidative killing mechanisms of the host. It has been shown that overexpression of a peroxidase can increase viability of certain *M. tuberculosis* strains in human monocytes (Manca, Paul et al. 1999).

While encapsulins are conserved among each other, there is no significant homology to structural virus proteins on the level of amino acids, which explains why this similarity was not obvious from sequence analysis. Nevertheless, it is possible that this protein originated from a viral capsid but this must have happened very early in evolution considering that no other genes of viral origin are found in the vicinity of encapsulin genes and the fact that encapsulins are found in bacteria and archaea from many different environments. This would support the hypothesis of a common viral ancestor that was present before the host organisms diverged (Bamford, Grimes et al. 2005), a possible candidate being the HK97-like virus ancestor, which has been previously proposed to have emerged before the separation of bacterial and archaeal kingdoms (Bamford, Grimes et al. 2005). Since proteins in a genome are not subjected to evolutionary change in the same way as viruses are, encapsulins

potentially provide a unique and exciting view into a minimal ancestor of the large and widespread HK97 lineage of viruses. Although encapsulins are most likely of viral origin, their herein described cellular function illustrates that viruses could have originated from a similar cellular assembly by switching its specificity from encapsulating proteins to encapsulation of nucleic acids.

When encapsulin is compared to bacterial microcompartments such as the carboxysome, several parallels become evident, although the monomers differ significantly. The recently published model of the carboxysome shows a large icosahedral body formed by hexamers with a different protein at the pentamer vertices (Tanaka, Kerfeld et al. 2008). The T=1 structure of *T. maritima* encapsulin is built from pentamers only, while in the T=3 homologue from *P. furiosus* the same protein forms both pentamers and hexamers. This flexibility of the encapsulin monomer structure may allow formation of even larger icosahedral particles. The substrates and products of bacterial microcompartments involve mainly small metabolites for which access to the microcompartment is limited by selective pores at the axis of symmetry (Tsai, Sawaya et al. 2007; Yeates, Tsai et al. 2007). Pores of similar dimensions are observed in the encapsulin. Furthermore, the sequestration of iron in a mineral form in the interior of the encapsulin can be compared to the proposed function of *pdu* organelles (with homology to the carboxysome microcompartment shell proteins) in sequestering propionaldehyde to reduce cytotoxicity (Havemann, Sampson et al. 2002). In light of these parallels, the molecular basis of enzyme packaging in other bacterial microcompartments may rely on similar principles as described here for the encapsulin.

Indeed, the β- and γ-subunits of the diol dehydratase which is packaged into the *pdu* microcompartments have flexible N-terminal regions which are not necessary for enzymatic function (Tobimatsu, Kawata et al. 2005). These display a similar pattern of a conserved sequence with a few hydrophobic residues followed by a less conserved linker region (Figure 47).
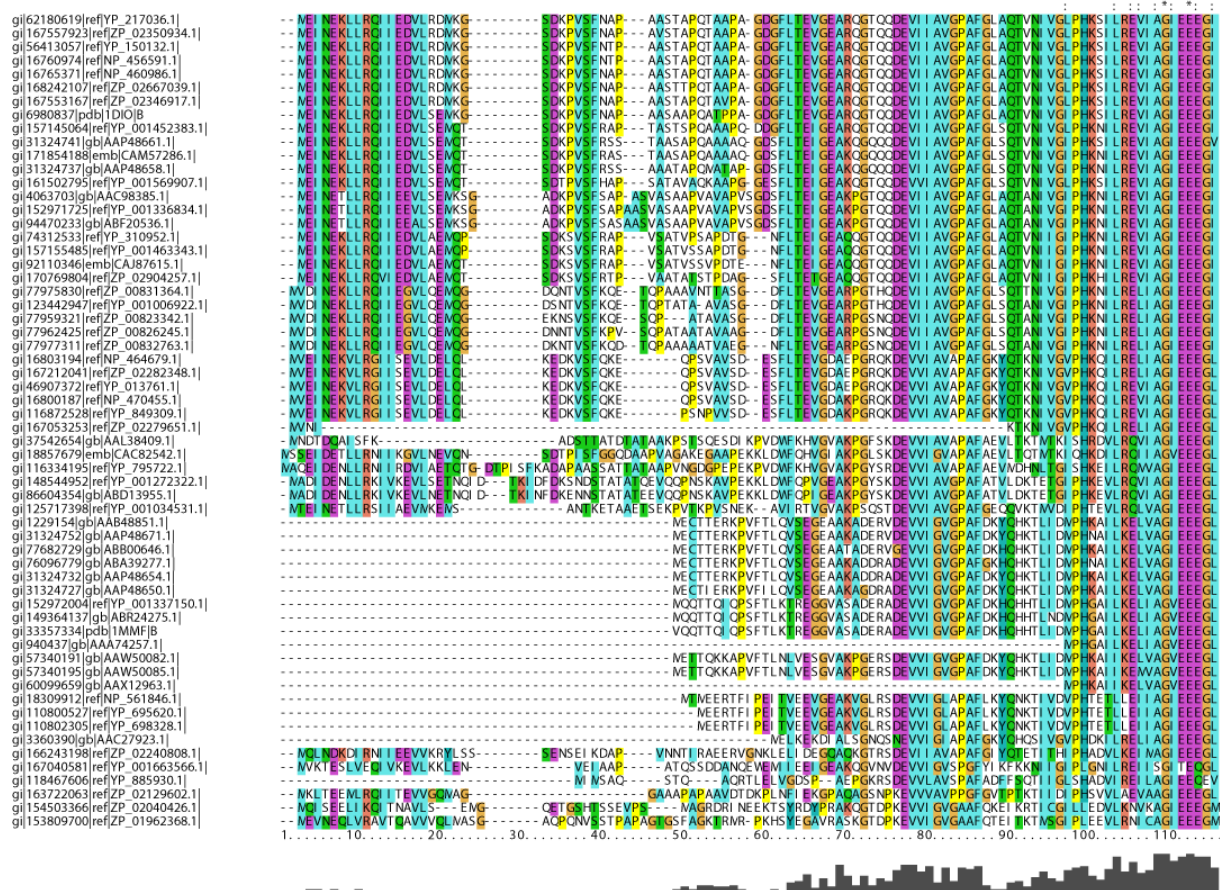
Figure 47: Alignment of the first part of the β subunit of diol dehydratases. The proteins found in the context of carboxysome-like microcompartments have an N-terminal extension. Coloring is according to physicochemical properties of the residues.

Packaging of enzymes into the encapsulin shell could have a protective role for the enzymes it carries. Due to the extreme structural stability over a wide range of temperatures and pH values, which was observed for the *T. maritima* and *B. linens* encapsulins (Stauber 1995), the shell would extend the life time of the enzyme considerably. Encapsulation does not prevent the enzymes from functioning since small molecular substrates such as iron ions or hydrogen peroxide can easily penetrate the shell. The pores in the shell even have the potential to modulate the substrate spectrum of the enzymes by imposing a size restriction on the molecules capable of entering the compartment. Small substrates such as guaiacol seem to penetrate the shell as no significant difference between free and encapsulated DyP was observed (Figure 43B).

The specific targeting sequence could be of use in biotechnological applications to package various proteins or enzymes inside the stable self-assembled icosahedral shell of encapsulin, to either protect the enzyme from proteolytic degradation or the cell from a toxic enzyme. With these applications in mind an icosahedral enzyme complex, lumazine synthase, was

engineered to encapsulate target molecules by means of charge complementarity (Seebeck, Woycechowsky et al. 2006).

Many more bacterial compartments might exist in nature but they are difficult to detect at the sequence level due to the diversity of building blocks that can form molecular shells. A combination of structural biology techniques such as X-ray crystallography and electron microscopy together with genomics/proteomics and recombinant expression methods provides a powerful approach to study such systems.

5 OUTLOOK

This work opens up many possibilities for further experiments regarding the *in vivo* roles of encapsulins and its cargo molecules. It would be interesting to study the growth of bacteria lacking the encapsulin gene or the packaged enzyme under various environmental conditions. Decreased growth of the knockout strains would provide hints that encapsulins are important in oxidative stress response. Alternatively, antibodies could be useful to test the effect of oxidative stress on the expression level of encapsulin *in vivo*.

The location of encapsulins in the cell is still an open question. While some encapsulins were found in the culture supernatant we have purified native *T. maritima* and recombinant *B. linens* encapsulin from the cytosol. Protein found in the culture supernatant could originate from lysed cells and encapsulin would be one of the few proteins stable against proteolytic digestion. Antibodies derived against encapsulins from various species and testing of the different cellular fractions could answer the localization question.

Another interesting aspect is the effect that introducing a stop codon in the extension sequence leads to significantly lower expression of encapsulin (Figure 42B). This might be due to some translational effect where the ribosome does not dissociate from the mRNA but reinitiates translation. Cloning the two proteins in a pETDuet vector or different vectors should lead to similar expression levels for the two proteins.

While the binding motif is very conserved in most species, the one found in of *T. maritima* Flp shows some deviations (Figure 29C). Mutational analysis of the more representative *B. linens* DyP could be useful to study the exact mode of binding in this organism. A tighter interaction is desired for packaging cargo which has only one binding peptide. The crystal structure of recombinant *B. linens* encapsulin would enable a more rational approach to mutation if the binding peptide is visible. This would have to be done with a different cargo than DyP though since only 3 bound sites out of 60 are unlikely to be seen in electron density maps. Maybe a chimeric protein of *T. maritima* Flp with the binding motif from *B. linens* DyP would bind efficiently on the inside of *B. linens* encapsulin.

Virus-like particles have been used in medical applications for gene delivery, bio-imaging, drug encapsulation and vaccine development (Lee and Wang 2006). Most of these particles were chemically modified on the inside or outside to fulfill this role. Encapsulins have an additional mode of binding protein cargo non-covalently on the inside. Moreover, encapsulins are easy to produce recombinantely in large amounts and are extermely stable which make them very interesting for biotechnological applications.

6 APPENDIX

6.1 Primers and plasmids

6.1.1 Primers for *Brevibacterium linens* cloning

Blin001   `5'- GAGC CATATG AATAACCTCTATCGCGAGCTTG -3'`
Blin002   `5'- GCAC GAATTC TCATCAGAGGCTCAAAGGCACAC -3'`
BL encapsulin with 5' NdeI site and 3' EcoRI site

Blin007b   `5'- GAGC CATATG GCAGATCAGTACCACGAACCAGTTTC -3'`
Blin008   `5'- GCAC GAATTC GATAGAGGTTATTCACTT -3'`
BlDyP with 5' NdeI site and 3' EcoRI site

Blin0016   `5'- CAGGATTTCCTCGACGATCCCGGGGGAGAGCTCGCCGCCGCCGAG -3'`
Blin0017a   `5'- CTCGGCGGCGGCGAGCTCTCCCCCGGGATCGTCGAGGAAATCCTG -3'`
Quickchange primers to introduce XmaI site into linker region of BlDyP (D311G)

Blin013   `5'- TCTAGAATGGGCCATCATCATCATCAT -3'`
Blin015   `5'- GATAACAATTCCCCTCTAGA -3'`
hisGFP with 5' XbaI site and 3' XmaI site from pET16bhisGFPssrA

Blin020   `5'- CCCGGGCTTATTATTTGTATAGTTCATCCATGCCAT -3'`
hisGFP 3' XmaI with 2x stop codon

Blin024   `5'- CCCGGGTTATCAATTGGGTGGGTCTGTTTCCT -3'`
BlDyP XmaI rev, 2xstop after N341 (P342Stop)

Blin026a   `5'- GACGACAGCCTCGGCCGCGGCAGCCTGAGAAGGAGAGAG-3'`
Blin026b   `5'- CTCTCTCCTTCTCAGGCTGCCGCGGCCGAGGCTGTCGTC-3'`
Quickchange primers to introduce R in binding motif (I357R)

6.1.2 DNA and Protein Sequence of *B. linens* M18 DyP

DNA sequence

```
  1    ATGGCACTTC CGAACGGAAA GACGCCGCAG CATGTGCTGG GTCCGCCGGC TCCGGCCGCC
 61    GTGTTCCTCG TCCTCACCGT GAGATCCGGC GCCGAGGCGG AGGCCAAGGA CTTCCTCGGC
121    GACATTGCGG GGGTGGTCAG GTCCGTGGGC TTTCGCGCTC GCGAGGATCA CCTCAGCTGC
181    GTGACCGGCA TCGGCGCCGA GCTCTGGGAC CGGATGTTCG ACGCCCCGCG CCCGGCCGGG
241    CTGCACCCCT TCATCGAACA GCGCGGAGAC GTCCACACCG CGCCGTCGAC CCCCGGTGAT
301    CTGCTCTTCC ACATCCGGGC CCGGCGGATG GACCTCTGCT TCGAACTCGC CCGCCAGCTC
361    GTCGGCGAAC TCGGTGACGC CGTGTCCGTC GTCGACGAGG TGCACGGTTT CCGCTACTTC
421    GACGAACGCG ACATCATGGG CTTCGTCGAC GGCACCGAGA ACCCCGAGGA TCAGGAGGCC
481    GTGGACTCGG TGTTCACGCC GACCGGTGGA GATGACCCGG CGAGCAGCAC CTATGTGATC
541    GTGCAGAAGT ATACCCACGA CATGGCCGCA TGGGAGGCGC TGAGCGTCGA GGACCAGGAG
601    GCGGCCTTCG GTCGGCACAA ACTCTCCGAT ATGGAGTTCC CCGACGAGGA CAAGGCGCCG
661    AACTCCCATC TCATCCTCAA CACGATCGAG GACGAAGACG GGACCGAGCA CAAGATTGTG
```

```
 721        CGTGACAATA  TGGTCTTCGG  ATCGGTCGAG  TCCGGCGAGT  TCGGGACCTA  CTTCATCGGC
 781        TACGCCGCCG  ACGTCAGCGT  GACCGAACAG  ATGCTCGAGA  ACATGTTCAT  CGGCAACCCC
 841        CGCGGCACCT  ACGATCGGAT  CCTCGACTTC  TCGACCGCGC  AGACCGGCGG  CCTCTTCTTC
 901        GTCCCCTCCC  AGGATTTCCT  CGACGATCCC  GACGGAGAGC  TCGCCGCCGC  CGAGCCATCC
 961        GATGCGCAGA  ACGACGACCC  GGCCTCCGCC  TCGGCGAGGG  TCGAGGAAAC  AGACCCACCC
1021        AATCCCGCAT  CAGCCGACGA  CCCCGCACCG  GCCGACGACA  GCCTCGGCAT  CGGCAGCCTG
1081        AGAAGGAGAG  ACCAGTGA
```

Protein sequence

```
  1        MALPNGKTPQ  HVLGPPAPAA  VFLVLTVRSG  AEAEAKDFLG  DIAGVVRSVG  FRAREDHLSC
 61        VTGIGAELWD  RMFDAPRPAG  LHPFIEQRGD  VHTAPSTPGD  LLFHIRARRM  DLCFELARQL
121        VGELGDAVSV  VDEVHGFRYF  DERDIMGFVD  GTENPEDQEA  VDSVFTPTGG  DDPASSTYVI
181        VQKYTHDMAA  WEALSVEDQE  AAFGRHKLSD  MEFPDEDKAP  NSHLILNTIE  DEDGTEHKIV
241        RDNMVFGSVE  SGEFGTYFIG  YAADVSVTEQ  MLENMFIGNP  RGTYDRILDF  STAQTGGLFF
301        VPSQDFLDDP  **D**GELAAAEPS  DAQNDDPASA  SARVEETDPP  N**P**ASADDPAP  ADDSLG**I**GSL
361        RRRDQ
```

6.1.3 Plasmids
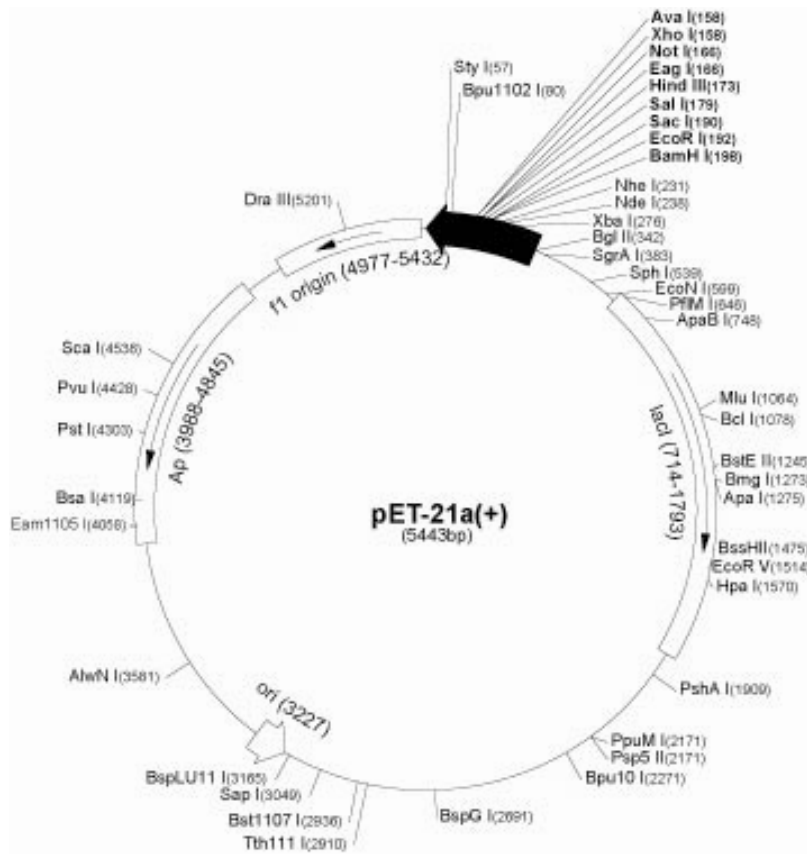
Figure 48: pET21a vector map (Novagen)
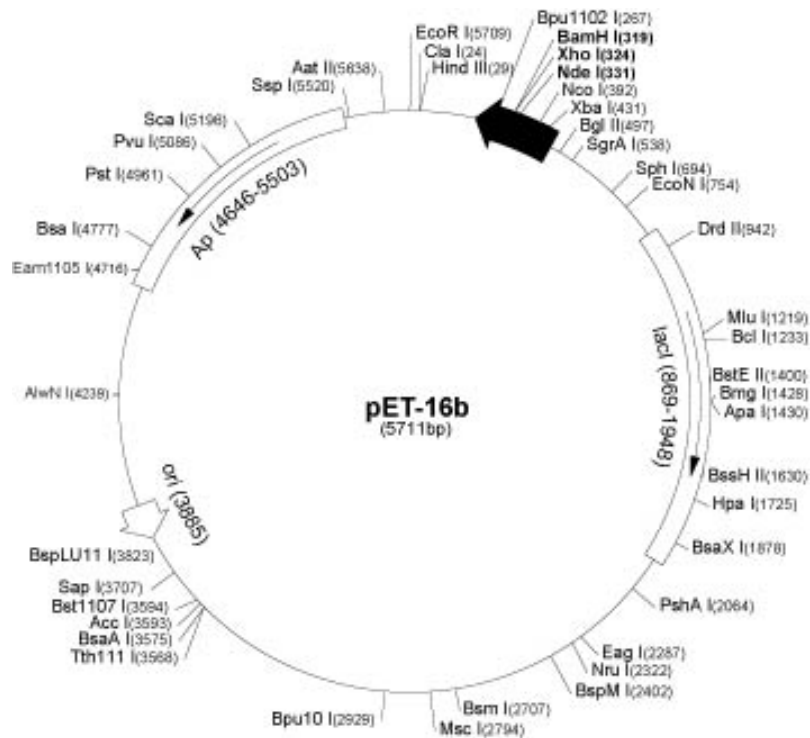
Figure 49: pET16b vector map (Novagen)
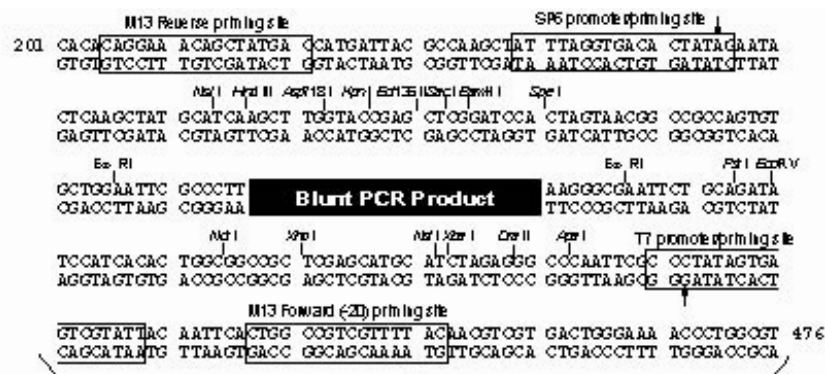
Figure 50: pCR-Blunt II-TOPO vector map (Invitrogen)

6.2 Media and buffers

All media recipes refer to a volume of 1 L.

PCB (agar)

  5.0 g peptone

  2.5 g yeast extract

  1.0 g glucose

  30.0 g NaCl

  (13.0 g agar)

LB (agar)

  10.0 g tryptone

  5.0 g yeast extract

  5.0 g NaCl

  (13.0 g agar)

BHI (agar)

  12.5 g brain infusion solids

  5.0 g beef heart infusion solids

  10.0 g peptocomplex

  2.0 g glucose

  5.0 g NaCl

  2.5 g Na_2HPO_4

  (8.0 g agar)

buffer A / final buffer (FB)

  20.0 mM Tris pH 7.4

  20.0 mM $MgCl_2$

  150.0 mM NH_4Cl

  6.0 mM β-me

6.3 Computer scripts

6.3.1 Calculation of monomer grid

This pascal program creates the monomer grid (pdb file format) which is then used for mask creation.

input: none

output: monomer pdb in standard orientation

74

```pascal
program planes;
uses WinDos,WinCrt;

var a1,a2,a3,b1,b2,b3,c1,c2,c3,d1,d2,d3,n1,n2: Real;
    Aac,Bac,Cac,Abc,Bbc,Cbc,Aad,Bad,Cad,Abd,Bbd,Cbd,D: Real;
    absac,absbc,absad,absbd: Real;
    dbc,dac,dad,dbd: Real;
    x,y,z: Real;
    a,b,c:Integer;
    i,res: Integer;
    test: Boolean;
    r: Real;
    ser:Integer;
    outfile,s1,s2,s3:String;
    grid:Integer;
    sc:Integer;
    rmin,rmax:Real;

procedure CreatePDB(var filename:String);
var P:text;
begin
  Assign(P,filename);
  Rewrite(P);
  Close(P);
end;

procedure                                                        WritePDB(var
f:String;aser:Integer;aname:String;resname:String;cid:Char;rseq:Integer;x:Real;y:Real;z:Real;b
fac:Real);
var P:text;
begin
 Assign(P,f);
 Append(P);
 Write(P,'ATOM  ',aser:5,' ',aname:4,' ',resname,' ',cid,rseq:4,' ','    ',x:8:3,y:8:3,z:8:3);
 WriteLn(P,'  1.00',bfac:6:2,'        ','A   ',' C',' ');
 Close(P);
end;

begin
/* three vectors a,b and c*/
  n1:=(Sqrt(5)+1)*Sqrt(3)/4;   /*1.401258538*/
  n2:=Sqrt(3)/2;              /*0.866025403*/
  b1:=0;
  b2:=(n1+n2)/2;           /*(1.401+0.866025)/2;*/
  b3:=-n1/3;        /*(-1.401)/3;*/            /*threefold*/
  a1:=-n2;
  a2:=n1;
  a3:=0;                     /*fivefold*/
  c1:=-n2/2;                /*-0.866025/2;*/
  c2:=(n1+n2)/2;         /*1.401/2+0.866025/2;*/
  c3:=-n1/2;   /*-1.401/2;            */  /*twofold*/
  d1:=0;
  d2:=n1;
  d3:=0;                     /*twofold*/
/* plane equations: Ax+By+Cz+D=0  D always=0 (through origin)*/
  Aac:=c2*a3-c3*a2;
  Bac:=c3*a1-c1*a3;
  Cac:=c1*a2-c2*a1;
  Abc:=c2*b3-c3*b2;
  Bbc:=c3*b1-c1*b3;
  Cbc:=c1*b2-c2*b1;
  Aad:=a2*d3-a3*d2;
  Bad:=a3*d1-a1*d3;
  Cad:=a1*d2-a2*d2;
  Abd:=d2*b3-d3*b2;
  Bbd:=d3*b1-d1*b3;
  Cbd:=d1*b2-d2*b1;
  D:=0;
/*' distances point to plane*/
  ser:=0;
/* origin:  Inc(ser);WritePDB(outfile,ser,'CA ','ALA','A',1,0,0,0,30);*/
/*  absac:=Sqrt(Sqr(Aac)+Sqr(Bac)+Sqr(Cac));
  absbc:=Sqrt(Sqr(Abc)+Sqr(Bbc)+Sqr(Cbc));
  absad:=Sqrt(Sqr(Aad)+Sqr(Bad)+Sqr(Cad));
  absbd:=Sqrt(Sqr(Abd)+Sqr(Bbd)+Sqr(Cbd));*/
  grid:=100;
  sc:=10;
```

```
  rmin:=60;
  rmax:=120;
  Str(rmin:2:0,s1);
  Str(rmax:3:0,s2);
  Str(grid:2,s3);
  outfile:='z:\'+s1+s2+'g'+s3+'.pdb';
  CreatePDB(outfile);
  i:=0;
  res:=0;
  for a:=1 to grid do
  begin
    WriteLn(a/grid*100:3:0,'% done.');
    for b:=1 to grid do
      begin
      for c:=1 to grid do
        begin
        test:=true;
        x:=-0.001-rmax/10*1.5+2*(rmax/10*1.5)*a/grid;
        y:=-0.001-rmax/10*1.5+2*(rmax/10*1.5)*b/grid;
        z:=-0.001-rmax/10*1.5+2*(rmax/10*1.5)*c/grid;
        dac:=(Aac*x+Bac*y+Cac*z+D)/*/absac*/;
        dbc:=(Abc*x+Bbc*y+Cbc*z+D)/*/absbc*/;
        dad:=(Aad*x+Bad*y+Cad*z+D)/*/absad*/;
        dbd:=(Abd*x+Bbd*y+Cbd*z+D)/*/absbd*/;
        r:=Sqrt(Sqr(x)+Sqr(y)+Sqr(z));
        if r<rmin/10 then test:=false;
        if r>rmax/10 then test:=false;
        if dac<=0 then test:=false;
        if dbc>=0 then test:=false;
        if dbd<=0 then test:=false;
        if dad<=0 then test:=false;
        if test then
          begin
          Inc(i);
           if i=10 then
            begin
              Inc(res);
              i:=0;
            end;
           Inc(ser);
           WritePDB(outfile,ser,'CA ','ALA','A',res,-x*sc,y*sc,z*sc,30);
          end;
        end;
      end;
    end;
end.
```

6.3.2 Calculation of masks

This script generates the monomer and 10-mer mask in the correct position and outputs the relating symmetry operators in formats for the various programs.

input: monomer.pdb in standard orientation

output: rotated/translated monomer pdb, monomer mask, 10-mer mask, 10-mer pdb, DM style operators, CCP4 style operators, O format operators


```
#!/bin/csh -f
###### transform atoms according to icosahedral symmetry and make masks
###### followed by operator generation script
source /xray/setup/ccp4.set
source /xray/setup/uppsala.set
setenv CCP4_OPEN /xray/ccp4

####### input as $3 is the monomer pdb file in origin position

set inputdir = $1
set hklin = $inputdir/$2.mtz      #$inputdir/nativehkl.mtz
set unitcell = `mtzcat $hklin |grep -m 1 "Cell constants"|tail -c 45|head -c 7`
```

```
set log = $inputdir/make_mask.log
set tmp = $inputdir/tmp

if (-e $tmp) then
  rm -rf $tmp
endif
mkdir $tmp
set expand   = -10
######## mask generation files
set mask = monomer
set mask_in = $mask.o
set mask_out = $mask.b.o
set map_out = $maskb.ccp4

######## for operator generation
set inputfile = $inputdir/full.pdb
set outfile = $inputdir/operators.txt
set rawout = $inputdir/matrices.txt
set ccp4out = $inputdir/ncs.ccp4
set oout = $inputdir/ncs.o
echo "#######################  10-mer and mask generation #######################"
echo "input file:    "$3  "(monomer pdb file)"
echo "output files: "full.pdb
echo "             "full.msk
echo "             "full.o
echo "             "monomer.pdb
echo "             "monomer.msk
echo "             "monomer.o
echo "             "monomerb.o"   (monomer omask with border)"
echo "log file      "$log
echo "##############################################################################"

cd /$inputdir
if (-e full.pdb) then
rm full.pdb
endif
if (-e full.msk) then
rm full.msk
endif
if (-e full.o) then
rm full.o
endif
if (-e monomer.pdb) then
rm monomer.pdb
endif
if (-e monomer.msk) then
rm monomer.msk
endif
if (-e monomer.o) then
rm monomer.o
endif

set skipsecond = '/bin/false'
set skipmask = '/bin/false'
echo "CRYST1  "$unitcell"  "$unitcell"  "$unitcell"  90.00  90.00  90.00 F 41 3 2" > full.pdb
echo "CRYST1  "$unitcell"  "$unitcell"  "$unitcell"  90.00  90.00  90.00 F 41 3 2" > in.pdb
cat $3|grep ATOM>>in.pdb
cp in.pdb monomer.pdb

echo "generating pdb files..."
#transformation 1
pdbset XYZIN $inputdir/in.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >$log
rotate -
    1 0 0 -
    0 1 0 -
    0 0 1
chain A A
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
#transformation 2
rm tmp2.pdb
pdbset XYZIN $inputdir/in.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
   0.500000  -0.309017    0.809017 -
  -0.309017   0.809017    0.500000 -
  -0.809017  -0.500000    0.309017
```

```
chain A B
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
#transformation 3
rm tmp2.pdb
pdbset XYZIN $inputdir/in.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
  -0.309017  -0.809017    0.500000 -
  -0.809017   0.500000    0.309017 -
  -0.500000  -0.309017   -0.809017
chain A C
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
#transformation 4
rm tmp2.pdb
pdbset XYZIN $inputdir/in.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
  -0.309017  -0.809017   -0.500000 -
  -0.809017   0.500000   -0.309017 -
   0.500000   0.309017   -0.809017
chain A D
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
#transformation 5
rm tmp2.pdb
pdbset XYZIN $inputdir/in.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
   0.500000  -0.309017   -0.809017 -
  -0.309017   0.809017   -0.500000 -
   0.809017   0.500000    0.309017
chain A E
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
#transformation whole 5 subunits
rm tmp2.pdb
pdbset XYZIN $inputdir/full.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
  -0.809017  -0.500000   -0.309017 -
  -0.500000   0.309017    0.809017 -
  -0.309017   0.809017   -0.500000
chain A F
chain B G
chain C H
chain D I
chain E J
end
eof-1
cat tmp2.pdb|grep ATOM>>full.pdb
# translation whole molecule by 1/8,1/8,1/8 and rotation
rm tmp2.pdb
pdbset XYZIN $inputdir/full.pdb XYZOUT $inputdir/tmp2.pdb << eof-1 >>$log
rotate -
   -0.596359207  -0.707106781   0.3799417      -
    0.537318829        0         0.8433792      -      # green
   -0.596359207   0.707106791   0.3799417
#rotate -
#    0.15942644   -0.707106781  0.6889000      -
#    0.974251823        0        0.2254626      -      # red
#    0.15942644    0.707106781  0.6889000
shift -
   fractional -
   0.125  0.125  0.125
end
eof-1
mv tmp2.pdb full.pdb

#compiling full.pdb, adding END
mv full.pdb tmp1.pdb
echo "CRYST1 "$unitcell" "$unitcell" "$unitcell"  90.00  90.00  90.00 F 41 3 2" > full.pdb
cat tmp1.pdb|grep ATOM>>full.pdb
cat tmp2.pdb|grep ATOM>>full.pdb
echo END>>full.pdb
# generate monomer
```

78

```
echo "CRYST1   "$unitcell"   "$unitcell"   "$unitcell"   90.00   90.00   90.00 F 41 3 2" >
monomer.pdb
cat full.pdb | grep "ALA A">>monomer.pdb
echo END>>monomer.pdb
######### mark origin of mask and clean up

echo REMARK $3 >> monomer.pdb
echo REMARK $3 >> full.pdb
rm tmp1.pdb
rm in.pdb

echo "generating mask files..."
echo "full mask..."
########### generate masks

# generating mask of full 10-mer
ncsmask XYZIN $inputdir/full.pdb MSKOUT $inputdir/full.msk << eof-1>>$log
grid 400 400 400
radius 10.0
peak 2
smooth 3
average 1
rota euler 0 0 0
tran 0 0 0
expand `echo $expand`
eof-1

echo "monomer mask..."
# generating mask of monomer
ncsmask XYZIN $inputdir/monomer.pdb MSKOUT $inputdir/monomer.msk << eof-1>>$log
grid 400 400 400
radius 10.0
peak 1
smooth 3
average 1
rota euler 0 0 0
tran 0 0 0
expand `echo $expand`
eof-1
lx_mapman << eof-1>>$log
read m1 full.msk ccp4
wr m1 full.o mask 0.0010
read m2 monomer.msk ccp4
wr m2 monomer.o mask 0.0010
quit
eof-1

####################################################################
# makes border around monomer mask

if (-e $inputdir/$mask_out) then
rm $inputdir/$mask_out
endif
if (-e $inputdir/$map_out) then
rm $inputdir/$map_out
endif

#read input mask
lx_mama << eof > $tmp/mama.log
re m1 $inputdir/$mask_in
wr m1 $tmp/mask.o
list m1
quit
y
eof

# get parameters
set mask_cell=`cat $tmp/mama.log | grep "Cell    =" | cut -c11-80`
set mask_grid=`cat $tmp/mama.log | grep "Grid    =" | cut -c11-80`
set mask_origin=`cat $tmp/mama.log | grep "Origin" | cut -c11-80`
set mask_extent=`cat $tmp/mama.log | grep "Extent" | cut -c11-80`
# change origin and extent and make new mask

echo re m1 $inputdir/$mask_in > $tmp/mama.inp
echo new same m1 >> $tmp/mama.inp
python << eof >> $tmp/mama.inp
from math import *
```

```
from Numeric import *
from string import *
from fpformat import *
from LinearAlgebra import *
border=3
origin="$mask_origin"
extent="$mask_extent"
o=split(origin)
e=split(extent)
origin=str(int(o[0])-border)+" "+str(int(o[1])-border)+" "+str(int(o[2])-border)
extent=str(int(e[0])+2*border)+" "+str(int(e[1])+2*border)+" "+str(int(e[2])+2*border)
print("new origin "+origin)
print("new extent "+extent)
eof
echo new make m2 >> $tmp/mama.inp
echo or m2 m1 >> $tmp/mama.inp
echo wr m2 $tmp/$mask_out >> $tmp/mama.inp
echo quit >> $tmp/mama.inp
echo y >> $tmp/mama.inp

lx_mama -b < $tmp/mama.inp >>$log

lx_mapman <<eof-1 >>$log
read m1 `echo $tmp/$mask_out` mask
write m1 `echo $inputdir/$map_out` ccp4
quit
eof-1

mv $tmp/$mask_out $inputdir

echo "################### mask program done."

set atomnr = `cat $inputfile|grep "ALA A"|grep -c "ATOM"`

echo " "
echo "##############  generate symmetry operators from starting model  ##############"
echo "input 10-mer:                  "$inputfile
echo "origin of input 20-mer         "`cat $inputfile|grep "REMARK"`
echo "output matrix (only numbers)   "$rawout
echo "output DM input matrices       "$outfile
echo "output CCP4 matrices           "$ccp4out
echo "output O format matrices (RAVE) "$oout
echo "###############################################################################"
echo "unit cell in A                 "$unitcell
echo "number of atoms to superpose   "$atomnr
echo "###############################################################################"

if (-e $outfile) then
rm $outfile
endif
if (-e $rawout) then
rm $rawout
endif
if (-e $ccp4out) then
rm $ccp4out
endif
if (-e $oout) then
rm $oout
endif

foreach chain(A B C D E F G H I J)
 echo "CRYST1   "$unitcell"   "$unitcell"   "$unitcell"  90.00   90.00   90.00 F 41 3 2" > $tmp/{$chain}.pdb
 cat $inputfile | grep "ALA "{$chain} >> $tmp/{$chain}.pdb
lsqkab  XYZIN2  $tmp/A.pdb  XYZIN1  $tmp/{$chain}.pdb  XYZOUT  $tmp/tmp.pdb  <<  eof  > $tmp/log{$chain}.out
 fit atom 1 to `echo $atomnr` -
 chain A
 match 1 to `echo $atomnr` -
 chain A
 output -
 xyz
 end
eof
# rawout, only numbers
cat $tmp/log{$chain}.out | grep "ROTATION MATRIX" -m 1 -A 3| tail -3 >> $rawout
cat $tmp/log{$chain}.out | grep "ROTATION MATRIX" -m 1 -A 4 |tail -1|tail -c 32>> $rawout
```

80

```
# formatted out, ready for dm
cat -E $tmp/log{$chain}.out | grep "ROTATION MATRIX" -m 1 -A 1| tail -1| tr '$' ' '|cat -E |tr
'$' '-'>> $tmp/rot{$chain}.txt
cat -E $tmp/log{$chain}.out | grep "ROTATION MATRIX" -m 1 -A 2| tail -1| tr '$' ' '|cat -E |tr
'$' '-'>> $tmp/rot{$chain}.txt
cat $tmp/log{$chain}.out | grep "ROTATION MATRIX" -m 1 -A 3| tail -1>> $tmp/rot{$chain}.txt
cat  $tmp/log{$chain}.out  |  grep  "ROTATION  MATRIX"  -m  1  -A  4  |tail  -1|tail  -c  32>>
$tmp/trans{$chain}.txt
echo "average " >> $outfile
echo "rota matrix - " >> $outfile
cat $tmp/rot{$chain}.txt>> $outfile
echo "tran - " >> $outfile
cat $tmp/trans{$chain}.txt>> $outfile

end
########################################################################
##reading rotation matrix from file matrices.txt (taking matrix and not inverse matrix !!!)

python << eof-1
from math import *
from Numeric import *
from string import *
from fpformat import *
from LinearAlgebra import *

# get NCS operators from file matrices.txt
RT=[]
f=open("`echo $rawout`")
for i in range(10):
        for j in range(3):
                line=f.readline()
                RT.append(line)
        line=line=f.readline()
        RT.append(line)
f.close()

########################################################################
# CCP4 format

# get translation
trans=[]
for i in range(10):
        trans.append(RT[i*4+3])

# get rotation
rot=[]
for i in range(10):
        r=[]
        for j in range(3):
                s=split(RT[i*4+j])
                r.append([float(s[0]),float(s[1]),float(s[2])])
        rot_array=array(((r[0][0],r[0][1],r[0][2]),(r[1][0],r[1][1],r[1][2]),(r[2][0],r[2][1],
r[2][2])),Float)
        #!!!! not transposed for DM program
        #rot_array=transpose(rot_array)
        rot.append(rot_array)
# write file for ccp4
f=open(r"`echo $ccp4out`","w")
line=""
for i in range(20):
        for j in range(3):
                for k in range(3):
                        line=line+rjust(fix(rot[i][j][k],7),11)
#       f.write(" NCS operator #*#*#*#*#\n")
#       f.write("ROTA MATRIX   "+line+"\n")
#       f.write("TRAN   "+trans[i]+"\n")
         f.write("rotate -"+"\n")
         f.write(line+"\n")
         f.write("shift -"+"\n")
         f.write(trans[i]+"\n")
        line=""
f.close()

########################################################################
# O format
# get translation
trans=[]
```

```
for i in range(10):
        trans.append(RT[i*4+3])
# get rotation
rot=[]
for i in range(10):
        r=[]
        for j in range(3):
                s=split(RT[i*4+j])
                r.append([float(s[0]),float(s[1]),float(s[2])])
        rot_array=array(((r[0][0],r[0][1],r[0][2]),(r[1][0],r[1][1],r[1][2]),(r[2][0],r[2][1],
r[2][2])),Float)
        #!!!! transposed for O program
        rot_array=transpose(rot_array)
        rot.append(rot_array)
# write file for O format (rave)
f=open(r"`echo $oout`","w")
line=""
for i in range(10):
        f.write(".LSQ_RT_1"+str(i+1)+" R 12 (3f15.8)"+"\n")
        for j in range(3):
                for k in range(3):
                        line=line+rjust(fix(rot[i][j][k],8),15)
                f.write(line+"\n")
                line=""
        st=split(trans[i])
        f.write(rjust(fix(st[0],8),15)+rjust(fix(st[1],8),15)+rjust(fix(st[2],8),15)+"\n")
f.close()
eof-1

echo "program done."
echo "##################################################################################"
echo ""

rm -rf $tmp
```

6.3.3 AVE averaging script

Script for averaging with Uppsala ave program.

Input: mtz file, monomer mask, operators, space group symmetry operators, starting model

Output: phased mtz file and corresponding map

Standard averaging script was only slightly modified (http://xray.bmc.uu.se/usf/rave.html)

6.3.4 DM averaging script

This script performs the DM averaging.

Input: mtz file with ave phases, monomer mask, 10-mer mask

Output: output mtz with dm phases, map file

```
#!/bin/csh -f
source /xray/setup/ccp4.set
set datapath = .
set inputfile = $datapath/input_ms041_2_test/ms041_2_ph_fom02.mtz
set inputF    = F_ms041_2
set inputSIGF = SIGF_ms041_2
set inputPHI  = PHICALC
set inputFOM  = FOM_02
set series    = 3
set run       = 36
set highres   = 3.0
set lowres    = 50
set solventc  = 0.88
set solupdate = 50
set cycles    = 300
set dmhklout  = $datapath/dmout/dmhklout.series.$series.run.$run.mtz
set logout    = $datapath/dmout/log.series.$series.run.$run.log
set mapout    = $datapath/dmout/dm2fofc.series.$series.run.$run.map
```

```
set ncsout    = $datapath/dmout/ncs.series.$series.run.$run.mask
set solin     = $datapath/input_ms041_2_test/full.msk
set monomerin = $datapath/input_ms041_2_test/monomer.msk
echo $inputfile
dm HKLIN $inputfile HKLOUT $dmhklout SOLIN $solin NCSIN1 $monomerin NCSOUT $ncsout << eof >
$logout
mode -
    HIST -
    SOLV -
    AVER
scheme RES -
    from 5.0
grid 800 800 800
ncycles -
    `echo $cycles`
solc `echo $solventc`
scale 1 0
resolution `echo $highres` `echo $lowres`
solmask -
    update `echo $solupdate`
`cat $datapath/input_ms041_2_test/ncs_clean.dm`
ncsmask NMER 10
LABIN  FP=`echo $inputF` SIGFP=`echo $inputSIGF` PHIO=`echo $inputPHI` FOMO=`echo $inputFOM`
LABOUT  FDM=FDM2 PHIDM=PHIDM2 FOMDM=FOMDM2
END
eof
fft HKLIN $dmhklout MAPOUT $mapout <<eof
scale F1 1.0
xyzlim asu
labin -
  F1=`echo $inputF` SIG1=`echo $inputSIGF` PHI=PHIDM2 W=FOMDM2
end
eof
```

7 REFERENCES

Afonine, P. V., R. W. Grosse-Kunstleve, et al. (2005). CCP4 Newsletter 42: contribution 8.

Akita, F., K. T. Chong, et al. (2007). "The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses." J Mol Biol 368(5): 1469-83.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol 215(3): 403-10.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res 25(17): 3389-402.

Andrews, S. C. (1998). "Iron storage in bacteria." Adv Microb Physiol 40: 281-351.

Babior, B. M. (1984). "The respiratory burst of phagocytes." J Clin Invest 73(3): 599-601.

Bamford, D. H., J. M. Grimes, et al. (2005). "What does structure tell us about virus evolution?" Curr Opin Struct Biol 15(6): 655-63.

Barefoot, S. F. and T. R. Klaenhammer (1983). "Detection and activity of lactacin B, a bacteriocin produced by Lactobacillus acidophilus." Appl Environ Microbiol 45(6): 1808-15.

Bickley, J. and R. J. Owen (1995). "Preparation of bacterial genomic DNA." Methods Mol Biol 46: 141-7.

Caspar, D. L. and A. Klug (1962). "Physical principles in the construction of regular viruses." Cold Spring Harb Symp Quant Biol 27: 1-24.

Chang, C., Evdokimova, E., Savchenko, A., Joachimiak, A., MCSG (2005). Crystal structure of protein NE0167 from Nitrosomonas europaea, PDB ID: 1ZPY.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr 50(Pt 5): 760-3.

Cowtan, K. D. and P. Main (1993). "Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints." Acta Crystallogr D Biol Crystallogr 49(Pt 1): 148-57.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

Dube, P., P. Tavares, et al. (1993). "The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry." EMBO J. 12(4): 1303-1309.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-32.

Eppert, I., N. Valdes-Stauber, et al. (1997). "Growth reduction of Listeria spp. caused by undefined industrial red smear cheese cultures and bacteriocin-producing Brevibacterium linens as evaluated in situ on soft cheese." Appl Environ Microbiol 63(12): 4812-7.

Finegold, S. M. and W. L. George (1989). Anaerobic infections in humans. San Diego, Calif. etc., Academic Press.

Finkel, T. and N. J. Holbrook (2000). "Oxidants, oxidative stress and the biology of ageing." Nature 408(6809): 239-47.

Gantt, E. and S. F. Conti (1969). "Ultrastructure of blue-green algae." J Bacteriol 97(3): 1486-93.

Gutmann, S. (2001). unpublished data.

Hahn, T. (1983). International tables for crystallography. Dordrecht, Reidel.

Harrison, P. M. and P. Arosio (1996). "The ferritins: molecular properties, iron storage function and cellular regulation." Biochim Biophys Acta 1275(3): 161-203.

Havemann, G. D., E. M. Sampson, et al. (2002). "PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2." J Bacteriol **184**(5): 1253-61.

Helgstrand, C., W. R. Wikoff, et al. (2003). "The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution." J Mol Biol **334**(5): 885-99.

Hicks, P. M., K. D. Rinker, et al. (1998). "Homomultimeric protease in the hyperthermophilic bacterium *Thermotoga maritima* has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria." FEBS Lett **440**(3): 393-8.

Holm, L. and C. Sander (1993). "Protein structure comparison by alignment of distance matrices." J Mol Biol **233**(1): 123-38.

Imlay, J. A., S. M. Chin, et al. (1988). "Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro." Science **240**(4852): 640-2.

Jancarik, J., W. G. Scott, et al. (1991). "Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of *Salmonella typhimurium*." J Mol Biol **221**(1): 31-4.

Jenni, S., M. Leibundgut, et al. (2007). "Structure of fungal fatty acid synthase and implications for iterative substrate shuttling." Science **316**(5822): 254-61.

Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47 (Pt 2)**: 110-9.

Kabsch, W. (1993). "Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants." Journal of Applied Crystallography **26**: 795-800.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-637.

Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." Protein Sci **12**(9): 1865-71.

Kerfeld, C. A., M. R. Sawaya, et al. (2005). "Protein structures forming the shell of primitive bacterial organelles." Science **309**(5736): 936-8.

Kim, S. J. and M. Shoda (1999). "Purification and characterization of a novel peroxidase from *Geotrichum candidum* dec 1 involved in decolorization of dyes." Appl Environ Microbiol **65**(3): 1029-35.

Kleywegt, G. J. and T. A. Jones (1995). "Where freedom is given, liberties are taken." Structure **3**(6): 535-40.

Kleywegt, G. J. and T. A. Jones (1999). "Software for handling macromolecular envelopes." Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 941-4.

Kleywegt, G. J. and R. J. Read (1997). "Not your average density." Structure **5**(12): 1557-69.

Kofoid, E., C. Rappleye, et al. (1999). "The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins." J Bacteriol **181**(17): 5317-29.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-68.

Ladenstein, R., K. Ritsert, et al. (1994). "The lumazine synthase/riboflavin synthase complex of Bacillus subtilis. X-ray structure analysis of hollow reconstituted beta-subunit capsids." Eur J Biochem **223**(3): 1007-17.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Lee, L. A. and Q. Wang (2006). "Adaptations of nanoscale viruses and other protein cages for medical applications." Nanomedicine **2**(3): 137-49.

Leibundgut, M. A. (2005). Crystal structures of the selenocysteine-tRNA specific elongation factor SelB and functional implications. Institute of Molecular Biology and Biophysics. Zurich, ETH Zurich. **Doctor of Natural Sciences:** 129.

Lenoir, J. (1984). "The surface flora and its role in the ripening of cheese." In. Dairy Fed. Annu. Bull. **171**: 3-20.

Macara, I. G., T. G. Hoy, et al. (1972). "The formation of ferritin from apoferritin. Kinetics and mechanism of iron uptake." Biochem J **126**(1): 151-62.

Manca, C., S. Paul, et al. (1999). "Mycobacterium tuberculosis catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro." Infect Immun **67**(1): 74-9.

Marcotte, E. M., M. Pellegrini, et al. (1999). "Detecting protein function and protein-protein interactions from genome sequences." Science **285**(5428): 751-3.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-7.

Milner-White, E., B. M. Ross, et al. (1988). "One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins." J Mol Biol **204**(3): 777-82.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.

Nelson, K. E., R. A. Clayton, et al. (1999). "Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*." Nature **399**(6734): 323-9.

Nickell, S., F. Forster, et al. (2005). "TOM software toolbox: acquisition and analysis for electron tomography." J Struct Biol **149**(3): 227-34.

Ohi, M., Y. Li, et al. (2004). "Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy." Biol Proced Online **6**: 23-34.

Orlova, E. V., P. Dube, et al. (1997). "Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 A resolution by electron cryomicroscopy and angular reconstitution." J Mol Biol **271**(3): 417-37.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode." Methods in Enzymology **276**(Macromolecular Crystallography): 307-326.

Persson, K., G. Schneider, et al. (1999). "Crystal structure analysis of a pentameric fungal and an icosahedral plant lumazine synthase reveals the structural basis for differences in assembly." Protein Sci **8**(11): 2355-65.

Rocchia, W., S. Sridharan, et al. (2002). "Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects." J Comput Chem **23**(1): 128-37.

Rocha, E. R., S. C. Andrews, et al. (1992). "Isolation of a ferritin from *Bacteroides fragilis*." FEMS Microbiol Lett **74**(2-3): 207-12.

Rose, G. D., L. M. Gierasch, et al. (1985). "Turns in peptides and proteins." Adv Protein Chem **37**: 1-109.

Rosenkrands, I., P. B. Rasmussen, et al. (1998). "Identification and characterization of a 29-kilodalton protein from *Mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells." Infect Immun **66**(6): 2728-35.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.

Sander, B., M. M. Golas, et al. (2003). "Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra." J Struct Biol **142**(3): 392-401.

Seebeck, F. P., K. J. Woycechowsky, et al. (2006). "A simple tagging system for protein encapsulation." J Am Chem Soc **128**(14): 4516-7.

Shively, J. M., F. Ball, et al. (1973). "Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus." Science **182**(112): 584-6.

Shively, J. M., G. L. Decker, et al. (1970). "Comparative ultrastructure of the thiobacilli." J Bacteriol **101**(2): 618-27.

Shuman, S. (1994). "Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase." J Biol Chem **269**(51): 32678-84.

Stauber, N. V. (1995). Isolierung, Charakterisierung und Nukleinsäuresequenz eines Bakteriozins aus *Brevibacterium linens* M18. Institut für Mikrobiologie. München, Technische Universität München-Weihenstephan. **Doktor der Naturwissenschaften:** 115.

Storz, G. and J. A. Imlay (1999). "Oxidative stress." Curr Opin Microbiol **2**(2): 188-94.

Switzer, R. C., 3rd, C. R. Merril, et al. (1979). "A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels." Anal Biochem **98**(1): 231-7.

Takahashi, T. and S. Kuyucak (2003). "Functional properties of threefold and fourfold channels in ferritin deduced from electrostatic calculations." Biophys J **84**(4): 2256-63.

Tanaka, S., C. A. Kerfeld, et al. (2008). "Atomic-level models of the bacterial carboxysome shell." Science **319**(5866): 1083-6.

Tang, G., L. Peng, et al. (2007). "EMAN2: An extensible image processing suite for electron microscopy." J Struct Biol **157**(1): 38-46.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.

Tobimatsu, T., M. Kawata, et al. (2005). "The N-terminal regions of beta and gamma subunits lower the solubility of adenosylcobalamin-dependent diol dehydratase." Biosci Biotechnol Biochem **69**(3): 455-62.

Tonami, H., H. Uyama, et al. (2004). "Guaiacol oxidation products in the enzyme-activity assay reaction by horseradish peroxidase catalysis." Chemistry Letters **33**(7): 796-797.

Tong, L. and M. G. Rossmann (1997). "Rotation function calculations with GLRF program." Methods Enzymol **276**: 594-611.

Tsai, Y., M. R. Sawaya, et al. (2007). "Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome." PLoS Biol **5**(6): e144.

Valdes-Stauber, N. (1995). Isolierung, Charakterisierung und Nukleinsaeuresequenz eines Bakteriozins aus *Brevibacterium linens* M18. Institut fuer Mikrobiologie. Muenchen, Technische Universitaet Muenchen-Weihenstephan. **Doktor der Naturwissenschaften**.

Valdes-Stauber, N. and S. Scherer (1994). "Isolation and characterization of Linocin M18, a bacteriocin produced by *Brevibacterium linens*." Appl Environ Microbiol **60**(10): 3809-14.

Valdes-Stauber, N. and S. Scherer (1996). "Nucleotide sequence and taxonomical distribution of the bacteriocin gene lin cloned from *Brevibacterium linens* M18." Appl Environ Microbiol **62**(4): 1283-6.

van Heel, M. (1987). "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction." Ultramicroscopy **21**(2): 111-123.

van Heel, M. and J. Frank (1981). "Use of multivariate statistics in analysing the images of biological macromolecules." Ultramicroscopy **6**(2): 187-194.

van Heel, M., G. Harauz, et al. (1996). "A new generation of the IMAGIC image processing system." <u>J. Struct. Biol.</u> **116**(1): 17-24.

Wikoff, W. R., L. Liljas, et al. (2000). "Topologically linked protein rings in the bacteriophage HK97 capsid." <u>Science</u> **289**(5487): 2129-33.

Wriggers, W., R. A. Milligan, et al. (1999). "Situs: A package for docking crystal structures into low-resolution maps from electron microscopy." <u>J Struct Biol</u> **125**(2-3): 185-95.

Yeates, T. O., Y. Tsai, et al. (2007). "Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells." <u>Biochem Soc Trans</u> **35**(Pt 3): 508-11.

Zubieta, C., S. S. Krishna, et al. (2007). "Crystal structures of two novel dye-decolorizing peroxidases reveal a beta-barrel fold with a conserved heme-binding motif." <u>Proteins</u> **69**(2): 223-33.

GLOSSARY

Bl	*Brevibacterium linens*
bp	Base pairs
cv	Column volume
DyP	Dye-decolorizing Peroxidase
E. coli	*Escherichia coli*
EtBr	Ethidium bromide
FT	Flowthrough
his	Histidine (tag)
kDa	Kilodalton
LB	Luria-Bertani medium
MPD	Methylpentanediol
Ni-NTA	Nickel-nitrilo-triacetic acid (resin)
O/N	Overnight
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
ROS	Reactive oxygen species
RT	Room temperature
SDS	Sodium dodecyl sulfate
SN	Supernatant
TM	*Thermotoga maritima*
wt	Wild type
β-me	β-mercaptoethanol
GFP	Green Fluorescent Protein
asu	asymmetric unit
PDB	Protein Data Bank (http://www.rcsb.org/)
BLAST	Basic Local Alignment Search Tool

ACKNOWLEDGEMENTS

CURRICULUM VITAE

Personal Information

Name	Sutter, Markus Dipl. Natw. ETH
Date of birth	December 16th, 1976
Place of birth	Wald, ZH, Switzerland
Nationality	Swiss
Address	Schreinerstr. 42 8004 Zurich Switzerland

Education

2002-2008	PhD in the Ban Group, Institute of Molecular Biology and Biophysics, ETH Zurich "*Structural basis of enzyme encapsulation into a bacterial nanocompartment*"
2001-2002	Diploma work in the Ban Group, ETH Zurich: "*Characterization and Crystallization of an RNA/Protein complex*"
1997-2002	Diploma in Biology ETH Zurich Type Ae
1994	Exchange year at Haileybury College, Melbourne, Australia
1991-1997	Matura Kantonsschule Wetzikon

Publications

Gutmann, S., P. W. Haebel, L. Metzinger, M. Sutter, B. Felden and N. Ban (2003). "Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB." <u>Nature</u> **424**(6949): 699-703.

Sutter, M., D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban (2008). "Structural basis of enzyme encapsulation into a bacterial nanocompartment" <u>Nat. Struct. Mol. Biol.</u> **15(9):** 939-947.