UiT The Arctic University of Norway

Faculty of science and Technology, The Institute of Chemistry

Structural and functional studies of Ectoine Synthase from *Chromohalobacter salexigens DSM 3043* and *Marinobacter sp. CK1*

Sigurd Eidem Gundesø

Master's thesis in Molecular Science, KJE3900, October 2021

Acknowledgements

First and foremost, I would like to extend a sincere thanks to my supervisor Prof. Ingar Leiros for all help, guidance, encouragement and tough love throughout this thesis period. It has been a pleasure working under your supervision, and I very much appreciate your help, vast knowledge and good humor throughout this last year.

Secondly, I want to say a huge thank you to my co-supervisor Heidi Therese Hillier. Your patience with me has been amazing, and your help in the lab has been invaluable. I also appreciate you letting me share your office space for the last year, and I'll miss our coffee breaks and our academic and not-so academic discussions.

I would also like to extend a thanks to Dr. Bernd Striberny. While not directly involved in this project, your influence on my academic and scientific development cannot go unmentioned.

Furthermore, I would like to thank the entire NorStruct centre for a welcoming and including environment. I appreciate all help and patience with me in the lab.

Finally, I would like to thank my friends and family for encouragement and support throughout the last year.

Sigurd Eidem Gundesø, October 2021, Tromsø

Abbreviations

UV	Ultra Violet
DNA	Deoxyribonucleic acid
EctB	L-2,4-diaminobutyric-2-oxoglutaric acid transaminase
PLP	pyridoxal-5-‘-phosphate
ASA	L-aspartate-β-semialdehyde
DABA	L-2,4-diaminobutyrate
ND	Not Determined
H. elongata	*Halomonas elongata*
C. salexigens	*Chromohalobacter salexigens*
P. lautus	*Paenibacillus lautus*
EctA	L-2,4-diaminobutyrate acetyltransferase
Acetyl-CoA	Acetyl coenzyme A
N-*γ*-ADABA	*N*-*γ*-L-2,4-diaminobutyrate
GNAT	GCN5-related *N*-acetyltransferases
PCR	Polymerase Chain Reaction
COVID-19	Coronavirus Disease of 2019
EctC	Ectoine Synthase
IUBMB	International union of Biochemistry and Molecular Biology
M. alcaliphilum	*Methylomicrobium alcaliphilum*
S. alaskensis	*Sphingopyxis alaskensis*
N. maritimus	*Nicrophorus maritimus*
A. cryptum	*Acidiphilium cryptum*
AMR	Antimicrobial Resistance
RNA	Ribonucleic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
FPLC	Fast Protein Liquid Chromatography
PDB	Protein Data Bank
DLS	Dynamic Light Scattering
DSC	Differential Scanning Calorimetry
ITC	Isothermal Titration Calorimetry
IMAC	Immobilized Metal Affinity Chromatography
Histag	Hexahistidine tag
MarEctC	*Marinobacter sp. CK1* Ectoine Synthase
CSEctC	*Chromohalobacter salexigens* Ectoine Synthase
TE	Tris-EDTA
LB	Luria-Bertoni
TB	Terrific Broth
Rpm	Rounds Per Minute
OD	Optical Density
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
LC-MS	Liquid Chromatography Mass spectrometry
CV	Column Volume
SEC	Size Exclusion Chromatography
TCEP	tris(2-carboxyethyl)phosphine
kDa	kiloDalton
PEG	Polyethylene Glycol
MSA	Multiple Sequence Alignment
GMQE	Global Model Quality Estimate
PhD	Philosophiae Doctor
HPLC	High-Pressure Liquid Chromatography
UiT	University of Tromsø
Lys	Lysate
FT	Flow Through
W	Wash
ESRF	European Synchrotron Radiation Facility
MBP	Maltose Binding Protein
SUMO	Small Ubiquitin-Like Modifier
GST	Glutathione-S-Trasferase
TSA	Thermal Shift Assay
NMR Spectroscopy	Nuclear Magnetic Resonance Spectroscopy

Abstract

Ectoine is a compatible solute found in many microorganisms adapted to survive in saline and other extreme environments. Here, it aids microorganisms to counter osmotic stress and protect their enzymes. Ectoine exhibit many interesting properties that is potentially commercially exploitable, and it is currently produced and found in several products on the market. While ectoine is produced by whole cell synthesis, the EctABC enzymes in the biosynthesis pathway of ectoine was currently not well described structurally or functionally. Here, we present structural and biochemical characterizations of ectoine synthase from two organisms, *Chromohalobacter salexigens DSM3043* and *Marinobacter sp. CK1*. We cloned, expressed and expression optimized both candidates, and purified them by immobilized metal affinity chromatography and gel filtration. *C. salexigens* EctC (CSEctC) yielded 14-18 mg/L cell culture while *Marinobacter sp. CK1* (MarEctC) yielded 0.75-1.5 mg/L culture. We then produced diffracting crystals of CSEctC and obtained a data set from which the structure of CSEctC was determined. We further obtained preliminary biochemical data relating to thermostability and activity from both candidates. The crystal structure from CSEctC shows that it is adapts a typical β-sandwich fold, consistent with earlier structural investigations of other EctC-type proteins. This study provides a solid foundation for further research on EctC from our model organisms, and protocols and techniques developed herein can be further optimized to obtain more biochemical data about this interesting enzyme.

Innhold

1 Introduction

1.1 Ectoine and the *ect* enzymes in the biosynthesis pathway of ectoine

Ectoine [(4S)-2- methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid] is an amino acid – it is a heterocyclic derivative of aspartate, the ionic state of aspartic acid (Asp, D) [1]. At physiological pH, ectoine is a zwitterionic compound, and it is highly water soluble (6 M at 4 °C) [2, 3]. The first organism in which ectoine was identified was the phototrophic purple sulphur bacteria *Ectothiorhodospira halochloris* belonging to the gammaproteobacterial class [4]. This is a species adapted to extremely saline environments, and the early discoveries of ectoine suggested that ectoine predominantly was found on rare occurrences in extremely halophilic to moderately halophilic baceteria [5, 6]. However, bioinformatical analysis and identification of genes associated with ectoine synthesis have revealed that ectoine producers are widely distributed [6, 7]. Today, we know that ectoine producers can be found across all three domains of life, but predominantly among bacteria [7-9]. In addition to ectoine, an ectoine derivative called 5-hydroxyectoine exists and is physiologically relevant for many species [10], but only ectoine will be discussed in this thesis.

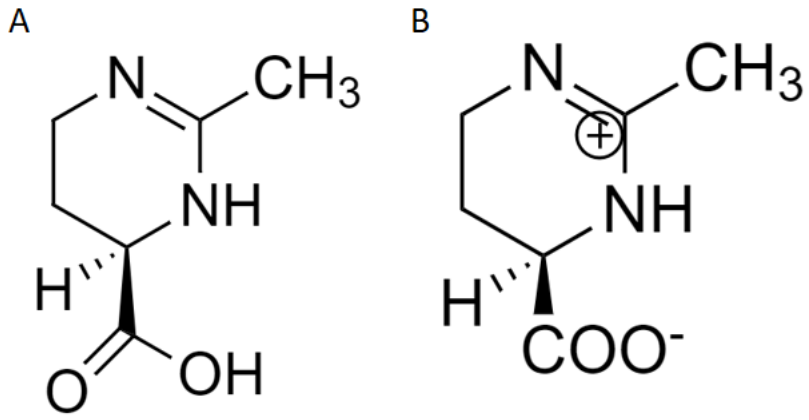


Figure 1-1: A) The structure of [(4S)-2- methyl-1,4,5,6-tetrahydropyrimidine-4carboxylic acid] ectoine. B) At physiological pH, ectoine is in a zwitterionic form, with a negatively charged carboxylate group and a delocalized positive charge. The figure was made with Chemdraw direct (PerkinElmer).

1.1.1 Ectoine and its role in extremophiles

In order to preserve their biological functionality during osmotic stress, microorganisms can follow one of two main strategies; the "salt-in" strategy or the "salt-out" strategy. There is no mechanism in microorganisms for actively transporting water across their cell membranes. Such organisms must instead manipulate the flow of water in and out of the cell by uptake or removal of ions to their cytoplasm [11]. In saline environments, many microorganisms accumulate salt ions (primarily potassium and chloride) to increase their osmotic potential and thereby the net water influx to the cell. This is the salt-in strategy. While this is an efficient way to combat hyperosmotic conditions, the cell may become susceptible to bursting under rapid decline in external salinity [6]. Thus, the salt-in strategy is mostly adapted by microorganisms which lives under permanent or stable saline conditions [12, 13]. A consequence of this however, is that the intracellular proteins must be adapted to function under permanently high ion-content of the cytoplasm. One way this is achieved is by increasing the number of negatively charged amino acids on the protein surface [12].

The salt-out strategy represents a more flexible approach to combat osmotic stress, and is adapted

widely by single-cell organisms [14]. Initially when exposed to hyperosmotic conditions, the cell will accumulate potassium and chloride ions, but then gradually replace these ions with self-synthesized or uptake of organic compounds; the compatible solutes. These types of osmolytes are so named because they can be found at high concentrations intracellularly without altering the vital processes of the cell [6, 15]. One of the advantages of the salt-out strategy is that the cell can maintain a relatively low ionic content while still containing water and turgor pressure of the cell. In addition, it avoids exposing its proteins to evolutionary pressure, as proteins does not need to adapt to function in a highly ionic cytoplasm. A disadvantage of the salt-out strategy is however that accumulation of compatible solutes is a highly energy-demanding process [6, 16]. Ectoine is one such compatible solute. Other examples of osmolytes are other amino acid derivatives like taurine and betaine, or carbohydrates and sugar-alcohols like trehalose, sorbitol and glycerol [17].

Ectoine serves two main purposes as a cell protectant when accumulated in high concentrations; as an organic osmolyte, and as a chemical chaperone. When accumulated in high concentrations, ectoine protects the cell against osmotic stress. Numerous studies have been conducted to illuminate the exact mechanics of how ectoine protects microorganisms from environmental stress [6, 7, 18]. While the mode of actions is not completely understood, ectoine is hypothesized to the preferential exclusion effect [19, 20]. This is in line with known mechanisms for other compatible solutes found in bacteria [21]. The preferential exclusion model describes a situation where the compatible solute is distributed unevenly throughout the cell and is excluded from the surface of intracellular proteins. This is thought to be because of unfavourable interactions between the ectoine molecule and exposed residues on the protein surface. This bulk accumulation of ectoine away from the protein hydration shell pushes water molecules to the protein surface. This effect is called preferential hydration [22] In addition to preferential hydration, unfavourable interactions between ectoine and the protein surface, which are also thermodynamically unfavourable, forces the protein to maintain its folded and well-ordered state [23, 24]. This effect complements the interactions normally associated with protein folding like hydrophobic interactions, hydrogen bonding and Wan der Waals forces [25-27].

In the same way that ectoine is excluded from the hydration shell of proteins, they are also displaced from the hydration shell of the cell membrane [28]. *In silico* molecular dynamics simulation has shown that ectoine is capable of hydrogen bonding with up to 7 water molecules simultaneously [6, 29]. Ectoine has also shown to accumulate in certain microorganisms as a response to extreme temperatures, and it can therefore also be seen as a thermo- and cryoprotectant [30, 31]. The mechanisms of how ectoine function in this aspect is currently not well understood. Other studies have found ectoine to be a protectant against UV radiation-induced breaks of single strand DNA [32]. Ectoine also shows protecting properties against ultraviolet protein damage and cellular stress [33]. In summary, ectoine exhibits multiple interesting properties in the context of cell protection against a variety of stress conditions.

1.1.2 Commercial production and applications of ectoine

The many fascinating properties of ectoine highlighted in section 1.1.1 has made ectoine a highly interesting molecule in terms of commercial exploitation. However, early *in vitro* attempts at chemical ectoine synthesis proved too costly, as ectoine precursors (mainly the EctB substrate ASA) was expensive to synthesize [34, 35]. This prompted industrial players to evaluate different halophilic bacteria as candidates for *in vivo* ectoine synthesis. Ectoine is currently produced by whole-cell synthesis from the halophilic *H. elongata* by the German company Bitop [36, 37]. This strain is a moderately halophilic, highly halotolerant γ-proteobacterium, and ectoine production is conducted under highly saline conditions (2.57 M NaCl) [1]. To extract the synthesized product, the cells are subjected to a hypo-osmotic downshock, which causes mechanosensitive channels to open and cause a mass efflux of ectoine into the medium. This process has been named bacterial milking [38]. Because there is no cell lysis involved in extraction, the cell mass can be reused in ectoine production by increasing the salinity of the medium again. Released ectoine can further be extracted and purified by e.g. acidification, chromatography and other techniques to yield highly pure ectoine [1, 6, 35]. Bitop sells and markets ectoine as an active ingredient under the registered name Ectoin® natural.

Today, the main use of ectoine is found in products aimed at skin protection and other cosmetic products for humans[20, 39]. The earlier discussed water binding properties of ectoine makes it well suitable to prevent dehydration in skin, and is therefore found in moisturising products. Advantages of ectoine in skin care is well documented, and existing products have been improved under clinical trials with a 2 % addition of ectoine [40]. This addition showed to promote greater moisturising effects of skin products, as well as improvement of skin elasticity and skin structure [41]. The UV-radiation protective properties of ectoine also makes it a suitable ingredient in sun protection products, and multiple products on the market for this use contain ectoine [40]. Ectoine is further found as an active ingredient in anti-inflammatory creams [42], anti-allergic nasal sprays [43] [44] and as stabilizing agent in cosmetics to reduce skin irritation [45].

While the main market for ectoine currently is the expanded beauty and skin care industry, advances in evaluating ectoine for other therapeutic applications have been made the last years. When found natively, ectoine and other compatible solutes aids in stabilizing proteins and act as a folding chaperone (see section 1.1.1). In that respect, ectoine is seen a potential candidate as a therapeutic agent against diseases associated with protein misfolding like amyloidosis [46]. Protein aggregation is also associated with neurodegenerative diseases like Alzheimer's disease, and studies of ectoine has shown positive effects in animal studies of Alzheimer's, although not at a clinical level [47]. Other protein aggregation and misfolding related diseases where ectoine has shown therapeutic potential includes prion peptide neurotoxicity and insulin amyloid formation [48].

Another highly interesting potential market for ectoine is found in the molecular biology segment. The stabilizing effect of ectoine on enzymes in solution makes ectoine a good candidate for stabilizing enzymes during biotechnological applications where higher temperatures or other environmental stresses is required [49]. Other such stresses includes freeze-thaw cycles which is a common treatment of enzymes in laboratories worldwide, and ectoine has shown stabilizing and activity-preserving properties under such treatments of enzymes [50]. While ectoine-mediated stabilizing of other macromolecules such as DNA is lesser studied, ectoine has been shown to inhibit

endonuclease activity of DNA by inducing changes in the DNA structure [51]. Other studies has suggested that supplementing Polymerase Chain Reactions (PCRs) with ectoine can enhance amplification reactions where the reaction template is Guanine-Cytosine rich [52]. Ectoine thereby shows potential both as an improver of PCR efficiency and as a stabilizer of enzymes, both of which has been increasingly relevant during extensive use of PCR-based diagnostics of diseases such as COVID-19 [53].

1.1.3 EctB

The first enzyme involved in ectoine synthesis is L-2,4-diaminobutyric-2-oxoglutaric acid transaminase EctB. EctB belongs to the transaminase class of enzymes (EC.2.6.1) and more specifically the class III ω-transaminases [54]. This class of enzymes catalyse the transfer of an amine group from a donor to an acceptor, with pyridoxal-5'-phosphate (PLP) as a required coenzyme [55]. Structural and chromatographical studies of this class of transaminases have shown that the enzyme operates as a functional dimer [56, 57]. This is also the case for EctB, although a structure solved for EctB from *C. salexigens* revealed that it is found as a tetramer of two functional dimers both in its soluble and crystalline states [58]. EctB catalyses a reversible reaction between L-aspartate-β-semialdehyde (ASA) and L-glutamate; an amino group is transferred from L-glutamate to the aldehyde group of L-aspartate-β-semialdehyde. This forms the product L-2,4-diaminobutyrate (DABA) and 2-oxoglutarate, a side product also involved in the synthesis of the ectoine derivative 5-hydroxyectoine [59].

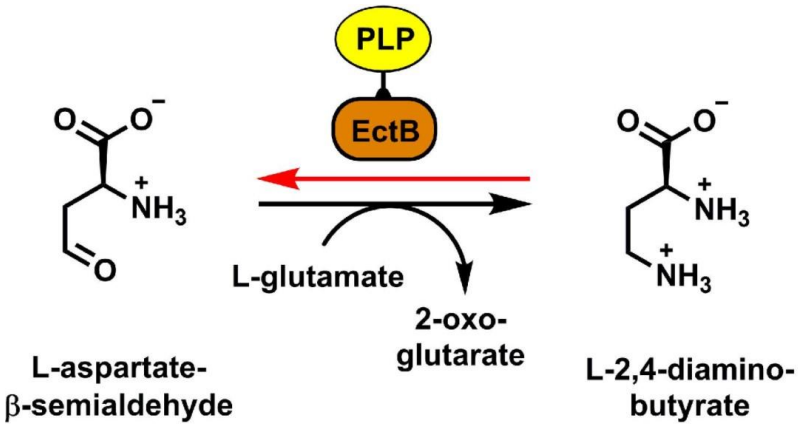


Figure 1-2: The reversible reaction catalyzed by EctB between L-aspartate-β-semialdehyde and L-glutamate. The figure was adapted from Richter et al. (2019) [60].

The best characterized EctB enzyme is from the moderately halophilic *Halomonas elongata [61]*. It was earlier reported that EctB from this organism seems to be K^+-dependent for both activity and stability, but EctB from *Chromohalobacter salexigens* is both active and stable with only NaCl in its buffer [58]. EctB from *H. elongata* has its pH optimum around 8.6-8.7, temperature optimum at 25 °C and 0.5 M NaCl concentration [61]. Similar characterizations have been conducted with EctB from *C. salexigens*, where the enzyme was active over a rather wide range of pH values, and optimum was found at pH 8.0. The optimum NaCl concentration was found from 0.3-0.7 M and the temperature optimum 60 °C with a melting temperature (T_m) at 71 °C [58]. A third study with EctB from *Paenibacillus lautus*, a thermo-tolerant bacterium revealed similar broad ranges of activity optima conditions [62].

Table 1-1: Activity optima and kinetic parameters for three characterized EctB variants. All values are for the forward reaction except kinetic values for P. lautus which is for the reverse reaction. Kinetic values are given for L-glutamate or DABA. Values in paranthesis are the range in which the enzyme are retains at least 50 % relative activity. ND: not determined. Values from references [58, 60, 61].

Organism	pH	[NaCl] (M)	Temperature (°C)	K_m (mM)	V_{max} (µmol min^{-1} mg^{-1})
H. elongata	8.6-8.7	0.5	25	4.5	12
C. salexigens	8.0 (7.0-10.0)	0.3 (0.1-0.7)	60 (30-60)	ND	ND
P. lautus	6.5 (6.0-9.0)	200-350 (0-1000)	45 (30-50)	0.4	3.1

1.1.4 EctA

The second enzyme in the ectoine synthesis pathway is L-2,4-diaminobutyrate acetyltransferase, or EctA. As the full name suggests, EctA catalyses the transfer of an acetyl group from the DABA product of the upstream EctB-catalysed reaction from an acetyl coenzyme A (acetyl-CoA) molecule (Figure 1-3). The reaction yields the product *N*-γ-L-2,4-diaminobutyrate (*N*-γ-ADABA). An isomer of this product, *N*-α-ADABA is physiologically relevant as an expression regulator for bacterial metabolism, but EctA exclusively produces *N*-γ-ADABA [63, 64] This illustrates the highly specific nature of the EctABC enzymes. Structurally, bioinformatic analysis reveals that EctA belongs to the GCN5-related *N*-acetyltransferases (GNAT) [64, 65].

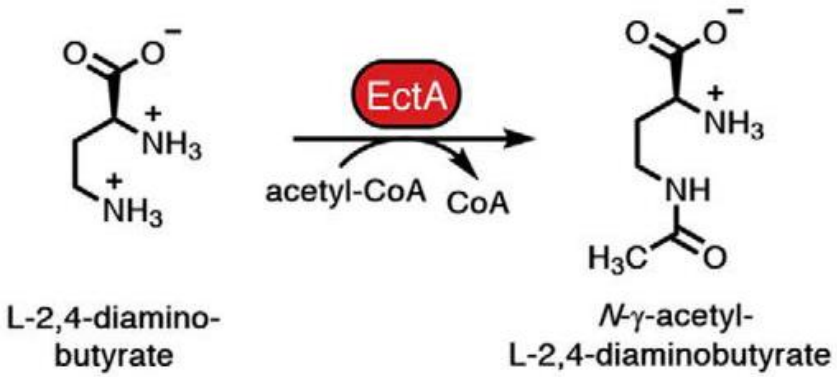


Figure 1-3: An overview of the EctA-catalysed reaction. The figure was adapted from Richter et al (2020) [64].

Gel filtration experiments have shown that EctA is most likely a homodimer in solution [6]. As with EctB, the first characterized EctA came from *H. elongata.* In addition, some biochemical characterizations has been done for EctA from a haloakalotolerant methanotroph, *Methylomicrobium alcaliphilum* [66]. The most recent studied EctA came from *P. lautus,* and a crystal structure in apo-form has been solved for this enzyme [64]. The crystal structure confirmed that EctA is a homodimer, and that it adopts a classical GNAT type fold [64]. Known activity optima and kinetics for EctA comes from enzymes from the three organisms mentioned above (Table 1-2), as well as two more methylotrophs [67]. Hence, much is still unknown of EctA activity and structure from a wider range of organisms.

Table 1-2: Overview of activity optima and kinetics parameters of characterized EctA variants. Values in parenthesis are the range in which the enzyme retains at least 60 % relative activity. Kinetic parameters are given for DABA. ND: not determined. Values from references [61, 64, 66].

Organism	pH	[NaCl] (mM)	Temperature (C °)	K_m (mM)	V_{max} (µmol min^{-1} mg^{-1})
H. elongata	8.2	0.4	20	ND	ND
M. alcaliphilum	9.5	0.2	ND	ND	ND
P. lautus	8.5-9.5	0-50(0-200)	50(40-60)	0.13	44.87

1.2 EctC

The enzyme responsible for converting N-γ-ADABA to ectoine is ectoine synthase, or EctC. Genes belonging to the ectABC gene cluster are usually found together, but not always. In this regard, EctC is viewed as the diagnostic enzyme for ectoine production in a genome [68]. As found in nature, EctC catalyzes the cyclocondensation reaction of the natural substrate N-γ-ADABA into ectoine.

According to the nomenclature developed by the International Union of Biochemistry and Molecular Biology (IUBMB), EctC is classified as a Hydro-lyase (EC.4.2.1) Hydro-lyases are enzymes which catalyse the cleavage of a carbon-oxygen bond to remove a water molecule from its substrate. EctC enzymes belongs structrually to the cupin superfamily of proteins. This is a highly diverse group of proteins, which shares a common structural domain; a β-barrel fold [69]. The cupin name comes from the latin term *cupa*, which means small barrel. While this structural domain is conserved and found in all three domains of life, they have adapted highly dissimilar functions. The cupin superfamily was first recognised as a conserved structural domain from sequence similarity studies of the wheat protein germin and spherulin; a protein found in the slime mold *Physarum polycephalun* where it is involved in spore formation during extreme environmental pressure [70]. Further studies

soon revealed a wide range of cupin proteins, which was identified to contain two highly conserved motifs. Motif 1 contained two conserved histidine residues in a 20 or 21-amino acid long sequence [$G(X)_5HXH(X)_{3,4}E(X)_6G$]. The second motif contained one conserved histidine residue, [$G(X)_5PXG(X)_2H(X)_3N$], with unconserved regions between the two motifs ranging from 15-50 amino acids [71]. In enzymes, both motifs are involved in the coordination of a catalytically essential metal. While the two motifs are considered a diagnostic feature of a cupin, the motifs are not as conserved as previously thought, and variations in the metal binding residues are well documented [72, 73].

The characteristic feature of the cupin superfamily is the β-barrel fold. A β-barrel is a structure where β-strands are aligned in repeats with flexible turns to form one or more β-sheet(s). The β-strands are often but not always aligned in an antiparallel fashion. EctC-type enzymes have been identified to adapt a jelly-roll fold type of a β-barrel. A jelly roll (sometimes referred to as a Swiss roll) consists of two antiparallel β-sheets, often with either 4 or 8 strands each. While classical β-barrel sheets interact with hydrogen bonds where the edges of two sheets meets, the β-sheets in jelly roll type proteins typically does not. In that regard, the jelly-roll fold is by many not considered a true β-barrel, but rather a β-sandwich [74]. Cupin-like β-sandwiches are generally known to be thermally stable and resistant against proteases [72].

1.2.1 Activity, structure and properties of EctC

EctC was originally identified as a cupin protein by sequence analysis, and later confirmed by Widderich et al. by the structural characterization of EctC from *Sphingopyxis alaskensis* by X-ray crystallography [75]. They further confirmed that EctC indeed is a divalent metal-dependent enzyme, and that iron (Fe^{2+}) is the metal most likely associated with EctC activity. As found in nature, EctC catalyses the cyclocondensation reaction of the natural substrate *N*-γ-ADABA into ectoine (Figure 1-4).

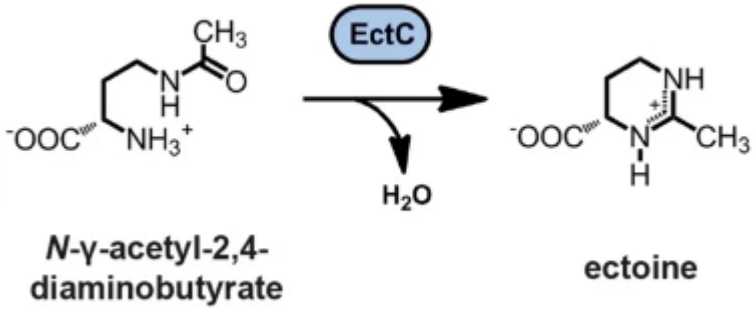


Figure 1-4: The EctC-mediated reaction from N-γ-ADABA to ectoine. The catalytic essential divalent metal (Fe^{2+}) is not shown. The figure was adapted from Czech et al. (2019) [76].

EctC is perhaps the best studied enzyme of the EctABCs thus far, with biochemical characterizations reported from 6 EctC variants (

Table 1-3), including one archeal organism (*Nitrospumilus maritimus)*. It should however be noted that several of the studies of different EctCs has been conducted across different buffer compositions and by different assay techniques. Thus, direct comparisons of kinetic parameters may not be accurate or advisable.

Table 1-3: Activity optima and kinetic parameters previously reported for EctC enzymes from various organisms. Values in parenthesis are value ranges in which the enzyme retained at least 60 % relative activity. ND: Not determined. Values were collected from references [8, 61, 67, 75-77].

Organism	pH	[NaCl] (M)	Temperature (C °)	K_m (mM)	V_{max} (µmol min^{-1} mg^{-1})
H.elongata	8.5-9.0	0.5 (at 15 °C)	15 (at 0.77 M [NaCl])	8.4 (at 0.77 M [NaCl])	56 (at 0.77 M [NaCl])
M. alcaliphilum	ND	ND	ND	28.7	64
S. alaskensis	8.5 (8-9)	0.25 (0.1-0.75)	15 (10-20)	5	25
N. maritimus	7.0	0.2	30	6.4	12.8
A. cryptum	8.5-9.0	0	ND	ND	ND
P. lautus	8.5(7.5-9.0)	0.05 (0-0.4)	30 (15-40)	7.8	16.0

There are currently two organisms from which EctC crystal structures have been solved: *S. alaskensis* and *P. lautus* [75, 76]. Both crystal structures show that EctC is a homodimer, aligned in an antiparallel fashion (head-to-tail alignment).

1.3 Extremophiles as model organisms for characterization of ectoine-related enzymes

Extremophiles are organisms adapted to live and thrive in harsh environments, with respect to temperature, salinity pH, pressure or other basic factors [78]. While the term extremophile mainly applies to microorganisms., the definition for extreme conditions holds an anthropocentric point of view. In fact, the abundance of species adapted to what humans would consider extreme or

uninhabitable is vast, and extremophiles is perhaps one of the most abundant lifeforms on the planet [79]. As already discussed, organisms living under environmental stress have developed strategies for surviving, like production of compatible solutes like ectoine. However, the enzymes of such organisms themselves must be adapted to retain activity when faced with extreme conditions like temperature or salinity. In this thesis we have selected EctC enzymes to be studied from two organisms which will be shortly presented.

1.3.1 Chromohalobacter salexigens DSM3043

The *Chromohalobacter* genus is a rod-shaped, gram-negative bacteria, belonging to the *Gammaproteobacteria* order and the *Halomonadaceae* family. Bacteria belonging to the *Chromohalobacter genus* are characterized as being moderately halophilic, chemoorganotrotrophic and usually facultatively anaerobic [80].

Chromohalobacter salexigens is perhaps the most well studied species of the *Chromohalobacter* genus, especially in the context of osmoregulation. This is because *C. salexigens* shows several interesting features to manage its osmosis. Firstly, it is highly halotolerant, while it also displays one of the widest ranges of salt concentrations in which it can grow. *C. salexigens* has shown to grow in media containing 0.9-25 % NaCl and under in media with pH 5-10 [81]. When compatible solutes are found in their external environment, *C. salexigens* can accumulate these compatible solutes to supplement ectoine synthesis (see section 1.1.1) [80]. This addition to their salt-coping mechanisms makes *C. salexigens* a species which can thrive both in high-salt environments, as well as environments where salt concentrations vary. Finally, they have been found to be efficient in synthesis of ectoine and 5-hydroxy-ectoine [82], and thereby a suitable model organisms for ectoine synthesis studies.

1.3.2 Marinobacter sp. CK1

The *Marinobacter* genus is a diverse group of bacteria found in a variety of marine locations, as well as some saline terrestrial environments. They are gram negative, flagella-motile and known to be

slightly to moderately halophilic [83]. Generally, *Marinobacter* species are mesophilic and neutrophilic, but there are known exceptions both in terms of temperature preference and pH. Interestingly, some species of *Marinobacter* is found in oil spills, where they contribute to biodegradation of hydrocarbons, and is as such a potential candidate for combatting toxic oil spills [84].

The *Marinobacter* species relevant for this thesis is *Marinobacter sp. CK1*. This is a psychrophilic bacterium first identified in the Kårvika (Ringvassøy in Tromsø municipality, Norway). It belongs to an in-house collection of bacteria, and its genome or general characterizations are currently not published. It was chosen as a model organism to evaluate potential cold adaptations in the ectoine biosynthesis pathway enzymes.

1.4 Functional and structural studies of proteins

1.4.1 Recombinant protein expression and purification

Early studies of enzymes relied on native protein production; letting the natural host of the protein of interest produce the protein and then extract and purify it based on its natural properties. With the advent of recombinant DNA technology in the 1970's, it became possible to make artificial DNA an insert in host organisms to express protein recombinantly [85]. Today, recombinant protein production is a core skill in the field of enzymology, and the first step in any characterization of an enzyme. The procedure of producing protein recombinantly is quite simple; the gene that encodes the protein of interest is inserted into a suitable expression vector, transformed into a host and the host is induced to synthesize the protein. The protein of interest can thereafter be purified by fractionating and selecting based on its biochemical properties. There are however many pitfalls and potential obstacles to obtaining a sufficient amount of protein for downstream characterization.

Firstly, a suitable expression vector needs certain attributes to make it effective (Figure 1-5). An expression vector is usually a gene modified bacterial plasmid in which a gene of interest can be inserted. While traditionally this was done by restriction enzymes and ligation reactions, more

sophisticated methods have largely replaced this [86]. One of the most popular systems to build a recombinant plasmid today because of its user-friendliness is with the Gateway™ cloning system [87]. Instead of a two-step reaction with restriction enzymes and ligases, the site-specific recombination is mediated by two recombinases originally found in bacteriophage γ. In the reaction, a pre-cloned entry vector that contains the gene of interest (which may be synthesized with traditional restriction enzyme and ligation technology) and an expression vector is combined. The gene of interest in the entry vector is flanked by two cloning sites; *attL1* and attL2. The expression vector usually contains a toxic gene flanked by compatible *attR1* and *attR2* cloning sites. During the gateway reactions, highly specific recombinases cuts and rejoins the DNA at the four recognition sites and transfer them to the cloning sites of the opposite vector [88]. The toxic gene is now transferred to the entry vector, and the gene of interest to the expression vector. This ensures that in downstream transformation, bacteria with the entry vector will be killed by the toxic gene [89].

An expression vector commonly carries several features to make expression of the gene of interest as effective as possible (Figure 1-5). In order to ensure that only bacteria with the plasmid present is grown in the culture, a selectable marker is present. This is usually a gene that encodes some sort of antimicrobial resistance (AMR) protein. Common selectable markers are β-lactamase-encoding genes (e.g. *AmpR*, *AmpC)* or Kanamycin resistance-encoding genes (e.g. *KanR*) [90, 91]. Bacteria that do not express these genes will be killed by adding the appropriate antibiotic to the growth medium. To control the timing of the synthesis of the protein of interest, an inducible promoter is usually also present in the expression vector. Gateway™ vectors usually use the T7 promoter system. This is a promoter sequence located upstream of the gene of interest in the plasmid and is recognized by the phage T7 RNA polymerase [92]. The Polymerase gene is encoded in the bacterial genome in most cell strains compatible with Gateway™ vectors, under the control of *lac*UV5 promoter, which can be induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) [93]. Under induction, the T7 RNA polymerase is produced and transcription of the gene of interest is initiated. This induction control is important to ensure that the bacterial culture is in the proper growth phase when expression of the protein of interest is induced. This, like a lot of other factors concerning protein expression can

be optimized experimentally, but a normal starting point for induction is during the mid-log phase (Figure 1-6: A visualization of how a bacterial culture grows over time. Initially, during the lag phase the bacteria adjust to the culture conditions, and growth is slow. During the log phase, the bacteria grow exponentially. When cell division and cell death is happening at equal pace, the culture enters the stationary phase. This is because the level of nutrition left in the media cannot sustain further growth. Finally, the culture enters the death phase when cell death exceeds cell division. The figure is reproduced from Wang et al. (2015) [94]).

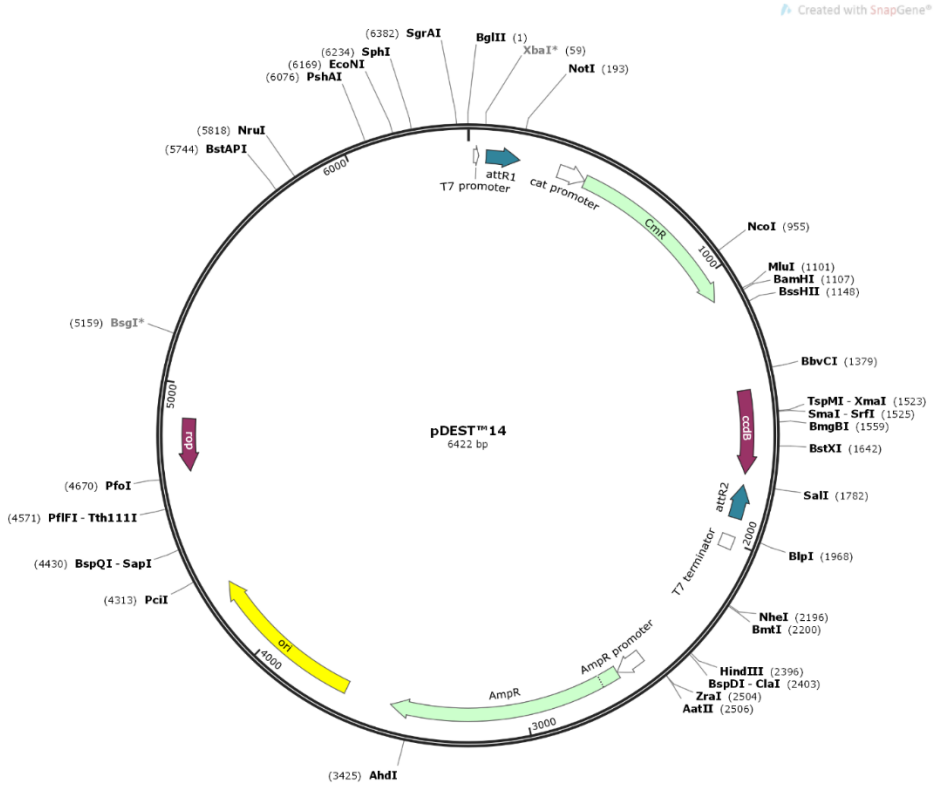


Figure 1-5: The pDEST™14 vector commonly used during Gateway™ Cloning.

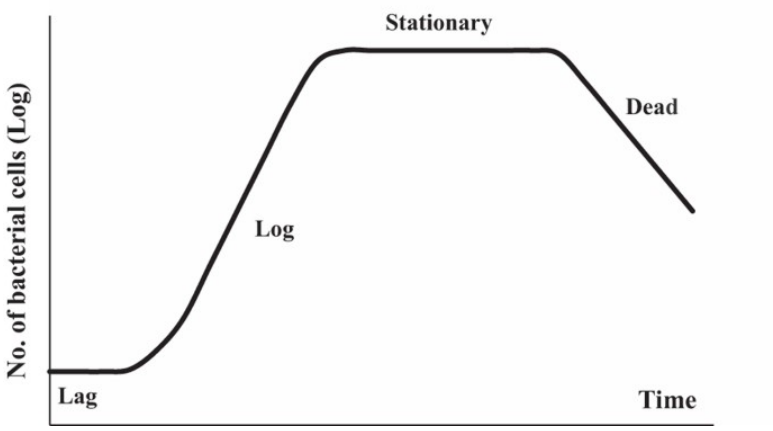


Figure 1-6: A visualization of how a bacterial culture grows over time. Initially, during the lag phase the bacteria adjust to the culture conditions, and growth is slow. During the log phase, the bacteria grow exponentially. When cell division and cell death is happening at equal pace, the culture enters the stationary phase. This is because the level of nutrition left in the media cannot sustain further growth. Finally, the culture enters the death phase when cell death exceeds cell division. The figure is reproduced from Wang et al. (2015) [94].

After successfully expressing the protein of interest there is a need to extract the protein from the bacterial culture and purify it by getting rid of other cell proteins. Extraction is usually solved by some sort of cell lysis, followed by a separation of cell debris and soluble protein by centrifugation. Today, the most popular techniques for protein purification are variants of Fast Protein Liquid Chromatography (FPLC). FPLC relies on separating soluble proteins by their chemical or physical attributes, like size, metal affinity, charge, recombinantly added tags and fusion parterns or hydrophobicity [95]. The basic FPLC setup consist of a pump, a column with a stationary phase and a solution of macromolecules one wishes to separate. This solution is referred to as the mobile phase. The pump injects the mobile phase into the column and the solution protein migrates through the stationary phase at different speeds, dependent on their ease of doing so.

1.4.2 Protein crystallization and structure determination

Since the first protein structure of myoglobin was solved by X-ray crystallography in 1958, X-ray crystallography has provided invaluable information of how proteins function at a molecular level [96]. X-ray crystallography as a method to determine three-dimensional structures of protein has since provided 158806 deposited structures to the Protein Data Bank (PDB), over 90 % of all

deposited structures (as of 30.07.21) [97]. While X-ray crystallography has some limitations, it is still the most common method for structural studies of proteins today.

While crystallization of proteins is a widely utilized technique and the field is well studied, the complexity of protein chemistry makes it difficult to describe a universal “recipe” for protein crystallization. Since even small proteins contains thousands of atoms which all interact with each other and their environment to varying degrees, there simply is too many variables to account for. This explains why protein crystallization is still a matter of trial and error, and is largely an empirical exercise [98]. Most protein crystal experiments rely on some sort of shotgun approach; testing a wide range of conditions which has earlier been known to crystallize other proteins, evaluate the results and then optimize conditions which produce encouraging results. This can be done more or less systematically, and commercial vendors offer many types of crystallization screens. These are based either on specific factors like pH, salt, buffer types or polymer types, or they can be a mix of historically successful crystallization conditions [99].

There are almost unlimited factors that contributes to protein crystallization, but a simplified way of describing how proteins crystallize is often visualized with the phase diagram (Figure 1-7) [99, 100]. For crystallization to occur, the protein needs to enter a supersaturated state. A supersaturated state is a non-equilibrium condition in which the molecule exist in a soluble state but is found in excess over the solubility limit for the particular solution [98]. During a vapor diffusion crystallization experiment, the protein is initially found in an undersaturated solution. The solution usually contains a buffer, a salt, a polymer or a combination of these. During the experiment, the protein and solution component concentrations will increase as a result of vapor diffusion between the protein droplet and a reservoir solution. This drives the protein into the supersaturated zones of the phase diagram. Here, it is no longer thermodynamically favourable for the protein to appear in solution. Protein crystallization occurs when the relationship between protein and precipitant is in the nucleation zone. The nucleation zone is found at higher levels of supersaturation than required for crystal growth. Here, protein molecules form aggregates, solid protein molecules which represent a transition from a

unordered molecule state to a somewhat ordered, solid state [99]. The exact way nucleation occurs is still poorly understood, and is challenging to detect because initial nuclei are undetectable by traditional microscopy [101]. As nucleation occur, the appearance of solid-state protein drives the protein concentration down and into the metastable zone. It is in this zone a nucleus which has been formed can grow by protein adhesion on the surface of the established nuclei. adding new protein molecules and form larger and larger aggregates, when crystallization is successful, these grow to large, highly ordered crystals with solvent-filled channels that are capable of diffracting when exposed to high-energy x-ray sources [99].

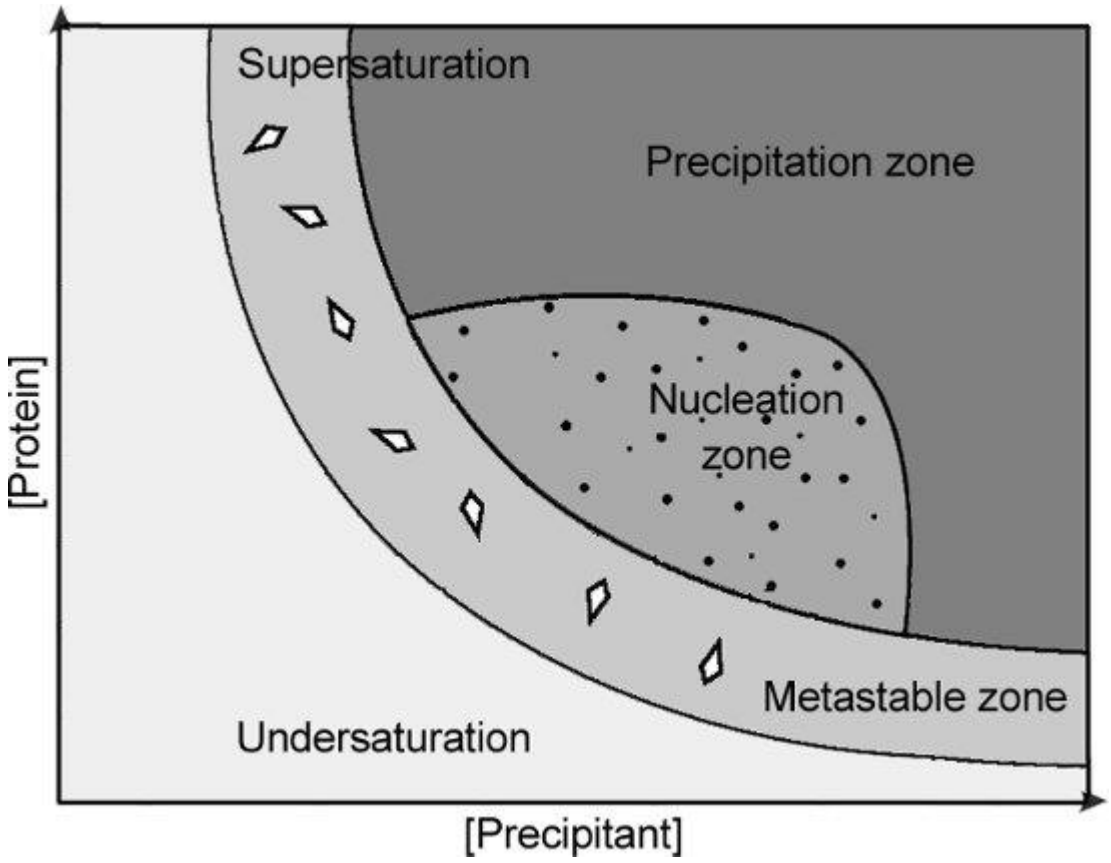


Figure 1-7: The phase diagram, a simplified way of describing protein behaviour during crystallization.The metastable zone, nucleation zone (sometimes called the labile zone) and the preceipitation zone are all representative of a state of supersaturation. Protein crystallization experiments aim to drive the protein-containing solution from a state of undersaturation, into the nucleation zone before it equilibrates in the metastable zone where crystal growth occur. A drop driven into the precipitation zone indicates that supersaturation could not stabilize and protein aggregation becomes unordered. The figure was reproduced from Pichlo et al. (2016) [102].

The presence of solvent in the cavities of protein crystals are one of the major obstacles for high resolution diffracting patterns. Conversely, they are also of high value when assessing the model of the protein, as the high solvent content ensures that protein structures resembles how they are found in solution naturally [99]. In fact, many protein crystals have been found to be catalytically active in crystal form, and ligands, substrates and co-factors may diffuse freely in and out of the crystal, making it possible to add such important chemicals to the crystal after crystallization [98]. This also highlights that protein structures are not rigid images of a structure, but rather snapshots of dynamic

structures that continuously move to some degree.

A protein crystal is often described geometrically using the unit cell. The unit cell is defined as the smallest repeating unit which can describe the entire crystal and its symmetry. The crystal lattice then becomes the repeating patterns of unit cells. A unit cell can be made up of several asymmetric units. An asymmetric unit is the smallest unit from which the entire crystal structure can be solved, and can consist of many molecules, but often only one single protein molecule [103].

In the context of structural biology, the end-goal of protein crystallization is to obtain a protein crystal that can yield high quality diffraction pattern from X-ray scattering. While many labs have in-house X-ray sources, it is often desirable to ship crystals to a synchrotron facility. Synchrotrons are particle accelerators which are capable of creating extremely high intensity X-ray beams. This allows for lower exposure times which in turn gives higher signal to noise ratios and as a result, usually data sets of higher quality [103]. During data collection, the crystal is mounted in front of the X-ray source, in a way that allows the crystal to be rotated. Behind the crystal there is a detector where the scattered X-rays are detected. When the crystal is being irradiated with high-energy X-rays, the X-rays collides the electrons in the ordered crystal, and diffracts in patterns of different intensities onto the detector. The crystal is then rotated according to an experimentally found oscillation range (usually 0.1-2 °) and a new diffraction pattern is collected. Collectively, these patterns, or reflections can be combined to create an electron density map. This electron density map is a mathematical reconstruction of all the reflections included in the data set, which describes each atoms position in the structure. To aid the researcher in placing the amino acids of the protein correctly, a previously solved structure with as high sequence similarity and coverage as possible may be used as a template. This technique is called molecular replacement, and is still a highly popular method for structure determination [103].

1.4.2.1 Dynamic Light Scattering

Dynamic light Scattering (DLS) is a simple and fast technique to determine the size and dispersity of

particles in a solution. It is widely used in several scientific fields like physics, biology and medicine. In the field of protein science, light scattering techniques has several applications and is most commonly used to confirm the size of the protein, as well as the oligomeric state [104]. In addition, DLS is useful to evaluate the protein stability during storage, as periodical measurements can identify potential increase of larger particles in the sample over time. If the population of larger particles increases, it would suggest that the protein aggregates [104]. As an extension, DLS is also useful for optimizing crystallization conditions. Firstly, protein crystallization is dependent on high purity samples to maximise the chance of growing high-quality crystals. DLS can also be used to evaluate factors that affect protein nucleation, which is a crucial step in protein crystallization

A typical DLS setup consist of a (single frequency) laser light source, which beams through a sample solution at a fixed angle. When the light from the laser hits the sample solution, molecules in the sample scatters the light and is collected by detectors. Usually DLS is set up with two detectors, one at a 90 ° angle from the light source and one at a 173 ° angle. These are called right angle and back angle detectors, respectively [105]. If there was no movement in the solution, the light would be diffracted at a constant angle over time. However, since molecules in a solution is subject to diffusion, the light scattering fluctuates. Smaller particles fluctuate more rapidly and can thereby be used to calculate the particle size in the solution. DLS measurements is taken at short time intervals, and faster moving particles will diffuse more rapidly in the solution, meaning that the correlation between measurements will be less than bigger, slower moving particles [104]. The time in which the measurements takes to no longer correlate can be used to determine the size of the particles through the Stokes-Einstein equation (E1):

E1 $$R_h \frac{kT}{6\pi n D_t}$$

Where R_h is the hydrodynamic diameter, k is the Boltzmann's constant, T is the temperature in Kelvin and n is the viscosity of the sample solvent [104]. Modern DLS machines automatically calculates the measurements and provides the size and relative distribution of particles in the sample.

While it is possible to directly calculate the molecular weight from DLS-derived hydrodynamic diameter, it is not considered a reliable or accurate method [106]. Interpretation of DLS data, especially when evaluating molecules with close molecular weights should be done with caution.

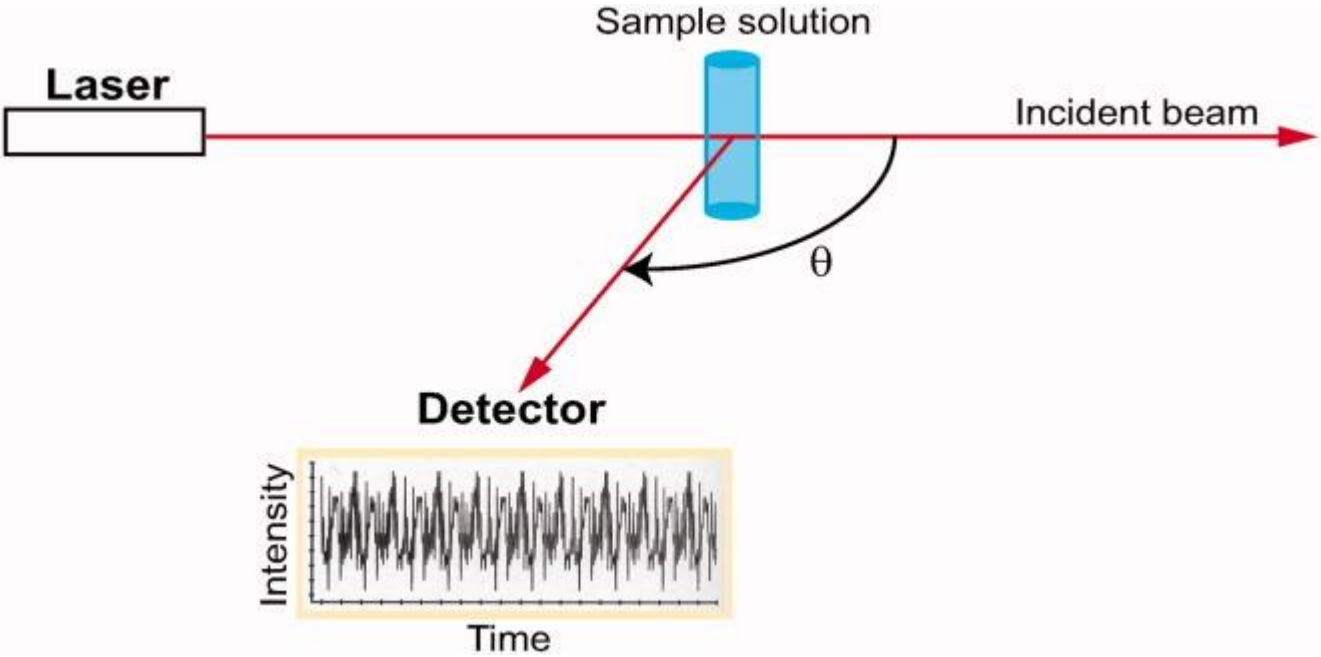


Figure 1-8: A basic representation of the components included in a DLS experiment. The angle ϑ is the angle between the laser and detector, here shown for the back angle detector. The figure is reproduced from Lorber et al. (2012) [104].

1.4.2.2 Differential Scanning Calorimetry

Calorimetry is the science of measuring heat change in a chemical reaction. Differential Scanning Calorimetry (DSC) is in this context a useful technique, with broad applications in physics, chemistry and life sciences in general. DSC is used to determine the enthalpy change (ΔH), heat capacity (C_p) and melting temperature (T_m) of macromolecules in solution. In protein chemistry, DSC is a useful method to evaluate protein stability under various conditions as well as protein folding [107].

The basis of a DSC experiment setup is to measure heat absorption differences in sample cells when the temperature of both cells is increased. The DSC apparatus consist of two sample cells, one for the

sample to be analysed and one with identical buffer as the sample to be analysed, but without the analyte. When the system increases the temperature of both cells, the sample solution and reference buffer will absorb the heat differently, according to their heat capacity. The DSC apparatus then provides additional heat energy to one of the cells to keep the temperature of each cell equal. This additional heat energy is what is measured, and since the sample cell contains protein but otherwise the cells are identical, the difference in heat applied is equal to the protein heat capacity [108]. The T_m of the protein during DSC experiments is commonly defined as the temperature where 50 % of the protein is unfolded and is found in the thermogram at the peak midpoint, when heat applied to the sample cell is plotted against the temperature [108].

1.4.2.3 Isothermal Titration Calorimetry

Another calorimetry-based technique is used to evaluate thermodynamic parameters of interactions between a macromolecule and a ligand is Isothermal Titration Calorimetry (ITC). As the name suggests, ITC involves titrating small aliquots of a ligand into a cell containing the sample of interest while keeping the sample and reference at a constant temperature. A big advantage of ITC experiments is that it enables the direct determination of reaction stoichiometry (*n),* enthalpy (*H)* and binding affinity (*K_a)[109].* In addition, changes in entropy (ΔS) and Gibbs' free energy (ΔG) can be derived by equation 2 and 3 (E2, E3):

E2 $$\Delta G = -RTln(K_a)$$

E3 $$\Delta S = \frac{(\Delta H - \Delta G)}{T}$$

Where T is the temperature in Kelvin and R is the ideal gas constant.

A typical ITC apparatus consists of two identical cells; one sample cell with an injection syringe and one reference cell which should contain the solvent. These cells are typically made of hastalloy or gold, or another efficient conductor of thermal energy, and are surrounded by an adiabatic jacket which helps keeping the cells at a constant temperature. Heat sensing devices monitors the

temperature of each cell, and heaters then offset any heat changes in the cells by either increasing or decreasing the power applied. When the ligand binds to the macromolecule, a release or absorbance of heat transfers into the ligand-protein complex, depending on whether the reaction is exothermic or endothermic[110]. This change in enthalpy, ΔH, is measured for each titration and generates either positive or negative spikes in the applied power to the sample cell to keep a constant temperature.

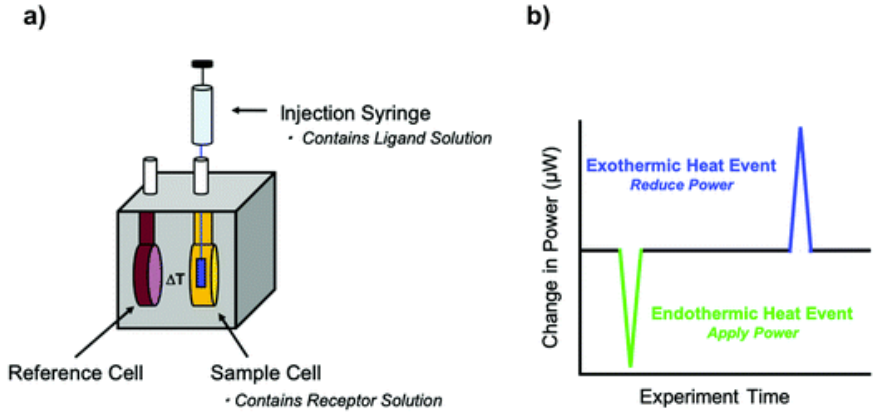


Figure 1-9: a) simplified overview over the parts of an isothermal titration calorimeter and b) a representation of experimental data spikes gathered when the apparatus increase (endothermic) or decrease (exothermic) the applied power The figure was reproduced from Archer et al. (2020) [110].

1.5 Project overview

This thesis and the work therein are part of the larger Ectoine project in the research group of Prof. Ingar Leiros. The overall goal is to functionally and structurally characterize the three main enzymes in the biosynthesis pathway of ectoine from the organisms described above.

2 Methods

2.1 Plasmids, gateway cloning and sequencing

As the basis for all downstream work with Ectoine synthase from either of our two model organisms, vectors carrying our genes with the appropriate modifications was required. To that end, the sequences for Ectoine Synthase (EctC) from *Chromhalobacter Salexigens DSM 3043* and *Marinobacter sp. CK1* were codon optimized for expression in *E. coli* expression systems. The amplicons were purchased from Gene Art and pre-cloned into pENTR221 vectors (Thermo Fisher). This vector contains attL cloning sites for recombination by the gateway cloning system and *kanR*, the kanamycin resistance gene as a selectable marker. To facilitate purification by immobilized metal affinity chromatography (IMAC) in downstream processing, a hexahistidine tag (Histag) was included at the 3'end for both constructs, corresponding to the C-termini ends for translated protein. The pDEST™14 plasmid was chosen as the expression vector. This vector contains the *AmpR* gene for antibiotic selection, *attR* recombination sites compatible with the *attL* sites in pENTR221 and an inducible *lac* operon with a T4 promoter.

2.1.1 Cloning EctC sequences into an expression vector, transformation and amplification

To produce the recombinant pDEST™14 vector with either MarEctC or CSEctC, the Gateway cloning system was used to transfer our amplicons from the entry vector to the expression vector. The entry and expression vectors were combined in Eppendorf tubes to a final concentration of 10 ng/µL, diluted in TE-buffer (pH 8) To start the reactions, 2 µL of the commercial LR-clonase enzyme mix (Invitrogen) was added in a final volume of 8 µL. The tubes were left at room temperature overnight and the reactions were terminated by adding 1 µL proteinase K (Invitrogen) and incubated at 37 °C for 10 minutes to digest active gateway cloning enzymes.

For transformation, One Shot™ TOP10 *E. coli chemically* competent cells were chosen for plasmid amplification. To single tubes containing barely thawed Top10 competent cells, 1 µL Gateway

reaction was added without mixing. The tubes were left to incubate on ice for 30 minutes and transformed into TOP10 cells by heat shock at 42 °C for 30 seconds. The tubes were left on ice for 2 minutes, before 250 µL SOC media was added, and the tubes were incubated at 37 °C, shaking at 200 rpm. The cells were plated on LB agar plates containing 100 µg/mL ampicillin and incubated at 37 °C for 16 hours. Subsequently, single colonies were picked and inoculated in 4 mL LB media with 100 ug/mL ampicillin and grown overnight at 27 °C shaking at 200 rpm. The cells were harvested by centrifugation at 4800 rpm for 15 minutes at room temperature.

2.1.2 Plasmid mini-prep and BigDye™ sequenicing

Harvested cells from overnight cultures were used to isolate plasmid DNA for verification of a successful cloning procedure. We used the E.Z.N.A.® Endo-Free Plasmid DNA Mini Kit II (Omega Bio-Tek) for plasmid isolation, following the manufacturers protocol which will be summarized here. Firstly, the cell pellets from the overnight cultures were resuspended by pipetting in 250 µL Solution 1 with RNase A added. The suspension was transferred to new Eppendorf tubes, and 250 µL Solution II was added and gently mixed. The addition of Solution II starts a lysis reaction which is then terminated by addition of 125 µL pre-chilled N3 Buffer after 3 minutes. This results in formation of a white precipitate, consisting of cell debris and lysate solution components. The tubes were subsequently centrifuged for 15 minutes at 13 800 rpm and the lysate were measured and transferred to new tubes. An equal volume ETR binding buffer were then added, and 700 µL lysate were added to a HiBind® DNA Mini Column with a 2 mL collection tube. The lysate was filtrated through the column by 1 minute centrifugation at 13 800 rpm until all lysate were loaded onto the column. The column was washed with 500 µL HBC Wash buffer by centrifugation. The wash step with 500 µL ETR buffer were omitted as this step is designed for preparing plasmid transfection by removal of endotoxins which often copurifies with plasmid DNA [111] Finally, 75 µL Milli-Q water was added to the column and incubated for 1 minute, before Plasmid DNA was eluted by centrifugation at 13 800 rpm for 1 minute into a clean Eppendorf tube. Concentration and purity were measured with a Nanodrop 2000 (Thermo Fisher) at 260 nm and by finding the ratio of absorbance at 260 nm and 280 nm.

2.2 Expression, solubility optimization and evaluation of expression output

2.2.1 EctC expression optimization

To obtain required amounts of target protein, we optimized our standard expression protocol (16 hour induction time at 25 °C with 1 mM IPTG) for the overexpression of EctC from both target organisms. Initially, we investigated protein expression as a function of time and growth temperature post-induction. The initial optimization was done identically for both target proteins.

From a pre-made glycerol stock, EctC-carrying *E. coli* was streaked onto an LB agar plate with 100 ug/mL ampicillin. The plates were incubated inverted at 37 °C for 18 hours, and then kept at 4 °C thereafter until inoculation. Two pre-cultures were established, using 50 mL TB media with 100ug/mL ampicillin. Each starter culture was inoculated with a single colony of BL21 cells with a pDEST14 expression vector carrying the EctC gene. The pre-cultures were incubated at 37 °C, shaking at 200 rpm for 18 hours, or until the mid-log phase was reached.

To evaluate the effect of expressing our proteins at different temperature post-induction, three replicates of 300 mL TB media with 100 ug/mL ampicillin in 2.5 L baffled plastic flasks was made. The media was inoculated with 3 mL pre-culture each and set to grow to an OD600 of 0.8-1.1 at 37 °C, shaking at 200 rpm. When the target OD was reached, 1 mM IPTG was added to each flask to induce production of EctC. The flasks were incubated at 15, 20 or 25 °C, to evaluate EctC expression at different temperatures, shaking at 200 rpm. At induction, 1 mL samples were taken, as well as after 1, 2.5, 4 and 18 hours post-induction, for downstream analysis of EctC expression at different timepoints. After 18 hours, all cultures were harvested by centrifugation at 6000 rpm for 30 minutes. The supernatants were sampled (1 mL) and discarded, and the pellets were resuspended in 40 mL lysis buffer (50 mM HEPES pH 7.5, 5000 mM NaCl, 15 mM Imidazole) by gentle pipetting. For storage, the resuspended pellets were flash frozen in liquid nitrogen, and stored at - 80 degrees until further use.

2.2.2 Evaluating EctC expression by SDS-PAGE

To evaluate the effect of time and temperature on the expression cultures of EctC, a number of gel electrophoresis analyses were conducted. Samples taken at different stages in the expression process were thawed on ice, and a suitable quantity (100-200 uL) was taken out into new 1.5 mL Eppendorf tubes. To quickly lyse a large number of cell cultures, the samples were put through a freeze-thaw cycle; each tube was incubated in liquid nitrogen (- 196 °C) for 30 seconds, then thawed in a water bath at 42 °C for 30 seconds. This procedure was repeated 5 times for each sample. To separate the lysate into soluble and insoluble fractions, the samples were centrifuged at 14 800 g for 15 minutes at 4 °C. The insoluble, pelleted fractions were resuspended in equal 3x or 5x volume lysis buffer (50 mM HEPES, 500 mM NaCl, 15 mM imidazole).

From each sample, 20 µL sample was mixed with 4 µL 6x loading dye, and the sample tubes were heated to 95 °C in a thermal block to denature any proteins in the tubes. The samples were loaded to 10-, 12-, or 15-well 4-20% Mini-Protean® TGX™ protein gels (Bio-Rad) and run at 240 V for 30 minutes, using a Tris-Glycine running buffer. The high voltage was due to the low theoretical molecular weight of our target proteins (15.4 and 15.6 kDa, respectively), to ensure large enough separation between small proteins to distinguish between our target and other small proteins from BL21 cells. The gels were washed with Milli-Q water and stained with Simply Blue™ Safe Stain Coomassie G-250 stain (Thermo Fisher) for 5 minutes. The gels were de-stained in Milli-Q water until protein bands was clear enough to be imaged. The gels were imaged with a Gel Doc XR, using the Quantity One software (both from Bio-Rad).

2.2.3 Optimization of MarEctC solubility during expression

In an effort to improve the solubility and overall production yield of MarEctC, we designed a solubility optimization experiment, based on the previous MarEctC expressions. Herein, the effect of the inducer concentration and the effect of an osmoprotectant (IPTG and sorbitol, respectively) were examined. From a MarEctC pre-culture (described in 2.2.1), four 300 mL TB media flasks (Table 2-1) were inoculated with 3 mL pre-culture. The cultures were incubated at 37 °C, shaking at 200

rpm until OD_{600} measured 0.7-0.8. At this point, MarEctC expression was induced by either 1 mM or 0.3 mM IPTG, and the cultures were incubated at 200 rpm, 20 °C for 18 hours. Samples from the induction timepoint and at harvest were analyzed by SDS-PAGE as described previously, to analyze the effect of inducer concentration and the presence of sorbitol.

Table 2-1: Cell cultures composition for initial solubility optimization trials. The culture were induced at 20 °C for 18 hours.

Culture ID	**E1**	**E2**	**E3**	**E4**
Conditions	TB-Media, 1 mM IPTG	TB, 0.3 mM IPTG	TB, 400 mM Sorbitol, 1 mM IPTG	TB, 400 mM Sorbitol, 0.3 mM IPTG
Induction OD	0.84	0.85	0.78	0.80

For a second solubility optimization experiment, a broader range of sorbitol and inducer concentration was tested at 15 °C, to evaluate if a lower expression temperature would influence the effect the inducer concentration or an osmoprotectant would have on solubility and expression of the target enzyme. Nine 300 mL cultures as listed in Table 2-2 was inoculated with 3 mL pre-culture grown overnight, and incubated shaking at 200 rpm at 37 °C until the OD_{600} reached 1-1.3. MarEctC transcription was induced by 1, 0.5 or 0.1 mM IPTG, and the cultures were harvested after 18 h induction time by centrifugation. At 6000 rpm for 40 minutes at 4°C. Samples were taken at harvest and evaluated by SDS-PAGE.

Table 2-2: Overview of sorbitol concentration, inducer concentration and OD_{600} at induction time for solubility optimization trials conducted at 15 °C for 18 hours.

Culture ID	G1	G2	G3	G4	G5	G6	G7	G8	G9
[Sorbitol] (M)	0	0	0	0.25	0.25	0.25	0.5	0.5	0.5
[IPTG] (mM)	1	0.5	0.1	1	0.5	0.1	1	0.5	0.1
Induction OD	1.12	1.14	1.12	1.21	1.18	1.16	1.20	1.21	1.15

2.2.4 Finding MarEctC migration distance on SDS-PAGE – HisTrap purification of pooled MarEctC cultures

Initially, MarEctC proved difficult to identify by SDS-PAGE, making any evaluation of temperature and time effect on its expression efficiency difficult. To identify the migration distance of MarEctC in our current experimental SDS-PAGE setup, we decided to pool three 300 mL cultures grown at different temperatures and remove contaminants by IMAC chromatography. We thus aimed to verify the presence of MarEctC in our expression cultures and identify it on acrylamide gels.

Six harvested cell culture pellets were thawed and resuspended in 15 mL modified lysis buffer (50 mM HEPES, 500 mM NaCl, 15 mM imidazole, 0.5 mg/mL lysozyme, and $1/6^{th}$ proteinase inhibitor cocktail tablet (EDTA free)) each on a horizontal shaker at 4 °C. A 50 mL pipette was used to resuspend any small lumps, before all resuspended pellets were pooled together. To lyse the cells, the pre-lysate was sonicated at 25 % amplitude for 40 minutes, with 9 seconds on and off cycles while placed on ice. After sonication, the lysate was centrifuged at 20 000 rpm for 30 minutes at 4 °C. The pelleted cell debris was discarded, and the lysate was kept at 4 °C until further processing.

A 1 mL HisTrap™ FF (GE Healthcare) column was pre-equilibrated with 5 column volumes (CV) Lysis buffer (50 mM HEPES, 500 mM NaCl, 50 mM imidazole) and approximately 100 mL lysate was loaded onto the column using an Äkta Pure (GE Healthcare). EctC was eluted with a 15-500

mM imidazole gradient over 40x CVs, collecting 0.5 mL fractions. The untypically long gradient was applied to maximize the resolution of the possible peaks, as multiple peaks were expected from previous test purification. To visualize the content of eluted protein, selected fractions were analysed by SDS-PAGE, as described in 2.2.2. A band with appropriate migration distance was used as a reference for further analysis of previous or future MarEctC gels.

2.2.5 Mass Spectrometry

To determine the identity of two bands produced by SDS-PAGE from MarEctC expression lysate, these bands were cut out of the gel and sent to digestion followed by liquid chromatography Mass spectrometry (LC-MS) at the Tromsø University Proteomics Platform (TUPP). The in-gel digestion was done enzymatically with trypsine and the resulting peptides were analysed with UHPLC-ESI-MS to identify the sample proteins. For possible verification of MarEctC, the amino acid sequence of the target protein was provided.

2.3 Protein purification

2.3.1 Purification of CSEctC

Frozen CSEctC pellets corresponding to 2 L cell culture were incubated with 50 % w/v adapted lysis buffer (50 mM HEPES, 15 mM imidazole, 500 mM NaCl and 1 protease inhibitor cocktail tablet/1 L cell culture) on a shaking incubator until the pellets were fully resuspended. To complete cell lysis, the resuspended pellets were sonicated with a sonicator at 25 % amplification for 40 minutes, with 9 seconds on/off intervals while on ice. To prevent overheating of the lysate, the maximum temperature was set at 12 °C, measured with a temperature probe continuously. Cell debris were removed by centrifugation at 14 000 xg for 30 minutes, keeping the centrifuge at 4 °C. The lysate was transferred to a new container and the cell debris were discarded.

A 5 mL HisTrap FF attached to an Äkta Pure FPLC machine was washed with water and equilibrated with 5x CV of lysis buffer. The lysate filtered through 0.22 µM sterile syringe filter (Merck) and then loaded onto the column at 1 mL/min, to ensure proper binding of the HisTag to the

nickel in the column resin. To remove weakly binding contaminating proteins, a wash step was introduced, using 5 CV wash buffer (50 mM HEPES, 75 mM imidazole, 500 mM NaCl). The remaining protein in the column was eluted with a 75-500 mM imidazole over a 12x CV gradient, collecting 1.5 mL fractions over the entire gradient. Relevant fractions according to the chromatogram were analysed by SDS-PAGE as described earlier, and the fractions of interest were combined for further purification with Size Exclusion Chromatography (SEC).

To further purify our target protein, a HiLoad 26/600 75 pg column (GE Healthcare) were equilibrated with 1x CV Milli-Q water followed by 1x CV storage buffer (500 mM NaCl, 50 mM HEPES pH 7.5, 0.4 mM TCEP. The combined HisTrap fractions with protein of interest were concentrated with a 3 kDa spin filter (Amicon) by centrifuging the filter at 4800 xg until the total volume was 10 mL. The sample was injected into a 10 mL loop and loaded onto the column. Fractions of 1 mL were collected while running a 2 mL/min flow of 1.2x CV storage buffer through the column. The peak fractions according to the chromatogram were analysed by SDS-PAGE, and fractions with protein of interest were combined for downstream analysis and handling. The protein concentration was calculated with Nano Drop 2000 by measuring absorbance at 280 nm and calculated with proprietary software using the equation (E4):

$$[Protein]\left(\frac{mg}{mL}\right) = \frac{Absorbance280 \times mW}{extinction\ coefficient(\Delta\varepsilon)} \quad \text{E4}$$

Where mW is the protein molecular weight. The theoretical molecular weight and extinction coefficient was found using the ExPasy ProtParam tool [112]. The protein was stored at 4 °C for further use.

2.3.2 Purification of MarEctC

Since the expression levels of MarEctC was less than ideal and optimization efforts were unsuccessful, it was decided to upscale production and purify from 6 L bacterial culture. MarEctC was used for purification was expressed at 20 °C for 18 h in TB media as described in section 2.2.1.

Frozen bacterial pellets with MarEctC were thawed at 4 °C on a horizontal shaker with 50 % w/v adapted lysis buffer (50 mM HEPES, 500 mM NaCl, 15 mM imidazole, 1 protease inhibitor cocktail tablet/1 L culture). When completely thawed, any lumps were resuspended with a 50 mL pipette. The cells were lysed by sonication at 30 % for 30 minutes with the pulse set to 5 seconds on/5 seconds off. The temperature was monitored and kept under 12 °C under sonication. The lysate was centrifuged at 14 000 xg for 30 minutes at 4 °C to remove cell debris. The supernatant from centrifugation was transferred to a clean container and the cell debris was discarded.

A 5 mL HisTrap FF (Cytiva Life Sciences) was washed with water and then pre-equilibrated with 5x CV lysis buffer. The lysate was loaded onto the column at 1 mL/min to ensure proper binding to the column. Because of the large lysate volume, the column was periodically washed with lysis buffer to flush out weak or non-bonded contaminants. To further remove contaminants, the column was washed with 10x CV wash buffer (50 mM HEPES, 500 mM NaCl, 50 mM Imidazole). The remaining protein in the column was eluted over a 50-500 mM imidazole gradient, over 20x CV while fractionating the elute in 1.5 mL fractions. Peak fractions were analysed by SDS-PAGE and relevant fractions were combined for downstream handling.

To further purify MarEctC, a HiLoad 26/600 200 pg Size Exclusion Column was washed with water and then equilibrated with 1x CV of storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP). The pooled fractions from HisTrap purification were concentrated with a spin filter (Amicon) with a 3 kDa cut off until the total concentration was 10 mL. The sample was added to a 10 mL loop with a syringe using drop-to-drop technique to avoid introducing air into the system, and the sample was subsequently included into the system flow. The flow was set at 2 mL/min and 1.2x CV storage buffer was passed through the column while collecting 5 mL fractions at the anticipated elution volume for MarEctC. Peak fractions were analyzed by SDS-PAGE and fractions with the protein of interest were combined in a 50 mL falcon tube. The protein concentration was calculated with Nano Drop 2000 as described above and stored at 4 °C until further use.

2.4 Protein crystallization and structure determination

2.4.1 Initial crystallization screening with EctC

Purified protein was divided into two fractions, and the fractions were concentrated with Amicon 3kDa cut off spin filters to a final concentration of 6 and 12 mg/mL, respectively. Prior to crystallization, EctC was incubated with $Fe(II)Cl_2$ to a final concentration of 4 mM for at least 30 minutes. The protein solution was centrifuged at 13 800 rpm for 10 minutes to separate $Fe(II)Cl_2$ precipitate, and soluble protein was transferred to a new Eppendorf tube. Commercial screens Pact Premier, SG1-Eco Screen and JCSG-plus from Molecular Dimensions were set up with the NT8 crystallization robot (Formulatrix) and Rock Maker Crystallization software. The vapor diffusion experiments were set up using the sitting drop method in 96-well MRC 2 well UV compatible plates (Swissci). The plates were incubated in a Rock Imager 54 (Formulatrix) at 22 °C, and images were taken of each drop periodically. This was done identically for both EctC candidates.

2.4.2 Crystallization optimization of initial hits from CSEctC

Promising results and hits from commercial screens were optimized by vapor diffusion in 24-well VDX hanging drop plates (Hampton Research). Table 2-3 summarizes the efforts to optimize native CSEctC crystals.

Table 2-3: Overview of the optimization trials conducted to improve size and shape of CSEctC crystals. SPG (Succinic Acid, Sodium Dihydrogen Phosphate and Glycine in a 2:7:7 Molar Ratio), MIB (Malonic Acid, Imidazole and Boric Acid in a 2:3:3 Molar Ratio) and MMT (Malic Acid, MES and TRIS in a1:2:2 Molar Ratio) are buffer systems which can be adjusted to a broad range of pH values without altering the chemical composition or ratios of components PP: Pact Premier.

Initial hit	Drop volume	Reservoir volume	Reservoir solution
PP1-1: 0.1 M SPG buffer (pH 4.0), 25 % w/v PEG 1500	4 µL	240 µL	0.1 M SPG buffer (pH 4.0), 17-30 % w/v PEG 1500
PP1-13: 0.1 M MIB buffer (pH 4.0), 25 % w/v PEG 1500	4 µL	240 µL	0.1 M MIB buffer (pH 4.0), 17-30 % w/v PEG 1500
PP-1-37: 0.1 M MMT buffer (pH 4.0), 25 % w/v PEG 1500	4 µL	240 µL	: 0.1 M MMT buffer (pH 4.0), 17-30 % w/v PEG 1500

Optimization experiments were designed with ROCK MAKER® software, and components for crystallization were dispensed with the FORMULATOR® screen builder. From the reservoir solution, one 2 µL drop was pipetted onto an 18 mm cover slide (Hampton Research) and mixed with an equal volume 11.6 mg/mL protein solution. The drops were mixed by slow pipetting, and the cover slide were sealed over the reservoir solution with the drop facing inwards. The plates were left at room temperature and inspected periodically.

In an effort to produce crystals with either N-γ-ADABA or ectoine bound in the active site, we attempted to co-crystallize CSEctC with the substrate or product, respectively at the optimum conditions. To achieve this, fractions of 11.6 mg/mL CSEctC was pre-incubated with N-γ-ADABA or ectoine in 1:10 and 1:30 molar ratios for both additives for at least 30 minutes prior to setting up crystallization plates. Crystallization experiments were set up in two 24-well plates, one with N-γ-ADABA and one with ectoine. The reservoir solution in each plate contained 0.1 M MIB buffer (pH

4.0) with 27-32 % w/v PEG 1500 in replicates of four. The molar ratios were set up in duplicates across the PEG 1500 gradient and the plates were kept at room temperature and inspected periodically.

2.4.3 Optimization of hits from MarEctC crystallization screening

Promising hits from the initial commercial screens with MarEctC was also attempted optimized. As the initial screens were set up only ten days before shipping of crystals from the NorStruct lab to a synchrotron facility, this was done under some time pressure. Potential conditions were identified manually by microscopy, and optimization experiments were designed in RockMaker® (

Table 2-4). Due to low remaining amounts of MarEctC left, we decided to use one concentration of 9.5 mg/mL for all experiments.

Table 2-4: Overview of the optimization trials conducted to improve size and shape of MarEctC crystals. Salt concentrations increased across columns and PEGs increased across rows. To maximise number of optimization trials with a small remaining batch of MarEctC, the bottom two experiments were set up with only 2 µL drop volume (1 µL reservoir solution and 1 µL protein solution). PP: Pact Premier.

Initial hit	Drop volume	Reservoir volume	Reservoir solution
PP 1-8: 0.1 M Sodium Acetate pH 5.0, 20 % w/v PEG 6000, 0.2 M NH_4Cl	4	240	0.1 M sodium acetate pH 7.0, 15-30 % PEG 6000, 0.1-0.3 M NH_4Cl
PP 1-33: 0.1 M HEPES pH 7.0, 20 % w/v PEG 6000, 0.2 M LiCl	4	240	0.1 M HEPES pH 7.0, 15-30 % PEG 6000, 0.1-0.3 M LiCl
PP 1-34: 0.1 M HEPES pH 7.0, 20 % w/v PEG 6000, 0.2 M $MgCl_2 \cdot 6H_2O$	4	240	0.1 M HEPES pH 7.0, 15-30 % PEG 6000, 0.1-0.3 M $MgCl_2 \cdot 6H_2O$
PP 1-44: 0.1 M Tris pH 8.0, 20 % w/v PEG 6000, 0.2 M NH4Cl	4	240	0.1 M Tris pH 8.0, 15-30 % w/v PEG 6000, 0.1-0.3 M NH4Cl

PP 2-29: 0.1 M Bis-Tris propane pH 7.5, 20 % w/v PEG 3350, 0.2 M $NaNO_3$	4	240	0.1 M Bis-Tris propane pH 7.5, 15-30 % w/v PEG 3350, 0.1-0.3 M $NaNO_3$
SG1 1-41: 0.1 M Hepes pH 7.5, 25 % w/v PEG 3350, 0.2 M $MgCl_2 \cdot 6H_2O$	4	240	0.1 M Hepes pH 7.5, 15-30 % w/v PEG 3350, 0.1-0.3 M $MgCl_2 \cdot 6H_2O$
SG1 1-42: 0.1 M Hepes pH 7.5, 25 % w/v PEG 3350, 0.2 M $LiSO_4$	4	240	0.1 M Hepes pH 7.5, 15-30 % w/v PEG 3350, 0.1-0.3 M $LiSO_4$
SG1 1-43: 30 % w/v PEG 8000, 0.2 M NH_4SO_4	4	240	20-37 % w/v PEG 8000, 0.1-0.25 M NH_4SO_4
SG1 2-2: 25 % w/v PEG 3350	2	240	15-30 % w/v PEG 3350
SG1 2-26: 20 % w/v PEG 3350, 0.2 M $NaSO_4$	2	240	15-30 % w/v PEG 3350, 0.1-0.3 M $NaSO_4$

Components for crystallization were dispensed with the FORMULATOR® screen builder. From the reservoir solution, one 1 or 2 µL drop was pipetted onto an 18 mm cover slide (Hampton Research) and mixed with an equal volume 9.5 mg/mL protein solution. The drops were mixed by slow pipetting, and the cover slide were sealed over the reservoir solution with the drop facing inwards. The plates were left at room temperature and inspected periodically.

2.4.4 Data collection, processing, refinement and structure determination

Selected crystals from optimization experiments with CSEctC were harvested with sample pins, washed in a cryo solution (reservoir solution with 20 % (v/v) glycerol) and mounted in CryoCaps (Molecular Dimensions). Mounted crystals were transferred to Universal V1-Pucks (Molecular

Dimensions) and stored in liquid nitrogen for preservation and preparation for shipping to the ESRF synchrotron facilities in Grenoble. Data sets from shipped crystals were collected at the ID30B beamline. From one native-form CSEctC crystal, data was collected to 1.2 Å resolution, and autoprocessed with XDSAPP (*Krug et al. 2012 and references therein)* [113]. The data was further processed using the CCP4 software suite [114]; The data was re-processed in-house using Aimless [115] in order to optimize the number of images and resolution range, truncate multiple observations and scale multiple observations of reflections.

The phase problem (section 1.3.5) was solved in MolRep [116] using molecular replacement with a crystal structure of EctC from *Paenibacillus lautus* (PDB: 5ONM) as the template [76]. During data processing, we determined that our data set represented one of two possible space groups: I 222 and I $2_12_12_1$. Since it is impossible to distinguish these from one another as body-centring (I-centring) hides the possible screw axis of I $2_12_12_1$. Thus, this was solved experimentally by trial and error, and both space groups were tried during molecular replacement. To refine the positional refinement prior to structure building, Refmac5 [117] was run using default settings. Using the Buccaneer software [118, 119], an initial structure with 116 amino acids was autobuilt with R_{work}/R_{free}- values of 26.86/29.09. The model was improved with manual model building in Coot [119], and then several cycles of Refmac5 refinement with manual model building interspersed. The last cycle of Refmac included refinement of anisotropic B-factors, bringing the final model to 117 amino acid residues and R_{work}/R_{free} of 13.32/16.82. The final model was evaluated and analysed with PYMOL [120]. Data re-processing, refinement and structure solving was jointly done with prof. Ingar Leiros.

2.5 Bioinformatical analysis and homology modelling of EctC

To identify universally conserved and other residues of interest in the EctC sequence, a multiple sequence alignment (MSA) was generated. The amino acid sequence of *S. alaskensis* EctC (SAEctC) was chosen as a template and uploaded to the ConSurf server, searching for homologs in the Uniprot database [121, 122]. Default settings were otherwise applied. Homologs of the uploaded sequence were automatically collected by the ConSurf server. The generated MSA were analysed with

UGENE and visualized with ESPript 3.0 [123, 124]

2.5.1 Homology modelling of MarEctC

A homology model of MarEctC was created with the open access server SWISS-MODEL[125]. The MarEctC sequence was uploaded to the server, and the in-house CSEctC structure was chosen as a template. The model was built according to the OpenStructure framework and the ProMod3 modelling engine which the SWISS-MODEL service relies upon [126]. The model was evaluated manually in PYMOL.

2.6 Biochemical characterization

2.6.1 Dynamic light scattering

The size and oligomeric state of the protein in solution was verified by DLS. A quartz cuvette with a 0.5 cm slit was filled with 16 µL CSEctC protein solution (1.5mg/mL) in the storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP). The cuvette was placed in a DynaPro-99-E-50 DLS module (Protein Solutions). For the measurement of light scattering, the scattering angle was set to 90 ° per the default setting of the apparatus, and the laser wavelength was set to 830 nm. Light scattering measurements was acquired ten times across three separate measurements, and the mean of each measurement was calculated. This measurement was only done for CSEctC due to time constraints.

2.6.2 Differential scanning calorimetry

Differential scanning caliometry was used to assess thermal stability of EctC in our chosen storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP). To conduct the experiments, an N-DSC III instrument (Calorimetry Scanning Corporation) and analysis software NanoAnalyzer 3.6 was used. All solutions were filtered through 0.22 µM sterile syringe filters. Buffers, protein solutions and Milli-Q water was degassed for 15 minutes prior to usage.

Firstly, the DSC apparatus were pre-washed with water. Both cells were washed with 1000 mL Milli-Q water, and Milli-Q water was then loaded in both cells. The apparatus was pressurized to 3

atm and the cells were equilibrated to the starting temperature for 10 minutes. The samples were then heated from 20-85 °C, with an increment of 1 °C/minute, while measuring energy input to both cells. This wash cycle was repeated 2 times overnight. To establish a buffer baseline, we ran an identical cycle the next day with Storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP) in both chambers. The baseline was established for instrument calibration and to verify suitability of the buffer system used.

Finally, EctC from both model organisms was tested in the storage buffer using the same experimental parameters as in the initial runs with water and buffer-buffer. The protein thermal stability experiments were done as single experiments only, with no replicates. During the experiment, the concentrations of CSEctC and MarEctC was 1.1 mg/mL and 1.4 mg/mL, respectively. The DSC instrument was cleaned with 50 % formic acid using a standardized protocol at our lab. The raw data from the DSC experiments were fitted using the two-state transition model, and the melting temperature (T_m) was found at the transition midpoint temperature according to the graph peak.

2.6.3 Preliminary isothermal titration calorimetry with CSEctC

To further evaluate the thermodynamic properties of EctC, we attempted to develop a protocol to perform ITC experiments with CSEctC. MarEctC was not evaluated with ITC due to time constraints, limited amounts of protein available and the lack of a standardized protocol with CSEctC. All buffers, protein solution, substrate solutions and water used was degassed for at least 10 minutes prior to usage.

Prior to producing a usable data set a series of optimization experiments were conducted, to find suitable concentration and molar ratios between protein and substrate (Appendix D). This optimization was partly done by PhD Heidi T. Hillier. All experiments were conducted using a MicroCal™ ITC200 system (GE Healthcare) and data was analysed with Origin® analysis software (OriginLab). The experiments were all conducted at 25 °C, while stirring the sample cell at 450 rpm.

Over 7 individual experiments we optimized protein and substrate concentrations, as well as number of substrate injections. CsEctC concentrations varied from 8-150 μM, ADABA concentrations from 2-6 mM and number of injections from 44-120 (Appendix D).

The protein was dialysed overnight in the sample buffer (50 mM Tris pH 9.0, 500 mM NaCl, in a 1 L buffer volume while slowly stirring with a magnet mixer. Prior to the experiments, the sample and buffer cells were washed with milli-Q water, then with the sample buffer (50 mM Tris pH 9.0, 500 mM NaCl. The glass syringe which holds the titrant was washed with milli-Q water according to the instrument standard protocol, and then filled with a sufficient amount of diluted substrate in the sample buffer. In each experiment, the initial delay was set to 250 seconds to equilibrate the protein and sample cells to the correct temperature (25 °C). In the final experiment the substrate was added in 0.5 μL aliquots with 350 seconds spacing between each injection, for 44 total injections. The data was fitted with the one-site binding model, and thermodynamic parameters was calculated by the Origin® software.

2.6.4 Evaluation of activity

Preliminary activity data from CSEctC and MarEctC was collected by PhD Heidi Hillier as a part of the overall Ectoine project during this thesis period and is thus presented here. Activity evaluation primarily focused on *in vitro* effect of pH in the reaction buffer.

In a final reaction volume of 200 μL, 10 mM ADABA, 500 mM NaCl, 10 μM $FeCl_2$ and 50 mM HEPES/CHES buffer with pH ranging from 6.5-9.5 were mixed to make up the reaction buffer with substrate. Reactions were initiated by adding CSEctC or MarEctC to a final concentration of 10 μM. The reactions were incubated at 30 °C for 2 minutes and terminated with 400 μL 50 % methyl cyanide (MeCN). The reactions were centrifuged at 21 000 rpm at 4 °C for 60 minutes to remove inactivated enzyme, and 450 μL of each reaction supernatant were transferred to a High-Pressure Liquid Chromatography (HPLC) vial for product detection.

Products were detected using a Trinity P3 HPLC column (Thermo Fisher). The supernatant was analysed with a HPLC-MS system at the department of Organic Chemistry, UiT. The mobile phase consisted of 7.3 mM ammonium formate in MeCN (pH 3.65), and column content was eluted by isocratic elution at 250 μL/min. Ectoine was detected at 360 nm, and the molecular weight of the peak were verified by Mass spectrometry. Product were quantified by integrating the peak area.

3 Results

3.1 Gateway cloning, plasmid miniprep and sequencing

After cloning, transformation and subsequent growth on LB agar plates, both candidates showed good bacterial growth, while a negative control plate showed no growth over the same time period. After isolation, the DNA concentration was measured by absorbance at 260 nm and the quality estimated by the 260/280 nm ratio (

Table 3-1). The concentrations measured at 14.4 and 14.8 for CSEctC and MarEctC respectively, was at the low end of what could be expected from 4 mL cell culture. The manufacturer (Omega Biotek) states that from 4 mL cell culture with bacteria containing a low-copy plasmid, one could expect yields from 1-15 µg. The spin columns are optimized for elution at pH over 8.5, and while the protocol suggest nuclease-free water as a appropriate elution agent, the pH of the water used may not have been at the ideal pH for this application. In addition, the protocol mentions an optional equilibration step of the spin columns, which was omitted. For future plasmid minipreps, this should be included.

Table 3-1: yield and quality of plasmid after miniprep isolation. The volume shown is the eluted volume.

Gene	Volume (µL)	concentration(ng/uL)	Total yield (µg)	$A_{260/280}$ nm ratio
CSEctC	75	14.4	1.08	1.88
MarEctC	75	14.8	1.1	1.91

The $A_{260/280}$ nm ratio is commonly used as a measure of purity, since it quantifies the amount of aromatic amino acids present in the sample compared to DNA. For DNA samples, a ratio of 1.80 is considered pure [127]. Our constructs were therefore considered to be of high enough quality for downstream use.

As a second quality control, we amplified the EctC amplicons and sent the samples for sequencing to

confirm the identity of our recombinant genes. The sequences retrieved confirmed successful cloning of both amplicons.

3.2 Protein expression, optimization and preliminary evaluation

3.2.1 Expression optimization of CSEctC and MarEctC

CSEctC was initially expressed in 300 mL cultures at 25, 20 and 15 °C. Samples taken at different timepoints were visualized by SDS-PAGE, showing proteins from both the soluble and insoluble fractions. In Figure 3-1, a protein band at the expected size of CSEctC is clearly visible at all temperatures after 18 hours induction. An overexpressed band is also present at 31 kDa, corresponding to the size of the *AmpR* product β-lactamase, which provides resistance against the chosen antibiotic ampicillin. While EctC is overexpressed at all evaluated temperatures, there seems to be a slight advantage in quantity of EctC when expressed at 25 °C. In addition, there is only negligible amounts of EctC present in the insoluble fraction after 18 hours at 25 °, which suggests that almost all of the target protein is soluble at this time and temperature. At 15 and 20 °C, there are visibly more target protein in the insoluble fractions.

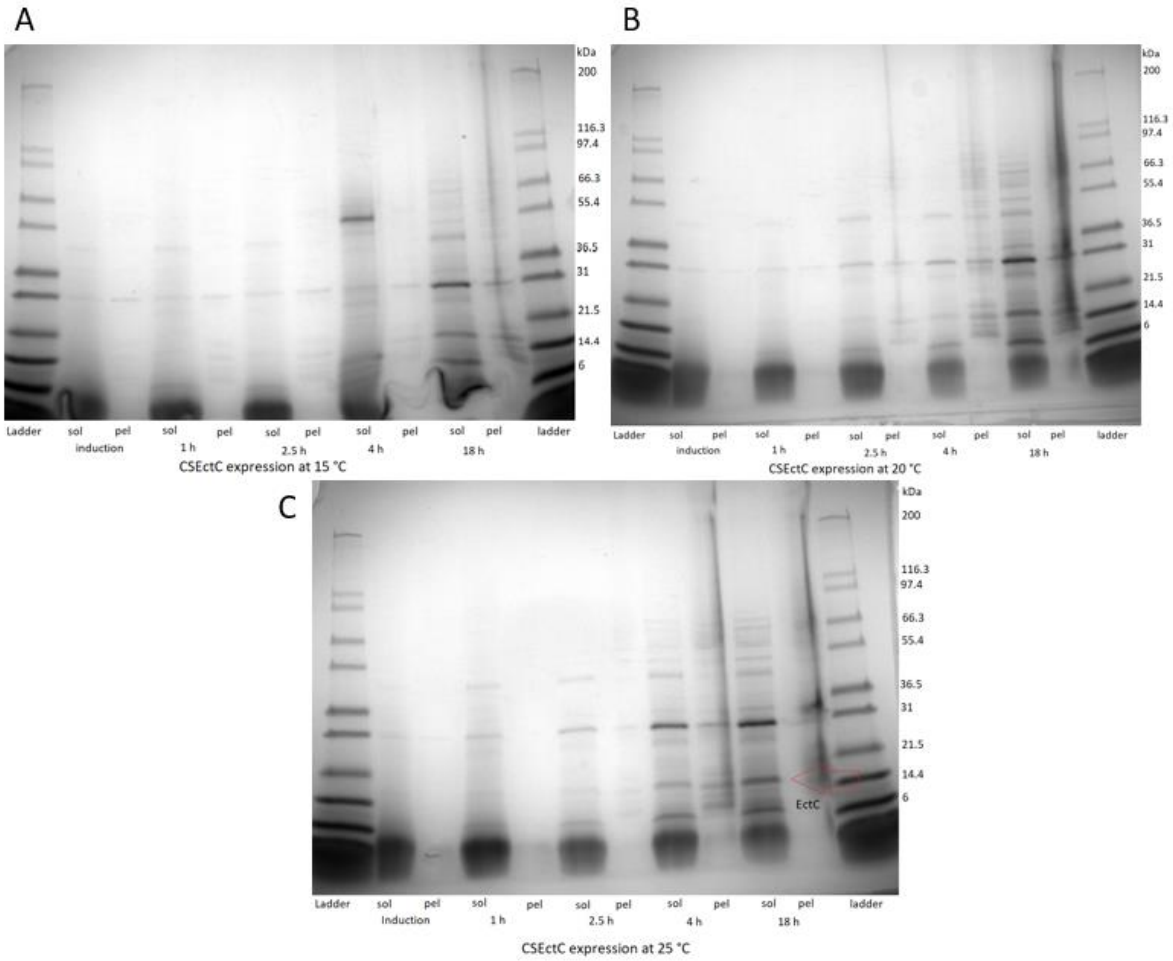


Figure 3-1: Panel view of soluble (sol) and unsoluble pellet fractions (pel) from expression of CSEctC at 15 °C (A), 20 °C (B) and 25 °C (C), with samples taken at five different time points. For reference, two Mark 12™ (Thermo Fisher) protein standards were loaded for each gel and annotated with their molecular weight (kDa).

When comparing the presence of EctC at different time points across the three temperatures, it is apparent that EctC increases in concentration at all time points as expected and is still induced after 18 hours. This allows CSEctC to be expressed overnight, and based on this expression optimization, it was decided to express CSEctC at 25 °C for 18 hours for large-scale production.

The initial optimization of MarEctC expression was done identically as with CSEctC in 3.2.1. The gels (Figure 3-2) show that MarEctC may have been expressed, but not to the degree desirable for

downstream production and purification. At the expected migration distance for MarEctC with a molecular weight of 15.4 kDa there are three bands in close proximity, who all show similar band intensity (Figure 3-3). At all three temperatures, β-lactamase seems to be overexpressed, which suggests that the expression vector is present and functional. There seems to be a slight but detectable difference in overall protein amount at increasing temperature, which is to be expected. The three bands close to the molecular weight of MarEctC seems to be quite equally expressed, but at 20 °C, there seems to be a slight but detectable increase in specific protein production of these three bands. It also seems to be a higher quantity of protein in the insoluble protein fractions than in the CSEctC optimization experiments.

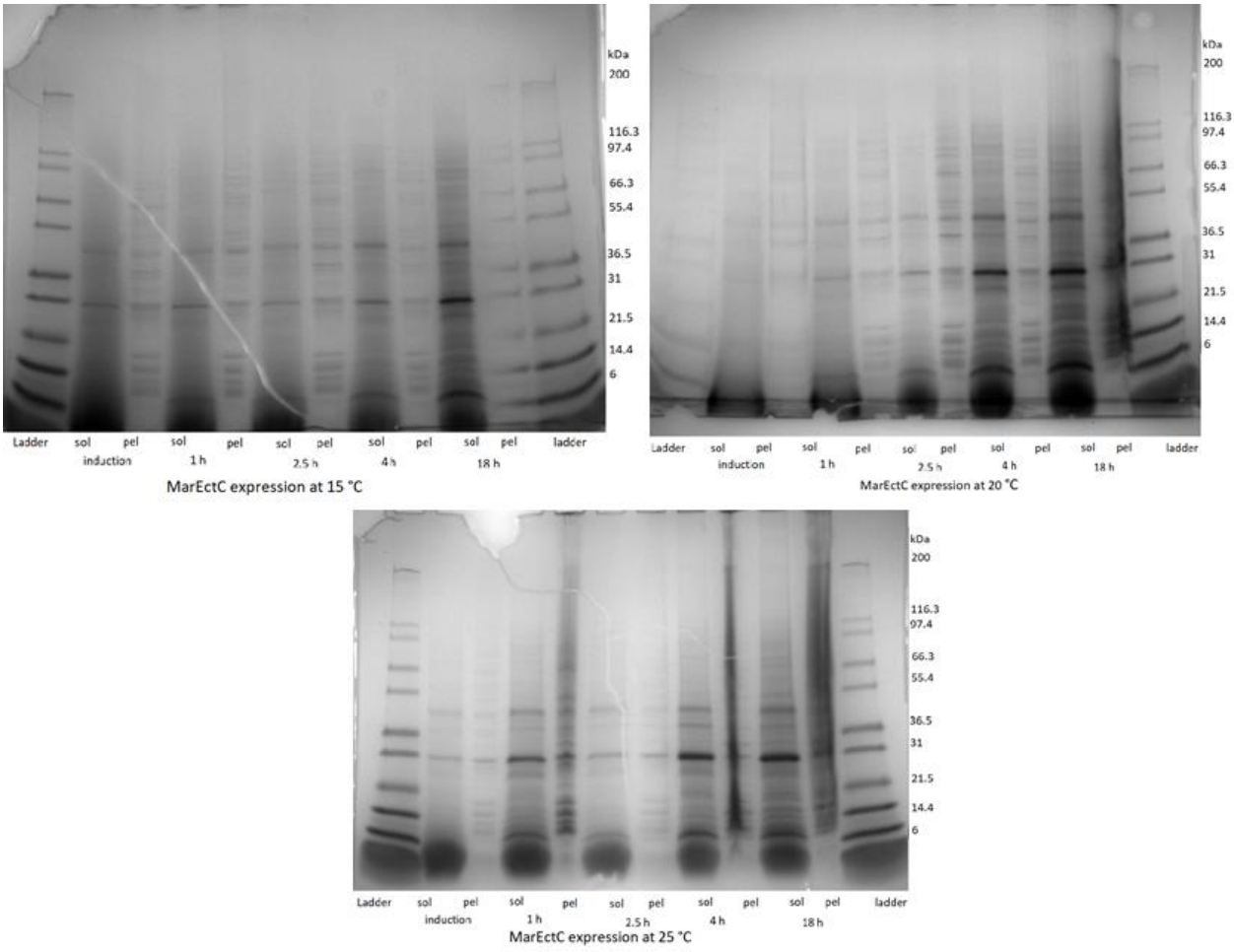


Figure 3-2: Panel view of soluble (sol) and insoluble pellet fractions (pel) from expression of MarEctC at three different temperatures, and samples taken at five different time points. For reference, two Mark 12™ (Thermo Fisher) protein standards were loaded for each gel and annotated with their molecular weight (kDa). In the gel with samples expressed at 15 °C, this protein standard is also present in the well marked for insoluble fraction after 18 hours.

As there is no significant difference in expression level between the three bands between 14.4-21.5 kDa, it is difficult to assess which one is in fact MarEctC. We therefore decided to purify a batch of bacterial culture with MarEctC, to hopefully separate these bands by their binding affinity to a HisTrap column. While it is most likely that one of the two bands closer to 14.4 kDa is MarEctC, abnormal gel-shifting of other cytosolic proteins is well known [128].

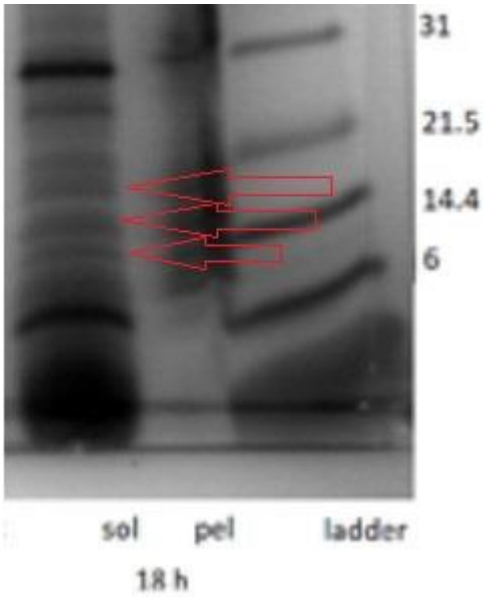


Figure 3-3: Three bands in close proximity close to the theoretical molecular weight of MarEctC, taken from expression at 18 h, 20 °C.

3.2.2 Solubility optimization of MarEctC

Since there seemed to be at least the same amount of protein in the insoluble fractions as the soluble fractions after MarEctC expression, we attempted to optimize the solubility of MarEctC during expression. We tested the effect of IPTG concentration and the presence of an osmoprotectant at two concentrations, to evaluate any potential effect on solubility or overall expression rate of MarEctC. After expression, lysis by sonication and separation by centrifugation, we ran samples from the soluble and insoluble fractions on separate gels (Figure 3-4). The soluble fractions show no significant difference in expression level of MarEctC. There is no apparent trend between the tested parameters in overall protein expression or specific MarEctC expression, or in the amount of insoluble protein compared with soluble protein in the samples. As no improvement of MarEctC solubility or expression was detected, we decided to continue with the standard TB media and IPTG concentration in downstream purification, and increase batch sizes to obtain sufficient amount of protein for crystallization and biochemical characterization.

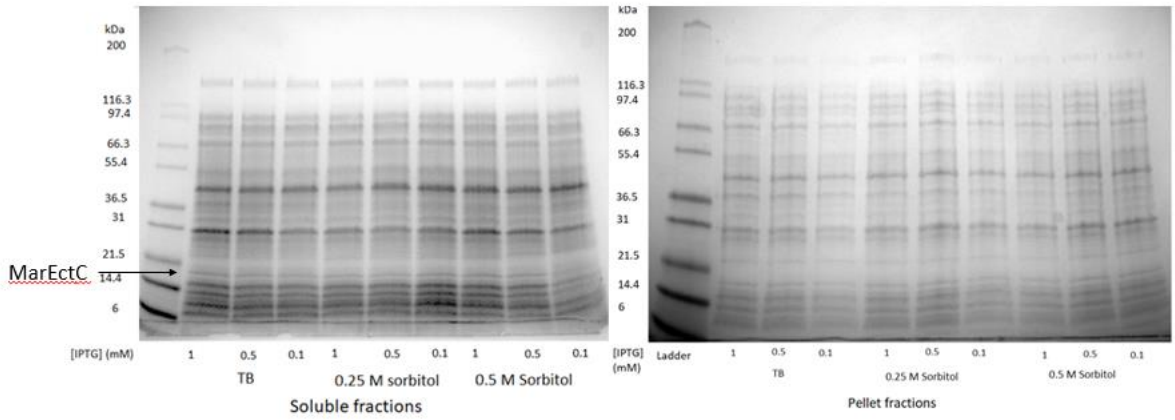


Figure 3-4: Influence of sorbitol at two concentrations and three concentrations of inducer (IPTG) on the expression and solubility of MarEctC when expressed at 15 °C. The probable position of MarEctC in the gel is marked on the soluble fraction gel (left).

3.2.3 Finding the migration distance of MarEctC on an acrylamide gel and mass spectrometry analysis

Early evaluations of MarEctC expression by SDS-PAGE revealed several bands in the area of the expected MarEctC migration distance. In addition, the addition of lysozyme in the lysis buffer, which is a normal procedure in protein production protocols [insert 2-5 references] added even more contamination, as lysozyme has a molecular weight of 14.4 kDa, close to the 15.4 kDa molecular weight of MarEctC. Thus, it was decided to omit lysozyme from the lysis buffer for both enzyme candidates from this point out.

Since the bacterial cells overexpressed the AmpR resistance gene and showed ampicillin resistance, we interpreted that the expression vector was successfully transformed into the cells, and that this was not the cause for lack of overexpression of MarEctC. We therefore decided to combine the culture from all three of the initial expression optimization experiments, grown at three different temperatures. We then attempted to remove contaminants by IMAC chromatography and gel filtration, with the idea to better identify MarEctC migration distance through an acrylamide gel. Two protein bands were cut out from an SDS-PAGE analysis after gel filtration, and sent to mass spectrometry to hopefully identify the protein (Figure 3-5). The bands migrated with a theoretical

molecular weight of 21.5 kDa according to the protein standard, while MarEctC has a theoretical molecular weight of 15.4 kDa (HisTag included). Sample #1 showed an approximately 70 % sequence coverage (Figure 3-6), meaning about 70 % of the amino acids in the MarEctC sequence were identified in the sample. There are no other proteins from the *E. coli* expression system used with any significant sequence similarity to MarEctC as evaluated with a BLASTp search [129, 130]. This identifies the sample as MarEctC and show that the protein migrates slower than its molecular weight suggests.

Figure 3-5: Protein bands cut out from a gel after preliminary IMAC and gel filtration purification of MarEctC.

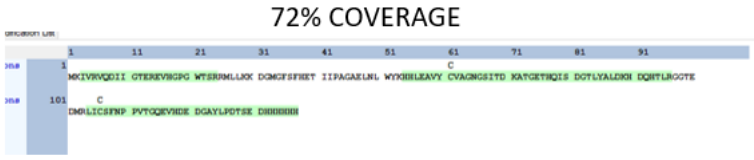


Figure 3-6: Sequence coverage in sample #1 sent for mass spectrometry. The amino acids highlighted in green shows the residues covered during sequencing.

3.3 Protein purification

3.3.1 Purification of CSEctC

CsEctC was purified in batches from 2 L bacterial culture by IMAC and gel filtration. The chromatogram and SDS-PAGE gel (Figure 3-7) from HisTrap purification (IMAC) shows the lysate protein behaviour in the column. We loaded the lysate (150-200 mL, with some inter-batch

variations) at 1 mL/min, to secure proper binding to the column. While there seems to be some protein at the molecular weight of CSEctC in the flow through fraction (FT), there is sufficient amount of protein binding to the column. To hopefully lose some of the more weakly binding contaminants in the column, we introduced a wash step prior to elution, with 75 mM imidazole. The wash sample in the gel (W) shows that this wash step cleared the column of various contaminants. There also looks to be a small amount of protein at the molecular weight of CSEctC, which affects overall CSEctC retention. However, the overall amount of CSEctC in the lysate makes this trade-off between a small loss of target protein versus loss of contaminants in fractions with substantial amounts of CSEctC beneficial.

After the wash step, we eluted the remaining column protein with a 75 mM-600 mM imidazole gradient. The elution peak appears around 45-55 %, which corresponds to around 236-288 mM imidazole concentration. The peak fractions (C6-C12) contain the majority of CSEctC protein. These fractions however also contain larger amounts of contaminants than fractions found in the “tail” of the peak (D2-D6). In addition, late fractions after the peak also seem to contain smaller amounts of CSEctC, while also being apparently free of most contaminants. To maximize the yield of CSEctC, we decided to include both peak fractions and late fractions, including the peak fractions with higher amounts of contaminants. Since the main contaminant looks to be the AmpR product β-lactamase, which is known to be a tetramer in solution [131], we anticipated that we would separate CSEctC, likely a dimer in solution from most contaminants during gel filtration.

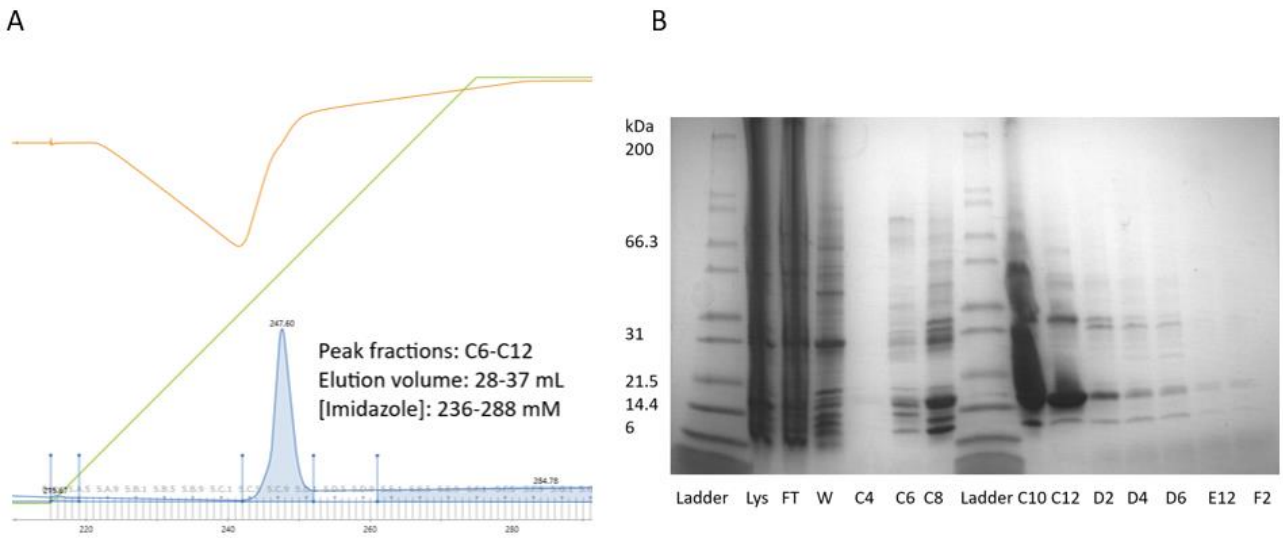


Figure 3-7: Chromatogram and SDS-PAGE analysis of CSEctC HisTrap purification. A: Chromatogram showing the elution process from the HisTrap column. The main peak is eluted approximately 50 % through the elution gradient and corresponds to the fractions C4-C12. The orange line shows the measured conductivity during elution. The green line shows the percentage of buffer B in the mobile phase. The peak protein concentration was measured to 247.40 mAu at 280 nm. B: SDS-PAGE analysis of relevant fractions from lysis and IMAC. Lys: Lysate, FT: flow-through during loading. W: Elute during the washing step. C4-D2 corresponds to the fractions according to the chromatogram.

We experienced some issues with column pressure during this step of purification. As seen in the lysate (Lys) and flow through (FT) samples, there are visible smearing of the protein bands present. This may stem from cell debris not fully separated during centrifugation, or traces of nucleic acids in the lysate. There were no nucleases like DNase or similar included in the lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 15 mM imidazole), and such contaminants could contribute to lysate viscosity and thereby be a factor in building up pressure in the column.

As a second purification step, we introduced gel filtration, which separates molecules based on their molecular weight. This step also served as a buffer-exchange step, as we did not perform dialysis prior to gel filtration, but rather started gel filtration as soon as possible after IMAC purification. After concentration of IMAC fractions, we aimed for an approximately 10 mL volume of our sample. This was passed through the Superdex 26/600 75 pg column, and produced a chromatogram and SDS-PAGE samples as seen in Figure 3-8. The majority of protein eluted in a single peak at

around 190-210 mL elution volume, corresponding to fractions 4.G.12 - 5.B.2. Both early, peak and tail fractions from this peak looks to contain CSEctC with very minor amounts of contaminants. From visual inspection of SDS-PAGE samples, we anticipate that the sample in total is over 95 % pure, which makes it suitable for crystallization purposes.

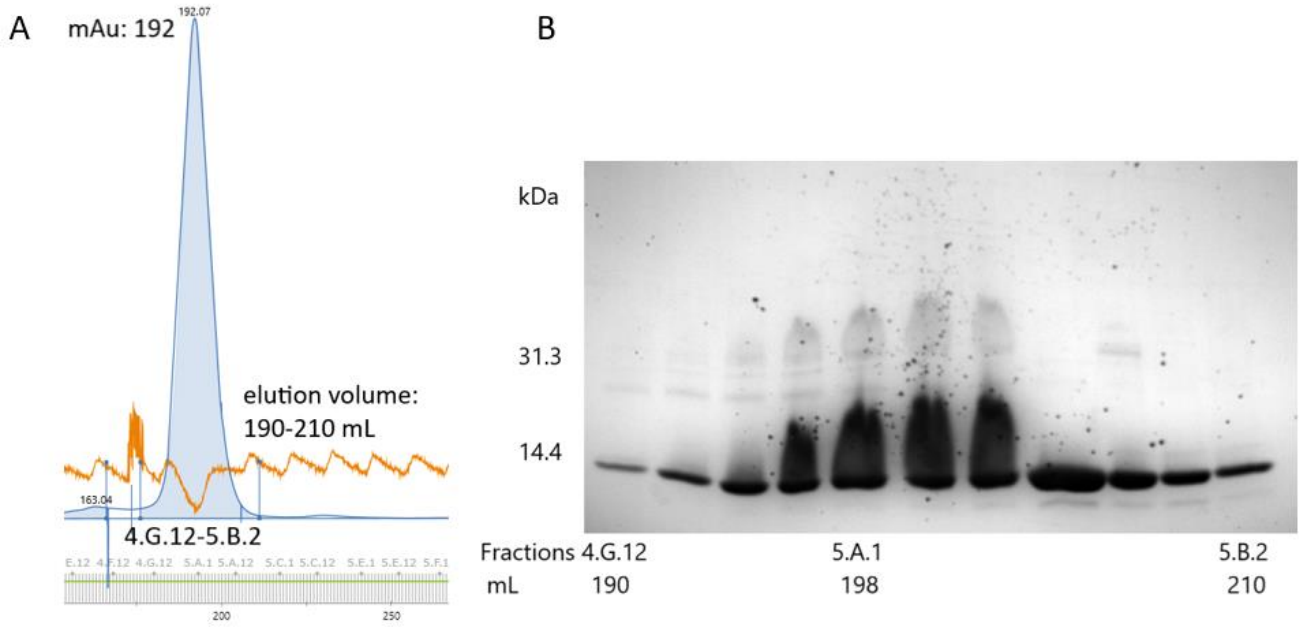


Figure 3-8: Chromatogram snippet and SDS-PAGE analysis of relevant fractions from gel filtration of CSEctC. A: A snippet of the main peak from the chromatogram from gel filtration. The blue line is protein concentration (absorption at 280 nm), annotated with the analysed fractions and elution volume at the peak. The orange line is conductivity measured in the elute. B: SDS-PAGE analysis of the early, peak and tail fractions of the eluted peak according to the chromatogram. Approximate molecular weights are annotated, as no protein marker was included in this gel.

In total, we produced around 13-18 mg CSEctC from 1 Litre bacterial culture. As we mainly purified from 2 L culture batches, we produced 26-36 mg CSEctC per production cycle. The final product was stored at 4 °C in the storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP) with 4-6 mg/mL concentration.

3.3.2 Purification of MarEctC

Since MarEctC expression produced less protein per litre bacterial culture during expression, we purified MarEctC in 6 L batches, with a goal to obtain at least 5 mg MarEctC protein per production cycle.

As with CSEctC we loaded the lysate at 1 ml/min to ensure proper binding to the column. The lysis and flow-through fractions of the SDS-PAGE analysis (Figure 3-9) shows proper binding to the column, but also highlights the relative amount of contaminants in the sample. After the wash step, we eluted protein over a 20 CV 50-500 mM imidazole gradient.The protein mainly eluted at two peaks, corresponding to about 110-130 mM imidazole and 175-190 mM imidazole, respectively. According to SDS-Page analysis of the peak fractions, the first peak mainly contained contaminants, including large amounts of a protein with a molecular weight of approximately 60 kDa.There looks to be small amounts of protein at the migration distance of MarEctC, but the amounts of contaminants makes the undesirable for further purification. The second peak contained MarEctC along with contaminants, the most noticable at 60 kDa, 30 kDa and 6 kDa. The tail fractions of the second peak were relatively purer compared to the peak fractions, but all fractions from the peak (B5-C2 according to Figure 3-9) were combined for further purification.

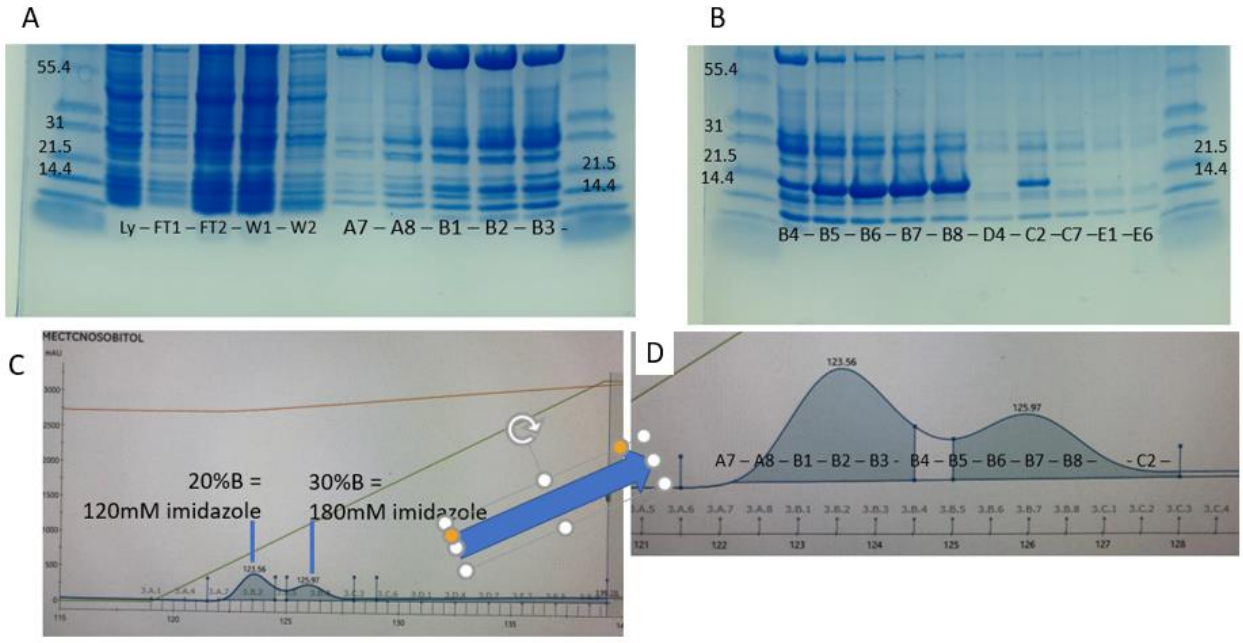


Figure 3-9: Representative SDS-PAGE analysis and chromatogram of elution from HisTrap purification of MarEctC. A-B: SDS-PAGE analysis of Lysate and early fractions of elution. Ly: Lysate, FT1-Ft2: flowthrough during loading. W1-W2: early and late flowthrough during the wash step, A7-B8: Fractions collected according to the annotated chromatogram (figure D). D4-E6: Late fractions during elution, not associated with any peaks in the chromatogram. C-D: Chromatogram of the elution of protein. Two peaks are visible in the chromatogram, and MarEctC looks to mainly be eluted at the second peak.

The MarEctC-containing fractions from IMAC were further purified by SEC. Protein eluted in three distinct peaks, one of which were significantly bigger than the other two (Figure 3-10: Chromatogram and SDS-PAGE analysis of SEC purification of MarEctC A: Chromatogram of the three main elution peaks. The second peak with fractions A2-A4 contained our target protein MarEctC. B: SDS-PAGE analysis of the fractions as annotated in the chromatogram. His: the sample which was loaded after Histrap purification. M12: the protein standard.). The first peak eluted at around 90 mL, the second peak at around 180-200 mL while the last peak eluted at 250 mL. The peaks were analysed by SDS-PAGE, revealing that peak 2 contained MarEctC along with smaller amounts of contaminants. The tail fraction of the peak looked to be over 95 % pure as estimated by visual inspection, while the peak fraction was estimated to be at least 70 % pure. Compared with the fractions loaded from after HisTrap purification (His), the SEC purification improved purity

significantly.

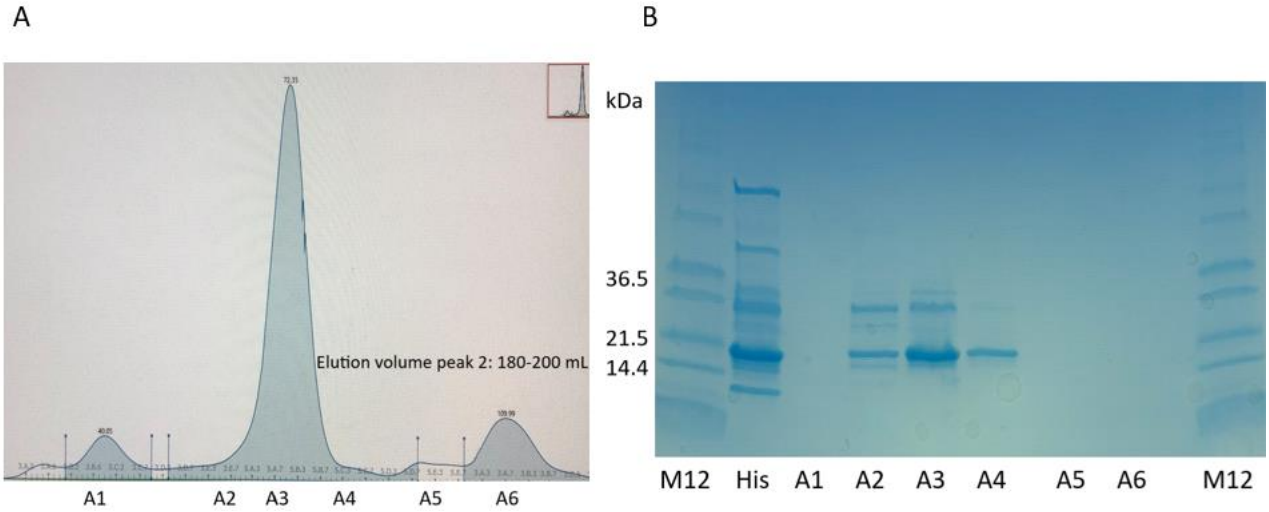


Figure 3-10: Chromatogram and SDS-PAGE analysis of SEC purification of MarEctC A: Chromatogram of the three main elution peaks. The second peak with fractions A2-A4 contained our target protein MarEctC. B: SDS-PAGE analysis of the fractions as annotated in the chromatogram. His: the sample which was loaded after Histrap purification. M12: the protein standard.

Fractions from the entire peak were combined and the protein concentration were measured at 280 nm. A typical purification yielded about 4-8 mg/mL, corresponding to 0.75-1.5 mg/L bacterial culture when purifying from 6 L culture.

3.4 Crystallization, data processing and structural investigations

3.4.1 Crystallization of CSEctC

Initially we screened for promising crystallization conditions for CSEctC with three commercial screens, using two different protein concentrations (6 and 11.6 mg/mL). From the initial screens, we identified several conditions which produced three-dimensional crystals with diameters of >0.05 mm. The three most promising crystallization conditions came from the Pact Premier Eco Screen (Molecular Dimensions). Interestingly, the pH of all three buffer systems (SPG, MMT and MIB) were identical, and the most acidic pH in the screening grid (pH 4.0). The polymer size and concentration were also identical (25 % w/v PEG-1500). Crystallization almost exclusively

happened at the highest tested protein concentration (11.6 mg/mL), and this was used later for optimization and co-crystallization.

Optimization of CSEctC crystals were conducted in 24-well plates with 17-30 % w/v PEG 1500, increasing the concentration across the rows, giving for replicates for each condition. The crystal size and apparent quality judged by appearance increased with higher polymer concentration, and crystals grown with 0.1 M MMT or MIB buffer (pH 4.0), 30 % w/v PEG 1500 and 11.6 mg/mL were preserved in a cryo-solution (reservoir solution with 20 % glycerol) and prepared for shipment to a synchrotron facility.

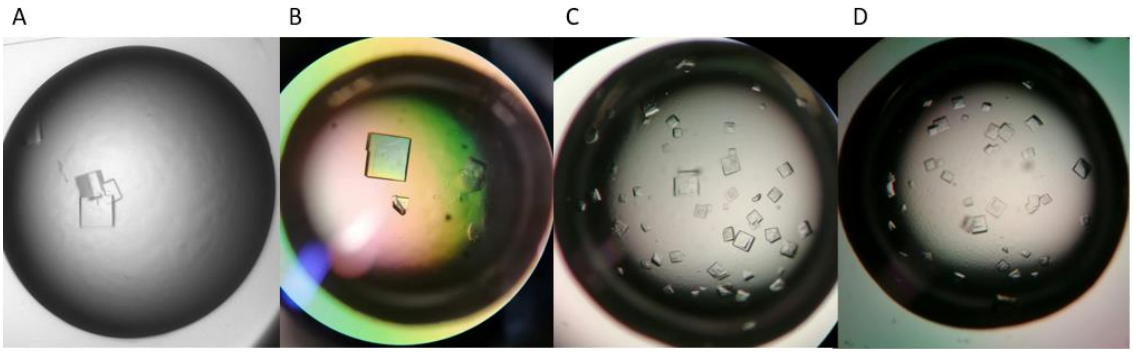


Figure 3-11: CSEctC Protein crystals produced by hanging drop vapor diffusion. A: CSEctC crystals grown in 0.1 M SPG buffer (pH 4.0), 27 % w/v PEG 1500. B: CSEctC/N-γ-ADABA crystal grown in 0.1 M MIB buffer (pH 4.0), 30 % w/v PEG 1500, 1:30 molar ratio CSEctC/N-γ-ADABA. C: crystals grown in 0.1 MIB buffer (pH 4.0), 30 % w/v PEG 1500, 1:30 molar ratio CSEctC/ectoine D: CSEctC crystals grown in 0.1 M MTT buffer (pH 4.0), 30 % PEG 1500. The protein concentration in all drops were 10-13 mg/mL.

As we produced crystals of satisfactory size and shape, we attempted to co-crystallize CSEctC with both the native substrate *N*-γ-ADABA and the product ectoine. CSEctC was pre-incubated with the substrate and product in 1:10 and 1:30 ratios with otherwise identical conditions as previously. Both ligand concentrations yielded crystals of similar size and shape as crystals grown in apo-form. 4-5 crystals of each type were preserved sent to ESRF in Grenoble. However, during data collection and processing, it was discovered that none of the co-crystallized crystals included a substrate or product molecule bound in the structure.

3.4.2 Crystallization of MarEctC

The initial commercial crystallization screens with MarEctC produced several promising results, although no signs of crystallization were identified. There were however many conditions that produced supersaturated droplets, granular precipitation and spherulites, which all has been known to be starting points for growing high-quality diffracting crystals [99, 132]. From a total 10 optimization experiments, the most promising result obtained was one-dimensional needle-shaped crystals (Figure 3-12). These were grown in 0.1 M Hepes pH 7.5, 25 % w/v PEG 3350 and 0.2 M $LiSO_4$. Thin, nearly one-dimensional crystals are generally known to be unsuitable for data collection, as they produce low-quality data sets [132]. These crystals were thus not prioritized for shipment to the synchrotron facility. However, they may serve as a promising starting point for further optimization, and alternative techniques like seeding could be introduced to improve the shape of the crystals.

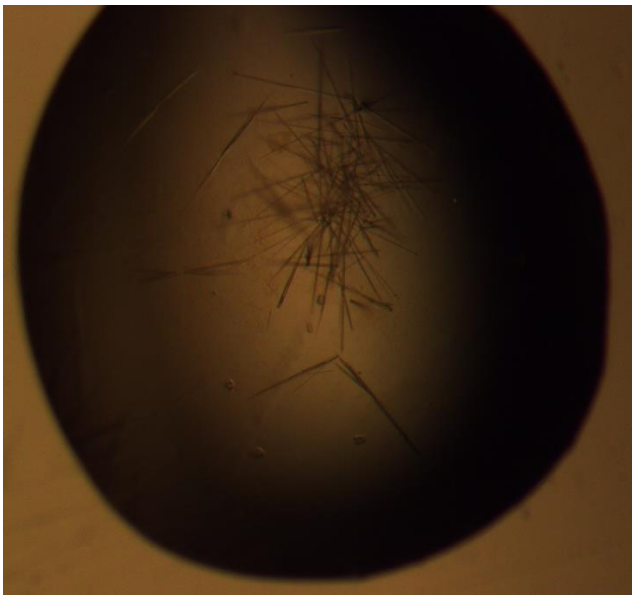


Figure 3-12: MarEctC crystal needles grown during optimization.

3.4.3 Dataprocessing, refinement and structure solving of a CSEctC crystal

Our crystals were sent to ESRF in Grenoble for data collection. Several CSEctC crystals diffracted

well and produced data sets of high quality. A data set from a CSEctC protein crystal grown in 0.1 M MIB buffer (pH 4.0), 30 % w/v PEG 1500 was used in downstream processing and refinement (Figure 3-13). Statistics gathered from data collection, processing and refinement is summarized in

Table 3-2.

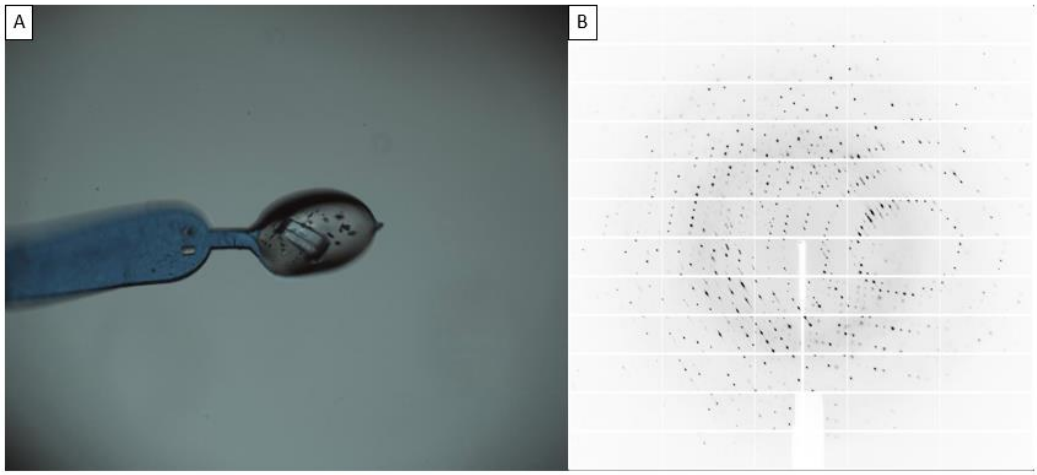


Figure 3-13: Representative images from data collection. A: A single crystal in cryo-solution, held in the cryo-solution (reservoir solution with 20 % w/v glycerol) in a sample pin. B: A diffraction pattern gathered from a CSEctC crystal. The resolution range of the collected data was 45.18-1.38.

Overall, we obtained a high-quality data set from our CSEctC crystal. The data processing revealed that our data set represented either the I 222 or the I $2_12_12_1$ space groups. By trial and error, we determined that the correct space group must be I 222 as this was the only one that yielded a usable model after molecular replacement. Data reduction, the removal of high-resolution data, also reduced the resolution from the initial 1.12 Å to 1.38 Å. This is done to reduce potential noise during model building, which may lead to incorrect placement of atoms in the structure.

The final model after refinement had a resolution of 1.38 Å and R_{work} /R_{free} values of 13.38/16.82 and were the main determinants of model quality. The R_{work} represent how well the x-ray data collected agrees with the structure model after building. R_{free} are data that are removed prior to refinement and is mainly used as a comparison to the R_{work}. In essence, R_{free} should ideally be as close to the R_{work} as

possible as they are based on the same data. Since R_{free} is excluded from the refinement, the R_{free} only decreases when refinement is successful, and is thus a good marker for a well-refined model.

Table 3-2: Data collection and refinement statistics from CSEctC data set. Values in parentheses represent the highest resolution shell.

Data collection	Values	Refinement	Values
Beamline	ID30B	**Resolution (Å)**	1.38
Detector	PILATUS3 6M	**R_{free} reflections**	1224
Wavelength (Å)	0.976	**R-factor**	13.38
Temperature (K)	100	**R_{free}**	16.82
Space group	I 2 2 2	**Number of non-H atoms**	1075
Unit Cell		**RMSD from ideal values**	
***a, b, c* (Å)**	59.243, 60.921, 67.340	**Bond lengths (Å)**	0.03
Molecules in the asymmetric unit	1	**Bond angles (°)**	2.36
Oscillation range (°)	0.1	**Ramachandran statistics (%)**	
Resolution range (Å)	45.18-1.38 (1.41-1.38)	**Favoured**	93.68
Unique reflections	25429	**Allowed**	98.95
Completeness	100 (99.9)	**Outliers**	1.05
$CC_{1/2}$	0.999 (0.886)	**Average B-factor**	24.776
Multiplicity	9.7 (9.9)		
Wilson B-factor ($Å^2$)	16.62		

R_{pim} (%)	0.034 (0.413)
<Mean I/σ(I)>	13.9 (2,7)

3.4.4 Bioinformatical analysis of EctC-type proteins

To identify interesting and conserved residues within the EctC structure, a MSA were generated with the ConSurf server, using the Uniprot database [121, 122] and the amino acid sequence from *S. alaskensis* EctC (SAEctC) as the reference sequence. From 159 sequences of EctC-type proteins we found 33 strictly conserved residues (Figure 3-14). Earlier bioinformatical analyses of 437 EctC-type proteins based on *P. lautus* EctC (PlEctC) of EctC-type proteins found 26 strictly conserved residues [76]. This was a surprisingly high amount, as EctC is a small enzyme and the average similarity score (including gaps) in our MSA was around 50 % as calculated by the UGENE software [123]. The conserved residues includes residues thought to be involved in coordinating an active site metal (SAEctC E57, Y84 and H92, corresponding to E59, Y87 and H95 for the CSEctC sequence), and several residues potentially involved in substrate binding (SAEctC W21, R25, T40, Y52 corresponding to CSEctC W23, R27, T42, Y54). There are also several strictly conserved glycine residues found between major secondary structure elements (e.g. CSEctC G37, G48, G85). These glycine residues may facilitate the folding of the β-sandwich domain as they are without side chains and allows for flexible and sharp turns. The typical β-barrel motifs are however not well conserved in the EctC sequences, in spite of the conserved topology and fold.

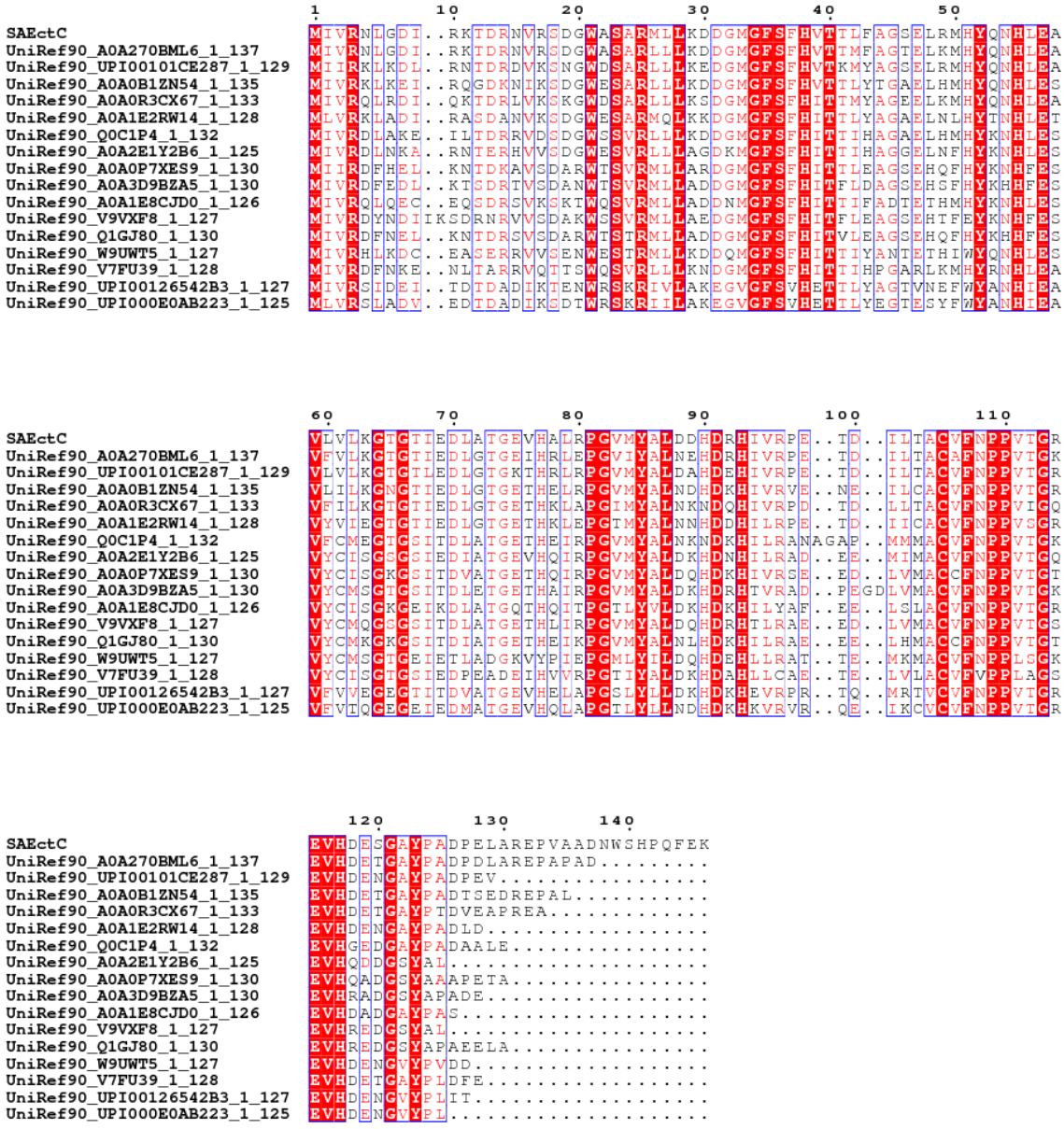


Figure 3-14: A representative selection of 16 random sequences in a MSA with SAEctC from a total of 159 sequences. Residues with red background are completely preserved. Red letters on white background shows residues with similar properties. The figure was prepared using UGENE and ESPript 3.0 [123, 124].

In addition, a smaller alignment of our two target EctCs along with the previously structurally

characterized EctC proteins from *Sphingopyxis alaskensis* and *Paenibacillus lautus* (SAEctC and PLEctC respectively) was generated (Figure 3-15). Here we also included the secondary structure elements using our CSEctC structure as a template, to describe the EctC topology and to use as reference when describing the EctC structure. Between the four sequences, there are 40 strictly conserved residues. The sequence similarity between them ranges from 42 (SAEctC), 48 (PLEctC) and 50 % (MarEctC). The sequence similarity as well as overall length of the sequences made the CSEctC structure the choice of template during homology modelling of MarEctC.

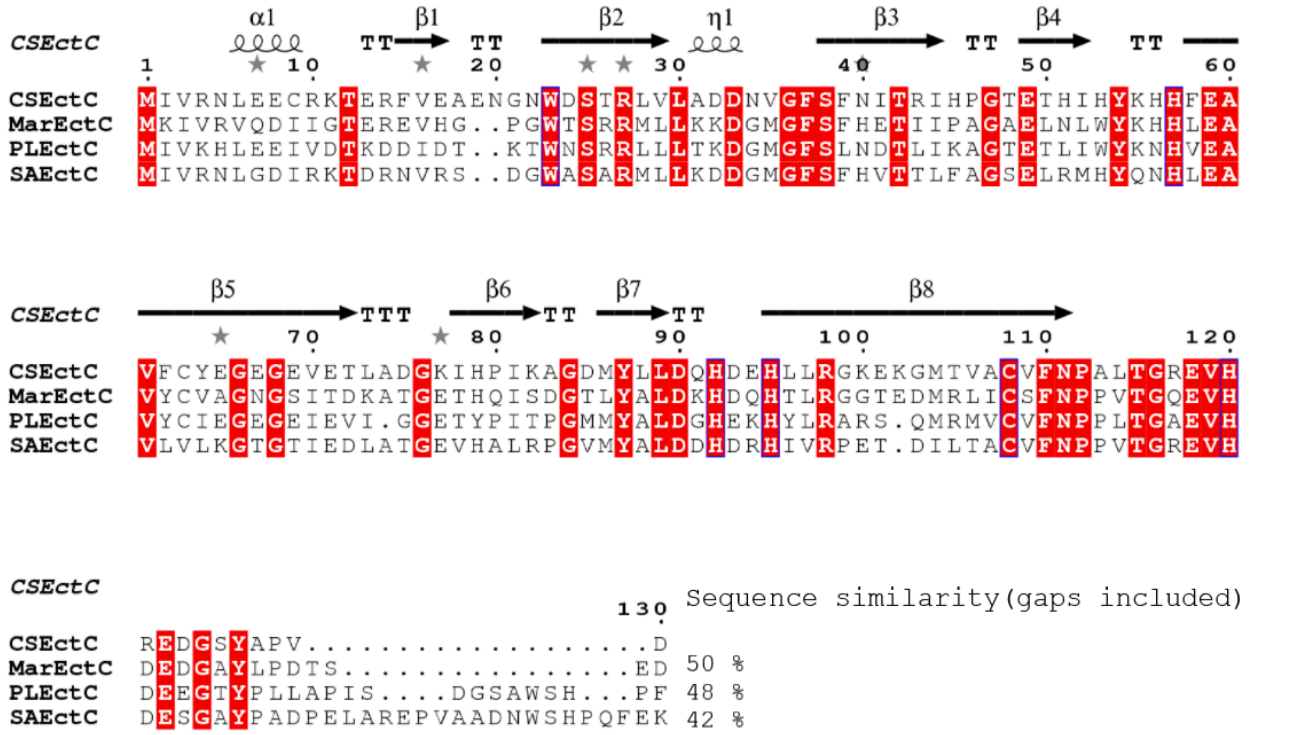


Figure 3-15: Sequence alignment of our two target structures along with the other structurally characterized EctCs. The secondary structure elements are shown for our CSEctC structure and shows the general topology of EctC. The secondary structure elements are numbered chronologically. α: alpha-helices, β: beta-strands, T: turns, η: 3-10 helices. Residues which are highlighted red are strictly conserved residues. The sequence similarity is compared to CSEctC as calculated by the UGENE software. The figure was prepared with ESPript 3.0 [124].

3.4.5 Structural investigations of CSEctC and the MarEctC homology model

After refinement, the final CSEctC structure contained117 of the 130 amino acids of the protein. The final 13 residues (excluding the six histidine residues from the hexahistidine tag) of the C-terminus are not included, likely because of high flexibility of the and therefore no clear electron density for these residues. The overall structure reveals that CSEctC is indeed an enzyme in the cupin superfamily. CsEctC adapts a classical jelly-roll shape, with two antiparallel β-sheets both consisting of 4 β-strands each and two peripheral α-helices. The first EctC crystal structure (Widderich et al. (2016) described two states of the protein which they termed the open and the semi-closed states [75]. The semi-closed state describes the structure in which the entrance to the cupin-barrel is covered by the C-terminus. In the open state, the C-terminus flexes away from the barrel entrance. Our CSEctC structure most closely resembles the open state, as the visible part of the C-terminus looks to extend up and away from one of the entrances to the barrel shape. Since the C-terminus of EctC proteins seldom gives any electron density of good resolution, it is suggested that it is very flexible [75]. Its amino acid sequence does however contain many well conserved residues (Figure 3-14), and could therefore still have an important function in enzyme activity.

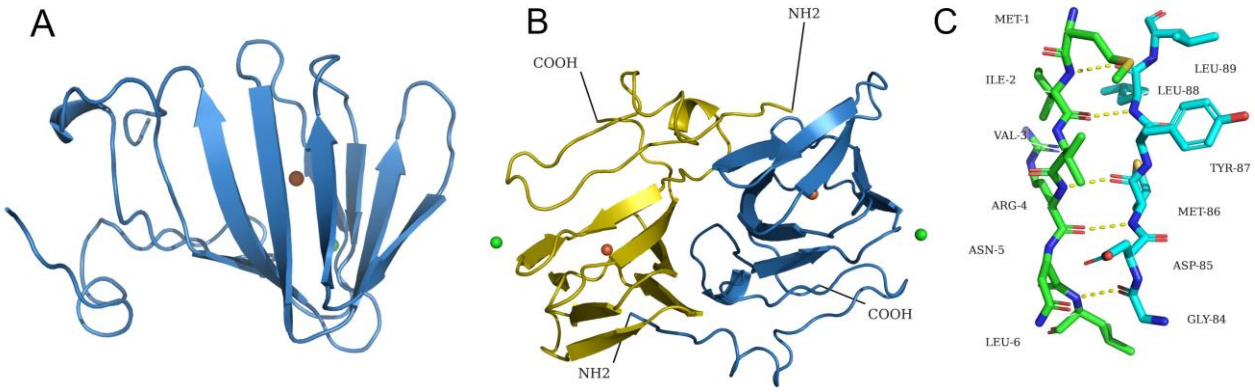


Figure 3-16: Structural overview of CSEctC. Orange spheres represent Fe^{2+}, green spheres represent Cl^-. A: Cartoon representation of the monomer. B: Head-to-tail dimer with oligomer interface at the N-terminal ends. C: Stick representation of one of the dimer interfaces between the N-terminus (green) and β1 (cyan). The yellow lines represent hydrogen bonding between main-chain nitrogen and main-chain oxygen atomes.

The N-terminus of EctC proteins is generally well conserved. The β7 strand which the N-terminus interacts with is however not as well conserved, as Figure 3-17 A illustrates. However, while the residues are not conserved, the hydrophobic properties of dimer interface residues looks to be retained. This is also observed in the PlEctC structure and the MarEctC homology model, with similar patterns in regard to hydrophobicity at the dimer interface (Figure 3-18).

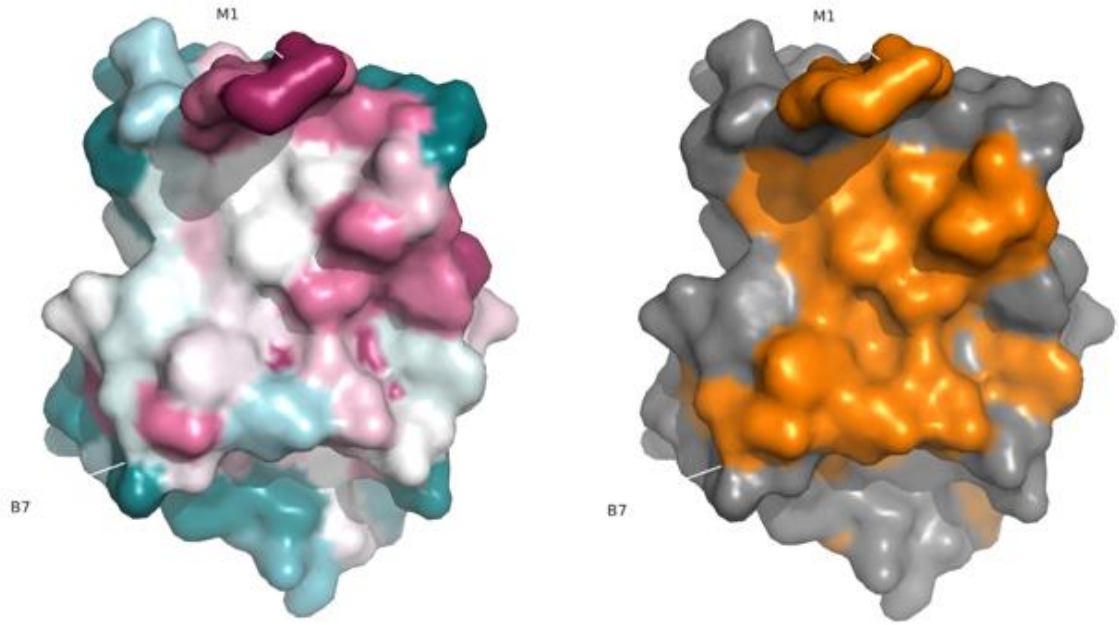


Figure 3-17: The dimer interface side of CSEctC using a surface representation, with conserved residues (A) and hydrophobic residues (B) shown in colours. A: This structure representation of CSEctC is colour coded after how conserved its surface residues are. Green-blue colour represent less conserved residues while pink and red colours represent well conserved residues. B: CSEctC in a surface representation, where hydrophobic residues are coloured orange. M1: N-terminus represented by the initial methionine residue. B7: β-strand 7 according to Figure 3-15.

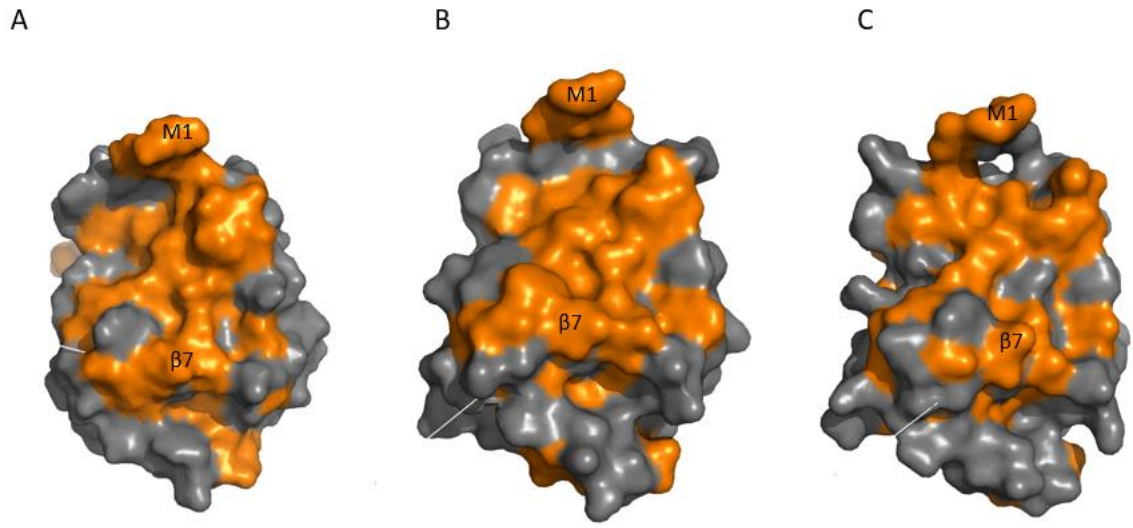


Figure 3-18: Surface representation of three EctC proteins with hydrophobic residues highlighted in orange. A: PLEctC B: CSEctC, C: MarEctC. The N-terminii are represented by the 1-methionine. β7: The β-strand interacting with the N-terminus in the dimer interface.

The modelled Fe^{2+} ion is located in the jelly roll fold's lumen, where it likely contributes to the catalytic activity of the enzyme. Figure 3-19 shows how the iron ion likely is coordinated, by E59, Y87 and H95, with distance between the iron and residue side chains of 2.8 Å, 3.1 Å and 2.9 Å, respectively. This modelled iron-coordination is near-identical through all known structures of EctC-type proteins, and suggest that it most likely serves a critical role during activity. However, refinement of EctC structures has yet to confirm an iron molecule at the position where it is fitted in the structures.

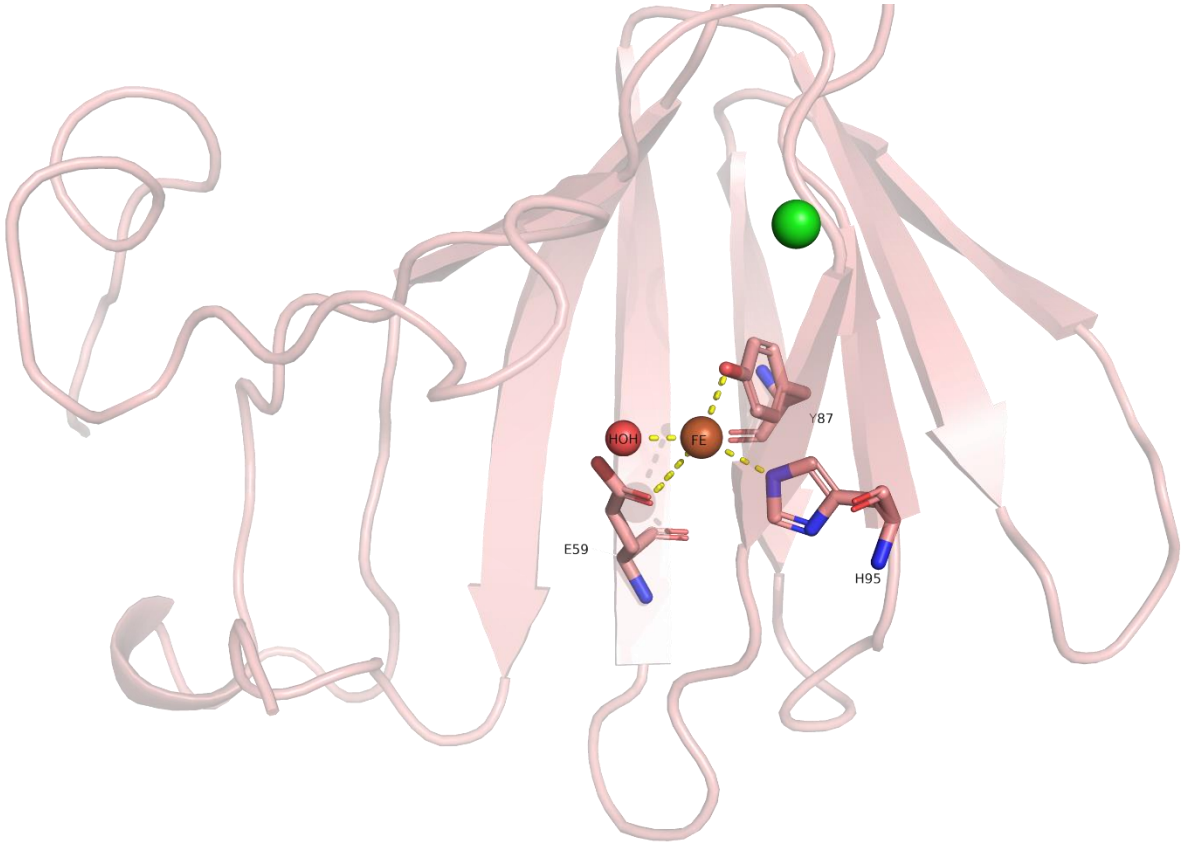


Figure 3-19:Coordination of the active site Fe^{2+} ion in CSEctC. The red sphere represents a water molecule, which is thought to be a placeholder for the substrate which interact directly with the iron ion during catalysis. The green sphere represent a chloride ion. Residues E59, Y87 and H95 are represented with sticks, and their position in the overall structure represented as a cartoon is shown.

3.5 Biochemical characterization of stability and activity

3.5.1 Dynamic light scattering

Prior to crystallization optimization, we performed a DLS experiment with a purified CSEctC batch, which at that point had been stored at 4 °C for two weeks. Since any potential protein aggregation could negatively impact the protein ability to develop high-quality crystals, we wanted to assess the dispersity of our sample after storage. The sample was tested in three separate measurements, which each contained ten measurements each. The particle size distribution of calculated and the mean

particle size is presented for all measurements (Figure 3-20). The results show that after two weeks, the sample was still highly monodisperse, with only one significant peak visible at 2-3 nm. This reinforces our assumption that our purified CSEctC was at least 95 % pure, and that the protein is stable when kept at 4 °C in our storage buffer (50 mM HEPES, 500 mM NaCl, 0.4 mM TCEP). A single peak would also signify that the protein is present in a single oligomeric state, which is in line with earlier studies of EctC by Widderich et al. (2016) [75]. It is highly likely that CSEctC is present as a dimer in solution, and while a mean radius of 2-3 nm aligns with a theoretical molecular weight of a CSEctC dimer at 31 kDa [106], DLS data is not considered an accurate method for molecular wight determination [106].

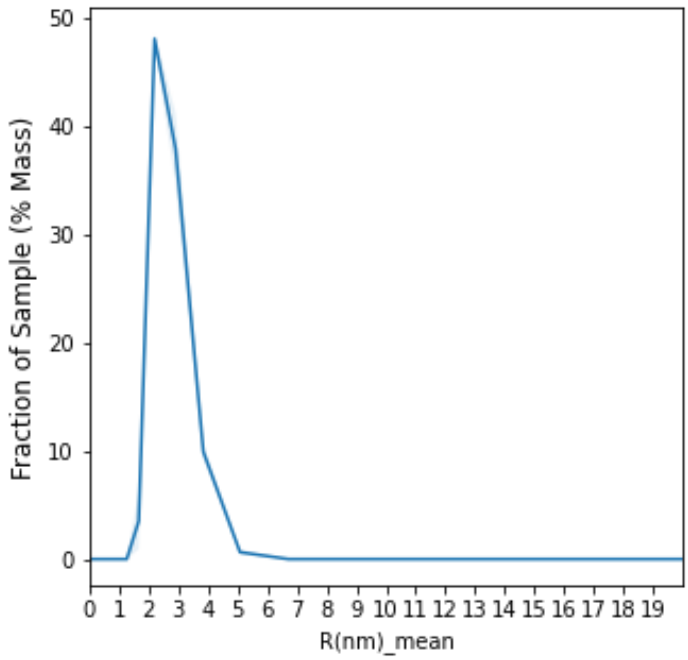


Figure 3-20: The mean size (radii) of particles in a purified CSEctC sample. A single peak shows that the sample is highly homogeneous.

3.5.2 Differential scanning calorimetry

As our thermal shift assay produced unreliable results, we used DSC to obtain preliminary values for melting temperature (T_m) and enthalpy (ΔH) for both EctC candidates (Table 3-3). T_m was found at the graph peak, which represents the transitional midpoint, where 50 % of the protein is unfolded (Figure 3-21). CSEctC and MarEctC was found to have melting temperatures in the storage buffer of 60.6 and 53.1 °C. This shows the typically thermostable nature of cupin superfamily proteins, and also an expected lower thermal stability of the enzyme from the psychrophile *Marinobacter sp. CK1*. While CSEctC has a higher melting temperature, MarEctC showed the higher experimental ΔH. The transition peak of CSEctC is significantly broader (8-10 °C) than the peak of MarEctC (5-7 °C), and the maximum heat capacity (Cp_{max}) which is found at the transition peak is also higher for CSEctC (32 kj/molK) than for MarEctC (26.5 kj/molK). The higher heat capacity of CSEctC is also expected, but the higher experimental ΔH of MarEctC is somewhat surprising. This could however be due to length of storage prior to the experiment, or difference in purity between the two enzymes, as we detected larger amounts of impurities in MarEctC than in CSEctC after gel filtration.

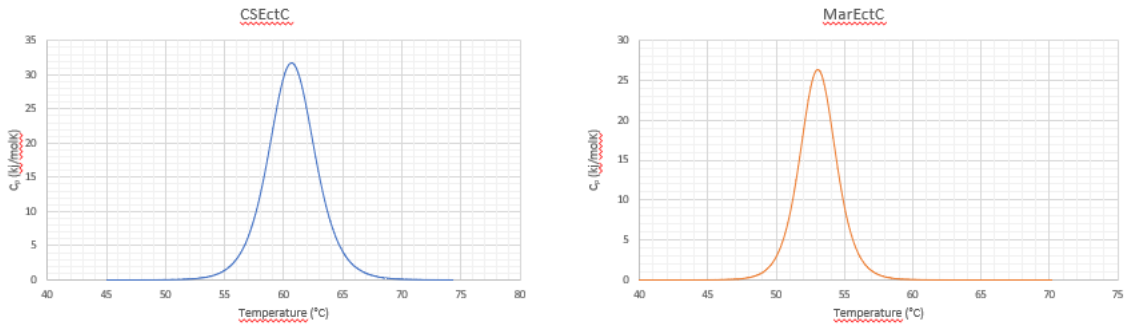


Figure 3-21:DSC Two-state scaled curves from experiments with CSEctC and MarEctC. The melting temperature was found at the peak transition point while ΔH was calculated by integrating the area underneath the curves.

Table 3-3: Experimental melting temperatures (T_m) and ΔH for CSEctC and MarEctC in our storage buffer (50 mM HEPES, 500 mM NacL, 0.4 mM TCEP). The melting temperature is found at the midpoint transition of the peaks, and ΔH is found by integrating the area underneath the thermogram peak.

Protein	Experimental T_m (°C)	Experimental ΔH (kj/mol)
CSEctC	60.6	723.03
MarEctC	53.1	1018.25

3.5.3 Preliminary isothermal titration calorimetry evaluation of CSEctC

While a number of optimization experiments with CSEctC and the ADABA substrate were conducted, we experienced difficulties establishing a standardized protocol for ITC to produce robust raw data. The most successful preliminary experiment is shown in (Figure 3-22: Experimental raw data and integrated curve data from one ITC experiment with EctC during optimization. A: Raw data. Initially during baseline establishment the sample heat decreased at each injection, meaning the titration of substrate initiated an endothermic reaction. At saturation, the sample heat increased at each injection meaning that each injection caused an exothermic reaction. This shift in enthalpy direction is unusual, and suggest that further optimization of experiment parameters is necessary to evaluate the substrate binding reaction thermodynamically properly. A possible explanation for this shift could be buffer or ligand-interactions with the iron ion, or displacement of the iron in the catalytic core of the enzyme. A possible way to evaluate if iron interactions have any impact on the quality of data is to titrate iron into the protein, without pre-incubating the enzyme with iron. The fit of the peaks using a one-site binding model does however yield a typical ITC S-shaped curve. We nevertheless evaluate the data as preliminary, and thus no thermodynamic parameters are presented. The current data do however represent a solid foundation for further optimization of the molar ratio between protein and ligand, as well as titration volume.

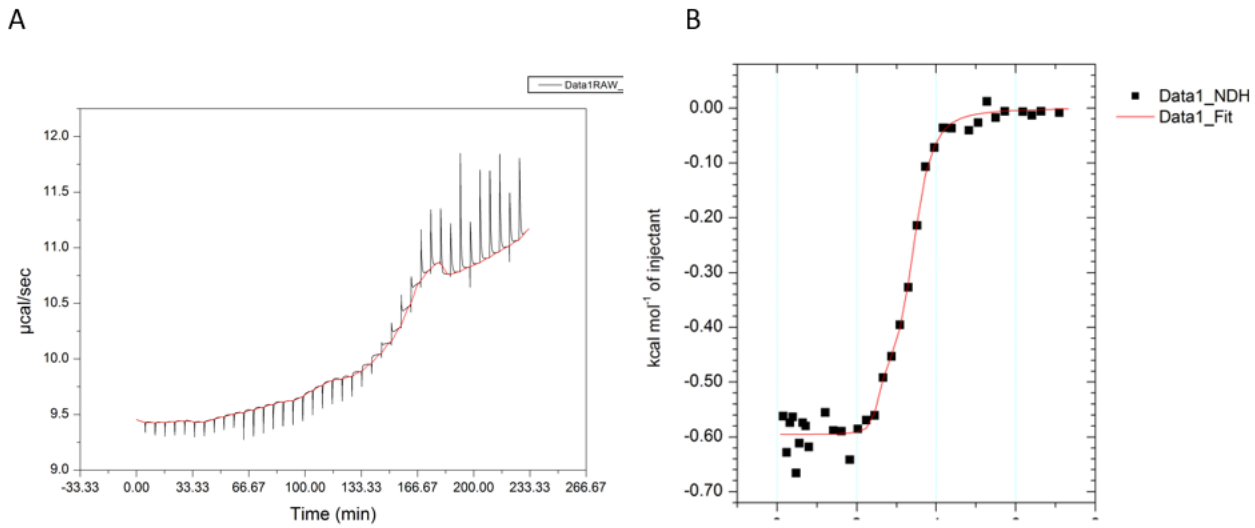


Figure 3-22: Experimental raw data and integrated curve data from one ITC experiment with EctC during optimization. A: Raw data collected, the spikes represent the change in power at each injection. B: Fitted curve of each measurement using the one-fit model. The red line represent the fit of least squares.

3.5.4 Activity

Earlier activity studies of EctC-type proteins from other organisms have revealed that EctCs usually displays a broad range of pHs where it is active, with optima ranging from pH 7-9 (

Table 1-3). Our preliminary activity assay showed that both CSEctC and MarEctC exhibit similar characteristics. The intra-assay variation between the triplicate reactions seems to be relatively low, and the trend is consistent with aforementioned earlier studies of EctC. Both enzymes show its highest activity at pH 9.0, similar to *H. elongata* and *A. cryptum.* Above 9.0, activity decreases with over 50 % for both enzymes, and the activity also declines with 40-50 % below pH 8.

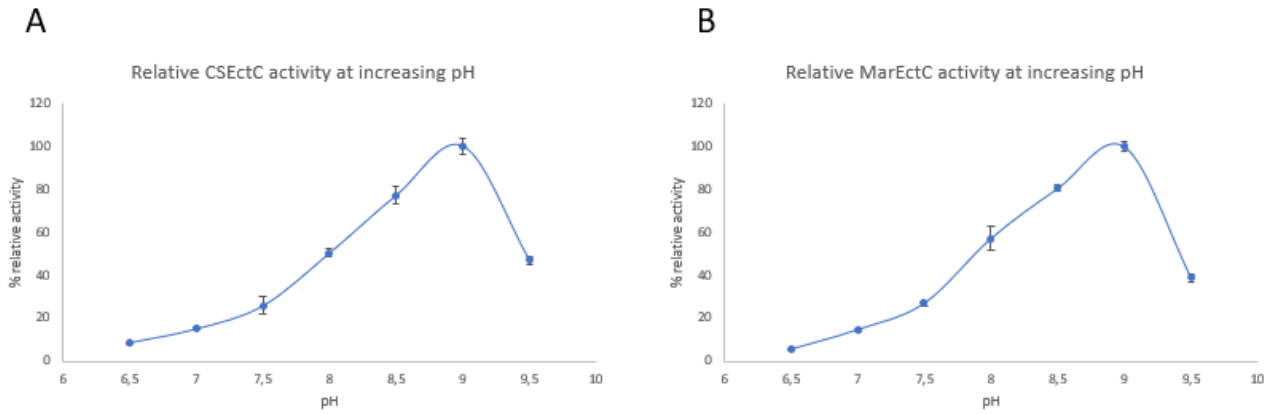


Figure 3-23: Effect of buffer pH on the activity of CSEctC (A) and MarEct (B). Activity is given as the retained relative activity compared with the highest measured activity. The error bars show the standard deviation between triplicate reactions.

4 Discussion

4.1 Expression optimization of MarEctC

While we were successful in expressing MarEctC for crystallization and biochemical characterization purposes, our yield of pure protein was less than ideal. We initially optimized the temperature and time of the bacterial culture post-induction but did not detect any significant differences between the test parameters. We did however detect higher amounts of insoluble protein at the theoretical molecular weight of MarEctC. *Marinobacter sp. CK1* is a psychrophilic bacterium, and MarEctC may require even lower temperatures during expression than our lowest tested temperature (15 °C). Lower temperature could in turn require longer expression periods, as *E. coli* expression systems are mesophilic and thereby not optimized for low-temperature expression [133]. Psychrophilic enzymes are generally less stable than mesophilic enzymes, as they have more flexible structures which aids them in function at low temperatures. This may lead to denaturing of synthesized protein or incorrect folding. One way to potentially solve this is to co-express MarEctC in *E. coli* expression systems with chaperonins that aid protein folding, which has been reported to improve psychrophilic protein expression [134]. Another potential solution is outlined by Bjerga et al. [135, 136]. They presented an alternative means of induction, by introducing a cold-shock promoter in a n *E. coli* expression-optimized vector. The *E. coli* cold-shock protein A (cspA) promoter is known to produce comparable levels of protein, but is induced by a downshift in temperature [137].

An alternative approach to improve protein solubility was also tried, as we optimized the inducer (IPTG) concentration and the presence of an osmoprotectant (sorbitol). Neither parameter showed any significant effect on protein solubility or overall expression level. Sorbitol as an osmoprotectant was mainly chosen due to availability, but several other chemical additives could potentially improve solubility. There was no apparent trend found relating to IPTG concentration at induction, and further optimizing efforts should focus on other parameters.

When looking at the MarEctC sequence (appendix B) and probable dimer interface, there is a high portion of hydrophobic residues present at the dimer interface and particularly the N-terminus (Figure 3-18 C). When the protein is expressed as a monomer, there is a risk that the dimer oligo does not form quick enough, and the hydrophobic N-terminus is exposed and tries to fold itself into the monomer to avoid surface exposure. While CSEctC has similar hydrophobic residues at its dimer interface areas, it has a more stable structure as seen during our DSC experiments which could explain why we did not have similar problems with the mesophilic CSEctC. A potential way to solve this is to introduce a fusing partner to co-express with MarEctC, which later can be cleaved off to yield pure MarEctC. Commonly used fusing partners to improve solubility includes the Maltose-Binding Protein (MBP), Small Ubiquitin-like Modifier SUMO) protein and glutathione-S transferase (GST) [95, 138-141]. These are highly soluble proteins that when co-expressed with a protein of interest helps the protein of interest to remain soluble in solution before the fusion partner is removed in a downstream purification step. One of more of these could be included in the expression vector to stabilize MarEctC during expression.

4.2 Crystallization and structure of CSEctC and MarEctC

Both CSEctC and MarEctC were attempted crystallized with commercial screens, with subsequent optimization in larger reservoir solution volumes for both candidates. With CSEctC, we were successful in producing high-quality crystals that diffracted well. Co-crystallization with substrate and product did however not yield any crystals with any significant electron density corresponding to either the substrate or product. An alternative to co-crystallization is soaking; a technique where crystals first are grown without ligand and the crystals are then soaked in a ligand-containing solution, usually in great molar excess [142]. A pre-requisite for soaking to be successful is that the active site is accessible once the protein is in its crystalline form. In addition, it is beneficial if ligand-binding does not induce large conformational changes to the overall or local structure, which would also limit the ligand access to the binding pocket [142]. Czech et al. (2019) produced crystals of PlEctC-ligand compexes with both *N*-γ.ADABA and ectoine. They discovered that when overlaid,

the active site-residues remained overall in the same position both spatially and rotationally, apart from Trp-21 [76]. This was also true for the overall structure, which could potentially mean that soaking is a sensible alternative for crystallizing CSEctC with substrate and product bound in the active site.

Interestingly, CSEctC crystallized at low pH, 4.0 during both initial screening and optimization. Low pH was also observed in crystallization conditions used to grow crystals from SAEctC and PLEctC [75, 76]. While it is generally difficult to draw a connection between crystallization conditions and biochemical properties of enzymes due to the sheer number of variables influencing crystal growth, this trend is noteworthy. A possible explanation can be found in the distribution of charged amino acids in EctC. In CSEctC, there are 24 negatively charged residues (aspartic acid and glutamic acid) (at physiological pH) and 14 positively charged residues (arginine and lysine). Enzymes living in acid environments are known to favour negatively charged amino acids as they contribute to improved protein stability [143]. This should be kept in mind for future crystallization trials with MarEctC and other EctC-type proteins.

While iron is likely catalytically essential for EctC activity, it is not necessary structurally essential. During purification we did not supplement the buffers with any iron, but EctC still remained soluble and folded as it was determined to be catalytically active after purification. This suggests that iron is not essential for structure integrity, and that the iron-coordinating position in the active site could potentially substitute a placeholder water molecule and remain well folded. In addition, we did not unequivocally identify an iron molecule in its likely position in the enzyme active site, but the enzyme still crystallized in its three-dimensional shape. This hypothesis is strengthened by Widderich et al. (2016) and Czech et al. (2019) who observed that the architecture of the iron-binding site remained virtually unchanged with and without and iron present [75, 76].

4.3 Thermostability of CSEctC and MarEctC

We managed to find experimental melting temperatures for both candidates in our storage buffer (50

mM HEPES pH 7.5, 500 mM NaCl, 0.4 mM TCEP). As expected, the mesophilic CSEctC showed a higher melting temperature than the psychrophilic MarEctC. According to their respective amino acid sequences, CSEctC has 38 charged residues (9 aspartic acids, 15 glutamic acids, 8 arginines and 6 lysines) while MarEctC has 33 (11 aspartic acids, 10 glutamatic acids, 6 arginines and 6 lysines), and this difference in charged amino acids could play a role in the higher stability of CSEctC since salt bridges is much stronger than hydrogen bonds and other weaker interactions. While the biggest contributors to protein stability in solution is considered to be hydrophobic interactions and hydrogen bonding, increased surface charges and salt bridges has been connected with increased thermal stability [144].

Generally, the protein environment, the buffer, salt content, pH and other physiological factors, can impact stability and fold integrity [145]. We tested thermostability with DSC in our storage buffer only, and experiments with varying buffer compositions could potentially shed more light on the mechanisms and differences in protein stability between our two targets. While DSC is considered a robust method for determining melting temperatures, it is time consuming and demands relatively large amounts of protein. An optional method for evaluating protein stability is fluorescence binding-based assays; thermofluor or thermal shift assays. Thermofluor assays are based on detection of a dye which binds to hydrophobic residues of a protein, which becomes gradually exposed during protein unfolding. When the dye binds to hydrophobic residues, it becomes strongly fluorescent, and the shift in fluorescence can be used to determine unfolding with a spectrophotometer [146]. The advantages of thermofluor are that it is much more suited to screen more variables, as one can screen many different conditions simultaneously. DSC however allows for only one condition per experiment, and an experiment can take a full day or more. A logical next step for evaluating thermal stability could therefore be to introduce a thermofluor based assay to evaluate effect of salt concentration, pH and buffer composition on CSEctC and MarEctC protein stability. It should be noted that during the thesis period we attempted to create a protocol for such an assay, but due to instrument and dye issues, we were not successful in collecting any reliable data from these experiments.

5 Conclusions and future work

The ectoine synthase protein from *Chromohalobacter salexigens DSM 3043* and *Marinobacter sp. CK1* were successfully cloned, expressed and purified. CSEctC proved to express well after optimization, while MarEctC expressed at relatively lower levels. Both EctC candidates crystallized, but only CSEctC yielded three-dimensional high quality crystals that diffracted and produced a usable data set. Structural determination of CSEctC revealed that it adapts a β-sandwich-type jelly roll fold, as observed in other EctC structures previously. We have applied several different techniques for biochemical characterization, and we obtained experimental melting temperatures for both enzymes, where the psychrophilic MarEctC predictably showed lower thermal stability than the mesophilic CSEctC. Overall, the work presented herein has given a thorough introduction to basic structural and functional studies of enzymes, and provides a solid foundation for further studies, both related to structure determination of MarEctC, crystallization of protein-ligand complexes and further stability and activity functional studies of both candidates.

A natural next step would be to further optimize crystallization of MarEct, to hopefully determine the structure of this EctC-type as well, and facilitate comparative studies between the psychrophilic MarEctC and the mesophilic CSEctC. If high quality crystals proves hard to obtain from MarEctC, it is an option to apply other structure determination techniques, like Nuclear Magnetic Resonance spectroscopy (NMR Spectroscopy. It would also be desirable to determine structures of both candidates with the substrate bound in the active site, to help better understand the mechanics of reaction in EctC. Furthermore, while we have applied several different biochemical techniques, most of our functional studies data is still quite preliminary. It would be natural to continue these efforts to further illuminiate the stability, thermodynamics and activity profiles our candidate enzymes.

References

1. Liu, M., et al., *Microbial production of ectoine and hydroxyectoine as high-value chemicals.* Microbial Cell Factories, 2021. **20**(1): p. 76.
2. Schuh, W., et al., *Die Kristallstruktur des Ectoin, einer neuen osmoregulatorisch wirksamen Aminosäure / The Crystal Structure of Ectoine, a Novel Amino Acid of Potential Osmoregulatory Function.* Zeitschrift für Naturforschung C, 1985. **40**(11-12): p. 780-784.
3. Hermann, L., et al., *The ups and downs of ectoine: structural enzymology of a major microbial stress protectant and versatile nutrient.* Biological Chemistry, 2020. **401**(12): p. 1443-1468.
4. Galinski, E.A., H.-P. Pfeiffer, and H.G. TrÜPer, *1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid.* European Journal of Biochemistry, 1985. **149**(1): p. 135-139.
5. Wohlfarth, A., J. Severin, and E.A. Galinski, *The spectrum of compatible solutes in heterotrophic halophilic eubacteria of the family Halomonadaceae.* Microbiology, 1990. **136**(4): p. 705-712.
6. Czech, L., et al., *Role of the Extremolytes Ectoine and Hydroxyectoine as Stress Protectants and Nutrients: Genetics, Phylogenomics, Biochemistry, and Structural Analysis.* Genes, 2018. **9**(4): p. 177.
7. Widderich, N., et al., *Biochemical Properties of Ectoine Hydroxylases from Extremophiles and Their Wider Taxonomic Distribution among Microorganisms.* PLOS ONE, 2014. **9**(4): p. e93809.
8. Widderich, N., et al., *Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon Nitrosopumilus maritimus.* Environmental Microbiology, 2016. **18**(4): p. 1227-1248.
9. Harding, T., et al., *Osmoadaptative Strategy and Its Molecular Signature in Obligately Halophilic Heterotrophic Protists.* Genome Biology and Evolution, 2016. **8**(7): p. 2241-2258.
10. Inbar, L. and A. Lapidot. *The Structure and Biosynthesis of New Tetrahydropyrimidine Derivatives in Actinomycin D Producer Streptomyces parvulus*. 1988.
11. Gunde-Cimerman, N., A. Plemenitaš, and A. Oren, *Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations.* FEMS Microbiology Reviews, 2018. **42**(3): p. 353-375.
12. Oren, A., *Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes.* Frontiers in microbiology, 2013. **4**: p. 315-315.
13. Talon, R., et al., *An experimental point of view on hydration/solvation in halophilic proteins.* Frontiers in microbiology, 2014. **5**: p. 66-66.
14. Brown, A.D., *Microbial water stress physiology. Principles and perspectives*. 1990, Chichester: John Wiley & Sons. xiii + 313 pp.
15. Bohnert, H.J., D.E. Nelson, and R.G. Jensen, *Adaptations to Environmental Stresses.* The Plant cell, 1995. **7**(7): p. 1099-1111.
16. Oren, A., *Bioenergetic aspects of halophilism.* Microbiology and molecular biology reviews : MMBR, 1999. **63**(2): p. 334-348.
17. Roberts, M.F., *Organic compatible solutes of halotolerant and halophilic microorganisms.* Saline Systems, 2005. **1**(1): p. 5.
18. Kuhlmann, A.U. and E. Bremer, *Osmotically regulated synthesis of the compatible solute ectoine in Bacillus pasteurii and related Bacillus spp.* Applied and environmental microbiology, 2002. **68**(2): p. 772-783.

19. Yu, I., Y. Jindo, and M. Nagaoka, *Microscopic Understanding of Preferential Exclusion of Compatible Solute Ectoine: Direct Interaction and Hydration Alteration.* The Journal of Physical Chemistry B, 2007. **111**(34): p. 10231-10238.
20. Graf, R., et al., *The multifunctional role of ectoine as a natural cell protectant.* Clinics in Dermatology, 2008. **26**(4): p. 326-333.
21. Crowley, E.L. *Compatible Solute Ectoine Review: Protection Mechanisms and Production Methods*. 2017.
22. Smiatek, J., et al., *Properties of compatible solutes in aqueous solution.* Biophysical Chemistry, 2012. **160**(1): p. 62-68.
23. Mojtabavi, S., N. Samadi, and M.A. Faramarzi, *Osmolyte-Induced Folding and Stability of Proteins: Concepts and Characterization.* Iranian journal of pharmaceutical research : IJPR, 2019. **18**(Suppl1): p. 13-30.
24. Bolen, D.W. and I.V. Baskakov, *The osmophobic effect: natural selection of a thermodynamic force in protein folding11Edited by D. Draper.* Journal of Molecular Biology, 2001. **310**(5): p. 955-963.
25. Li, H., C. Tang, and N.S. Wingreen, *Nature of Driving Force for Protein Folding: A Result From Analyzing the Statistical Potential.* Physical Review Letters, 1997. **79**(4): p. 765-768.
26. Camilloni, C., et al., *Towards a structural biology of the hydrophobic effect in protein folding.* Scientific Reports, 2016. **6**(1): p. 28285.
27. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884-890.
28. Harishchandra, R.K., et al., *The effect of compatible solute ectoines on the structural organization of lipid monolayer and bilayer membranes.* Biophysical Chemistry, 2010. **150**(1): p. 37-46.
29. Eiberweiser, A., et al., *Hydration and Ion Binding of the Osmolyte Ectoine.* The Journal of Physical Chemistry B, 2015. **119**(49): p. 15203-15211.
30. Bursy, J., et al., *Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by Streptomyces coelicolor A3(2) in response to salt and heat stresses.* Applied and environmental microbiology, 2008. **74**(23): p. 7286-7296.
31. Omara, A.M.A., et al., *Optimizing ectoine biosynthesis using response surface methodology and osmoprotectant applications.* Biotechnology Letters, 2020. **42**(6): p. 1003-1017.
32. Schröter, M.A., et al., *Ectoine protects DNA from damage by ionizing radiation.* Scientific reports, 2017. **7**(1): p. 15272-15272.
33. Sajjad, W., et al., *Ectoine: a compatible solute in radio-halophilic Stenotrophomonas sp. WMA-LM19 strain to prevent ultraviolet-induced protein damage.* Journal of Applied Microbiology, 2018. **125**(2): p. 457-467.
34. Mitsuhiko Mano, K.M., Tatsuo Nishimura, Yoshio Sugiyama, *Production of tetrahydropyrimidine derivatives*, T.C.I. LTD, Editor. 1991: Japan.
35. Hans Jorg, K., L. Georg, and A.G. Erwin, *Industrial Production of the Cell Protectant Ectoine: Protection Mechanisms, Processes, and Products.* Current Biotechnology, 2014. **3**(1): p. 10-25.
36. VREELAND, R.H., et al., *Halomonas elongata, a New Genus and Species of Extremely Salt-Tolerant Bacteria.* International Journal of Systematic and Evolutionary Microbiology, 1980. **30**(2): p. 485-495.
37. Cantera, S., et al., *Novel haloalkaliphilic methanotrophic bacteria: An attempt for enhancing methane bio-refinery.* Journal of Environmental Management, 2019. **231**: p. 1091-1099.

38. Vandrich, J., et al., *Contribution of mechanosensitive channels to osmoadaptation and ectoine excretion in Halomonas elongata.* Extremophiles : life under extreme conditions, 2020. **24**(3): p. 421-432.
39. Bownik, A. and Z. Stępniewska, *Ectoine as a promising protective agent in humans and animals.* Archives of Industrial Hygiene and Toxicology, 2016. **67**(4): p. 260-265.
40. Buenger, J. and H. Driller, *Ectoin: An Effective Natural Substance to Prevent UVA-Induced Premature Photoaging.* Skin Pharmacology and Physiology, 2004. **17**(5): p. 232-237.
41. Heinrich, U., B. Garbe, and H. Tronnier, *In vivo Assessment of Ectoin: A Randomized, Vehicle-Controlled Clinical Trial.* Skin Pharmacology and Physiology, 2007. **20**(4): p. 211-218.
42. Marini, A., et al., *Ectoine-Containing Cream in the Treatment of Mild to Moderate Atopic Dermatitis: A Randomised, Comparator-Controlled, Intra-Individual Double-Blind, Multi-Center Trial.* Skin Pharmacology and Physiology, 2014. **27**(2): p. 57-65.
43. Werkhäuser, N., A. Bilstein, and U. Sonnemann, *Treatment of allergic rhinitis with ectoine containing nasal spray and eye drops in comparison with azelastine containing nasal spray and eye drops or with cromoglycic Acid containing nasal spray.* Journal of allergy, 2014. **2014**: p. 176597-176597.
44. Dwivedi, M., et al., *Biophysical investigations of the structure and function of the tear fluid lipid layers and the effect of ectoine. Part B: Artificial lipid films.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2014. **1838**(10): p. 2716-2727.
45. Bujak, T., M. Zagórska-Dziok, and Z. Nizioł-Łukaszewska, *Complexes of Ectoine with the Anionic Surfactants as Active Ingredients of Cleansing Cosmetics with Reduced Irritating Potential.* Molecules (Basel, Switzerland), 2020. **25**(6): p. 1433.
46. Kanapathipillai, M., et al., *Ectoine and hydroxyectoine inhibit aggregation and neurotoxicity of Alzheimer's β-amyloid.* FEBS Letters, 2005. **579**(21): p. 4775-4780.
47. Bazazzadegan, N., et al., *Effects of Ectoine on Behavior and Candidate Genes Expression in ICV-STZ Rat Model of Sporadic Alzheimer's Disease.* Advanced pharmaceutical bulletin, 2017. **7**(4): p. 629-636.
48. Kanapathipillai, M., et al., *Small stress molecules inhibit aggregation and neurotoxicity of prion peptide 106–126.* Biochemical and Biophysical Research Communications, 2008. **365**(4): p. 808-813.
49. Knapp, S., R. Ladenstein, and E.A. Galinski, *Extrinsic protein stabilization by the naturally occurring osmolytes beta-hydroxyectoine and betaine.* Extremophiles, 1999. **3**(3): p. 191-8.
50. Piszkiewicz, S. and G.J. Pielak, *Protecting Enzymes from Stress-Induced Inactivation.* Biochemistry, 2019. **58**(37): p. 3825-3833.
51. Malin, G., R. Iakobashvili, and A. Lapidot, *Effect of Tetrahydropyrimidine Derivatives on Protein-Nucleic Acids Interaction: TYPE II RESTRICTION ENDONUCLEASES AS A MODEL SYSTEM *.* Journal of Biological Chemistry, 1999. **274**(11): p. 6920-6929.
52. Schnoor, M., et al., *Characterization of the synthetic compatible solute homoectoine as a potent PCR enhancer.* Biochemical and Biophysical Research Communications, 2004. **322**(3): p. 867-872.
53. Mardian, Y., et al., *Review of Current COVID-19 Diagnostics and Opportunities for Further Development.* Frontiers in Medicine, 2021. **8**(562).
54. Steffen-Munsberg, F., et al., *Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications.* Biotechnology Advances, 2015. **33**(5): p. 566-604.

55. Coscolín, C., et al., *Bioprospecting Reveals Class III ω-Transaminases Converting Bulky Ketones and Environmentally Relevant Polyamines.* Applied and Environmental Microbiology, 2019. **85**(2): p. e02404-18.
56. Haruyama, K., et al., *Structures of Escherichia coli Histidinol-Phosphate Aminotransferase and Its Complexes with Histidinol-Phosphate and N-(5'-Phosphopyridoxyl)-l-Glutamate: Double Substrate Recognition of the Enzyme.* Biochemistry, 2001. **40**(15): p. 4633-4644.
57. Binter, A., et al., *Characterization of the PLP-dependent aminotransferase NikK from Streptomyces tendae and its putative role in nikkomycin biosynthesis.* The FEBS Journal, 2011. **278**(21): p. 4122-4135.
58. Hillier, H.T., B. Altermark, and I. Leiros, *The crystal structure of the tetrameric DABA-aminotransferase EctB, a rate-limiting enzyme in the ectoine biosynthesis pathway.* Febs j, 2020. **287**(21): p. 4641-4658.
59. Reuter, K., et al., *Synthesis of 5-hydroxyectoine from ectoine: crystal structure of the non-heme iron(II) and 2-oxoglutarate-dependent dioxygenase EctD.* PloS one, 2010. **5**(5): p. e10647-e10647.
60. Richter, A.A., et al., *Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB.* Frontiers in Microbiology, 2019. **10**(2811).
61. Ono, H., et al., *Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, Halomonas elongata.* Journal of bacteriology, 1999. **181**(1): p. 91-99.
62. Richter, A.A., et al., *Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB.* Frontiers in microbiology, 2019. **10**: p. 2811-2811.
63. Tramonti, A., et al., *The MocR-like transcription factors: pyridoxal 5'-phosphate-dependent regulators of bacterial metabolism.* The FEBS Journal, 2018. **285**(21): p. 3925-3944.
64. Richter, A.A., et al., *The architecture of the diaminobutyrate acetyltransferase active site provides mechanistic insight into the biosynthesis of the chemical chaperone ectoine.* The Journal of biological chemistry, 2020. **295**(9): p. 2822-2838.
65. Neuwald, A.F. and D. Landsman, *GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein.* Trends in Biochemical Sciences, 1997. **22**(5): p. 154-155.
66. Reshetnikov, A.S., V.N. Khmelenina, and Y.A. Trotsenko, *Characterization of the ectoine biosynthesis genes of haloalkalotolerant obligate methanotroph "Methylomicrobium alcaliphilum 20Z".* Archives of Microbiology, 2006. **184**(5): p. 286-297.
67. Reshetnikov, A.S., et al., *Chapter Two - Genes and Enzymes of Ectoine Biosynthesis in Halotolerant Methanotrophs*, in *Methods in Enzymology*, A.C. Rosenzweig and S.W. Ragsdale, Editors. 2011, Academic Press. p. 15-30.
68. Czech, L., et al., *Ectoine Synthase: An Iron-Dependent Member of the Cupin Superfamily*, in *Encyclopedia of Inorganic and Bioinorganic Chemistry*. p. 1-14.
69. Dunwell, J.M., A. Purvis, and S. Khuri, *Cupins: the most functionally diverse protein superfamily?* Phytochemistry, 2004. **65**(1): p. 7-17.
70. Dunwell, J.M., *Cupins: A New Superfamily of Functionally Diverse Proteins that Include Germins and Plant Storage Proteins.* Biotechnology and Genetic Engineering Reviews, 1998. **15**(1): p. 1-32.

71. Uberto, R. and E.W. Moomaw, *Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily.* PLOS ONE, 2013. **8**(9): p. e74477.
72. Uberto, R. and E.W. Moomaw, *Protein similarity networks reveal relationships among sequence, structure, and function within the Cupin superfamily.* PloS one, 2013. **8**(9): p. e74477-e74477.
73. Woo, E.-J., et al., *Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities.* Nature Structural Biology, 2000. **7**(11): p. 1036-1040.
74. Hashimoto, H., et al., *The first crystal structure of a family 31 carbohydrate-binding module with affinity to β-1,3-xylan.* FEBS Letters, 2005. **579**(20): p. 4324-4328.
75. Widderich, N., et al., *Biochemistry and Crystal Structure of Ectoine Synthase: A Metal-Containing Member of the Cupin Superfamily.* PloS one, 2016. **11**(3): p. e0151285-e0151285.
76. Czech, L., et al., *Illuminating the catalytic core of ectoine synthase through structural and biochemical analysis.* Scientific Reports, 2019. **9**(1): p. 364.
77. Moritz, K.D., et al., *The hydroxyectoine gene cluster of the non-halophilic acidophile Acidiphilium cryptum.* Extremophiles, 2015. **19**(1): p. 87-99.
78. Rampelotto, P.H., *Extremophiles and extreme environments.* Life (Basel, Switzerland), 2013. **3**(3): p. 482-485.
79. Merino, N., et al., *Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context.* Frontiers in microbiology, 2019. **10**: p. 780-780.
80. Ventosa, A. and R.R. de la Haba, *Chromohalobacter*, in *Bergey's Manual of Systematics of Archaea and Bacteria*. p. 1-16.
81. Arahal, D.R., et al., *Chromohalobacter salexigens sp. nov., a moderately halophilic species that includes Halomonas elongata DSM 3043 and ATCC 33174.* International Journal of Systematic and Evolutionary Microbiology, 2001. **51**(4): p. 1457-1462.
82. Vargas, C., et al., *Ectoines as compatible solutes and carbon and energy sources for the halophilic bacterium Chromohalobacter salexigens.* Journal of Applied Microbiology, 2006. **100**(1): p. 98-107.
83. Handley, K. and J. Lloyd, *Biogeochemical implications of the ubiquitous colonization of marine habitats and redox gradients by Marinobacter species.* Frontiers in Microbiology, 2013. **4**(136).
84. Raddadi, N., et al., *Marinobacter sp. from marine sediments produce highly stable surface-active agents for combatting marine oil spills.* Microbial Cell Factories, 2017. **16**(1): p. 186.
85. Itakura, K., et al., *Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin.* Science, 1977. **198**(4321): p. 1056.
86. Roberts, R.J., *How restriction enzymes became the workhorses of molecular biology.* Proceedings of the National Academy of Sciences, 2005. **102**(17): p. 5905.
87. Walhout, A.J.M., et al., *[34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes*, in *Methods in Enzymology*, J. Thorner, S.D. Emr, and J.N. Abelson, Editors. 2000, Academic Press. p. 575-IN7.
88. Hartley, J.L., G.F. Temple, and M.A. Brasch, *DNA cloning using in vitro site-specific recombination.* Genome research, 2000. **10**(11): p. 1788-1795.
89. Reece-Hoyes, J.S. and A.J.M. Walhout, *Propagating Gateway Vectors.* Cold Spring Harbor protocols, 2018. **2018**(1): p. pdb.prot094920-pdb.prot094920.

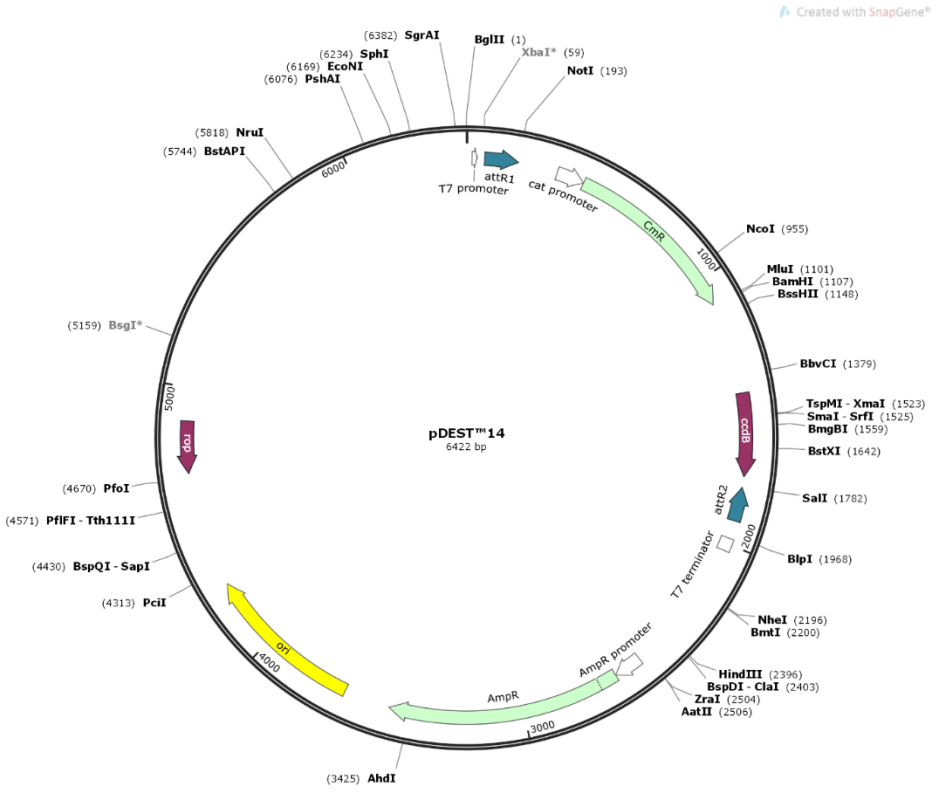
90. Sutcliffe, J.G., *Nucleotide Sequence of the Ampicillin Resistance Gene of Escherichia coli Plasmid pBR322.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(8): p. 3737-3741.
91. Kokotek, W. and W. Lotz, *Construction of a lacZ-kanamycin-resistance cassette, useful for site-directed mutagenesis and as a promoter probe.* Gene, 1989. **84**(2): p. 467-471.
92. Tabor, S., *Expression Using the T7 RNA Polymerase/Promoter System.* Current Protocols in Molecular Biology, 1990. **11**(1): p. 16.2.1-16.2.11.
93. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in microbiology, 2014. **5**: p. 172-172.
94. Wang, L., et al., *Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces.* Scientific Reports, 2015. **5**(1): p. 15159.
95. Young, C.L., Z.T. Britton, and A.S. Robinson, *Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications.* Biotechnology Journal, 2012. **7**(5): p. 620-634.
96. Kendrew, J.C., et al., *A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis.* Nature, 1958. **181**(4610): p. 662-666.
97. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
98. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization.* Acta crystallographica. Section F, Structural biology communications, 2014. **70**(Pt 1): p. 2-20.
99. McPherson, A., *Protein Crystallization*, in *Protein Crystallography: Methods and Protocols*, A. Wlodawer, Z. Dauter, and M. Jaskolski, Editors. 2017, Springer New York: New York, NY. p. 17-50.
100. Haas, C. and J. Drenth, *Understanding protein crystallization on the basis of the phase diagram.* Journal of Crystal Growth, 1999. **196**(2): p. 388-394.
101. McGinty, J., et al., *CHAPTER 1 Nucleation and Crystal Growth in Continuous Crystallization*, in *The Handbook of Continuous Crystallization*. 2020, The Royal Society of Chemistry. p. 1-50.
102. Pichlo, C., et al., *Production, Crystallization and Structure Determination of C. difficile PPEP-1 via Microseeding and Zinc-SAD.* Journal of Visualized Experiments, 2016. **2016**.
103. Smyth, M.S. and J.H. Martin, *x ray crystallography.* Molecular pathology : MP, 2000. **53**(1): p. 8-14.
104. Lorber, B., et al., *Protein analysis by dynamic light scattering: Methods and techniques for students.* Biochemistry and Molecular Biology Education, 2012. **40**(6): p. 372-382.
105. Proteau, A., R. Shi, and M. Cygler, *Application of Dynamic Light Scattering in Protein Crystallization.* Current Protocols in Protein Science, 2010. **61**(1): p. 17.10.1-17.10.9.
106. Stetefeld, J., S.A. McKenna, and T.R. Patel, *Dynamic light scattering: a practical guide and applications in biomedical sciences.* Biophysical reviews, 2016. **8**(4): p. 409-427.
107. Jelesarov, I. and H.R. Bosshard, *Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition.* Journal of Molecular Recognition, 1999. **12**(1): p. 3-18.
108. Johnson, C.M., *Differential scanning calorimetry as a tool for protein folding and stability.* Archives of Biochemistry and Biophysics, 2013. **531**(1): p. 100-109.

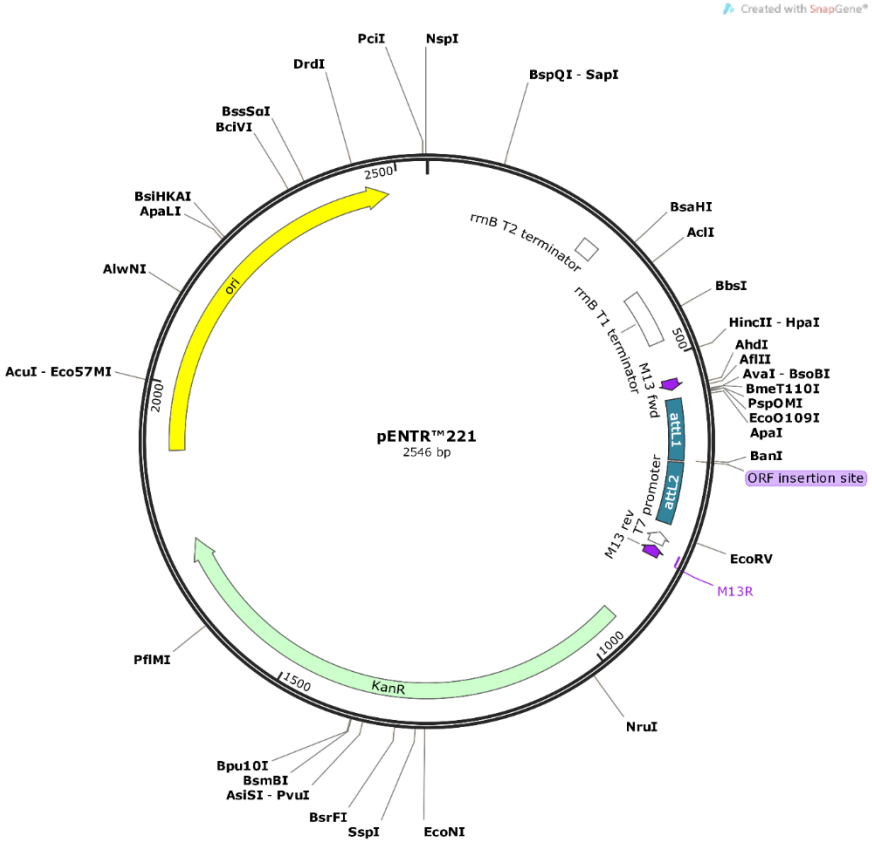
109. Damian, L., *Isothermal Titration Calorimetry for Studying Protein–Ligand Interactions*, in *Protein-Ligand Interactions: Methods and Applications*, M.A. Williams and T. Daviter, Editors. 2013, Humana Press: Totowa, NJ. p. 103-118.
110. Archer, W.R. and M.D. Schulz, *Isothermal titration calorimetry: practical approaches and current applications in soft matter.* Soft Matter, 2020. **16**(38): p. 8760-8774.
111. Wicks, I.P., et al., *Bacterial Lipopolysaccharide Copurifies with Plasmid DNA: Implications for Animal Models and Human Gene Therapy.* Human Gene Therapy, 1995. **6**(3): p. 317-323.
112. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.
113. Krug, M., Weiss, M. S., Heinemann, U., & Mueller, U.. (2012). XDSAPP: a graphical user interface for the convenient processing of diffraction data usingXDS. Journal of Applied Crystallography, 45(3), 568–572. https://doi.org/10.1107/s0021889812011715.
114. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta crystallographica. Section D, Biological crystallography, 2011. **67**(Pt 4): p. 235-242.
115. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta crystallographica. Section D, Biological crystallography, 2013. **69**(Pt 7): p. 1204-1214.
116. Vagin, A. and A. Teplyakov, *MOLREP: an Automated Program for Molecular Replacement.* Journal of Applied Crystallography, 1997. **30**(6): p. 1022-1025.
117. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta crystallographica. Section D, Biological crystallography, 2011. **67**(Pt 4): p. 355-367.
118. Cowtan, K., *TheBuccaneersoftware for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D Biological Crystallography, 2006. **62**(9): p. 1002-1011.
119. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D Biological Crystallography, 2004. **60**(12): p. 2126-2132.
120. Schrodinger, L.T.P.M.G.S., Version 2.4.
121. Consortium, T.U., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Research, 2018. **47**(D1): p. D506-D515.
122. Ashkenazy, H., et al., *ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.* Nucleic Acids Research, 2016. **44**(W1): p. W344-W350.
123. Okonechnikov, K., et al., *Unipro UGENE: a unified bioinformatics toolkit.* Bioinformatics, 2012. **28**(8): p. 1166-1167.
124. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Research, 2014. **42**(W1): p. W320-W324.
125. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic acids research, 2018. **46**(W1): p. W296-W303.
126. Biasini, M., et al., *OpenStructure: an integrated software framework for computational structural biology.* Acta crystallographica. Section D, Biological crystallography, 2013. **69**(Pt 5): p. 701-709.
127. Lucena-Aguilar, G., et al., *DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis.* Biopreservation and biobanking, 2016. **14**(4): p. 264-270.

128. Shi, Y., et al., *Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding.* Protein Science, 2012. **21**(8): p. 1197-1209.
129. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-410.
130. Jeong, H., H.J. Kim, and S.J. Lee, *Complete Genome Sequence of Escherichia coli Strain BL21.* Genome announcements, 2015. **3**(2): p. e00134-15.
131. Vadlamani, G., et al., *The β-lactamase gene regulator AmpR is a tetramer that recognizes and binds the D-Ala-D-Ala motif of its repressor UDP-N-acetylmuramic acid (MurNAc)-pentapeptide.* The Journal of biological chemistry, 2015. **290**(5): p. 2630-2643.
132. McPherson, A. and B. Cudney, *Optimization of crystallization conditions for biological macromolecules.* Acta crystallographica. Section F, Structural biology communications, 2014. **70**(Pt 11): p. 1445-1467.
133. Chen, R., *Bacterial expression systems for recombinant protein production: E. coli and beyond.* Biotechnology Advances, 2012. **30**(5): p. 1102-1107.
134. Ferrer, M., et al., *Chaperonins govern growth of Escherichia coli at low temperatures.* Nature Biotechnology, 2003. **21**(11): p. 1266-1267.
135. Bjerga, G.E.K. and A.K. Williamson, *Cold shock induction of recombinant Arctic environmental genes.* BMC Biotechnology, 2015. **15**(1): p. 78.
136. Bjerga, G.E.K., R. Lale, and A.K. Williamson, *Engineering low-temperature expression systems for heterologous production of cold-adapted enzymes.* Bioengineered, 2016. **7**(1): p. 33-38.
137. Tanabe, H., et al., *Identification of the promoter region of the Escherichia coli major cold shock gene, cspA.* Journal of bacteriology, 1992. **174**(12): p. 3867-3873.
138. Planson, A.-G., et al., *Assistance of Maltose Binding Protein to the in Vivo Folding of the Disulfide-Rich C-Terminal Fragment from Plasmodium falciparum Merozoite Surface Protein 1 Expressed in Escherichia coli.* Biochemistry, 2003. **42**(45): p. 13202-13211.
139. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
140. Kapust, R.B. and D.S. Waugh, *Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.* Protein science : a publication of the Protein Society, 1999. **8**(8): p. 1668-1674.
141. Zuo, X., et al., *Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli.* Journal of structural and functional genomics, 2005. **6**(2-3): p. 103-111.
142. Müller, I., *Guidelines for the successful generation of protein-ligand complex crystals.* Acta crystallographica. Section D, Structural biology, 2017. **73**(Pt 2): p. 79-92.
143. Reed, C.J., et al., *Protein adaptations in archaeal extremophiles.* Archaea (Vancouver, B.C.), 2013. **2013**: p. 373275-373275.
144. Chan, C.-H., T.-H. Yu, and K.-B. Wong, *Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding.* PloS one, 2011. **6**(6): p. e21624-e21624.
145. Brudar, S. and B. Hribar-Lee, *Effect of Buffer on Protein Stability in Aqueous Solutions: A Simple Protein Aggregation Model.* The Journal of Physical Chemistry B, 2021. **125**(10): p. 2504-2512.
146. Bai, N., et al., *Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities.* Scientific Reports, 2019. **9**(1): p. 2650.

Appendices

Appendix A –Entry and expression vectors used for CSEctC an MarEctC





Appendix B – Amino acid sequences of CSEctC and MarEctC

CSEctC:

MIVRNLEECRKTERFVEAENGNWDSTRLVLADDNVGFSFNITRIHPGTETHIHYKHHFEAVFCYEGEGEVETLADGK

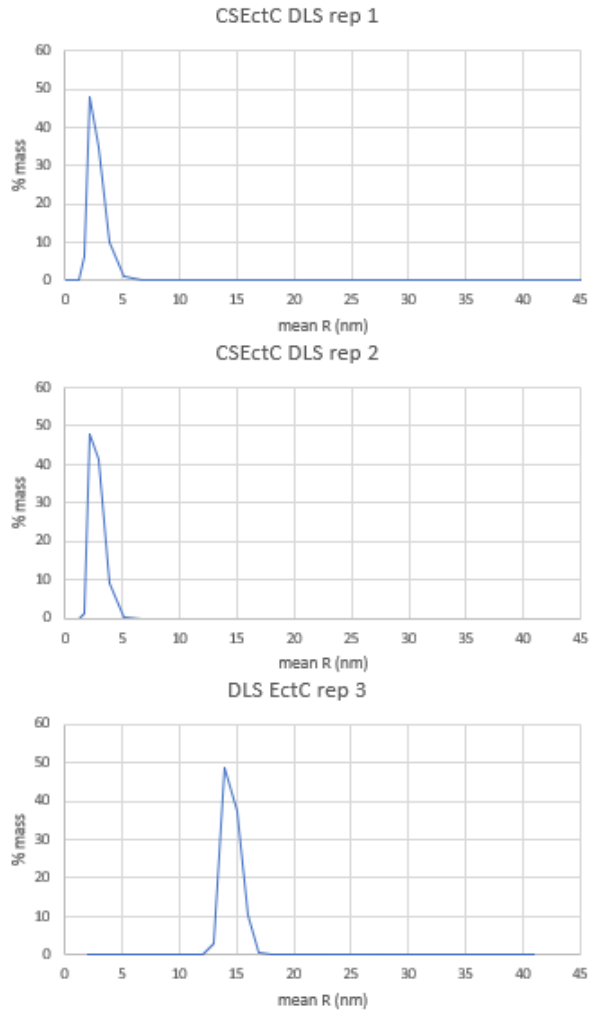
IHPIKAGDMYLLDQHDEHLLRGKEKGMTVACVFNPALTGREVHREDGSYAPVD

MarEctC:

MKIVRVQDIIGTEREVHGPGWTSRRMLLKKDGMGFSFHETIIPAGAELNLWYKHHLEAVYCVAGNGSITDKATGET

HQISDGTLYALDKHDQHTLRGGTEDMRLICSFNPPVTGQEVHDEDGAYLPDTSED

Appendix C – DLS data from CSEctC



Appendix D – ITC experiments

Table: Summary of all ITC experiments conducted with EctC to find a suitable molar ratio between enzyme and substrate.

Experiment ID	[CSEctC] (µM)	[ADABA] (mM)	Injections (no.)

ITC1	150	6	48
ITC2	50	2	48
ITC3	8	3	?
ITC4	8	4	120
ITC5	8	4	100
ITC6	8	5	48
ITC7	8	6	44

Appendix E – Activity of CSEctC and MarEctC at increasing pH