Production and Purification of the Proposed Bacteriocin Receptor YthA in Preparation for NMR Analysis

Thesis for the Degree of Master of Science

Alec Spera

Supervisor: Per Eugen Kristiansen

Genetics and Developmental Biology

60 Study Credits

Department of Biosciences

The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

2023

Dedicated to my family, especially my mother: Yvonne Mikitzel-Spera, who believed in me when I couldn't believe in myself; and to my supervisor, Per Eugen Kristiansen, who put up with my incessant requests, delays and stupid questions.

Acknowledgements

The author of this thesis would like to acknowledge the University of Oslo for the providing the equipment and reagents used in this thesis, as well as access to the scientific literature for the planning, design and writing of this thesis; Per Eugen Kristiansen for guidance on experimental design and thesis writing; and my friends in Norway, Canada, and else who supported me and allowed me to bounce ideas off them.

Table of Contents

Abstract

The emergence of multi-antibacterial resistant pathogens worldwide has become a crisis and new therapies are needed to combat this threat. One potential solution is the use of bacteriocins: bacterially produced polypeptides which killing competing species of bacteria. Here we describe an attempt to purify and determine the structure of putative bacteriocin receptor YthA. We successfully produced 320mg of inclusion bodies per litre of growth media, most of which was recombinant protein. Fusion protein precipitation caused the loss of much of the recovered fusion protein, though ultimately purification and refolding attempts were successful.

1 Introduction

1.1 The Threat of Antimicrobial Resistance

Among the most serious issues currently facing mankind is the emergence of 'super-bugs': multi-antibacterial resistant pathogens. Of particular concern are those pathogens capable of infecting humans, as many treatments are becoming increasingly ineffective against them due to the development of resistance. This issue is so severe that the World Health Organization has declared it one of the largest challenges humanity faces and the United Nations have described antimicrobial resistance as a 'fundamental threat' to global health and safety. Indeed, all 193 UN member states have signed a declaration to combat the proliferation of microbial resistance (Holpuch, 2016).

The WHO has identified 12 pathogenic bacteria which are in urgent need of novel antibiotics to combat and pose a world-wide threat to public health (WHO, 2017). Table 1 highlights bacteria that are resistant to multiple antibiotics and can spread resistance to other bacteria. The priority was determined by the severity of infections caused by the bacteria and the vulnerability of those who are likely to fall ill by it. For example, the critical priority class includes those bacteria that are resistant to several drugs; cause severe and deadly infections; and infect those who are particularly vulnerable, such as hospital patients, nursing home patients and those who require devices such as catheters and ventilators (WHO, 2017).

Table 1: Antimicrobial-resistant microorganisms listed by the WHO and CDC as having the potential to cause a health crisis. Backgrounds denote the highest priority between CDC and WHO concerns. A red background denotes critical priority, yellow backgrounds are high priority, green backgrounds are medium priority and blue backgrounds are only on a watch list. The classification; result of the gram test; and the physiological location they attack in humans is also listed. Data from WHO, 2017 and HHS, 2019.

Bacteria	Causes for Concern	Gram status	Infection Location	Authority
Acinetobacter baumannii	Carbapenem-resistant	Negative	Blood, Urinary Tract, Lungs, Wounds	CDC, WHO
Enterobacteriaceae	Carbapenem-resistant, Extended-spectrum β-Lactamase producing	Negative	Soft Tissues and Blood, especially Urinary and Respiratory Tracts	CDC, WHO
Neisseria gonorrhoeae	Cephalosporin-resistant, Fluoroquinolone-resistant	Negative	Primarily Mucosal Membranes of the Reproductive Tract, but also Eyes, Throat, and Joints	CDC, WHO
Pseudomonas aeruginosa	Carbapenem-resistant	Negative	Opportunistic, infects many bodily locations	CDC, WHO
Campylobacter spp.	Fluoroquinolone-resistant	Negative	Stomach and Intestinal Tract	CDC, WHO
Helicobacter pylori	Clarithromycin-resistant	Negative	Stomach	WHO
Salmonellae	Fluoroquinolone-resistant	Negative	Intestinal Tract	CDC, WHO
Shigella spp.	Fluoroquinolone-resistant	Negative	Intestinal Tract	CDC, WHO
Haemophilus influenzae	Ampicillin-resistant	Negative	Ears, Blood, and Respiratory Tract	WHO
Bordetella pertussis	Multiple Drug-resistant	Negative	Lungs	CDC
Mycoplasma genitalium	Multiple Drug-resistant	Negative	Urinary and Genital Tracts	CDC
Clostridioides difficile	Vancomycin-resistant	Positive	Colon and Intestinal Tract	CDC
Enterococcus faecium	Vancomycin-resistant	Positive	Blood, Urinary Tract, Wounds	CDC, WHO
Staphylococcus aureus	Methicillin-resistant, Vancomycin-resistant	Positive	Skin, Blood, Joints, and many Soft Tissues	CDC, WHO
Streptococcus pneumoniae	Penicillin-non-susceptible	Positive	Respiratory Tract	CDC, WHO

Current estimates place over 4.95 million deaths annually associated with antimicrobial-resistant pathogens (Murray *et al.,* 2022). This belies the ongoing threat of AMR, as the annual death toll could reach 10 million and cost upwards of 8 trillion USD in global GDP every year by 2050 (O'Neill, 2014). For context, this is over 13 times the death toll from COVID-19 as of November 3rd, 2021 (WHO, 2021) and the same annual economic cost as COVID-19 (Cutler & Summers, 2020). Already more than 2.8 million AMR infections occur each year within the United States of America, a

number which is expected to increase greatly without a drastic response. This is despite the hundreds of millions of USD invested by the CDC in preventing and containing infections, and research in new treatments (HHS, 2019)

1.1.1 Bacteria and The Emergence of Antibacterial Resistance

Bacteria are classically grouped into two broad groups, gram-positive and gram-negative, based on the structure of the cell membrane and cell wall. In gram-positive bacteria, a single cell membrane is surrounded by a thick, rigid layer of sugars and peptide called peptidoglycan, as seen in Figure 1. This layer allows the bacteria to retain its shape and prevents the cell from lysing. The peptidoglycans are also extensively cross-linked to further improve stability. Gram-negative bacteria have only a thin, comparatively flexible peptidoglycan layer sandwiched between two lipid membranes to provide the necessary stability. Unlike gram-positive bacteria, additional or lengthened proteins are needed to relay signals and molecules through the intermembrane space as in Figure 1. In either case, the cell membrane is a critical component of all bacteria as it protects the cell components from the environment and prevents the components and molecules within from escaping the cell. This makes the stability and synthesis of the membrane an obvious target for antibiotics and naturally occurring antimicrobial peptides (AMPs) to neutralize the bacteria. However, it should be noted that gram status does not allow classification into different phylogenetic groups, as members of the same group are not necessarily more closely related to each other than they are to members of the other group.

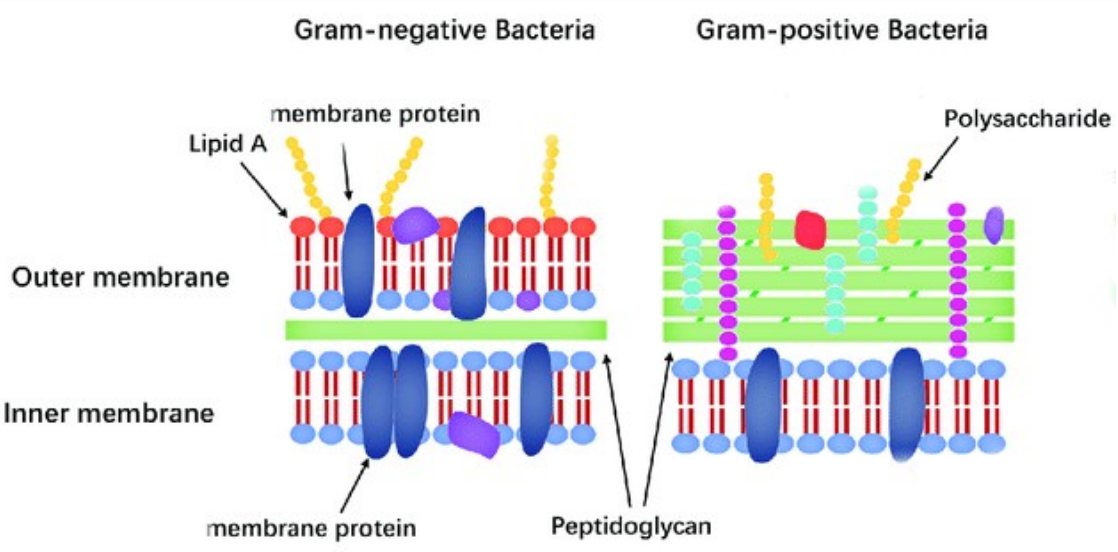


Figure 1: Diagram of Gram-negative cell membrane and Gram-positive cell membrane and their components. Adapted from Huan et al. *2020.*

Bacteria consist of several components that function together to allow the cell to grow, survive and reproduce. The DNA of the cell contains the information required for the cell to function effectively, including the sequences of all proteins found in the cell. However, RNA is transcribed to perform several functions including carrying the information from DNA to the ribosomes, which produce proteins according to the RNA sequence. Finally, proteins perform a wide variety of tasks necessary for the cell to function and reproduce. In addition to the membrane, all these components are necessary for the cell to operate and must be replicable for the bacteria to reproduce. As such, a disruption to any of these systems can kill bacteria or prevent its reproduction, and so there are several available targets for an outside force or molecule to attack the bacteria. Antimicrobials are molecules that act to disrupt one or more of these systems with minimal effect to the bacterial host.

Bacteria can adapt to these antimicrobials, and mutations to their genes can reduce the effectiveness of the outside agent. This reduction in effectiveness of antimicrobials is termed 'resistance'. The speed at which resistance to typical antibiotics can occur is another factor which challenges pharmaceutical treatments. Resistance has developed to several antibiotics within the same year of the antibiotic being released or even earlier. Famously, penicillin resistant strains of *Staphylococcus aureus* were reported before the drug was released to the public (Abraham & Chain, 1940). More recently, Fluconzole was approved for use two years after a resistant strain was discovered (HHS, 2019).

Given the threat of antibiotic resistance poses, it has become critical to ensure that effective treatments remain available. Continued exposure to a given antibiotic speeds the development of resistance to that antibiotic, and potentially others from cross-resistance (Nathan & Cars, 2014). While reducing non-essential antibiotic usage, tailoring antibiotic use to need, prioritizing human use over animal use, preventing over-the-counter sales and other such measures should improve an individual antibiotic's effective lifespan, developing new classes of antibiotics will be necessary to meet that goal (Nathan & Cars, 2014).

1.2 Antibiotics in Current Use

Antimicrobials function by several mechanisms and are classed by their mechanism of action. β-lactams and glycopeptides antibiotics both target the cell wall of bacteria. β-lactams bind to penicillin binding proteins (PBPs) weakening the cell wall cross-linking and lysing the cell, whereas glycopeptides bind to D-alanyl D-alanine peptide precursors and prevent cell wall synthesis (Kapoor *et al.*, 2017). Aminoglycosides, Tetracyclines, Chloramphenicol, Macrolides and Oxazolidinones all attack the ribosomes of the cell to prevent protein synthesis. Aminoglycosides target the 30S subunit

of bacterial ribosomes, bind near the A site and cause translation errors and premature terminations. Conversely, Tetracyclines bind to the 16S RNA of the subunit and prevent t-RNA binding, stopping synthesis altogether (Kapoor *et al.*, 2017). Chloramphenicol, Macrolides, and Oxazolidinones target the 50S subunit of the ribosomes and interfere with its function, also preventing protein synthesis. Other antibiotic classes and strategies disrupt DNA replication, such as quinolones, sulfonamides, and trimethoprim. Quinolones attack bacterial DNA gyrases and cause super-winding of DNA, and sulfonamides and trimethoprim inhibit the folic acid metabolism, thereby preventing the synthesis of nucleic acids and thus the duplication of DNA (Kapoor *et al.*, 2017).

Antimicrobial resistance occurs via numerous mechanisms which further complicates the issue. Any mutation in the cell that causes an antibiotic to have a reduced effect can be considered resistance, and one strain may possess multiple resistance mechanisms and limit effective strategies for designing antibiotics targeting the pathogen. Changes to the cell wall structure or its charge could prevent β-lactams from binding to PBPs and prevent membrane disruption. Overexpression of efflux pumps removes the antibiotic from the cell before it can reach a functional concentration. Expression of enzymes degrading antibiotics simply inactivate the molecule and have a similar effect. Mutations to DNA gyrase and topoisomerases prevent the antibiotics from interacting with them at all allowing DNA to be replicated (Bax & Griffin, 2012).

One potential solution to these issues is the use of antimicrobial peptides instead of traditional small-molecule antimicrobial drugs. One promising class of molecules are bacteriocins. Low effective doses, potentially in the nanomolar range, general lack of toxicity to humans and the existence of both broad and narrow spectrum bacteriocins provide notable advantages (Lei *et al.,* 2019) to their use. It should also be noted that their conservation within species suggest long-term viability. Though low yields, poor shelf life, stability at appropriate pH and temperature, and expensive equipment and expertise that are necessary for production may so far have been a hinderance to its wide-spread use (Abriouel *et al.,* 2011). Modification of the molecule or designing a new molecule with the same mechanism of action may mitigate these issues by increasing the yield, shelf-life and allowing chemical synthesis of the molecule (Ageitos *et al.,* 2017).

1.3 Novel Antibiotic Development

The high-throughput screening broad-spectrum antibiotic pipeline is a development scheme to identify promising antimicrobial treatments and prepare them for market; and consists of eight to nine phases (Payne *et al.,* 2007). First, a potential target protein is identified by comparing sequences from numerous targeted bacteria to identify those with a conserved sequence. Those targets with

human homologues, those difficult to reach or which were found to be unnecessary for viability are excluded. Next molecules potentially affecting the target are identified from a library of compounds and the most promising molecules are identified based on selectivity, sensitivity, stability, and potential for production (Payne *et al.,* 2007). These molecules then undergo further development in those areas to produce a candidate for testing. After this point, clinical trials may begin at phase I starting with a small group of healthy volunteers given a variety of doses to assess safety. Should the drug pass phase I, phase II is started with volunteers infected by a targeted pathogen to assess effectiveness, safety, determine side effects and identify an appropriate dosage. Phase III is much like phase II, having a larger patient pool and testing for effectiveness. Finally, the drug is submitted for approval to treat the targeted pathogens and for approval to market. Figure 2 is a visual representation of this process.

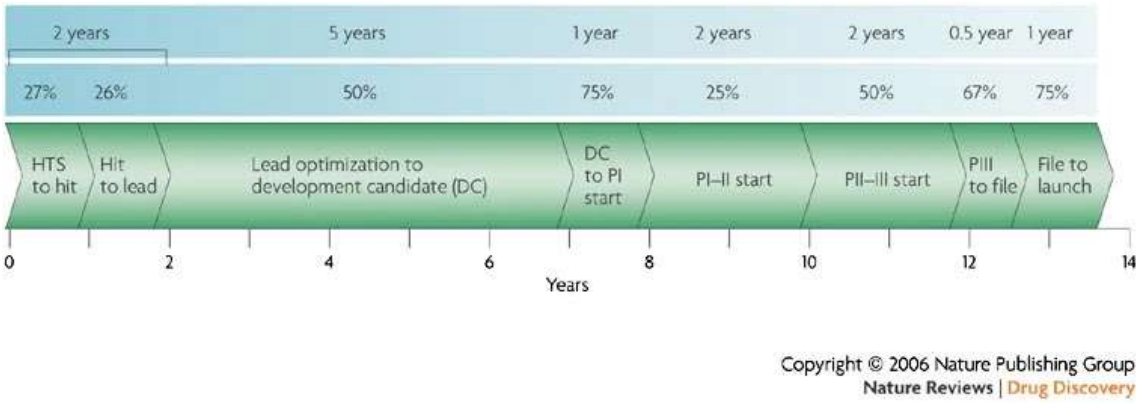


Figure 2: Flowchart describing the drug development pipeline, chances of success of each step and expected time necessary for each step. Figure courtesy of Payne *et al.,* 2007.

Small molecule antimicrobial drugs may not be able to keep pace with the rate of advancement of antibiotic resistance for a variety of reasons (Payne *et al.*, 2007). Antibiotics take years and often decades between the identification of a promising molecule and its release as a new antibiotic; additionally, less than 0.2% of such molecules pass all stages of development and testing and make it to market as antibiotics (Payne *et al.*, 2007). Worse, resistant strains can develop within hours and days of exposure to an antibiotic, limiting the antibiotic's effective lifespan (Payne *et al.,* 2007). This last point has also led to new antibiotics being used as treatments of last resort in order to extend the effectiveness of the antibiotic but leads to low unit sales of those antibiotics. As low-cost antibiotics remain effective against most infections, the current market for new antibiotics is small. These factors have led to many large companies leaving the market and focusing on other diseases with better returns, restricting the development pipeline further and leading to a reduction in the number of novel antibiotics brought to market (Cooper, M., & Shlaes, D., 2011). This is despite initiatives to encourage private companies to come to market, such as the innovative medicines initiative (IMI) injecting billions of Euros into antibiotic R&D (IMI, 2022).

1.4 Bacteriocins

Bacteriocins are bacterially produced and ribosomally synthesized polypeptides which act to kill closely related species of bacteria to the producer (Nissen-Meyer *et al.*, 1992). All bacteriocins are translated with an immunity protein (to ensure the producing bacteria is not killed by the produced peptide) and an ABC-transporter (to ensure export from the cell) and frequently are translated with an accessory protein (Stephens *et al.*, 1998). The exact function and mechanism of action of this accessory protein is not known but is believed to be involved in exporting the bacteriocin from the cell (Oppegård *et al.* 2007). Additionally, bacteriocins are most commonly synthesized with a leader sequence, which shows similarity within the bacteriocin class. Few bacteriocins, such as Garvicin KS (Dubey *et al.,* 2022) and those produced by *Halobacteriaceae* lack a leader sequence (Riley & Chavan, 2007).

Though the application of bacteriocins in the treatment of AMR pathogens has not yet become a front-line treatment, bacteriocins hold may be a future alternative or complement to traditional, small-molecule antibiotics. Encouragingly, bacteriocins and probiotics have been shown to improve health and mortality in aquaculture against a variety of pathogens, both reducing our reliance on traditional antibiotics and suggesting uses in similar settings (Pereira *et al,* 2022). Bacteriocins can also be modified through protein engineering or interaction with other molecules such as chitosan, nanoparticles or nano-fibers, liposomes and others (Naskar & Kim, 2021). These properties make bacteriocins attractive subjects for medical research.

Multiple bacteriocin classification schemes have been proposed, including those by Heng and Tagg in 2006 (Heng & Tagg, 2006) and Klaenhammer in 1993 (Klaenhammer, 1993). In the following, we use the scheme by Cotter *et al.* in 2003 (Cotter *et al.,* 2003) which currently is the most used.

In Class I the polypeptides contain modified amino acids such as lanthionine, α-methyllanthionine, dehydroalanine, or dehydrobutyrine, this class is thus often referred to as lantibiotics. Nisin belongs to Class I; post-translationally modified bacteriocins. Members of Class Ia tend to be amphiphilic, cationic, and elongated, and act by compromising the selective permeability of the cell membrane and destroying the electrochemical gradient across the membrane (Cotter *et al.,* 2003). As Class Ia bacteriocins target protein and lipid features common to many bacteria, they have broad antimicrobial activity. Class Ib bacteriocins are small globular proteins with a negative or neutral charge that have specific enzymatic targets (Cotter *et al.,* 2003).

Class II bacteriocins do not contain modified amino acids and the class has more members than Class I (Ennahar *et al.*, 2000). The Class II bacteriocins are typically grouped into 4 subclasses: Class IIa pediocin-like bacteriocins, Class IIb the two-peptide bacteriocins, Class IIc cyclic bacteriocins and Class IId linear, non-pediocin-like peptides. Class IIa are defined by their similarity to pediocin, which can be seen in Table 2, by having the N-terminal consensus sequence YGNGV (Ennahar *et al.* 2000) and a β-sheet like N-terminal followed by an α-helix (Oppegård *et al.,* 2007).

Table 2: Sequence alignment of selected Class IIa bacteriocins. The consensus sequence YGNGV is highlighted in yellow and other conserved residues are highlighted in purple.

Bacteriocin	Sequence Alignment	Reference
Lactococcin MMFII	--TSYGNGVHCNKSKCWIDVSELETYKAGTVSNPKDILW----------	(Ferchichi *et al.*, 2001)
Leucocin A	--KYYGNGVHCTKSGCSVNWGEAFSAGV--------HRLANGGNGFW---	(Fregeau *et al.*, 1997)
Enterocin P	ATRSYGNGVYCNNSKCWVNWGEAKENIAGI----VISGWASGLAGMGH--	(Cintas *et al.*, 1997)
Curvacin A	-ARSYGNGVYCNNKKCWVNRGEATQSIIGG----MISGWASGLAGM----	(Haugen *et al.*, 2005)
Pediocin PA-1	--KYYGNGVTCGKHSCSVDWGKATTCIINN----GAMAWATGGHQGNHKC	(Henderson *et al.*, 1992)
Listeriocin 743A	--KSYGNGVQCNKKKCWVDWGSAISTIGNN----SAANWATGGAAGWKS-	(Kalmokoff *et al.*, 2001)
Sakacin P	--KYYGNGVHCGKHSCTVDWGTAIGNIGNN----AAANWATGGNAGWNK-	(Uteng *et al.*, 2003)

A common mechanism of action for Class I and Class II bacteriocins is the destruction of the membrane potential through leakage of ions (Moll *et al.,* 1996). Nisin (Weidemann *et al.* 2001), Pediocin-PA1 (Rodriguez *et al.,* 2002) and other bacteriocins produce the effect by binding to a receptor in the cell membrane and either inhibiting synthesis of cell membrane components, causing lysis, or by forming a pore in the membrane (Kjos *et al.,* 2014). This leakage of ions and small molecules results in the instantaneous termination of cellular processes and cell death.

Unfortunately, the receptors necessary for bacteriocin function have been poorly studied until recently (Cotter *et al.*, 2013). Of those bacteriocins with a known receptor, many belonging to Class I use Lipid II as a receptor, while Class IIa and some Class IId bacteriocins use components of the Mannose Phosphotransferase System (PTS) (Cotter *et al.*, 2013). Other receptors known include Maltose ABC transporters for bacteriocin Garvicin ML and Zn-metallopeptidase for LsbB (Cotter *et al.*, 2013).

Below we will discuss two component bacteriocin PlnS, and the three peptide bacteriocin Garvicin KS, as these are the bacteriocins which became less active when mutations appeared in the C-terminal of YthA.

1.4.1 Class IIb; Two-Component Bacteriocins

Class IIb bacteriocins require two distinct polypeptides to function optimally, though one or both components may display low antimicrobial activity individually (Oppegård *et al.,* 2007). These bacteriocins share several properties with the Class IIa bacteriocins in that they are usually amphiphilic or hydrophobic, and are cationic (Oppegård *et al.* 2007). The genes involved in Class IIb bacteriocin systems are usually found in one or two operons (Cotter *et al.,* 2003). An immunity protein and both bacteriocin peptides are found on one operon, and the ABC transporter protein and an additional accessory protein may be found on a separate operon or the same operon as the other components (Oppegård *et al.* 2007). It should be noted that only a single immunity protein is necessary for the cell to become immune to a Class IIb system.

The two peptides of a two peptide bacteriocin are typically of similar length, while different two peptide bacteriocins have greatly varying length. For instance, Plantaricin S (PlnS) contains two peptides of lengths 27AA and 26AA while Lactoccocin-G (LcnG) 39AA and 35AA respectively. Table 3 shows selected two peptide bacteriocin sequences. The bacteriocins are unstructured when in an aqueous solution, but become α-helical when they interact with micelles, TFE, or DPC (Rogne *et al.,* 2008; Ekblad *et al.,*2016; Ekblad & Kristiansen, 2019).

Table 3: Alignments and sequences of selected class IIb bacteriocins. Sequences with similar sequence are aligned. A vertical line is inserted between a pair of the residues in the same position, a colon is inserted when the atoms are similar residues, and a period indicates a dissimilar residue in that position. GxxxG and GxxxG-like motifs are highlighted in green.

Bacteriocin	Sequence and Alignment	Reference
Plantaricin S α	RNKLAYNMGHYAGKATIFGLAAWALLA--	(Stephens et al., 1998)
	.|:||..:|||.|||.|.|.....:.:	
Muricidin α	YNRLAGQIGHYTGKAVIVGATVLGIASLF	(Ekblad & Kristiansen, 2019)
Plantaricin S β	KKKKQSWYAAAGDAIVSFGEGFLNAW	(Stephens et al., 1998)
	|:...::.. ||:||||:|||:|:	
Muricidin β	-KRGLGYHIV--DAVVSFGKGFLDAF	(Ekblad & Kristiansen, 2019)
Lactococcin-G α	-GTWDDIGQGIGRVAYWVGKALGNLSDVNQASRINRKKKH	(Nissen-Meyer *et al.*, 1992)
	..:..||..:|..|||:.|.|||:||||||.||||||.	
Enterocin 1071 α	ESVFSKIGNAVGPAAYWILKGLGNMSDVNQADRINRKKH-	(Franz *et al.*, 2002)
Lactococcin-G β	KKWG---WLAWVDPAYEFIKGFGKGAIKEGNKDKWKNI	(Nissen-Meyer *et al.*, 1992)
	| ||.|:.|||:|:.|..||..|||||:||||:	
Enterocin 1071 β	---GPGKWLPWLQPAYDFVTGLAKGIGKEGNKNKWKNV	(Franz *et al.*, 2002)
Plantaricin E	FNRGGYNFGKSVRHVVDAIGSVAGIRGILKSIR	(Anderssen *et al.*, 1998)
Plantaricin F	VFHAYSARGVRNNYKSAVGPADWIISAVRGFIHG	(Anderssen *et al.*, 1998)

In transmembrane proteins there is a highly frequent motif: GxxxG, in which two Glycine residues are separated by three amino acids (Senes *et al.,* 2000). The GxxxG motif or GxxxG like motifs, motifs in which the Glycines are replaced by other small amino acids such as Alanine and Serine, are typically found in the interaction surface between transmembrane helices (Senes *et al.*, 2004). The short sidechains allow helices to come close together and make hydrogen bonding possible between backbone residues of the two helices (Senes *et al.,* 2000).

The two peptide bacteriocins do not display a high sequence similarity, but all contain at least one GxxxG or GxxxG-like motif in each peptide as shown in Table 3 (Ekblad & Kristiansen, 2019). The GxxxG motifs in the two peptide bacteriocins are believed to be the interaction motifs between the peptides constituting the bacteriocin and are critical to the bacteriocin function (Oppegård *et al.*, 2008). Ekblad *et al.* performed single amino acid substitutions to several GxxxG and GxxxG-like motifs and observed that in certain motifs the replacements were detrimental to the bacteriocins activity (Ekblad *et al.,* 2016). In PlnS, single amino acid replacements to the G_9xxxG_{13} of PlnSα and $S_{17}xxxG_{21}$ of PlnSβ motifs respectively reduced the antimicrobial activity (Ekblad & Kristiansen, 2019). Single amino acid replacements to the G_5xxxG_9 motif of PlnE and $S_{26}xxxG_{30}xxxG_{34}$ motif of PlnF had a similar result (Ekblad *et al.*, 2016).

The reduced antimicrobial activity is likely due to a reduced ability of the two bacteriocins peptides to dimerize and activate the receptor (Ekblad & Kristiansen, 2019). Similarly, Oppegård *et al.* found that reduced antimicrobial activity was observed when single amino acid replacements were done to $G^{7}xxxG^{11}$ of LcnGα and $G^{18}xxxG^{22}$ of LcnGβ but not in the other GxxxG or GxxxG-like motifs. A model of the PlnS dimer embedded in a membrane is shown in Figure 3; similar models have been published for LcnG (Oppegård *et al.,* 2008) and PlnEF (Ekblad *et al.,* 2016).

PlnS is produced naturally by the gram-positive bacteria *Lactobacillus plantarum*, specifically those stains involved in olive fermentation (Ekblad & Kristiansen, 2019). The system functions by a receptor mediated mode of action and likely due to this property is active in the nanomolar range (Ekblad & Kristiansen, 2019). Another class IIb bacteriocin, Muricidin (Mur), shares ~%40 sequence similarity with PlnS, and displays antimicrobial behaviour when a Mur peptide is present with the complementary peptide of PlnS. Single chains of neither Mur nor PlnS display such behaviour when the complementary chain is absent (Ekblad & Kristiansen, 2019). This cross-compatibility suggests that the bacteriocins descended from a recent common ancestral gene, as well as share a similar structure and mechanism of action.

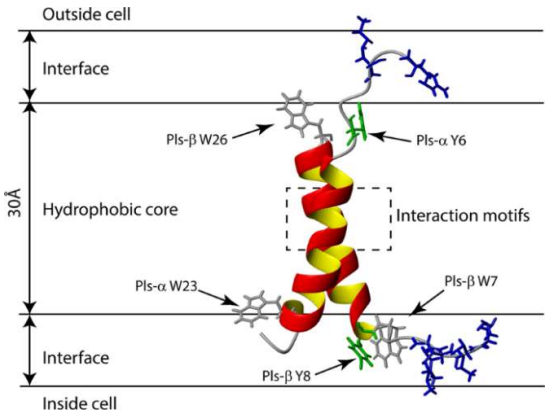


Figure 3: Proposed model of PlnS α and β chains in the membrane of a target cell. Note the antiparallel orientation of the dimer. Tryptophan residues are shown in grey, Tyrosine residues are shown in green, and positively-charged residues are shown in blue. Figure courtesy of Ekblad & Kristiansen, 2019.

1.4.2 Unmodified, Unclassified Bacteriocins

Not all bacteriocins can be classified by Cotter's scheme. Garvicin KS (GarKS) for example is a three component bacteriocin that requires all three peptides in relatively equal concentrations to function effectively (Dubey *et al.,* 2022). This bacteriocin has been placed in a 'multi-peptide, leaderless bacteriocin group' to classify it in some way (Dubey *et al.*, 2022). The peptides composing GarKS show strong sequence similarity to each other, but not to other multi-peptide bacteriocins (Table 4).

Table 4: Sequence alignment of Garvicin KS, a three-peptide leaderless bacteriocin. Perfectly conserved residues are highlighted in purple.

Bacteriocin	Sequence Alignment	Reference
Garvicin KS A	MGAIIKAGAKIVGKGVLGGGASWLGWNVGEKIWK	(Ovchinnikov *et al.*, 2016)
Garvicin KS B	MGAIIKAGAKIIGKGLLGGAAGGATYGGLKKIFG	(Ovchinnikov *et al.*, 2016)
Garvicin KS C	MGAIIKAGAKIVGKGALTGGGVWLAEKLFGGK--	(Ovchinnikov *et al.*, 2016)

GarKS is a particularly promising bacteriocin for use in the food and health sectors as it has been shown to have a wide spectrum of activity, being active against the CDC and WHO high priority bacteria: methicillin resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *enterococci* (VRE), Penicillin-non-susceptible *Streptococcus* (Table 1) (Ovchinnikov *et al.,* 2016; Dubey *et al.,* 2022), and other pathogenic bacteria. GarKS inhibition against various pathogens is shown in Figure 4. Additionally, GarKS displays synergy with other bacteriocins, allowing it to affect several gram-positive species (Kranjec *et al.,* 2021). Furthermore, the production of GarKS has already been optimized and it is possible to produce as much as 1.2g/L, an amount among the highest production per litre known for any bacteriocin (Telke *et al.*, 2019).

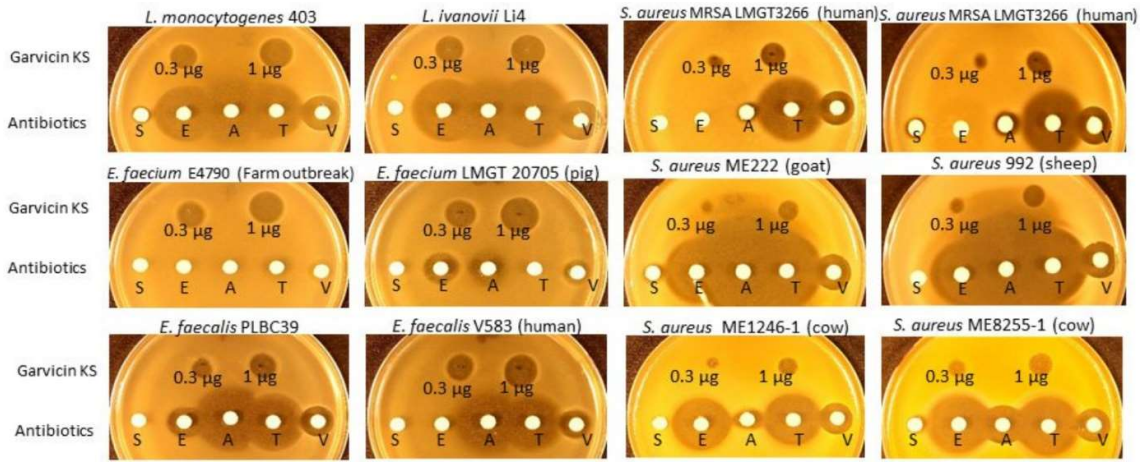


Figure 4: Susceptibility assay of Listeria monocytogenes, L. ivanovii, Staphylococcus aureus, Enterococcus faecium, *and* E. faecalis *to 0.3µg and 1µg GarKS and selected antibiotics: Streptomycin (S; 5µg), erythromycin (E; 5 µg); ampicillin (A, 25 µg); tetracycline (T, 5 µg) and vancomycin (V; 5 µg). Clinical strains are indicated in parentheses. Figure courtesy of Dzung B Diep (unpublished data).*

1.5 Bacteriocin receptor YthA

Prior to commercial use of a bacteriocin it is important to understand how the bacteriocins function and the potential for resistance development.

Receptor-mediated functionality is a common trait in bacteriocins, these receptors are usually distinct from those targeted by traditional antibiotics (Cotter *et al.*, 2013), and so cross-resistance is unlikely to occur. Resistance to bacteriocins often occurs due to mutations in the receptors to which bacteriocins bind, so understanding these receptors and mutations remain an important step in utilizing bacteriocins to their full potential.

To determine the receptor of PlnS, susceptible cells were treated with the bacteriocin at a level below the lethal concentration to develop resistance and the resistant mutants were sequenced to find the mutations responsible for the increased resistance. Similar experiments were performed on GarKS (Per Eugen Kristiansen, unpublished data), LacG (Kjos *et al.*, 2014), enterocin K1 (EntK1), enterocin EJ97 (EntEJ97) (Ovchinnikov *et al.*, 2017), among others, and the receptors of these bacteriocins were determined in this way. In both cases it was observed that the only mutations leading to increased resistance were in the C-terminal of the *YthA* gene, specifically the insertion of a stop codon into the C-terminal portion of *YthA* of the DNA sequence (Per Eugen Kristiansen, unpublished data).

Genes in the Carboxylesterase (Ces) family are increased in Δ*YthA* mutants. *Ces* genes are known to, among other effects, increase the expression of various proteins involved in the stabilization and synthesis of the cell wall (Rodriguez *et al*., 2002).

While several models have been proposed to explain how bacteriocins interact with YthA, the mechanism and exact nature of this interaction has yet to be determined experimentally. *YthA* is found in the *Yth* operon, together with genes *YthB* and *YthC*. At the time of writing, no structures are available for YthA, despite its importance in susceptibility to bacteriocins. However, little information has otherwise been elucidated regarding YthA, including its mechanism of action, despite its reception of multiple bacteriocins and presence in many gram-positive bacteria (Huvet *et al.*, 2011).

Furthermore, the amino acid sequence of YthA has high sequence similarity/homology in the N-terminal to the Phage Shock Protein C (PspC). Phage Shock Proteins are part of the bacterial virus protection system, reinforce the cell wall, and switch the cell to anaerobic respiration in response to various stresses. The levels of PspA, PspB, and PspC increase drastically in response to extreme stress, triggering a cascade (Brissette *et al.*, 1990) that reinforces the cell wall and induces more Psp's to be produced. PspD and PspG are also effectors in the system and act to subtly affect the redox state of the cell (Jovanovic *et al.*, 2006).

PspB and PspC are membrane bound proteins that have regulatory and effector roles within the Psp system (Maxson & Darwin, 2006). PspA is bound to PspF, inactivating both proteins. During acid, thermal, and other stress however, the PspAF complex disassociates and PspA begins reinforcing the cell wall while PspF induces the *Pspabc* operon. This results in more PspA being produced and reinforcing the cell wall and potentially preventing cell lysis (Maxson & Darwin, 2006), as shown in Figure 5. The Yth system has been characterized as holding a similar role to the Psp system (Wu *et al.,* 2018).

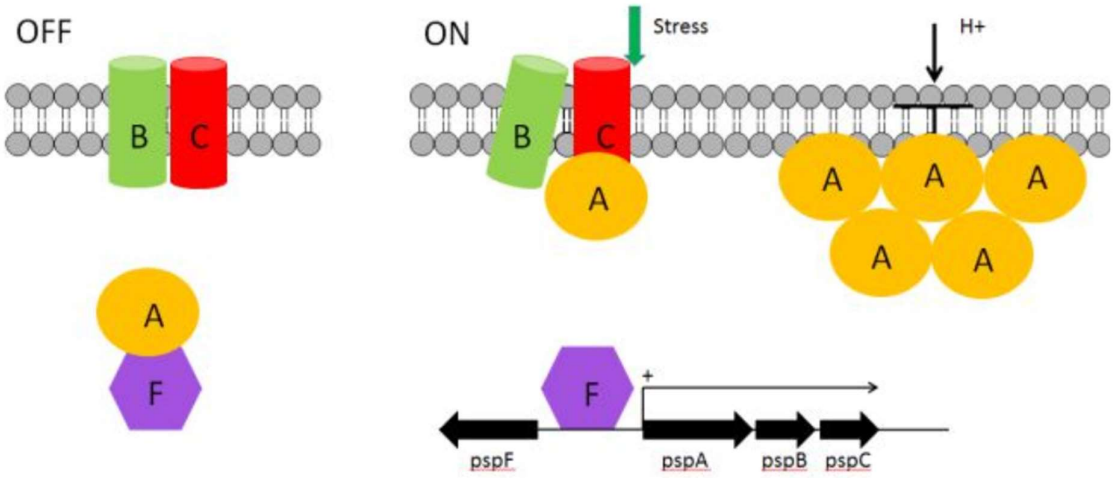


Figure 5: Psp system in the off and on states. In the off state, PspA and PspF are bound together and inactivate each other. When stress occurs PspB and PspC are activated and disassociate PspA from PspF. PspA reinforces the cell wall by binding to it and other PspA subunits, and PspF activates the psp *operon. Figure courtesy of Per Eugen Kristiansen.*

One model as to the mechanism and function of YthA is based off the Lia system. In the Lia system (shown in Figure 6) LiaS, LiaF and LiaI are membrane-bound proteins, and LiaR and LiaH are cytosolic proteins. Under non-stress conditions the C-terminal of LiaF blocks the active site of LiaS, preventing it from phosphorylating LiaR. Under inducing conditions, the C-terminal of LiaF changes conformation and allows LiaS to phosphorylate LiaR, in turn LiaR forms a homo-tetramer and binds to the *liaIH* operon (Wolf *et al.* ,2010). This induces production of LiaI and LiaH. LiaI acts as a membrane anchor for LiaH which is believed to reinforce the cell membrane (Domínguez-Escobar *et al.*, 2014).

However, the sequence of YthA shares a higher degree of similarity to PspC than to any component of the Lia system. A model of the Yth system is shown in Figure 7. YthA is 154AA long with an isoelectric point of 6.62, has a mass of 18348.7 Daltons, and is hydrophobic.

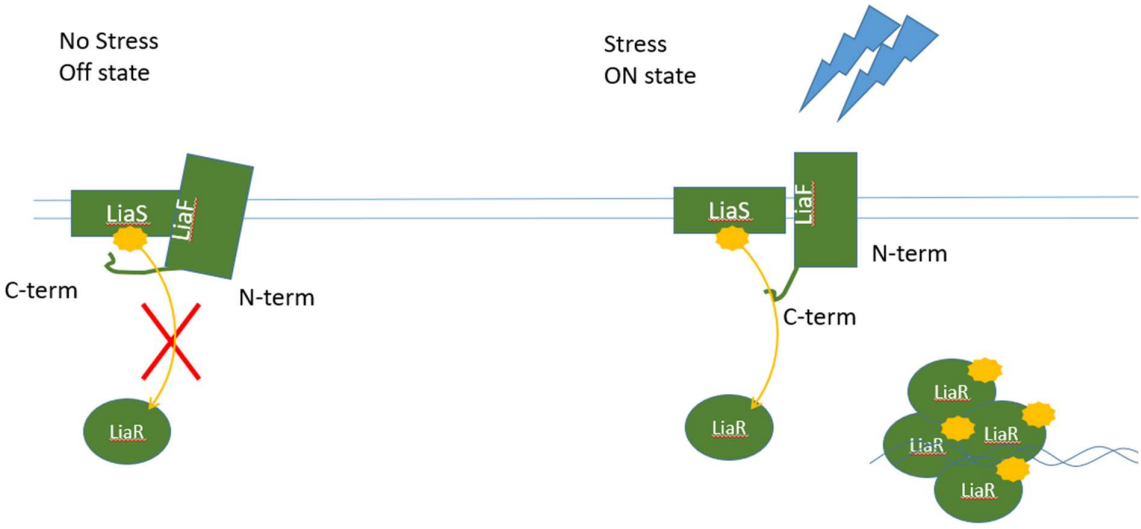


Figure 6: Model of the mechanism of the LiaSFR components of the Lia system. Under non-inducing conditions, the C-terminal of LiaF blocks LiaR from being phosphorylating by LiaS. Under stress conditions the C-terminal of LiaF changes conformation and allows LiaS to phosphorylate LiaR. LiaR then forms a tetramer and binds to the liaIH *operon and induces production of the LiaI and LiaH proteins. Figure courtesy of Per Eugen Kristiansen.*

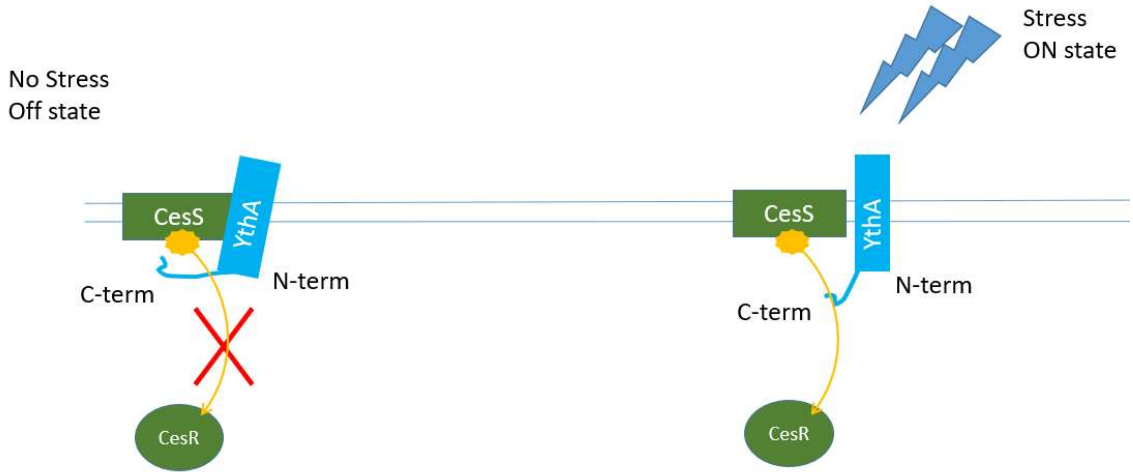


Figure 7: Proposed model of the function of YthA. In the model, YthA rests in the membrane and blocks another membrane bound protein, CesS, from phosphorylating CesR under normal conditions. Under stress conditions YthA changes conformation and CesS can activate CesR. Figure courtesy of Per Eugen Kristiansen.

1.6 Objective

Problematically, small membrane proteins, such as YthA, are grossly underrepresented in the Protein Data Bank (PDB), having 64 entries as of April 5th, 2022 compared to the nearly 200 000 total structures. Thus, the structure of YthA may contribute to our overall knowledge as to how these proteins function. The N-terminal is also of interest due to its sequence similarity to PspC, which may further aid in the elucidation of the YthA function and importance, as well as understanding these small proteins' structure and function.

YthA, and proteins in general, can be produced chemical protein synthesis by chemically building the protein chain. An amino acid with a blocked C-terminal is reacted to the C-terminal of the previous amino acid, the remainder of the unreacted amino acid is washed away, and the chemical blockage removed before the process is repeated with the next amino acid in the protein chain. This process allows for potentially toxic or easily degraded proteins to be produced and purified without the need for a cell. However, the cost for this method makes it restrictive compared to the yield and larger proteins are more prone to errors in the sequence compared to production using cells. Few vendors will accept orders for synthetic proteins above 50 amino acids. Instead, proteins are more commonly produced in a cellular system, but then require purification to acquire pure samples. Due to the other proteins produced in cellular protein production systems and the extremely high cost of labelled amino acids, the system must first be optimized for YthA production before labelling can occur to be practical.

Here we describe an attempt to produce and purify the protein at least the 0.5µmol of YthA necessary in order to obtain structural information by CD spectroscopy and NMR spectroscopy with an aim to identify active sites, domains, and structural motifs to allow further investigation into its role in the Yth system and immune response. Since it is believed that the N-terminal of YthA is involved in bacteriocin interactions, while the C-terminal of YthA is involved in regulating the activity of CesS and CesR. These reasons make YthA of the highest interest in our work which may provide information as to YthA's function.

We first modelled and analyzed YthA using structure prediction tools and physical property prediction tools to aid in the construct designs and in experimental design. We investigated various protein tags and host cells under several conditions to maximize the experimental yield of the protein in preparation of the analysis, attempted to re-solubilize the tagged protein, and to purify YthA. All this was done to prepare the protein for N^{15}- and C^{13}-labelling for use in NMR structural analysis. This is necessary to determine the structure of the protein in this fashion as only these isotopes of nitrogen and carbon can be detected by NMR and are too scarce in a natural sample to be effectively detected.

2 Results and Discussion

Prior to the start of this thesis work, attempts were made to produce a construct producing a His-tagged, full-length YthA sequence in BL21(DE3) *E. coli* cells. These attempts were unsuccessful as the cell growth ceased immediately after induction and little recombinant protein was produced. Toxicity of YthA to the cell was the likely cause of the production failure (Kristiansen, P.,E., personal communication, October 30, 2018).

2.1 Scheme Development for the Bacterial Protein Production of YthA

In this thesis we worked to produce sufficient protein for the structure determination of YthA by solution NMR. Figure 8 outlines a typical scheme for the production of purified proteins in bacteria (Ferrer-Miralles *et al.,* 2022). For bacteria to produce a protein, the gene coding for the protein must be present in the cell, such as by the inclusion of a construct containing the protein's code. One also needs to be able to select for bacterial cells producing the product, if the gene coding for the protein is present in a vector this is typically done by including additional genes providing resistance to certain antibiotics, thus enabling selection by the addition of those antibiotics. Making the production inducible rather than constitutive allows for the rapid synthesis of large amounts of protein (Figure 8, Step D) without slowing cell growth prior to induction. Production tags can be used to further increase protein yields, the selection of which specific tags were selected for use in this thesis are discussed in detail in sections 2.4 and 2.5.

High performance liquid chromatography (HPLC) methods were to be used to purify YthA (Figure 8, Step E) both from other proteins natively present in the cell, and then from the production and purification tags produced with YthA and any proteases necessary to cleave YthA from the tags. Several purification tags have been previously described and characterized including maltose binding protein (MBP) tags (Raran-Kurussi & Waugh, 2016), poly-lysine tags, poly-histidine tags (His-tags) and similar repeating residue tags (Wijekoon *et al.,* 2016). The success of His-tags previously in the thesis laboratory and access to the equipment needed for His-tag purification made it the natural choice.

Initially, the experiment required us to design one or more vector constructs which would produce YthA in a non-toxic state, efficiently produce YthA or the YthA fusion protein, and allow for efficient purification of YthA from other cellular proteins and any tags used to produce YthA (Figure 8, Step A).

After designing the constructs, we would need to have them manufactured and used GenScript Biotech Corp's plasmid preparation service (Research-Grade) for this purpose (Figure 8, Step B). Following the manufacture of the constructs and transformation of the cells, protein production would need to be optimized to ensure that sufficient amounts of YthA were produced with the minimum amount of media due to the expense of labelled media. Protein NMR structure determination experiments typically require 0.5mM-1mM protein concentration and a sample volume of 0.550mL (Teng, 2005), making our goal to reliably produce that much labelled YthA in a minimal volume of media.

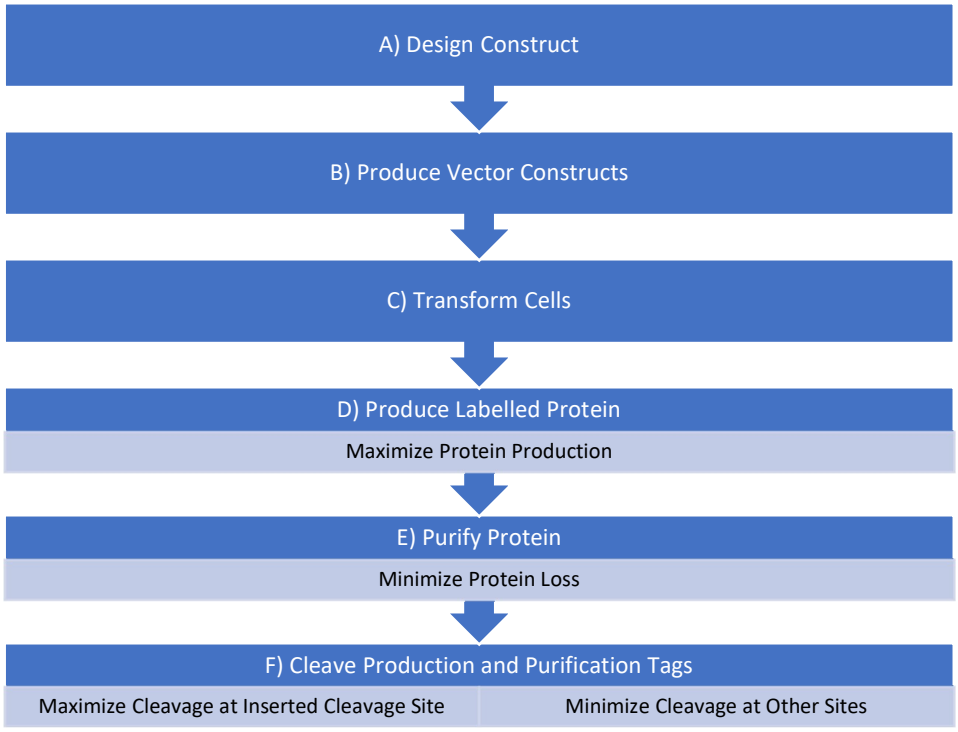


Figure 8: YthA structure determination project workflow. Subordinate points are important goals within the step.

2.2 *In silico* YthA Modelling and Sequence Analysis to Gain Structural Data and Determine Optimal YthA Splitting Position

Beginning with the design of YthA constructs (Figure 8, Step A) splitting YthA into two fragments and producing each fragment separately disrupts the ability of YthA to harm the producing cell. The splitting strategy should perform well as multi-domain proteins can usually be studied domain-by-domain (Card & Gardner, 2005). The decision to produce two YthA fragments covering the entire full-length protein required us to determine the best position to 'split' YthA. Maximizing fidelity of each structural element to the same element in the full-length YthA protein was critical to the success of the NMR studies to follow, and so was the primary concern on the position of the split.

Splitting the protein in the middle of a structural element could disrupt the structures normally present and reduce the accuracy of YthA's determined structure. As membrane and cytosolic fragments are expected to have minimal impact on each other's structures, and we expected that splitting the protein near the membrane interface would not affect the structures obtained of either the membrane or soluble domains.

PspC is a protein with high homology to YthA (As discussed in section 1.5) and a similar peptide sequence, with the exception that YthA has an elongated C-terminal (Figure 9). Previous studies have determined that the C-terminal end of PspC is cytosolic and loosely structured (Flores-Kim & Darwin, 2012).

```
YthA   1 MSQRQLTKSV----TNRRVSGVIAGIAEYFGLGRDV-VTILRILFVV-LA     44
           |:...|.|.:    ....|.||.||||.||    || |.::|||.|: :.
PspC   1 MAGINLNKKLWRIPQQGMVRGVCAGIANYF----DVPVKLVRILVVLSIF     46


YthA  45 FGSWGGLIPLYFVASW/IIPSARPRNYYDDSEDD--YQEKWNRKAQHFDE-    91
           ||     :.|:.:.:: ||.|..    .|...|:  :.|:....::..||
PspC  47 FG-----LALFTLVAY IILSFA----LDPMPDNMAFGEQLPSSSELLDEV    87


YthA  92 ------------KMDRWSERYSDKMNNWARRYEDKGRQNQQDSNQWGNPW    129
                       :|:|:....:..:.:..|:.
PspC  88 DRELAASETRLREMERYVTSDTFTLRSRFRQL------------------    119


YthA 130 DEPKSRKTKEAQPVEKEKEDDWSDF    154

PspC 120 -------------------------    119
```

Figure 9: Alignment of proteins YthA and PspC. Indicated are identical (|), similar (:), and dissimilar (.) residues and gaps ( ) at each position.

In an attempt to determine the optimal position to divide YthA into an N-terminal fragment and a C-terminal fragment, we used TOPCONS to predict which regions of YthA and related protein PspC form a membrane domain. Five sub-methods (Philius, SCAMPI-multi, OCTOPUS, SPOCTOPUS, and PolyPhobius) are used by TOPCONS to predict the membrane topology of a given sequence. TOPCONS then correlates the results of these sub-methods into a single membrane topology model (Tsirigos *et al.*, 2015). Results of the TOPCONS analysis of YthA, shown in Figure 10, predicted one internal, one external, and one inter-membranal region. The membrane helix of YthA is predicted to start at R37 and end at V57, cytosolic to the membrane helix is predicted to be the C-terminal ending at F154 while N-terminal to the helix, residues M1 to L36 would be extracellular.

```
MSQRQLTKSVTNRRVSGVIAGIAEYFGLGRDVVTILRILFVVLAFGSWGGLIPLYFVASWIIPSAR
oooooooooooooooooooooooooooooooooooooMMMMMMMMMMMMMMMMMMMMiiiiiiiii

PRNYYDDSEDDYQEKWNRKAQHFDEKMDRWSERYSDKMNNWARRYEDKGRQNQQDSNQWGNPWDEP
iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii

KSRKTKEAQPVEKEKEDDWSDF
iiiiiiiiiiiiiiiiiiiii
```

Figure 10: TOPCONS consensus prediction of YthA topology. Residue predicted to be extracellular are labeled o, cystolic i, and inter membrane M.

TOPCONS analysis of PspC had similar results to the YthA analysis, though the direction of the peptide through the membrane was inverted: an internal region of PspC is predicted to exist from peptides M1 through approximately R38, at which point a membrane helix is predicted to occur until approximately peptide L60, after which an extra-cellular region is predicted until the end of the protein. The TOPCONS analysis of PspC is incorrect as previous studies have shown that PspC is polytopic with cystolic C- and N-termini and two membrane helices (Flores-Kim & Darwin, 2012). Additionally, positively charged regions of transmembrane proteins are typically localized cytoplasmically and negatively charged regions are typically localized externally to the cell (Baker *et al.,* 2017), and is inverted from the homologous PspC protein.

In an attempt to gain more structural information about YthA, a SWISS-MODEL (Waterhouse *et al.,* 2018) structure prediction of YthA and PspC were performed to suggest potential protein conformations and membrane orientation using a database of experimentally determined protein structures (Bienert *et al.,* 2017). Models with the highest general model quality estimate (GMQE) were chosen for further consideration. A higher GMQE is a better prediction according to the algorithm, with a score of 1 being the highest possible score. The highest GMQE of the PspC models was Autophagy protein 16, with a GMQE of only 0.08, a sequence similarity score of 0.31 and a coverage of 0.25. Comparatively, the most reliable model of YthA was slightly better with a GMQE of

0.09, sequence similarity score of 0.27 and coverage of 0.33 assigned to the model of a fragment of the Dopamine Transporter (Biasini *et al., 2014)*. The low GMQE scores of the YthA and PspC models indicated that the models were unreliable and likely inconsistent with the actual structures of both YthA and PspC, so the SWISS-MODEL models were discarded from further consideration.

Based on the limited information that was obtained, it was decided to split the protein close to the interface between cytosolic and membrane interface as determined by TOPCONS, more specifically between W60 and I61. Thus two fragments were devised, a N-terminal fragment containing the putative external region and membrane helix, and a C-terminal fragment containing the predicted internal region. Any structures in the fragments should represent those in the complete YthA protein, as the membrane interface should prevent the cytosolic, periplasmic and membranal structural elements from affecting each other.

2.3 Design of Vectors for the Production of YthA-Fragment Fusion Proteins

In addition to toxicity and structural considerations, the exposed regions of the C-terminal of YthA may be cleaved by native proteases during or after protein production due to its loose structure (Koenig *et al.,* 2003), reducing the amount of labeled YthA produced. Bacterial peptide production of small peptides typically is largely reduced by enzymatic degradation, due to their natural tendency to be unstructured in the cytosol of bacteria. However, producing them fused to a fusion tag has repeatedly been shown to yield non-degraded peptides (Fimland *et al.*, 2008; Rogne *et al.,* 2008; Cheng & Patel, 2004; Koenig *et al.,* 2003).

Once expressed, the fusion proteins need to be purified. To simplify purification, all fusion proteins were produced with a purification tag, specifically a His-tag, to allow purification by nickel-nitrilotriacetic acid (Ni-NTA) column affinity. The inclusion of a His-tag was also necessary as Ni-NTA ion affinity HPLC was the chosen method of purification due to the high purity and potential yields of the technique (Walker *et al.,* 2019). All affinity purification tags are known to bind to a specific substrate. This allows a polypeptide fused to an affinity tag to bind to an immobilized substrate without binding contaminants. Which tag is the best choice is determined based on the volume and purity requirements of the purified protein, as there is generally an inverse correlation between purity and yield among highly used affinity tags (Lichty *et al.,* 2005). Protein yields have been shown to decrease as much as 6-fold depending on the length and position of the His-tag (Mohanty & Weiner, 2004), so care must be taken when integrating a His-tag coding sequence into a construct design.

To be able to remove the tags from the YthA fragment, cleavage sites were introduced. Factor Xa was the preferred choice of protease as it cleaves at the end of its I(E/D)GR recognition site, preventing undesirable amino acids from remaining attached to the N-terminal end of YthA fragment, though it's shorter recognition site reduces cleavage specificity (Bronsoms & Trejo, 2015). Tobacco etch virus protease (TEVp) was also chosen as a second protease with a unique recognition site was needed, and due to its efficient cleavage at its recognition site ENLYFQG/S (Tropea *et al.,* 2009). TEVp remains active across a range of temperatures, pH values, and concentrations (Tropea *et al.,* 2009), making it a versatile protease for removing protein tags, and inactivating or activating engineered proteins.

2.4 GB1–YthA Fragment Constructs and Fusion Proteins

To combat the degradation problem, various protein tags were considered to add structure to the overall fusion protein and prevent degradation. GB1 was considered and accepted as the protein tag is known to be highly soluble; causing as much as 90% of the recombinant protein to be found in the soluble phase; and produce more than 20mg of recombinant protein per litre of media (Cheng & Patel, 2004). GB1 was an obvious choice for the high yield tag as other researchers in our lab have successfully used GB1 to produce several labelled peptides, including LacG (Rogne *et al.,* 2008) and PlnEF (Fimland *et al.,* 2008). A GB1 protein tag was thus the first tag to be tested with the new shortened constructs.

GB1- and His-tags are sufficiently small to design an insert containing the tag. The construct coding for the full-length fusion protein containing a GB1 coding sequence and a stop codon was inserted into a pET22(+) vector between the NdeI and NcoI restriction sites. As conditions may prevent either TEV or Factor Xa from functioning, both ENLYFQS and IEGR sites were inserted into the fusion proteins to maximize the chances of at least one protease cleaving the fusion protein efficiently. pET22(+) was chosen because it has previously been used successfully in our lab for several projects. Vector maps of pET22(+) and the other vectors used in this thesis work are available in Appendix 6.4. The map of the GB1-constructs and selected properties is given in Figure 11.

GB1[N]; N-terminal YthA fragment tagged with GB1
His tag – Linker – GB1 – TEV cut site – Factor Xa cut site – N-terminal YthA Fragment
pI=8.04 Mw= 20356.81Da

GB1[C]; C-terminal YthA fragment tagged with GB1
His tag – Linker – GB1 – TEV cut site – Factor Xa cut site – C-terminal YthA Fragment
pI=5.88 Mw=25461.49Da

Figure 11: Designations, descriptions, maps, isoelectric points, and molecular weights of YthA fusion proteins tagged with GB1.

The vectors containing the YthA constructs were produced by Genscript Biotech Corp (Figure 8; Step B).

2.4.1 GB1 Fusion Protein Production

Congruent to step C in Figure 8, the GB1[C] and GB1[N] expression vectors were transformed into BL21(DE3) *E. coli* cells prior to cell plating, colony picking and testing growth and inducible protein production.

Protein production attempts using the GB1[C] construct caused rapid cell growth (Figure 12) and succeeded in producing recombinant protein but resulted in modest yields (Figure 13). GB1[C] was observed on SDS-page with an apparently greater mass than expected for its actual molecular weight (MW), however, this has been observed in previous studies where the 10kDa protein appeared to be approximately 14kDa in SDS analysis (Hartl *et al.,* 2010). Very low amounts of the protein was present in the cytosol as most was in the non-soluble fraction under both 18°C and 37°C induction conditions. This despite that both the His-tag and cytosolic YthA fragment were predicted to be soluble. As GB1[C] was found to not be soluble, we believed that the fusion protein formed inclusion bodies (IBs). We based our general IB purification scheme on IB purification schemes in Ferrer-Miralles *et al.,* 2022 (Figure 14).

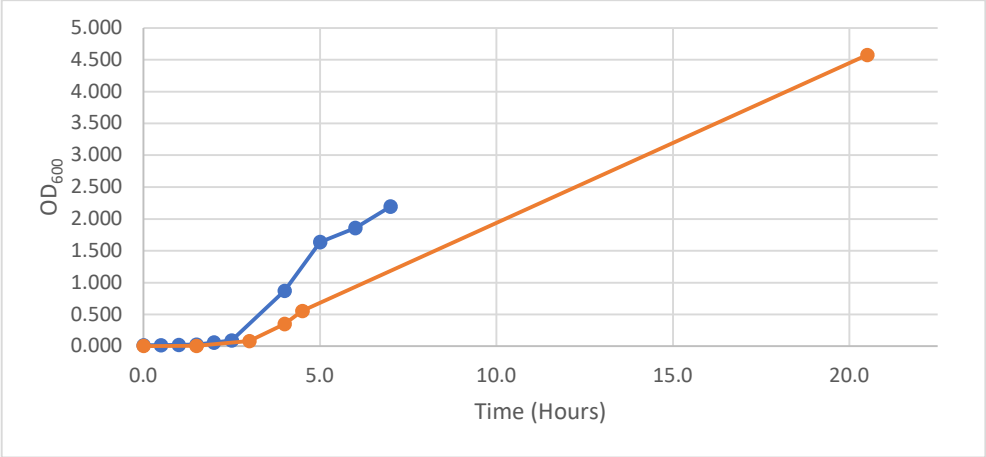


Figure 12: Growth Curve of GB1[C] producing cells induced at 37°C (Blue Line) and 18°C (Orange Line).

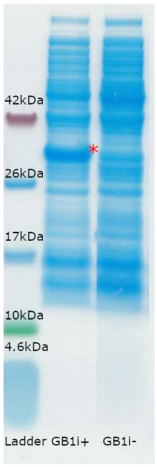


Figure 13: SDS-Page of induced (i+) and uninduced (i-) GB1[C]-producing cells. A protein ladder is included for size comparison. Red asterisks indicate GB1[C] fusion protein. SDS-page edited for clarity; unedited SDS-page in ***Appendix 6.6.1.***

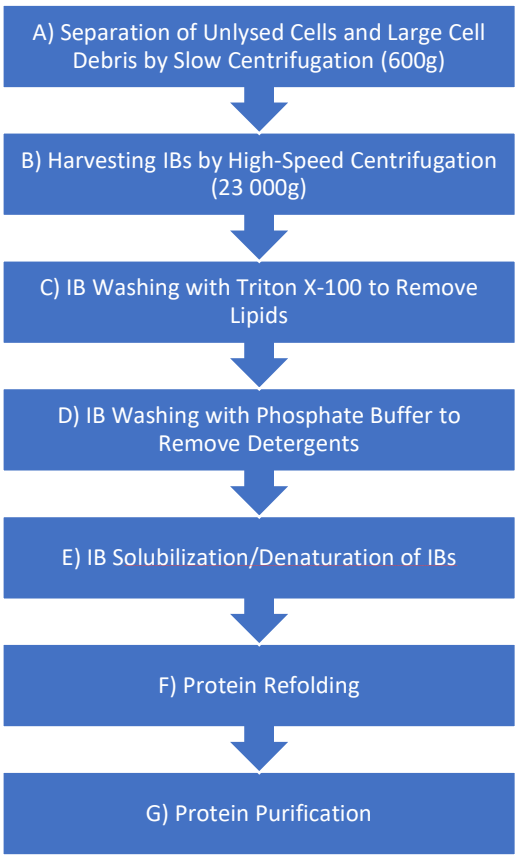


Figure 14: Inclusion body purification scheme based on a scheme from Ferrer-Miralles et al.*, 2022. Optimization of each step is required for each protein.*

Following the removal of large cell debris and unlysed cells in the lysate by low-speed centrifugation (600g) (Figure 14, Step A), IBs were harvested by high-speed centrifugation (Figure 14, Step B). Then membrane fragments and lipids were removed by washing the harvested IBs with detergent Triton X-100 (Figure 14, Step C) and the detergent removed by washing the IBs without detergent (Figure 14, Step D). The remainder of the non-soluble fraction was confirmed by SDS-page to be insoluble proteins with yields typically less than 3mg protein per litre of LB media, even following optimization of induction temperature. Further, great difficulty was encountered solubilizing the protein even in 8M urea.

2.4.2 Testing Urea and Trifluoroethanol to Maximize GB1[N] Solubility

GB1[N] fusion protein was also observed to form IBs and in order to continue using GB1[N] in the project, solubilization should be maximized. Typically, an Ni-NTA column is used to purify proteins of interest which form IBs by designing a fusion protein containing a His-tag and the protein of interest (Ortega *et al.,* 2022). However, inclusion bodies must first be solubilized before protein purification can take place (Figure 14, Step E). Recent studies have shown that IBs contain a significant amount of protein with a native-fold or like-native-fold (Upadhyay *et al.,* 2016). Two commonly used solubilizing agents, TFE and urea, are known to solubilize IBs and have previously shown success in the solubilization for purification of several proteins in non-denaturing conditions. They were thus tested to determine the conditions necessary for the solubilization of GB1[N]. Table 5 summarizes the results of the solubility test and Figure 15 is an image of the stained SDS gel.

Table 5: Conditions and results of solubilization test on GB1[N] fusion protein pellet.

TFE Content (%)	Urea Content (M)	Dilution Factor	Amount of Protein Solubilized
30	1.5	1	None detected
30	3	10	None detected
30	3	1	Very low amounts of protein detected
0	2	1	None detected
0	4	1	Low amounts of protein detected
0	6	10	Minimal protein detected, almost nondetectable
0	6	1	Moderate amounts of protein detected
0	8	10	Moderate amounts of protein detected
0	8	1	Very high amounts of protein detected

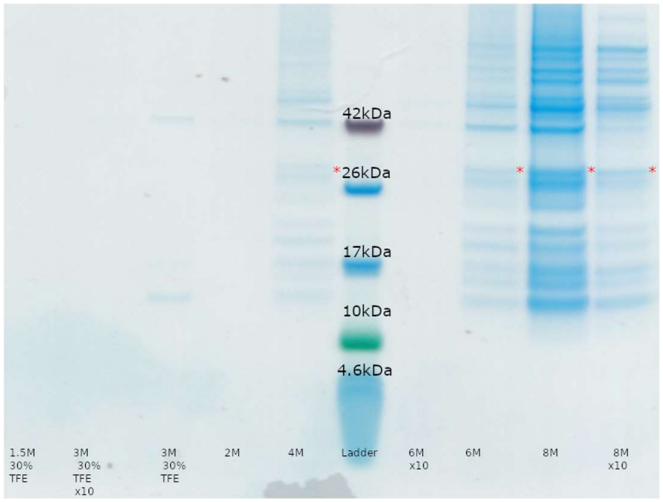


Figure 15: SDS-page of solubility tests of GB1[N]-containing protein pellets in 1.5M, 2M, 3M, 4M, 6M or 8M of urea, 0% or 30% TFE (v/v), and undiluted or 10-fold diluted. A protein ladder is included for size comparison. Red asterisks indicate GB1[N] fusion protein.

TFE and low urea concentrations had no apparent effect on the solubility of GB1[N], indicating there was little natively folded protein in the IBs that could be recovered with mild solubilizing conditions. However, protein solubility increases noticeably as urea concentration increased to the maximum tested concentration of 8M, indicating that strong denaturant is required for the solubilization of GB1[N].

2.4.3 Solubilizing GB1[C] with Guanidine Hydrochloride (GHCl)

Due to the difficulty dissolving GB1[C] during handling in an 8M urea solution, other solubilizing agents were considered (Figure 8, Step E; Figure 14, Step G). As GHCl has been used extensively for protein solubilization, measuring protein stability and is considered to be a stronger denaturant than urea (Smith & Scholtz, 1996) it was chosen. The concentration of GHCl used was 6M, the highest concentration tolerated by HisTrap purification columns.

On-column refolding was initially considered for the refolding step (Figure 14, Step F) as protein losses from refolding were expected to be minimal while the fusion proteins remain bound to the Ni-NTA column matrix. By binding the denatured fusion protein to the column then slowly reducing the concentration of the denaturant in the solution, the fusion protein would adopt a native conformation while remaining bound to the Ni-NTA column (Figure 14, Step F), and thus having limited ability to interact with other proteins to form precipitates (Ortega *et al.,* 2022). An elution step could then be performed to separate the fusion protein from other proteins (Figure 8, Step E; Figure 14, Step G).

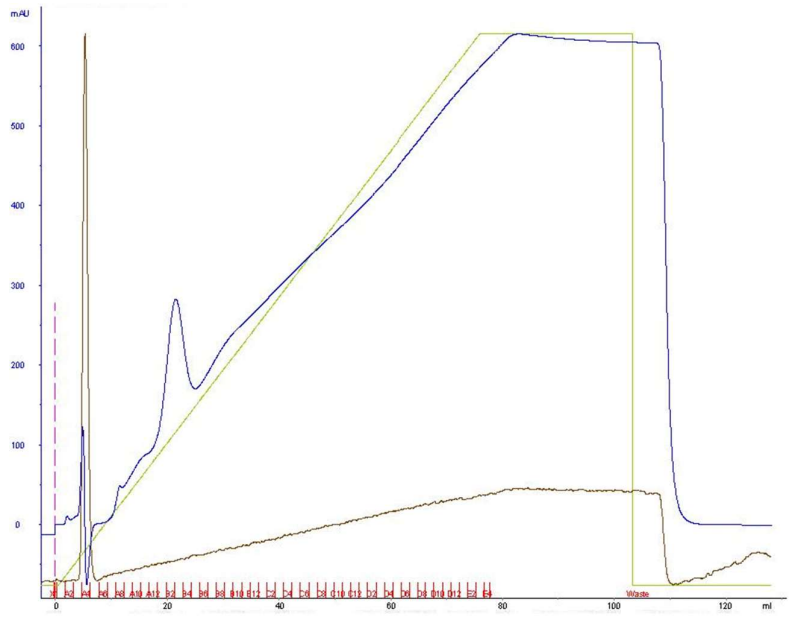


Figure 16: Elution chromatograph of GB1[C] fusion protein purification attempt. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters).

The loading (Appendix 6.5.1) and refolding (Appendix 6.5.2) chromatographs showed no unusual features. The elution chromatograph had a peak stretching between fractions A12 and B4, believed to be GB1[C] (Figure 16). SDS-page analysis of the fractions showed no detectable protein (Appendix 6.6.2). The SDS-page result was not surprising considering that the peak displayed in the elution chromatograph is approximately 100 mAU tall, further indicating that there was little protein bound to the column. Poor yields, IB formation and difficulty purifying the product led us to abandon production of the C-terminal fragment of YthA using this construct after several repeats.

2.4.4 Refolding and Purification Attempts on 8M Urea-Solubilized GB1[N]

On-column refolding (Figure 14, Step F) was performed by applying solubilized fusion protein to the ion affinity column, reversibly binding to it. The buffer was changed to 8M urea and a 100%-0% gradient was applied, ideally causing the bound fusion protein to refold into a native conformation. Lastly the protein was eluted by running a 5mM-500mM imidazole gradient, eluting the recombinant protein in a fraction, which is retained, separate from any other bound proteins (Figure 8, Step E; Figure 14, Step G). Low-imidazole (5mM imidazole) buffers were used in the purification attempts involving GB1[N] as comparisons of 20mM imidazole and 5mM imidazole had no noticeable effect on purification.

Initial purification attempts appeared promising as a peak was present in fractions C4 through C8 (Figure 17). An SDS-Page confirmed the presence of relatively pure protein in the fractions (Figure 18).

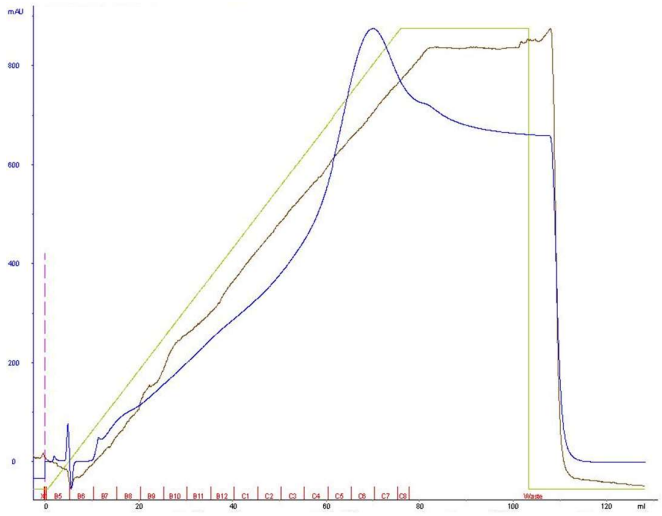


Figure 17: Elution chromatograph of GB1[N] fusion protein purification attempt using low-imidazole refolding buffer. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole, (green line), sample injection (dotted line) and sample fractions (red letters).

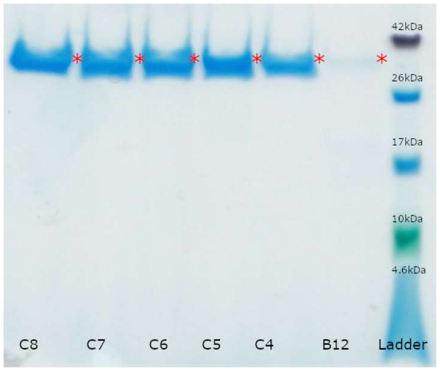


Figure 18: SDS-page of elution fractions B12, C4, C5, C6, C7, and C8 from the purification of GB1[N]. A protein ladder is included for size comparison. Red asterisks indicate GB1[N].

However, during an attempt to make a phosphate-based digestion buffer, white insoluble crystals formed in solution. These crystals were consistent with calcium phosphate formed by a reaction between the calcium chloride present in the solution and the phosphate buffer the CaCl was to be dissolved in. Because Factor Xa cleavage protocols call for $CaCl_2$ to be present in the reaction solution, $CaCl_2$ could not be omitted from the reaction buffer. This required the buffering agent of the solution to be changed. Tris(hydroxymethyl)aminomethane (Tris) buffers have been successfully prepared with calcium chloride without chemical reactions (Kamaraj & Pasupathi, 2021), making it an attractive option. The buffer exchange to a tris buffer was performed on the purified GB1[N] sample using anion exchange HPLC, and all further experiments involving buffers used a tris buffer (Appendix 6.5.3).

Prior to the buffer exchange, fractions B12, and E4-E8 were pooled, and protein concentration measured by OD_{280}. The OD_{280} of the pooled samples was 0.0621, the calculated extinction coefficient of GB1[N] was 58 900$M^{-1}cm^{-1}$, and the path length of the cuvettes used was 1cm. The volume of the fractions totalled 20mL. Based on this we estimated GB1[N] yield per 500mL growth media is 0.542mg. The 0.550mL, 1mM protein solution required for NMR experimentation would require at least 14.00mg of pure, uncleaved GB1[N], likely more as this number assumes that no further GB1[N] losses would be incurred. The volume of labelled minimal media required to produce sufficient labelled GB1[N] would be impractically expensive. The GB1[N] buffer exchange was successfully completed, however an SDS-Page of fractions with an A_{280} above 50mAU revealed no detectable protein (Appendix 6.6.3).

Low protein yields, poor solubility, expected proteins losses during each experimental step and the cost of producing labelled proteins for use in NMR experiments lead to the exploration of other constructs for the production of YthA fusion proteins which may improve protein production and yield of pure protein.

2.5 New Approaches and Constructs for YthA Production

When looking for alternative production methods two general approaches were considered for the production of YthA: the first maximized solubility and reduced additional experimental steps necessary to produce the YthA fragments in their native conformation, thus reducing or eliminating protein loss from aggregation and purification. The second approach was to use an established IB production system known to maximize protein production and IBs that are readily solubilize, thus offsetting losses from the necessary processing steps.

IBs may contain more than 80% pure recombinant protein, and the quantity of recombinant protein produced by the cell is usually much higher if the protein forms IBs than if the recombinant protein is soluble (Rodríguez *et al.,* 2014; Hoffmann *et al.,* 2018). However, purification of IBs require several new steps as compared to proteins in the supernatant, as discussed in Figure 14. This is due to the steps needed for removing unlysed cells and cellular debris from the IBs (Figure 14, Step A), removing lipids and soluble proteins from the IBs (Figure 14, Step C) prior to resolubilizing (Figure 14, Step E) and refolding of the proteins (Figure 14, Step F). Further, the resolubilization and refolding step results in protein loss, although the steps can be optimized to reduce the loss. The high yield and comparatively pure protein of IBs suggested to us that an optimized IB purification scheme combined with a high-yield production tag could overcome the protein loss caused by the extra steps processing the IBs, thereby making IB production beneficial.

Keto-Steroid Isomerase (KSI) was considered to be a strong candidate for tagging the YthA fragments as it has been known to produce 50mg of fusion protein per litre of growth media and it readily causes the formation of IBS that can be shown to be effectively refolded (Morreale *et al.,* 2004; Rodriguez *et al.,* 2014). Because the GB1-tagged fusion proteins also formed IBs, this was considered acceptable if it produced other benefits. It was also thought that the strong structure of the KSI tag and IB formation should have a synergistic effect in preventing protein degradation (Schrader *et al.,* 2009). Additionally, several well tested protocols for KSI solubilization and purification exist (Morreale *et al.,* 2004). Maltose-binding protein (MBP) – YthA constructs were also designed as they are reported to readily solubilize in mild conditions, increase protein production and improve yields (Bronsoms & Trejo, 2015), however, these constructs were not produced, as MBP-tag purification requires purification systems not currently available in our laboratory.

KSI constructs were inserted between the XhoI and BlpI restriction sites of the pET31b(+) vector to produce a KSI-YthA construct. Upon induction, the fusion protein produced would have the YthA fragment linked to the KSI tag. A similar strategy was to be used to join the MBP tag to the remainder of the fusion protein by inserting the construct into the pMAL-c5X vector, where the construct was to be inserted between the Ava I and Nde I restriction sites. The protein map and selected properties of each of the KSI- and MBP-tagged fusion proteins are shown in Figure 19.

KSI[N, Variant 1]; N-terminal YthA fragment tagged with KSI
KSI tag – Factor Xa cut site – N-terminal YthA Fragment – TEV cut site – His tag
pI=5.96 Mw= 23801.32Da

KSI[N, Variant 2]; N-terminal YthA fragment tagged with KSI (Internal histadine tag)
KSI tag – His tag – TEV cut site – N-terminal YthA Fragment
pI=6.20 Mw= 23894.37Da

MBP[N]; N-terminal YthA fragment tagged with MBP
MBP – Linker – TEV cut site – Factor Xa cut site – N-terminal YthA Fragment
pI=5.70 Mw= 49044.00Da

KSI[C, Variant 1]; C-terminal YthA fragment tagged with KSI
KSI tag – Factor Xa cut site – C-terminal YthA Fragment – TEV cut site – His tag
pI=5.22 Mw=28906.01Da

KSI[C, Variant 2]; C-terminal YthA fragment tagged with KSI (Internal histadine tag)
KSI tag – His tag – TEV cut site – C-terminal YthA Fragment
pI=5.49 Mw= 28999.06Da

MBP[C]; C-terminal YthA fragment tagged with MBP
MBP – Linker – TEV cut site – Factor Xa cut site – C-terminal YthA Fragment
pI=5.16 Mw= 54148.69Da

Figure 19: Designations, descriptions, maps, isoelectric points and molecular weights of YthA fusion proteins tagged with KSI or MBP.

YthA fragments without a production tag were also designed to reduce the amount of amino acid necessary for tag production and potentially removing the need for a second cleavage step. These fusion proteins should be the easiest to purify as no additional tag would interfere with the His-tag's function and only a single cleavage step would be necessary to free the YthA fragment from any tags, if production was successful, and protein degradation was avoided (Koenig *et al.,* 2003).

The inserts coding for YthA fragments lacking production tags were inserted into the pET16(+) vector between the Nde I and Xho I restriction sites to produce a construct coding for His_{10}-tag on the N-terminal of the YthA fragment. The protein map and selected properties of each of His[N] and His[C] are shown in Figure 20.

His[N]; N-terminal YthA fragment tagged with a His-tag only
His tag – TEV cut site – Factor Xa cut site - N-terminal YthA Fragment
pI=9.98 Mw= 9527.96Da

His[C]; C-terminal YthA fragment tagged with a His-tag only
His tag – TEV cut site – Factor Xa cut site – C-terminal YthA Fragment
pI=6.10 Mw= 14632.65Da

Figure 20: Designations, descriptions, maps, isoelectric points, and molecular weights of YthA fusion proteins tagged with only a His-tag.

2.5.1 Expression Tests on New Constructs

Expression vectors were successfully transformed into BL21(DE3) cells and protein was successfully produced after IPTG induction. (Figure 8, Steps C and D) KSI[N, Variant 2] was the exception, no fusion protein was detected following induction of KSI[N, Variant 2]-transformed cells (Appendix 6.6.4). Higher OD_{600} measurements were obtained when induction occurred at 18°C compared to 37°C (Figure 21) and large amounts of fusion protein were produced by the induction of transformed cells as confirmed by SDS-page comparing induced and uninduced cells (Figure 22). SDS-Page shows that production of KSI[C, Variant 1] fusion protein far exceeded the production of GB1[C] and His[C] fusion proteins (Figure 22). The high protein yields of KSI[C, Variant 1] is consistent with results from others, (Moreale *et al.,* 2004), whereas GB1 yields were less than half of what was expected (Cheng & Patel, 2004).

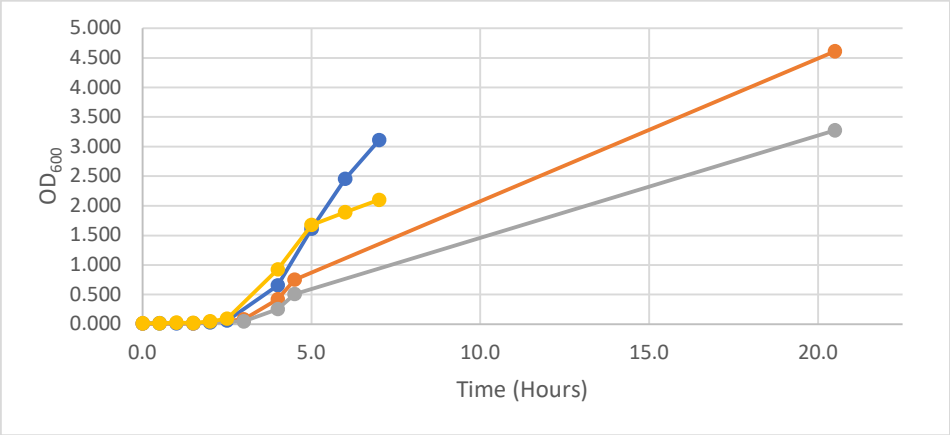


Figure 21: Growth profiles typical of KSI[C, Variant 1]-transformed and His[C]-transformed BL21(DE3) cells under 37°C (Blue Line for KSI[C, Variant 1], Yellow Line for His[C]) and 18°C (Orange Line for KSI[C, Variant 1], Grey Line for His[C]) induction temperatures.

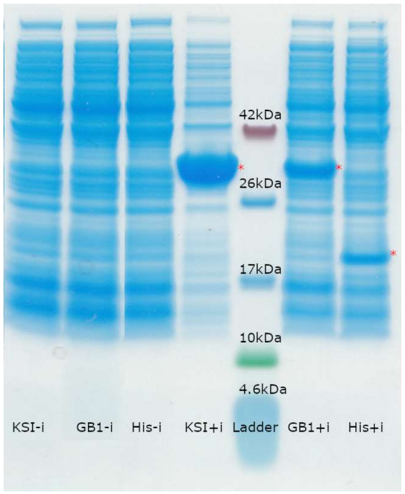


Figure 22: SDS-page of induced (+i) and uninduced (-i) His[C]- (His), GB1[C]- (GB1) or KSI[C, Variant 1]-transformed cells. A protein ladder is included for size comparison. Red asterisks indicate fusion protein. SDS-page edited for clarity; unedited SDS-page ***in Appendix 6.6.1.***

Similar to previous literature (Moreale *et al.,* 2004), the KSI-tagged fusion protein (KSI[C, Variant 1]) was almost exclusively detected in the protein pellet and virtually undetectable in the soluble fraction.

2.5.1.1 Quantification of KSI[N, Variant 1] and KSI[C, Variant 1] Yield by Bradford Assay

To determine if further optimization of KSI[N, Variant 1] and KSI[C, Variant 1] was an advantageous use of time and resources, we decided to estimate the current protein yield per litre of growth media using a Bradford Assay (Bradford, 1976). Washed KSI[N, Variant 1] and KSI[C, Variant 1] protein pellets each produced from 0.5L LB media induced at 18°C were solubilized in 50mL of buffer to maximize solubilized protein and improve accuracy. The protein estimation is based on a standard curve of bovine serum albumin (BSA) concentration and absorbance in the same buffer as KSI[N, Variant 1] and KSI[C, Variant 1] (Figure 23), and as such is a notable source of error in the estimate. Another notable source of error are the protein impurities present in the protein pellets, though this is expected to be a much lesser error source due to the comparative purity of the protein pellets, as seen in previous SDS-pages. Repeats were performed to minimize any other random sources of error.

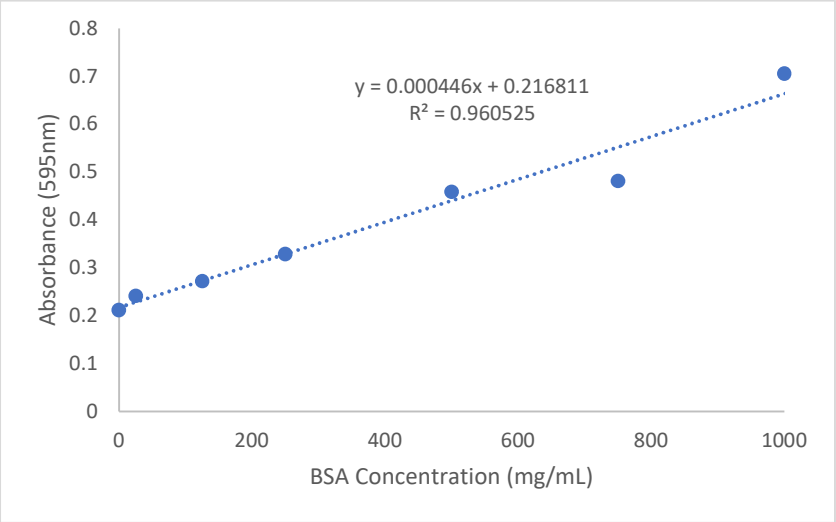


Figure 23: A_{595} absorbance curve of BSA using the average of 3 repetitions. Measured data points are indicated by blue dots and the line-of-best-fit is indicated as a dotted line. The R-squared value and equation of the line-of-best-fit are also given.

Dilutions of KSI[N, Variant 1] and KSI[C, Variant 1] with an absorbance closest 0.440, the expected absorbance of the mid-point of the linear range of BSA, were used to determine protein concentration in the samples and protein yield per litre of growth media. After accounting for the absorbance of the blank samples, the relevant dilution factors and the volume of growth media relative to 1L, KSI[N, Variant 1] was estimated to yield 50.77mg/L of LB growth media and KSI[C, Variant 1] was estimated to yield 319.64mg/L of LB growth media.

Both KSI[N, Variant 1] and KSI[C, Variant 1] produced very high yields of proteins relative to the volume of growth media. Though high, the yield of KSI[N, Variant 1] remains below some previously reported KSI-tagged recombinant protein yields, however KSI[C, Variant 1] yielded far more protein than those yields (Rodríguez *et al.,* 2014). However, direct comparison of the yields presented here and those presented elsewhere in literature should be undertaken with caution, as the yields reported in other literature are typically measured after purification and tag-cleavage. Based on the high yield it is clear that further optimization of KSI[C, Variant 1] protein production is not worthwhile and resources when the issues of solubility, purification and refolding remain unresolved. However, further optimization of KSI[N, Variant 1] appeared to be worthwhile as the construct already displayed high production levels during our experiments. KSI[C, Variant 1], a highly similar construct to KSI[N, Variant 1], was able to achieve yields over 6 times greater than KSI[N, Variant 1] indicating that KSI[N, Variant 1] yield might still be increased drastically and potentially offset processing losses.

2.5.1.3 Optimization of KSI[N, Variant 1] Growth

Among the easiest and most important parameters in protein production optimization is heat. Although the optimal growth rate of *E. coli* is usually achieved at 37°C, recombinant protein solubility and yields can often be maximized by reducing the colony's temperature following induction (Ferrer-Miralles *et al.,* 2022). To test for the optimal induction temperature, we split a 1L LB media growth once it achieved an OD_{600} of 0.5, leaving one in a 37°C shaking incubator for 4 hours prior to harvesting and placing one in a pre-cooled 18°C shaking incubator for 20 hours prior to harvesting. The cell pellets were lysed, washed and then completely resuspended in 10mL of tris loading buffer before SDS-analysis (Figure 24).

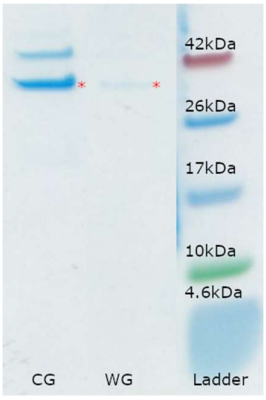


Figure 24: SDS-page of KSI[N, Variant 1] fusion protein production following induction at 37°C for 4 hours (WG) or at 18°C for 20 hours (CG). A protein ladder is included for size comparison. Red asterisks indicate KSI[N, Variant 1]. SDS-page edited for clarity; unedited SDS-page in ***Appendix 6.6.5****.*

As Figure 24 clearly shows, KSI[N, Variant 1] induced at 18°C for 20 hours yields several times the protein compared to the same colony induced at 37°C for 4 hours. The KSI[N, Variant 1] sample tested had yields (313mg/L media) similar to those of KSI[C, Variant 1] suggesting that the induction temperature optimization was a great success. All future protein productions on KSI[N, Variant 1] were performed at 18°C for 20 hours.

2.5.2 Solubilization Tests on KSI Fusion Proteins

Solubilizing KSI fusion proteins was an expected and necessary step in producing the YthA fragments with KSI constructs (Figure 14, Step E). Previous studies indicated that IBs produced at lower temperatures are typically more soluble than those produced at 37°C (Ferrer-Miralles *et al.*, 2022). We tested the solubility of IBs grown at both 18°C and 37°C to ensure the highest yield of recombinant protein possible following processing. More of the 37°C induced KSI[C, Variant 1] fusion

protein solubilized compared to its 18°C induction counterpart, as evidenced by SDS-Page analysis although most KSI[C, Variant 1] was particulate (Figure 25). The result more 37°C-induced KSI[C, Variant 1] solubilized compared to 18°C-induced KSI[C, Variant 1] under otherwise identical growth and solubilization conditions led to us inducing KSI[C, Variant 1] at 37°C in all future experiments. 8M urea and 6M GHCl were chosen to solubilize KSI fusion proteins instead of milder methods as several previous studies have found such conditions necessary for KSI fusion protein solubilization (Morreale *et al.,* 2004; Rodríguez *et al.,* 2014; Hoffmann *et al.,* 2018).

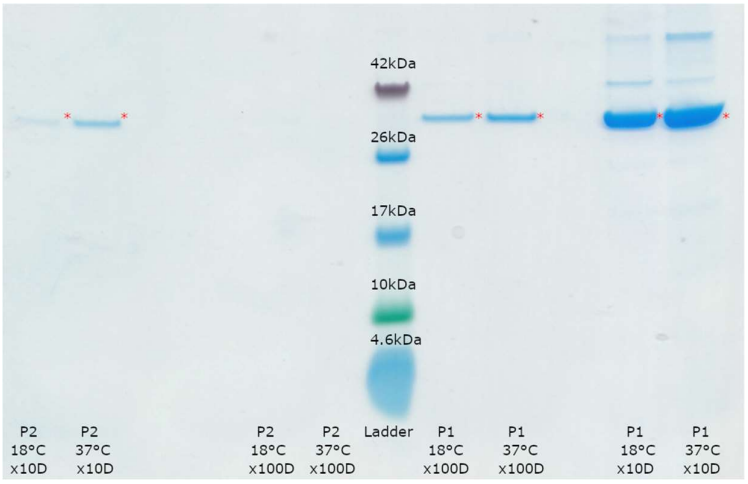


Figure 25: SDS-Page of the resuspended KSI[C, Variant 1] IBs grown at 18°C and 37°C, centrifuged at 600g (P1) or 23 000g (P2) and diluted 10-fold (X10D) or 100-fold (X100D). A protein ladder is included for size comparison. Red asterisks indicate KSI[C, Variant 1] fusion protein.

KSI[N, Variant 1] was successfully produced at 37°C and washed (Figure 26), though the protein pellet containing KSI[N, Variant 1] was notably smaller than those of KSI[C, Variant 1]. Additionally, a large protein of approximately 40kDa, likely a membrane protein, was present in the KSI[N, Variant 1] sample (Figure 26), similar to those seen in the KSI[C, Variant 1] samples (Figure 25). While solubilization of the 37°C-induced KSI[N, Variant 1] was noticeably easier during handling compared to the 18°C-induced KSI[N, Variant 1], the improved solubilization did not offset the lower yield. Another solubilization protocol could also be attempted on the undissolved 18°C-induced KSI[N, Variant 1] pellet after the solution containing the dissolved protein was removed. Repeating the solubilization protocol on the pellet would maximize solubilization yields and minimize protein loss at this step.

It should be noted that the band intensities of KSI[N, Variant 1] (Figure 26) and KSI[C, Variant 1] (Figure 25) in their respective post-wash SDS-pages cannot be compared as the KSI[N, Variant 1] sample was taken by scraping the protein pellet.

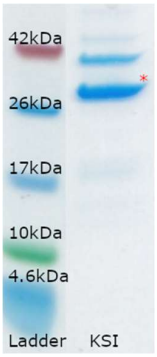


Figure 26: SDS-page of 37°C-induced KSI[N, Variant 1] washed IBs. A protein ladder is included for size comparison. Red asterisks indicate KSI[N, Variant 1]. SDS-page edited for clarity; unedited SDS-page in Appendix 6.6.6.

2.5.3 Growth and Production of KSI[N, Variant 1] ^{15}N-labelled M9 Media

Labelled M9 media typically produces approximately half the recombinant protein compared to unlabelled LB media due to the lower nutrient and energy levels available in M9 minimal media compared to a rich media, such as LB (Marley *et al.*, 2001). The changes in the average atomic weights of nitrogen and, in future experiments, carbon between labelled and unlabelled media also stress the cell and may impact cell growth and protein production, leading to reduced yield. Protocols exist for the mitigation of these issues (Sivashanmugam *et al.,* 2009), though experimentation is required to optimize the protocols for each recombinant protein and production system. To determine if such optimization was necessary, we grew KSI[N, Variant 1]-transformed BL21(DE3) cells in ^{15}N-labelled M9 media to determine the decrease in cell density, and thus presumably in protein production.

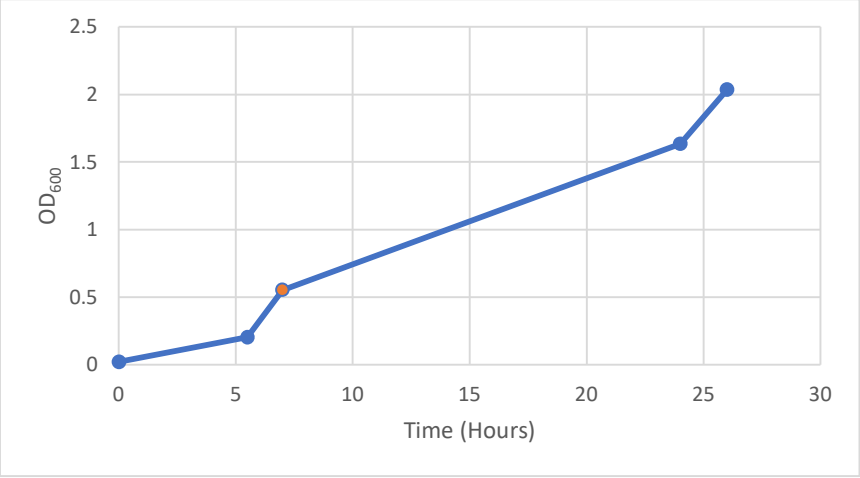


Figure 27: Growth curve of KSI[N, Variant 1]-transformed BL21 (DE3) cells in ^{15}N-labelled M9 media as measured by OD_{600}. The orange point is induction time (7 hours).

KSI[N, Variant 1]-transformed cells in labelled media grew to approximately half the cell density of those cells grown in unlabelled LB media (Figures 21 and 27), as is typical (Marley *et al.*, 2001). Using these results, we determined that protocol optimization would benefit the production of KSI-tagged proteins in labelled M9 media, though due to the already high yields of KSI[N, Variant 1] and especially KSI[C, Variant 1] optimization was unnecessary and other experiments took priority.

2.5.4 Comparing Refolding by Dialysis and Rapid Dilution KSI[N, Variant 1]

We attempted to optimize the refolding process of KSI[N, Variant 1] (Figure 14, Step F) by comparing the estimated KSI[N, Variant 1] yield remaining following rapid dilution refolding or dialysis. Measurements were taken at each stage of the refolding process to determine where protein loss was occurring.

Table 6: Estimates of KSI[N, Variant 1] remaining following refolding by rapid dilution or dialysis and subsequent sample concentration.

Most Recent Step Completed	KSI[N, Variant 1] Yield (mol) estimate per L of LB media
IB Washing	1.315×10^{-5}
Rapid Dilution	1.364×10^{-5}
Concentration following Rapid Dilution	4.381×10^{-6}
Dialysis	Precipitate prevented accurate reading
Concentration following Dialysis	1.108×10^{-7}

The rapid dilution refolding of KSI[N, Variant 1] appeared to have negligible protein loss, and the slight difference in protein concentration is within experimental error (Table 6). Concentration of the rapid dilution sample caused a loss of approximately 2/3 of the protein present. The sample lost was due to protein precipitation against the centrifugation filter and small protein impurities passing through the filter, which should be possible to redissolve and refold again to maximize the yield of this step.

99% of the protein present in the washed IB was lost following concentration of the dialysis sample. What remained in solution was not sufficient for concentration determination. Table 6 above shows that rapid dilution refolding of KSI[N, Variant 1] protein is superior to refolding through dialysis.

An SDS-page was performed on the samples following each stage of processing to ensure that the results were not artifacts of the A_{280} measuring process. SDS-page analysis of the KSI[N, Variant 1] refolding samples (Figure 28) reinforces the conclusion that rapid dilution refolding lost some amount of KSI[N, Variant 1] but retained some soluble KSI[N, Variant 1]. However, the SDS-page confirms that nearly all KSI[N, Variant 1] was lost from the dialysis sample. The presence of a strong band at the expected size of KSI[N, Variant 1] dimer suggests that the x1 LDS buffer was insufficient to denature fusion protein macromolecules. However, no changes were made to the buffer as the dimer did not interfere with the SDS-page.

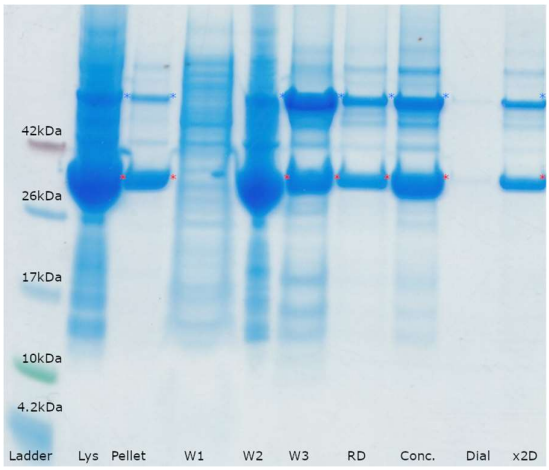


Figure 28: SDS-page of impure KSI[N, Variant 1] cell lysate (Lys), IB pellet (Pellet), supernatant of wash 1 (W1), pellets of washes 2 (W2), and 3 (W3), rapid dilution refolding supernatent (RD), concentration following rapid dilution refolding (Conc.), dialysis supernatant (Dial) and x2 diluted concentrated rapid dilution refolding sample (x2D). A protein ladder is included for size comparison. Red asterisks indicate KSI[N, Variant 1] monomer and blue asterisks indicate KSI[N, Variant 1] dimer.

The concentrated, rapid diluted sample was purified by ion affinity HPLC (Figure 8, Step E; Figure 14, Step G).

There were no detectable peaks in the purification chromatograph (Appendix 6.5.4). This indicates that no protein bound to the column, and refolded KSI[N, Variant 1] flowed through the column.

2.5.5 Ion Affinity Chromatography Purification of KSI[C, Variant 1]

Initial attempts at ion affinity HPLC purification were initially unsuccessful with the resolubilized KSI[C, Variant 1], as the fusion protein was believed to fail to bind to the column during injection, and absorbance during elution gave no indication that protein was eluted (Appendix 6.5.5). The lack of protein binding may be due to imidazole present in the buffers outcompeting the fusion protein for matrix binding. We thus reattempted the experiment with decreased concentration of imidazole in the loading and refolding buffers.

Repeating the on-column refolding and elution of KSI[C, Variant 1] with a reduced concentration of imidazole had no effect (Appendix 6.5.6). This suggests KSI[C, Variant 1] has very poor binding affinity to the Histrap column. To test whether KSI[C, Variant 1] passed through the column, the flowthrough was concentrated in a 20 000 MWCO Amicon tube to approximately 5mL. The concentrated flowthrough was compared to the sample injected into the HPLC apparatus by SDS-Page, which revealed that KSI[C, Variant 1] had flowed through (Figure 29). Considering the numerous other potential causes of the protein binding failure, including an inaccessible His-tag and insufficient ionic interaction to immobilize the fusion protein, an attempt was made to use anion exchange chromatography to exchange the buffers of the recovered KSI[C, Variant 1] to a 5mM imidazole refolding buffer (Low-imidazole refolding buffer) prior to refolding and additional purification. These experiments also failed to give any indication that KSI[C, Variant 1] was successfully refolded or eluted (Appendix 6.5.7). The necessity of exchanging buffers into an NMR buffer at later stages of the project provides an additional benefit to this approach.

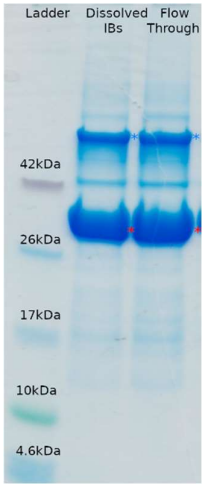


Figure 29: SDS-Page of dissolved and washed KSI[C, Variant 1] fusion protein IBs and concentrated flowthrough from the low-imidazole purification attempt of KSI[C, Variant 1]. A protein ladder is included for size comparison. Red asterisks indicate KSI[C, Variant 1] monomer and blue asterisks indicate KSI[C, Variant 1] dimer.

We tested if KSI[C, Variant 1] purification would be more successful in a pH 8 tris buffer because pH 8 is further from KSI[C, Variant 1]'s isoelectric point. An additional benefit would be that as the digestion buffer is a tris buffer as well, the exchange is less likely to shock KSI[C, Variant 1] into precipitating. A new growth of KSI[C, Variant 1]-producing cells was harvested, lysed, washed, and solubilized using the appropriate pH 8 tris buffers. It appeared as though a greater amount of the KSI[C, Variant 1] IBs had been dissolved in the tris loading buffer compared to attempts to dissolve it into the standard loading buffer. Detectable column binding did occur under these conditions, though the KSI[C, Variant 1] bound was minimal and eluted over very wide peaks (Figure 30). The KSI[C, Variant 1] that did bind was eluted at comparatively high imidazole concentration and though little KSI[C, Variant 1] was recovered, it was highly pure (Figure 31).

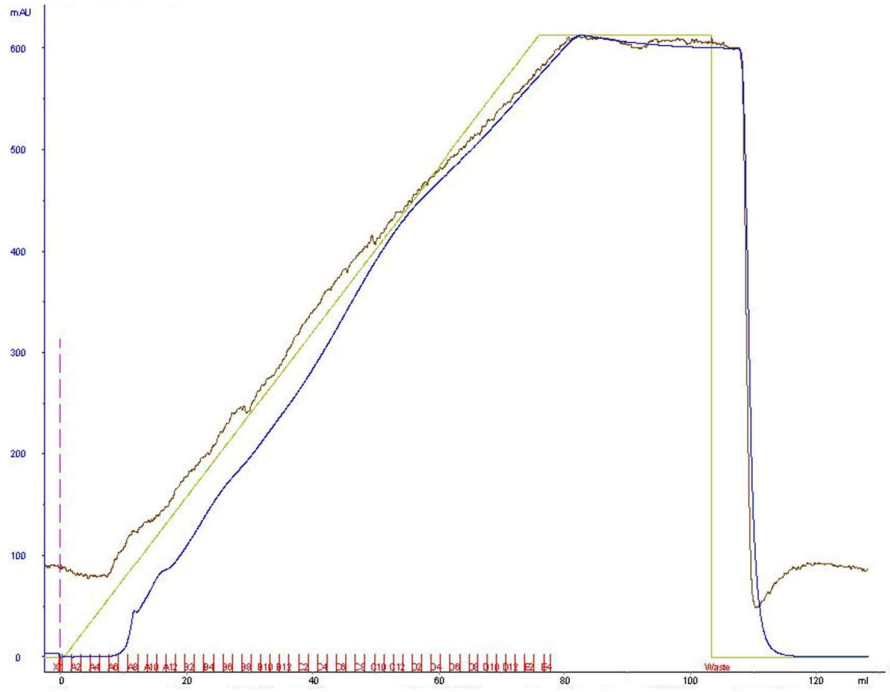


Figure 30: Elution chromatograph of KSI[C, Variant 1] fusion protein purification attempt using tris buffers. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters).

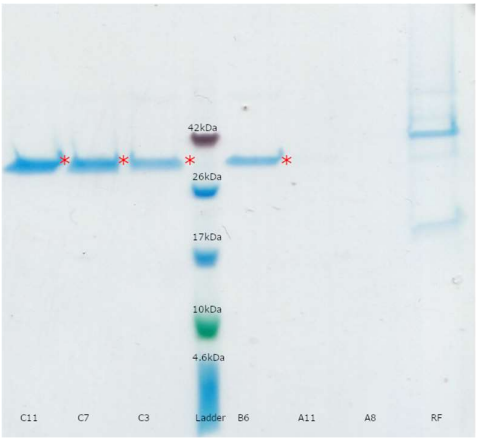


Figure 31: SDS-page of concentrated KSI[C, Variant 1] elution fractions A8, A11, B6, C3, C7, C11; and refolding flow-through (RF) following an on-column refolding test using tris buffers. A protein ladder is included for size comparison. Red asterisks indicate KSI[C, Variant 1] fusion protein.

2.5.6 Production and Rapid Dilution Refolding of KSI[C, Variant 2]

Deciding to test the efficacy of a larger, internal His-tag, reasoning that even if protein yield were adversely affected, it may still be the preferred construct if purification efficiency was improved. We transformed BL21(DE3) cells with the KSI[C, Variant 2] construct successfully, as well as grew and induced production of KSI[C, Variant 2]. SDS-page of washed IBs confirmed the production of KSI[C, Variant 2] in induced cells (Figure 32). The yield of KSI[C, Variant 2] appeared very similar to that of the KSI[C, Variant 1] construct.

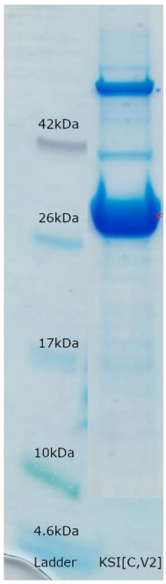


*Figure 32: SDS-page of washed KSI[C, Variant 2] fusion protein IBs. A protein ladder is included for size comparison. Red asterisks indicate KSI[C, Variant 2] monomer and blue asterisks indicate KSI[C, Variant 2] dimer. SDS-page has been edited for clarity; unedited SDS-page in A**ppendix 6.6.7.***

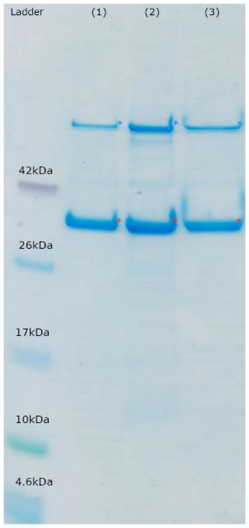


Figure 33: Comparison of solubilized KSI[C, Variant 2] protein samples with putative KSI[C, Variant 2] fusion protein monomer and dimer indicated. (1) prior to rapid dilution refolding, (2) rapid diluted into an 8-fold greater volume, (3) rapid diluted into a 20-fold greater volume. Samples were normalized to putatively the same protein concentration. A protein ladder is included for size comparison. Red asterisks indicate KSI[C, Variant 2] monomer and blue asterisks indicate KSI[C, Variant 2] dimer.

Rapid dilution refolding on KSI[C, Variant 2] was successful in increasing protein solubility (Figure 33) with an 8-fold dilution providing the best results. Additionally, there was no apparent change in the ratio of monomer to dimer in any of the samples, though this may be an artifact of the normalization process.

The purification of refolded KSI[C, Variant 2] was inconclusive due to HPLC apparatus malfunction. Distinct peaks in the conductivity measurements indicate that the buffers were not mixed properly and the conditions inside the ion affinity column were less controlled than expected. Regardless of the cause, no clear elution peaks were detected and there was no detectable protein binding.

2.6 Production of His YthA Constructs

Reconsidering our use of KSI- and GB1-tagged proteins, we decided to test the ability of His[N] and His[C] to bind to the ion affinity column. Should we successfully be able to purify the proteins, production optimization might increase the protein yield enough for use in NMR experiments. Both His fusion proteins were successfully produced (Figure 8, Step D), production of His[C] is shown in Figure 22, and the His[N]-induced cells consistently produced a weak, smeared band around 10kDa in size; this was interpretted as a production success. Both fusion proteins formed IBs and were processed accordingly.

2.6.1 Purification Attempt on His[N]

By detecting absorbance at 260nm in addition to absorbance at 280nm, we hoped to increase the sensitivity of our measurements. This is because that while tryptophan and especially tyrosine have a high extinction coefficient at 280nm, the extinction coefficient at 260nm is far smaller as the absorbance curves of both residues begin to approach a local minimum. However, the absorbance curve of phenylalanine contains a local maximum at approximately 260nm and has a higher extinction coefficient at this wavelength than either tryptophan or tyrosine (Teale & Weber, 1957). Because His[N] contains 6 tyrosine residues, 6 tryptophan residues and 3 phenylalanine residues (Bolotin *et al.,* 2001) a strong peak in A_{280} and a moderate peak in A_{260} simultaneously during elution suggests that the fraction contains His[N].

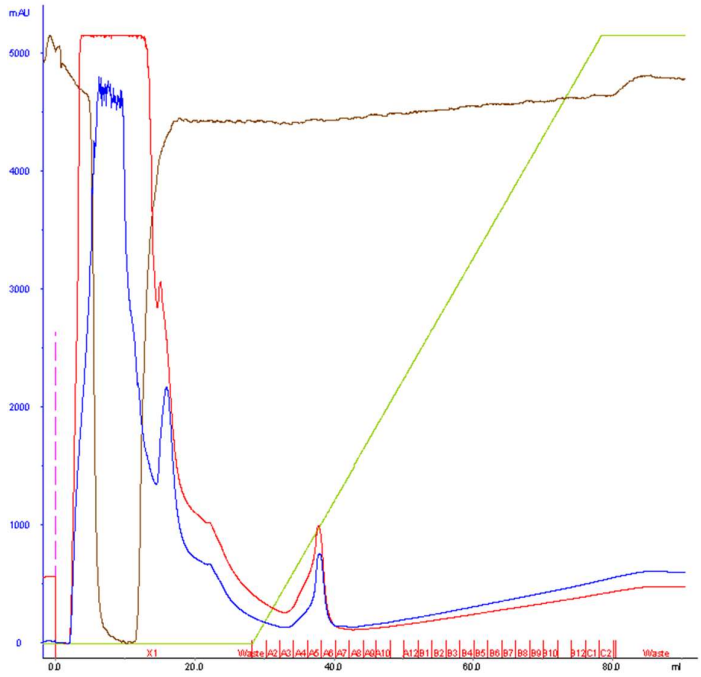


Figure 34: Elution chromatograph of His[N] fusion protein purification attempt. Indicated on the chromatograph are absorbance at 280nm (blue line), absorbance at 260nm (red line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters).

The purification of His[N] was believed to be successful with a clear peak in fractions A5 and A6, and possibly A4 (Figure 34). Fraction A6 was selected for buffer exchange to a digestion buffer. The buffer exchange of His[N] was successful, and fractions A1-A3, A8 and A10 were sampled for SDS-page analysis (Figure 35).

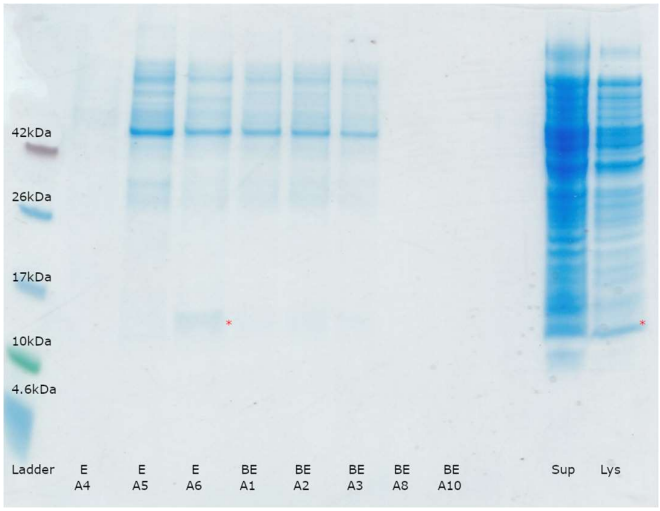


Figure 35: SDS-page of His[N] elution fractions (E) A4, A5, A6; buffer exchange fractions (BE) A1, A2, A3, A8, A10; His[N] supernatant following cell lysis (Sup) and 5x diluted cell lysate (Lys). A protein ladder is included for size comparison. Red asterisks indicate His[N] fusion protein.

Curiously, a large amount of high MW protein co-eluted with His[N], the cause is unclear as so many different proteins would have, by chance, been able to bind successfully to the Ni-NTA column and elute in a similar fraction to His-tagged protein. No His[N] was detected in the A4 and A5 fractions, though other high MW proteins were eluted in these fractions (Figure 35) and thus caused the large shoulder at A4 and A5 from the NiNTA purification chromatograph (Figure 34). Additionally, putative His[N] was only detected in elution fraction A6 and in barely detectable amounts (Figure 35). The His[N] fusion protein was abandoned due to low yield and purification failures.

2.6.2 Purification Attempt on His[C]

In this thesis, we have found that inducing the production of proteins at lower temperatures increases the protein yield (Figure 24), and in previous literature, proteins produced at such temperatures were more soluble than those produced at higher temperatures (Holmes, 1999), thus we chose to produce His[C] at 18°C. Growth and induction of His[C] succeeded (Figure 22). His[C] formed IBs, similar to other fusion proteins produced in this thesis, though solubilized more easily and completely than the KSI- and GB1-fusion proteins.

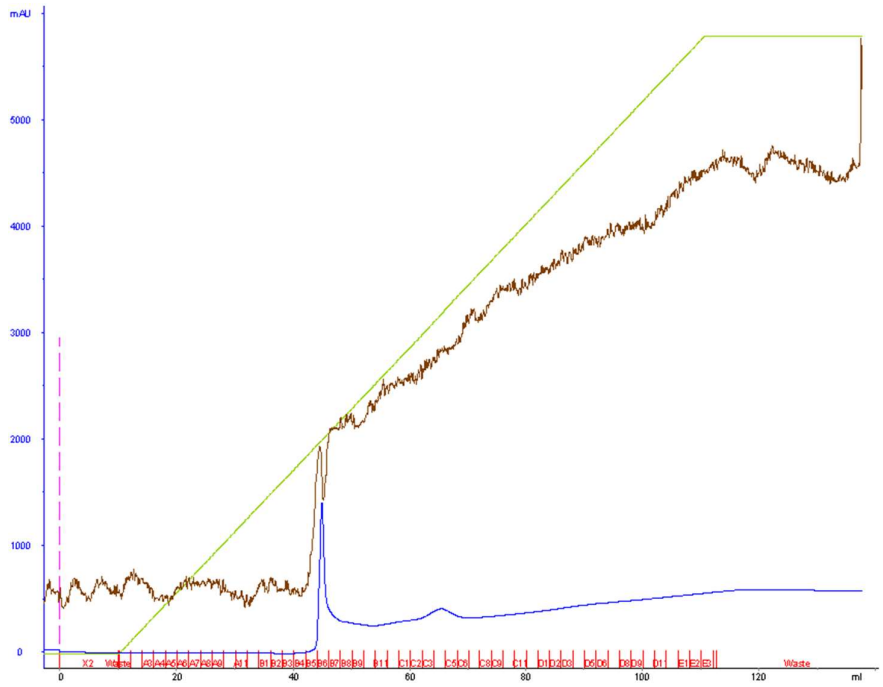


Figure 36: Elution chromatograph of 18°C-induced His[C] fusion protein purification attempt using tris buffers. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters).

In the elution step (Figure 36), a strong, narrow peak of absorbance at fraction B6 was observed, possibly indicating purification success. The putatively pure His[C] rapidly flowed through the column during buffer exchange and was collected. His[C] fractions A1 and A2 were analysed by SDS-page to ensure the protein is pure and of high yield.

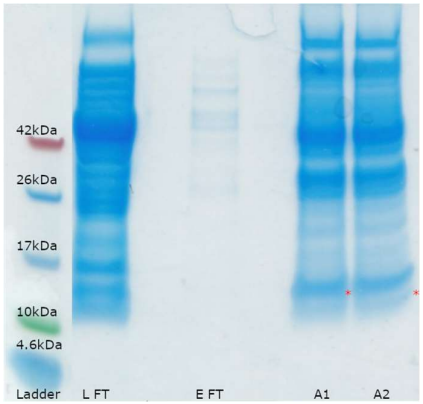


Figure 37: SDS analysis of His[C] purification fraction A1 and A2, elution flowthrough (E FT), and loading flowthrough (L FT). A protein ladder is included for size comparison. Red asterisks indicate His[C] fusion protein. SDS-page gel has been edited for clarity, unedited SDS-page in ***Appendix 6.6.8.***

SDS-page confirmed the presence of His[C] in the putatively pure sample (Figure 37), however, the purification failed and His[C] appears to have co-eluted with many contaminating proteins. Due to the very high levels of contaminating protein following purification and buffer exchange, purification of His[C] failed. His[C] was abandoned following low yields and purification failures.

2.7 Production and Purification of TEVp

Due to the continued difficulties with purifying the YthA fusion proteins led us to test our protocols and equipment. TEVp is necessary for the removal of production and purification tags from the YthA fragments and can be produced and purified with protocols similar to ours (Tropea *et al.,* 2009), making it an excellent candidate for these tests.

Isolated pRK793 vector was gifted from Daniel Hatlem, ready to be transformed into cells. pRK793 is an expression vector containing an MBP-tagged TEVp insert and an ampicillin resistance marker (Appendix 6.4.4). pRK793 was successfully transformed into BL21(DE3) cells and TEVp successfully induced (Figure 38). The tagged protease is self-cleaving due to the inclusion of a TEVp cleavage site, ENLYFQG, linker sequence between the tag and the protease.

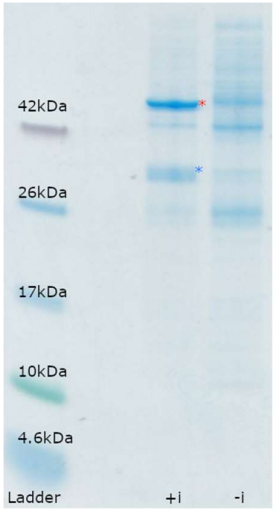


Figure 38: SDS-page of cell lysate of TEVp-transformed cells prior to induction (-i) and 4 hours after induction (+i). A protein ladder is included for protein size comparison. Red asterisks indicate the cleaved MBP-tag and blue asterisks indicate untagged TEVp.

TEVp and the cleaved MBP-tag were detected in the induced cells (Figure 38), indicating production was successful.

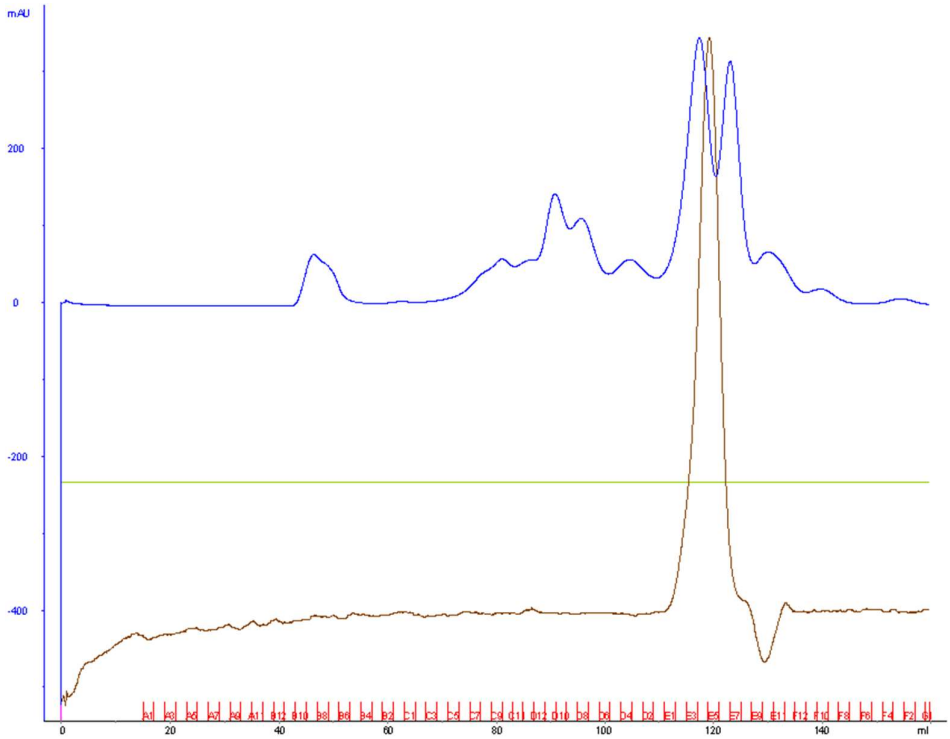


Figure 39: Size exclusion chromatograph of TEVp-containing lysate. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), flow rate (green line), and sample fractions (red letters).

SDS-page of the pooled fractions E6-E8 and E1-E4 shows the active TEVp was eluted in fractions E6-E8, and the fractions E1-E4 contained the larger cleaved MBP-tag, and the uncleaved MBP-TEVp fusion protein, ie. the larger peptides (Figures 39 and 40).

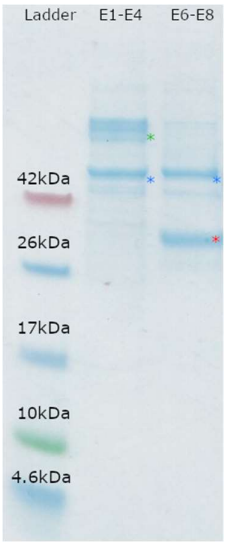


Figure 40: SDS-page of size exclusion purification of TEVp fractions E6-E8 (5x diluted, pooled) and E1-E4 (5x diluted, pooled). A protein ladder is included for size comparison. Red asterisks indicate untagged TEVp, blue asterisks indicate cleaved MBP-tag, and green asterisks indicate uncleaved MBP-TEVp fusion protein.

TEVp was successfully purified in an active conformation as proven by Figure 40. Two bands are clearly visible in the well containing fractions E6-E8 corresponding to an untagged TEVp and a cleaved MBP-tag, indicating that the TEV-MBP fusion protein was successfully cleaved at the inserted cleavage site. The very high molecular weight (MW) bands present in the E1-E4 fractions indicate TEVp with an uncleaved MBP-tag.

2.8 Cleavage Testing of TEVp

During the planning of this project, TEVp was selected as one of the proteases used in the project due to its high specificity though it does cleave degenerate sites, especially when the only degenerate amino acid is in the C-terminal end of the recognition site (Kapust *et al.,* 2002). All the fusion proteins designed in this thesis contain a TEVp cleavage site (ENLYFQ/S) to allow separation of the YthA fragment from purification or production tags. However, non-specific cleavage may still occur. For these reasons optimization involves both maximizing cleavage at the inserted cleavage site and minimizing cleavage at other sites. By varying the temperature, duration and concentration of protein present in the reaction, selective cleavage at the inserted cleavage sites was optimized. Due to promising yields and the purification success of KSI[C, Variant 1], KSI fusion proteins were the chosen for this cleavage test. KSI[C, Variant 2] was chosen for the test as it would only require a single cleavage reaction to obtain the untagged, C-terminal YthA fragment, and it was hoped KSI[C, Variant 2] yield could be increased above 300mg/L, similar to KSI[C, Variant 1] and KSI[N, Variant 1]. KSI[N, Variant 1] was chosen for the cleavage reaction due to the failure of KSI[N, Variant 2]-transformed cells to produce recombinant protein and for the reasons listed above. However, difficulty was expected in differentiating uncleaved from cleaved KSI[N, Variant 1] by SDS-page as the difference in MW is only approximately 1kDa.

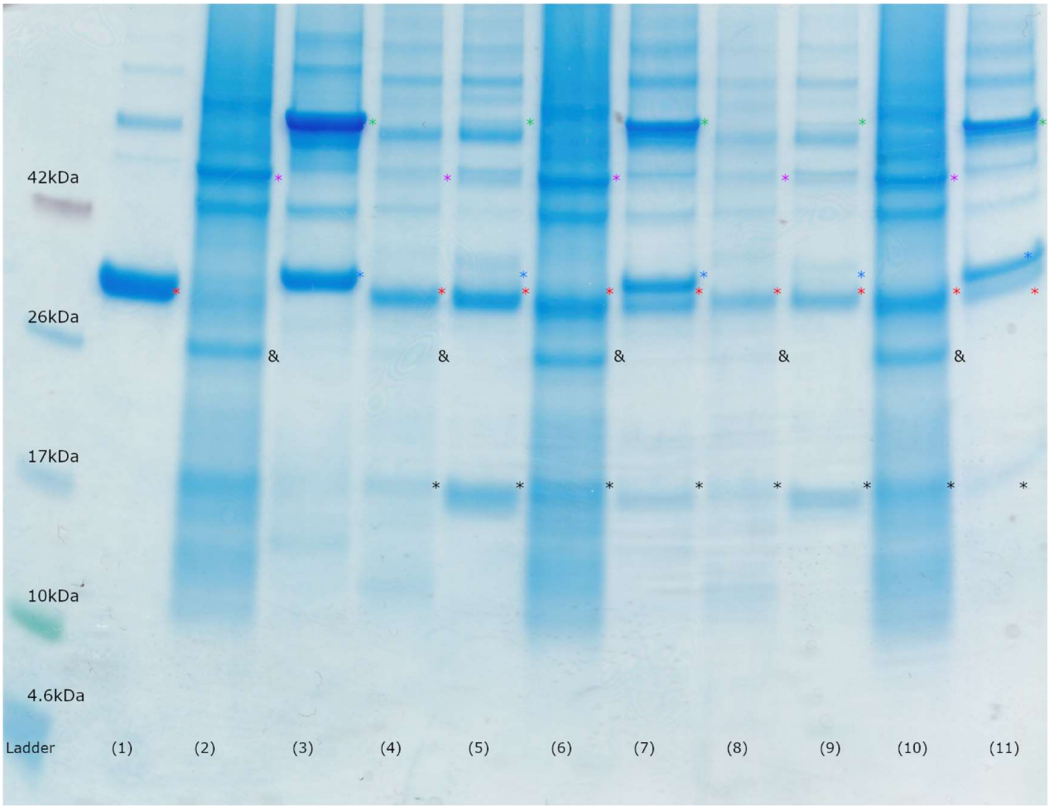


Figure 41: SDS-Page of TEVp cleavage protein components and reactions on KSI[N, Variant 1] and KSI[C, Variant 2] under various conditions. Identification of the contents of each well is as follows: (1) TEVp, (2) KSI[N, Variant 1], (3) KSI[C, Variant 2], (4) Cleavage of KSI[N, Variant 1] performed at 4°C for 24 hours, (5) Cleavage of KSI[C, Variant 2] performed at 4°C for 24 hours, (6) Cleavage of KSI[N, Variant 1] performed at room temperature for 48 hours, (7) Cleavage of KSI[C, Variant 2] performed at room temperature for 48 hours, (8) Cleavage of 10-fold diluted KSI[N, Variant 1] performed at room temperature for 48 hours, (9) Cleavage of 10-fold diluted KSI[C, Variant 2] performed at room temperature for 48 hours, (10) Cleavage of KSI[N, Variant 1] performed at room temperature for 24 hours, (11) Cleavage of KSI[C, Variant 2] performed at room temperature for 24 hours. Putative identifications of specific bands have been highlighted. Black asterisks indicate cleaved KSI-tag dimer, red asterisks indicate TEVp, blue asterisks indicate uncleaved KSI[C, Variant 2] monomer, ampersands indicate uncleaved KSI[N, Variant 1] monomer, purple asterisks indicate uncleaved KSI[N, Variant 1] dimer and green asterisks indicate uncleaved KSI[C, Variant 2] dimer.

Based on the SDS-page of the cleavage reactions, the reactions proceeding at 4°C produced the most complete cleavage and the highest yield of properly cleaved product in KSI[C, Variant 2], though no KSI[N, Variant 1] cleavage could be observed in any reaction (Figure 41). It was hoped that the SDS-page would produce distinct bands for cleaved and uncleaved KSI[N, Variant 2], though this was not the case. It is unclear as to whether the cleaved KSI[N, Variant 2] could not be differentiated from the uncleaved fusion protein, whether cleavage failed, or proceeded to completion as the SDS-gel can't distinguish between these results.

TEVp also cleaved sequences outside of the inserted cleavage site in both fusion proteins more frequently at 4°C than other cleavage conditions. No reaction condition displayed any further activity after 24 hours, likely due to inactivation of the TEVp. Due to the difficulty expected in separating the cleaved YthA fragments from peptides and other cleavage products, we saw it as preferable that a solution contain large, uncleaved proteins which might be recovered for another reaction than smaller degradation products which are more difficult to separate from YthA fragments. Reducing the reaction time below 24 hours or reducing the TEVp concentration in a 4°C cleavage reaction might lead to preferable results.

Due to time limitations further experiments became impossible to perform.

3 Concluding Remarks

The final results of the thesis work were the production of mostly pure KSI[C, Variant 1] and KSI[N, Variant 1] yields per litre of growth media in excess of 300mg/L, well above the expected yields of similar fusion proteins. We successfully purified fusion proteins GB1[N] and KSI[C, Variant 1] by HPLC was also achieved. Finally, successful solubilization and refolding of GB1[N] and KSI[C, Variant 1] was achieved.

In the continued work to produce pure YthA in sufficient quantities to determine its structure, priority should be given to solubilizing produced IBs, as precipitation and difficulty solubilizing the fusion proteins was consistently the greatest difficulty faced in this work. Different solubilization agents and techniques than 6M GHCl and 8M urea should be examined than those described here, as though these solubilizing agents were largely successful, precipitation occurred and difficulty was encountered solubilizing the IBs. Refolding of GB1[N] and KSI[N, Variant 1] were also successful with yields similar to those described in literature (Singhvi & Panda, 2022). Compared to all the difficulties and failures we have had with NiNTA YthA fusion protein purification, size exclusion chromatography should be assessed due the ease of use and success of TEVp purification by this technique.

While the author of this thesis work will be unable to continue the research presented here, we believe it has produced a solid foundation for further experiments and devised protocols and construct designs which will benefit further work with YthA protein production and similar studies.

4 Methods

4.1 Modelling the Structure of YthA and PspC from Sequence Data

TOPCONS and SWISS-MODEL services were used to predict YthA and PspC conformation and orientation relative to the membrane. No specific structure was chosen as a template for YthA in TOPCONS analysis.

4.2 Construct Design

```
         10         20         30         40         50         60
MSQRQLTKSV TNRRVSGVIA GIAEYFGLGR DVVTILRILF VVLAFGSWGG LIPLYFVASW

          70         80         90        100        110        120
/IIPSARPRNY YDDSEDDYQE KWNRKAQHFD EKMDRWSERY SDKMNNWARR YEDKGRQNQQ

        130        140        150
DSNQWGNPWD EPKSRKTKEA QPVEKEKEDD WSDF
```

Figure 42: Protein sequence of YthA (Bolotin et al., *2001), the '/' indicates where the sequence was split into C- and N-terminal fragments.*

YthA was split into two fragments between W60 and I61 (Figure 42). Four constructs were designed for each peptide, each construct containing a sequence coding for a His-tag, Tobacco Etch Virus (TEV) cut site and a Factor Xa cleavage site. DNA coding for a His_{10}-tag, Protein GB1 domain (GB1), TEV cut site, Factor Xa cut site, C-terminal *YthA* fragments and termination sequence were inserted into a pET-22b(+) vector between the 220 (Nco I restriction site) and 288 (Nde I restriction site) in that order. The construct was termed the C-terminal GB1 construct. The construct termed the C-terminal His-only construct created using a similar design, except that no GB1 was present, and the construct was inserted into a pET-16b(+) vector at position 331 (Nde I restriction site). The C-terminal KSI construct was created by inserting a factor Xa cut site, C-terminal *YthA* fragment, and a TEV cut site into a pET31b(+) vector at the Xho I restriction site (position 158) and BlpI restriction site (position 80). A variation of the C-terminal KSI construct was also designed by inserting a His_{10}-tag coding sequence followed by a TEV cleavage site coding sequence and the C-terminal *YthA* sequence into the pET-31b(+) vector and including a termination sequence. The N-terminal constructs were created using the same vectors, designs and vector insertion sites as their C-terminal counterparts, except that the C-terminal *YthA* sequence was replaced with an N-terminal *YthA* sequence. The vector constructs were ordered from Genscript. The MBP constructs were not produced. The fusion proteins produced by the vectors are given in Figure 43.

KSI[N, Variant 1]
N-Terminal YthA KSI-tagged Construct
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVEDPFGQPPIHGREQIAAFYRQGLGGGKVRACLTGPVRASHNGCGAMPFRVEMV
WNGQPCALDVIDVMRFDEHGRIQTMQAYWSEVNLSVRECQMLLEIEGRMSQRQLTKSVTNRRVSGVIAGIAEYFGLGRDVVTILRILFVVLAFGS
WGGLIPLYFVASWENLYFQSHHHHHH

KSI[N, Variant 2]
N-Terminal YthA KSI-tagged Construct (Internal His-tag)
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVEDPFGQPPIHGREQIAAFYRQGLGGGKVRACLTGPVRASHNGCGAMPFRVEMV
WNGQPCALDVIDVMRFDEHGRIQTMQAYWSEVNLSVRECQMLLEHHHHHHHHHHENLYFQSMSQRQLTKSVTNRRVSGVIAGIAEYFGLGR
DVVTILRILFVVLAFGSWGGLIPLYFVASW

GB1[N]
N-Terminal YthA GB1-tagged Construct
MHHHHHHHHHHNSSSNNNNQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNG
VDGEWTYDDATKTFTVTEENLYFQSIEGRMSQRQLTKSVTNRRVSGVIAGIAEYFGLGRDVVTILRILFVVLAFGSWGGLIPLYFVASW

MBP[N]
N-Terminal YthA MBP-tagged Construct
MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYA
QSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK
YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAG
INAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPNSSSNNNNNNNNNNLGENLYFQSIEGR
MSQRQLTKSVTNRRVSGVIAGIAEYFGLGRDVVTILRILFVVLAFGSWGGLIPLYFVASW

His[N]
N-Terminal YthA His-tag only Construct
MGHHHHHHHHHHENLYFQSIEGRMSQRQLTKSVTNRRVSGVIAGIAEYFGLGRDVVTILRILFVVLAFGSWGGLIPLYFVASW

KSI[C, Variant 1]
C-Terminal YthA KSI-tagged Construct
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVEDPFGQPPIHGREQIAAFYRQGLGGGKVRACLTGPVRASHNGCGAMPFRVEMV
WNGQPCALDVIDVMRFDEHGRIQTMQAYWSEVNLSVRECQMLLEIEGRIIPSARPRNYYDDSEDDYQEKWNRKAQHFDEKMDRWSERYSDK
MNNWARRYEDKGRQNQQDSNQWGNPWDEPKSRKTKEAQPVEKEKEDDWSDFENLYFQSHHHHHH

KSI[C, Variant 2]
C-Terminal YthA KSI-tagged Construct (Internal His-tag)
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVEDPFGQPPIHGREQIAAFYRQGLGGGKVRACLTGPVRASHNGCGAMPFRVEMV
WNGQPCALDVIDVMRFDEHGRIQTMQAYWSEVNLSVRECQMLLEHHHHHHHHHHENLYFQSIIPSARPRNYYDDSEDDYQEKWNRKAQHFD
EKMDRWSERYSDKMNNWARRYEDKGRQNQQDSNQWGNPWDEPKSRKTKEAQPVEKEKEDDWSDF

GB1[C]
C-Terminal YthA GB1-tagged Construct
MHHHHHHHHHHNSSSNNNNQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNG
VDGEWTYDDATKTFTVTEENLYFQSIEGRIIPSARPRNYYDDSEDDYQEKWNRKAQHFDEKMDRWSERYSDKMNNWARRYEDKGRQNQQDS
NQWGNPWDEPKSRKTKEAQPVEKEKEDDWSDF

MBP[C]
C-Terminal YthA MBP-tagged Construct
MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYA
QSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK
YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAG
INAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPNSSSNNNNNNNNNNLGENLYFQSIEGRII
PSARPRNYYDDSEDDYQEKWNRKAQHFDEKMDRWSERYSDKMNNWARRYEDKGRQNQQDSNQWGNPWDEPKSRKTKEAQPVEKEKEDD
WSDF

His[C]
C-Terminal YthA His-tag only Construct
MGHHHHHHHHHHENLYFQSIEGRIIPSARPRNYYDDSEDDYQEKWNRKAQHFDEKMDRWSERYSDKMNNWARRYEDKGRQNQQDSNQW
GNPWDEPKSRKTKEAQPVEKEKEDDWSDF

Figure 43: Designations, descriptions, and sequences of YthA fusion proteins. Yellow highlighting indicates a Factor Xa cleavage site, cyan highlighting indicates a TEV cleavage site, green highlighting indicates YthA fragment, dark grey highlighting indicates a linker sequence, red highlighting indicates a His-tag and light grey highlighting indicates a production tag.

4.3 Preparation of Stock Solutions, Buffers and Growth Media

4.3.1 0.2M, 7.34pH Phosphate Stock Buffer Solutions

17.8g $Na_2HPO_4 \bullet 2H_2O$ was dissolved in MilliQ water to a total volume of 500mL to form a 0.2M basic stock solution. 15.7g of $NaH_2PO_4 \bullet 2H_2O$ was dissolved in MilliQ water to a total volume of 500mL to form a 0.2M acidic stock solution. 100mL of acidic stock solution was added to 300mL of basic stock solution, then pH adjusted to a value of 7.4pH using NaOH or HCl as appropriate, to form a 0.2M stock buffer solution.

4.3.2 1M Tris(hydroxymethyl)aminomethane (Tris) Stock Solutions

60.568g Tris-HCl powder was added to a 0.5L flask and MilliQ water was added to a total volume of 0.45L. The solution was vigorously shaken to mix and pH was adjusted with NaOH or HCl as appropriate to reach pH 8. MilliQ water was added to a total volume of 0.5L.

4.3.3 0.1g/ml Ampicillin Stock Solution

In a chemical hood, 1.00g of ampicillin powder and 9ml of MilliQ water was added to a 15mL falcon tube and the solution vortexed on high until the ampicillin was completely dissolved. MilliQ water was then added to a total volume of 10.0mL. The 0.1g/ml solution was aliquoted in 1.0mL fractions in 1.5mL microcentrifuge tubes and stored at -4^{o}C until use.

4.3.4 1M Isopropyl β-D-1-Thiogalactopyranoside (IPTG) Stock Solution

In a chemical hood, 4.77g of IPTG powder was added to a 50mL falcon tube. 15mL MilliQ water was added to the powder and vortexed until the IPTG was completely dissolved. MilliQ was then added to a total volume of 20mL and the mixture was mixed briefly. The solution was then aliquoted into 20 1mL aliquots and stored at -20°C until use.

4.3.5 4M Sodium Chloride Stock Solution

116.9g of NaCl powder was added to 0.4L of MilliQ water and shaken vigorously to dissolve the NaCl. MilliQ water was added to a total volume of 0.5L and again shaken vigorously.

4.3.6 0.5M $CaCl_2$ Solution

13.87g of $CaCl_2$ powder was added to a 250mL reagent flask with 200mL MilliQ water. The mixture was shaken until the $CaCl_2$ dissolved completely. MilliQ water was added to a total volume of 250mL.

4.3.7 5x M9 Stock Salt Solution

15g $KHPO_4$, 41.3g $Na_2HPO_4 \bullet 2H_2O$, 2.5g NaCl, and 5.0g NH_4Cl were added to 900mL MilliQ water and stirred by magnetic stir rod for 2 hours to dissolve. MilliQ was added to a volume of 1L and briefly stirred to mix evenly.

4.3.8 Lysis Buffer

10mL of 0.2M, 7.4 pH $NaPO_4$ buffer solution was added to 12.5mL NaCl stock solution and 25mL 40% glycerol solution. MilliQ water was added to a volume of 100mL before being mixed again. The solution was used or disposed of within one month of assembly.

4.3.9 Initial IB Wash Buffer

The initial wash buffer was prepared using the same procedure as the lysis buffer described in section 4.3.8 with the exception that 20mL of 10% Triton-X100 solution was added prior to the addition of MilliQ water to a total volume of 100mL.

4.3.10 Second IB Wash Buffer

The second wash buffer was prepared using the same procedure as the lysis buffer described in section 4.3.8 with the exception that no glycerol was added to the solution. MilliQ water was still added to a total volume of 100mL.

4.3.11 10M Urea Stock Solution

60.06g of urea crystals was added to 60mL of MilliQ water and mixed vigorously. A magnetic stir-rod was added to the solution and placed on a heated stir plate set on medium-high temperature and 5/10 stir. MilliQ water was added to a total volume of approximately 95mL. The crystals and water were gently heated and stirred vigorously for 1 hour in a sealed container to form a solution. MilliQ water was added to a final volume of 100mL.

4.3.12 1M Glycine Stock Solution

In a sterile biohood and in a sterilized reagent flask, 3.754g of glycine powder was added to 90mL MilliQ water and inverted repeatedly to mix until the glycine was completely dissolved. MilliQ water was then added to a total volume of 100mL. The solution was used or disposed of within one month of mixing.

4.3.13 2M Imidazole Stock Solution

13.6154g of Imidazole was added to a 100mL flask, 90mL MilliQ water was added and the mixture was shaken vigorously to dissolve. MilliQ water was added to a total volume of 100mL.

4.3.14 HPLC Loading Buffer

360.36g of urea; 3.7535g of glycine; 10mL of 2M Imidazole stock solution; 125mL of 4M NaCl stock solution; 100mL of 0.2M, 7.4pH phosphate buffer; was added to a 1L flask and MilliQ water was added to a total volume of 900mL. The mixture was swirled vigorously until the components were dissolved. pH adjustment of the solution was performed using NaOH or HCl as appropriate to reach a pH of 7.4. MilliQ water was added to a total volume of 1L.

4.3.15 Low-Imidazole HPLC Loading Buffer

Low-imidazole HPLC loading buffer was produced exactly as written in section 4.3.14, except that 2.5mL of 2M imidazole stock solution was added in place of the 10mL of 2M imidazole solution.

4.3.16 Refolding Buffer

Refolding buffer was prepared as the loading buffer in section 4.3.14, except that no glycine and no urea was added to the buffer. Similar to the preparation of HPLC loading buffer in section 4.3.13, volume was brought to a total of 900mL before pH adjustment and to a total of 1L after pH adjustment.

4.3.17 Low-Imidazole Refolding Buffer

Low-imidazole refolding buffer was produced exactly as written in section 4.3.16 except that 2.5mL of 2M imidazole stock solution was added in place of the 10mL of 2M imidazole solution.

4.3.18 HPLC Elution Buffer

Elution buffer was prepared as the loading buffer in section 4.3.14 except that 250mL of 2M imidazole stock solution was added to the buffer instead of 10mL. Volume was still brought to a total of 900mL before pH adjustment and to a total of 1L after pH adjustment.

4.3.19 Digestion Buffer

12.5mL of 4M NaCl stock solution, 50mL 0.2M 7.4pH phosphate buffer, 5mL 0.5M $CaCl_2$ stock solution and 432.5mL MilliQ water was added to a 0.5L flask and shaken vigorously to mix.

4.3.20 10x Tobacco Etch Virus Protease (TEVp) Reaction Buffer

0.5mL 0.5M ethylenediaminetetraacetic acid (EDTA) was added to 25mL of 1M 8pH Tris-HCl buffer stock (from section 4.3.2) then brought to a total volume of 45mL with MilliQ water. pH was adjusted to 8 with KOH or HCl as appropriate. MilliQ water was added to a total volume of 50mL.

4.3.21 Tris-based buffers

Lysis (4.3.8), initial wash (4.3.9), second wash (4.3.10), loading (4.3.14 and 4.3.1d), refolding (4.3.16 and 4.3.17), elution (4.3.18), and digestion (4.3.19) buffers were also produced using Tris. Lysis, initial wash, and second wash buffers substituted 2mL of 1M, pH 8 Tris buffer in place of 10mL phosphate buffer. 20mL of 1M, pH 8 Tris buffer replaced 100mL phosphate buffer in loading, refolding and elution buffers. Digestion buffer was made with 10mL of 1M, pH 8 Tris buffer in place of 50mL phosphate buffer. All Tris buffers were adjusted to pH 8 using NaOH or HCl as appropriate.

4.3.22 Guanidine Hydrochloride (GHCl) Buffer

286.59g GHCl was added to a 0.5L reagent flask with 62.5mL of 4M NaCl, 10mL of 1M Tris HCl, 1.25mL of 2M imidazole, and 35.2µL of 2-mercaptoethanol. MilliQ water was added to a volume of 450mL and swirled vigorously to mix. MilliQ water was added to a volume of 0.5L.

4.3.23 10x Phosphate-Buffered Saline (PBS) Stock Solution

80g NaCl, 2g KCl, 14.4g Na_2HPO_4 and 2.4g KH_2PO_4 were added to a 1L reagent flask and MilliQ water was added to a volume of 800mL. The solution was pH-adjusted to 7.4 (for a final diluted pH of 7.2). Additional MilliQ water was added to a volume of 1L.

4.3.24 1x PBS 6M Urea solution

10mL of 10x PBS stock from section 4.3.23 was added to 60mL of 10M Urea stock solution from section 4.3.11 and MilliQ water added to a volume of 100mL.

4.3.25 LB media

LB media was prepared by addition of 10g each of meat peptide and NaCl, and 5g yeast extract with approximately 0.9L of MilliQ water. The mixture was shaken vigorously until the solutes were dissolved. MilliQ water was then added to a total volume of 1L. Following this, the media was autoclaved as described in section 4.4.3. The media was stored at 4^{o}C until needed and was used or disposed of within 2 months.

4.3.26 ^{15}N-Labelled M9 Minimal Media

5x M9 salt stock solution, 1M $MgSO_4$ stock solution, and 0.5M$CaCl_2$ stock solution, in addition to all handheld tools and glassware used in the preparation of ^{15}N-labelled M9 minimal media were sterilized by autoclave as described in section 4.4.3. 200g/L D-glucose stock solution was sterilized by vacuum filtration as described in section 4.4.4. 200mL 5x M9 salt stock solution, 20mL 200g/L D-glucose stock solution, 10mL 100x minimum essential medium (MEM) stock solution, 2mL 1M $MgSO_4$ stock solution, 200µL 0.5M $CaCl_2$ stock solution, and 20mL 50g/L $^{15}NH_4Cl$ stock solution were added to 700mL MilliQ water and mixed by magnetic stir rod to mix. MilliQ water was added to a total volume of 1L. ^{15}N-labelled M9 minimal media was stored at 4°C until used or disposed of within 2 weeks.

4.3.27 Agar Plates

4.3.27.1 Agar Plates Containing No Antibiotics

Agar plates were prepared by mixing 2g of powdered agar with 100mL of sterile LB media. The mixture was then microwaved on high, stirring every 30 seconds, until all the agar had dissolved. In a sterile biosafety hood the mixture was poured into plastic petri dishes and allowed to cool to room-temperature. The plates were covered with the plastic lid, placed into a resealable plastic bag and stored at 4°C. The plates were used within 2 weeks or disposed of.

4.3.27.2 Preparation of Ampicillin-Containing Agar Plates

Ampicillin agar plates were produced as on 4.3.27.1 except that 0.1mL 100mg/L ampicillin was added when the solution cooled to a temperature between 50°C and 60°C.

4.4 Solution Treatments

4.4.1 Cooling Solutions

4.4.1.1 Cooling Solutions to Ice-Cold Temperatures

Solutions to be cooled ice-cold were cooled by aliquoting no more than 20mL of solution into a 50mL falcon tube and storing it at -20°C for 1 hour. Solutions of up to 250mL in volume were cooled by storing at -20°C for a minimum of 2 hours. Solutions were kept on ice after removal from -20°C storage during thawing and use.

4.4.1.2 Cooling Solutions to 4°C

Solutions to be cooled to 4°C were stored at 4°C overnight and were kept at 4°C until used, disposed of, or no longer required to be at 4°C.

4.4.2 Degassing Solutions

Solutions were degassed by pressure reduction using a water-powered vacuum degassing apparatus. The solution was cooled to 4°C as described in section 4.4.1.2. The apparatus was attached to the flask containing the solution and the faucet. The apparatus was operated for 2 hours per litre of solution, during which a stir-rod was mixing the solution at medium-high intensity. The solution was considered to be degassed for 2 weeks after degassing if it remained sealed or for 2 days after unsealing the solution.

4.4.3 Sterilizing Solutions by Autoclave

The autoclaved was prepared by ensuring the water level was within the acceptable parameters and that a thin layer was covering the bottom of the autoclave chamber. The solutions were added to the autoclave chamber and the autoclave chamber sealed. The autoclave was set to 120°C and turned on. The autoclave reached was allowed to reach 120°C for 20 minutes and begin cooling. The chamber was unsealed once the temperature had fallen to below 60°C. The solutions were considered sterile for 2 months or until 1 month after it was first opened following autoclaving.

4.4.4 Filtration by Vacuum Filter

A water-powered vacuuming apparatus was attached to the filter. MilliQ water was added to the filter until the entire surface of the filter was wet. The vacuuming apparatus was turned on briefly to flush the MilliQ water through. A flask of the appropriate volume was attached to the filter and the solution to be filtered was poured slowly into the filter until the filter was almost full and the vacuum apparatus turned back on. Solution was poured into the filter at a rate such that the filter never overflowed or went dry until all the solution was filtered.

4.5 Determining Optical Density

4.5.1 Determining Optical Density at 600nm (OD_{600})

OD_{600} was determined using one plastic cuvette loaded with 1mL of sample and two cuvettes were loaded with 1mL each of blank; the same solution lacking the substrate to be measured. The blanks were loaded into the spectrophotometer. The parameters were set to measure at a wavelength of 600nm and the spectrophotometer zeroed. The blank was removed from the experimental sample reader and replaced with the experimental sample and the OD_{600} recorded.

4.5.2 Determining Protein Yield by OD_{280}

4.5.2.1 Preparation of Cuvettes

The interior and exterior of quartz cuvettes were rinsed twice each with water, then twice each with rectified spirit, ensuring that the clear sides of the cuvettes are free of any dirt, smudges, or scratches. The cuvettes were allowed to dry completely at room temperature.

4.5.2.2 Measuring OD_{280}

OD_{280} was measured similarly to OD_{600} except that quartz cuvettes prepared as described in 4.5.2.1 were used instead of plastic cuvettes and that the spectrophotometer was set to measure a wavelength of 280nm.

4.5.2.3 Calculating Protein Concentration

Equation 1 was used to determine the estimated concentration of protein in a solution.

$$\frac{A}{\varepsilon \times l}$$

Equation 1: Equation for the calculation of dissolved protein concentration in a sample. A is the measured absorbance of the sample at 280nm, ε is the extinction coefficient of the protein at 280nm and l is the path length.

4.6 Sonication Procedure

The sample was kept on ice for the duration of the procedure. The microtip probe was attached securely to the sonicator. The probe was inserted into the sample without touching the walls of the tube containing the sample. The sonicator was set to 39% intensity with a cycle of 10 seconds on followed by 20 seconds rest and repeated for the duration of the sonication.

4.7 Centrifugation Procedure

The ultra-centrifuge was set to spin at 23 000g at 4°C for 20 minutes and allowed to cool to 4°C if necessary. The samples were inserted into the ultra-centrifuge, ensuring that the rotor was balanced, using dummies consisting of water if necessary.

4.8 Cell Growth and Expression

4.8.1 Preparative Culture Growth

4.8.1.1 Preparative Culture Growth for Antibiotic-Susceptible Cells

In a sterile bio-hood, a scraping of DH5α or BL21(DE3) *Escherichia coli* (*E. coli)* cells from frozen cell cultures was taken using a sterile pipette tip. Alternatively, if the preparative culture was produced from plated cells, then a single, distinct colony was chosen from the plate and gently scraped using a sterile pipette tip in a sterile bio-hood. Regardless of the source of the cells, the scraping was inoculated into 10mL of LB media in a sterile 50mL falcon tube with a loose-fitting lid taped on. The inoculated media was placed into a shaking incubator for 16 hours at 37°C and 250 rpm.

4.8.1.2 Preparative Culture Growth for Ampicillin-Resistant Cells

Ampicillin-resistant preparative cell cultures were produced as 4.8.1.1 except that the scraping of cells was taken from an ampicillin-resistant frozen or plated culture, and 10µL of 100mg/L ampicillin was added to the 10mL of LB media prior to the inoculation with the cell scraping.

4.8.2 Generating Competence in *E.coli* Cells

0.5mL DH5α or BL21(DE3) *E. coli* preparative culture was transferred into 50mL of LB media and grown at 37°C and 250rpm, testing OD_{600} until it is approximately 0.4. Centrifugation was performed in a sterile tube at 5 000g at 4°C for 10 minutes and the supernatant discarded. The cell pellet was immediately resuspended in 20mL of sterile, ice-cold 50mM $CaCl_2$ solution and incubated for 20 minutes. Following this incubation, the suspension was centrifuged at 5 000g at 4°C for 10 minutes and the supernatant discarded. The cells were immediately resuspended in 2.5mL of sterile, ice-cold 50mM $CaCl_2$, 10% glycerol solution then split into 100µL aliquots and used immediately or stored at -80°C.

4.9 Cell Transformation and Incubation Test

4.9.1 Cell Transformation

In a biosafety hood, 2 100µL aliquots of competent cells purchased or created as in 4.8.2 were thawed on ice, then 2.5µL of vector solution (5ng of DNA) were added to each aliquot and gently swirled to mix with the cells. The cells were incubated on ice for 15 minutes. Immediately following this, the cells were heat-shocked in a water-bath at 42°C for exactly 45 seconds and then placed on ice for 2 minutes. 0.9mL room temperature SOC media was added to each aliquot.

4.9.2 Plating of Transformed Cells

A spreading tool consisting of an 'L'-shaped piece of glass was submerged in rectified spirit to sterilize the tool and ignited to remove the spirit. The tool was allowed to cool before use. 100µL of the sample from section 4.9.1 was added to an agar plate containing LB-100µg/mL ampicillin. The spreading tool was used to spread the cell solution evenly over the plate by placing the bottom of the tool on the plate with the end of the tool touching the wall of the plate and the tool facing the centre of the plate. The plate was then rotated to spread the cells.

4.9.3 Comparison of Shaking Incubation to Non-Shaking Incubation

The sample from section 4.9.2 (Sample 1) was immediately placed into a 37°C, non-shaking incubator for 1 hour. Another 100µL sample from section 4.9.1 was taken (Sample 2) and was incubated for 1 hour in a 225rpm shaking incubator at 37°C prior to re-sterilizing the spreading tool and being spread on another LB-100µg/mL ampicillin agar plate.

4.9.4 Incubation of Inoculated Agar Plates

Plates were then incubated in a 37°C non-shaking incubator overnight before qualitative visual assessment was performed on Samples 1 and 2 to determine whether the shaking incubation step provided any benefit. Plates were kept at 4°C for 1 week after removal from the incubator, before being disposed of.

4.10 Preparing Cell Cultures for -80°C Storage

Any culture to be frozen was first grown as a preparative culture with regular spectrophotometry checks at 600nm until an OD_{600} of at least 0.6 was achieved. The culture was centrifuged at 600g for 5 minutes at 4°C. Media was then decant and discarded. 1.5mL of lysis buffer as described in section 4.3.8 was added to the cell pellet and gently mixed. The cell solution was transferred to a freezer safe tube 2mL tube and immediately stored at -80°C.

4.11 Induction and Expression Culture Growth

Preparative culture was added in a 1:100 ratio to inoculate 0.5L LB-amp media in a 2L baffled flask. OD_{600} was measured at 2 hours after inoculation and every 20 minutes thereafter until an OD_{600} of approximately 0.5 was obtained. The culture was induced with 0.5mL 1M isopropyl β-d-1-thiogalactopyranoside (IPTG) and inserted into the incubator at either 18°C or 37°C and induction continued for overnight or for 4 hours respectively. Cells grown at 37°C had their OD_{600} measured as described in section 4.5.1 every 30 minutes until induction and again 3-4 hours after induction.

4.12 Harvesting of Protein-Producing Cells

Cell cultures were transferred to 370mL centrifugation bottles prior to centrifuging at 12 000g, 4°C for 20 minutes. Spent media was decanted, and the cell paste transferred to a pre-weighed 50mL sterile falcon tube. The paste was weighed to determine cell yield. The paste was then processed immediately or stored at -20°C.

4.13 Cell Lysis

4.13.1 Resuspension of the cells in lysis buffer

10mL of lysis buffer was mixed with a protease inhibitor tablet by vortexing on high. From this point onward, all work with the lysis buffer was performed on ice. The lysis buffer was then added to the cell paste from 0.5L of LB media and the cells resuspended by vortexing of the falcon tube containing the cells.

4.13.2 Cell lysis by Sonication

The lysis buffer containing cells from section 4.13.1 underwent sonication as described in section 4.6.

4.13.3 Cell Lysis by French Press

Components of the French press were sterilized with an 80% alcohol solution, dried, and reassembled. The resuspended cell solution from section 4.13.1 was passed through the French press at 1250 pounds per square inch (PSI), approximately 17500 PSI cell pressure.

4.13.4 Separation of Particulate and Soluble Fractions

Regardless of the method used to lyse the cells, the lysate from sections 4.13.2 or 4.13.3 was centrifuged at 22 000rpm at 4°C for 15 minutes. The supernatant was decanted and retained, and the pellet remained in the centrifugation tube. Samples were either stored at -20°C or processed further immediately.

4.13.5 Test for Protease Inhibitor Necessity

Cell pellet produced from section 4.13.2 or 4.13.3 was dissolved into the supernatant by high intensity vortexing, then split into two 4.5mL fractions. The fractions then underwent another cycle of centrifugation (section 4.7). Following this, the fractions underwent resuspension, sonication and centrifugation as described in 4.13.1, 4.6 and 4.7 respectively 3 additional times, except one of the samples was not treated with any protease inhibitor during the 3 repetitions of the procedure. Both samples were centrifuged as described in section 4.7. Supernatant was from the final centrifugation step was retained. The pellet samples were then resuspended in 5mL of initial wash buffer (4.3.9) by sonication (4.13.2), and again centrifuged as described in section 4.7. Wash buffer was again decanted and retained, and the pellet resuspended in 3mL of lysis buffer.

4.14 Performance of SDS Page

4.14.1 Preparation of Samples

4.14.1.1 Preparation of Samples in Pellet Form

Samples were thawed on ice if necessary. 75µL of lysis buffer as described in section 4.3.8 was mixed with 25µL of 4x LDS buffer to make a 1x LDS buffer solution. A scraping of the sample was taken with a pipette tip and added to the 1x LDS buffer solution. Following this, the sample was vortexed until solids were dissolved and suspended. Samples were heat-shocked at 95°C for 5 minutes and centrifuged briefly in a small desktop centrifuge, inverted repeatedly to mix, and were centrifuged in a small desktop centrifuge again.

4.14.1.2 Preparation of Samples in Solution

Liquid samples were prepared by adding 25µL 4x LDS buffer to 75µL sample. The samples were vortexed, heat-shocked, and spun as described in 4.14.1.1

4.14.2 Loading of Electrophoresis Apparatus and Gel

An SDS gel was loaded onto an electrophoresis apparatus and the outer section filled with 1x SDS running buffer to loading level and inner section completely filled. The apparatus was set to 200V, 90mA to run for 40 minutes. Avoiding spillage, 5µL of protein ladder was loaded into a well by pipette and 20µL of each sample loaded into its own well before the protocol was executed.

4.14.3 Staining and Destaining SDS Gel

The SDS gel was removed from the electrophoresis apparatus and its plastic casing. The gel was gently laid in a container without folding or wrinkling the gel and 50mL instant-blue stain was added. The container was placed on a shaking table for 4 hours. The stain was decanted and replaced with 50mL distilled water, the container was placed on shaking table overnight.

4.14.4 Imaging SDS Gel

The gel was then laid on transparent plastic film and covered with the plastic film. Any bubbles present were massaged out from between the gel and film. The exposed areas of the film were then dried thoroughly with a paper towel before the gel was loaded onto a photo-scanner and a picture of the gel was taken.

4.15 Detection of C-terminal Fusion Proteins in Soluble and Insoluble Fractions

KSI[C, Variant 1], GB1[C], and His[C] proteins were produced at 37°C as described in sections 4.11, harvested as described in section 4.12, and sonicated as described in sections 4.6. Soluble and pelleted fractions were prepared for analysis by SDS-page as described in section 4.14.1.

4.16 Solubility Testing of N-Terminal GB1 YthA Construct

4.16.1 Preparation of the Protein Pellets

Protein pellets from lysed N-terminal GB1 YthA producing cells were thawed on ice. 4mL of 20mM 7.4pH phosphate buffer was added to the sample sonicated as described in section 4.6 until the pellet was fully resuspended in the buffer. The suspension was then split into 6 1mL samples and the samples were centrifuged as described in section 4.7 and the supernatant discarded.

4.16.2 Determining Minimum Urea Concentration Necessary for Solubilization

Testing condition solutions were mixed in the proportions listed in Table 7.

Table 7: Reagents mixed together to produce the listed urea and TFE concentrations for use in solubility tests.

Final Urea Concentration (M)	Final TFE Concentration (%)	99% TFE Added (mL)	10M Urea Added (mL)	1M Glycine Added (mL)	7.4 pH 0.2M Phosphate Buffer Added (mL)	MilliQ Water Added (mL)
1.5	30	1.5	0.6	0.2	0.4	1.3
2	0	0	0.8	0.2	0.4	2.6
3	30	1.5	1.2	0.2	0.4	0.7
4	0	0	1.6	0.2	0.4	1.8
6	0	0	2.4	0.2	0.4	1.0
8	0	0	3.2	0.2	0.4	0.2

The solutions were made by adding 0.2mL 1M glycine and 0.4mL 7.4pH, 0.2M phosphate buffer to the appropriate volume of MilliQ water and swirled gently to mix. An appropriate volume of 10M urea was then added to the solution and again swirled gently to mix. If trifluoroethanol (TFE) was to be added to the solution, it was added to the solution and swirled gently to mix.

4.16.3 Solubilization of the Protein Pellets in Experimental Solutions

4mL of each experimental solution from Table 7 was added to different protein pellets produced from 4.13.4. Each sample was then sonicated for 5 minutes as described in section 4.6. The suspensions were then centrifuged as described in section 4.7. An SDS page was performed on the soluble fractions as described on 4.14.

4.17 Testing the Effects of Vortexing on Proteins

The soluble fractions of the 3M urea with 30% TFE, 6M urea, and 8M urea from the solubility tests from section 4.16 were thawed on ice. 15µL of each sample was reserved and not vortexed. The remaining samples were vortexed on the highest setting for 5 minutes each. An SDS page was conducted on the vortexed and unvortexed samples as described in section 4.14.

4.18 Test for Protein Degradation in Denaturing Solutions

4.18.1 Preparation of Samples

N-terminal GB1 unpurified protein pellet was thawed on ice. 20mL 20mM, 7.4pH phosphate buffer was added to the pellet and sonicated as described in section 4.6 until the pellet was resuspended. The sample was then split into 2 10mL samples and centrifuged as described in section 4.7. The supernatant was discarded. 4mL of the 6M urea solution from 4.16.2 was added to one protein pellet, and 4mL of the 3M urea, 30% TFE solution from 4.16.2 was added to the other pellet. The samples were then again sonicated as described in section 4.6 for 5 minutes.

4.18.2 Test for Protein Degradation

Beginning immediately and for every 30 minutes for 4 hours, 75µL of each sample was taken and prepared for analysis by SDS-page as in section 4.14.1. The experimental samples were left in sterile bio-hood at room-temperature for the duration of the experiment. The SDS-prepared samples were stored at -20°C until the SDS-page was preformed.

4.19 Test for Uninduced Expression of Recombinant Proteins

4.19.1 Preparation of Cultures

A preparative culture of untransformed BL21(DE3) cells and a preparative culture of N-terminal GB1-transformed cells were prepared as described in section 4.8.1.1 and 4.8.1.2 as appropriate.

4.19.2 Test for Expression

A 0.5L expression culture was grown for each cell type as described in section 4.11, except that the media intended for the transformed cells contained ampicillin. 0.5mL of 1M IPTG was added to each culture when OD_{600} was approximately 0.5. A 1mL sample was immediately taken from each culture and prepared for SDS-page as described in section 4.14.1. The cultures were allowed to continue growing in a 37°C, 225rpm shaking incubator for 3 hours before the next sample was taken. An SDS-page was performed to test for expression.

4.20 Washing Protein Pellets

4.20.1 Initial Wash

Protein pellets were thawed on ice if necessary. 20mL of initial wash buffer as described in section 4.3.9 was added to the protein pellet. The pellet was sonicated as described in section 4.6 for 5 minutes and centrifuged as described in section 4.7. The supernatant was decanted.

4.20.2 Second Wash

Protein pellets from section 4.20.1 were washed as described in section 4.20.1, except that 20mL of second wash buffer as described in section 4.3.10 was used instead of 20mL initial wash buffer.

4.20.3 Third Wash

Protein pellets from section 4.20.2 were washed in 20mL second wash buffer as described in section 4.20.2.

4.21 NiNTA Column Purification

4.21.1 Preparation and Parameters

Loading (section 4.3.14) and Elution (4.3.18) buffers were filtered (section 4.4.4) and degassed (section 4.4.2), if not already and Buffer A and Buffer B input tubes were inserted into the respective buffers. The washed protein pellet was thawed, if necessary, and resuspended in 4mL of loading buffer by sonication as described in section 4.6. The machine was set to capture elution fractions in the test tubes, the equilibration flow-through in a Falcon tube and wash flowthrough in another 50mL Falcon tube. The output tube was positioned to capture the equilibration flowthrough into a 50mL Falcon tube. Akta superloop was filled with loading buffer and attached to HPLC machine. 4mL of loading buffer was injected into the superloop. Gas was removed from inside the machine with a 5mL syringe and a pump wash was performed on pumps A and B. The machine was set to the parameters listed on Table 8.

Table 8: HPLC column machine settings for HPLC purification

Parameter	Setting
Run Type	Ion Affinity
Column	HisTrap 5mL HP
Flow Rate	3mL/min
Wavelength Detected	280nm
Equilibration Volume	2 Column Volumes (CV) (10mL)
Buffer A	Loading Buffer
Buffer B	Elution Buffer
Injection Volume	4mL
Wash Volume	2CV (10mL)
Flowthrough Capture	11mL
Elution Fraction Size	1.5mL
Initial Gradient	0%
Final Gradient	100%
Gradient Length	15CV (75mL)
Cleaning Volume	5CV (25mL)
Re-equilibration Volume	5CV (25mL)

4.21.2 Insertion of Sample and Execution of the Protocol

4mL of sample was loaded into the Akta superloop by using a 5mL sterile syringe to force the sample through a 22µm filter and into the superloop. The protocol described in Table 8 was then executed on the injected sample.

4.21.3 NiNTA Column Purification Using Low-Imidazole Buffers

NiNTA column-based purification using low-imidazole buffers was performed as written in sections 4.21.1 and 4.21.2, except that the loading buffer was replaced with the low-imidazole loading buffer.

4.22 Protein Refolding

4.22.1 On-Column Refolding Using Standard Buffers

4.22.1.1 On-Column Refolding

Samples, buffers, columns and the machine were prepared, and the protocol run as described in sections 4.21.1 and 4.21.2, except that refolding buffer was used in place of elution buffer and no capture of flowthrough occurred.

4.22.1.2 Elution of Putatively Refolded Protein

Elution was performed as described in sections 4.21.1 and 4.21.2 except that no sample insertion occurred, and refolding buffer was used in place of loading buffer.

4.22.2 On-Column Refolding Using Low-Imidazole Buffers

On-column refolding with low-imidazole buffers was performed as described in section 4.21.1, except that the loading buffer was replaced with the low-imidazole loading buffer and the refolding buffer was replaced with the low-imidazole refolding buffer.

4.22.3 On-Column Refolding Using GHCl Buffer

On-column refolding with GHCl buffer was performed as described in section 4.21.1, except that the sample was resuspended in GHCl buffer and the column was equilibrated with GHCl buffer. Also, the loading buffer was replaced with the GHCl buffer and the refolding buffer was replaced with the low-imidazole Tris refolding buffer.

4.22.4 Rapid Dilution Refolding

Samples were thawed on ice, if necessary, and resuspended in 5mL 6M Urea PBS solution by sonication as described in section 4.6. The sample drip-fed by a sterile peristaltic pump at a rate of approximately 2mL/hour into 5mL PBS (x10) and 45mL MilliQ water which was continually stirred. 15mL of the rapidly diluted sample was loaded into a 10 000 MW cut-off Amicon ultracentrifugation tube and centrifuged at 4500g, 4°C for 5 minutes at a time, adding sample to the chamber and disposing of the flowthrough to ensure there was always between 7.5mL and 15mL sample in the tube during the entire process. The process was repeated until all sample was concentrated.

4.22.5 Refolding by Dialysis

Samples were thawed on ice, if necessary, and resuspended in 7mL 6M Urea PBS from section 4.3.22 by sonication as described in section 4.6. Resuspended protein was injected into a 10 000 MW cut-off dialysis apparatus and any air in the apparatus gently removed to prevent strain on the membrane. The sample was inserted into a flask of 750mL, 4°C, 0M Urea, x1 PBS solution. A stir bar was added to the flask and set to gently stir the solution. Dialysis proceeded for 2 hours at 4°C. The 0M Urea PBS solution was replaced with 750mL fresh x1 PBS solution and dialysis continued for 16 hours.

4.23 Buffer Exchange

4.23.1 Preparation of Sample, Buffer, Fraction Collector, Column and HPLC Machine

The buffer and sample were prepared exactly as described in section 4.21.1, except that only the buffer into which the protein is to be dissolved was filtered and degassed, and had Buffer input tube A inserted into it. The fraction collector was set up exactly as described in section 4.21.1, except that only one falcon tube was used, and the equilibration flowthrough was not retained. Preparation of the column and HPLC machine performed as described on 4.21.1, except that the super-loop was filled with the buffer the protein is to be suspended in, and the same buffer was injected into the HPLC machine. Also, the machine was set to the parameters in Table 9, not Table 8.

Table 9: HPLC column machine settings for buffer exchange

Parameter	Setting
Run Type	Anion Exchange
Column	HiTrap 5mL HP
Flow Rate	1mL/min
Wavelength Detected	280nm
Equilibration Volume	2CV (10mL)
Injection Volume	4mL
Wash Volume	2CV (10mL)
Flowthrough Capture	11mL
Fraction Size	1.5mL
Run Length	6CV (30mL)
Cleaning Volume	5CV (25mL)

4.23.2 Insertion of Sample and Execution of the Protocol

Insertion of the sample was performed exactly as described in section 4.21.2 prior to the execution of the protocol.

4.24 Size Exclusion Chromatography Protein Purification

The buffer and sample were prepared exactly as described in section 4.21.1, except that only the buffer into which the protein is dissolved in was filtered and degassed, and had Buffer input tube A inserted into it. The fraction collector was set up exactly as described in section 4.21.1, except that only one falcon tube was used, and the equilibration flowthrough was not retained. Preparation of the column and HPLC machine performed as described on 4.21.1. Also, the machine was set to the parameters in Table 10, not Table 8.

Table 10: HPLC column machine settings for size exclusion chromatography

Parameter	Setting
Run Type	Size Exclusion
Column	Acclaim™ SEC-300
Flow Rate	0.35mL/min
Wavelength Detected	280nm
Equilibration Volume	2CV (40mL)
Injection Volume	10mL
Flowthrough Capture	11mL
Fraction Size	2mL
Run Length	8CV (160mL)
Cleaning Volume	10CV (200mL)

4.25 Bradford Assay

4.25.1 Preparation of Bovine Serum Albumin (BSA) Dilutions

2mg/mL BSA was diluted with loading buffer into 6 concentrations with loading buffer according to Table 11.

Table 11: Dilutions of BSA into specified concentrations by the addition of loading buffer to 2mg/mL BSA.

Volume and source of BSA added	Volume of Loading Buffer added (µL)	Final BSA Concentration (µg/mL)	Total Solution Volume (µL)
0	100	0	100
20 µL (125µg/mL solution)	80	25	100
50 µL (250µg/mL solution)	50	125	100
50 µL (500µg/mL solution)	50	250	100
25 µL (2mg/mL stock)	75	500	100
37.5 µL (2mg/mL stock)	62.5	750	100
50µL (2mg/mL stock)	50	1000	100

Dilutions of BSA were mixed in microwell plates by gentle pipetting. 3 10µL replicates of each dilution and the loading buffer without BSA were transferred to wells in another microplate, and 200µL of Bradford assay 1x dye reagent was added to each well. Solutions were homogenized by gentle pipetting and incubated at room temperature for 5 minutes.

4.25.2 Preparation of YthA protein solutions

Washed KSI[N, Variant 1] and KSI[C, Variant 1] pellets grown from 0.5L of LB media from section 4.20.3 were resuspended in 50mL 8M urea, 20mM Tris, 250mM NaCl pH8 buffer, sonicated as in section 4.6 and centrifuged as in section 4.7. Dilutions of the KSI solutions were prepared according to Table 12.

Table 12: Dilutions of KSI[N, Variant 1] and KSI[C, Variant 1] protein solutions into specified dilutions by the addition of buffer to resuspended protein solution.

Volume and source of KSI solution added	Volume of Loading Buffer added (µL)	Dilution Factor	Total Solution Volume (µL)
50 µL (25x Dilution)	50	50	100
20 µL (10x Dilution)	80	25	100
10 µL (1x Dilution)	90	10	100
20 µL (1x Dilution)	80	5	100
50µL (1x Dilution)	50	2	100

Further preparation and insertion of KSI[N, Variant 1] and KSI[C, Variant 1] samples and dilutions was performed as described in section 4.25.1, except KSI[N, Variant 1] and KSI[C, Variant 1] protein solutions were used instead of BSA solutions.

4.25.3 Determining absorbance of BSA and KSI protein solutions

Microwell plates containing the BSA and KSI[N, Variant 1] and KSI[C, Variant 1] solutions from sections 4.25.1 and 4.25.2 were inserted into a microplate reader to measure absorbance at 595nm.

4.25.4 Determining the Standard Absorbance Curve of BSA Dilutions

The repetitions of each BSA dilution sample were averaged and the average of the 0µg/mL samples were subtracted from each BSA dilution sample absorbance average. The average absorbance of each dilution sample was plotted against its BSA concentration, and the line of best fit with an origin of 0 was calculated.

4.25.5 Estimating the Protein Concentration of KSI[N, Variant 1] and KSI[C, Variant 1] Solutions

The repetitions of each KSI[N, Variant 1] and KSI[C, Variant 1] dilution were averaged and the average of the blank samples were subtracted from each protein solution average. The KSI[N, Variant 1] and KSI[C, Variant 1] dilutions closest in absorbance to that of the 250µg/mL BSA solution's average absorbance were used to estimate the protein concentration, in conjunction with the equation of the line of best fit of the BSA standard curve. The result was then multiplied by the dilution factor of the KSI[N, Variant 1] and KSI[C, Variant 1] solutions chosen for analysis. Equation 2 describes the formula for the calculation.

$$\boldsymbol{A_K \times s \times d = c}$$

Equation 2: Equation used to estimate protein concentration of KSI[N, Variant 1] and KSI[C, Variant 1] resuspensions. A_k is the absorbance of the solution at 595nm, s is the slope of the line of best fit from the BSA standard curve, d is the dilution factor of the analyzed KSI sample and c is the estimated protein concentration in the sample (µg/mL).

4.26 TEVp Cleavage

4.26.1 Reaction Preparation

Reaction replicates of 50µL 10x TEVp reaction buffer (section 4.3.20), 50µL 50mg/mL purified TEVp and 400µL of solution containing solubilized KSI[N, Variant 1] or KSI[C, Variant 2] were prepared by gently mixing the reagents together in a microcentrifuge tube. 10x diluted replicates were prepared in the same manner as above, except that only 40µL of protein solution was added to the reaction and 360µL of MilliQ water was added as well.

4.26.2 Cleavage Reaction

Room temperature replicates were left sealed on the workbench away from any sources of warmth or cold during the reaction period, and 4°C replicates were left sealed in a 4°C cold room during the reaction period.

4.26.3 Quenching the Cleavage Reaction and SDS-Page Analysis

After the prescribed reaction period, reaction samples were immediately prepared for SDS-page analysis as described in section 4.14.1. Prepared SDS-page samples were stored at 4°C until all samples were prepared. Further SDS-page analysis on the samples followed sections 4.14.2-4.14.4.

5 References

Abraham, E. P., & Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, 146, 837.

Abriouel, H., Franz, C.M.A.P., Omar, N.B., Gálvez, A. (2011) Diversity and applications of Bacillus bacteriocins, *FEMS Microbiology Reviews*, 35(1), Pages 201–232, https://doi.org/10.1111/j.1574-6976.2010.00244.x

Ageitos, J. M., Sánchez-Pérez, A., Calo-Mata, P., & Villa, T. G. (2017). Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria. *Biochemical Pharmacology*, 133, 117–138. https://doi.org/10.1016/j.bcp.2016.09.018

Anderssen, E. L., Diep, D. B., Nes, I. F., Eijsink, V. G. H., & Nissen-Meyer, J. (1998). Antagonistic activity of Lactobacillus plantarum C11: Two new two- peptide bacteriocins, plantaricins EF and JK, and the induction factor plantaricin A. *Applied and Environmental Microbiology*, 64(6), 2269–2272. https://doi.org/10.1128/aem.64.6.2269-2272.1998

Baker, J. A., Wong, W. C., Eisenhaber, B., Warwicker, J., & Eisenhaber, F. (2017). Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule." *BMC Biology*, 15(1), 1–29. https://doi.org/10.1186/s12915-017-0404-4

Bax, R., & Griffin, D. Introduction to antibiotic resistance. Editor: Coates, A., R., M. *Antibiotic Resistance*, Springer (2012). Pg 1-12, ISBN 978-3-642-28950-7, DOI 10.1007/978-3-642-28951-4

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(W1), 252–258. https://doi.org/10.1093/nar/gku340

Bienert, S., Waterhouse, A., De Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*, 45(D1), D313–D319. https://doi.org/10.1093/nar/gkw1132

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D., & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium lactococcus lactis ssp. lactis IL1403. *Genome Research*, 11(5), 731–753. https://doi.org/10.1101/gr.GR-1697R

Bradford, M., M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry.* 72(1), 248-254. https://doi.org/10.1016/0003-2697(76)90527-3

Brissette, J. L., Russel, M., Weiner, L., & Model, P. (1990). Phage shock protein, a stress protein of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 87(3), 862–866. https://doi.org/10.1073/pnas.87.3.862

Bronsoms, S., & Trejo, S. A. (2015). Applications of mass spectrometry to the study of protein aggregation. *Methods in Molecular Biology*. (Vol. 1258). https://doi.org/10.1007/978-1-4939-2205-5_19

Card, P. B., & Gardner, K. H. (2005). Identification and optimization of protein domains for NMR studies. *Methods in Enzymology*, 394, 3–16. https://doi.org/10.1016/S0076-6879(05)94001-9

Cheng, Y., & Patel, D. J. (2004). An efficient system for small protein expression and refolding. *Biochemical and Biophysical Research Communications*, 317(2), 401–405. https://doi.org/10.1016/j.bbrc.2004.03.068

Cintas, L. M., Casaus, P., Håvarstein, L. S., Hernández, P. E., & Nes, I. F. (1997). Biochemical and genetic characterization of enterocin P, a novel sec- dependent bacteriocin from Enterococcus faecium P13 with a broad antimicrobial spectrum. *Applied and Environmental Microbiology*, 63(11), 4321–4330. https://doi.org/10.1128/aem.63.11.4321-4330.1997

Cooper, M. A., & Shlaes, D. (2011). Fix the antibiotics pipeline. *Nature*, 472(7341), 32. https://doi.org/10.1038/472032a

Cotter, P., D., Hill, C., & Ross, R., P. (2003). Bacteriocins, Developing Innate Immunity for Food. *Nature Reviews Microbiology,* 3(10), 777-788. DOI: 10.1038/nrmicro1273

Cotter, P. D., Ross, R. P., & Hill, C. (2013). Bacteriocins-a viable alternative to antibiotics? *Nature Reviews Microbiology*, 11(2), 95–105. https://doi.org/10.1038/nrmicro2937

Cutler, D., M., & Summers, L., H. (2020). The COVID-19 Pandemic and the $16 Trillion Virus. *The Journal of the American Medical Association*, 324(15), 1495-1496.

Domínguez-Escobar, J., Wolf, D., Fritz, G., Höfler, C., Wedlich-Söldner, R., & Mascher, T. (2014). Subcellular localization, interactions and dynamics of the phage-shock protein-like Lia response in Bacillus subtilis. *Molecular Microbiology*, 92(4), 716–732. https://doi.org/10.1111/mmi.12586

Dubey, S.; Diep, D.B., Evensen, Ø., Munang'andu, H.M. (2022) Garvicin KS, a Broad-Spectrum Bacteriocin Protects Zebrafish Larvae against Lactococcus garvieae Infection. *Int. J. Mol. Sci*, 23, 2833. https://doi.org/10.3390/ijms23052833

Ekblad, B., & Kristiansen, P. E. (2019). NMR structures and mutational analysis of the two peptides constituting the bacteriocin plantaricin S. *Scientific Reports*, 9(1), 1–10. https://doi.org/10.1038/s41598-019-38518-6

Ekblad, B., Kyriakou, P. K., Oppegård, C., Nissen-Meyer, J., Kaznessis, Y. N., & Kristiansen, P. E. (2016). Structure-Function Analysis of the Two-Peptide Bacteriocin Plantaricin EF. *Biochemistry*, 55(36), 5106–5116. https://doi.org/10.1021/acs.biochem.6b00588

Ennahar, S., Sashihara, T., Sonomoto, K., & Ishizaki, A. (2000). Class IIa bacteriocins: biosynthesis, structure and activity. *FEMS Microbiology Reviews*, 24(1), 85–106. https://doi.org/10.1111/j.1574-6976.2000.tb00534.x

Ferchichi, M., Frère, J., Mabrouk, K., & Manai, M. (2001). Lactococcin MMFII, a novel class IIa bacteriocin produced by Lactococcus lactis MMFII, isolated from a Tunisian dairy product. *FEMS Microbiology Letters*, 205(1), 49–55. https://doi.org/10.1016/S0378-1097(01)00435-9

Ferrer-Miralles, N., Saccardo, P., Corchero, J., L. and Garcia-Fruitós, E. Recombinant Protein Production and Purification of Insoluble Proteins. *Insoluble Proteins: Methods and Protocols.* Editted by Garcia-Fruitos, E. & Giralt, A., A. Springer Science + Business Media. (2022). Pg. 1-31. ISBN: 978-1-0716-1859-2

Fimland, N., Rogne, P., Fimland, G., Nissen-Meyer, J., & Kristiansen, P. E. (2008). Three-dimensional structure of the two peptides that constitute the two-peptide bacteriocin plantaricin EF. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1784(11), 1711–1719. https://doi.org/10.1016/j.bbapap.2008.05.003

Flores-Kim, J., & Darwin, A. J. (2012). Phage shock protein C (PspC) of Yersinia enterocolitica is a polytopic membrane protein with implications for regulation of the Psp stress response. *Journal of Bacteriology*, 194(23), 6548–6559. https://doi.org/10.1128/JB.01250-12

Franz, C. M. A. P., Grube, A., Herrmann, A., Abriouel, H., Stärke, J., Lombardi, A., Tauscher, B., & Holzapfel, W. H. (2002). Biochemical and genetic characterization of the two-peptide bacteriocin enterocin 1071 produced by Enterococcus faecalis FAIR-E 309. *Applied and Environmental Microbiology*, 68(5), 2550–2554. https://doi.org/10.1128/AEM.68.5.2550-2554.2002

Fregeau Gallagher, N. L., Sailer, M., Niemczura, W. P., Nakashima, T. T., Stiles, M. E., & Vederas, J. C. (1997). Three-dimensional structure of leucocin a in trifluoroethanol and dodecylphosphocholine micelles: Spatial location of residues critical for biological activity in type IIa bacteriocins from lactic acid bacteria. *Biochemistry*, 36(49), 15062–15072. https://doi.org/10.1021/bi971263h

Hartl, M. J., Mayr, F., Rethwilm, A., & Wöhrl, B. M. (2010). Biophysical and enzymatic properties of the simian and prototype foamy virus reverse transcriptases. Retrovirology, 7, 1–10. https://doi.org/10.1186/1742-4690-7-5

Haugen, H. S., Fimland, G., Nissen-Meyer, J., & Kristiansen, P. E. (2005). Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide curvacin A. *Biochemistry*, 44(49), 16149–16157. https://doi.org/10.1021/bi051215u

Henderson, J. T., Chopko, A. L., & van Wassenaar, P. D. (1992). Purification and primary structure of pediocin PA-1 produced by Pediococcus acidilactici PAC-1.0. *Archives of Biochemistry and Biophysics*, 295(1), 5–12. https://doi.org/10.1016/0003-9861(92)90480-K

Heng, N., C., K., Tagg, J., R. (2006). What's in a name? Class distinction for bacteriocins. *Nature Microbiology Review.* 4(2), 160. DOI: 10.1038/nrmicro1273-c1

Hoffmann, D., Ebrahimi, M., Gerlach, D., Salzig, D., & Czermak, P. (2018). Reassessment of inclusion body-based production as a versatile opportunity for difficult-to-express recombinant proteins. *Critical Reviews in Biotechnology*, 38(5), 729–744. https://doi.org/10.1080/07388551.2017.1398134

Holmes, K. C. (1999). Structural biology. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 354(1392). https://doi.org/10.1098/rstb.1999.0537

Holpuch, Amanda. “UN meeting tackles the ‘fundamental threat’ of antibiotic-resistant superbugs.” *The Guardian*, 21 September 2016.

Huan, Yuchen & Kong, Qing & Mou, Haijin & Yi, Huaxi. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Frontiers in Microbiology.* (2020). 11. 582779. 10.3389/fmicb.2020.582779.

Huvet, M., Toni, T., Sheng, X., Thorne, T., Jovanovic, G., Engl, C., Buck, M., Pinney, J. W., & Stumpf, M. P. H. (2011). The evolution of the phage shock protein response system: Interplay between protein function, genomic organization, and system function. *Molecular Biology and Evolution*, 28(3), 1141–1155. https://doi.org/10.1093/molbev/msq301

Innovative Medicines Initiative. “The IMI funding model”. https://www.imi.europa.eu/about-imi/imi-funding-model. Retrieved March 17th, 2022.

Jovanovic, G., Lloyd, L. J., Stumpf, M. P. H., Mayhew, A. J., & Buck, M. (2006). Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. *Journal of Biological Chemistry*, 281(30), 21147–21161. https://doi.org/10.1074/jbc.M602323200

Kalmokoff, M. L., Banerjee, S. K., Cyr, T., Hefford, M. A., & Gleeson, T. (2001). Identification of a New Plasmid-Encoded sec-Dependent Bacteriocin Produced by Listeria innocua 743. *Applied and Environmental Microbiology*, 67(9), 4041–4047. https://doi.org/10.1128/AEM.67.9.4041-4047.2001

Kamaraj, C., & Pasupathi, G. (2021). Investigation on certain optical, thermal, and Z-scan studies of nonlinear optical tris–sarcosine calcium chloride single crystal. *Journal of Materials Science: Materials in Electronics*, 32(23), 27362–27372. https://doi.org/10.1007/s10854-021-07109-5

Kapoor, G., Saigal, S., & Elongavan, A. (2017). Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology Clinical Pharmacology,* 33(3):300-305. doi: 10.4103/joacp.JOACP_349_15.

Kapust, R. B., Toözseór, J., Copeland, T. D., & Waugh, D. S. (2002). The P1′ specificity of tobacco etch virus protease. Biochemical and Biophysical Research Communications, 294(5), 949–955. https://doi.org/10.1016/S0006-291X(02)00574-0

Klaenhammer, T., R. (1993). Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbio. Rev.* 12(1). 39-85. DOI: 10.1016/0168-6445(93)90057-G

Kjos, M., Oppegård, C., Diep, D. B., Nes, I. F., Veening, J. W., Nissen-Meyer, J., & Kristensen, T. (2014). Sensitivity to the two-peptide bacteriocin lactococcin G is dependent on UppP, an enzyme involved in cell-wall synthesis. *Molecular Microbiology*, 92(6), 1177–1187. https://doi.org/10.1111/mmi.12632

Koenig, B. W., Rogowski, M., & Louis, J. M. (2003). A rapid method to attain isotope labeled small soluble peptides for NMR studies. *Journal of Biomolecular NMR*, 26(3), 193–202. https://doi.org/10.1023/A:1023887412387

Kranjec, C., Kristensen, S. S., Bartkiewicz, K. T., Brønner, M., Cavanagh, J. P., Srikantam, A., Mathiesen, G., & Diep, D. B. (2021). A bacteriocin-based treatment option for Staphylococcus haemolyticus biofilms. *Scientific Reports*, 11(1), 1–14. https://doi.org/10.1038/s41598-021-93158-z

Lei, J., Sun, L. C., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D. H., & He, Q. Y. (2019). The antimicrobial peptides and their potential clinical applications. *American Journal of Translational Research*, 11(7), 3919–3931.

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification*, 41(1), 98–105. https://doi.org/10.1016/j.pep.2005.01.019

Marley, J., Lu, M., & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20, 71–75.

Maxson, M. E., & Darwin, A. J. (2006). PspB and PspC of Yersinia enterocolitica are dual function proteins: Regulators and effectors of the phage-shock-protein response. *Molecular Microbiology*, 59(5), 1610–1623. https://doi.org/10.1111/j.1365-2958.2006.05047.x

Millipore-Sigma. "pET-22b(+) DNA – Novagen". *EMDMillipore.com*. Retrieved February 23, 2023. https://www.emdmillipore.com/CA/en/product/pET-22b-DNA-Novagen,EMD_BIO-69744

Millipore-Sigma. "pET-31b(+) DNA – Novagen". *EMDMillipore.com*. Retrieved February 23, 2023. https://www.emdmillipore.com/CA/en/product/pET-31b-DNA-Novagen,EMD_BIO-69952

Mohanty, A. K., & Wiener, M. C. (2004). Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expression and Purification*, 33(2), 311–325. https://doi.org/10.1016/j.pep.2003.10.010

Moll, G., Ubbink-kok, T., Hildeng-hauge, H., Nissen-meyer, J. O. N., Nes, I. F., Konings, W. I. L. N., & Driessen, A. J. M. (1996). Lactococcin G is a potassium ion-conducting, two-component bacteriocin. *Journal of Bacteriology*. 178(3), 600–605.

Morreale, G., Lee, E. G., Jones, D. B., & Middelberg, A. P. J. (2004). Bioprocess-centered molecular design (BMD) for the efficient production of an interfacially active peptide. *Biotechnology and Bioengineering*, 87(7), 912–923. https://doi.org/10.1002/bit.20209

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet, 399(10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Naskar, A., & Kim, K. S. (2021). Potential novel food-related and biomedical applications of nanomaterials combined with bacteriocins. Pharmaceutics, 13(1), 1–15. https://doi.org/10.3390/pharmaceutics13010086

Nathan, C., & Cars, O. (2014). Antibiotic Resistance – Problems, Progress and Prospects. *The New England Journal*, 371(19), 689–691.

Nissen-Meyer, J., Holo, H., Havarstein, L. S., Sletten, K., & Nes, I. F. (1992). A novel lactococcal bacteriocin whose activity depends on the complementary action of two peptides. *Journal of Bacteriology*, 174(17), 5686–5692. https://doi.org/10.1128/jb.174.17.5686-5692.1992

NovoPro. "pMAL-c5X vector (V012449#)". *NovoProLabs.com*. Retrieved February 23, 2023. https://www.novoprolabs.com/vector/V12449

O'Neill, J. (2014). Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. *The Review on Antimicrobial Resistance.*

Oppegård, C., Rogne, P., Emanuelsen, L., Kristiansen, P. E., Fimland, G., & Nissen-Meyer, J. (2007). The two-peptide class II bacteriocins: Structure, production, and mode of action. *Journal of Molecular Microbiology and Biotechnology*, 13(4), 210–219. https://doi.org/10.1159/000104750

Oppegård, C., Schmidt, J., Kristiansen, P. E., & Nissen-Meyer, J. (2008). Mutational analysis of putative helix-helix interacting GxxxG-motifs and tryptophan residues in the two-peptide bacteriocin lactococcin G. *Biochemistry*, 47(18), 5242–5249. https://doi.org/10.1021/bi800289w

Ortega, C., Oppezzo, P. & Correa, A. Overcoming the Solubility Problem in E. coli: Available Approaches for Recombinant Protein Production. *Insoluble Proteins: Methods and Protocols.* Editted by Garcia-Fruitos, E. & Giralt, A., A. Springer Science + Business Media. (2022). Pg. 35-64. ISBN: 978-1-0716-1859-2

Ovchinnikov, K. V., Chi, H., Mehmeti, I., Holo, H., Nes, I. F., & Diep, D. B. (2016). Novel group of leaderless multipeptide bacteriocins from Gram-positive bacteria. *Applied and Environmental Microbiology*, 82(17), 5216–5224. https://doi.org/10.1128/AEM.01094-16

Ovchinnikov, K. V., Kristiansen, P. E., Straume, D., Jensen, M. S., Aleksandrzak-Piekarczyk, T., Nes, I. F., & Diep, D. B. (2017). The leaderless bacteriocin enterocin K1 is highly potent against Enterococcus faecium: A study on structure, target spectrum and receptor. *Frontiers in Microbiology*, 8(MAY), 1–12. https://doi.org/10.3389/fmicb.2017.00774

Payne, D. J., Gwynn, M. N., Holmes, D. J., & Pompliano, D. L. (2007). Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery*, 6(1), 29–40. https://doi.org/10.1038/nrd2201

Pereira, W. A., Mendonça, C. M. N., Urquiza, A. V., Marteinsson, V. Þ., LeBlanc, J. G., Cotter, P. D., Villalobos, E. F., Romero, J., & Oliveira, R. P. S. (2022). Use of Probiotic Bacteria and Bacteriocins as an Alternative to Antibiotics in Aquaculture. Microorganisms, 10(9). https://doi.org/10.3390/microorganisms10091705

Raran-Kurussi, S., & Waugh, D. S. (2016). A dual protease approach for expression and affinity purification of recombinant proteins. *Analytical Biochemistry*, 504, 30–37. https://doi.org/10.1016/j.ab.2016.04.006

Riley, M.A., & Chavan, M.A. (2007). Bacteriocins: Ecology and Evolution. *Springer*. ISBN: 978-3-540-36604-1. https://doi-org.ezproxy.uio.no/10.1007/978-3-540-36604-1

Rodríguez, V., Asenjo, J. A., & Andrews, B. A. (2014). Design and implementation of a high yield production system for recombinant expression of peptides. *Microbial Cell Factories*, 13(1), 1–10. https://doi.org/10.1186/1475-2859-13-65

Rodriguez, J., M., Martinez, M., I., & Kok, J. (2002). Pediocin PA-1, a Wide-Spectrum Bacteriocin from Lactic Acid Bacteria. *Critical Reviews in Food Science and Nutrition.* 42(2). 91-121.

Rogne, P., Fimland, G., Nissen-Meyer, J., & Kristiansen, P. E. (2008). Three-dimensional structure of the two peptides that constitute the two-peptide bacteriocin lactococcin G. *Biochimica et Biophysica Acta - Proteins and Proteomics,* 1784(3), 543–554. https://doi.org/10.1016/j.bbapap.2007.12.002

Schrader, E. K., Harstad, K. G., & Matouschek, A. (2009). Targeting proteins for degradation. *Nature Chemical Biology*, 5(11), 815–822. https://doi.org/10.1038/nchembio.250

Senes, A., Engel, D. E., & Degrado, W. F. (2004). Folding of helical membrane proteins: The role of polar, GxxxG-like and proline motifs. *Current Opinion in Structural Biology*, 14(4), 465–479. https://doi.org/10.1016/j.sbi.2004.07.007

Senes, A., Gerstein, M., & Engelman, D. M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and association with β-branched residues at neighboring positions. *Journal of Molecular Biology*, 296(3), 921–936. https://doi.org/10.1006/jmbi.1999.3488

Singhvi, P., & Panda, A., K. Solubilization and Refolding of Inclusion Body Proteins. *Insoluble Proteins: Methods and Protocols.* Editted by Garcia-Fruitos, E. & Giralt, A., A. Springer Science + Business Media. (2022). Pg. 371-388. ISBN: 978-1-0716-1859-2

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., & Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Science*, 18(5), 936–948. https://doi.org/10.1002/pro.102

Smith, J. S., & Scholtz, J. M. (1996). Guanidine hydrochloride unfolding of peptide helices: Separation of denaturant and salt effects. *Biochemistry*, 35(22), 7292–7297. https://doi.org/10.1021/bi960341i

Stephens, S. K., Floriano, B., Cathcart, D. P., Bayley, S. A., Witt, V. F., Jiménez-Díaz, R., Warner, P. J., & Ruiz-Barba, J. L. (1998). Molecular analysis of the locus responsible for production of plantaricin S, a two-peptide bacteriocin produced by Lactobacillus plantarum LPCO10. *Applied and Environmental Microbiology,* 64(5), 1871–1877. https://doi.org/10.1128/aem.64.5.1871-1877.1998

Teale, W. W., & Weber, G. (1957). Ultraviolet fluorescence of the aromatic amino acids. *The Biochemical Journal*, 65(3), 476–482. https://doi.org/10.1042/bj0650476

Telke, A. A., Ovchinnikov, K. V., Vuoristo, K. S., Mathiesen, G., Thorstensen, T., & Diep, D. B. (2019). Over 2000-fold increased production of the leaderless bacteriocin garvicin Ks by increasing gene dose and optimization of culture conditions. *Frontiers in Microbiology*, 10(MAR), 1–9. https://doi.org/10.3389/fmicb.2019.00389

Teng, Q. (2005). Structural Biology: Practical NMR Applications. *Springer Science+Business Media*. ISBN: 0-387-24368-2

Tropea, J., E., Cherry, S., and Waugh, D., S. (2009). Expression and Purification of Soluble His_6-Tagged TEV Protease. *Methods in Molecular Biology: High Throughput Protein Expression and Purification*, 498. 297-307. doi: 10.1007/978-1-59745-196-3

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research*, 43(W1), W401–W407. https://doi.org/10.1093/nar/gkv485

United States Department of Health and Human Services. (2019). Antibiotic resistance threats in the United States. *Centers for Disease Control and Prevention*, 1–150.

Upadhyay, V., Singh, A., Jha, D., Singh, A., & Panda, A. K. (2016). Recovery of bioactive protein from bacterial inclusion bodies using trifluoroethanol as solubilization agent. *Microbial Cell Factories*, 15(1), 1–13. https://doi.org/10.1186/s12934-016-0504-9

Uteng, M., Hauge, H. H., Markwick, P. R. L., Fimland, G., Mantzilas, D., Nissen-Meyer, J., & Muhle-Goll, C. (2003). Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide sakacin P and a sakacin P variant that is structurally stabilized by an inserted C-terminal disulfide bridge. *Biochemistry*, 42(39), 11417–11426. https://doi.org/10.1021/bi034572i

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wijekoon, C. J. K., Ukuwela, A. A., Wedd, A. G., & Xiao, Z. (2016). Evaluation of employing poly-lysine tags versus poly-histidine tags for purification and characterization of recombinant copper-binding proteins. Journal of Inorganic Biochemistry, 162, 286–294. https://doi.org/10.1016/j.jinorgbio.2015.12.009

Wolf, D., Kalamorz, F., Wecke, T., Juszczak, A., Mäder, U., Homuth, G., Jordan, S., Kirstein, J., Hoppert, M., Voigt, B., Hecker, M., & Mascher, T. (2010). In-depth profiling of the LiaR response of bacillus subtilis. *Journal of Bacteriology*, 192(18), 4680–4693. https://doi.org/10.1128/JB.00543-10

World Health Organization. (2017). WHO publishes list of bacteria for which new antibiotics are urgently needed. *World Health Organization.* Online

World Health Organization. (2021). WHO Coronavirus (COVID-19) Dashboard. *World Health Organization.* Online. https://covid19.who.int/ Retrieved November 3rd, 2021.

Wu, H., Liu, J., Miao, S., Zhao, Y., Zhu, H., Qiao, M., Saris, P., E., J., & Qiao, J. (2018). Contribution of YthA, a PspC Family Transcriptional Regulator of Lactococcus lactis F44 Acid Tolerance and Nisin Yield: a Transcriptomic Approach. *Applied and Environmental Microbiology.* 84(6). e02483-17. https://doi.org/10 .1128/AEM.02483-17

6 Appendices

6.1 Abbreviations and Acronyms

AA	Amino Acid
AMP	Antimicrobial Peptide
AMR	(Multi-)Antimicrobial Resistant
BSA	Bovine Serum Albumin
CDC	Centres for Disease Control (and Prevention)
CD	Circular Dichromism
CV	Column Volume
DC	Drug Candidate
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESBL	Extended-spectrum β-lactamase
FDA	Food and Drug Administration
GDP	Gross Domestic Product
GHCl	Guanidine Hydrochloride
GRAS	Generally-recognized-as-safe
HHS	United States Department of Health and Human Services
HPLC	High Pressure Liquid Chromatography
HTS	High Throughput Screening
INI	Innovative Medicines Initiative
LDS	Lithium dodecyl sulfate
LB	Lysogeny Broth
MEM	Minimum Essential Medium
MES	2-(N-morpholino)ethanesulfonic acid
MGM	Minimal Growth Media
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
OD_x	Optical Density at x nanometers
PBP	Penicillin Binding Protein
pI	Isoelectric Point
PlnS	Plantaricin S
PTS	Mannose Phosphotransferase System
RNA	Ribonucleic Acid
SDS	Sodium dodecyl sulfate
TEV	Tobacco Etch Virus
TFE	Trifluoroethanol
Tris	tris(hydroxymethyl)aminomethane
UV	Ultraviolet
USD	United States Dollars

6.2 Equipment

Equipment	Manufacturer	Model
AutoClave	TOMY	SS-325
Block Heater	Grant	QBT2
Centrifugation Rotor	Beckman Coulter	JA-10
Centrifugation Rotor	Beckman Coulter	JA-25.50
Centrifuge	Eppendorf	Centrifuge 5804 R
Centrifuge	Eppendorf	Centrifuge 5415 R
Electrophoresis Unit	Invitrogen	PowerEase 500
HPLC System	General Electric Healthcare	ÄKTApurifier 10
Incubator	Termaks	B8023
Microplate Reader	Tecan	Sunrise
Scanner	Canon	Canoscan LiDE 90
Shaking Incubator	New Brunswick Scientific	Innova 4340
Shaking Incubator	Eppendorf	Innova 40
Shaking Table	Heidolph	Duomax 1030
Sonicator	Sonics	VCX-500
Sonicator Probe	Sonics	CV33
Ultracentrifuge	Beckman Coulter	Optima MAX
Ultracentrifuge	Beckman Coulter	Avanti J-20 XP
Ultra-Low Temperature Freezer	Thermo-Fischer	700 Series
UV-Vis Spectrometer	Shimadzu	UV-160A
Vortexing Unit	Marshall Scientific	MS2 Minishaker
Water Purification System	Millipore	MilliQ Advantage A10
Weighing Scale	Mettler	AT250
Weighing Scale	Mettler	BB2400

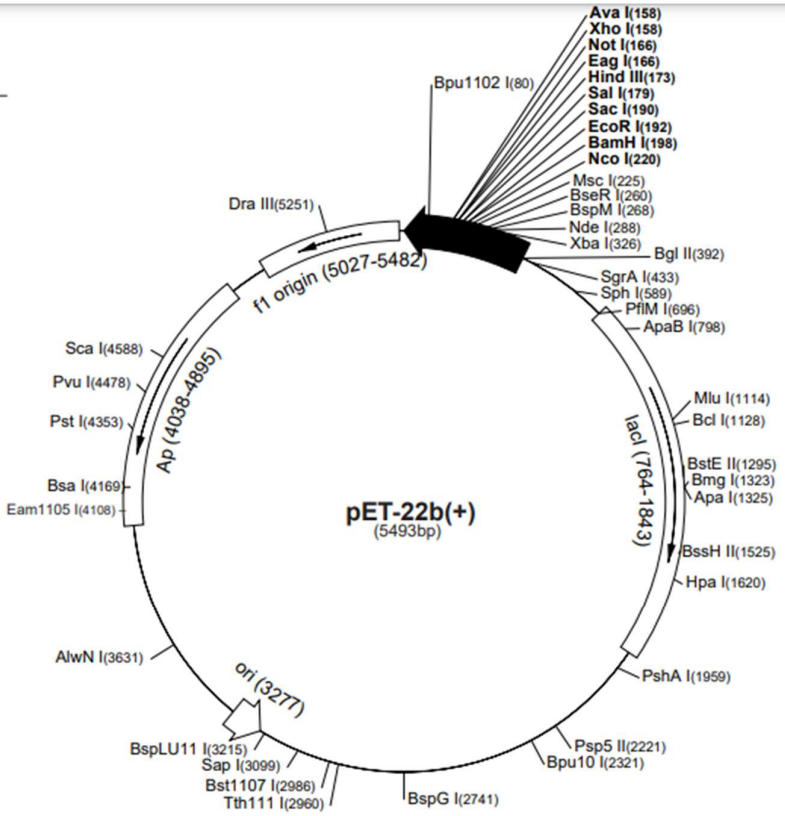
6.3 Reagents and Disposables

Reagent	Manufacturer	Product Code
12% Bis-Tris running gel	Invitrogen	NW00122BOX
2-Propanol	Sigma	33539
Agar	Invitrogen	30391-023
Bovine Serum Albumin, 2mg/mL Standard	Biorad	5000206
Bradford Assay 1x Dye Reagent	Biorad	5000205
Centrifugation Filters (50 000 MWCO)	Thermo Scientific	UFC905008
D(+)-Glucose-monohydrate	Merck	1.08342.1000
Desalting Column	General Electric Healthcare	17-1408-01
di-Sodium hydrogen phosphate dihydrate	Merck	1.06580.1000
EDTA	AppliChem	A1103
Ethanol	Arcus	
Glycerol	VWR	24388.295
Glycine	Sigma	G8898
Guanidine hydrochloride	Sigma	177253
Histadine Purification Column	General Electric Healthcare	17-5248-02
Imidazole	Sigma	56750
L-Arginine	Sigma	W381918
LDS Sample Buffer	Invitrogen	NP0007
MES SDS Running Buffer	Invitrogen	NP0002-02
Ni-NTA Resin	Thermo Scientific	88222
Peptone from meat, peptic digest	Fluka	70174
Plasmid Midiprep Kit	Invitrogen	K210005
Potassium Chloride	Merck	1.04936.1000
Size Exclusion Column	Thermo-Fischer	079723
Sodium Chloride	VWR	27810.295
Sodium dihydrogen phosphate dihydrate	Merck	1.06342.1000
Spectra™ Multicolor Low Range Protein Ladder	Thermo Scientific	26628
Tris-(hydroxymethyl) aminomethane	VWR	33621.260
Triton™ X-100	Sigma	X100
Urea	Merck	1.08488.9010
Yeast Extract	Merck	1.03753.0500

6.4 Vector Maps

6.4.1 pET-22b(+)

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites (*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482

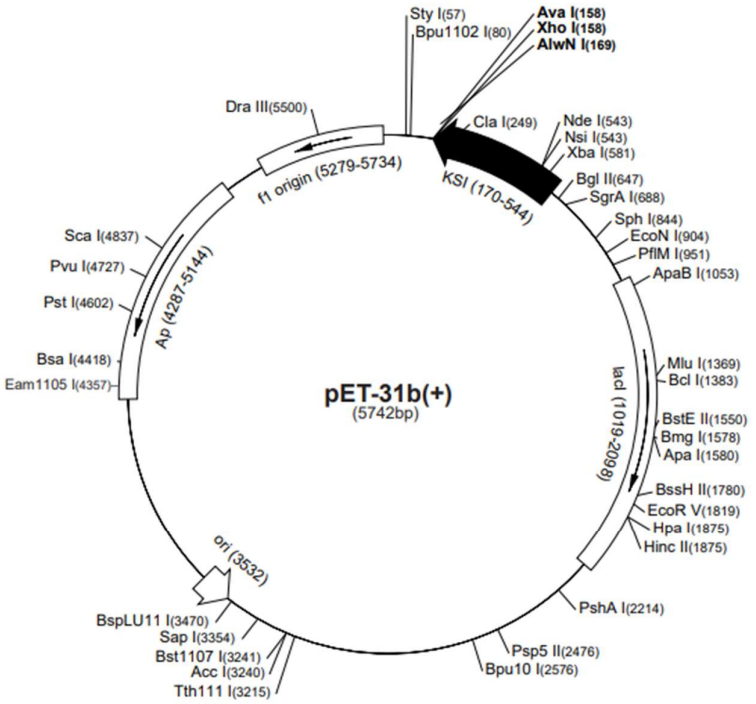


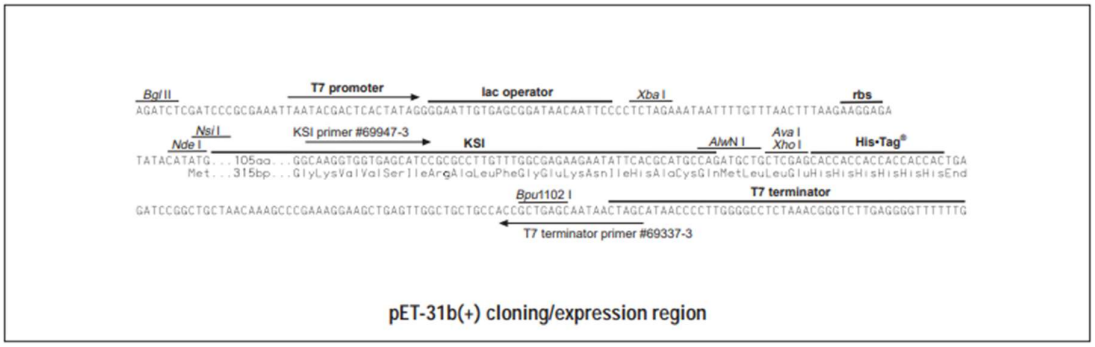
pET-22b(+) cloning/expression region

(Millipore-Sigma. “pET-22b(+) DNA – Novagen”, Retrieved February 23, 2023)

6.4.2 pET-31b(+)

pET-31b(+) sequence landmarks	
T7 promoter	616-632
T7 transcription start	615
KSI coding sequence	170-544
AlwN I site	169
His•Tag® coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1019-2098
pBR322 origin	3532
bla coding sequence	4287-5144
f1 origin	5276-5731

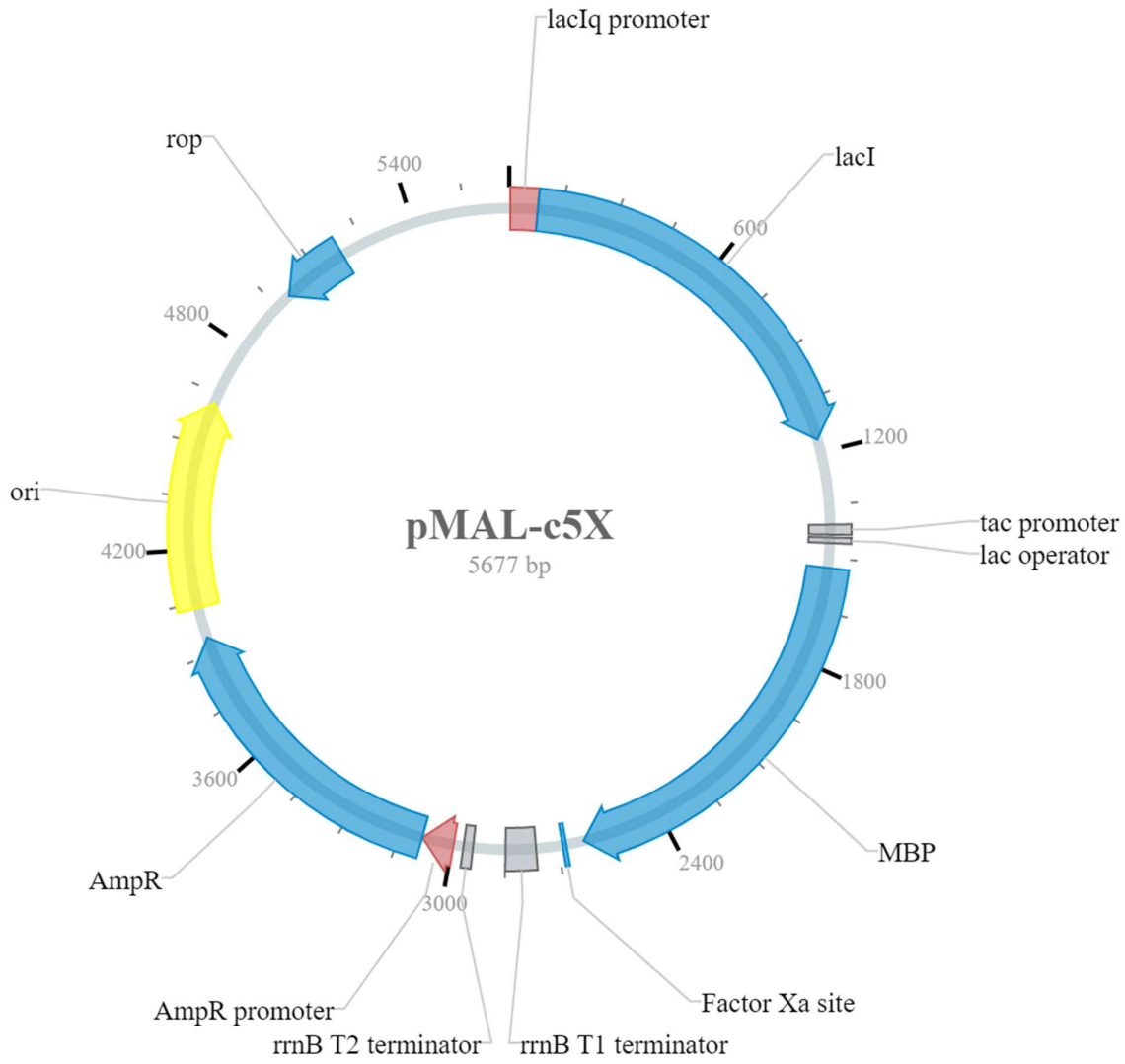




pET-31b(+) cloning/expression region

(Millipore-Sigma. "pET-31b(+) DNA – Novagen". Retrieved February 23, 2023)

6.4.3 pMAL-c5X



(NovoPro, "pMAL-c5X vector (V012449#)".Retrieved February 23, 2023)

6.4.4 pRK793

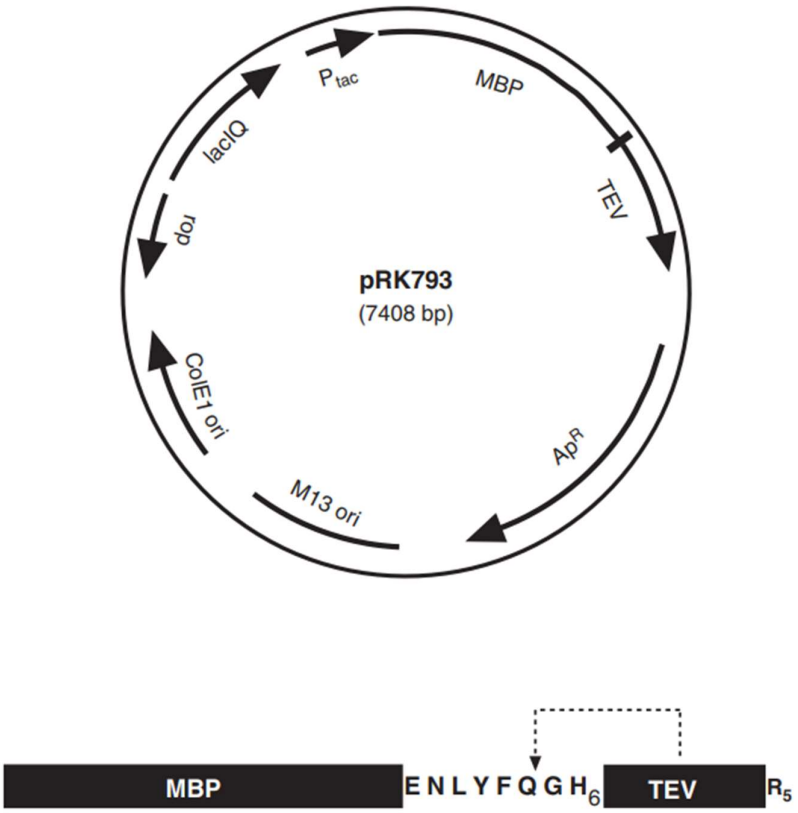


Fig. 19.1. Schematic representation (not to scale) of the TEV protease expression vector pRK793 and its fusion protein product. Further information about this plasmid can be found at http://mcl1.ncifcrf.gov/waugh_tech.html.

(Tropea *et al.,* 2009)

6.5 IMAC Graphs

6.5.1 Loading Chromatograph of GB1[C] Purification

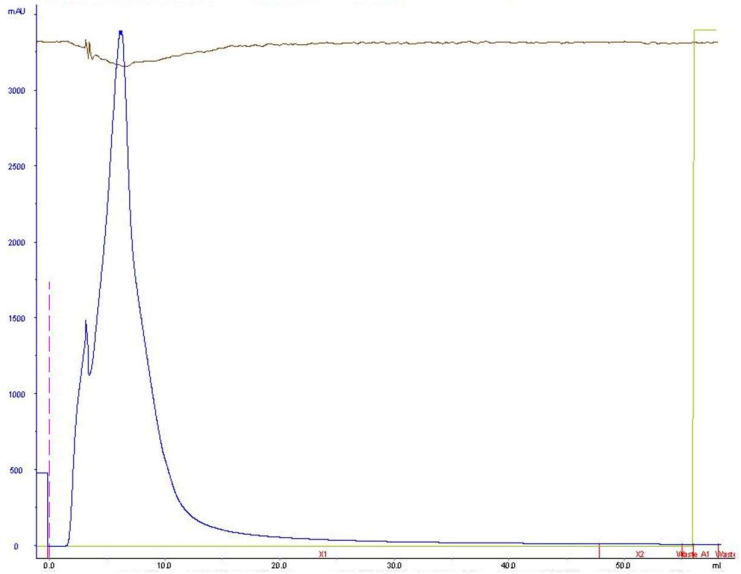


Figure 44: Loading chromatograph of GB1[C] fusion protein purification attempt. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), sample injection (dotted line) and sample fractions (red letters).

6.5.2 Refolding Chromatograph of GB1[C] Purification

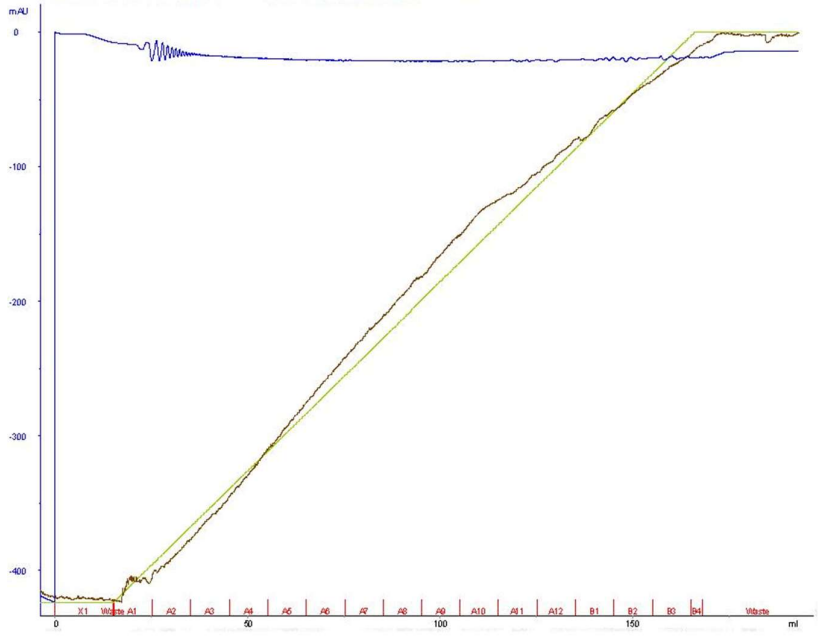


Figure 45: Refolding chromatograph of GB1[C] on-column refolding during purification. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient from 6M GHCl to 0M GHCl (green line), sample injection (dotted line) and sample fractions (red letters).

6.5.3 Buffer Exchange Chromatograph of GB1[N] Purification

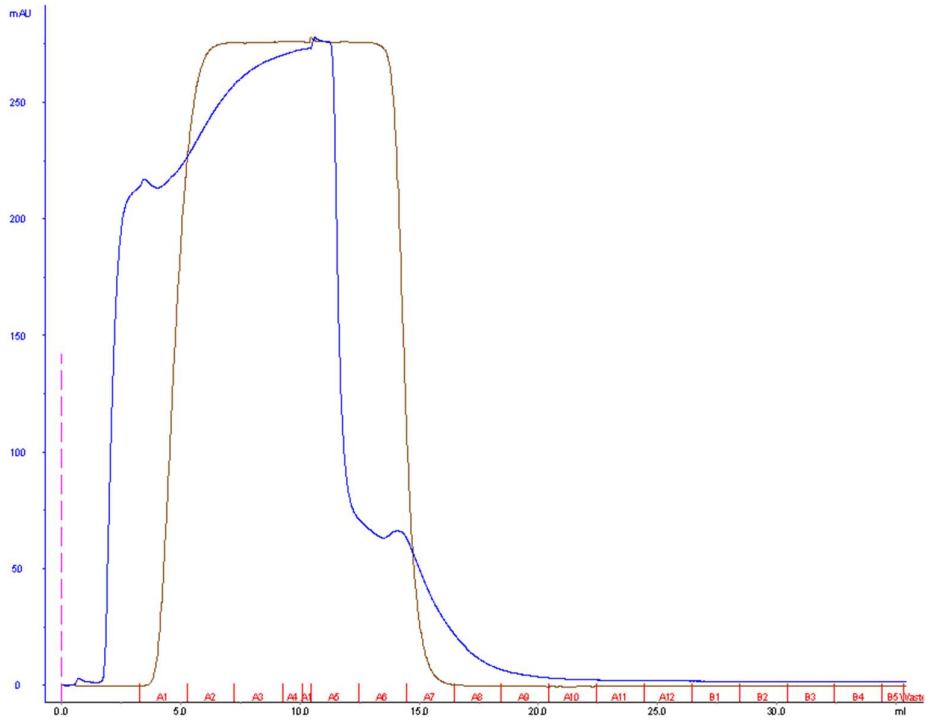


Figure 46: Buffer exchange Chromatograph of GB1[N] dissolved in phosphate-based buffer to tris-bases buffer. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), sample injection (dotted line) and sample fractions (red letters).

6.5.4 Elution Chromatograph of Refolded KSI[N, Variant 1]

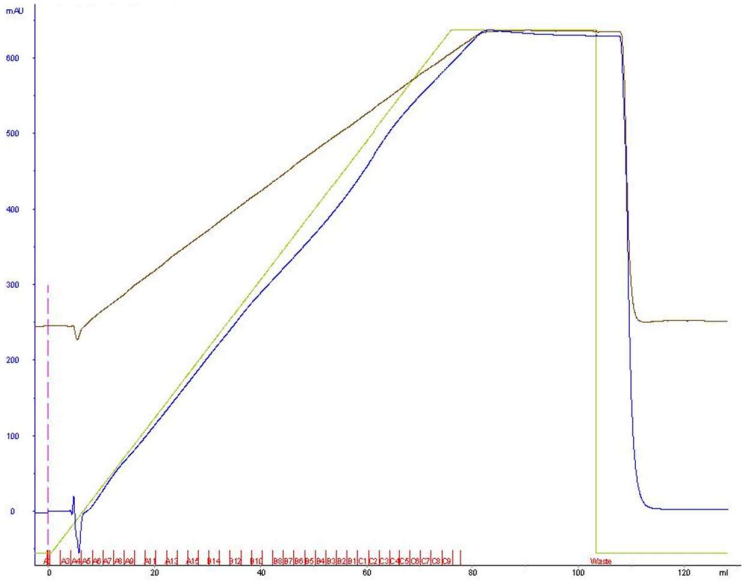


Figure 47: Elution chromatograph of refolded KSI[N, Variant 1] fusion protein purification attempt. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 20mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters). No absorbance peaks were observed.

6.5.5 Initial Purification Elution Chromatograph of KSI[C, Variant 1]

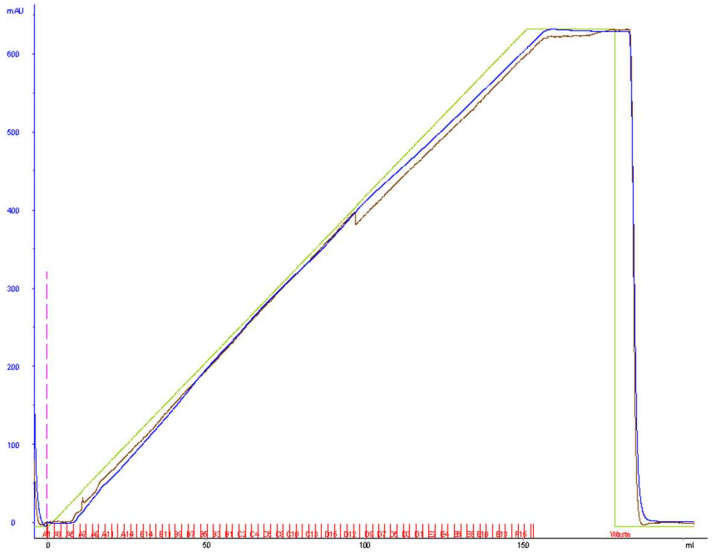


Figure 48: Elution chromatograph of KSI[C, Variant 1] fusion protein purification attempt. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 20mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters). No absorbance peaks were observed.

6.5.6 Low-Imidazole Elution Chromatograph of KSI [C, Variant 1]

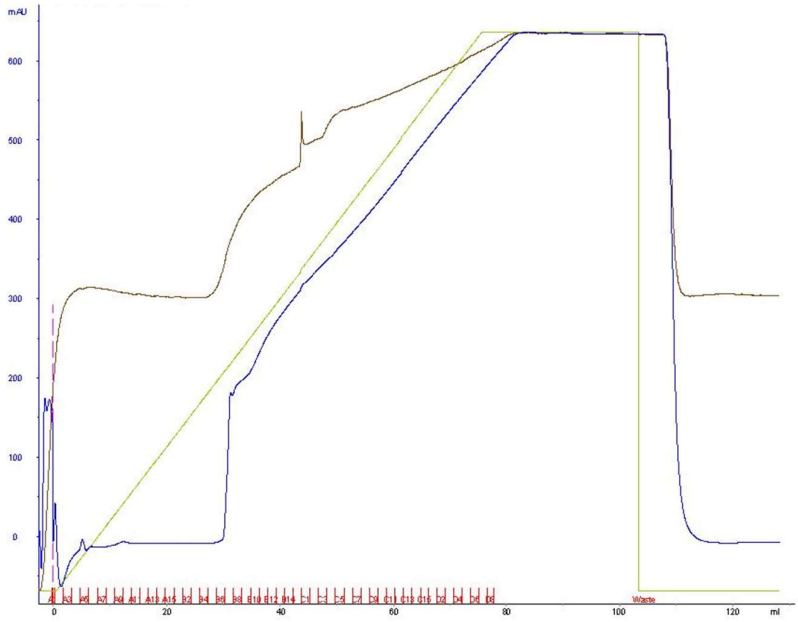


Figure 49: Elution chromatograph of KSI[C, Variant 1] fusion protein purification attempt using a low-imidazole loading buffer. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters). No absorbance peaks were observed.

6.5.7 Low-Imidazole Elution Chromatograph of KSI [C, Variant 1], 2nd Attempt

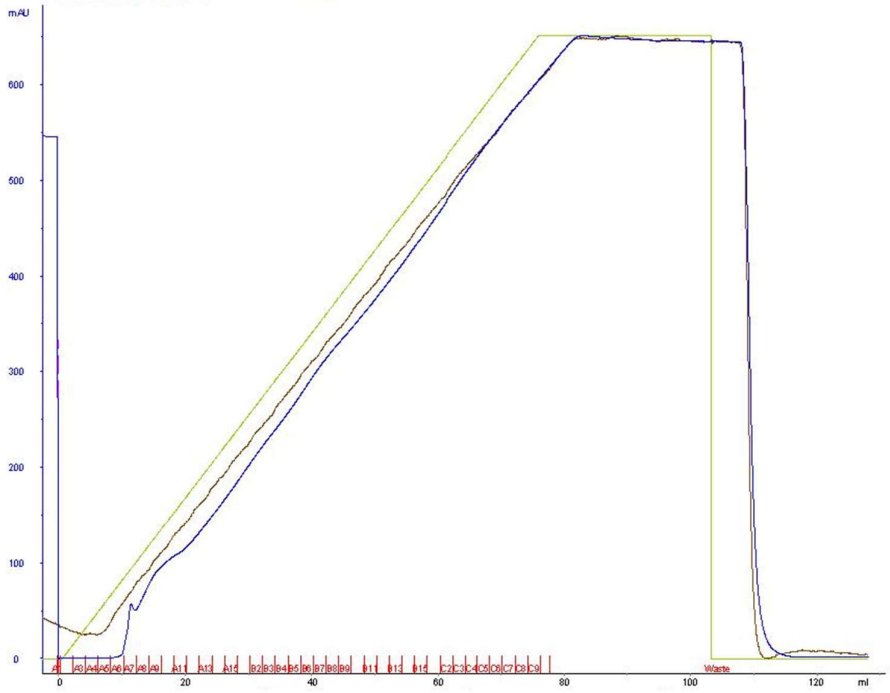


Figure 50: Elution chromatograph of KSI[C, Variant 1] fusion protein purification attempt using a low-imidazole loading buffer. Indicated on the chromatograph are absorbance at 280nm (blue line), conductivity (brown line), buffer gradient of 5mM to 500mM imidazole (green line), sample injection (dotted line) and sample fractions (red letters). No absorbance peaks were observed.

6.6 SDS-Page Gels

6.6.1 Unedited Induction Tests of KSI[C, Variant 1], GB1[C] and His[C]

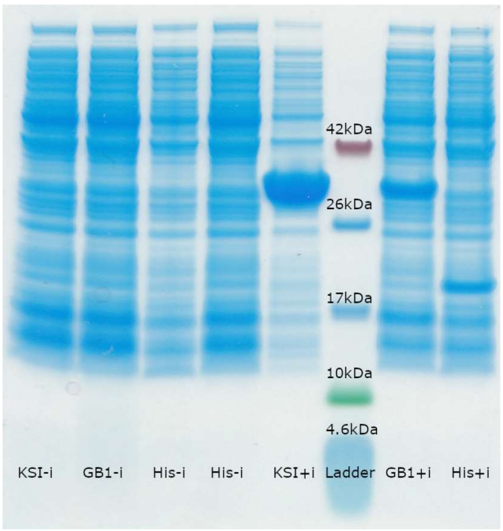


Figure 51: Cropped and labelled, but otherwise unedited SDS-page of induced (+i) and uninduced (-i) His[C]- (His), GB1[C]- (GB1) or KSI[C, Variant 1]-transformed cells. A protein ladder is included for size comparison. Lysate of uninduced His[C]-transformed cells was accidentally loaded twice.

6.6.2 Analysis of GB1[C] Initial Purification Attempt

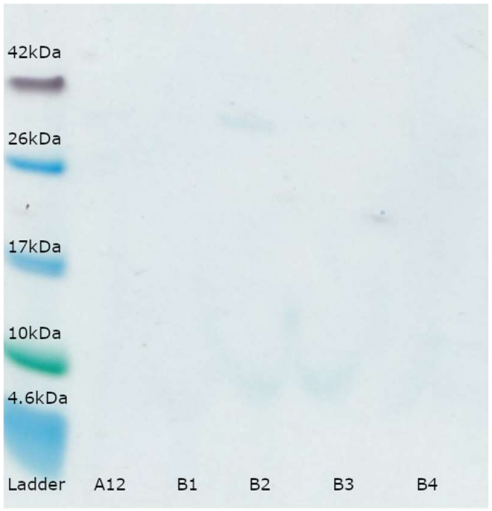


Figure 52: Cropped and labelled, but otherwise unedited SDS-page of the peak elution fractions of the initial purification attempt GB1[C]. A protein ladder is included for size comparison. No bands could be definitively distinguished.

6.6.3 Analysis of GB1[N] Buffer Exchange

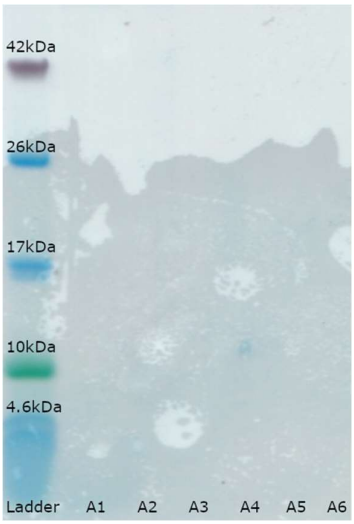


Figure 53: Cropped and labelled, but otherwise unedited SDS-page of the peak fractions of the buffer exchange on GB1[N]. A protein ladder is included for size comparison. No bands could be definitively distinguished.

6.6.4 Analysis of KSI[N, Variant 2] Induction Failure

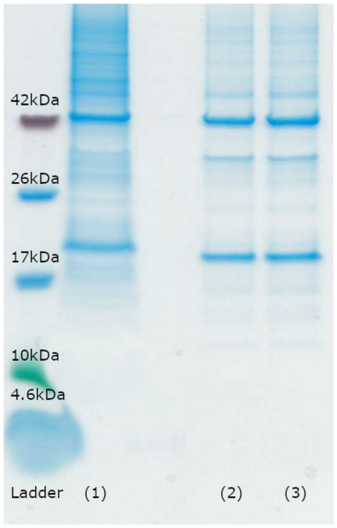


Figure 54: Cropped and labelled, but otherwise unedited SDS-page of uninduced (2) and induced (3) KSI[N, Variant 2]-transformed cell lysate. A protein ladder is included for size comparison. No KSI[N, Variant 2] could be distinguished. (1) was mistakenly loaded with an irrelevant sample.

6.6.5 Unedited Induction Temperature Optimization of KSI[N, Variant 1] Growth

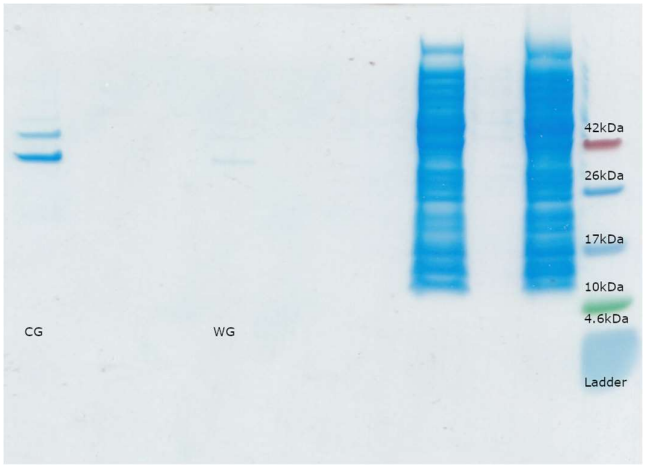


Figure 55: Cropped and labelled, but otherwise unedited SDS-page of KSI[N, Variant 1] fusion protein production following induction at 37°C for 4 hours (WG) or at 18°C for 20 hours (CG). A protein ladder is included for size comparison. Other wells were used for testing other experiments to conserve resources.

6.6.6 Unedited Solubilization Test of KSI[N, Variant 1] IBs Induced at 37°C

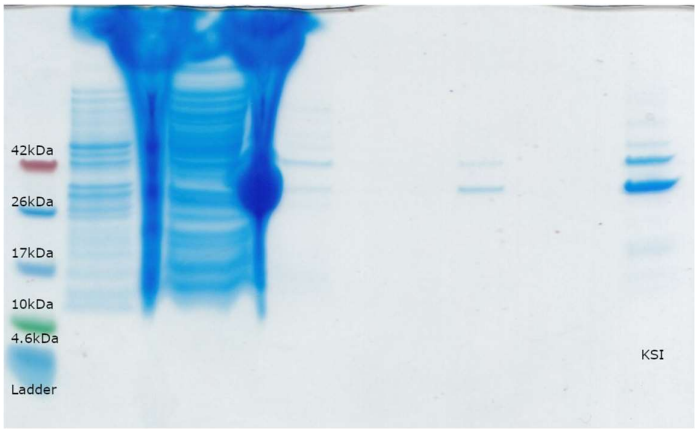


Figure 56: Cropped and labelled, but otherwise unedited SDS-page of 37°C-induced KSI[N, Variant 1] washed IBs. A protein ladder is included for size comparison. Other wells were used for testing other experiments to conserve resources.

6.6.7 Unedited Production of KSI[C, Variant 2]

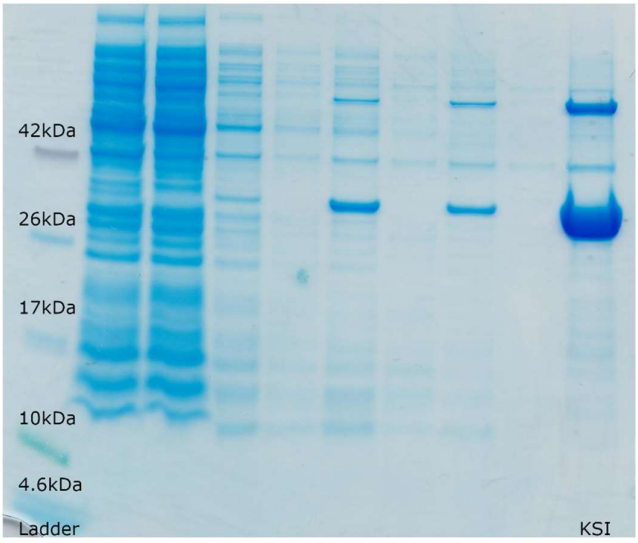


Figure 57: Cropped and labelled, but otherwise unedited SDS-page of washed KSI[C, Variant 2] fusion protein IBs. A protein ladder is included for size comparison. Other wells were used to check for protein loss during processing.

6.6.8 Unedited Analysis of His[C] Purification

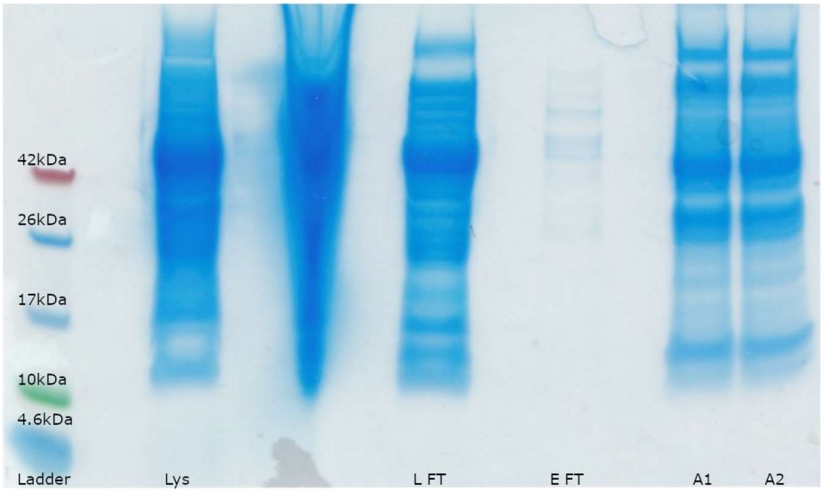


Figure 58: Cropped and labelled, but otherwise unedited SDS-page analysis of His[C] purification fraction A1 and A2, elution flowthrough (E FT), and loading flowthrough (L FT). A protein ladder is included for size comparison. SDS-page of His[C] lysed soluble fraction failed.