Characterisation of a Gram-positive secreted protein signal sequence (SPss): localisation and processing during transport in a Gram-negative bacterium

by

Matthew Fisher

Student no. 11272164

Supervisor: Prof. Jacques Theron

Co-supervisors: Dr. Robyn Roth,

Dr. Michael Crampton

Submitted in fulfilment of the requirements for the degree

Master of Science

in the Faculty of Natural and Agricultural Sciences

Department of Biochemistry, Genetics, and Microbiology

University of Pretoria

April 2023

Declaration

I, Matthew Fisher, declare that the dissertation, which I hereby submit for the degree MSc (Microbiology) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this, or any other, tertiary institution.

SIGNATURE:  DATE: 21/4/2023

Acknowledgements

I would like to thank my family and friends for their support during a time where things seemed to grind to a halt and delays abounded. Through a global pandemic, lockdowns, strains in finance and health this project is finally done.

Thanks to the my supervisors and the staff at the University of Pretoria and the Council for Scientific and Industrial Research who persisted in trying times.

Summary

Characterisation of a Gram-positive secreted protein signal sequence (SPss): localisation and processing during transport in a Gram-negative bacterium

by

Matthew Fisher

Supervisor: Prof. J. Theron
Department of Biochemistry, Genetics and Microbiology
University of Pretoria

Co-supervisors: Dr. R. Roth
Dr. M. Crampton
Biosciences Unit
Council for Scientific and Industrial Research

for the degree MSc Microbiology

Secreted proteins offer the potential to reduce purification cost to enzyme and therapeutic proteins. While many bacterial transport signal sequences have been tested, few have been used across species. This study explores the effectiveness of a Gram-positive secreted protein signal sequence (SPss) used in a Gram-negative production platform (*Escherichia. coli* BL21(DE3)).

A pET bacterial expression plasmid was constructed with a secreted protein signal sequence (SPss) from a secreted protein originating in *Clostridium perfringens* and combined with a common marker protein (Green Fluorescent Protein). The construct (SPss-GFP) was expressed in *E. coli* and tested under a variety of temperatures (25°C, 30°C, 37°C), IPTG concentrations (1µM, 10µM, 100µM, and 1000µM), and two growth media (Lysogeny Broth and an enzyme-release complex medium) to ascertain the ideal expression conditions.

The results indicated that the SPss-GFP protein was present in the periplasm after expression, but transport was sporadic with most experiments showing only cytoplasmic GFP. More periplasmic protein was present when a control plasmid, expressing the full-length secreted

protein from *C. perfringens* was used. The parameters that resulted in the highest levels of GFP was 100µM IPTG at 30°C grown in the enzyme-release medium. High levels of expression and transport were also observed at 1000µM, 37°C and at 1000µM, 30°C. Reduced cell concentration and rate of growth was observed during expression of SPss-GFP, but no parameter was found to be the cause and requires further investigation. Mass spectrophotometry indicated cleavage of SPss but much of the GFP remained in the cytoplasm.

Based on the results obtained in this study, it is concluded that protein folding kinetics and limited interaction with accessory transport factors are the main problems with protein transport in this system.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

List of Abbreviations

ABC:	ATP-binding cassette
Ala:	Alanine
AP:	Alkaline Phosphatase
ASP:	Accessory Secretory Proteins
ATP:	Adenosine Triphosphate
°C:	Degrees Celsius
Cys:	Cystine
dH_2O:	Distilled water
DNA:	Deoxyribonucleic acid
dNTP:	Deoxynucleotide triphosphates
EDTA:	Ethylenediaminetetraacetic acid or 2,2′,2″,2‴-(Ethane-1,2-diyldinitrilo) tetraacetic acid
FRAP:	Fluorescence Recovery After Photobleaching
FRET:	Fluorescence Resonant Energy Transfer
g:	Gram
GFP:	Green Florescent Protein
Gly:	Glycine
hrs:	Hours
Ile:	Isoleucine
IPTG:	Isopropylthio-β-galactoside
kb:	Kilobase pairs (1000 base pairs)
L:	Litres
M:	Moles
Met:	Methionine
Min:	Minutes
mL:	Millilitres
µL:	Microlitres
µg:	Micrograms
ng:	Nanograms
nm:	Nanometres

OD_{600}:	Optical Density at light wavelength of 600nm
RCF:	Relative Centrifugal Force
PAGE:	Polyacrylamide Gel Electrophoresis
PBS:	Phosphate Buffered Saline
PCR:	Polymerase Chain Reaction
Phe:	Phenylalanine
Pro:	Proline
Ser:	Serine
Sec:	General secretion system
SDS:	Sodium Dodecyl Sulphate
SPss:	Secreted Protein Signal Sequence
SPss-GFP:	SPss joined with GFP
SRP:	Signal Recognition Particle
Tat:	Twin-arginine transport
TAE:	Tris, acetic acid and EDTA buffer
Tris:	Tris(hydroxymethyl)aminomethane or 2-Amino-2-(hydroxymethyl)propane-1,3-diol
Trp:	Tryptophane
V:	Volt
w/v:	Weight per volume

Chapter 1: Literature Review

1.1 Overview of microbial biotechnology

The complex world of micro-organisms has been explored in-depth but remains a highly fascinating field. Microbes occupy most known environmental niches and maintain an enormous range of biological activity; everything from disease causing and curing proteins, structural components of a cell, to enzymes that can alter and degrade a wide variety of compounds. This activity depends on the protein's conformation and its expression. One well-studied example is the production of antibodies in a range of host organisms (Elander, 2003). Using similar production methods of microbial fermentation, proteins can be produced, theoretically, from any organism in a chosen host cell. The proteins produced in this way are termed heterologous proteins and are defined as genes (encoding proteins) that are sourced from other organisms and expressed inside a specific cellular host to increase expression and yield of the desired protein. These heterologous genes are introduced into the host cell using genetic manipulation, while the heterologous protein is separated from the other host cell proteins using various purification methods (Shokri et al., 2003).

Of notable interest are heterologous proteins that can have a direct therapeutic effect for humans (*i.e.* hormones, cytokines, antibodies, and other proteins with medicinal applications) or they can have catabolic or anabolic activity (*i.e.* enzymes) which is useful in many instances where a specific chemical change is needed in a molecule. The main advantages of using microbes to produce such heterologous proteins are mainly because microbes grow quickly on a relatively cheap growth medium. However, depending on the organism and/or heterologous protein, the ease of use, activity, and expense of these proteins can change drastically (Chhetri et al., 2015). Some drawbacks of these systems are incompatibilities such as with the host producing the protein (de Marco, 2009) or with a therapeutic protein and a patient (Riemann & Schröder, 2005), and lack of specific metabolic pathways (*i.e.* post-translational modifications, or no enzymes to produce the metabolites). These drawbacks are interrelated because they rely on the presence of specific genes, their expression, and the correct protein folding conformation to produce a functional product.

Challenges faced during the production of heterologous proteins revolve around the host being in optimal conditions, which range from growth factors required by the host in the growth medium, optimum temperatures, gas exchange, as well as the stress of the host cell due to overexpression of the heterologous protein (Makrides, 1996). Any of these factors can be a detriment to protein expression and cellular growth which result in several problems including reduced cell viability, low target protein yield, or loss of the heterologous plasmid (Overton, 2014).

Separation of the heterologous protein from the host proteins and other cellular material can be a significant cost of production. Keeping the protein functional or returning the protein to a functional conformation is the priority after production in the microbial host. At each stage of these processes, a loss of protein can occur which affects the overall yield of functional protein. Reducing the steps in the purification process would, therefore, be a major benefit to the cost-effectiveness of heterologous protein production.

As time has gone by, more and more information and resources are available to the modern biotechnologist to deal with some of these challenges. But, as with most biological systems, these elements are not always interchangeable between cellular systems, and combinations of genetic and biological elements do not always work in harmony with each other. Such elements are the native protein transport systems within all cellular life (Yen et al., 2002). These transport systems allow the cell to localise proteins where they are required for general host functioning or specialised features of a cell. In the model organism *Escherichia coli,* transport systems exist for proteins in different states (folded vs unfolded, cytoplasmic, periplasmic, extracellular, or integrated into the cell membrane, as well as proteins involved in pathogenesis, antibiotic resistance, biofilm formation or intercellular communication) (Green & Mecsas, 2015). While these transport systems are not completely understood, some peptide signals have been identified that facilitate the transport of a protein through some specific transport systems. However, whether these systems recognise peptide signals from other domains of bacteria is not a well-studied area. While these transport systems exist in most cellular organisms, differences do exist and whether signal peptides are conserved enough to be used in different organisms is often not clearly defined. The approach of producing a heterologous protein often requires a trial and error approach where combinations of host, target protein, and gene expression must be carefully considered, while

tracking changes to expression of the host genes to asses any consequences to foreign protein expression (Berlec & Štrukelj, 2013). As more elements of protein transport and expression are discovered, more systematic approaches can be used. For example, combining the use of signal peptides, predictive protein folding, co-expression of chaperones, membrane disruption and selective protein separation strategies may aid in purification of heterologous proteins.

In this study a signal peptide derived from the epsilon toxin of the Gram-positive bacterium *Clostridium perfringens* was used to assess the effects of protein transport in a Gram-negative host (*E. coli),* and whether the efficacy of protein transport changes when this signal peptide is used to express and transport a heterologous protein in *E. coli.*

1.2 Bacterial transport systems

Transporting proteins throughout, within, and outside of the bacterial cell requires complex interaction of multiple protein systems. A large variation exists in the functions, structures and recognition of these systems; not all of which are very well understood. Much of the understanding of these systems revolves around what bacteria use them, the protein subunits that form the transport structure and what triggers their formation, how the target protein is recognised and processed during transport, and any accessory proteins that form part of the transport process (and their interaction with the system as a whole). These aspects are not always clearly defined and many gaps in the literature remain but systematic investigation has given a large amount of information, some of which will be discussed below.

1.2.1 Protein transport in Gram-negative bacteria

The secretion systems in Gram-negative bacteria, like *E. coli*, exist in a variety of forms. Focusing solely on protein transport, there are several notably conserved different peptide selective transport systems (Shokri et al., 2003). Some span the double membrane, allowing transport from the cytoplasm to the extracellular matrix (*i.e.* into the growth medium or external environment). Some systems have evolved to deliver pathogenic peptides into another cell. Other systems only transport peptides across the cytoplasmic membrane, while

others still, only transport proteins across the periplasmic membrane (Kleiner-Grote et al., 2018). These transported peptides can either reside in a specific cellular compartment or be integrated into the membrane. A bacterium would use these systems to transport enzymes, quorum sensing molecules, pathogenic proteins, or any potential protein used in the functioning or metabolism of the cell. The peptide sequence and conformation dictates which protein system is utilised and where the final location of the protein will be (Costa et al., 2015).

The most ubiquitous transport systems are the General Secretion system (also known as the SecYEG, or SRP pathway (Chatzi et al., 2014; Green & Mecsas, 2015; Paetzel et al., 2002)) and the Twin Arginine Translocation system (abbreviated Tat system) (Costa et al., 2015; Nivaskumar & Francetic, 2014). Both systems possess an N-terminal peptide signal (also referred to as a secretion signal) that allows the peptide to be recruited and transported through the respective systems (Natale et al., 2008). The Tat system transports folded proteins, while the Sec transports unfolded proteins. Both systems transport proteins across the cytoplasmic membrane, from there they can be transported into the extracellular space, remain in the periplasm, or integrate within a membrane.

The Sec system also transports proteins that integrate into the cytoplasmic membrane. Both transport across, and integration with, the cytoplasmic membrane is mediated by SecYEG. However, differences in the N-terminal sequence will dictate what chaperones bind and the destination of the protein. The Signal Recognition Particle (SRP) is a ribonucleoprotein that binds to the N-terminal sequence in a similar way to the SecB chaperone (Schneewind & Missiakas, 2014). The SRP specifically binds to a hydrophobic region of a protein as it is translated; this region will eventually be the anchor of the peptide when it is integrated into the membrane. Recruitment of another chaperone, Ftsy, then helps shuttle the protein to the SecYEG complex where the unfolded peptide is threaded through until the SRP is released and the hydrophobic section embeds in the cytoplasmic membrane (Papanikou et al., 2007).

1.2.2. Sec- and Tat-dependent extracellular protein transport

Once a protein has been transported through the cytoplasmic membrane the Type II and Type V secretion systems in *E. coli* facilitate transport of proteins from the periplasmic space to the extracellular space (Korotkov et al., 2012; Van Ulsen et al., 2014). Proteins transported through either the Sec or Tat transport system will be folded and used as substrates for the Type II transport system. The Type II system has a barrel structure spanning the periplasmic space, a base attached to the cytoplasmic membrane and a selective pore on the periplasmic membrane (Korotkov et al., 2012). A key feature that is not fully understood is the assembly of the barrel and integration of the target protein for transport. What is known is that this system accepts folded proteins from the Sec and Tat systems (Sandkvist, 2001).

Proteins transported through the Sec system (but not the Tat system) can be transported through Type V secretion system. This system is characterised by a β-barrel that forms a channel on the periplasmic membrane. The formation of this channel can occur in three different ways and so are divided into three classes of Type V secretion: Autotransporters, Two-partner secretion, and Chaperone/usher secretion (Costa et al., 2015). The notable characteristic of the autotransporters is that the peptide sequence of a Type V autotransporter can be contained within the target protein sequence and does not require additional proteins to be transported through the outer membrane (Van Ulsen et al., 2014). The autotransporter domain is a C-terminal region that usually contains a translocator domain (which forms the pore in the membrane), a linker domain, the passenger domain (the protein to be transported), and a protease to cleave the passenger domain from the remaining peptide (Green & Mecsas, 2015).

Other protein transport systems can be generally classified into Types I, III - VII (Figure 1.1), which often involve the transfer of proteins to facilitate some external cell function (such as biofilm formation, infection, or cellular sensing). This is done by transporting the proteins either directly from the cytoplasm into the extracellular space or across external cellular membranes during infection or cellular communication (Costa et al., 2015; Holland, 2004).

1.2.3. Sec-independent extracellular protein transport

Type I secretion systems transport proteins from the cytoplasm to the extracellular space using a transmembrane channel that spans both the cytoplasmic and periplasmic membranes. This transmembrane region consists of three parts: the cytoplasmic ATP-binding cassette (ABC), a membrane fusion protein, and an outer membrane factor (Green & Mecsas, 2015). To be transported through this system a C-terminal amino acid sequence is required (Holland et al., 1990). When the peptide is recognised and binds to the ABC region the other elements of the Type I system are recruited to form the complete transport system (Masi & Wandersman, 2010). The assembly of the transport channel is transient and is thought to be initiated when a peptide binds to the ABC region in the cytoplasm (Holland et al., 2005). Other elements of peptide recognition have been identified in this secretion system that include the binding of the chaperone SecB to keep the target protein in a transportable state (Masi & Wandersman, 2010).

Although the C-terminal region of the peptide is important for transport recognition, it is not the only element required for successful transport through the Type I secretion system (Masi & Wandersman, 2010). Recognition of a transported protein, the assembly and disassembly of the system is not completely understood. However, multiple proteins such as TolC, an ABC-transporter, and a few others, assemble into a membrane-spanning protein complex through which a protein can be transported (Masi & Wandersman, 2010). A C-terminal signal is required for transport and some signals have been elucidated (Holland et al., 2005). However, other sections of the transported protein interact with the Type I system and chaperones like SecB; transport then utilises ATP hydrolysis and the C-terminal region is thought to act as a disassembly signal, returning the Type I to a pre-transport state (Masi & Wandersman, 2010).

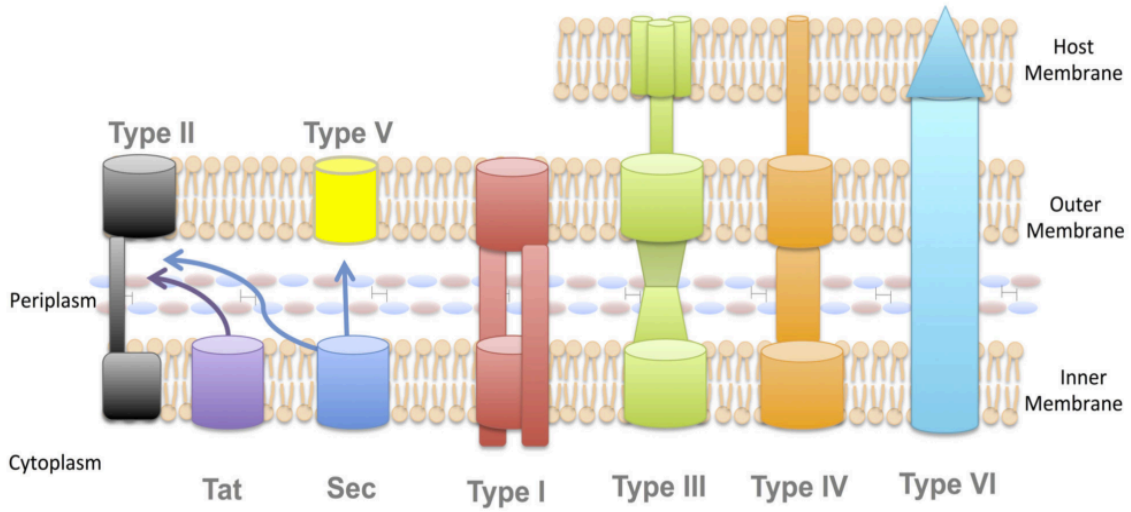


Figure 1.1: Secretion systems of Gram-negative bacteria (Green & Mecsas, 2015). *Inner and outer membranes refer to the cell system in which the transport systems are embedded, while the host membrane is an external cell. The host cell membrane shows which systems typically transport protein effectors as part of pathogenic infection of other cells.*

Type III, IV and VI secretion systems are all large assemblies of protein channels that facilitate protein transport from the cytoplasmic space, across the outer periplasmic membrane, and into a neighbouring cell; this often facilitates the transport of protein effectors needed during infection of a host cell by a given microbe (Green & Mecsas, 2015). Differences in the protein structure of the channel, transported proteins, and function of these systems characterise these secretion systems into their respective groups.

Type III secretion systems are sometimes referred to as injectosomes or needle-syringe apparatus. Although initially the transport was considered to take place in one step, evidence of multiple steps have arisen (Green & Mecsas, 2015). Closely related to flagellin, the Type III system has nine core proteins with approximately an additional 20 proteins that are also required for complete functioning. Most of the substrates for the Type III system are considered to be effector proteins that aid in infection or host disruption. Some of these effectors require chaperones to guide them to the transport apparatus and are thought to be transported in an unfolded conformation (Green & Mecsas, 2015; Schechter et al., 2004). Notably the N-terminal signal sequence of these effector proteins are not cleaved after or during transport (Galan & Wolf-Watz, 2006).

The Type IV secretion system has a large variation in transported substrates (both protein and DNA), and is present in a large variety of Gram-negative and Gram-positive bacteria. This

secretion system is used for both pathogenesis and cell-cell communication, but has been prominently studied in pathogenic organisms, and as a mode of plasmid exchange (Costa et al., 2015; Green & Mecsas, 2015). A C-terminal sequence is sufficient for protein transport in some cases, but other, poorly defined, conserved regions are also required (Hohlfeld et al., 2006).

Type VI is the most recently discovered secretion system and is thought to play a role in virulence and microbial niche competition (Green & Mecsas, 2015). A cluster of 13 genes encode the Type VI transport apparatus with several other genes encoding pathogenic effectors or a protein that prevents self-toxicity from those same effectors used in infection or competition (Green & Mecsas, 2015).

1.2.4. Protein transport in Gram-positive bacteria

While Gram-positive bacteria do not have a second cell membrane, they do possess a layer of peptidoglycan as a barrier to the external environment. Protein transport across these layers occurs through the Sec and Tat systems (similar to Gram-negative organisms), but other mechanisms exist for protein transport, such as SecA2, various sortases, and the Type VII secretion system found in some *Mycobacterium* species (Feltcher & Braunstein, 2012; Freudl, 2013; Green & Mecsas, 2015).

The SecA2 is a co-factor that aids in transport of proteins through the SecYEG complex (the most common form of transport), but can also facilitate transport through another transmembrane channel (SecY2). The SecA2 and SecY2 factors can transport glycosylated proteins, but still transport the protein in an unfolded conformation before they are folded outside of the cytoplasm. Transport of glycosylated proteins also requires additional co-factors termed Accessory Secretory Proteins (ASP; of which five have been studied); however, their exact mechanisms of action are not well understood. It is thought that the ASP 4 and 5 co-factors share similar roles as the SecE and SecG proteins in that they are associated with the membrane channel, and aid in transport (Feltcher & Braunstein, 2012). The substrates recognised by SecA2 can also facilitate the transport of unglycosylated proteins through the

SecYEG complex, but glycosylated proteins will utilise the SecY2 channel (Feltcher & Braunstein, 2012).

Sortases are another route a protein can take during transport. While initially the Sec system is used for transport through the cytoplasmic membrane, a C-terminal transport signal is then recognised by a given sortase and then anchored to the cell wall using isopeptide bonds (Hendrickx et al., 2011). This route is often utilised by proteins that become pili or a variety of other external pathogenic or housekeeping proteins, such as proteins involved in cellular adhesion, heme uptake and immune system evasion (Schneewind & Missiakas, 2014).

A type VII secretion system, also referred to as an injectosome, has been identified in several Gram-positive bacteria such as mycobacteria. Although functionally like Gram-negative injectosomes, there are differences in protein structure and homology as these structures must span a highly lapidated cell wall (Green & Mecsas, 2015). Both are used for injecting pathogen proteins into a host cell during infection.

1.3 Signal peptide recognition and processing

1.3.1 Elements of transported peptides

Studies into the General Secretion system and the Tat system showed conserved N-terminal sequences in the peptides that were transported using these secretion systems (Berks et al., 2000; Knoblauch et al., 1999; Natale et al., 2008). The peptide motifs have a similar sequence and composition of amino acids (*i.e.* regions of charged, hydrophobic, and polar amino acids in the N-terminal region before the cleavage site). These motifs allow recognition of the peptides by co-factors as well as the transport apparatus itself and facilitate the ushering of either unfolded proteins (in the case of the General Secretion system), or folded proteins (in the case of the Tat system) (Palmer & Berks, 2012). Notably, the Tat system recognises a pair of arginine residues (used to name the Twin arginine transport system) in the signal sequence that is highly conserved, and is one such element that distinguishes recognition and transport of proteins labelled with the Sec or Tat signal peptides (Natale et al., 2008; Palmer & Berks, 2012).

The signals that allow transport of proteins through Type I - VII secretion systems are not as clearly defined as those in the General and Tat secretion systems. Type I, for example, uses a C-terminal region of amino acids that does not have a clear motif, with evidence that additional elements on the polypeptide are required for successful transport through the Type I apparatus (Holland et al., 2005; Masi & Wandersman, 2010). Secretion systems Type III - VII are mainly associated with pathogenesis and the transport of effector proteins, but can also be involved in other protein transport as well as DNA transfer or a combination of all three (Hohlfeld et al., 2006; Holland, 2004; Samudrala et al., 2009; Schechter et al., 2004). It has been very difficult to detect the exact signals that dictate which secretion system a given protein will be transported through. In general, signal-sequence analysis is focused primarily on N- or C-terminal peptide regions but internal regions can also play a role, as well as the recruitment of specific protein chaperones that help direct the transport of proteins to specific secretion systems (Hohlfeld et al., 2006; Holland et al., 2005; Masi & Wandersman, 2010; Samudrala et al., 2009; Schechter et al., 2004). Type II, IV, and V secretion systems can transport proteins already in the periplasmic space (for Type V this is a prerequisite – Types II and IV can transport both cytoplasmic or periplasmic proteins). In this case, type V transport requires two signals, the N-terminal domain (*i.e.* general secretion signal) and a C-terminal "helper" signal that allows transport through the Type V apparatus, located on the outer-membrane (Henderson et al., 2004).

The way type VI and VII systems recognise and facilitate transport of a given protein is not well understood. The function of these systems tends to be related to pathogenesis and effector transport (Freudl, 2013). Originally thought to only exist in mycobacteria, the Type VII secretion system appears to have similar genes encoding for the transport apparatus, co-factors, and effector proteins in other Gram-positive bacteria such as *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis* (Costa et al., 2015; Freudl, 2013; Green & Mecsas, 2015).

The signal peptides in proteins transported by the Tat and Sec systems (as well as other systems that utilise N-terminal signal sequences) facilitates binding to the transport apparatus, as well as recognition by a separate cleavage and degradation protein embedded in the membrane called a signal peptidase. These peptidases exist in bacteria, mammals and archaea, and occur in all instances where protein transport occurs (Paetzel et al., 2002).

Recognition by the peptidase requires the specific signal sequence motifs discussed earlier (the charged N-, hydrophobic H-, and carboxy C- domains prior to the cleavage site). The hydrophobic domain seems to be of particular importance during the recognition step for transport and subsequent processing (Hoyt & Gierasch, 1991).

Experiments with Alkaline Phosphatase (AP), a periplasmic protein in *E. coli,* have shown that the few amino acids before the cleavage site are also particularly important (Karamyshev et al., 1998). Mutations at positions -1 and -3 totally prevented AP transport. Only Ala, Gly and Ser were tolerated in position -1 and -3, with the addition of Leu and Cys at position -3 only (*i.e.* neutrally charged amino acids). Mutations at position -4 were inconsequential, but large- and medium-sized residues at positions -2 and -5 affected the rate of transport (Karamyshev et al., 1998). Alterations of position +1 interfered less with transport, as Trp, Ile, Phe, Met and Pro were not found in transported peptides but only Pro that replaced an Arg prevented cleavage but allowed for recognition (Karamyshev et al., 1998).

1.3.2 Signal peptide processing in the General Secretion system

While it is possible for certain proteins to be transported without alteration, it is common for transported proteins to undergo alterations or processing before, during, or after transport. The N-terminal region associated with the General Secretion system is known to be cleaved off during the transport process. Simply put, before transport is complete, chaperones are recruited as the N-terminal sequence is being translated, to ensure the peptide remains in an unfolded state before being shuttled to the secretion system, located in the cytoplasmic membrane, where it is transported through the secretion system into the periplasmic space. The N-terminal region is cleaved and degraded during recognition and transport, while the mature peptide is released into the periplasm (Freudl, 2013; Tsirigotaki et al., 2017).

The process of transporting a protein requires several co-factors that alter the properties of the peptide that they bind to. In the case of the General Secretion system, the co-factor SecB binds to the peptide during ribosomal translation and ensures the peptide does not fold. The SecA co-factor then binds to the peptide and directs the peptide to the SecYEG translocase (a transmembrane protein spanning the inner membrane of the cell). The SecA co-factor then

utilises ATP to provide the energy needed for transport (Chatzi et al., 2014; Green & Mecsas, 2015).

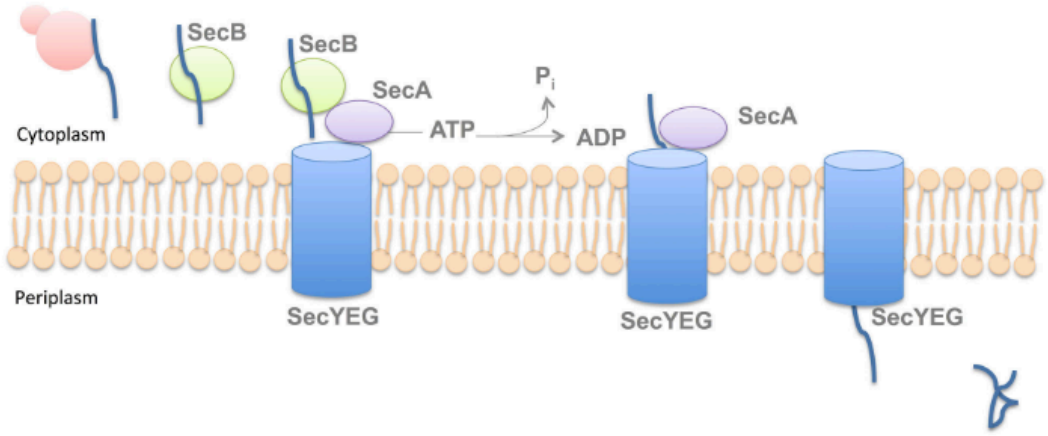


Figure 1.2: General Secretion system and associated co-factors SecA and SecB that facilitates translocation of a peptide (Green & Mecsas, 2015).

During protein transport, proteins in the membrane interact with the signal peptide to either anchor the protein into the membrane (in the case of SRP and membrane protein transport) or cleave and degrade the signal peptide (in more common SecYEG transport) (Schneewind & Missiakas, 2014).

While the protein is transported (or shortly after), a signal peptidase cleaves the signal peptide from the remainder of the transported protein. It must be degraded, or else it is thought to cause interference with protein transport and membrane stability (Martoglio et al., 1997). The signal peptidase (commonly Signal Peptidase I in bacteria) cleaves the signal peptide from the protein, which can occur in parallel with translation and transport (Auclair et al., 2012). The signal peptide is then rapidly degraded by peptide hydrolases (Paetzel et al., 2002).

1.3.3 Signal peptide processing in the Twin Arginine Translocation (Tat) system

The Twin Arginine Translocation system also transports proteins across the cytoplasmic membrane but in a folded state. The structure of the N-terminal signal sequence on both Sec and Tat systems is similar in that they both have a sequence of charged, hydrophobic, and

polar regions (designated n-, h-, and c-regions). Differences in length and the presence of two arginine residues at the boundary of the n- and h-region are key differentiating factors for these transport systems (Natale et al., 2008).

The transmembrane protein of the Tat system (TatA) serves a similar purpose as the SecYEG complex in the Sec system, namely to transport the protein across the cytoplasmic membrane. The energy source for Tat transport is from an ion gradient of H^+ (similar to that found in mitochondria and chloroplasts) (Müller, 2005). This differs from the Sec system which uses ATP, recruited through the SecA co-factor (Figure 1.2).

The most likely route for Tat transport of a folded peptide begins with the binding of TatB and TatC co-factors to the signal peptide region of the folded peptide. The protein-co-factor complex then binds to TatC to form a transmembrane transport complex on the cytoplasmic membrane. The folded protein is then transported into the periplasm and the signal peptide is cleaved (Palmer & Berks, 2012).

1.4 Utilising signal peptides in heterologous protein production

Bacterial secretion systems are undoubtedly complex. With a wide range of transported proteins, DNA and other molecules, the applications of signal peptides are extensive. The most straightforward application of these signal peptides is using them to transport heterologous proteins in highly productive strains of protein-producing cells. The combination of a signal peptide to a heterologous protein, if successful, would improve the efficiency of production and purification by reducing the need for complex protein separation procedures. This is due to the reduced amount of protein in the periplasmic space and extracellular space, as well as the potential for improved folding of the heterologous protein in the periplasmic space of *E. coli* (Hsu et al., 2016). In common practice, purification of proteins via insoluble inclusion bodies is sometimes preferred because of high yield and simple purification; however, refolding of the target protein is required and the inclusion bodies can incorporate other cytoplasmic proteins, making further processing essential (Wurm et al., 2018). Using a signal peptide to transport a target protein will reduce the purification and refolding processes downstream of production.

In most examples of protein transport in literature a native protein that is transported in the host cell is identified. The signal sequence is isolated and combined with a target peptide sequence, and then expressed in a plasmid. In some cases, the entire coding sequence of a transported peptide is combined with the target sequence and expressed, but this carries with it its own set of problems, such as the expression of a very large peptide in which only a small portion is transported (Campo et al., 2004; Choi & Lee, 2004; Kleiner-Grote et al., 2018; Natale et al., 2008; Osborne et al., 2005; Shokri et al., 2003).

When combining a signal peptide found in another bacterium there may be problems of compatibility or recognition of the foreign signal peptide and the native transport apparatus of the host cell. While there are many common elements and pathways for protein transport, there is no guarantee that a signal peptide will function effectively in a different cellular environment. This is equally true when considering the compatibility of the target protein with a signal peptide. Recognition elements on a peptide are essential for correct and efficient protein translocation in a cell (Freudl, 2018; Gao et al., 2016; Heijne, 1983; Simmons & Yansura, 1996). When these factors are considered together with more common problems of heterologous protein production, such as expression rates, medium composition and managing the cellular stress response, the challenge of effective production becomes even more difficult to solve and more rewarding when these challenges are managed correctly.

The gene fusion process can be as simple as isolating a gene of a known, transported protein and combining it with a gene for a target heterologous protein, and is employed when attempting to track protein localisation (Hynes et al., 2004). This can be good for answering research questions in terms of transport routes and proof-of-concept expression systems. However, the production of more proteins, especially large ones, can be detrimental to production efficiency due to the increased burden that protein production has on the host cellular systems, including transport systems (Baumgarten et al., 2018; Schlegel et al., 2013; Wagner et al., 2008).

Combining a signal peptide with a heterologous protein requires fusing the genes together using biotechnology techniques, commonly using polymerase chain reaction (PCR) and restriction enzymes, and N-terminal sequences are more commonly combined with a target protein than C-terminal signal peptides (Gao et al., 2016; Pechsrichuang et al., 2016;

Samudrala et al., 2009). The more common utilisation of N-terminal secretion signals is likely because these signals and transport systems (specifically the General Secretion system) are common in all cellular life and account for the majority of transported proteins (Costa et al., 2015; Orfanoudaki & Economou, 2014). Many *in-silico* analysis techniques have also been developed to detect N-terminal regions that code for signal peptides (Käll et al., 2004, 2007; Orfanoudaki & Economou, 2014).

When it comes to assessing the effects of taking a signal peptide from one organism and expressing it in another organism, there can be more issues that arise. While the General Secretion system is universal, evolutionary divergence could make these signals incompatible when transferred between species (Low et al., 2013). Mutations in genes, while likely conserved, could make certain protein interactions less effective and thus affect transport efficiency (Caspers et al., 2010).

Signal peptide gene fusion is only the first step in utilising a transport system during protein production. Optimal expression, protein interactions and processing need to occur for a protein to be effectively transported. As discussed in Section 1.3, the recognition and cleaving of a signal peptide occurs at the transport apparatus and any interruption of this process can cause blockages in the transport system, resulting in reduced protein yields and cellular stress that reduces growth (Berlec & Štrukelj, 2013).

Transport of *E. coli* proteins in other Gram-negative bacteria has been successfully observed, but experiments utilising Gram-positive proteins and their transport in *E. coli* are less prominent (Wong & Buckley, 1993). Even when utilising Gram-positive secretion signals in a Gram-positive host such as *Bacillus subtills,* there can be incompatibility at the protein level which prevents transport (Brockmeier et al., 2006). It is important, then, to test the combinations of signal peptides and target proteins in addition to host attributes such as different cell wall types in bacteria. Presently the only methods for characterising signal peptide efficiency are *in-silico* analysis (using tools like SignalP or other computer-based tools), and experimental testing. Experiments involve combining multiple promoters, signal peptides and target proteins and testing transport efficiencies, either manually or using automated, high throughput screening (Wu et al., 2020; W. Zhang et al., 2016). Mutating the

signal peptide can also provide additional ways to improve transport efficiency (Caspers et al., 2010).

1.5 Other challenges associated with heterologous protein production

Heterologous proteins produced in excess can lead to other challenges when trying to produce, isolate and utilise the target protein. When produced, the heterologous protein accumulates within a cellular compartment (typically the cytoplasm) to high levels, as much as 50% of total cellular protein (Studier & Moffatt, 1986). This excessive expression, while desirable, can disrupt cellular processes and cause stress to the host cell (Schweder et al., 2002). The cellular stress response can reduce overall yield and lead to target protein degradation as misfolded and aggregated proteins tend to be the first targets for degradation (Maurizi, 1992). Along with protein degradation, cellular growth is down-regulated while stress response factors are upregulated (Maurizi, 1992). These factors must be managed correctly to effectively produce high yields of heterologous protein.

Hosts with knocked out genes coding for proteases has been a common strategy for improving heterologous protein yields and such *E. coli* strains already exist (Makrides, 1996; Meerman & Georgiou, 1994). However, protease activity can still be detected in some strains expressing heterologous proteins (Ellis et al., 2017). Protease activity is linked with the stress response, but less activity is observed in the periplasmic space and extracellular space (Choi & Lee, 2004; Harcum & Bentley, 1999). This means that successful transport of a peptide can bypass most proteases, and if the expression is managed well, stress responses could be kept to a minimum. Harmonising all of these factors could provide optimal expression and high yields of heterologous protein.

1.5.1 Heterologous protein purification

Arguably the costliest part of heterologous protein expression is separating the target protein from other host proteins and other metabolites (like lipids) and the growth medium. The separation requires multiple steps and specialised materials and reagents, all of which must not be harsh enough to damage or disrupt the protein (Lightfoot & Moscariello, 2004). The

goal of protein purification is to obtain a specific protein at a high purity. This is a standard set out and can differ based on the application of the protein and its origin. For example, some bacterial expression systems produce toxins that must be separated out and require extremely precise and effective methods to achieve a pure product (Nfor et al., 2008).

Purification (also called downstream processing), broadly speaking, follows a common workflow: recovery of the protein source, release of the proteins from the source, separation, and concentration of the target protein from other metabolites, removal of further impurities and then formulation of the target into a stable form that can be used, for example, for further analysis or application as a medicine (Zydney, 2016). A review of common procedures is outlined in Nfor *et. al.* (2008) and will be expanded on below. It is cited that a combination of these procedures requires extensive testing and a good knowledge base, or even heuristics (rules of thumb) to get an efficient purification process (Lightfoot & Moscariello, 2004).

The application of procedures depends on the target protein and the host in which it is produced. In a cellular suspension, that is grown in bioreactor, the target protein can reside within the cell (or cellular compartment) or in the growth medium (extracellular space). Different forms of centrifugation or filtration are then applied to concentrate the protein-containing cells, or the protein residing in the growth medium (Jungbauer, 2013; Zydney, 2016). The cell needs to be disrupted to release the protein, at which point various extraction techniques are used to separate proteins from other cell debris, and other types of proteins. So far, this is a common procedure for more soluble proteins, but insoluble proteins often form inclusion bodies, which are aggregates of protein. These inclusion bodies are often easier to purify and require fewer steps of separation from other metabolites, but they do need to be denatured and then refolded (Singh et al., 2015).

Once the proteins have been recovered from the bioreactor, using some of the processes described above, a variety of methods are available for concentrating the proteins into a smaller volume. These include evaporation, reverse osmosis filtration, precipitation, or crystallization (Jungbauer, 2013; Zydney, 2016). Once concentrated the target protein can be purified further, commonly using some form of chromatography. Once purified to a satisfactory level, the target protein can then be formulated into a tablet, crystal, lyophilised, or in other ways that preserve its function and activity (Jungbauer, 2013; Zydney, 2016).

In some cases, processing steps must be repeated to obtain good results. Repeated filtration, dialysis, precipitation, or chromatography may be needed to properly recover and purify a protein (Jungbauer, 2013; Zydney, 2016). Paying close attention to the characteristics of the target protein as well as related proteins produced in the host is important for this. Proteins that have similar characteristics of size, charge, and hydrophobicity, for example, will mean that it will be more difficult to separate and require different types of filters and columns (Wilken & Nikolov, 2012; Zydney, 2016).

To aid in separation and purification, a tag can be attached to a target protein during the design phase of heterologous protein production. This tag can aid in purification or localisation of the protein that can help with isolating it from other proteins or metabolites (Wingfield, 2015). For example, a poly-histidine tag can be used in a column containing a Ni^{2+} -nitrilotriacetic acid ligand to which the histidine tag can bind (Demurtas et al., 2013; Gräslund et al., 2008; Makrides, 1996). Some tags can also aid in improving yields and provide protection from proteolysis, and be cleaved off during production (Makrides, 1996). This depends on the fusion partner and must be taken into consideration for the entire process, because of cost impacts on production and downstream processing. This includes the number and type of columns, filters, centrifuges and associated reagents and skilled staff needed by a given facility to successfully purify a given protein (Zydney, 2016).

Some important factors when considering a complete purification process can be summarised as follows: remove the most plentiful impurities first, start with the easiest impurity to remove, the most expensive or difficult purification step should be last, make use of methods that take advantage of the biggest differences in target and impurity properties, and select for a sequence of methods that use a variety of different separation mechanisms (Harrison et al., 2015).

As an example, since chromatography techniques can be the most expensive step, centrifugation or filtration processes should be used first. Preferably, any one step of filtration, precipitation, or solvent use should not occur twice together. These steps should have intermediate purification methods that differ significantly enough to maximise the efficiency of the entire purification process (Zydney, 2016). A process for purifying an insoluble protein produced in *E. coli,* would start with concentrating the intact cells using

centrifugation, lysing the cells to release the proteins, recovering the insoluble fraction, and redissolving the proteins. Then, at this stage, filtration and solvent separation can be performed followed by column chromatography to gain a purer form of protein (Wilken & Nikolov, 2012; Zydney, 2016). This process can be improved with addition of specific protein tags, mentioned earlier to improve performance of the column (Zydney, 2016).

A soluble protein may present a greater purification challenge than an insoluble protein. While it may not require refolding, a soluble protein is present in solution, either in a cellular compartment or growth medium, along with all other soluble proteins (Wingfield, 2015). A soluble protein is also likely to be in a lower concentration than an insoluble protein either due to its expression levels (i.e., a highly expressed protein is more likely to form inclusion bodies), or because it resides in the external cellular space which has a greater volume than the internal cell volume (either cytoplasmic or periplasmic). This larger volume results in a longer running time for filters and columns. This can be overcome by running several columns/filters in parallel which reduces downtime and fowling (Jungbauer, 2013). Knowing the size, charge, hydrophobicity, and other peptide characteristics (such as the addition of an affinity tag) before production will greatly improve the efficiency of the purification process.

1.6 *E. coli* – a well-studied heterologous protein expression platform

1.6.1 *E. coli* characteristics

The bacterium *E. coli* can be considered one of the best understood organisms. It can occupy a vast variety of environmental niches and has been cultured and manipulated for decades (Blount, 2015; Shiloach & Fass, 2005; Van Elsas et al., 2011). While the initial focus was on selecting productive strains, advances in genetic engineering meant that novel mutants could be better identified and characterised. These strains often possessed useful characteristics, for example, protease-deficient strains, lower acetate-producing strains, better stress tolerance, and additional protein folding factors (Meerman & Georgiou, 1994; Shiloach & Fass, 2005). The factors listed mainly relate to protein expression and are particularly useful to heterologous protein expression, but are by no means an exhaustive representation of the diverse metabolic functions of *E. coli.*

Selecting *E. coli* strains specifically to express heterologous proteins is complex and time-consuming because excessive protein production can interfere with basal metabolic processes and host fitness; yet high yields are an extremely useful trait (Schlegel et al., 2015; Schweder et al., 2002). Therefore, strains that do have useful characteristics are usually retained and shared with laboratories across the world. Initially, *E. coli* isolates were used for the study of bacteriophages and differentiation between strains was understood by the ability of a strain to host and replicate a variety of phages (Daegelen et al., 2009). After many decades of research and further understanding of genetic regulation, strains were kept and mutated, and desirable traits were selected for. These mutations eventually lead to the discovery of the phage T7 RNA polymerase promoter, which is used extensively in heterologous protein expression because of its high affinity for its associated RNA polymerase and resulting high gene expression levels (Daegelen et al., 2009). The *E. coli* strain BL21(DE3) was developed alongside expression vectors containing the T7 RNA polymerase promoter and both have become some of the most common elements in bacterial protein production (Daegelen et al., 2009).

1.6.2 The pET expression system

The pET expression system comprises genetic elements used to express heterologous genes at high levels under various means of control. Many variations exist but with common core elements. In general, gene expression is linked to a promoter region upstream of a multiple cloning site (Rosano & Ceccarelli, 2014). Other elements include origins of replication for both bacterial and phage replication of the plasmid, one or multiple antibiotic resistance genes, as well as other gene control elements like the *lacI* repressor gene (Rosano & Ceccarelli, 2014). Multiple restriction enzyme cut sites are present on the plasmid that can be used to insert DNA and this provides different options for cloning DNA inserts. Certain pET plasmids are optimised for expression in particular hosts, such as the plasmid pET26 and *E. coli* strain BL21(DE3) (Agilent Technologies, 2014; Daegelen et al., 2009).

Variations of plasmid expression systems allow for differential control of protein expression. The pET plasmids commonly utilise IPTG or lactose to induce protein expression in a way that allows very high levels of expression or no expression at all (some basal levels of expression

can occur but mutations to the *lacI* repressor allow for very tight control of expression) (Marschall et al., 2017). In contrast, the arabinose operon can be used to control heterologous gene expression similar to the *lac* operon, but the control mechanism differs slightly (*i.e.* a DNA loop is formed in the transcription region of the leading region of the arabinose operon that represses expression) (Matthews, 1992). The addition of arabinose triggers the expression of genes under the control of a subsection of the arabinose operon. In the wild-type, this facilitates arabinose transport into the cell and its metabolism for energy purposes but isolated promoters from this operon provide another mechanism of gene control that can be modulated to different levels of gene expression (Guzman et al., 1995; Marschall et al., 2017; Schleif, 2010).

The promoters mentioned above are only a small set of examples of regulated promoters and are meant to illustrate two methods by which heterologous protein expression can be controlled. Other examples of regulated gene expression often relate to metabolic regulation or shifts in environmental conditions and investigations into these control mechanism lead to the identification of new regulatable promoters (Anilionyte et al., 2018; Marschall et al., 2017). One example is the P_{bad} promoter isolated from the *araBAD* operon, involved in arabinose metabolism (Guzman et al., 1995). The CYC6 promoter, that can be induced by nickel and repressed with copper, is an example found in *Chlamydomonas reinhardtii* (Ferrante et al., 2008). Other promoters can react to temperature, osmolarity, oxygen (or lack of), phosphate starvation, glucose starvation, or the more conventional lactose substitute, IPTG (Makrides, 1996).

1.7 Advantageous mutations in *E. coli* BL21(DE3)

The BL21(DE3) strain of *E. coli* commonly used for heterologous protein production has been selected for and mutated in a variety of ways that aid in protein production, adaptation to culturing conditions of the laboratory, and host stability while under the stress of excessive protein production (Daegelen et al., 2009; Kwon et al., 2015). These and other mutations have increased the capabilities of *E. coli* BL21(DE3) to produce heterologous proteins that are functional or in extremely high yields. However, limits still exist in these systems, some of them still amenable to genetic alterations (*i.e.* addition of chaperone proteins to improve

protein folding or reduce cell stress), while others are physical limits, for example oxygen uptake (Shiloach & Fass, 2005).

While many factors can influence protein production, some small alterations can result in dramatic improvements. In particular, the BL21(DE3) strain of *E. coli* possesses a mutation in the lacUV5 promotor, which results in reduced activity in the associated inhibitor, *lacI*, which then results in reduced toxicity for membrane proteins and proteins transported using the Sec secretion system (Kwon et al., 2015; Marschall et al., 2017). In this case, the lacUV5 mutation reduces the rate of transcription of any expressed heterologous protein but still results in overall improved heterologous protein expression because of reduced stress in the host. Analogous effects were observed when reducing levels of T7 RNA polymerase separately, which resulted in lower levels of toxicity when the expressed heterologous protein was a membrane protein (Wagner et al., 2008). This seems to suggest that the expression rate is very important in avoiding a stress response during high-level protein expression, especially for membrane and transported protein expression.

Other mutations that have been selected for in *E. coli* BL21(DE3) are knockout proteases, such as OmpT as well as Lon protease, which is an ATP dependent protease active in the cytoplasm of many organisms, including *E. coli* BL21(DE3) (Maurizi, 1992). The knockout of these proteases reduces the degradation of proteins being expressed. While a large variety of proteases exist within *E. coli,* knockout of the OmpT and Lon proteases are sufficient to significantly reduce heterologous degradation and effectively increases yields of the target protein (Maurizi, 1992; Waegeman et al., 2013).

Derivatives of *E. coli* BL21(DE3) strains exist that have been selected to enable membrane protein overexpression (Wagner et al., 2008). The mutations that exist in strains C41(DE3) and C43(DE3) show a better tolerance for membrane protein expression. Experiments have determined that this tolerance is mainly due to a mutation in the lacUV5 promoter region, which reduces its activity and the rate of protein expression and restoration of function of Lon protease in the C41(DE3) strain (Kwon et al., 2015). This results in lower toxicity to the cell, specifically when membrane proteins are over-expressed. The restored functionality of the Lon protease is thought to degrade the misfolded proteins that block the Sec transport proteins in the membrane (Kwon et al., 2015).

Yields of heterologous proteins expressed in *E. coli* can reach a large proportion of the total cellular protein; however, this depends heavily on the construct being expressed as well as any stress that it places on the host during expression. When a protein is successfully expressed, further complications can occur, such as increased protease expression in the host (Harcum & Bentley, 1999; Ramchuran et al., 2003). During the expression of heterologous proteins, the proteins can aggregate into inclusion bodies. While this is advantageous for purification (inclusion bodies are insoluble, making them easier to separate from the native, soluble proteins), the inclusion bodies often require refolding into a native, and active conformation (Singh et al., 2015; Wurm et al., 2018). These additional processes come at a cost; the re-solubilisation of the aggregated proteins as well as turning the proteins into correctly folded proteins in solution can lower the yields expected during production (de Marco, 2013; Mergulhão et al., 2005).

1.8 Green Fluorescent Protein (GFP) as an intracellular transport reporter

The Green Fluorescent Protein, derived from *Aequorea victoria,* is a commonly used reporter protein in biotechnology in bacteria, mammalian and cyanobacterial systems due to its ability to form a chromophore and fluoresce without the need for additional chaperones, or substrates (other than oxygen) (Dammeyer & Tinnefeld, 2012; Mullineaux et al., 2006; Seibel et al., 2007; Spence et al., 2003). Its utility lies in its simple fluorescent protein structure, which consists of a barrel structure made from β-sheets. When exposed to ultraviolet light a green colour fluoresces, but multiple mutations has given a variety of different colour options (Shaner et al., 2005). Applications in tracking subcellular localisation in microscopy experiments is well documented in timelapse experiments that utilise Fluorescence Resonant Energy Transfer (FRET), and Fluorescence Recovery After Photobleaching (FRAP) (Meyer & Dworkin, 2007). GFP can be attached to other proteins as a reporter for subcellular localisation (Hynes et al., 2004). It can also be used to test the effects of signal peptides or entire proteins that are suspected of being transported (Dinh & Bernhardt, 2011). The prolific use, versatile and simple use makes GFP an ideal candidate for use in this study.

1.9 Aims

To investigate the properties of a novel signal peptide, SPss, and the effects it has on protein transport when joined to the common reporter protein Green Fluorescent Protein.

1.10 Objectives

1. Fuse signal peptide with reporter protein in chimeric plasmid
2. Transform *E. coli* with chimeric plasmid
3. Test various parameters of protein expression and cellular growth:
    a. Various induction strengths using IPTG (*e.g.* 0, 1, 10, 100, 1000µM)
    b. Induction temperature (*e.g.* 25°C, 30°C and 37°C)
    c. Standard compared to enzyme-release growth medium (*e.g.* LB and M2P)
4. Asses protein transport and localisation of chimeric protein

Chapter 2: Materials and Methods

2.1. *In-silico* analysis

Initial analysis of the ETX DNA sequence from the Gram-positive bacterium *Clostridium perfringens* was done using online tools Phobius (Stockholm Bioinformatics Centre), SignalP and TMHMM, both of which are hosted by the Technical University of Denmark's Centre for Biological Sequence Analysis. The ETX protein was analysed for signal sequence topologies to identify a signal peptide, designated as SPss in this study. The SignalP tool has options to limit its search to signal sequence characteristics of eukaryotic, Gram-positive or Gram-negative organisms. The ETX DNA sequence was obtained from the GenBank entry M80837 (accessed through the National Centre for Biotechnology Information), generated from the publication by *Hunter et al.* (1992). The deduced amino acid sequence of SPss-GFP (which was later confirmed by DNA sequencing of the plasmid) was run through the online Phyre2 tool (Protein Homology/analogY Recognition Engine V2.0), hosted by the Imperial College of London's Structural Bioinformatics Group. The Phyre2 tool predicts the folding structure of a given protein sequence and gives a resulting match to other proteins.

URLs for the tools are listed as follows:

Phobius - http://phobius.sbc.su.se/

SignalP - http://www.cbs.dtu.dk/services/SignalP/

TMHMM - http://www.cbs.dtu.dk/services/TMHMM/

Phyre2 – http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

2.2. Bacterial cultivation

To isolate plasmid DNA, or for inoculation of larger volumes of growth medium for subsequent analysis, *E. coli* BL21(DE3) (Novagen - *Sigma-Aldrich Corp.*) bacterial cultures were grown overnight at 37°C in an orbital shaking incubator (rotating at 175rpm) in 5mL of LB (10g Tryptone (*Oxoid Ltd*), 10g NaCl (*Glassworld & Chemical cc*), 5g Yeast Extract (*Merck Pty Ltd*), in 1L of distilled water), containing 50µg/mL kanamycin (*Sigma-Aldrich Corp.*). Growth of *E. coli* expressing SPss-GFP, ETX, GFP and a plasmid with no insert were all cultured in this way.

Competent *E. coli* BL21(DE3) (*Thermo Fisher Scientific*) cells were prepared, as follows, for use in plasmid transformation experiments. A streaked *E. coli* colony (grown on LB agar at 37°C) was used to inoculate 10mL of LB and incubated overnight at 37°C. A volume of 2mL of the overnight culture was used to inoculate 200mL of LB, which was incubated at 37°C on an orbital shaker until an OD_{600} of 0.6 was reached. The culture was cooled on ice for 30min before transferring to 50mL centrifuge tubes. The tubes were centrifuged for 15min at 4000rpm at 4°C. The bacterial cell pellets were washed twice with 10% (v/v) Glycerol (*Associated Chemical Enterprises Pty Ltd)* and then resuspended in 4mL of ice-cold GYT broth (10mL Glycerol (*Associated Chemical Enterprises Pty Ltd)*, 0.125g Yeast Extract (*Merck Pty Ltd)*, 0.25g Tryptone (*Oxoid Ltd)* in 100mL distilled water). The resuspended cells were then transferred to 1.5mL centrifuge tubes as 80µL aliquots that were stored at -80°C until needed.

2.2.1. Optical Density measurements (OD_{600}) of bacterial cultures

Samples of *E. coli* cultures were taken at regular intervals and the OD_{600} was measured using a 96-well plate and a spectrophotometer. Optical density measurements of bacterial cultures were done by removing 1mL of culture from each flask at the designated time point. The 1mL sample was then subdivided into three wells of a 96-well plate. Samples were diluted, if necessary, but the total volume in each well was kept at 300µL. These samples were then read with a microplate spectrophotometer (*Hidex Sense Microplate Reader*). Error bars on graphs represent the standard deviation of the samples from the triplicate samples obtained from the three wells used for each sample (n = 3). If the OD_{600} measurements exceeded 0.8, the samples were diluted and re-measured. The data was corrected for the control and dilution factor before being presented in graphical format.

After IPTG induction the bacterial cultures were initially sampled at 0, 4, 8 and 24 hours before changing the sampling range to 0, 6 and 24 hours. This was done to more easily compare protein samples to one another when analysed using the limited capacity of the SDS-PAGE gels. The OD_{600} readings were taken as described above for the pre-induction samples. In all cases, duplicate 1.5mL samples were taken for further processing and analysis. Processing included fractionation of cells, cell lysis, protein precipitation, SDS-polyacrylamide gel

electrophoresis, Bradford protein quantification and mass spectrometry, as described in the sections below.

2.2.2. Optimisation of heterologous protein expression

The growth temperature (after inducing protein expression), concentration of induction reagent (IPTG), and growth medium were tested in various combinations to ascertain the optimum conditions for expressing SPss-GFP in *E. coli* BL21(DE3). The chosen temperatures were 25°C, 30°C, and 37°C, while the IPTG concentrations were chosen as 0µM, 1µM, 10µM, 100µM, and 1000µM IPTG (*Thermo Fisher Scientific*). The growth media chosen were a Lysogeny Broth and a proprietary growth medium (M2P) that uses a polysaccharide as the carbon source, which is made available to *E. coli* through the addition of an enzyme that breaks down the polysaccharide into a useable carbon source, at a steady rate. The polysaccharide and enzyme was obtained from M2P labs (subsidiary of Beckman Coulter GmbH). Additional nutrients were mixed from laboratory stocks according to the manufacturer's instructions:

A 100mL volume of M2P medium consisted of:

- 48mL Salt mix [50mL stock solution: 4.19g MOPS (*Roche Diagnostics Ltd.*), 1g Ammonium Sulphate (*Merck Pty Ltd.*), 0.3g Dipotassium Phosphate (*Sigma-Aldrich Corp.*), 0.3g Tri-sodium citrate dihydrate (*Merck Pty Ltd.*), 0.2g Sodium Sulphate (*Merck Pty Ltd.*), 0.1g Ammonium Chloride (*Merck Pty Ltd.*)]
- 0.9mL Glucose solution [12.22g Glucose Monohydrate (*Merck Pty Ltd.*) in 100mL distilled water]
- 1ml Magnesium solution [1g Magnesium Sulphate Heptahydrate (*Sigma-Aldrich Corp.*) in 20mL distilled water]
- 1mL Vitamin solution [0.2g Thiamine Hydrochloride (*Sigma-Aldrich Corp.*) in 20mL distilled water]
- 0.1mL Trace elements [0.011g Zinc Sulphate Heptahydrate (*SAAR-Chem Pty Ltd.*), 0.010g Copper Sulphate Pentahydrate (*Rochelle Chemicals*), 0.006g Manganese Sulphate Monohydrate (*SAAR-Chem Pty Ltd.*), 0.835g Iron(III) Chloride Hexahydrate (*Sigma-Aldritch Corp.*), 0.668g Triplex III (*Merck Pty Ltd.*), 0.011g

Cobalt Chloride Hexahydrate (*SAAR-Chem Pty Ltd.*), 0.04g Calcium Chloride Dihydrate (Minema Chemicals Pty Ltd.), in 20mL distilled water]

- 19mL distilled water to reach 100mL total volume of M2P growth-medium.

The *E. coli* cultures were grown in temperature-controlled orbital shakers (*New Brunswick I24R* for temperatures at 30°C and above or *Thermoline TLM-570* for lower incubation temperatures), rotating at 200rpm. Pre-cultures of the chosen recombinant *E. coli* were grown overnight in 5mL of LB before being used to inoculate 30mL of either LB or M2P, in a baffled flask with a nominal capacity of 500mL. Inoculation was done by taking 300µL of the pre-culture (1% of final culture volume) and adding it to the medium in the baffled flask. The enzyme required for the M2P medium was added immediately before inoculation. Flasks were sealed with a double layer of autoclaved filter paper, secured with a rubber band. This was repeated for each sample. Multiple flasks were prepared and batched together according to the chosen test condition. In practice, this meant preparing all IPTG induction conditions in individual flasks and placing them together in an incubator at the chosen temperature. This batch method was repeated for each temperature and growth-medium type.

Once inoculated, the flasks were incubated at 37°C until an adequate cell density (OD_{600} 0.6 - 0.8) was achieved. In LB this was approximately 3-4 hours and for M2P this was 5-7 hours, but some recombinant *E. coli* cultures needed to grow more than 7 hours before reaching the required OD_{600}. In the case of M2P, these slow-growing cultures were left to grow overnight before inducing protein expression with IPTG.

2.2.3. DNA isolation, purification and agarose gel electrophoresis

All plasmids used in this study were isolated from bacterial cultures using a Zippy Plasmid purification kit (Zymo Research). Overnight bacterial cultures (5mL) were centrifuged at 13,000rpm (16060 RCF) for 10min in a benchtop centrifuge (*Eppendorf Centrifuge 5417R* – Rotor: FA45-24-11, 24x2mL) at 4°C. The manufacturer's instructions were then followed to

isolate purified plasmid DNA. DNA purity and concentration were determined with a Nanodrop ND1000 spectrophotometer (*Thermo Fisher Scientific*).

Products from PCR reactions and restriction enzyme digestions were isolated from agarose gels using a Gel DNA Recovery Kit (*Zymo Research*). The desired band was cut out of the agarose gel using a sterilised blade and placed into a pre-weighed microfuge tube. The manufacturer's instructions were followed to obtain the desired DNA. The purity and concentration of the DNA were analysed using a spectrophotometer, as above.

The agarose gel was made up of 1% (w/v) agarose in 1xTAE buffer (diluted from 50x stock TAE: 2M Tris (*Sigma-Aldrich Corp.*), 1M Glacial Acetic Acid (*Sigma-Aldrich Corp.*), 50mM EDTA (*SAAR-Chem Pty Ltd*) in 1L distilled water). The agarose was weighed and placed in a flask with a measured volume of 1xTAE (enough to make a small gel is approximately 50mL). This mixture was heated to ensure dissolution of the agarose. Once cooled, the indicator solution Pronasafe (*Condalabs*) was added to the gel, which was then cast in a mould with the appropriate comb for the desired number of wells.

Once the gel was set, the gel was placed in the electrophoresis tank with enough 1xTAE to cover the gel. Samples were prepared so that 300-500ng of DNA was loaded into each well. Electrophoresis was performed at a constant voltage of 120V using a BioRad power supply (*Bio-Rad PowerPac Universal Power Supply (1645070)*). When adequate migration of the dye occurred (40-60min), the gel was removed and visualised in a gel imaging apparatus (*Bio-Rad ChemiDoc MP Imaging System (1708280)*).

2.3. DNA modification and pET26-SPssGFP plasmid construction

A basic outline of the workflow to construct the pET26-SPssGFP plasmid (Figure 2.1) was performed as follows: The plasmid pET26-SPss was digested with *Bam*HI and *Xho*I (both supplied by *Thermo Fisher Scientific Inc.*) to create two sticky ends downstream of the SPss sequence. The GFP gene was amplified from the plasmid pET28-GFP via PCR to modify its original C-terminal *Bam*HI to an *Xho*I so that ligation would be in the correct reading frame and orientation (i.e. all sequences are in an N-terminal to C-terminal direction). The modified GFP PCR product was blunt-end ligated to pBluescript SK (PBSK), linearised with *Eco*RV

(*Thermo Fisher Scientific Inc.*). From that construct, the GFP was isolated by digestion with *Bam*HI and *Xho*I and ligated to pET26-SPss to generate the pET26-SPssGFP construct using a DNA ligase (Fast Link DNA ligase - *Epicentre Health Research*). This was then used to transform *E. coli* competent cells (DH10B and BL21(DE3)) for later experiments.

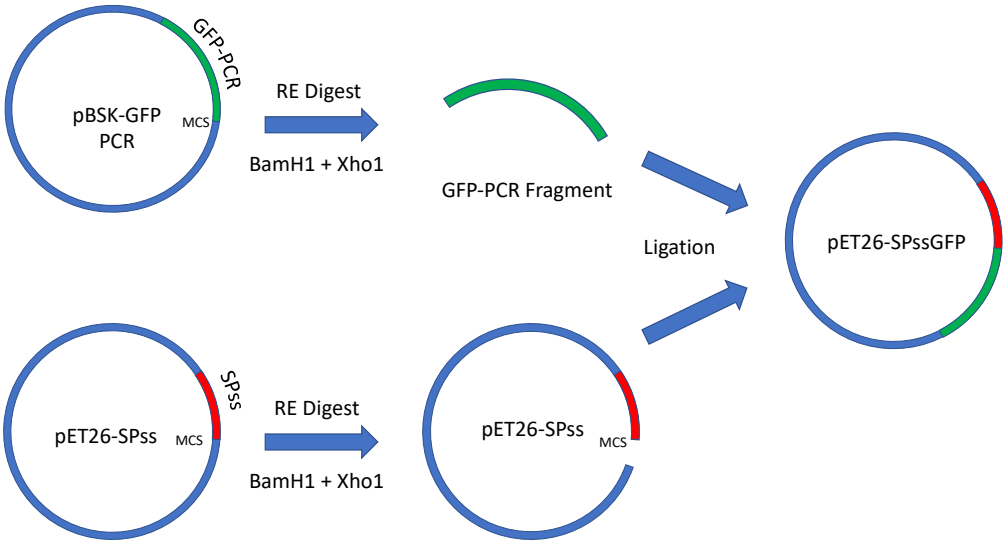


Figure 2.1: Basic outline of pET26-SPssGFP construction. Plasmids were replicated in E. coli, isolated and digested using restriction enzymes of both plasmids containing the modified GFP PCR product and pET-26b plasmid containing the SPss.

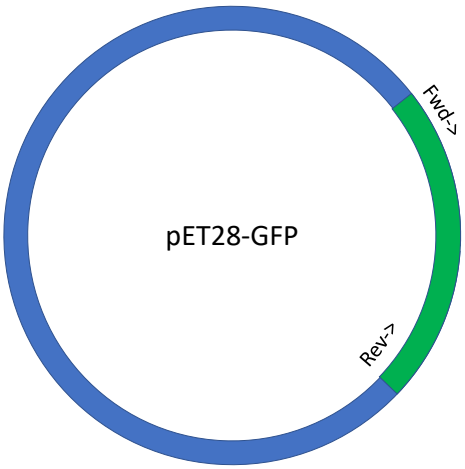
The pET26-SPss plasmid was already present at the CSIR and was constructed by amplification of the signal peptide sequence of the ETX gene by PCR, and insertion into pET-26. The pET28a-GFP plasmid, available at the CSIR, contained a GFP gene that was modified with N- and C-terminal Histidine tags, as well as *Bam*HI sites at the 5' and 3' ends. PCR was used to amplify the GFP gene while altering the 3' restriction site, converting the *Bam*HI site into an *Xho*I site. This PCR product was inserted into a pBSK plasmid (pBSK-GFP), which allowed for blue-white selection of positive transformants. The recombinant pBSK plasmid was then digested with *Bam*HI and *Xho*I restriction enzymes. The DNA insert was isolated and then inserted into a pre-digested pET26-SPss plasmid. The pET26-SPss possessed a *Bam*HI site in frame with the 3' end of the SPss sequence, while the downstream *Xho*I region was in-frame with the poly-His tag and the stop codon. Once ligated to generate the pET26-SPssGFP construct, the presence of a cloned insert was confirmed by restriction enzyme digestion reactions and nucleotide sequencing.

2.3.1. Polymerase Chain Reaction (PCR)

All reactions were supplied with both forward and reverse primers, a buffer, a mixture of deoxynucleotide triphosphates (dNTPs), as well as the DNA polymerase. Taq polymerase was used in all reactions, but a High Fidelity Taq (KAPA HiFi - *Roche Diagnostics Ltd*) was used when the PCR product needed replication with minimal DNA amplification errors. The high fidelity Taq polymerase incorporates fewer errors during replication and is better suited to applications where sequence accuracy is important, such as sequencing and gene expression.

To experimentally obtain the optimal annealing temperature for the primers used, a range of temperatures were tested using a gradient PCR. A gradient PCR was performed as follows: A PCR master solution was prepared to a total volume of 50µL and comprised of 25µL dH_2O, 10µL of 10x O-Buffer (*Thermo Fisher Scientific Inc*), 6µL of 2.5µM Forward Primer, 6µL of 2.5µM Reverse Primer, 1.5µL of 10mM dNTPs, and 1µL of DNA polymerase (KAPA HiFi or Taq). Addition of 1µL of diluted bacterial colony, 0.5µL of template DNA, or 1µL of dH_2O (control) took place once the Master Mix had been subdivided into 10µL aliquots in PCR tubes. Each tube was added to the PCR Thermo-cycler (*Eppendorf Mastercycler 5340 EP Gradient S*) and subjected to the following conditions and cycles: 95°C – 2 minutes, 30x cycles (98°C – 20 seconds, gradient temp. – 15-30 seconds, 72°C – 3 minutes), 72°C – 5min. Once complete, the apparatus was set to hold at 4°C. The gradient temperatures were chosen as 52°C, 55°C, 58°C, 61°C and 64°C, where 58°C was the calculated annealing temperature of the primers using the New England Biolab Tm calculator (https://tmcalculator.neb.com). The thermocycler had the function of running the range of temperatures in one operation as the heating bed could be differentially heated. The reactions were then electrophoresed on an agarose gel. An annealing temperature of 64°C was chosen for subsequent PCR reactions.

PCR was used to construct a fusion gene of GFP and the Secreted Protein Signal Sequence (SPss). The GFP gene was modified using an PCR to alter the 3' restriction site for later fusion. PCR primers were designed to amplify the GFP region contained in a pET28 plasmid. The reverse primer was designed with an overhang containing an *Xho*I restriction site (Figure 2.2).



ATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGGCTAGCATGACTG**GTGGACAGCAAATGGGTCGC***GG*
*ATCC*AGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCA
ATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTT
TCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA
AAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCT
GTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTT
CAATGCTTTTCAAGATACCCAGATCATATGAAGCGGCACGA
CTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAG
AGGACCATCTCTTTCAAGGACGACGGGAACTACAAGACAC
GTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGTCAACAG
GATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAA
CATCCTCGGCCACAAGTTGGAATACAACTACAACTCCCACA
ACGTATACATCACGGCAGACAAACAAAAGAATGGAATCAA
AGCTAACTTCAAAATTAGACACAACATTGAAGATGGAAGC
GTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGG
CGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCAC
ACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGAC
CACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATT**AC**
ACATGGCATGGATGAACTATACAAATAAA*GATCC*GAATTC
GAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAG**CACCAC**
CACCACCACCACTGAGATCCGGCTG

Figure 1: GFP gene and primer sequences. Forward and Reverse primer binding sites are highlighted in green and yellow, respectively. The Reverse primer (*5'- GTGCTCGAGTTTGTATAGTTCATCCATGCCATGTGT -3'*) contains an extended DNA sequence at the 5' end that contains an *Xho*I restriction site (underlined). The restriction site was inserted to the 3' end of the GFP gene. The Forward primer sequence used was (*5'- GTGGACAGCAAATGGGTCGC -3'*).

Once the PCR was complete, the GFP PCR product was isolated from an agarose gel using a kit (Zymoclean Gel DNA Recovery Kit - *Zymo Research*) and then ligated into a Bluscript plasmid to generate pBSK-GFP-PCR. The derived transformants were selected on LB agar containing Ampicillin (100µg/mL), X-Gal (40µg/mL), and IPTG (1mM) for blue-white selection. Colonies were screened by colony PCR, and then grown in LB broth for plasmid isolation and subsequent restriction enzyme analysis.

Colony PCR was performed in the same way as a regular PCR with a difference in the source of template DNA. A colony of *E. coli,* grown on LB agar with 50µg/mL kanamycin, was scraped off the agar surface using a sterile pipette tip. The colony was mixed with a small volume of distilled water (5µL) and used in a PCR.

2.3.2. Restriction enzyme digestion

Restriction enzymes were used to digest the isolated plasmids containing the SPss sequence as well as the modified GFP sequence. The restriction enzymes were evaluated for compatibility and required buffers on the New England Biolabs website (https://nebcloner.neb.com). The plasmid DNA was diluted to appropriate concentrations (between 0.5µg/µL and 1µg/µL) and a 50µL reaction was assembled using 1µL or 10U of the respective enzymes (*e.g.* for a double digest 1µL of *Bam*HI and 1µL of *Xho*I was used (*Thermo Fisher Scientific Inc.*), 10µL of *Bam*HI buffer was used and the volume was made up with dH_2O). The tube was incubated for 2 hours in a water bath at 37℃. The tube was then removed, DNA loading dye was added to the reaction mixture, and electrophoresed on a 1% agarose gel.

2.3.3. DNA ligation of pET26-SPss and modified GFP PCR product

To fuse the SPss plasmid and the modified GFP gene, both digested sequences were excised from an agarose gel, purified using a gel DNA recovery kit (Zymoclean Gel DNA Recovery Kit - *Zymo Research*) and the concentration of the DNA fragments were determined using the NanoDrop ND1000 spectrophotometer. The ratio of insert to be used in relation to the plasmid was calculated as follows:

$$Mass\ of\ insert\ (ng) = \frac{vector\ mass\ (ng)\ \times\ insert\ size\ (kb)}{vector\ size\ (kb)} \times 3$$

The ligation reaction (using the Fast Link DNA ligation kit - *Epicentre Health Research*) was prepared as follows: 1µL of Ligase, 0.75µL of 10mM of ATP, 1.5µL of 10x Buffer, 6µL of Vector (pET26-SPss – 33ng), 5.75µL of Insert (GFP-PCR – 14ng). The ligation reaction mixture was subdivided into two reactions; one was incubated at room temperature for 15min, while the other was incubated overnight at 4°C. Both reactions were inactivated at 70°C for 15 minutes and then used to transform *E. coli* competent cells. Positive transformants were selected from agar plates containing kanamycin and a colony PCR was performed to confirm successful cloning the GFP gene.

2.4. Bacterial transformation

Transformation of competent *E. coli* cells with the desired plasmid or ligation reaction was performed using electroporation. Competent cells were mixed with isolated plasmid in an electroporation cuvette (1mm gap) and exposed to 1.6kV (200ohm, 25µF) using a BioRad Gene Pulser and pulse controller. The culture was then washed out of the cuvette using sterile SOC growth medium (200µL with a composition of 2% (w/v) Tryptone (*Oxoid Ltd*), 0.5% (w/v) Yeast Extract (*Merck Pty Ltd*), 10mM NaCl (*Glassworld & Chemical cc)*, 2.5mM KCl (*Sigma-Aldrich Corp.*), 10mM MgCl (*Sigma-Aldrich Corp.*) and 2mM glucose (*Sigma-Aldrich Corp.*) in dH_2O). The cells were incubated at 37°C, in a glass tube, for 2hrs. The culture (800µL) was then spread on an agar plate of LB containing 50µg/mL Kanamycin and incubated overnight at 37°C. Colonies were selected and inoculated into 5mL of LB, while, in parallel, a PCR was performed to confirm the presence of the plasmid insert. This is specific to pET26/28 plasmids; for pBSK plasmids the ampicillin resistance gene is present, which required 100µg/mL ampicillin to be added to the growth medium for selection.

2.5. Storage of bacterial seed stocks

Bacterial cultures were stored at -80°C in cryoprotective liquid (25% glycerol [*Associated Chemical Enterprises Pty Ltd*] - final concentration). Cultures were stored in cryotubes or microfuge tubes to a volume of 1.5mL for master seed and working culture stocks. The stocks were prepared as follows: an overnight culture of *E. coli* was mixed, in equal volumes, with sterile 50% glycerol (*i.e.* 750µL of overnight culture, added to 750µL glycerol in a 2mL microfuge tube). The microfuge tubes were labelled and sealed with parafilm. To start an overnight culture from the seed stocks, the stock microfuge tubes were taken from the -80°C freezer and placed on ice. Then, 5mL of LB growth medium (in a sealable glass vial) was inoculated with the seed stock by scraping a sterile pipette tip through the surface of the frozen seed stock and placing the pipette tip into the 5mL of LB broth. The inoculated culture was incubated, while the seed stock was immediately returned to the freezer.

2.6. Protein fractionation

E. coli samples taken at different time points during the comparative cultivations were used to fractionate proteins into extracellular, periplasmic and cytoplasmic fractions. These fractions were then analysed independently. The fractions were obtained using a combination of cold osmotic shock and centrifugation. The process was performed as follows:

A sample of culture (1mL) was centrifuged in a benchtop centrifuge for 10min at 13,000rpm (16060 RCF). The cell pellet and supernatant (extracellular fraction) were separated. The cell pellet was then subjected to cold osmotic shock using a modified protocol from Borjaliloo *et al.* (2003). In brief, the pelleted sample was suspended in 60µL of TES buffer (0.2M Tris (*Sigma-Aldrich Corp.*), 0.5M EDTA (*SAAR-Chem Pty Ltd*), 0.5mM sucrose (*Sigma-Aldrich Corp.*) in dH_2O). The sample was repeatedly vortexed and placed on ice for 20min. The sample was then centrifuged at 4°C, 10,000rpm (9391 RCF) for 10min. The supernatant was removed and discarded, and the pellet was resuspended in 30-60µL (depending on the size of the pellet) of ice-cold dH_2O and left on ice for 30min. The sample was then centrifuged once more at 4°C, 16,000rpm (24041 RCF) for 20min, and the supernatant was recovered as the periplasmic fraction. The remaining pellet was labelled as the cytoplasmic fraction. The cytoplasmic fraction was further processed by suspending in 300µL of Bugbuster (Novagen - *Sigma-Aldrich Corp.*) lysis reagent and incubating at 37°C, 180rpm, for 30min. All samples were frozen at -20°C in between processing and analysis. During preparation, all samples were thawed but kept on ice to reduce any possible degradation of the protein samples.

2.7. Protein analysis

Bacterial proteins were prepared by isolating cellular fractions or whole-cells through methods of fractionation or centrifugation, respectively. The samples were then analysed on SDS-PAGE gels for visual examination, or subjected to alternate analysis such as the Bradford assay to determine total protein concentration.

2.7.1. SDS-PAGE

Analysis of the bacterial proteins was performed using a BioRad Mini-Protean Tetra Cell. Samples were prepared with 1x loading buffer (diluted from a 6x stock: 45%(v/v) Glycerol, 6% (w/v) SDS, 9% (v/v) β-mercaptoethanol, 0.03% (w/v) Bromophenol Blue, dissolved in 375mM Tris-HCl (pH 6.8)) and loaded into the gel (Stacking gel (4% Acrylamide): 1.3mL Acrylamide (30% - *BioRad*), 2.5mL 0.5M Tris-HCl pH 6.8, 6.1mL dH_2O, 20µL 10% SDS (*BioRad*), 50µL 10% Ammonium Persulphate (*BioRad*), 8.5µL TEMED (*BioRad*)); Resolving gel (12% Acrylamide): 3.4mL Acrylamide (30%), 2.125mL 1.5M Tris-HCl (pH 8.8), 2.89mL dH_2O, 20µL 10% SDS, 50µL 10% Ammonium Persulphate, 8.5µL TEMED). The gels were electrophoresed at 200V until the dye had migrated from the top to the bottom of the gel. The gel was then removed from the Tetra Cell cassette, carefully peeled away from the glass plates and placed in the SDS-PAGE Staining solution (destaining solution with 0.25g/L Coomassie Brilliant Blue R250) until the dye had permeated the gel completely (4hrs – overnight). The stain was then removed from the gel by incubating in destaining solution (4hrs – overnight; 10%(v/v) Isopropanol (*Associated Chemical Enterprises Pty Ltd*), 10%(v/v) Ethanol (*Associated Chemical Enterprises Pty Ltd*), 10%(v/v) Acetic Acid (*Associated Chemical Enterprises Pty Ltd*), in dH_2O) leaving the gel almost clear with the proteins in the gel stained and amenable to visualisation. The gel was then visualised using the ChemiDoc and annotated using Image Lab software (Bio-Rad Lab – Version 6.1 for Windows).

2.7.2. Bradford protein concentration quantification – microplate assay

To determine the concentration of total protein in different samples, a method of dying the proteins and measuring the absorbance at 595nm, commonly referred to as the Bradford Protein Assay, was used. Quick Start Bradford Dye Reagent from BioRad was used to perform the assay. The assay was conducted using a 96-well plate. A protein standard of 2mg/mL of Bovine Serum Albumin (BSA, *BioRad*) was serially diluted in 1xPBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4 (*Sigma-Aldrich Corp.*), 1.8mM KH_2PO_4 (*Sigma-Aldrich Corp.*)) to form a range of standards, while 1xPBS was used as a blank control. The experimental samples were also serially diluted in order to fall within the standard curve range. A volume of 10µL of standard/sample was pipetted into each well and 200µL of Bradford Dye Reagent was added

to each well and left to incubate at room temperature for 5min before being placed in a microplate spectrophotometer and the absorbances read at 595nm. The BSA standards were used to generate a standard curve to which the sample readings were compared, and protein concentration determined.

2.7.3. Protein quantification – SDS-PAGE gel densitometry

Quantifying protein concentration by gel densitometry was performed by diluting a BSA standard to known concentrations and loading those onto an SDS-PAGE gel, along with total protein fractions of *E. coli* BL21(DE3) expressing SPss-GFP. Using Image Lab Software (Bio-Rad Lab – Version 6.1 for Windows) the BSA standards were analysed and used to generate a standard curve to which the SPss-GFP protein bands were compared and the concentration was determined after correcting for dilution.

2.7.4. Analysis of extracellular protein fractions

To visualise proteins in the extracellular fraction, acetone was used to precipitate the proteins and purify them away from the growth medium. This was done using acetone (kept at a temperature of -20°C) and centrifugation. Samples of supernatant from an experiment (separated from bacterial pellets via centrifugation) were mixed in a 2mL microfuge tube in a ratio of 1:5 (300µL sample added to 1.2mL of chilled acetone, total volume 1.5mL). The samples were incubated overnight at -20°C and then centrifuged at 16,000rpm for 10min at 0°C (Eppendorf Centrifuge 5417R – Rotor: FA45-24-11, 24x2mL). The supernatant was removed, and the remaining pellet was air-dried in a laminar flow cabinet, taking care not to completely desiccate the pellet. The pellet was then suspended in 60µL of 1xPBS buffer. The sample was analysed immediately on an SDS-PAGE gel or stored at -20°C before analysis.

2.7.5. Mass spectrometry of SPss-GFP protein

Mass spectrometry was conducted internally at the CSIR using multiple protein samples (obtained under conditions to maximise SPss-GFP protein expression levels – 1000µM 30°C

and 37°C; 100µM 37°C). The SPss-GFP was extracted from an SDS-PAGE gel and analysed using a Dionex Ultimate 3000 RSLC system coupled to a QSTAR ELITE mass spectrometer. The Protein Pilot software (v4.0.8085) was used to compare data to the peptide sequence of SPss-GFP, obtained from the sequenced and translated DNA sequence of the pET26-SPssGFP construct.

Chapter 3: Results

3.1. Protein and DNA sequence analyses

3.1.1. *In-silico* analysis of peptide sequences

Analysis of the full-length *Clostridium perfringens* ETX peptide sequence by SignalP (restricted to Gram-negative bacteria) indicated a reduced analysis score of the 1-32 amino acid region (Figure 3.2) compared to the higher score when the analysis was restricted to Gram-positive bacteria (Figure 3.1). The probability of a cleavage site peaks at amino acid 33 (*i.e.* cleavage occurs between amino acid 32 and 33) but this score is below the probability cut-off of 0.5. This analysis therefore indicated that there is a low possibility that ETX possesses a signal peptide that is recognised in Gram-negative bacteria. Analysis by the TMHMM tool showed that there is an N-terminal region of ETX that has a moderate probability of being a transmembrane region, while the remainder of the peptide has a high probability of having a final localisation outside the cytoplasmic membrane (Figure 3.3).

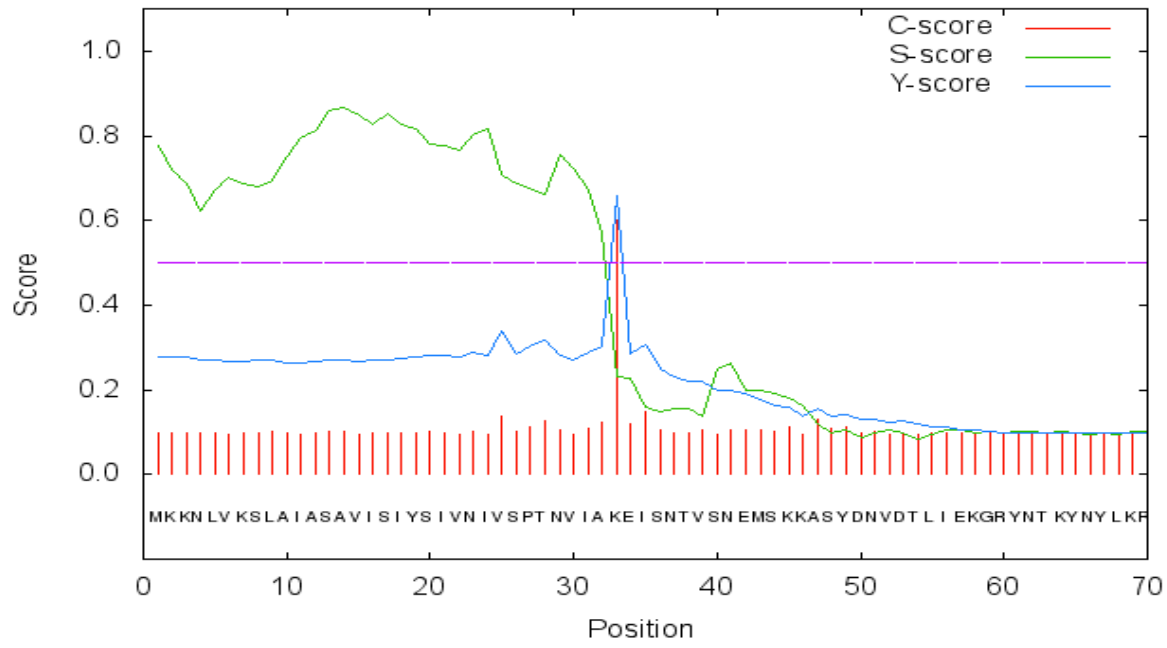


Figure 3.1: SignalP analysis of the ETX peptide restricted to Gram-positive bacteria. The analysis indicated that the first 32 amino acids of ETX can function as a signal peptide sequence with a predicted signal peptidase cleavage site between amino acids 32 and 33. C-score indicates the probability of a signal peptidase cleavage site; S-score indicates the probability of a signal peptide domain; Y-score is a combined C and S score.

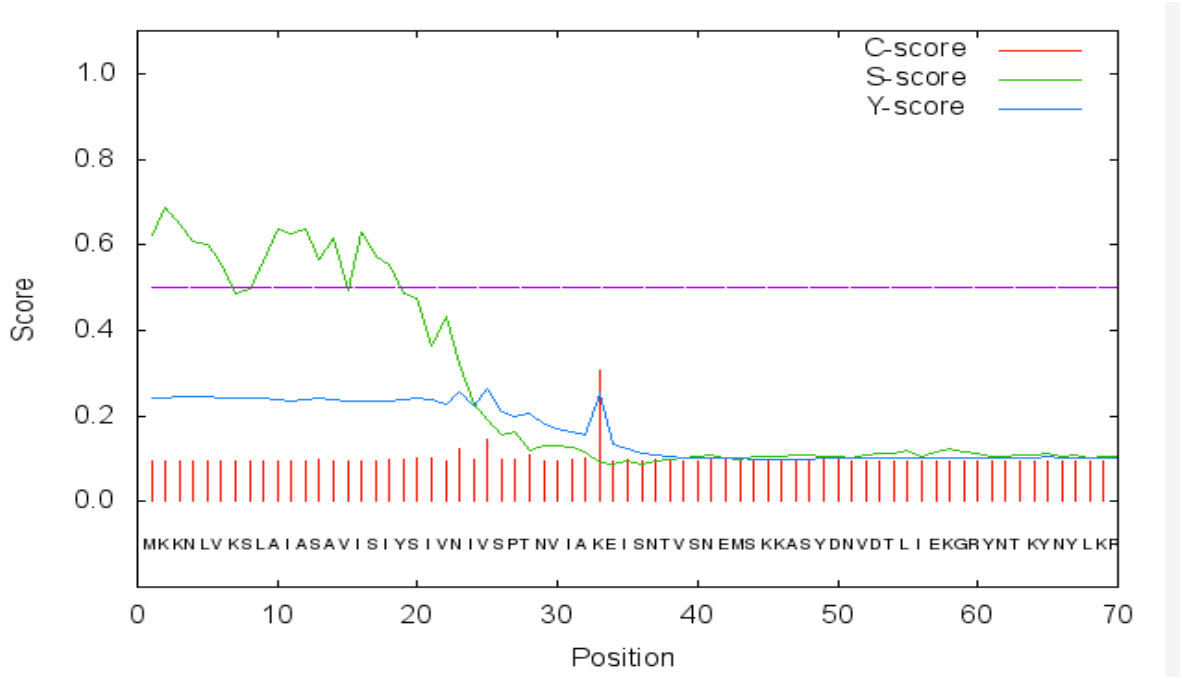


Figure 3.2: SignalP analysis of the ETX peptide restricted to Gram-negative bacteria. Low cleavage and signal peptide domain scores can be seen for a large portion of the 1-32 amino acid region. These low scores are below the cut-off value of 0.5, and SignalP indicates that the first 32 amino acids of ETX has a low probability of functioning as a signal peptide sequence in Gram-negative bacteria C-score indicates the probability of a cleavage site; S-score indicates the probability of a signal peptide domain; Y-score is a combine C and S score.

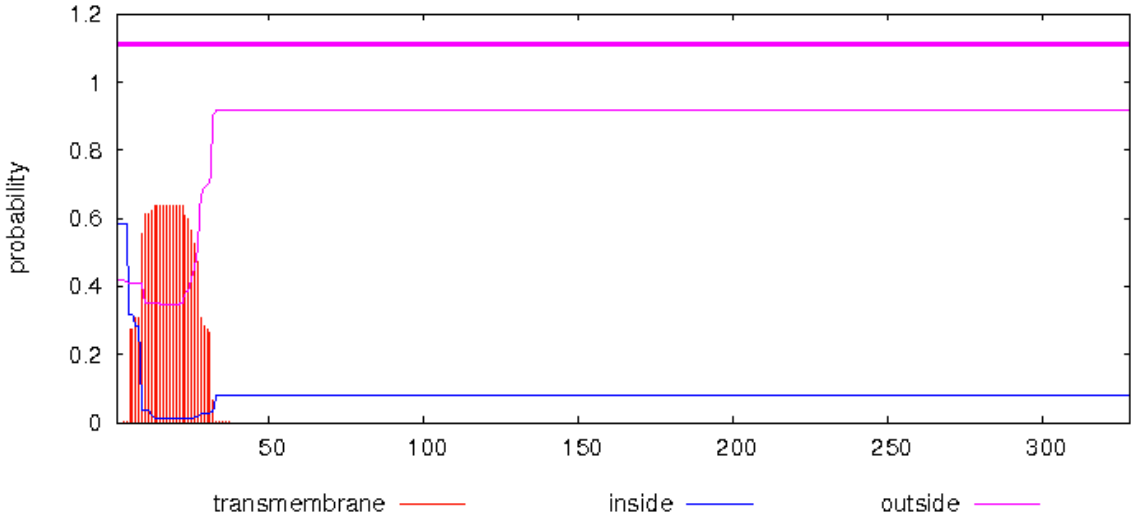


Figure 3.3: TMHMM analysis of the ETX peptide indicating the probability of the ETX peptide possessing transmembrane regions. An N-terminal transmembrane domain is predicted and the peptide has a high probability (~0.9) of being located outside of the cytoplasmic membrane.

To further analyse ETX for signal peptide regions, the online tool Phobius was used. Phobius utilises a combination analysis methods to predict transmembrane or signal peptide regions (Käll et al., 2007). Combining the analytical methods of SignalP and TMHMM, Phobius claims to reduce false positives by as much as 5-10%. The resulting analysis of ETX using Phobius showed that the first N-terminal 32 amino acids possess a signal sequence domain and that the rest of the protein is localised outside of the cytoplasm (Figure 3.4). This is in agreement with previous research done on the ETX peptide (Hunter et al., 1992). Within the signal peptide sequence, the Phobius tool indicated that there are regions within the sequence consistent with Sec-mediated transport (namely the N-, H- and C-domains) (Natale et al., 2008). This 32 amino acid signal peptide sequence of the ETX protein was designated SPss.

Prediction of ETX

```
ID   ETX
FT   SIGNAL          1      32
FT   REGION          1       8        N-REGION.
FT   REGION          9      20        H-REGION.
FT   REGION         21      32        C-REGION.
FT   TOPO_DOM       33     328        NON CYTOPLASMIC.
//
```

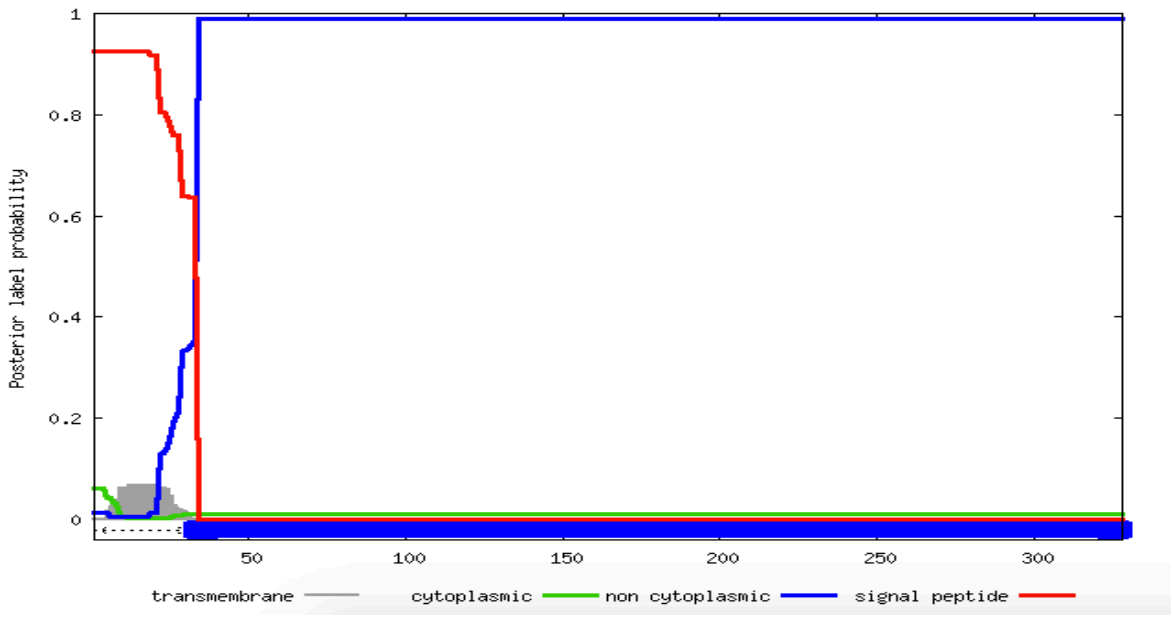


Figure 3.4: Phobius analysis of the ETX peptide sequence. The graph shows the probability of ETX amino acid regions functioning as a signal peptide sequence and its subcellular locations. The presence of a charged region (N-region; peptides 1-8), hydrophobic region (H-region; peptides 9-20) and polar region (C-region; peptides 21-32) are consistent with signal peptides associated with Sec-mediated transport, while the remaining, mature peptide likely has a non-cytoplasmic localisation.

Analysis of the sequence of the SPss signal peptide fused to GFP (SPss-GFP) using SignalP, TMHMM and Phobius are shown below. These results indicated that there is a reduced probability for the SPss to function as a signal peptide sequence when the analysis performed by SignalP was restricted to Gram-negative bacteria (Figure 3.5) and when restricted to Gram-positive bacteria (Figure 3.6). Moreover, SignalP predicted additional signal peptidase cleavage sites in the SPss region, albeit with low probability scores. The TMHMM analysis indicated the presence of a transmembrane domain at the N-terminus of SPss-GFP, while the remainder of the peptide has a localisation outside the cytoplasmic membrane (Figure 3.7). Analysis of the SPss-GFP sequence with Phobius indicated that SPss had an altered secretion signal domain when combined with GFP (Figure 3.8). Not only was the predicted signal peptide domain shorter (N-terminal 20 amino acids), but the sizes and composition of the N-, H- and C-regions associated with Sec-mediated transport was different when compared to ETX predictions using the same analysis (Figure 3.4).

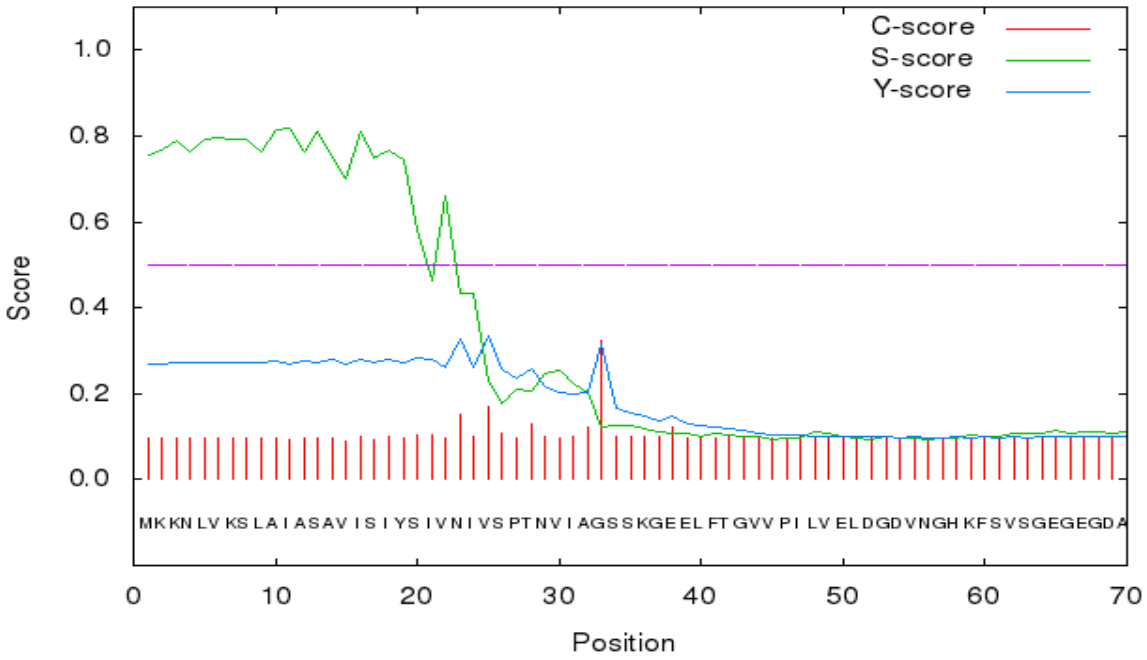


Figure 3.5: SignalP analysis of the SPss-GFP peptide sequence, limited to Gram-negative bacteria. The most likely cleavage site is predicted at amino acid 32, with low-probability cleavage sites also predicted at amino acid 23 and 25. C-score indicates the probability of a signal peptidase cleavage site; S-score indicates the probability of a signal peptide domain; Y-score is a combined C and S score.

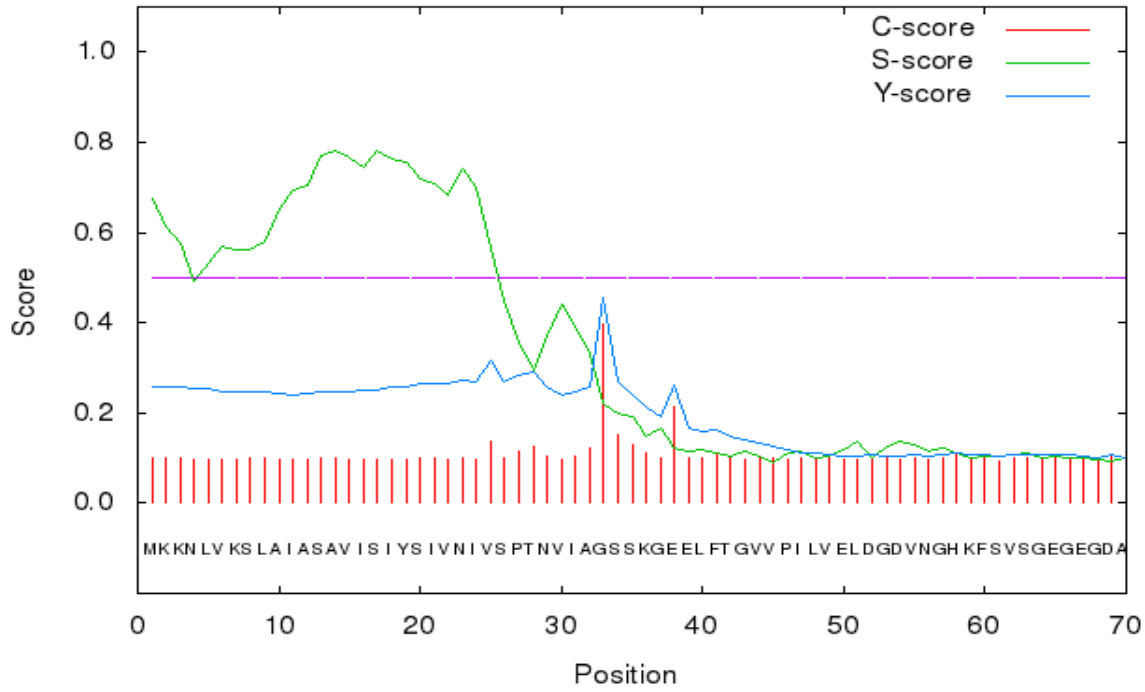


Figure 3.6: SignalP analysis of the SPss-GFP peptide sequence, limited to Gram-positive bacteria. The most likely cleavage site is predicted at amino acid 32 and a secondary possible cleavage site is predicted at amino acid 39. C-score indicates the probability of a signal peptidase cleavage site; S-score indicates the probability of a signal peptide domain; Y-score is a combined C and S score.

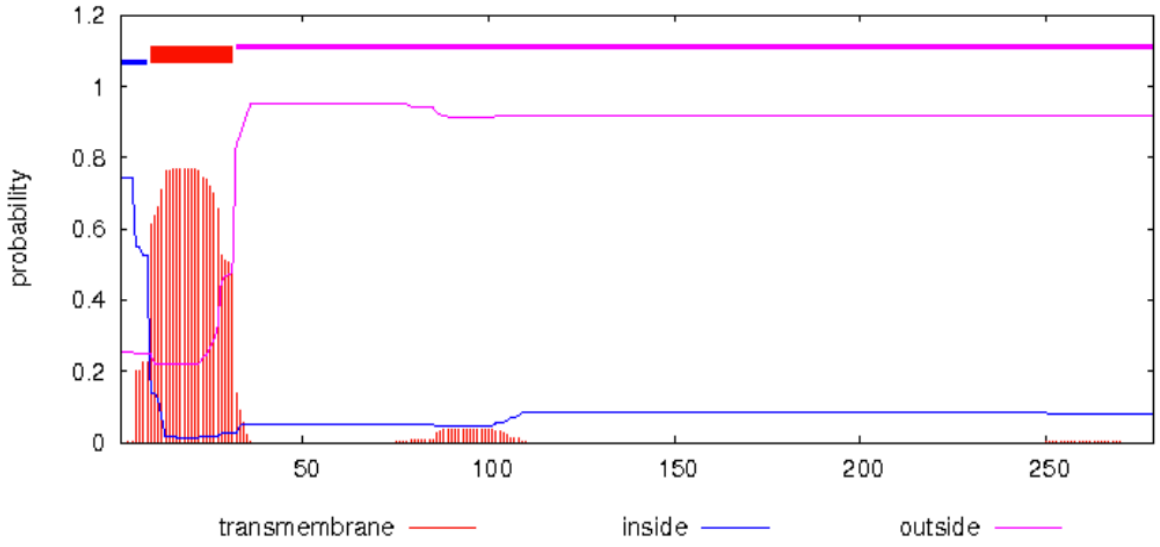


Figure 3.7: TMHMM analysis of the SPss-GFP peptide sequence. In this case, TMHMM returned values indicating that amino acids 1-8 are likely inside the cell, amino acids 9-31 are a transmembrane helix region, and amino acids 32-279 (the remainder of the peptide) lies outside the cytoplasmic membrane.

```
ID   SPssGFP
FT   SIGNAL        1     20
FT   DOMAIN        1      7       N-REGION.
FT   DOMAIN        8     16       H-REGION.
FT   DOMAIN       17     20       C-REGION.
FT   DOMAIN       21    279       NON CYTOPLASMIC.
//
```

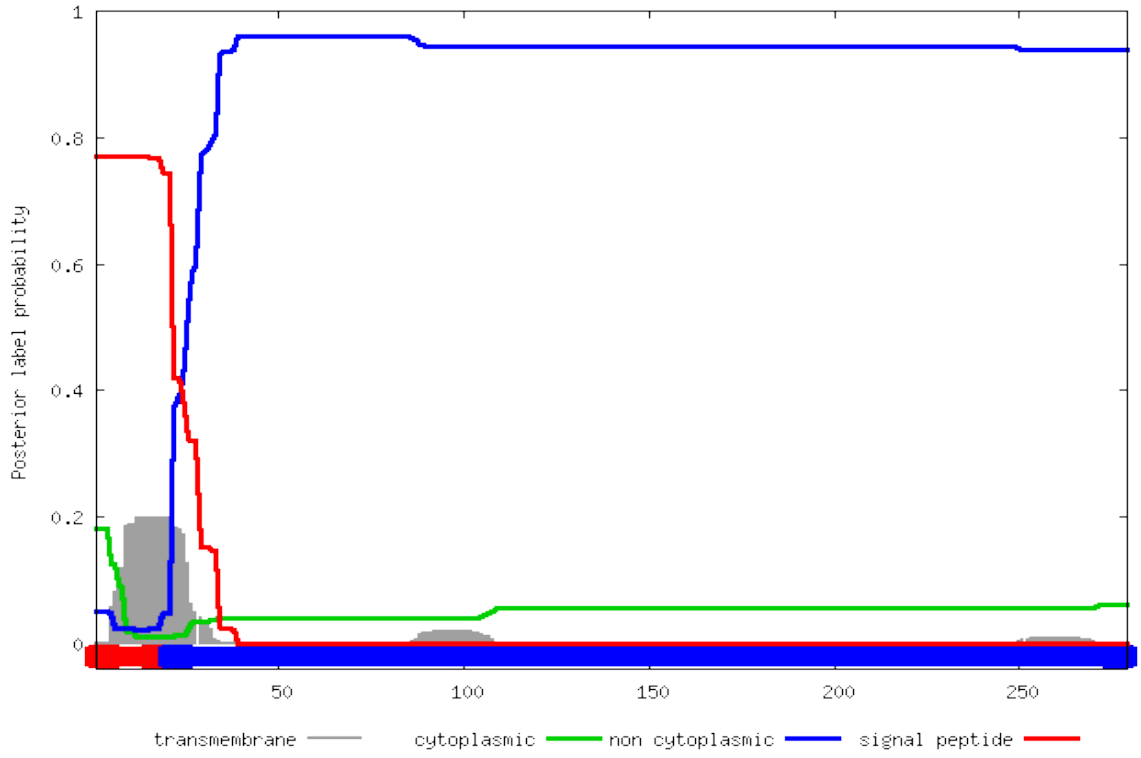


Figure 3.8: Phobius analysis of SPss-GFP peptide sequence. A difference in signal sequence domains can be observed when compared to Phobius analysis of ETX (Figure 3.4). The predicted signal peptidase cleavage site at amino acid 20 is significantly different than the predicted cleavage site ETX (amino acid 32). Although a charged region (N-region; peptides 1-7), hydrophobic region (H-region; peptides 8-16) and polar region (C-region; peptides 17-20) was predicted in the signal peptide sequence, the sizes and amino acid sequence composition of these region differed from predicted for ETX (Fig. 3.4).

The potential of the SPss-GFP protein to fold into a functional conformation was analysed using the Phyre2 online protein homology tool to visualise potential folding patterns of this protein (Figure 3.9). The SPss-GFP nucleotide sequence was determined and confirmed to be identical to the reference sequence. The DNA sequence was then translated using the ExPASy online Translate tool (Swiss Institute of Bioinformatics - www.expasy.org) and the theoretical molecular weight and pI of SPss-GFP was determined to be 31,214.27 Da and 6.34, respectively. The prediction indicated that a stable GFP chromophore would form and that the SPss sequence did not affect chromophore formation.

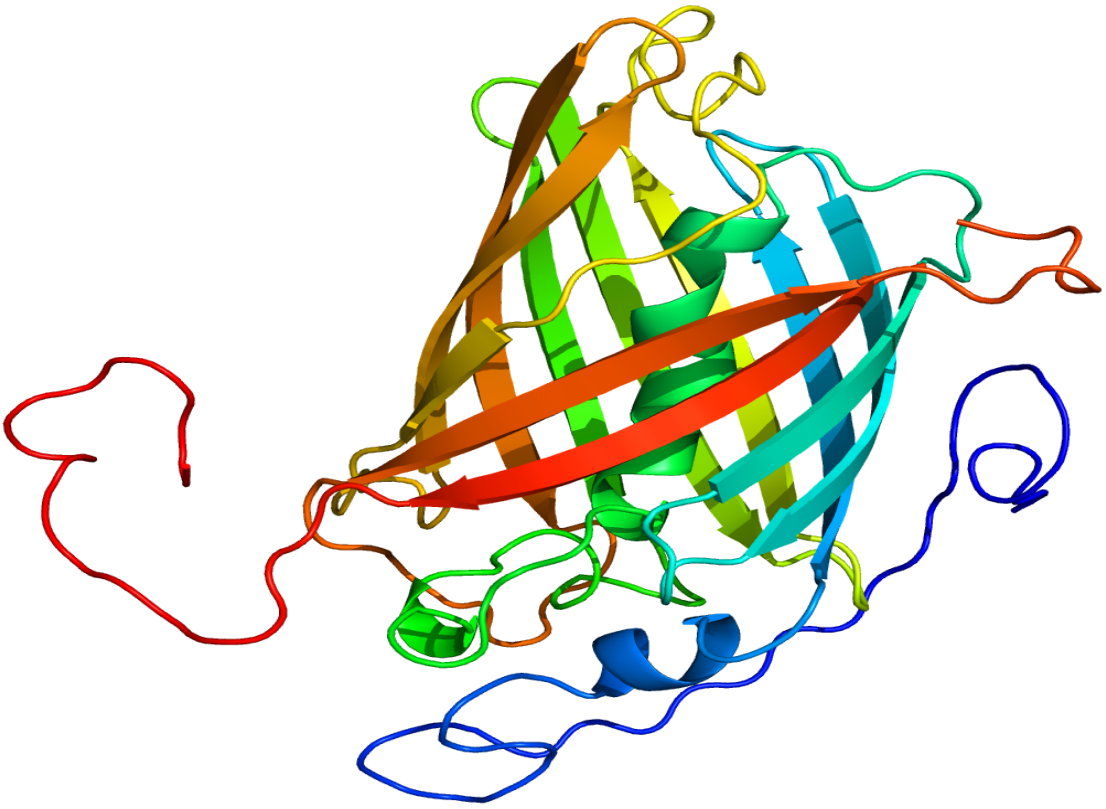


Figure 3.9: Phyre2 analysis of SPss-GFP peptide sequence and the resulting three-dimensional model.

3.1.2. Construction and characterisation of the pET26-SPssGFP plasmid

The GFP DNA sequence in pET28-GFP was obtained by PCR amplification, and the resulting PCR product was purified from an agarose gel and then cloned into pBluescriptS SKII+. The GFP gene was subsequently excised from this construct using *Bam*HI and *Xho*I and then ligated into an identically digested plasmid containing the SPss sequence (pET26-SPss) to generate the recombinant pET26-SPssGFP construct. This cloning strategy ensured that the SPss and GFP were cloned in-frame with each other. Colony PCR was carried out to identify positive transformants. The pET26-SPssGFP plasmid DNA was then isolated, digested with restriction enzymes and sequenced to confirm the integrity of the recombinant plasmid construct. The pET26-SPssGFP plasmid was used to transform *E. coli* BL21(DE3) to enable expression analyses of the SPss-GFP peptide.

Colony PCR with a GFP-specific primer pair was carried out on *E. coli* DH10B transformants. The PCR results of colonies 2 and 3 in Figure 3.10 showed a positive result, as evidenced by the similar-sized band (752bp) compared to the positive control (GFP). The isolated plasmids from both colonies were digested with a combination of the restriction enzymes *Bam*HI and *Xho*I. The results of the restriction digestion of plasmid DNA extracted from colony 3 are shown in Figure 3.11. Comparison of the DNA fragment sizes indicated that the double digestion excised a DNA fragment of approximately 750bp, which corresponds in size with the GFP insert (752bp). The DNA sequence of SPss-GFP is provided in Figure 3.12, as confirmed by DNA sequencing by Inqaba Biotech (Pty) Ltd, and the deduced amino acid sequence is provided in Figure 3.13.

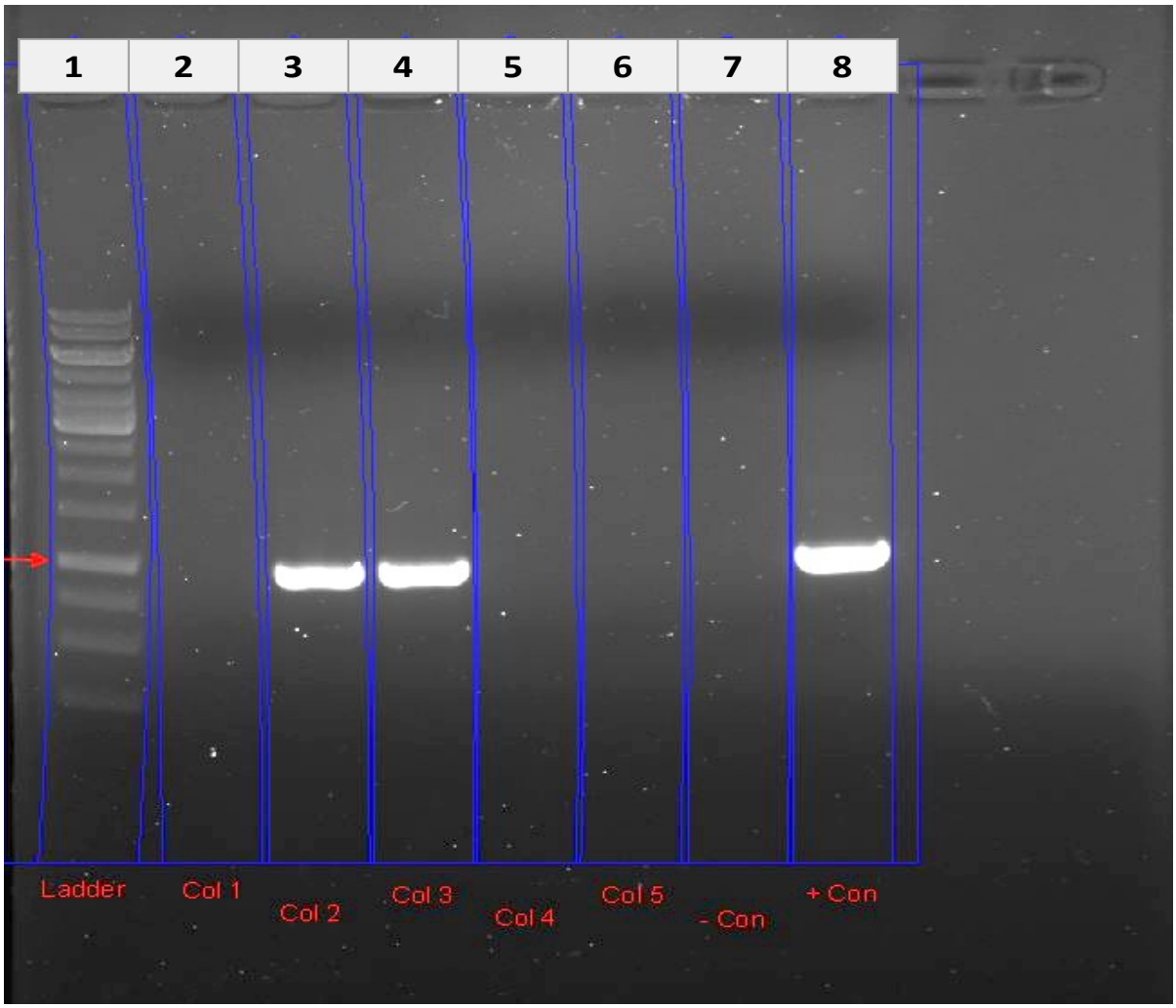


Figure 3.10: Colony PCR of selected E. coli colonies. Lanes 3 and 4 show a band slightly below the 1kb marker (red arrow in Lane 1 DNA ladder) indicating the bacteria was successfully transformed with a recombinant pET26-SPssGFP plasmid. The template control in Lane 8 is a pET28-GFP plasmid; Lane 7 is a non-template control.

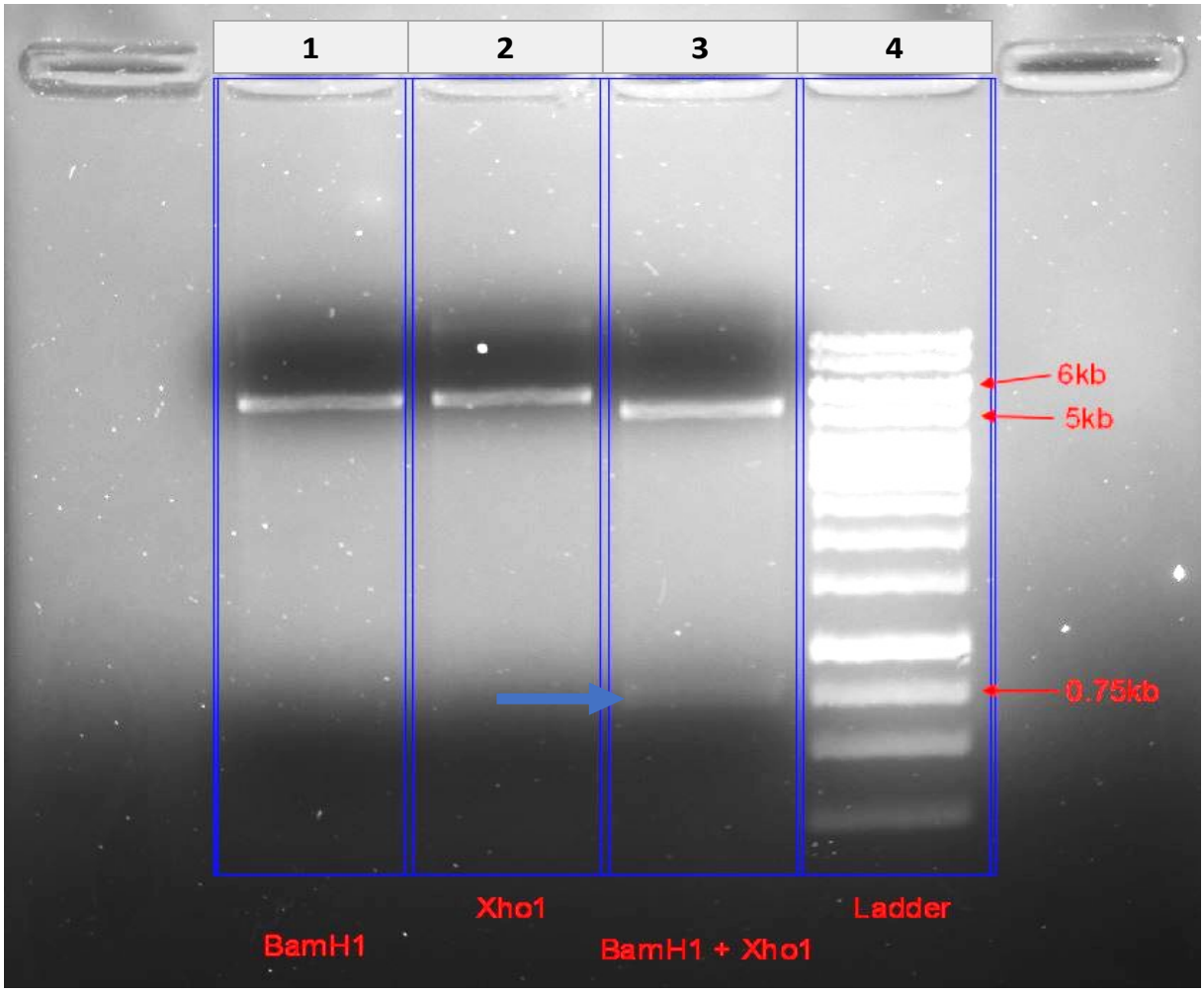


Figure 3.11: Restriction enzyme digestion of purified pET26-SPssGFP plasmid DNA using a combination of BamHI and XhoI restriction enzymes. Lanes 1 and 2 are single digests; Lane 3 is a double digest that resulted in excision of the GFP insert DNA from the pET26-SPssGFP plasmid. The expected band size is 752bp and is indicated by the blue arrow in Lane 3.

CATATGAAAAAAAACCTGGTTAAATCTCTGGCGATCGCGTCTGCGGTTATCTCTATCTACTCTATC
GTTAACATCGTTTCTCCGACCAACGTTATCGCG**GGATCC**AGTAAAGGAGAAGAACTTTTCACTGG
AGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAG
AGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTA
CCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAG
ATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGA
CCATCTCTTTCAAGGACGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACAC
CCTCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCAC
AAGTTGGAATACAACTACAACTCCCACAACGTATACATCACGGCAGACAAACAAAAGAATGGAA
TCAAAGCTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTAT
CAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAA
TCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGC
TGCTGGGATTACACATGGCATGGATGAACTATACAAA**CTCGAG**CACCACCACCACCACCAC**TGA**

Figure 3.12: Nucleotide sequence of SPss-GFP (840bp). The SPss signal sequence (red); GFP sequence (green); and poly-His tag (black) are shown. The restriction sites NdeI (CATATG) includes the start codon; BamHI (GGATCC); and XhoI (CTCGAG) are underlined and highlighted. The NdeI site marks the start of the SPss sequence; the BamHI site is in between the SPss and GFP sequences; and the XhoI site marks the end of the GFP sequence. A poly-His tag is present after the XhoI site before a stop codon (TGA).

MKKNLVKSLAIASAVISIYSIVNIVSPTNVIAGSSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA
TYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLE
HHHHHH

Figure 3.13: Deduced amino acid sequence of SPss-GFP. The SPss region is highlighted in red, the GFP region in green and the poly-His tag in black.

3.1.3. Mass spectrometry data

Mass spectrometry (MS) analysis was used to confirm that SPss-GFP was being expressed, and to indicate any changes to the protein sequence of SPss-GFP (*i.e.* cleavage of the SPss

from GFP). Comparison of the sequence from MS and the deduced amino acid sequences of SPss and GFP showed differences. The MS data (Figure 3.14) did not show the presence of a large section of N-terminal amino acids, which includes the SPss amino acid sequence. However, the section of absent amino acids extended beyond the known cleavage sight of SPss, namely four additional amino acids were absent from the GFP N-terminal region. These difference were observed in three independent samples submitted for MS analysis. These results indicate that the processing of SPss-GFP might have occurred but that cleavage did not occur strictly as predicted by the *in-silico* analysis. The absence of other peptide sequences in MS data is discussed further in the Discussion chapter (Chapter 5).

MKKNLVKSLAIASAVISIYSIVNIVSPTNVIAGSSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDA
TYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLE
HHHHHH

Figure 3.14: Mass spectrometry (MS) data of the SPss-GFP peptide. The SPss sequence is underlined. Grey-coloured sequences indicate no match of MS and sequence data. The green-coloured sequence represents a >=95% confidence match of MS and compared sequence data. The red-coloured sequence represents a confidence interval of between 0% and 50%, while the sequence in yellow shows a confidence interval between 50% and 95% The results shown are from one of three SPss-GFP protein samples extracted from SDS-PAGE gels and submitted for MS analysis.

3.2. GFP fluorescence

The fluorescence of GFP and the SPss-GFP proteins were compared by observing bacterial cell pellets under UV light exposure on a light table. The cell pellets were obtained by centrifugation of 1.5mL aliquots of bacterial cultures that were grown for 24hrs in M2P growth medium and then treated with 1mM IPTG to induce protein expression. Fluorescence could be observed in the bacterial cell pellets of *E. coli* BL21(DE3) containing pET28-GFP, but not from *E. coli* BL21(DE3) containing pET26-SPssGFP (Figure 3.15). This comparison clearly illustrates that chromophore formation of GFP is altered to the extent of being non-fluorescent when the SPss sequence is present together with GFP.

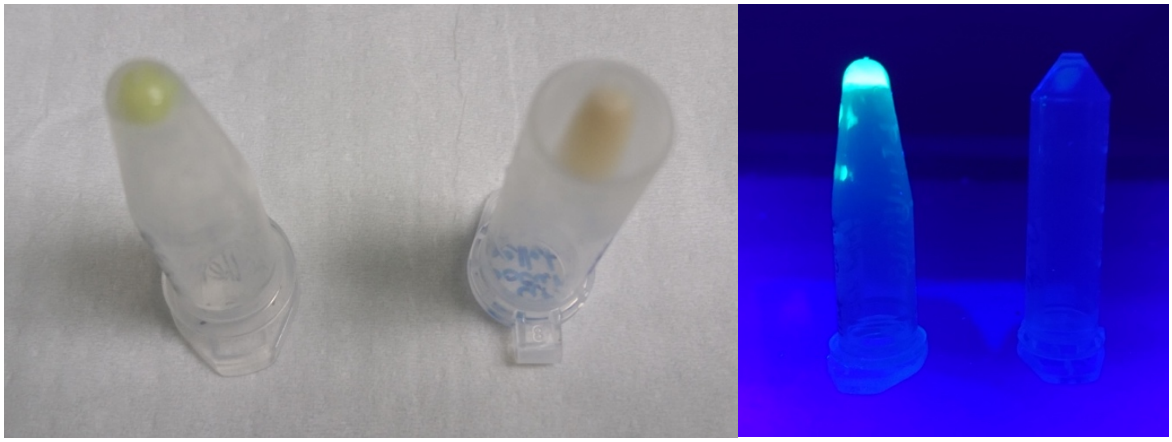


Figure 3.15: Tubes containing E. coli cell pellets, observed under visible light (left) and UV light (right). The left tube in each pair contains E. coli BL21(DE3) harbouring pET28-GFP, while the right tube in each pair contains E. coli BL21(DE3) harbouring pET26-SPssGFP. Both bacterial cell cultures were grown in M2P growth medium and induced with 1mM IPTG after overnight growth.

3.3. Growth of *E. coli* BL21(DE3) containing pET26-SPssGFP

3.3.1. Pre-induction cell densities

The OD_{600} readings over a 6 hrs time course of various *E. coli* BL21(DE3) cultures containing plasmids pET26-SPssGFP, pET28-GFP, pET26-ETX, or a pET26 plasmid without a DNA insert (labelled 'Empty Plasmid') are shown in Figure 3.16. Optical density measurements of the respective bacterial cultures indicated a gradual increase in OD_{600} readings between 1 to 3 hrs of incubation, reaching an OD_{600} of approximately 0.2 after 2 hrs of growth. However, from 3 to 6 hrs of incubation, there was a steep increase in OD_{600} readings. The final OD_{600} (at 6 hrs) was between 1.1 to 1.2 for the bacterial cultures, except for the *E. coli* pET-GFP culture, which was only marginally lower than 1.1. These results indicated that the growth of *E. coli* BL21(DE3) was not adversely affected by the recombinant expression vector constructs.

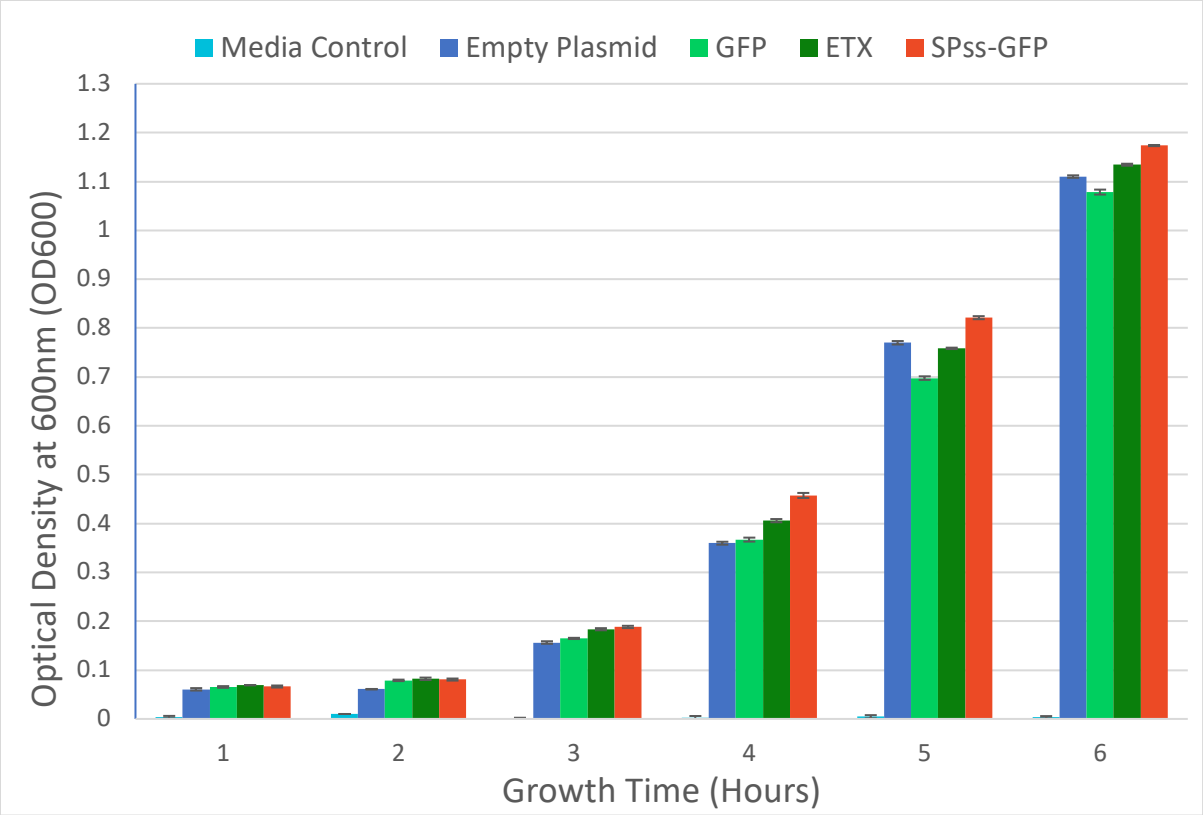


Figure 3.16: OD_{600} readings of various E. coli BL21(DE3) cultures harbouring different plasmid constructs grown in M2P growth medium. The E. coli containing the 'Empty Plasmid' pET26 with no DNA insert (dark blue); the pET28-GFP (light green); the pET26 ETX D plasmid (dark green); and the pET26-SPssGFP plasmid (red) showed similar OD_{600} readings. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

3.3.2. Effects of IPTG on the growth of *E. coli* BL21(DE3) containing pET26-SPssGFP

A comparison of the different growth temperatures and IPTG inducer concentrations on the growth of *E. coli* BL21(DE3) with pET26-SPssGFP is shown below. Surprisingly, at 37°C (Figure 3.17), IPTG had a positive effect on cell density but a detrimental effect at other temperature points (Figure 3.18 and Figure 3.19) except for 1µM, which showed slightly higher cell density levels at all temperatures. At lower temperatures (30°C and 25°C), higher IPTG concentrations (particularly 10µM and higher) showed a negative effect on cell density with maximal repression of growth at both high IPTG concentration and low temperatures (25°C). Similar cell densities were observed at 37°C at both 1µM and 1000µM IPTG with an OD_{600} of approximately 20 at 24 hrs post-induction. At 30°C and 25°C, 0µM and 1µM IPTG treatments showed higher OD_{600} values than other treatments after 24 hours of treatment and growth

temperature change. All differences in treatments were observed more clearly at least 8 hours after IPTG induction.

The effects of IPTG concentration and growth temperature on *E. coli* BL21(DE3) harbouring pET26-ETX and Empty pET26 constructs were more pronounced in LB growth studies compared to M2P (see Appendix – Graphs, Figures 6.1 – 6.10). After 24 hours post-induction with IPTG, the high IPTG concentrations (above 100µM) had an associated reduction in OD_{600} values, with this reduction most pronounced at 37°C, at 100µM and 1000µM IPTG. Lowering the growth temperature during induction was effective in minimising the deleterious effects of IPTG on cell density.

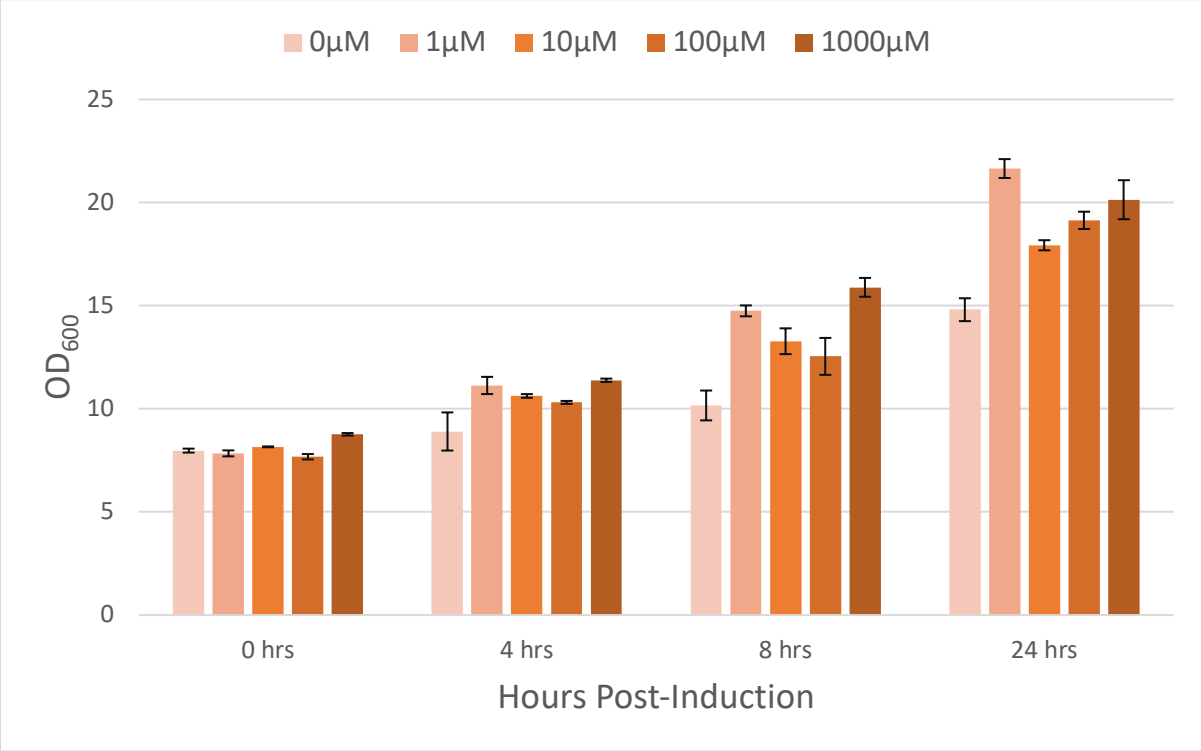


Figure 3.17: Post-induction OD_{600} measurements of E. coli BL21(DE3) containing pET26-SPssGFP at various IPTG concentrations, at 37°C, grown in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

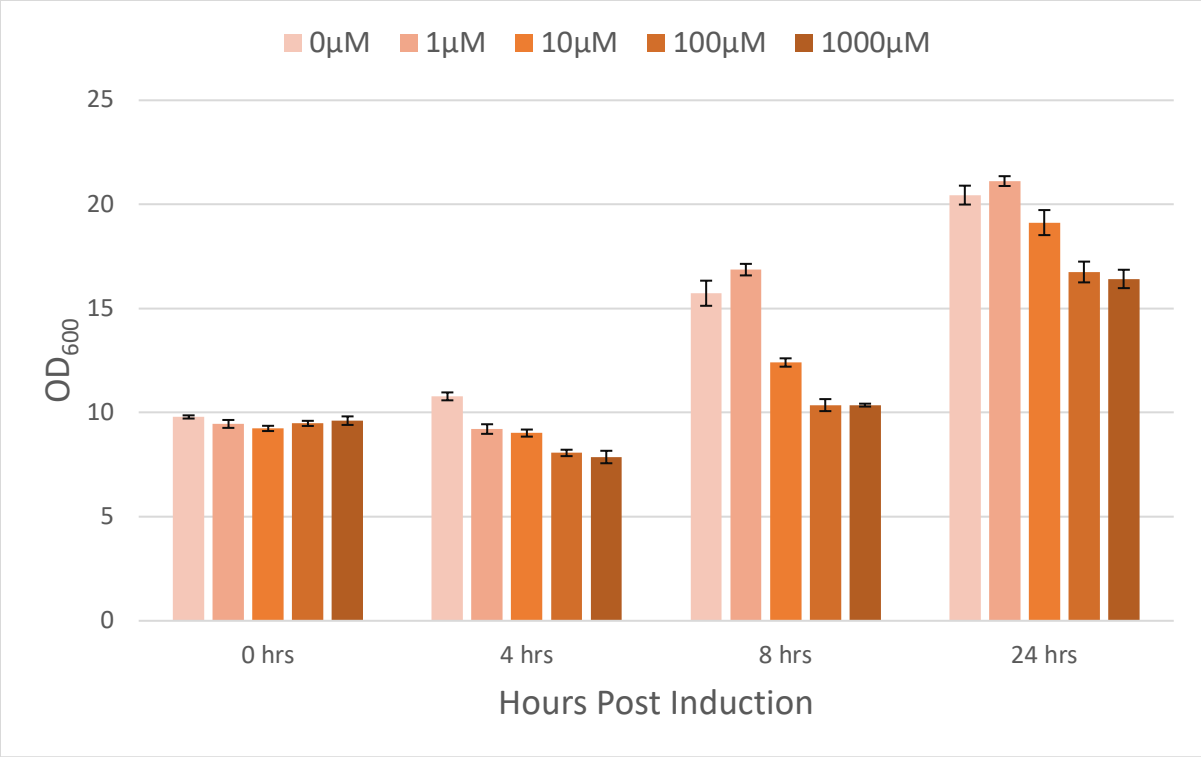


Figure 3.18: Post-induction OD_{600} measurements of E. coli BL21(DE3) containing pET26-SPssGFP at various IPTG concentrations, at 30°C, grown in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

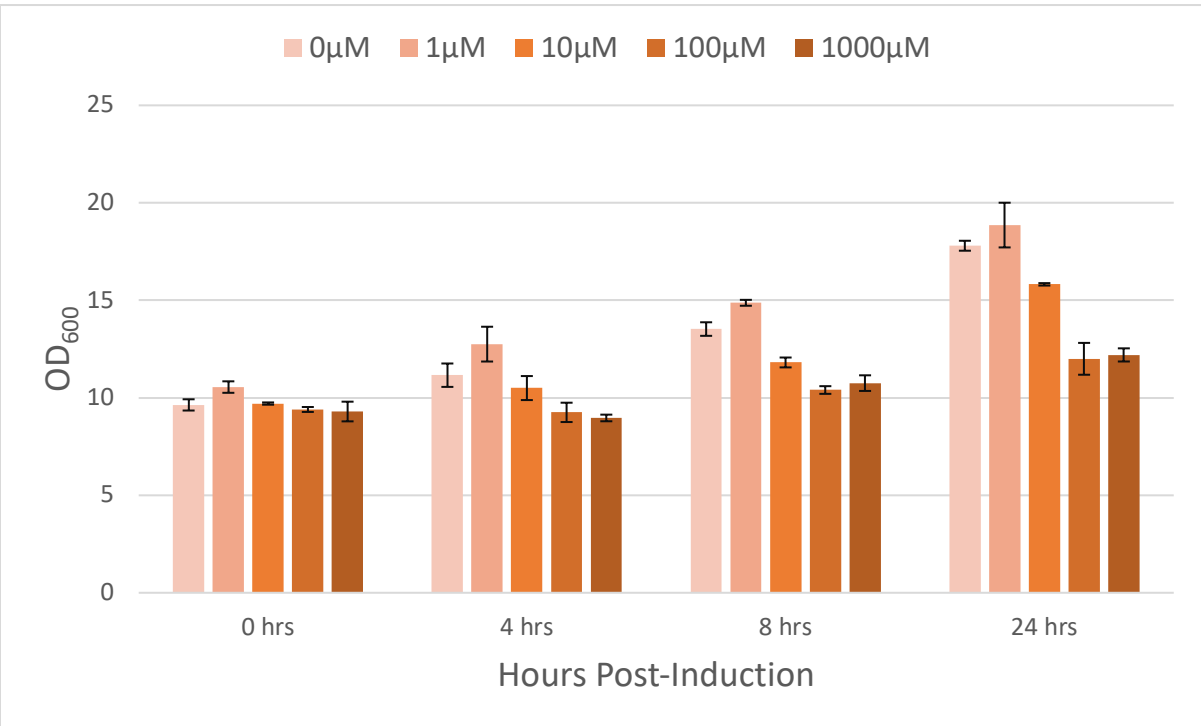


Figure 3.19: Post-induction OD_{600} measurements of E. coli BL21(DE3) containing pET26-SPssGFP at various IPTG concentrations, at 25°C, grown in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

3.3.3. Total protein concentration of *E. coli* expressing SPss-GFP grown in M2P growth medium

The concentration of total cellular protein (*i.e.* unfractionated protein samples), as determined by Bradford assays of *E. coli* BL21(DE3) containing pET26-SPssGFP grown at 37°C, 30°C and 25°C in M2P growth medium, is shown in Figure 3.20, Figure 3.21 and Figure 3.22, respectively. In all assays, a high total protein concentration was observed in the bacterial cultures treated with 10µM IPTG. An anomaly was observed in bacterial cultures grown at 25°C, where the 10µM IPTG treatment showed a high protein concentration before the addition of IPTG, and protein levels were elevated compared to every other time point. The high total protein concentrations did not reflect high levels of SPss-GFP protein in the cells. Total protein SDS-PAGE gels showed that observable levels of SPss-GFP only occurred in bacterial cultures induced with 100µM and 1000µM IPTG (see Section 3.4.2, Figure 3.32 and Figure 3.33).

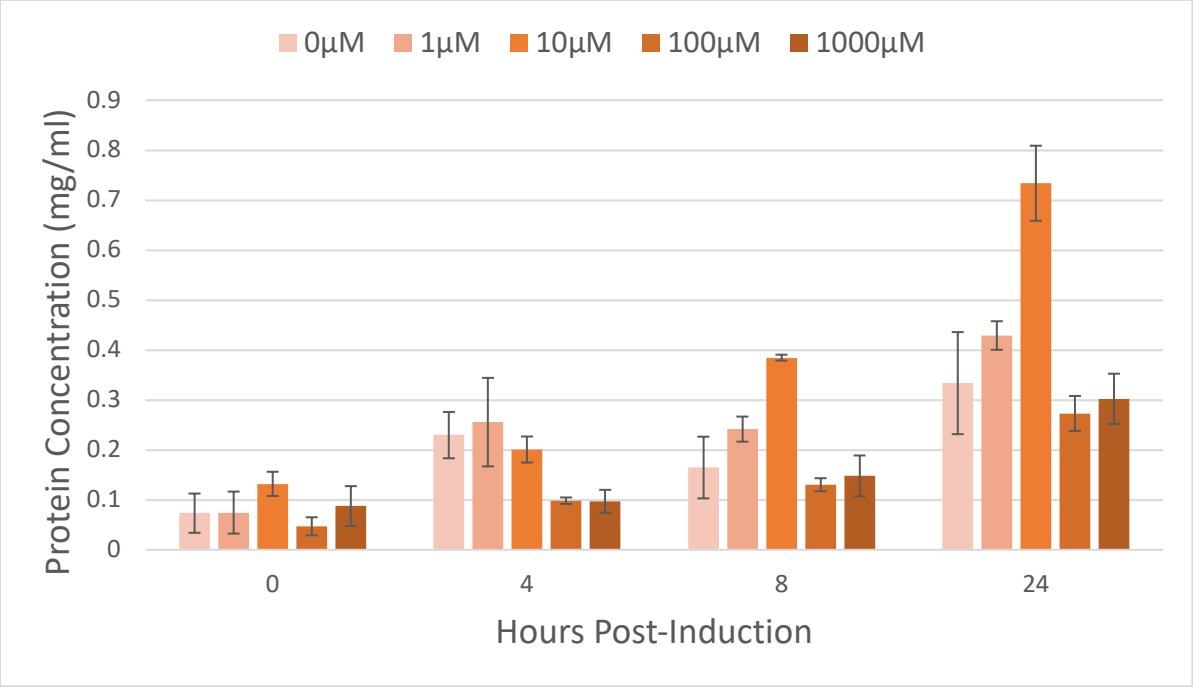


Figure 3.20: Protein concentration (mg/mL) of E. coli BL21(DE3) containing pET26-SPssGFP, at various time-points (0, 4, 8 and 24 hours) after induction of protein expression with IPTG and grown in M2P growth medium at 37°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

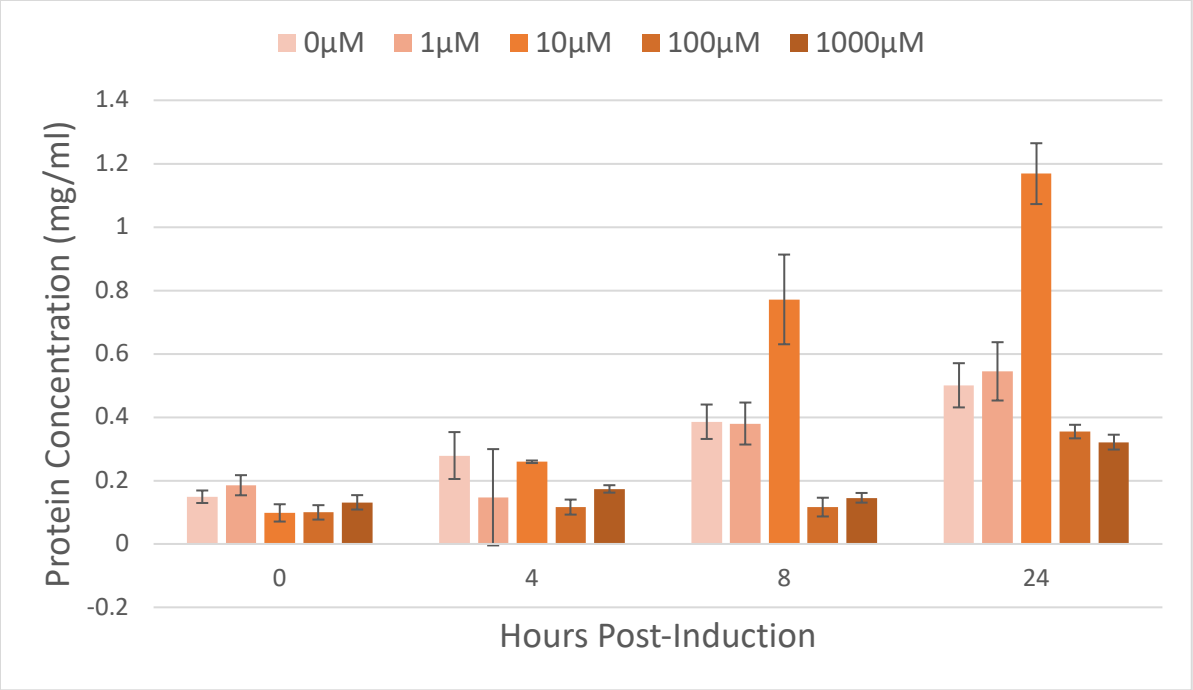


Figure 3.21: Protein concentration (mg/mL) of E. coli BL21(DE3) containing pET26-SPssGFP, at various time-points (0, 4, 8 and 24 hours) after induction of protein expression with IPTG and grown in M2P growth medium at 30°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

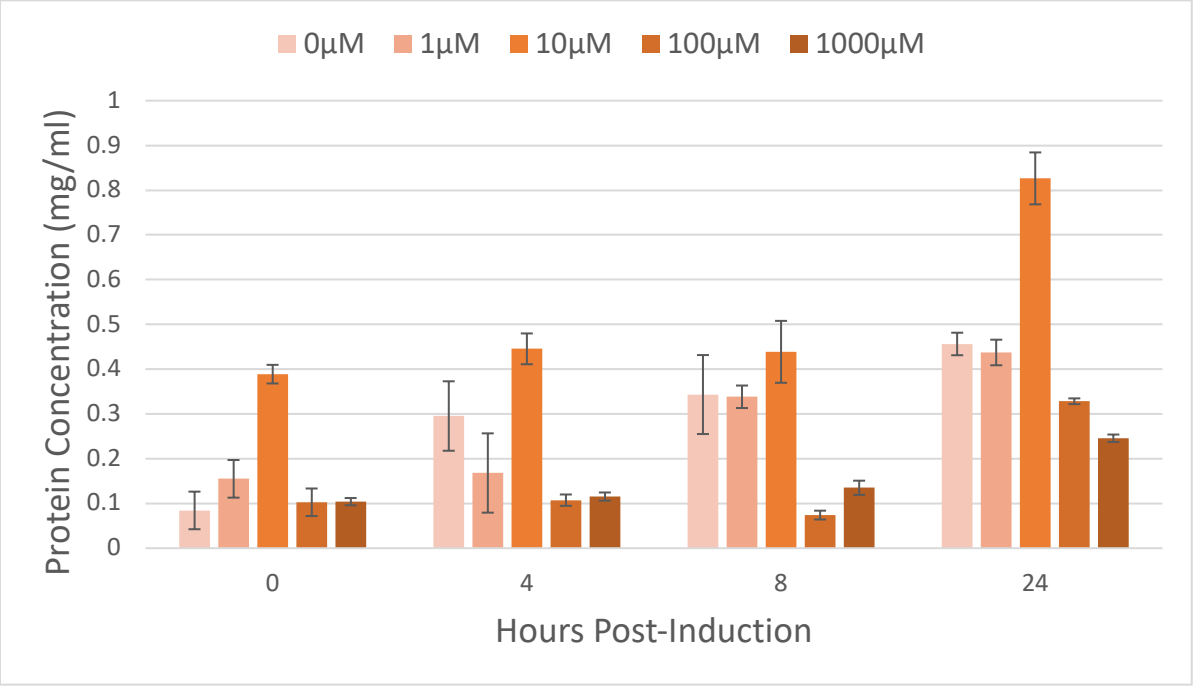


Figure 3.22: Protein concentration (mg/mL) of E. coli BL21(DE3) containing pET26-SPssGFP, at various time-points (0, 4, 8 and 24 hours) after induction of protein expression with IPTG and grown in M2P growth medium at 25°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

3.4. SDS-PAGE analysis of protein samples

3.4.1. Subcellular protein fractions of bacterial cultures grown in LB

Fractions of subcellular locals revealed that ETX produced by pET26-ETX was consistently transported into the periplasmic space (Figure 3.23 and Figure 3.25) and resided in the extracellular space in most cases (Figure 3.26 and Figure 3.29). The pET26-SPssGFP construct, on the other hand, did not show much, if any, presence of SPss-GFP in the periplasmic or extracellular fractions (Figure 3.25 and Figure 3.26; Figure 3.28 and Figure 3.29), and the presence of SPss-GFP was difficult to observe consistently in the cytoplasmic fraction under a variety of test conditions when the samples were grown in LB growth medium (no/low expression in Figure 3.24 compared to visible expression in Figure 3.27). To visualise the proteins from the extracellular space on the SDS-PAGE gels, the supernatant (obtained from centrifuging the bacterial culture) was subjected to a protein precipitation technique and resuspended in a reduced volume of buffer, concentrating the precipitated protein into a volume 10 times smaller than the source supernatant (*i.e.* protein from 300µL supernatant was concentrated into a 30µL volume).

The control *E. coli* BL21(DE3) pET26-Empty cultures for these experiments performed well and showed consistent growth throughout experiments. The pET26-ETX construct showed good growth characteristics, but inconsistent protein expression was observed in some experiments. Some inconsistencies manifested as the absence of the ETX protein in either the extracellular fraction or the periplasmic fraction. However, this was not the norm and the ETX protein was, more often than not, found in the periplasmic space in higher abundance than in either the cytoplasmic or extracellular space. The ETX protein was also observed in the cytoplasmic space (Figure 3.24 and Figure 3.27), but commonly in less abundance than in the periplasmic space (Figure 3.25 and Figure 3.28). The ETX protein commonly occurred in the extracellular space mainly at 24 hours after induction and to a lesser extent at 6 hours after induction (Figure 3.26 and Figure 3.29). Periplasmic fractions for the same time-points showed a similar accumulation of ETX at both 6 and 24 hours after induction, with slightly higher levels being observed at 24 hours post-induction (Figure 3.25 and Figure 3.28).

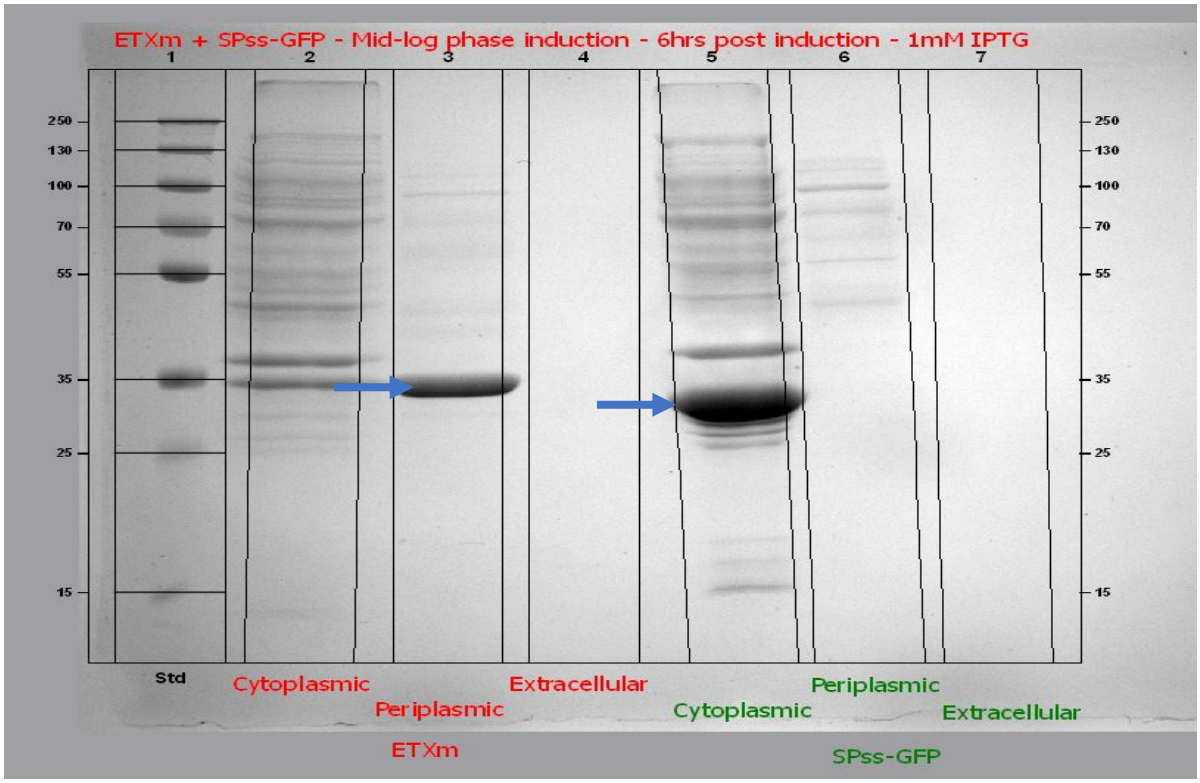


Figure 3.23: Comparison of cellular fractions of ETX toxin and SPss-GFP expression at 1mM IPTG, 6 hours post-induction, at 30°C, grown in LB growth medium. The 36kDa ETX protein can be observed in the periplasmic fraction (Lane 3), while the 31kDa SPss-GFP protein can be observed in the cytoplasmic fraction (Lane 5).

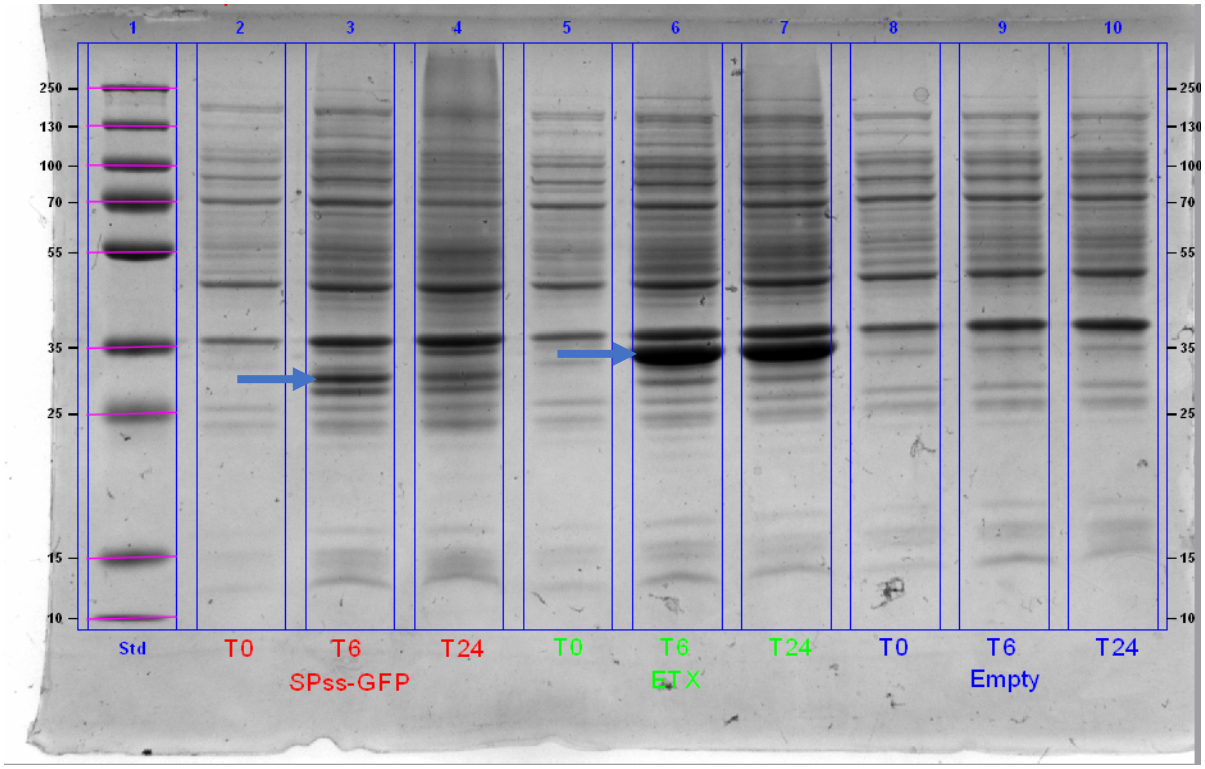


Figure 3.24: Cytoplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 37°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrows indicate SPss-GFP and ETX protein.

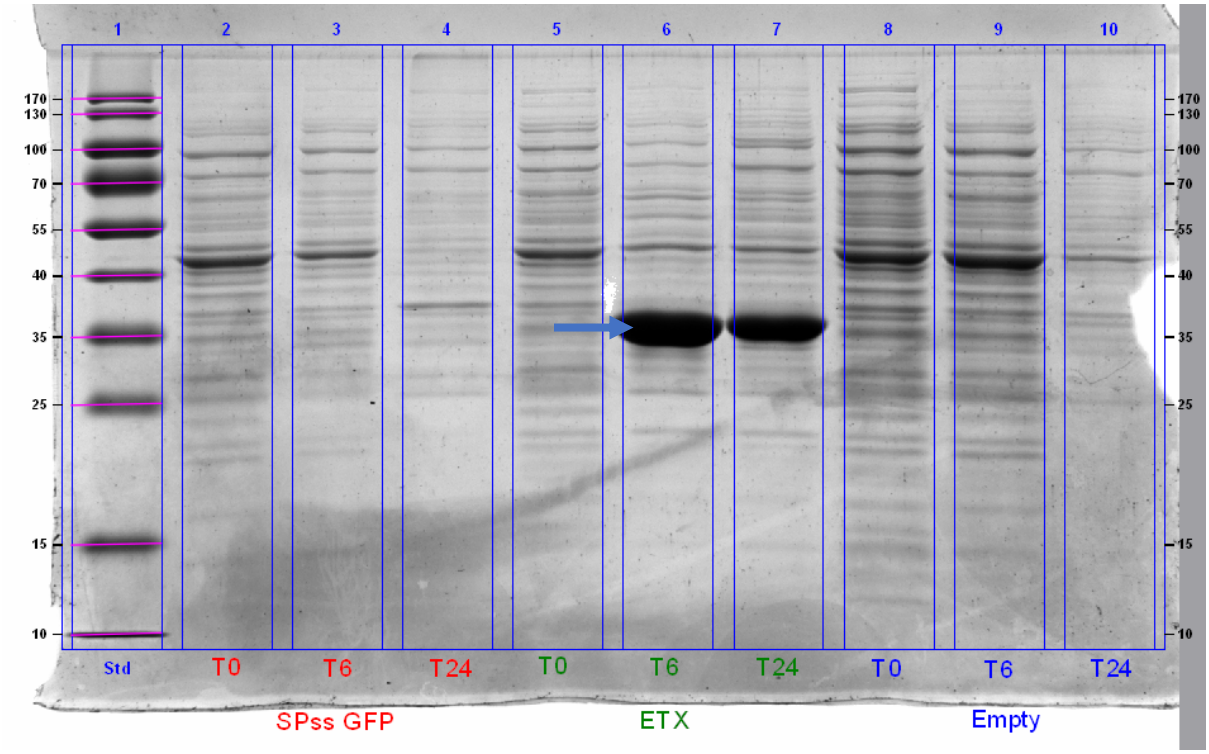


Figure 3.25: Periplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 37°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrow indicates ETX protein band at 6 and 24 hours after induction.

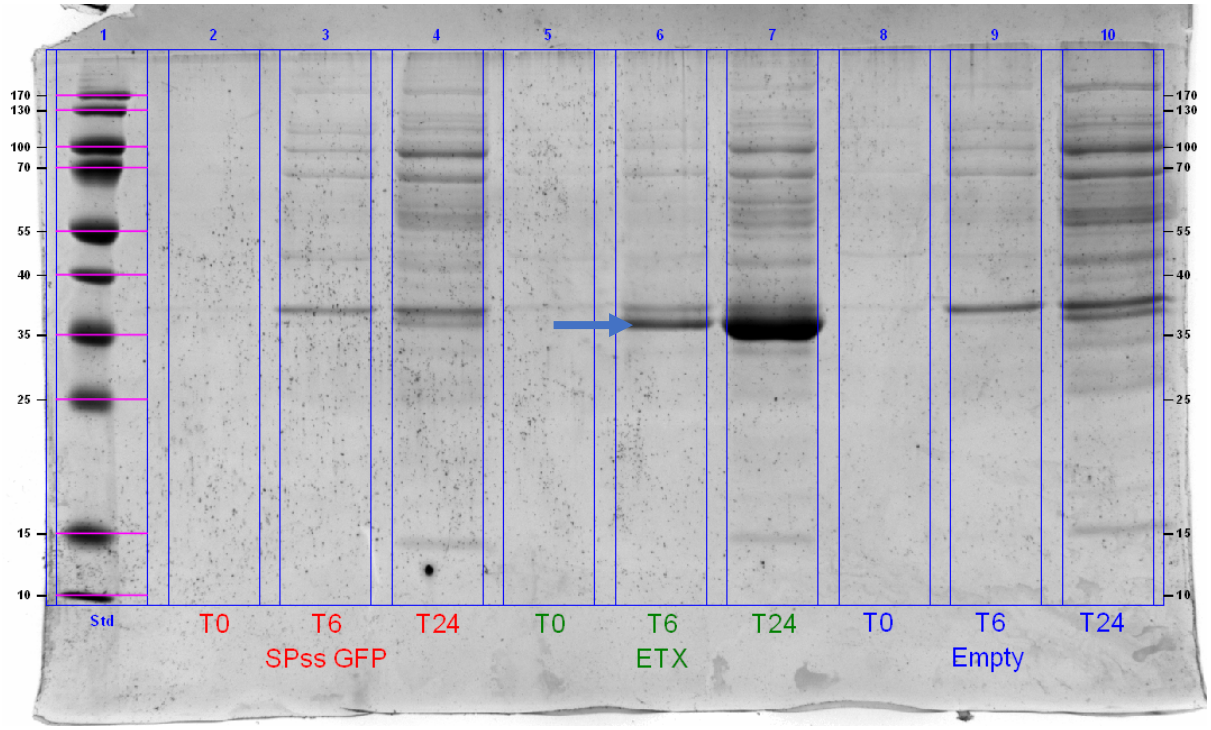


Figure 3.26: Extracellular protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 37°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrow indicates ETX protein band at 6 and 24 hours after induction.

Reducing the temperature of induction reduced the observed total protein expression in Empty, ETX and SPss-GFP constructs, as well as final OD_{600} readings (*i.e.* lower total bacterial biomass). This effect was most prominent at 25°C with minimal effect at 30°C (Figure 3.19 and Figure 3.18, respectively). The reduction in temperature did improve the expression of SPss-GFP (Figure 3.27), but only minor SPss-GFP translocation into the periplasmic (Figure 3.28) or extracellular (Figure 3.29) spaces was observed.

Although ETX protein expression was reduced, periplasmic (Figures 3.36) and extracellular (Figure 3.29) translocation was still observed. For periplasmic expression of ETX both 6 hours and 24 hours post-induction showed similar expression levels (Figure 3.28), while extracellular expression was reduced at 6 hours, and maximal at 24 hours (Figure 3.29).

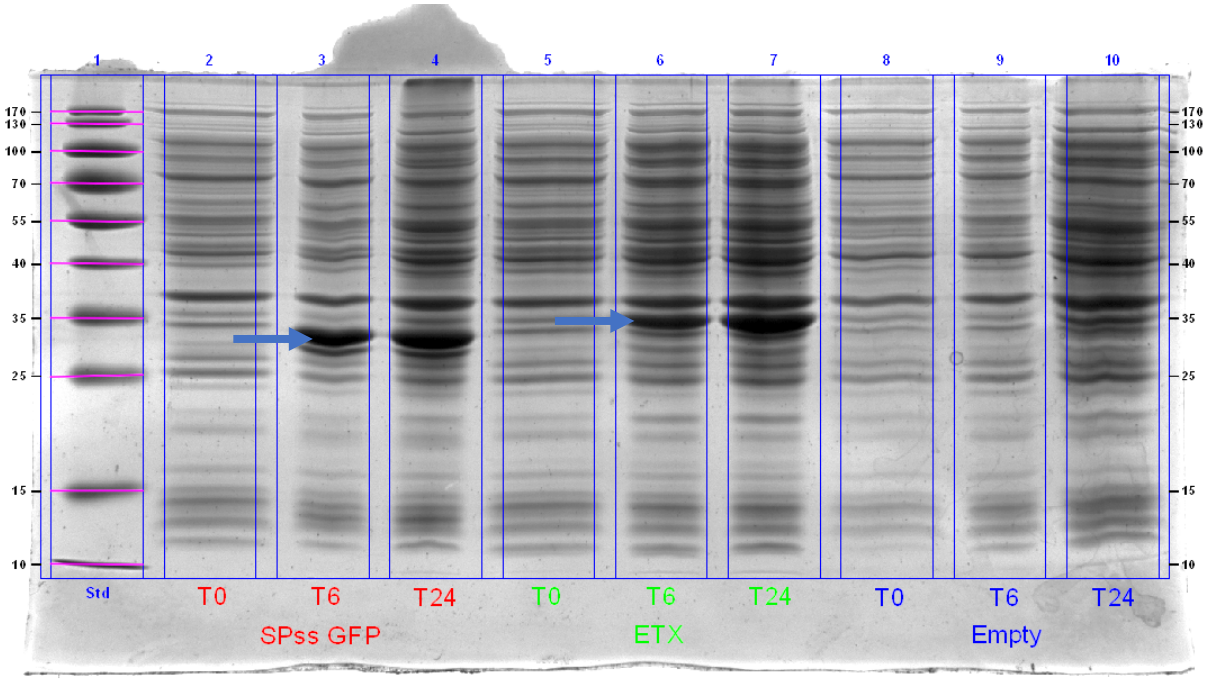


Figure 3.27: Cytoplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 25°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrows indicate SPss-GFP and ETX protein bands.

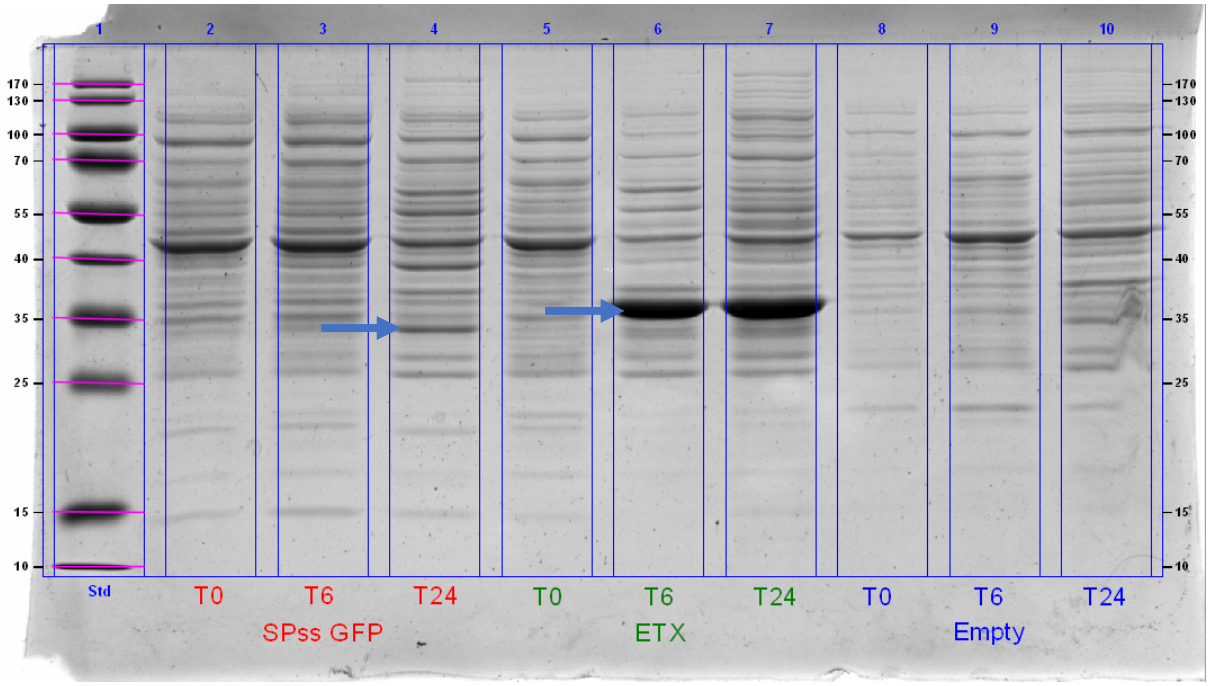


Figure 3.28: Periplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 25°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrow indicates the ETX protein band. Faint bands in the SPss-GFP (T24) sample may be SPss-GFP protein, but are very similar to bands observed in the Empty control making a conclusive identification difficult.

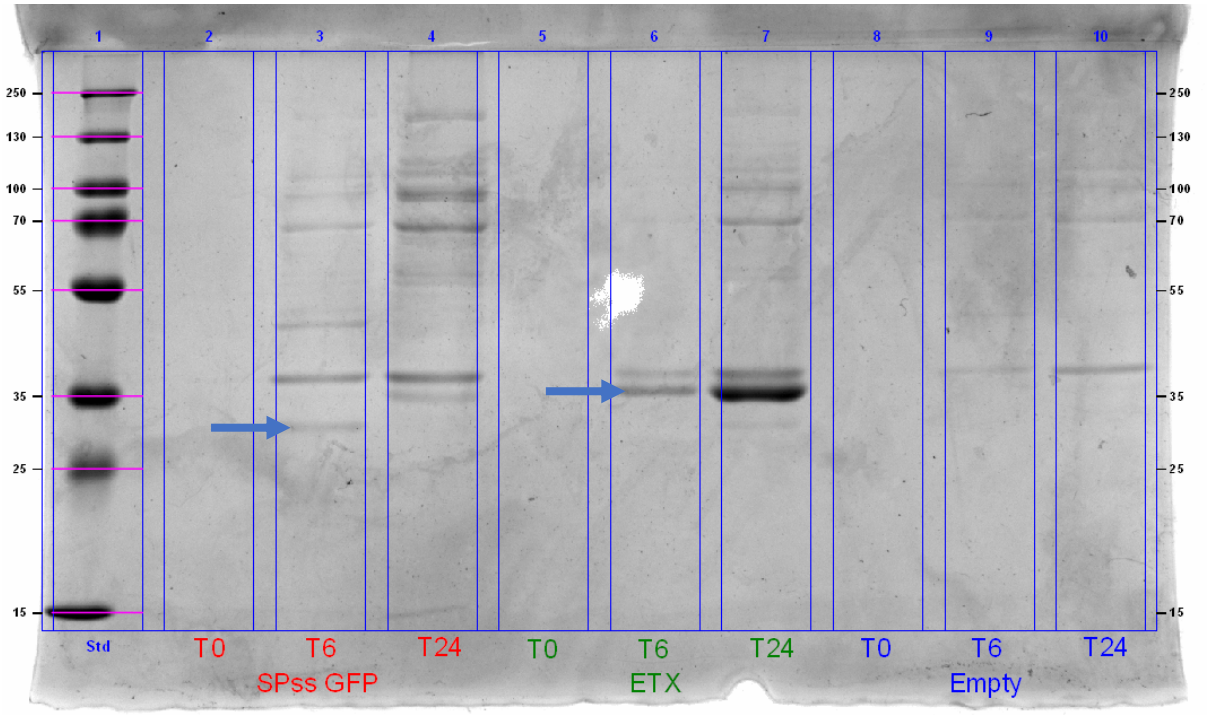


Figure 3.29: Extracellular protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µM IPTG, at 25°C, grown in LB growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrows indicate SPss-GFP and ETX protein bands. The SPss-GFP band at T6 is faint and not observed at T24.

3.4.2. Total protein concentration of *E. coli* expressing SPss-GFP grown in M2P

The threshold for SPss-GFP expression, in M2P growth medium, was 100µM and 1000µM IPTG at a minimum of 4 hours after induction (Figure 3.30 to Figure 3.33 shows total protein for all test conditions in M2P growth medium). Reducing the temperature during protein induction led to reduced levels of SPss-GFP. This was observed as extended time was required for SPss-GFP to accumulate to significantly observable levels (Figure 3.32 and Figure 3.33, T4 at 37°C compared to T8/T24 at 25°C), and lower levels of protein at terminal stages of induction when the expression is at lower temperatures. Background levels of protein did not appear to be affected by IPTG at any stage of induction, regardless of temperature.

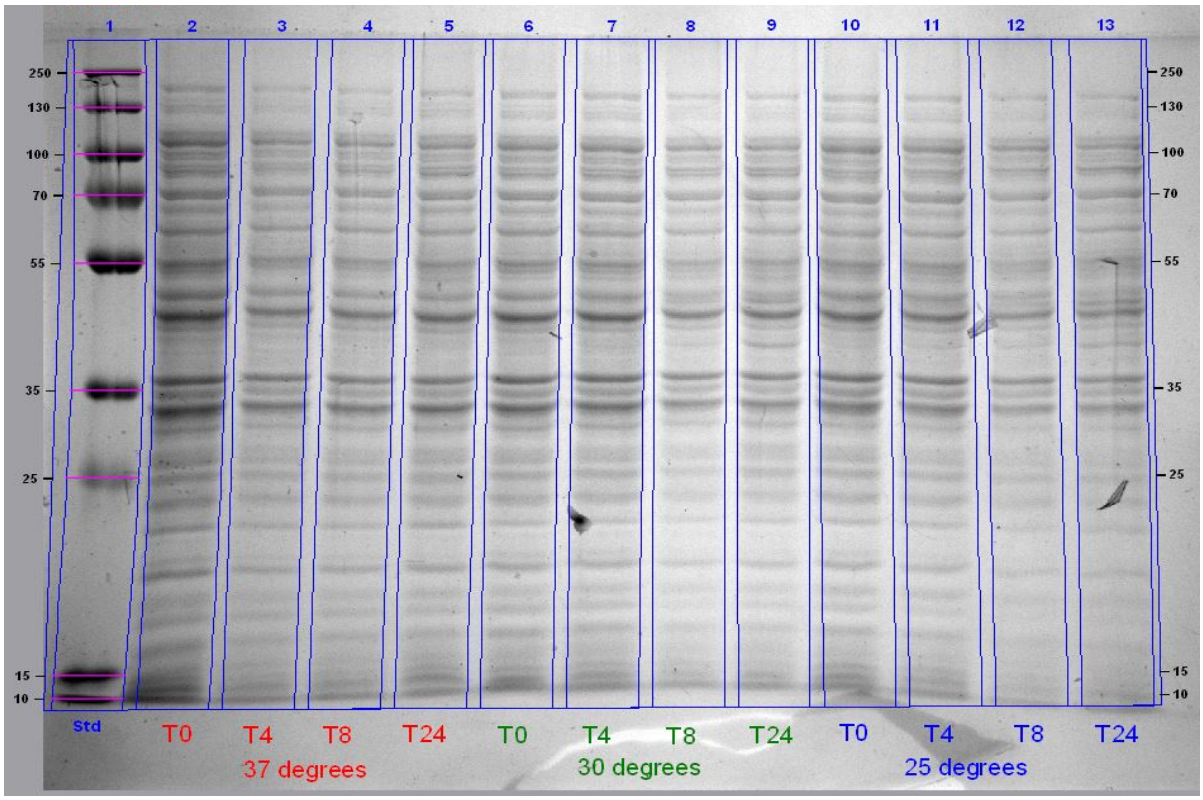


Figure 3.30: Total protein of E. coli BL21(DE3) expressing SPss-GFP. The bacterial cultures were induced with 1µM IPTG, at three induction temperatures (37°C, 30°C and 25°C), and grown in M2P growth medium. Lanes show samples at 0, 4, 8 and 24 hours post-induction. No SPss-GFP protein bands were observed.

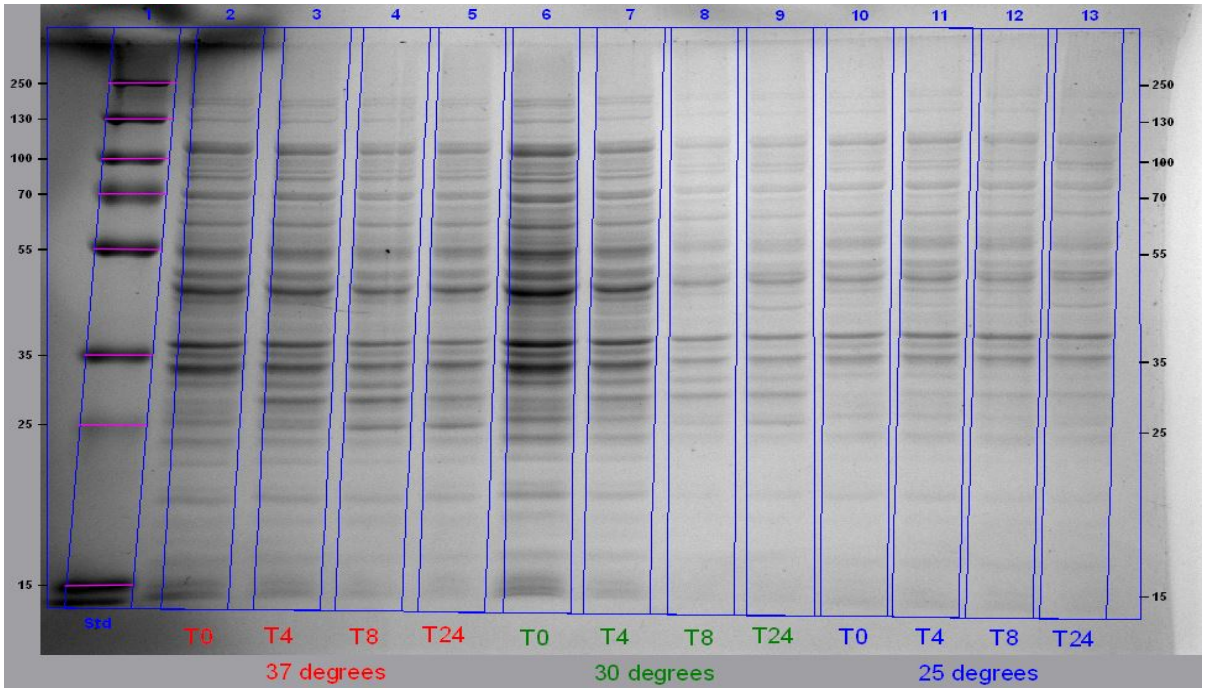


Figure 3.31: Total protein of E. coli BL21(DE3) expressing SPss-GFP. The bacterial cultures were induced with 10µM IPTG, at three induction temperatures (37°C, 30°C and 25°C), and grown in M2P growth medium. Lanes show samples at 0, 4, 8 and 24 hours post-induction. No SPss-GFP protein bands were observed.

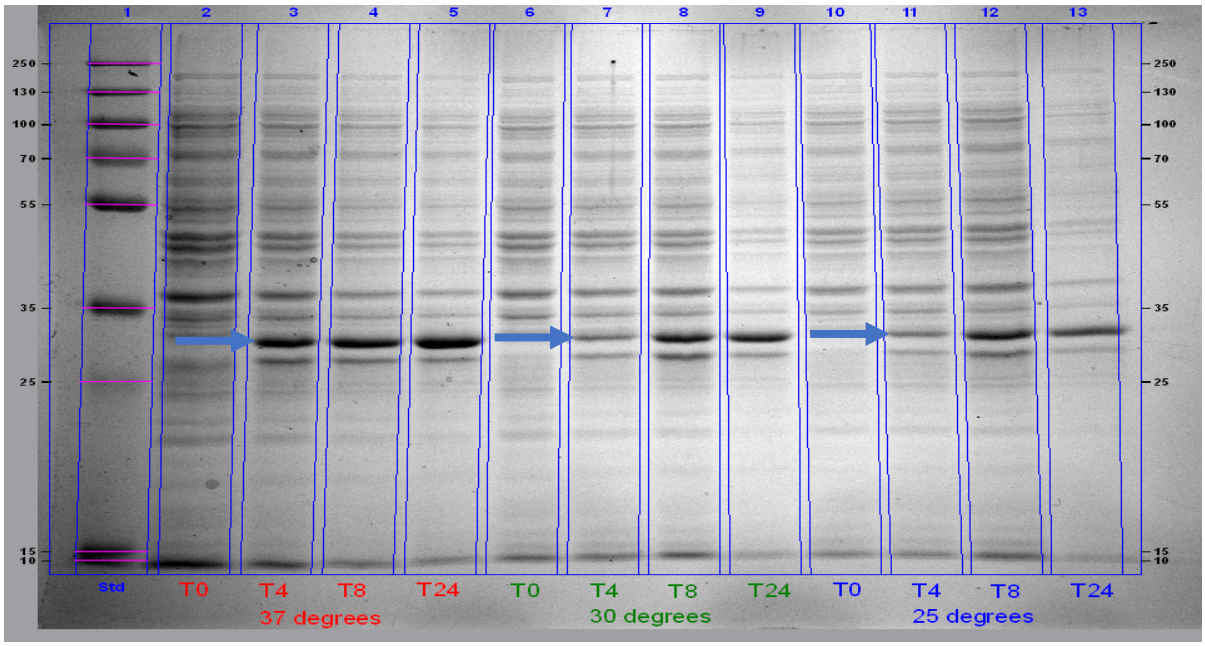


Figure 3.32: Total protein of E. coli BL21(DE3) expressing SPss-GFP. The bacterial cultures were induced with 100µM IPTG, at three induction temperatures (37°C, 30°C and 25°C), and grown in M2P growth medium. Lanes show samples at 0, 4, 8 and 24 hours post-induction. Arrows indicate SPss-GFP protein bands, visble after 4 hours of induction at all temperature conditions, but with variable protein levels.

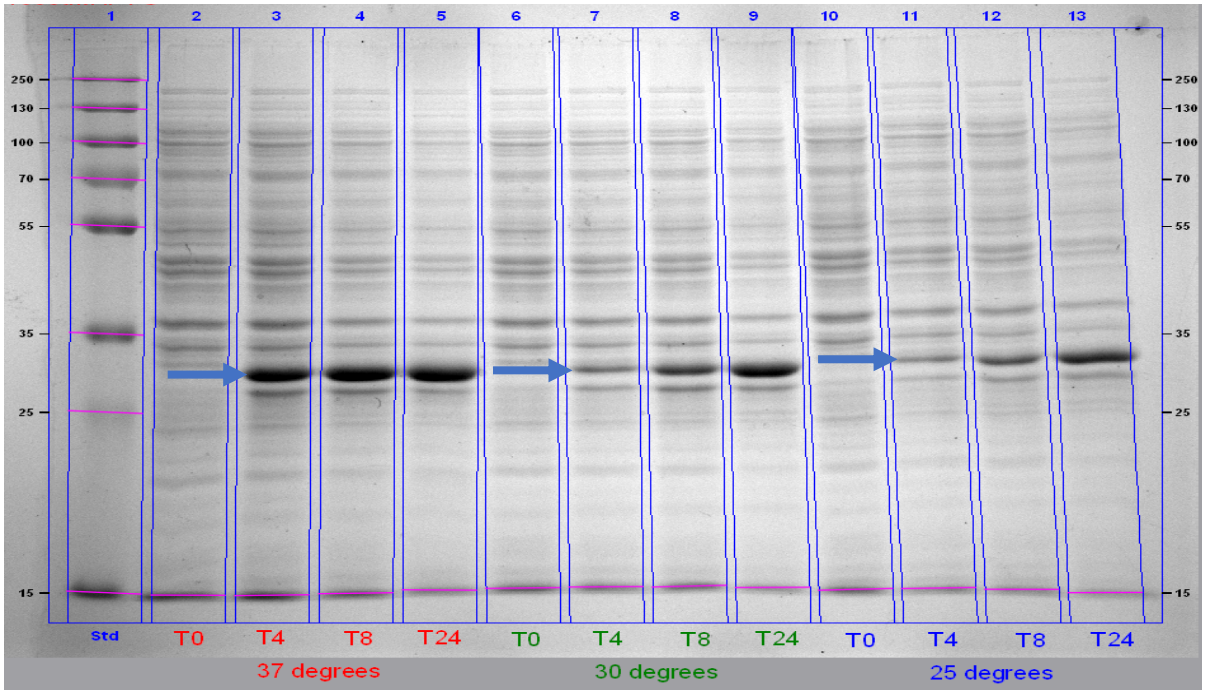


Figure 3.33: Total protein of E. coli BL21(DE3) expressing SPss-GFP. The bacterial cultures were induced with 1000µM IPTG, at three induction temperatures (37°C, 30°C and 25°C), and grown in M2P growth medium. Lanes show samples at 0, 4, 8 and 24 hours post-induction. Arrows indicate SPss-GFP protein bands, appearing after 4 hours of IPTG induction.

3.4.3. Optimal conditions for expression of SPss-GFP in M2P growth medium

Samples that were observed on the SDS-PAGE gels to express the most SPss-GFP were analysed and quantified further by gel densitometry using a BSA standard curve on an SDS-PAGE gel, and comparing the visual intensity of target protein bands using Image Lab software. The three highest concentration samples were quantified using various dilutions (Figure 3.34). Dilution-corrected absorbances showed that the conditions that produced the highest levels of SPss-GFP were 30°C at 100µM IPTG, followed closely by 30°C at 1000µM IPTG (Table 3.1).

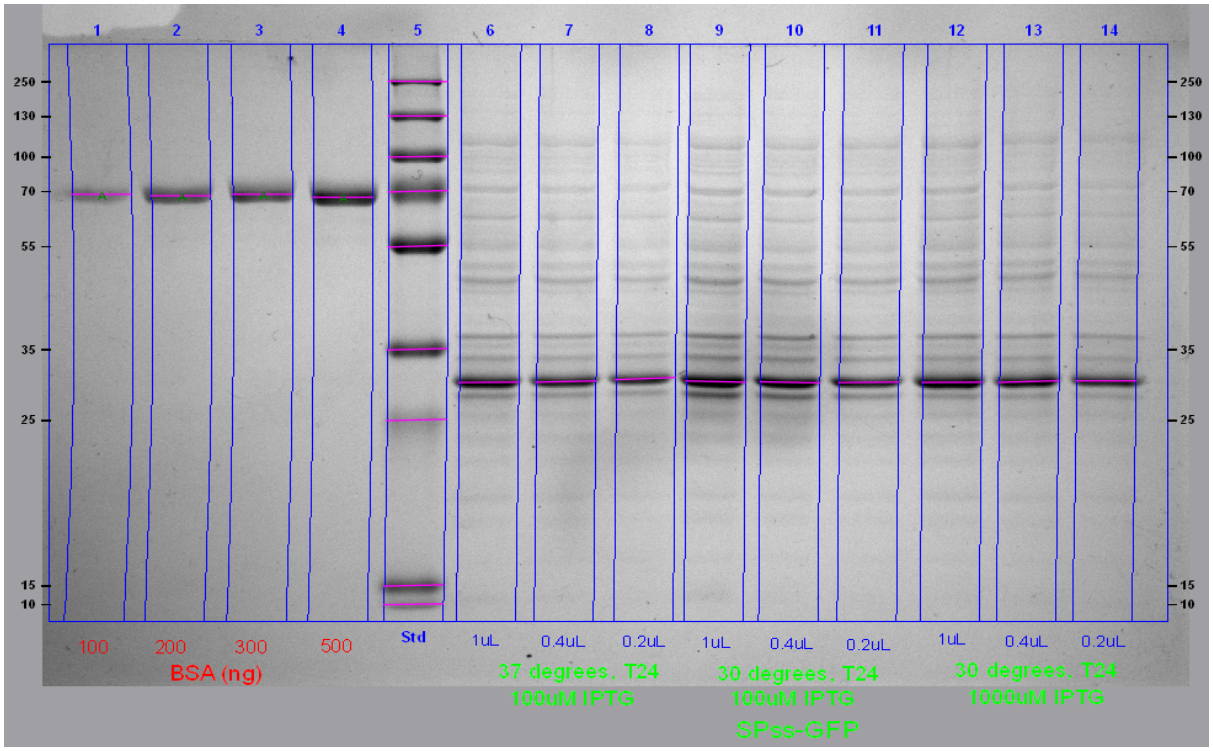


Figure 3.34: Optimal expression conditions of SPss-GFP in M2P. Unfractionated SPss-GFP protein samples were separated by SDS-PAGE and the intensity of the protein bands was measured using the Bio-Rad ChemiDoc and Image Lab software. A BSA standard curve was used to determine highest concentration of SPss-GFP and the corresponding culture conditions.

Table 3.1: Protein quantification results for SPss-GFP

Sample Dilutions	37°C, T24, 100µM IPTG			30°C, T24, 100µM IPTG			30°C, T24, 1000µM IPTG		
	1uL	0.4uL	0.2uL	1uL	0.4uL	0.2uL	1uL	0.4uL	0.2uL
Protein Quant. (ng)	381.7	298	204.8	461.8	439.3	305.5	453.5	349	271.5
Concentration (ng/uL)	381.7	745	1024	461.8	1098.3	1527.5	453.5	872.5	1357.5
Average (ng/uL)	716.9			1029.2			894.5		
Standard Deviation	322.1			536.2			452.4		

3.4.4. Subcellular protein fractions of bacterial cultures grown in M2P

Subcellular fractionation of bacterial cultures grown in M2P growth medium showed little to no periplasmic (Figure 3.36) or extracellular (Figure 3.37) translocation of SPss-GFP at any temperature. In the case of the ETX protein, cytoplasmic expression (Figure 3.35) as well as periplasmic translocation (Figure 3.36) were observed, but definitive ETX extracellular translocation was difficult to observe (Figure 3.37). The observation of ETX in the extracellular

fraction may be due to cell-lysis or the translocation of ETX protein. The band appears in the expected area (at or slightly below the 35kDa marker), but significant increases over background protein levels were not observed like those observed in the LB tests (Figure 3.26). Some native proteins that appear on the gel at a similar size range and a similar intensity to other, non-target proteins can be observed in the control samples (pET26-Empty), and cast some doubt as to whether the ETX protein was translocated during culturing. Compared with the periplasmic fraction (Figure 3.36), showing ETX accumulation as expected, the expected accumulation of ETX in the extracellular fraction was reduced and in many cases absent from the extracellular space (Figure 3.37).

Observations of the SPss-GFP protein was also difficult during this stage of experiments due to unknown factors and large variations of band intensity was observed. Although clear protein bands were observed (Figure 3.35), compared to samples analysed in previous experiments, these bands are of similar intensity to other, native cellular protein bands (Figure 3.33). Despite repeated testing, initial expression levels were not observed. Other tests were performed using a freshly transformed *E. coli* BL21(DE3) strain, but the results were the same as before. A possible reason was not accurately identified that could explain the differences in expression.

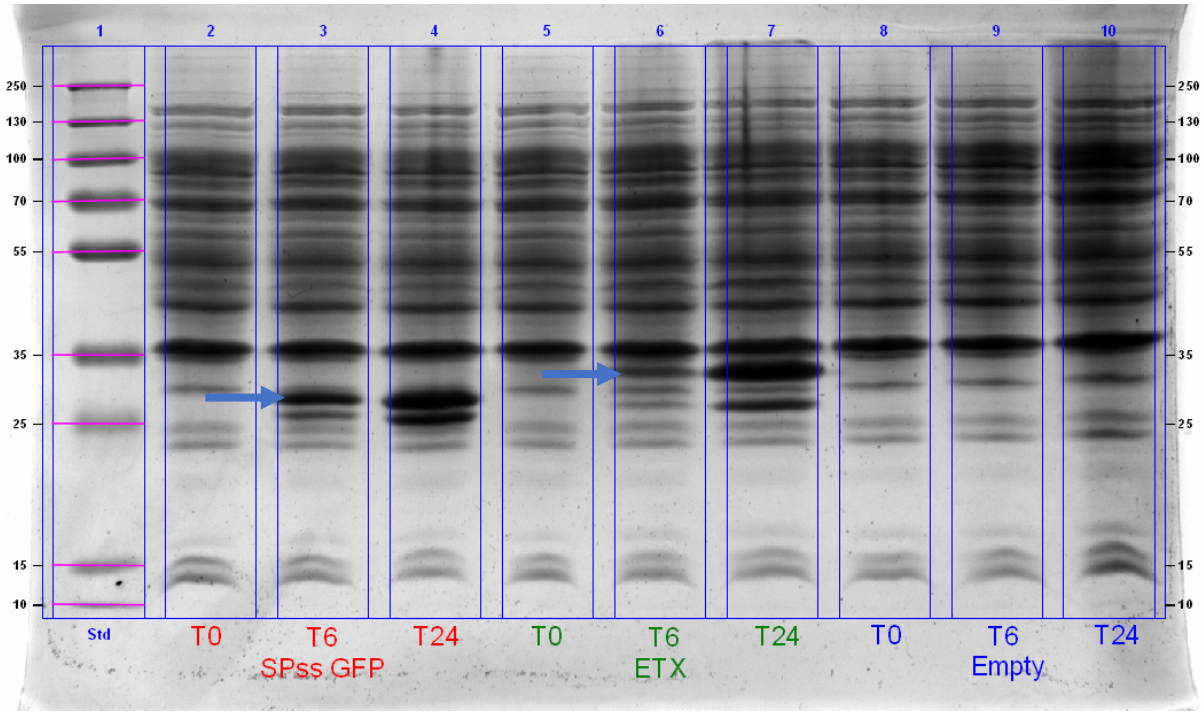


Figure 3.35: Cytoplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmids. The bacterial cultures were induced with 1000µM IPTG, at 25°C, and grown in M2P growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrows indicate SPss-GFP and ETX protein bands at 6 hours post-induction. SPss-GFP and ETX protein bands are also present at 24 hours after induction.

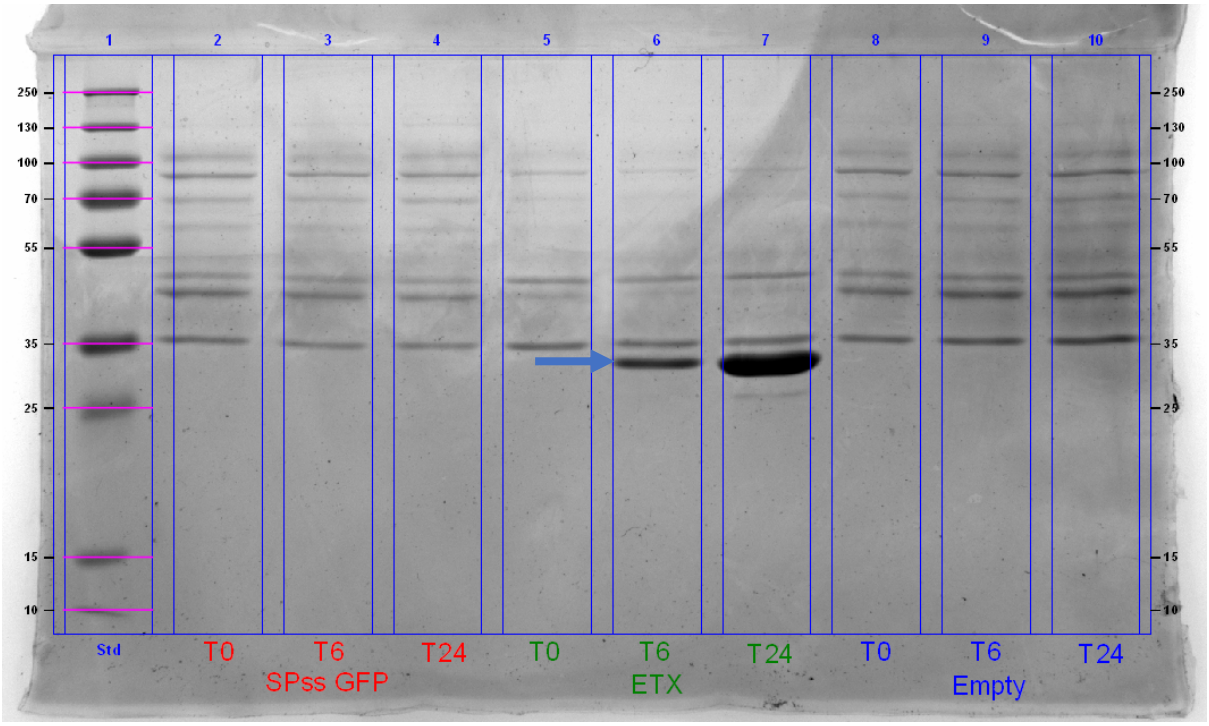


Figure 3.36: Periplasmic protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmid. The bacterial cultures were induced with 1000µM IPTG, at 25°C, and grown in M2P growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected. Arrow indicates ETX protein at 6 hours after induction. ETX protein is also observed at 24 hours after induction.

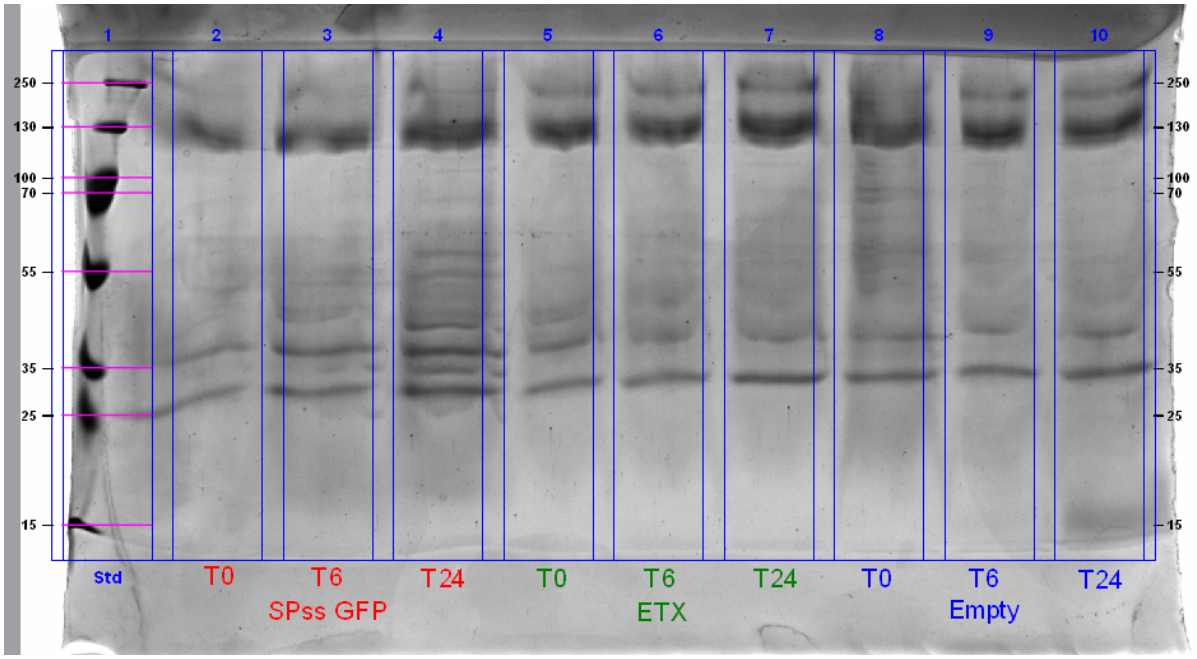


Figure 3.37*: Extracellular protein fraction of E. coli BL21(DE3) containing SPss-GFP, ETX or 'Empty' plasmid. The bacterial cultures were induced with 1000μM IPTG, at 25°C, and grown in M2P growth medium. Time-points at 0, 6 and 24 hours post-induction (T0, T6, T24 respectively) were selected.*

Chapter 4: Discussion

Although expression of the SPss-GFP protein was inconsistent, a clear link between protein induction (*i.e.* IPTG concentration) and cell biomass was observed. This link was negatively correlated, *i.e.* a higher IPTG concentration leads to reduced biomass accumulation. Temperature, bacterial biomass accumulation, and target protein levels were also correlated. Reducing the induction temperature reduced the overall biomass of the *E. coli* culture; this was only significant at 25°C, while 30°C and 37°C induction temperatures had similar biomass levels after 24 hours of induced growth. Significant reductions in biomass were observed at IPTG concentrations above 10µM (*i.e.* 100µM and 1000µM) when inducing SPss-GFP expression, but not in either positive (ETX) or negative ('Empty') controls. At IPTG concentrations above 10µM, significant levels of SPss-GFP were observed in SDS-PAGE gels.

The lack of SPss-GFP in the periplasmic fraction, in conjunction with mass spectrometry showing that the SPss region was removed, indicates that the SPss-GFP peptide was likely recognised by the cellular protein transport apparatus but no translocation occurred. These findings also seem to suggest that partial processing of the SPss-GFP peptide did occur, but the peptide likely became stuck or was rejected from the transport apparatus and left to accumulate in the cytoplasm, as shown by SDS-PAGE gel results.

The unsuccessful translocation of the SPss-GFP protein as well as its high levels of expression is the likely cause of the observed cellular stress and reduction in biomass accumulation. The reason for these observations may be due to a variety of factors, which could be broken down into the following areas:

- The SPss did not effectively interact with the *E. coli* General Secretion system.
- Incompatibility of the GFP peptide with the SPss, General Secretion system or associated chaperone proteins.
- Expression levels of the GFP and the overwhelming or blockage of the transport apparatus.

4.1. Bioinformatics of ETX and SPss-GFP

Bioinformatic analysis of the full-length ETX peptide sequence from the Gram-positive bacterium *Clostridium perfringens* showed evidence for a signal peptide in the 1-32 amino acid region. Although the analysis also showed weak evidence for this signal sequence being used in Gram-negative bacteria, research suggests that the 1-32 amino acid region of ETX functions as a transport signal in *E. coli* (Hunter et al., 1992). Further analysis of SPss-GFP showed that SPss-GFP is less likely to be transported compared to the ETX sequence. Experimental results indicated that ETX is localised in the periplasm of *E. coli* and only partially transported in the growth medium at the conditions tested here. Whether or not the extracellular ETX is due to active transport, passive diffusion or cell lysis should be investigated further.

Experimental results do indicate that the SPss functions as a signal peptide in its native protein form (*i.e.* ETX) but not when joined to GFP, in the form of SPss-GFP, a novel recombinant peptide. Bioinformatic analysis of SPss-GFP also suggests there is a change in the likelihood of SPss functioning as a signal peptide when joined to GFP. This suggests that there is some kind of interference to the functioning of SPss and further analysis is needed to narrow down what factors cause this interference and what would make an ideal fusion partner for SPss. Additional factors that should be considered are the amino acid sequences flanking the cleavage region of SPss, folding kinetics of the recombinant protein, the necessity for interaction of chaperones or usher proteins to help with protein folding and transport, changes to the expression system to allow variable control recombinant protein levels, and changes to the growth medium to aid in the periplasmic release of the target protein or reductions in cell stress.

4.1.1. Sequence motifs of signal sequences

Given the nature of the SPss peptide sequence observed through experimental results, *in-silico* analysis, and the fact that protein transport is observed in both *E. coli* (Gram-negative) and *C. perfringens* (Gram-positive), it is likely that the General Secretion (Sec) transport system is responsible for these observations (Green & Mecsas, 2015; Natale et al., 2008). The

Sec transport system is highly conserved in many organisms and transports proteins in an unfolded state via the binding of chaperons (*i.e.* SecB or signal recognition particles), which keeps the peptide unfolded and guides it to the membrane-bound SecYEG complex, allowing transport across the cytoplasmic membrane or integration into the membrane (Natale et al., 2008). For a peptide to be recognised by the Sec transport system, it requires a conserved motif to be present in the N-terminal region of the peptide, as well as the interaction of the chaperone-bound peptide and SecYEG complex (Chatzi et al., 2014; Tsirigotaki et al., 2017). The observed cleavage of the N-terminal sequence suggests that recognition of the signal sequence occurs but lack of periplasmic SPss-GFP indicates that transport stalls, or is not completed. This may indicate a problem with the recognition of SecB-bound peptide and the SecYEG complex (Fekkes et al., 1999; Kim & Kendall, 1998, 2000; Knoblauch et al., 1999; Randall, 1992). This is assumed to be one of two potential problems: SecB is not binding to the GFP peptide, or the SecB-bound peptide is not being recognised and processed.

4.1.2. Mass spectrometry (MS) data and protein processing

The MS data shows that the N-terminal region of SPss-GFP was processed or cleaved in some way. This cleavage site was four amino acids downstream of the predicted cleavage site. This is likely due to the loss of sample during the processing of protein samples for MS analysis. During the preparation of MS samples, the peptide is digested with trypsin, which cleaves amino acids arginine and lysine. Some of these sites within SPss-GFP form small fragments when digested with trypsin. These small peptide fragments might be degraded or otherwise undetected, as is the case for some MS methods and apparatus (Steen & Mann, 2004; Zhang et al., 2010). The MS data indicates that the regions of low probability of sequence match are flanked by lysine or arginine residues. It must be noted, however, that this is not the case for all peptide regions with flanking arginine and lysine sites in the SPss-GFP sequence. The loss of small peptides during MS sample processing is one area that could be investigated further to assess the accuracy of these results.

Another reason for the observed difference in predicted and measured cleavage site could be a shift in the cleavage site due to changes in amino acids that form the restriction enzyme site in the signal peptide region. This region was altered in previous work so that the SPss region

could be digested and ligated to other heterologous protein sequences. The ligation of GFP and SPss sequences could also alter the cleavage site, which could explain the difference in measured and predicted cleavage site.

4.2. Protein transport and cell stress

The cellular stress response can be triggered by a variety of factors, such as chemical or protein toxins, pH, heat, cold, nutrition and competition (Chung et al., 2006; Gerdes et al., 2005; Jozefczuk et al., 2010). While specific stress responses exist, such as the toxin-antitoxin response, others are more general such as heat and metabolic stress responses that result in reduced protein expression, increase in chaperone production, as well as increased protease activity (Harcum & Bentley, 1999; Schweder et al., 2002). In general, the bacterial stress response leads to an overall reduction in growth while the bacteria adjust to the present stress.

Oversaturation of the protein transport apparatus is a common cause of cellular stress during protein expression (Khokhlova & Nesmeyanova, 2004; Ytterberg et al., 2019). The main factor that contributes to this stress is the blockage of the transport apparatus by misfolded proteins. While the transport stalls, proteins vital to cell homeostasis cannot reach their destinations; this leads to an increase in protein degradation and heat-shock proteins, which are involved in protein folding and cell stability (Harcum & Bentley, 1999).

An additional factor is the binding of chaperone proteins during the expression of the SPss-GFP peptide. While SecB does not bind directly to the signal peptide, it binds instead to the central region of the polypeptide and assists in recognising and threading the bound polypeptide through the secretion channel, specifically SecA, an ATP-powered nanomotor that provides the motive force for protein translocation (Chatzi et al., 2014; Khokhlova & Nesmeyanova, 2004; Smith et al., 1997). The interaction between SecB and SecA requires a divalent cation (like Cu^{2+} or preferably Zn^{2+}) (Fekkes et al., 1999).

Some factors that contribute to the blockage of the Sec transport apparatus are the folding rate of the transported protein and the aggregation of proteins before transport (Low et al., 2013). The folding rate is critical to the binding of SecB to the protein; this keeps it in an

unfolded, transport-ready state before being shuttled through the SecYEG complex. While SecB binds a large range of peptide sequences, the folding rate and the presence of positively charged peptides in the middle of the sequence are two important factors when considering the binding of SecB and the transport of the peptide through the SecYEG complex (Randall, 1992; Smith et al., 1997).

4.2.1. Protein folding versus transport

Protein folding prediction of the SPss-GFP peptide sequence indicated that a stable GFP chromophore would form and the SPss sequence would not affect chromophore formation during expression. This assumption was contradicted by experimental results. Further scrutiny of the online tool used for protein folding prediction (Phyre2) revealed that the peptide sequence is matched to existing databases of known proteins and their folded conformations and not re-simulated for variations in folding kinetics due to changes in amino acid sequences. Because the SPss sequence is very small in comparison to the GFP sequence, the match of SPss-GFP with the unmodified GFP sequence was greater than 95%. What is likely is that when SPss-GFP is being expressed, the SPss signals the newly forming peptide to bind to the SecYEG complex. As the GFP section of SPss-GFP is translated, it begins folding before SecB can bind to the majority of the peptide and keep it in an unfolded state. The GFP section begins to fold while also being transported.

Observations of GFP fluorescence and the lack of SPss-GFP fluorescence indicate that the SPss-GFP chromophore does not correctly fold when expressed. This indicates that the addition of a transport tag prevents correct protein folding. The function of SPss is to transport any attached peptide in an unfolded state, but this is in direct competition with the GFP sequence when it tries to fold to form a functional chromophore. This competition in the folding of conformational states is likely the reason for both the non-functional GFP chromophore and the cellular stress observed in the form of reduced cellular biomass, which is due to the inability to complete the transport of the SPss-GFP peptide. The SPss sequence region is likely in an unfolded state and is amenable to processing. After the SPss region has been accepted by the transport apparatus, it is cleaved and degraded, hence why there is no evidence of SPss in the MS results (Kim & Kendall, 2000). Once the GFP region begins being

threaded through the transport apparatus, an already folded region of GFP is encountered and blocks the transport apparatus. While the apparatus is blocked, no other proteins can be transported which leads to the observed cellular stress.

4.3. Effects of growth medium

The growth medium used to culture *E. coli* is the basis for all metabolic activity that can occur during the growth of the microbe and expression of the heterologous protein. Not only does the growth medium supply all the necessary primary metabolites for the base cellular structure, but it can also contains a large variety of secondary metabolites that can influence growth rates, cellular kinetics (such as membrane stability) and the ability of the cell to handle stress (Jensen & Carlsen, 1990; Krause et al., 2016; Losen et al., 2004; Sánchez et al., 2009; Shokri et al., 2002). A common approach to growth medium composition utilises a carbon source (often glucose or other saccharides), nitrogen, phosphate and various combinations of buffer salts and micronutrients (including sulphur, magnesium, potassium, iron, manganese, zinc and copper)(Pelczar et al., 1960; Shiloach & Fass, 2005). In addition, secondary metabolites, produced during metabolises, can interfere with growth rate and recombinant protein production. These factors must be taken into account for optimal production, but it is assumed that these have been optimised in commercially available growth media, such as M2P.

4.3.1. Growth kinetics

The rate of growth of microbes can be simplified to how effectively the carbon source is taken up and utilised by the cells. Glucose is one of the most common carbon sources for *E. coli* cell culture and is utilised readily during growth and protein expression, in addition to nitrogen, phosphorous, potassium and various other micronutrients (Bren et al., 2016; Harrison et al., 1997; Jensen & Carlsen, 1990; Panula-Perälä et al., 2008). The rate at which the carbon source is utilised is varied but mainly consists of the type of carbon source (*i.e.* glucose is easy to take up and breakdown into useable energy, while cellulose is not), the rate at which *E. coli* can transport the carbon source into the cell, and the concentration of the carbon source (Bren et al., 2016; Shokri et al., 2003). Growth conditions such as temperature and metabolic

limitations due to reduced accessory nutrients, also play a central role in metabolic activity and therefore carbon uptake and utilisation (Gadgil et al., 2005; Wurm et al., 2016).

One major factor that is focussed on during bacterial cultivation is carbon source concentration and feed rate during growth. Traditionally there have been two approaches to bacterial cultivation, namely batch or fed-batch systems. Batch cultivation involves preparing a set volume of growth medium with all required nutrients, sterilising and inoculating it with an active culture. The growth conditions such as temperature and induction time are chosen, and the batch of bacterial culture is then harvested at the desired time. The batch method is usually optimised for final cell density or target protein yield (or both). In contrast, fed-batch systems also begin with a batch of defined growth medium but additional nutrients and metabolites are added at specific times and specific rates to the bacterial culture (Glazyrina et al., 2010; Horn et al., 1996; Krause et al., 2016; Liew et al., 2010; Panula-Perälä et al., 2008; Ramchuran et al., 2003; Rosano & Ceccarelli, 2014; Shiloach & Fass, 2005). Additions to the growth medium can include glucose, lactose, yeast extract, arabinose and many others. Care must be taken as the metabolic pathways shift when using different carbon sources and changes in growth kinetics can be observed (Bren et al., 2016). It is important to carefully select growth medium, or any additives, for the desired application and organism.

Another mechanism that is used to control the rate of growth, through the limitation of carbon source, is using an indigestible polysaccharide which is only accessible to the *E. coli* when digested by an external enzyme. This is the basis for the M2P growth medium used in these experiments. The rate at which the enzyme can digest the polysaccharide is, therefore, the limiting factor in the rate at which the *E. coli* can grow (Krause et al., 2016). In comparison, *E. coli* grown in LB will have access to all available nutrients and grow as rapidly as possible. This rapid growth can easily overtake the oxygen transfer rate of the growth medium. When oxygen is limited and nutrients are abundant, it leads to acetate accumulation, which inhibits further growth (Losen et al., 2004; Shiloach & Fass, 2005). This inhibition is reduced in fed-batch and enzyme-release growth media. This allows for a far higher final cell density than in batch media.

In some scenarios, it is possible to remove growth-inhibiting metabolites from the growth medium using dialysis. This allows for very high levels of bacterial growth but can require high volumes of the solution used in the dialysis process (*i.e.* the solution used to dissolve the metabolites from the culture must always be at a lower concentration than the metabolites in the culture itself) (Shiloach & Fass, 2005). This means that a dialysis technique may use a large amount of solution to keep the level of inhibitory metabolites low.

4.3.2. Release of proteins into the extracellular space

The observation of low levels of target protein could indicate a variety of mechanisms of protein release. These mechanisms could involve active transport through a transport apparatus, membrane disruption (or cell leakiness), or as a result of complete cell lysis which is likely more prominent at the later stages of growth (*i.e.* late stationary and death phases).

The mechanism of active transport of a protein requires recognition of a transport machinery associated with transport from the cytoplasm, periplasm (or both) to the extracellular space (or growth medium). This could be associated with the activity of Type I, II, or III secretion systems (or other systems mainly associated with pathogenesis); however, efficient transport of recombinant proteins through these systems is only commonly observed with the aid of chaperone or usher proteins (Yoon et al., 2010).

If the protein resides in the periplasmic space, the addition of certain molecules can destabilise the outer membrane of *E. coli* enough to release the proteins from the periplasmic space. This can be done in the process of osmotic shock described in the Methods section, or it can be done while the bacteria is still in the culture using more subtle forms of membrane disruption such as 350mM Tris incubation for several hours, followed by mild heat-shock of 38°C (Wurm et al., 2017). A variety of chemicals have been found to disrupt the outer membrane of *E. coli* and release proteins contained in the periplasmic space and some can even be applied without killing cells (Bao et al., 2016; Jalalirad, 2013; Tang et al., 2008). This has applications for use in continuous bioprocesses, but more investigation is needed into the effects of continuous membrane disruptions as well as other factors of cellular stress and growth in this context.

4.4. Possible improvements

A variety of challenges arose during experimentation and could be improved on. The number of test conditions, although seemingly low, resulted in a large sample set that prolonged experimental time. The test conditions comprised temperature (*i.e.* 25, 30, and 37°C), and induction conditions (*i.e.* IPTG concentration 0, 1, 10, 100, 1000µM) and the need for observation of three cellular fractions (cytoplasmic, periplasmic, and extracellular) of three different expression systems (SPss-GFP, ETX and Empty plasmids – test, positive control, and negative control, respectively). All of these tests were sampled, initially at four time-points (0, 4, 8, and 24 hours after induction), later reduced to three time-points. This resulted in 540 samples generated for one media type, all of which were analysed using several methods (*i.e.* OD_{600}, SDS-PAGE, BSA Quantification, etc.). Repetition of experiments due to inconsistent expression of SPss-GFP exacerbated this problem. This becomes an even bigger issue when considering that many other factors were excluded from analysis, such as pH, induction time, acetate levels, dissolved oxygen, elements of growth medium, and the addition of protein chaperones.

One solution would be to use a form of multivariate analysis. However, this would require more specialised skills and careful attention to the experimental set-up and data analysis. This would allow for the testing of more variables with fewer samples. This does beg the question of how does one determine the correct or manageable number of variables to analyse in an experiment. Having more data is often considered good in an experimental setting, but challenges of complexity, time and resource constraints can be a critical factor in determining the scope of research. The availability of specific skills and methods are also a limiting factor in experimentation and data analysis. If the skills, or apparatus, is not available the tests cannot be done and the results cannot be analysed.

It is also recommended to include more than one signal peptide in further experiments as well as other fusion partners, because the compatibility of peptides may be a simple fix that would result in reduced cellular stress without the need for additional changes.

The bio-informatic analysis gives a rough prediction of efficacy and should be the first type of analysis done. The addition of any additional genetic elements, such as chaperones, should also be taken into account.

Chapter 5: Concluding Remarks

The production of heterologous proteins is an important part of industrial and medicine production. The discovery of new biological tools and target molecules is ever-expanding and new methods are sought after to improve the efficiency of production. One such bottleneck with efficient production is the separation of target peptide from the rest of the cellular structures and native peptides. Using the endogenous transport machinery of the host cell to help separate the target peptide from non-target molecules was explored in this study.

Peptide compatibility is the most challenging aspect of this work. While *in-situ* and *in-vivo* experiments showed the presence of a functional signal peptide, making practical use of this sequence remains challenging. The signal sequence itself is not the only factor to take into account when experimenting with heterologous protein transport. The host organism, the peptide bound to the signal peptide, how the heterologous protein is expressed and the interaction of the peptide with accessory elements during transport are all critical factors that should be addressed before experimentation.

Without careful consideration and optimisation of protein expression, the addition of a signal peptide to the expression system can complicate optimisation. The use of an extensive amount of data points and analytic methods is unavoidable, unless careful considerations are given to the genetic elements and tools used. This consideration will hopefully reduce the amount of measurement and optimisation needed but the addition of every genetic element can alter the ideal expression parameters for a given construct. This makes the optimisation methods effective primarily for a given construct and less useful when trying to expand these genetic elements to other target proteins. This means that each new heterologous construct has the potential need for optimisation specifically suited to the expression conditions and signal peptide used. While there is experimental literature that shows that certain signal peptides can be utilised in *E. coli,* this does not mean that a target heterologous protein will be compatible with the transport pathway being utilised. In addition, chaperones and other protein elements should be considered when looking at the transport and folding of proteins in *E. coli*, as these accessory elements can play a significant role in the efficient expression and purification of a target protein.

While the considerations and preparations for the expression of a heterologous protein are complex and highly specific, continued exploration of genetic elements, and the conditions under which they are effectively utilised, is an integral part of biotechnology and the discovery of new tools and elements to streamline this process are highly valuable, making these challenges well worth the effort.

References

Agilent Technologies. (2014). *pET System Vectors and Hosts*.

Anilionyte, O., Liang, H., Ma, X., Yang, L., & Zhou, K. (2018). Short, auto-inducible promoters for well-controlled protein expression in *Escherichia coli*. *Applied Microbiology and Biotechnology*, *102*(16), 7007–7015.

Auclair, S. M., Bhanu, M. K., & Kendall, D. A. (2012). Signal peptidase I: Cleaving the way to mature proteins. *Protein Science*, *21*(1), 13–25.

Bao, R. M., Yang, H. M., Yu, C. M., Zhang, W. F., & Tang, J. B. (2016). An efficient protocol to enhance the extracellular production of recombinant protein from *Escherichia coli* by the synergistic effects of sucrose, glycine, and Triton X-100. *Protein Expression and Purification*, *126*, 9–15.

Baumgarten, T., Ytterberg, A. J., Zubarev, R. A., & de Gier, J.-W. (2018). Optimizing recombinant protein production in the *Escherichia coli* periplasm alleviates stress. *Applied and Environmental Microbiology*, *84*(12), e00270-18.

Berks, B. C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Molecular Microbiology*, *35*(2), 260–274.

Berlec, A., & Štrukelj, B. (2013). Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology*, *40*(3–4), 257–274.

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *ELife*, *4*, 1–12.

Borjaliloo S., Zomorodipour A.R., Yakhchali B., S. S. (2003). Comparison of T7- and Lac-based systems for the periplasmic expression of human granulocyte macrophage colony stimulating factor in *Escherichia coli*. *Iranian Journal of Biotechnology*, *1*(2), 101–108.

Bren, A., Park, J. O., Towbin, B. D., Dekel, E., Rabinowitz, J. D., & Alon, U. (2016). Glucose becomes one of the worst carbon sources for E. coli on poor nitrogen sources due to suboptimal levels of cAMP. *Scientific Reports*, *6*(1), 24834.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic screening of all signal peptides from *Bacillus subtilis*: A powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. *Journal of Molecular Biology*, *362*(3), 393–402.

Campo, N., Tjalsma, H., Buist, G., Stepniak, D., Meijer, M., Veenhuis, M., Westermann, M., Müller, J. P., Bron, S., Kok, J., Kuipers, O. P., & Jongbloed, J. D. H. (2004). Subcellular sites for bacterial protein export. *Molecular Microbiology*, *53*(6), 1583–1599.

Caspers, M., Brockmeier, U., Degering, C., Eggert, T., & Freudl, R. (2010). Improvement of Sec-dependent secretion of a heterologous model protein in *Bacillus subtilis* by saturation mutagenesis of the N-domain of the AmyE signal peptide. *Applied Microbiology and Biotechnology*, *86*(6), 1877–1885.

Chatzi, K. E., Sardis, M. F., Economou, A., & Karamanou, S. (2014). SecA-mediated targeting and translocation of secretory proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1843*(8), 1466–1474.

Chhetri, G., Kalita, P., & Tripathi, T. (2015). An efficient protocol to enhance recombinant protein expression using ethanol in *Escherichia coli. MethodsX*, *2*, 385–391.

Choi, J. H., & Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Applied Microbiology and Biotechnology*, *64*(5), 625–635.

Chung, H. J., Bang, W., & Drake, M. A. (2006). Stress response of *Escherichia coli*. *Comprehensive Reviews in Food Science and Food Safety*, *5*(3), 52–64.

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, *13*(6), 343–359.

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 634–643.

Dammeyer, T., & Tinnefeld, P. (2012). Engineered fluorescent proteins illuminate the bacterial periplasm. *Computational and Structural Biotechnology Journal*, *3*(October), e201210013.

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli. Microbial Cell Factories*, *8*(1), 26.

de Marco, A. (2013). Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. *Microbial Cell Factories*, *12*(1), 101.

Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., & Giuliano, G. (2013). A Chlamydomonas-derived Human Papillomavirus 16 E7 vaccine induces specific tumor protection. *PloS One*, *8*(4), e61473.

Dinh, T., & Bernhardt, T. G. (2011). Using superfolder green fluorescent protein for periplasmic protein localization studies. *Journal of Bacteriology*, *193*(18), 4984–4987.

Elander, R. P. (2003). Industrial production of β-lactam antibiotics. *Applied Microbiology and Biotechnology*, *61*(5–6), 385–392.

Ellis, M., Patel, P., Edon, M., Ramage, W., Dickinson, R., & Humphreys, D. P. (2017). Development of a high yielding *E. coli* periplasmic expression system for the production of humanized Fab' fragments. *Biotechnology Progress*, *33*(1), 212–220.

Fekkes, P., De Wit, J. G., Boorsma, A., Friesen, R. H. E., & Driessen, A. J. M. (1999). Zinc stabilizes the SecB binding site of SecA. *Biochemistry*, *38*(16), 5111–5116.

Feltcher, M. E., & Braunstein, M. (2012). Emerging themes in SecA2-mediated protein export. *Nature Reviews Microbiology*, *10*(11), 779–789.

Ferrante, P., Catalanotti, C., Bonente, G., & Giuliano, G. (2008). An optimized, chemically regulated gene expression system for Chlamydomonas. *PloS One*, *3*(9), e3200.

Freudl, R. (2013). Leaving home ain't easy: Protein export systems in Gram-positive bacteria. *Research in Microbiology*, *164*(6), 664–674.

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems.

Microbial Cell Factories, *17*(1), 1–10.

Gadgil, M., Kapur, V., & Hu, W. S. (2005). Transcriptional response of *Escherichia coli* to temperature shift. *Biotechnology Progress*, *21*(3), 689–699.

Galan, J., & Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, *444*(7119), 567–573.

Gao, D., Luan, Y., Liang, Q., & Qi, Q. (2016). Exploring the N-terminal role of a heterologous protein in secreting out of *Escherichia coli*. *Biotechnology and Bioengineering*, *113*(12), 2561–2567.

Gerdes, K., Christensen, S. K., & Løbner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. *Nature Reviews Microbiology*, *3*(5), 371–382.

Glazyrina, J., Materne, E.-M., Dreher, T., Storm, D., Junne, S., Adams, T., Greller, G., & Neubauer, P. (2010). High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor. *Microbial Cell Factories*, *9*(1), 42.

Gräslund, S., Nordlund, P., Weigelt, J., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., Dhe-Paganon, S., Park, H., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.-H., … Gunsalus, K. C. (2008). Protein production and purification. *Nature Methods*, *5*(2), 135–146.

Green, E. R., & Mecsas, J. (2015). Bacterial Secretion Systems: An Overview. In *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition* (Vol. 4, Issue 1, pp. 215–239). American Society of Microbiology.

Guzman, L.-M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P BAD promoter. *Journal of Bacteriology*, *177*(14), 4121–4130.

Harcum, S. W., & Bentley, W. E. (1999). Heat-shock and stringent responses have overlapping protease activity in *Escherichia coli*. Implications for heterologous protein yield. *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*, *80*(1), 23–37.

Harrison, J. S., Keshavarz-Moore, E., Dunnill, P., Berry, M. J., Fellinger, A., & Frenken, L. (1997). Factors affecting the fermentative production of a lysozyme-binding antibody fragment in *Escherichia coli*. *Biotechnology and Bioengineering*, *53*(6), 611–622.

Harrison, R. G., Todd, P. W., Rudge, S. R., & Petrides, D. P. (2015). *Bioseparations science and engineering* (2nd ed.). Oxford University Press.

Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *European Journal of Biochemistry*, *133*(1), 17–21.

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., & Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiology and Molecular Biology Reviews*, *68*(4), 692–744.

Hendrickx, A. P. A., Budzik, J. M., Oh, S. Y., & Schneewind, O. (2011). Architects at the bacterial surface-sortases and the assembly of pili with isopeptide bonds. *Nature Reviews Microbiology*, *9*(3), 166–176.

Hohlfeld, S., Pattis, I., Püls, J., Plano, G. V., Haas, R., & Fischer, W. (2006). A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the *Helicobacter pylori* CagA protein. *Molecular Microbiology*, *59*(5), 1624–1637.

Holland, I. B. (2004). Translocation of bacterial proteins - An overview. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1694*(1-3 SPEC.ISS.), 5–16.

Holland, I. B., Kenny, B., & Blight, M. (1990). Haemolysin secretion from *E. coli*. *Biochimie*, *72*(2–3), 131–141.

Holland, I. B., Schmitt, L., & Young, J. (2005). Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. *Molecular Membrane Biology*, *22*(1–2), 29–39.

Horn, U., Strittmatter, W., Krebber, A., Knüpfer, U., Kujau, M., Wenderoth, R., Müller, K., Matzku, S., Plückthun, A., & Riesenberg, D. (1996). High volumetric yields of functional dimeric miniantibodies in *Escherichia coli*, using an optimized expression vector and high-cell-density fermentation under non-limited growth conditions. *Applied Microbiology and Biotechnology*, *46*(5–6), 524–532.

Hoyt, D. W., & Gierasch, L. M. (1991). A peptide corresponding to an export-defective mutant OmpA signal sequence with asparagine in the hydrophobic core is unable to insert into model membranes. *Journal of Biological Chemistry*, *266*(22), 14406–14412.

Hsu, C. C., Thomas, O. R. T., & Overton, T. W. (2016). Periplasmic expression in and release of Fab fragments from *Escherichia coli* using stress minimization. *Journal of Chemical Technology and Biotechnology*, *91*(3), 815–822.

Hunter, S. E., Clarke, I. N., Kelly, D. C., & Titball, R. W. (1992). Cloning and nucleotide sequencing of the *Clostridium perfringens* epsilon-toxin gene and its expression in *Escherichia coli*. *Infect. Immun.*, *60*(1), 102–110.

Hynes, T. R., Hughes, T. E., & Berlot, C. H. (2004). Cellular localization of GFP-tagged α subunits. In *G Protein Signaling* (Vol. 237, pp. 233–246). Humana Press.

Jalalirad, R. (2013). Selective and efficient extraction of recombinant proteins from the periplasm of *Escherichia coli* using low concentrations of chemicals. *Journal of Industrial Microbiology and Biotechnology*, *40*(10), 1117–1129.

Jensen, E. B., & Carlsen, S. (1990). Production of recombinant human growth hormone in *Escherichia coli*: Expression of different precursors and physiological effects of glucose, acetate, and salts. *Biotechnology and Bioengineering*, *36*(1), 1–11.

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., Selbig, J., & Willmitzer, L. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Molecular Systems Biology*, *6*(364), 1–16.

Jungbauer, A. (2013). Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology*, *31*(8), 479–492.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, *338*(5), 1027–1036.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research*, *35*(SUPPL.2), 429–

432.

Karamyshev, A. L., Karamysheva, Z. N., Kajava, A. V., Ksenzenko, V. N., & Nesmeyanova, M. A. (1998). Processing of *Escherichia coli* alkaline phosphatase: Role of the primary structure of the signal peptide cleavage region. *Journal of Molecular Biology*, *277*(4), 859–870.

Khokhlova, O. V., & Nesmeyanova, M. A. (2004). Interdependent effects of the charge of the N-terminal region of the signal peptide, SecA, and SecB on secretion of alkaline phosphatase in *Escherichia coli*. *Molecular Biology*, *38*(2), 239–246.

Kim, J., & Kendall, D. A. (1998). Identification of a sequence motif that confers SecB dependence on a SecB-independent secretory protein *in vivo*. *Journal of Bacteriology*, *180*(6), 1396–1401.

Kim, J., & Kendall, D. A. (2000). Sec-dependent protein export and the involvement of the molecular chaperone SecB. *Cell Stress & Chaperones*, *5*(4), 267–275.

Kleiner-Grote, G. R. M., Risse, J. M., & Friehs, K. (2018). Secretion of recombinant proteins from *E. coli*. *Engineering in Life Sciences*, *18*(8), 532–550.

Knoblauch, N. T. M., Rüdiger, S., Schönfeld, H. J., Driessen, A. J. M., Schneider-Mergener, J., & Bukau, B. (1999). Substrate specificity of the SecB chaperone. *Journal of Biological Chemistry*, *274*(48), 34219–34225.

Korotkov, K. V., Sandkvist, M., & Hol, W. G. J. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, *10*(5), 336–351.

Krause, M., Neubauer, A., & Neubauer, P. (2016). The fed-batch principle for the molecular biology lab: Controlled nutrient diets in ready-made media improve production of recombinant proteins in *Escherichia coli*. *Microbial Cell Factories*, *15*(1), 1–13.

Kwon, S.-K., Kim, S. K., Lee, D.-H., & Kim, J. F. (2015). Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Scientific Reports*, *5*(October), 16076.

Liew, M. W. O., Rajendran, A., & Middelberg, A. P. J. (2010). Microbial production of virus-like particle vaccine protein at gram-per-litre levels. *Journal of Biotechnology*, *150*(2), 224–231.

Lightfoot, E. N., & Moscariello, J. S. (2004). Bioseparations. *Biotechnology and Bioengineering*, *87*(3), 259–273.

Losen, M., Frölich, B., Pohl, M., & Büchs, J. (2004). Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnology Progress*, *20*(4), 1062–1068.

Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology*, *97*(9), 3811–3826.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, *60*(3), 512–538.

Marschall, L., Sagmeister, P., & Herwig, C. (2017). Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. *Applied Microbiology and Biotechnology*, *101*(2), 501–512.

Martoglio, B., Graf, R., & Dobberstein, B. (1997). Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. *EMBO Journal*, *16*(22), 6636–6645.

Masi, M., & Wandersman, C. (2010). Multiple signals direct the assembly and function of a type 1 secretion system. *Journal of Bacteriology*, *192*(15), 3861–3869.

Matthews, K. S. (1992). DNA looping. *Microbiological Reviews*, *56*(1), 123–136.

Maurizi, M. R. (1992). Proteases and protein degradation in *Escherichia coli*. *Experientia*, *48*(2), 178–201.

Meerman, H. J., & Georgiou, G. (1994). Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted Rrcombinant proteins. *Bio/Technology*, *12*(11), 1107–1110.

Mergulhão, F. J. M., Summers, D. K., & Monteiro, G. A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, *23*(3), 177–202.

Meyer, P., & Dworkin, J. (2007). Applications of fluorescence microscopy to single bacterial cells. *Research in Microbiology*, *158*(3), 187–194.

Müller, M. (2005). Twin-arginine-specific protein export in *Escherichia coli*. *Research in Microbiology*, *156*(2), 131–136.

Mullineaux, C., Nenninger, A., & Ray, N. (2006). Diffusion of green fluorescent protein in three cell environments in <i>Escherichia coli<i>. *Journal of Bacteriology*, *188*(10), 3442–3448.

Natale, P., Brüser, T., & Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta - Biomembranes*, *1778*(9), 1735–1756.

Nfor, B. K., Ahamed, T., van Dedem, G. W. K., van der Wielen, L. A. M., van de Sandt, E. J. A. X., Eppink, M. H. M., & Ottens, M. (2008). Design strategies for integrated protein purification processes: Challenges, progress and outlook. *Journal of Chemical Technology and Biotechnology*, *83*(2), 124–132.

Nivaskumar, M., & Francetic, O. (2014). Type II secretion system: A magic beanstalk or a protein escalator. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1843*(8), 1568–1577.

Orfanoudaki, G., & Economou, A. (2014). Proteome-wide subcellular topologies of *E. coli* polypeptides database (STEPdb). *Molecular and Cellular Proteomics*, *13*(12), 3674–3687.

Osborne, A. R., Rapoport, T. A., & van den Berg, B. (2005). Protein Translocation by the Sec61/SecY Channel. *Annual Review of Cell and Developmental Biology*, *21*(1), 529–550.

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590–601.

Paetzel, M., Karla, A., Strynadka, N. C. J., & Dalbey, R. E. (2002). Signal peptidases. *Chemical Reviews*, *102*(12), 4549–4579.

Palmer, T., & Berks, B. C. (2012). The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology*, *10*(7), 483–496.

Panula-Perälä, J., Šiurkus, J., Vasala, A., Wilmanowski, R., Casteleijn, M. G., & Neubauer, P. (2008). Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks. *Microbial Cell Factories*, *7*(1), 31.

Papanikou, E., Karamanou, S., & Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol*, *5*(11), 839–851.

Pechsrichuang, P., Songsiriritthigul, C., Haltrich, D., Roytrakul, S., Namvijtr, P., Bonaparte, N., & Yamabhai, M. (2016). OmpA signal peptide leads to heterogenous secretion of *B. subtilis* chitosanase enzyme from *E. coli* expression system. *SpringerPlus*, *5*(1), 1200.

Pelczar, M. J., Gunsalus, I. C., & Stanier, R. Y. (1960). The bacteria, a treatise on structure and function. Volume I: Structure. *AIBS Bulletin*, *10*(6), 37.

Ramchuran, S. O., Nordberg Karlsson, E., Velut, S., De Maré, L., Hagander, P., & Holst, O. (2003). Production of heterologous thermostable glycoside hydrolases and the presence of host-cell proteases in substrate limited fed-batch cultures of *Escherichia coli* BL21(DE3). *Applied Microbiology and Biotechnology*, *60*(4), 408–416.

Randall, L. L. (1992). Peptide binding by chaperone SecB: Implications for recognition of nonnative structure. *Science*, *257*(5067), 241–245.

Riemann, T., & Schröder, F. (2005). More effective prevention of incompatibility reactions through the use of four lumen central venous catheters in critically ill patients. *PflegenIntensiv*, *2*(1), 57.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, *5*(APR), 1–17.

Samudrala, R., Heffron, F., & McDermott, J. E. (2009). Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for Type III secretion systems. *PLoS Pathogens*, *5*(4), e1000375.

Sánchez, B., Chaignepain, S., Schmitter, J.-M., & Urdaci, M. C. (2009). A method for the identification of proteins secreted by lactic acid bacteria grown in complex media. *FEMS Microbiology Letterscrobiology Letters*, *295*(2), 226–229.

Sandkvist, M. (2001). Biology of type II secretion. *Molecular Microbiology*, *40*(2), 271–283.

Schechter, L. M., Roberts, K. A., Jamir, Y., Alfano, J. R., & Collmer, A. (2004). *Pseudomonas syringae* Type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *Journal of Bacteriology*, *186*(2), 543–555.

Schlegel, S., Genevaux, P., & de Gier, J. W. (2015). De-convoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. *Cell Reports*, *10*(10), 1758–1766.

Schlegel, S., Rujas, E., Ytterberg, A. J., Zubarev, R. A., Luirink, J., & de Gier, J.-W. (2013). Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microbial Cell Factories*, *12*(1), 24.

Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiology Reviews*, *34*(5), 779–796.

Schneewind, O., & Missiakas, D. (2014). Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1843*(8), 1687–1697.

Schweder, T., Lin, H., Jürgen, B., Breitenstein, A., Riemschneider, S., Khalameyzer, V., Gupta, A., Büttner, K., & Neubauer, P. (2002). Role of the general stress response during strong overexpression of a heterologous gene in *Escherichia coli*. *Applied Microbiology and Biotechnology*, *58*(3), 330–337.

Seibel, N. M., Eljouni, J., Nalaskowski, M. M., & Hampe, W. (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Analytical Biochemistry*, *368*(1), 95–99.

Shaner, N. C., Steinbach, P. A., & Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nature Methods*, *2*(12), 905–909.

Shiloach, J., & Fass, R. (2005). Growing *E. coli* to high cell density - A historical perspective on method development. *Biotechnology Advances*, *23*(5), 345–357.

Shokri, A., Sandén, A., & Larsson, G. (2003). Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. *Applied Microbiology and Biotechnology*, *60*(6), 654–664.

Shokri, A., Sanden, A. M., & Larsson, G. (2002). Growth rate-dependent changes in Escherichia coli membrane structure and protein leakage. *Applied Microbiology and Biotechnology*, *58*(3), 386–392.

Simmons, L. C., & Yansura, D. G. (1996). Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nature Biotechnology*, *14*(5), 629–634.

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41.

Smith, V. F., Hardy, S. J. S., & Randall, L. L. (1997). Determination of the binding frame of the chaperone SecB within the physiological ligand oligopeptide-binding protein. *Protein Science*, *6*(8), 1746–1755.

Spence, E., Sarcina, M., Ray, N., Møller, S. G., Mullineaux, C. W., & Robinson, C. (2003). Membrane-specific targeting of green fluorescent protein by the Tat pathway in the cyanobacterium *Synechocystis* PCC6803. *Molecular Microbiology*, *48*(6), 1481–1489.

Steen, H., & Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*, *5*(9), 699–711.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130.

Tang, J., Yang, H., Song, S., Zhu, P., & Ji, A. (2008). Effect of Glycine and Triton X-100 on secretion and expression of ZZ-EGFP fusion protein. *Food Chemistry*, *108*(2), 657–662.

Tsirigotaki, A., De Geyter, J., Šoštarić, N., Economou, A., & Karamanou, S. (2017). Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology*, *15*(1), 21–36.

Van Elsas, J. D., Semenov, A. V., Costa, R., & Trevors, J. T. (2011). Survival of *Escherichia coli* in the

environment: Fundamental and public health aspects. *ISME Journal*, *5*(2), 173–183.

Van Ulsen, P., Rahman, S. ur, Jong, W. S. P., Daleke-Schermerhorn, M. H., & Luirink, J. (2014). Type V secretion: From biogenesis to biotechnology. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1843*(8), 1592–1611.

Waegeman, H., De Lausnay, S., Beauprez, J., Maertens, J., De Mey, M., & Soetaert, W. (2013). Increasing recombinant protein production in *Escherichia coli* K12 through metabolic engineering. *New Biotechnology*, *30*(2), 255–261.

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O., & de Gier, J.-W. (2008). Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14371–14376.

Wilken, L. R., & Nikolov, Z. L. (2012). Recovery and purification of plant-made recombinant proteins. *Biotechnology Advances*, *30*(2), 419–433.

Wingfield, P. T. (2015). Overview of the Purification of Recombinant Proteins. *Current Protocols in Protein Science*, *158*(October 2013), 6.1.1-6.1.35.

Wong, K. R., & Buckley, J. T. (1993). *Aeromonas* spp. can secrete *Escherichia coli* alkaline phosphatase into the culture supernatant, and its release requires a functional general secretion pathway. *Molecular Microbiology*, *9*(5), 955–963.

Wu, F., Ma, J., Cha, Y., Lu, D., Li, Z., Zhuo, M., Luo, X., Li, S., & Zhu, M. (2020). Using inexpensive substrate to achieve high-level lipase A secretion by Bacillus subtilis through signal peptide and promoter screening. *Process Biochemistry*, *99*(August), 202–210.

Wurm, D. J., Quehenberger, J., Mildner, J., Eggenreich, B., Slouka, C., Schwaighofer, A., Wieland, K., Lendl, B., Rajamanickam, V., Herwig, C., & Spadiut, O. (2018). Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in *E. coli*. *Applied Microbiology and Biotechnology*, *102*(2), 667–676.

Wurm, D. J., Slouka, C., Bosilj, T., Herwig, C., & Spadiut, O. (2017). How to trigger periplasmic release in recombinant *Escherichia coli*: A comparative analysis. *Engineering in Life Sciences*, *17*(2), 215–222.

Wurm, D. J., Veiter, L., Ulonska, S., Eggenreich, B., Herwig, C., & Spadiut, O. (2016). The E. coli pET expression system revisited—mechanistic correlation between glucose and lactose uptake. *Applied Microbiology and Biotechnology*, *100*(20), 8721–8729.

Yen, M. R., Peabody, C. R., Partovi, S. M., Zhai, Y., Tseng, Y. H., & Saier, M. H. (2002). Protein-translocating outer membrane porins of Gram-negative bacteria. *Biochimica et Biophysica Acta - Biomembranes*, *1562*(1–2), 6–31.

Yoon, S. H., Kim, S. K., & Kim, J. F. (2010). Secretory production of recombinant proteins in *Escherichia coli*. *Recent Patents on Biotechnology*, *4*(1), 23–29.

Ytterberg, A. J., Zubarev, R. A., & Baumgarten, T. (2019). Posttranslational targeting of a recombinant protein promotes its efficient secretion into the Escherichia coli periplasm. *Applied and Environmental Microbiology*, *85*(13), e00671-19.

Zhang, G., Annan, R. S., Carr, S. A., & Neubert, T. A. (2010). Overview of peptide and protein analysis by

mass spectrometry. *Current Protocols in Protein Science*, *SUPPL.62*, 1–30.

Zhang, W., Yang, M., Yang, Y., Zhan, J., Zhou, Y., & Zhao, X. (2016). Optimal secretion of alkali-tolerant xylanase in *Bacillus subtilis* by signal peptide screening. *Applied Microbiology and Biotechnology*, *100*(20), 8745–8756.

Zydney, A. L. (2016). Continuous downstream processing for high value biological products: A Review. *Biotechnology and Bioengineering*, *113*(3), 465–475.

Appendix

6.1. Additional OD_{600} readings of *E. coli* cultures

The growth of *E. coli* BL21(DE3) in M2P growth-medium was measure is shown below for various temperatures and IPTG concentrations. The pET plasmids contained no gene insert ('Empty') and an ETX gene in the multiple cloning site.

Only minor differences in OD_{600} measurements were observed between samples indicating a low impact of the gene insert on overall growth rate of the cell culture.

6.1.1. pET26-ETX contained in *E. coli* BL21(DE3) at various IPTG concentrations and temperatures grown in M2P

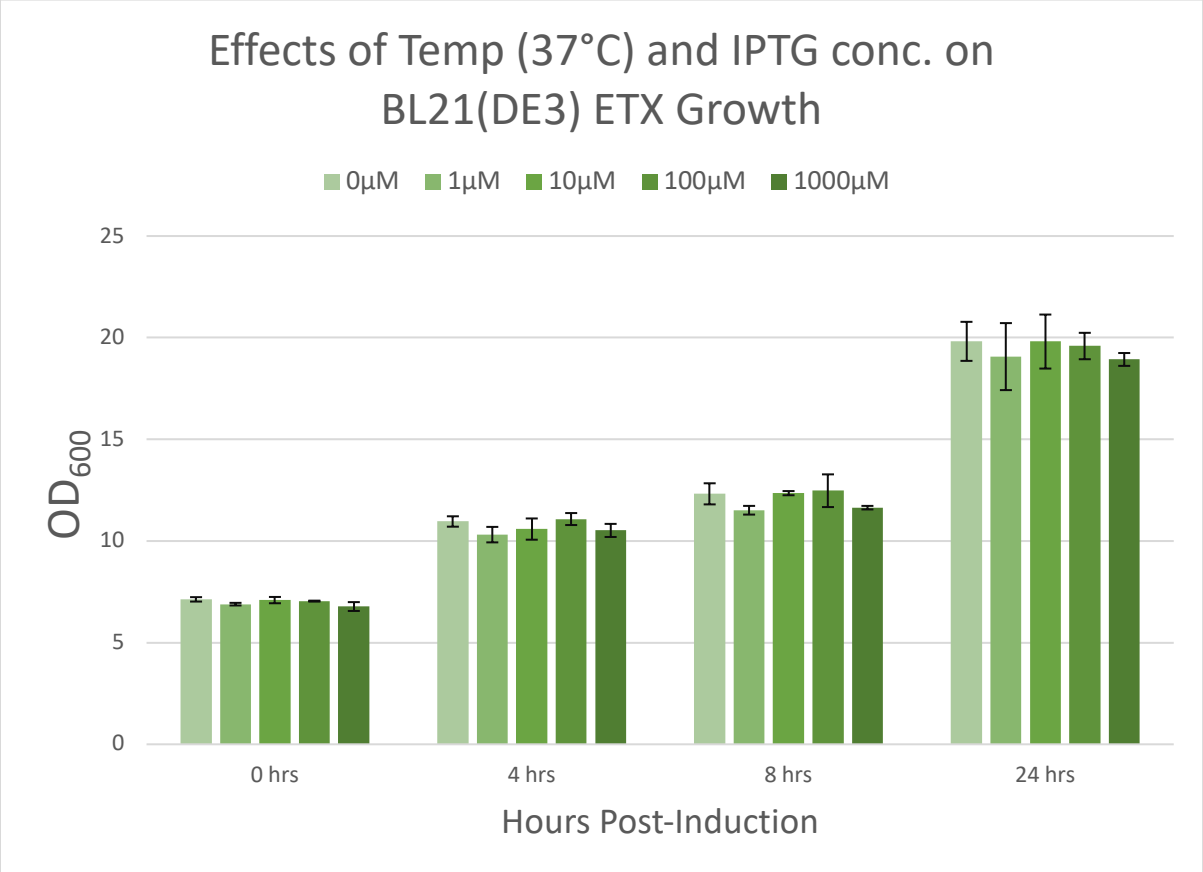


Figure 6.1: Post-induction OD_{600} measurements of E. coli BL21(DE3) expressing pET28-ETX at various IPTG concentrations, at 37°C, in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

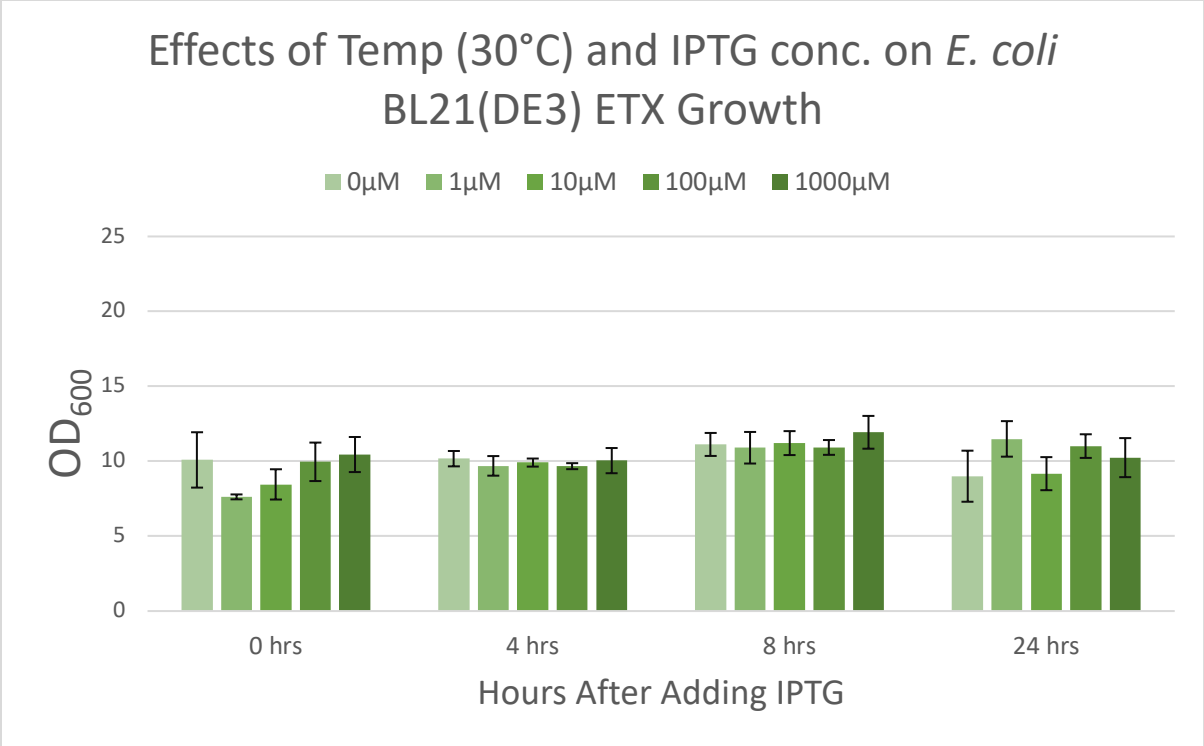


Figure 6.2: Post-induction OD_{600} measurements of BL21(DE3) expressing pET28-ETX at various IPTG concentrations, at 30°C, in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

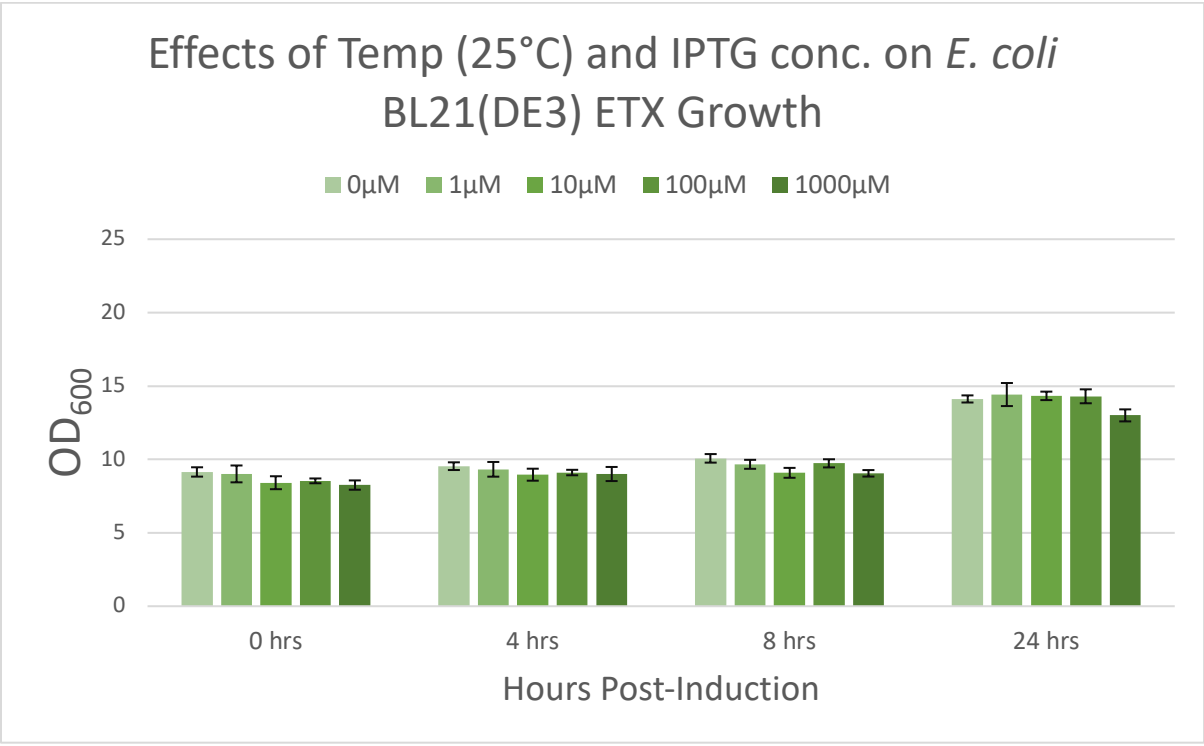


Figure 6.3: Post-induction OD_{600} measurements of E. coli BL21(DE3) expressing pET28-ETX at various IPTG concentrations, at 25°C, in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

6.1.2. pET26-Empty contained in *E. coli* BL21(DE3) at various IPTG concentrations and temperatures grown in M2P

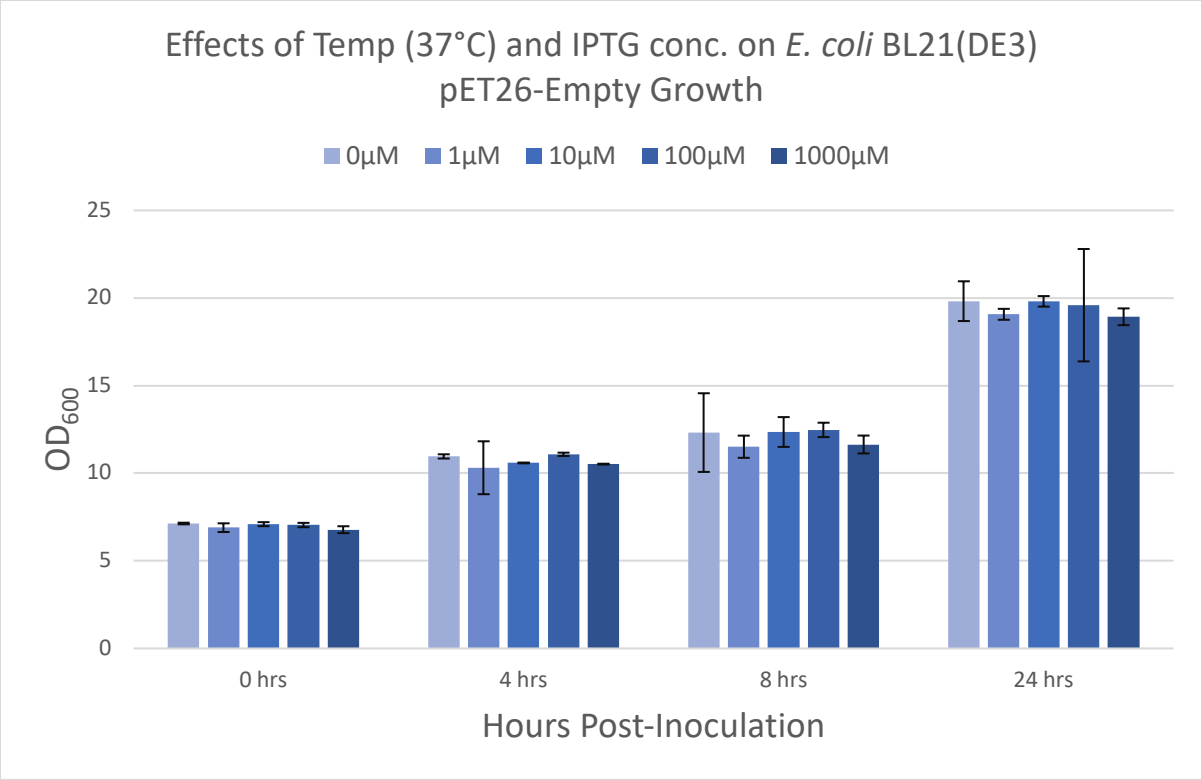


Figure 6.4: Post-induction OD_{600} measurements of E. coli BL21(DE3) expressing pET26 with no gene inserted into the multiple cloning site (Empty), at various IPTG concentrations, at 37°C, grown in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

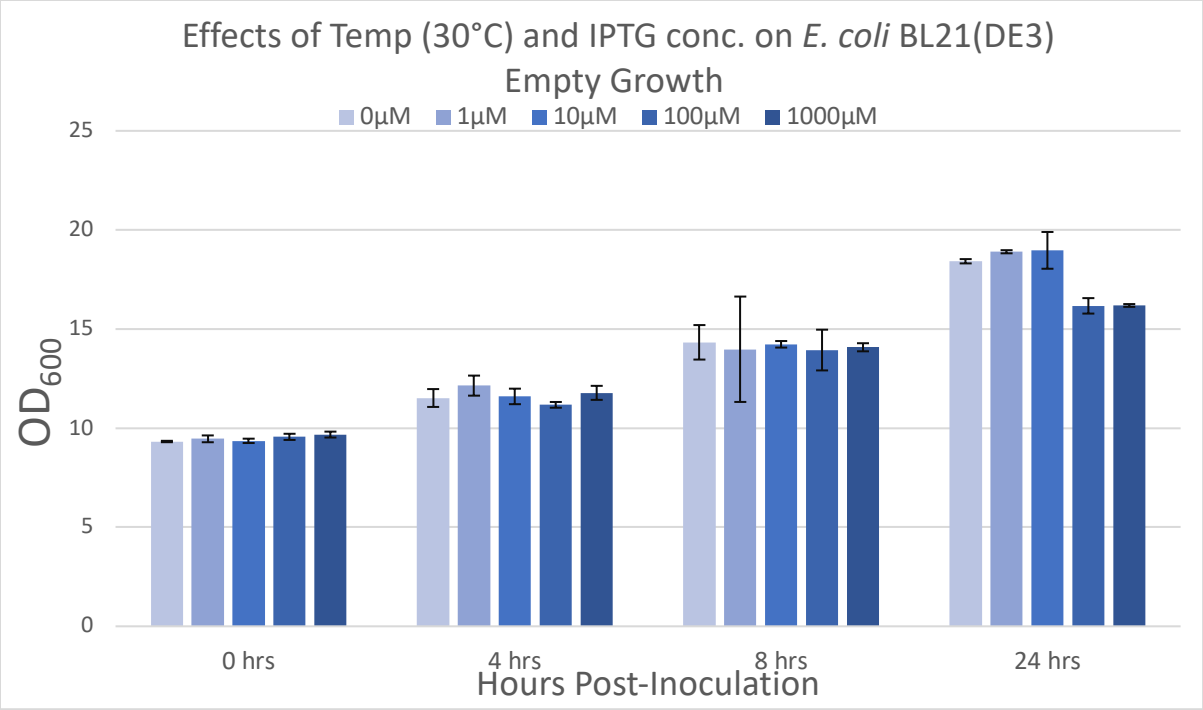


Figure 6.5: Post-induction OD_{600} measurements of E. coli BL21(DE3) expressing pET26 with no gene inserted into the multiple cloning site (Empty), at various IPTG concentrations, at 30°C, in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

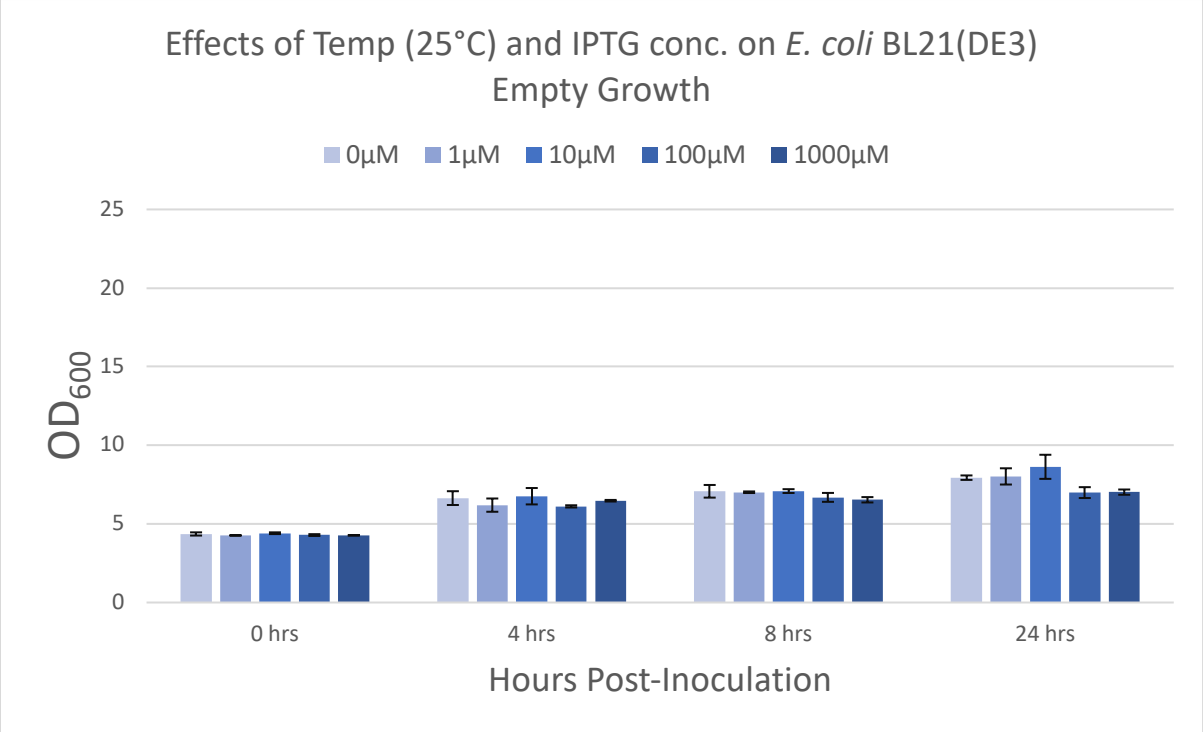


Figure 6.6: Post-induction OD_{600} measurements of BL21(DE3) expressing pET26 with no gene inserted into the Multiple Cloning Site (Empty), at various IPTG concentrations, at 25°C, in M2P growth medium. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

6.1.3. Pre-IPTG induction of *E. coli* cultures grown on LB

Comparison of the OD_{600} readings of each *E. coli* culture (grown in LB), indicated that before IPTG induction the different plasmids constructs did not play a significant role in *E. coli* growth. Lower culture temperatures reduced this effect.

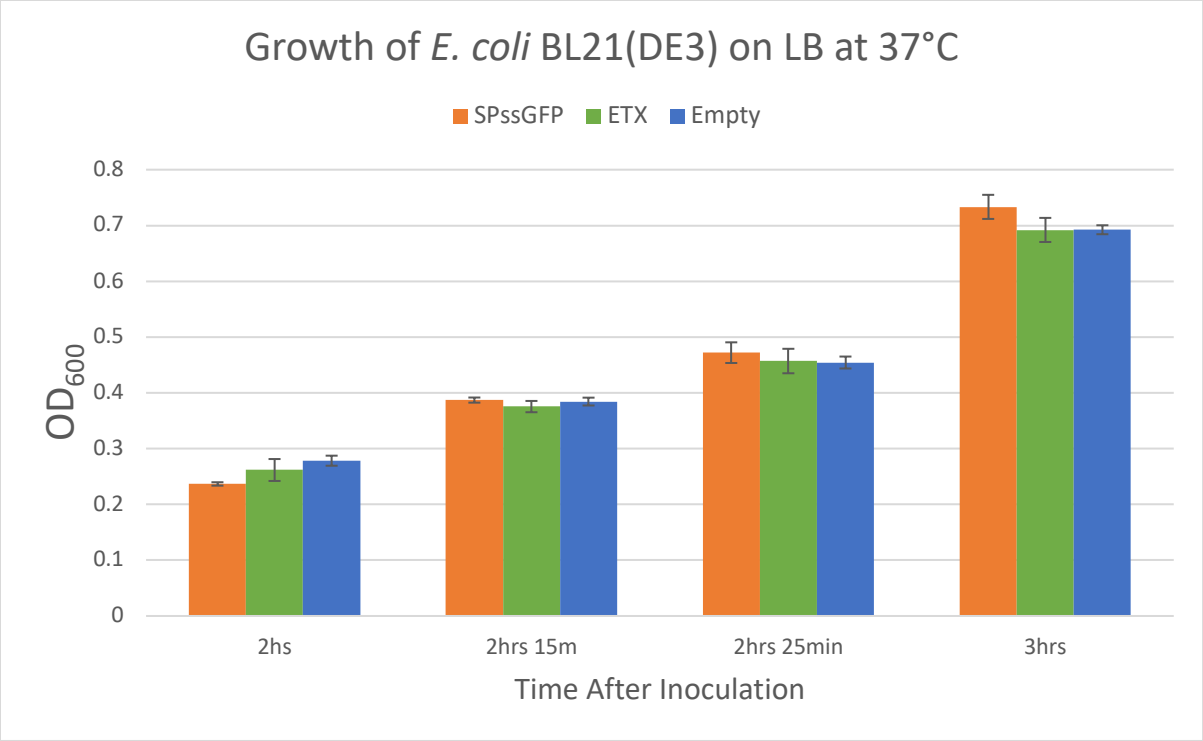


Figure 6.7: OD_{600} readings of selected samples of E. coli BL21(DE3) cultures containing plasmids pET26-SPssGFP, -ETX, and -Empty. Inoculation was done with 1% culture volume (i.e. 300uL pre-culture was added to 30mL LB). Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

6.1.4. Post-IPTG induction of *E. coli* BL21(DE3) cultures grown on LB

The OD_{600} readings of *E. coli* cultures after IPTG induction are shown below. The addition of IPTG (above 100uM) reduced the maximum cell mass compared to the control.

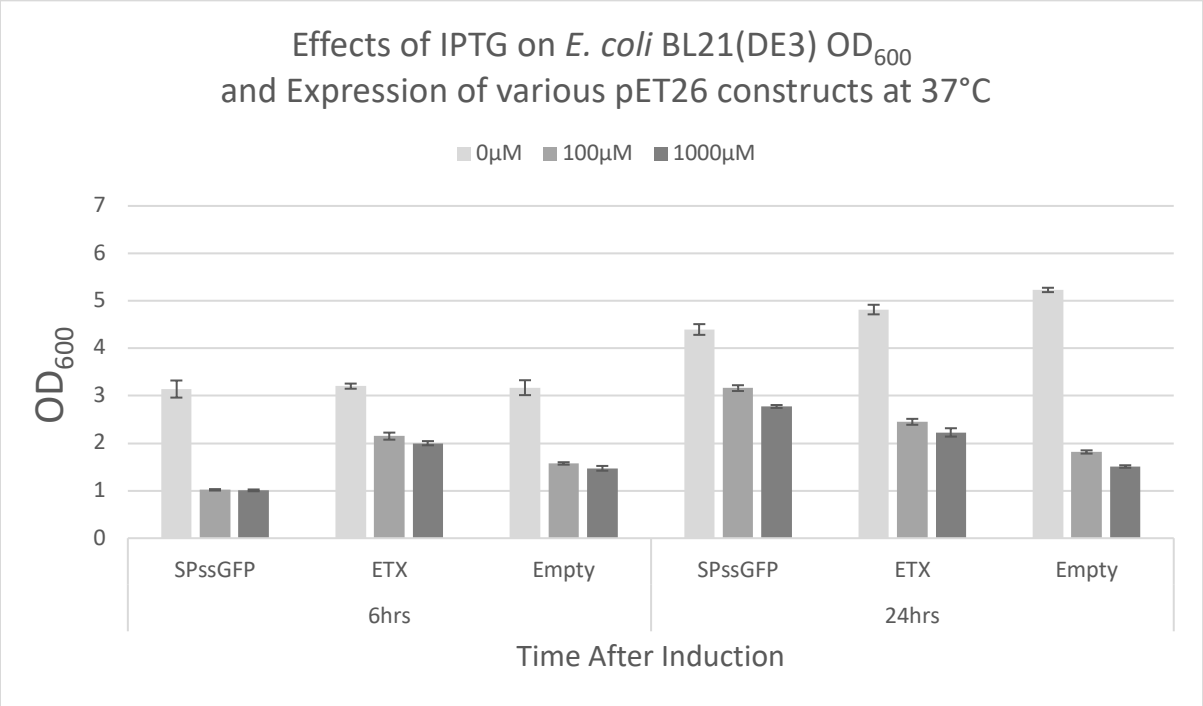


Figure 6.8: OD_{600} measurements of E. coli BL21(DE3) expressing pET26 constructs SPss-GFP, ETX and Empty. Grouped by sample time, 6 hours and 24 hours after induction with IPTG. Samples grown at 37°C in LB. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

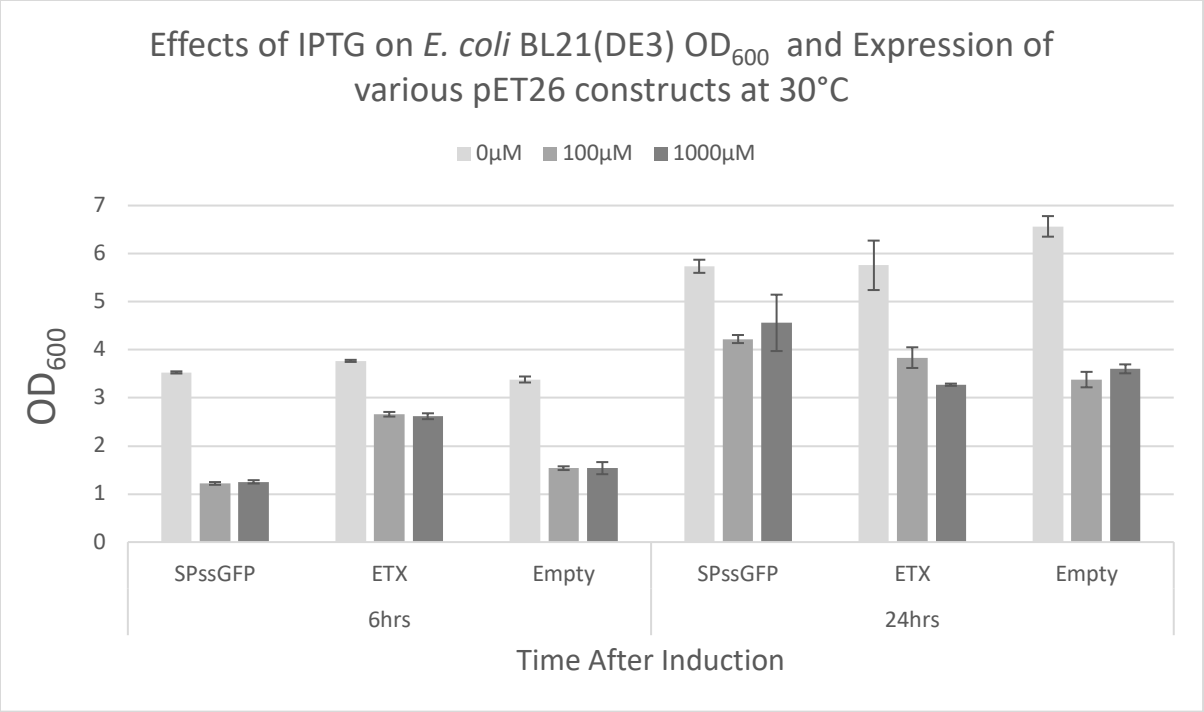


Figure 6.9: OD_{600} measurements of E. coli BL21(DE3) expressing pET26 constructs SPssGFP, ETX and Empty. Grouped by sample time, 6 hours and 24 hours after induction with IPTG. Samples grown at 30°C in LB. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

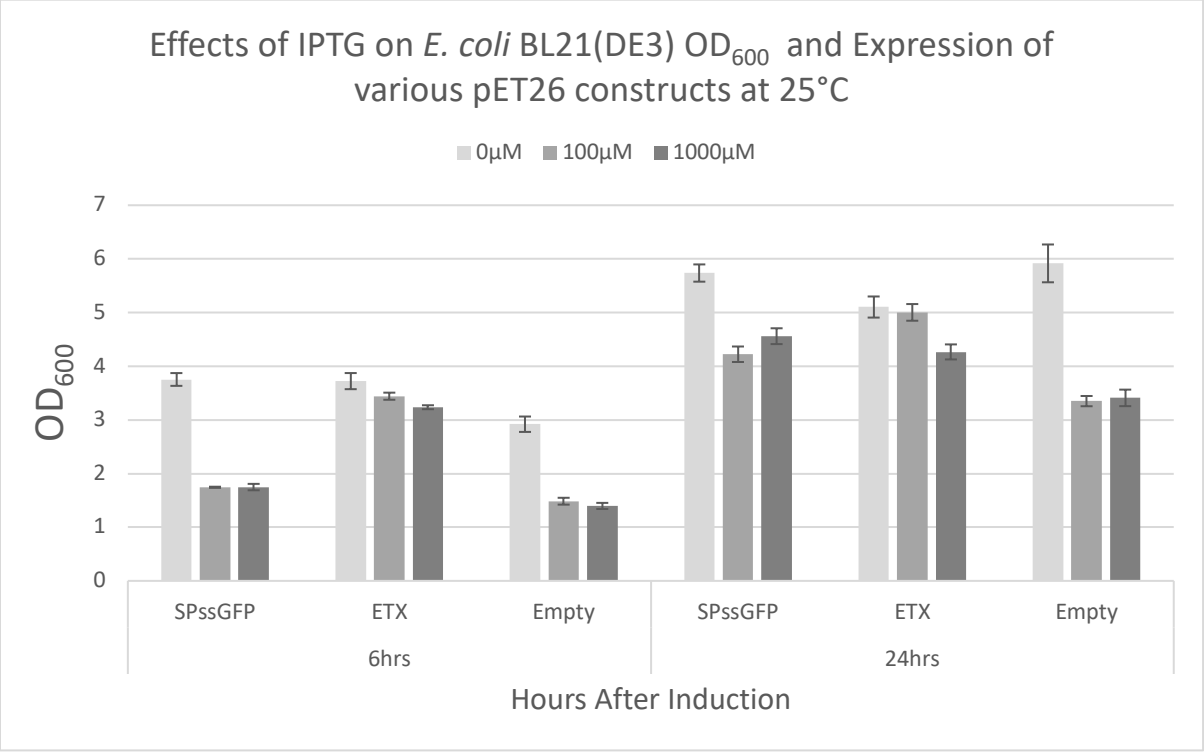


Figure 6.10: OD_{600} measurements of E. coli BL21(DE3) expressing pET26 constructs SPss-GFP, ETX and Empty. Grouped by sample time, 6 hours and 24 hours after induction with IPTG. Samples grown at 25°C in LB. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

6.2. Additional total protein concentration measurements of *E. coli* cultures

6.2.1. pET28-ETX plasmid contained in *E. coli* BL21(DE3) protein concentration at various temperatures and IPTG concentrations

Total protein measurements indicated that expression of ETX did not lead to a significant change in overall protein concentration of the cell.

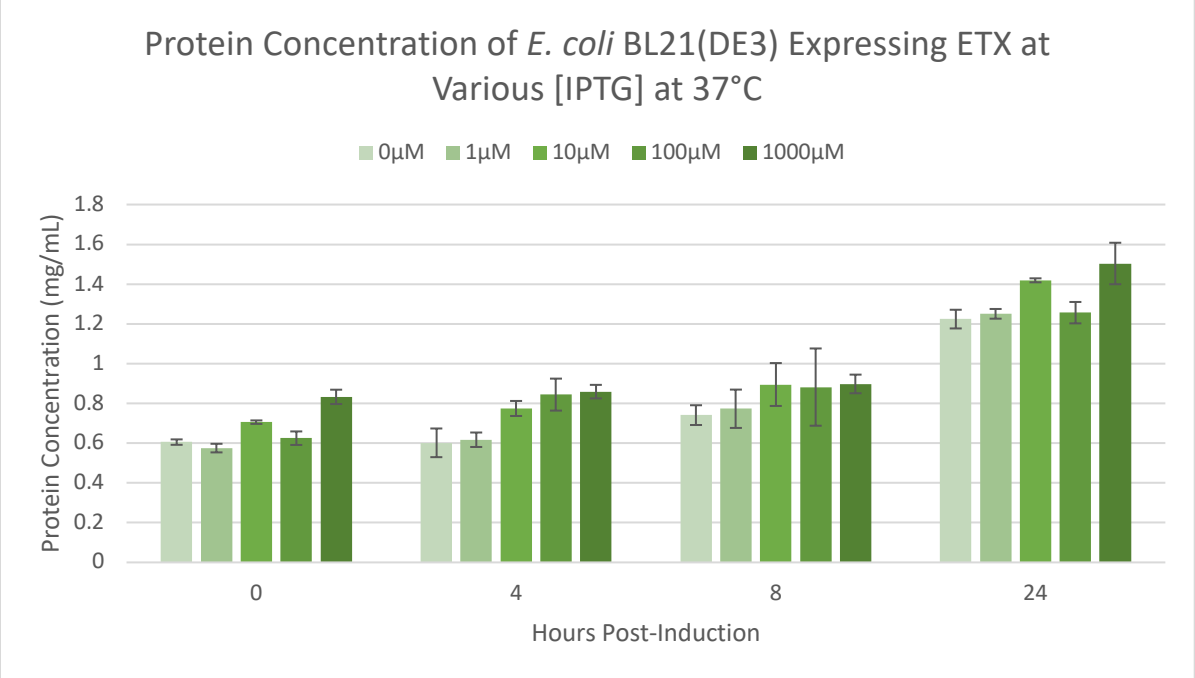


Figure 6.11: Protein concentration (mg/mL) of BL21(DE3) total protein extract, expressing pET26-ETX, at various time points (0, 4, 8 and 24 hours) after protein induction with IPTG and grown in M2P at 37°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

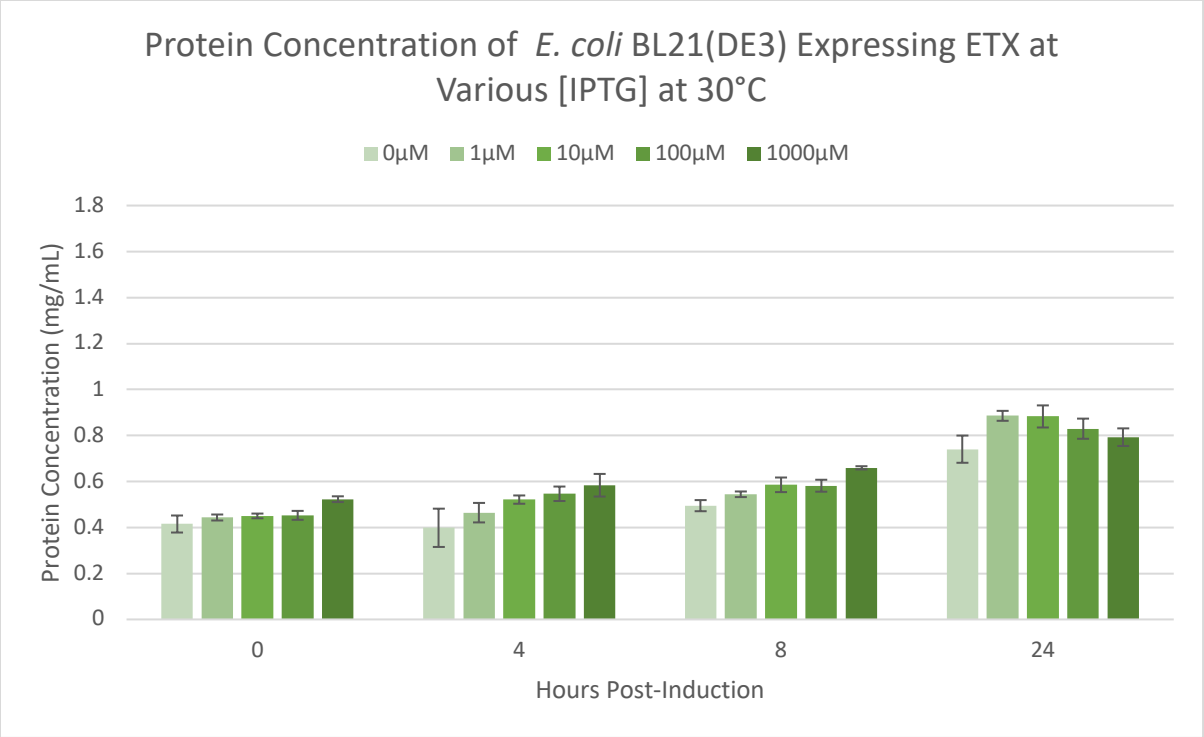


Figure 6.12: Protein Concentration (mg/mL) of BL21(DE3) total protein extract, expressing pET26-ETX, at various time points (0, 4, 8 and 24 hours) after protein induction with IPTG and grown in M2P at 30°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

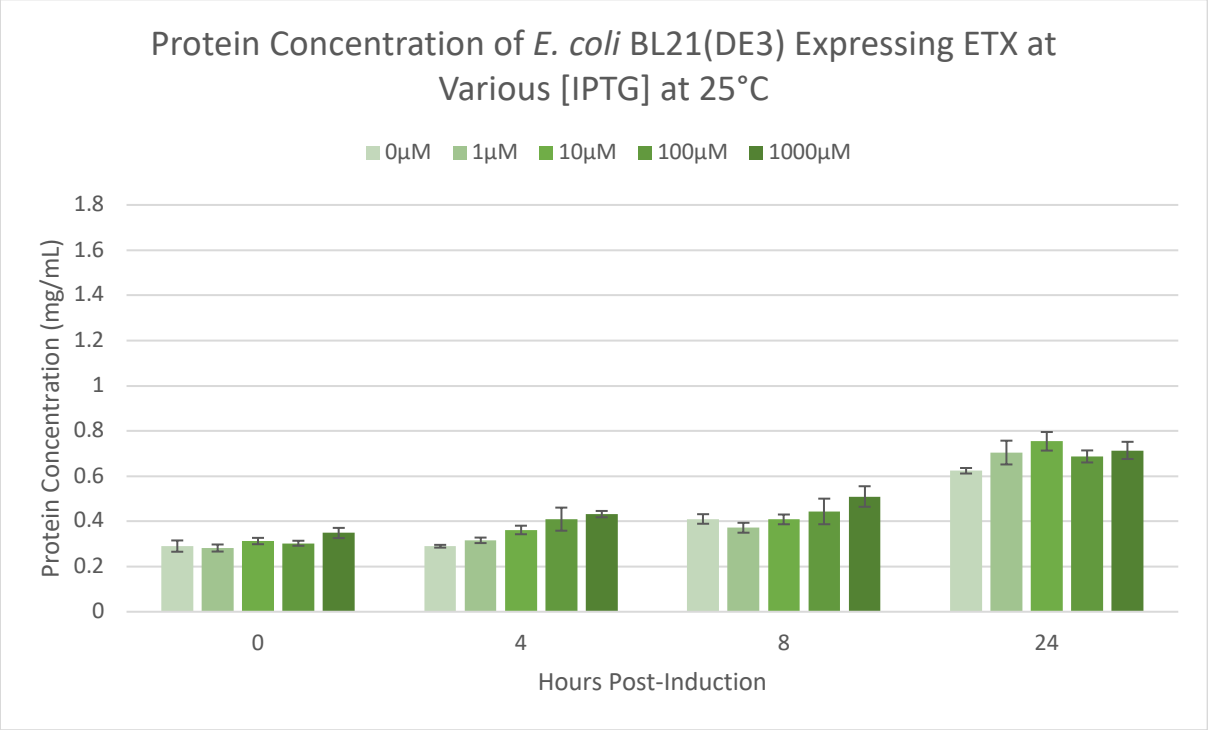


Figure 6.13: Protein Concentration (mg/mL) of BL21(DE3) total protein extract, expressing pET26-ETX, at various time points (0, 4, 8 and 24 hours) after protein induction with IPTG and grown in M2P at 25°C. Error bars represent one standard deviation from the mean of one sample from each test condition, measured in triplicate.

6.3. Additional PAGE gels of total and fractionated proteins of *E. coli* cultures grown in LB

6.3.1. Cytoplasmic fractions: 25°C IPTG induction temperature

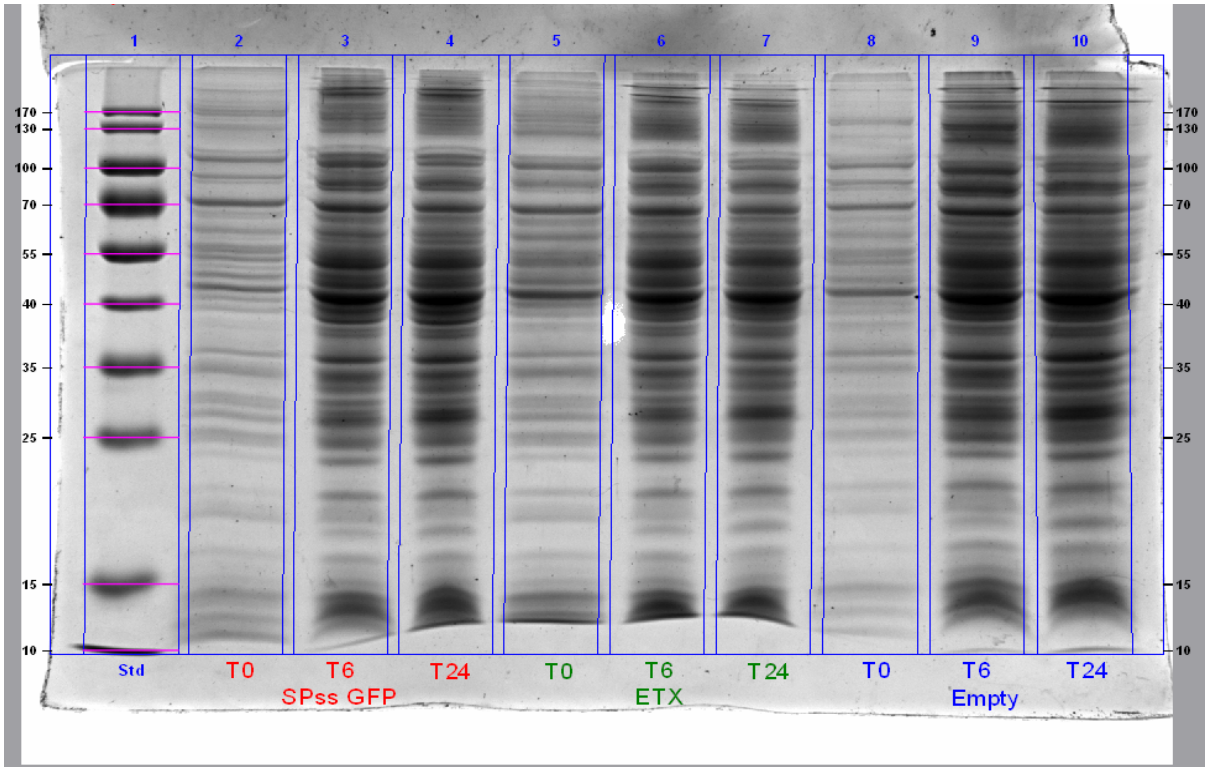


Figure 6.14: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µM IPTG, at 25°C, and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

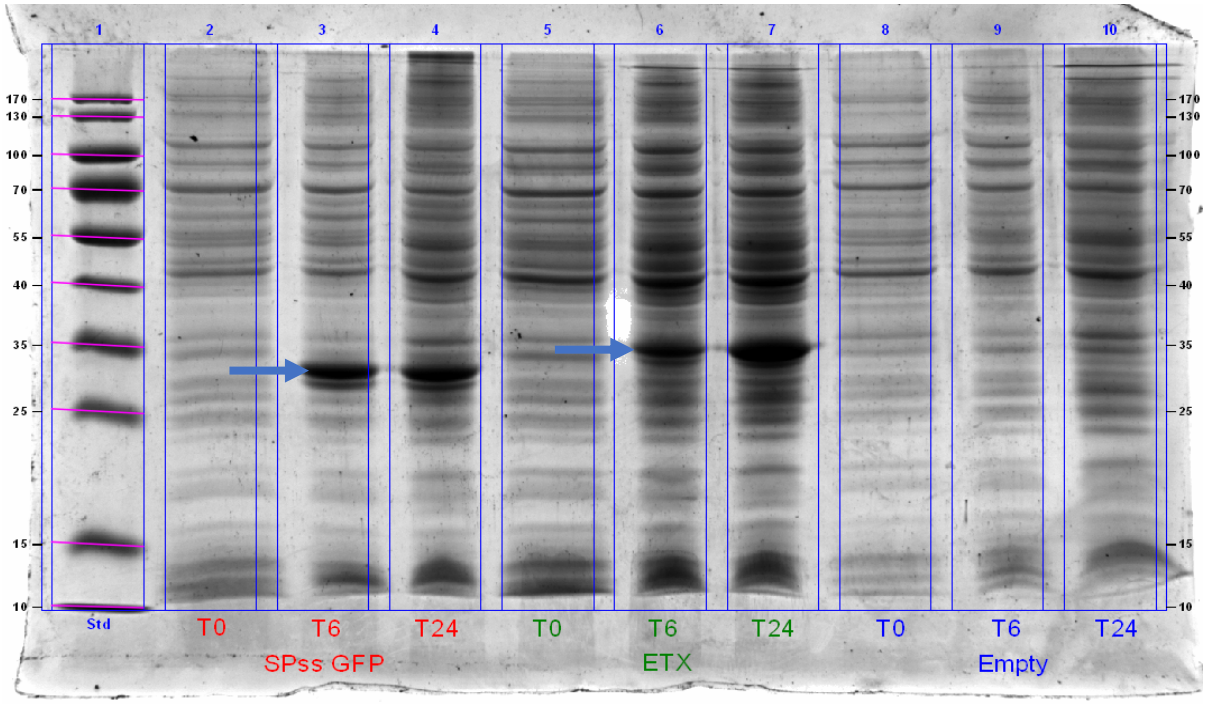


Figure 6.15: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µM IPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrows indicate SPssGFP and ETX protein bands.

6.3.2. Periplasmic fraction: 25°C IPTG induction temperature

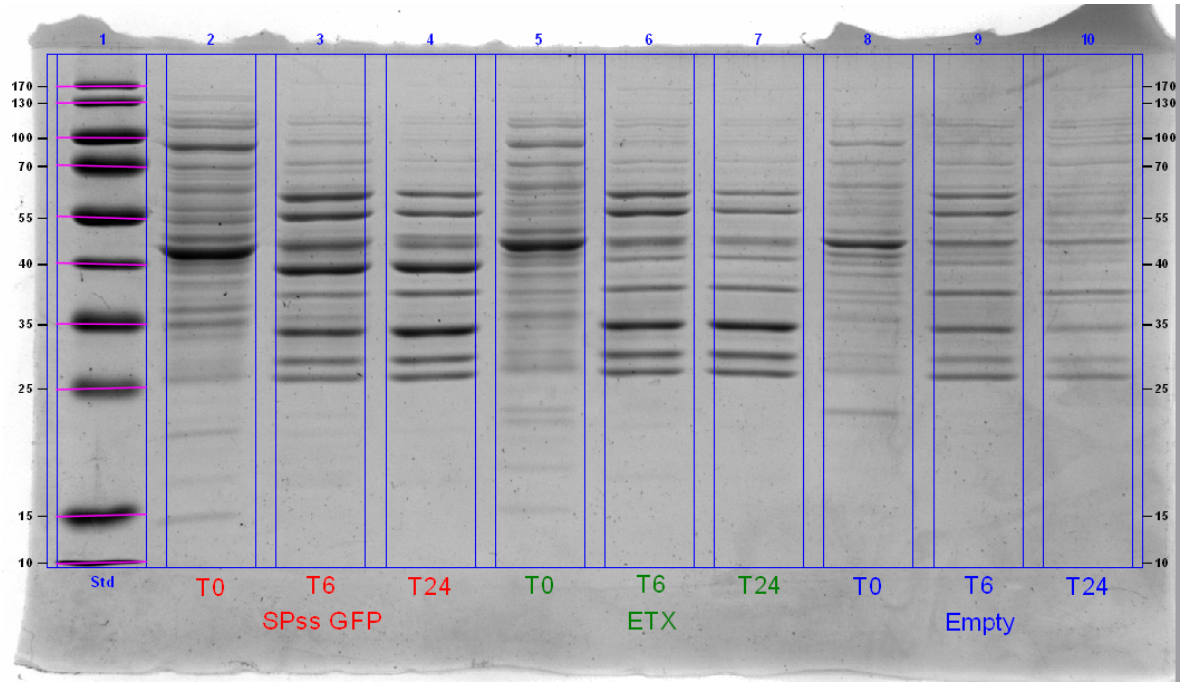


Figure 6.16: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µM IPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

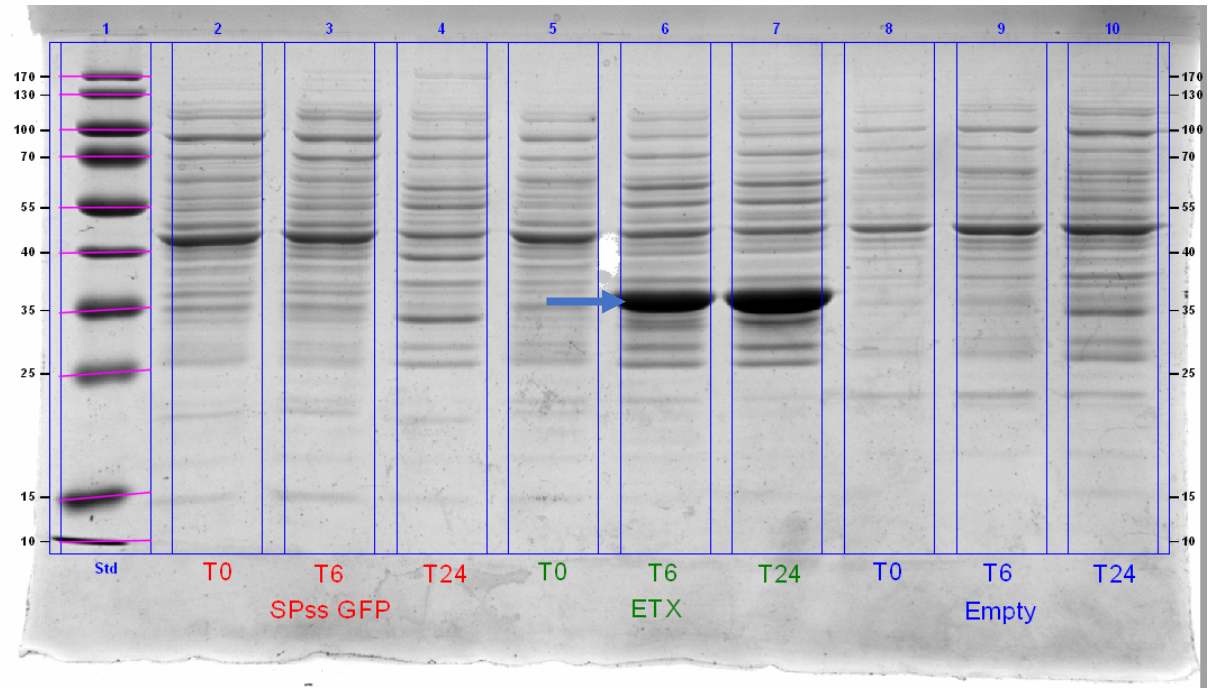


Figure 6.17: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µMIPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band.

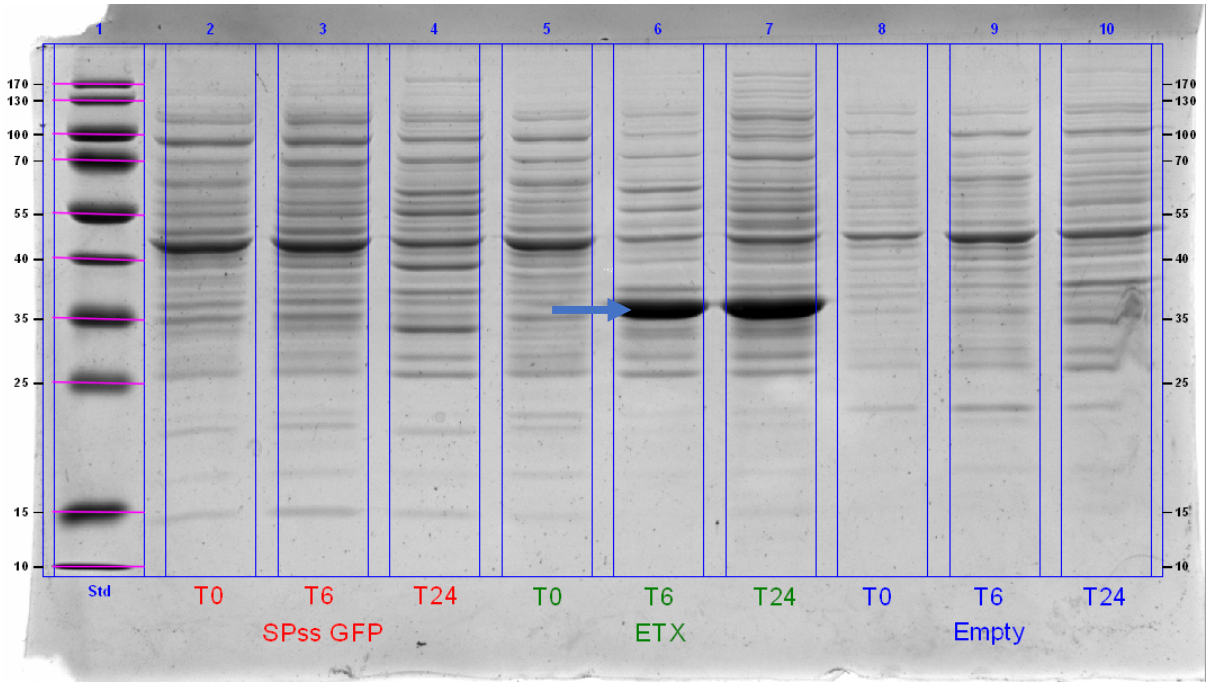


Figure 6.18: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µMIPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band. Feint bands in SPssGFP (T24) sample may be SPssGFP protein but are very similar to bands observed in the Empty control making a conclusive identification difficult.

6.3.3. Extracellular fraction: 25°C IPTG induction temperature

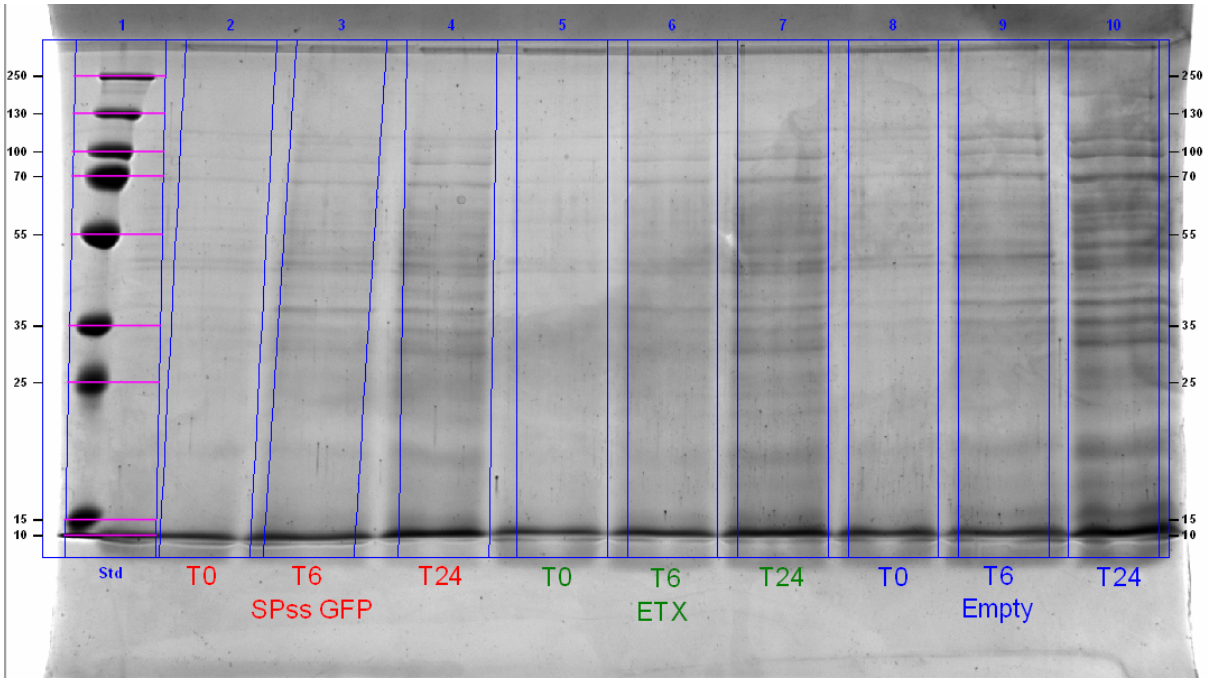


Figure 6.19: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µMIPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

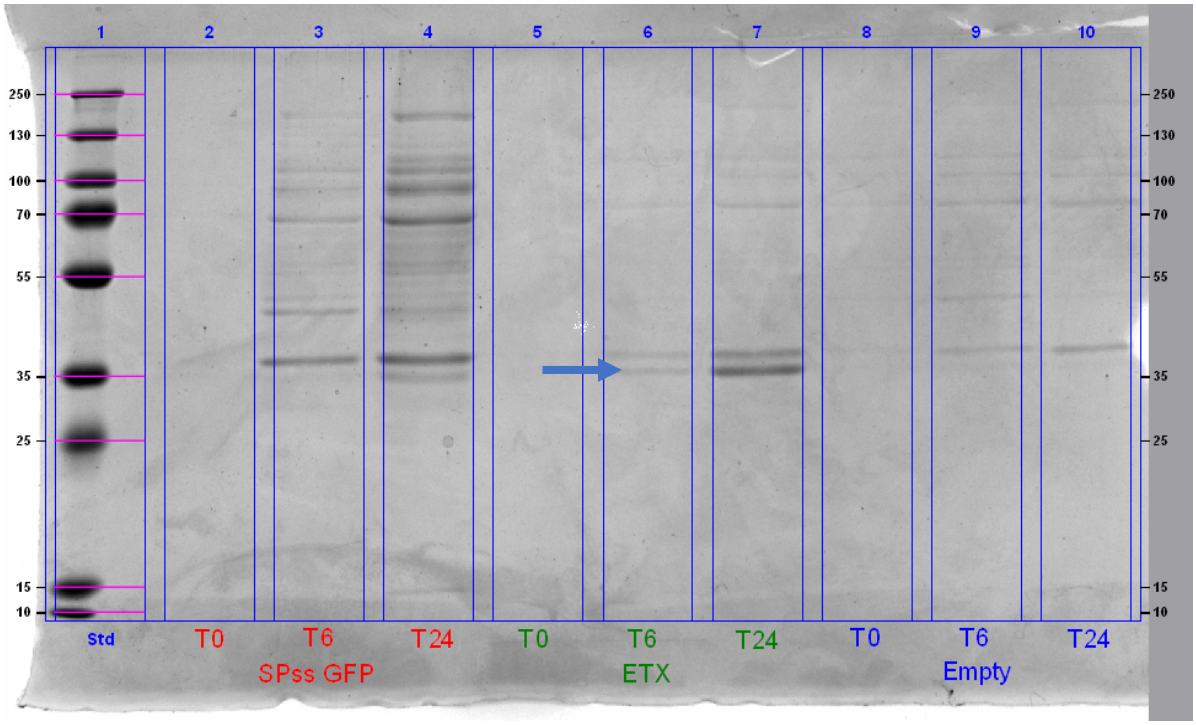


Figure 6.20: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000μM IPTG, at 25°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band.

6.3.4. Cytoplasmic fraction: 30°C IPTG induction temperature

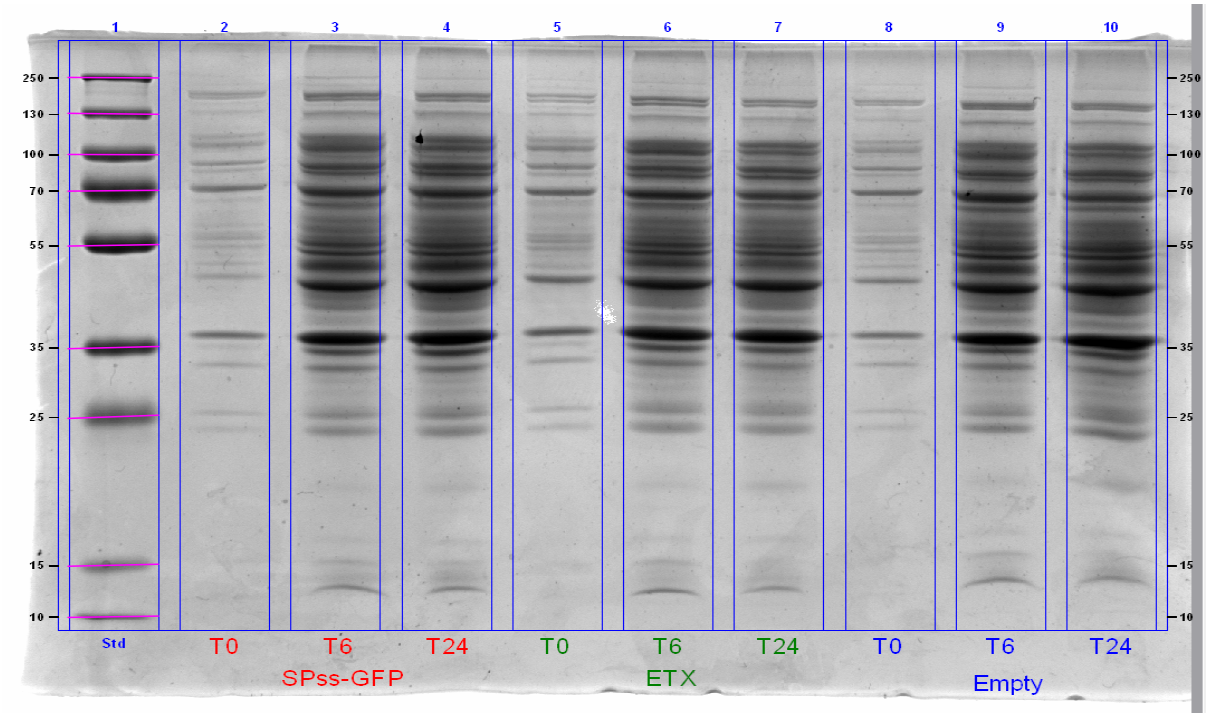


Figure 6.21: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

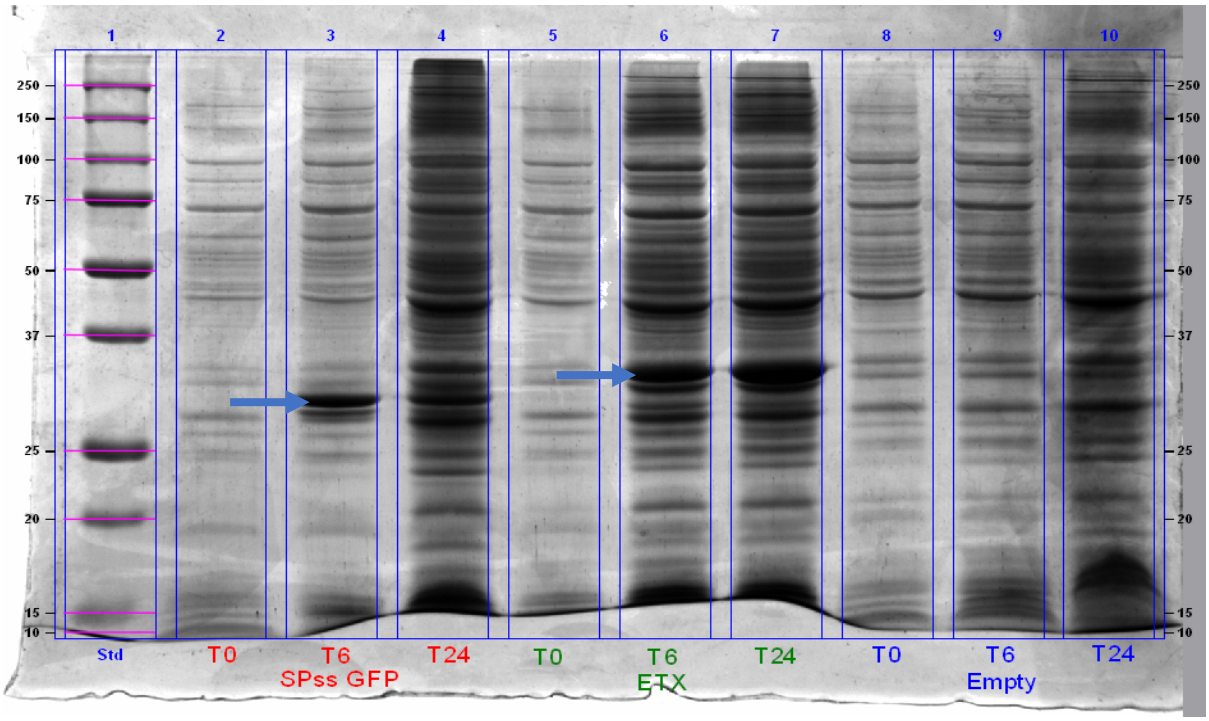


Figure 6.22: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrows indicate SPssGFP and ETX protein bands. Arrows indicate SPssGFP and ETX protein bands.

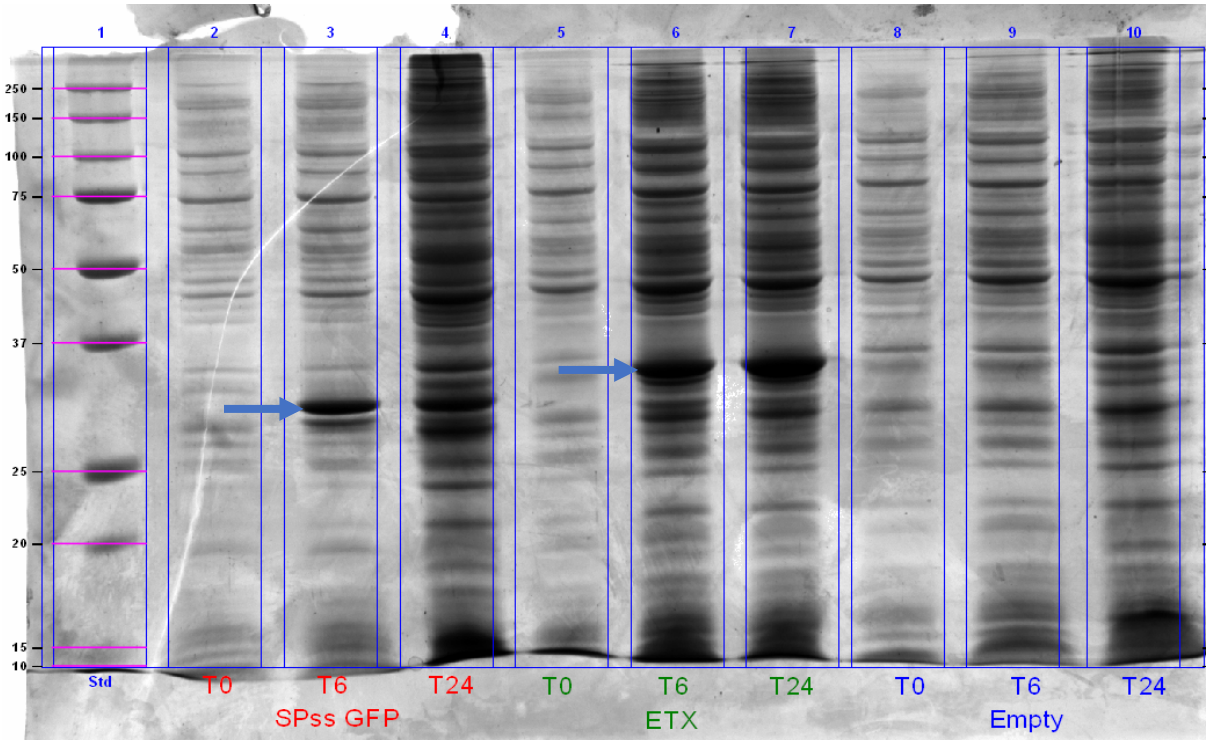


Figure 6.23: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrows indicate SPssGFP and ETX protein bands.

6.3.5. Periplasmic fraction: 30°C IPTG induction temperature

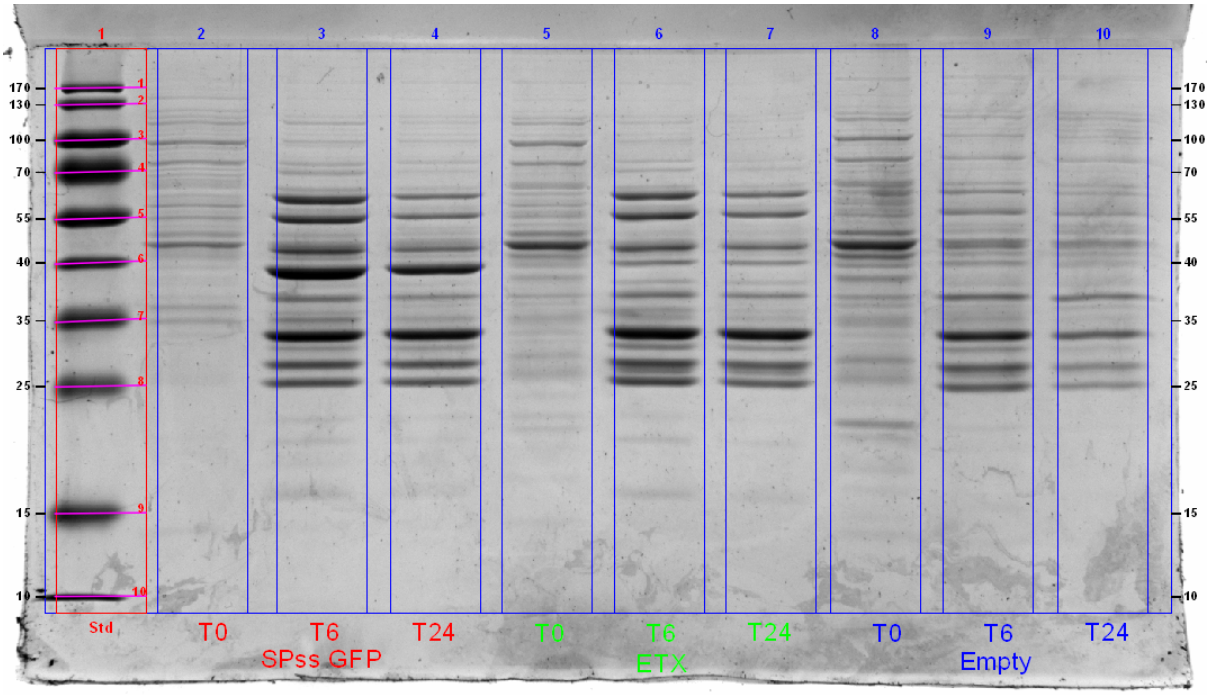


Figure 6.24: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µM IPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

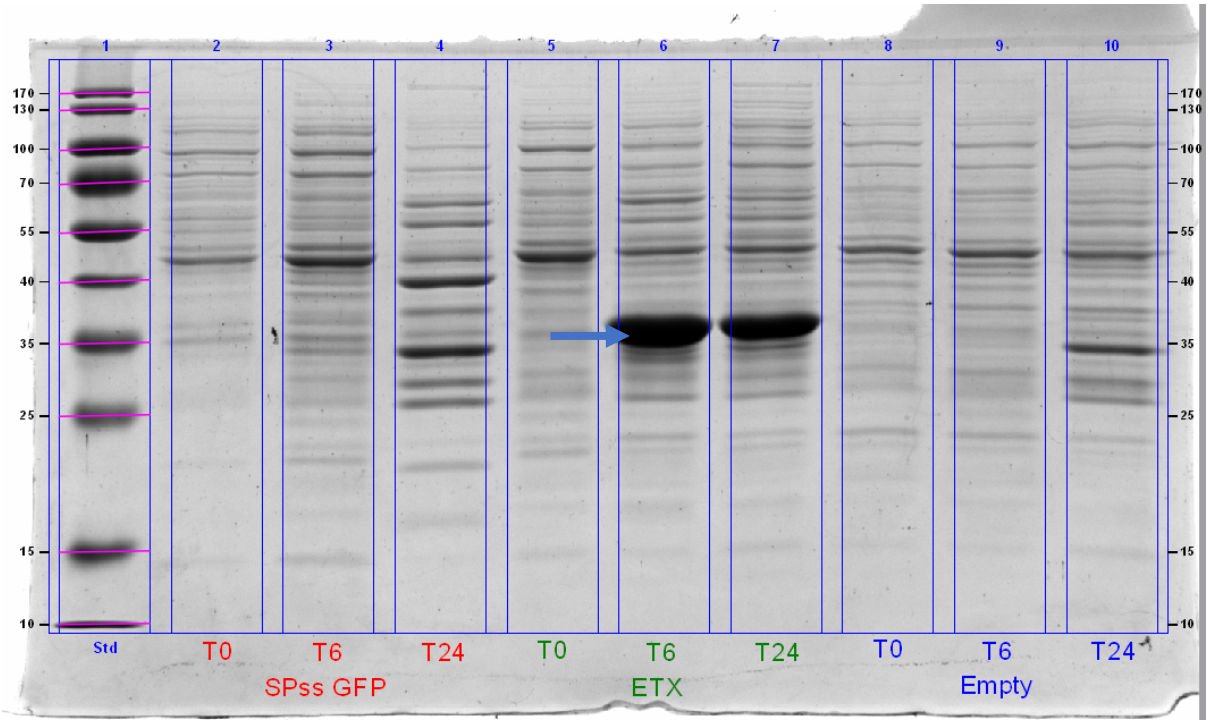


Figure 6.25: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band. The presence of SPssGFP at T24 is likely but looks extremely similar to the band pattern of the Empty control at T24.

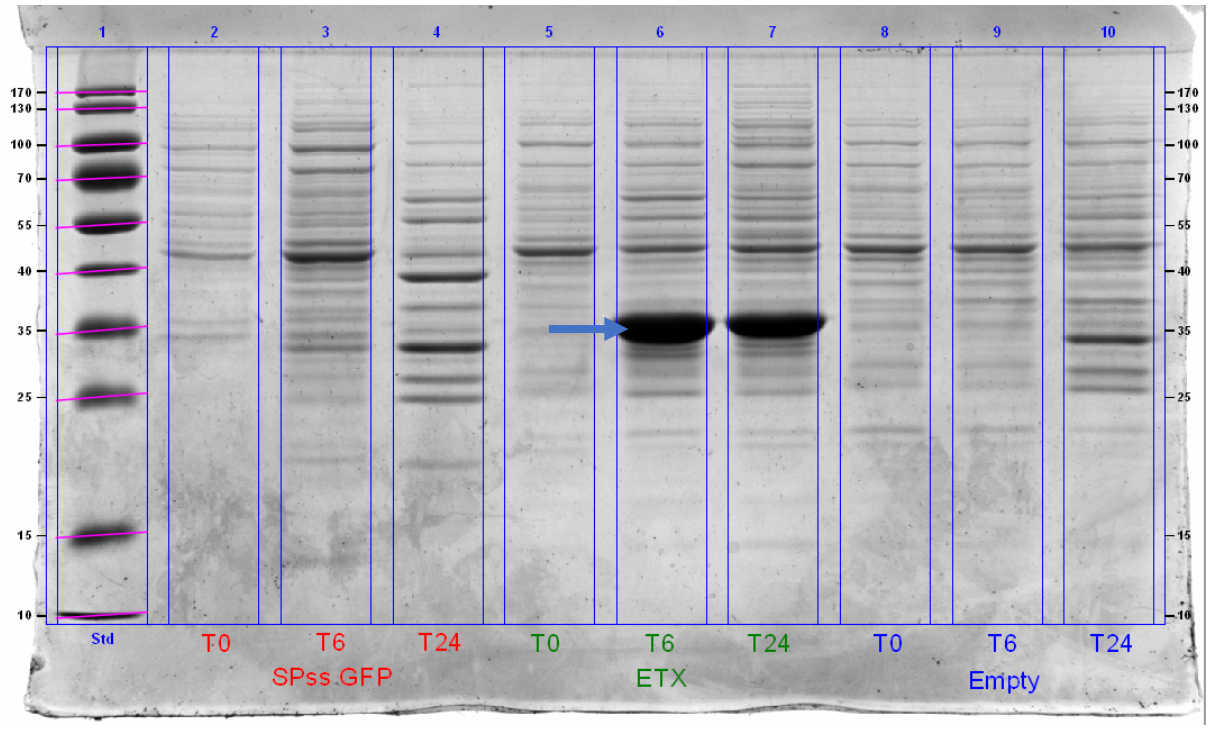


Figure 6.26: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band. The presence of SPssGFP at T24 is likely but looks extremely similar to the band pattern of the Empty control at T24.

6.3.6. Extracellular fraction: 30°C IPTG induction temperature

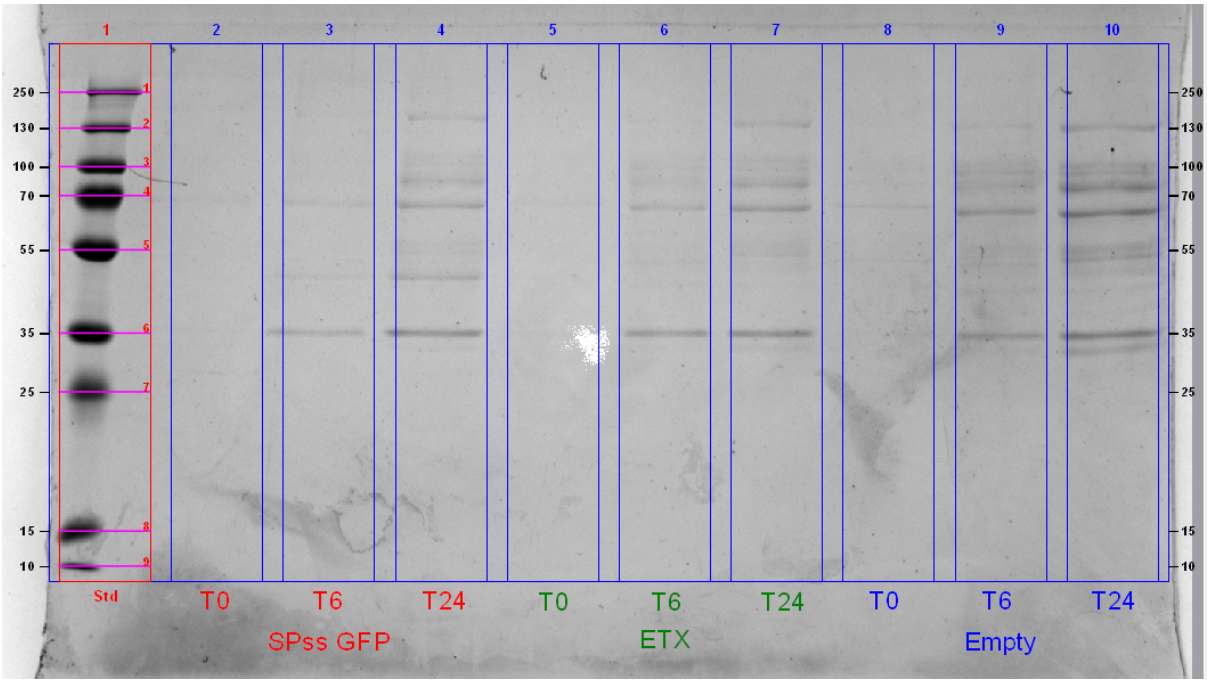


Figure 6.27: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µM IPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

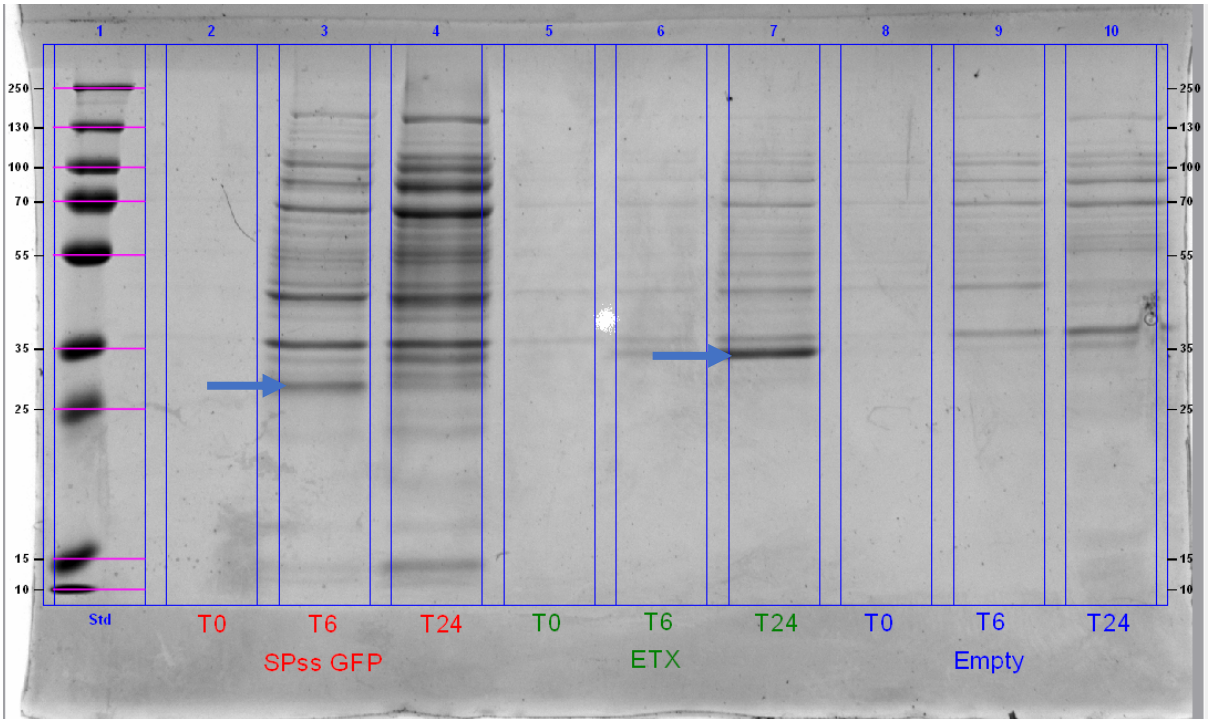


Figure 6.28: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µMIPTG, at 30°C grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrows indicate SPssGFP and ETX protein bands.

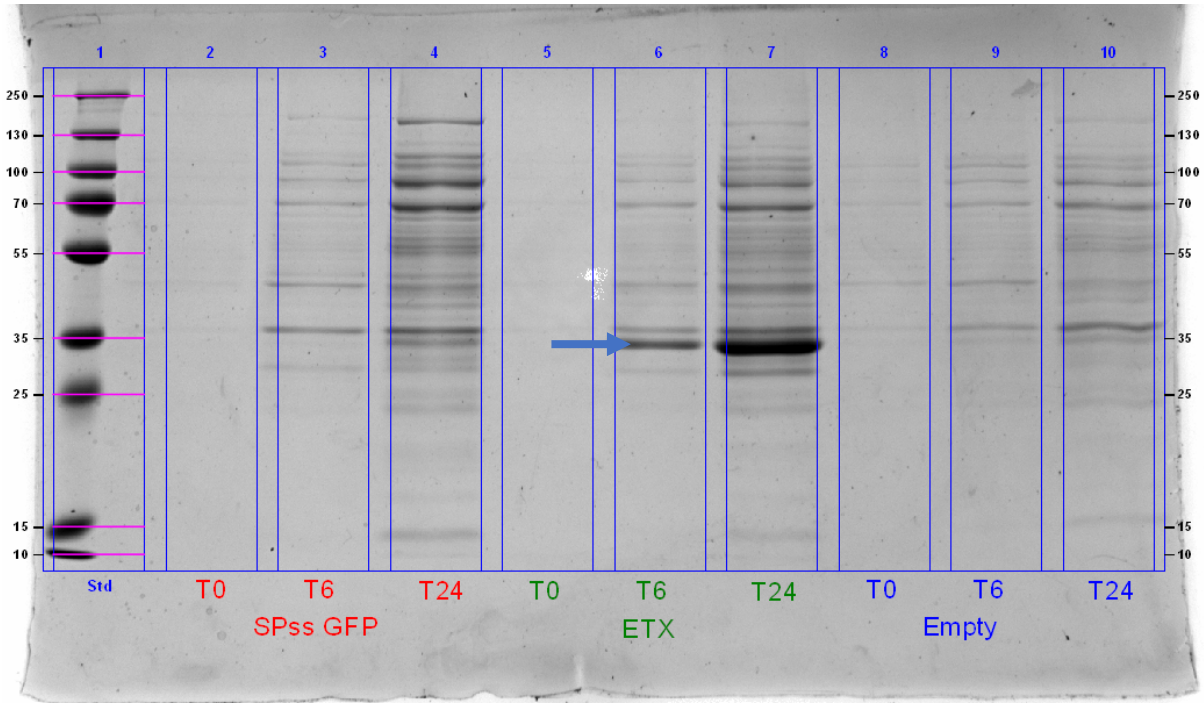


Figure 6.29: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µMIPTG, at 30°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein bands.

6.3.7. Cytoplasmic fraction: 37°C IPTG induction temperature

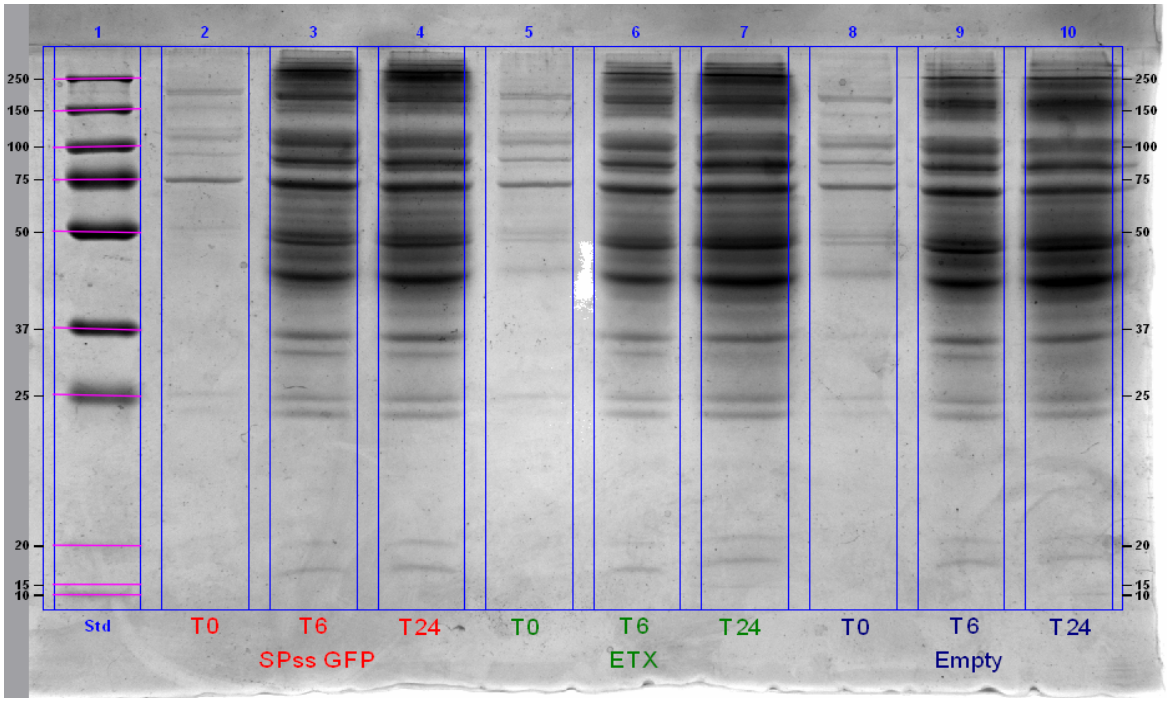


Figure 6.30: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µMIPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

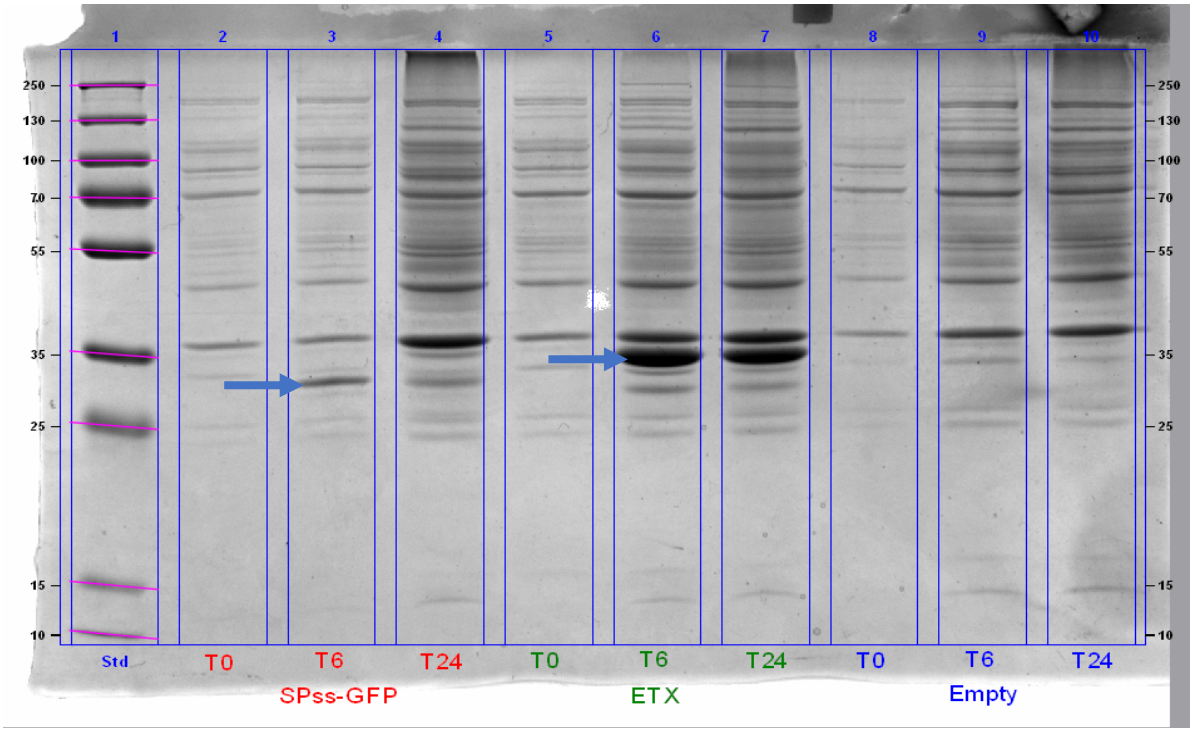


Figure 6.31: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100μM IPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrows indicate SPssGFP and ETX protein bands.

6.3.8. Periplasmic fraction: 37°C IPTG induction temperature

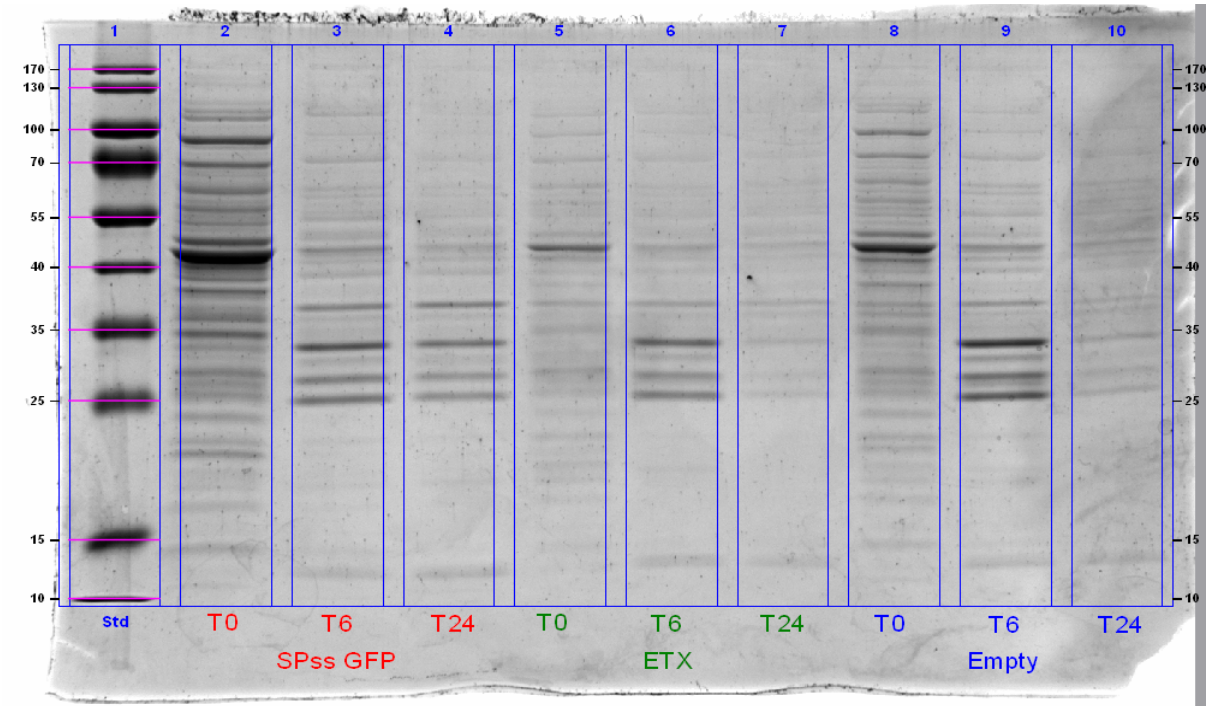


Figure 6.32: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0µM IPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

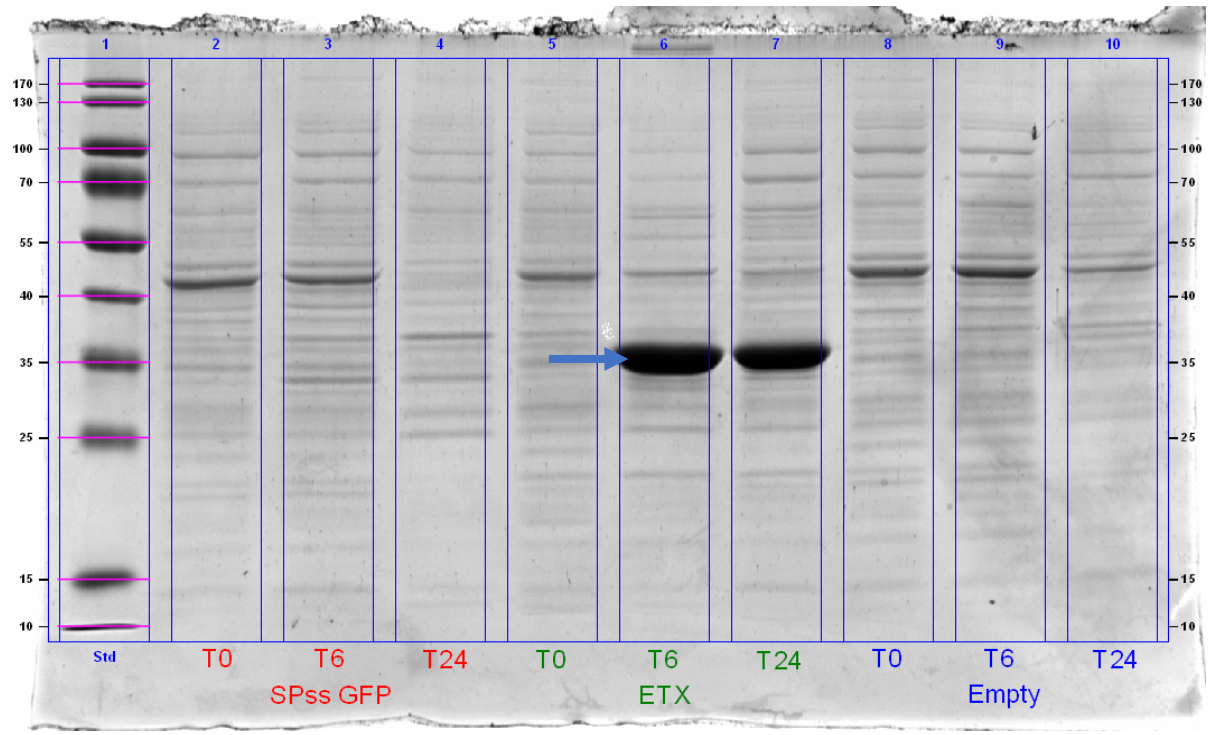


Figure 6.33: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100µMIPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). ETX protein bands are clearly visible at 6 and 24 hours after induction and indicated by an arrow.

6.3.9. Extracellular fraction: 37°C IPTG induction temperature

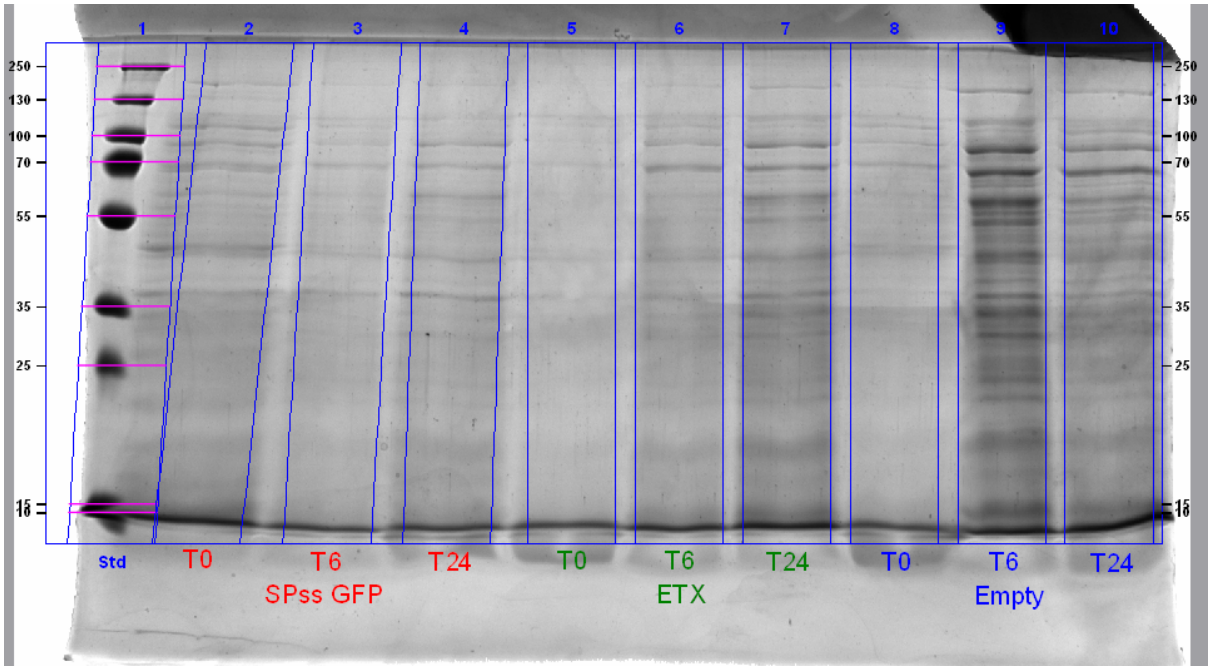


Figure 6.34: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 0μM IPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).

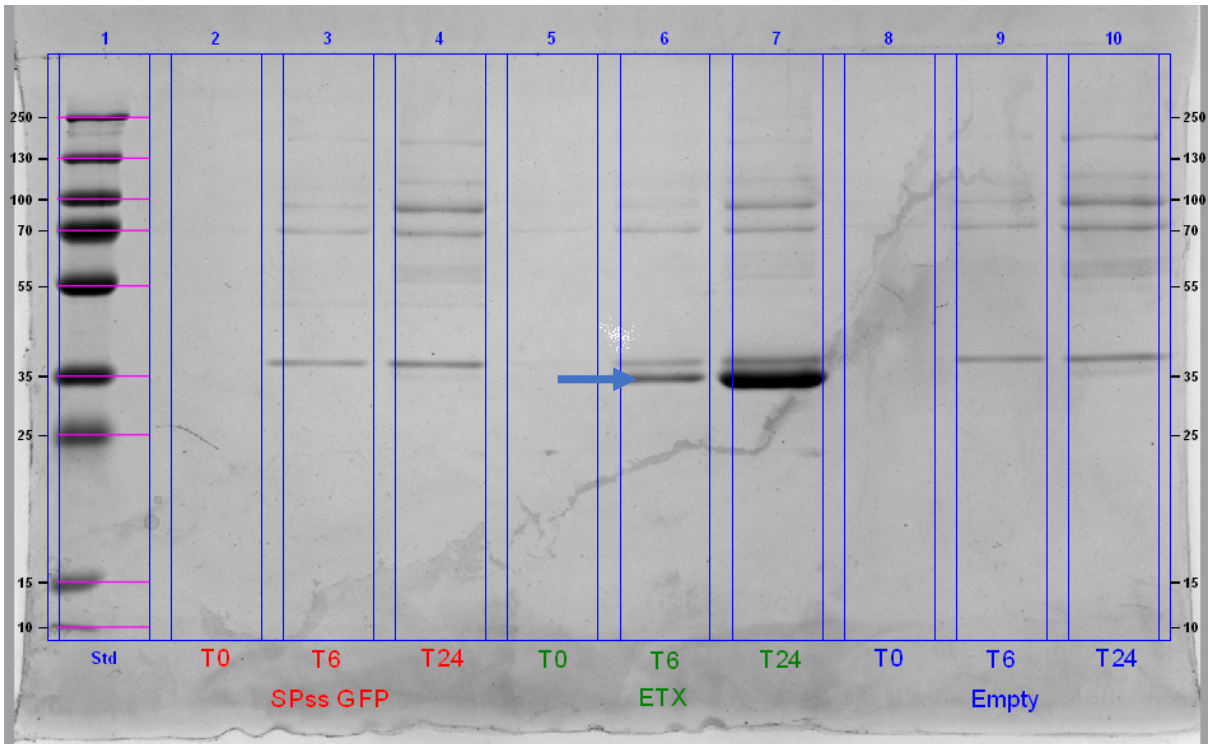


Figure 6.35: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 100μM IPTG, at 37°C and grown in LB. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band, present at 6 and 24 hours after induction.

6.4. Additional PAGE gels of total and fractionated proteins of *E. coli* cultures grown on M2P

The SDS-PAGE gels, shown below, are grouped by induction temperature. All cultures were grown overnight at 37°C before expression induction with IPTG. Threshold for protein expression was above 100µM IPTG. A reduction in temperature was correlated with a reduction in protein band intensity. No extracellular SPss-GFP protein was observed in these experiments.

6.4.1. Total protein fraction of SPss-GFP under different temperatures and IPTG concentrations.

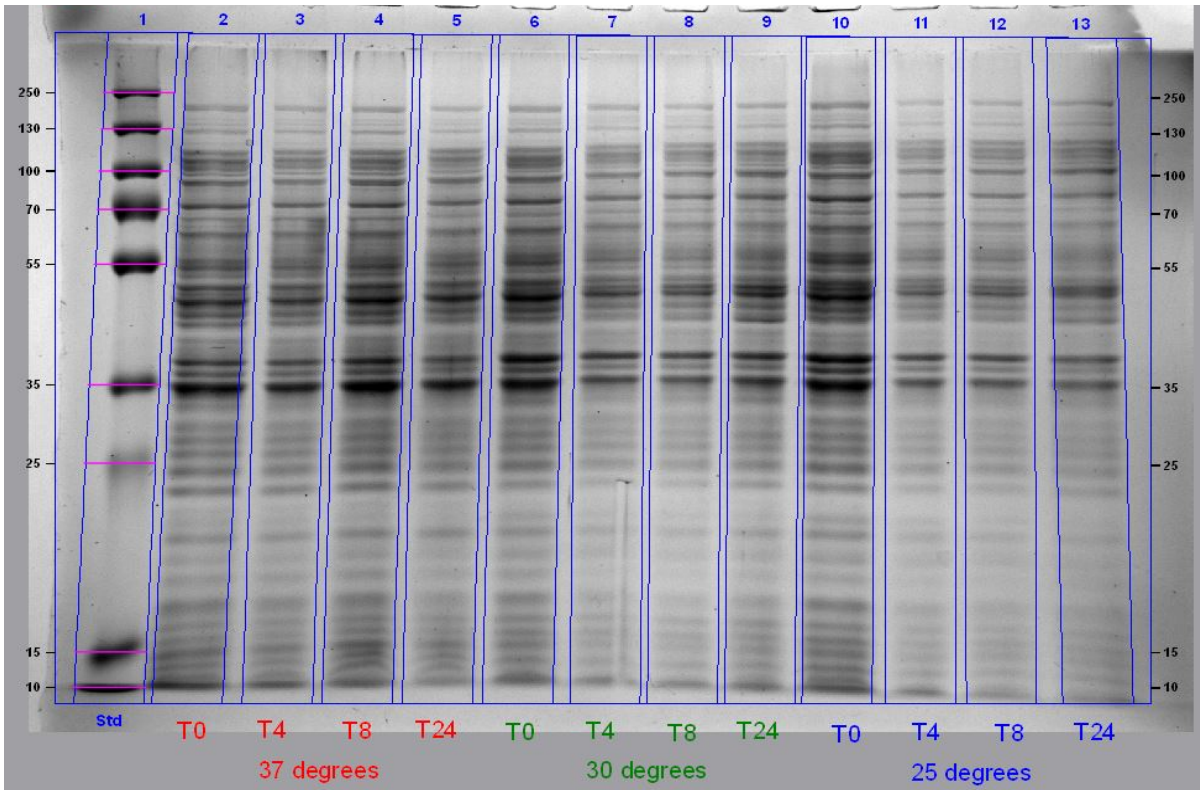


Figure 6.36: Total protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 0µM IPTG, at three induction temperatures (37°C, 30°C and 25°C). Lanes show samples at 0, 4, 8 and 24 hours post-induction and grown in M2P.

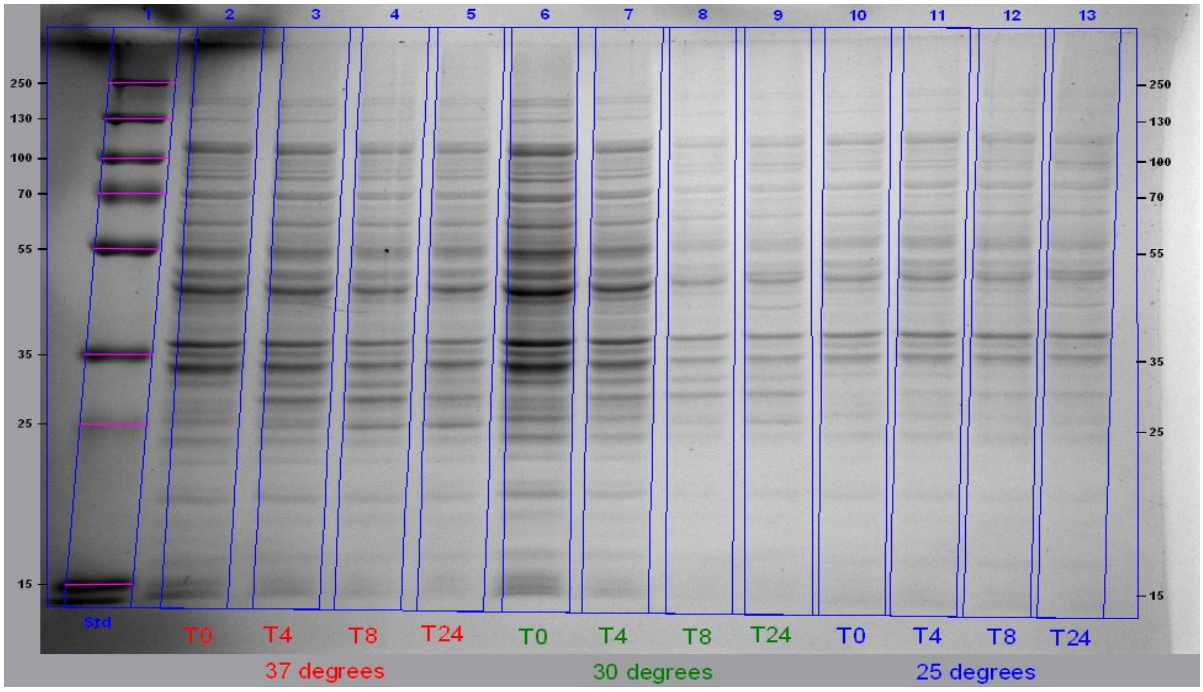


Figure 6.37: Total protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 10µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction. No SPssGFP protein bands were observed.

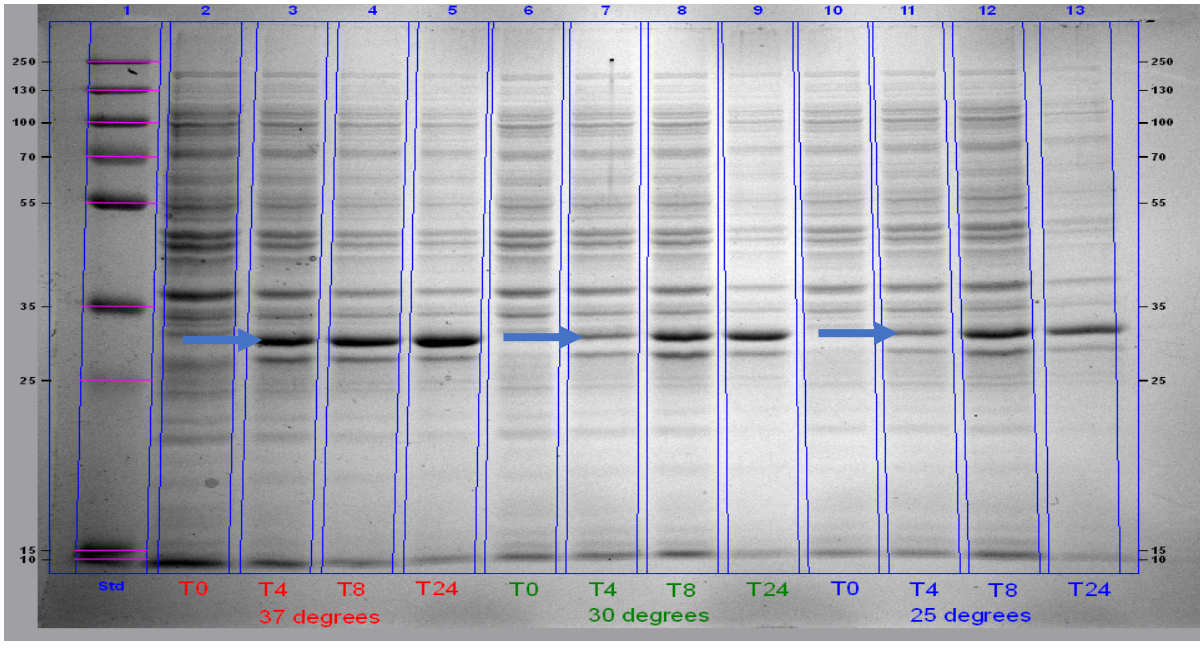


Figure 6.38: Total protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 100µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction. Arrow indicates feint SPssGFP protein band, other prominent protein bands are present as larger and darker bands of the same size. SPssGFP is present after 4 hours of induction at all temperature conditions, but with variable protein levels.

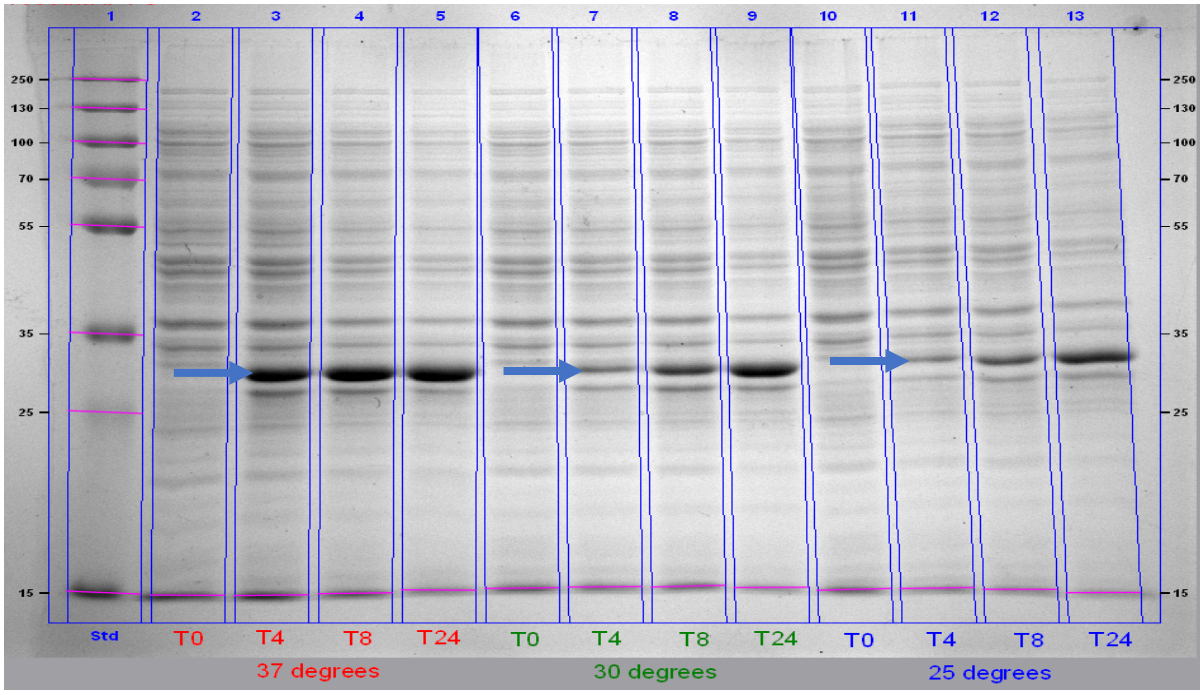


Figure 6.39: Total protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 1000µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction. Arrows indicate feint protine bands of SPssGFP, more priminant protein bands of SPssGFP appear as dark bands of the same size, appearing after 4 hours of IPTG induction.

6.4.2. Extracellular fraction of SPss-GFP grown in M2P growth medium

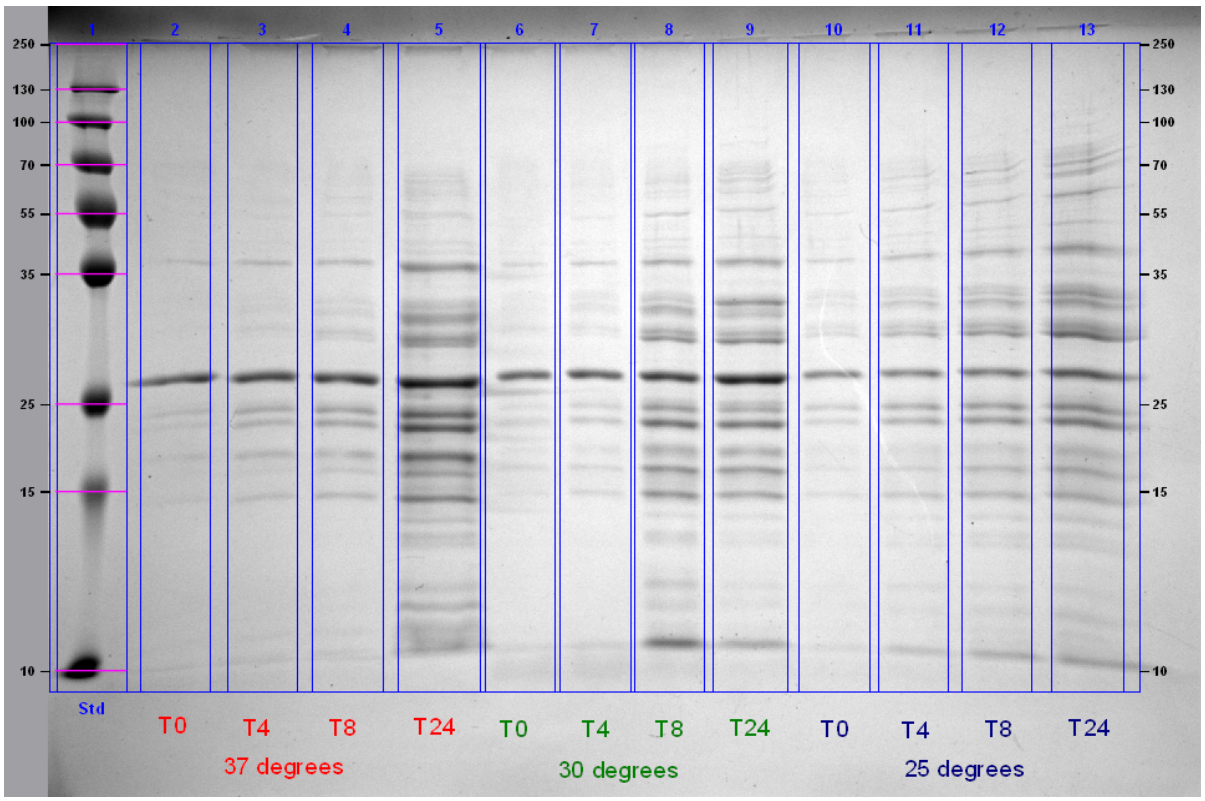


Figure 6.40: Extracellular protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 0µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction.

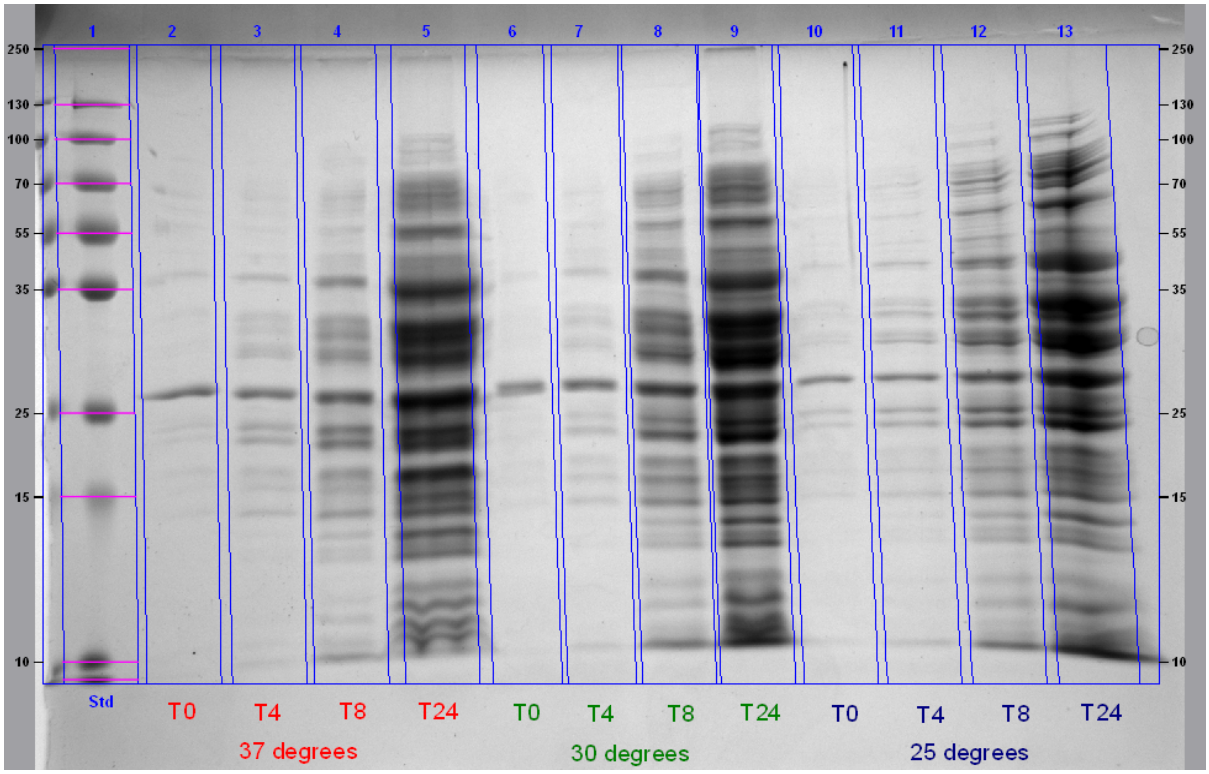


Figure 6.41: Extracellular protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 1µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction.

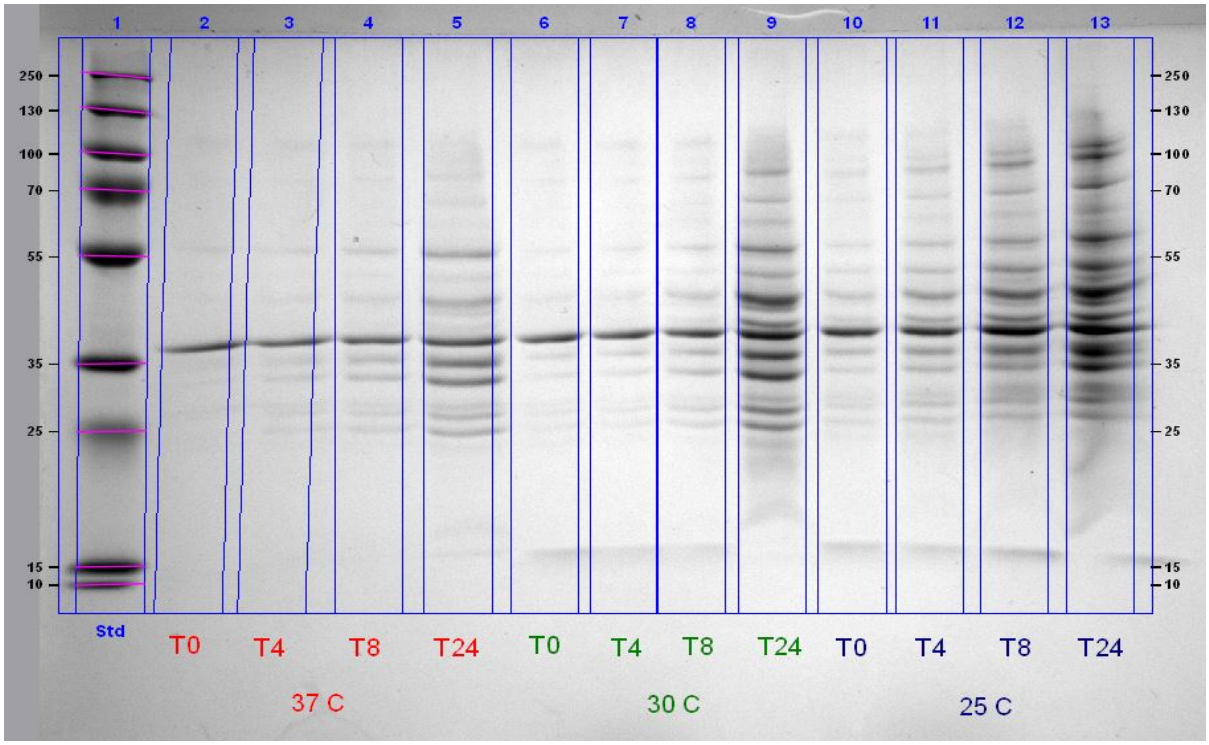


Figure 6.42: Extracellular protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 10µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction.

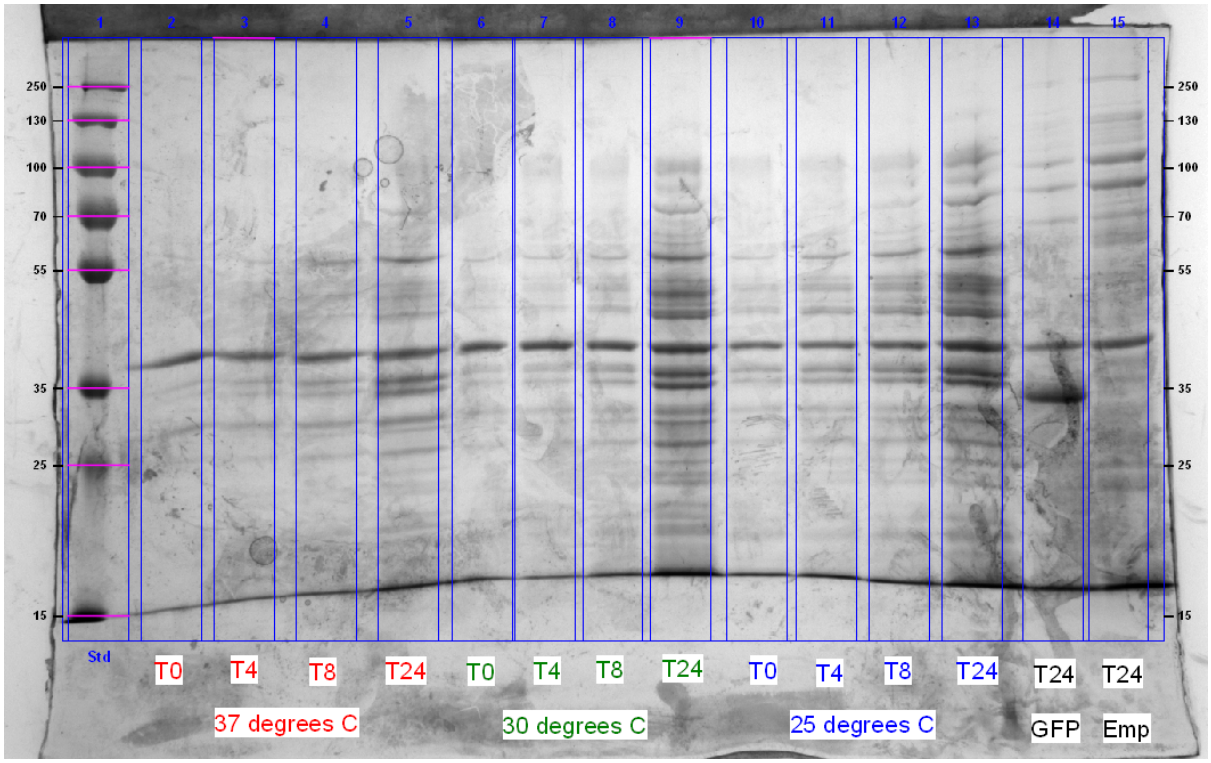


Figure 6.43: Extracellular protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 100µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction. Additional controls were also loaded in lanes 14 and 15, showing Green fluorescent protein and 'Empty' expression vector, both in BL21(DE3).

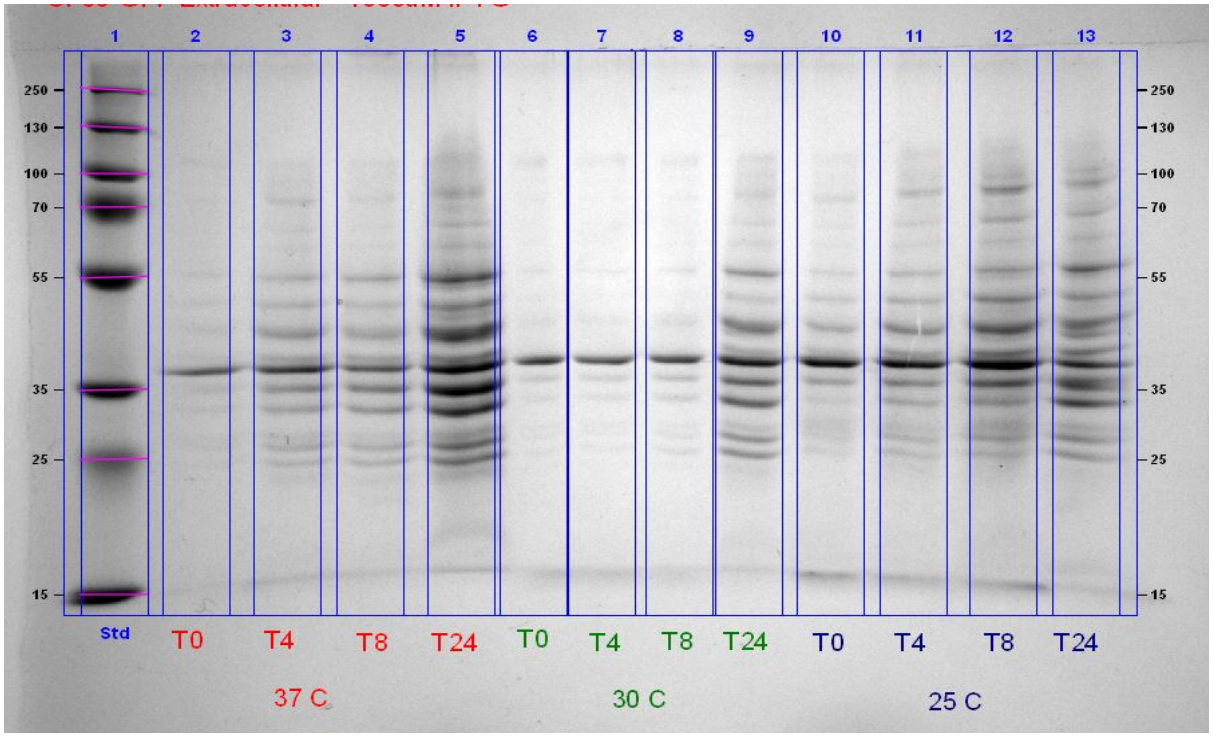


Figure 6.44: Extracellular protein of SPss-GFP construct expressed in E. coli BL21(DE3), at 1000µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction.

6.5. Bacterial culture fractions under selected conditions grown in M2P growth-medium

6.5.1. Cytoplasmic fraction

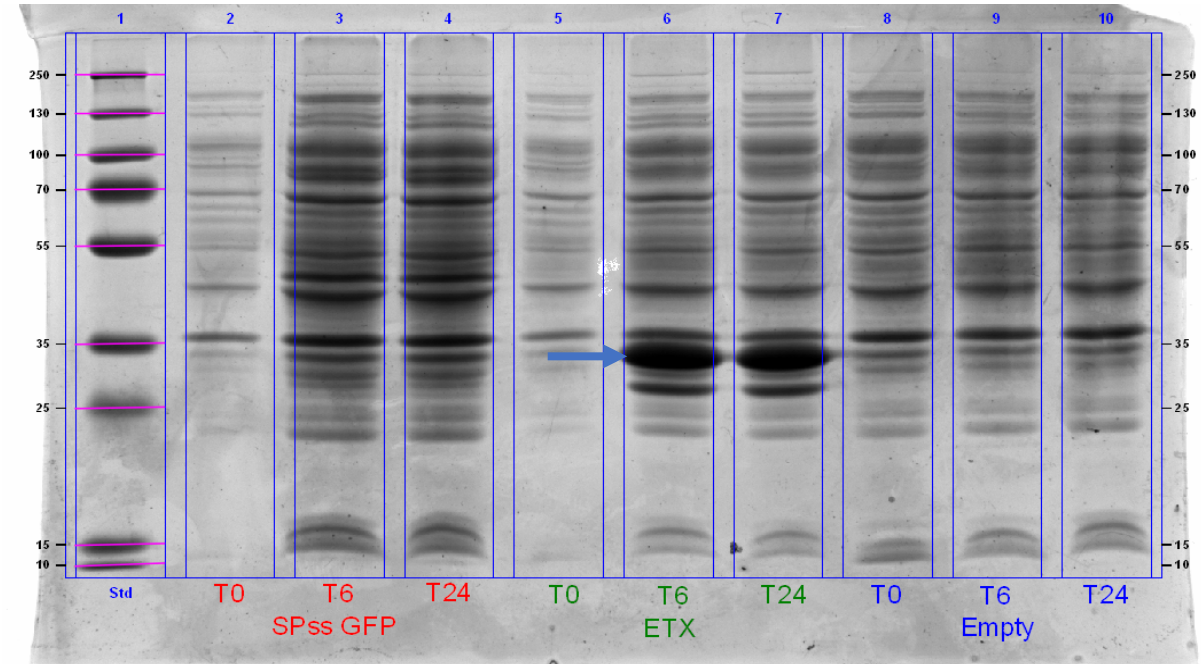


Figure 6.45: Cytoplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 30°C and grown in M2P. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). Arrow indicates ETX protein band.

6.5.2. Periplasmic fraction

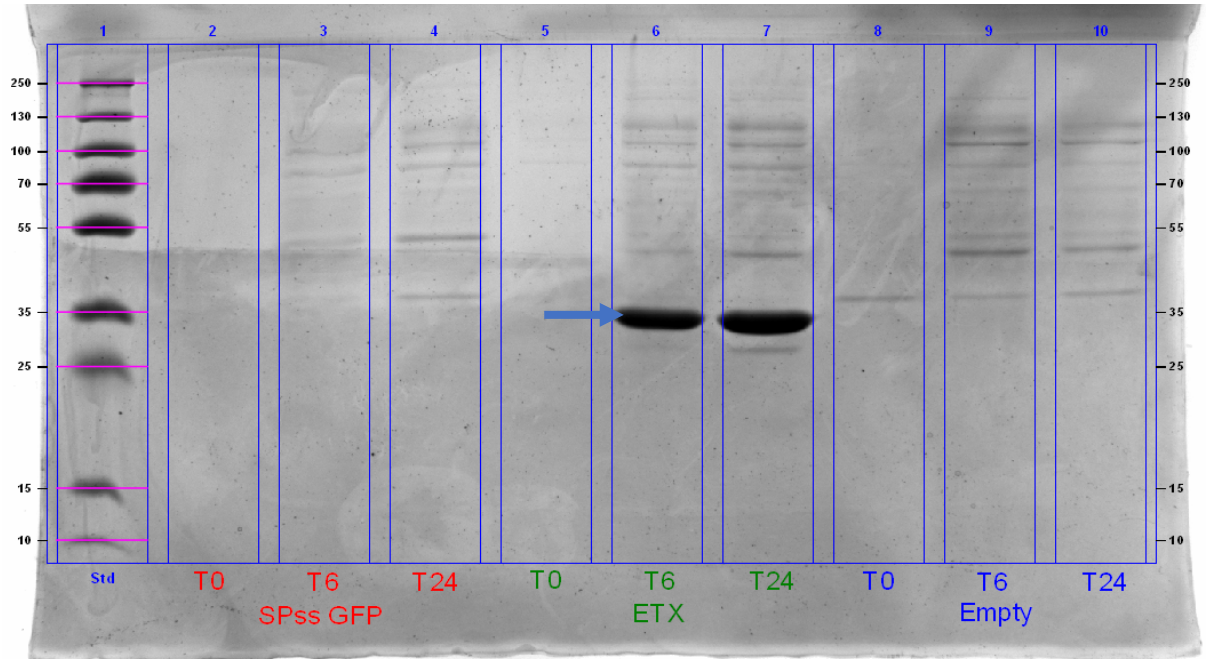


Figure 6.46: Periplasmic protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmids, induced with 1000µM IPTG, at 30°C and grown in M2P. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively). ETX bands are clearly present at 6 and 24 hours after induction, indicated by arrow.

6.5.3. Extracellular fraction

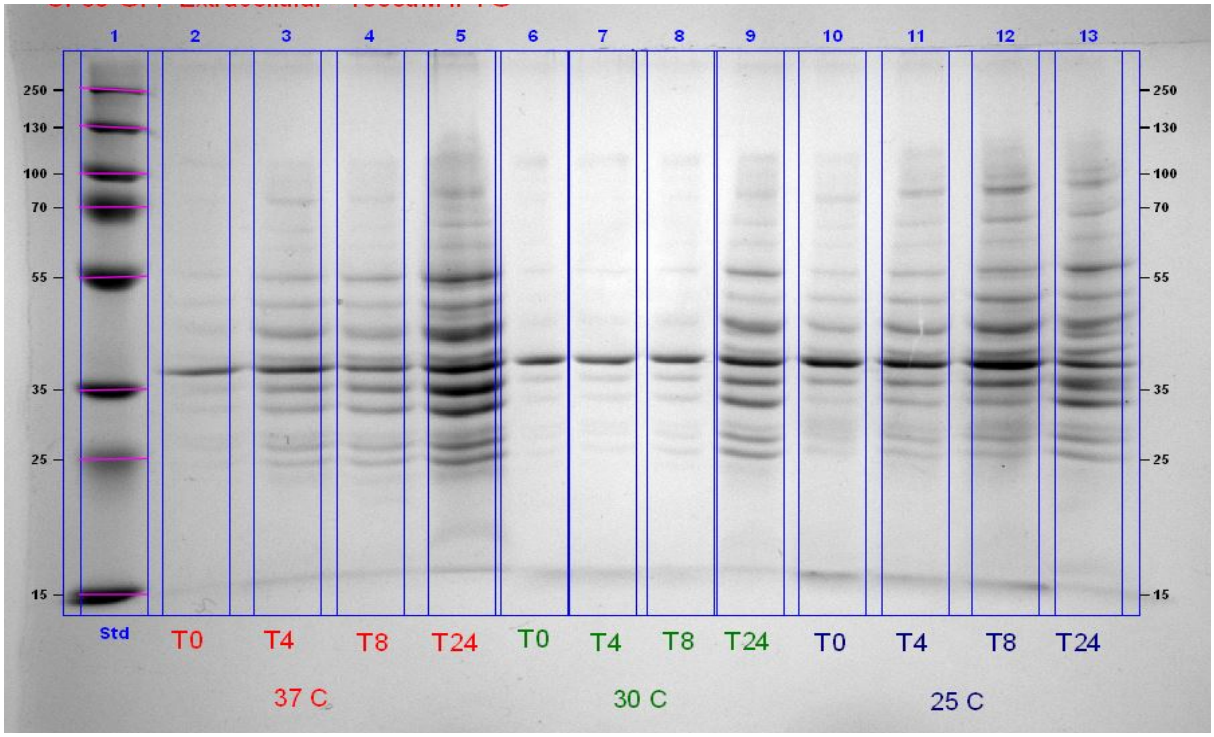


Figure 6.47: Extracellular protein of SPss-GFP construct, expressed in E. coli BL21(DE3), at 1000µM IPTG, at three induction temperatures (37°C, 30°C and 25°C) and grown in M2P. Lanes show samples at 0, 4, 8 and 24 hours post-induction.

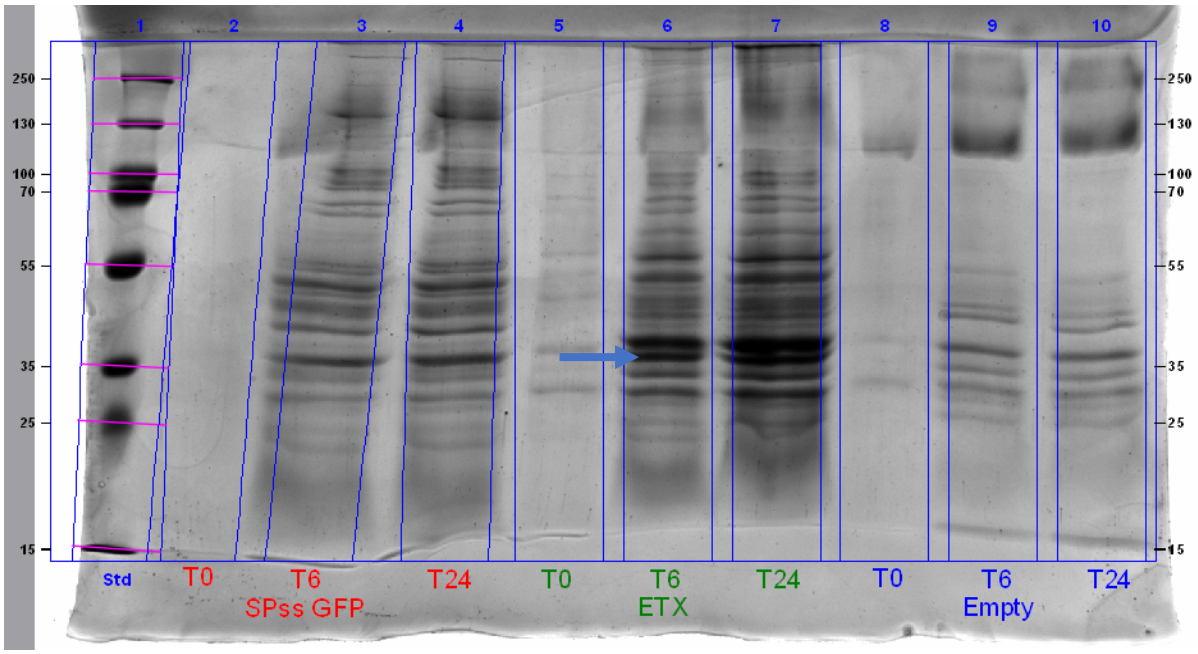


Figure 6.48: Extracellular protein fraction of E. coli BL21(DE3) expressing SPss-GFP, ETX or 'Empty' plasmid, induced with 1000µM IPTG, at 30°C and grown in M2P. Time points at 0, 6 and 24 hours post induction (T0, T6, T24 respectively).