Optimisation of feedstock utilisation by Geobacillus thermoglucosidasius

Holland, Alex

Award date:
2017

Awarding institution:
University of Bath

Optimisation of feedstock utilisation by *Geobacillus thermoglucosidasius*

Alexandria Tamsyn Norain Holland

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Pharmacy and Pharmacology

February 2017

Abstract

Geobacillus thermoglucosidasius (*GT*) is a thermophilic, ethanol-producing bacterium capable of utilising both hexose and pentose sugars for fermentation. One strategy to improve fermentation yields would be to engineer *GT* strains to secrete hydrolases to increase the amount of available sugars from various feedstocks. Therefore, optimised protein secretion would be vital to improve feedstock utilisation. Secretion in the related mesophile *Bacillus subtilis* (*BS*) has been well studied, and several strategies have been developed to improve secretion of heterologous proteins in *BS*, one such strategy being the manipulation or changing of the signal peptide.

One aim is to identify any differences in the secretion machinery and signal sequences between *GT* and *BS*. Another aim is to analyse any effects of overproduction of hydrolases and to identify any bottlenecks in protein secretion in GT.

Using bio-informatics tools we find that although *GT* is a thermophile, the signal peptides in this organism do not differ significantly from those in *BS*. From a shotgun mass spectrometry approach it was also observed that unlike BS, GT undergoes significant cell lysis during growth releasing cytoplasmic proteins into the extracellular milieu, which could have implications on the levels of secreted hydrolases.

A model enzyme was selected and over-produced at high levels in order to stress the secretion system in GT so as to identify any bottlenecks in secretion. The results thus far indicate that the rate limiting step in secretion could be post-translocation where the enzyme is degraded by proteases in the cell wall and extracellular milieu. The addition of protease inhibitor to growth media, increases the activity and abundance of the enzyme, suggesting that proteolysis may be a major factor when over-producing secreted enzymes at high levels.

List of abbreviations

APS	Ammonium persulphate
ASM	Ammonia Salts medium
AU	Arbitrary units
BLAST	Basic Local Alignment Search Tool
BS	*Bacillus subtilis*
BSA	Bovine Serum Albumin
C5	Pentose sugars
C6	Hexose sugars
CAZy	Carbohydrate-Active enZYmes
DNS	Dinitrosalicylic acid
dNTP	deoxyribonucleoside triphosphates
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionisation
ExPASy	Expert Protein Analysis System
FPLC	Fast protein liquid chromatography
GRAVY	Grand average of hydropathy
GT	Geobacillus thermoglucosidasius
IMAC	Immobilised-Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	Michaelis constant
LB	Luria broth
LCMS	Liquid Chromatography-Mass Spectrometry

LDH	Lactate dehydrogenase
Ni-NTA	Nickel-nitrilotriacetic acid
OD	Optical density
PCR	Polymerase Chain Reaction
PDH	Pyruvate dehydrogenase
pI	Isoelectric point
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	Super Optimal with Catabolite repression medium
SP	signal peptide
SPase	signal peptidase
SRP	Signal recognition particle
TAE	Tris-acetate-EDTA
TAT	twin-arginine targeting
TEMED	Tetramethylethylenediamine
TGP	Tryptone glycerol pyruvate
Tm	Melting temperature
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
USM	Urea Salts medium
Vmax	Maximum velocity of reaction

Acknowledgements

Firstly, I would like to thank my supervisor Dr Albert Bolhuis, for the amazing support and guidance throughout my PhD, always having time for discussion and demonstration of laboratory techniques, with patience, understanding and encouragement. Without his guidance, I would not be the scientist I am today. I could not have asked for a better supervisor during the last four years. I would also like to thank Professor Michael Danson for his input and supervision during the project; he has always made time for discussion, with patient and thorough explanations.

My deepest gratitude goes to everyone past and present in the Danson and Leak labs who over the years have watched me present at lab talks and given me feedback, advice and input on my work, and presentation skills! I would like to thank you all for being so welcoming and accepting, and treating me like one of your own, even though I am all the way in 7W! Especially Chris Hills, Emanuele Kendrick, Leann Bacon, Lisa Budruss, Charlie Bennet, Alice Marriott and my morning cheerleader Micaela Chacon.

I would also like to give a big thank you to everyone who has occupied the PhD office in 5W 2.48, for the top bants, board game nights, camping trips, and nights out, especially Emma Robson, Annelisa Sadler, Robin Alfred Wickens, Matt Udakis, Laura Newton, and everyone else. I would like to thank my office and lab-mate, Helen Ji Yuan, for her help, invaluable friendship, beautiful paintings and relentless questioning, without which, I am sure the lab would not have been as colourful and fun. I would like to thank the technical staff in P&P, who have considered me an honorary member of the technical team, treated me with friendship and kindness, and provided great banter throughout my time at Bath.

I have experienced numerous highs and lows, my struggle with depression being the biggest low. I would like to thank everyone mentioned above for your help, in whatever way, for helping me through some of my darkest of times. I will forever be grateful to my friends and colleagues who have been there for me, be it to have an existential crisis conversation, to a beautiful painting, to just having a laugh. Depression seems to be a bit of a taboo subject but I wanted to do my small part, in making the conversation about mental illness, a little more open and accepted.

I am eternally grateful for the financial contributions from the BBSRC, and TMO Renewables Ltd, without which I would not have embarked on this mad journey. I wish to thank the Microbiology society, the biochemical society, the society for applied microbiology and CBMnet for awarding me grants to present my work at conferences.

I am so grateful to my Mum and Dad, and the rest of my family and in-laws, for being supportive and proud of me. And most important of all, Lyes Badaoui, my husband, best friend, partner, thank you for supporting me through my PhD, thank you for putting up with me, my mood swings, my tiredness, my neediness and my erratic behaviour during my PhD. The end is in sight my love.

“It's the job that's never started as takes longest to finish”

Ham Gamgee, Samwise Gamgee’s Old Gaffer

Table of Contents

List of figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Appendix One

List of tables

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Appendix One

CHAPTER ONE: GENERAL INTRODUCTION

1.1 BIOFUELS AND BIO-ETHANOL

A finite supply of fossil fuels, energy security issues, fluctuating and increasing oil prices, environmental concerns, and rapid growth in energy demands, are just some of the reasons that have driven the search for alternative and renewable sources of energy. While several different types of renewable fuel are being considered for long term, lignocellulosic biomass as a resource for the production of biofuels and other chemicals is certainly feasible in the near future.

The term biofuel describes carbon-based fuels, either produced by or derived from a living organism, typically plants or plant matter. Biofuels such as bioethanol, bio-butanol, biodiesel and bio-hydrogen have great potential as renewable alternatives to fossil fuels as they are derived from plant biomass, which is an abundant and renewable source of carbon for microbial conversion of carbohydrate into biofuels such as bioethanol, or even other organic compounds, by bacteria, algae, yeasts and even archaea (Lan and Liao, 2013).

Bioethanol has been produced for the last three decades and is the most popular biofuel, with global bioethanol production at over 25 billion gallons in 2015, with the USA alone producing almost 15 billion gallons as seen in Figure 1.1. This is chiefly due to microorganisms that can have been found to naturally produce ethanol, and have been exploited and engineered to produce ethanol at high levels. Mature technologies for ethanol production are therefore mainly crop-based; typical crops include sugar cane, corn, beets, wheat, sorghum, sunflower, soybean, cassava, etc. These types of feedstocks contain high levels of starch or sucrose, which can be fermented to ethanol by microorganisms (Sanchez and Cardona, 2008); these are known as a first-generation biofuels. First-generation biofuels have been commercialised worldwide with established technologies and mature markets. However, this is to some extent controversial due to numerous socio-economic and environmental impacts caused by the utilisation of precious farmland for fuel production rather than food production (Haber, 2007, Tenenbaum, 2008, Stoeglehner and Narodoslawsky, 2009). There is therefore much interest towards exploiting the less expensive, and readily available,

biomass such as municipal, agricultural and industrial waste products and thus second-generation biofuels were developed.

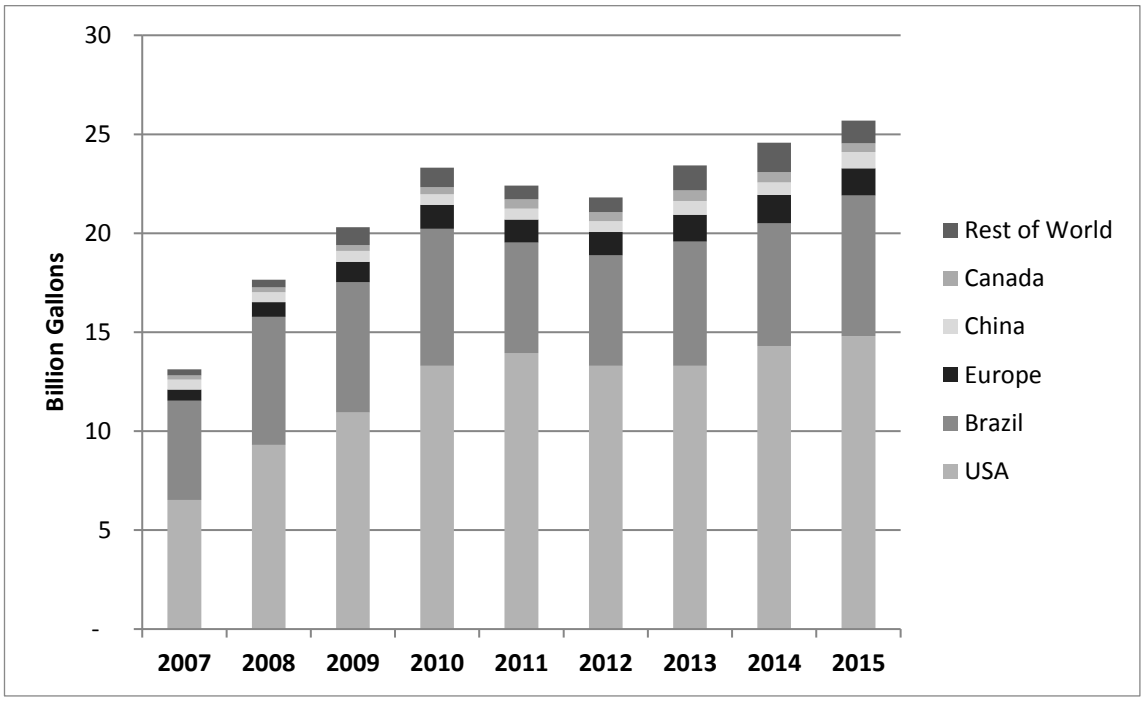

Figure 1.1: World Fuel Ethanol Production by Country or Region (Million Gallons). Data from Renewable fuels association (www.afdc.energy.gov/data)

Second-generation biofuels are derived from lignocellulosic feedstocks instead of food crops. This process utilises and exploits readily available organic material such as agricultural or municipal wastes and forestry residues, or fast growing grasses such as those grown on marginal cropland or land unsuitable for food crop production. Production of fuels from feedstocks of this nature enhances the value of waste products, while avoiding the use of farmland for food production, reduces landfill and therefore greenhouse gas emissions, therefore making it environmentally friendly (Liao et al., 2016). However, to release simple sugars from the lignocellulose, thermal, chemical and enzymatic processing is required prior to fermentation by micro-organisms (Peralta-Yahya et al., 2012), as can be seen in the simplified workflow in Figure 1.2, which adds to production costs.

Aside from biofuels like bioethanol, a range of green building-block chemicals such as lactic acid or butanol can be produced from biomass through microbial fermentation, but in order to be a large-scale alternative to petrochemicals, their production must

become more competitive in terms of cost, and be based on sustainable and renewable resources.

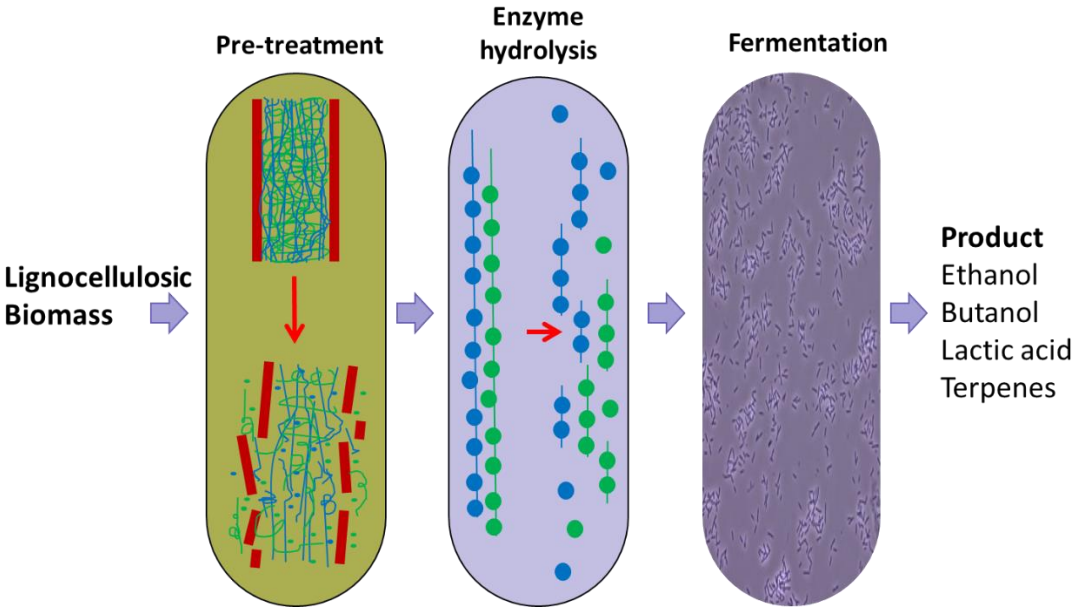

Figure 1.2: Simplified typical workflow of bio-ethanol production from lignocellulosic biomass.

1.2 LIGNOCELLULOSIC BIOMASS AND HYDROLYTIC ENZYMES

Biomass and biomass-derived materials are considered to be to be one of the most promising alternatives to fossil fuels (Zabed et al., 2016). Simply, these resources are generated through photosynthesis using available atmospheric carbon dioxide, water and light from the sun, making this type of resource a sustainable alternative to petroleum for the production of fuels and other organic chemicals

Lignocellulosic biomass typically describes plant matter and, in the context of this research, is the main carbon source for bio-ethanol production. Lignocellulosic biomass is mainly composed of three polymers: cellulose, hemicellulose and lignin. Depending on the source of the lignocellulosic biomass, these polymers are organized in complex, irregular, three-dimensional structures in variable relative composition. Lignocellulose has a structural function in plants, and has thus evolved to resist degradation. This recalcitrance to degradation is largely due to the crystallinity of cellulose, hydrophobicity of lignin, encapsulation of cellulose by the lignin-hemicellulose matrix, and the heterogeneous nature of hemicellulose.

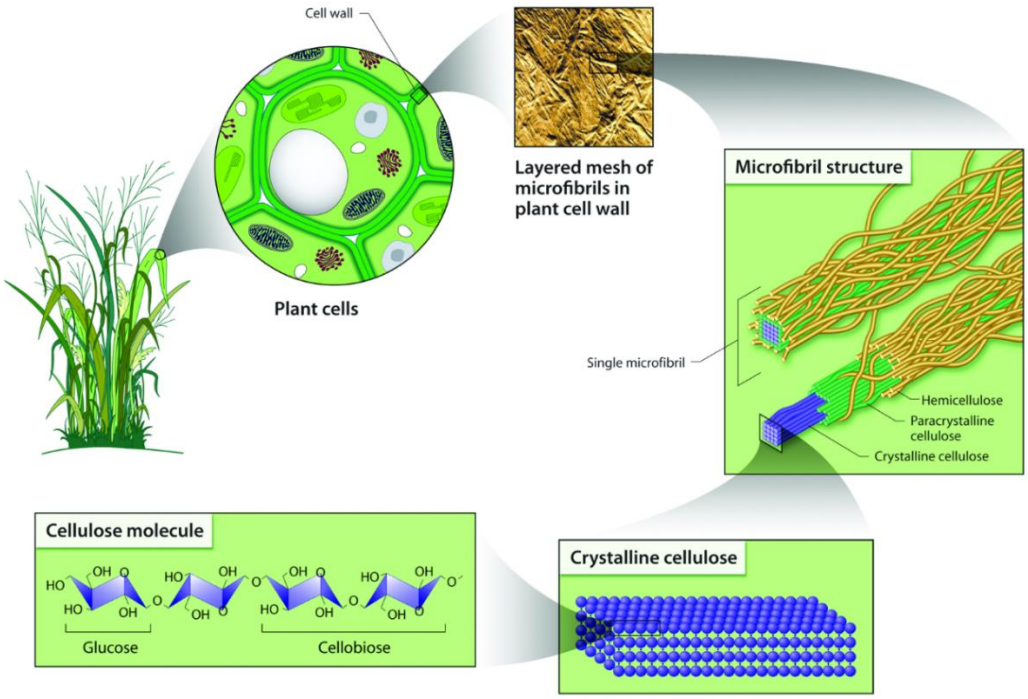

Figure 1.3: Organisation of plant cell wall material showing crystalline and non-crystaline cellulose and hemicellulose. Lignin is not shown here. The structure of crystalline cellulose is shown here to highlight the challenges faced for hydrolysis of crystalline cellulose. Image source: https://public.ornl.gov/site/gallery/detail.cfm?id=181&topic=&citation=&general=hemicellulose&restsection=all

Cellulose is the primary constituent in lignocellulosic biomass, and provides the rigidity in the architecture of the primary plant cell wall. Its structure is crystalline in nature, and consists of extensive intramolecular and intermolecular hydrogen bonding networks, which tightly bind the glucose units. These linkages result in the structural rigidity of cellulose, and confer significant recalcitrance to chemical or enzymatic hydrolysis. The enzymes responsible for the degradation of cellulose are known as cellulases, which are a type of glycoside hydrolase that hydrolyse β-1,4-glucosidic bonds between glucosyl residues.

In contrast to the homogenous composition of cellulose, hemicellulose is a heterogeneous and amorphous polysaccharide composed of a variety of C5 and C6 sugars such as xylose, arabinose, glucose, galactose and many others, depending on the actual source of the hemicellulose. The sugars within the hemicellulose are organised in tight polysaccharide chains, linked together by ß-1-4 glycosidic linkages. Hemicelluloses differ in composition depending on the source; for example, xylans are predominant in hardwood and grass hemicelluloses, while softwood hemicelluloses contain mostly glucomannans, and cereal grains commonly contain mostly arabinoxylans (Perez et al., 2002). Hemicelluloses are embedded in the plant cell walls to form a complex network of bonds, providing structural integrity by linking cellulose fibres into microfibrils and cross-linking with lignin. The xylan backbone is highly substituted with arabinose, glucuronic acid, and acetic, ferulic, and p-coumaric acids, all of which can be stearic obstacles to the action of xylanases and β-xylosidases, and thus limit the hydrolysis of the xylan backbone. Therefore, for complete hydrolysis to occur, the side chains must be cleaved by several auxiliary debranching hemicellulases as seen in Figure 1.4.

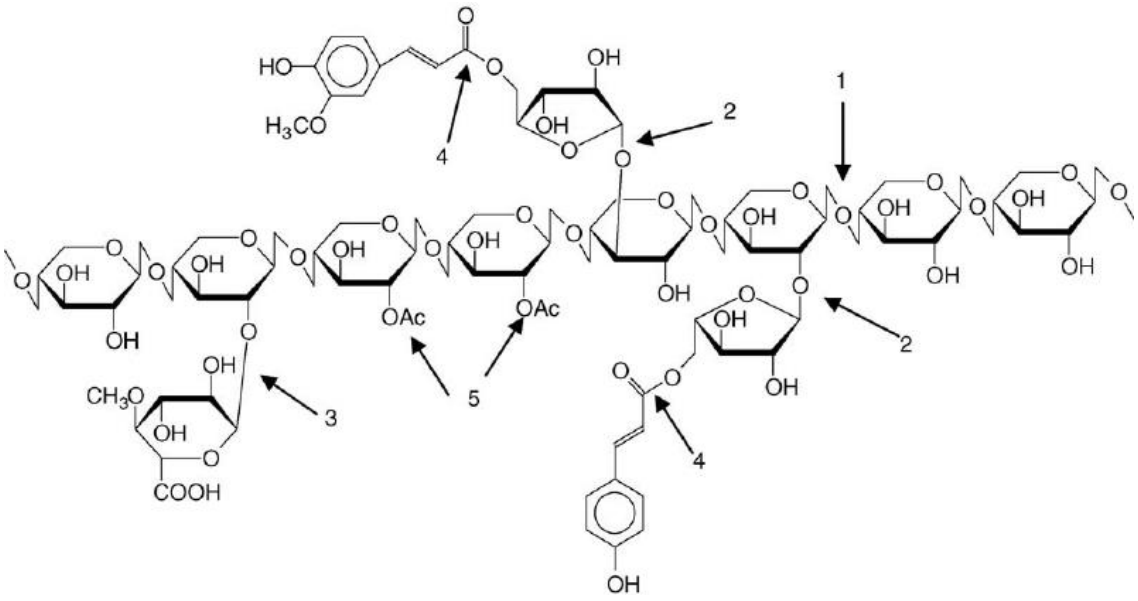

Figure 1.4 The structure of xylan and site of action of the enzymes of the xylanase complex. 1: endoxylanases; 2: arabinofuranosidases; 3: glucuronidases; 4: feruloyl and coumaroyl esterases; 5: acetyl xylan esterases. Image obtained from (Chavez et al., 2006)

As the sugars are locked in a polymer formation, the lignocellulosic biomass is recalcitrant in nature, thus requiring extensive pre-treatment before it can be used as feedstock for fermentation. These pre-treatment steps include physical and chemical pre-treatments, and more importantly, enzymatic pre-treatment to reduce the chain lengths, producing oligosaccharides which are more manageable. This enzyme pre-treatment step is the most costly step, so reduction or elimination of this step would increase cost efficiency of biofuel production (Alfani et al., 2000, Parisutham et al., 2014).

1.3 ETHANOL PRODUCING ORGANISMS

The yeast *Saccharomyces cerevisiae* is the traditional alcohol-producing microorganism, known for ethanol production in the brewing industry. However, in the past 30 years, several ethanol-producing bacteria have been described and developed including *E. coli (Ingram et al., 1987)* and *Zymomonas mobilis* (Vanvuuren and Meyer, 1982, Fein et al., 1983). Another group of organisms that are of interest are thermophiles, which belong to a sub-category of extremophilic microorganisms that are found in and grow at temperatures between 40 and 70°C. They are potentially valuable as microbial cellular factories, as they have a number of advantages over their mesophilic counterparts in industrial-scale bioethanol production. By and large, thermophiles are robust organisms that are able to withstand fluctuations in their environment, such as changes in pH or temperature. Importantly, they are also a valuable source of thermostable enzymes for biotechnology, such as glycosyl hydrolases, proteases, DNA polymerases and DNA restriction enzymes (Vieille and Zeikus, 2001, Turner et al., 2007).

Several thermophiles have also been found to be able to ferment both pentose and hexose sugars found in lignocellulosic biomass (Shaw et al., 2008), and in some cases are able to break down crystalline cellulose (Hirano et al., 2016). This capacity to utilise a wide range of substrates is especially valuable in the production of second-generation biofuels. Furthermore, the use of thermophilic organisms in industrial fermentations also has several advantages due to the increased temperature. For instance, the inhibition of mesophilic contamination reduces the need for the addition of antibiotics, which is costly and has negative environmental consequences. Higher bioprocessing temperatures result in accelerated chemical reaction rates and reduced energy input for refrigeration for example. Higher temperatures also promote improved solubility of substrates, and also facilitate the removal of volatile end products such as ethanol which can vaporise at 50˚C; therefore, applying a mild vacuum might allow continuous “stripping”, thereby reducing the build-up of ethanol to toxic levels (Cripps et al., 2009). Gas solubility decreases as the temperature is increased, which results in a more easily maintained anaerobic environment. Furthermore, thermophiles pose less of an issue if contaminating the environment, as they cannot grow at body or ambient temperatures.

Many of these advantages also translate into monetary savings, thus increasing the cost effectiveness of the fermentation process.

Thermophilic ethanol production has been reported using *Clostridium thermocellum* (Argyros et al., 2011), *Thermoanaerobacterium saccharolyticum* (Shaw et al., 2008, Lin et al., 2014) and *Geobacillus thermoglucosidasius* (*GT*) (Cripps et al., 2009). N-butanol and isobutanol have also been shown to be produced using *Thermoanaerobacterium saccharolyticum* and *Geobacillus thermoglucosidasius (Shaw et al., 2008, Lin et al., 2014)*. Thermophilic *Clostridia* sps. such as *Clostridium thermocellum* are potentially suitable candidates for use in the biofuel production process as they are both cellulolytic and ethanologenic, and therefore they have the potential to be model organisms for consolidated bioprocessing. *C. thermocellum* is able to degrade crystalline cellulose via expression of a diverse set of hydrolase enzymes that form a multi-enzyme complex known as a cellulosome (Bayer et al., 2004, Hirano et al., 2016, Fontes and Gilbert, 2010). Some *Thermoanaerobacter* spp. are also able to utilise both pentose and hexose sugars for ethanol fermentation and are also able to hydrolyse xylan (Shaw et al., 2009). Similarly, several *Geobacilli* are also able to produce ethanol, among other organic compounds such as lactate and acetate, using a wide range of substrates such as glucose, xylose and arabinose, and are able to utilise short oligomers of the same, while some have been shown to be able to degrade more complex polymers such as xylan.

Despite the advantages associated with using thermophiles for biofuel production, there are some limitations that currently prevent an efficient, economically profitable process. High ethanol yields are typically lacking as thermophilic fermentation usually results in a mixture of products, such as other organic acids, which is effectively a waste of carbon utilisation. Furthermore, mixed acid production may also lead to retarded growth of the cell culture due to inhibitory activity and changes in pH. Other limitations include poor genetic accessibility, hindering the genetic manipulation of these organisms, including bacterial transformation which is due to both lack of reported techniques and barriers caused by the physical nature of the cell. Many thermophilic bacteria have been reported to have a robust cell envelope, and a weakly permeable cytoplasmic membrane (Silhavy et al., 2010). The lack of genetic toolkits has until recently limited the use of thermophilic bacteria in industrial processes. However, significant advances

have been made in the development of a genetic toolbox for some thermophilic bacteria, such as *Geobacillus* spp. A number of thermostable plasmids have been reported that allow the expression of both foreign and native genes in thermophilic hosts (Reeve et al., 2016). Furthermore, several plasmids have been developed that allow chromosomal interruption and insertion of genes (Reeve et al., 2016, Cripps et al., 2009, Taylor et al., 2008) and several thermostable antibiotic selection markers, counter-selection methods, and transformation protocols have also supported the manipulation of *Geobacilli* (Tominaga et al., 2016, Bosma et al., 2015, Kananaviciute and Citavicius, 2015, Blanchard et al., 2014, Daas et al., 2016).

1.4 GEOBACILLUS THERMOGLUCOSIDASIUS

Geobacillus thermoglucosidasius is a Gram-positive thermophilic, facultatively anaerobic, spore forming bacterium that was discovered to be able to metabolise both pentose and hexose sugar monomers and oligomers (Nazina et al., 2001). Furthermore, it is naturally able to produce valuable organic compounds such as ethanol and lactic acid making it a suitable candidate for industrial bio-ethanol production. The establishment of a genetic tool kit and transformation protocols made this organism genetically tractable and allowed metabolic engineering through over-expression of genes on the plasmid pUCG18 or creating insertions and deletions using pTMO31 (Taylor et al., 2008). TMO Renewables Ltd. have engineered this organism to maximise ethanol production by knocking out carbon-consuming pathways such as lactate dehydrogenase [LDH] and pyruvate formate lyase [PFL], and up-regulating the pyruvate dehydrogenase pathway [PDH] as seen in Figure 1.5.

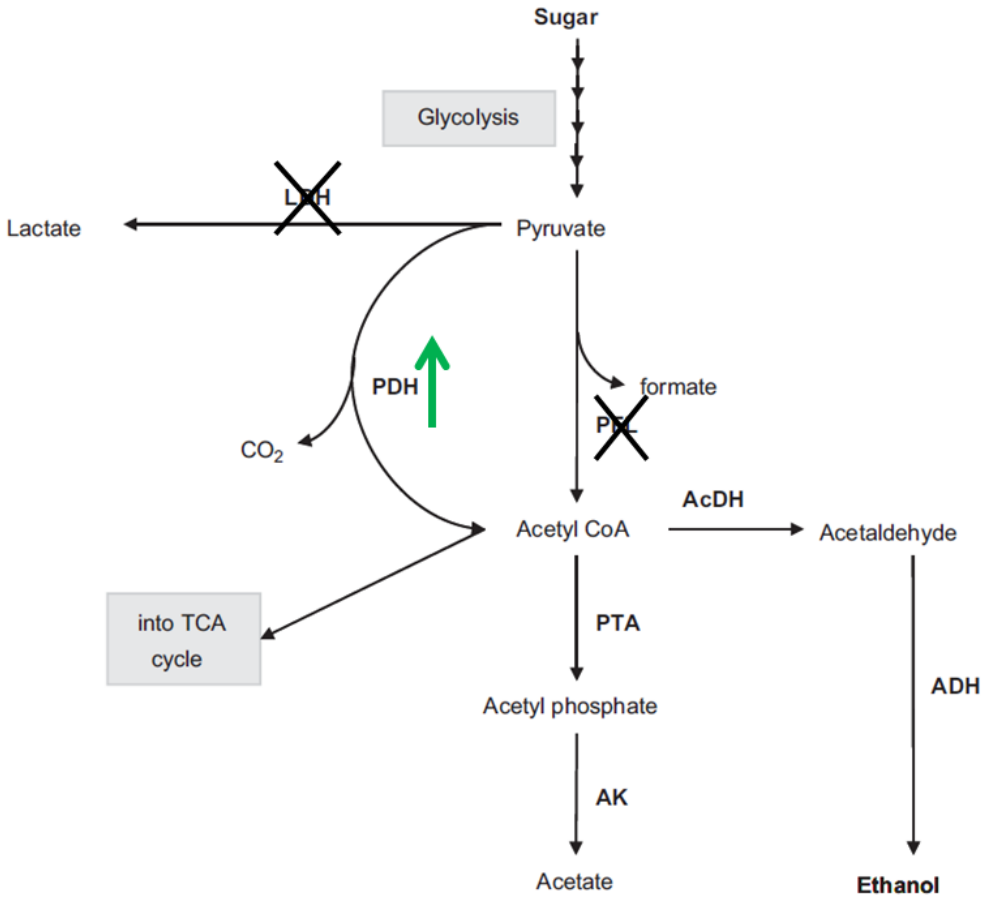

Figure 1.5: TM242 strain from TMO renewables. The genes encoding Lactate dehydrogenase (LDH) and Pyruvate formate lyase (PFL) have been knocked out, while those for the pyruvate dehydrogenase complex (PDH) have been up-regulated. Other enzymes shown are alcohol dehydrogenase (ADH), phosphate acetyltransferase (PTA) and acetate kinase (AK)

Other work is currently in progress to further optimise the fermentation process, such as identifying enzymes suitable for production towards the degradation of lignocellulosic biomass. One important optimization strategy would be to optimise the secretion of the enzymes that are used to degrade biomass.

1.5 PROTEIN SECRETION

Protein secretion is a process that is carried out in all living organisms. In eukaryotes, proteins are transported between both intracellular membranes and exported outside the cell. In prokaryotes, proteins are transported across the cell membrane, into the periplasm, cell wall or into the extracellular space. Prokaryotes have developed several systems of transporting protein cargo between locations, which fundamentally involve the assistance of dedicated protein secretion systems. In addition to several highly specialised transport mechanisms, prokaryotes contain two main systems for the general transport of proteins across the cytoplasmic membrane in bacteria, which are called the Sec and Tat pathways. These pathways are the most conserved mechanisms of protein secretion, and have been identified in all three domains of life (Papanikou et al., 2007, Robinson and Bolhuis, 2004).

Other specialized systems, especially in Gram-negative bacteria, have evolved to process the secretion of toxins or components of extracellular organelles such as flagella, across the outer membrane or across the entire cell envelope with no periplasmic intermediates. These specialised systems usually secrete only one or a few substrates; this is in contrast to the Sec and Tat systems, which are capable of secreting a wide variety of substrates.

Many industrial enzymes are produced in *B. subtilis* (BS) and its close relatives, for food, detergent, paper and research purposes due to a number of reasons. *BS* has the capacity to produce and secrete large quantities (20-25 g/L) of extracellular enzymes into the culture medium (Schallmey et al., 2004) and, as such, is regarded as a prolific cell factory for industrial enzymes and biopharmaceuticals. As a result, *BS* and protein secretion by *BS* are well described in the literature, and a great deal of research is being carried out to improve the organism in its use in microbial fermentations. The extensive literature and the close relation to *GT* (compared to *E. coli*) makes *BS* a good candidate with which to compare protein secretion in *GT.*

1.5.1 The Tat pathway

The Tat pathway is the alternative pathway, transporting mature, folded proteins across the cytoplasmic membrane and is found in bacteria, archaea and in chloroplasts. This pathway is utilised primarily for a subset of secretory proteins that are incompatible with the Sec pathway. Such reasons include: the protein has a co-factor that is incorporated during assembly within the cytoplasm, the substrate is only able to fold into its native conformation in the cytoplasm, or the kinetics of folding are too rapid resulting in a folded protein prior to exportation (Natale et al., 2008, Robinson and Bolhuis, 2004). Tat stands for twin arginine translocation, and is named as such due to the presence of twin arginine residues in the N-region of signal peptides (See section 4.4.1) targeted to the Tat machinery. The typical N-terminal twin-arginine sequence motif is S/T-R-R-X-F-L-K, where X is a polar amino acid. The core components of the Tat translocation machinery in Gram-positive bacteria are TatA and TatC, whereas in many Gram-negative bacteria a third component, TatB, is also critical for function (Palmer and Berks, 2012). Translocation is initiated once a cargo protein with the correct signal peptide interacts with the docking complex composed of TatC and TatA (Robinson and Bolhuis, 2004). The *B. subtilis* Tat machinery is composed only of TatA and TatC (Jongbloed et al., 2006) proteins although other Tat systems in other organisms may contain other components (Goosens et al., 2014). The Tat pathway will not be discussed in detail here as very few proteins in *BS,* and even fewer *GT,* are predicted to be translocated via this pathway.

1.5.2 The Sec Pathway

The major bacterial secretion pathway is the Sec pathway (de Keyzer et al., 2003, Tjalsma et al., 1998), which is involved in transporting proteins across the cytoplasmic membrane and into the surrounding extracellular milieu in an unfolded state. The Sec pathway is subdivided into co-translational secretion of proteins and post-translational secretion of proteins, both mediated by the recognition of N-terminal signal peptides that are recognised by different chaperones that mediate the targeting to the cytoplasmic membrane. The Sec machinery is involved in not only exporting secretory proteins, but also the translocation of transmembrane proteins, lipoproteins and cell

wall anchored surface proteins. The latter are characterized by the presence of a conserved N-terminal lipid-modified cysteine residue that allows the hydrophilic protein to anchor onto the bacterial cytoplasmic membrane by sortases (Paterson and Mitchell, 2004, Schneewind and Missiakas, 2012). Lipoproteins are anchored to membrane phospholipids, and are recognised and cleaved by type 2 signal peptidases.

The Sec machinery is composed of three main parts: the translocon channel, the motor, and the protein targeting component. Several other accessory components also play a crucial part in the protein secretion process, including cytoplasmic chaperones, signal peptidases, signal peptide peptidases, and folding factors (Figure 1.6).

1.5.3 Sec complex

The Sec complex comprises six main proteins. SecA is the motor component of the complex, which is an ATP-dependent protein that provides the energy to drive translocation through the SecYEG membrane pore (Lill et al., 1990). The SecYEG is a hetero-trimeric complex composed of SecY, E and G, which form an integral part of the hydrophilic pore that conducts secretory proteins and through which translocation occurs (Lycklama and Driessen, 2012).

SecYEG is essential, ubiquitous and conserved in all three domains of life and is located in the cytoplasmic membrane in bacteria or archaea, or the endoplasmic reticulum in eukaryotes (Osborne et al., 2005). SecY is the largest subunit of the translocation channel and it interacts with SecA, SecE and SecG. SecY forms a stable complex with SecE that does not dissociate *in vivo*. The association with SecE protects SecY from degradation by the membrane-bound protease FtsH that is involved in the degradation of unassembled membrane protein complexes (Akiyama et al., 1998, Kihara et al., 1995). SecG is not essential for protein translocation, but a knockout of the *secG* gene results in a cold-sensitive phenotype due to a reduced proton motive force (PMF) that is important for many cellular processes including protein translocation (van Wely et al., 1999).

The driving force for protein translocation is provided by ATP hydrolysis at SecA (Zimmer et al., 2008) and the PMF, which play a role at different stages of translocation. ATP is essential for the initiation of protein translocation. SecA is the central component of

the bacterial Sec system as it interacts with almost all other components of the translocase, and is classed as a molecular motor that drives protein translocation (Sianidis et al., 2001). SecA can interact with the membrane surface through two mechanisms whereby it can associate with low affinity with negatively-charged phospholipids at the cytoplasmic face of the cytoplasmic membrane (Lill et al., 1990), and can bind with high-affinity to the protein translocon (Hartl et al., 1990), binding of SecA is thought to prime the SecYEG channel for the arrival of a secretory protein (Li et al., 2016). SecA is not only located at the membrane, but is also found free in the cytoplasm, where it has a role in chaperoning and targeting secretory substrates from their site of synthesis to the Sec translocase (Chatzi et al., 2014a). It has recently been shown that successive rounds of ATP hydrolysis by SecA causes conformational changes in SecY causing the channel to open, and also directly bias the direction of polypeptide translocation in a so called 'Brownian ratchet' fashion (Allen et al., 2016). SecA has been shown to bind signal peptides as they emerge from the ribosome, and also to the mature domain, which has been shown to be involved in targeting, independent of their signal peptides (Gouridis et al., 2009).

SecDF is a membrane-integrated chaperone that is implicated in the final steps in translocation, promoting the release into the periplasm, and is driven via a PMF (Tsukazaki et al., 2011a, Tsukazaki et al., 2011b). SecDF has been shown to be required to maintain a high capacity for protein secretion. Unlike in *E. coli* and in archaea, where SecD and SecF are two distinct proteins, in *Bacillus* spp. the proteins are expressed as one protein (Bolhuis et al., 1998). In *E. coli*, the genes for SecD and SecF are co-transcribed with that of YajC. These three proteins do form a complex, but the role of YajC is not clear. In *B. subtilis*, YrbF is the functional homolog of YajC, but unlike in *E. coli,* the gene is not co-transcribed with SecDF (Tsukazaki et al., 2011b, Bolhuis et al., 1998).

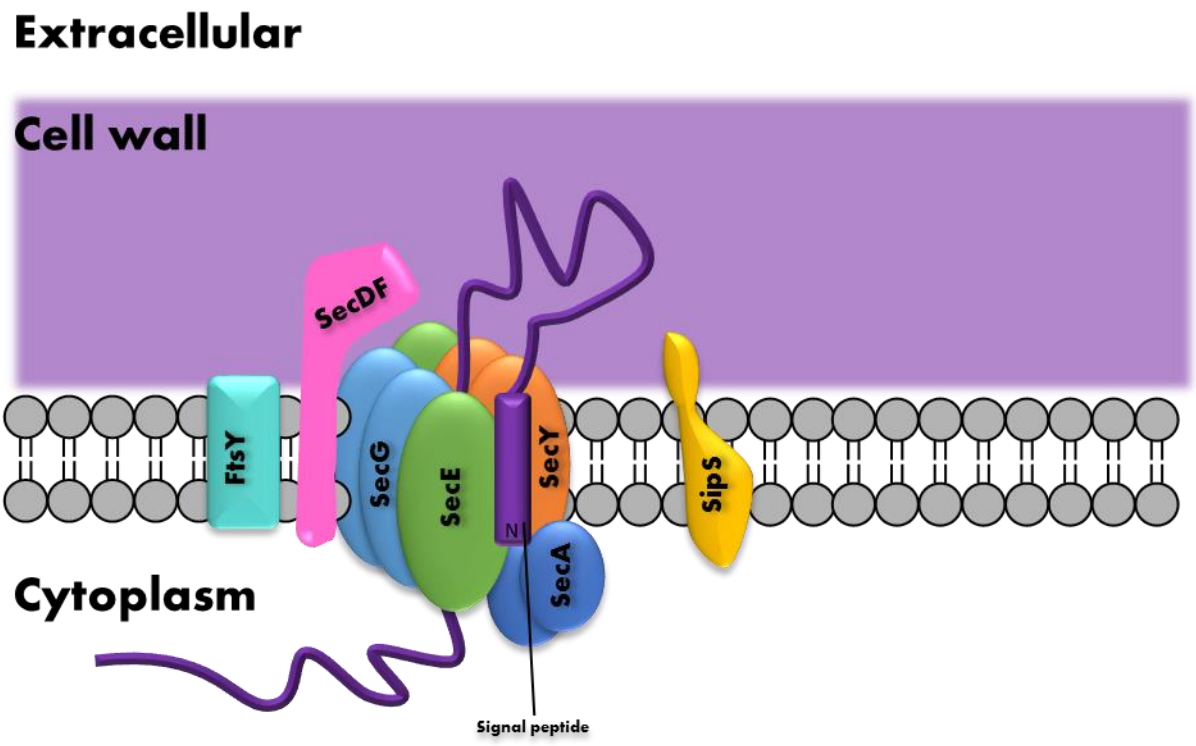

Figure 1.6: The Sec pathway machinery and accessory proteins with a secretory protein mid-translocation. The secretory protein (in purple) can be seen in the pore created by SecYEG. SecA binds to SecY, resulting in conformational changes and priming of the SecYEG channel for the arrival of a secretory protein. The signal peptide is inserted into the SecYEG channel as a hairpin loop and docks outside the lateral gate of SecY, with the N terminal end facing the cytoplasm. The signal peptide is then cleaved by a signal peptidase (yellow).

1.5.4 Signal peptides

In 1999, Gunter Blobel was awarded a Nobel prize for the discovery (in the 1970s) that proteins have intrinsic signals that govern their transport and localisation within the cell. Since his discovery, much has been revealed around the different pathways a protein can take for its translocation within and outside the cell. One class of targeting signals is the short, transient signal peptides at the N-terminus of proteins that are to be secreted. Signal peptides are required for the targeting of nascent pre-proteins to the secretion machinery at the cytoplasmic membrane, and the commencement of translocation across the membrane. They are generally composed of three characteristic domains, namely the positively-charged N-region, the hydrophobic H-region and the more polar C-region which is followed by a cleavage site (Vonheijne, 1990). They are cleaved by signal peptidases during, or shortly after the translocation through the secretion machinery.

The N-region is typically two to eight residues in length, with one or two positively charged residues such as arginine (R) or lysine (K). This domain is involved in targeting, although the exact mechanism in *BS* is still unclear, as the positively charged residues have been shown not to be strictly required for protein translocation (Chen and Nagarajan, 1994, Gennity et al., 1990). The N-region has been suggested to interact with the negatively charged lipid head groups of the cytoplasmic side of the cell membrane (Devrije et al., 1990, Deuerling et al., 1997), which is important for orientation of the signal peptide when embedded in the membrane, so the N-region is on the cytoplasmic side and not the extracellular side. It has also been shown to interact with the translocase, SecA (Akita et al., 1990, Bhanu et al., 2013). An increase in positive charge has been shown to improve the interaction with SecA which implies a direct link between the charged amino acids in the N-region and targeting to the translocon machinery (Akita et al., 1990).

The H-region, so named because of its hydrophobic nature, is the hydrophobic core of the signal peptide, which can be between 8 and 15 amino acids in length, and has been shown to form an α-helical structure within the cytoplasmic membrane (Briggs et al., 1986) to facilitate anchorage of the pre-protein to the secretion machinery. Furthermore, the H-region has been shown to be involved in targeting, through binding to the Signal Recognition Particle (SRP) that mediates the co-translational targeting pathway (Hatsuzawa et al., 1997, Goldstein et al., 1990). Insertion of the signal peptide into the membrane has been explained by an unlooping model, which proposes that the signal peptide forms a hairpin-like structure that is facilitated by α-helix destabilising amino acids in the middle of the H-region of the signal peptide (Shinde et al., 1989), and as it unloops, the signal peptide is inserted into the membrane, with the N-region on the cytoplasmic side of the membrane (Fekkes and Driessen, 1999). It has been shown that, when two cysteine residues are introduced into the signal peptide using mutagenesis, effectively inhibiting unlooping due to the formation of a disulphide bridge, translocation is hampered (Nouwen et al., 1994).

The third domain of the signal peptide, the C-region, is so named due to the presence of the cleavage site. The cleavage site is distinguished by the amino acids at the -1 and -3 position relative to the cleavage site. For proteins secreted via the Sec Pathway, type I

signal peptidases recognise and cleave the signal peptide from the mature sequence. The amino acid residues at these sites are normally residues with small and neutral side chains, such as alanine, glycine, serine and threonine, with a preference for alanine, giving rise to the A-X-A consensus sequence (Von Heijne, 1984, Tjalsma et al., 2000). However, this is the not the case for lipoprotein signal peptides, which are cleaved by type II signal peptidases and the consensus sequence for the cleavage site is L-A-G/A-C with the cysteine residue at the +1 position relative to the cleavage site. For both pre-lipoproteins and pre-proteins to be secreted, the position relative to the H-region is also significant, as the active site of the signal peptidase is located near the surface of the cytoplasmic membrane (Tjalsma et al., 1997, Pragai et al., 1997).

Signal peptides are different for different export pathways (as shown in Figure 1.7): the Tat pathway, the Sec pathway via SecA, the Sec pathway via the SRP, and lipoproteins (Sargent, 2001). Generally, the H-regions of Tat signal peptides are longer and less hydrophobic than that of Sec signal peptides (Cristobal et al., 1999) and signal peptides directed by the SRP are usually more hydrophobic and are sometimes uncleaved and remain in the membrane as an anchor.

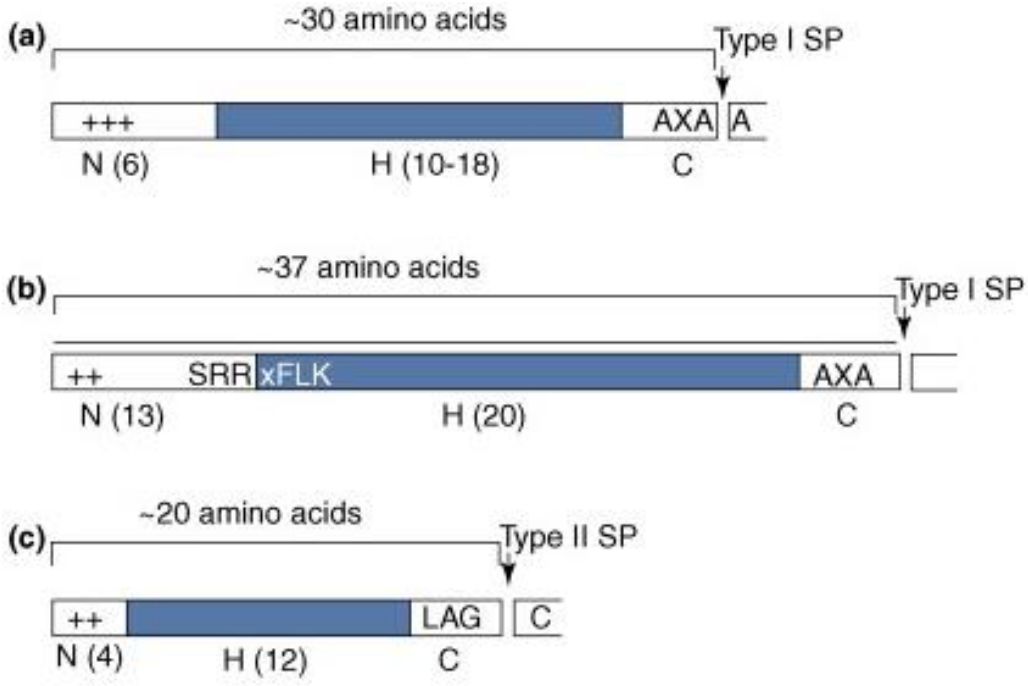

Figure 1.7: General features of the signal peptides of Bacillus secretory proteins. The N-terminal (N), hydrophobic (H) and cleavage (C) regions are identified by contrasting shading and their lengths (amino acid residues) are indicated in brackets. Cleavage sites are indicated by arrows. ***(a)*** *Sec-dependent signal peptide cleaved by a type I signal peptidase (SP) at the A-X-A cleavage site.* ***(b)*** *Tat-dependent signal peptide with a twin arginine motif (S-R-R-X-F-L-K), also cleaved by a type I SP.* ***(c)*** *Lipoprotein signal peptide cleaved by the type II SP. Image adapted from (Harwood and Cranenburgh, 2008)*

1.5.5 Signal peptidases

Signal peptidases (SPases) are a class of proteases that cleave the signal peptide from the secretory pre-proteins, releasing the mature domain of secretory proteins from the cytoplasmic membrane into the cell wall and extracellular milieu. There are two known classes of SPases: type I which process secretory protein type SPs, and type II which process lipoprotein type SPs. SPases process and remove signal peptides from pre-proteins when the C-domain of the signal peptide emerges at the extra-cytoplasmic side of the membrane. The signal peptide cleavage site specificity is often designated A-X-A rule due to the presence of Alanine at the −3 and −1 position relative to the cleavage site. Despite having no other apparent consensus sequences, signal peptides are recognized by SPase I with high fidelity.

In *BS*, seven type 1 signal peptidase genes have been identified, *sipS*, *sipT*, *sipU*, *sipV*, and *sipW,* on the chromosome of *BS* and a further two *sipP* genes have been found on plasmids identified in *natto* producing strains of *BS* (Tjalsma et al., 1998) (a type of Japanese food made from soybeans fermented with *BS).* However, only SipS and SipT are of major importance for secretory pre-protein processing and cell viability, and the other SPases play a minor role, and have different substrate specificities (Antelmann et al., 2001, Bron et al., 1998). Multiple type I SPases are also found in other prokaryotes such as *Archaeoglobus fulgidus*, *B. japonicum*, and *B. amyloliquefaciens*. In contrast, several other bacteria, such as *E. coli, Helicobacter pylori* and *Mycobacterium* only contain one solitary type I SPase gene, which is the case for most other bacteria (Tuteja, 2005).

Type II signal peptidases are signal peptidases that specifically process and cleave signal peptides from lipoproteins. *BS* contains only one gene for a type II SPase, *lspA,* which is specifically required for the processing of lipid-modified pre-proteins. However, strains in which *lspA* has been inactivated are still viable under laboratory conditions. This indicates that *lspA* is not strictly required for lipoprotein function, as at least one known lipoprotein, PrsA, is required for cell viability (Kontinen and Sarvas, 1993).

1.5.6 Signal peptide peptidases

Once the signal peptide has been cleaved and the pre-protein released from the translocation complex, the signal peptide is then rapidly degraded by signal peptide proteases. Hussain et al. (Hussain et al., 1982) were the first to identify SppA as an enzyme involved in signal peptide digestion when they observed, in an *in vitro* experiment, that *E. coli* lipoprotein signal peptides were digested upon the addition of a membrane extract containing SppA. Bolhuis et al. (Bolhuis et al., 1999a) were the first to report an SppA from *B. subtilis.*

1.5.7 Molecular chaperones

As nascent polypeptides emerge from the ribosome as they are being transcribed, they are often assisted by a class of proteins known as molecular chaperones, to facilitate protein folding and targeting to their specific sites such as the cytoplasm or membrane. These chaperones are proteins that catalyse protein folding and assist in the construction or assembly of multi-protein complexes (Wild et al., 1992, Schroder et al., 1993, Kusukawa et al., 1989). They inhibit aggregation by binding to exposed hydrophobic patches, preventing the formation of non-functional inclusion bodies. Some may also play a role in rescuing and refolding of misfolded polypeptide chains. Most proteins intended for translocation can only be translocated in a translocation-competent state, which is they are relatively unfolded, or bound to chaperones to prevent misfolding or aggregation. Some chaperones are secretion-dedicated, while others are general chaperones that assist in folding of many types of proteins but also have a role in protein secretion.

Secretion-dedicated chaperones in bacteria include SecB, for which a homologue is not found in *BS* or other Gram-positive bacteria. SecB facilitates protein translocation in *E. coli* by binding to unfolded precursor protein, and maintains them in a translocation-competent state, for delivery to the translocon where it interacts with SecA. *E. coli* SecB binds to the mature region of SecB-dependent pre-secretory proteins. The resulting binary complex interacts with a specific site within the C-terminal region of SecA to form a tertiary complex that, in turn, interacts with the membrane-located secretory translocase. Conformational changes that result from the interaction of the tertiary

complex with the secretory translocase lead to the release and recycling of SecB. SecB-dependent substrates have been identified, and heterologous production of *E. coli* SecB has been shown to facilitate secretion of some heterologously produced SecB-dependent proteins in *BS* (Collier, 1994). In bacteria, the SecB-binding domain of SecA is located at the C-terminus of SecA. The SecB-binding domain of *E. coli* SecA is highly conserved in the SecA protein of *B. subtilis*. This binding domain could possibly function as a docking site for another SecB analogue (Fekkes et al., 1997). Another study has shown that replacing the C-terminal of the *BS* SecA protein with that of *E. coli* facilitates binding of SecA to SecB, and when co-expressed, result in functional implementation of the SecA-SecB post-translational secretion of heterologous SecB dependent *E. coli* proteins in *BS* (Diao et al., 2012).

In the absence of SecB in *BS* and other Gram-positive bacteria, CsaA is a good candidate for a SecB analogue in *BS.* It has been demonstrated that CsaA has chaperone-like activity in *BS* (Muller et al., 2000a) and that CsaA has an affinity for the SecA translocase and pre-proteins, which strongly suggests that CsaA has a secretion-related function in *BS*. However, CsaA does not seem to bind to the conserved SecB-binding domain in SecA, and therefore the exact role of CsaA in protein secretion in *BS* remains to be elucidated.

Another secretion-dedicated chaperone is the Ffh protein (Fifty four homologue), which is the only secretion-specific protein found in *BS* and other Gram-positive bacteria to date. As the name suggests, Ffh is homologous to the 54kDa subunit, which is an essential part of the signal recognition particle (SRP) which is a ribonucleoprotein complex. The SRP is involved in co-translational targeting in protein secretion in both prokaryotes and eukaryotes (Zanen et al., 2006b). The SRP is a complex composed of protein and RNA and, although the function is analogous in all organisms, the composition of the complex varies greatly. In prokaryotes, one polypeptide chain is bound to one RNA molecule. In eukaryotes, there are 6 polypeptide chains and one RNA molecule. The protein chain in the prokaryotic version is known as Ffh and is crucial to binding of the targeting signals. The SRP binds the signal peptide at the N-terminus of the nascent peptide as it emerges from the ribosome. This forms a complex that is known as the ribosome nascent chain (RNC) complex, which then in turn interacts with a membrane bound SRP receptor FtsY (Angelini et al., 2005). In eukaryotic organisms,

the Alu domain in the SRP domain causes elongation arrest by blocking the elongation factor entry site and thus prevents membrane proteins from being prematurely released from the ribosome before the RNC has docked at the translocation machinery at the endoplasmic reticulum membrane. This elongation arrest was previously not thought to occur in prokaryotes, but recent studies in the field have shown that the *Alu* domain is indeed present in the RNA component of the SRP in prokaryotes, suggesting that elongation arrest may indeed occur during translation of membrane or secretory proteins in prokaryotes (Kempf et al., 2014, Beckert et al., 2015). However, it must be noted that in *E. coli,* while many inner membrane proteins are targeted via the SRP, only a small number of secretory proteins are dependent on this pathway (Huber et al., 2005).

Recently, SecA has been thought to play a much larger role in protein secretion than originally understood. SecA has been shown to bind not only the translocon machinery SecYEG and the chaperone SecB, but also to the ribosome, signal peptide sequences, and mature domain sites of pre-proteins (Huber et al., 2011, Huber et al., 2017, Wu et al., 2012). The SRP has a low cellular concentration relative to SecA and is extremely low in stoichiometry compared to ribosomes. The SRP has a very high affinity for nascent hydrophobic transmembrane sequences and highly hydrophobic signal peptides (Grudnik et al., 2009, Zhang et al., 2010). It is thought that, due to high affinity, the SRP is likely to binds its substrates first which would result in the sequestering of those proteins away from the post-translational secretion pathway as they would be obscured from post-translational chaperones, which would prevent SecA from binding proteins targeted to the co-translational pathway.

General chaperones in *BS* include GroEL, GroES, DnaK, DnaJ, GprE, and trigger factor. GroEL and GroES are homologues of eukaryotic Hsp60 and Hsp10, respectively. In *E. coli,* it has been shown that a subset of proteins are dependent on GroEL for effective translocation (Kusukawa et al., 1989) and it has been suggested that GroEL interacts with SecA (Bochkareva et al., 1998), although a defined role in protein secretion in *BS* or other Gram-positive organisms has not been elucidated. DnaK and DnaJ are homologues of the eukaryotic Hsp70 and Hsp40, respectively. These two chaperones work together with another chaperone known as GprE to mitigate stress-induced protein damage. In

E. coli, the trio have also been shown to be involved in the secretion of several SecB-independent proteins (Wild et al., 1992, Schroder et al., 1993) and some Tat pathway-dependent proteins (Perez-Rodriguez et al., 2007).

Trigger factor is a *cis-trans* proline isomerase that scans the nascent proteins when bound to the ribosome, and interacts with both cytoplasmic proteins and secretory proteins. The SRP is proposed to compete with trigger factor for binding of the signal sequence domain of the nascent chain (Hesterkamp et al., 1996) and has been found to retard protein export in *E. coli*, as interruption of the gene results in improved protein secretion (Lee and Bernstein, 2002) .

1.5.8 Extracellular proteases and chaperones

Once translocation has terminated, the secretory protein then finds itself on the extracellular side of the cell membrane and in the cell wall where it then has to fold into its native conformation. Here, the secretory proteins also encounter several extracellular proteases to which an unfolded protein is susceptible to degradation. As such, protein folding must occur rapidly and correctly, lest the secreted protein be degraded. Folding can occur spontaneously, or require the help of folding catalysts or chaperones. *BS* secretes high levels of extracellular proteases into the cell wall and extracellular milieu, to enable the degradation of misfolded or aggregated secreted proteins. These "quality-control" proteases include HtrA, HtrB and WprA; these proteases alleviate secretion stress, which occurs when proteins misfold or aggregate and accumulate at the cytoplasmic membrane – cell wall interface. *BS* also secretes numerous feeding proteases (to obtain nutrients from the environment), namely NprB, AprE, Epr, Bpr, NprE, Mpr and VprA, all of which contribute toward proteolytic degradation of extracellular proteins, native or heterologous, with the latter being especially susceptible to degradation.

In *BS*, a two-component system (CssRS) comprising CssR and CssS (Control of secretion stress Regulator and Sensor) performs an essential role in the response to secretion stress. The CssRS system, when stimulated by secretion stress, upregulates membrane-bound serine proteases, HtrA and HtrB, with the active sites located in the cell wall (Westers et al., 2006, Gullon et al., 2012). HtrA-type proteins have also been found to possess chaperone-like activity and are implicated in quality control of secretory proteins as well as the protein degradative role (Malet et al., 2012). HtrA has also been found in the extracellular milieu of *BS*, not bound to the cell wall, and not together with HtrB, which suggests that HtrA may have some other role in the extracellular milieu (Antelmann et al., 2003).

Another extracellular protease involved in extra-cytoplasmic protein quality control in *BS* is WprA, a cell-wall-bound protease. WprA has been shown to be processed into two separate cell wall proteins, one with a serine protease domain, and the other with putative chaperone-like activity (Stephenson and Harwood, 1998, Babe and Schmidt, 1998, Margot and Karamata, 1996).

One of the most well described extracellular protein folding factors in *BS* is PrsA, a lipoprotein anchored to the cytoplasmic membrane. PrsA has been shown to be essential for cell viability, and reduced levels of PrsA have been shown to result in increased degradation of a subset of proteins, thought to be PrsA dependent (Kontinen et al., 1991, Jacobs et al., 1993). Furthermore, PrsA shows sequence similarity to peptidyl-prolyl cis–trans isomerases (PPIases) of the parvulin family (Vitikainen et al., 2004, Tossavainen et al., 2006) which increase the rate of folding of proteins with cis-prolyl residues, which is consistent with the role of PrsA in assisting the folding of secreted proteins and reducing their susceptibility to proteolysis.

Four extra-cytoplasmic thiol-disulphide oxidoreductases, BdbA, BdbB, BdbC and BdbD, are another type of folding catalyst that have been implicated in the formation of disulphide bonds in exported proteins in *BS* (Bolhuis et al., 1999c). These proteins catalyse disulphide bond formation, and are thought to promote extra-cytoplasmic protein folding. However, disruptions in one or all four of the *bdb* genes in *BS* do not result in any significant change to the extracellular proteome of *BS,* suggesting that their activity is not critical to protein folding of natively secreted proteins.

1.6 POTENTIAL BOTTLENECKS IN PROTEIN SECRETION

In the context of this research, which is to investigate potential bottlenecks in secretion of glycosyl hydrolases, proteins can be either heterologous, or over-expressed and over-produced native proteins. Bottlenecks can occur at any stage of protein secretion, from the transcription level through to the extracellular milieu. These can be briefly broken down into the following categories: gene transcription, protein translation, protein targeting, translocation across the membrane, signal peptide processing, and extracellular folding and proteolysis.

Regulation of gene expression is controlled at the transcription level and expression levels are determined by a number of factors such as the type of promoter, sigma factor, gene copy number and other transcription factors. Codon harmonisation of the target gene sequence may also improve translation of heterologously produced proteins, as the speed of translation is linked to the rate of folding, and could have a link with chaperone binding as the nascent chain emerges from the ribosome (Angov et al., 2008, Welch et al., 2011).

Protein translation bottlenecks occur at the ribosome; for example, secretory proteins need to be in a translocation competent state, which is devoid of tight folding and which is facilitated by intracellular chaperones. Heterologous proteins may form insoluble aggregates in the cytoplasm due to limited activity of intracellular molecular chaperones. For heterologous protein production, an increased level of endogenous molecular chaperones has been shown help to increase heterologous protein production and secretion in *BS* (Wu et al., 1998).

Targeting of the protein to the translocation machinery is directed by the signal peptide, to which targeting chaperones such as SecB, Ffh or CsaA will bind and direct to the translocation machinery. Extensive work in *BS* has shown that there is no 'one size fits all' signal peptide for optimum secretion of proteins. Furthermore, the mechanism of the relationship between signal peptide and mature protein sequence is still poorly understood. It has been proposed that the N-terminus of the mature protein and the C-region of a signal peptide have co-evolved as a 'signal peptide-mature protein' junction. However, signal peptide libraries have been constructed containing signal peptides from

different organisms, and successfully used to screen for optimal protein secretion efficiency of desired protein (Brockmeier et al., 2006, Degering et al., 2010, Hemmerich et al., 2016). It has also been shown that mutations in the different domains of the signal peptide have also improved secretion efficiency, such as increased hydrophobicity of the H-region, or increased positively charged residues in the N-region of the signal peptide (Goldstein et al., 1990, Caspers et al., 2010, Low et al., 2012, Jonet et al., 2012, Low et al., 2013). It has also been shown that codon optimisation of the signal peptide sequences can also enhance targeting of heterologously produced proteins (Humphreys et al., 2000)

The actual translocation across the cell membrane into the extracellular milieu is carried out at the Sec translocon at the membrane. As secretory proteins are over-produced, it is feasible that overexpression of secreted proteins can cause congestion at the membrane due to the shortfall of Sec pathway components. This ultimately can result in the proteins being degraded, and thus a waste of energy in producing them in the first place. Jamming of the translocation machinery may also result in gross growth defects. Increased expression of the SecYEG genes have been shown to improve heterologous secretory protein translocation (Mulder et al., 2013, Chen et al., 2015b). Translocation is terminated when the signal peptide is processed and cleaved by type I signal peptidase for secretory proteins. It has been shown that when secretory proteins are over-produced, the rate-limiting factor can be the rate of processing by signal peptidases (Malten et al., 2005) and over expression of signal peptidase genes can result in improved protein secretion (Bolhuis et al., 1996).

Finally, once translocation across the membrane has occurred, the protein must then fold in the extra cytoplasmic space, the cell wall. Over-production of the lipoprotein PrsA has been shown to improve protein secretion of both native, and heterologous proteins (Chen et al., 2015b, Chen et al., 2015c, Vitikainen et al., 2005, Vitikainen et al., 2001, Wu et al., 1998) while in *prsA* mutants, the secretion and stability of some model proteins has been shown to be hampered. Furthermore, if folding occurs incorrectly, or too slowly, secretory proteins, especially heterologous proteins, are susceptible to proteolysis by quality control proteases. Work carried out in *BS*, in the creation of multiple-protease deficient strains, including a strain lacking eight extracellular

proteases, has highlighted the negative effect proteases can have on over-production of secretory proteins (Wu et al., 2002b). Even over-production of native proteins was shown to be improved in protease-deficient strains (Wu et al., 1991b)

1.7 PROJECT AIMS

As mentioned earlier, TMO Renewables Ltd. have developed and modified *G. thermoglucosidasius* (TM242) to produce ethanol from lignocellulosic feedstocks, with the aim to utilise waste materials to generate bioethanol as a renewable fuel source. The most expensive step during this process is the enzyme pre-treatment step, which uses commercial enzymes to hydrolyse polymers in the lignocellulosic material prior to fermentation. One objective of this project is to engineer TM242 to secrete those enzymes necessary to break down the recalcitrant polymers, increasing the efficiency and thus reducing the cost of the overall process. To achieve this, it is vital that the secretion pathways and their kinetics are better understood, and that the effects of over-production of secretory enzymes on the secretion machinery, growth and overall ethanol yields are addressed and analysed. The reason for this is that it is not simply a straightforward matter of placing a gene encoding a secretory protein behind a strong promoter and then expecting good secretion. Depending on the signal peptide/protein combination used, different bottlenecks can be encountered in protein transport, including in the early stages (targeting to the membrane), middle stages (translocation through the membrane) or late stages (release from the membrane). All these stages are potential bottlenecks, as shown in *Bacillus subtilis* for example (Bolhuis et al., 1999b). Importantly, such bottlenecks are particularly problematic with heterologous proteins. Improving our knowledge of secretion pathways and understanding the potential bottlenecks may thus provide information that can be used to improve and maximise the secretion potential. It is also important to note that *G. thermoglucosidasius* is a thermophile and the process of protein translocation may differ from that of *B. subtilis*, such as in the composition of the translocation machinery, the kinetics of translocation, and the composition of signal peptides. This is why this study is important for both a fundamental understanding of secretion in *G. thermoglucosidasius* and its application in the production of bioethanol. *G. thermoglucosidasius* is an ideal candidate for thermophilic secretion studies because it is a moderate thermophile and is genetically amenable, unlike some other extreme thermophiles that are more difficult to grow and manipulate.

The project is broken down into a number of aims that are listed below. The first three of these focus on fundamental aspects of protein translocation in *G. thermo*glucosidasius, whereas the last two are aimed more at the application of *G. thermoglucosidasius* in the production of bioethanol.

- Investigate and characterize any adaptations in secretion in the thermophile *G. thermoglucosidasius*, in particular in comparison to knowledge available on *B. subtilis*, which is a well-described mesophilic relative.
- Analysis of the kinetics of protein translocation and identification of rate-limiting steps. To this purpose a model enzyme will be overproduced, which will facilitate the identification of bottlenecks in the secretion process.
- Analysis of the effects of overproduction of the model enzyme on secretion of other proteins.
- Based upon information from the previous aims and knowledge of protein translocation in other bacteria, strategies will be designed to optimise levels of protein secretion of hydrolases.

CHAPTER TWO: METHODS AND MATERIALS

2.1 MEDIA AND STRAINS

Table 2.1: List of strains used in this study

Geobacillus thermoglucosidasius C56-YS93	Originally supplied by Dr David Mead from Lucigen Corporation. This strain has its genome sequence publicly available.
Geobacillus thermoglucosidasius TM242	(Cripps et al., 2009) supplied by TMO Renewables Ltd. This strain is the ldhA−pfl−P_ldh/pdhup variant of G. thermoglucosidasius NCIMB 11955 described in Chapter 1
E. coli Neb5α competent cells	(New England Biolabs, UK) This strain is a non-expression host for general purpose cloning and plasmid propagation as it is endonuclease (endA) and recombinase (recA) deficient. Resistant to phage T1 (*fhuA2)*
Geobacillus thermoglucosidasius WT11955	(Cripps et al., 2009) supplied by TMO Renewables Ltd. This strain is the wild-type variant of the TM242 working strain.
E. coli JM109	(Promega, Southampton, UK). This strain is a non-expression host for general purpose cloning and plasmid propagation as it is endonuclease (endA) and recombinase (recA) deficient, which ensures DNA stability and results in high-quality plasmid. These cells are also deficient in β-galactosidase activity due to deletions in both genomic and episomal copies of the lacZ gene.
Chemically competent *E. coli* BL21(DE3)	(Novagen®, Merck Millipore, Watford, UK). This strain is a general-purpose expression host as it is deficient in proteases (Ion and ompT) to favour protein expression. It possesses a lysogen of bacteriophage DE3 and it contains the gene for T7 RNA polymerase under control of the lacUV5 promoter.

2.2 BACTERIAL GROWTH MEDIA

Media components were dissolved in distilled water and were sterilised by either autoclaving at 121 °C for 20 min, or using 0.22 µm Steritop® filter units (Merck Millipore, Darmstadt, Germany). Solid plates of the various liquid media were made by adding 1.5% (w/v) agar prior to autoclaving. Where required, media were supplemented with antibiotics at the following concentrations: 30 µg/ml kanamycin (12.5 µg/ml for Geobacillus strains carrying plasmids with kanamycin resistance markers), 100 µg/ml ampicillin. All media used for *G. thermoglucosidasius* cultures were pre-warmed to 60 °C prior to inoculation.

2.2.1 Tryptone Glycerol Peptone (TGP) media

16 g/l tryptone, 10 g/l Soy peptone, 5 g/l NaCl, 4 g/l Sodium Pyruvate, 0.4 % (v/v) glycerol

2.2.2 Lysogeny Broth (LB)

LB consisted of 1 % (w/v) NaCl (Fisher Scientific, Loughborough, UK), 1 % (w/v) tryptone and 0.5 % (w/v) yeast extract.

2.2.3 Super Optimal broth with Catabolite repression (SOC)

SOC consisted of 2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 10 mM NaCl and 2.5 mM KCl. This solution was autoclaved and allowed to cool before adding 100× concentrated forms of filter sterilised magnesium and glucose solutions to final concentrations of 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose.

2.2.4 Soy Peptone Yeast Extract (No Glycerol) (2SPYNG)

2SPYNG consisted of 1.6 % (w/v) soy peptone , 1 % (w/v) yeast extract, and 0.5 % (w/v) NaCl, adjusted to pH 7.0 with 10M KOH. 2SPY medium is 2SPYNG with the addition of 1.0 % (w/v) glycerol.

2.2.5 Tryptone Soya Broth (TS)

TS Broth (Oxoid, Basingstoke, UK) was purchased pre-prepared as a dehydrated medium containing glucose that was dissolved as recommended, 30 g/L. This results in a final glucose concentration of 2.5 % (w/v)

2.2.6 Ammonium salts medium (ASM)

ASM consisted of 0.5 % (w/v) yeast extract (Oxoid, Basingstoke, UK), 25 mM NaH2PO4, 25 mM $(NH_4)_2SO_4$, 25 mM K2SO4, 5 mM citric acid, 3.125 mM MgSO4, 50 µM CaCl2, 2.5 µM Na2MoO4, 2.5 µM biotin, 2.5 µM thiamine and 12.5 ml/L of trace elements, pH 6.7. The medium was buffered with KOH to pH 7. Xylose and glucose were added at concentrations of 0.5 % (w/v) each.

2.2.7 Trace Elements

The trace elements solution consisted of 60 mM H2SO4, 0.144 % (w/v) ZnSO4.7H2O, 0.556 % (w/v) FeSO4.7H2O, 0.169 % (w/v) MnSO4.H2O, 0.025 % (w/v) CuSO4.5H2O, 0.0562 % (w/v) CoSO4.7H2O, 0.006 % (w/v) H3BO3, and 0.0886 % (w/v) NiSO4.6H2O, dissolved in MilliQ water.

2.2.8 Glycerol stocks

Strains were stored long-term in glycerol suspensions at -80 °C in cryogenic vials. E. coli glycerol stocks were made by mixing healthy cultures and sterile 70 % (w/v) glycerol to a final concentration of 20 % (v/v) glycerol. *G. thermoglucosidasius* glycerol stocks were made by mixing healthy cultures and sterile 70 % (w/v) glycerol to a final concentration of 25 % (v/v) glycerol. Glycerol stocks were snap-frozen in liquid nitrogen and immediately transferred to storage at -80 °C.

2.3 GROWTH CONDITIONS

2.3.1 E. coli

E. coli strains were cultured in either liquid or solid media. Liquid media was inoculated using a scraping from either colony or frozen glycerol stock.

2.3.2 Geobacillus thermoglucosidasius

Overnight cultures were prepared by inoculating liquid media from glycerol freezer stocks, followed by plating out onto agar plates. Single colonies were then used to inoculate liquid media such as TGP or 2SPYNG.

For aerobic growth, 50 ml media in 250 ml baffled conical flask were inoculated with 500 µl overnight culture (1 in 100). ASM media was inoculated by scraping an agar plate with confluent growth and transferring to 20ml of pre-warmed media. This was then diluted to OD600=0.1 for aerobic growth.

2.3.3 Quantification of bacterial cell density

Bacterial cell culture samples were diluted in growth media depending on the growth stage, and the optical density at 600 nm (OD600nm) analysed in a cuvette using an Eppendorf Biophotometer. The reading was corrected against growth medium as the background. For *GT,* an OD of 1 correlated to 0.25 g/L dry weight (Taylor 2008).

2.4 OPTIMISATION OF HETEROLOGOUS EXPRESSION AND PURIFICATION OF XYLANASE IN *E. COLI*

2.4.1 Heterologous expression

E. coli BL21(DE3) cells containing pET28a-β-xylanase1 were grown in 5ml Luria Bertani [LB] media overnight at 37 ˚C, which was then used to inoculate 500ml LB. The 500ml culture was incubated at 37 ˚C in a shaking 250 rpm incubator until the OD_{600} was approximately 0.6, at which point the culture was induced with IPTG to 100 µM and cultured for a further 3 hours. The cells were harvested by centrifugation at 7000 x *g* and stored at -20 ˚C until required.

2.4.2 Cell lysis

The frozen cell pellets were re-suspended in re-suspension buffer [20mM sodium phosphate buffer, 0.5 M NaCl, 500 mM Imidazole, Roche protease inhibitor EDTA free]. Lysozyme was added to 0.2 mg/ml, triton x-100 to 0.1 % and incubated for approximately 5 minutes. The cell suspension was then sonicated on ice until the cell lysate was clear and free flowing. At this step, the cell lysate is representative of the total cell protein content (T) and a sample taken for SDS-PAGE. The cell lysate was then centrifuged at 7000 x *g* to remove all insoluble protein. The clarified cell lysate

containing the soluble fraction (S) was then placed on ice in preparation for column loading.

2.4.3 Ni-NTA affinity purification using FPLC

A 1ml Hi-Trap Chelating HP was charged with 0.1 M $NiSO_4$ and then equilibrated with low imidazole buffer [20 mM Tris-HCl, 0.5 M NaCl, 10mM Imidazole] followed by high imidazole buffer [20 mM Tris-HCl, 0.5 M NaCl, 500mM Imidazole], followed by low imidazole buffer. The clarified cell lysate was then loaded onto the column and washed with 10 column volumes [cv] of low imidazole buffer. The protein was then eluted over a 30 ml imidazole gradient and collected in 1 ml fractions. The fractions corresponding to the chromatogram peaks were analysed by SDS-PAGE and the fractions with the highest yields were pooled in preparation for the next step.

2.4.4 Optimisation of Ion exchange chromatography using FPLC

2.4.4.1 Anion exchange chromatography

A 1 ml Q-sepharose column was equilibrated with Low salt buffer (10 mM Tris-HCl, 50 mM NaCl, pH 8) followed by high-salt (10 mM Tris-HCl, 4 M NaCl, pH 8) followed by low salt again. The pH was selected based on the theoretical pI of 6.25 as calculated using ExPASy ProtParam. The pooled protein from the affinity chromatography was then loaded onto the column and eluted over a 30 ml salt gradient.

2.4.4.2 Cation exchange chromatography

A 1 ml SP-sepharose column was equilibrated with Low salt buffer (10 mM NaOAc, 50 mM NaCl, pH 6) followed by high-salt (10 mM NaOAc, 4 M NaCl, pH 6). The pooled protein from the affinity chromatography was then loaded onto the column and eluted over a 30 ml salt gradient.

2.5 MOLECULAR BIOLOGY

2.5.1 Plasmid purification

E. coli strains containing plasmid were cultured overnight at 37˚C in LB with 20 µg/ml kanamycin sulphate. The cultures were then pelleted and a plasmid prep kit (Machery Nagel) was used according to the manufacturer's instructions and the plasmid DNA eluted in 50 µl of elution buffer.

2.5.2 Chromosomal DNA extraction

Chromosomal DNA from *GT* was extracted by the phenol-chloroform extraction method as previously described (Shankar et al. 1999) with some modifications. 5ml overnight cultures of *GT* in TGP or TSB were pelleted by centrifugation, washed, and re-suspended in 1 ml TES (50 mM NaCl, 100 mM Tris, 70 mM EDTA) containing 25 % sucrose. This was then incubated with 1 mg/ml lysozyme for 30 minutes at 37 °C. This was followed by incubation with 0.2 mg/ml proteinase K and 0.5 % (w/v) SDS at 55 °C. DNA was then extracted from the cell lysate by adding one volume of phenol - chloroform - isoamylalcohol (25:24:1), followed by mixing and centrifugation for 5 min. The aqueous top layer, containing the DNA fraction, was removed and extracted twice further. The aqueous fraction was then extracted with chloroform – isoamylalcohol twice to remove phenol. The DNA was precipitated with the addition of 0.1 volume of 3 M NaOAc. The DNA pellets were washed with 80 % ethanol and re-suspended in TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 8).

2.5.3 Polymerase chain reaction

To amplify the genes of interest, PCR reactions were performed using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) or Q5 High-Fidelity DNA polymerase (New England Biolabs) following the suppliers recommended protocol. A standard PCR reaction (50 µL) contained 1x HF Phusion buffer, 200 µM dNTPs (dATP, dCTP, dGTP, dTTP) (Promega), 0.25 µM of each forward and reverse specific primers, an appropriate amount of DNA template (1 pg–10 ng for plasmid DNA and 50–250 ng for genomic DNA), and 0.5 U PhusionHF DNA polymerase. The reactions were carried out in thin-walled PCR tubes placed into an Eppendorf Mastercycler® gradient PCR thermocycler (Eppendorf,

Cambridge, UK), following the programme recommended by the manufacturer. Annealing temperatures were calculated using the NEB Tm calculator tool (http://tmcalculator.neb.com/#!/)

For colony PCR, instead of purified DNA, 10 µl of crude cell extract was added to a total of 50 µl reaction. Crude cell extract was obtained by re-suspending a single colony in 50 µl of H2O, boiling for 5 minutes and centrifuged to remove cell debris.

2.5.4 Restriction digest

Restriction enzyme digests were performed as recommended by the manufacturer, New England BioLabs. Typical reactions were carried out in 50 µl volumes as per the manufacturer's protocol. Double digests were performed in the manufacturer's suggested buffer. Where appropriate, digestion was halted using heat-treatment and alkaline phosphatase (AP) was used to inhibit self-ligation of vectors. After digestion, DNA was purified using PCR clean-up gel extraction kit, NucleoSpin Gel and PCR clean-up (Macherey-Nagel).

2.5.5 Ligation reactions

DNA was ligated using T4 DNA ligase as per the manufacturer's instructions (Thermo).

2.5.6 Transformation of chemically competent *E. coli* cells

E. coli JM109 or NEB-5α cells were transformed by heat shock. 20 µl of cells were incubated on ice with either 100 ng of purified plasmid DNA or 2 µl ligation mixture for >30 min. Samples were then heat shocked at 42 °C for 45 s in a heat block or water bath, before being incubated on ice for >2 min. 1 ml of SOC medium was added and cells were incubated at 37 °C for 1 h before 50 µl of transformants were spread on LB plates containing appropriate antibiotics. Plates were incubated at 37 °C overnight to allow colony growth.

2.5.7 Blue-white screening

LB agar plates were prepared with initially 50 µg/ml Kanamycin that was then reduced to 20 µl/ml due to slow growth. 40 µl of 40 mg/ml X-gal and IPTG was spread onto the surface of each plate followed by 100 µl of the transformation culture after outgrowth.

The remaining cultures were then pelleted, re-suspended in 100 µl SOC media and plated onto the plates. The plates were then incubated at 37˚C for 24 hours.

2.5.8 Repression of the Lac operon

LB agar containing 2% glucose and 20 µg/ml Kanamycin were prepared and 100 µl transformation cultures plated onto them. The plates were then incubated at 37 ˚C for 24 hours.

2.5.9 **Preparation of electro-competent G. thermoglucosidasius**

G. thermoglucosidasius cells were revived from glycerol stocks by growing overnight in either 2SPYNG or TGP at 60°C in a shaking incubator set to 220rpm. Cells from 1ml of this culture were harvested by centrifugation at 3,400 × g and used to inoculate 50ml of 2SPYNG or TGP in a sterile 250ml baffled flask. Cultures were grown at 60°C in a shaking incubator set to 220rpm until an OD600 of 1 was reached. Cultures were cooled on ice for 10min prior to centrifugation at 2,000 × g for 20min at 4°C. The supernatant was discarded and cell pellets were re-suspended in ice-cold electroporation medium containing 10% (w/v) glycerol, 0.5M sorbitol, and 0.5M mannitol. Cells were centrifuged and washed with ice-cold electroporation medium a further four-times before final resuspension in electroporation medium to 2% of the original culture volume.

2.6.10 Transformation of electro-competent *G. thermoglucosidasius*

Aliquots (60 µl) of electro-competent *G. thermoglucosidasius* cells were incubated on ice with the desired plasmid. Aliquots were then transferred to Genepulser™ cuvettes of 1 mm path-length, before transformation using a Bio-Rad Genepulser™ electroporator at 2.5 kV, 25 µFD, 600 Ω and typical time constants of 5 ms. Transformed cells were recovered in 1 ml of TGP for 2 h at between 55 °C and 60 °C in a shaking incubator. Cells were harvested by centrifugation at 1,800 × g for 2 min. The majority of the supernatant was discarded before cells were re-suspended in approximately 100 µl of supernatant before spreading out onto TGP agar plates containing 12.5 µg/ml kanamycin.

2.5.10 Gel electrophoresis

To visualise DNA in the form of PCR products or restriction enzyme digests, agarose gels were prepared by dissolving 1 % (w/v) agarose in TAE buffer (40 mM Tris-acetate pH 8.0, 1 mM EDTA) by heating in a microwave until boiling. The solution was then allowed to cool and ethidium bromide (Sigma-Aldrich) was added to 0.001 % (v/v). This solution was poured into a gel cassette, a comb was positioned and the gel allowed to polymerise before being placed in a gel tank and covered with TAE buffer. DNA samples were prepared in 6x DNA loading buffer (50 % (v/v) glycerol, 50 mM EDTA pH 8.0, 0.05 % (w/v) bromophenol blue) and the required volume (5-60 µL) was loaded onto the gel. To determine the approximate size of the DNA, 6 µL of 1 kb DNA ladder (Thermo) was loaded with the samples. The electrophoresis was performed at a constant 80-100 V and monitored by following the dye front. The gels were run until the DNA bands were correctly separated. The DNA intercalated with ethidium bromide was visualised using an UV trans illuminator.

2.5.11 DNA sequencing

Sequencing was performed by Eurofins MWG Operon, UK. Samples were prepared as per Eurofins instructions.

2.6 SECRETOME ANALYSIS

2.6.1.1 Protein precipitation optimisation

Three different techniques of protein precipitation were trialled and are described in the Figure 2.1 below.

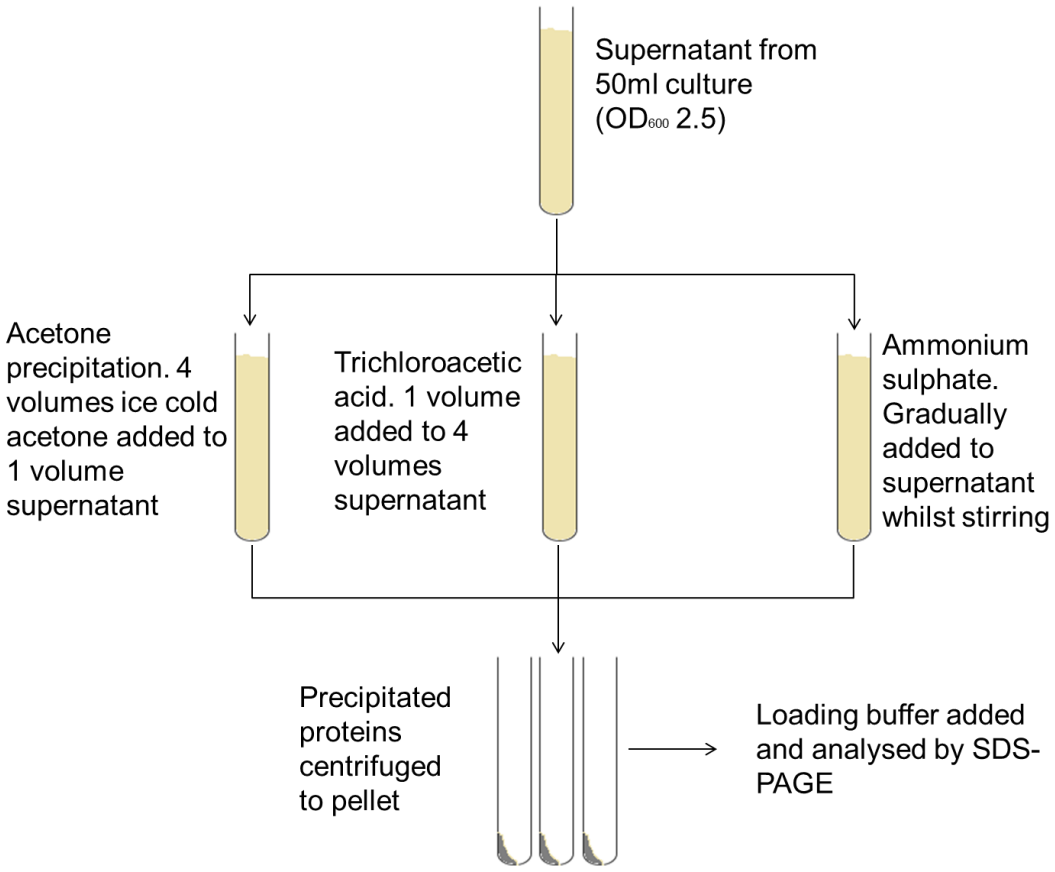

Figure 2.1 Simplified workflow of methods used to identify optimal technique for protein precipitation.

2.6.1.2 Trichloroacetic acid precipitation of proteins

Once the desired OD_{600} was obtained, the cells are removed by centrifuging at 7000x *g* followed by syringe filter sterilisation .To this, 100% TCA was added to a final concentration of 20% and incubated on ice for 2 hours. The supernatant was then centrifuged at 7000x *g* for 20 minutes to pellet the precipitated proteins. The pellet was then re-suspended in 100% acetone, and centrifuged again followed by three more washes in 80% acetone. Finally, the protein pellet was air dried overnight at room temperature.

2.7 SDS-PAGE

2.7.1 One dimensional SDS-PAGE

Samples were prepared by adding a 1:1 volume Laemmli loading dye and boiling for 3 minutes.

2.7.1.1 Stacking gel buffer

4xTris-HCl/SDS, pH 6.8 (0.5M Tris-Cl/ 0.4% SDS)

Dissolve 12.1 g Tris base in 80 ml dH_2O. Adjust pH to 6.8 with 1 M HCl. Add dH_2O to 200 ml total. Filter through 0.4 µm filter, add 0.8 g electrophoresis grade SDS and store at 4˚C.

2.7.1.2 Separating gel buffer

4xTris-Cl/SDS, pH 8.8 (1.5M Tris-Cl/ 0.4% SDS)

Dissolve 182 g Tris base in 600 ml dH_2O. Adjust pH to 8.8 with 1 M HCl. Add dH_2O to 1000 ml total. Filter through 0.4 µm filter, add 4 g electrophoresis grade SDS and store at 4˚C.

2.7.1.3 Sample loading buffer 2X

Component	**Amount**
1M Tris-Cl pH 6.8	1.6 ml
10% SDS	4 ml
87% Glycerol	2 ml
ß-mercaptoethanol	1 ml
Bromophenol blue	4 mg
H_2O	Up to 10ml

2.7.1.4 Laemmli buffer 10X

Component	**Amount**
Tris base	30.3g
Glycine	144g
SDS	10g
H_2O	Up to 1l

2.7.1.5 SDS-PAGE gels

Separating and stacking gels were prepared in varying concentrations as per the table below.

Separating gel

	7.5 %	10 %	12.5 %	15 %	17.5 %
30% acrylamide (ml)	3.6	4.8	6	7.2	8.4
H_2O (ml)	7.2	6	4.8	3.6	2.4
Separating buffer (ml)	3.6	3.6	3.6	3.6	3.6

Stacking gel

30% acrylamide (ml)	0.6
H_2O (ml)	3
Separating buffer (ml)	1.2

2.7.1.6 Coomassie staining

Proteins were separated by SDS-PAGE, and the gel was fixed (40 % methanol and 10 % acetic acid) for 25-20 minutes on a rocker at room temperature. Next, the gel was stained for 1 hour (in 0.025 % Coomassie brilliant blue G250 in 10 % acetic acid), and then de-stained (20 % methanol and 10 % acetic acid) until the bands were clearly visible. For long-term storage gels were dried in cellophane wrap (BioDesign GelWrap).

2.7.2 Western blot

2.7.2.1 Protein concentration

Protein concentration was determined using bicinchoninic acid (BCA) assay (Thermo Fisher Scientific) as per the manufacturer's instructions.

2.7.2.2 Transfer of protein to PVDF membrane

Proteins were firstly resolved by SDS-PAGE then transferred onto a polyvinylidene difluoride membrane (PVDF; Immobilon-P; Millipore) using the semi-dry transfer method The gel and polyvinylidene difluoride membrane (PVDF; Immobilon-P; Millipore) were sandwiched in between several layers of Whatman 3MM chromatography paper (Schleicher & Schuell). The buffers used for transfer were Towbin transfer buffer (25 mM Tris, 192 mM glycine, 20 % Methanol). The membrane was blotted for 2 hours at 0.8 mA/cm^2, and then removed and washed.

2.7.2.3 Membrane probing

Once blotted and the proteins transferred, the membrane was blocked in Phosphate-buffered saline containing Tween 20 (PBST; 150 mM NaCl, 20 mM KCl, 25 mM Tris pH 7.4 and 0.05 % Tween-20/litre) and 5 % skimmed milk (Sigma) overnight at 4 ˚C or at room temperature for 1 hour. The membranes were then incubated with primary antibody or pre-immune sera at the appropriate concentration (as stated in the chapter) in PBS-T and 0.1 % milk at room temperature for 1 hour or overnight at 4 ˚C followed by washing in PBS-T, The membrane was washed and the secondary antibody, horseradish peroxidase- conjugated to goat anti-rabbit (Promega), added, in all cases at concentration of 1 in 10000, for 2 hours at room temperature or overnight at 4 °C. Blots were washed with PSBT prior to performing signal detection with Pierce ECL western blotting substrate, following the manufacturer's guidelines. The blots were then visualised using exposure films which were then developed. The exposure time for each blot varied depending on the antibody and sample.

2.7.3 Cell Fractionation

Geobacillus thermoglucosidasius cells were grown on TGP agar plates to produce a thick lawn, which was then scraped off and added to 20 ml pre-warmed ASM. The cells were recovered by incubating at 60 °C and 220 rpm for 1 hour, which was then used to inoculate 20 ml fresh pre-warmed ASM in 250 ml baffled conical flasks to OD600 of around 0.1. The culture was then grown to an OD600 of around 1.5 and 2 ml harvested.

The cells were centrifuged and the supernatant collected as the media fraction. The pellet was then re-suspended in 2 ml pre-warmed protoplast buffer (20% sucrose, 50 mM Tris-HCl pH 7.5, 15 mM Mgcl2, 5 µg/ml lysozyme) at 37°C for 30 minutes. Protoplasts were then centrifuged at 700x *g* for 10 minutes. The supernatant was collected as the cell wall fraction, and the pellet the protoplasts. The protoplasts were then lysed by re-suspension in 2ml 50 mM Tris-HCL pH 7.5 and sonicated. The suspension was then centrifuged at 50,000 x *g* for 1 hour (Beckman coulter benchtop ultracentrifuge) and the supernatant collected as the cytoplasmic fraction and the pellet re-suspended in 2 ml 50 mM Tris-HCL pH 7.5 and collected as the cell wall fraction.

CHAPTER 3: CHARACTERISATION OF THE G. *THERMOGLUCOSIDASIUS* C56-YS93 SECRETOME AND COMPARISON WITH *B. SUBTILIS*

3.1 INTRODUCTION

3.1.1 The secretome

The term 'secretome' was first coined by Tjalsma et al (2000) in their studies into *Bacillus subtilis* (*BS*) and was defined as all the secreted proteins and the secretion machinery of the bacterium (Tjalsma et al., 2000). In this paper, they reviewed the secretion machinery, chaperones, folding factors, signal peptidases, and other proteins involved in protein secretion in *BS*, in particular those involved with the Sec pathway. In 2004, they published another review (Tjalsma et al., 2004), analysing signal peptides and other retention signals in depth, limitations of the Sec translocation machinery and related proteins, quality control factors, and covered the Tat translocation machinery as well. Since then, the term secretome has been used to define secretory proteins of not just *BS*, but many other types of cells and tissues including human bone tissue, tumour cells, plant cells, and many other eubacteria. As such, characterising the secretome of organisms is of great value, as it can confer information about the types of proteins secreted, in particular the enzyme subset, and also provide an insight into proteins that are hyper-secreted by a particular organism; this, in turn, may unlock secrets about the evolution of efficient secretion.

Significant strides in the understanding of protein secretion machinery in *BS* has led to numerous opportunities for biotechnological exploitation in several areas, such as the optimisation and overproduction of native and heterologous enzymes for industrial use in paper and detergent industries. Other examples include the overproduction and secretion of heterologous proteins for the production of vaccines, biologically relevant proteins like antibody fragments and other biopharmaceuticals (Kober et al., 2013, Pohl et al., 2013).

As mentioned in the general introduction, in order to improve feedstock utilisation by *GT* by production of hydrolases, an analysis of the secretion system by identifying differences in secretion components, could allow us to identify specific features that could be improved upon. Because *GT* is a thermophile, the secretion mechanisms may not be the same as those in a mesophilic system such as in *BS*. This could be due to

higher temperatures affecting the kinetics of protein folding and protein targeting, how the signal peptide interacts with the membrane, and membrane fluidity differences.

3.1.2 Signal peptides

Gunter Blobel and David Sabitini first proposed the signal hypothesis in 1971, suggesting that information coded at the N-terminal region of a protein directed nascent polypeptides to the endoplasmic reticulum of eukaryotic cells to be secreted, stored or disposed (Blobel and Sabatini, 1971). Subsequent research identified several key components of the translocation machinery, including the Signal Recognition Particle (Walter and Blobel, 1980) and Sec61 (Deshaies and Schekman, 1987) (the latter is a homologue of the prokaryotic SecY protein). In 1999 Gunter Blobel received the Nobel prize in Physiology or Medicine for his discovery that proteins have intrinsic signals that govern their transport and localisation in the cell (Makarow, 1999). The signal hypothesis has since been demonstrated in not only multicellular eukaryotes, but also in single-celled eukaryotes and prokaryotes.

In all living cells, proteins gain entry to the Sec or Tat secretory pathway if they bear a specific targeting signal sequence, the signal peptide [SP], which is typically a short stretch of 20 to 40 amino acids in length present at the N-terminus of secreted proteins. Sec-type SPs are characterised by their three-domain structure (Figure 3.1): a positively charged N-region, a hydrophobic H-region and a C-region that precedes the all-important cleavage site. The N-region usually contains positively-charged residues (lysine or arginine) which are thought to target the pre-protein to the membrane by interacting with the negatively-charged phospholipids and through interaction with the signal recognition particle Ffh by interacting with the negatively-charged backbone of the RNA (Janda et al., 2010). The H-region contains a long stretch of hydrophobic amino acids, which when embedded in the cell membrane form an α-helical shape. Finally, the C-region contains the cleavage site, which is recognised by type 1 signal peptidases, and which is usually preceded by a consensus sequence of A-X-A (van Roosmalen et al., 2004). These peptidases have thus far been described as serine proteases and their active sites, which catalyse the cleavage of the signal peptide from the full-length polypeptide, are located on the extra cytoplasmic side of the cell membrane. After

removal from the full-length polypeptide, signal peptides linger in the membrane where they are degraded by Signal peptide peptidases, which belong to another class of serine protease (Bolhuis et al., 1999a) .

Figure 3.1: Schematic representation of the signal peptide

3.1.3 Signal peptide prediction

From protein sequences alone, signal peptides, and therefore the corresponding secreted protein, can be predicted using bio-informatics. The field of bio-informatics has provided a host of powerful tools enabling scientists to make useful predictions in proteomics. In the past 20 years, a number of tools have been developed to make useful predictions ranging from analysis of primary sequences to prediction of tertiary structures. As mentioned previously in this chapter and in the general introduction, there is no sequence homology between signal peptides, although there is structural similarity in the tripartite structure of all signal peptides.

Gunnar von Heijne was the first to recognise the tripartite nature of signal peptides. In particular, he identified the key hydrophobic nature of the N terminus of secretory proteins (Von Heijne, 1981) and, at a later stage, recognised the charged N-region and the more polar C-region (Von Heijne, 1985, Von Heijne, 1984, Von Heijne, 1982). He was also among the first to develop a method for predicting signal peptides using weight matrices based on particular amino acids at specific locations, particularly at the -1 and -3 sites relative to the cleavage site (Vonheijne, 1986). A weight matrix is the probability of finding a particular residue at each position in the amino acid sequence of the signal peptide. To identify a potential new signal peptide sequence, the amino acid sequence is scanned by a moving window, and a score is calculated using the weight matrix. Position weight matrices were first described by Stormo et al in 1982 (Stormo et al., 1982), and were used to predict translation initiation sites in *E. coli*, as an alternative to consensus sequences, and have since been applied to identification of DNA motifs and

of protein domains. One such tool that uses this approach to predict signal peptides is PrediSi (Hiller et al., 2004).

Another computational method of predicting signal peptides is by applying machine-learning algorithms. These methods include a 'training' phase, where the algorithm is presented with both known typical signal peptide sequences and known non-signal peptides, in order for it to 'learn' from the sequences to build a classification model by tuning its specific parameters. This model is then used to categorise novel peptide sequences. One such prediction server that is based on learning algorithms is SignalP, which is based on artificial neural networks or hidden Markov models, developed by von Heine's group at the Technical University of Denmark . SignalP was the first web server, first released in 1996, that predicted signal peptides, and therefore, secreted proteins (Kall et al., 2004, Nielsen and Krogh, 1998, Nielsen et al., 1997). The first version was based on artificial neural networks, and later versions incorporated hidden Markov models. Subsequently, the server was refined further to discriminate between transmembrane helices and signal peptides that also form a transmembrane helix when inserted into the membrane. The group have also developed prediction servers for transmembrane helices (Kall et al., 2004, Krogh et al., 2001), lipoproteins (Juncker et al., 2003), proteins secreted without signal peptides via the general secretory pathway (Bendtsen et al., 2005a), and proteins secreted via the Tat pathway (Bendtsen et al., 2005b).

3.1.4 Proteomics techniques to identify the secretome

In order to identify the most highly abundant secreted proteins and their signal peptides, two main methods are used. Firstly, shotgun proteomics of the secretory fraction can be used, which refers to the use of bottom-up proteomics whereby a heterogeneous protein mixture undergoes a trypsin digest, and the peptides are then subjected to liquid chromatography (LC) coupled to mass spectrometry (Wu and MacCoss, 2002). The resulting mass spectra for each peptide are then used to identify the protein from which they are derived by searching against a database containing protein sequences.

Secondly, (secretory) proteins can be identified using two-dimensional gel electrophoresis, in which a heterogeneous mixture of proteins is separated in the first

dimension by their isoelectric point using isoelectric focusing. Then, in the second dimension (i.e. at a 90° angle from the first field), proteins are separated by size as in conventional SDS-PAGE. The resulting gel is then stained so each protein spot becomes visible. Automatic software is commonly used to identify the spots by size and isoelectric point, with the caveat that the probability of two proteins sharing the exact same two properties is very low. Specific spots can then be picked out and identified using mass spectrometry.

Both techniques obviously have their own advantages and drawbacks. The shotgun method is more high throughput, and can identify proteins in the sample even at very low levels, but is only semi-quantitative at best. The 2-D method is more quantitative, as spots can be quantified and compared, but is low-throughput, time consuming, and cannot identify proteins secreted at extremely low levels, despite sample enrichment.

In this project, we have chosen to get an overview of the secretome using shotgun mass spectrometry due its sensitivity and its semi-quantitative nature.

3.1.5 Signal peptide modification and libraries

As discussed previously, modification or replacing the Sec signal peptide has been shown to be a successful strategy in optimising protein secretion via the Sec pathway. For example, signal peptide libraries containing signal peptides from specific organisms have been created by several groups, with a view to screening for optimal secretion of heterologous proteins in the host organism (Brockmeier et al., 2006). This type of screening is required as it is has been demonstrated that there is no 'one size fits all' signal peptide, in that a signal peptide that results in efficient secretion for one polypeptide, may not be as efficient for another polypeptide

In several organisms, including *Bacillus subtilis*, it has frequently been shown that modifying the signal peptide may improve the secretion of specific proteins. Caspers et al (Caspers et al., 2010) modified the signal peptide from an alpha-amylase using saturation mutagenesis. This ultimately resulted in point mutations, which they demonstrated to improve secretion of another protein, a heterologous cutinase. However, there is yet to be found a perfect artificial signal peptide, although several parameters have been identified such as the preference by the signal peptidase for

smaller residues at positions -1 and -3 relative to the cleavage site (Borchert and Nagarajan, 1991) and the requirement for a positively-charged residue in the N-region (Tjalsma et al., 2000, Chen and Nagarajan, 1994).

3.2 AIMS AND OBJECTIVES

The characterisation of the secretome of *Geobacillus* will involve a two-pronged approach. The first approach will be to use bioinformatics to predict secreted proteins of two strains *GT* C56-YS93 and TM242 by their putative signal peptides and to compare them to those of *BS*.

The second approach, which involves the C56 strain, will be to perform a physical study of all the secreted proteins found in the extracellular media using protein separation techniques and mass spectrometry for identification. At the time the experiment was carried out, the C56 strain was selected as its genome sequence was the only one publically available (unpublished) at that time (2012). The genome of TM242 has since been sequenced and made publically available (Chen et al., 2015a).

Another aim is to use the Basic Local Alignment Search Tool (BLAST) to compare between the sequences in *BS* and *GT,* the secretion machinery, and to determine if all the components of the secretion machinery found in *BS* are also found in *GT*, including components like chaperones, folding factors and quality control proteases.

3.3 METHODS

3.3.1 Screening for signal peptide containing sequences

The ORFs for *Geobacillus thermoglucosidasius* C56 and TM242 strains were downloaded from NCBI and ERGO respectively. They were then analysed using the SignalP 4.1 server (Petersen et al., 2011) to identify the ORFs that contain putative signal peptides.

The ORFs that were positive for signal peptides were then analysed using the TMHMM server (Krogh et al., 2001), which predicts transmembrane domains. Those proteins containing two or more transmembrane domains were excluded from the list of secreted proteins.

The LipoP server was then used to identify lipoprotein signal peptides; proteins that scored positive for both LipoP and SignalP were excluded from the final list of secretory proteins.

3.3.2 Grand average of hydropathy (GRAVY) score calculation

GRAVY scores were calculated using the sequence manipulation site GRAVY calculator (Stothard, 2000).

3.3.3 Identifying sequence homology and determining correct annotation of ORFs

The compiled list of putative secreted proteins was individually analysed using BLAST software to determine the potential function of these proteins.

3.3.4 Growth of bacterial strains

Overnight cultures were prepared by inoculating 5ml TGP medium with a glycerol stock scraping followed by incubation at 60˚C in a shaking incubator overnight. The overnight cultures were then used to inoculate 50ml media in 200ml baffled flasks followed by incubation at 60˚C in a shaking incubator until the desired OD_{600}.

3.3.5 TCA precipitation of secreted proteins

Once the desired OD_{600} was obtained, the cells were removed by centrifuging at 7000 x *g* followed by 0.2 µm syringe filter sterilisation (Merck Milipore). To this, 100% TCA was added to a final concentration of 20 % (v/v) and incubated on ice for 2 h. The supernatant was then centrifuged at 7000 x *g* for 20 min to pellet the precipitated proteins. The pellet was then re-suspended in 100 % acetone and centrifuged again, followed by three more washes in 80 % (v/v) acetone. Finally, the protein pellet was air dried overnight at room temperature.

3.3.6 SDS-PAGE

See General Methods (Chapter 2)

3.3.7 In-gel digestion

The gel chunk of interest was excised and cut into 1mm cubes. These were then subjected to in-gel digestion, using a ProGest Investigator in-gel digestion robot (Digilab) using standard protocols (Shevchenko et al., 2006) . Briefly, the gel cubes were de-stained by washing with acetonitrile and subjected to reduction and alkylation before digestion with trypsin at 37°C. The peptides were extracted with 10% formic acid.

3.3.8 Mass spectrometry

Peptides were concentrated, if necessary, using a SpeedVac (ThermoSavant). They were then separated on an Acclaim PepMap 100 C18 trap and an Acclaim PepMap RSLC C18 column (ThermoFisher Scientific), using a nanoLC Ultra 2D plus loading pump and nanoLC as-2 auto sampler (Eskigent). The peptides were eluted with a gradient of increasing acetonitrile containing 0.1 % (v/v) formic acid (5-40% acetonitrile in 5 min, 40-95% in a further 1 min, followed by 95% acetonitrile to clean the column, before re-equilibration to 5 % acetonitrile). The eluent was sprayed into a TripleTOF 5600 electrospray tandem mass spectrometer (ABSciex) and analysed in Information Dependent Acquisition (IDA) mode, performing 250 msec of MS followed by 100 msec MS/MS analyses on the 20 most intense peaks seen by MS. The MS/MS data file generated was analysed using the Mascot algorithm (Matrix Science) against the NCBInr database Aug 2013 with no species restriction, trypsin as the cleavage enzyme,

carbamidomethyl as a fixed modification of cysteines, and methionine oxidation and deamidation of glutamines and asparagines as variable modifications.

3.4 RESULTS AND DISCUSSION

3.4.1 Secreted protein prediction and Signal Peptide comparison

Signal peptide libraries have been created for *B. subtilis* as a tool to aid the improvement of heterologous protein secretion by *B. subtilis* (Brockmeier et al., 2006). Although the general tripartite structure of signal peptides is known, there is no sequence homology, and there is no 'one size fits all' signal peptide (Zheng and Gierasch, 1996). However, there are discernible differences between Gram-positive and Gram-negative signal peptides, and differences between signal peptides that are used for targeting to different pathways; for example, secreted proteins targeted to the Tat pathway have larger N-regions than those targeted to the Sec pathway, and lipoproteins targeted via the sec or the tat pathway are generally shorter than their secreted counterparts.

Furthermore, structural differences have been found between thermophilic and mesophilic proteins that correlate with their environmental temperatures (Szilagyi and Zavodszky, 2000, McDonald, 2010, Sadeghi et al., 2006), such as the frequency of specific amino acids within α-helices (Warren and Petsko, 1995), and interactions such as salt bridges (Vogt et al., 1997) and internal hydrogen bonds that increase with increased protein thermostability (Vogt and Argos, 1997).

With this in mind, it is not inconceivable that there might be differences between the signal peptides of a Gram-positive thermophile and a Gram-positive mesophile. The first step was then to identify the predicted secretome of *GT* and *BS* using SignalP and TMHMM, and compile a list of secreted proteins.

Several prediction servers are available such as SignalP, PrediSi, and Phobius. For the purposes of this analysis, SignalP was selected due to its higher sensitivity and accuracy. Researchers at the Technical University of Denmark developed SignalP 4.1, which predicts secreted proteins using two different predictors based on neural networks and hidden Markov models (Bendtsen et al., 2004). To separate transmembrane proteins that can occasionally be predicted as secreted proteins, another prediction server was used, TMHMM 2.0, also developed by the Technical University of Denmark. Although transmembrane proteins are thought to be transported to the membrane via the sec

pathway, and bear signal peptides, it is unclear if intramembrane folding of transmembrane proteins is mutually exclusive from the Sec targeting and translocation process, and if the signal peptides are cleaved or remain one of the membrane-spanning domains (Craney et al., 2011). For this, reason, signal peptides borne by transmembrane proteins are excluded for the compilation of bio-informatically predicted signal peptide libraries.

Once the sequence was processed by SignalP and identified to be a protein bearing a signal peptide with a cleavage site, the results were then processed by TMHMM to exclude any transmembrane proteins that were mistakenly identified as secreted proteins.

To separate predicted secretion proteins from predicted transmembrane (TM) proteins, which both encode N-terminal signal peptides, two prediction servers were used to predict the presence of transmembrane helices in translated ORFs. TMHMM 2.0, developed by the Technical University of Denmark, predicts transmembrane helices using a Hidden Markov Model (HMM). Figure 3.2 shows an output example of a protein predicted to be secreted by SignalP, but is actually a transmembrane protein.

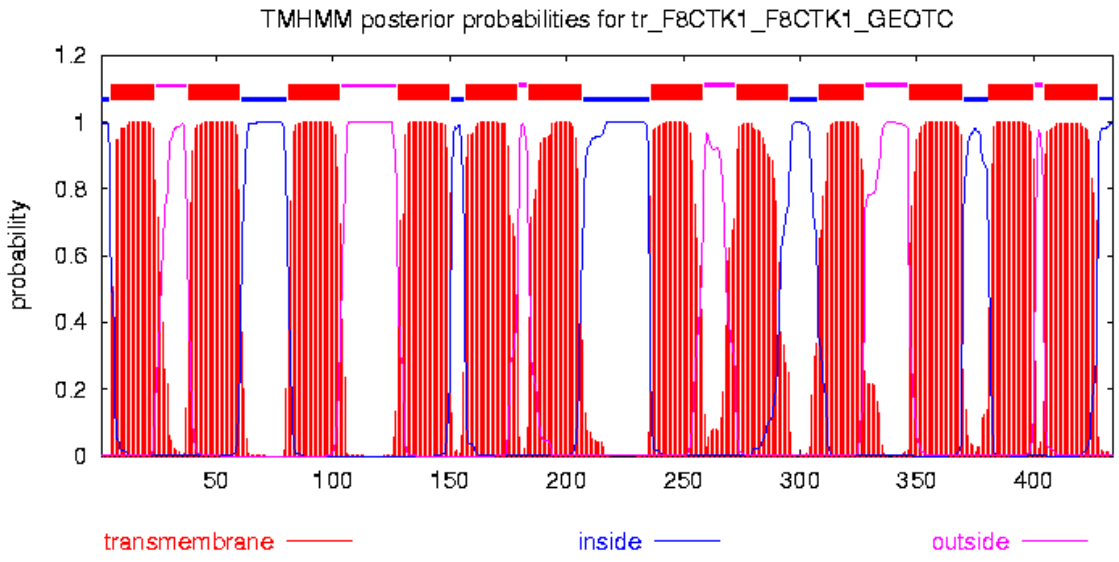

Figure 3.2: TMHMM output example plot of posterior probabilities of inside/outside/ transmembrane helix. The server produces this plot by calculating the total probability that a specific residue sits within one of the possible paths through the model. This is an example of MATE efflux family protein, a protein predicted to be secreted, but is clearly a membrane protein due to the transmembrane domains.

As can be seen in Table 1, of the 3656 possible proteins and peptides produced by *Geobacillus thermoglucosidasius* C56-YS93, 78 are predicted to be secreted. Of these, 22 were hypothetical proteins. As for the rest, their functions were known and correctly annotated based on sequence identity. Note that there are a number of proteins that, either because of a very short signal peptide or based on their function, are unlikely to represent genuine secretory proteins. As seen in Table 1, the *BS* genome encodes more than double the number of secreted proteins than in *GT*. Furthermore, the proportion of secreted proteins relative to the total proteome is double in *BS* (4.2%) compared to *GT* (2.1%). It is not clear why there is such a difference, but one might speculate that, as *B. subtilis* is a mesophile, it is found in a more diverse range of ecological niches, thus perhaps requiring a wider range of extracellular proteins to survive and compete, compared to the thermophilic *G. thermoglucosidasius*. Note, however, that the majority of putative extracellular proteins of *B. subtilis* have not been shown experimentally to be secreted and it is likely that the true number of secretory proteins is lower, which may also be the case in *G. thermoglucosidasius*. A protein that is predicted to bear a signal peptide is not necessarily secreted into the extracellular milieu, as in the case of Gram-positive bacterial proteins, they may bear a C-terminal cell wall anchor sequence or could remain anchored within the cell membrane due to the signal peptide being uncleaved.

Table 3.1: Number of signal peptides in GT and BS and hydrophobicity comparison. Genomes of BS and GT were screened as described in section 3.3.1

Strain	*Total*	*Secreted proteins*	*% secreted*	*Hypothetical proteins*	*GRAVY(hydrophobicity)*
BS 168	4244	178	4.2%	71	0.97
GT C56	3656	78	2.1%	22	1.00

The next step was to identify any sequence differences between the two groups of signal peptides. Signal peptides have been shown to form an α-helical structure within the cell membrane during the translocation process (Briggs et al., 1986). As such, the signal peptide fragment of the protein temporarily functions as a transmembrane domain. It has been shown that transmembrane proteins in thermophiles possess adaptations that confer thermostability, such as increased hydrophobicity. One study comparing

mesophilic and thermophilic transmembrane proteins observed that the most striking difference between the two is the increased hydrophobicity of the thermophilic transmembrane helices (Meruelo et al., 2012). For this reason, the hydrophobicity of signal peptides in *BS* and *GT* were compared using a GRAVY calculator, which calculates the grand average hydropathy of the amino acids in the sequence, which is the sum of all the hydropathy values of all the amino acids divided by the sequence length. Despite the significant differences in growth temperature of the two organisms, the GRAVY score calculations show no significant difference in the hydrophobicity of the signal peptides of *GT* and *BS* (p=0.621). The longest predicted signal peptide from *BS* is 47 amino acids in length while the longest from *GT* is 42 amino acids. However, the average length is 26 and 25 respectively, indicating no significant difference between the overall lengths of the signal peptides (p=0.165).

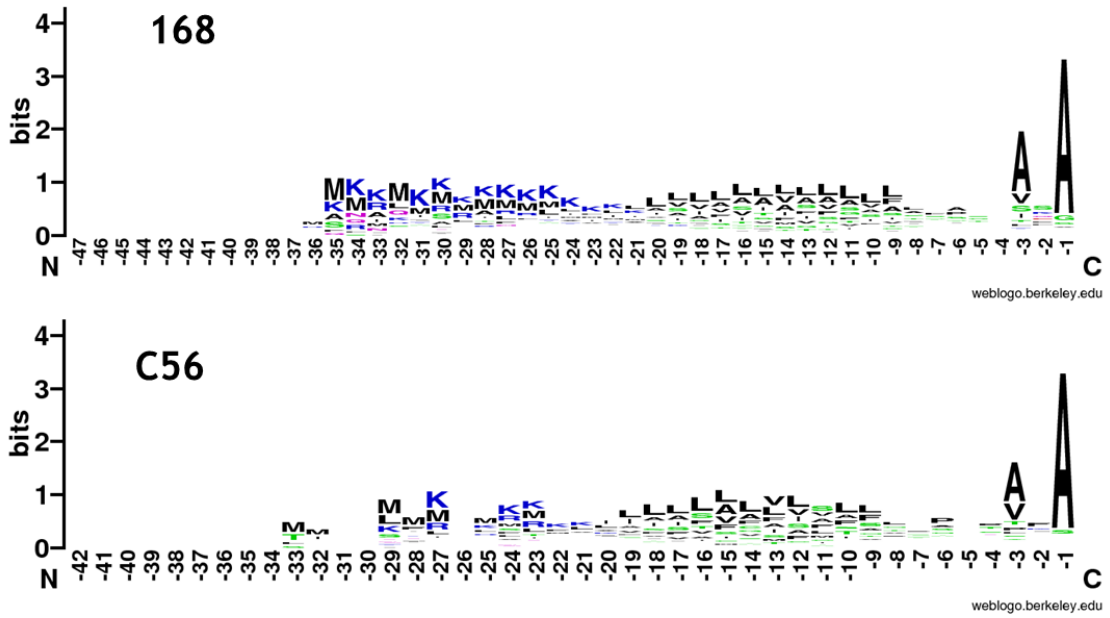

Figure 3.3: Weblogo sequence alignment of signal peptides from GT C56-YS93 (C56) and BS 168 (168) aligned at the signal peptidase cleavage site.

Considering the Weblogo sequence alignment in Figure 3.3, it can be observed that there is also little difference in the three residues immediately prior to the cleavage site, even though it remains to be shown experimentally if type 1 signal peptidases from *GT* are able to recognise and cleave signal peptides from BS or other organisms. In both

organisms they appear to use most frequently the A-X-A sequence, which has been shown to be the consensus sequence at this position (van Roosmalen et al., 2004) . Thus, the lack of differences between signal peptides of the two organisms indicate that signal peptides from both could be used interchangeably, and signal peptide libraries from either *BS* or *GT* could be used to screen for optimal secretion of heterologous proteins in *GT.*

Signal peptides from different species have been utilised to screen for optimal signal peptides for heterologous expression, for example *Bacillus subtilis* and *Bacillus licheniformis* signal peptide libraries have been used to screen for optimised heterologous protein secretion in *Bacillus subtilis* (Degering et al., 2010). In *Pichia pastoris,* the use of a *Saccharomyces cerevisiae* signal peptide results in efficient secretion. In *E. coli,* optimised *BS* signal peptides have been used for recombinant protein secretion. For example, Brockmeier at al used two different enzymes, a cutinase and a lipase, as reporters for heterologous protein secretion in *BS*. They demonstrated that signal peptides that resulted in optimal secretion for the cutinase, did not confer the same levels when used with the lipase, and *vice versa* (Brockmeier et al., 2006). This is thought to be due to the possibility that the sequence specificity does not end at the cleavage site of the signal peptide, but that it is the combination of amino acids both before and after the cleavage site that are crucial to the efficiency of signal peptide cleavage. With this in mind, it would be interesting to investigate if signal peptide libraries created to include several amino acids after the putative cleavage site would improve so-called hits of suitable signal peptides.

3.4.2 Secretion machinery components

One other thing to consider is the secretion pathway components, specifically the molecular machinery that is involved in the Sec secretion pathway. Between Gram-negative and Gram-positive bacteria there have been identified some important differences. One example is that *Bacillus subtilis* and most other Gram-positive bacteria lack a protein known as SecB, which in *E. coli* acts as a chaperone to keep secretory proteins in a translocation-competent state. Another notable difference is the membrane proteins SecD and SecF in *E. coli* are present as one single polypeptide in *BS* and most other Gram-positives.

Table 3.2 : Sec machinery components

***B. subtilis* 168**	***G. thermoglucosidasius* C56-YS93**
Protein name	*Genome annotation number*
SecY	Geoth_0150
SecE	Geoth_0115
SecG	Geoth_0448
SecDF	Geoth_1057
SecA	Geoth_0393
SipV	Geoth_3148
SipS	Geoth_2719
LspA	Geoth_2778

Using the sequences of BS proteins involved in protein transport, a BLAST search was carried out against the *GT* genome to identify members of the Sec pathway. As can be seen in Table 3.2, all the major components of the Sec pathway are present. These are SecYEG, which forms the translocon; SecA, which drives the translocation process and has also been shown to interact with the nascent chain (Zimmer et al., 2008, Huber et al., 2011, Chatzi et al., 2014b); SecDF, which may function as a chaperone; and type I and type II signal peptidases. The only difference is that in BS there are five genes for signal peptidases, namely *sipS, sipT, sipU, sipV and sipW*; while *GT* possesses only two genes for type I signal peptidases. This could be correlated with the fact that *BS* has double the secreted proteins compared to *GT,* and the 5 different signal peptidases appear to have different substrate specificities or preferences (Bron et al., 1998). However, it should be noted that in BS only *sipS* and *sipT* are key in protein secretion and cell viability, as inactivating mutations in both genes resulted in a non-viable strain(Tjalsma et al., 1997). Lipoproteins, which are also translocated across the membrane via the Sec pathway, are cleaved by a type II signal peptidase that is present in both *BS* and *GT*.

Apart from the actual translocon machinery, several other proteins also have vital roles in efficient protein secretion via the Sec pathway. These include chaperones, proteases and proteins involved in targeting to the membrane (Table 3.3).

Table 3.3: Secretion process accessory proteins

***B. subtilis* 168**	***G. thermoglucosidasius* C56-YS93**
Protein name	*Genome annotation number*
DnaK	Geoth_1172
DnaJ	Geoth_1173
PrsA	Geoth_3173
HtrA	NO
HtrB	NO
WprA	NO
GrpE	Geoth_1171
FtsY	Geoth_2728
HslO	Geoth_0087
Tig	Geoth_0987
NO	Geoth_2193
SpoIIIJ	Geoth_3947
GroEL	Geoth_3677
GroES	Geoth_3678
CsaA	NO

Molecular chaperones are proteins that assist in folding or unfolding of other molecular structures, such as other proteins, DNA, RNA, or combinations these macromolecules. One major function of this class of proteins is to prevent protein aggregation of newly synthesised polypeptides. DnaK and DnaJ, respectively known as Hsp70 and Hsp40 in eukaryotes, are molecular chaperones that are present in almost all prokaryotes and eukaryotes. These chaperones, as well as another co-chaperone GprE, were discovered to mitigate heat damage to proteins by preventing aggregation and have also been found to prevent damage due to stress (Schroder et al., 1993). GroEL and GroES, known as Hsp60 and Hsp10 respectively in eukaryotes, belong to the chaperonin family of molecular chaperones, and are another example of cytoplasmic chaperones involved in preventing protein aggregation due to misfolding of proteins. Although not exclusively chaperones for secreted proteins, as they are involved in prevention of misfolding of all proteins in the cytoplasm, they are considered to have an important role in protein secretion as well.

The protein PrsA is a membrane bound lipoprotein which is thought to be involved in post-translocational folding of secreted proteins and has been shown to be both an essential and rate-limiting factor in protein secretion in *BS* (Wahlstrom et al., 2003, Kontinen et al., 1991, Jacobs et al., 1993). Furthermore, it has also been shown that over-expression of the *prsA* gene results in improved secretion of heterologous protein in *BS.* As *prsA* is present in *GT*, over-expression of this gene may be a viable strategy to improve protein secretion in *GT.*

HtrA and HtrB are highly similar extracellular serine proteases involved in folding and proteolysis of misfolded secretory proteins (Darmon et al., 2002), and these proteins belong to a widely conserved set of proteins present in both prokaryotes and eukaryotes. When a BLAST search was performed against both protein sequences for *GT*, the returned results were only serine proteases with less than 50% identity, suggesting there may not be true HtrA and HtrB homologues in *GT*. However, it could be that one of these annotated serine proteases is a functional analogue of both HtrA and HtrB. However, in *BS* expression of *htrA* and *htrB* is tightly regulated by a two-component regulatory system named CssR-CssS, and these also appeared to be absent from *GT*. CssS is a sensor histidine kinase, and CssR is a transcription regulating protein,

both being involved in sensing secretion stress caused by e.g. protein misfolding(Westers et al., 2006). When a BLAST search was carried out for CssS and CssR respectively, a sensor histidine kinase was predicted to have less than 25% sequence identity, and several DNA binding response regulators were predicted to have less than 50% sequence identity. Bearing in mind that *GT* appears to lack both HtrA/B and CssS/R, it seems unlikely that the identified serine protease from the BLAST search is a true homolog of HtrA or HtrB. Further experimental work would need to be carried out in order to determine the function of the putative serine protease. In *BS*, a double knockout of *htrA* and *htrB* result in growth defects and temperature sensitivity (Darmon et al., 2002), so it would be interesting to identify what protein fulfils the role of HtrA and HtrB in *GT.*

Another protein of note that was not identified in *GT* is WprA, which is a cell-wall associated protein that after translocation across the membrane is processed into two separate cell wall proteins; one with a serine protease domain and the other protein with a putative chaperone activity. Similar to HtrA/B, these proteins may be involved in quality control and degradation and/or folding of extra cytoplasmic proteins. In BS, a *wprA* knockout strain results in improved production of a heterologous amylase (Stephenson and Harwood, 1998); for a similar strategy to be employed in *GT*, the functional analogues of these proteins would need to be identified first.

As mentioned before, Gram-positive bacteria lack the *E. coli* chaperone SecB. However, it has been suggested that this function is taken over in BS by the protein CsaA, as in a *secB* knockout of *E. coli,* the growth defects and stress conditions are restored to normal with the production of CsaA (Muller et al., 2000a). Surprisingly, CsaA is not present in *GT* despite being present in several other *Bacillus sp.*

3.4.3 Shotgun mass spectrometry

A shotgun mass spectrometry approach was used to identify the most abundant proteins in the extracellular media, in order to obtain an indication of the most highly secreted proteins. This was done by growing *GT* C56-YS39 aerobically in TGP media to an optical density of 2.5, which is just after log phase (Figure 3.4), and it can be observed that the growth rate is declining at this optical density. In *BS*, late log phase or early stationary phase, also known as deceleration phase or post-exponential phase, is the stage when most secreted proteins are produced (Hirose et al., 2000, Tjalsma et al., 2004, Antelmann et al., 2001). In this section, comparative *in silico* predictions for proteins secreted via the Sec pathway were combined with a set of experimental data derived from a shotgun mass spectrometry analysis of C56 extracellular-enriched fractions. This analysis provides insight into the nature of the extracellular milieu of *GT* C56-YS93 and direct experimental evidence of the secretion of proteins predicted to be secreted.

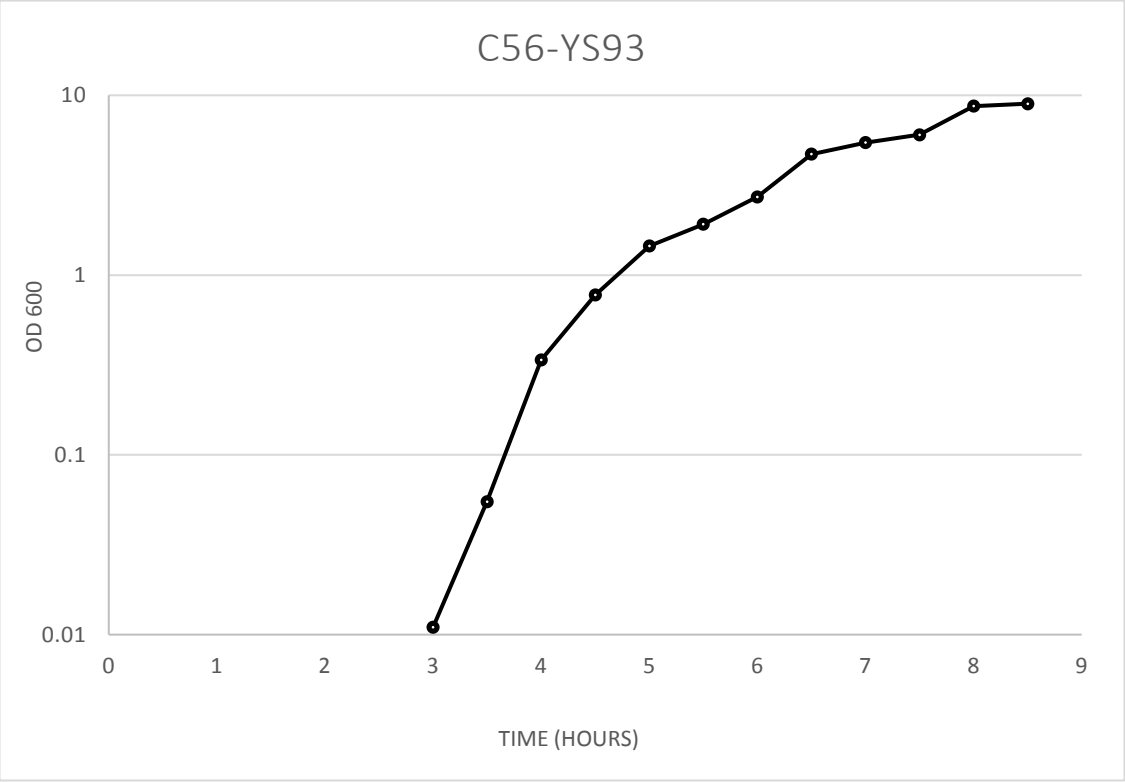

Figure 3.4 Growth curve of GT C56-YS93 on TGP medium of the optical density at 600nm over time

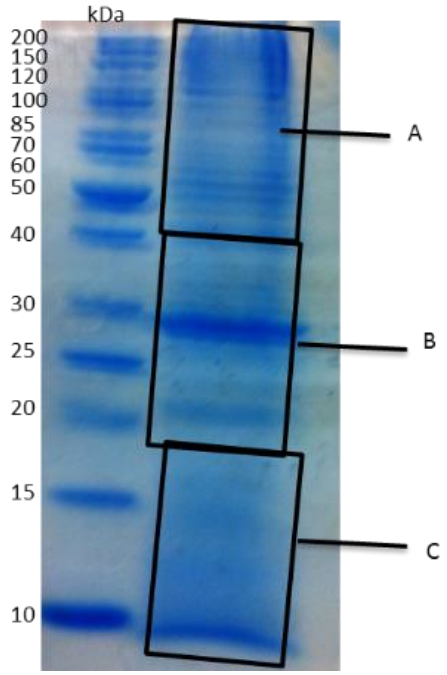

Figure 3.5 Segmented SDS-PAGE gel for shotgun mass spectrometry analysis. The gel slices were sent to St Andrews Mass Spectrometry and Proteomics Facility where in-gel trypsin digest was carried out, followed by analysis by MS/MS. The resulting data was then returned and analysed using Protein Pilot software.

The extracellular milieu sample was separated and resolved using SDS-PAGE. The resulting lane was then divided into three 2cm sections (Figure 3.5). These were then analysed by LC-MS/MS mass spectrometry. The results yielded a list of proteins (Table 3.4), ranked in order of confidence of the evidence of the protein based on the peptides. This technique is not a truly quantitative technique but, based on the frequency of peptides and therefore the proteins from which they are derived, we can obtain an indication of the proteins that are more abundant in the sample. Figure 3.6 is the example output when the results are exported from ProteinPilot.

	N	Unused	Total	%Cov	%Cov(50)	%Cov(95)	Accession	Name	Species	Peptides(95%)
2	1	328.8	328.8	89.5099998	89.50999975	89.5099998	F8CX47	S-layer domain-containing protein OS=Geobacillus tl	GEOTC	364
3	2	149.02	149.02	74.119997	69.27999854	68.5100019	F8CX44	Subtilisin (Precursor) OS=Geobacillus thermoglucosi	GEOTC	121
4	18	63.68	63.68	55.4099977	49.64999855	43.5799986	F8CX05	Mannosyl-glycoprotein endo-beta-N-acetylglucosan	GEOTC	33
5	33	46.33	46.33	38.319999	35.62999964	33.2599998	F8CSQ3	Ig domain protein group 2 domain protein (Precurso	GEOTC	24
6	58	32.43	32.43	60.0499988	52.28000283	50.4599989	I0UCA4	Extracellular zinc metallopeptidase, M23 family OS=(	BACTR	18
7	64	29.25	29.25	41.870001	40.27999938	40.2799994	F8CX92	Cell wall hydrolase/autolysin (Precursor) OS=Geoba	GEOTC	15
8	69	25.38	25.38	57.4999988	43.95999908	39.7899985	I0UCA7	Carboxyl-terminal protease OS=Geobacillus thermo	BACTR	15

Figure 3.6: Protein Pilot output example and data headings. The unused score is a measurement of the confidence in protein identification and reflects the amount of total unique peptide evidence related to that protein. The total score is a sum of all the peptide evidence related to the protein. The %Cov, %Cov(50) and %Cov(95) are the percentages of amino acid sequence covered by the peptides that correlate to a protein in the searched database, with the number in parentheses referring to the confidence level. Peptides (95%) is the actual number of distinct peptides seen in the MS data from the listed protein at 95% confidence interval; it is correlated to the coverage by database searching.

The raw data from the shotgun mass spectrometry contained hits for almost the entire sample contained in the gel slices, including some contamination, which were excluded by sorting by species type. For this analysis, we chose only proteins from *G. thermoglucosidasius* species, to exclude any human contamination. In addition to this, the data had to be manipulated further, as a huge number of cytoplasmic proteins were also present in the extracellular milieu fraction. This is most likely attributed to cell lysis during growth. There is little published works on the cell lysis phenomenon, but private communications with TMO Renewables and other members of the lab have revealed that this is typical for *GT*. The experimental results were then combined with the *in-silico* analysis screening for secreted proteins, which yielded a list of secreted proteins, ranked by abundance. However, as mentioned previously, this is not a truly quantitative method of estimating a protein's abundance in the sample, but can give an indication of which proteins are more abundant, and therefore more highly secreted. Table 3.4 is a list of the proteins, containing the UniProt accession number, protein function, and the corresponding signal peptide.

Table 3.4: List of secreted proteins from shotgun mass spectrometry analysis of Geobacillus thermoglucosidasius C56-YS93

UniProt accession number	*Function*	*Signal peptide*
F8CX47	S-layer domain-containing protein	MAYQPKSYRKFLAGSVSAALVATAVGPVVANA
F8CX44	Subtilisin (Precursor)	MKKWKKTAVSLGLASALVLPSFAQA
F8CX05	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	MRIGVQIRKFAALLSVLILLVSYAISPAYA
F8CSQ3	Ig domain protein group 2 domain protein	MAEKKKFLWLLMALLLCVAFGSVPAVAFG
F8CXX0	Peptidase M23	MALAAATVLSIGVFPHFADA
F8CX92	Cell wall hydrolase/autolysin	MQPLRLLLLCLLMFFGYSSGTYA
F8CXX3	Carboxyl-terminal protease	MNKKTTAMLMVLSMLIGAGGTYA
F8CVX6	S-layer domain-containing protein	MRKFYRFVLVFSLLVSIVFPGVVTEAKS
F8CWQ2	Aspartate transaminase	MKLAKRVASLTPSTTLAITAKA
F8CUA0	Serine-type D-Ala-D-Ala carboxypeptidase	MKRIKQKIIIIFLMIGLCFYFLPFHAAKA
F8CX45	S-layer domain-containing protein	MRKYSFLLFFAIAFIFGGKTVDA
F8CX46	NLP/P60 protein	MRRQLVLALLLGGSVFAAGARAEQAEA
F8CUA0	Serine-type D-Ala-D-Ala carboxypeptidase	MKRIKQKIIIIFLMIGLCFYFLPFHAAKA
F8CUX3	Stage II sporulation protein	MKRMKPLIALFS
F8CVA0	Uncharacterized protein	MKRMLTGCLLASLLFAFPAMA
F8D0U9	NLP/P60 protein	MKKSFILTGTIISSLLAGQTAFA
F8CXG1	Sporulation uncharacterized protein YkwD	MNKKIVFSLAASLAIVGASFTAKA
F8CUM8	Copper amine oxidase-like domain-containing protein	MRKIAFGLCVCFLIFTAYSSQIFPVYA
F8CSQ3	Ig domain protein group 2 domain protein	MAEKKKFLWLLMALLLCVAFGSVPAVAFG
F8CX79	S-layer domain-containing protein	MKRTFLHIALSLLAAMLALPAMNASA
F8CSH6	Alkaline phosphatase	MDKKKFFRGLTAFTLASSLGVSSLVTNHDVVKA
F8CYV4	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	MKKFIAVILLLAITGYGIWNALA
F8D1X4	Uncharacterized protein	MFKKGYLSILSLVMGFTFFSTNTFA
F8CYE7	Cytochrome c oxidase subunit II	MHLHKYEKIWLAFGIGCLFVFLTVIGVSAFA

3.5 CONCLUSIONS

The composition of the secretion machinery in *GT* is similar to that in *BS*, with a few exceptions such as the absence of HtrA/B, CssR/S and CsaA. This suggests that many of the modifications made in *BS* to improve heterologous protein secretion may also be employed to improve protein secretion in *GT*. Furthermore, this work shows that signal peptides in GT, when compared to those in *BS,* appear to have no significant differences in size length or hydrophobicity. This could mean that signal peptide libraries, which as discussed have been compiled for *BS* and several other *Bacilli*, may be also used to screen for optimal signal peptides for protein secretion of different hydrolases in *GT*.

CHAPTER 4: CHARACTERISATION OF XYLANASE

4.1 INTRODUCTION

Endo-1,4-β-D-xylanases, commonly known as xylanases, are glycosidases, which are a class of hydrolases that catalyse the endo-hydrolysis of 1,4-β-D-xylosidic linkages in xylan, at random positions along the xylan backbone, yielding short, xylo-oligomers. Xylanases do not, however, catalyse the hydrolysis of xylan side-chains, which can be substituted with residues such as arabinose, glucuronic acid, and acetic, ferulic and p-coumaric acids. Xylanases can be found in several glycoside hydrolase families in the CAZY database (Chakdar et al., 2016, Collins et al., 2005).

4.1.1 Xylanase as an enzyme to improve feedstock utilisation by *GT*

Biomass from waste sources is an abundant and renewable resource, which makes it an ideal candidate as a substrate from which to produce ethanol, other alcohols, or other high-value products. However, as mentioned in chapter 1, these types of biomass are rich in carbohydrate polymers such as cellulose and hemi-cellulose that are quite resistant to degradation. Sources of lignocellulosic biomass can be divided into several groups, which include energy crops, aquatic plants, forest biomass and waste, agricultural residues, and the organic fraction of municipal solid wastes, each of which may contain varying amounts of cellulose and hemicellulose.

In recent years, interest in the biotechnological use of biomass rich in xylan has increased and has intensified studies on xylanase and on xylose utilization. Furthermore, many attempts have been made to utilize xylose as a cost-effective carbon source to produce bioethanol. In a typical industrial process to produce ethanol from lignocellulosic biomass, the biomass is first pre-treated, which in itself may contain several steps. Then, the biomass can either be converted into ethanol, or another organic compound, by separate hydrolysis and fermentation (SHF) or by simultaneous saccharification and fermentation (SSF), or a combination of the two in combined hydrolysis and fermentation (Li et al., 2014, Acharya and Chaudhary, 2012, Alfani et al., 2000, Hetenyi et al., 2011). Either method still usually requires pre-treatment step or steps. One step is the de-lignification of the biomass, followed by a hydrolysis step. The hydrolysis step is designed to de-polymerise the cellulose and hemicellulose to liberate monomeric sugars. This step can be carried out using acid, alkali or enzymes. Acid and

alkaline hydrolysis are both generally nucleophilic substitution reactions by attacking the glycosidic bonds in the cellulose or hemi-cellulosic backbone. Acid or alkaline hydrolysis can be costly due to the requirement of acid or alkaline tolerant or resistant equipment, which increases running costs. Enzymatic hydrolysis is much more low maintenance and lower cost. A combination of mild acid hydrolysis and enzymatic hydrolysis is the most commonly employed approach as the mild acid hydrolysis disrupts the lignin, and increases the cellulosic and hemi-cellulosic components susceptibility to enzymatic hydrolysis (Taherzadeh and Karimi, 2008).

The total enzymatic biodegradation of hemicellulose to its simple sugars involves the synergistic action of several enzymes known as hemicellulases, amongst which xylanase is essential for the depolymerisation of the xylan backbone. Owing to its complexity and heterogeneity, several auxiliary hemicellulases would be required to act synergistically in order for complete de-polymerisation to occur. The composition of xylan varies from source to source with varying concentrations of the different side-chain residues. For example, Birchwood xylan contains less glucuronic acid compared to Beachwood xylan, while in both sources, xylose is the predominant (>80%) residue of which the hemicellulose is composed. However in the case of *GT*, complete hydrolysis may not be necessary, as *GT* is able to utilise oligomeric sugars (TMO Renewables, personal communication).

Furthermore, although hemicellulose is significantly less abundant than cellulose, which is made up of glucose monomers, the cellulosic component is significantly more recalcitrant due to the bulk of cellulose being in a crystalline form. Thus, although over-producing a xylanase by itself would not be sufficient to completely eliminate the need for a hydrolysis step, enzymatic or otherwise, its production in a high ethanol producing strain could indeed improve feedstock utilisation. It is important to bear in mind that we are not suggesting that xylanase be the only hydrolase to be over-expressed in *GT* toward consolidated bioprocessing, but it would be a good start to improving feedstock utilisation for second-generation biofuel production. In order to achieve improved feedstock utilisation, simple overproduction of hydrolytic enzymes may not be sufficient. These proteins are secreted, and the secretion process itself may be rate-limiting in the production of these enzymes. Thus, to improve feedstock utilisation it is

important to understand and optimise the secretion of hydrolytic enzymes such as xylanases.

4.1.2 Xylanase as a model enzyme to study secretion

To optimise protein secretion in *GT*, the rate-limiting steps in the secretion process need to be understood. For this, an easily tractable model enzyme is required. Over the years, several heterologous proteins were typically selected to study bottlenecks in protein secretion in *BS*, namely those that have industrial relevance such as α-amylase (Bolhuis et al., 1999b). One important requirement for the scope of this study is to select a protein that is natively secreted.

In the current work, an endo-1,4-beta-xylanase (GEOTH_2250) from *GT* C56-YS93 was selected as the initial model protein to study protein secretion in *G. thermoglucosidasius* for several reasons, namely because of its industrial relevance, because it is a natively-secreted protein, and because the gene is derived from a closely-related strain, which may mean that transcription and translation speed is likely to be somewhat optimal (codon usage). Another enzyme that was also a candidate for a model to study protein secretion was the α-amylase from *Geobacillus stearothermophilus*, which had already been cloned into the chromosome of *GT,* producing a new strain TM333 (TMO renewables). The strategy was to clone the gene onto a high copy number plasmid such as pUCG3.8 or pUCG4.8, downstream of a strong constitutive promoter

As mentioned in the previous section and in the Introduction, the secretion of xylanase would be of value during the fermentation process, with or without a pre-treatment step, to continue to liberate xylose oligosaccharides to be converted into product by *GT* during fermentation. Endo-1,4-β-D-xylanase from *GT* C56-YS93 is a predicted secreted xylanase, due to the presence of a putative signal peptide. The TM242 strain contains no predicted xylanase, nor does it appear to have any xylanolytic activity as shown by previous work in the Danson group (Dr Giannina Espina-Silva, personal communication). For this reason the xylanase from *GT* C56-YS93 was selected to be used as a model protein to study protein secretion in TM242, the working strain.

As discussed in the general introduction chapter, bottlenecks in protein secretion can occur at any point from gene expression, to the actual translocation of the protein across

the membrane, to proteolytic degradation in the extracellular milieu. Bottlenecks at transcription level can be caused by, for instance, mRNA instability, when it comes to heterologous protein expression. To reduce the possibility of there being a bottleneck at the translation level, the source of the gene would be optimal if from a closely-related organism, as differences in codon frequencies are often an issue when producing heterologous protein (Angov et al., 2008, Welch et al., 2011). Codon usage may also play a role in protein folding as the protein emerges from the ribosome, although it is unclear what secretion chaperones are involved in protein secretion in *GT*, or if the protein is folded in a translocational competent state; therefore, it would be decided to use a gene from another *Geobacillus sp.* to circumvent any issues that may occur at the translational level.

4.2 CHAPTER AIMS

- To achieve heterologous expression and production of the xylanase gene from *GT* in *EC*, and the optimisation of protein production and purification.
- Using purified xylanase protein, to raise antibodies suitable for immunoprecipitation and Western blotting.
- To characterise the enzymatic properties of the xylanase enzyme to ensure typical Michaelis-Menten kinetics towards product formation, with no substrate inhibition, in order for the AZCL xylanase enzyme assay to be used.
- To optimise antibody concentration for western blot and other western blot parameters.
- To demonstrate experimentally that xylanase is indeed secreted natively in the native C56-YS93 strain.

4.3 METHODS AND MATERIALS

4.3.1 Heterologous expression of xylanase in *E. coli*

E. coli BL21(DE3) cells containing pET28a-β-xylanase1 (from C56) (kindly provided by Dr Giannina Espina-Silva, University of Bath) were grown in LB medium and incubated at 37˚C until the OD_{600} was approximately 0.6, at which point the culture was induced with 100µM IPTG and cultured for a further 3 hours. The cells were collected by centrifugation at 7000x*g* and stored at -20˚C until required.

The plasmid pET28a-β-xylanase1 contained the modified GEOTH_2250 gene, with the part that encodes the native signal peptide removed and replaced with an ATG start codon.

4.3.2 Cell lysis

The frozen cell pellets were re-suspended in buffer [20mM sodium phosphate buffer, pH 7, 0.5M NaCl, 50mM Imidazole, Roche protease inhibitor EDTA free]; Lysozyme was added to 0.2mg/ml, and triton X-100 to 0.1%, and the cells were incubated for approximately 5 minutes. The suspension was then sonicated on ice until the cell lysate was clear and free flowing.

4.3.3 Ni-NTA affinity purification using FPLC

Fast protein liquid chromatography (FPLC) was carried out using an AktaPrime (GE Lifesciences). A 1ml Hi-Trap Chelating HP column (GE Lifesciences) was charged with 0.1M $NiSO_4$ and then equilibrated with low imidazole buffer (20mM Tris-HCl, 0.5M NaCl, 10mM Imidazole, pH7) followed by high imidazole buffer (20mM Tris-HCl, 0.5M NaCl, 500mM Imidazole, pH7), followed by low imidazole buffer. The clarified cell lysate was then loaded onto the column and washed with 10 column volumes (CV) of low imidazole buffer. The protein was then eluted over a 30ml imidazole gradient to 500mM imidazole and collected in 1ml fractions. Fractions corresponding to the protein peaks were analysed by SDS-PAGE and the fractions with the highest yields of the recombinant protein were pooled in preparation for the next step.

4.3.4 Optimisation of Ion exchange chromatography using FPLC

4.3.4.1 Theoretical isoelectric point calculation

The isoelectric point (pI) was calculated using the ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/) or the ExPASy Prot Param tool (http://web.expasy.org/protparam/).

4.3.4.2 Anion-exchange chromatography

A 1 ml HiTrap Q HP sepharose column was equilibrated with no salt buffer (10mM Tris-HCl, pH 8), followed by high-salt buffer (10mM Tris-HCl, 1M NaCl, pH 8), followed by low salt buffer again. The pH was selected based on the protein's theoretical pI of 6.25 as calculated using ExPASy ProtParam (based on predicted amino acid sequence of modified GEOTH_2250 without its signal peptide). The pooled protein from the affinity chromatography was then loaded onto the column and eluted over a 30ml salt gradient to 1M NaCl.

4.3.4.3 Cation-exchange chromatography

A 1 ml HiTrap SP HP sepharose column was equilibrated with low salt buffer (10mM NaOAc, pH 6) followed by high-salt buffer (10mM NaOAc, 1M NaCl, pH 6), followed by another wash in low salt buffer. The pooled protein from the affinity chromatography was then loaded onto the column and eluted over a 30ml salt gradient to 1M NaCl.

4.3.5 Protein dialysis

Fractions from FPLC purification containing xylanase (as determined by UV absorbance and confirmed by SDS-PAGE) were pooled and dialysed using Snakeskin dialysis membrane (10kDa MWCO, Pierce) overnight in 50mM Tris, pH 8.

4.3.6 Raising polyclonal antibodies against xylanase

Prior to immunisation, pre-immune sera from five potential donor rabbits were tested for cross-reactivity to xylanase, or any other proteins in the cell lysate and media from *GT* cultures. The two rabbits with the lowest cross-reactivity to proteins of M_r 40-50kDa were selected for immunisation.

Polyclonal antibodies were raised by immunizing two rabbits with SDS-PAGE gel slices containing 200µg of the purified heterologous xylanase protein per injection (Eurogentec, Belgium). A 3 month programme was used, with the rabbits being injected at day 0, 14, 28 and 56 days, with a final bleed on day 87. This was performed according to regulations on animal experiments.

4.3.7 Xylanase activity assays

4.3.7.1 Dinitrosalycylic acid (DNS) assay

Purified xylanase was incubated with different concentrations of xylan substrate (Xylan from birchwood, Sigma) in McIlvaine buffer (see section 3.4.3) at 60°C in a water bath. 1ml of substrate was incubated with 1ml of purified xylanase at the enzyme concentration as indicated. Samples were taken at selected time points and 200µl was added to 400µl DNS reagent and immediately incubated in a heat block at 100°C for exactly 20 minutes to stop the enzyme reaction, at which point they were then immediately placed on ice to stop the DNS reaction. Samples were then measured at 540nm in a 96 well plate. The amount of reducing sugar liberated was calculated against a standard curve of varying xylose concentrations.

A 0.1 M xylose stock solution was used to make standards of known concentrations (0-20mM). One unit (U) of xylanase activity was defined as the amount of enzyme that catalyses the release of 1 µmol of reducing sugar as xylose equivalents per minute, based on the xylose calibration curve, under the specified assay conditions.

Xylan substrate was prepared by homogenising xylan powder, up to 5g for 5% (w/v), in 80ml of McIlvaine buffer at 60°C. This was then heated to boiling point while stirring, then cooled with continual stirring overnight, then made up to 100ml with buffer.

4.3.7.2 AZCL-xylan assay

Column fractions of purified xylanase were incubated in 2ml microfuge tubes in a final concentration of 2mg/ml AZCL xylan at 60°C for one hour. Tubes were then centrifuged briefly to separate out any remaining non-soluble xylan. 150µl of each was then transferred to 96-well plates and the absorbance measured at 495nm (FLUOstar omega, BMG Labtech)

4.3.7.3 McIlvaine buffer

McIlvaine buffer was used for all enzyme assays and dialysis for enzyme assays. McIlvaine buffer prepared by combining 0.2M Na_2HPO_4 and 0.1M Citric acid in different volumes to obtain the desired pH.

4.3.7.4 Congo red assay

Colonies were grown on agar plates containing 0.1% (w/v) xylan and then dyed with 10% Congo red in water, which binds to xylan. Zones of clearing indicate xylan hydrolysis and thus xylanase activity.

4.3.8 Determination of kinetic parameters

Substrate saturation curves were obtained by plotting initial rates against substrate concentration. Analysis of the enzymatic assay results was carried out using the Enzyme Kinetics module in the SigmaPlot 12 Software (Systat Software, Hounslow, England). The kinetic parameters V_{max} and for each substrate were determined by non-linear regression, fitting the data to the Michaelis-Menten equation: Michaelis-Menten Equation: $v = \frac{Vmax\ [S]}{Km+[S]}$ where v is the initial velocity, V_{max} is the maximum enzyme velocity, [S] is the substrate concentration and K_M is the Michaelis-Menten constant.

4.3.9 Cloning GEOTH_2250 (xylanase) gene into pUCG4.8

The initial strategy was to clone the GEOTH_2250 gene downstream of the native promoter sequence as predicted by the BPROM server (Softberry). Amplification of the gene was carried out using primers AHfw3 and AH2 to amplify the gene from purified C56-YS93 DNA and digesting the resulting product with *Sac*I and *Xma*I. The fragment was then ligated into pUC19 digested with the same enzymes, and the ligation mixture was used to transform *E. coli* JM109. The fragment was then cut out of the pUC19 plasmid by digesting with *Hin*dIII and *Eco*RI and ligated into pUCG4.8 digested with the same restriction enzymes, followed again by transformation of *E. coli* JM109 with the ligation mixture.

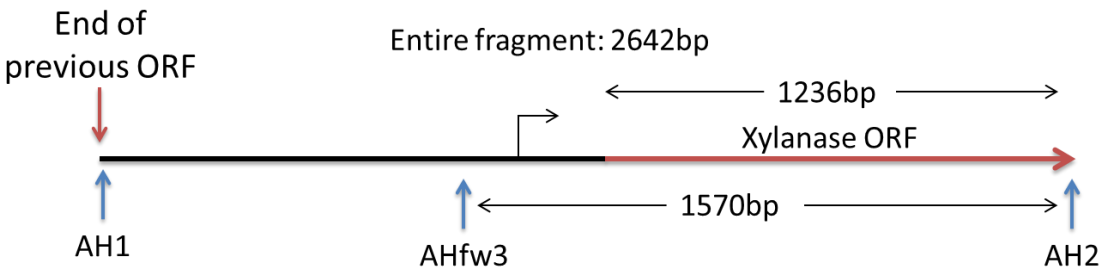

Figure 4.1 ***Genomic organisation of the xylanase gene on the genome of C56. The locations and names of the primers used for PCR amplification are indicated.***

The second strategy was to clone the GEOTH_2250 xylanase gene into pUCG4.8 downstream of the constitutive uracil promoter. The gene was amplified using Xylbstbfor and Xylsac1rev primers to yield a 1260bp product, which was then digested with *Bst*BI and *Sac*I. The pCEX3 plasmid (pUCG4.8 containing the uracil promoter and a cellulase gene; Dr Jeremy Bartosiak-Jentys, unpublished) was digested with *Cla*I and *Sac*I to cleave the existing inserted fragment, and the new insert ligated into the plasmid. Digestion with *Cla*I and *Bst*BI yield complementary sticky ends compatible for ligation.

Table 4.1 Xylanase cloning primers with upstream region.

Primer name	Primer sequence	Feature
AHfw3 (Forward primer)	aaaaGAGCTCGCTCACCGCGCAAATGGCCAG	*Sac*I site
AH2 (Reverse primer)	aaaaCCCGGGCAGCCCGATTGTGTTGGCGAACAG	*Xma*I site
Xylbstbfor	aaaaaTTCGAATGCGGAACGTTTTACGC	*Bst*BI site
Xylsac1rev	gcggacGAGCTCTTATTTATGATCGATAATGGC	*Sac*I site

4.4 RESULTS AND DISCUSSION

4.4.1 Heterologous Xylanase production in *E. coli* and purification

The *xylanase-1* gene without the signal peptide sequence was cloned into pET28c by a colleague (Dr Giannina Espina Silva). The signal peptide at this stage was not included as recombinant xylanase was to be produced in *E. coli,* a Gram-negative bacterium, and the presence of the signal peptide when over-expressing and over-producing a protein may result in deleterious effects such as limiting the growth rate or the formation of inclusion bodies. The resulting gene product contains a poly-histidine tag at the N-terminus, permitting purification on a chelating column charged with nickel.

In order to monitor its location and quantitatively determine its secretion levels, using for instance cell fractionation and pulse-chase analysis, antibodies are required. A first aim was therefore to purify xylanase, and use the purified protein to raise antibodies.

After protein expression, the cells are harvested and lysed to obtain a soluble cell extract, from which the recombinant protein with the histidine tag can be purified through Immobilised-Metal Affinity Chromatography (IMAC). The Nickel-nitrilotriacetic acid (Ni–NTA) matrix within the column selectively binds the poly-histidine affinity tag attached to the N-terminus of the xylanase protein. The purification process involves loading the soluble fraction of the cell lysate onto the column which has been primed using the same buffer as the re-suspension buffer, and the elution buffers, which are at a pH at which the nitrogens in the imidazole ring are in the non-protonated form, and of a relatively high ionic strength in order to reduce non-specific binding of proteins to the resin due to electrostatic interactions. Once the proteins have been loaded onto the column, the poly-histidine tagged target protein binds to the Ni-NTA matrix by interacting with the Nickel. The protein can then be eluted after washing by a ligand exchange step with imidazole, which binds competitively to the Ni-NTA matrix, eluting the target protein. This type of purification is commonly used as a single step purification process, but for applications where purity is vital, such as raising antibodies, a second purification step such as ion exchange chromatography would need to be used.

Ion exchange chromatography is based on the principle that the relationship between the net surface charge and pH is specific for each individual protein, and at a pH value either above or below the proteins pI, the protein of interest may bind to a positively or negatively charged matrix, respectively. As the ionic strength of the elution buffer increases, the salt ions in the buffer compete with the bound proteins, and displace them causing the bound protein to elute and move out of the column.

4.4.2 Affinity Ni-NTA chromatography

E. coli cells containing pET28c-*xylanase1* were grown and induced at OD_{600} of 0.6 with 0.1mM IPTG, and then grown for a further 3 hours. The cells were then harvested and lysed to obtain the soluble fraction containing the heterologous xylanase.

Recombinant protein from the soluble fraction of lysed *E. coli* cells was purified on a nickel charged column and eluted using an imidazole gradient. Figures 4.2 and 4.3 shows the chromatogram of the FPLC and the corresponding SDS-PAGE gel of the chromatogram peak fractions, confirming that the main chromatogram peak represents eluted protein of the correct size (MW 45kDa approximately) and demonstrates the purity of the eluted protein. The fraction corresponding to the smaller peak at around 13ml appears to contain a small amount of protein of the correct size, along with several larger bands. These could be other unrelated proteins, or they could be aggregates of xylanase - hence the larger size. Despite this, it was superfluous to optimise the affinity purification further due to the very high yields present in the fractions corresponding to the larger peak on the chromatogram. Nevertheless, further purification would need to be carried out. Even though the fractions appear to be relatively pure, for the purposes of raising antibodies, high levels of purity are essential, which is why ion exchange chromatography was selected and carried out for the next step.

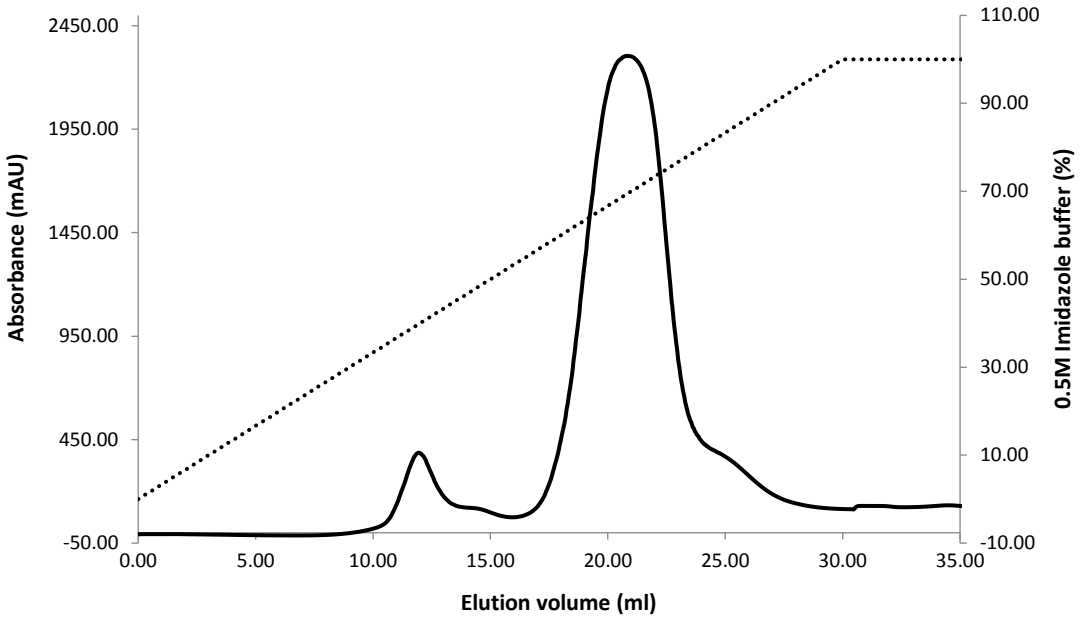

Figure 4.2: Chromatogram of affinity Ni-NTA chromatography of soluble cell lystae from E. coli expressing xylanase. The solid black line represents absorbance at 280nm corresponding to eluted protein and the dotted line represents the gradient of high imidazole buffer.

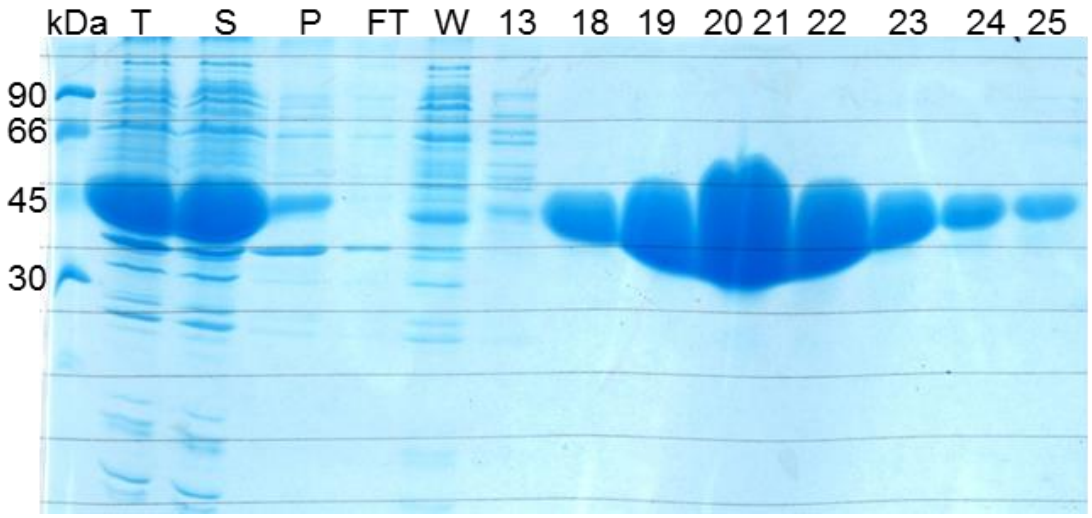

Figure 4.3: SDS-PAGE of the elution peaks corresponding to elution volumes 13 and 18-25 from the Ni-NTA chromatography column. Fractions: cell lysate total (T), soluble(S), and insoluble (P); column flow-through (FT) and column wash (W). The lanes T and S shows a large amount of soluble protein at 45kDa, which corresponds to the correct predicted MW of the heterologous xylanase. There is significantly less of the corresponding band in lane P lane, indicating that very little target protein was insoluble.

4.4.3 Ion-exchange chromatography

Ion-exchange chromatography was selected for the next step in purifying the recombinant xylanase protein. The theoretical pI of the expressed xylanase-1 was calculated to be 5.87 (Expasy). As such, the pH for the buffers for purification should be around 2 pH units greater or lower than that of the calculated pI (Roe, 2001).

Initial runs of ion-exchange chromatography using a HiTrap Q HP column (GE Lifesciences) and pH8 buffers resulted in poor binding of the protein to the column and, if any, the protein eluted during the wash or early on in the elution gradient (data not shown). One possible reason for poor binding is a pH close to the actual pI of the protein (which is not necessarily the same as the calculated theoretical pI), resulting in poor binding to the column). As such, it was decided to utilise a HiTrap SP HP (GE Lifesciences) cation exchange column instead.

The pH selected for the cation exchange was pH 4. At this pH however, the protein bound tightly to the column and an elution peak could only be seen at 1M salt concentration (data not shown). Different pH values were tested and finally pH 6 was selected. Despite being close to the theoretical pI, the protein bound tightly and eluted in a sharp peak at 1M salt as can be seen in Figure 4.4. The fractions corresponding to the peak were pooled, dialysed, and analysed by SDS-PAGE and Coomassie staining to assess the purity of the yield. Different amounts were run on the gel so as not to overload each lane, and it was determined that the absence of any additional bands in any of the lanes indicated sufficient purity to raise antibodies (Figure 4.5).

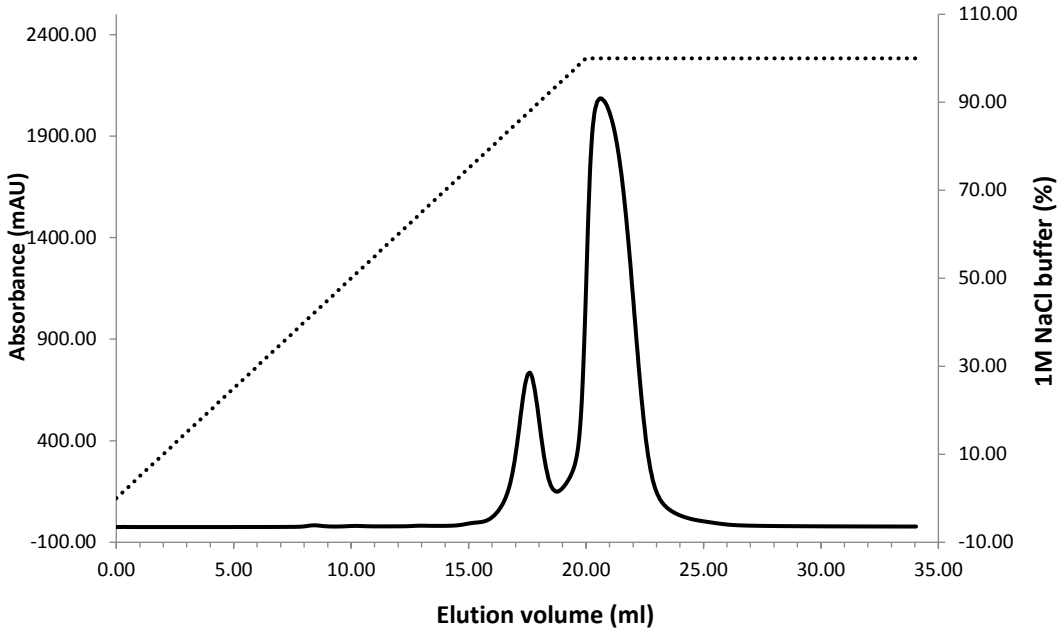

Figure 4.4: Chromatogram of cation-exchange chromatography of pooled and dialysed protein from Ni-affinity chromatography. The solid black line represents the absorbance at 280nm and the dotted line represents the elution gradient of high salt buffer.

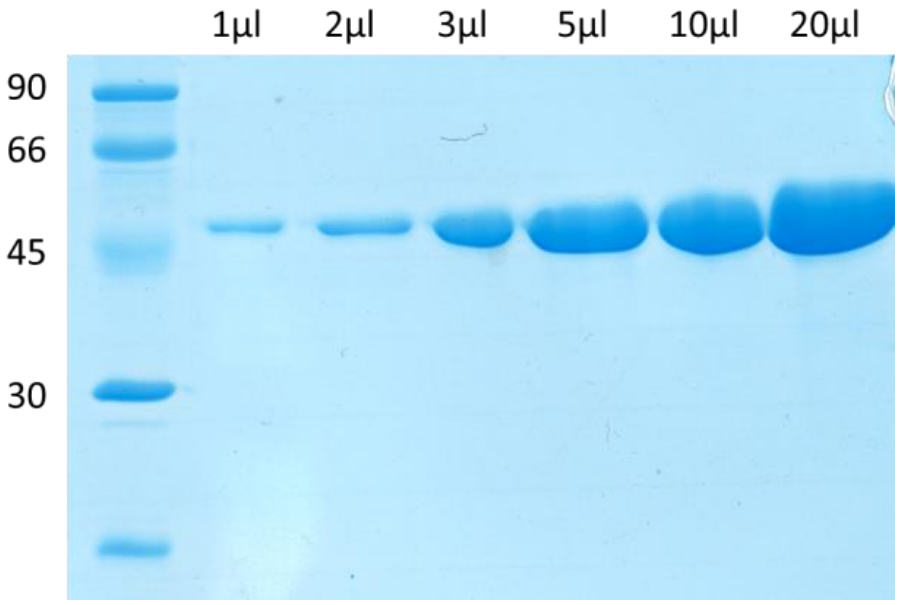

Figure 4.5: SDS-PAGE of purified protein from pooled 20-23ml fractions from IEX chromatography at different amounts showing purity of the protein necessary for raising polyclonal antibodies.

4.4.4 Activity of heterologous xylanase

The purified xylanase was assayed by measuring the formation of reducing sugars using the DNS assay, whereby the production of reducing ends from the hydrolysis of the xylan polymer react with the DNS, resulting in a change or shift in the absorption spectrum from yellow to red. In this reaction 3,5 dinitrosalicylic acid is reduced to 3-amino, 5-nitro salicylic acid. The purified enzyme was firstly assayed at different concentrations as shown in Figure 4.6. A linear relationship between the rate of hydrolysis and the amount of enzyme in the assay was evident. This suggests that the rate limiting factor of the assay was the amount of enzyme present in the assay and, that initial rates were being measured.

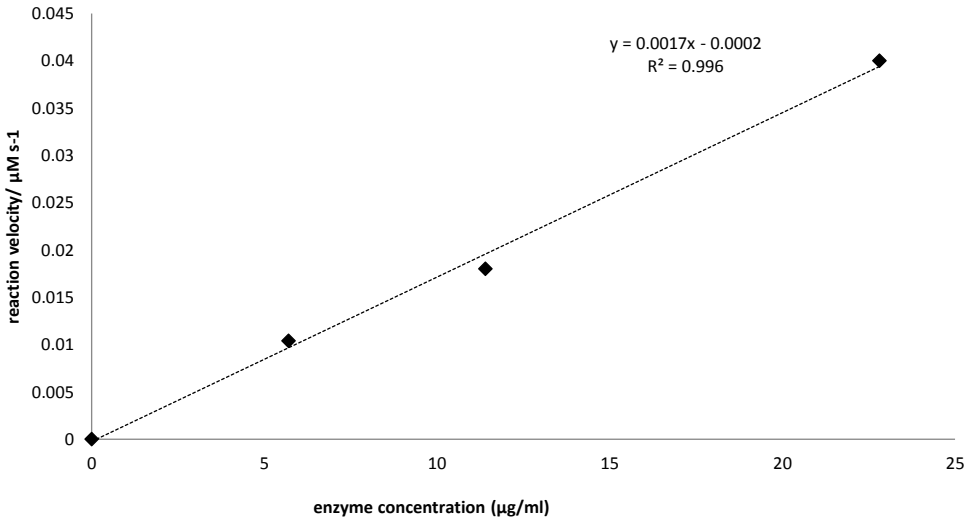

Figure 4.6: Initial rates of xylanase activity at different purified enzyme concentrations.

Further characterisation of the heterologous xylanase included determining the kinetic properties using the DNS assay where reaction velocity rates at different substrate concentrations were obtained. The kinetic parameters were then calculated using the Hanes Woolfe plot (Figure 4.7), yielding a V_{max} of 0.063µM/s and a K_M of 0.02 g/ml xylan or 20mg/ml xylan at 60°C.

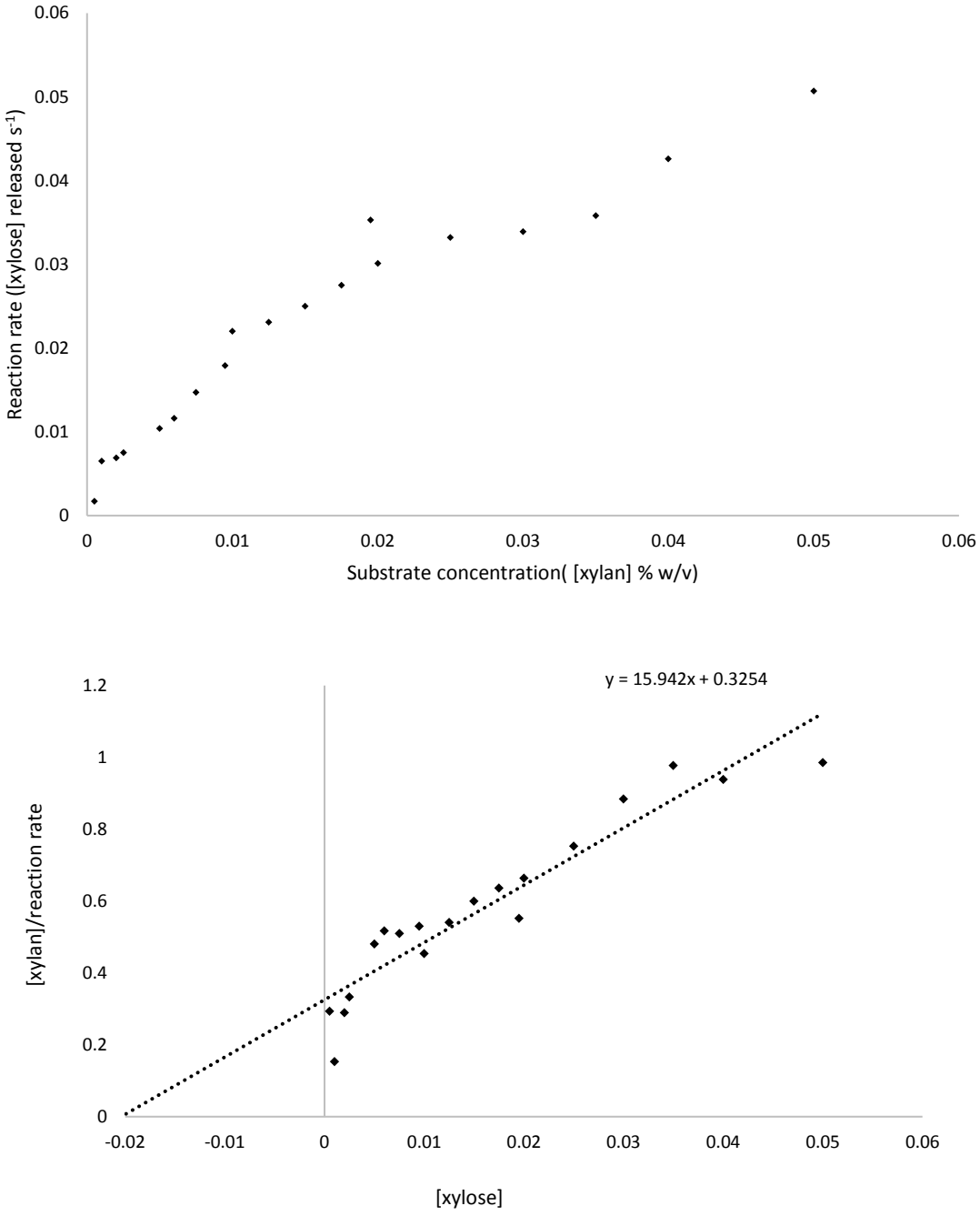

Figure 4.7: Michaelis Menten graph (top) and Hanes-Woolfe plot (bottom) of heterologous xylanase activity at 60°C.

Optimal pH and temperature for the heterologous xylanase were determined by assaying the enzyme using AZCL xylan substrate, which is an insoluble xylan substrate, cross-linked with blue molecules, which when hydrolysed by endo-xylanase, liberates water-soluble dyed fragments which absorb at 590nm. The absorbance of the liberated dye in the supernatant from the assay was used as an indicator of relative activity. Purified heterologous xylanase was incubated at different pH values and temperatures as indicated in Figures 4.8 and 4.9. The results indicate that the optimal pH for xylanase activity is between pH8 and pH9, and the optimal temperature for xylanase activity around 50°C.

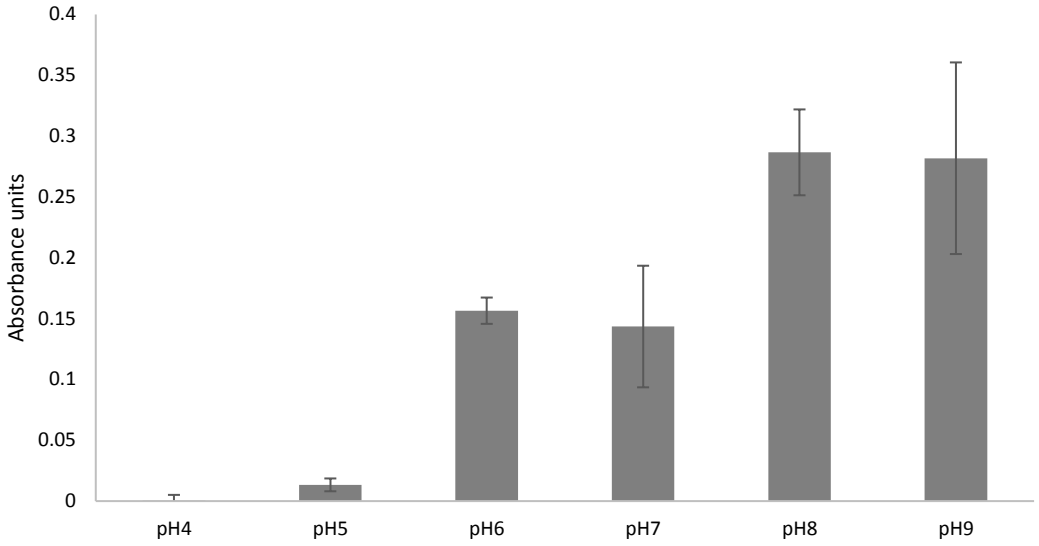

Figure 4.8 Dependence of xylanase activity on pH. Purified xylanase samples were incubated at 60°C with 2mg/ml AZCL-xylan for one hour at different pH in McIlvaine buffer. After incubation, absorbance was measured to determine relative levels of xylanase activity.

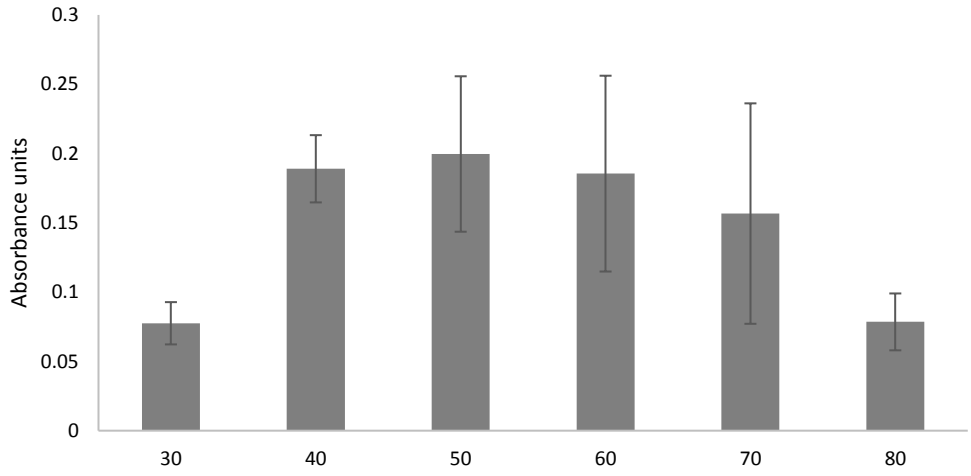

Figure 4.9: Dependence of xylanase activity on temperature. Purified xylanase samples were incubated at 60°C with 2mg/ml AZCL-xylan for one hour at different temperatures at pH 7 in McIlvaine buffer. After incubation, absorbance was measured to determine relative levels of xylanase activity.

The enzyme characteristics were compared with those of other xylanases from other *Bacilli sp.* and it was found that several other xylanases have a similar optimal pH value to that of the xylanase reported here (table 4.2).

Table 4.2: Optimal pH, optimal temperature and Km of xylanase from some Bacilli and Geobacilli

Organism	pH	Temp	Km	reference
Bacillus sp. SN5	7	40	0.6mg/ml	(Bai et al., 2012)
Geobacillus sp. WSUCF1	6.5	70	1.75mg/ml	(Bhalla et al., 2014)
Geobacillus sp. 71	8	75	0.425mg/ml	(Canakci et al., 2012)
Bacillus circulans	7	60	9.9mg/ml	(Heck et al., 2006)
Bacillus arseniciselenatis **DSM 15340**	8	50	5.26mg/ml	(Kamble and Jadhav, 2012)
Geobacillus sp. WBI	7	70	0.9mg/ml	(Kamble and Jadhav, 2012)
Paenibacillus macquariensis	9	50	2.2mg/ml	(Sharma et al., 2013)
Geobacillus thermodenitrificans TSAA1	9	70	0.625mg/ml	(Verma et al., 2013)
Geobacillus thermoleovorans	8.5	80	2.6mg/ml	(Verma and Satyanarayana, 2012)
Geobacillus thermodenitrificans **AK53**	5	70	4.34mg/ml	(Irfan et al., 2016)
Bacillus sp. **JYM1**	5	50	Not reported	(Lee et al., 2016)
Geobacillus thermoglucosidasius C56-YS93	8	50	20mg/ml	This study

4.4.5 Xylanase secretion by C56

Prior to the purification of heterologously produced xylanase in *E. coli*, and subsequent raising of antibodies, xylanase production and activity in the C56-YS93 strain was first tested using a simple plate assay using Congo red as a stain to reveal zones of xylan hydrolysis. As shown in Figure 4.10, a zone of clearing can be seen around the streak of colonies indicating xylanase activity.

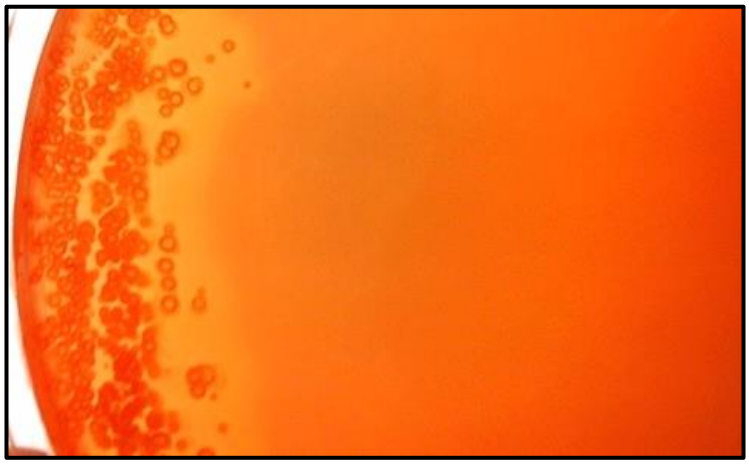

Figure 4.10 Congo red stained agar plate containing 0.1%(w/v) xylan with GT C56-YS93

Once xylanase activity in *GT* C56-YS93 was confirmed by the Congo red plate assay, the protein was then heterologously produced in *E. coli* and purified as described above. The purified protein was analysed by SDS-PAGE to verify the absence of any other bands (Figure 4.4).The purified protein was then used to raise antibodies (Eurogentec) for Western-blot analysis as per Eurogentec's instructions.

Xylanase production by C56 was detected using Western blotting with the antibodies raised as discussed in the previous section. The antibody concentration for Western-blot analysis was optimised by varying samples and antibody concentration. The ideal concentration for both primary and secondary antibody is 1:10,000 and exposure time ranges from 10 seconds to 1 minute for manual exposure with photographic film.

The blot shown in Figure 4.10 shows the presence of xylanase in the extracellular milieu from C56 but not in TM242, confirming the affinity of the antibody to xylanase that is natively produced in C56. Furthermore, it also shows the absence of any xylanase in either cell pellet, indicating that at the levels of xylanase produced, the secretion of this protein is efficient. However, the blot also shows a strong band at a size (47kDa) larger

than the secreted xylanase (44.6kDa); this is likely to be the result of non-specific binding to a similar protein. Note that this band was absent in western blots using pre-immune sera to determine any cross-reactivity prior to immunisation with purified xylanase (not shown).

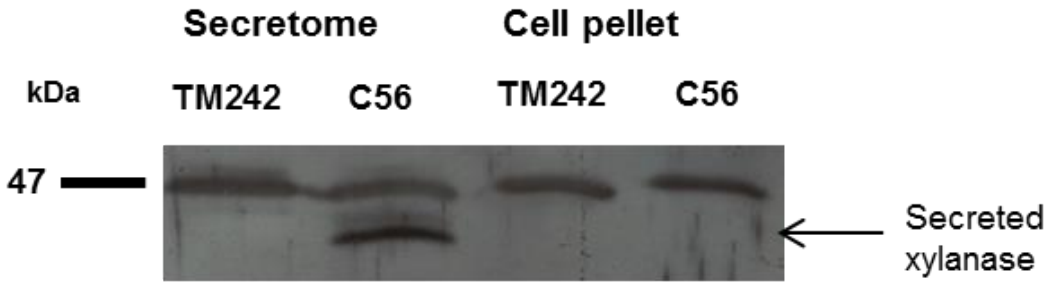

Figure 4.11: Western blot analysis comparing supernatant (secretome) and cell pellet fractions from TM242 and C56YS93 strains. The arrow indicates the band representing secreted xylanase (44.6 kDa) which is only present in the supernatant (secretome) from C56-YS93. The band corresponding to a protein of larger size (47kDa) is due to non-specific binding and can be seen in both the secretome and cell pellet of both TM242and in C56-YS93

It was also decided to characterise native xylanase production in *GT* with either xylan or glucose as the carbon source. It was during this process that it was discovered that the promoter appeared to be subject to catabolite repression. Figure 4.11 shows a western blot analysis of the C56-YS93 strain grown in ASM medium with different concentrations of xylan and glucose. As can be seen here, in the wild-type xylanase-producing strain grown in ASM medium with 1% glucose, xylanase is not produced and secreted. However, when grown with no additional sugar (only yeast extract as the carbon source), or with xylan as the additional sugar source, xylanase is produced, and secreted into the extracellular milieu. This suggests that the native promoter is subject to catabolite repression, and is non-inducible by xylan, making the promoter unsuitable for over-expression and production of xylanase to investigate secretion bottlenecks, as the mechanism of catabolite repression of this promoter is not understood.

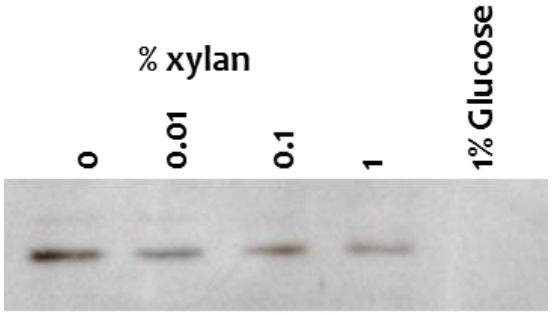

Figure 4.12: Western blot of supernatant fraction from GT C56 YS93 strain grown in ASM medium to OD_{600} of1.5, with varying concentrations of xylan and 1% glucose.

4.4.6 Construction of xylanase producing TM242 strains

The TM242 strain, which was the strain used by TMO renewables and is a derivative of the WT11955 strain, does not encode any secreted hemicellulases, nor does it display any xylanolytic activity despite being able to grow on pentose sugars such as xylose, and on xylose oligomers, or even xylan. This is most likely due to the presence of xylose monomers and oligomers present in the purified xylan substrate, which WT11955 strain and its derivatives can metabolise, and not due to xylanase activity.

Initially, the xylanase gene was to be cloned into the pUCG3.8 vector, including the promoter and upstream region, which may contain essential regulatory elements. The promoter region was identified using the BPROM promoter prediction online programme by Softberry (Salamov, 2011), the intention being that when expressing from a high-copy number plasmid, the expression levels would be higher than those if expressed from the chromosome with a single copy. The xylanase gene, located on a 2642 bp fragment, was cloned into the *E. coli* puC19 vector. The genomic organisation of the gene as in the native organism is shown in Figure 4.1 along with the locations of the primers used to amplify the larger fragment. However, we were unable to clone the 2642 bp fragment or a smaller fragment containing the region upstream of the xylanase gene into the shuttle vector pUCG3.8 vector.

Therefore, it was decided to clone the xylanase open reading frame downstream of a constitutive promoter. The promoter was obtained from Dr Ben Reeve, Imperial College London, and its sequence was modified from the region upstream of the Uracil phosphoribosyl transferase (involved in the uracil salvage pathway) in *GT* NCIMB 11955; it was shown to be constitutively active with moderate expression levels, with consensus -10 and -35 boxes (Dr Ben Reeve, personal communication).

Production of xylanase using the newly constructed plasmid (denoted pUCG4.8xyl) in TM242 was compared with the endogenous levels of expression in C56. This was done by Western blotting, the result of which was quantified using Image Studio Lite software (version 5.2). As can be seen in Figure 4.13, when the xylanse gene is expressed under the control of the uracil promoter on the pUCG4.8 plasmid in TM242, the xylanase levels when grown in ASM medium with 1% xylose are over 50% higher than the xylanase

levels when in the same growth conditions in C56-YS93. This is probably due to not only the promoter being a strong constitutive promoter, but also the high copy number plasmid results in an increased copy number of the gene, and consequently, an increased level of expression. This presents a satisfactory starting point to investigate secretion bottlenecks in *GT* TM242.

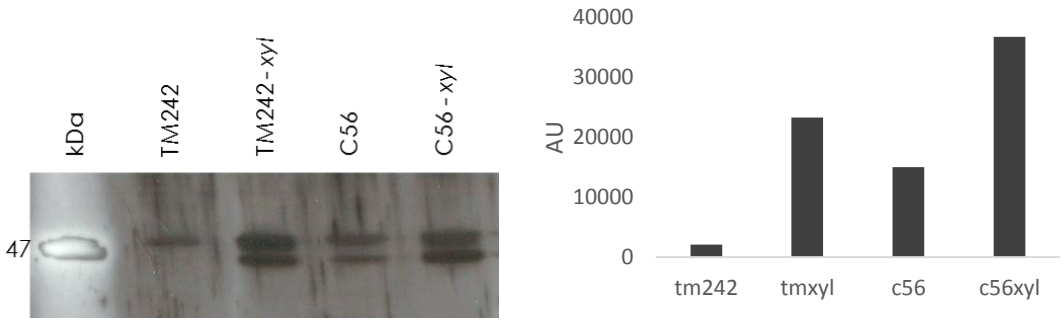

Figure 4.13: Western blot analysis of the media fraction of TM242 and C56-YS93 (left) and densitometry analysis of the western blot (right). The y-axis of the densitometry data represents arbitrary signal units. Lane 1: MW ladder; Lane 2 TM242; Lane 3: TM242 with pUCG4.8 with uracil promoter expressing xylanase gene (TM242-xyl); Lane 4: C56-YS93 strain expressing xylanase natively; Lane 5: C56-YS93 with pUCG4.8 with uracil promoter expressing xylanase gene (C56-xyl).

4.5 CONCLUSIONS

The xylanase enzyme heterologously produced in *E. coli* (without the signal peptide) was successfully characterised, and found to have optimal temperature between 50°C and 60°C, which is ideal for production in an ethanol producing strain with fermentation temperatures around 60°C. The enzyme also appears to retain around 50% activity even at 80°C. In terms of optimal pH, the heterologously produced xylanase was found to have optimal pH at between pH 8 and9, which is similar to several other xylanases found in other *Geobacilli.* This is ideal if pre-treatment selected for the lignocellulosic feedstock is mild alkaline treatment, as the xylanse produced would be active at mildly alkaline pH. However, the activity of the xylanase was not characterised at pH higher than 9, so it may be worthwhile investigating activity at higher pH.

The xylanase gene (with the signal peptide sequence) from *GT* C56-YS93 was successfully cloned into the pUCG4.8 vector downstream of the uracil promoter and subsequently inserted into *GT* TM242. This provides an ideal starting point to investigate secretion bottlenecks in *GT* TM242 through over-production of secreted xylanase.

CHAPTER 5: ANALYSIS OF XYLANASE SECRETION BY *GEOBACILLUS THERMOGLUCOSIDASIUS* TM242

5.1 INTRODUCTION

5.1.1 Protein secretion in *Geobacillus thermoglucosidasius*

Geobacillus thermoglucosidasius is a Gram-positive (monoderm), thermophilic bacterium, which is part of the same family, *Bacillaceae*, as the well-studied *Bacillus subtilis* (*BS*). Several *Geobacilli sp* have been of interest for industrial purposes, especially as a potential source of thermostable enzymes for various industrial applications such as for detergents, paper bleaching, baking, brewing, animal feed, and biofuel industries. The genes for these enzymes have been successfully expressed heterologously in mesophilic hosts such as *BS* (Zouari Ayadi et al., 2008, Finore et al., 2011, Canakci et al., 2012) or *Pichia pastoris* (Sun et al., 2007, Yamada et al., 2016) .

In terms of protein secretion in *GT*, however, very little has been described in the literature, although *BS* is very well studied in comparison, and is a model for protein secretion in Gram-positive organisms. As shown in Chapter 3, the bio-informatics comparison of the secretory machinery components and the related proteins of *GT* and *BS* did not show any difference in terms of the secretion machinery components. A few studies have described using *Geobacillus sp.* as a host for protein secretion such as glycosyl-hydrolase secretion in *GT (Bartosiak-Jentys et al., 2013)* and heterologous cellulase production in *Geobacillus kaustophilus (Suzuki et al., 2013),* which when expressed in *EC* were insoluble. As such, it is important that protein secretion, and the potential bottlenecks in *GT* is better understood. As it stands, *GT* is a good candidate for the production and secretion of industrially relevant thermophilic enzymes, but also a possible platform organism for consolidated bioprocessing for the production of organic compounds from waste products.

5.1.2 Potential bottlenecks in protein secretion

Protein secretion is a multistep process that begins with transcription of DNA coding for a secreted protein. This is transcribed into an mRNA that contains the sequence for the secreted protein including the signal peptide, which as discussed is a stretch of amino acids at the N-terminus of a protein destined to be translocated. The mRNA is then translated at the ribosome where it is targeted to the secretion machinery by one of two

pathways of which we are currently aware: the co-translational pathway and the post-translational pathway.

Co-translational translocation is mediated by the signal recognition particle (SRP) which binds to the signal peptide of the nascent chain as it emerges from the ribosome. The SRP then guides the nascent chain – ribosome complex to the SRP docking protein, FtsY, which is adjacent to the Sec machinery. Once docked, the emerging polypeptide can then be translocated through the SecYEG translocon. This pathway is better understood and better described in the literature than the post-translational translocation pathway. However, this does not necessarily indicate that it the more predominantly utilised pathway. In eukaryotic cells, translation is arrested until the SRP is bound to the docking protein, whereby translation resumes and is carried out concurrently with translocation (Luirink and Sinning, 2004, Zanen et al., 2006a, Shan and Walter, 2005). In bacteria, this translation arrest does not occur and the SRP pathway is predominantly utilised for transmembrane proteins where the signal peptide may or may not be cleaved(Shan and Walter, 2005). As translocation is coupled to translation, the energy for translocation is driven by translation.

The post-translational pathway is where the nascent chain emerging from the ribosome is kept in a translocation-competent state by chaperones. That is, the chaperones prevent folding so it can be translocated as a polypeptide chain through the SecYEG translocon. In *E. coli*, this process is relatively well understood. This process is mediated by a protein known as SecB, which interacts with the nascent chain as it emerges from the ribosome, and then guides the unfolded polypeptide to the Sec machinery whereby SecB interacts with SecA. In *B. subtilis*, where SecB is absent, it has been shown that SecA may be involved in post-translational targeting to the Sec translocon (Huber et al., 2011, Muller et al., 2000b). Another protein that is thought to be a chaperone involved in targeting is the CsaA protein, which as briefly mentioned in Chapter 3, is a chaperone that relieves secretion stress in a *secB* knockout in *E. coli.* The mechanism of action of CsaA is poorly understood thus far, and its role in protein secretion is yet to be defined in *B subtilis* (Muller et al*.,* 2000a*)*. That being said, it is thought that many secreted proteins fold in the cytoplasm, but unfold, in an unravelling fashion, as they are being pushed through the SecYEG translocon (Lycklama and Driessen, 2012). The energy for

this pathway is provided by SecA, which is an ATPase that hydrolyses ATP, which results in conformational changes that push the protein through the translocon in a ratchet like fashion.

It is thought that each pathway is specific for different subsets of proteins, although there is little evidence to date that indicates what characteristics of the signal peptide or polypeptide confer the correct signposting to each pathway. As such, each pathway poses unique challenges if the secretion system were to become stressed due to over-secretion of a protein or several proteins.

Once the polypeptide begins translocation through the translocon, the signal peptide is then shunted sideways into the cell membrane, where the hydrophobic H-region forms an α-helix. The signal peptide is then cleaved by a signal peptidase, releasing the protein into the cell wall, where it folds. Signal peptidases, like all other enzymes, can be substrate limited (Tjalsma et al., 1997). The journey does not end there, however; some proteins undergo further processing, some are bound to the cell wall, and some diffuse into the extracellular milieu. However, in some cases, especially when over-produced, the secreted protein could aggregate, and become a target for hydrolysis by extracellular and cell wall bound proteases (Schallmey et al., 2004, Margot and Karamata, 1996). Nevertheless, chaperones and foldases play a role in both helping the folding of secreted proteins, and also rectifying incorrect folding of misfolded proteins, preventing the protein from being degraded. Even post-translocation, several challenges are posed for the over-produced protein, whether that is native or heterologous. Signal peptidases, signal peptide peptidases, and chaperones such as PrsA, all have maximum capacities, which could cause a bottleneck if over-burdened (Chen et al., 2015b).

Once in the cell wall, secreted proteins can be anchored to the cell wall via specific mechanisms like sortases, which recognise and cleave the L-P-X-T-G motif in the C-terminal part of specific proteins and covalently attach this to the cell wall peptidoglycan via a trans-peptidation reaction (Ton-That et al., 2004, Paterson and Mitchell, 2004, Schneewind and Missiakas, 2014). It remains unclear how extracellular proteins make their way through the cell wall into the extracellular milieu, as although the cell wall is porous in nature, a study using purified peptidoglycan estimated the permeability of *BS*

peptidoglycan is limited to globular proteins with mass of approximately 25kDa (Demchick and Koch, 1996). With this is mind, it is difficult to envisage proteins larger than 25kDa being transported through the cell wall passively and unaided. However, several factors will influence the permeability of the cell wall during the life cycle of a bacterial cell: the level of cross-linking in the peptidoglycan and whether or not there are bridges between the cross-linking peptides, and what growth stage the cell is in, as that will directly affect the cell wall condition. For example, during exponential growth, the cell wall is regularly renovated during binary fission, possibly making the cell wall at these locations more permeable to secreted protein diffusion (Silhavy et al., 2010). Furthermore, the localisation of the SecYEG machinery appears to be spatially organised in a spiral-like fashion around the cell. This is not synchronised with the formation of the cytoskeletal structure formed by MreB and MbL proteins, but rather more reminiscent of the cable-like structure the cell wall in *BS* adopts (Campo et al., 2004). Moreover, it has been observed that, in the coccus-shaped *S. pyogenes*, the secretion machinery localises in clusters around a so-called 'ex portal' region which is at the nascent septum, suggesting that protein secretion in Gram-positive (monoderm) bacteria occurs in regions where the cell wall is less rigid, thus expediting diffusion of proteins into the extracellular milieu (Rosch and Caparon, 2004).

Proteins that make it to their final destination of the extracellular milieu, then encounter an environment starkly different from that within the cytoplasm, or even the cell wall. Depending on the growth media, the pH and the salt concentration could be significantly different, and the presence of other proteins such as proteases and the presence of enzyme inhibitors may play a role in whether the protein survives, and for how long. In the case of industrial production of secreted proteins, or other organic chemicals, the medium is reasonably well regulated for salt concentration and buffered for pH.

5.1.3 Cell fractionation

Cell fractionation, as a technique to study protein secretion, that is the movement of proteins within and outside of the cell, was first described by George E. Palade, who later went on to win the Nobel prize for his ground-breaking work in cell biology (Monneron et al., 1972). He had originally used cell fractionation to separate out organelles in eukaryotic cells in order to elucidate their function (Monneron et al., 1972).

Since then, methods have been developed to isolate and purify the different cellular sub-compartments of various types of cells, including Gram-negative and Gram-positive bacteria (Chassy, 1976, Zuobi-Hasona and Brady, 2008).

Cell fractionation of Gram-positive bacterial cells involves splitting the cells up into the extracellular milieu, cell wall, cytoplasmic and cell membrane fractions, as depicted in Figure 5.2-. It involves creating protoplasts by digesting the cell wall with lysozyme, an enzyme that hydrolyses the 1,4-β-linkages between N-acetylmuramic acid and N-acetylglucosamine in peptidoglycan. This step will also liberate any cell wall associated proteins, or proteins that have misfolded and become trapped within the cell wall, or proteins that are being degraded by cell wall proteases. The protoplasts are then lysed and separated into the cytoplasmic fraction, containing cytoplasmic proteins and proteins that have yet to be targeted to the membrane for translocation, and the cell membrane fraction, which contains membrane proteins, and proteins associated or coupled to membrane proteins, such as proteins mid-translocation. For the purposes of the cell fractionation analysis carried out in this work, it is not essential to keep the integrity of the membrane proteins, such as the SecYEG translocon, as the target protein here is xylanase mid-translocation. The membrane fraction is then re-suspended in a mild detergent, which releases any membrane-associated proteins.

5.1.4 Pulse-chase analysis

Pulse-chase analysis refers to a technique whereby a cellular process, such as protein translocation, can be examined over time, by exposing the cells to a labelled compound that is to be incorporated into the molecule of interest, followed by addition of an excess of the same compound, but unlabelled. In the context of protein secretion, the cells are exposed to a radiolabelled amino acid for a set period of time, known as the pulse, followed by the addition of an excess of the non-radiolabelled version of the amino acid, known as the chase. The amino acid is utilised by the cell for protein synthesis. Therefore, in the set time where the cells are exposed to the radiolabelled amino acid, every protein synthesised should incorporate the radiolabelled amino acid, and thus be radioactive. Following this, an excess of the non-radiolabelled amino acid is added, effectively stopping further incorporation of the radiolabelled amino acid into newly synthesised proteins.

Pulse chase has been used for several applications, including determining the half-life of proteins(Simon and Kornitzer, 2014), studying protein folding kinetics (Nissley et al., 2016), or determining the localisation of protein folding or assembly (Kim and Arvan, 1991, Woolhead et al., 2000). Pulse chase experiments in *BS* have been well described in the literature, and have been used successfully to investigate bottlenecks in protein secretion, by determining processing times for secreted proteins to be cleaved and processed from the precursor state, and the mature state without its signal peptide(Bolhuis et al., 1999b), or to evaluate secretion kinetics when genes for secretion machinery components are over-expressed or knocked out.

In the first step in a pulse chase experiment the cells are cultured to mid-log phase, and they are then harvested by centrifugation, and re-suspended in a defined medium that lacks the amino acid which is to be used as the label. This is usually methionine as the sulphur isotope is radioactive, and provides good resolution during subsequent fluorography steps. Methionine is also present in almost every single protein. The cells are incubated in the starvation media for typically an hour to deplete any existing methionine in the cells. The next step is the incubation with the radioactive methionine for a short specific time, followed by the addition of an excess of non-radioactive methionine. Samples are then taken at specific intervals and immediately TCA precipitated to abruptly stop protein secretion and all other cellular processes. The precipitated samples are then re-solubilised, and immune-precipitated with the appropriate bait antibody followed by protein-A affinity beads. The immune-precipitate is then separated using SDS-PAGE, and the gel then dehydrated and exposed to photographic film and analysed.

The samples for pulse chase are not separated into cell and media fractions, rather they are whole samples. To analyse the kinetics of protein secretion, the ratio between precursor protein, with signal peptide intact, and mature protein, with the signal peptide cleaved, is measured. The ratio changes with each time-point, as the labelled species of the target protein is cleaved by signal peptidase and translocation has terminated.

5.2 AIMS

The aim of the work described in this chapter is to identify potential bottlenecks in secretion in *GT* by over-expressing and over-producing the model secreted enzyme, xylanase. The enzyme will be expressed both with and without the native signal peptide. This is to demonstrate the levels of the xylanase enzyme and its activity in the different fractions when secreted, and for the strain expressing xylanase without the signal peptide, xylanase levels within the cytoplasm, and representative of total levels of xylanase. Figure 5.1 depicts a simplified version of the planned workflow showing xylanase being produced with and without its signal peptide, and being targeted to the secretion machinery, or not.

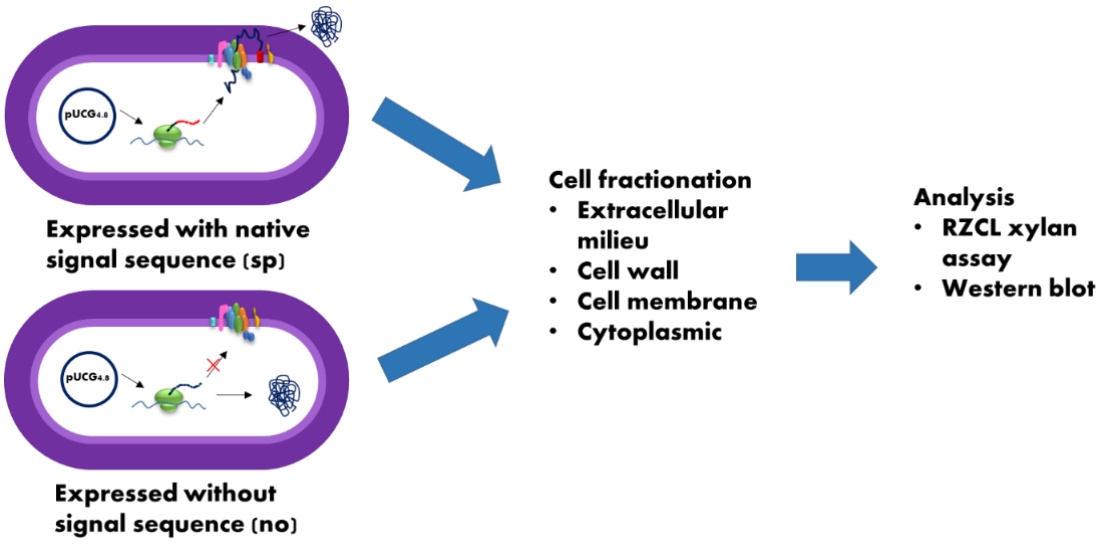

Figure 5.1: Workflow depicting xylanase production and translocation

5.3 METHODS AND MATERIALS

5.3.1 Pulse chase analysis

5.3.1.1 Radiolabelling

Cells of *G. thermoglucosidasius* were grown in rich ASM medium supplemented with 0.5% yeast extract until an optical density at 660 nm (OD_{660}) of 1.0-1.5 was reached. Cells were collected by centrifugation, washed in minimal medium, and then re-suspended in minimal medium (OD_{660} ~0.8). Cells were incubated for 1 hour at 45 ˚C in a shaking incubator. Cells were pulsed for 5 minutes with 40 μCi [^{35}S]-methionine/ cysteine mixture (Perkin Elmer, Waltham, Massachusetts, USA) per ml culture medium. Next, an excess of non-radioactive methionine was added (1 mg/ml), and 1 ml samples were taken after 0, 10, and 30 minutes. Samples were immediately mixed with cold trichloroacetic acid (TCA; final concentration 15%), and kept on ice for at least 30 minutes.

5.3.1.2 Immunoprecipitation

Cells and proteins were pelleted by centrifugation, and washed twice with ice-cold acetone. Pellets were re-suspended in 50 μl buffer (50 mM Tris-HCl pH 8, 1% SDS, and 1 mM EDTA) and boiled for 10 minutes. Next, 1 ml Triton buffer (2% Triton X-100, 50 mM Tris-HCl pH 8, 150 mM NaCl, and 0.1 mM EDTA) was added, and insoluble precipitates were removed by centrifugation. Samples were incubated for 2 hours at room temperature in the presence of XylA-specific polyclonal antibodies. Next, 5 mg protein A sepharose washed in Triton buffer was added, and the samples were incubated for a further 2 hours.

The protein A sepharose beads were washed three times with Triton buffer and boiled in 40 μl SDS-PAGE loading buffer. Samples were visualised using SDS-PAGE and a Fuji FLA-5000 phosphorimager.

5.3.2 Cloning the xylanase gene from C56-YS93 into puCG4.8 vector

The xylanase gene was amplified from *Geobacillus thermoglucosidasius* C56-YS93 chromosomal DNA by PCR using specific primers with a *Bst*BI site at the 5' end and a *Sac*I site at the 3' end. Purified PCR products were then digested with *Bst*BI and *Sac*I, while purified target vector pUCG4.8-RPLS-sfGFP (with RPLS constitutive promoter and superfolder GFP) was digested with *Cla*I and *Sac*I. The fragments were then purified and ligated to produce pUCG4.8-rpls-xyl and pUCG4.8-rpls-xyl-sp$^-$. The ligation mixture was then used to transform chemically competent *E. coli* JM109 cells. The colonies were screened using UV light to determine undigested pUCG4.8-sfGFP and the desired ligation product. Selected colonies were then screened by colony PCR using the M13 universal primers.

Table 5.1 List of primers to amplify Xylanase-1 gene from GT C56-YS93

Primer name	**Primer sequence**	**Feature**
Ahfw3	AAAAGAGCTCGCTCACCGCGCAAATGGCCAG	*Sac*I site
AH2	AAAACCCGGGCAGCCCGATTGTGTTGGCGAACAG	*Xma*I site
GHspF	TTCGAAATGGCAGATACGGCTTCCTAT	*Bst*BI site
Xylsac1rev	gcggacGAGCTCTTATTTATGATCGATAATGGC	*Sac*I site
M13 forward -21	GTAAAACGACGGCCAGTG	Universal primers
M13 reverse -48	GGAAACAGCTATGACCATG	Universal primers

5.3.3 Cloning the *prsA* gene

The *prsA* gene, including the native ribosome binding site, was amplified using primers listed in table 5.2 as described in section 2.7. The PCR product was then cleaned up and digested with SacI and EcoR1. The plasmid, pUCG4.8-rpls-xyl was digested with the same enzymes to produce compatible sticky ends and the two fragments ligated to produce pUCG4.8-rpls-xyl-prsa. The ligation mixture was then used to transform chemically competent *E. coli* JM109 cells. Selected colonies were then screened by colony PCR using the M13 universal primers.

Table 5.2 List of primers to amplify prsA gene from GT C56-YS93

Primer name	**Primer sequence**	**Feature**
prsASac1for	AATATGgagctcAATTGGCGTAGGAGTTGTGGAACAAATG	*Sac*I site
prsAEcoR1rev	AAGTAAgaattcACGATTTGCAGGACATTGCCGCAACAATTC	*Eco*R1 site
M13 forward -21	GTAAAACGACGGCCAGTG	Universal primers
M13 reverse -48	GGAAACAGCTATGACCATG	Universal primers

5.3.4 Cell fractionation

GT cells were grown on TGP agar plates to produce a thick lawn, which was then scraped off and added to 20ml pre-warmed ASM. The cells were recovered by incubating at 60°C and 220rpm for 1 hour, and were then used to inoculate 20ml fresh pre-warmed ASM in 250ml baffled conical flasks to OD_{600} of around 0.1. The culture was then grown to an OD_{600} of around 1.5, and the cells in 2ml of the culture were harvested.

The cells were centrifuged at 2000xg and the supernatant collected as the medium fraction. The pellet was then re-suspended in 2ml pre-warmed protoplast buffer (20% sucrose, 50mM Tris-HCl pH 7.5, 15mM Mgcl2, 5µg/ml lysozyme) at 37°C for 30 minutes. Protoplasts were then centrifuged at 700xg for 10 minutes. The supernatant was collected as the cell wall fraction, and the pellet as the protoplasts. The protoplasts were then lysed by re-suspension in 2ml 50mM Tris-HCL pH 7.5 and sonicated. The suspension was then centrifuged at 50,000x*g* for 1 hour (Beckman coulter benchtop ultracentrifuge) and the supernatant collected as the cytoplasmic fraction; the pellet was re-suspended in 2ml 50mM Tris-HCL pH 7.5 and collected as the cell membrane fraction.

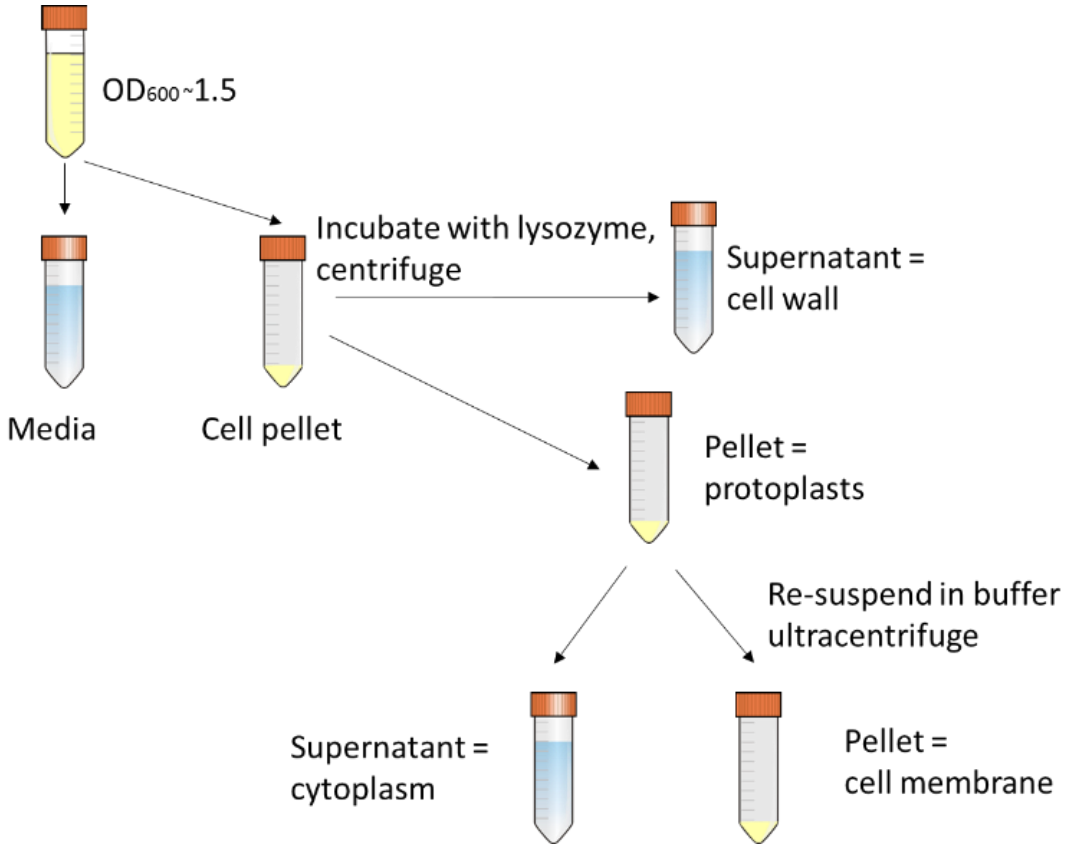

Figure 5.2: Simplified workflow of cell fractionation.

5.3.5 RZCL-xylan activity assay

Unless otherwise stated, 8mg/ml AZCL-xylan in phosphate citrate buffer at pH7 was incubated with equal volumes (0.5ml) of cell fractions for 1h at 60°C in 2ml microfuge tubes. The tubes were then briefly centrifuged to remove insoluble xylan, and the absorbance of the supernatant was measured at 595nm (BMG labtech platereader). The results were analysed using the Mars analysis suite (BMG Labtech).

5.3.6 Western blot analysis of cell fractionation samples

Samples were loaded onto SDS-PAGE gel in appropriate volumes with the OD_{600} corrected to 1.0 to ensure equal loading. Western blots were then carried out as described in chapter 2.

5.4 RESULTS AND DISCUSSION

5.4.1 Optimisation of Pulse chase analysis of xylanase secretion in *Geobacillus thermoglucosidasius*

The first aim in optimising the pulse chase experiments was to determine the adequate level of radioactivity to sufficiently label proteins in *GT.* To do this, the cells over-producing xylanase were grown to log phase, incubated in starvation media, and different aliquots incubated with different amounts of radioactivity for different periods of time. Figure 5.3A is an autoradiograph of SDS-PAGE-separated samples from different strains incubated with 25µCi methionine/cysteine label for 5 or 10 minutes. The result shows that TM242 is poorly radiolabelled in these conditions while C56 appears to have been radiolabelled well, and WT11955 labelling intensity was between the two. The intensity of the bands on the autoradiograph directly reflect radio-labelling, as the same cell density was loaded into each well. It was decided that the pulse chase experiments would be carried out in WT11955 and C56-YS93.

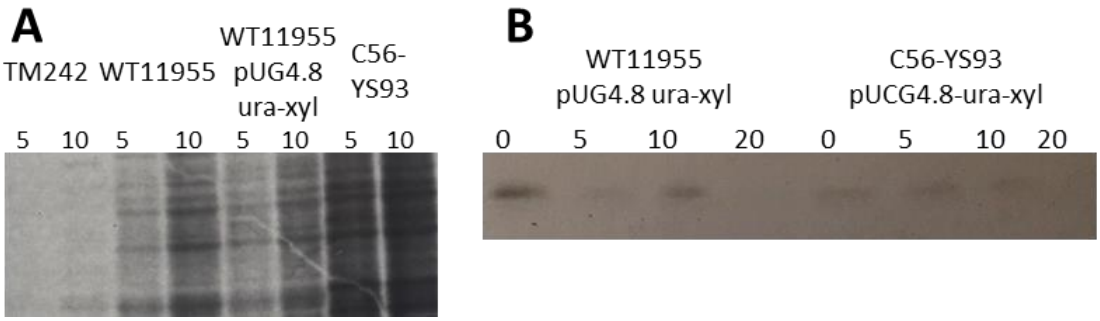

Figure 5.3: A: Two-week exposure autoradiography film with whole culture (cells and media) samples from TM242, WT11955, WT11955 pUCG4.9-uracil-xylanase and C56-YS93 incubated with 25µCi for 5 and 10 minutes showing the highest radiolabelling in C56-YS93 strain with the highest signal, and least radiolabelling in TM242 with the weakest signal overall. B: two-week exposure of pulse-chase autoradiograph after labelling and immunoprecipitation of xylanase protein steps showing weak signals in each lane.

However, further experiments to optimise the pulse-chase labelling revealed that, after immunoprecipitation, mature secreted protein was only weakly visible, but no precursor protein was observed (Figure 5.3B). Thus, the levels of xylanase produced and labelled were not sufficient for effective analysis. The next step was then to insert a constitutive

promoter, the RPLS promoter (Dr Ben Reeve, Imperial College), that is stronger than the uracil promoter, upstream of the xylanase gene in pUCG4.8, with the expectation that with increased expression of xylanase, and increased protein synthesis of xylanase, radiolabelling of the strain would result in detectable levels of radiolabelled xylanase. Unfortunately, even with the stronger promoter, xylanase levels were still not high enough for effective quantification (result not shown).

In the experiments above, a mixture of ^{35}S methionine/cysteine was used. As an alternative, we tested the use of ^{14}C labelled amino acids. ^{14}C provides a weaker signal than ^{35}S, but all amino acids would be labelled instead of just a small portion the amino acids in a protein, and this might thus improve the signal obtained in pulse-chase analysis. Unfortunately, this also did not provide the intensity of bands required for effective quantification (data not shown). Interestingly, TMO Renewables reported that *GT* TM242 actually utilised available amino acids in the culture media as a carbon source, rather than a supply of amino acids for protein synthesis (TMO Renewables, personal communication). They also reported that serine, threonine and glutamic acid were the only three absolutely essential amino acids required for growth. Furthermore, cysteine, methionine, glycine, serine and threonine biosynthesis are linked, which together with not utilising free amino acids for protein synthesis, may explain why *GT* TM242 did not label well with S^{35} labelled cysteine/methionine.

5.4.2 Xylanase (GEOTH_2250) secretion by TM242 with and without the signal peptide

Figure 5.4 is a western-blot detection of xylanase and shows the cell and secreted fractions of the TM242 strain expressing xylanase (GEOTH_2250) with the signal peptide (TMSP) and the strain expressing xylanase (GEOTH_2250) without any signal peptide (TMno). In the secreted fraction of the strain producing xylanase with the signal peptide, a band representing xylanase at 45kDa was observed, which is the protein without the signal peptide; the band highlighted by the red box represents xylanase with the signal peptide uncleaved. The presence of the precursor of xylanase (with the SP uncleaved) in the secreted or supernatant fraction is normal. As discussed in Chapter 3, significant cell lysis occurs during the growth of *GT*, which accounts for why there is precursor protein in the extracellular fraction. This cell lysis phenomenon will be discussed further in Chapter 6.

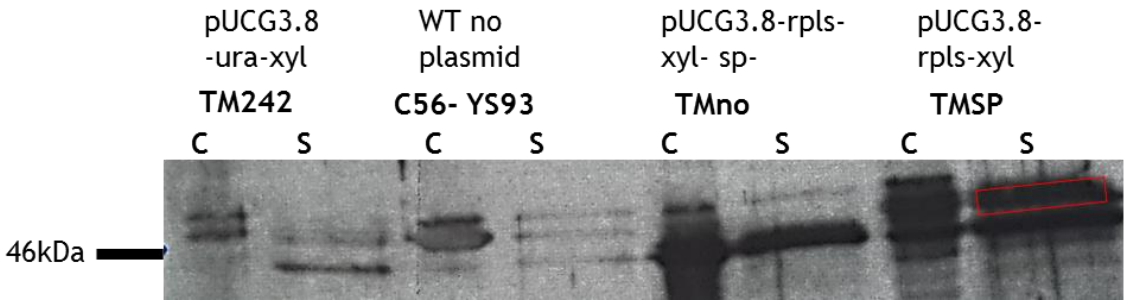

Figure 5.4 Western blot showing xylanase from cell and secreted fractions from TM242 producing xylanase with and without a signal peptide. C stands for cell fraction and S for secreted fraction. The red box highlights the position of the precursor protein, with uncleaved signal peptide in the secreted fraction.

As can be seen in Figure 5.5, over-production of xylanase does not appear to affect the growth of *GT* in terms of final optical density and growth rate during log phase in TGP culture medium. The rate of growth during log phase appears to be no different between the three strains, suggesting that the xylanase gene, over-expressed constitutively at this level, does not negatively impact growth. The growth curve also shows that maximum optical density, which correlates to biomass, remains the same between the three strains. This suggests that there is no discernible burden to over-production of xylanase at this level.

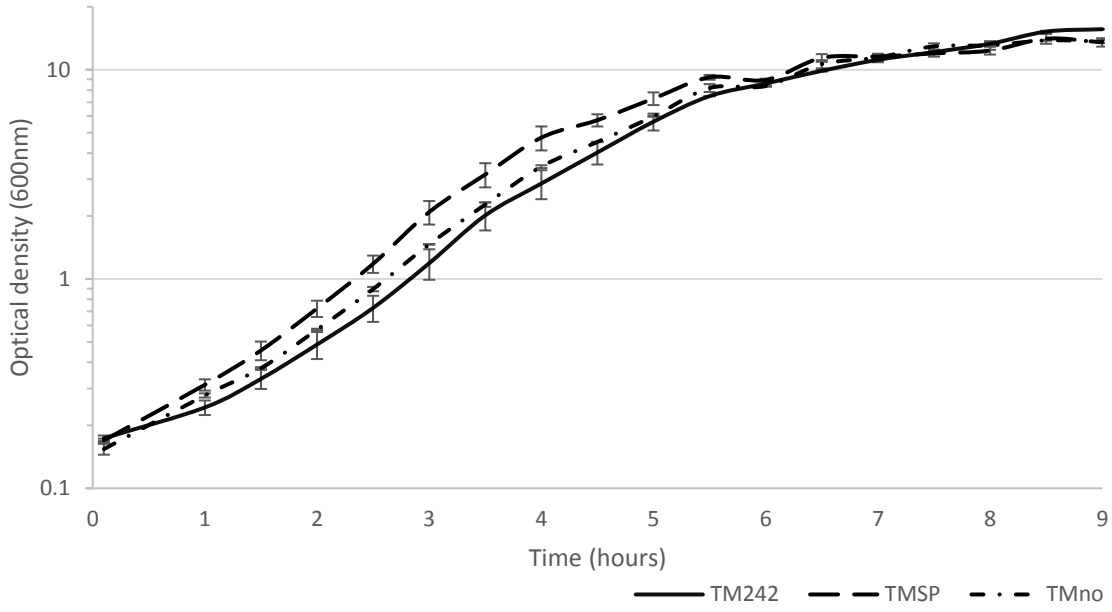

Figure 5.5 Optical densities over time of TM242 (TM242, solid black line), TM242-SP (TMSP, dashed line) and TM242-NoSP TMNo, dot and dash line) strains grown in TGP media showing no difference in growth rates between the strains. n=6

5.4.3 Cell fractionation of TM242 producing xylanase with and without the signal peptide

Cell fractions were obtained from *GT* TM242 expressing xylanase with the signal peptide (TMSP) in order to examine the relative xylanase activity exhibited by each fraction. This would reveal if there were any bottlenecks in protein translocation. For instance, as the protein bears a signal peptide, theoretically the bulk of the protein and enzyme activity should be in the media fraction. As can be seen in Figure 5.6, this is the case, as two thirds of the relative activity is found in the media fraction, and relatively little found in the cytoplasmic, wall and membrane fractions.

When the same analysis was carried out on the strain expressing xylanase without the signal peptide, the expected result would be to find most of the activity in the cytoplasmic fraction due to the protein lacking a signal peptide to target it to the secretion machinery. However, the actual result showed a significant amount of the activity in the media fraction. As seen in the western-blot analysis in Figure 5.4, a significant amount of xylanase is indeed found in the media fraction, which is presumed to be a result of significant cell lysis, also discussed in Chapter 3. Furthermore, there are significant levels of activity present in the wall and membrane fractions, but it is thought that this is partially due to cell lysis during collection of the fractions. This is unfortunate but not essential for this study, as the most important fractions are the secreted fractions.

When the total activity of all the fractions of TMSP and TMNo are considered and compared, the total enzyme activity from TMNo was almost double that of TMSP. This could be due to a number of reasons, such as secreted xylanase misfolding post-translocation and not being active, or secreted xylanase being degraded due to non-specific proteolysis in the extracellular milieu. Another possible explanation for the discrepancy in total xylanase activity levels is a difference in mRNA levels, due to differences in expression or mRNA stability, for example. However, as both plasmids were identical except for the signal peptide sequence, this is not particularly likely. As such, it is more likely that the cause of the discrepancy is at the protein level, due to either inactivity or degradation of xylanase. When the actual xylanase protein levels are

considered, as determined using western-blot analysis (Figure 5.7), the levels reflect a similar trend to that of the xylanase activity (Figure 5.6), suggesting that loss of activity of intact secreted protein is unlikely. Therefore, it was decided to investigate the effect of reducing proteolytic activity in the extracellular milieu, on extracellular xylanase activity.

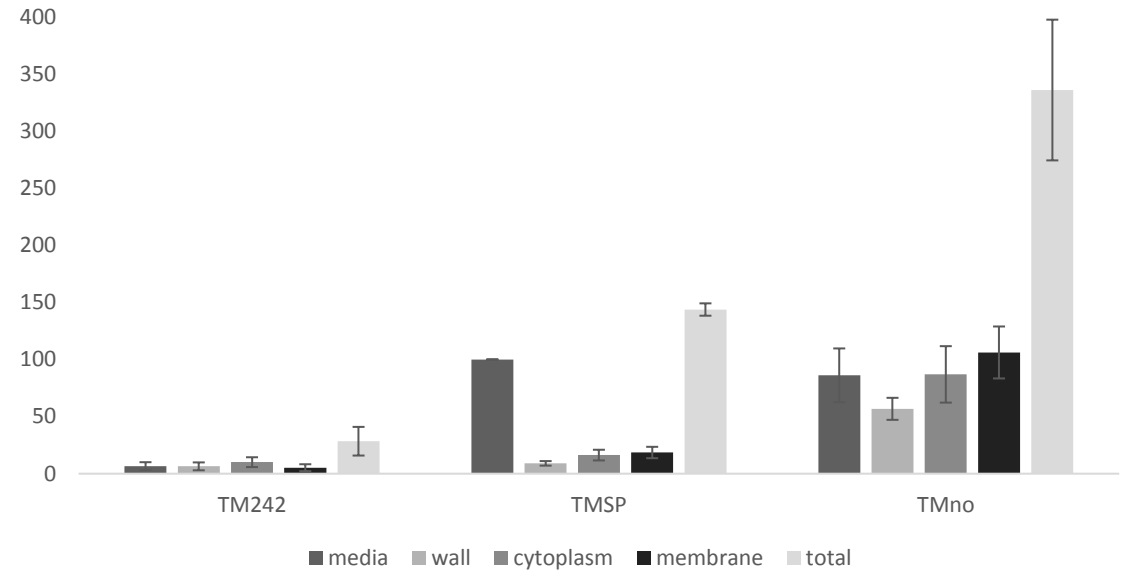

Figure 5.6: Xylanase assay using AZCL xylan from media, cell wall, cytoplasmic and membrane fractions of TM242, and the strains expressing the xylanase gene with (TMSP) and without (TMno) the signal peptide. The activity values are corrected for differences in OD_{600} between the different cultures. n=6

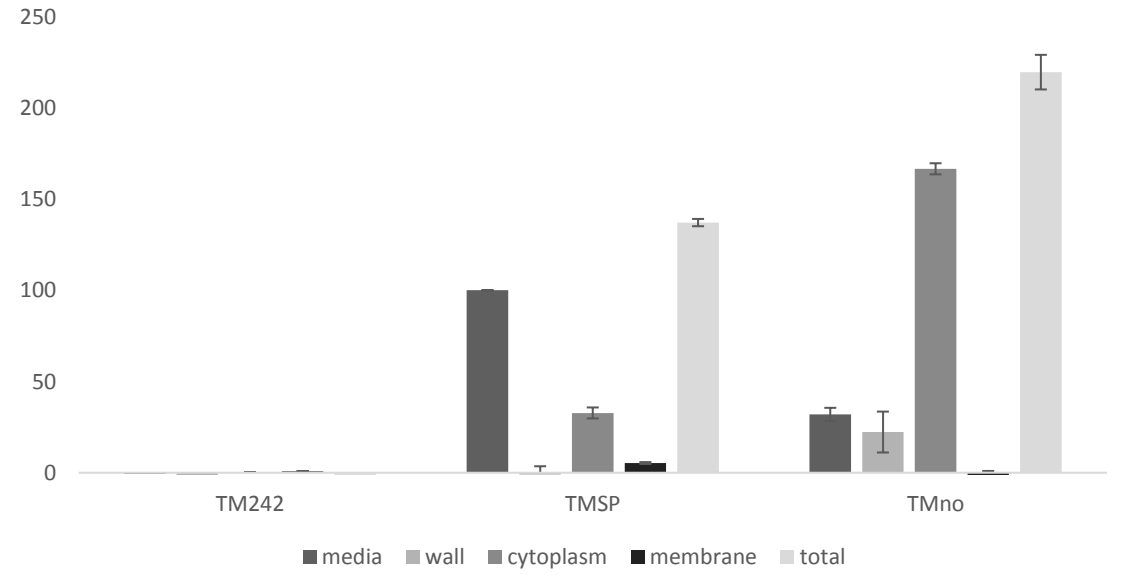

Figure 5.7: Western-blot densitometry of xylanase levels from the media, cell wall, cytoplasmic and membrane fractions of TM242, and the strains expressing the xylanase gene with (TMSP) and without (TMno) the signal peptide. The band intensity values are corrected for differences in OD_{600} between the different cultures n=6

5.4.4 The effect of the addition of protease inhibitors on xylanase secretion

Post-translocation, secreted proteins fold in the cell wall where they encounter a microenvironment that contains several quality control proteins, many of them proteases. In some circumstances, when proteins are over-expressed, the cell will then up-regulate the production of various proteases, which may result in higher levels of non-specific proteolysis of secreted proteins (Westers et al., 2006, Clausen et al., 2011). There are a number of ways in which to reduce proteolytic activity in the extracellular milieu, either to inactivate proteases by chemical means, or to inactivate genes encoding proteases at the genome level.

In *BS,* the phenomenon of proteolytic degradation of industrially produced enzymes is one that is well described (Stephenson and Harwood, 1998, Li et al., 2004, Delic et al., 2014). This organism secretes several proteases, leading to high levels of extracellular proteases which, in turn, degrade secreted proteins, especially those vulnerable to proteolytic attack. In *BS*, inactivating proteases at the genome level has been successfully accomplished in order to improve protein production of heterologous proteins (Wu et al., 1991b, Yang et al., 2004, Pohl et al., 2013, Wu et al., 2002a). Several studies have performed knock-outs of several key extracellular proteases in a single strain, in order to enhance heterologous protein secretion (Krishnappa et al., 2013, Pohl et al., 2013, Stephenson and Harwood, 1998).

Here we used a protease inhibitor cocktail (cOmplete, Roche), which was added to the media and the cultures then grown to an OD_{600} of around 1.5. The cOmplete protease inhibitor tablets were selected as they are readily available, and contain a cocktail of protease inhibitors that inhibit both cysteine and serine proteases, although the majority of extracellular proteases in *BS* and *GT* are serine proteases. Figure 5.8 shows the log phase growth curve of the two TM242 strains producing xylanase with and without the signal peptide, grown in defined media, with and without protease inhibitor. The growth rates show no deleterious effect as a result of incubation with protease

inhibitor, during log phase of the culture growth. This suggests that samples taken for xylanase activity at similar cell densities should be reliably comparable.

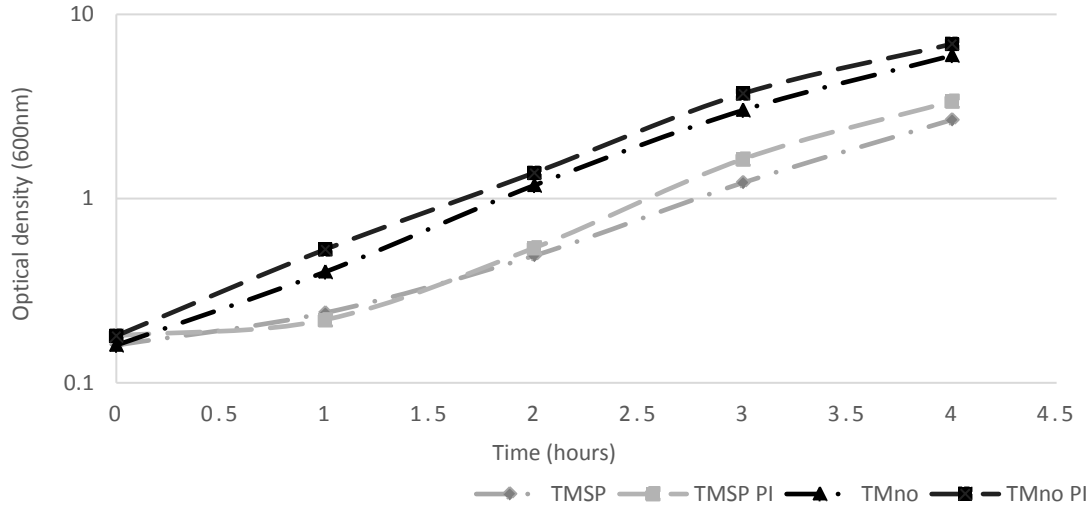

Figure 5.8 Growth curves of TMSP and TMNoSP strains grown in defined ASM media (0.5%glucose 0.5% xylose 0.2% yeast extract) with (indicated with PI) and without protease inhibitor. n=6

Figure 5.9 shows the relative xylanase activity of the different fractions from samples of TM242 producing xylanase with and without signal peptide, grown in the presence or absence of protease inhibitor. The data show that the addition of the protease inhibitor cocktail significantly ($p=0.022$) increases the xylanase activity in the media and wall fraction of the SP strain. The western-blot densitometry analysis corroborates these results, also showing increased xylanase levels in the media fraction (Figure 5.10). However, it is important to take into consideration that western-blot densitometry analysis is only semi-quantitative, due to the lack of loading controls suitable for cell fractionation analysis. These results, when considered together with the discrepancy in total xylanase activity between the two strains, suggests that there is a loss of xylanase post-translocation, which is most likely due to proteolysis in the extracellular milieu.

The control experiment, which was incubating the TM242-NoSP strain with and without protease inhibitor, was carried out to investigate if the addition of protease inhibitor has any significant impact on xylanase activity, even when not secreted. The results showed no significant change in xylanase activity in any of the fractions, nor any significant change in xylanase levels. This confirms that the discrepancy in xylanase activity and levels are a result of an extracellular event, such as proteolysis in the extracellular milieu.

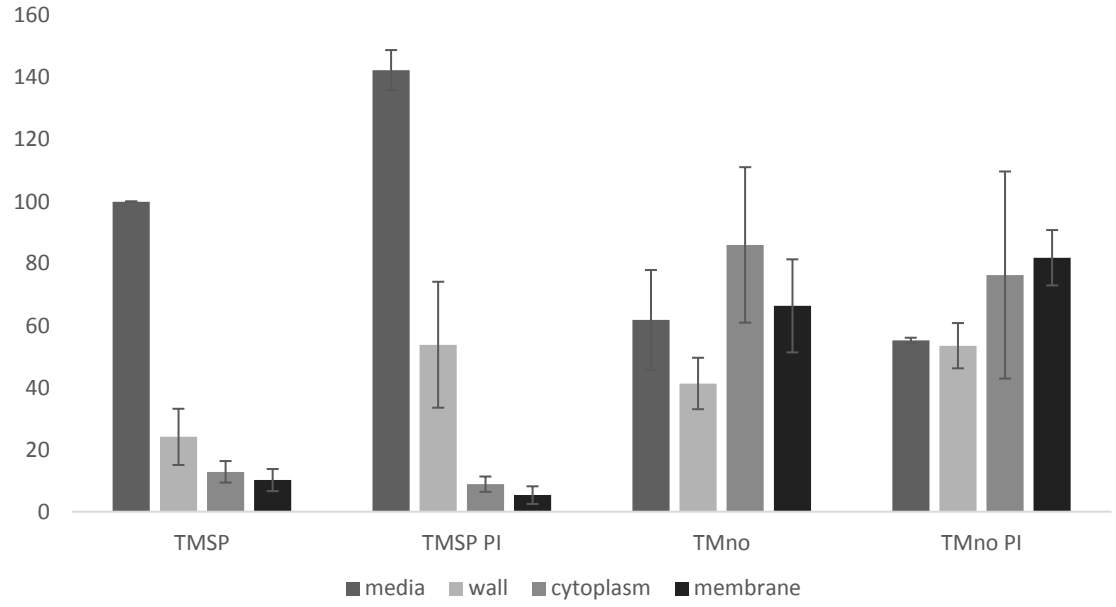

Figure 5.9 Relative xylanase activity in media, cell wall, cytoplasmic and cell membrane fractions taken from GT TM242 strains TMSP and TMno at OD_{600} 1.5. PI indicates the addition of protease inhibitor. The activity values are corrected for differences in OD_{600} between the different cultures. The addition of protease inhibitor to the TMSP strain resulted in significant increase (p=0.022) in xylanase activity in the media fraction, while the TMno strain was unaffected. n=6

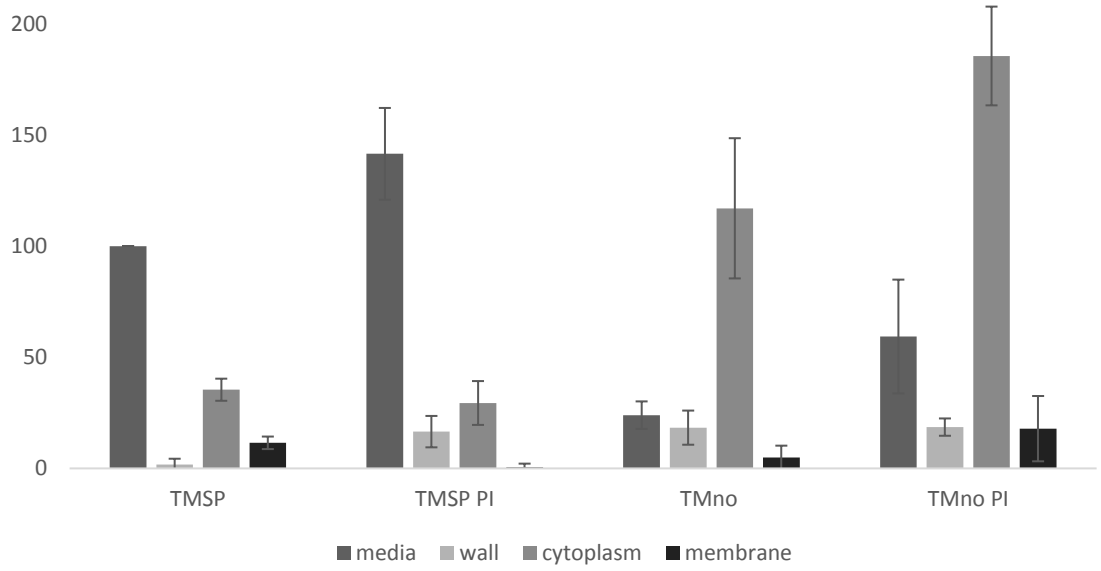

Figure 5.10 Western-blot densiometry analysis of media, cell wall, cytoplasmic and cell membrane fractions taken from GT TM242 strains TMSP and TMno at OD_{600} 1.5. PI indicates the addition of protease inhibitor. The band intensity values are corrected for differences in OD_{600} between the different cultures. n=3

5.4.5 The effect of over-expression of PrsA on xylanase secretion

As shown above, degradation of xylanase is likely to occur, and this may be related to the rate of folding directly after translocation. An important factor in this could be PrsA, which is a membrane-associated lipoprotein with peptidyl-prolyl *cis-trans* isomerase activity. PrsA is present in almost all Gram-positive bacteria; it does not influence the expression or the translocation of secretory proteins, but it is required for their folding and stability in the post-translocational phase of secretion at the membrane–cell wall interface. In Gram-positive bacteria, which do not have a periplasm, secreted proteins emerge from the translocase to the area immediately between the cell membrane and the cell wall. This is a demanding environment for protein folding and stability due to the high density of negative charge, high concentration of cations, and low pH immediately outside the membrane. These factors most likely pose constraints for the kinetics of folding of secreted proteins. Native proteins compatible with the conditions at the membrane–wall interface fold with fast kinetics into their normal conformation. However, heterologous proteins produced in *BS* have been shown to be more susceptible to proteolytic degradation than native proteins (Bolhuis et al., 1999b). In *B. subtilis*, it has been suggested that over-expression of PrsA may be advantageous when expressing heterologous proteins, both at levels that saturate the secretion translocon machinery, and at lower levels (Vitikainen et al., 2001) . In several studies, increased levels of PrsA lipoprotein have resulted in increased levels of secreted protein (Kakeshita et al., 2011, Vitikainen et al., 2001), which suggest that processing by PrsA is the rate-limiting step in protein secretion of those proteins (Kontinen and Sarvas, 1993). It has also been shown to be essential not only for secretion, but also for cell viability (Vitikainen et al., 2001, Jacobs et al., 1993, Kontinen and Sarvas, 1993).

As in the previous experiments investigating the addition of protease inhibitor on culture growth, Figure 5.11 shows the growth curve during log phase comparing TM242, TM242-SP, TM242-SP with protease inhibitor, and TM242-SP-PrsA (expressing both xylanase with the signal peptide, and *prsA* genes). The growth curves confirm no noticeable effect on culture growth when *prsA* is expressed as well as xylanase.

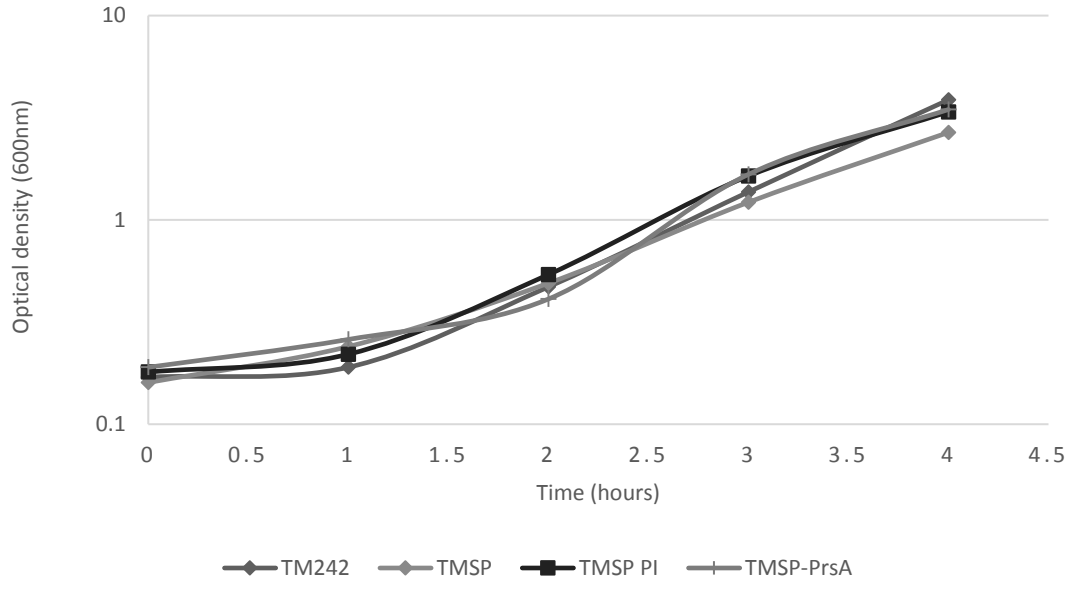

Figure 5.11 Culture growth curves of TM242, TM242-SP, Tm242-SP with protease inhibitor, and TM242-SP-prsA in TGP medium. n=3

The cell fractionation analysis results indicate that overproduction of PrsA does not significantly change xylanase activity (measured using the RZCL-activity assay) or xylanase protein levels (determined by western blotting) in any of the fractions. The fractions of most interest in this case would be the media and cell wall fractions, both of which, when the enzyme activity data are considered, show no significant difference from that of TM242-SP or TM242-SP incubated with protease inhibitor. The western-blot densitometry data also reflect the same trend.

This suggests that, in the case of xylanase secretion in TM242, PrsA activity is not rate limiting. It could even suggest that post-translocational folding is not the rate-limiting step in this case. Although the xylanase is technically a heterologously produced protein, as it originates from *GT* C56-YS93, the two strains are very closely related, suggesting that xylanase should be able to fold efficiently after translocation in TM242. However, this is not to suggest that over-expression of *prsA* is of no benefit for secretion of heterologous proteins in *GT*, but the effect of *prsA* over-expression would need to be investigated with a protein from a more distantly related strain.

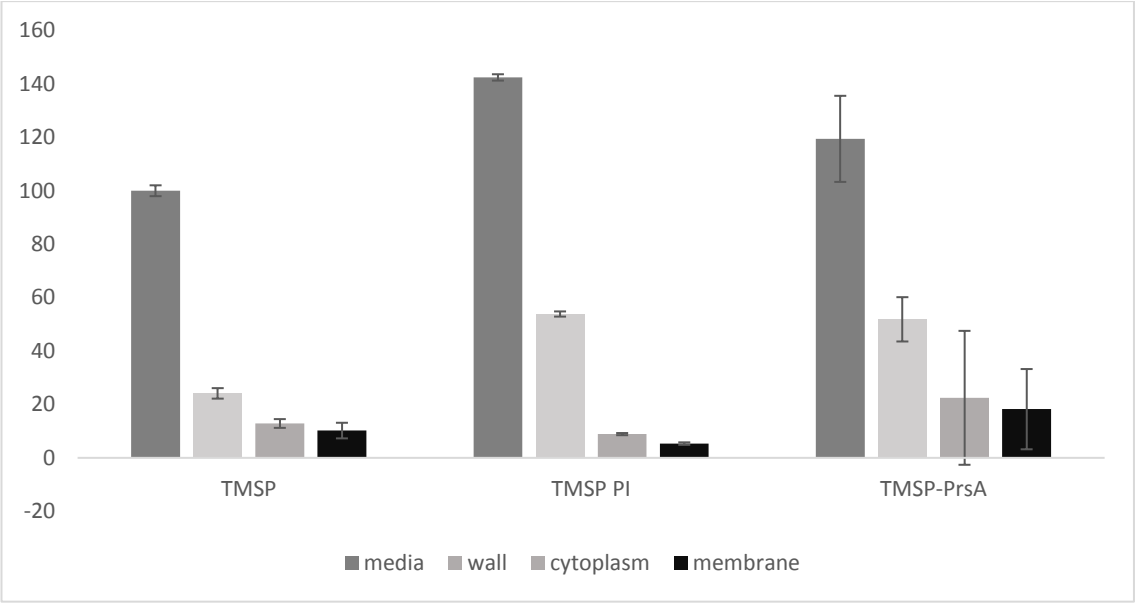

Figure 5.12 Xylanase activity in media, wall,cytoplasm and membrane fractions of GT TM242 strains TMSP and TMSP-PrsA. PI indicates the addition of protease inhibitor. The activity values are corrected for differences in OD_{600} between the different cultures. n=3

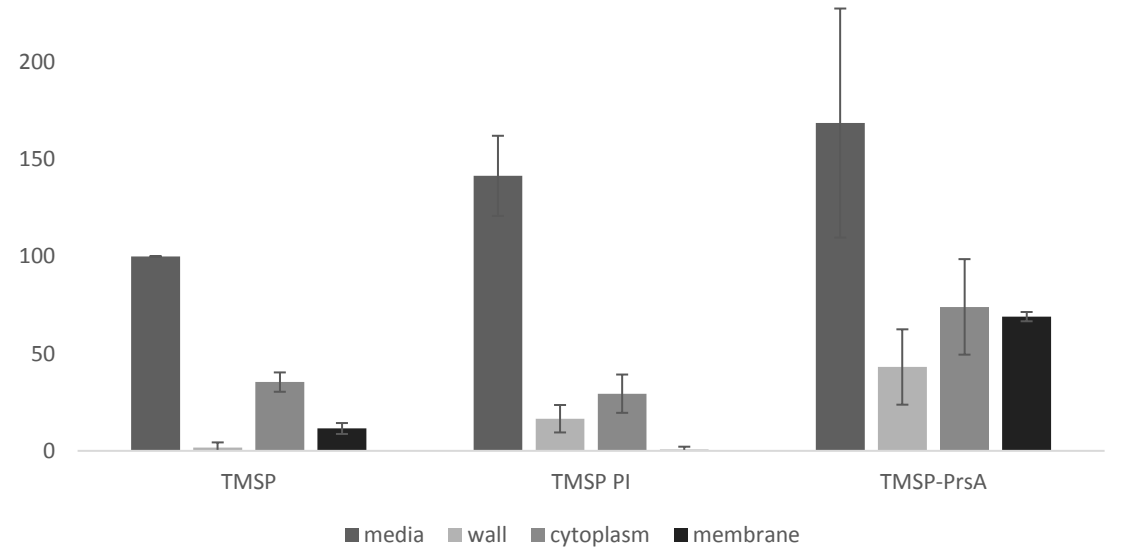

Figure 5.13 Western-blot densiometry of xylanase levels in media, cell wall, cytoplasm and cell membrane fractions from GT TM242 strains TMSP and TMSP-PrsA. PI indicates the addition of protease inhibitor. The band intensity values are corrected for differences in OD_{600} between the different cultures .n=3

5.5 CONCLUSIONS

The results from the cell fractionation experiments shown here strongly suggest that the bottleneck when over-producing and secreting xylanase at the levels conferred by the RPLS promoter, is post-translocational, namely in the cell wall and/or extracellular milieu, and that xylanase activity is lost due to proteolytic activity. These results suggest that inactivating extracellular proteases would be of benefit, especially for the purposes of over-producing secreted hydrolases, to increase lignocellulosic feedstock utilisation. However, the addition of protease inhibitors is costly, both in monetary terms, and also in terms of the metabolic burden placed on the cell. Similar to *BS,* it would be beneficial to create strains of *GT* lacking key extracellular proteases. Thus, creating strains of *GT* that have key proteases inactivated, or not produced at all, may be of significant benefit to the production of secreted hydrolases, and therefore feedstock utilisation.

CHAPTER SIX: GENERAL CONCLUSION AND FUTURE PERSPECTIVES

6.1 GENERAL DISCUSSION

Geobacillus thermoglucosidasius is naturally able to ferment a range of substrates, including both pentose and hexose sugars, and produce a number of organic compounds such as ethanol or lactic acid. *G. thermoglucosidasius* TM242 has been genetically engineered to divert the fermentation pathway away from the natural mixed acid fermentation, to generate ethanol as the main product. This makes it a good platform candidate for the production of second generation bioethanol from lignocellulosic feedstocks. Furthermore, with the rapid development of genetic tools to further engineer the organism, *Geobacillus* could be utilised to produce several other value-added organic compounds as well. However, Lignocellulosic feedstocks require pre-treatment steps before they are suitable for fermentation, one of which is a hydrolysis step with commercial enzymes. The latter remains a major cost to the production of bioethanol or other organic compounds, which significantly affects cost-efficiency of the whole process. Reduction or elimination of this enzyme hydrolysis step would render the whole process much more cost effective, and make the production of bioethanol from lignocellulose more lucrative. This could be achieved by engineering *GT* to produce its own enzymes that hydrolyse the lignocellulosic material. However, the enzymes would need to be secreted due to the polymeric nature of lignocellulose. The ultimate goal would be to engineer *GT* to produce a cocktail of enzymes depending on the feedstock, towards consolidated bioprocessing. Even with partial elimination of the hydrolysis pre-treatment, costs could be significantly reduced. However, before this is done, it would be prudent to characterise the protein secretion pathways of *GT* and determine the effects of over-production of secretory proteins, and from there, devise strategies to optimise protein secretion of hydrolases in *GT*. The work presented in this thesis aims to characterise the secretion machinery of *GT* and elucidate potential differences in this with that of the well described mesophilic relative *BS.* This characterisation and comparison is not only focussed on the secretion machinery itself, but also accessory factors and signal peptides. In this work, an endo-xylanase enzyme from *GT* C56-YS93 was selected as the model enzyme to be over-produced in the working strain *GT* TM242 for a number of reasons. Xylanase randomly cleaves the 1-4-xylosidic linkages in xylan, which is a major component of lignocellulosic feedstocks.

Although heterogeneous in nature, xylan is amorphous, unlike cellulose, which although more abundant, usually occurs in a recalcitrant crystalline form, making hemicellulose the more accessible option as a substrate for hydrolysis and subsequent fermentation. Another important reason as to why this xylanase was selected is because it is demonstrably a natively secreted protein. Furthermore, the gene is from a closely related strain, which may reduce the chance of encountering bottlenecks at transcription and translation level.

In chapter 4, the purification and characterisation of xylanase produced heterologously in *E. coli* is described. In order to investigate bottlenecks in protein secretion, antibodies were required to quantify levels of xylanase present in the various cellular fractions. To this purpose, polyclonal antibodies were raised against purified xylanase and western blot optimised. Furthermore, xylanase was characterised in terms of enzyme activity, and optimal pH and temperature. It was found to have an optimal pH between pH8 and pH9, which makes the xylanase from *GT* C56-YS93 a potentially suitable choice for improving feedstock utilisation, especially if alkaline treatment is used to pre-treat the biomass prior to fermentation, due to the elevated pH of the biomass. The optimum temperature is between 50 and 60°C, which is consistent with the organism's optimal growth temperature and ideal for ethanol fermentation at high temperatures.

Over-production of xylanase at these levels do not appear to hamper the growth of TM242. However, it would be interesting to test whether TM242 shows improved growth on lignocellulosic substrates. TM242 is actually able to grown on xylan substrate, but this is most likely due to the presence of xylose monomers and oligomers that *GT* is able to utilise.

6.1.1 The Sec machinery and signal peptides in *GT* and *BS*

Signal peptides are required for all secretory proteins to gain entry into the secretory pathway with very few known exceptions. They all have a characteristic tripartite structure, although do not possess sequence identity between them, apart from the consensus A-X-A sequence at the signal peptide cleavage site. It is very likely that there is no one size fits all signal peptide, in any organism, as the combination of signal peptide with the N-terminal domain of the mature protein is important for efficient secretion.

Nevertheless, optimising or changing signal peptides to improve protein secretion of particular proteins has been accomplished in *BS* and other mesophilic species. Signal peptide libraries have been used interchangeably between different organisms to achieve improved protein secretion. Thermophilic proteins have been produced heterologously in *BS* with their own native signal peptide, but there is no published work on the differences (or even if there are any) between mesophilic and thermophilic signal peptides. More generally, studies on comparing mesophilic and thermophilic proteins have shown that there are key features that thermophilic proteins have adapted which confer thermostability, such as increased hydrophobicity in alpha-helical regions or increased salt bridges to confer tighter folding. It is therefore conceivable that there are also differences between thermophilic and mesophilic signal peptides. The results in chapter 3 show however that there is no significant difference in average length or hydrophobicity between the sets of signal peptides from the predicted secretory proteins of *GT* and *BS*. This suggests that signal peptide libraries from either organism, or even other related organisms could be used interchangeably to screen for optimal protein secretion of heterologous enzymes in *GT.* However, very little is known or understood on the relationship between the signal peptide and the mature protein sequence, and future work elucidating this relationship will be an advantage towards the intelligent design of optimal signal peptides.

Signal peptides direct secretory proteins to the protein translocation machinery, and for the purposes of this research, only the Sec system was investigated, due to it being the predominant secretion system, and the fact that very few proteins in *GT* are transported via the Tat pathway. The Sec pathway of protein secretion is one that occurs in all domains of life: bacteria, archaea and eukaryotes. In chapter 3, the genes for secretion machinery and accessory protein in *BS* were listed and compared to the genome of *GT*. Here we found that the main components of the Sec machinery in *GT* were homologous to those in the well described *BS*, such as the SecYEG translocon, SecA motor protein, SecDF chaperone and the SRP and its receptor FtsY. We found that similar to *BS, GT* also lacks the SecB protein that in *E. coli* is involved in targeting and acts as a chaperone. However, the CsaA protein, which in *BS* is thought to be implicated in targeting and acts as a chaperone with activity complementary to SecB, is also absent in *GT*. It is therefore

unclear how non-SRP proteins in *GT* are targeted to the membrane. One option could be that all secretory proteins (whether translocated post- or co-translational) are SRP dependent, but it is also conceivable that there are other so far unidentified targeting factors that are perhaps specific to thermophilic organisms.

Of the accessory proteins in *BS*, such as chaperones, foldases and proteases, several extracellular quality control proteases have putative chaperone activity. HtrA, HtrB and their corresponding two-component histidine kinase/response regulatory proteins CssR and CssS, not found to be annotated in the genome, nor did a protein BLAST search yield any proteins with identity suggesting homology. This was also the case for another cell wall associated protease WprA. However, this is not to say that the roles of these proteins in *GT* are unfulfilled, but their roles may be performed by other proteins with a similar, but yet not described function. Although their absence, or absence of proteins with a similar function, may offer some explanation as to why the potential bottleneck found in protein secretion of xylanase is post translocational (see chapter 5). One method to identify proteins in *GT* which may be involved in the secretion stress response is to determine which genes are overexpressed (through transcriptome analysis) in different secretion-stress conditions, which can be achieved by over-producing secretory proteins at different levels.

6.1.2 Secretion bottlenecks caused by over-production of xylanase in *Geobacillus thermoglucosidasius TM242*

In chapter 5, the investigation into the effects of overproduction of secreted xylanase in *GT* TM242 was described. From the results, no apparent bottlenecks could be observed during the actual translocation process. However when comparing strains over-producing xylanase with or without a signal peptide, we found that total xylanase activity and protein levels were higher in the strain producing xylanase without a signal peptide. It is possible that without signal peptide, the xylanase folds rapidly in the cytoplasm and remains active, whereas when it is secreted there may be a loss of xylanase either in the cell wall, or in the extracellular milieu. This is most likely due to protein degradation, either due to misfolding or aggregation caused by slow folding, or non-specific proteolysis in the cell wall or extracellular milieu. From the shotgun mass

spectrometry data shown in chapters three and six, it was found that one third of secretory proteins are proteases, which is consistent with the non-specific proteolysis sustained by the over-produced secreted xylanase. Studies in *BS* revealed that slow folding of heterologously produced and secreted proteins at the cell membrane-cell wall interface leaves them susceptible to hydrolysis by cell-wall associated proteases as slowly folding proteins expose protease-sensitive sites that are not exposed in the correctly folded protein (Williams et al., 2003, Wu et al., 1991a). Misfolded or slowly folding proteins are rapidly degraded to prevent interference with cell wall growth and renovation, and to prevent blockages at the translocase (Sarvas et al., 2004, Jensen et al., 2000). In *BS*, several strains have been engineered by interrupting or deleting one, or combinations of extracellular proteases with up to 11 extracellular proteases inactivated (Pohl et al., 2013). However, although the genes encoding proteases can be interrupted individually without major effects on cell physiology, strains in which both the *htrA* and *htrB* genes were interrupted demonstrated a significant reduction in viability, which may suggest that HtrA and/or HtrB perform a role that is vital for protein secretion. This presents a number of options for engineering *GT* for reduced proteolysis of secreted proteins. Strains could be engineered where extracellular proteases, and combinations of the proteases, could be interrupted or deleted to investigate the effect on over-production and secretion of xylanase. With the rapid development of genome editing CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technologies, this may soon become a much easier task than previously, as using CRISPR technology we would no longer have to rely on multi-step and time consuming knockout strategies previously employed to engineer *GT* strains (Peters et al., 2015, Singh et al., 2017). Currently the knock-in of heterologous genes into *GT* is carried out using a double recombination approach. In this approach, the locus of a non-essential or unwanted gene in the genome of the microorganism is used for the insertion of the new heterologous gene. First, the heterologous gene needs to be cloned with a suitable promoter upstream into a vector such as pUC19. Following this, a short DNA sequence from the locus where the gene is going to be inserted is cloned either side of the heterologous gene, the cassette is then cloned into a knock-in plasmid, which is then transformed into TM242. Primary integrants (single cross-overs) are then selected, from

which the finished knock-in (double cross-overs) containing the heterologous gene but not the other DNA from the plasmid, can be selected.

Furthermore, as discussed previously, the *prsA* gene in *BS* has been over expressed to improve secretion of heterologous proteins (Wu et al., 1998, Vitikainen et al., 2005). One notable study investigating the effects of PrsA depletion or over-production on the secretion of the *B. amyloliquefaciens* α-amylase in *BS* showed that depletion of PrsA resulted in reduction (and also cell death) and upregulation of the gene led to significant increase in α-amylase production (Chen et al., 2015c). However, the work presented here showed that the over-production of PrsA in *GT* did not lead to increased secretion of xylanase. Actual expression levels of prsA would need to be verified in order to confirm this. However, it has been shown that in *BS* at least, not all secretory proteins are dependent on PrsA for post-translocational folding (Vitikainen et al., 2004), and the results here certainly suggest that xylanase may be one of those proteins. However, it remains a potentially useful strategy, as other heterologous enzymes selected for consolidated bioprocessing in *GT* may depend on PrsA for folding. The prospect of xylanase folding being PrsA independent also supports the indication that xylanase is subject to non-specific proteolytic degradation.

In Appendix One, the cell lysis phenomena is delved into further; looking at the shotgun mass spectrometry data it was found that the vast majority of proteins found in the extracellular milieu of *GT* C56-YS93 are cytoplasmic in origin (the assumption that the proteins are cytoplasmic is based on lack of predicted signal peptides, and annotated function). This suggests that cell lysis is rampant during the exponential growth phase. Further evidence of cell lysis was demonstrated with western blots of GroEL, a cytoplasmic chaperone. Interestingly, when the xylanase gene was expressed in TM242 without the signal peptide, a considerable amount of active enzyme was found in the extracellular milieu as evidenced by the xylanase activity, and by western blot. This draws attention to the possibility that enzymes required for lignocellulosic hydrolysis, may not need to be secreted into the extracellular milieu, rather they can be delivered there via naturally occurring cell lysis. However, referring back to the suggestion of deleting or interrupting genes for extracellular proteases, this brings up two considerations. 1) cell lysis will also result in cytoplasmic proteases being released into

the extracellular milieu, which will probably also contribute towards non-specific proteolysis. 2) It has been observed in *BS* that strains with several protease genes deleted, tend to lyse more readily, so whether or not this is the case in *GT* will have an impact on the balance that would need to be struck between cell lysis, and intact cells manufacturing the desired product.

We also briefly touched on cell lysis as a link to cannibalistic behaviour. One study done in *BS* investigated the transient heterogeneity of bacillopeptidase and subtilisin, and found that transcriptome levels of these genes were heterogeneous throughout the population (Veening et al., 2008). As protease levels were high in the extracellular milieu, this suggested that all cells in the population would benefit from protease production, even the cells not (or poorly) expressing those genes. This leads to a further suggestion that *BS* displays co-operative behaviour in a heterogeneous population of vegetative cells or dividing cells. The protease secreted also is able to hydrolyse proteins released from dead cells, the products of which can be scavenged by growing cells. AprE (subtilisin) and Bpr are scavenging proteins that are secreted into the growth medium and degrade (large) proteins into smaller peptides, which can be taken up and used as an alternative nutrient source. Currently, social behaviours in microbial populations are very poorly understood, both in natural environments but also to some extent in laboratory conditions. Bacterial populations are almost always heterogeneous in nature, in terms of cell cycle stage, and also potentially the role of the cell with respect to the rest of the population.

6.2 FUTURE PERSPECTIVES

Could we use cell lysis as a means to deliver hydrolytic proteins? This may be useful in some cases, but it would not be a good solution if the aim is to produce cellulosomes that break down and utilise crystalline cellulose, as these are multi-enzyme structures that are anchored to the cell surface. In addition, productivity of e.g. bioethanol is possibly reduced if there is a significant amount of cell lysis, and it may therefore be important to engineer strains with reduced levels of cell lysis, or possibly optimise growth conditions in fermenters that reduce cell lysis to obtain a balance of biomass and product production.

One limitation of this study in terms of identifying bottlenecks in secretion is that the selected protein (xylanase) was from a very closely related organism (*GT* C56-YS93), so although technically a heterologous protein, it may be beneficial if an investigation was done into investigating posttranslational proteolysis of a protein from a more distantly related organism. However, for the purposes of engineering *GT* to hydrolyse, utilise and ferment lignocellulosic feedstocks, the source of enzymes selected is ideally from closely related species due to simpler legislation issues (GMO vs non-GMO), better compatibility in terms of gene expression, and simply because related *Geobacilli* contain many of the genes responsible for the potential complete hydrolysis of lignocellulosic feedstocks .

Another limitation of the work investigating bottlenecks of protein secretion was the lack of a range of promoters. Better availability of various types of promoters, including those that lead to very high expression levels and/or those that are inducible, will help in maximising secretion and production of hydrolases and determining the rate-limiting steps in the secretion process. Having a range of promoter strengths would also be useful if a transcriptome analysis study of *GT* at different secretion stress levels were to be carried out.

It has been shown in *BS* that the relationship between the heterologous secreted protein and the absence or presence of proteases and foldases is not straightforward. For example, strains lacking combinations of extracellular proteases have been helpful in improving the productivity of *BS* for the production of single-chain antibodies against some antigens but not others (Wu et al., 1998), suggesting that there is a delicate balance between folding and structure and the secretion yield of different proteins. As such, the most ideal route to take may be a synthetic biology approach for the development of commercial strains of *GT*. Understandably, it may be quite some time before the genetic tools for *GT* will be developed enough for such an approach, so an approach somewhat in between the conventional genetic engineering and a truly synthetic approach may be the best strategy forward.

As such, it may be useful to have a strain with increased expression levels of PrsA in *GT* so future work on cloning secretory hydrolases can compare with normal PrsA and increased PrsA. In the current strain shown in this work, the *prsA* gene is over-expressed

in an operon under control of Rpls promoter, and after xylanase gene on the same plasmid, but it would be useful to engineer a strain with a stronger constitutive promoter controlling the chromosomal copy of *prsA.*

All experiments presented in this thesis are of log phase growth, and not necessarily reflective of what would occur in industrial fermenters. However, in terms of identifying bottlenecks in protein secretion, the work presented here shows that loss of activity of over-produced secreted proteins is due to extracellular proteolysis, and is an excellent starting point towards the reduction of loss of secretory proteins.

REFERENCES

ACHARYA, S. & CHAUDHARY, A. 2012. Bioprospecting thermophiles for cellulase production: A review. *Brazilian Journal of Microbiology,* 43**,** 844-856.

AKITA, M., SASAKI, S., MATSUYAMA, S. & MIZUSHIMA, S. 1990. Seca interacts with secretory proteins by recognizing the positive charge at the amino terminus of the signal peptide in *Escherichia coli*. *Journal of Biological Chemistry,* 265**,** 8164-8169.

AKIYAMA, Y., EHRMANN, M., KIHARA, A. & ITO, K. 1998. Polypeptide binding of Escherichia coli FtsH (HflB). *Molecular Microbiology,* 28**,** 803-812.

ALFANI, F., GALLIFUOCO, A., SAPOROSI, A., SPERA, A. & CANTARELLA, M. 2000. Comparison of SHF and SSF processes for the bioconversion of steam-exploded wheat straw. *Journal of Industrial Microbiology & Biotechnology,* 25**,** 184-192.

ALLEN, W. J., COREY, R. A., OATLEY, P., SESSIONS, R. B., BALDWIN, S. A., RADFORD, S. E., TUMA, R. & COLLINSON, I. 2016. Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. *Elife,* 5.

ANGELINI, S., DEITERMANN, S. & KOCH, H. G. 2005. FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *Embo Reports,* 6**,** 476-481.

ANGOV, E., HILLIER, C. J., KINCAID, R. L. & LYON, J. A. 2008. Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. *Plos One,* 3.

ANTELMANN, H., DARMON, E., NOONE, D., VEENING, J.-W., WESTERS, H., BRON, S., KUIPERS, O. P., DEVINE, K. M., HECKER, M. & VAN DIJL, J. M. 2003. The extracellular proteome of *Bacillus subtilis* under secretion stress conditions. *Molecular Microbiology,* 49**,** 143-156.

ANTELMANN, H., TJALSMA, H., VOIGT, B., OHLMEIER, S., BRON, S., VAN DIJL, J. M. & HECKER, M. 2001. A proteomic view on genome-based signal peptide predictions. *Genome Research,* 11**,** 1484-1502.

ARGYROS, D. A., TRIPATHI, S. A., BARRETT, T. F., ROGERS, S. R., FEINBERG, L. F., OLSON, D. G., FODEN, J. M., MILLER, B. B., LYND, L. R., HOGSETT, D. A. & CAIAZZA, N. C. 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl Environ Microbiol,* 77**,** 8288-94.

BABE, L. M. & SCHMIDT, B. 1998. Purification and biochemical analysis of WprA, a 52-kDa serine protease secreted by *B. subtilis* as an active complex with its 23-kDa propeptide. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology,* 1386**,** 211-219.

BAI, W., XUE, Y., ZHOU, C. & MA, Y. 2012. Cloning, expression and characterization of a novel salt-tolerant xylanase from *Bacillus* sp. SN5. *Biotechnol Lett,* 34**,** 2093-9.

BARTOSIAK-JENTYS, J., HUSSEIN, A. H., LEWIS, C. J. & LEAK, D. J. 2013. Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology,* 159**,** 1267-75.

BAYER, E. A., BELAICH, J. P., SHOHAM, Y. & LAMED, R. 2004. The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology,* 58**,** 521-554.

BECKERT, B., KEDROV, A., SOHMEN, D., KEMPF, G., WILD, K., SINNING, I., STAHLBERG, H., WILSON, D. N. & BECKMANN, R. 2015. Translational arrest by a prokaryotic signal recognition particle is mediated by RNA interactions. *Nature Structural & Molecular Biology,* 22**,** 767-773.

BENDTSEN, J. D., KIEMER, L., FAUSBOLL, A. & BRUNAK, S. 2005a. Non-classical protein secretion in bacteria. *BMC Microbiol,* 5**,** 58.

BENDTSEN, J. D., NIELSEN, H., VON HEIJNE, G. & BRUNAK, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology,* 340**,** 783-795.

BENDTSEN, J. D., NIELSEN, H., WIDDICK, D., PALMER, T. & BRUNAK, S. 2005b. Prediction of twin-arginine signal peptides. *BMC Bioinformatics,* 6**,** 167.

BHALLA, A., BISCHOFF, K. M., UPPUGUNDLA, N., BALAN, V. & SANI, R. K. 2014. Novel thermostable endo-xylanase cloned and expressed from bacterium *Geobacillus* sp. WSUCF1. *Bioresour Technol,* 165**,** 314-8.

BHANU, M. K., ZHAO, P. & KENDALL, D. A. 2013. Mapping of the SecA Signal Peptide Binding Site and Dimeric Interface by Using the Substituted Cysteine Accessibility Method. *Journal of Bacteriology,* 195**,** 4709-4715.

BLANCHARD, K., ROBIC, S. & MATSUMURA, I. 2014. Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. *Applied Microbiology and Biotechnology,* 98**,** 6715-6723.

BLOBEL, G. & SABATINI, D. D. 1971. *Ribosome membrane interaction in eukaryotic cells*.

BOCHKAREVA, E. S., SOLOVIEVA, M. E. & GIRSHOVICH, A. S. 1998. Targeting of GroEL to SecA on the cytoplasmic membrane of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 478-483.

BOLHUIS, A., BROEKHUIZEN, C. P., SOROKIN, A., VAN ROOSMALEN, M. L., VENEMA, G., BRON, S., QUAX, W. J. & VAN DIJL, J. M. 1998. SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins. *Journal of Biological Chemistry,* 273**,** 21217-21224.

BOLHUIS, A., MATZEN, A., HYYRYLAINEN, H. L., KONTINEN, V. P., MEIMA, R., CHAPUIS, J., VENEMA, G., BRON, S., FREUDL, R. & VAN DIJL, J. M. 1999a. Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *Journal of Biological Chemistry,* 274**,** 24585-24592.

BOLHUIS, A., SOROKIN, A., AZEVEDO, V., EHRLICH, S. D., BRAUN, P. G., DEJONG, A., VENEMA, G., BRON, S. & VANDIJL, J. M. 1996. *Bacillus subtilis* can modulate its capacity and specificity for protein secretion through temporally controlled expression of the sipS gene for signal peptidase I. *Molecular Microbiology,* 22**,** 605-618.

BOLHUIS, A., TJALSMA, H., SMITH, H. E., DE JONG, A., MEIMA, R., VENEMA, G., BRON, S. & VAN DIJL, J. M. 1999b. Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*. *Applied and Environmental Microbiology,* 65**,** 2934-2941.

BOLHUIS, A., VENEMA, G., QUAX, W. J., BRON, S. & VAN DIJL, J. M. 1999c. Functional analysis of paralogous thiol-disulfide oxidoreductases in *Bacillus subtilis*. *Journal of Biological Chemistry,* 274**,** 24531-24538.

BORCHERT, T. V. & NAGARAJAN, V. 1991. Effect of signal sequence alterations on export of levansucrase in *Bacillus subtilis*. *Journal of Bacteriology,* 173**,** 276-282.

BOSMA, E. F., VAN DE WEIJER, A. H. P., DAAS, M. J. A., VAN DER OOST, J., DE VOS, W. M. & VAN KRANENBURG, R. 2015. Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of *Bacillus smithii* ET 138 as a New Biocatalyst. *Applied and Environmental Microbiology,* 81**,** 1874-1883.

BRIGGS, M. S., CORNELL, D. G., DLUHY, R. A. & GIERASCH, L. M. 1986. Conformations of signal peptides induced by lipids suggest initial steps in protein export. *Science,* 233**,** 206-208.

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. *J Mol Biol,* 362**,** 393-402.

BRON, S., BOLHUIS, A., TJALSMA, H., HOLSAPPEL, S., VENEMA, G. & VAN DIJL, J. M. 1998. Protein secretion and possible roles for multiple signal peptidases for precursor processing in *Bacilli*. *Journal of Biotechnology,* 64**,** 3-13.

CAMPO, N., TJALSMA, H., BUIST, G., STEPNIAK, D., MEIJER, M., VEENHUIS, M., WESTERMANN, M., MULLER, J. P., BRON, S., KOK, J., KUIPERS, O. P. & JONGBLOED, J. D. H. 2004. Subcellular sites for bacterial protein export. *Molecular Microbiology,* 53**,** 1583-1599.

CANAKCI, S., CEVHER, Z., INAN, K., TOKGOZ, M., BAHAR, F., KACAGAN, M., SAL, F. A. & BELDUZ, A. O. 2012. Cloning, purification and characterization of an alkali-stable endoxylanase from thermophilic *Geobacillus* sp. 71. *World J Microbiol Biotechnol,* 28**,** 1981-8.

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. *Appl Microbiol Biotechnol,* 86**,** 1877-85.

CHAKDAR, H., KUMAR, M., PANDIYAN, K., SINGH, A., NANJAPPAN, K., KASHYAP, P. L. & SRIVASTAVA, A. K. 2016. Bacterial xylanases: biology to biotechnology. *3 Biotech,* 6.

CHASSY, B. M. 1976. A gentle method for the lysis of oral *Streptococci*. *Biochemical and Biophysical Research Communications,* 68**,** 603-608.

CHATZI, K. E., SARDIS, M. F., ECONOMOU, A. & KARAMANOU, S. 2014a. SecA-mediated targeting and translocation of secretory proteins. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1843**,** 1466-1474.

CHATZI, K. E., SARDIS, M. F., ECONOMOU, A. & KARAMANOU, S. 2014b. SecA-mediated targeting and translocation of secretory proteins. *Biochim Biophys Acta,* 1843**,** 1466-74.

CHAVEZ, R., BULL, P. & EYZAGUIRRE, J. 2006. The xylanolytic enzyme system from the genus *Penicillium*. *Journal of Biotechnology,* 123**,** 413-433.

CHEN, J., ZHANG, Z., ZHANG, C. & YU, B. 2015a. Genome sequence of *Geobacillus thermoglucosidasius* DSM2542, a platform hosts for biotechnological applications with industrial potential. *J Biotechnol,* 216**,** 98-9.

CHEN, J. Q., FU, G., GAI, Y. M., ZHENG, P., ZHANG, D. W. & WEN, J. P. 2015b. Combinatorial Sec pathway analysis for improved heterologous protein secretion in Bacillus subtilis: identification of bottlenecks by systematic gene overexpression. *Microbial Cell Factories,* 14.

CHEN, J. Q., GAI, Y. M., FU, G., ZHOU, W. J., ZHANG, D. W. & WEN, J. P. 2015c. Enhanced extracellular production of alpha-amylase in *Bacillus subtilis* by optimization of regulatory elements and over-expression of PrsA lipoprotein. *Biotechnology Letters,* 37**,** 899-906.

CHEN, M. & NAGARAJAN, V. 1994. Effect of alteration of charged residues at the n-termini of signal peptides on protein export in *Bacillus subtilis*. *Journal of Bacteriology,* 176**,** 5796-5801.

CLAUSEN, T., KAISER, M., HUBER, R. & EHRMANN, M. 2011. HtrA proteases: regulated proteolysis in protein quality control. *Nature Reviews Molecular Cell Biology,* 12**,** 152-162.

COLLIER, D. N. 1994. Expression of *Escherichia coli secb* in *Bacillus subtilis* facilitates secretion of the Secb-dependent maltose-binding protein of *Escherichia coli*. *Journal of Bacteriology,* 176**,** 4937-4940.

COLLINS, T., GERDAY, C. & FELLER, G. 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev,* 29**,** 3-23.

CRANEY, A., TAHLAN, K., ANDREWS, D. & NODWELL, J. 2011. Bacterial transmembrane proteins that lack N-terminal signal sequences. *PLoS One,* 6**,** e19421.

CRIPPS, R. E., ELEY, K., LEAK, D. J., RUDD, B., TAYLOR, M., TODD, M., BOAKES, S., MARTIN, S. & ATKINSON, T. 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab Eng,* 11**,** 398-408.

CRISTOBAL, S., DE GIER, J. W., NIELSEN, H. & VON HEIJNE, G. 1999. Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. *Embo Journal,* 18**,** 2982-2990.

DAAS, M. J. A., VAN DE WEIJER, A. H. P., DE VOS, W. M., VAN DER OOST, J. & VAN KRANENBURG, R. 2016. Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. *Biotechnology for Biofuels,* 9.

DARMON, E., NOONE, D., MASSON, A., BRON, S., KUIPERS, O. P., DEVINE, K. M. & DIJL, J. M. V. 2002. A Novel Class of Heat and Secretion Stress-Responsive Genes Is Controlled by the Autoregulated CssRS Two-Component System of *Bacillus subtilis*. *Journal of Bacteriology,* 184**,** 5661-5671.

DE KEYZER, J., VAN DER DOES, C. & DRIESSEN, A. J. 2003. The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci,* 60**,** 2034-52.

DEGERING, C., EGGERT, T., PULS, M., BONGAERTS, J., EVERS, S., MAURER, K. H. & JAEGER, K. E. 2010. Optimization of protease secretion in *Bacillus subtilis* and *Bacillus*

*licheniformi*s by screening of homologous and heterologous signal peptides. *Appl Environ Microbiol,* 76**,** 6370-6.

DELIC, M., GONGRICH, R., MATTANOVICH, D. & GASSER, B. 2014. Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxid Redox Signal,* 21**,** 414-37.

DEMCHICK, P. & KOCH, A. L. 1996. The permeability of the wall fabric of *Escherichia coli* and *Bacillus subtilis*. *Journal of Bacteriology,* 178**,** 768-773.

DESHAIES, R. J. & SCHEKMAN, R. 1987. A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic-reticulum. *Journal of Cell Biology,* 105**,** 633-645.

DEUERLING, E., MOGK, A., RICHTER, C., PURUCKER, M. & SCHUMANN, W. 1997. The ftsH gene of *Bacillus subtilis* is involved in major cellular processes such as sporulation, stress adaptation and secretion. *Molecular Microbiology,* 23**,** 921-933.

DEVRIJE, G. J., BATENBURG, A. M., KILLIAN, J. A. & DEKRUIJFF, B. 1990. Lipid involvement in protein translocation in *Escherichia coli*. *Molecular Microbiology,* 4**,** 143-150.

DIAO, L. Y., DONG, Q. L., XU, Z. H., YANG, S., ZHOU, J. H. & FREUDL, R. 2012. Functional Implementation of the Posttranslational SecB-SecA Protein-Targeting Pathway in Bacillus subtilis. *Applied and Environmental Microbiology,* 78**,** 651-659.

FEIN, J. E., CHARLEY, R. C., HOPKINS, K. A., LAVERS, B. & LAWFORD, H. G. 1983. Development of a simple defined medium for continuous ethanol-production by *Zymomonas mobilis*. *Biotechnology Letters,* 5**,** 1-6.

FEKKES, P. & DRIESSEN, A. J. M. 1999. Protein targeting to the bacterial cytoplasmic membrane. *Microbiology and Molecular Biology Reviews,* 63**,** 161-+.

FEKKES, P., VANDERDOES, C. & DRIESSEN, A. J. M. 1997. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *Embo Journal,* 16**,** 6105-6113.

FINORE, I., KASAVI, C., POLI, A., ROMANO, I., ONER, E. T., KIRDAR, B., DIPASQUALE, L., NICOLAUS, B. & LAMA, L. 2011. Purification, biochemical characterization and gene sequencing of a thermostable raw starch digesting α-amylase from *Geobacillus thermoleovorans* subsp. *stromboliensis* subsp. *nov*. *World Journal of Microbiology and Biotechnology,* 27**,** 2425-2433.

FONTES, C. & GILBERT, H. J. 2010. Cellulosomes: Highly Efficient Nanomachines Designed to Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *In:*

KORNBERG, R. D., RAETZ, C. R. H., ROTHMAN, J. E. & THORNER, J. W. (eds.) *Annual Review of Biochemistry, Vol 79.*

GENNITY, J., GOLDSTEIN, J. & INOUYE, M. 1990. Signal peptide mutants of *Escherichia coli*. *Journal of Bioenergetics and Biomembranes,* 22**,** 233-269.

GOLDSTEIN, J., LEHNHARDT, S. & INOUYE, M. 1990. Enhancement of protein translocation across the membrane by specific mutations in the hydrophobic region of the signal peptide. *Journal of Bacteriology,* 172**,** 1225-1231.

GONZALEZ-PASTOR, J. E. 2011. Cannibalism: a social behavior in sporulating *Bacillus subtilis*. *FEMS Microbiol Rev,* 35**,** 415-24.

GOOSENS, V. J., MONTEFERRANTE, C. G. & VAN DIJL, J. M. 2014. The Tat system of Gram-positive bacteria. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1843**,** 1698-1706.

GOURIDIS, G., KARAMANOU, S., GELIS, I., KALODIMOS, C. G. & ECONOMOU, A. 2009. Signal peptides are allosteric activators of the protein translocase. *Nature,* 462**,** 363-U132.

GRUDNIK, P., BANGE, G. & SINNING, I. 2009. Protein targeting by the signal recognition particle. *Biological Chemistry,* 390**,** 775-782.

GUIRAL, S., MITCHELL, T. J., MARTIN, B. & CLAVERYS, J. P. 2005. Competence-programmed predation of noncompletent cells in the human pathogen *Streptococcus pneumoniae:* Genetic requirements. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 8710-8715.

GULLON, S., VICENTE, R. L. & MELLADO, R. P. 2012. A novel two-component system involved in secretion stress response in *Streptomyces lividans*. *PLoS One,* 7**,** e48987.

HABER, W. 2007. Energy, food, and land — The ecological traps of humankind. *Environmental Science and Pollution Research - International,* 14**,** 359-365.

HARTL, F. U., LECKER, S., SCHIEBEL, E., HENDRICK, J. P. & WICKNER, W. 1990. The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *Escherichia coli* plasma-membrane. *Cell,* 63**,** 269-279.

HARWOOD, C. R. & CRANENBURGH, R. 2008. *Bacillus* protein secretion: an unfolding story. *Trends Microbiol,* 16**,** 73-9.

HATSUZAWA, K., TAGAYA, M. & MIZUSHIMA, S. 1997. Hydrophobic region of signal peptides is a determinant for SRP recognition and protein translocation across the ER membrane. *Journal of Biochemistry,* 121**,** 270-277.

HECK, J. X., SOARES, L. H. D. B., HERTZ, P. F. & AYUB, M. A. Z. 2006. Purification and properties of a xylanase produced by *Bacillus circulans* BL53 on solid-state cultivation. *Biochemical Engineering Journal,* 32**,** 179-184.

HEMMERICH, J., ROHE, P., KLEINE, B., JURISCHKA, S., WIECHERT, W., FREUDL, R. & OLDIGES, M. 2016. Use of a Sec signal peptide library from *Bacillus subtilis* for the optimization of cutinase secretion in *Corynebacterium glutamicum*. *Microbial Cell Factories,* 15.

HESTERKAMP, T., HAUSER, S., LUTCKE, H. & BUKAU, B. 1996. *Escherichia coli* trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 4437-4441.

HETENYI, K., NEMETH, A. & SEVELLA, B. 2011. Investigation and modeling of lactic acid fermentation on wheat starch via SSF, CHF and SHF technology. *Periodica Polytechnica-Chemical Engineering,* 55**,** 11-16.

HILLER, K., GROTE, A., SCHEER, M., MUNCH, R. & JAHN, D. 2004. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research,* 32**,** W375-W379.

HIRANO, K., KUROSAKI, M., NIHEI, S., HASEGAWA, H., SHINODA, S., HARUKI, M. & HIRANO, N. 2016. Enzymatic diversity of the *Clostridium thermocellum* cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. *Scientific Reports,* 6.

HIROSE, I., SANO, K., SHIODA, I., KUMANO, M., NAKAMURA, K. & YAMANE, K. 2000. Proteome analysis of Bacillus subtilis extracellular proteins: a two-dimensional protein electrophoretic study. *Microbiology-Uk,* 146**,** 65-75.

HOFLER, C., HECKMANN, J., FRITSCH, A., POPP, P., GEBHARD, S., FRITZ, G. & MASCHERT, T. 2016. Cannibalism stress response in *Bacillus subtilis*. *Microbiology-Sgm,* 162**,** 164-176.

HUBER, D., BOYD, D., XIA, Y., OLMA, M. H., GERSTEIN, M. & BECKWITH, J. 2005. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *Journal of Bacteriology,* 187**,** 2983-2991.

HUBER, D., JAMSHAD, M., HANMER, R., SCHIBICH, D., DORING, K., MARCOMINI, I., KRAMER, G. & BUKAU, B. 2017. SecA Cotranslationally Interacts with Nascent Substrate Proteins In Vivo. *Journal of Bacteriology,* 199.

HUBER, D., RAJAGOPALAN, N., PREISSLER, S., ROCCO, M. A., MERZ, F., KRAMER, G. & BUKAU, B. 2011. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. *Mol Cell,* 41**,** 343-53.

HUMPHREYS, D. P., SEHDEV, M., CHAPMAN, A. P., GANESH, R., SMITH, B. J., KING, L. M., GLOVER, D. J., REEKS, D. G. & STEPHENS, P. E. 2000. High-level periplasmic expression in *Escherichia coli* using a eukaryotic signal peptide: Importance of codon usage at the 5 ' end of the coding sequence. *Protein Expression and Purification,* 20**,** 252-264.

HUSSAIN, M., OZAWA, Y., ICHIHARA, S. & MIZUSHIMA, S. 1982. Signal peptide digestion in *Escherichia-coli* - effect of protease inhibitors on hydrolysis of the cleaved signal peptide of the major outer-membrane lipoprotein. *European Journal of Biochemistry,* 129**,** 233-239.

HYYRYLAINEN, H. L., SARVAS, M. & KONTINEN, V. P. 2005. Transcriptome analysis of the secretion stress response of *Bacillus subtilis*. *Applied Microbiology and Biotechnology,* 67**,** 389-396.

INGRAM, L. O., CONWAY, T., CLARK, D. P., SEWELL, G. W. & PRESTON, J. F. 1987. Genetic-engineering of ethanol-production in *Escherichia coli*. *Applied and Environmental Microbiology,* 53**,** 2420-2425.

IRFAN, M., GULER, H. I., BELDUZ, A. O., SHAH, A. A. & CANAKCI, S. 2016. Cloning, purification and characterization of a cellulase-free xylanase from *Geobacillus thermodenitrificans* AK53. *Applied Biochemistry and Microbiology,* 52**,** 277-286.

JACOBS, M., ANDERSEN, J. B., KONTINEN, V. & SARVAS, M. 1993. *Bacillus subtilis* PrsA is required invivo as an extracytoplasmic chaperone for secretion of active enzymes synthesized either with or without prosequences. *Molecular Microbiology,* 8**,** 957-966.

JANDA, C. Y., LI, J., OUBRIDGE, C., HERNANDEZ, H., ROBINSON, C. V. & NAGAI, K. 2010. Recognition of a signal peptide by the signal recognition particle. *Nature,* 465**,** 507-U139.

JENSEN, C. L., STEPHENSON, K., JORGENSEN, S. T. & HARWOOD, C. 2000. Cell-associated degradation affects the yield of secreted engineered and heterologous proteins in the Bacillus subtilis expression system. *Microbiology-Uk,* 146**,** 2583-2594.

JONET, M. A., MAHADI, N. M., MURAD, A. M., RABU, A., BAKAR, F. D., RAHIM, R. A., LOW, K. O. & ILLIAS, R. M. 2012. Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*. *J Mol Microbiol Biotechnol,* 22**,** 48-58.

JONGBLOED, J. D. H., VAN DER PLOEG, R. & VAN DIJL, J. M. 2006. Bifunctional TatA subunits in minimal Tat protein translocases. *Trends in Microbiology,* 14**,** 2-4.

JUNCKER, A. S., WILLENBROCK, H., VON HEIJNE, G., BRUNAK, S., NIELSEN, H. & KROGH, A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science,* 12**,** 1652-1662.

KAKESHITA, H., KAGEYAMA, Y., ENDO, K., TOHATA, M., ARA, K., OZAKI, K. & NAKAMURA, K. 2011. Secretion of biologically-active human interferon-beta by *Bacillus subtilis*. *Biotechnol Lett,* 33**,** 1847-52.

KALL, L., KROGH, A. & SONNHAMMER, E. L. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol,* 338**,** 1027-36.

KAMBLE, R. D. & JADHAV, A. R. 2012. Isolation, purification, and characterization of xylanase produced by a new species of bacillus in solid state fermentation. *Int J Microbiol,* 2012**,** 683193.

KANANAVICIUTE, R. & CITAVICIUS, D. 2015. Genetic engineering of *Geobacillus* spp. *J Microbiol Methods,* 111**,** 31-9.

KEMPF, G., WILD, K. & SINNING, I. 2014. Structure of the complete bacterial SRP Alu domain. *Nucleic Acids Research,* 42**,** 12284-12294.

KIHARA, A., AKIYAMA, Y. & ITO, K. 1995. FTSH IS REQUIRED FOR PROTEOLYTIC ELIMINATION OF UNCOMPLEXED FORMS OF SECY, AN ESSENTIAL PROTEIN TRANSLOCASE SUBUNIT. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 4532-4536.

KIM, P. S. & ARVAN, P. 1991. Folding and assembly of newly synthesized thyroglobulin occurs in a pre-golgi compartment. *Journal of Biological Chemistry,* 266**,** 12412-12418.

KOBER, L., ZEHE, C. & BODE, J. 2013. Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol Bioeng,* 110**,** 1164-73.

KONTINEN, V. P., SARIS, P. & SARVAS, M. 1991. A gene (prsA) of *Bacillus subtilis* involved in a novel, late stage of protein export. *Molecular Microbiology,* 5**,** 1273-1283.

KONTINEN, V. P. & SARVAS, M. 1993. The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Molecular Microbiology,* 8**,** 727-737.

KRISHNAPPA, L., DREISBACH, A., OTTO, A., GOOSENS, V. J., CRANENBURGH, R. M., HARWOOD, C. R., BECHER, D. & VAN DIJL, J. M. 2013. Extracytoplasmic proteases determining the cleavage and release of secreted proteins, lipoproteins, and membrane proteins in *Bacillus subtilis*. *J Proteome Res,* 12**,** 4101-10.

KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology,* 305**,** 567-580.

KUSUKAWA, N., YURA, T., UEGUCHI, C., AKIYAMA, Y. & ITO, K. 1989. Effects of mutations in heat-shock genes *groES* and *groEL* on protein export in *Escherichia coli*. *Embo Journal,* 8**,** 3517-3521.

LAN, E. I. & LIAO, J. C. 2013. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. *Bioresource Technology,* 135**,** 339-349.

LEE, C. K., JANG, M. Y., PARK, H. R., CHOO, G. C., CHO, H. S., PARK, S. B., OH, K. C., AN, J. B. & KIM, B. G. 2016. Cloning and characterization of xylanase in cellulolytic *Bacillus* sp strain JMY1 isolated from forest soil. *Applied Biological Chemistry,* 59**,** 415-423.

LEE, H. C. & BERNSTEIN, H. D. 2002. Trigger factor retards protein export in *Escherichia coli*. *Journal of Biological Chemistry,* 277**,** 43527-43535.

LI, J., LIN, J., ZHOU, P., WU, K., LIU, H., XIONG, C., GONG, Y., XIAO, W. & LIU, Z. 2014. One-pot simultaneous saccharification and fermentation: a preliminary study of a novel configuration for cellulosic ethanol production. *Bioresour Technol,* 161**,** 171-8.

LI, L., PARK, E., LING, J., INGRAM, J., PLOEGH, H. & RAPOPORT, T. A. 2016. Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature,* 531**,** 395-9.

LI, W., ZHOU, X. & LU, P. 2004. Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res Microbiol,* 155**,** 605-10.

LIAO, J. C., MI, L., PONTRELLI, S. & LUO, S. S. 2016. Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology,* 14**,** 288-304.

LILL, R., DOWHAN, W. & WICKNER, W. 1990. The ATPase activity of SecA is regulated by acidic phospholipids, secy, and the leader and mature domains of precursor proteins. *Cell,* 60**,** 271-280.

LIN, P. P., RABE, K. S., TAKASUMI, J. L., KADISCH, M., ARNOLD, F. H. & LIAO, J. C. 2014. Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metab Eng,* 24**,** 1-8.

LOW, K. O., JONET, M. A., ISMAIL, N. F. & ILLIAS, R. M. 2012. Optimization of a *Bacillus sp* signal peptide for improved recombinant protein secretion and cell viability in *Escherichia coli* : Is there an optimal signal peptide design? *Bioengineered,* 3**,** 334-8.

LOW, K. O., MUHAMMAD MAHADI, N. & MD ILLIAS, R. 2013. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol,* 97**,** 3811-26.

LUIRINK, J. & SINNING, I. 2004. SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta,* 1694**,** 17-35.

LYCKLAMA, A. N. J. A. & DRIESSEN, A. J. 2012. The bacterial Sec-translocase: structure and mechanism. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1016-28.

MAKAROW, M. 1999. Nobel prize in physiology and medicine to Gunter Blobel. *Duodecim; laaketieteellinen aikakauskirja,* 115**,** 2711-4.

MALET, H., CANELLAS, F., SAWA, J., YAN, J., THALASSINOS, K., EHRMANN, M., CLAUSEN, T. & SAIBIL, H. R. 2012. Newly folded substrates inside the molecular cage of the HtrA chaperone DegQ. *Nature Structural & Molecular Biology,* 19**,** 152-157.

MALTEN, M., NAHRSTEDT, H., MEINHARDT, F. & JAHN, D. 2005. Coexpression of the type I signal peptidase gene sipM increases recombinant protein production and export in *Bacillus megaterium* MS941. *Biotechnology and Bioengineering,* 91**,** 616-621.

MARGOT, P. & KARAMATA, D. 1996. The *wprA* gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease. *Microbiology-Uk,* 142**,** 3437-3444.

MCDONALD, J. H. 2010. Temperature Adaptation at Homologous Sites in Proteins from Nine Thermophile-Mesophile Species Pairs. *Genome Biology and Evolution,* 2**,** 267-276.

MERUELO, A. D., HAN, S. K., KIM, S. & BOWIE, J. U. 2012. Structural differences between thermophilic and mesophilic membrane proteins. *Protein Science,* 21**,** 1746-1753.

MONNERON, A., PALADE, G. E. & BLOBEL, G. 1972. Fractionation of nucleus by divalent cations - Isolation of nuclear membranes. *Journal of Cell Biology,* 55**,** 104-&.

MUESCH, A., HARTMANN, E., ROHDE, K., RUBARTELLI, A., SITIA, R. & RAPOPORT, T. A. 1990. A novel pathway for secretory proteins. *Trends in Biochemical Sciences,* 15**,** 86-88.

MULDER, K. C. L., BANDOLA, J. & SCHUMANN, W. 2013. Construction of an artificial secYEG operon allowing high level secretion of alpha-amylase. *Protein Expression and Purification,* 89**,** 92-96.

MULLER, J. P., BRON, S., VENEMA, G. & VAN DIJL, J. M. 2000a. Chaperone-like activities of the CsaA protein of *Bacillus subtilis*. *Microbiology,* 146 (Pt 1)**,** 77-88.

MULLER, J. P., OZEGOWSKI, J., VETTERMANN, S., SWAVING, J., VAN WELY, K. H. M. & DRIESSEN, A. J. M. 2000b. Interaction of Bacillus subtilis CsaA with SecA and precursor proteins. *Biochemical Journal,* 348**,** 367-373.

NANDY, S. K., BAPAT, P. M. & VENKATESH, K. V. 2007. Sporulating bacteria prefers predation to cannibalism in mixed cultures. *Febs Letters,* 581**,** 151-156.

NATALE, P., BRUSER, T. & DRIESSEN, A. J. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-distinct translocases and mechanisms. *Biochim Biophys Acta,* 1778**,** 1735-56.

NAZINA, T. N., TOUROVA, T. P., POLTARAUS, A. B., NOVIKOVA, E. V., GRIGORYAN, A. A., IVANOVA, A. E., LYSENKO, A. M., PETRUNYAKA, V. V., OSIPOV, G. A., BELYAEV, S. S. & IVANOV, M. V. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp nov and *Geobacillus uzenensis* sp nov from petroleum reservoirs and transfer of *Bacillus stearothermophilus Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius and Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius and G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology,* 51**,** 433-446.

NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & VONHEIJNE, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering,* 10**,** 1-6.

NIELSEN, H. & KROGH, A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol,* 6**,** 122-30.

NISSLEY, D. A., SHARMA, A. K., AHMED, N., FRIEDRICH, U. A., KRAMER, G., BUKAU, B. & O'BRIEN, E. P. 2016. Accurate prediction of cellular co-translational folding indicates proteins can switch from post- to co-translational folding. *Nature Communications,* 7.

NOUWEN, N., TOMMASSEN, J. & DEKRUIJFF, B. 1994. Requirement for conformational flexibility in the signal sequence of precursor protein. *Journal of Biological Chemistry,* 269**,** 16029-16033.

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.*

PALMER, T. & BERKS, B. C. 2012. The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol,* 10**,** 483-96.

PAPANIKOU, E., KARAMANOU, S. & ECONOMOU, A. 2007. Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol,* 5**,** 839-51.

PARISUTHAM, V., KIM, T. H. & LEE, S. K. 2014. Feasibilities of consolidated bioprocessing microbes: from pretreatment to biofuel production. *Bioresour Technol,* 161**,** 431-40.

PATERSON, G. K. & MITCHELL, T. J. 2004. The biology of Gram-positive sortase enzymes. *Trends Microbiol,* 12**,** 89-95.

PERALTA-YAHYA, P. P., ZHANG, F. Z., DEL CARDAYRE, S. B. & KEASLING, J. D. 2012. Microbial engineering for the production of advanced biofuels. *Nature,* 488**,** 320-328.

PEREZ-RODRIGUEZ, R., FISHER, A. C., PERLMUTTER, J. D., HICKS, M. G., CHANAL, A., SANTINI, C. L., WU, L. F., PALMER, T. & DELISA, M. P. 2007. An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *Journal of Molecular Biology,* 367**,** 715-730.

PEREZ, J., MUNOZ-DORADO, J., DE LA RUBIA, T. & MARTINEZ, J. 2002. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol,* 5**,** 53-63.

PETERS, J. M., SILVIS, M. R., ZHAO, D. H., HAWKINS, J. S., GROSS, C. A. & QI, L. S. 2015. Bacterial CRISPR: accomplishments and prospects. *Current Opinion in Microbiology,* 27**,** 121-126.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8**,** 785-786.

POHL, S., BHAVSAR, G., HULME, J., BLOOR, A. E., MISIRLI, G., LECKENBY, M. W., RADFORD, D. S., SMITH, W., WIPAT, A., WILLIAMSON, E. D., HARWOOD, C. R. & CRANENBURGH, R. M. 2013. Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics,* 13**,** 3298-308.

PRAGAI, Z., TJALSMA, H., BOLHUIS, A., VANDIJL, J. M., VENEMA, G. & BRON, S. 1997. The signal peptidase II (lsp) gene of *Bacillus subtilis*. *Microbiology-Uk,* 143**,** 1327-1333.

REEVE, B., MARTINEZ-KLIMOVA, E., DE JONGHE, J., LEAK, D. J. & ELLIS, T. 2016. The Geobacillus Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synth Biol*.

RHEE, S. K., JEON, C. O., BAE, J. W., KIM, K., SONG, J. J., KIM, J. J., LEE, S. G., KIM, H. I., HONG, S. P., CHOI, Y. H., KIM, S. M. & SUNG, M. H. 2002. Characterization of *Symbiobacterium toebii*, an obligate commensal thermophile isolated from compost. *Extremophiles,* 6**,** 57-64.

RHEE, S. K., LEE, S. G., HONG, S. P., CHOI, Y. H., PARK, J. H., KIM, C. J. & SUNG, M. H. 2000. A novel microbial interaction: obligate commensalism between a new gram-negative thermophile and a thermophilic *Bacillus* strain. *Extremophiles,* 4**,** 131-136.

ROBINSON, C. & BOLHUIS, A. 2004. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim Biophys Acta,* 1694**,** 135-47.

ROE, S. 2001. *Protein purification techniques: a practical approach,* Oxford, UK, Oxford University Press.

ROSCH, J. & CAPARON, M. 2004. A Microdomain for protein secretion in Gram-positive bacteria. *Science,* 304**,** 1513-1515.

RUBARTELLI, A., COZZOLINO, F., TALIO, M. & SITIA, R. 1990. A novel secretory pathway for interleukin-1-beta, a protein lacking a signal sequence. *Embo Journal,* 9**,** 1503-1510.

SADEGHI, M., NADERI-MANESH, H., ZARRABI, M. & RANJBAR, B. 2006. Effective factors in thermostability of thermophilic proteins. *Biophysical Chemistry,* 119**,** 256-270.

SALAMOV, V. S. A. A. 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. *In:* LI, R. W. (ed.) *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies.* Nova Science Publishers.

SALZBERG, L. I., POWELL, L., HOKAMP, K., BOTELLA, E., NOONE, D. & DEVINE, K. M. 2013. The WalRK (YycFG) and sigma(I) RsgI regulators cooperate to control CwlO and LytE expression in exponentially growing and stressed *Bacillus subtilis* cells. *Molecular Microbiology,* 87**,** 180-195.

SANCHEZ, O. J. & CARDONA, C. A. 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresour Technol,* 99**,** 5270-95.

SARGENT, F. 2001. A marriage of bacteriology with cell biology results in twin arginines. *Trends in Microbiology,* 9**,** 196-198.

SARVAS, M., HARWOOD, C. R., BRON, S. & VAN DIJL, J. M. 2004. Post-translocational folding of secretory proteins in Gram-positive bacteria. *Biochim Biophys Acta,* 1694**,** 311-27.

SCHALLMEY, M., SINGH, A. & WARD, O. P. 2004. Developments in the use of *Bacillus* species for industrial production. *Canadian Journal of Microbiology,* 50**,** 1-17.

SCHNEEWIND, O. & MISSIAKAS, D. 2014. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochim Biophys Acta,* 1843**,** 1687-97.

SCHNEEWIND, O. & MISSIAKAS, D. M. 2012. Protein secretion and surface display in Gram-positive bacteria. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1123-39.

SCHRODER, H., LANGER, T., HARTL, F. U. & BUKAU, B. 1993. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *Embo Journal,* 12**,** 4137-4144.

SHAN, S. O. & WALTER, P. 2005. Co-translational protein targeting by the signal recognition particle. *FEBS Lett,* 579**,** 921-6.

SHARMA, M., MEHTA, S. & KUMAR, A. 2013. Purification and Characterization of Alkaline Xylanase Secreted from *Paenibacillus macquariensis*. *Advances in Microbiology,* 03**,** 32-41.

SHAW, A. J., PODKAMINER, K. K., DESAI, S. G., BARDSLEY, J. S., ROGERS, S. R., THORNE, P. G., HOGSETT, D. A. & LYND, L. R. 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proc Natl Acad Sci U S A,* 105**,** 13769-74.

SHAW, A. J., PODKAMINER, K. K., DESAI, S. G., BARDSLEY, J. S., ROGERS, S. R., THORNE, P. G., HOGSETT, D. A. & LYND, L. R. 2009. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *International Sugar Journal,* 111**,** 164-171.

SHEVCHENKO, A., TOMAS, H., HAVLIS, J., OLSEN, J. V. & MANN, M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc,* 1**,** 2856-60.

SHINDE, U. P., ROW, T. N. G. & MAWAL, Y. R. 1989. Export of proteins across membranes - the helix reversion hypothesis. *Bioscience Reports,* 9**,** 737-745.

SIANIDIS, G., KARAMANOU, S., VRONTOU, E., BOULIAS, K., REPANAS, K., KYRPIDES, N., POLITOU, A. S. & ECONOMOU, A. 2001. Cross-talk between catalytic and regulatory elements in a DEAD motor domain is essential for SecA function. *Embo Journal,* 20**,** 961-970.

SILHAVY, T. J., KAHNE, D. & WALKER, S. 2010. The bacterial cell envelope. *Cold Spring Harb Perspect Biol,* 2**,** a000414.

SIMON, E. & KORNITZER, D. 2014. Pulse-Chase Analysis to Measure Protein Degradation. *In:* LORSCH, J. (ed.) *Laboratory Methods in Enzymology: Protein Pt A.*

SINGH, V., BRADDICK, D. & DHAR, P. K. 2017. Exploring the potential of genome editing CRISPR-Cas9 technology. *Gene,* 599**,** 1-18.

STEPHENSON, K. & HARWOOD, C. R. 1998. Influence of a cell-wall-associated protease on production of alpha-amylase by *Bacillus subtilis*. *Applied and Environmental Microbiology,* 64**,** 2875-2881.

STOEGLEHNER, G. & NARODOSLAWSKY, M. 2009. How sustainable are biofuels? Answers and further questions arising from an ecological footprint perspective. *Bioresour Technol,* 100**,** 3825-30.

STORMO, G. D., SCHNEIDER, T. D., GOLD, L. & EHRENFEUCHT, A. 1982. Use of the perceptron algorithm to distinguish translational initiation sites in *Escherichia coli*. *Nucleic Acids Research,* 10**,** 2997-3011.

STOTHARD, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques,* 28**,** 1102-+.

SUN, J.-Y., LIU, M.-Q., WENG, X.-Y., QIAN, L.-C. & GU, S.-H. 2007. Expression of recombinant *Thermomonospora fusca* xylanase A in *Pichia pastoris* and xylooligosaccharides released from xylans by it. *Food Chemistry,* 104**,** 1055-1064.

SUZUKI, H., YOSHIDA, K. & OHSHIMA, T. 2013. Polysaccharide-Degrading Thermophiles Generated by Heterologous Gene Expression in *Geobacillus kaustophilus* HTA426. *Applied and Environmental Microbiology,* 79**,** 5151-5158.

SZILAGYI, A. & ZAVODSZKY, P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure,* 8**,** 493-504.

TAHERZADEH, M. J. & KARIMI, K. 2008. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: A review. *International Journal of Molecular Sciences,* 9**,** 1621-1651.

TAYLOR, M. P., ESTEBAN, C. D. & LEAK, D. J. 2008. Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid,* 60**,** 45-52.

TENENBAUM, D. J. 2008. Food vs. fuel: Diversion of crops could cause more hunger. *Environmental Health Perspectives,* 116**,** A254-A257.

TJALSMA, H., ANTELMANN, H., JONGBLOED, J. D., BRAUN, P. G., DARMON, E., DORENBOS, R., DUBOIS, J. Y., WESTERS, H., ZANEN, G., QUAX, W. J., KUIPERS, O. P., BRON, S., HECKER, M. & VAN DIJL, J. M. 2004. Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. *Microbiol Mol Biol Rev,* 68**,** 207-33.

TJALSMA, H., BOLHUIS, A., BRON, S., JONGBLOED, J., MEIJER, W. J. J., NOBACK, M., VANROOSMALEN, M., VENEMA, G. & VANDIJL, J. M. 1997. Type I signal peptidases of *Bacillus subtilis*. *In:* HOPSUHAVU, V. K., JARVINEN, M. & KIRSCHKE, H. (eds.) *Proteolysis in Cell Functions.*

TJALSMA, H., BOLHUIS, A., JONGBLOED, J. D. H., BRON, S. & VAN DIJL, J. M. 2000. Signal Peptide-Dependent Protein Transport in *Bacillus subtilis*: a Genome-Based Survey of the Secretome. *Microbiology and Molecular Biology Reviews,* 64**,** 515-547.

TJALSMA, H., BOLHUIS, A., VAN ROOSMALEN, M. L., WIEGERT, T., SCHUMANN, W., BROEKHUIZEN, C. P., QUAX, W. J., VENEMA, G., BRON, S. & VAN DIJL, J. M. 1998.

Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases. *Genes & Development,* 12**,** 2318-2331.

TOMINAGA, Y., OHSHIRO, T. & SUZUKI, H. 2016. Conjugative plasmid transfer from *Escherichia coli* is a versatile approach for genetic transformation of thermophilic *Bacillus* and *Geobacillus* species. *Extremophiles,* 20**,** 375-381.

TON-THAT, H., MARRAFFINI, L. A. & SCHNEEWIND, O. 2004. Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta,* 1694**,** 269-78.

TOSSAVAINEN, H., PERMI, P., PURHONEN, S. L., SARVAS, M., KILPELAINEN, I. & SEPPALA, R. 2006. NMR solution structure and characterization of substrate binding site of the PPIase domain of PrsA protein from *Bacillus subtilis*. *Febs Letters,* 580**,** 1822-1826.

TSUKAZAKI, T., MORI, H., ECHIZEN, Y., ISHITANI, R., FUKAI, S., TANAKA, T., PEREDERINA, A., VASSYLYEV, D. G., KOHNO, T., MATURANA, A. D., ITO, K. & NUREKI, O. 2011a. Structure and function of a membrane component SecDF that enhances protein export. *Nature,* 474**,** 235-238.

TSUKAZAKI, T., MORI, H., ECHIZEN, Y., ISHITANI, R., FUKAI, S., TANAKA, T., PEREDERINA, A., VASSYLYEV, D. G., KOHNO, T., MATURANA, A. D., ITO, K. & NUREKI, O. 2011b. Structure and function of a membrane component SecDF that enhances protein export. *Nature,* 474**,** 235-8.

TURNER, P., MAMO, G. & KARLSSON, E. N. 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories,* 6.

TUTEJA, R. 2005. Type I signal peptidase: An overview. *Archives of Biochemistry and Biophysics,* 441**,** 107-111.

VAN ROOSMALEN, M. L., GEUKENS, N., JONGBLOED, J. D., TJALSMA, H., DUBOIS, J. Y., BRON, S., VAN DIJL, J. M. & ANNE, J. 2004. Type I signal peptidases of Gram-positive bacteria. *Biochim Biophys Acta,* 1694**,** 279-97.

VAN WELY, K. H. M., SWAVING, J., BROEKHUIZEN, C. P., ROSE, M., QUAX, W. J. & DRIESSEN, A. J. M. 1999. Functional identification of the product of the *Bacillus subtilis* yvaL gene as a SecG homologue. *Journal of Bacteriology,* 181**,** 1786-1792.

VANVUUREN, H. J. J. & MEYER, L. 1982. Production of ethanol from sugar-cane molasses by *Zymomonas mobilis*. *Biotechnology Letters,* 4**,** 253-256.

VEENING, J. W., IGOSHIN, O. A., EIJLANDER, R. T., NIJLAND, R., HAMOEN, L. W. & KUIPERS, O. P. 2008. Transient heterogeneity in extracellular protease production by Bacillus subtilis. *Molecular Systems Biology,* 4.

VERMA, D., ANAND, A. & SATYANARAYANA, T. 2013. Thermostable and alkalistable endoxylanase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1: cloning, expression, characteristics and its applicability in generating xylooligosaccharides and fermentable sugars. *Appl Biochem Biotechnol,* 170**,** 119-30.

VERMA, D. & SATYANARAYANA, T. 2012. Cloning, expression and applicability of thermo-alkali-stable xylanase of *Geobacillus thermoleovorans* in generating xylooligosaccharides from agro-residues. *Bioresour Technol,* 107**,** 333-8.

VIEILLE, C. & ZEIKUS, G. J. 2001. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews,* 65**,** 1-+.

VITIKAINEN, M., HYYRYLAINEN, H. L., KIVIMAKI, A., KONTINEN, V. P. & SARVAS, M. 2005. Secretion of heterologous proteins in *Bacillus subtilis* can be improved by engineering cell components affecting post-translocational protein folding and degradation. *J Appl Microbiol,* 99**,** 363-75.

VITIKAINEN, M., LAPPALAINEN, I., SEPPALA, R., ANTELMANN, H., BOER, H., TAIRA, S., SAVILAHTI, H., HECKER, M., VIHINEN, M., SARVAS, M. & KONTINEN, V. P. 2004. Structure-function analysis of PrsA reveals roles for the parvulin-like and flanking N- and C-terminal domains in protein folding and secretion in *Bacillus subtilis*. *Journal of Biological Chemistry,* 279**,** 19302-19314.

VITIKAINEN, M., PUMMI, T., AIRAKSINEN, U., WAHLSTROM, E., WU, H., SARVAS, M. & KONTINEN, V. P. 2001. Quantitation of the capacity of the secretion apparatus and requirement for PrsA in growth and secretion of alpha-amylase in *Bacillus subtilis*. *J Bacteriol,* 183**,** 1881-90.

VOGT, G. & ARGOS, P. 1997. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design,* 2, Supplement 1**,** S40-S46.

VOGT, G., WOELL, S. & ARGOS, P. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. *Journal of Molecular Biology,* 269**,** 631-643.

VON HEIJNE, G. 1981. The hydrophobic nature of signal sequences. *European Journal of Biochemistry,* 116**,** 419-422.

VON HEIJNE, G. 1982. Signal sequences are not uniformly hydrophobic. *Journal of Molecular Biology,* 159**,** 537-542.

VON HEIJNE, G. 1984. How signal sequences maintain cleavage specificity. *Journal of Molecular Biology,* 173**,** 243-252.

VON HEIJNE, G. 1985. Signal sequences the limits of variation. *Journal of Molecular Biology,* 184**,** 99-106.

VONHEIJNE, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research,* 14**,** 4683-4690.

VONHEIJNE, G. 1990. THE SIGNAL PEPTIDE. *Journal of Membrane Biology,* 115**,** 195-201.

WAHLSTROM, E., VITIKAINEN, M., KONTINEN, V. P. & SARVAS, M. 2003. The extracytoplasmic folding factor PrsA is required for protein secretion only in the presence of the cell wall in *Bacillus subtilis*. *Microbiology,* 149**,** 569-77.

WALTER, P. & BLOBEL, G. 1980. Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic-reticulum. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 77**,** 7112-7116.

WARREN, G. L. & PETSKO, G. A. 1995. Composition analysis of alpha-helices in thermophilic organisms. *Protein Engineering,* 8**,** 905-913.

WEI, H. & HAVARSTEIN, L. S. 2012. Fratricide is essential for efficient gene transfer between *Pneumococci* in biofilms. *Applied and Environmental Microbiology,* 78**,** 5897-5905.

WELCH, M., VILLALOBOS, A., GUSTAFSSON, C. & MINSHULL, J. 2011. Designing genes for successful protein expression. *Methods Enzymol,* 498**,** 43-66.

WESTERS, H., WESTERS, L., DARMON, E., VAN DIJL, J. M., QUAX, W. J. & ZANEN, G. 2006. The CssRS two-component regulatory system controls a general secretion stress response in *Bacillus subtilis*. *FEBS J,* 273**,** 3816-27.

WILD, J., ALTMAN, E., YURA, T. & GROSS, C. A. 1992. DnaK and DnaJ heat-shock proteins participate in protein export in *Escherichia coli*. *Genes & Development,* 6**,** 1165-1172.

WILLIAMS, R. C., REES, M. L., JACOBS, M. F., PRAGAI, Z., THWAITE, J. E., BAILLIE, L. W. J., EMMERSON, P. T. & HARWOOD, C. R. 2003. Production of Bacillus anthracis protective antigen is dependent on the extracellular chaperone, PrsA. *Journal of Biological Chemistry,* 278**,** 18056-18062.

WOOLHEAD, C., BOLHUIS, A. & ROBINSON, C. 2000. Novel mechanisms for the targeting of proteins into and across chloroplast membranes. *Biochemical Society Transactions,* 28**,** 491-494.

WU, C. C. & MACCOSS, M. J. 2002. Shotgun proteomics: Tools for the analysis of complex biological systems. *Current Opinion in Molecular Therapeutics,* 4**,** 242-250.

WU, S. C., YE, R. Q., WU, X. C., NG, S. C. & WONG, S. L. 1998. Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones. *Journal of Bacteriology,* 180**,** 2830-2835.

WU, S. C., YEUNG, J. C., DUAN, Y., YE, R., SZARKA, S. J., HABIBI, H. R. & WONG, S. L. 2002a. Functional Production and Characterization of a Fibrin-Specific Single-Chain Antibody Fragment from *Bacillus subtilis*: Effects of Molecular Chaperones and a Wall-Bound Protease on Antibody Fragment Production. *Applied and Environmental Microbiology,* 68**,** 3261-3269.

WU, S. C., YEUNG, J. C., DUAN, Y. J., YE, R. Q., SZARKA, S. J., HABIBI, H. R. & WONG, S. L. 2002b. Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: Effects of molecular chaperones and a wall-bound protease on antibody fragment production. *Applied and Environmental Microbiology,* 68**,** 3261-3269.

WU, X. C., LEE, W., TRAN, L. & WONG, S. L. 1991a. ENGINEERING A BACILLUS-SUBTILIS EXPRESSION-SECRETION SYSTEM WITH A STRAIN DEFICIENT IN 6 EXTRACELLULAR PROTEASES. *Journal of Bacteriology,* 173**,** 4952-4958.

WU, X. C., LEE, W., TRAN, L. & WONG, S. L. 1991b. Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in 6 extracellular proteases. *Journal of Bacteriology,* 173**,** 4952-4958.

WU, Z. C., DE KEYZER, J., KEDROV, A. & DRIESSEN, A. J. M. 2012. Competitive Binding of the SecA ATPase and Ribosomes to the SecYEG Translocon. *Journal of Biological Chemistry,* 287**,** 7885-7895.

YAMADA, R., KIMOTO, Y. & OGINO, H. 2016. Combinatorial library strategy for strong overexpression of the lipase from *Geobacillus thermocatenulatus* on the cell surface of yeast *Pichia pastoris*. *Biochemical Engineering Journal,* 113**,** 7-11.

YANG, C. K., EWIS, H. E., ZHANG, X., LU, C. D., HU, H. J., PAN, Y., ABDELAL, A. T. & TAI, P. C. 2011. Nonclassical protein secretion by *Bacillus subtilis* in the stationary phase is not due to cell lysis. *J Bacteriol,* 193**,** 5607-15.

YANG, M. J., JUNG, S. H., SHIN, E. S., KIM, J., YUN, H. D., WONG, S. L. & KIM, H. 2004. Expression of a *Bacillus subtilis* endoglucanase in protease-deficient *Bacillus subtilis* strains. *Journal of Microbiology and Biotechnology,* 14**,** 430-434.

ZABED, H., SAHU, J. N., BOYCE, A. N. & FARUQ, G. 2016. Fuel ethanol production from lignocellulosic biomass: An overview on feedstocks and technological approaches. *Renewable & Sustainable Energy Reviews,* 66**,** 751-774.

ZANEN, G., ANTELMANN, H., MEIMA, R., JONGBLOED, J. D., KOLKMAN, M., HECKER, M., VAN DIJL, J. M. & QUAX, W. J. 2006a. Proteomic dissection of potential signal recognition particle dependence in protein secretion by *Bacillus subtilis*. *Proteomics,* 6**,** 3636-48.

ZANEN, G., ANTELMANN, H., MEIMA, R., JONGBLOED, J. D. H., KOLKMAN, M., HECKER, M., VAN DIJL, J. M. & QUAX, W. J. 2006b. Proteomic dissection of potential signal recognition particle dependence in protein secretion by Bacillus subtilis. *Proteomics,* 6**,** 3636-3648.

ZHANG, X., RASHID, R., WANG, K. & SHAN, S. O. 2010. Sequential Checkpoints Govern Substrate Selection During Cotranslational Protein Targeting. *Science,* 328**,** 757-760.

ZHENG, N. & GIERASCH, L. M. 1996. Signal sequences: The same yet different. *Cell,* 86**,** 849-852.

ZIMMER, J., NAM, Y. & RAPOPORT, T. A. 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature,* 455**,** 936-43.

ZOUARI AYADI, D., BEN ALI, M., JEMLI, S., BEN MABROUK, S., MEZGHANI, M., BEN MESSAOUD, E. & BEJAR, S. 2008. Heterologous expression, secretion and characterization of the *Geobacillus thermoleovorans* US105 type I pullulanase. *Appl Microbiol Biotechnol,* 78**,** 473-81.

ZUOBI-HASONA, K. & BRADY, L. J. 2008. Isolation and Solubilization of Cellular Membrane Proteins from Bacteria. *In:* POSCH, A. (ed.) *2D PAGE: Sample Preparation and Fractionation.* Totowa, NJ: Humana Press.

APPENDIX 1: CELL LYSIS IN GEOBACILLUS THERMOGLUCOSIDASIUS

INTRODUCTION

The cell lysis phenomenon is one that has been a common theme throughout this work. The strains used in this work *GT* C56-YS93 and TM242 appear to lyse, releasing cytoplasmic protein into the culture medium. One possible explanation for the cell lysis in batch culture is that this is an artefact of the laboratory conditions, which are far removed from the organism's natural or native environment or ecosystem. However, it is also conceivable that cell lysis is a natural phenomenon. For instance, one study discovered the apparently symbiotic relationship of *Symbiobacterium toebii* and *Geobacillus toebii,* where *S. toebii* feeds on the lysis products from *G. toebii (Rhee et al., 2000, Rhee et al., 2002)* in the native environment. Furthermore, such "cannibalistic" behaviour have been observed and well documented in *BS* and other sporulating bacteria (Nandy et al., 2007, Gonzalez-Pastor, 2011, Hofler et al., 2016, Guiral et al., 2005, Wei and Havarstein, 2012). This cannibalistic activity has been purported to be a result of exhaustion of nutrients, and a means to delay sporulation which is an energy-intensive process. Similar to *BS* and other mesophilic bacilli, *GT* and other *Geobacilli* encode the majority of essential sporulation genes such as *spo0A*, which has been implicated in cannibalistic behaviour in *BS* (Gonzalez-Pastor, 2011), which is a further suggestion that *Geobacilli*, like their other spore-forming counterparts, may also display predation/cannibalistic behaviour. One study in *Geobacillus thermoleovorans* observed that high growth rates and substrate exhaustion resulted in cell lysis, while this was less with slower growth rates in continuous culture. Throughout this work, however, cell lysis has been observed in mid-log-phase, and not in stationary phase where nutrients would be most likely to be running low, which presents an argument in itself: can cell lysis in *GT* be compared to that in *BS,* in that it is a form of self-sacrifice as a means to delay sporulation in the population?

Another explanation for the apparent cell lysis is non-classical protein secretion, a term that describes the translocation of proteins to the extracellular milieu, all the while lacking a classical signal peptide. This has been observed in several species of bacteria and from intact cells, suggesting that non-classical secretion is not a consequence of cell lysis. Furthermore, functions of several proteins found to be non-classically secreted have been established to be separate from their role in the cytoplasm, and have been termed moonlighting proteins (Bendtsen et al., 2005a), as they appear to have distinct and different functions in the different locations. The term was first coined in 1990, when a group working on human monocytes found that interleukin-1 was found to be present in the extracellular medium, in the absence of evidence of cell lysis or other cytoplasmic proteins (Rubartelli et al., 1990, Muesch et al., 1990). Examples of so-called non-classical secretion have also been found in bacteria, initially in *Mycobacterium sp.,* and later in other pathogenic bacteria. In *B. subtilis* even, a very well studied organism, there has been examples of non-classical protein secretion not due to cell lysis (Yang et al., 2011, Bendtsen et al., 2005a, Antelmann et al., 2001). The detection of non-classically secreted protein in the extracellular milieu could easily be attributed to cell lysis, especially during experimental handling, and could be an artefact of laboratory conditions. However, there is some evidence of proteins being secreted into the extracellular milieu from intact cells, which still needs to be considered.

Methodologies to investigate cell lysis

One simple method of investigating levels of cell lysis is to analyse levels of a cytoplasmic protein in the culture medium. In this work, GroEL was selected as it is present in *GT,* and the antibody to GroEL is commercially available. GroEL, as mentioned in the General Introduction (Chapter 1), is a cytoplasmic chaperone involved in protein folding, and prevention of protein aggregation after synthesis (Schroder et al., 1993). GroEL also does not bear a predictable signal peptide, and plays an important role in the cytoplasm of all bacteria. Using Western blots, we can estimate levels of GroEL protein in the whole cell fraction, and the extracellular fraction, as an indicator of cell lysis.

METHODS AND MATERIALS

Western blot

Rabbit anti-GroEL polyclonal antibodies (Enzo Life Sciences) were used to probe for GroEL protein in different fractions as indicated. Membranes were incubated overnight at 4°C in a concentration of 1 in 1000 primary anti-GroEL antibody in PBS-T, followed by washing and secondary antibody incubation as described in the General Methods and Materials (Chapter 2).

Western blot intensity signals were quantified using Image Studio Lite Ver5.2 (LI-COR Bioscience).

Mass spectrometry

As described in Chapter 3.

RESULTS AND DISCUSSION

Shotgun mass spectrometry analysis of *GT* C56-YS93

The shotgun mass spectrometry analysis was initially carried out to explore the secretome of *GT*, and identify the most abundantly secreted proteins and their corresponding signal peptides. However, from the mass spectrometry analysis, more information can be gleaned on the extracellular milieu of *GT C56-YS93* in batch culture, in a rich medium (TGP medium), other than just the most abundantly secreted proteins. The first and most striking point is the high abundance of purportedly cytoplasmic proteins in the extracellular milieu as can be seen in Table A, which shows a fraction of the proteins identified using the shotgun mass spectrometry approach. After removing duplicates between the three mass spectrometry samples, and removal of obvious contaminants from other species, 540 proteins were identified. The *GT* C56-YS93 predicted proteome has 3656 potential ORFs that could be transcribed into proteins. The mass spectrometry analysis combined with the in-silico prediction show that of the 540 proteins identified using shotgun mass spectrometry, 29 bear signal peptides, which are recognised by type 1 signal peptidases, meaning they are secreted. The proteins are listed in order of abundance based on the number of unique peptides, but it should be noted that the technique is only partially quantitative. The table also shows whether the protein is predicted to be secreted or not, and the sample here highlights the relatively low amount of secreted proteins from the sample.

Table A: A sample of some of the proteins identified using the shotgun mass spectrometry technique. The proteins are ranked by relative abundance in the sample and the relatively small number of secreted proteins for an extracellular fraction is highlighted.

UniProt	Protein description	Peptides	Signal peptide
F8CX47	S-layer domain-containing protein	364	Y
E3IAX4	Flagellin domain protein	166	N
F8CX44	Subtilisin	121	Y
F8CVJ1	L-iditol 2-dehydrogenase	81	N
F8CW93	Formate acetyltransferase	77	N
I0U7E7	Aconitate hydratase 1	71	N
I0U3N1	Elongation factor G	70	N
F8CTR1	Bifunctional purine biosynthesis protein PurH	58	N
I0U5G5	60 kDa chaperonin	53	N
F8D1G3	Phage major capsid protein, HK97 family	51	N
I0U3Y6	Inosine-5'-monophosphate dehydrogenase	50	N
F8CV27	DNA-directed RNA polymerase	48	N
F8CX05	Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	48	Y
I0U5D6	Phosphoribosylformylglycinamidine synthase 2	44	N
I0UCD4	Thioredoxin reductase	42	N
I0U3W8	Cysteine synthase	38	N
I0U606	Isoleucine-tRNA ligase	35	N
F8CV77	NADPH dehydrogenase	27	N
I0UBA4	6-phosphofructokinase	26	N
I0UBF6	Thioredoxin	20	N
I0UAC4	DNA-binding protein HU 1	17	N
I0U3N2	Elongation factor Tu	16	N
I0U4J1	50S ribosomal protein L9	15	N
F8CXU3	Flagellin domain protein	13	N
F8CXW3	Sigma 54 modulation protein	10	N
I0U700	Histidine triad (HIT) protein	10	N

I0U692	Peroxiredoxin	9	N
I0UAE9	Menaquinol-cytochrome c reductase	9	N
I0UA52	Thiol-disulfide oxidoreductase	8	Y

The comparison is shown in Table B, which displays the total number of proteins, those predicted to be secreted proteins of that population, and the percentage. The results show that the experimental secretome is somewhat enriched in putative secretory proteins: 5.3% secretory proteins compared to 2.1% of the predicted secretome. Nonetheless, the amount of cytoplasmic protein in the extracellular milieu is higher than expected, compared to *BS* for example, where only 26% of the extracellular proteome is attributed to cell lysis (Tjalsma et al., 2004) and over 70% of the proteins present in the extracellular milieu were predicted to be secreted. This begs the question, how did these supposedly cytoplasmic proteins end up in the extracellular milieu? Was it through cell lysis, or some other mechanism?

Table B: Number of proteins identified using shotgun mass spectrometry compared to the predicted proteome, and predicted secreted protein.

	Total	**Secreted proteins**	**% predicted secreted**
Shotgun mass spectrometry	540	29	5.3%
Predicted proteome	3656	78	2.1%

Of the total number of proteins identified, 44 are proteases or peptidases, and 10 of these are predicted (using SignalP) to be extracellular proteases, and a further three proteases predicted (using TMHMM) to contain transmembrane domains as shown in Table C. Therefore, 10 of that number is almost one third of the total secreted proteins, a significant proportion. The most highly abundant protease is subtilisin, a serine protease, equivalent to the aprE gene product in *BS.* This would suggest that the organism has some intrinsic need for products of proteolysis in the extracellular medium, and suggests that the organism may utilise peptides and amino acids as a source of nutrition. TMO Renewables have also reported that *GT 11955* and its derivatives also utilise amino acids in a defined growth medium as a carbon source. With this in mind, it is not unlikely that *GT* C56-YS93 also utilises protein hydrolysis products

as a major carbon source, which may be one explanation as to why so many proteases are secreted; this would make sense if cell lysates from sister cells are a source of nutrients.

Table C: Extracellular proteases identified using the mass spectrometry analysis combined with the in-silico prediction (SignalP)

UniProt	Description	Predicted fraction
F8CX44	Subtilisin (Precursor)	secreted
I0UCA4	Extracellular zinc metallopeptidase, M23 family	secreted
F8CX92	Cell wall hydrolase/autolysin (Precursor)	secreted
I0UCA7	Carboxyl-terminal protease	secreted
I0U981	Cell wall-associated hydrolase, NLP/P60 family	secreted
C5D336	Serine-type D-Ala-D-Ala carboxypeptidase	secreted
I0U3B6	Cell-wall bound hydrolase, containing NLP/P60 domain	secreted
I0U3Y7	D-Ala-D-Ala carboxypeptidase, serine-type	secreted
I0UA72	D-alanyl-D-alanine carboxypeptidase, vanY family	secreted
Q5KVA9	Carboxyl-terminal processing protease	secreted
F8CX77	Peptidase S1 and S6 chymotrypsin/Hap	transmembrane
I0U3E7	Extracellular peptidase, trypsin-like family	transmembrane
F8D2J7	HtrA2 peptidase	transmembrane

Cell lysis analysis

Anti-GroEL antibodies were used to detect GroEL in cell lysates and in the extracellular fraction of *GT* TM242. This was initially conducted to corroborate findings from Chapter 5, that XylA^{sp-} was present in the culture medium as a result of cell lysis. As can be seen in Figures A and B, significant levels of GroEL can be detected in the culture medium. Taken at face value, this would suggest that around 30% of the cell density is lysed. The cells were grown to mid-log phase and harvested at an optical density of around 1.5. As such, a large portion of the cells would be undergoing binary fission towards exponential growth, so although during this stage there is significant remodelling of the cell wall architecture, cell lysis would not normally be expected, as cell lysis would normally occur during stationary phase, death phase, or during sporulation.

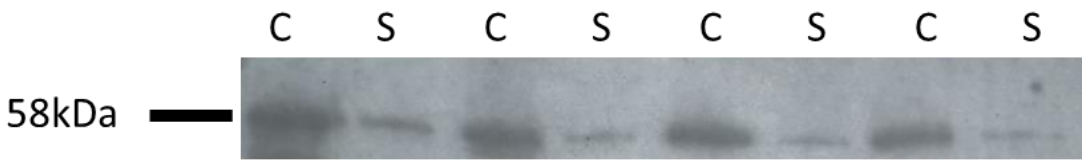

Figure A: Western blot of GroEL in the cell pellet (C) fraction and extracellular milieu (S) fractions of GT TM242

Figure B: Western blot densitometry of GroEL in cell and media fractions from GT TM242 grown in ASM media with 0.2% yeast extract. (n=4)

Figure C shows Western blot densitometry analysis comparing GroEL levels in a culture of TM242-*pUCG4.8xylA* and TM242-*pUCG4.8xylA*$^{-sp}$, to investigate if there is any difference in the ratio of GroEL in the cell pellet and extracellular fractions. The result here shows that there is twice as much GroEL in the extracellular fraction of the strain producing secreted xylanase, compared to the strain producing xylanase lacking the signal peptide. This could suggest that over-producing a secreted protein leads to an increase in cell lysis, although the mechanism is unclear. The work in the previous chapter did not show any evidence of bottlenecks in protein secretion at the membrane, and if cell lysis was triggered by obstruction and congestion of the secretion machinery, secreted protein (i.e. xylanase) would be seen in the membrane fraction of the cell fractionation experiments, unless the accumulated proteins were rapidly degraded and could not be visualised using the methods employed to investigate bottlenecks. Furthermore, in *BS,* secretion stress is detected by the CssRS sensor-histidine kinase, and modulated by inducing the upregulation of extracellular proteases that degrade misfolded protein in the cell wall (Darmon et al., 2002, Westers et al., 2006, Hyyrylainen et al., 2005). This system or a functional homolog has not yet been identified in *GT* so the stress response is still unknown. Furthermore, in *BS*, autolysin production has been shown to be linked to stress conditions (Hyyrylainen et al., 2005, Salzberg et al., 2013).

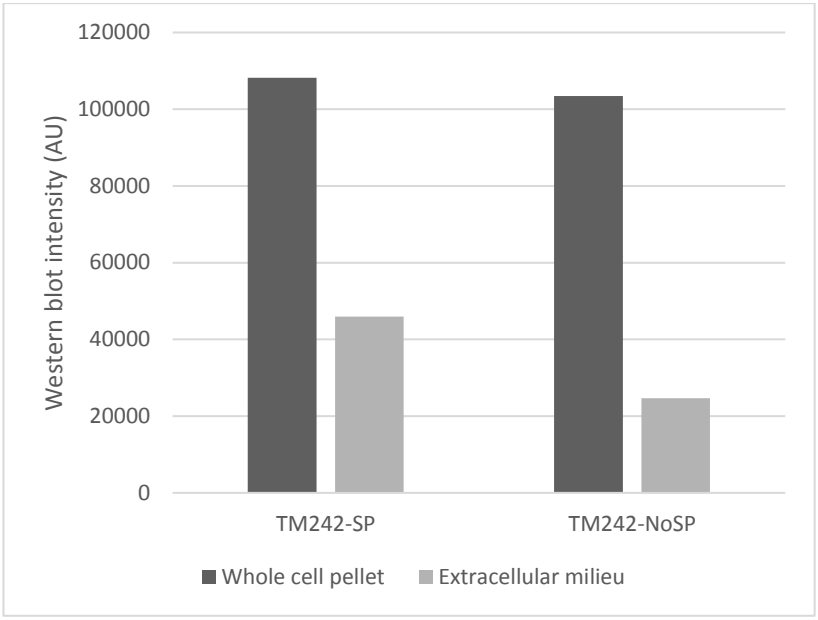

Figure C: Western blot intensity densitometry analysis of GroEL levels in extracellular milieu and whole cell pellet of GT

Conclusions

Could the cell lysis phenomenon be exploited to the advantage of growing *GT* as a means to deliver enzymes such as hydrolases into the extracellular milieu? This is not to say that protein secretion as a means to translocate heterologous proteins is obsolete, as protein secretion is still required to assemble multi-enzyme complexes such as cellulosomes to the cell wall for attachment, which is why it is still important to further investigate bottlenecks in protein secretion in *GT*.

APPENDIX 2: LIST OF PREDICTED SECRETORY PROTEINS OF *GEOBACILLUS THERMOGLUCOSIDASIUS TM242*

Name	ERGO function	Signal sequence and cleavage site
RTMO03117	3D domain protein	**MILLKNIVRRITMSLLFAMALLTTFQAISGVEA**/KVI
RTMO04689	ABC transporter substrate-binding protein	**MMKKRIQQIAMLCSASFLLLSGCGA**/KET
RTMO01915	ABC transporter substrate-binding protein	**MKKQLSIILAFLLSFGILAACGNKETASNA**/AED
RTMO00145	ABC transporter substrate-binding protein	**MKKAIKRLAVPMLVGMLALSGCTKEKT**/ATK
RTMO04256	ABC transporter substrate-binding protein	**MKKTAIVSLFLFLLIIPLAACNQQA**/NKE
RTMO02289	ABC transporter substrate-binding protein	**MKKWLSALFAVVLVLALAACGGNNNA**/SDG
RTMO02471	ABC transporter substrate-binding protein	**MEDWPMKAHKILFSFIALAFLVLSGCSSLTQNTNSS**/ATK
RTMO00453	ABC transporter substrate-binding protein	**MFGAMKKLYVLALFTVLFGILIGCGKNEA**/SDN
RTMO00456	ABC transporter substrate-binding protein	**MREVGKMVKKAFISILAFILVFSLAGCGKTAGSEA**/KDK
RTMO03923	Aldehyde dehydrogenase (EC 1.2.1.3)	**MSLSSKILLKSLKSFA/**APK
RTMO01667	Alkaline phosphatase (EC 3.1.3.1)	**MKEDAKLKNIGRKIISFALIGSLTAGSFAFA**/ARE
RTMO05800	Alkanesulfonates-binding protein	**MISKIYKAAFIFLIIAAIGVLSSCGRSA**/STA
RTMO02464	Alkanesulfonates-binding protein	**MRSMKRSKHILWIIHFLVFSLLLSSCGKAE**/ETG
RTMO00265	Alkanesulfonates-binding protein	**MEMEFKNAHLGERKLKQFFRISVKPVLAVLIASLALTGCGTDAEK**/ANS
RTMO02765	Alpha-amylase (EC 3.2.1.1)	**MGNRVFALFILPCLLFYAFPVQA**/AEK
RTMO01873	Amide-urea binding protein	**MMKGKIYRIFLVVMTIMMILSACSNSSSG**/NST
RTMO01393	Amine oxidase family protein	**MRKIAFGLCVCFLIFTAYSSQIFPVYAD**/DHE
RTMO04269	Arginine-binding protein	**MLKLFNFLRIERERRCMKKLLSLLVSSILLIGLLSACGAGS**/EEK
RTMO04309	Aspartate aminotransferase (EC 2.6.1.1)	**MKLAKRVASLTPSTTLAITAKAK**/ELK
RTMO02744	ATP-dependent nuclease subunit B (EC 3.1.-.-) / DNA helicase (EC 3.6.4.12)	**MKVMPMSLRFL**/LGR
RTMO01464	Beta-lactamase family protein	**MSNRFVSVVLLSVMLSSAIFFSPPSVLA**/TSH
RTMO01465	Beta-N-acetylhexosaminidase (EC 3.2.1.52)	**MLSFYKKITVILVAVVMLFVPWTSPQA**/HTE
RTMO05116	Carboxy-terminal processing protease precursor (EC 3.4.21.102)	**MNKKTTAMLMVLSMLIGAGGTYA**/GMQ
RTMO00896	Cell wall hydrolase family protein	**MLASVFCGAFFLGSHAYA**/ATT
RTMO01036	Cobalamin-binding protein	**MKKWKRYVMLLVFALVFGLMFGCSGENA**/SKE
RTMO02164	Cobalt transport protein cbiN	**MKRSLLLLVVAVLLTAAPLLFIPHS**/DFG
RTMO04909	ComE operon protein 2	**MTLCRLRHYTKSMC**/TSI
RTMO04657	Cu-containing dissimilatory periplasmic nitrite reductase (EC 1.7.2.1)	**MKRKFYTFMSIVVAALFLTACEHTGGKEA**/EKE
RTMO04721	Cystine-binding protein	**MKRFFHKSLLLLLTASILLLAACGNQQSNE**/KSG

RTMO02091	Cytochrome c oxidase Cu(A) center assembly protein	**MPKQRCWAGSRKNKKRVKRKMKRMIVLLAIVLLAACGKTIPDA**/KNW
RTMO04339	Cytochrome c oxidase polypeptide II (EC 1.9.3.1)	**MHLHKYEKIWLAFGIGCLFVFLTVIGVSAFA**/EGN
RTMO05119	Cytochrome c551	**MQLIGLGGVCFMKWKLASLFIGASLLLAACGGGNDA**/SNN
RTMO01064	D-alanyl-D-alanine serine-type carboxypeptidase (EC 3.4.16.4)	**MAHTLVQKSKRDGDYMKLWKLIVLFIVAVAMLFSCIPDQAKA**/MNE
RTMO01094	D-alanyl-D-alanine serine-type carboxypeptidase (EC 3.4.16.4)	**MKKQLIRLLLFASVFLFTISSYVHA**/EEK
RTMO02580	D-ribose-binding protein	**MNNKYKERKRTMKKLASMWLSFLLVIGVLAGCSLDNG**/ATS
RTMO01968	Fe3+-siderophore binding protein	**MKIIIKNEGGYIMFKSKLSFLITAILTLVIILAGCGKNEKA**/EPK
RTMO03667	Fe3+-siderophore binding protein	**MIIIINIRSRKGDLRSMMKRKWLYFSLIALLILILTACGAKQSSA**/PDK
RTMO02349	Ferric anguibactin-binding protein	**MMLKKRWLPIFVAFFTAILLAACGNEDNAKN**/ASS
RTMO03250	Ferrichrome-binding protein	**MKKLLIPFIVLIVLVMSACGGKTTENKDNSAA**/KEK
RTMO04015	Flagellar biosynthetic protein fliZ	**MLQSRIIALFLCIVVAIAAQTEFPVFA**/EQS
RTMO04185	Gamma-D-glutamyl-meso-diaminopimelate peptidase (EC 3.4.19.11)	**MNLKPRHIVLTSAFASALFWMPDDGKA**/AEW
RTMO02485	Gamma-D-glutamyl-meso-diaminopimelate peptidase (EC 3.4.19.11)	**MKKWKWYLTAFLCFCMVFGFLLPANA**/KTD
RTMO04121	Germination protein germ	**MFNRGARKLAASVAALLLLLSGCGLFG**/KDG
RTMO02375	Glucose-binding protein	**MKKKRLWLSLALVAGLALSGCNSDSAS**/NSN
RTMO05345	Glutamyl endopeptidase precursor (EC 3.4.21.19)	**MKKIGFIILVMVGFIISPIINAPETVNA**/QKN
RTMO01484	Glycerol-3-phosphate-binding protein	**MKKGIFALFLFIFVTLTACSSESNEAA**/ATP
RTMO02444	High-affinity zinc uptake system protein znuA precursor	**MLVKHFDRKVLNMKAKSFILSLLLVISAFLYGCNA**/EKN
RTMO04518	Hydrolase (HAD superfamily)	**MKEMGSVIISALLVLLALVVGAVVGFFVRKSIAEA**/KIG
RTMO04111	Hypothetical cytosolic protein	**MAEKRKFLWLLMALLLCVAFGNVPAVAFG**/ADN
RTMO01466	Hypothetical exported protein	**MRNRWWMVCLAVILGLSLFTGVLA**/KGP
RTMO01132	Hypothetical exported protein	**MAHNLCPFASHYTKEKQRVRVMRWILAAMLVLSSFFSISASAAA**/ETQ
RTMO00612	Hypothetical exported protein	**MKAVMERSERLMKKWRICLCIGVLMMYATTFTADAA**/SRQ
RTMO01939	Hypothetical lipoprotein	**MMKWKGILMTMFAILVLAVAGCSKK**/EVK
RTMO02353	Hypothetical membrane associated protein	**MKGRRRLMMFCFPFLCSVLAAMG**/MTV
RTMO00351	Hypothetical membrane associated protein	**MRGNVFSVFLCAILLIALAGCGAKS**/QED
RTMO00908	Hypothetical protein	**MKKMKKVYAFLALLMPSLFLFAACA**/QEK
RTMO04588	Hypothetical protein	**MKKLLSPFFAFVLLLVAATGCSSEQSSS**/SNK
RTMO04497	Hypothetical protein	**MKIKKTLLLTMTVIVLLAACSTKQD**/AVQ
RTMO01449	Hypothetical protein	**MKGWSKFFICLCLLFAFHLPVQA**/QHV
RTMO02116	Hypothetical protein	**MKLPKWLRKVLVVTITVCTFGLVTPPASLMA**/ADE
RTMO03498	Hypothetical protein	**MKKTLLRLLLLLSVFILAVGCSSK**/IHD
RTMO02631	Hypothetical protein	**MLKRAIYRSLYLCAMFVFLFTLPFHA**/EET
RTMO03728	Hypothetical protein	**MSTPSYIVVVNGGGTRMKRIWLLAFI**/AFI
RTMO00164	Hypothetical protein	**MCYCKYSKTLMR**/CTL
RTMO00163	Hypothetical protein	**MKRKPWKVMTAAALTSSLLLASACTSSG**/KET
RTMO03084	Hypothetical protein	**MKGVFAMRKALLAVTLSAATLAGCAQ**/TAQ

RTMO03242	Hypothetical protein	**MLPLAVVFGLAFSSATITKAEA**/VQT
RTMO00237	Hypothetical protein	**MTKKKWLLLKLFGAFVAIVVATGCNA**/NNN
RTMO00393	Hypothetical protein	**MKRCLIAMSAACLFIGGCMHENKQQA**/PEP
RTMO00492	Hypothetical protein	**MNAVKATIPVLTAATLLLSSATGTYAA**/APD
RTMO03575	Hypothetical secreted protein	**MKKMAKAVMITSAILLLSACSSSNEKKQA**/FIN
RTMO00227	Intracellular proteinase inhibitor	**MMGKGKTLGLAGMIAGAAAVSMLFASNSGEQPK**/AKD
RTMO01730	L-arabinose-binding protein	**MYMRKRLLFLVTIIFAFSMILA**/GCS
RTMO03075	Leucine-, isoleucine-, valine-, threonine-, and alanine-binding protein	**MKKKKLAGAFLSLMVTAGIMAGCGAQK**/DST
RTMO03610	Leucine-, isoleucine-, valine-, threonine-, and alanine-binding protein	**MRRFLSAMISIFCVFILASCGKEPSNA**/SKS
RTMO01970	Lipoprotein	**MKRTAVMAACLLSFGIIMGACS**/DDK
RTMO00399	Lipoprotein (pheromone precursor)	**MRQDFLKKLNKVSRCGTMKKKMILFAALLLFLSSCAPK**/FGE
RTMO05083	LysM domain protein / 3D domain protein	**MKKLLLSITSSFFLAFGFSGAASA**/AGT
RTMO03502	LysM domain protein / NlpC/P60 family protein	**MKKSFILTGTIISSLLADQTAFA**/SSY
RTMO03847	Lytic transglycosylase homolog yjbJ	**MDVSTLKLLLELQALQTFTPARA**/NTV
RTMO00446	Methionine sulfoxide-binding protein	**MGGKRMNKKLSLIVVLLLTFFLAACSSKEGA**/TST
RTMO00875	Methyl-accepting chemotaxis protein	**MFKTLKAKLIALMALLMIVSLMITQIVGV**/VET
RTMO03578	Multicopper oxidase family protein	**MKKLLFGTILAGVVAIGAACSNNASQSSM**/QGH
RTMO01722	Multiple sugar-binding protein chvE	**MKRFLSVLVLLTFVFTLSA**/CSG
RTMO00287	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	**MCLRRMRLLLFLFCLSMVVGMVLPVLA**/AKN
RTMO01717	Nitropropane dioxygenase / Trans-enoyl-CoA reductase family	**MRKVLNTISVPIIQAPMAGGVSTPALA**/AAV
RTMO02136	NlpC/P60 family protein	**MSISLSVPKWLLTVLSILSLVVAFIFGTVSNASAT**/INY
RTMO03798	NlpC/P60 family protein	**MRKYSFLLFFAIAFIFGGKTVDA**/HVV
RTMO00581	NlpC/P60 family protein	**MKQFVTLVSLSFLVVFSSLFAHTSSAEA**/AVN
RTMO03974	Nucleoside-binding protein	**MKKRFGFALSLVLTAGMLLSACGGQGGDNA**/GGK
RTMO04406	Oligoendopeptidase F (EC 3.4.24.-)	**MMKKQLYVWLMIVLLLVPWHASAE**/QTK
RTMO03828	Oligopeptide-binding protein oppA	**MKKRSFMLLSFMLALSLFLSACGGFQKGNESA**/GEK
RTMO01709	Oligopeptide-binding protein oppA	**MKKTFASIFALLLLVSAVLTGCGSKG**/TSG
RTMO00439	Oligopeptide-binding protein oppA	**MALMIKSRKKKKTNFMKGLWLSISLVLLLTACDSQK**/ETA
RTMO04258	Peptidase, M16 family	**MCTMSRKRERKKRKSF**/RCM
RTMO04175	Peptidoglycan anchor protein	**MKRFCIAIITCFFFATAHGAAPAFA**/QVD
RTMO03290	Peptidoglycan endo-beta-N-acetylglucosaminidase (EC 3.2.1.96) / N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	**MTTFLWRCFEVRIGVQIRKFAALLSVLILLVSYAISPAYA**/ANA
RTMO03082	Peptidoglycan-specific endopeptidase, M23 family	**MHPFIIAIVTTAVIFLSPKPIFA**/QEK
RTMO02183	Peptidoglycan-specific endopeptidase, M23 family	**MMKRRKVMALAAATVLSIGVFPHFADA**/VSD
RTMO00094	Phosphate-binding protein	**MWKKSIKFGVAALLITGMLAGCGKS**/DNN
RTMO02511	Polysaccharide deacetylase	**MKSVLFAFLLFIPFFSFFNHHVKA**/AEL

RTMO01761	PTS system, cellobiose-specific IIB component (EC 2.7.1.69)	**MKRILLACSSGMSTSLLVAKM**/QEY
RTMO04318	SCP domain protein family	**MNKKIVFSLAASLAIVGASFTAKA**/AEA
RTMO00768	S-layer protein	**MRKFYSFILVFSLLVSIVFPGVVTEAKS**/KFK
RTMO02216	S-layer protein	**MGYIIKPRTGGYSMKRTFLHIALSLLAAMLALPAMNASA**/ATR
RTMO01808	S-layer protein	**MKQHKGIGGDNMFKHFKIWVGVLMAAFICVSVMHPHKAKA**/EEK
RTMO01801	S-layer protein	**MAYQPKSYRKFLAGSVSAALVATAVGPVVANA**/ASF
RTMO01308	Spermidine/putrescine-binding protein	**MGIMKRWMMTGFLMLIMALAGCGVPDA**/KPP
RTMO00500	Spermidine/putrescine-binding protein	**MRKLISLFAAVFFA**/SFV
RTMO00616	Spore coat N-acetylmuramic acid deacetylase (3.5.1.-)	**MKNATIWLFFSAVILSFIPVSAEA**/ASN
RTMO00221	Stage II sporulation protein D	**MKRMKPLIALF**/SFL
RTMO03947	Sugar-binding protein	**MKRWLTAVGITSVLMGSILAGCGGGDEKA**/ANK
RTMO01490	Taurine-binding protein	**MKKMVFKKSEINAILIILLLLIFSVITGCSSPKTSTAK**/NGE
RTMO03803	Thermitase (EC 3.4.21.66)	**MKKWKKTAVSLGLASALVLPSFAQA**/STM
RTMO03149	Thermonuclease (EC 3.1.31.1)	**MPHISRHSLKEDGIMKKFVSALAIIVSTAIFPGNSFA**/HPG
RTMO04351	Thiol:disulfide interchange protein tlpA	**MKKFIAVILLLAITGYGIWNALA**/AEK
RTMO05278	Thioredoxin	**MKKLLIFGSIIVALFAALAFVTSYQQKEA**/VKN
RTMO01492	Transcription antiterminator, BglG family / PTS system, mannitol (Cryptic)-specific IIA component (EC 2.7.1.69)	**MRLIKIVRKMVCKSVSECRFL**/CPS
RTMO00278	Trehalose/maltose-binding protein	**MCIKRGRKNMKKKGFTKLIAALLVVALIGTGCQGQNEGKNA**/KGD
RTMO01319	unassigned	**MGKCMKKFLSALLLFSFIISFWSIGNLTFA**/AST
RTMO05618	unassigned	**MKRILTAWMLFPLISACS**/AET
RTMO01192	unassigned	**MKKILLASAAVSLMFLAGCQNDQP**/EVK
RTMO05744	unassigned	**MTFSKKKKNLLIILASLVLSIITITSAYA**/AVL
RTMO03883	unassigned	**MSKKHRPFLPVKINKVCAKKAVSAYLFVYVASRNVSA**/AMH
RTMO00878	unassigned	**MMVKQRTTRRLCFVIALKA**/ADR
RTMO00826	unassigned	**MIPMNKTKSYLSFLLSFVLVLSTLGGAGIAQA**/QAE
RTMO04621	unassigned	**MMSLPKQIMILFLFLFAIFGAWTPKA**/KVF
RTMO05808	unassigned	**MTSNFFSSITPFIFIKSFSASLPPPSGA**/SSL
RTMO05856	unassigned	**MKKKKFAVLGLAVGLMAFGGAVQA**/GTS
RTMO05861	unassigned	**MLVYLPSKLLALLSPTVNKA**/KPF
RTMO03412	unassigned	**MNKTKKMMVGVLSTLMAASLAACS**/DES
RTMO04373	unassigned	**MAFTCLARKRSG**/KRC
RTMO05903	unassigned	**MSWIHVQSKTMHKLLRKVMTLAGVLVLAITAFSLVNPNQAAA**/WLH
RTMO00133	unassigned	**MLTKTNNQLRLSMFVIISSLVLVFSTLLAPLKSEA**/VTS
RTMO03032	unassigned	**MKCLPLYLGKMLSNAAA**/SSQ
RTMO04985	unassigned	**MRRFFLFFSLALIFILNSTPISVFA**/YSY
RTMO05003	unassigned	**MMCTKHIRRMAV**/LCC
RTMO05216	unassigned	**MLAGSFLIASATSALRASACLA**/ANV
RTMO03262	unassigned	**MFRHLLLFSLNGILGSANA**/AEN
RTMO04831	unassigned	**MEYRLRKKCKKQ**/KTS

RTMO00308	unassigned	**MKRVVMLIMGIVFYFVSGCSIVNE**/NNN
RTMO00312	unassigned	**MNEEEMVLKKILSGVLGLSLLLGGTNFAFA**/KDG
RTMO00325	unassigned	**MKKKFAVLTLAAGLLAAGGLAQA**/GTM
RTMO00330	unassigned	**MMKLNKNLKTIAMSLGIGLTLLAGANVYA**/ATQ
RTMO00403	unassigned	**MFKKGYLSILSLVMGFTFFSTNTFAA**/TDI
RTMO00404	unassigned	**MSALKEKFLIAGVASVVLAVSLVVYNGTDIAGN**/QDN
RTMO00584	unassigned	**MKTRWLFLAAALMLMLPTGTLAA**/QRA
RTMO01740	Xylooligosaccharide-binding protein	**MHCSKNKGGLDLLKKAHSLLCIMIIIFALVLTGCSGTA**/NEG