# A *CORYNEBACTERIUM GLUTAMICUM* EXPRESSION SYSTEM FOR GLYCOSIDE HYDROLASE ANALYSIS

by

Hirak Saxena

B.Sc., Ryerson University, 2013

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in

Molecular Science

Toronto, Ontario, Canada, 2016

**AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

**A *Corynebacterium glutamicum* expression system for glycoside hydrolase analysis**

Hirak Saxena, 2016

Master of Science, Molecular Science

Ryerson University

## Abstract

The biological hydrolysis of glycosidic linkages in complex sugars is facilitated by glycoside hydrolases. These enzymes are ubiquitous across all domains of life, playing significant roles in important biological processes like the degradation of cellulosic biomass, viral pathogenesis, antibacterial defense, and normal cellular functions. The potential industrial applications of highly efficient glycoside hydrolases, as well as the fact that a number of lysosomal storage diseases have been attributed to deficiencies in these enzymes [43, 22], merits further study into their structure and activity. For this reason, a handful of novel glycoside hydrolases from *Cellulomonas fimi*, a Gram-positive *Actinobacteria* known for its ability to degrade cellulose [39], will be cloned, expressed and biochemically analyzed.

**Acknowledgements**

I would like to extend my deepest and sincerest gratitude to my supervisor and mentor Dr. Warren W. Wakarchuk. I am indebted to him for all of the opportunities he has afforded me in such a short time and for the chance to grow not only as a scientist, but also as a person. His unfaltering patience and passion for science continue to motivate me and those who are lucky enough to work with him.

I would like to thank my committee members, Dr. Anthony J. Clarke and Dr. Jeffery Fillingham for their valuable opinions and guidance throughout the duration of my graduate degree.

I would like to thank my lab mates, new and old, for sharing their knowledge, skills, and comradery during our time together.

I would also like to thank the following individuals for their valuable contributions: Dr. Hugo Menzella and Pablo Ravasi for their generous gift of the synthetic shuttle vector, Dr. Harry Brumer and Nicholas McGregor for the various endoglucanase substrates, and Dr. Lyann Sim for her insights and assistance.

Also, I would like to extend a heartfelt thank you to my family for their support and encouragement, in addition to instilling in me the value of pursuing my passions.

Finally, I would like to thank my friends. If not for them, I would have easily completed my thesis six months in advance.

**Table of Contents**

**List of Figures**

## List of Tables

# Chapter 1

# Introduction

## 1.1 Glycoside hydrolases

Glycoside hydrolases (GHs) are a broad group of enzymes that catalyze the hydrolysis of *O*-glycosidic bonds between two carbohydrate residues or between a carbohydrate and non-carbohydrate moiety [42] (Figure 1.1). These enzymes are ubiquitous in nature, playing significant roles in the degradation of cellulosic biomass for the production of biofuels and various cellular functions [30]. They are primarily utilized in nutrient acquisition among prokaryotes, being found as both intracellular and extracellular enzymes [7]. Of significant historical importance is the bacterial GH β-galactosidase which, in concert with its constitutive genes, has functioned as an invaluable tool for molecular biologists. Among higher organisms GHs can be found within the endoplasmic reticulum and Golgi apparatus where they are involved in the processing of *N*-linked glycoproteins [55], and in the lysosome as enzymes involved in the degradation of carbohydrate structures [18]. Deficiencies of specific lysosomal GHs have been attributed to the development of a range of lysosomal storage (e.g., Tay-Sachs, Hunter, Fabry, Gaucher, and Krabbs) diseases resulting in developmental issues or death [3]. GHs are even involved in processes like viral pathogenesis (eg., viral neuraminidase). Therefore, novel and engineered GHs both have significant industrial and therapeutic potential [65].

### *1.1.1 Originating from soil bacteria*

Cellulose is the most abundant organic polymer on Earth, being composed of β-1, 4-linked glucose monomers (in the form of the repeating disaccharide cellobiose) [32]. The degradation of cellulose by soil microorganisms is a vital step in carbon cycling, releasing $CO_2$ aerobically and various fermentation products anaerobically [65, 5]. Though fungal GHs are preferred for industrial applications, there is still a strong interest in the search for novel bacterial GHs. The diversity of cellulolytic enzymes expressed by

various species of soil bacteria makes them an enticing reservoir of potentially novel GHs [64]. Enhanced thermostability is one of the most desirable properties for a novel GH to possess, since large scale saccharification reactions are commonly carried out at elevated temperatures [56]. Additionally, the synergistic actions of various GHs, which can originate from a wide range of microorganisms, are required for the production of fermentable sugars from cellulosic substrates [5, 56]. For these reasons, novel GHs originating from soil bacteria could be utilized synergistically alongside existing enzyme cocktails to facilitate saccharification of cellulosic substrates.

*1.1.2 CAZy database annotation*

The CAZy (**C**arbohydrate **A**ctive en**zy**mes) database classifies known CAZymes into various families defined by amino acid sequence similarity and fold, with at least one biochemically characterized founding member [8, 26]. Currently, there are only five enzyme classes annotated by CAZy: glycoside hydrolases (GHs) (EC 3.2.1.-), glycosyltransferases (GTs) (2.4.x.y), polysaccharide lyases (PLs) (EC 4.2.2.-), carbohydrate esterases (CEs) (EC 3.1.1.-), and auxiliary activities (AAs) (EC 1.1.3.- and 1.11.1.-). The CAZy database also annotates associated carbohydrate binding modules (CBMs). Currently, there are 135 GH families annotated by CAZy. GH families are systematically updated resulting in the modification, addition, or deletion of entire families. By nature of their classification, members within families will have similar mechanisms, folds, and catalytic residues [37]. GH families can be grouped together in clans where members contain significant tertiary structure similarity, which is more conserved than sequence similarity [23]. The CAZy database allows users to make educated predictions about the fold, mechanism, and substrate of novel CAZymes with only the amino acid sequence.

*1.1.2.1 Mechanism*

Most GHs hydrolyze glycosidic bonds through the use of two catalytic amino acid residues: a general acid (proton donor) and a nucleophile/base [17]. Depending on the positioning of these catalytic residues, hydrolysis occurs via overall retention or overall inversion in configuration around the anomeric carbon (Figure 1.2). The CAZy database indicates both the stereochemical outcome of the reaction as well as the amino acid residues acting as a proton donor and nucleophile (when known) [58]. Though not relevant to this work, a completely unrelated mechanism has been recently demonstrated for two families of glycoside hydrolases utilizing NAD+ as a cofactor [48, 36].

*1.1.2.2 Classes of glycoside hydrolases*

GHs can be classified according to a number of unique schemes in addition to their sequence based annotation in the CAZy database. The utility of these classification schemes depends on the context in which the classification is both made and used. Among these schemes are the Enzyme Commission (EC) number, a mechanistic classification, and the exo/endo classification (Figure 1.3). Exo- refers to the ability of a GH to cleave a substrate at the non-reducing end (most frequently, but sometimes the reducing end as well), while endo- refers to the ability of a GH to cleave a substrate within the chain generating new chain termini [17]. The new chain termini are then accessible for further cleavage by exo-acting GHs.

*1.1.2.2.1 Glycoside hydrolase* family 5

GH family 5 (GH5) is one of the largest CAZy GH families. Previously known as "Cellulase Family A" [25, 20], this group is comprised of enzymes with a generally robust range of known activities including endo-β-1,4-glucanase/cellulase (EC 3.2.1.4), endo-β-1,4-xylanase (EC 3.2.1.8), β-glucosidase (EC 3.2.1.21), β-mannosidase (EC 3.2.1.25), β-glucosylceramidase (EC 3.2.1.45), and glucan β-1,3-glucosidase (EC

3.2.1.58). There are no human enzymes in GH5, though they are widely distributed across archaea, bacteria, and eukaryotes, most notably fungi and plants [37]. GH5 enzymes are retaining enzymes and follow a classical Koshland double-displacement mechanism, utilizing glutamate residues found at the C-terminal ends of β-strands 4 and 7 acting as proton donor and nucleophile, respectively (Table 1.1) [24, 28].

1.1.2.2.2 Glycoside hydrolase family 6

GH family 6 (GH6) was one of the first glycoside hydrolase families classified by hydrophobic cluster analysis, and was previously known as "Cellulase Family B" [25, 20]. The only reported activities for GH6 are endoglucanase (EC 3.2.1.4), and cellobiohydrolase (EC 3.2.1.91) [37]. GH6 enzymes are inverting enzymes with strong evidence supporting an aspartic acid (Table 1.1) playing the role of the proton donor within the enclosed tunnel of the active site [16]. The identification of the nucleophile is far less simple, though there is mounting evidence that it is also an aspartic acid residue located within the active site tunnel [51].

*1.1.3 A brief history of glycoside hydrolases from Cellulomonas*

*Cellulomonas* spp. are Gram-positive coryneform rod-shaped soil bacteria within the *Actinobacteria* phylum that are best known for their ability to degrade complex polysaccharides [1, 10]. *Cellulomonas fimi* ATCC 484 was extensively investigated in the 1980's resulting in the identification of a handful of novel GHs using activity based screening for recombinant enzymes [19]. However, the recent sequencing of the *C. fimi* genome as well as the analysis of the total CAZome (the collection of CAZymes encoded by the genome of an organism) [37] revealed a large number of putative GH genes, many of which have yet to be biochemically characterized [19]. This provided an impetus to revisit this organism in order to identify and characterize these previously overlooked GHs.

*1.1.3.1 Endoglucanases Celf_3184 and Celf_1925*

The first two secreted GH6s isolated and characterized from *C. fimi* ATCC 484, Celf_3184 (CfCel6A) and Celf_1925 (CfCel6B), exemplify how numerous GHs were overlooked due to technical limitations. The identification of these enzymes resulted from the activity based cloning and screening of recombinant libraries as well as purifying enzymes from the host organism [70, 44]. With improved technology giving better sensitivity, a direct proteomic analysis of the secretome of *C. fimi* and *C. flavigena* [66] revealed a number of GHs not detected before in addition to the previously characterized enzymes (Figure 1.4). Many of the newly identified GHs were hypothesized to be novel cellulases or xylanases and a handful of them were identified as being unique to *C. fimi*. Of interest to this study (Figure 1.5) are the GH5, Celf_2403 (which was recently removed from GH5 and currently remains unclassified [37], but for the purposes of this work will be treated as a GH5) and the GH6s, Celf_1230 (CfCel6C), and Celf_0233 (CfCel6D).

*1.1.4 Potential applications of novel glycoside hydrolase family 6s*

The identification and characterization of novel GH6s could serve to improve the efficiency of the saccharification of cellulosic biomass for the production of biofuels. Novel, thermostable GH6s could work synergistically alongside current enzyme cocktails to release more fermentable sugars from various cellulosic substrates. Moreover, understanding the structures and/or amino acid residues responsible for conferring unique traits to particular GHs of interest would greatly facilitate the engineering of these enzymes to improve their efficiency.

**1.2 Problematic expression in *Escherichia coli***

*Escherichia coli* expression systems represent the gold standard in heterologous protein expression, as they offer a cost effective approach to producing heterologous proteins on a large scale. Unfortunately,

the expression of genes originating from *C. fimi* has lead to many complications like irregular folding and aggregation into inclusion bodies (Figure 1.6), resulting in non-functional proteins that are difficult to harvest [33]. The possibility that the difference in GC% content between *E. coli* and *C. fimi* (50.8% and 74.7%, respectively) [46, 11] genes being a contributing factor to expression issues cannot be overlooked, however it is much more probable that there are minor to significant differences in the cellular machinery utilized by these two organisms during the translation, folding, or secretion of proteins. Also, true secretion from the cell is difficult in *E. coli*. The *E. coli* expression issues of secreted, native *C. fimi* GHs could also be linked to the fact that the enzymes don't fold correctly when not destined for secretion. For these reasons, this work seeks to appropriate an alternative expression host in tandem with secretion linked expression.

**1.3 Twin-arginine-translocon dependent secretion**

Protein export from the cytoplasm of a cell to other organelles or out of the cell is a common phenomenon that occurs in both prokaryotes and eukaryotes [54, 49, 47, 67]. In the general secretory (Sec) pathway, pre-protein is transported from the cytoplasm to the extracellular space in an unfolded form and the signal sequence is cut off by a type I signal peptidase (SPase) [14, 15, 60, 45]. The twin-arginine-translocon (TAT) pathway utilizes signal sequences containing two conserved arginine residues "-RR-" in the region near their N-terminus in addition to a type I SPase. TAT translocation is Sec-independent and employs a cytoplasmic chaperone to secrete proteins in a fully folded form (Figure 1.7) [4, 13, 41, 71, 67].

*1.3.1 TorA signal peptide*

The TAT signal peptide of the molybdopterin-containing protein trimethylamine N-oxide (TMAO) reductase (TorA), a protein known to be exported via the TAT pathway in *E. coli*, has been utilized to localize heterologous proteins to the periplasm in *E. coli* as well as for secretion in *Corynebacterium*

*glutamicum* [52, 53]. It has been demonstrated that the fusion of the TorA signal peptide to green fluorescent protein (GFP) simultaneously improves the cytoplasmic folding of heterologous GFP and localizes it to the periplasmic space when expressed in *E. coli* (Figure 1.8) [59]. In *C. glutamicum*, the use of a TorA-GFP fusion resulted in the secretion of active-form GFP into the culture supernatant at quantities approaching 20 mg/L [57].

## 1.4 *Corynebacterium glutamicum*

*C. glutamicum* are Gram positive non-pathogenic, non-sporulating, non-motile rod shaped bacteria that have been widely utilized in industrial applications such as the production of amino acids, nucleotides and enzymes [41, 31, 61]. *C. glutamicum* and *C. fimi* are both members of the *Actinomycetales* Figures 1.9 and 1.10) order [29] and both have genomes that are high in GC% content [69]. Since *C. glutamicum* are widely utilized, fairly well-studied, and bear a genetic resemblance to *C. fimi*, this organism serves as an ideal candidate for a Gram positive protein expression/secretion system.

### *1.4.1 Use as an expression host in tandem with TAT secretion*

The utilization of *C. glutamicum* and TAT dependent secretion is expected to facilitate both the expression and purification of novel GHs originating from *C. fimi*. The employment of an expression host more closely related to *C. fimi* should be able to overcome the aforementioned issues that arise during expression in *E. coli*. Moreover, the secretion of heterologous proteins into the culture supernatant completely eliminates the need for cell lysis during protein purification.

## 1.5 Overall goal

The objective of this project is to generate a stable Gram-positive expression system based in *C. glutamicum* for the expression and subsequent analysis of various novel GHs originated from the Gram-positive soil bacterium, *C. fimi.*

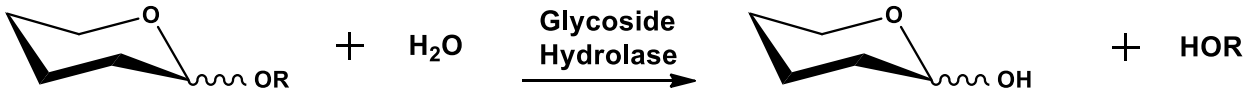



**Figure 1.1: Hydrolysis of carbohydrate polymers by glycoside hydrolases.** Glycoside hydrolases catalyze the hydrolysis of *O*-glycosidic bonds between two carbohydrate residues or between a carbohydrate and non-carbohydrate moiety.

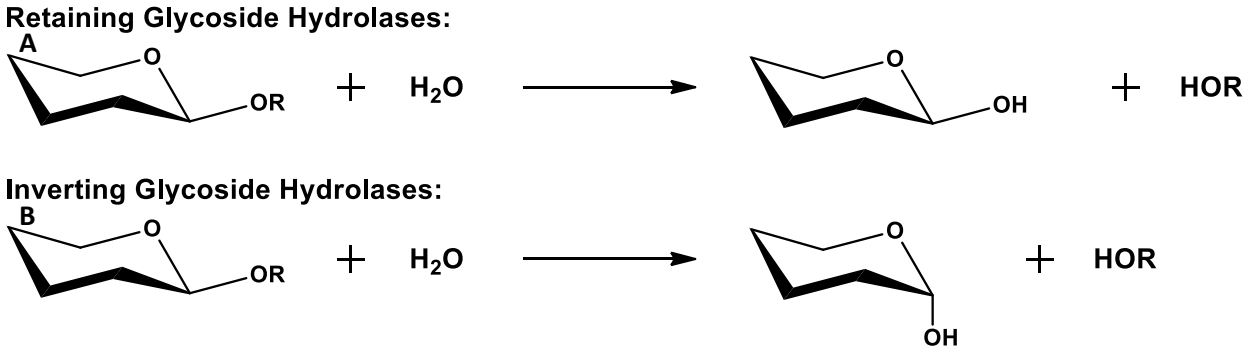



**Figure 1.2: Mechanisms of glycoside hydrolases.** Retaining glycoside hydrolases **(A)** hydrolyze glycosidic linkages via a retaining mechanism that retains the stereochemical orientation around the anomeric carbon, while inverting glycoside hydrolases **(B)** invert the stereochemical orientation around the anomeric carbon.

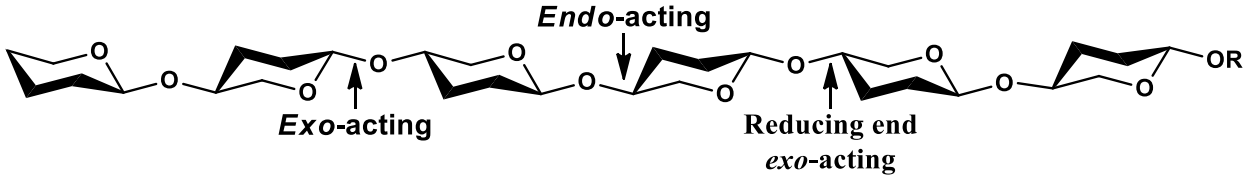



**Figure 1.3: Exo-/endo- classification of glycoside hydrolases.** Exo-acting glycoside hydrolases cleave glycosidic linkages from either the reducing or non-reducing end of a carbohydrate chain. Endo-acting glycoside hydrolases cleave internal sites, generating new chain termini.

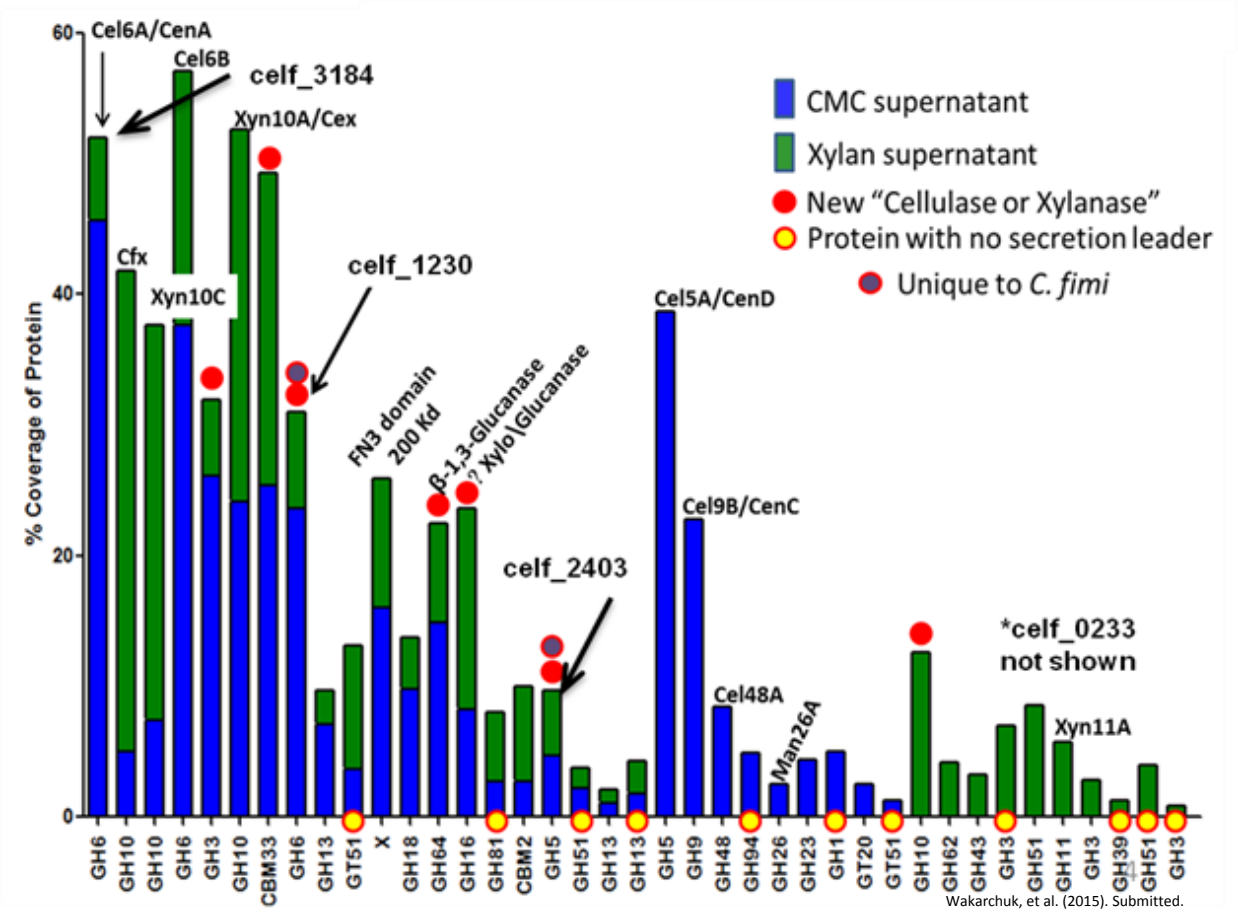



**Figure 1.4: *C. fimi* secretome analysis for cellulases or xylanases by mass-spectrometry.** Proteins with a MASCOT score of less than 50 are not shown. The % coverage of protein is a correlation of the relative amount of protein. Proteins that are unique to *C. fimi* were qualitatively determined by their amino acid sequence. Proteins identified in both supernatant samples are indicated by stacked bars.

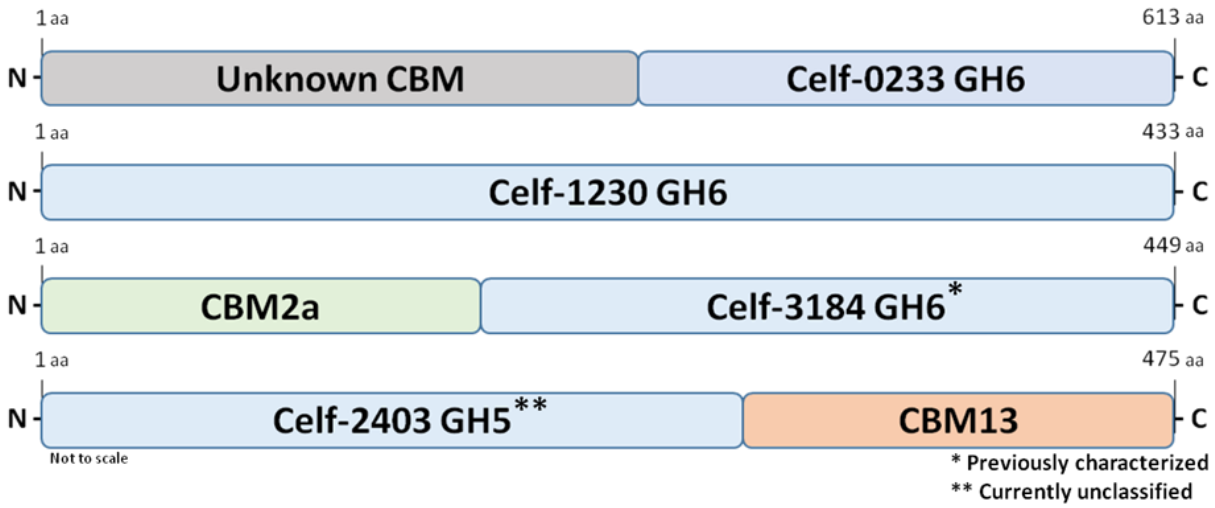



**Figure 1.5: Putative native glycoside hydrolase domain architecture as determined by BLAST.** The GH5, Celf_2403 contains a CBM13. The GH6s, Celf_0233, Celf_1230, and Celf_3184 contain an unknown CBM, no identifiable CBM, and a CBM2a, respectively.

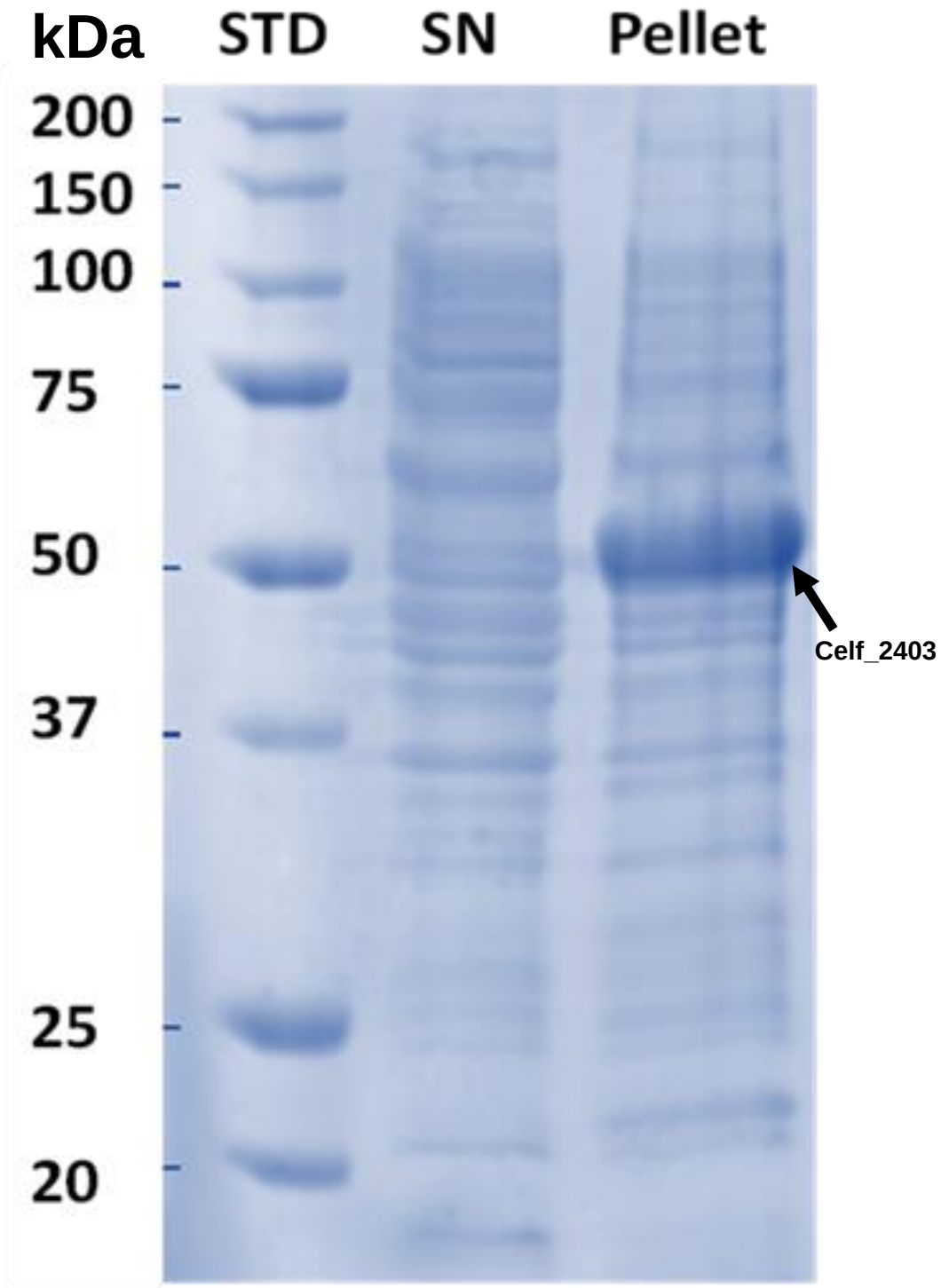



**Figure 1.6: Expression of native *C. fimi* glycoside hydrolase Celf_2403 in *E. coli*.** The native Celf_2403 protein is not expressed in a soluble form in *E. coli*. Any expressed protein forms insoluble inclusion bodies with no discernible activity.

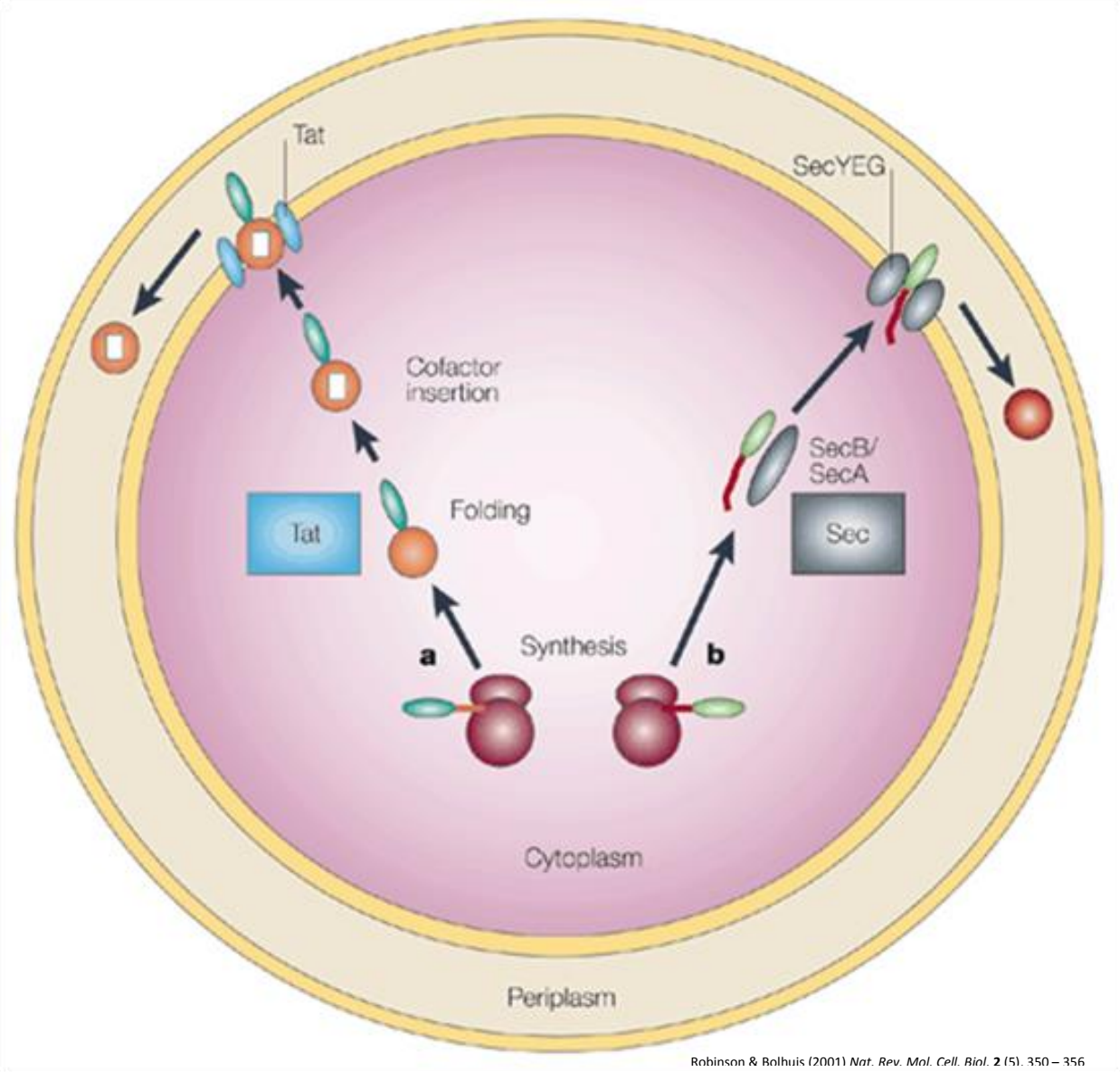
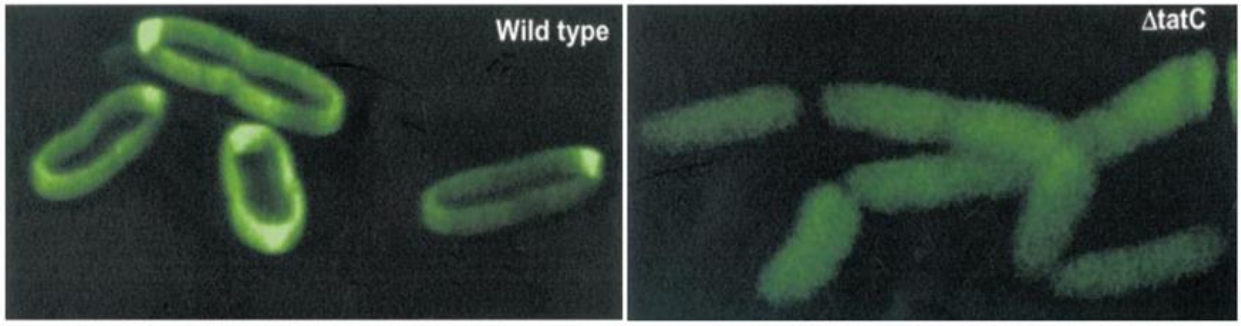
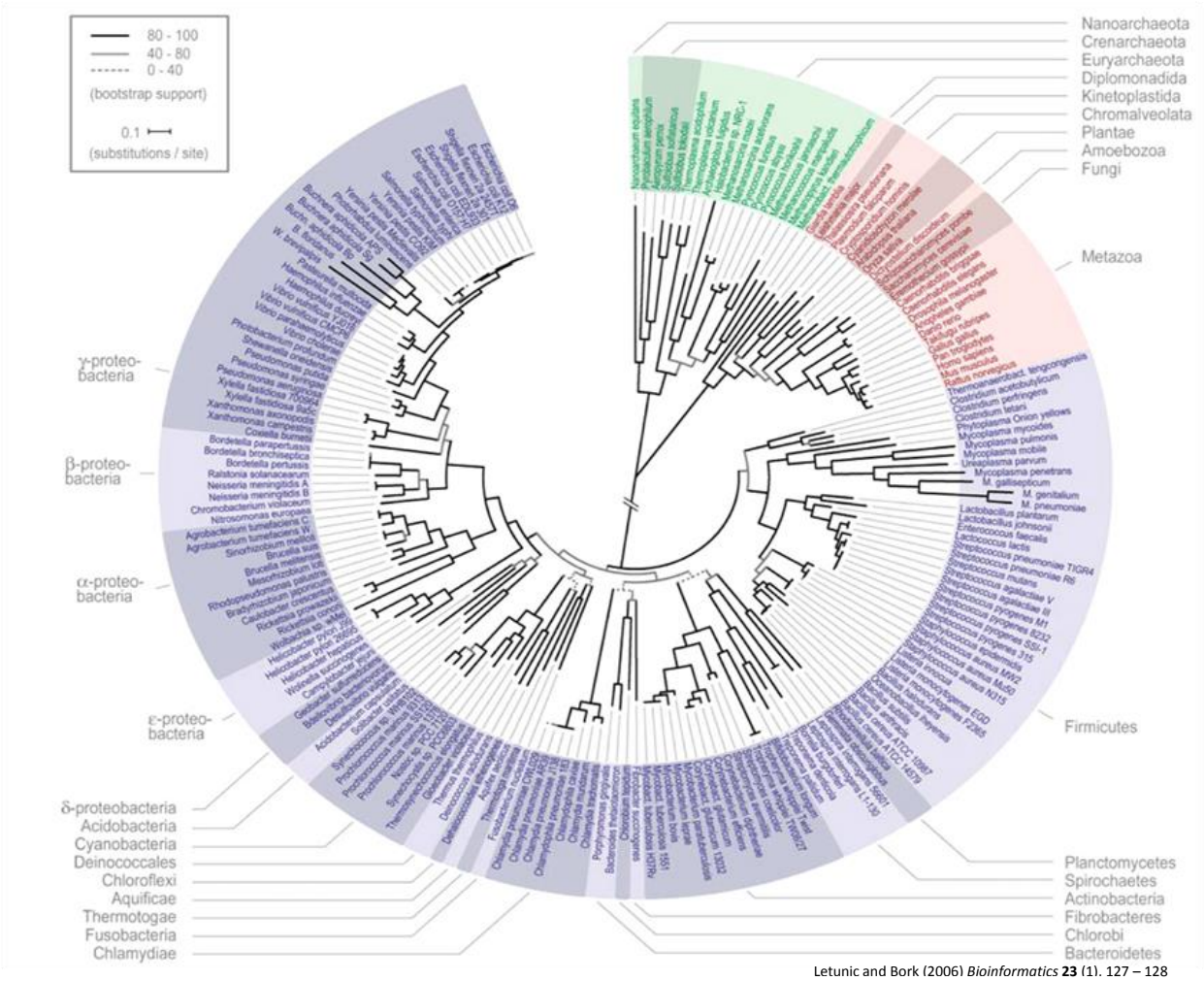



**Figure 1.7: Periplasmic localization via Sec-dependent and Tat-dependent transport.** The Tat pathway **(A)** allows pre-proteins to bind redox cofactors (white square) following synthesis in the cytoplasm. Pre-proteins fold cytoplasmically before being transferred through the Tat apparatus and undergoing processing. The Sec pathway **(B)** associates pre-proteins with the SecB protein (which prevents cytoplasmic folding of the mature domain) before being transferred to SecA. This drives ATP-dependent translocation through the SecYEG complex, threading the pre-protein into the periplasm where it is processed and allowed to refold.



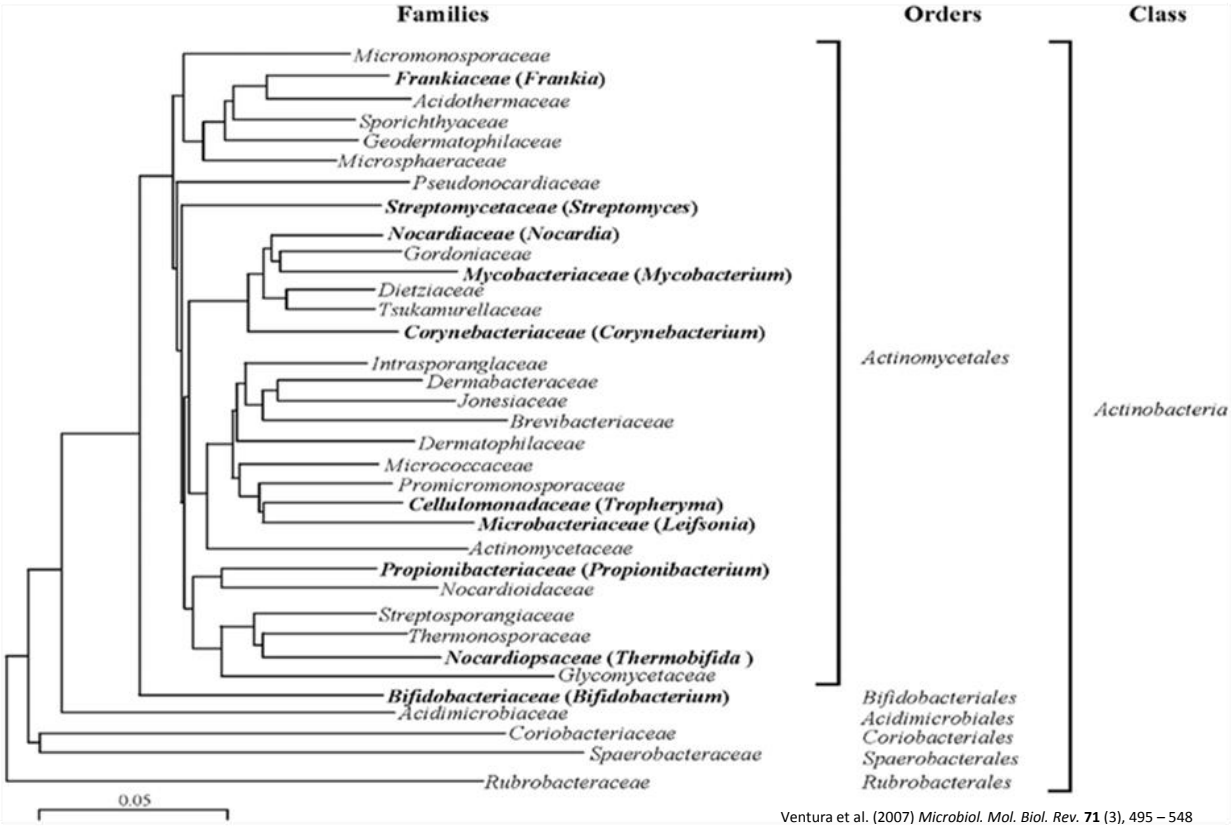

Thomas, et al. (2001) *Molecular Microbiology* **39** (1), 47 – 53

**Figure 1.8: Periplasmic localization of GFP in *E. coli* with TorA signal peptide.** In cells with functional Tat machinery (left panel), TorA localizes GFP to the periplasm. In cells lacking a functional tatC gene (right panel) all localization is lost and mature GFP remains in the cytoplasm.




Letunic and Bork (2006) *Bioinformatics* **23** (1). 127 – 128

**Figure 1.9: Genetic relationship of *E. coli* and *C. fimi*.** A large evolutionary gap exists between *C. fimi* (*Actinobacteria*) and *E. coli* (*γ-proteobacteria*). This distance likely implies a significant difference in the cellular machinery utilized by these two organisms during the translation, folding, or secretion of proteins.




Ventura et al. (2007) *Microbiol. Mol. Biol. Rev.* **71** (3), 495 – 548

**Figure 1.10: Genetic relationship of *C. glutamicum* and *C. fimi*.** *C. glutamicum* and *C. fimi* share both their class (*Actinobacteria*) as well as their order (*Actinomycetales*). This close relationship implies close homology in the cellular machinery utilized during the translation, folding, or secretion of proteins.

**Table 1.1: Characteristics of GH5 and GH6.**

| Family | Mechanism | Proton Donor | Nucleophile or Base | Select Known Activities |
|---|---|---|---|---|
| **GH5** | Retaining | Glutamic acid (experimental) | Glutamic acid (experimental) | endo-β-1,4-glucanase/cellulase (EC 3.2.1.4), endo-β-1,4-xylanase (EC 3.2.1.8), β-glucosidase (EC 3.2.1.21), β-mannosidase (EC 3.2.1.25) |
| **GH6** | Inverting | Aspartic acid (experimental) | Aspartic acid (experimental) | endoglucanase (EC 3.2.1.4), cellobiohydrolase (EC 3.2.1.91) |

# Chapter 2

# Materials and Methods

## 2.1 Strains

Genotypes of bacterial strains can be found in Table 2.7. *C. fimi* ATCC 484 was the source for genomic DNA used during cloning. *C. glutamicum* ATCC 13032 was utilized as an expression host. Both of these strains are type strains for their respective species. ElectroMAX™ DH10B™ (ThermoFisher Scientific) cells were used for routine cloning and the generation of the construct library used in this work. ElectroMAX™ DH10B™ cells are a derivative of the *E. coli* K-12 strain MG1655 optimized for improved transformation efficiency. This strain also contains engineered versions of Endonuclease I (endA1) and RecA (recA1) that are inactive. These modifications help to improve plasmid yield and stability within cells [9]. Shuffle Express™ (New England Biolabs) cells were used for protein expression requiring the formation of disulfide bonds. Shuffle Express™ cells are an *E. coli* BL21 derivative that constitutively expresses cytoplasmic disulfide bond isomerase DsbC, which also acts as a chaperone for protein folding. Additionally, Shuffle Express™ cells contain deletions for gultaredoxin reductase (*Δgor*) and thioredoxin reductase (Δ*trxB*) to prevent the reduction of cysteine residues in the cytoplasm [6]. Strains such as DH5α, BL21, and AD202 were occasionally used for troubleshooting.

## 2.2 Shuttle vector Design

### *2.2.1 Synthetic gene design and synthesis*

Synthetic genes (Table 1) containing the TorA TAT signal sequence were synthesized and supplied by Biobasic Canada (Markham, Ontario). A mutant TorA TAT signal sequence was designed to include a 3' $His_6$ tag which would be present after processing the leader and then used for subsequent protein purification (Figure 2.5). Synthetic genes were designed with a 5' AseI site (VspI/degenerate NdeI) and a 3' NdeI site. Codon usage was optimized to normalize regional GC content.

*2.2.2 Cloning*

Synthetic genes were excised from their parent pUC57 vector in a two-step digestion. First, the pUC57 MCS was excised via digestion with EcoRI and HindIII. The multiple cloning site (MCS) fragment containing the synthetic gene was extracted from the gel and purified using the GeneClean® Turbo Kit (MP Biomedicals) following manufacturer specifications. The purified MCS fragment was further digested with AseI and NdeI to liberate the synthetic gene, which was again purified using the GeneClean® Turbo Kit (MP Biomedicals). Vector DNA (pTGR5) was linearized by digestion with NdeI and gel purified in a similar manner to the synthetic genes. Digestions were allowed to incubate at 37°C for 10 – 30 minutes before heat inactivation at 65° – 85°C. Enzyme concentration in reactions was maintained at 1 µL enzyme (1 U) per 1 µg plasmid DNA and 1/10 total reaction volume according to manufacturer specifications (ThermoFisher Scientific) to minimize star activity. Ligation reactions were prepared with varying molar ratios of insert to vector (3:1 – 10:1) and 10 pmol vector. T4 DNA Ligase was purchased from New England Biolabs. Transformations were carried out via electroporation into the *E. coli* strain DH10B (New England Biolabs) using a 0.2 cm gap electroporation cuvette and a MicroPulser (Bio-Rad) set to deliver a single 2.5 kV pulse. Outgrowth was performed by shaking the cells at 37°C for 1 hour in recovery media (SOC :2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM $MgSO_4$, 20 mM D-glucose)prior to plating onto selective media containing 50 µg/mL kanamycin (BioShop) and overnight incubation at 37°C.

*2.2.3 Verification of TorA modification*

Individual colonies were picked and used to inoculate 2 mL LB (1.0% w/v tryptone 0.5% w/v yeast extract, 1.0% w/v NaCl) containing 50 µg/mL kanamycin (BioShop) and grown overnight with shaking at 37°C. Plasmids were then prepared from cultures following the standard High-Speed Plasmid Mini Prep Kit (FroggaBio) protocol. Clones were screened for inserts with a diagnostic double digestion using

BamHI and NdeI (ThermoFisher Scientific). Clones with the correct insert were further screened for correct orientation with a second diagnostic double digestion using NdeI and EcoRI (ThermoFisher Scientific). Digestions were analysed by agarose gel electrophoresis to confirm their orientation in the vector. Purified plasmids containing correctly oriented inserts were prepared with specific sequencing primers (Table 2.6) and sequenced by The Centre for Applied Genomics (TCAG, Toronto). The shuttle vectors that were generated can be found in Table 2.3 and vector maps can be found in Figures 2.1-2.3.

## 2.3 Generation of glycoside hydrolase construct library

### *2.3.1 Genomic DNA extraction*

*C. fimi* gDNA was harvested from saturated 40 hour cultures grown in low salt LB (1.0% w/v tryptone 0.5% w/v yeast extract, 0.5% w/v NaCl) using the DNeasy Blood & Tissue kit (Qiagen).

### *2.3.2 Primer design and PCR*

Primers (Table 2.2) were designed to include a 5' NdeI site and a 3' EcoRI site and were optimized for GC content, length, and to minimize undesired secondary structures. Primers were synthesized by The Centre for Applied Genomics (TCAG, Toronto). Since primer melting temperatures could not be optimized due to the intrinsically high GC content of the *Cellulomonas* genome, various annealing temperatures were used during PCR amplification (Table 2.2). Cloning was conducted using Q5 High-Fidelity DNA Polymerase and High GC Enhancer buffer (NEB) in the presence of DMSO to aid in the denaturation of high GC content templates.

### *2.3.3 Verification of cloning and sequencing*

Vectors and inserts were digested using NdeI and EcoRI and subsequently purified as described in Section 2.2.2. Clones were prepared as described in Section 2.2.3 and screened for inserts with a

diagnostic double digestion using NdeI and EcoRI. Possible multiple insertion events were ruled out with single digestions using only EcoRI. Digestions were subjected to agarose gel electrophoresis for confirmation. Purified plasmids containing correctly sized inserts were prepared with specific sequencing primers (Table 2.6) and sequenced by The Centre for Applied Genomics (TCAG, Toronto). A list of generated constructs can be found in Table 2.4.

## 2.4 Preparation and transformation of competent cells

### *2.4.1 Chemically competent* E. coli

A 2 mL culture of the *E. coli* strain Shuffle Express was grown in LB media for 16 hours at 30°C with shaking. This was used to inoculate 100 mL LB media at a 1:100 dilution, which was subsequently incubated at 30°C with shaking until early exponential growth was observed, indicated by an $OD_{600}$ measurement of 0.35 – 0.4. The culture was chilled on ice for 10 minutes prior to centrifugation at 3,000 *g* for 15 minutes at 4°C. Cells were gently resuspended in 50 mL ice-cold sterile 100 mM $MgCl_2$. The cells were centrifuged as before and gently resuspended in 25 mL ice-cold sterile 100 mM $CaCl_2$ and incubated on ice for 20 minutes. Following another centrifugation as above, cells were resuspended in 10 mL ice-cold sterile 85 mM $CaCl_2$ supplemented with 15% glycerol. After a final centrifugation, cells are resuspended in 2 mL ice-cold sterile 85 mM $CaCl_2$ and 15% glycerol. Cells were divided into 50 µL portions and snap frozen with a mixture of dry ice and ethanol prior to storage at -80°C. Frozen cells were thawed on ice prior to the addition of 1 µL containing ≈ 100 ng of plasmid DNA. Cells and DNA were incubated on ice for 30 minutes, then heat shocked at 42°C for 30 seconds. Transformations were incubated on ice for 5 minutes before the addition of 950 µL SOC media. Outgrowth was carried out at 30°C for 1 hour with shaking. Transformations were then plated onto selective media containing 50 µg/mL kanamycin (BioShop) and grown overnight at 30°C. Transformants were confirmed via High-Speed Plasmid Mini Prep Kit (FroggaBio) and diagnostic digestion.

*2.4.2 Electrocompetent* E. coli

A 10 mL culture of the *E. coli* strain DH10B was grown in LB media for 16 hours at 37°C with shaking. This was used to inoculate 500 mL LB media at a 1:100 dilution, which was subsequently incubated at 37°C with shaking until early logarithmic growth was observed, indicated by an $OD_{600}$ measurement of 0.35 – 0.4. The culture was chilled on ice for 10 minutes prior to centrifugation at 1,000 *g* for 20 minutes at 4°C. Cells were gently resuspended in 400 mL ice-cold sterile MilliQ $H_2O$ prior to a second centrifugation. Cells were gently resuspended in 200 mL ice-cold sterile MilliQ $H_2O$ prior to a third centrifugation. Cells were then gently resuspended in 40 mL ice-cold sterile 10% glycerol and centrifuged as before. Cells were gently resuspended in 1 mL ice-cold sterile 10% glycerol resulting in a final $OD_{600}$ of 200 – 250. Cells were divided into 50 µL aliquots and snap frozen with a mixture of dry ice and ethanol prior to storage at -80°C. Frozen cells were thawed on ice prior to the addition of 1 µL containing ≈ 10 ng of plasmid DNA. Cells were transformed as detailed in Section 2.2.2. Transformants were confirmed via High-Speed Plasmid Mini Prep Kit (FroggaBio) and diagnostic digestion.

*2.4.3 Electrocompetent* C. glutamicum

A 5 mL culture of *C. glutamicum* was grown overnight in MB media (0.5% w/v yeast extract, 1.5% w/v tryptone, 0.5% w/v soytone, 0.5% w/v NaCl) at 30°C with shaking. This was used to inoculate 200 mL MB media supplemented with 3.5% w/v glycine at a 1:100 dilution and incubated at 30°C with shaking until an $OD_{600}$ measurement of 0.2 – 0.25 was reached. Ampicillin (BioBasic) was then added to a final concentration of 0.5 µg/mL and the culture was further incubated under the same conditions for 1.5 h. The culture was chilled on ice for 10 minutes, then centrifuged at 3,000 *g* for 10 minutes at 4°C. Cells were gently resuspended in 30 mL ice-cold sterile 20 mM HEPES, 5% v/v glycerol (pH 7.2) and centrifuged as before. Resuspension and centrifugation was repeated two more times before the final cell pellet was resuspended in 1 mL ice-cold sterile 5 mM HEPES, 15% v/v glycerol (pH 7.2). Cells were

divided into 100 µL portions and snap frozen with a mixture of dry ice and ethanol prior to storage at -80°C. Frozen cells were thawed on ice prior to the addition of 1 µL containing ≈ 50 ng of plasmid DNA. Cells and DNA were electroporated using a 0.2 mm gap electroporation cuvette and a MicroPulser (Bio-Rad) set to deliver a single 2.5 kV pulse. Outgrowth was carried out with shaking at 30°C for 2 hours in LB media supplemented with 0.5% w/v glucose prior to plating onto selective media containing 25 µg/mL kanamycin (BioShop) and overnight growth at 30°C. Transformants were confirmed via High-Speed Plasmid Mini Prep Kit (FroggaBio), plasmid rescue by DH10B, and diagnostic digestion.

**2.5 TorA-dependant localization of recombinant proteins**

*2.5.1 E. coli periplasmic extraction of recombinant proteins (adapted [2])*

2 mL cultures of Shuffle Express with various constructs (containing GFP, CFI-65, CFI-75, CFI-85 or CFI-93) were grown overnight in LB media with 50 µg/mL kanamycin (BioShop) at 30°C with shaking. This was used to inoculate 50 mL 2YT media (Sigma-Aldrich) with 50 µg/mL kanamycin (BioShop) at a 1:500 dilution and incubated at 30°C with shaking until exponential growth was observed, indicated by an $OD_{600}$ measurement of 0.5 – 0.7. Sterile IPTG (isopropyl β-D-1-thiogalactopyranoside, BioVectra) was added to a final concentration of 0.5 mM and induction was continued overnight at 25°C with shaking. Cells were harvested by centrifugation at 1,100 *g* for 10 minutes at 4°C. The supernatant was decanted and cell pellets were gently resuspended in the residual media and 500 µL of $CHCl_3$ was added. The samples were maintained at room temperature for 15 minutes with occasional mixing, then 5 mL 10 mM Tris-HCl (pH 8.0) was added. Protoplasts were separated from periplasmic contents by centrifugation at 6,000 *g* for 20 minutes at 4°C. The supernatant fraction containing periplasmic proteins was carefully removed and stored at 4°C. Remaining protoplasts were further lysed using CelLytic B™ lysis reagent (Sigma-Aldrich) to analyze both the soluble and pellet fractions. Protoplasts were resuspended in 5 mL of a 0.2X CelLytic B™ solution containing 0.2 mg/mL lysozyme (BioShop), 50 U/mL

Benzonase® (Sigma-Aldrich), and a complete™ Mini EDTA protease inhibitor tablet (Roche) and incubated at room temperature on a rotator for 15 minutes. Extracts were then centrifuged at 16,000 *g* for 10 minutes at 4°C and the supernatant containing soluble proteins was removed and stored at 4°C. Insoluble pellets were resuspended in 5 mL TBS (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) and stored at 4°C. Samples were subjected to SDS-PAGE and analysed by Western blotting using anti-$His_6$ peroxidase (Roche).

**2.6 Expression and purification of recombinant glycoside hydrolases**

*2.6.1 Expression and purification in* Escherichia coli

2 mL cultures of Shuffle Express with various constructs (CFI-65, CFI-75, CFI-85, or CFI-93) were grown overnight in LB media with 50 µg/mL kanamycin (BioShop) at 30°C with shaking. This was used to inoculate 200 mL 2YT media (Sigma-Aldrich) with 50 µg/mL kanamycin (BioShop) at a 1:500 dilution and incubated at 30°C with shaking until logarithmic growth was observed, indicated by an $OD_{600}$ measurement of 0.5 – 0.7. Sterile IPTG (BioVectra) was added to a final concentration of 0.5 mM and induction was continued overnight at 25°C with shaking. Cells were harvested by centrifugation at 9,000 *g* for 15 minutes at 4°C and frozen at -20°C until needed. Cell pellets were thawed on ice and mechanically lysed in a mortar and pestle with Celite® (diatomaceous earth, Sigma-Aldrich). Lysed $His_6$ constructs (CFI-65, CFI-75, and CFI-85) were resuspended in 5 mL binding/wash buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 15 mM imidazole, pH 8.0) containing a cOmplete™ Mini EDTA protease inhibitor tablet (Roche), and clarified via centrifugation at 13,000 *g* for 20 minutes at 4°C. The clarified lysate was added to 1 mL of a 50% Ni-NTA agarose (Qiagen) slurry that had been previously pre-equilibrated in binding buffer and incubated at 4°C for 1 h on a rotator. The lysate-Ni-NTA agarose mixture was loaded onto a disposable column and washed twice with 5 mL binding/wash buffer. Proteins were eluted from the matrix twice in 1 mL elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 300 mM imidazole, pH 8.0). Samples

of each fraction were analyzed via SDS-PAGE, elutions were pooled and buffer exchanged to 50 mM HEPES, 300 mM NaCl (pH7.0). Following lysis, CFI-93 was resuspended in 5 mL preparation buffer (50 mM ammonium bicarbonate, pH9.0) containing a cOmplete™ Mini EDTA protease inhibitor tablet (Roche), and clarified via centrifugation at 13,000 *g* for 20 minutes at 4°C. The clarified lysate was subjected to anion-exchange chromatography on SourceQ (Bio-Rad). Protein fractions were applied to the column using a peristaltic pump at a flow rate of 1 mL/min and eluted with a linear gradient from 0 to 1.0 M NaCl over 30 minutes [12]. Fractions were analyzed by SDS-PAGE and the most pure fractions were pooled and buffer exchanged to 50 mM HEPES, 300 mM NaCl (pH7.0).

*2.6.2 Expression and purification in* C. glutamicum

2 mL cultures of *C. glutamicum* with various constructs (CFI-65, CFI-75, CFI-85, or CFI-93) were grown overnight in LB media with 25 µg/mL kanamycin (BioShop) at 30°C with shaking. This was used to inoculate 50 mL 2YT media (Sigma-Aldrich) with 25 µg/mL kanamycin (BioShop) at a 1:100 dilution and incubated at 30°C with shaking overnight. Sterile IPTG (BioVectra) was added to a final concentration of 0.5 mM and induction was continued over 48 h at 30°C with shaking. Supernatants were clarified via centrifugation at 9,000 *g* for 20 minutes at 4°C. Supernatants were stored at 4°C and cell pellets were frozen at -20°C. A 10X stock solution of binding buffer was diluted with the supernatants of $His_6$ constructs (CFI-65, CFI-75, and CFI-85) to a final concentration of 50 mM $NaH_2PO_4$, 300 mM NaCl, 15 mM imidazole (pH 8.0) and contained a Complete™ Mini EDTA protease inhibitor tablet (Roche). 1 mL of a 50% Ni-NTA agarose (Qiagen) slurry that had been previously pre-equilibrated was added to the supernatant and incubated overnight at 4°C on a rotator. The supernatant-Ni-NTA agarose mixture was loaded onto a disposable column and washed four times with 5 mL binding/wash buffer. Proteins were eluted from the matrix twice in a total of 1 mL elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 300 mM imidazole, pH 8.0). Samples of each fraction were analyzed via SDS-PAGE, elutions were pooled and

buffer exchanged to 50 mM HEPES, 300 mM NaCl (pH7.0). A 10X stock solution of preparation buffer was diluted with the supernatant of CFI-93 to a final concentration of 50 mM ammonium bicarbonate (pH9.0) and contained a cOmplete™ Mini EDTA protease inhibitor tablet (Roche). The supernatant was subjected to anion-exchange chromatography on SourceQ (Bio-Rad). Protein fractions were applied to the column using a peristaltic pump at a flow rate of 1 mL/min and eluted with a linear gradient from 0 to 1.0 M NaCl over 30 minutes [12]. Fractions were analyzed by SDS-PAGE and the most pure fractions were pooled and buffer exchanged to 50 mM HEPES, 300 mM NaCl (pH7.0).

**2.7 Enzyme activity assays**

*2.7.1 PAHBAH*

Increasing concentrations of reducing ends produced by glycoside hydrolases acting on various carbohydrate substrates were monitored using 4-hydroxybenzoic acid hydrazide (PAHBAH) (Sigma-Aldrich). The original protocol by Lever [35] was modified as follows. The PAHBAH reagent (5% w/v PAHBAH, 5% HCl) and the alkali solution (50 mM citrate, 10 mM $CaCl_2$, 0.5M NaOH, pH 8.0) were prepared separately in advance. Prior to use, 1 part PAHBAH reagent was combined with 9 parts alkali solution and the mixture was kept on ice during use. Various concentrations of carbohydrate substrates (generally 0.5% in assay) were combined with 2.5 µg/mL glycoside hydrolase in 50 mM MOPS, 50 mM NaCl (pH 5.5 for celf_3184, pH 6.0 for celf_1230) and allowed to incubate at optimal enzyme temperature. At designated time points 100 µL of the assay were quenched with the addition of 50 µL 0.5 M NaOH. 30 µL of the quenched reaction were placed into a 96-well plate and had 200 µL of the PAHBAH working reagent added to it. After the completion of all time points, 96-well plates were incubated for 10 minutes at 100°C then allowed to slowly cool to room temperature before being read at 410 nm.

*2.7.2 Dye-conjugated substrates*

Azurine-crosslinked-substrates (AZCL, Megazyme) were used to determine the relative activity of purified glycoside hydrolases in a colorimetric assay. AZCL-Hydroxyethylcellulose (AZCL-HEC) and AZCL-Barley-β-Glucan (AZCL- β-Glucan) were utilized to measure cellulase activity. Assays were prepared to a total volume of 1 mL in 50 mM MOPS, 50 mM NaCl (pH 6.0), 5 mg/mL AZCL-substrate, and 100 µg/mL purified glycoside hydrolase. Substrates were preheated and assays were incubated at optimal temperatures of previously characterized enzymes (65°C for CFI-75, 55°C for CFI-93). At designated time points, 150 µL of the assay was removed and centrifuged at 5,000 *g* for 5 minutes to pellet the insoluble material. 100 µL of the chromogenic supernatant was then moved to a 96-well plate and read at 595 nm.

**2.8 Thin-layer chromatography**

The determination of hydrolysis products for celf_1230 was achieved via thin-layer chromatography (TLC). Sufficient separation of standard cellodextrins (G1-G5, 2 mg/mL each) was observed when using a silica gel 60 plate on a plastic backing (Merck). The hydrolysis of cellotetraose (Megazyme) and 1,3:1,4-β-glucotetraose (Megazyme) (Figure 2.4) by celf_1230 was conducted at 65°C (5 mg/mL substrate, 2.5 µg/mL celf_1230, 50 mM MOPS, 50 mM NaCl, pH6.0). Reactions were quenched at various time points by the addition of 0.5 M NaOH before spotting onto the silica plate and being exposed to the liquid phase solvent (ethyl acetate:methanol:$H_2O$:acetic acid in a 4:2:1:0.1 ratio by volume). Carbohydrates were visualized on the TLC plate by spraying with 5% (v/v) $H_2SO_4$ in ethanol and charring over a hot plate.

**2.9 Determination of enzyme concentration**

Enzyme concentrations were determined using $A_{280}$ readings obtained when using a 50 µL microcell and a spectrophotometer. The theoretical molar extinction coefficients ($M^{-1}$ $cm^{-1}$) used for each protein can be found in Table 2.5 along with their molecular weights. Parameters were determined by ExPASy ProtParam [68].

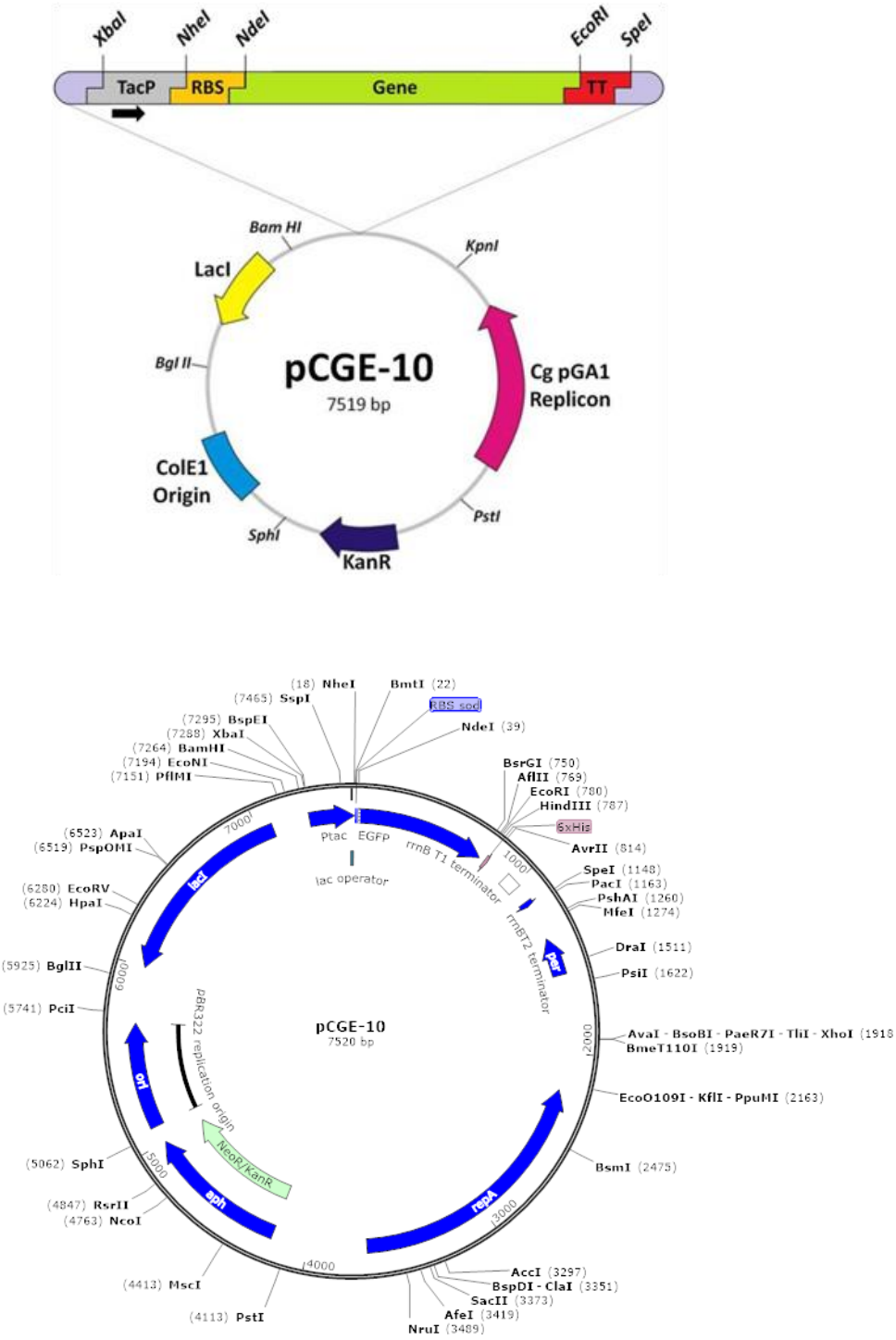
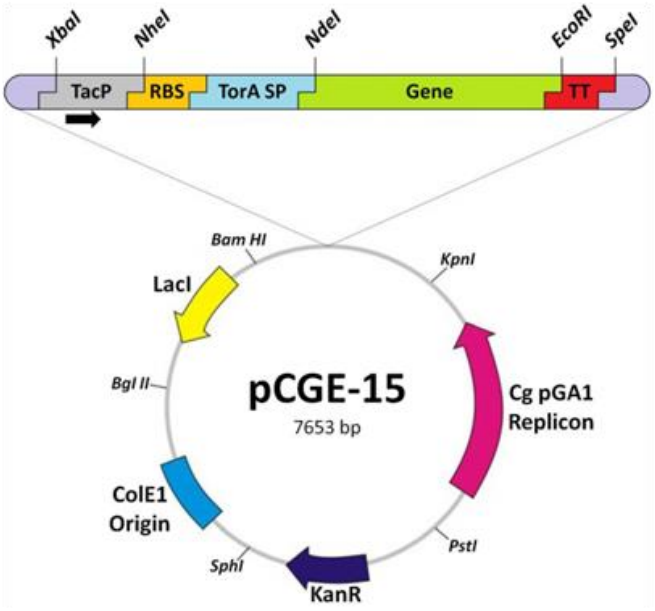



**Figure 2.1: Simplified and detailed vector maps of pCGE-10.**

Vector pCGE-10 features kanamycin resistance (KanR), Lac repressor (LacI) for IPTG induction, ColE1 origin for replication in *E. coli*, and Cg pGA1 replicon for replication in *C. glutamicum*.



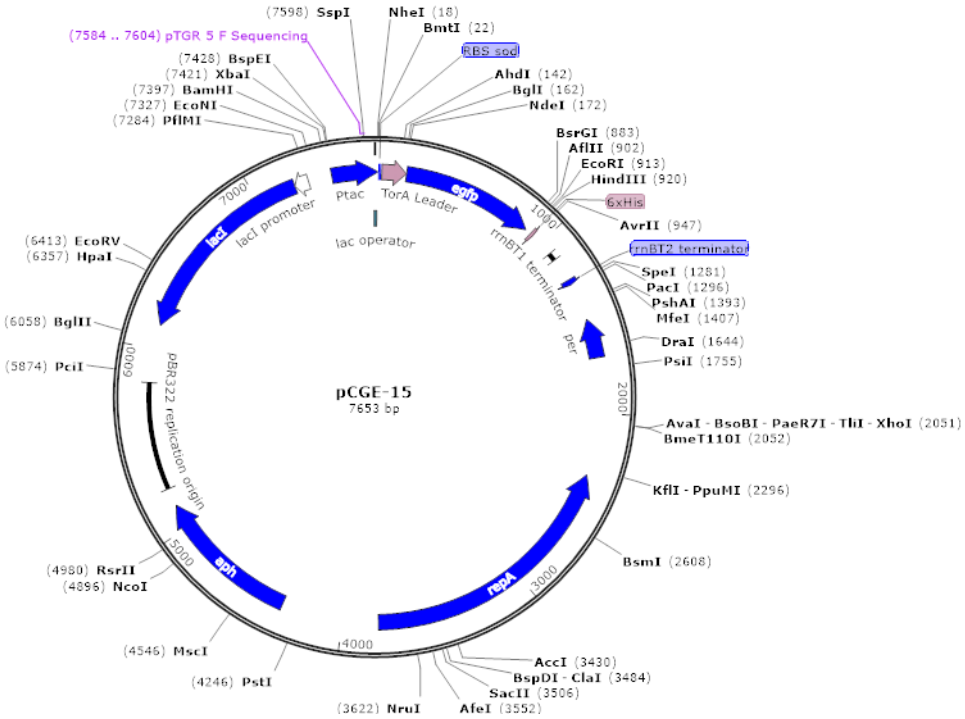




**Figure 2.2: Simplified and detailed vector maps of pCGE-15.**

Vector pCGE-15 features are identical to pCGE-10, including the addition of the TorA TAT signal sequence directly upstream of the MCS.

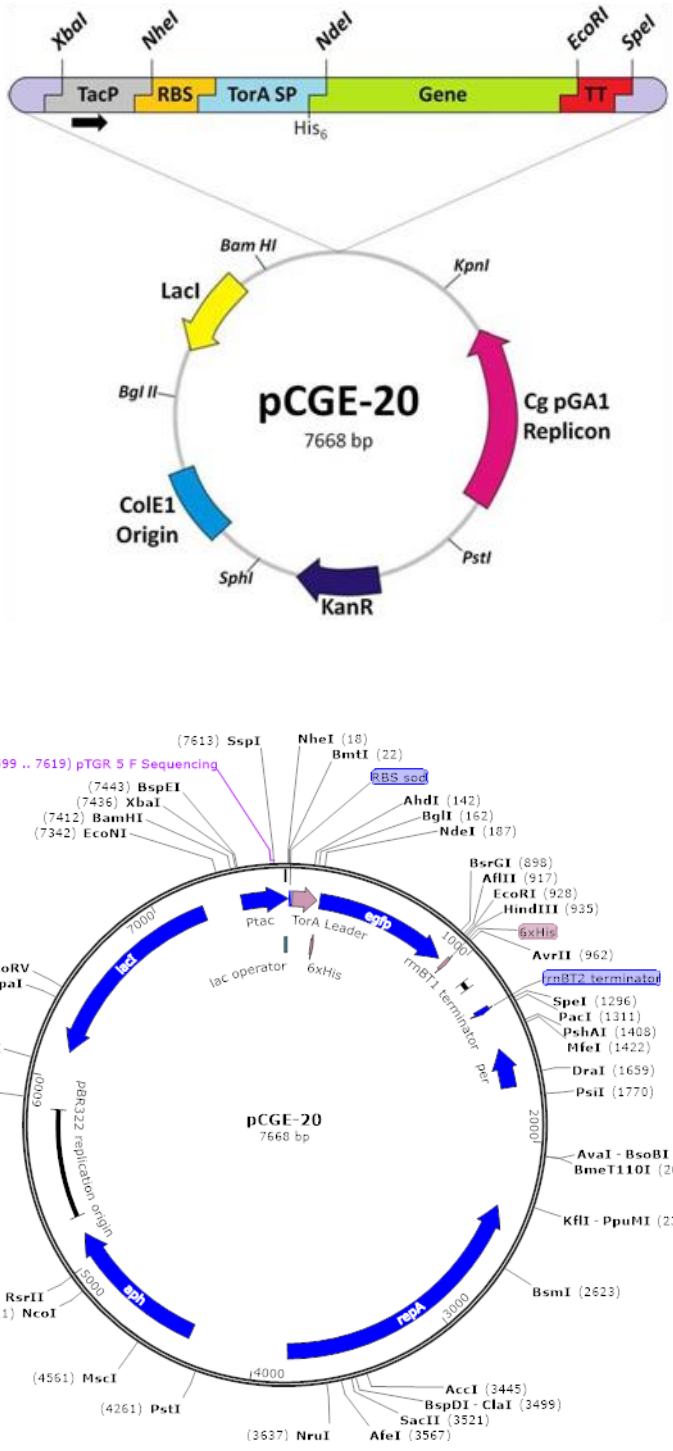



**Figure 2.3: Simplified and detailed vector maps of pCGE-20.**

Vector pCGE-20 features are identical to pCGE-15, including the addition of $His_6$ tag at the C-terminus of the TorA TAT signal sequence.

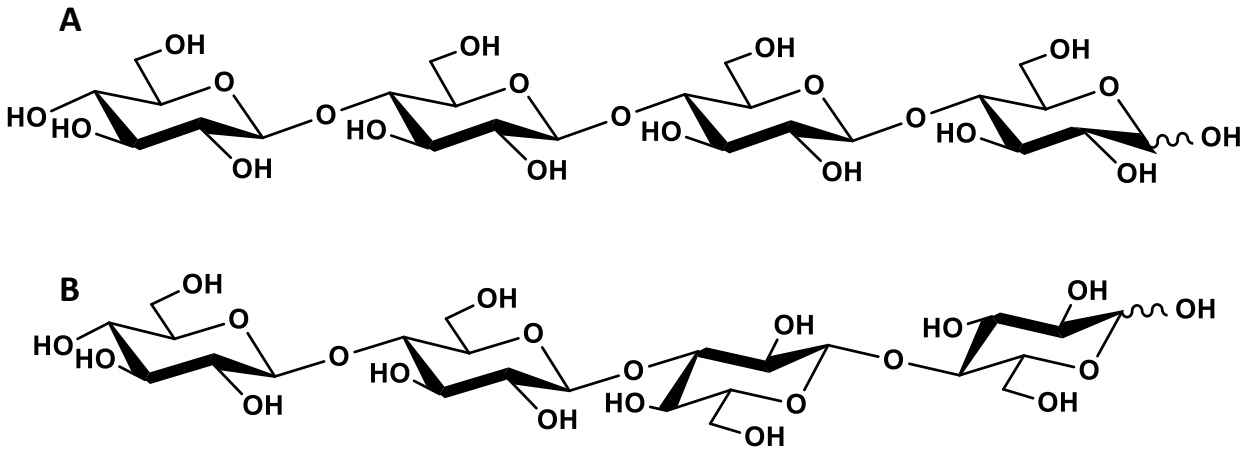



**Figure 2.4: Structural comparison of cellotetraose (A) and 1,3:1,4-β-glucotetraose (B).**

Cellotetraose (A) is composed of 4 glucose β-1,4-linked glucose monomers, while 1,3:1,4-β-glucotetraose (B) is composed of two β-1,4-linkages and a β-1,3-linkage in the middle.

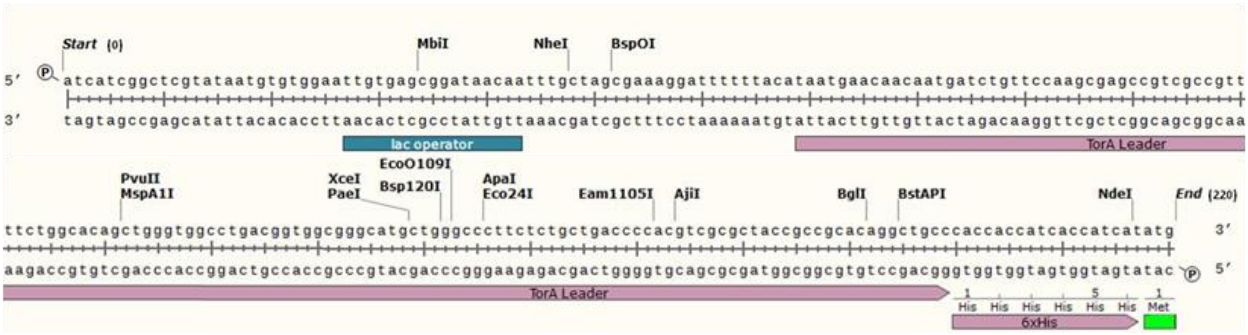



**Figure 2.5 Sequence of TorA signal peptide within expression vector pCGE-20.**

Located between the ribosome binding site and the cloning site of expression vectors pCGE-15 (lacking 6xHIS tag) and pCGE-20, the TorA signal peptide directs expressed proteins to the periplasm in Gram negative bacteria and allows for secretion in Gram positive bacteria.

**Table 2.1: TorA synthetic gene sequence and features.**

| Gene Name | Sequence (5'-3') | Features |
|---|---|---|
| **TorA** | ATTAATGAACAACAATGATCTGTTCCAAGCGAGCCGTCGCC GTTTTCTGGCACAGCTGGGTGGCCTGACGGTGGCGGGCAT GCTGGGCCCTTCTCTGCTGACCCCACGTCGCGCTACCGCCG CACAGGCTGCCCATATG | Encodes TorA TAT signal peptide. Contains 5' AseI (degenerate NdeI) site and 3' NdeI site. |
| **TorA N-His$^6$** | ATTAATGAACAACAATGATCTGTTCCAAGCGAGCCGTCGCC GTTTTCTGGCACAGCTGGGTGGCCTGACGGTGGCGGGCAT GCTGGGCCCTTCTCTGCTGACCCCACGTCGCGCTACCGCCG CACAGGCTGCCCACCACCATCACCATCATATG | Encodes TorA TAT signal peptide. Contains 5' AseI (degenerate NdeI) site, 3' NdeI site, and C-terminal $His_6$ tag. |

**Table 2.2: Glycoside hydrolase cloning primers.**

| Primer Name | Target Gene | Sequence (5'-3') | Tm (°C) | Annealing Temp (°C) | Features |
|---|---|---|---|---|---|
| **1 F** | celf_2403 | GGG GTA TTC **CAT ATG** AGA GCA GCG ACC CGT ACG GCG | 71.42 | 66 | 5' primer with **NdeI** site. |
| **1 FL** | celf_2403 | GGG GTA TTC **CAT ATG** GCC TCC CCC GCG GTC GGT CCC | 78.61 | 74 | 5' primer with **NdeI** site. Lacking native leader. |
| **1 R** | celf_2403 | GGG **GAA TTC** TCA TCA GCC GCG CGT CCA CTG CTG | 74.84 | 70 | 3' primer with **EcoR**I site. |
| **2 F** | celf_1230 | GGG GTA TTC **CAT ATG** GCC CGA CCC TTC CGC ACC CCT CTC GCC | 83.07 | 78 | 5' primer with **NdeI** site. |
| **2 FL** | celf_1230 | GGG GTA TTC **CAT ATG** GGC GGC CCC CCG CAC GGG TAC CCG | 84.75 | 80 | 5' primer with **NdeI** site. Lacking native leader. |
| **2 R** | celf_1230 | GGG **GAA TTC** TCA TCA GCT GCG CGG GCC CCC GAT CGG | 84.13 | 79 | 3' primer with **EcoRI** site. |
| **3 F** | celf_0233 | GGG GTA TTC **CAT ATG** CCC CGT CCC GGT CCG GTC CCG GCC | 89.54 | 85 | 5' primer with **NdeI** site. |
| **3 FL** | celf_0233 | GGG GTA TTC **CAT ATG** ACG ACC GCA CCG GAC CGC GTG CTC | 86.56 | 82 | 5' primer with **NdeI** site. Lacking native leader. |
| **3 R** | celf_0233 | GGG **GAA TTC** TCA TCA CCA GGA CGC GGC CGC GGC CAG | 87.73 | 83 | 3' primer with **EcoRI** site. |
| **4 F** | celf_3184 | GGG GTA TTC **CAT ATG** TCC ACC CGC AGA ACC GCC GCA GCG | 87.09 | 82 | 5' primer with **NdeI** site. |
| **4 R** | celf_3184 | GGG **GAA TTC** TCA TCA CCA CCT GGC GTT GCG CGC CAT C | 85.55 | 81 | 3' primer with **EcoRI** site. |

**Table 2.3: Shuttle vectors.**

| Vector Name | Resistance | Tags | Features |
|---|---|---|---|
| **pCGE-10** | Kanamycin | N/A | N/A |
| **pCGE-15** | Kanamycin | N/A | TorA TAT signal peptide |
| **pCGE-20** | Kanamycin | $His_6$ | TorA TAT signal peptide |

**Table 2.4: Glycoside hydrolase construct library.**

| Construct Name | Gene | Tags | Features |
|---|---|---|---|
| **CFI-63** | celf_0233 | N/A | Amplified from *C. fimi* GH6. |
| **CFI-64** | celf_0233 | N/A | Amplified from *C. fimi* GH6. Lacking native leader. |
| **CFI-65** | celf_0233 | $His_6$ | Amplified from *C. fimi* GH6. Lacking native leader. |
| **CFI-73** | celf_1230 | N/A | Amplified from *C. fimi* GH6. |
| **CFI-74** | celf_1230 | N/A | Amplified from *C. fimi* GH6. Lacking native leader. |
| **CFI-75** | celf_1230 | $His_6$ | Amplified from *C. fimi* GH6. Lacking native leader. |
| **CFI-83** | celf_2403 | N/A | Amplified from *C. fimi* GH5. |
| **CFI-84** | celf_2403 | N/A | Amplified from *C. fimi* GH5. Lacking native leader. |
| **CFI-85** | celf_2403 | $His_6$ | Amplified from *C. fimi* GH5. Lacking native leader. |
| **CFI-93** | celf_3184 | N/A | Amplified from *C. fimi* GH6. |

**Table 2.5: Theoretical molar extinction coefficients of expressed glycoside hydrolases.**

| Gene | Length (AA) | Molecular Weight (kDa) | Extinction Reduced Cys ($M^{-1}$ $cm^{-1}$) | Extinction Non-Reduced Cys ($M^{-1}$ $cm^{-1}$) |
|---|---|---|---|---|
| **celf_0233** | 613 | 61.78 | 86,400 | 86,650 |
| **celf_1230** | 433 | 46.31 | 107,830 | 108,205 |
| **celf_2403** | 475 | 49.83 | 121,350 | 121,725 |
| **celf_3184** | 449 | 46.70 | 107,370 | 107,745 |

**Table 2.6: Shuttle vector and construct sequencing primers.**

| Primer Name | Target | Sequence (5'- 3') | Tm (°C) | Features |
|---|---|---|---|---|
| **pTGR 7451F** | pCGE-10<br>pCGE-15<br>pCGE-20 | ACGGTTCTGGCAAATATTCTG | 54 | Anneals upstream of *lac* operator and sequences towards MCS. Capable of confirming operator, RBS, and signal peptide sequence (if applicable). |

**Table 2.7: Genotypes of bacterial strains.**

| Strain | Genotype |
|---|---|
| ***Cellulomonas fimi* ATCC 484** | Wild-type |
| ***Escherichia coli* (DH10B)** | F$^-$*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139Δ(*ara*, *leu*)7697 *gal*U *gal*K λ$^-$*rps*L *nup*G |
| ***Escherichia coli* (Shuffle Express)** | *fhuA2 [lon] ompT ahpC gal* λatt::pNEB3-r1-*cDsbC* (SpecR, *lacI$^q$) ΔtrxB sulA11* R(*mcr-73::miniTn10*--Tet$^S$)*2 [dcm] R(zgb-210::Tn10* --Tet$^S$) *endA1 Δgor Δ(mcrC-mrr)114::IS10* |
| ***Corynebacterium glutamicum* ATCC 13032** | Wild-type |

# Chapter 3

# Results

## 3.1 Molecular biology

### *3.1.1 Vector modification*

Synthetic TorA genes ordered from Biobasic Canada (Markham, Ontario) were contained within the MCS of pUC57. Genes were excised using a two-step double digestion. The entire MCS was first excised via digestion with both EcoRI and HindIII, then purified by gel extraction (as described in Section 2.2.2). The MCS fragment (containing the TorA synthetic gene) was then further digested using AseI and NdeI and purified with a secondary gel purification (Appendix Figure 5.1). The TorA synthetic genes were then ligated into the NdeI site of the pCGE-10 plasmid. Since NdeI and AseI sites are compatible, orientation of the TorA gene inserts needed to be determined. Since AseI creates a degenerate site, it was possible to determine insert orientation by digestion with NdeI and another enzyme. To identify the presence of an insert for both pCGE-15 and pCGE-20 (Appendix Figure 5.2 A and 5.3 A, respectively) purified plasmid DNA from clones was digested using BamHI and NdeI. To rule out multiple insertions and confirm orientation purified plasmid DNA from clones was digested using NdeI and EcoRI for both pCGE-15 and pCGE-20 (Appendix Figure 5.2 B and 5.3 B respectively). Vector modifications were confirmed via sequencing as described in Section 2.3.3.

### *3.1.2 Glycoside hydrolase construct library*

Glycoside hydrolase genes were amplified from *C. fimi* genomic DNA using synthetic primers (Table 2.2). Genes were amplified, digested with NdeI and EcoRI, and purified as described in Section 2.2.2 and 2.3.2 (Appendix Figure 5.4 A) prior to cloning into pCGE-15 and pCGE-20. Generated constructs (Table 2.4) were confirmed using diagnostic single and double digestions (to check for multiple insertion events) for

pCGE-10 constructs (Appendix Figure 5.4 B) in addition to pCGE-15 and pCGE-20 constructs (Appendix Figure 5.4 C). Constructs were also confirmed via sequencing as described in Section 2.3.3.

## 3.2 Recombinant protein production

### *3.2.1 Test expression of eGFP*

The *E. coli* strain Shuffle Express was transformed with pCGE-10 and pCGE-15 (unmodified vector, and vector containing the TorA synthetic gene). Both transformed strains were grown, induced to express eGFP, and protein was extracted from various cellular fractions (as described in Section 2.5.1) in order to determine the functionality of the TorA modification. Fractions were first analyzed by visualization with blue light tray to observe differences in fluorescence, then subjected to SDS-PAGE analysis and visualized by UV tray to determine any difference in total eGFP expression. Increased fluorescence was observed in the periplasmic fraction of cells containing pCGE-15, and significantly more fluorescence was also observed in the soluble fraction of these cells (Figure 3.1 A). SDS-PAGE analysis showed that the quantity of eGFP expressed by cells transformed by pCGE-10 and pCGE-15 were equivalent.

### *3.2.2 Expression of glycoside hydrolases in Escherichia coli*

Glycoside hydrolase constructs were transformed into *E. coli* strain Shuffle Express. These cells were grown, induced, and protein was extracted from various cellular fractions as described in Section 2.6.1. Samples were subjected to SDS-PAGE analysis and western blot to identify the production of recombinant proteins (Figure 3.2). As seen previously, CFI-65 and CFI-85 were only produced in an insoluble form in *E. coli*. Both CFI-75 and CFI-93 were produced in a soluble form (CFI-93 does not produce a visible band, so recombinant protein was confirmed via an enzyme activity assay) and CFI-75 was also seen to be exported to the periplasmic fraction (Figure 3.2 B).

Active CFI-75 and CFI-93 were purified from *E. coli* cell lysates (Figure 3.3 A and Figure 3.4 A, respectively) via affinity or anion exchange chromatography. Following the comparison of activity with control proteins, it was determined that 780 µg of CFI-75 and 43 µg of CFI-93 were purified (Figure 3.3 B and Figure 3.4 B, respectively). Celf_1230 [27] and Celf_3184 [12] (used as controls in gels or assays) were produced solely in *E. coli* and originated from previously utilized constructs. Though similar to the methods described in Section 2.6.1, the purification of these proteins is not described here.

*3.2.3 Expression of glycoside hydrolases in Corynebacterium glutamicum*

Glycoside hydrolase constructs were transformed into *C. glutamicum*. These cells were grown, induced, and protein was purified from the culture supernatants as described in Section 2.6.2. Expression and purification of CFI-65 and CFI-85 was not successful with *C. glutamicum* and the data is not shown. Active CFI-75 and CFI-93 were able to be purified from *C. glutamicum* culture supernatants (Figure 3.5 A and Figure 3.6 A, respectively) via affinity or anion exchange chromatography. Following the comparison of activity with control proteins, it was determined that 25 µg of CFI-75 and 350 µg of CFI-93 were purified (Figure 3.5 B and Figure 3.6 B, respectively).

**3.3 Biochemical characterization**

*3.3.1 Control protein activity on cellulose derivatives*

Though the activity of Celf_1230 and Celf_3184 produced in *E. coli* had previously been assayed on CMC [12, 27], their activities on cellulose derivatives remained to be determined. For this reason, the activity of Celf_1230 and Celf_3184 were assayed on wheat arabinoxylan, 1,3:1,4-Barley β-glucan, and tamarind xyloglucan alongside various concentrations of CMC (Figure 3.7). Each glycoside hydrolase was assayed at its optimal temperature and pH (65°C and pH 6.0 for Celf_1230, 50°C and pH 5.5 for Celf_3184) [12, 27] as described in Sections 2.7.1, 2.7.2, and 2.8. Both glycoside hydrolases exhibited significantly improved

activity on 1,3:1,4-Barley β-glucan in comparison to all concentrations of CMC assayed. The hydrolysis products of both Celf_1230 and Celf_3184 on CMC and 1,3:1,4-Barley β-glucan was assayed by TLC as described in Section 2.8 (Figure 3.8) to identify if the β-1,3 glycosidic bond (Figure 2.4) was being cleaved, resulting in the observed increase in activity on the mixed linkage substrate. The result of this assay shows that the β-1,3 linkage is in fact not being cleaved.

The thermostabilities of Celf_1230 and Celf_3184 were assayed on both CMC and 1,3:1,4-Barley β-glucan to observe any trends in activity over time for both glycoside hydrolases and substrates at both their optimal temperatures and +5°C of their optimal temperatures (Figure 3.9). Assays were performed at the optimal pH of each enzyme as described in Section 2.7.2. Both glycoside hydrolases showed improved activity on 1,3:1,4-Barley β-glucan (compared to CMC) over time and across both assayed temperatures. These results show that both glycoside hydrolases are not only more active, but also more thermostable on the 1,3:1,4-Barley β-glucan substrate than on CMC.

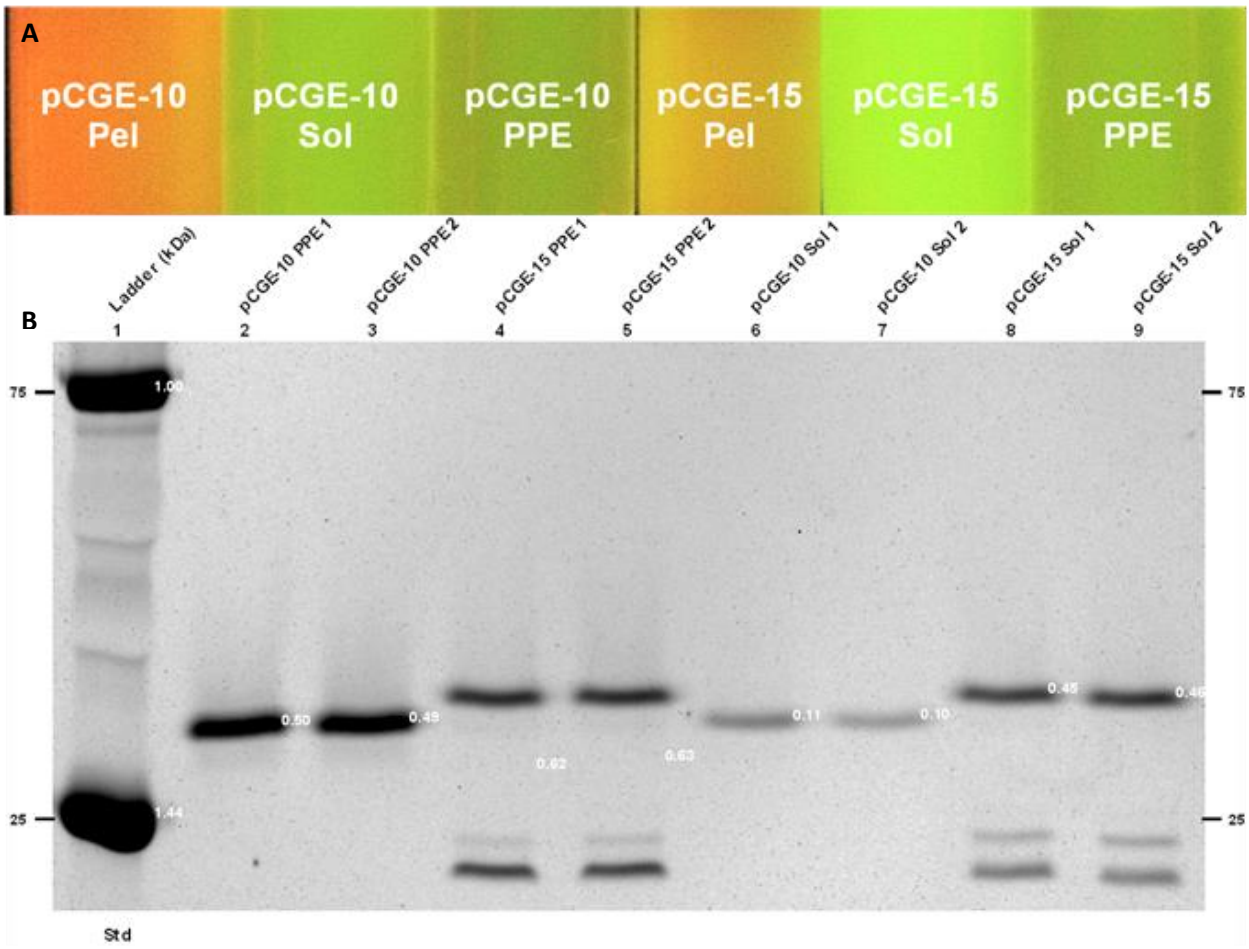



**Figure 3.1: Comparison of GFP activity and localization in expression vectors with and without TorA TAT signal peptide (pCGE-15 and pCGE-10, respectively).** Comparison of GFP fluorescence **(A)** in periplasmic (PPE), soluble (Sol), and pellet (Pel) fractions. GFP expressed with TorA signal peptide shows more fluorescence when visualized with a shortwave blue light box in all fractions, especially the soluble fraction. Increased fluorescence in the soluble fraction of GFP coupled to TorA is the result of a greater proportion of correctly folded and active protein in comparison to GFP without the TAT signal sequence. Fractions were run in side-by-side duplicates on a 12% SDS non-denaturing PAGE to preserve GFP fluorescence and visualized via UV tray **(B)**, relative band intensities are shown. The TorA signal peptide improved GFP activity by a factor of 1.26 in the periplasmic fraction and by a factor of 4.33 in the cytoplasmic fraction.

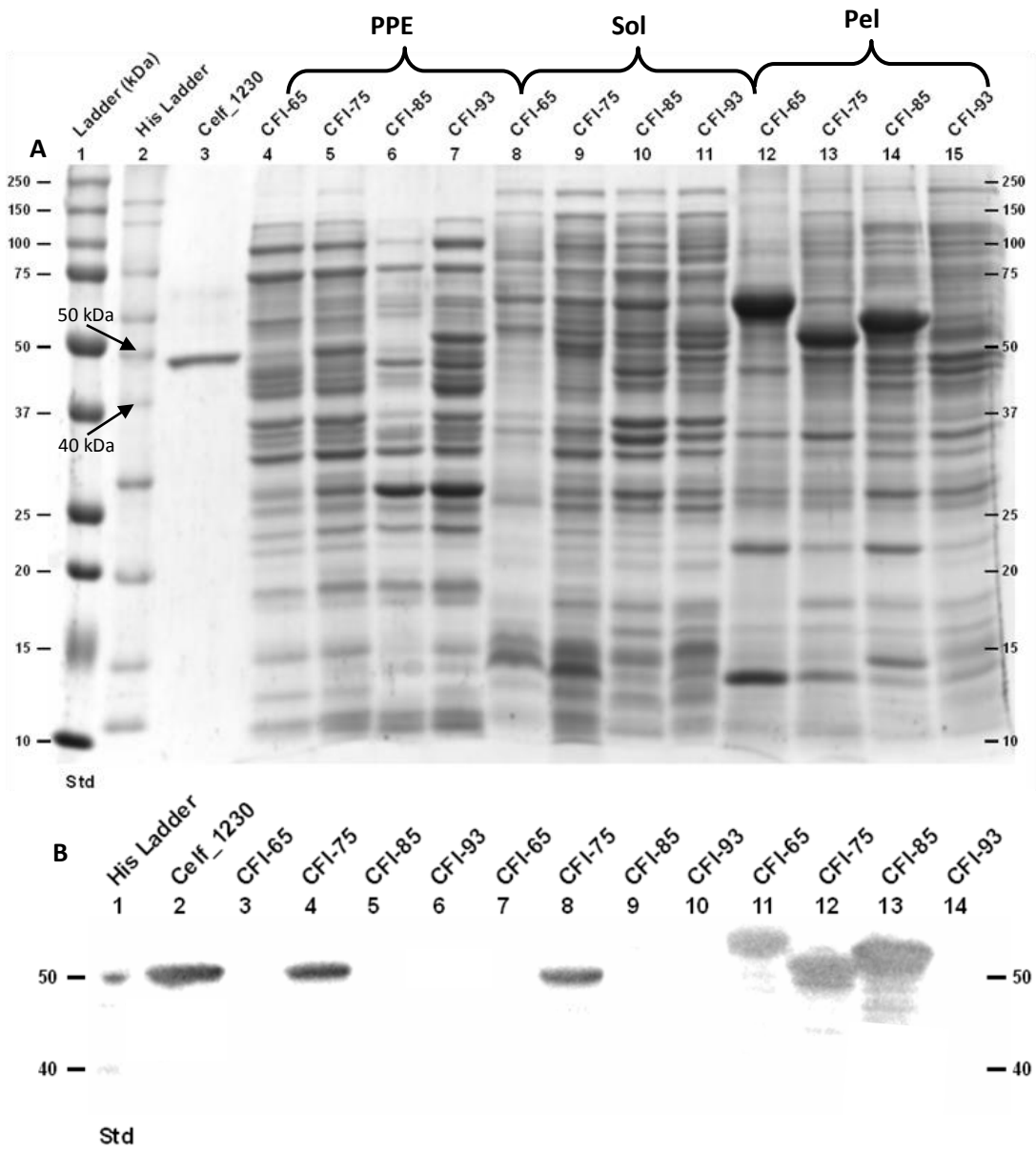



**Figure 3.2: Coomassie stained 12% SDS-PAGE showing periplasmic (PPE), soluble (Sol), and pellet (Pel) fractions of $His_6$ tagged constructs and corresponding Western blot.** In *E. coli*, only CFI-75 and CFI-93 are produced in a soluble form, but at such low levels they are unidentifiable via Coomassie staining **(A)**. The TorA signal peptide is shown to localize soluble CFI-75 into the *E. coli* periplasm **(B)**, while CFI-65 and CFI-85 remain totally insoluble **(B)**. A total of 300 ng of purified 6His tagged Celf_1230 was loaded as a western blot control, while 25 µg of protein from cell lysates was loaded per lane.

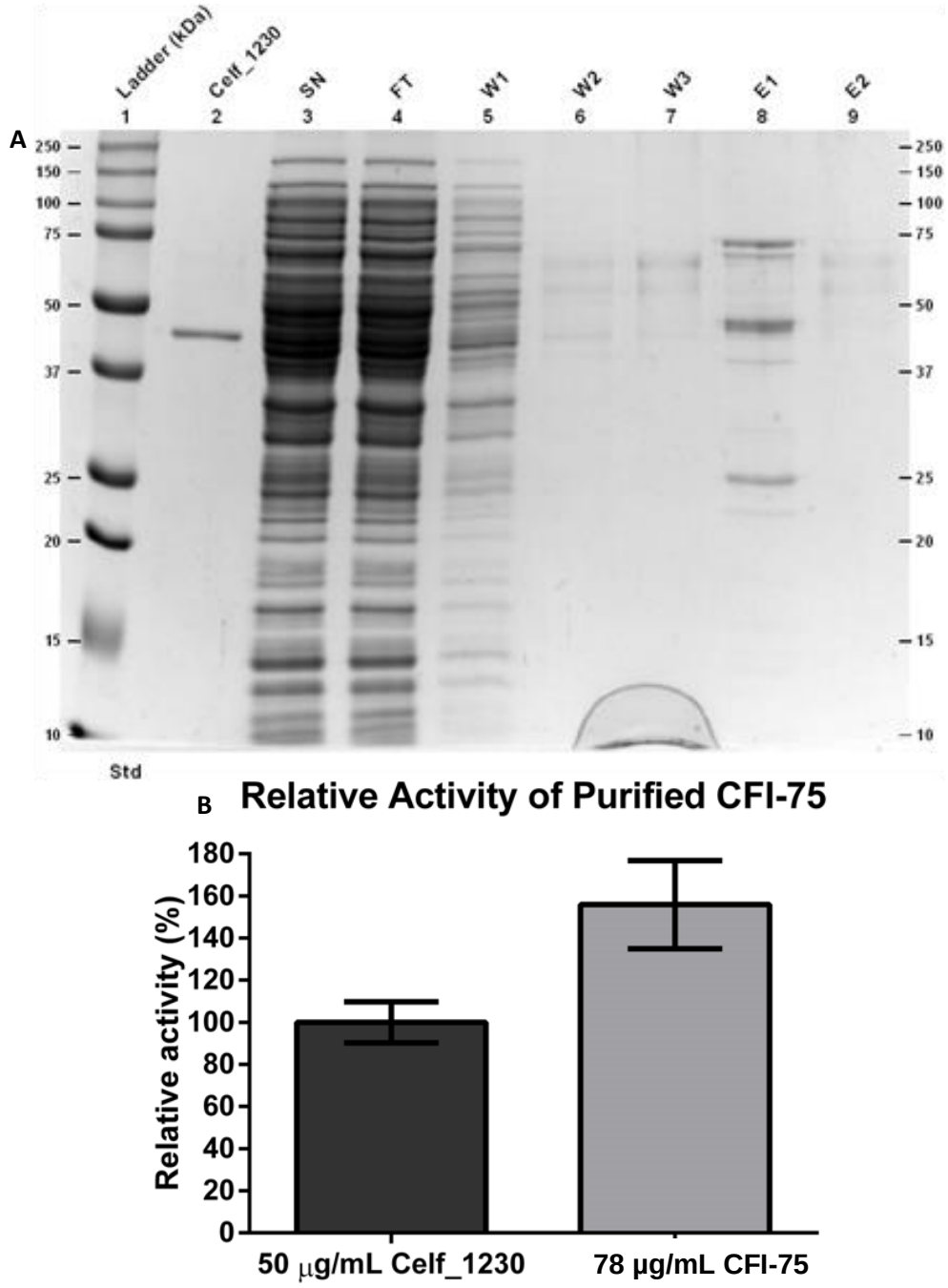



**Figure 3.3: Batch purification of *E. coli* produced CFI-75 with $Ni^{+}$ resin and relative activity on AZCL-Barley β-glucan.** Following purification **(A)** and buffer exchange, protein recovery was determined by testing relative activity against purified Celf_1230 **(B)** with no detectable activity in either the flow-through or wash fractions. In total 780 µg of CFI-75 was purified. SN: lysate supernatant, FT: column flow-through, W1 – W3: wash 1 – wash 3, E1 – E2: elution 1 – elution 2.

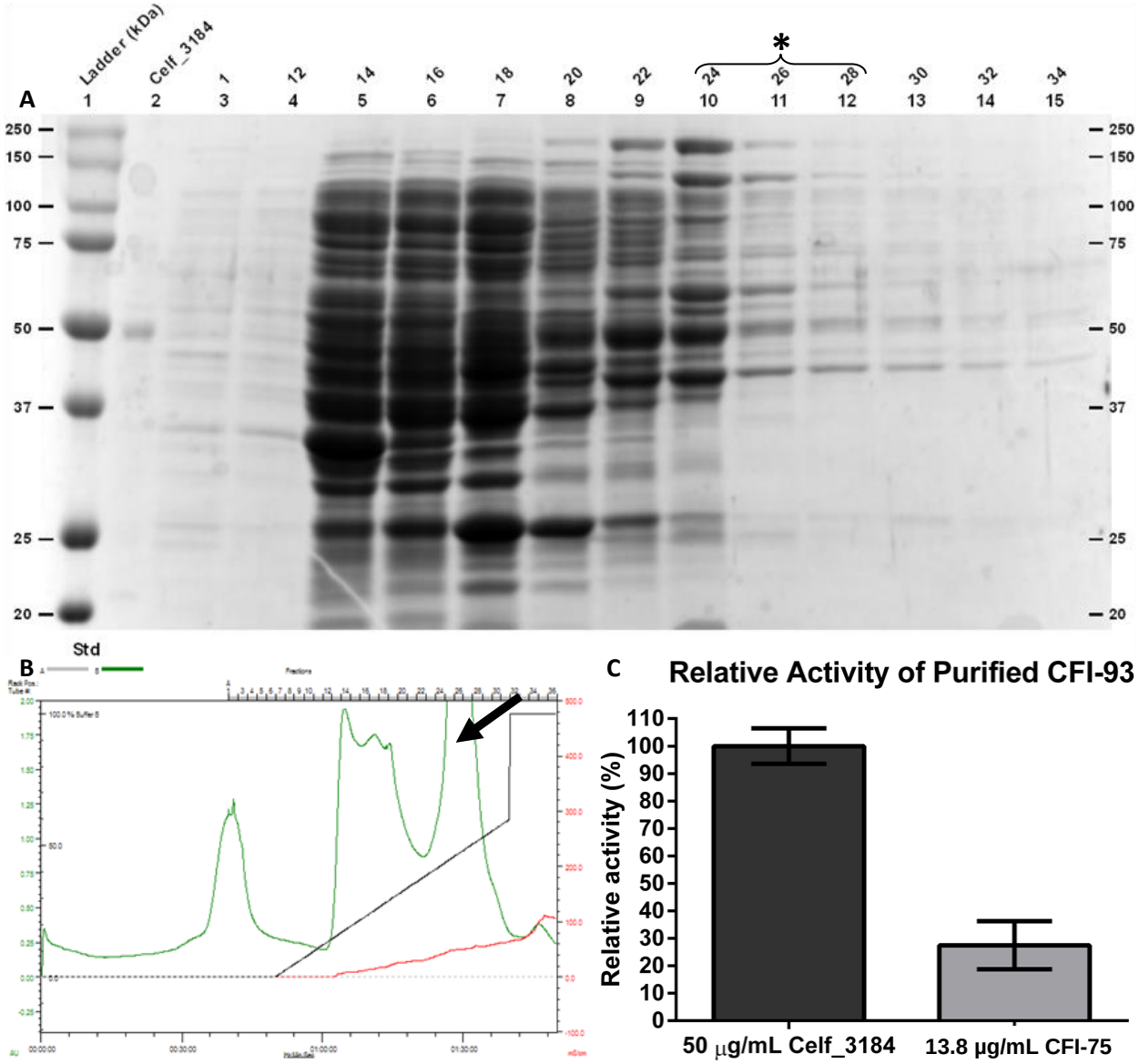



**Figure 3.4: Anion exchange of *E. coli* expressed CFI-93 using 5 mL cartridge of SourceQ resin and its relative activity on AZCL-Barley β-glucan.** Following purification **(A and B)** fractions with activity were pooled and buffer exchanged. Protein recovery was determined by testing relative activity against Celf_3184 **(C)**. In total 43 µg of CFI-93 was purified. Fractions marked by an asterisk contained activity and were subsequently pooled. Black arrow denotes peak on chromatogram containing activity.

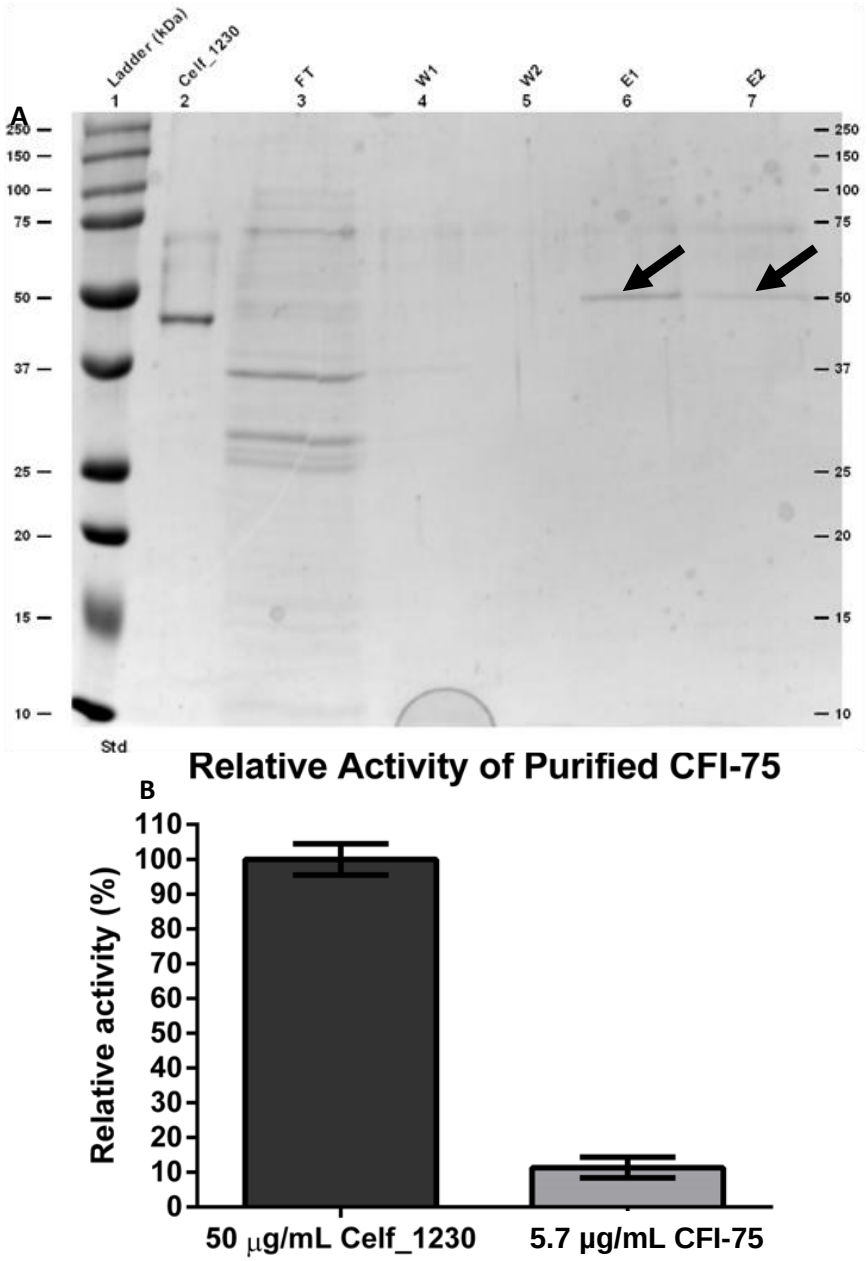



**Figure 3.5: Batch purification of CFI-75 from *C. glutamicum* culture supernatant with $Ni^{+}$ resin and its relative activity on AZCL-Barley β-glucan.** Following purification **(A)** and buffer exchange, protein recovery was determined by testing relative activity against Celf_1230 **(B),** in total 25 µg of CFI-75 was purified. Activity in culture supernatants was only detectable after 32 hour reactions. Black arrows denote purified CFI-75. FT: column flow-through, W1 – W2: wash 1 – wash 2, E1 – E2: elution 1 – elution 2.

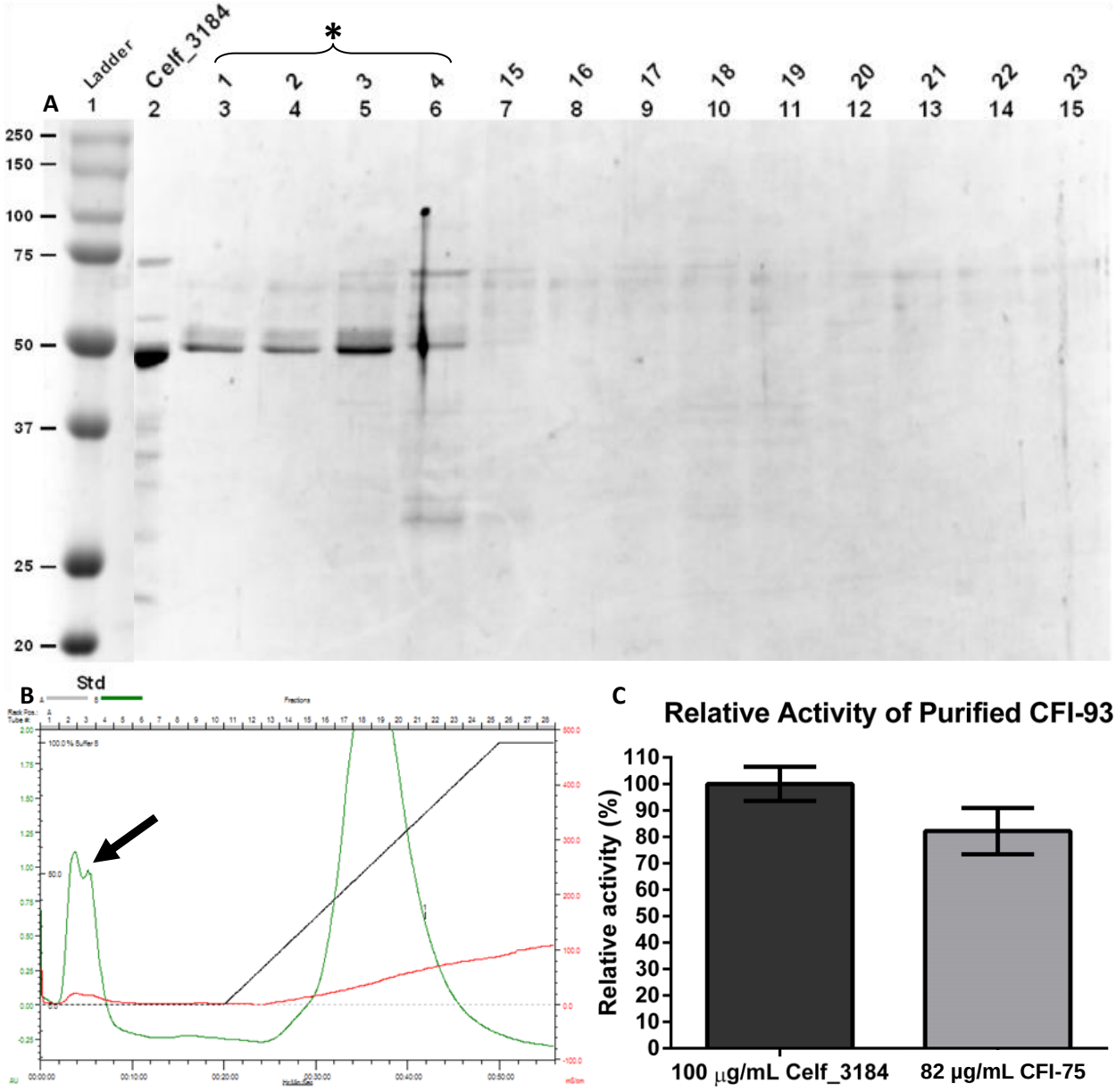



**Figure 3.6: Anion exchange of CFI-93 from *C. glutamicum* culture supernatant using 5 mL cartridge of SourceQ resin and its relative activity on AZCL-Barley β-glucan.** Following purification **(A and B)** fractions with activity were pooled and buffer exchanged. Protein recovery was determined by testing relative activity against Celf_3184 **(C).** In total 350 µg of CFI-93 was purified. Fractions marked by an asterisk contained activity and were subsequently pooled. Black arrow denotes peak on chromatogram containing activity.

A

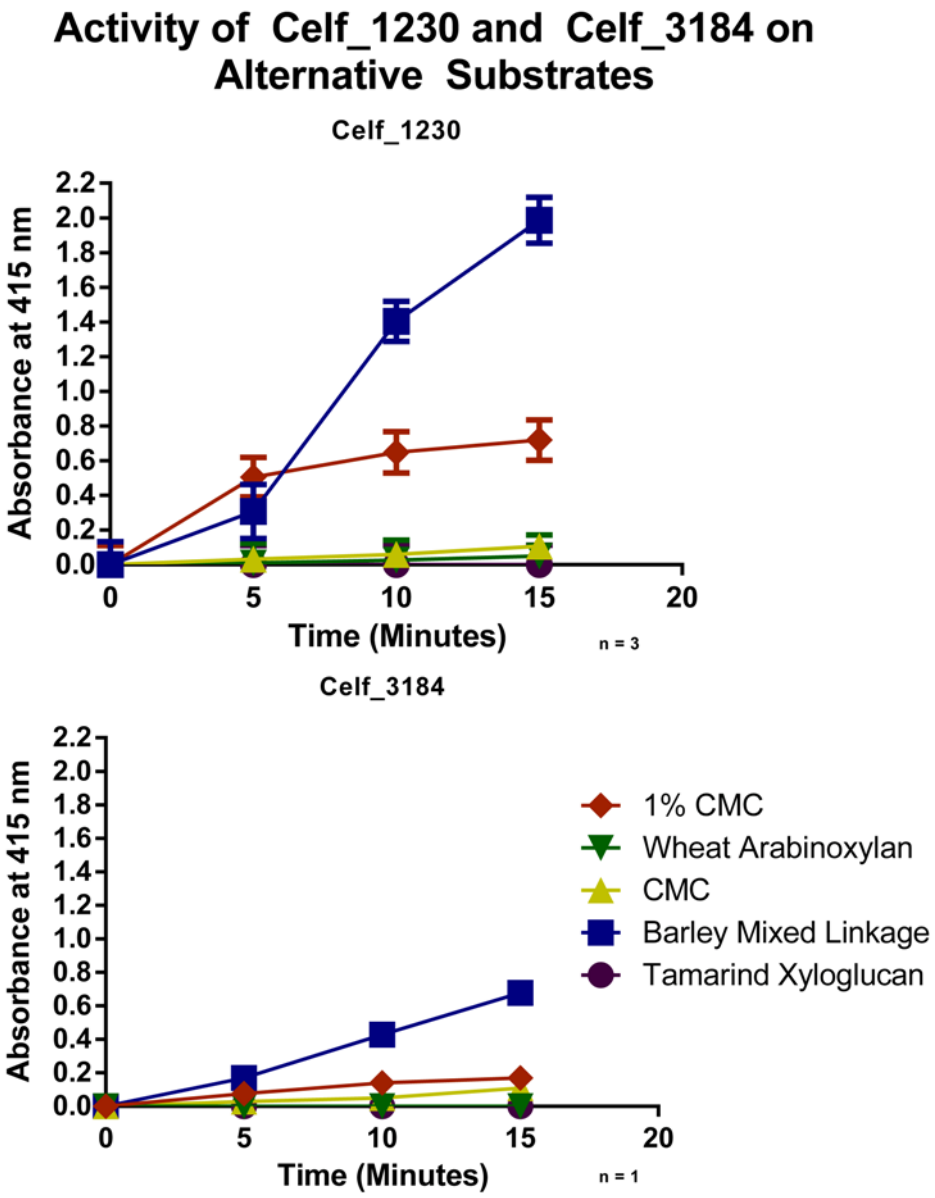



**Figure 3.7: Activity of Celf_1230 and Celf_3184 on alternative substrates.** All substrates assayed at 0.5% unless otherwise specified. Both Celf_1230 **(A)** and Celf_3184 **(B)** display activity on 1,3:1,4-Barley β-glucan that is significantly greater than on the classical glycoside hydrolase substrate CMC. Only one replicate is shown for Celf_3184, but the observed trend was similar across all data sets.

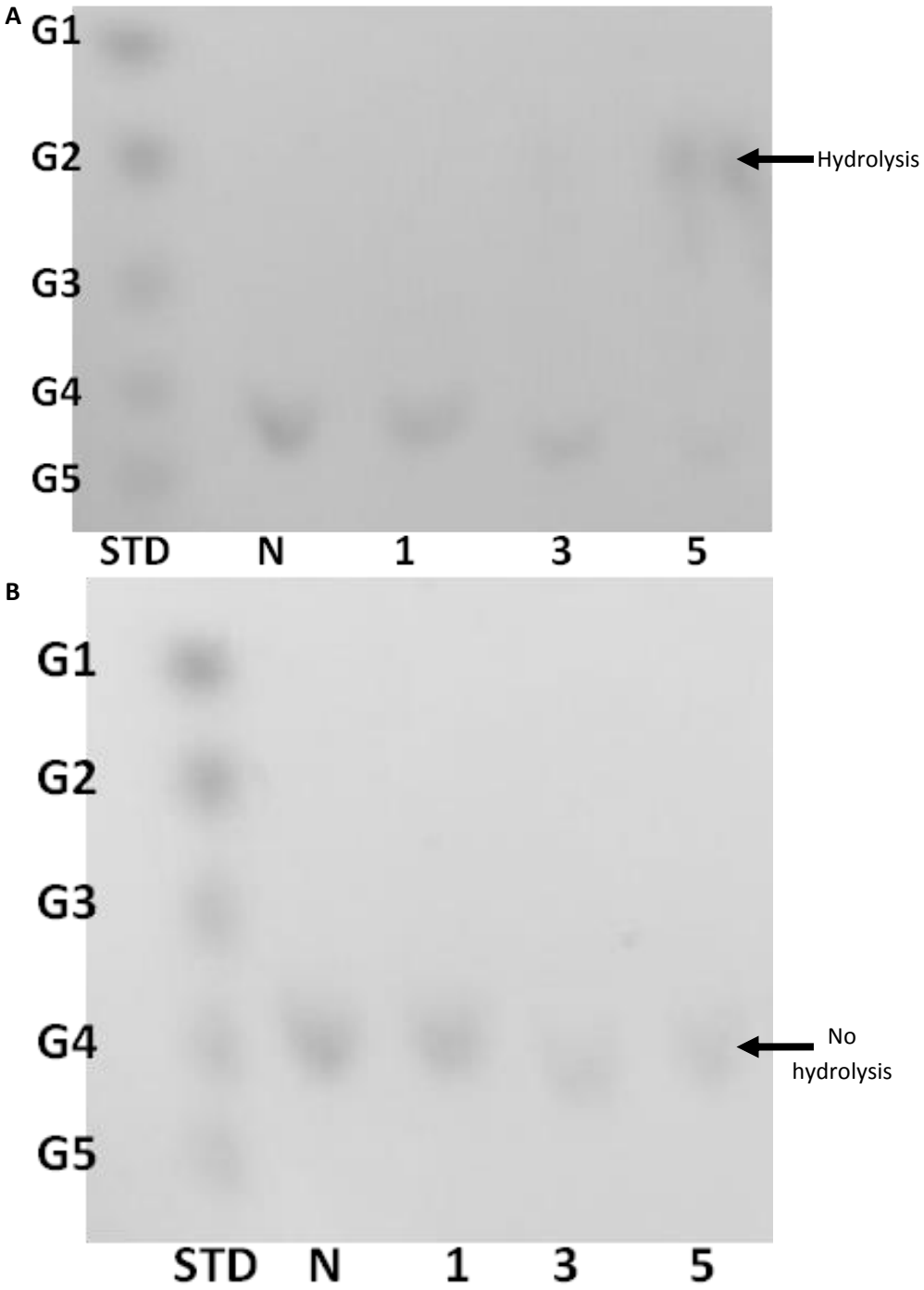



**Figure 3.8: TLC of Celf_1230 hydrolysis products on cellotetraose and 1,3:1,4-β-glucotetraose.**

Celf_1230 hydrolyzes cellotetraose (G4) to cellobiose (G2) rapidly **(A)** but does not hydrolyze 1,3:1,4-β-glucotetraose **(B)**. Time-points in minutes, N: negative reaction, G1: glucose, G2: cellobiose, G3: cellotriose, G4: cellotetraose, G5: cellopentaose.

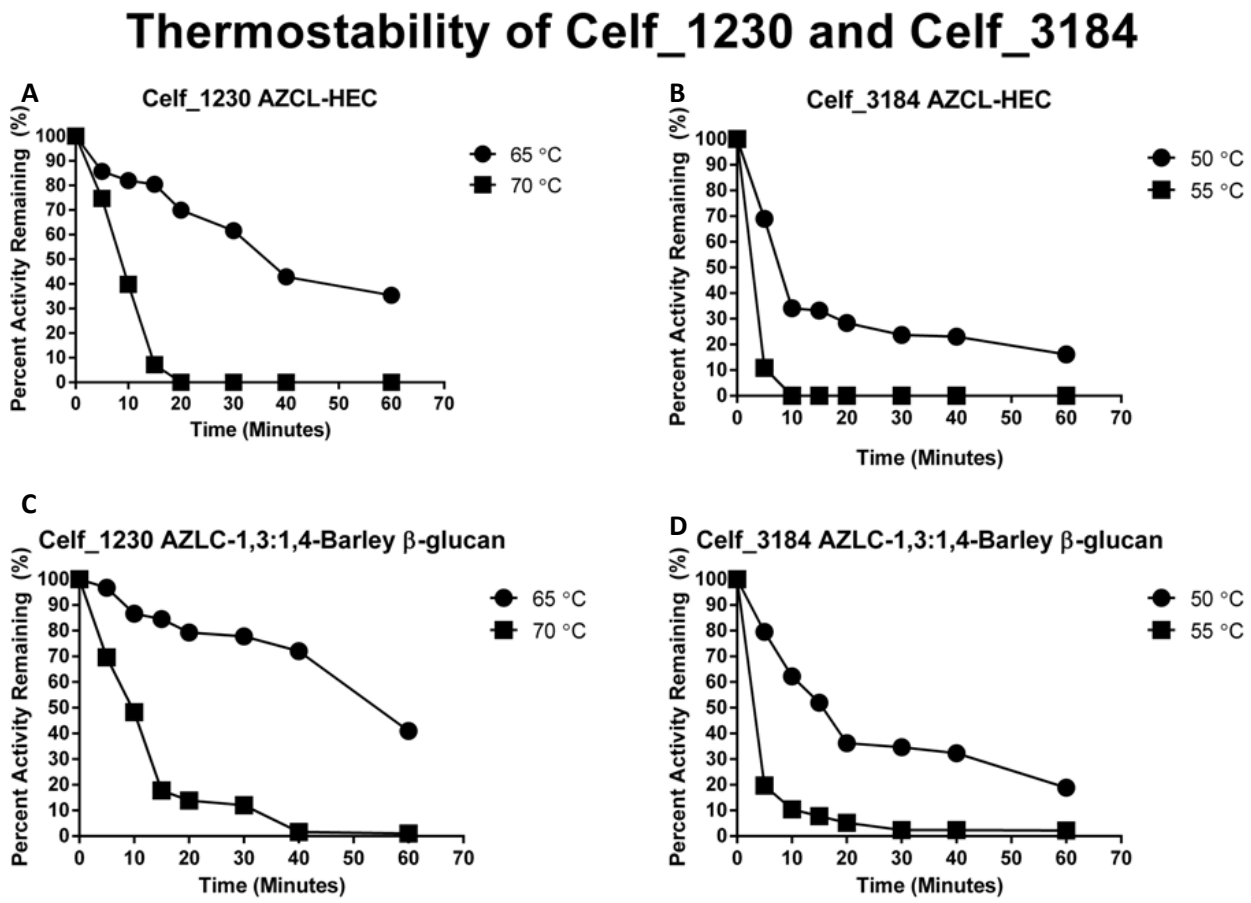



**Figure 3.9: Thermostability of Celf_1230 and Celf_3184 at optimal temperature on CMC and 1,3:1,4-Barley β-glucan.** The thermostabilities of Celf_1230 (left panels) and Celf_3184 (right panels) on both AZCL-HEC (top panels) and AZCL-1,3:1,4-Barley β-glucan (bottom panels). The thermostability profile of Cel6C is unique among characterized glycoside hydrolases originating from *C. fimi*. Displayed data is of a representative data set; however the observed trend is similar across all sets.

# Chapter 4

# Discussion

## 4.1 TAT dependent export

The utilization of a TAT signal peptide during heterologous protein expression in *E. coli* is in itself not novel; however, co-opting this export pathway is an emerging facet of certain Gram-positive expression systems [61, 67, 59, 57]. This work demonstrated two key advantages in using TAT dependent export when expressing heterologous proteins in both *E. coli* and *C. glutamicum*. It is understandable that individuals unfamiliar with TAT dependent export may expect most or all expressed heterologous protein to be located in the periplasm. However, as demonstrated here and in other works [50, 59], that is not the case. In reality, the overall quantity and distribution of expressed protein is nearly identical to the control vector lacking the TAT signal peptide. The primary difference is in the proportion of soluble active protein between vectors with and without the TAT signal peptide. In the case of eGFP expression with the TorA TAT signal peptide, significantly more fluorescence was detected in the cytoplasmic fraction than in the corresponding fraction of the control vector. The TAT export pathway accomplishes this by folding expressed proteins cytoplasmically, where they are able to associate with (predominately) cytoplasmic chaperones and cofactors [50]. The association of cytoplasmic cofactors and chaperones is an important factor in maximizing activity in disulfide bond containing proteins (like eGFP) expressed in *E. coli* strains like Shuffle Express™. This strain has mutations which provide an oxidizing redox potential in the cytoplasm (ΔtrxB and Δgor) in addition to expressing a cytoplasmic disulfide bond isomerase (DsbC) which helps promote the formation of disulfide bonds. Therefore, it is beneficial to have multi-disulfide bond containing proteins fold in the oxidative cytoplasm of similar strains [6]. The improved cytoplasmic fluorescence of expressed eGFP shows the TAT pathway folds proteins cytoplasmically prior to export. Furthermore, the utilization of TAT dependent export in a Gram-positive expression host allows for expressed proteins to be secreted out of the cell (reaching concentrations of 21 mg/L in the case of

GFP), directly into the culture supernatant since Gram-positive organisms lack a traditional periplasmic space [61, 57]. This greatly simplifies the harvest and purification of expressed proteins.

**4.2 Glycoside hydrolase analysis**

*4.2.1 Celf_0233 and Celf_2403*

Neither Celf_0233 nor Celf_2403 were produced in a soluble, active manner in *E. coli*, the entirety of the expressed protein is completely insoluble and inactive. Unfortunately, no heterologous protein production was observed in *C. glutamicum* either. Though more information needs to be gathered in order to accurately determine why these proteins would not express in the Gram-positive expression system, a few hypotheses are currently possible. It was initially thought that the secretion signal differences between the native protein and the TAT sequence utilized by the expression system could account for the shortcomings. It is possible that a protein destined for Sec dependent secretion is not compatible with TAT dependent secretion. However, this is likely not the case since the native Celf_1230 protein has a Sec leader sequence and was produced in detectable quantities in *C. glutamicum* (Figure 3.5). It is much more probable that the expression issues arose from the fact that the *C. glutamicum* expression system still requires a significant amount of optimization. Just like in *E. coli*, growth and induction conditions are intimately related to the stability and structure of individual proteins. Conditions such as optical density at induction, IPTG concentration, induction length and temperature, growth supplements, and using other stains of *C. glutamicum* still need to be checked when attempting to express these proteins before it can be stated that they do not express in the *C. glutamicum* expression system.

*4.2.2 Celf_1230*

The native celf_1230 gene expressed well in both *E. coli* and *C. glutamicum*. Though a significant portion of the *E. coli* produced protein was insoluble, some active protein was still produced. Celf_1230 itself is an intriguing protein, with an optimal temperature of 65°C (Figure 3.9) and originating from a soil bacteria whose optimal growth temperature is only 30°C. This could make Celf_1230 a good candidate for future study and possible engineering for industrial purposes. Mixtures of Celf_1230 and Celf_3184 were previously shown to only have additive activity on cellulose [27] , further study on other types of activity would need to be assayed for synergistic action. Other than the activity of Celf_1230 on 1,3:1,4-Barley β-glucan, its lack of activity on alternate substrates was not surprising as the GH6s only had endoglucanase and cellobiohydrolase activities reported. It was initially thought possible that the significantly improved activity of Celf_1230 on 1,3:1,4-Barley β-glucan compared to the traditional endoglucanase/cellulase substrate, carboxymethylcellulose (CMC) (Figure 3.7), could have been explained by a novel β-1, 3 hydrolysis activity, though this was later rejected after the analysis of Celf_1230 hydrolysis products of 1,3:1,4-β-glucotetraose. This analysis showed that the enzyme only cleaved at the β-1,4 glucose linkage and not at the β-1,3 glucose linkage (Figure 3.8).

*4.2.3 Celf_3184*

The native Celf_3184 protein has been remarkably troublesome to express in our lab in *E. coli*, where it arbitrarily produced active protein. A significantly low induction temperature of 15°C (in tandem with late induction) were previously used [12] to produce soluble protein in *E. coli*. However, with the same construct in *C. glutamicum,* induction during early exponential growth carried out at 30°C was more than sufficient to produce easily detectable quantities of protein (Figure 3.6). This is likely due to the fact that the native Celf_3184 TAT leader sequence is much more compatible with the *C. glutamicum* machinery than in *E. coli*. Regardless, the amount of produced protein in *E. coli* was so insignificant it

was undetectable as a strongly induced protein by SDS-PAGE analysis (Figure 3.2 A). In *C. glutamicum* however, the protein expresses much better even at induction temperatures 15°C higher than that used in *E. coli* (Figure 3.6 A). Activity data for Celf_3184 falls in line with the literature, as this GH has been previously characterized. Celf_3184 exhibited the same characteristic of improved activity on 1,3:1,4-Barley β-glucan compared to CMC.

**4.3 Carboxymethylcellulose is not a preferred substrate**

Cellulose is itself a crystalline water-insoluble polymer, therefore rendering the traditional reducing sugar assay virtually useless for the measurement of cellulase activity from enzymes that do not hydrolyze crystalline cellulose. For this reason, the highly substituted and water-soluble substrate CMC became standard for assaying cellulase activity. The comparison of activity of Celf_1230 and Celf_3184 on both CMC and 1,3:1,4-Barley β-glucan contrast the traditional zeitgeist of CMC being the "ideal" substrate for assaying endoglucanase activity [62]. In fact, there is an emerging trend of moving away from the utilization of CMC due to its high degree of substitution to maximize its solubility. The current observation is that the level of substitution of CMC adversely affects the activity on enzymes acting on it. It is possible that the introduction of alternating β-1,3 and β-1,4 glycosidic linkages alters the stereochemistry of the carbohydrate polymer, making the substrate more accessible to enzymatic cleavage. This "kinked" carbohydrate polymer is much more similar in structure to hemicellulose, which has a random, amorphous structure. Not to be overlooked however, is the possibility that 1,3:1,4-Barley β-glucan is an actual preferred substrate for these particular GH6s from *C. fimi*.

**4.4 Applications of a Gram-positive expression system**

The utilization of this Gram-positive expression system (in concert with a TAT signal peptide) for the production of heterologous proteins has a number of practical applications. Most notably, its use could

decrease the labour and cost of purifying heterologous proteins, since they would be pumped directly into the culture supernatant. From there, the expressed proteins could easily be purified using a variety of chromatography methods including affinity or ion exchange chromatography or even a simple gel filtration. In the case of *C. glutamicum*, there is the added benefit of the organism not producing any secreted proteases, so heterologous proteins would be safe from degradation in the culture supernatant. This expression system also virtually eliminates the need for cell lysis during the harvest of heterologous proteins.

Our lab plans on using this expression system for genes originating from Gram-positive organisms that are troublesome to produce in *E. coli*. Research in the Wakarchuk lab has recently shown that *C. glutamicum* may be an expression platform to understand Gram positive actinobacterial protein O-mannosylation which was first observed with *C. fimi* cellulases in the 1980's (Nakita Buenbrazo, MSc thesis 2016). So far there have not been many other examples of enzymes from Gram-positive bacteria, but examples in other organisms include cellulases from *Clostridium thermocellum*, and *Streptomyces lividans* [38]. Genes from the companion organism *C. flavigena* which is also being examined in our laboratory are also targets for GH and glycoprotein research.

This expression system is also from an organism "generally regarded as safe" or GRAS. This means that proteins produced in this organism are not contaminated with endotoxin that is a serious concern for proteins produced in *E. coli*. Our lab is very interested in the expression of other CAzymes, notably glycosyltransferases, for the modification of therapeutic proteins and mammalian cell surface glycan remodeling. The use of a GRAS expression system could aid with the production of these enzymes in a form which would be better for biomedical applications than the corresponding *E. coli* produced enzymes.

## 4.5 Future directions

### *4.5.1 Optimization of TAT signal sequence*

In order to optimize both the folding and localization of heterologous proteins in both *E. coli* and *C. glutamicum* further study could be directed towards identifying and using a TAT signal sequence that shows reliable functionality in both organisms across a wide variety of protein types. This work brings to light the Celf_3184 TAT leader sequence as a potential avenue for further study in *C. glutamicum* expression and secretion.

### *4.5.2 Activity of glycosylated glycoside hydrolases and other proteins*

While glycosylation of *C. fimi* GHs has not been shown to be required for enzymatic activity, there is evidence for it affecting protein stability [21].Further comparison of production and stability between the glycosylated and non-glycosylated forms of Celf_1230 and Celf_3184 could determine if the modification has any impact on their application to biomass hydrolysis. Furthermore, a significant amount of protein engineering could be directed towards these two GHs to capitalize on their thermostability and possibly even improve their respective activities. Other possibilities are that engineered glycosylation sites might aid in the secretion and stability of other enzymes produced in *C. glutamicum*.

### *4.5.3 Optimization of Corynebacterium glutamicum expression*

A significant amount of work remains to optimize both the growth and expression conditions for *C. glutamicum*. Primarily, identifying the optimal strain for expression and combination of media conditions like buffers to maintain culture pH, amino acids, urea as a nitrogen source, vitamins like biotin and thiamine, glucose as an energy source, and essential anions and cations as well as trace elements like $FeSO_4$ and $MnSO_4$ could greatly increase the levels of heterologous protein production.

This optimization could easily increase protein production many times over as was seen with the addition of $Ca^{+2}$. Calcium sufficient cultures routinely secreted heterologous proteins at levels over 30-fold greater than in comparison to calcium deficient cultures [57]. Furthermore, identifying the optimal induction conditions would also be beneficial. And finally, determining the kinetics of heterologous protein production following induction and how long it maintains its integrity in the culture supernatant (although this is likely protein dependant) are important parameters to consider.

**4.6 Concluding remarks**

This work provides evidence that the expression of heterologous proteins in *E. coli* (in this case for GHs originating from *C. fimi*) does not always produce proteins with the same characteristics as when they are produced in their native organisms (eg., solubility, activity, stability, and expression levels). It supports the thought that it may be time to revisit some previously characterized GHs from other Gram-positive organisms, since they would have all been produced in *E. coli*. Further modifications to the expression vector, such as the utilization of a *C. fimi* TAT sequence and the addition of a tandem promoter could help to produce secreted soluble protein from troublesome genes. In addition, other stains of *C. glutamicum* could also be investigated in an attempt to increase the efficiency of heterologous protein expression. Following optimization, this expression system could serve as an important adjunct to *E. coli* for characterizing new GHs being identified by genome sequencing of environmental or biomedical isolates of Gram-positive bacteria.

**Chapter 5**

**Appendix**

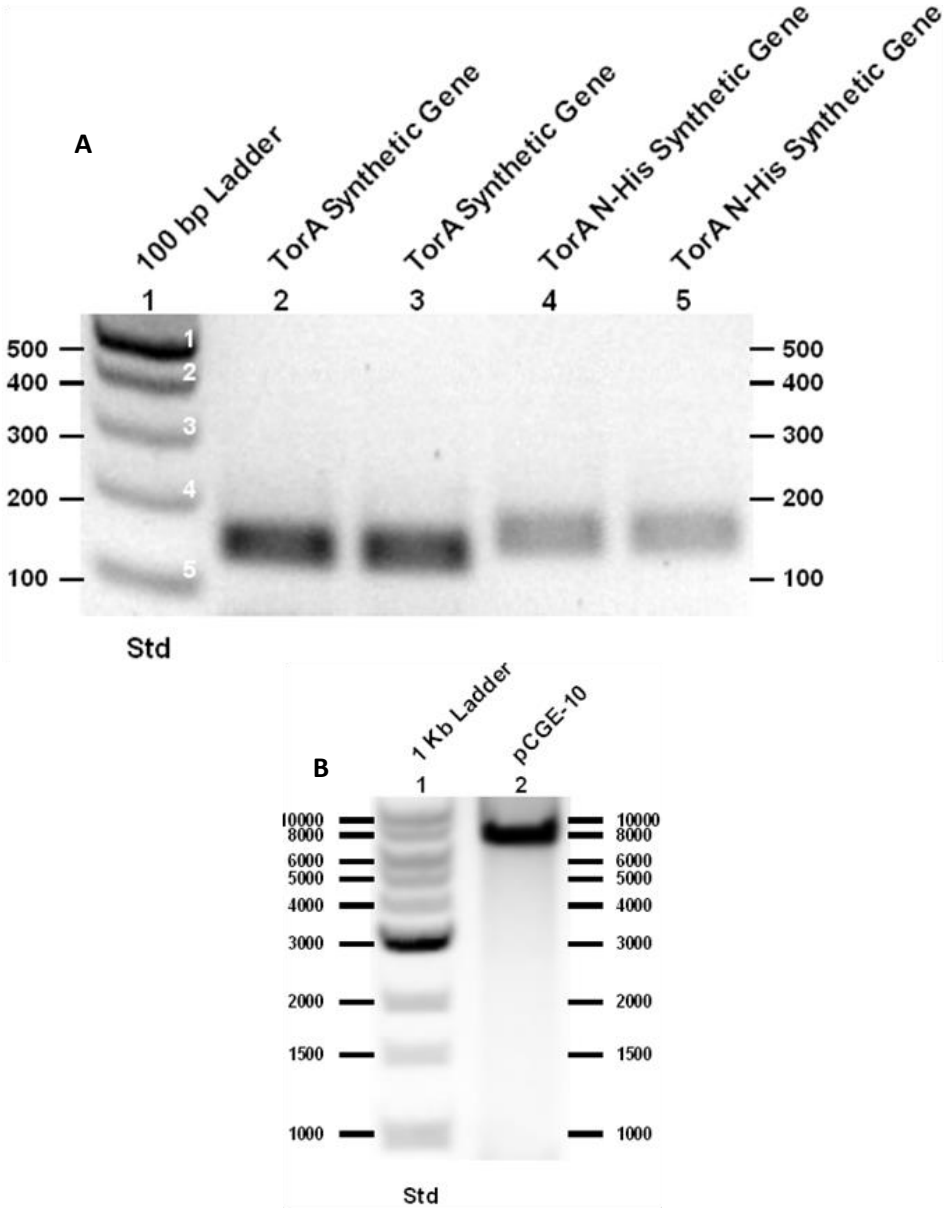



**Figure 5.1: 2.5% agarose gel showing digested TorA synthetic genes (A) and 0.8% agarose gel showing unmodified pCGE-10 (B).** TorA (133 bp) and TorA N-$His_6$ (148 bp) synthetic genes **(A)** that were ligated into the NdeI site of the unmodified pCGE-10 (7519 bp) shuttle vector **(B)**.

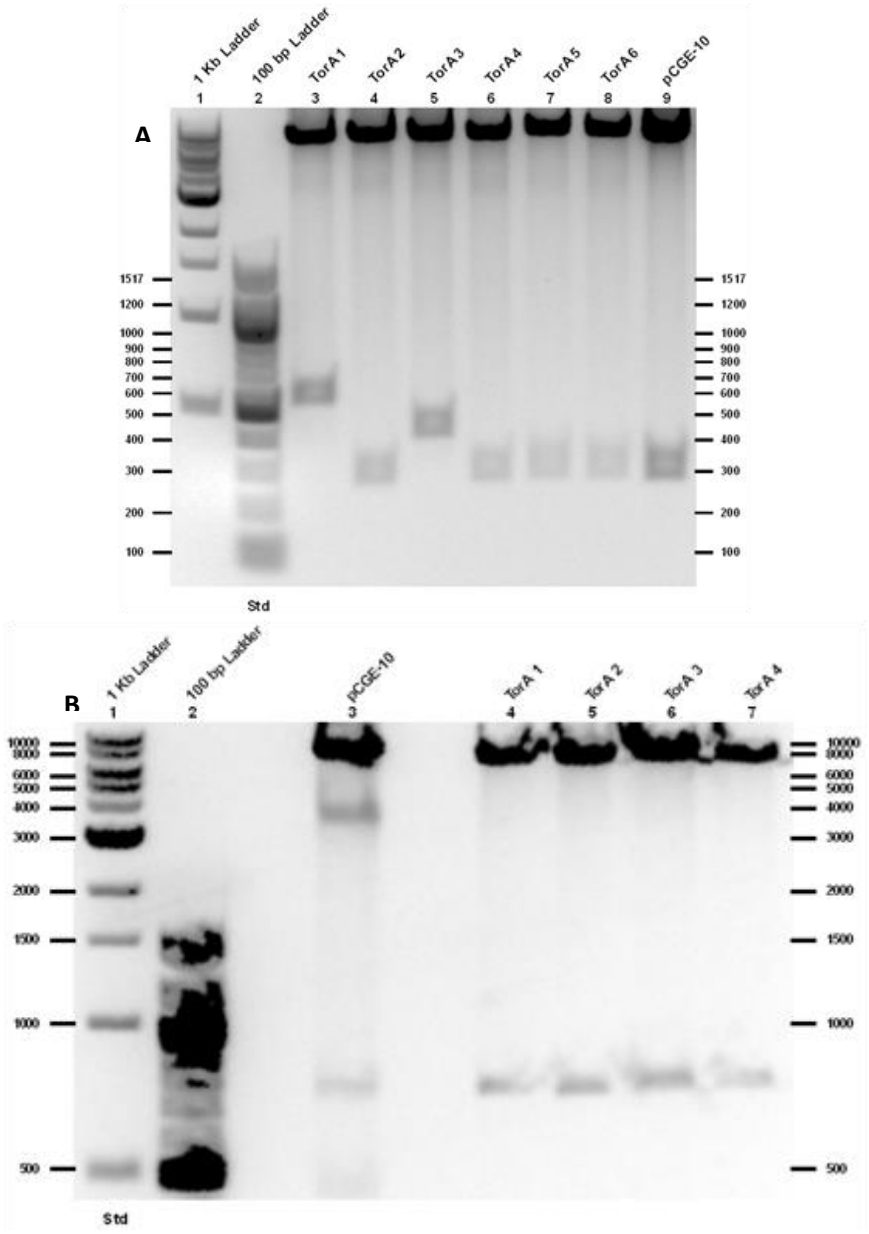



**Figure 5.2: 1.5% agarose gels checking pCGE-15 clones for TorA insertion and synthetic gene orientation.** TorA insertions checked via diagnostic digest with BamHI and NdeI **(A)**. Lane 3 shows a double insertion event and lane 5 shows an insert of the correct size. Orientation of synthetic gene was confirmed by diagnostic digest with NdeI and EcoRI **(B)**. The clone designated "TorA 3" was shown to have the correct insert in the correct orientation and was sequenced for confirmation. Accessory bands in lane 3 are due to incomplete digestion.

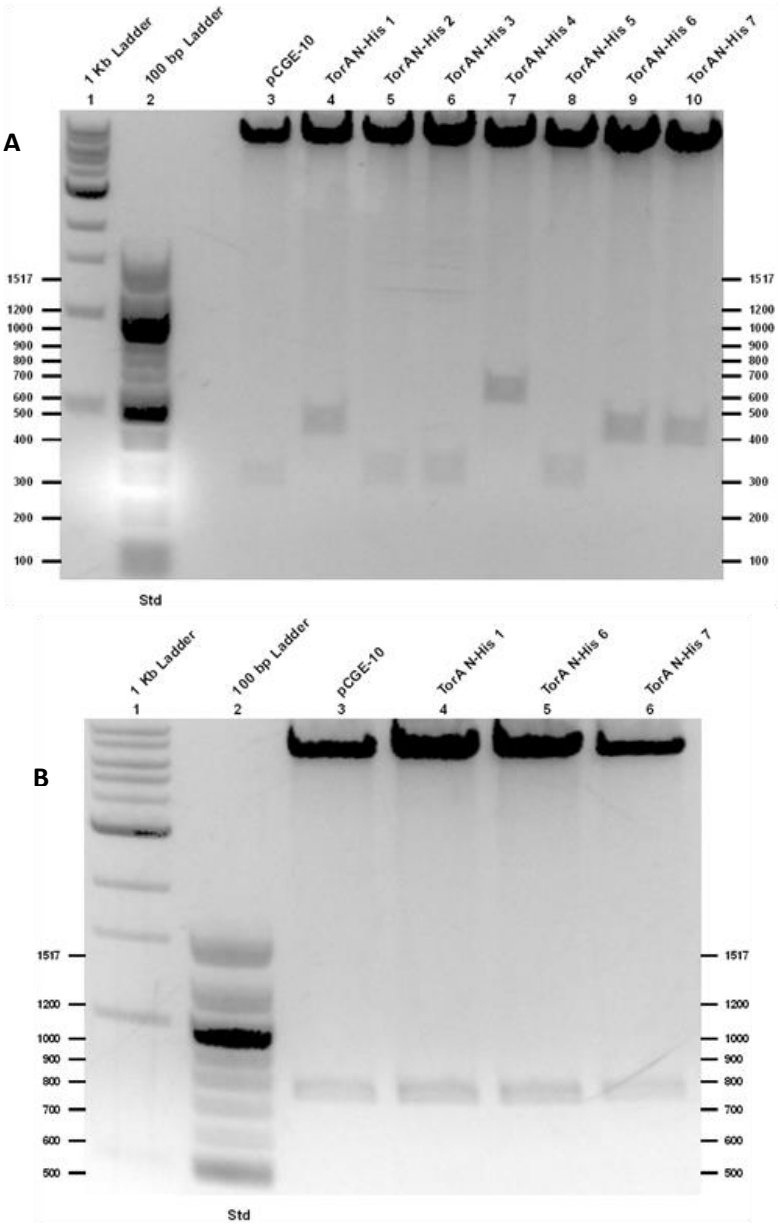



**Figure 5.3: 1.5% agarose gels checking pCGE-20 clones for TorA N-His$_6$ insertion and synthetic gene orientation.** TorA N-His$_6$ insertions checked via diagnostic digest with BamHI and NdeI **(A)**. Lane 7 shows a double insertion event while lanes 4, 9, and 10 show an insert of the correct size. Orientation of synthetic gene was confirmed by diagnostic digest with NdeI and EcoRI **(B)**. The clones designated "TorA N-His$_6$ 1, 6, and 7" were shown to have the correct insert in the correct orientation and were sequenced for confirmation.

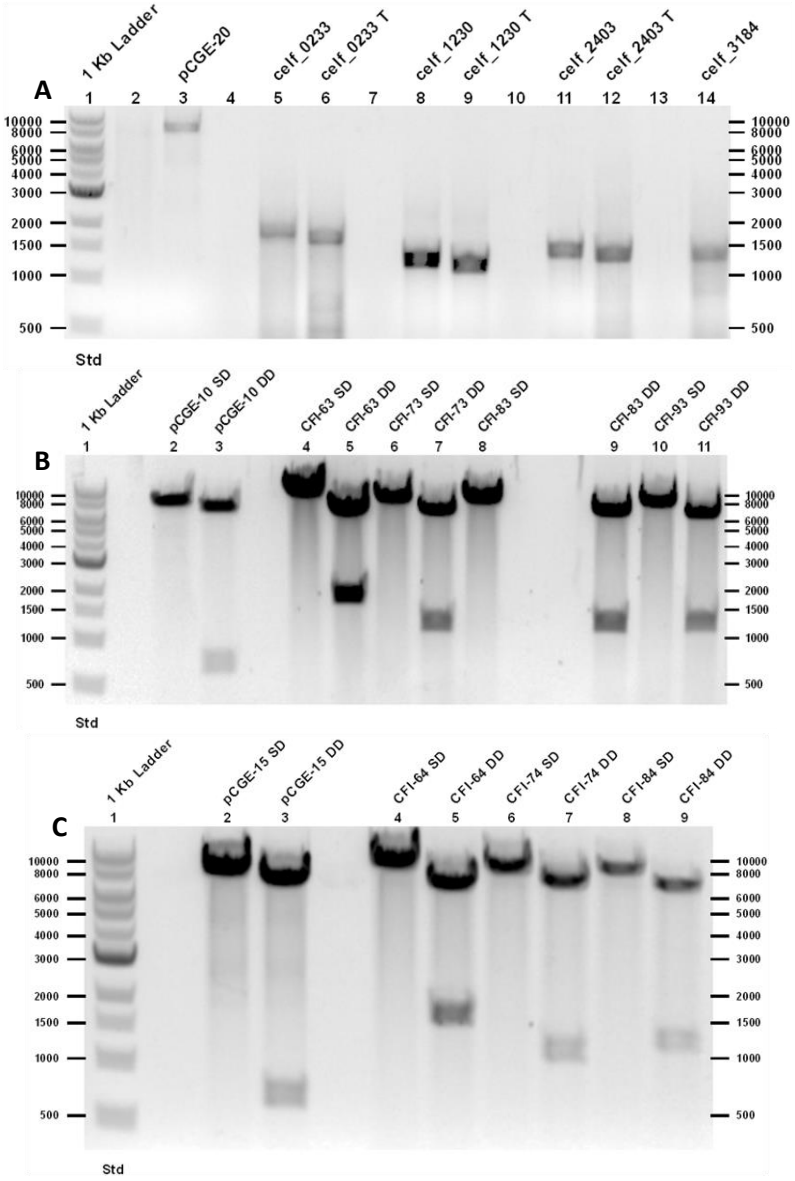



**Figure 5.4: Glycoside hydrolase PCR products and construct library in pCGE-10, pCGE-15 and pCGE-20.**

Ligation reactants **(A)** show amplified native glycoside hydrolases and truncated (T) glycoside hydrolases lacking native leader sequences. Single and double digestions of glycoside hydrolase constructs in pCGE-10 **(B)** with native gene leaders and in pCGE-15 **(C)** lacking native gene leaders. Constructs in pCGE-20 are not shown as they appear identical to pCGE-15 constructs (the only difference being the addition of an N-terminal $His_6$ on inserted gene). The 750 bp band observed in lane 3 of **(A)** and **(B)** is the eGFP stuffer sequence intrinsic to the vector.

**Copyright permissions for figures in this thesis.**

Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

AMERICAN SOCIETY FOR MICROBIOLOGY

**Title:** Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum

**Author:** Marco Ventura, Carlos Canchaya, Andreas Tauch et al.

**Publication:** Microbiology and Molecular Biology Reviews

**Publisher:** American Society for Microbiology

**Date:** Sep 1, 2007

Copyright © 2007, American Society for Microbiology

LOGIN

**If you're a copyright.com user,** you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

**Permissions Request**

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

BACK | CLOSE WINDOW

Copyright © 2016 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS
Jan 27, 2016

This is a License Agreement between Hirak Saxena ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

| | |
|---|---|
| License Number | 3797311336117 |
| License date | Jan 27, 2016 |
| Licensed content publisher | Nature Publishing Group |
| Licensed content publication | Nature Reviews Molecular Cell Biology |
| Licensed content title | Protein targeting by the twin-arginine translocation pathway |
| Licensed content author | Colin Robinson, Albert Bolhuis |
| Licensed content date | May 1, 2001 |
| Volume number | 2 |
| Issue number | 5 |
| Type of Use | reuse in a dissertation / thesis |
| Requestor type | academic/educational |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Figures | Figure 4, Page 354 |
| Author of this NPG article | no |
| Your reference number | None |
| Title of your thesis / dissertation | Development of a Corynebacterium glutamicum Expression System for Glycoside Hydrolase Analysis |
| Expected completion date | Jan 2016 |
| Estimated size (number of pages) | 80 |
| Total | 0.00 CAD |

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the

conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   **Note: For republication from the *British Journal of Cancer*, the following credit lines apply.**
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:

Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   **Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.**
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   **Note: For translation from the *British Journal of Cancer*, the following credit line applies.**
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

**Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Jan 27, 2016

---

---

This Agreement between Hirak Saxena ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

| | |
|---|---|
| License Number | 3797320289360 |
| License date | Jan 27, 2016 |
| Licensed Content Publisher | John Wiley and Sons |
| Licensed Content Publication | Molecular Microbiology |
| Licensed Content Title | Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli |
| Licensed Content Author | Joanna D. Thomas,Richard A. Daniel,Jeff Errington,Colin Robinson |
| Licensed Content Date | Dec 21, 2001 |
| Pages | 7 |
| Type of use | Dissertation/Thesis |
| Requestor type | University/Academic |
| Format | Print and electronic |
| Portion | Figure/table |
| Number of figures/tables | 1 |
| Original Wiley figure/table number(s) | Figure 4 |
| Will you be translating? | No |
| Title of your thesis / dissertation | Development of a Corynebacterium glutamicum Expression System for Glycoside Hydrolase Analysis |
| Expected completion date | Jan 2016 |
| Expected size (number of pages) | 80 |
| Requestor Location | Hirak Saxena<br>92 Muirlands Drive<br><br>Scarborough, ON M1V2B4<br>Canada |

| | |
|---|---|
| | Attn: Hirak Saxena |
| Billing Type | Invoice |
| Billing Address | Hirak Saxena<br>92 Muirlands Drive<br>Scarborough, ON M1V2B4<br>Canada<br>Attn: Hirak Saxena |
| Total | 0.00 CAD |

Terms and Conditions

**TERMS AND CONDITIONS**

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

**Terms and Conditions**

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or

proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

**WILEY OPEN ACCESS TERMS AND CONDITIONS**

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

**The Creative Commons Attribution License**

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

**Creative Commons Attribution Non-Commercial License**

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

**Creative Commons Attribution-Non-Commercial-NoDerivs License**

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial

purposes and no modifications or adaptations are made. (see below)

**Use by commercial "for-profit" organizations**

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

**Other Terms and Conditions:**

**v1.10 Last updated September 2015**

**Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

---

---

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS
Jan 27, 2016

---

---

This is a License Agreement between Hirak Saxena ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

| | |
|---|---|
| License Number | 3797320579208 |
| License date | Jan 27, 2016 |
| Licensed content publisher | Oxford University Press |
| Licensed content publication | Bioinformatics |
| Licensed content title | Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation |
| Licensed content author | Ivica Letunic, Peer Bork |
| Licensed content date | January 1, 2007 |
| Type of Use | Thesis/Dissertation |
| Institution name | None |
| Title of your work | Development of a Corynebacterium glutamicum Expression System for Glycoside Hydrolase Analysis |
| Publisher of your work | n/a |
| Expected publication date | Jan 2016 |
| Permissions cost | 0.00 USD |
| Value added tax | 0.00 USD |
| Total | 0.00 USD |
| Total | 0.00 USD |
| Terms and Conditions | |

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL**

1. Use of the material is restricted to the type of use specified in your order

details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press

reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

**Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

---

## References


1. Abt, B., Foster, B., Lapidus, A., Clum. A., Sun, H., Pukall, R., Lucas, S., Glavina Del Rio, T., Nolan, M., Tice, H., Cheng, J.F., Pitluck, S., Liolios, K., Ivanova, N., Mavromatis, K., Ovchinnikova, G., Pati, A., Goodwin, L., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.J., Jeffries, C.D., Rohde, M., Goker, M., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., & Klenk, H.P. (2010). Complete genome sequence of *Cellulomonas flavigena* type strain (134). *Stand. Genomic Sci.* **3** (1), 15 – 25.

2. Ames, G.F., Prody, C., & Kustu, S. (1984). Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J. Bacteriol.*, **160**, 1181 – 1183.

3. Ardevol, A., & Rovira, C. (2015). Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolasesand glycosyltransferases. Insights from *ab Initio* quantum mechanics/molecular mechanisms dynamic simulations. *J. Am. Chem. Soc.* **137** (24), 7528 – 7547.

4. Berks, B.C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Mol Microbiol* **35**, 260 – 274.

5. Berlemont, R., & Martiny, A.C. (2013). Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* **79** (5), 1545 – 1554.

6. Bessette, P.H., Aslund, F., Beckwith, J., & Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *PNAS*, **96** (24), 13703 – 13708.

7. Blesak, K., & Janecek, S. (2013). Two potentially novel amylolytic enzyme specificities in the prokaryotic glycoside hydrolase α-amylase family GH57. *Microbiology* **159**, 2584 – 2593.

8. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37,** 233 – 238.

9. Casali, N. (2003). *Escherichia coli* host strains. *Methods Mol. Biol.,* **235**, 27 – 48.

10. Chaudhary, P., Kumar, N.N., & Deobagkar, D.N. (1997). The glucanases of *Cellulomonas*. *Biotechnology advances* **15**, 315 – 331.

11. Christopherson, M.R., Suen, G., Bramhacharya, S., Jewell, K.A., Aylward, F.O., Mead, D., & Brumm, P.J. (2013). *PLoS One* **8** (1), e53954.

12. Cockburn, D.W., Vandenende, C., & Clarke, A.J. (2010). Modulating the pH-activity profile of cellulase by substitution: replacing the general base catalyst aspartate with cysteinesulfinate in cellulase A from *Cellulomonas fimi*. *Biochemistry,* **49**, 2042 – 2050.

13. Cline, K., Henry, R., Li, C., & Yuan, J. (1993). Multiple pathways for protein transport into or across the thylakoid membrane. *EMBO J* **12**, 4105 – 4114.

14. Dalbey, R.E., & Von Heijne, G. (1992). Signal peptidases in prokaryotes and eukaryotes – a new protease family. *Trends Biochem Sci* **17**, 474 – 478.

15. Dalbey, R.E., Lively, M.O., Bron, S., & van Dijl, J.M. (1997). The chemistry and enzymology of the type I signal peptidases. *Protein Sci* **6**, 1129 – 1138.

16. Damude, H.G., Withers, S.G., Kilburn, D.G., Miller, R.C., & Warren, R.A (1995). Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. *Biochemistry* **34** (7), 2220 – 2224.

17. Davies, G., & Henrissat B. (1995). Structures and mechanisms of glycoside hydrolases. *Structure* **3**, 853 – 859.

18. Durand, P., Fabrega, S., Henrissat, B., Mornon, J.P., & Lehn, P. (2000). Structural features of normal and mutant human lysosomal glycoside hydrolases deduced from bioinformatics analysis. *Hum. Mol. Genet.* **9** (6), 967 – 977.

19. Gao, J., & Wakarchuk, W. (2014). Characterization of five β-glycoside hydrolases from *Cellulomonas fimi* ATCC 484. *J. Bacteriol* **196** (23), 4103 – 4110.

20. Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., & Warren, R.A. (1991). Domains in microbial β-1, 4-glycanases: sequence conservation, function and enzyme families. *Microbiol Rev.* **55** (2), 303 – 315.

21. Gilkes, N.R., Kilburn, D.G., Miller, R.C., & Warren, R.A. (1989). Structural and functional analysis of a bacterial cellulase by proteolysis. *J. Biol. Chem.* 264 (30), 17802 – 17808.

22. Glazer, A.N., & Nikaido, H. (2007). *Microbial Biotechnology*, 2nd edition. United Kingdom: Cambridge University Press.

23. Henrissat, B., & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J.* **316** (Pt 2), 695 – 696.

24. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., & Davies, G. (1996). Conserved catalytic machinery and the prediction of a common fold for several familes of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA* **93** (11), 5674.

25. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., & Mornon, J.P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81** (1), 83 – 95.

26. Henrissat, B., & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinions in Structural Biology* **7,** 637 – 44.

27. Hsu, B., & Wakarchuk W. (2015). Unsubmitted.

28. Jenkins, J., Lo Leggio, L., Harris, G., & Pickersgill, R. (1995). Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett.* **362** (3), 281 – 285.

29. Jones, D. (1975). A numerical taxonomic study of coryneform and related bacteria. *Journal of General Microbiology* **87** (1), 52 – 96.

30. Khan, N.A. (2009). *Acanthamoeba: Biology and Pathogenesis*. Norfolk, UK: Horizon Scientific Press.

31. Khurana, S., Sanli, G., Powers, D.B., Anderson, S., & Blaber, M. (2000). Molecular modeling of substrate binding in wild-type and mutant Corynebacteria 2, 5-diketo-D-gluconate reductases. *Proteins: Structure, Function, and Bioinformatics* **39** (1), 68 – 75.

32. Klemm, D., Heublein, B., Fink, H.P., & Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angew. Chem. Int. Ed.* **44** (22), 3358 – 3393.

33. LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., & McCoy, J.M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* **11** (2), 187 - 93.

34. Letunic, I., & Bork, P. (2007). Interactive tree of life (iTOL). *Bioinformatics* **23** (1), 127 – 128.

35. Lever, M. (1973). Colorimetric and fluorometric carbohydrate determination with *p*-hydroxybenzoic acid hydrazide. *Biochemical Medicine,* **7**, 274 – 281.

36. Liu, Q.P., Sulzenbacher, G., Yuan, H., Bennet, E.P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Levery, S.B., White, T., Neveu, J.M., Lane, W.S., Bourne, Y., Olsson, M.L., Henrissat, B., Clausen, H. (2007). Bacterial glycosidases for the production of universal red blood cells. *Nature biotechnol.* **25**, 454 – 464.

37. Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., & Bernard, H. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**, 490 – 495.

38. Lommel, M. & Strahl, S. (2009). Protein *O*-mannosylation: Conserved from bacteria to humans. *Glycobiology* **19** (8), 816 – 828.

39. Madigan, M.T., Martinko, J.M., Dunlap, P.V., & Clark, D.P. (2009). *Brock Biology of Microorganisms*, 12th edition. San Francisco: Pearson.

40. Martin, J.F. (1989). Molecular genetics of amino acid-producing *corynebacteria*. In: Microbial Products. New Approaches (Braumberg, S., Hunter, J., & Rhodes, M., Eds.) 25 – 59, Society for General Microbiology Symposium **44**, Cambridge: Cambridge University Press.

41. Mori, H., & Cline, K. (2001). Post-translational protein translocation into thylakoids by the Sec and DpH-dependent pathways. *Biochim Biophys Acta* **1541**, 80 – 90.

42. Naumoff, D.G. (2011). Hierarchical classification of glycoside hydrolases. *Biochemistry* **76** (6), 622 – 635.

43. Neufeld, E. F. (1991). Lysosomal storage diseases. *Annu. Rev. Biochem.*, **60**, 257-280.

44. Owolabi, J.B., Beguin, P., Kilburn, D.G., Miller, R.C., & Warren, R.A.J. (1988). Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. *Applied and Environmental Microbiology* **54** (2), 518 – 523.

45. Paetzel, M., Karla, A., Strynadka, N.C., & Dalbey, R.E. (2002). Signal peptidases. *Chem Rev* **102**, 4549 – 4580.

46. Paradis, F.W., Warren, R.A., Kilburn, D.G., & Miller, R.C. (1987). The expression of *Cellulomonas fimi* cellulase genes in *Brevibacterium lactofermentum*. *Gene* **61** (2), 199 – 206.

47. Pohlschroder, M., Prinz, W.A., Hartmann, E., & Beckwith, J. (1997). Protein translocation in the three domains of life: variations on a theme. *Cell* **91**, 563 – 566.

48. Rajan, S.S., Yang, X., Collart, F., Yip, V.L., Withers S.G., Varrot, A., Thompson, J., Davies, G.J., & Anderson W.F. (2004). Novel catalytic mechanism of glycoside hydrolysis bases on the structure of an NAD+/Mn2+ -dependent phopho-alpha-glucosidase from *Bacillus subtilis*. *Structure* **12**, 1619 –

1629.

49. Riezman, H. (1997). The ins and outs of protein translocation. *Science* **278**, 1728 – 1729.

50. Robinson, C., & Bolhuis, A. (2001). Protein targeting by the twin-arginine traslocation pathway. *Nat. Rev. Mol. Cell. Biol.* **2** (5), 350 – 356.

51. Rouvinen, J., Jones, T.A., Bergfors, T., Knowles, J.K., & Teeri, T. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* **249** (4967), 380 – 386.

52. Santini, C.L., Ize, B., Chanal, A., Müller, M., Giordano, G., & Wu, LF. (1998). A novel sec-independent periplasmic protein translocation pathway in Escherichia coli. *The EMBO Journal* **17** (1), 101 – 112.

53. Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B.C., & Palmer, T. (1998) Overlapping functions of components of a bacterial Secindependent export pathway. *EMBO J* **17**, 3640 – 3650.

54. Schatz, G., & Dobberstein, B. (1996). Common principles of protein translocation across membranes. *Science* **271**, 1519 – 1526.

55. Shah, N., Kuntz, D.A., & Rose, D.R. (2008). Golgi α-mannosidase II cleaves two sugars sequentially in the same catalytic site. *PNAS* **105** (28), 9570 – 9575.

56. Telke, A.A., Zhuang, N., Ghatge, S.S., Lee, S.H., Shah, A.A., Khan, H., Um, Y., S, H.D., Chung, Y.R., Lee, K.H., & Kim, S.W. (2013). Engineering of family-5 glycoside hydrolase (Cel5A) from an uncultured bacterium for efficient hydrolysis of cellulosic substrates. *PLoS One* **8** (6), e65727.

57. Teramoto, H., Wantanabe, K., Suzuki, N., Inui, M., & Yukawa, H. (2011). High yield secretion of heterologous proteins in *Corynebacterium glutamicum* using its own Tat-type signal sequence. (2011). *Appl Microbiol Biotechnol* **91**, 677 – 687.

58. Terwisscha van Sheltinga, A., Armand, S., Kalk, K.H., Isogai, A., Henrissat, B., & Dijkstra, B.W. (1995). Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and x-ray structure of a complex with allosamidin. Evidence for substrate assisted catalysis. *Biochemistry* **34**, 15619 – 15623.

59. Thomas, J.D., Daniel, R.A., Errington, J., & Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. *Molecular Microbiology* **39** (1), 47 – 53.

60. Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., & van Dijl, J.M. (2000). Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. Microbiol Mol Biol Rev **64**, 515 – 547.

61. Tsuchidate, T., Tateno, T., Okai, N., Tanaka, T., Ogino, C., & Kondo, A. (2011). Glutamate production from β-glucan using endoglucanase-secreting *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **90**, 895 – 901.

62. Uhlig, H. (1998). *Industrial Enzymes and Their Applications*. New York: John Wiley & Sons.

63. Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., & van Sinderen, D. (2007). *Microbiol. Mol. Biol. Rev.* **71** (3), 495 – 548.

64. Verastegui, Y., Cheng, J., Engel, K., Kolczynski, D., Mortimer, S., Lavigne, J., Montalibet, J., Romantsov, T., Hall, M., MConkey, B.J., Rose, D.R., Tomashek, J.J., Scott, B.R., Charles, T.C., & Neufeld, J.D. (2014). Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. *mBio* **5** (4), e01157-14.

65. Vuong, T.V., & Wilson, D.B. (2010). Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnology and Bioengineering* **107** (2), 195 – 205.

66. Wakarchuk, W.W., Brochu, D., Foote, S., Robothan, A., Saxena, H., Erak, T., & Kelly, J. (2015). Submitted.

67. Wantanabe, K., Tsuchida, Y., Okibe, N., Teramoto, H., Suzuki, N., Inui, M., & Yukawa, H. (2009). Scanning the *Corynebacterium glutamicum* R genome for high-efficiency secretion signal sequences. *Microbiology* **155**, 741 – 750.

68. Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams K.L., Appel, R.D., & Hochstrasser D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology*, **112**, 531 – 552.

69. Woese, C.R. (1987). Bacterial evolution. *Microbiological reviews* **51** (2), 221.

70. Wong, W.K.R., Gerhard, B., Guo, Z.M., Kilburn, D.G., Warren, R.A.J., & Miller, R.C. (1986). Characterization and structure of an exoglucanase gene *cenA* of *Cellulomonas fimi*. *Gene* **44**, 315 – 324.

71. Yen, M.R., Tseng, Y.H., Nguyen, E.H., Wu, L.F., & Saier, M.H. (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch Microbiol* **177**, 441 – 450.