THE UNIVERSITY *of* EDINBURGH

The Degradation of Cellulosic Material by *Cellulomonas fimi*

Steven Kane

BSc (Hons), MRes

PhD School of Biological Sciences
The University of Edinburgh
2015

Acknowledgements:

I would like to thank:
the BBSRC for the funding to allow me to do this research; Dr Chris French, my supervisory committee, Dr Bruce Ward, Dr Andrew Free and Dr Maurice Gallagher for constructive feedback and help when asked; members of any lab who has kindly lent me their time and resources in aid. Special thanks to those friends who have kept me sane and happy. I'd like to especially thank my wife for putting up with me throughout.

I declare that this work is entirely my own unless explicitly stated as being that of another.

Steven Kane

ABSTRACT

The world stocks of fossil fuels are dwindling and may be all but out before the end of the century. Despite this there is increasing demand for them to be used for transport, and the ever increasing green house gases which their use produces. Renewable and less environmentally damaging forms of fuel are needed. Biofuels, particularly bioethanol, are a possibility to subsidise or replace fossil fuels altogether. Ethanol produced from fermentation of starch sugars from corn are already in wide use. As this bioethanol is currently produced from crops such as corn and sugar cane, that puts fuel crops in direct competition for space and resources with food crops. This has led to increases in food prices and the search for more arable land. Hydrolysis of lignocellulosic biomass, a waste by-product of many industries, to produce the sugars necessary for ethanol production would ease many of the problems with current biofuels. Degradation of lignocellulose is not simple and requires expensive chemical pre-treatments and large quantities of enzymes usually from fungal species making it about 10 times more expensive to produce than corn starch bioethanol. The production of a consolidated bioprocessor, an organism able to degrade, metabolise and ferment cellulosic material to produce ethanol or other useful products would greatly reduce the cost currently associated with lignocellulosic biofuel.

Cellulomonas fimi ATCC 484 is an actinomycete soil bacterium able to degrade efficiently cellulosic material. The US Department of Energy (DOE) released the genome sequence at the start of 2012. In this thesis the released genome has been searched, for genes annotated as encoding polysaccharide degrading enzymes as well as for metabolic pathways. Over 100 genes predicted to code for polysaccharide hydrolysing enzymes were identified. Fifteen of these genes have been cloned as BioBricks, the standard synthetic biology functional unit, expressed in *E. coli* and *C. freundii* and assayed for endo β-1,4-glucanase activity using RBB-CMC, endo β-1,4-xylanase activity using RBB-xylan, β-$_D$-xylosidase activity using ONPX, β-$_D$-cellobiohydrolase activity using ONPC and α-$_L$-arabinofuranosidase activity using PNPA. Eleven enzymes not previously reported from *C. fimi* were identified as active on a substrate with the strongest activities being for 2 arabinofuranosidases (AfsA+B), 4 β-xylosidases (BxyC, BxyF, CelE and XynH), an endoglucanase (CelA), and 2 multifunctional enzymes CelD and XynF, active as cellobiohydrolases, xylosidases and

endoxylanases. Four enzymes were purified from *E. coli* cell lysates and characterised. It was found that AfsB has an optimum activity at pH 6.5 and 45ºC, BxyF has optimum activity at pH 6.0 and 45ºC and XynH has optimum activity at pH 9.0 and 80ºC. XynF exhibited different optima for the 3 substrates with pH 6.0 and 60ºC for ONPC, pH 4.5 and 50ºC for ONPX and pH 5.5 and 40ºC for RBB-xylan.

Searching the genome and screening genes for activities will help genome annotation in the future by increasing the number of positively annotated genes in the databases. The BioBrick format is well suited for rapid cloning and expression of genes to be classified. Searching and screening the genome has also given insights into the complex and large network of enzymes required to fully hydrolyse and metabolise the sugars released from lignocellulose. These enzymes are spread across many different glycosyl hydrolase families conferring different catalytic activities. The characterisation of these novel enzymes points towards a system adapted to not only a broad specificity of substrate but also environmental factors such as high temperature and pH. Genomic analysis revealed gene clusters and traits which could be used in the design of a synthetic cellulolytic network, or for the conversion of *C. fimi* into a consolidated bioprocessor itself.

TABLE OF CONTENTS

List of Figures and tables

Figures

Tables

LIST OF ABBREVIATIONS

Amp – ampicillin
BB – Binding buffer
BSA – Bovine serum albumin
CBM – Carbohydrate binding module
CBP – Consolidated bioprocessor
CMC – Carboxymethyl cellulose
Cml – Chloramphenicol
Contig – Contiguous DNA fragment
CV – Column volume
DMSO – Dimethyl sulfoxide
DSM – Dubos salt medium
DTT – Dithithreitol
EB – Elution buffer
EDTA – Ethylenediaminetetraacetic acid
Erm – Erythromycin
EtOH – Ethanol
GH – Glycosyl hydrolase
goi – Gene of interest
HIC – Hydrophobic interaction chromatography
His – Histidine
IPTG – Isopropyl β-$_D$-1-thiogalactopyranoside
Kan – Kanamycin
LB – Luria broth
MUC – 4-methylumbelliferyl β-$_D$-cellobioside
MUG – 4-methylumbelliferyl β-$_D$-glucopyranoside
NaI – Sodium iodide
NaOAc – Sodium acetate
NaOH – Sodium hydroxide
ONPC – 2-nitrophenyl β-$_D$-cellobioside
ONPGal – 2-nitrophenyl β-$_D$-galactopyranoside
ONPX – 2-nitrophenyl β-$_D$-xylopyranoside
ORF – Open reading frame
PAGE – Poly acrylamide gel electrophoresis
PBS – Phosphate buffered saline
PCR – Polymerase chain reaction
PNPA – 4-nitrophenyl α-$_L$-arabinofuranoside
RBB – Remazol brilliant blue
RBS – Ribosome binding site
rcf – Relative centrifugal force
SDS – Sodium dodecyl laurel sulfate
Spec – Spectinomycin
TAE – Tris-acetate-EDTA
Tris-HCl – Tris(hydroxymethyl)aminomethane hydrochloride
US DOE – US Department of Energy
X-gal – 5-bromo-4-chloro-3-indolyl-β-$_D$-galactopyranoside
YE – Yeast extract

CHAPTER 1: INTRODUCTION

A recent report by the UKERC predicts that the world production of oil will peak before 2030 and could reach its limits even before 2020 [1]. Fossil fuels pose major environmental problems with the release of greenhouse gases (GHGs) through their transport, refining and use, and are now generally regarded as the major contributor to global warming. The UK government has pledged to reduce GHG emissions to 80% of 1990 levels through several strategies (www.gov.uk). An over-dependence on crude oil may also lead to uncertainty about the price stability and security of the national energy supply, not just for transportation fuels but also for oil-derived chemical products. Stocks are dwindling, and worldwide 27% of primary energy is earmarked for transportation use. It is becoming clear that there is a need for finding renewable alternatives to fossil fuel which offer a reduction in GHG levels/emissions, to supplement the current and future stocks of oil.

Alternatives to oil-derived fuels for transport energy are realistically limited to either hydrogen (requiring new technologies to produce effectively and conversion of transport pipelines and engines) or biofuels produced from plant sugars, such as ethanol or biodiesel, requiring no major changes to transport infrastructure or vehicles. When ethanol is mixed up to 30% with petrol there is no need for engine alterations, and bioethanol global production has increased markedly since 2000 [2]. For 30 years Brazil has been producing bioethanol to blend with their gasoline to make “gasohol” [3]. Of the corn in the US, 20% is currently grown for ethanol production, with the Energy Independence and Security Act (EISA) mandating a 5 fold increase in bioethanol production to 117 billion litres by 2022 [4]. More than half of this is to come from corn ethanol, with a quantity equivalent to 43% of corn grown in 2004 being produced for ethanol in 2016 [5].

There are several drawbacks however. Bioethanol is currently produced from the harvesting of starch sugars from crops and their fermentation to ethanol. von Blottnitz and Curran in their review of studies into the effects of biofuel use from a net energy, GHG and environmental life cycle perspective, report that the use of

sugar crops could replace fossil fuel use by up to 250 GJ per hectare of land turned over for this crop, but requires tropical climates to produce. Starch crops on the other hand would only produce 35-50 GJ per hectare of the crop [6]. The increased demand for bioethanol has led to substitution of food crops for fuel crops, creating an increase in food costs: 40% for corn, 20% for soybean and 17% for wheat [5]. This in turn would, and has, led to increased land usage for food crop growth to replace that turned over to biofuel crops which may also lead to water shortages due to increased irrigation of drought-ridden areas for cultivation into arable land [4]. Further likely side effects are the increased run-off from fertilisers into rivers and the ocean and increased NO_2 levels. The increased clearance of land for crop growth, both fuel and food, could also greatly increase GHGs by release of the carbon stored in soils, and a decrease in CO_2 sequestration due to deforestation and clearance of grasslands [5, 7]. Finding a production method for ethanol that does not displace the existing crops or lead to increased deforestation or destruction of grasslands, could lead to a decrease in atmospheric GHGs of 72 $PgCO_{2equiv}$. Deforestation or changes in crop-land use however, could lead to an average 92 $PgCO_{2equiv}$ increase in GHG emissions between 2000 and 2100 [8]. von Blottnitz concludes that using lignocellulose or other industrial waste products (such as molasses) can replace between 25-95 GJ per hectare of fossil fuel, without the need for increased land usage, and this is not taking into account the possibilities from grass lands [6].

Lignocellulosic biomass is the only real renewable energy source for biofuels that would not displace existing crops or lead to increased arable land and deforestation. This biomass is a waste produced from several industries including forestry and agriculture. Worldwide annual biomass production has been estimated to be 200 x 10^{12} kg, with total energy content, estimated from heat combustion analysis, 5 times that of the total worldwide crude oil consumption [9]. However, the decomposition of biomass is not an easy process and is the main reason why ethanol production from starch costs \$0.01-0.05 per gallon and a starting cost of \$5 per gallon from lignocellulosic biomass [10].

1.1 Lignocellulosic Biomass

1.1.1 *Cellulose*

Cellulose is the most abundant biopolymer on earth and forms a large part of societal waste [11]. There are two types of plant cell wall, comprising ~40% plant biomass, of which cellulose is the major component (figure 1.1). The primary cell wall is formed as cells develop and is the only wall type for some cells such as some parenchyma [12]. The secondary cell wall is deposited onto the primary cell wall after cells are fully expanded and is usually lignified [12]. Cellulose itself is comprised of thousands of glucose molecules linked through β-1,4-glycosidic bonding with the smallest repeating subunit being cellobiose. These chains lie parallel to one another allowing for hydrogen bonds and van der Waal's interactions to form between the chains creating mostly crystalline, but also containing amorphous soluble regions, aggregates called microfibrils. These microfibrils are present in primary and secondary cell walls, but have been shown to aggregate in secondary cell walls to form larger fibrils [12]. These cellulose fibrils and microfibrils are interwoven with non-cellulosic polysaccharides generally known as hemicellulose and lignin which bind cellulose through covalent and non-covalent interactions, stabilising further the plant cell walls and increasing recalcitrance to enzymatic attack (figure 1.1 B).

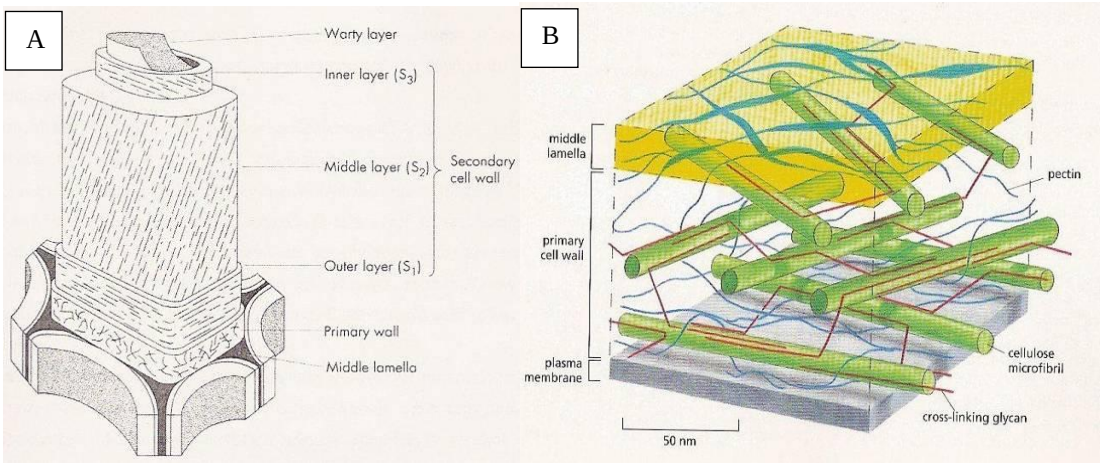


Figure 1.1. Primary and secondary plant cell wall structures. A: Primary cell wall composition [13]. B: Secondary cell wall composition [9]

1.1.2 *Hemicellulose*

Hemicelluloses are comprised of several types of sugar backbone with variable degrees of branching and interaction with other sugar residues giving heterogeneous polymers [14]. They comprise 15-35% of plant biomass and may contain pentoses (xylose and arabinose), hexoses (mannose, glucose and galactose) and uronic acids (glucuronic, methylgalacturonic and galacturonic acids). The most abundant hemicellulose residues are xylans and glucomannans.

Heteroglucans (xyloglucans) have a repeating backbone unit of 4 β-1,4-linked glucopyranosyl molecules with 75% of these residues substituted for by xylopyranosyl units [12, 14]. Arabinose and fucose residues can be attached to xylose to form di- or tri- residue side chains in these polymers. Heteroxylans (glucuronoxylans) are the most abundant non-cellulosic polysaccharide in hardwoods [12]. Heteroxylans consist of a linear backbone of β-xylopyranosyl units linked by β-1,4-glycosidic bonds. These units can be acetylated and roughly every tenth xylopyranosyl sugar can be joined to a glucopyranosyl uronic acid residue. These glucopyranosyl uronic acid residues increase the resistance of glucuronoxylan to acids and may contain rhamnose and galacturonic acid, the latter of which increases resistance to alkali treatment [12]. Galactoglucomannans (heteromannans) are the most abundant hemicelluloses in softwoods [9, 12, 14]. They consist of a linear chain of β-mannopyranosyl and β-glucopyranosyl residues joined by β-1,4-glycosidic bonds and arranged in a random fashion. These chains are also partially acetylated with α-galacturopyranosyl residues attached to both the glucose and mannose sugars by α-1,6-bonds. The galactose residues govern their tendency to associate with the cellulose chain and therefore the cellulose's extractability [14]. In hardwoods the glucomannans are the second most abundant hemicellulose and are comprised of the same backbone as galactomannans, but without the galactose residues [12].

Source	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Grasses	25-40	25-50	10-30
Softwoods	45-50	25-35	25-35
Hardwoods	45-55	24-40	18-25

Table 1.1. Cellulosic percentage composition of different woods [9]

Other hemicellulose components include arabinoglucuronoxylans and arabinoxylans. Arabinoglucuronoxylans consists of a xylopyranose backbone with attached glucopyranosyl uronic acid residues and α-1,2- and -1,3- glycosidic bonding of arabinofuranosyl residues. Arabinoxylans consist of a β-1,4-xylopyranose backbone linked to an α-arabinofuranosyl unit side chain and α-glucopyranosyl uronic acid unit. The arabinofuranosyl residues may also be esterified with ferulic and coumaric acids. Dimerisation of esterified phenolic compounds can lead to intra and inter molecular interactions increasing xylan resistance to extraction and hydrolysis [14].

Pectin, a polymer of negatively charged galacturonic acid units, is also present in primary cell walls [13]. It interacts with the other polysaccharides present through both covalent and non-covalent means by forming a gel-like composition thus giving the primary cell wall structure and rigidity, and is also found in the middle lamella that joins adjacent cells [13].

Lignin is present in the cell walls of hardwoods and softwoods and always comprises the lowest fraction of lignocellulosic biomass (table 1.1). It is a random co-polymer of phenylpropane units that binds covalently to polysaccharides through direct ester and ether linkages [9, 12].

With their varied and complex structures, a range of enzymes have evolved to degrade hemicellulose, lignin and cellulose. These enzymes include xylanases, mannanases, arabinofuranosidases, β-glucosidases, cellobiosidases, lignin peroxidise and pectate lyases. Some work has gone into the harnessing of xylanase enzymes for the hydrolysis of hemicellulose to release surface area for cellulolytic attack [14-18]. Industrially however, the focus of much research has centred on the chemical and/or physical removal of cell material that hinders enzymatic hydrolysis of cellulose.

1.2 Current Production of Lignocellulosic Biofuel

The hydrolysis of lignocellulosic biomass is a multistep process. It is currently still necessary to pre-treat biomass to expose the cellulose to enzymatic attack. The pre-treatment must be balanced however to be just severe enough to expose the cellulose without conversion of the sugars to non-fermentable or toxic compounds [10]. There are several currently used methods of pre-treatment with physical, chemical and biologically based approaches, each with their own effects on sugar availability and lignocellulosic structure [19].

1.2.1 ***Pre-treatment***

Physical pre-treatments reduce the particle size, degree of polymerisation and crystallinity of cellulose by mechanical stress [20]. Pre-treatments which increase the surface area for enzymatic degradation include milling, steam explosion and irradiation.

Chemical pre-treatments comprise the use of acidic, alkali and oxidative agents to solubilise lignin and hemicellulose [21]. Dilute acids (0.5-1% sulphuric acid) at moderate temperatures (140-190ºC) are used to solubilise hemicellulose, but leave lignin relatively untouched, increasing the glucose recovery yields from cellulose to nearly 100% with complete hemicellulose removal. Alkaline pre-treatments (usually sodium hydroxide or ammonia) do not hydrolyse hemicellulose but cause swelling of the cellulose fibres, decreasing their degree of polymerisation and crystallinity which disrupts the linkages between lignin and carbohydrates and lignin structure [19]. Solvent treatment works on the basis of differential solubilisation of sugars, including cellulose, within the cell wall [20]. Organic solvents such as alcohols and phosphoric acid, combined with acetone, can be used to dissolve the various components (lignin with alcohol, cellulose with phosphoric acid) and increase their availability to enzymatic attack.

Biological pre-treatment is also a viable option. White–rot fungi are typically utilised as they produce a wealth of enzymes necessary to degrade lignin. Unfortunately, incubation periods can last from days to weeks. However, the energetic requirement of chemical pre-treatment can be reduced if coupled to a fungus [19]. It has also been shown that there is a potential decrease in the sugars available for fermentation, due to the metabolism of the fungi.

The pre-treatment processes all have advantages and disadvantages and are more effective on some biomass types than others [20, 21]. They also differ in the production of potentially fermentation-inhibiting compounds such as phenols, furans, furfurals and carboxylic acids, and have been demonstrated to affect directly enzyme hydrolysis of cellulose [22, 23]. This is also one of the costliest steps in cellulose bioprocessing due to the construction and materials needed to withstand the extremes of temperature and pH. A reduction in the need for pre-treatment through a wider variety of robust enzymes would remove much of the cost associated with cellulosic biofuel production.

The next step in the biofuel process is the release of sugars from cellulose which is generally achieved through enzymes produced by the filamentous fungus *Trichoderma reesei*. This fungus has been shown to produce 20 g/l of cellulase protein and secretes a broad range of cellulases, including 2 cellobiohydrolases which release cellobiose from crystalline cellulose, 5 endoglucanases which primarily degrade soluble cellulose and 2 β-glucosidases to cleave cellobiose to produce glucose [24, 25]. This is in addition to hemicellulolytic enzymes including 4 endo-1,4-β-xylanases, a mannanase, acetylxylan esterase, α-galactosidase and an arabinofuranosidase. Production of the necessary enzymes is another costly step in the production of biofuels. Production of hypersecretory strains is a focus for certain lab groups, including those in a large drive funded by the US Department of Energy, in an attempt to reduce cost greatly [10, 26]. The released sugars are then fermented in a traditional manner by *Saccharomyces cerevisiae* to produce ethanol [27].

With much of the cost for biofuel production associated with pre-treatment strategies and cellulase production, alternatives have been sought that aim to consolidate the cellulolytic and fermentative steps.

1.3 Microbial Degradation of Lignocellulose and Biofuel Production

After pre-treatment of lignocellulose, as described above, there are a number of possibilities for cellulose degradation. The first is separate hydrolysis and fermentation (SHF). This is the addition of cellulolytic enzymes to release the sugars stored as cellulose first, followed by fermentation of the sugars released. This requires the production of enzyme mixtures, and reports of functional cellulase enzymes produced by *E. coli* have been published [28-33]. However, this strategy still relies on the addition of a fermentative organism, typically *Saccharomyces cerevisiae*, after enzymatic hydrolysis. The main problem with this is that enzymatic hydrolysis of lignocellulosic biomass is inhibited by the resultant products cellobiose and glucose, reducing the conversion rate of cellulose to glucose [34]. One solution for this problem is simultaneous saccharification and fermentation (SSF), where lignocellulosic biomass is degraded by cellulolytic enzymes, and the sugars released immediately used for microbial fermentation to ethanol [34-36]. This should prove economically favourable as enzyme inhibition by-product is reduced leading to lower enzyme levels being needed. There is still the cost of enzyme production however, and the side effect of the enzymes being inactivated by the ethanol produced requiring more enzymes, while pentose sugars are fermented in a separate reaction. Consolidated bioprocessing is the integration of cellulolytic enzyme production, saccharification of the cellulose and hemicellulose to fermentable sugars and their conversion to a useful product such as ethanol in a one-step process. Several extensive reviews have been published recently on this endeavour [27, 37, 38].

1.3.1 *Consolidated Bioprocessing*

According to Lynd et al development of a single organism to produce cellulolytic enzymes and ethanol could lead to a significant reduction in production costs,

estimated at ≥$0.5 per gallon ethanol produced by SSF [27]. This has been termed a consolidated bioprocessor (CBP). There are two main strategies to producing a CBP, by either enhancing the ability of a naturally cellulolytic organism to produce the desired products, or by introducing cellulolytic enzymes into non-cellulolytic organisms capable of producing the product to a high yield [27]. The majority of efforts to produce heterologous cellulases have thus far occurred in 3 main species: *Saccharomyces cerevisiae*, *Zymomonas mobilis* and multiple species of enteric bacteria.

The introduction of cellulase enzymes into *S. cerevisiae* has been demonstrated to produce functional enzymes. The use of *S. cerevisiae* is desirable as it can tolerate high levels of ethanol and produces it through a homoethanol pathway, so carbon isn't siphoned off into unwanted by-products [39]. Recombinant hydrolysis of filter paper and pre-treated wood chips by *S. cerevisiae* expressing Cex and CenA from *Cellulomonas fimi* was reported by Wong et al [30]. McBride et al introduced recombinant genes coding for BGL1 (β-glucosidase) from *Saccharomycopsis fibuligera* and EG1 (endoglucanase) from *T. reesei* into *S. cerevisiae* [40]. This strain was then able to grow on cultures with cellobiose as the sole carbon source, which the wild type is unable to do, with growth seemingly limited by enzyme saturation of the secretory system limiting β-glucosidase activity. Lee et al co-expressed XynA, a xylanase from *Bacillus spp*, and Egl6, an endoglucanase from *Trichoderma spp*. in *S. cerevisiae* [41]. They were able to show functional secretory production of the 2 enzymes, and cell growth on YPD medium, however as part of their screen they grew the yeast on carboxymethyl cellulose containing plates and growth was visibly lower than on xylose containing plates.

Zymomonas mobilis is a facultative anaerobic ethanologenic bacterium from which the production of ethanol (PET) operon is derived. Just like *S. cerevisiae*, this organism has been used as a host for cellulolytic enzymes. Initially expression of heterologous endoglucanases from either *Pseudomonas fluorescens* or from *Bacillus subtilis* resulted in poor expression and a lack of extracellular activity [42, 43]. Brestic-Goachet however produced a strain capable of producing and secreting an

endoglucanase (CelZ) from *Erwinia chrysanthemi* [44]. Introduction and expression of a β-glucosidase from *Ruminoccocus albus* conferred the efficient hydrolysis and fermentation of cellobiose [45].

Despite the ability to produce cellulolytic enzymes and the high production of ethanol, *S. cerevisiae* and *Z. mobilis* both lack the natural capabilities to ferment pentose sugars. The ability to hydrolyse and ferment both hexoses (glucose) and pentoses (xylose and arabinose) released from biomass is essential for maximal biofuel yield from the substrate [46]. Considerable effort has gone into producing capable strains of both organisms with some success [46-48]. It has been shown however that sugar transport can be a problem in *S. cerevisiae*, with glucose and xylose uptake through the same transporter channel, but xylose having a very low affinity and glucose inhibiting the uptake of xylose [49]. Although hydrolysis of soluble substrates is demonstrably effective, degradation of crystalline cellulose is still problematic and relatively unsuccessful [27, 38].

Enteric bacteria such as *Escherichia coli* or *Klebsiella oxytoca* offer an alternative to the previous organisms due to their ability to utilise glucose and pentose sugars naturally. Ethanol is already a product of anaerobic respiration in these bacteria and through the introduction of the PET operon consisting of *Z. mobilis* pyruvate decarboxylase and alcohol dehydrogenase, they have been pushed to produce relatively high levels of ethanol above those of pentose-fermenting yeasts [48, 50, 51]. The *E. chrysanthemi celZ* gene was shown to be expressed in *E. coli* producing CelZ inclusion bodies, but was successfully exported out of the cell by the additional expression of the *out* operon genes of *E. chrysanthemi* [27]. The same cloning strategy also resulted in similar levels being produced in *K. oxytoca*. Zhou and Ingram were able to produce synergistic hydrolysis of carboxymethyl cellulose (CMC) by the expression of CelY, another endoglucanase, along with CelZ [52]. Expressing these enzymes in *K. oxytoca* has resulted in the production of ethanol from CMC and crystalline cellulose, but the addition of extra cellulases to the growth medium was necessary for crystalline [53]. Enteric bacteria appear to have several advantages over yeast as potential CBPs, however crystalline cellulose hydrolysis is

still an issue and the ethanol tolerance levels are much lower than those of traditional industrial fermentative strains [54].

Seemingly most studies involve the expression of *T. reesei* or other fungal genes in *S. cerevisiae*, which tends to yield low amounts of enzyme and lacks an ability to grow on crystalline or complex forms of cellulose [38]. *T. reesei* being the major contributor to cellulolytic genes so far has led to a lack of suitable genes for expression in prokaryotes. Expression of eukaryotic proteins in bacteria is problematic due to potential for necessary post-translational modifications and protein folding not being carried out correctly. A deeper study of bacterial cellulolytic systems for a broader range of cellulase and hemicellulase genes is therefore necessary. Synergy between enzymes is clearly an important factor too for the efficient hydrolysis of cellulose, as demonstrated by the effect of *celZ* and *celY* when expressed in enteric bacteria, pointing towards the need for more characterised enzymes and more complex gene combinations.

1.4. Synthetic Biology

Synthetic biology is an emerging field in the area of genetic engineering. The basic aim is to apply the methods of engineering and system design to the production of novel biological functions and the optimisation of existing biological systems such as metabolism. In engineering a system can be designed and characterised efficiently and carefully prior to it being built. This is largely due to the components being used. If designing a radio, each transistor, resistor, screw, switch and wire is characterised and produced to an industry-wide standard so it is already known how they work together, that they can fit together, and how to achieve the desired outcome. To carry on the analogy to biological systems the components will be made up of promoters and genes. These components in synthetic biology are known as BioBricks [55]. This particular form of synthetic biology was first proposed by Tom Knight with their main feature being the ease by which networks can be constructed and parts interchanged.

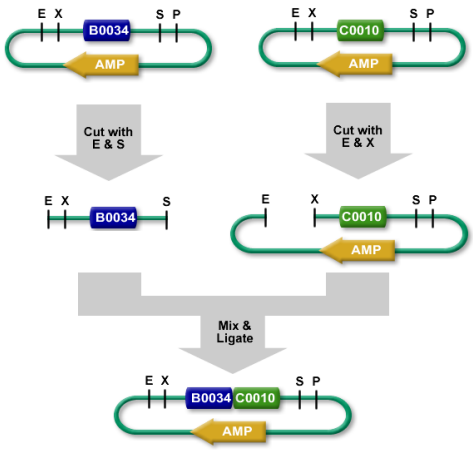


Figure 1.2. Diagram of standard BioBrick assembly
(http://parts.igem.org/wiki/images/1/1a/StdAss.png)

Figure 1.2 displays the basic premise of the BioBrick and its assembly. The BioBrick part (gene coding sequence, ribosome binding site, or promoter for example) is cloned by PCR with specific forward and reverse primers. The forward primers contain the prefix sequence which has *Eco*RI and *Xba*I restriction sites (E and X), the reverse primers contain the suffix sequence with *Spe*I and *Pst*I restriction sites (S and P). These are then digested into a standard BioBrick plasmid (eg. pSB1A2), which codes for a specific antibiotic resistance (AMP in figure 1.2 representing ampicillin resistance). BioBricks can then be combined directionally by digestion of plasmids to open the plasmid backbone, or excise the BioBrick part, followed by ligation. Figure 1.2 shows BioBrick B0034 being excised by digestion with E and S, and the target plasmid digested using E and X. *Xba*I and *Spe*I produce complimentary overhangs so B0034 will ligate upstream of C0010. B0034 similarly could have been inserted downstream of C0010 by digestion with X and P, and C0010 digested with S and P.

This simple assembly method and the BioBrick concept of interchangeable parts lends itself to the rapid creation of synthetic networks and operons. The International Genetically Engineered Machine (iGEM) competition asks teams of undergraduates worldwide to make systems using the BioBrick format (www.igem.org). Projects started in the iGEM competition have gone on to be published with three examples of this from 2007 alone. The Edinburgh University team produced *E. coli* that was able to detect arsenate ions in water, a toxic

by-product of contamination by arsenic, with applications in the detection of contaminated water supplies in developing countries [56]. The entry produced by the team from Brown University, USA, was a tri-switch able to control 3 reporter genes by repressor feedback in the promoter regions – each promoter expressed a repressor for the other 2 genes – which could potentially be used to control artificial systems [57]. The team from Davidson College, USA, produced a switch that works by inverting DNA sequences between sites recognised by Hin, an invertase, which they hoped in a multicellular format could be used for performing calculations, specifically solving the "burnt pancake" mathematical problem [58].

By conforming to a standard in this way there are 2 main advantages over traditional cloning methods [59]. The first is that these parts all conform to the same standard so any biological part should fit together with any other in any combination no matter who has produced it. This naturally leads to the second important feature, that these BioBricks, once created, can be assembled in an automation-friendly way opening the possibilities for high-throughput network assembly of all possible combinations of the system parts. BioBricks and synthetic biology therefore offer a viable option in the search for consolidated bioprocessing [38].

1.5 Cellulolytic Bacteria

There are two types of bacterial cellulase systems denoted by their environment. The first belongs primarily to the anaerobic bacteria and is known as a cellulosome. The cellulosome is a nanostructure attached to the cell surface of cellulolytic anaerobic bacteria such as *Clostridium thermocellum*, *Clostridium spp.*, *Bacterioides cellulosolvens* and *Ruminococcus flavefaciens* [60]. It is arranged into a high molecular weight complex (>2 MDa) consisting of scaffoldin protein (CipA) which has a cellulose binding domain and nine type I cohesion modules to which dockerin containing cellulolytic enzymes can bind [61]. The production of cellulosomes is thought to be an evolutionary advantage to counteract the low energy yields produced by fermentative respiration, through fine control of metabolic activities [62]. By grouping together cellulolytic enzymes, synergistic activity can be

increased and the released sugars concentrated close to the bacterium. This close proximity to substrate and released product also allows the cell to have fine control of the concentration of potentially inhibitory released products (such as sugar monomers), turning on and off genes as and when required which eliminates wasteful energy expenditure caused by the continuous expression of large amounts of cellulolytic enzymes.

The second system is of free cellulase, typically produced by fungi and aerobic bacteria. Cellulolytic enzymes that work synergistically to break down cellulose and hemicellulose are secreted, and the released sugars taken up from the environment [9]. The cellulase enzymes are modular in structure, generally comprising a carbohydrate binding module (CBM), catalytic domain and linker peptide, and are grouped into glycosyl hydrolase families based on sequence homologies of their catalytic domains [63]. Free cellulase system bacteria include *Paenibcaillus spp.*, *Thermobifida* (formerly *Thermomonospora*) *fusca* and *Cellulomonas spp.* [62]. A paradigm of their synergistic function is given by Mansfield and Meder based on the cellulase system of *Cellulomonas fimi* [64]. Endoglucanases bind preferentially to amorphous cellulose and cleave internal β-1,4-glycosidic bonds to create more ends for exoglucanases to act upon. Exoglucanases, typically cellobiohydrolases, processively remove cellobiose from the cellulose chain ends, with different exoglucanases showing catalytic activity towards the reducing and non-reducing ends. This simplified system demonstrates that at least 4 synergistically working enzymes are required for the hydrolysis of cellulose alone (figure 1.3), generating cellobiose, and then the hydrolysis of cellobiose to glucose by a β-glucosidase. For the efficient hydrolysis of lignocellulose, other synergistically working enzymes are required to remove hemicellulose (figure 1.4). These enzymes can be just as important even when hydrolysing chemically pre-treated biomass, as when hydrolysing raw materials [23, 65].

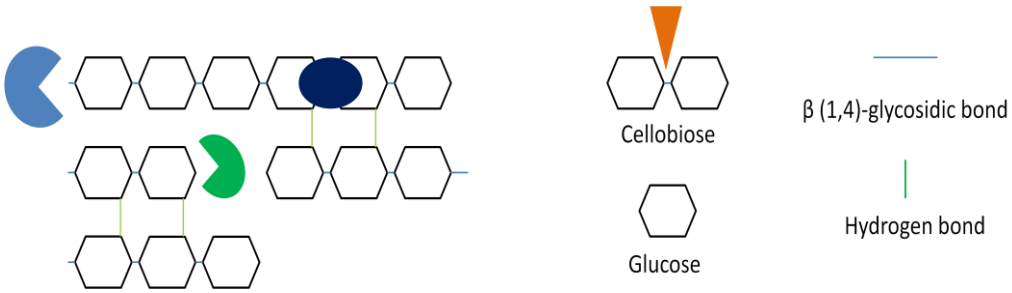


Figure 1.3. Cartoon representation of the cell-free hydrolysis of cellulose
: Endo-1,4-β-glucanase; : Non-reducing end cellobiohydrolase; : Reducing end cellobiohydrolase; : 1,4-β-glucosidase

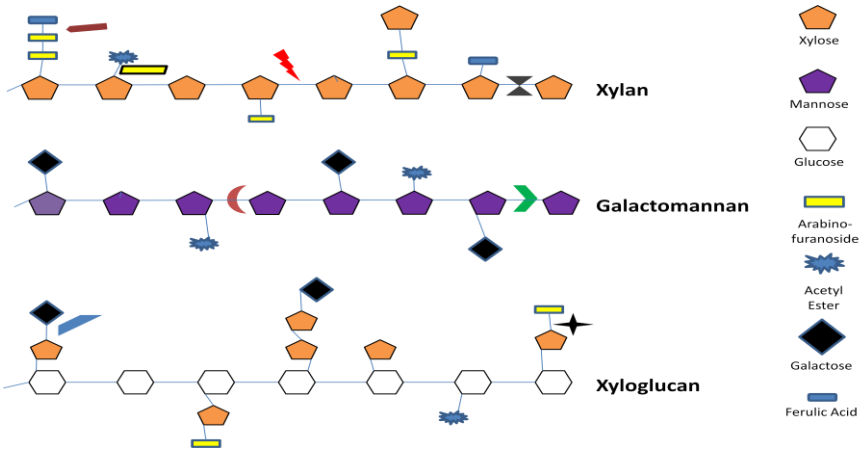


Figure 1.4. Cartoon representation of the cell-free hydrolysis of hemicellulose
: Endo-1,4-β-xylanase; : 1,4-β-xylosidase; : Mannan endo-1,4-β-mannanase; : β-1,4-mannosidase; : α-$_L$-arabinofuranosidase; : β-galactosidase; : Acetyl xylan esterase; : Feruloyl esterase

For the construction of a consolidated bioprocessing organism, it therefore makes sense to start with a series of enzymes that are fairly well characterised from an organism that is also well studied. *Cellulomonas fimi*, an organism that has been studied for the past 30 years with several well characterised glucanase enzymes is a prime example.

1.5.1 *Cellulomonas fimi*

Cellulomonas fimi is a soil dwelling, Gram positive, rod-shaped, mesophilic, facultative anaerobe with a GC content of 75% [66]. It grows optimally at 30ºC, in

aerobic conditions and can use cellulose as the major or sole carbon source, with growth demonstrated on wood, agricultural wastes and crystalline cellulose [67].

Protein Name	Function	GH family	CBM family	Catalytic Domain	Reference and Accession Number
CenA	Endoglucanase (EC 3.2.1.4)	6	2	Inverting	Wong et al, 1986 M15823
CenB	Endoglucanase	9	2	Inverting	Meinke et al, 1991 M64644
CenC	Endoglucanase	9	4 and 9	Inverting	Coutinho et al, 1991 X57858
CenD	Endoglucanase	5	2	Retaining	Meinke et al, 1993 L02544
CbhB	Exocellobiohydrolase (EC 3.2.1.176; reducing end)	48	2	Inverting	H Shen et al, 1995 L38827
CbhA	Exocellobiohydrolase (EC 3.2.1.91 non-reducing end)	6	2	Inverting	Meinke et al, 1994 L25809
XynB (Cex)	Endoxlanase/exoglucanase (EC 3.2.1.8)	10	2	Retaining	O'Neill et al, 1986 L11080
XynC	Endoxylanase	10	9	Retaining	J.H. Clarke, 1995 Z50866
XynD	Endoxylanase	11	2	Retaining	Millward-Sadler et al, 1994 X76729
XynE (Cfx)	Endoxylanase	10	2	Retaining	Hekmat et al, 2005 DQ146941
ManD (Man26A)	β-mannanase (EC 3.2.1.78)	26	23	Retaining	Stoll et al, 1999 AF126471
ManA (Man2A)	β-mannosidase (EC 3.2.1.25)	2		Retaining	Stoll et al, 1999 AF126472

Table 1.2. Known *Cellulomonas fimi* cellulases and hemicellulases. The catalytic domain refers to the mechanism of inversion or retention of the anomeric carbon configuration at the site of catalysis.

There are 12 known extracellular enzymes involved in cellulose and hemicellulose degradation produced by *Cellulomonas fimi*. The enzyme mix is given in table 1.2. There are four endoglucanase enzymes present each with different activities on

cellulose and structural conformations, but each contains a catalytic domain that hydrolyses β-1,4-glycosidic bonds, a proline/threonine-rich linker sequence and a carbohydrate binding domain (CBD). Endoglucanase A (CenA) is a 418 amino acid enzyme, more active on amorphous and soluble cellulose than crystalline [68, 69]. Endoglucanase B (CenB), is much larger than CenA at 1012 amino acids, due to fibronectin-type III (fnIII) repeats in the linker region, and is also more active on soluble cellulose [70]. Endoglucanase C (CenC) has a 1101 amino acid sequence, with seemingly 2 CBDs that work in tandem at the N-terminal to allow for binding to amorphous cellulose to which it has shown a preference, and exhibits both endo- and exoglucanase activities [71-73]. Endoglucanase D (CenD) is 747 amino acids in length and is preferentially active on crystalline cellulose, and contains fnIII repeats in the linker sequence like CenB [70].

XynB, formerly Cex, has exoglucanase activity on cellulose but is characterised as a true xylanase which preferentially binds to xylose and contains 2 CBDs [74, 75]. To complete the degradation of cellulose, cellobiohydrolases are also required and two have been found to be secreted by *C. fimi*. CbhA and CbhB are both exo-acting cellobiohydrolases with some degree of endoglucanase activity that act processively to remove cellobiose from the non-reducing and reducing ends of the cellulose chain respectively [76, 77].

To add to *C. fimi's* arsenal for lignocellulosic hydrolysis, several hemicellulase enzymes have been reported. Three additional endoxylanase proteins (XynC, XynD and XynE) have been found in the *C. fimi* secretome as well as one mannanase (ManD) and one mannosidase (ManA). Two of these xylanases are from glycosyl hydrolase family 10, XynC and XynE, as is XynB, which has catalytic activity towards xylose and cellulose [78-80]. Polysaccharide hydrolysing enzymes and carbohydrate binding domains can all be grouped into families dependent on their amino acid sequences, with members of families containing the same conserved regions [81]. Members of the same glycosyl hydrolase families often have the same or similar catalytic functions (such as GH10), where as other families have more enzymatic activities associated with them (such as GH5), a full list of families is

available at www.cazy.org. They all contain carbohydrate binding domains with specificities to xylose and cellulose. Two mannan degrading proteins have been found in the secretome. A β-mannosidase (ManA) of GH family 2 and a β-mannanase (ManD) of family 26 were cloned and characterised by Stoll et al and were the first to be reported in *C. fimi* [82]. Proteolytic attack on ManD produced two discrete modules, one of which was the catalytic domain, the other was of unknown function; however an S-layer homology domain was identified by sequence analysis and at least 2 other unknown modules were present. Further study of ManD led to the discovery of a mannose binding domain that binds to soluble mannans only, Ig-like folds and a family 23 CBD [83, 84]. Man2A has only one module, that of a family 2 catalytic domain [82].

With the number of carbohydrate binding domains present in the major components of the cellulase system they clearly play an important role in hydrolysis. It has been suggested that CBDs aid hydrolysis by increasing the concentration of enzymes on the substrate, but the act of binding itself might disrupt the crystalline structure of cellulose allowing for enzymatic attack [85]. Carrard et al showed through the modular nature of CBDs that they are intrinsic in promotion of site specific hydrolysis by enzymes, with little overlap being seen between different hydrolases [86]. Lateral diffusion of enzymes across cellulose has also been demonstrated, mediated by CBDs, ensuring enzymes can reach their target substrate and this plays an important role in substrate inhibition of enzyme activity [87, 88]. The linker sequence cannot be overlooked either as this allows the catalytic and binding domains to work in synergy – the CBD attaches to the substrate and the catalytic domain is able to move from site to site while still being tethered to the substrate [89].

The enzymes and linker sequences are glycosylated, which protects them from protease attack when bound to cellulose but not when they are in solution [89-91]. However, glycosylation is not necessary for the enzymes to be functional [92]. Proteases are simultaneously secreted by *C. fimi* along with endoglucanases [67, 91]. Proteases have been suggested as one way of enzymatic control by releasing more of

the catalytic domains in to the soluble medium as non-cellulosic nutrients increase [67]. This all in all points to a more complex cellulase system than the early views of Langsford et al, warranting a much closer look [93].

1.6 Project Aims

Damian Barnard, a previous member of the French lab has already cloned the known cellulases and hemicellulases of *C. fimi* as BioBricks and has combined them to identify hidden synergies when hydrolysing cellulosic substrates (currently unpublished). Damian first assayed the individual genes using standardised conditions so as to accurately compare the activities of the enzymes. Within the set of genes tested were *bxyF* and *xynF*, 2 genes cloned by myself which exhibited β-xylosidase activity (chapters 4 and 5). He expressed these BioBricks in *Escherichia coli* MG1655, as *E. coli* is a likely host for a consolidated bioprocessor, being a well characterised bacterium with the ability to ferment sugars. He also expressed the constructs in *Citrobacter freundii* strains SBS197 and NCIMB11490, due to its relatedness to *E. coli* and its ability to utilise cellobiose as a carbon source, which *E. coli* is unable to do, and he demonstrated its ability to grow under saline conditions (in lab observations). This would be beneficial for sea based fermentation plants that would not require the repurposing of land. He successfully expressed these genes with no apparent problems in all three host strains, although it has been previously noted that over expression of XynB (Cex) in *E. coli* can be toxic [94]. His next step was to then join combinations of these BioBricks to test for synergistic behaviours with cultures expressing these constructs grown on mannan, xylan, carboxymethyl cellulose (CMC, soluble cellulose), avicel (crystalline cellulose powder) or filter paper made of cellulose.

Through these combinations he found that different enzymes did in fact work synergistically and preferentially on certain substrates and that there were also differences in activities of enzymes between strains. A construct expressing both Man2A and Man26A was able to give MG1655 the ability to grow using mannan as sole carbon source, but this was not observed in the *C. freundii* strains despite both

species having mannose utilisation pathways [95]. Of particular note was the expression of constructs consisting of *bxyF+xynB* seemingly being better for the hydrolysis of birch wood xylan than constructs with *xynD+xynB*. Constructs expressing *cenA* or *cenC* with *xynB* were always the best at facilitating the utilisation of cellulose, soluble or otherwise. The addition of *cbhA+B* to these constructs only had a positive effect when *C. freundii* NCIMB11490 was the host. Also of interest is that *cenB+cenD* would act synergistically together, and *cenA+cenC* but not other combinations

When hosts were grown with filter paper as the carbon source expressing constructs containing *cenA+xynB*, appreciable difference in growth rate from that of hosts with vector only was only observed after 3 days of growth, and other constructs containing *cenB* or *cenD* with *xynB* had little effect on growth. The greatest differences were seen when using *C. freundii* SBS197 as host, as growth rate differences were far more pronounced than in *E. coli* MG1655 or *C. freundii* NCIMB11490. Constructs expressing *cfbglu+endoglucanase+cbhA+cbhB* and *cfbglu+cenA+xynB* were used to visually detect the destruction of filter paper. Each day the filter paper was inspected and it was only the *cfbgl+cenA+cbhA+cbhB* that showed appreciable physical degradation after 9 days for MG1655. With NCIMB11490 as host *cenA+xynB+cbhA+cbhB* showed the most destruction after 9 days and for SBS197 *cenA+xynB* was enough for filter paper destruction. *C. freundii* SBS197 co-cultures expressing *cenA/B/C/D+xynB* had their ability to physically degrade filter paper compared to that of *C. fimi*. With the *C. freundii* co-cultures, degradation was only beginning to be visible after 7-8 days. In comparison *C. fimi* cultures displayed almost total destruction of filter paper after a mere 2 days of growth. It was suggested that the possible reason for this is the relative inefficiencies in protein production due to the high GC content of the genes, and the possibility of poor secretion rates of the recombinant proteins. While these are quite likely to account for some of the discrepancy between destruction and growth rates observed, it may also be indicative that the known cellulase enzymes are not the sole or main reason for *C. fimi*'s apparent efficiency at cellulose utilisation and certainly that more than a couple of enzymes are necessary, especially as the main known cellulases

(*cenA-D*), do not universally work well together. Add to this the observation that constructs of *bxyF+xynB* work best together for the hydrolysis of xylan and it strongly suggests there are hitherto unknown enzymes that play an important synergistic role in cellulose hydrolysis, or that uptake of released sugars is extremely efficient which would facilitate growth.

With the recent release of the *C. fimi* genome by the US DOE, the full extent of the cellulolytic system can be identified. It is my aim to interrogate the genome to identify novel putative enzymes utilised by *C. fimi* in the hydrolysis of complex polysaccharides such as unprocessed lignocellulose. The enzymes in table 1.2 were identified using traditional methodologies of genomic library screening and by specifically looking at the secretome when grown on cellulose, potentially underestimating the enzymic repertoire of *C. fimi*. A more thorough overview of these methods and their drawbacks is presented in chapter 3. By screening the genome I hope to be able to add some understanding to how *C. fimi* is able to utilise so effectively a wide variety of complex substrates and their released sugars, the range of enzymes involved and the possible methods of control. This in turn will hopefully lead back to production of better heterologous cellulolytic systems. I will clone a subset of these putative genes in the BioBrick format, express them in *E. coli* and assay them for common cellulolytic functionalities before further kinetic characterisation. In doing so I hope to expand the repertoire of bacterial cellulolytic enzymes available for the potential production of a consolidated bioprocessor.

CHAPTER 2: MATERIALS AND METHODS

2.1 Genome Analysis

Prior to the release of the *Cellulomonas fimi* ATCC484 genome by the US DOE, the same strain was sequenced for our lab by Edinburgh GenePool using a combination of Illumina, Roche 454 and Sanger platforms. This sequencing effort led to the *C. fimi* genome being assembled into ~2700 contiguous DNA sequences with ambiguous overlapping ends which could not be automatically assembled. These contigs varied in size from several kilobases to less than 50 base pairs. Where possible the DNA was divided into potential ORFs by GenePool using RAST (Rapid Annotation using Subsystem Technology) [96]. Where not possible to identify an ORF due to the sequence being truncated as found at the end of a contig, or if the contig was very short, these ends or contigs were translated into the 6 potential reading frames using an in-house programme coded by Dr. Chris French. These reading frames were then manually BLAST searched and their putative annotations based on E-value and homology between hit and query, with hits of over 70% identity considered the most likely. The top or most frequent hit in the NCBI database was then used to annotate the putative gene.

Once the full genome was released by the US Department of Energy (US DOE), our putative annotations were mapped to the genome by nucleotide BLAST, generating full sequences and eliminating any duplicated gene annotations if originally spread over several contigs. Using the NCBI genome browser and graphical setting for display, it is possible to search the genome using keywords. As the uploaded genome was automatically annotated, common generic terms were searched for ("glycoside hydrolase", "carbohydrate binding", "cellulose", "xylan", "pectin", "mannose", "arabinose"), as well as more specific terms ("cellulase", "glucanase", "cellobiohydrolase", "xylanase", "pectate lyase", "arabinofuranosidase", "xylosidase", "mannosidase", "mannanase"). Any ORFs annotated as having one of these terms as part of their description was highlighted by the software. I checked each one for relevance, and BLAST searched all identified using their translated

protein sequence and annotated them as previously described based on the top or most frequent NCBI database hit.

Following identification of putative genes encoding polysaccharide hydrolysing enzymes, the translations of these genes were analysed using InterProScan [97] and PsortB [98] to identify protein domains and potential cellular localisations, respectively. Flanking genes were also noted and their distances from upstream and downstream genes calculated (tables 3.2, 3.3) to help identify potential operons or gene clusters of similar function.

To identify potential gene regulatory motifs associated with these genes, upstream DNA of the putative genes was collated. The 200 bases upstream (5`) of the A in the first ATG of the gene, and the first 10 bases downstream (3`) were used in alignments and input into the predictive software as outlined in Chapter 3.

2.2 Cloning and Expression

2.2.1 *Chemicals and Reagents*

All chemicals and reagents were purchased from Sigma-Aldrich unless otherwise stated. All cloning was performed using either KOD or KOD Xtreme Hot Start DNA Polymerase kits from Novagen (product numbers 71086 and 71975, respectively). All primers were purchased from Invitrogen. Growth medium was Luria Broth or Luria agar. For transformant selection Luria agar plates were supplemented with 100 μg/ml ampicillin for *E. coli* or 100 μg/ml carbenicillin for *C. freundii*, 90 μg/ml IPTG and 40 μg/ml X-gal. For the activity assays 2-nitrophenyl xylopyranoside, cellobioside, β-galactopyranoside, 4-nitrophenyl α-$_L$-arabinofuranoside, 4-methylumbelliferyl cellobioside and 4-methylumbelliferyl glucopyranoside (ONPX, ONPC, ONPGal, PNPA, MUC and MUG, respectively) were used and purchased from Sigma-Aldrich. Puregene yeast/bacterial genomic DNA extraction, PCR solution purification, Gel-extraction and mini-prep kits were purchased from

Qiagen. Restriction enzymes were purchased from New England Biolabs (NEB). T4 DNA ligase and T4 Polynucleotide Kinase were purchased from Promega.

2.2.2 *PCR Cloning of Target* Cellulomonas fimi *Genes*

The standard protocols supplied with the polymerases were followed, with minor adjustments. When using KOD Hot Start Polymerase, the reaction mixture was supplemented with 5% (v/v) DMSO, and the initial denaturation time was extended to five minutes at 95ºC, with cycling denaturation times extended to one minute. For PCR reactions using KOD Xtreme Hot Start Polymerase, the supplied GC buffer was used, and the denaturation steps extended as above. Genomic DNA was extracted using the Puregene Qiagen kit and used as PCR template. Annealing temperatures varied dependent on the Tm of the primers used. Thirty cycles were used to amplify the genes, with a 10 minute chase incubation at the polymerase optimum temperature after these cycles.

Five microlitres of the PCR reaction were analysed by gel electrophoresis using 0.8% agarose and TAE buffer with 1 kb DNA ladder from NEB (N3232). If one band was observed the PCR product was purified from solution using either a Qiagen PCR purification kit, or by glass beads. The glass beads were prepared as follows: 1 g of silica beads was suspended in 10 ml of 1X PBS (phosphate buffered saline, pH 7.4) and allowed to settle for 2 hours before removing the supernatant containing the finest particles, and then repeated. The remaining beads were then centrifuged at low speed 4000 rcf and the remaining supernatant removed. The final pellet was suspended in 5 ml of 3 M sodium iodide (NaI) and stored at 4ºC. Glass bead purification: to the remaining PCR mixture 150 µl of 6 M NaI was added followed by 20 µl of the glass bead suspension and mixed by pipetting. The mixture was incubated on ice for 15 minutes with mixing by agitation every 5 minutes. The suspension was centrifuged at high speed for 30 seconds to pellet the beads and the supernatant was discarded. The beads were then washed three times with ice-cold wash buffer (10 mM Tris-HCl, pH 7.5, 2.5 mM EDTA, 50 mM NaCl, 50% v/v ethanol), centrifuging to pellet and discarding the supernatant each time. The DNA

was then eluted from the beads with EB buffer (10 mM Tris-HCl, pH 8) and a 10 minute incubation at 55ºC, mixing by flicking the tube at 5 minutes. The suspension was spun at high speed again to pellet the beads and the DNA-containing supernatant was transferred to a clean microcentrifuge tube and stored at -20ºC.

2.2.3 *List of Primers and Plasmids Used*

Primer Name	Sequence (5`-3`)	Purpose
BioBrick Prefix	tctgaattcccttctagatg...	To lie upstream of coding sequence and include *Eco*RI and *Xba*I restriction sites, to replace native start codon with ATG
BioBrick Suffix	tctctgcagctactagtatta,,,	To lie downstream of coding sequence and include *Spe*I and *Pst*I restriction sites, to replace stop codon with TAA
AfsAf1	...cctcagcgcgccaccgtcc	For cloning the named genes coding sequence
AfsAr1	,,,gcgggtcagggtcgccgggaggg	
AfsBf1	...ttccccgtccgcctcaccc	
AfsBr1	,,,ggcgcccaggcggacgacgttcc	
AfsCf1	...acccaggcacgactcgtcctc	
AfsCr1	,,,gcgggtcagggtcgccgggagg	
BxyCf1	...tcgacgacgaccccgcccg	
BxyCr1	,,,ggacgcgcgtcccgccgctcc	
BxyDf1	...acggaacgcgacccccgg	
BxyDr1	,,,tggcctcgtcccttcgaggtgctcg	
BxyEf1	...acgctcctggaacccgccgc	
BxyEr1	,,,tgccgtccgggtgccgaaggcggtgcgcccgga	
BxyFf1	...acgagacgactggtgcccgagc	
BxyFr1	,,,cccctgcctcgatgatgtgggtga	
CelAf1	...cggtccgggcggttcc	
CelAr1	,,,gccggctgtgccggtcacc	

CelDf1	...tcacggaacacccaccaccgaggg	minus start and stop codons
CelDr1	,,,ggcgcagggcgacccgttgacgg	
CelEf1	...gtcccctcacccgccacgcg	
CelEr1	,,,gctgctcgcgcgcgccctccgg	
CelNf1	...acgcgacccagaacc	
CelNr1	,,,gccgcgcgtccactgctgc	
XynFf1	...acgaccccacgtctccacc	
XynFr1	,,,ggatgccgcgcacgtgacg	
XynGf1	...ccgcaccgccccatcg	
XynGr1	,,,ggcgggcaggcgcacctcgac	
XynHf1	...aaccgcgtcgtcgtccc	
XynHr1	,,,ccctgccggaccgtcgtcgc	
pSB1A2insF2	aataggcgtatcacgacg	Sequencing of pSB1A2 insert site, forward primer
pSB1A2insR2	tttgagtgagctgatacc	Sequencing of pSB1A2 insert site, reverse primer
LacZf3	gggtcgacaggtttcccgactg	Binds at position 215 of BBa_J33207 for sequencing of downstream inserts
rbs2f	ccttgaattcgcggccgcttctagagctcaaggaggtactagtagcg	Construct of RBS Biobrick BBa_B0034
rbs2r	reverse compliment of rbs2f	
Stevetagr1	gtgatggtggtggtgatgcatatggtactccttgagctctagtat	For inserting a 6X His-tag after the first ATG of any gene downstream of RBS B0034
celAhisf1	agacgggctgtccgcaccacc	Gene specific forward primers for cloning the DNA after the first ATG for construction of N-terminal His-tagged enzymes
xynIhisf1	agcgtggtcgacggctacctg	
celDhisf1	tcacggaacacccaccaccgag	
afsChisf1	acccaggcacgactcgtcctcg	
celEhisf1	gtcccctcacccgccacgcgc	
bxyDhisf1	acggaacgcgacccccgggcg	
bxyEhisf1	atgacgctcctggaacccgccc	
celFhisf1	cacacgctgcacctcgacgacg	
bxyChisf1	tcgacgacgaccccgcccgtg	
celNhisf1	acgcgacccagaaccgcgcg	
SpecF3	aacgagctcaacgaggtgaaatcatg	Cloning of spectinomycin gene in
Spec R1	aaggctgcagcggccgctactagtaagcacctgttattgc	

		pVK168?
Edinbrick1	pSB1A2 backbone with BBa_J33207 insert	Insertion of *goi* in pSB1A2
pSB1C3	pSB1C3 backbone containing BBa_J33207-RBS Biobrick	Insertion of P_{lac}-RBS upstream of *goi*

Table 2.1. List of primers used for gene of interest (goi) cloning, inserting 6XHis-tags, and sequencing of inserts, as well as the plasmids used.

2.2.4 *BioBrick Construction*

This was a 2-step process, first to insert the PCR product into the plasmid backbone, removing the *lacZα* marker with the gene BioBrick, then adding P_{lac}, *lacZα* and a ribosome binding site (RBS) upstream of the gene BioBrick insert to allow inducible expression in a manner similar to figure 1.2.

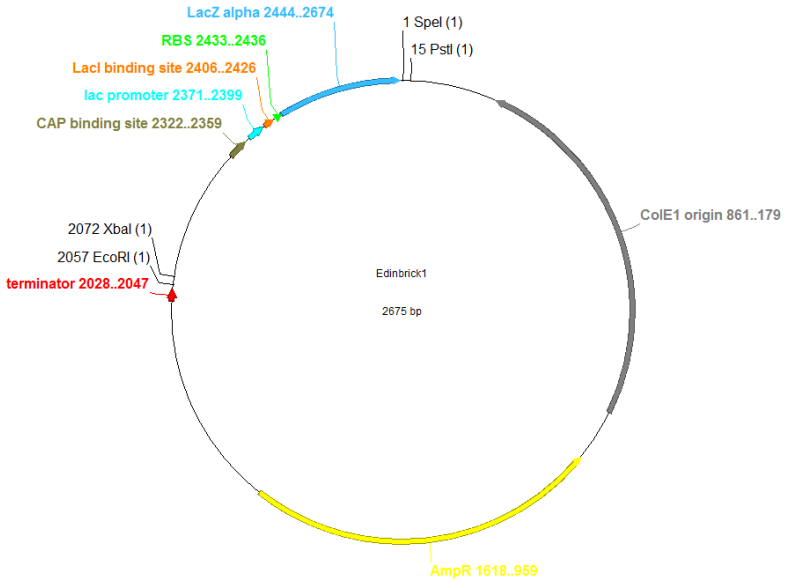


Figure 2.1. Graphic map of plasmid Edinbrick1 with all salient features, including the standard BioBrick restriction sites, highlighted. Arrows indicate the coding direction of the feature. Plasmid backbone pSB1A2 was ligated to BBa_J33207 (from *Eco*RI site, CAP binding site...*lacZα* – *Pst*I site) to create it. Map generated using ApE – a plasmid editor (http://biologylabs.utah.edu/jorgensen/wayned/ape/).

BioBrick-compatible plasmid Edinbrick1 (figure 2.1) was used to make the constructs as outlined in Introduction. PCR products and plasmid were digested with *Eco*RI-HF and *Spe*I for 2 hours at 37ºC in a 20 µl volume with appropriate NEB

supplied buffer, then purified from solution using the glass bead procedure, except adding 65 µl 6 M NaI, 5 µl glass bead suspension and eluting in 10 µl EB. To the eluate 1.2 µl T4 DNA ligase buffer was added followed by 1.2 µl T4 DNA ligase. The mixture was then incubated at 16ºC overnight to allow ligation to occur. Five microlitres of the ligation were used to transform competent *E. coli* JM109 cells. The cells were prepared according to the protocol of Chung et al [99], stored at -80ºC and thawed on ice for use. After the addition of the ligation mixture the cells were incubated on ice for one hour before being heat shocked at 42ºC for 90 seconds, and then back on ice for 90 seconds. The cell/ligation mixture was made up to 1 ml with LB and incubated at 37ºC for 1 hour before spreading on antibiotic selective plates, which were then incubated at 37ºC overnight. White colonies, those that have had the lac promoter and *lacZα* replaced with the gene of interest (figure 2.2), were patched to fresh plates of the same medium type and screened for the correct inserts by restriction digestion of plasmid recovered by miniprep.

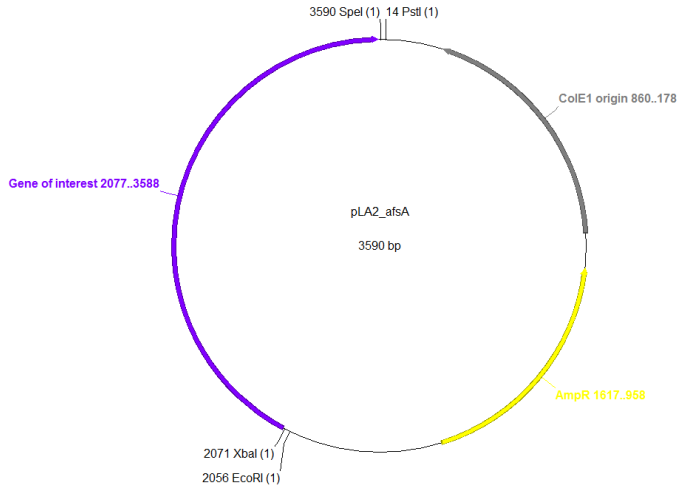


Figure 2.2. Graphic map of intermediate plasmid pA2_*goi* with the gene of interest (goi) inserted, replacing BBa_J33207. Map generated using ApE – a plasmid editor.

Plasmids were recovered either using the Qiagen miniprep kit or with the following protocol: 3 ml LB supplemented with antibiotic was grown overnight at 37ºC. Of this 1.3 ml was transferred to a microcentrifuge tube and the cells pelleted at 6000 rcf for 4 minutes and the supernatant discarded. The pellet was suspended by pipetting in 100 µl Solution 1 (5 µl of 5 mg/ml RNaseA in 1 ml H_2O). Cells were lysed by addition of 200 µl Solution 2 (0.2 M NaOH, 1% SDS) and mixing by inversion until

the solution became clear. Selective precipitation of cellular components and genomic DNA was achieved with the addition of 150 µl Solution 3 (3 M potassium acetate mixed with 2 M acetic acid), and immediate mixing by inversion. The mixture was incubated on ice for 10 minutes before centrifugation at 14000 rcf for 3 minutes. The supernatant was transferred to a clean microcentrifuge tube and 920 µl 100% ethanol was added. The mixture was incubated on ice for 10 minutes before centrifugation as before for 10 minutes. The supernatant was discarded and the pellet washed with 200 µl 70 % ethanol, centrifuged again for 1 minute, supernatant discarded and then centrifuged again for 1 minute and the last remnants of the ethanol removed. Pelleted DNA was solubilised in EB and stored at -20ºC.

Digested plasmids of the correct size were sequenced by the Edinburgh GenePool facility, School of Biological Sciences, University of Edinburgh, by the Sanger method using primers pSB1A2insf2 and r2 (table 2.1). Once confirmed by sequencing, the plasmid was digested with *Eco*RI-HF and *Xba*I, while a second plasmid (pLC3, figure 2.3 A) containing the P_{lac}-RBS construct was digested with *Eco*RI-HF and *Spe*I as previously described. The digests were purified from solution and 5 µl of each were mixed together and T4 DNA ligase added, followed by incubation. JM109 cells were transformed as described above, spread on ampicillin selective plates, incubated, and blue colonies patched and screened for correct plasmid size as above. Correctly sized plasmids were sent to GenePool for Sanger sequencing to confirm the upstream presence of P_{lac}-RBS and gene of interest using primers LacZf3 and pSB1A2insr2 (table 2.1).

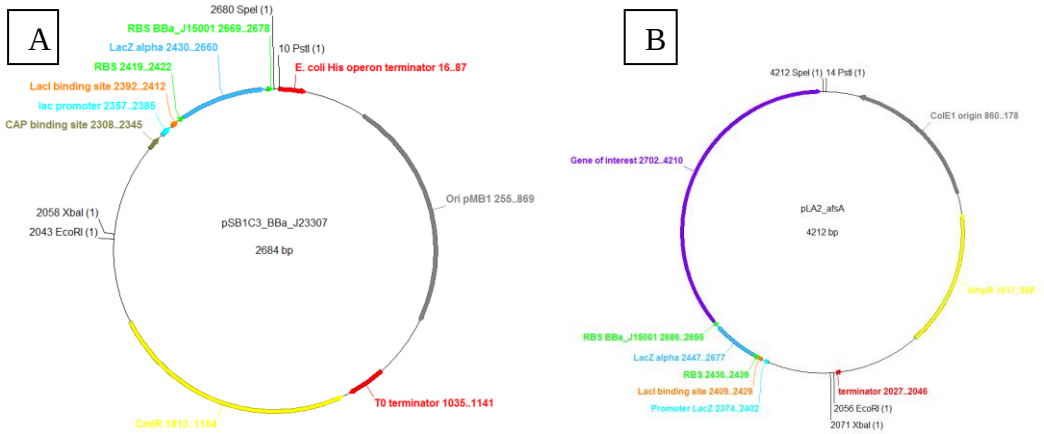


Figure 2.3. A) pLC3 - pSB1C3 with P_{lac}-RBS; B) Final construct after digestion and ligation of pLC3 and pA2_*goi* to give pLA2_*goi* as outlined above in materials and methods. Maps generated using ApE – a plasmid editor.

2.2.5 *Activity Assays*

2.2.5.1 *Protein Extraction*

Fifteen millilitre cultures of LB supplemented with IPTG and either ampicillin for *E. coli* or carbenicillin for *C. freundii* were inoculated and incubated overnight at 37ºC with shaking. The pSB1A2 plasmid contains the ampicillin resistance gene which also confers resistance to carbenicillin. *C. freundii* is naturally resistant to ampicillin but sensitive to carbenicillin. The cell suspension was centrifuged at 7000 rcf for 5 minutes to pellet the cells. The cells were then suspended in 500 µl of 1x PBS and 500 µl of 50% (v/v) glycerol. The cell suspension was lysed by sonication by 10 rounds of 5 seconds of sonication at 10 kHz, followed by at least 30 seconds of cooling in ice, with the cells being mixed by inversion between cycles. The lysate was then centrifuged at high speed for 10 minutes to pellet the cell debris and the soluble protein fraction transferred to a fresh centrifuge tube. Lysates were kept in ice and stored at -20ºC.

2.2.5.2 *Assay Procedures*

A reaction mixture of 15 µl 50 mM sodium acetate buffer pH, 5.0, 5 µl cell lysate and 5 µl of nitrophenyl substrate was routinely used. 2-nitrophenyl β-$_D$-xylopyranoside (ONPX), 2-nitrophenyl β-$_D$-cellobioside (ONPC), 2-nitrophenyl β-$_D$-galactopyranoside (ONPGal) and 4-nitrophenyl α-$_L$-arabinofuranoside (PNPA)

were used to assay for β-xylosidase, β-cellobiosidase, β-galactosidase and α-arabinofuranosidase activities, respectively. The twenty five microlitre reaction mixture was incubated for 1 hour at 37ºC, then placed on ice to stop the reaction. Immediately before taking the spectrophotometric reading at 405 nm, an equal volume (25 µl) of 1 M Na_2CO_3 was added and thoroughly mixed. All readings were taken using a NanoDrop 2000.

Endoglucanase and endoxylanase activity was assayed for using remazol brilliant blue (RBB) labelled carboxymethyl cellulose (CMC) and birch wood xylan, respectively. The RBB-CMC was purchased from Megazyme. RBB-xylan was produced as follows, as described at www.openwetware.org/wiki/Xylanase_Protocols#Preparing_the_Dyed_Xylan:

1. Five hundred milligrams of birch wood xylan (Sigma) and 500 mg of remazol brilliant blue were added to 12 ml of deionised sterile water and mixed thoroughly.
2. While being stirred four millilitres of (135 mg in 4 ml) sodium acetate (NaOAc) solution was added slowly over a 5 minute period.
3. Four millilitres of sodium hydroxide (NaOH, 300 mg in 4 ml) was added and the solutions mixed for 90 minutes at room temperature.
4. Two volumes of 96% (v/v) ethanol were added to precipitate the RBB-xylan.
5. This was then filtered with a vacuum flask and Whatman filter paper.
6. The solid material was washed with wash solution (132 ml EtOH, 66 ml H_2O, 270 mg NaOAc) until the flow through ran clear.
7. The precipitate was then washed with 20 ml 75% (v/v) ethanol, and 4 ml of 100% acetone.
8. The RBB-Xylan was left to air dry overnight. The dry weight was measured, autoclaved and dissolved in sterile de-ionised water to a final concentration of 5% (w/v) and stored at 4ºC.

Reaction volumes of one hundred microlitres were used for the assays containing 0.2% (w/v) RBB-CMC or RBB-xylan, 20 µl cell lysate, and then made up to 100 µl with 50 mM sodium acetate buffer, pH 5.0. The mixture was thoroughly mixed by

pipetting and incubated for 2 hours at 37ºC. The reaction was stopped with the addition of three volumes of 100% ethanol and thoroughly mixed. After a one minute incubation at room temperature the mixture was centrifuged at high speed for 10 minutes to pellet the precipitated polysaccharides. The soluble fraction was spectrophotometrically measured using a NanoDrop 2000 at 590 nm.

Lysates were assayed for β-glucosidase activity with 4-methylumbelliferyl β-$_D$-glucopyranoside (MUG). Ten microlitres of cell lysate was added to 85 µl of 50 mM sodium acetate, pH 5.0 and 5 µl of MUG (5 mg /ml in H_2O). The lysates were then incubated at 37ºC for 1 hour and the presence of released methylumbelliferol was measured by absorbance at 348 nm using a NanoDrop 2000.

2.2.5.3 *Protein Estimation*

Protein levels for the cell lysates were estimated by Bradford assay using a BSA standard curve. Pierce Coomassie Reagent (Thermo Scientific 23200) was mixed with protein sample and absorbance measured at 595 nm using a NanoDrop 2000. For concentrations estimated to be between 100 and 2000 µg/ml, 5 µl of sample was added to 250 µl of Coomassie reagent. For concentrations estimated to be lower than 100 µg/ml, equal volumes of reagent and sample were mixed, typically 20 µl plus 20 µl.

2.2.6 *His-Tag Purifications*

2.2.6.1 *Tagging and Protein Extraction*

Proteins had 6xHis-tags added to their N-terminus by PCR amplification of the whole plasmid using Kod Xtreme and the following primers. Gene-specific forward His primers and RBS-HisbluntR1 reverse primer were used to insert the 6xHis tag between amino acid 1 (ATG, methionine) and amino acid 2 in the protein (table 2.1). PCR products of the correct size were purified and blunt-end ligation performed (6 µl H_2O, 1 µl PCR product, 1 µl Promega T4 DNA ligase buffer (+ATP already), 1 µl T4 DNA ligase, 1 µl T4 polynucleotide kinase) at 16ºC overnight. Five microlitres of the ligation were used to transform *E. coli* JM109 as before and transformants

were patched to fresh selective plates. After confirmation by sequencing the transformant was cultured in a 100 ml volume plus appropriate antibiotic and expression induced with IPTG. The soluble protein fraction was obtained as before (2.2.5.1) with one exception – cells were lysed in 25 mM Tris-HCl, pH 8.0, 50 mM imidazole, 0.3 M NaCl, 10% (v/v) glycerol and 8 M urea (lysis buffer). For native protein purifications the above steps were followed, but the lysis buffer did not contain urea.

2.2.6.2 *Purification*

A column was packed using Ni-NTA (Qiagen 30210) to give a column volume (CV) of 1 ml. The column was washed with 10 CVs of H_2O and equilibrated using 6 CVs of lysis buffer. The one millilitre cell lysate volume was loaded on the column and the flow through and all subsequent fractions kept for analysis. The column was then washed three times with 2 CVs of lysis buffer. Elutions were performed in a stepwise fashion with increasing concentrations of imidazole. First three times 1 ml of lysis buffer made up to 0.5 M imidazole, and then three times 1 ml lysis buffer made up to 1 M imidazole. The column was then regenerated by loading 0.4 M NaOH for 30 minutes, rinsing thoroughly with a large volume of water and then storing the column at 4ºC in 20% (v/v) ethanol. Fractions were analysed by 4-12% SDS-PAGE, and stained with Coomassie Blue and, if needed silver stain following the protocol of Gromova and Celis [100].

2.3 Purification and Characterisation

2.3.1 *Cell Cultures*

Escherichia coli JM109 transformed with pLA2*_afsB/xynF/xynH/bxyF* (pLA2 = pSB1A2*_*P_{lac}*_RBS* construct) and *Citrobacter freundii* NCIMB11490 transformed with pLA2*_celD/celA* were grown in 50 ml volumes of LB containing 100 µg/ml ampicillin (100 µg/ml carbenicillin for *C. freundii*) and 90 µg/ml IPTG overnight. The cells were recovered by centrifugation at 7000 rcf for 5 minutes and then the

pellet suspended in 1 ml of Binding Buffer (25 mM Tris-HCl, pH 8.0, 25% glycerol). The cells were lysed by sonication as previously described (chapter 2.1.5.1).

2.3.2 *Anion Exchange Chromatography*

All purification steps were performed at 4ºC. A column was packed with DEAE-sepharose to a volume of 4 ml. This was first washed with a large volume of water. The column was equilibrated using 5 column volumes (CV) of Binding Buffer, until the pH of buffer eluted from the column equals that of the buffer applied. The sonicated cell lysate was loaded on the column and washed through with 3 ml of BB. The flow through (FT) and all subsequent fractions were collected. The column was washed three times with 1 CV of BB. The protein was eluted in a stepwise fashion, 1 CV of BB containing NaCl in 0.1, 0.2, 0.3, 0.4 and 0.5 M concentrations unless otherwise stated. The column was regenerated with one CV of 5 M NaCl, one CV of 1 M NaOH, rinsed thoroughly with water and then stored in 20% ethanol at 4ºC. Fractions were assayed for enzyme activity as previously described. Elution fractions showing activity were made up to a concentration of 1 M NaCl and used for the next purification step.

2.3.3 *Hydrophobic Interaction Chromatography (HIC)*

A column was packed with phenyl-agarose to a volume of 2 ml. This was washed with a large volume of water. The column was equilibrated with 5 CV of BB2 (BB with 1 M NaCl), and the DEAE-fractions exhibiting enzyme activity were passed through the column. The column was washed three times with one CV of BB2. The proteins were then eluted in a stepwise fashion, with 3*1 ml of BB with 0.5, 0.1, 0.08, 0.06, 0.04, 0.02 and 0 M concentrations of NaCl. Fractions were assayed for enzyme activity as previously described and all fractions were stored at -20ºC.

2.3.4 *Cation Exchange Chromatography*

XynF was purified on CM-sepharose after being passed through DEAE-sepharose. The active DEAE fractions were unaltered before being passed through the column. The column was washed in 3 times 1 CV of BB end eluted using 3 times 1 ml BB plus 0.05 M, 0.1 M and 1 M NaCl. Fractions were assayed for enzyme activity as before. Fractions were analysed by SDS-PAGE to observe the quantities of contaminating proteins present, and the denatured molecular weight of the protein of interest where possible. All gels were stained using 0.1% (w/v) Coomassie Blue in 40% (v/v) methanol, 10% (v/v) glacial acetic acid.

2.3.5 *Enzyme Characterisation*

All fractions were assayed for protein concentration using the Bradford assay as previously described (2.2.5.2). After HIC purification if more than one fraction demonstrated enzyme activity, the fraction exhibiting the highest activity:protein ratio was used for characterisation assays.

2.3.5.1 *pH Optima*

Firstly, the optimum pH for the enzymes was measured for. Enzymes were assayed as previously described but with increasing pH levels and appropriate buffers. For pH 4.0 - 5.5, 50 mM sodium acetate buffer, pH 6.0 - 7.5 50 mM potassium phosphate buffer, pH 8.0 – 9.0 20 mM Tris-HCl buffer and pH 9.5-10.5 50 mM glycine buffer was used. All assays were performed at 37ºC.

2.3.5.2 *Temperature Optima*

The optimum pH for the enzyme was used for all temperature experiments. The assays were performed as previously described but at increasing temperatures from 15ºC - 70ºC (or higher if needed) in 5ºC steps using a DNA Engine PCR machine (MJ Research) and 0.2 ml thin walled PCR tubes (Axygen).

2.3.5.3 *Effect of Supplements*

The optimum pHs and temperatures were used for all subsequent assays. The assays were performed as previously described but with 14 µl of buffer. The remaining 1 µl was one of the following supplements at a final concentration of 1 mM – $CaCl_2$, $CoCl_2$, $CuSO_4$, $FeSO_4$, $MgCl_2$, $MnCl_2$, $ZnSO_4$, DTT, EDTA, glucose, cellobiose, xylose or arabinose. $FeSO_4$ was made fresh on the day of use as precipitate would form and obfuscate the OD 405 nm readings. The effects of NaCl were assayed for using non-purified enzymes at optimum pH and temperatures as the purified proteins are already in the presence of high concentrations of salt. All incubations were performed in duplicate with enzyme negative controls. Activity was measured relative to the maximum activity observed for a reaction, minus any enzyme negative background observed. For the purification tables, standard curves of 2- and 4-nitrophenol at 0, 50, 100, 250, 500, 750 and 1000 µM concentrations at pH 5.5 were used to estimate extinction coefficients and fraction activities. A unit was defined as the amount of protein required to release 1 µmol of nitrophenol from its bound sugar per minute.

2.4 Transformation of *Cellulomonas fimi*

2.4.1 *Plasmids Used*

The following plasmids were used in the transformation experiments of *C. fimi*.

Plasmid	Resistance Marker	References
pVK168	Spectinomycin	Joshi et al, 2009 [56]
pTG262	Chloramphenicol	Shearman et al 1994 [101], parts.igem.org/Part:BBa_I742123
pCJW93	Apramycin, Thiostrepton	Wilkinson et al, 2002 [102]

Table 2.2. Plasmids used within this study, the antibiotic resistance conferred and relevant references where available

2.4.2 *Transformation Procedures*

2.4.2.1 *Electroporation*

2.4.2.1.1 Electrocompetent Cell Preparation:

Minimal M9 medium with 10 mM cellobiose as carbon source and 50 μg/ml kanamycin was inoculated with a single colony of *C. fimi* and grown to a dense culture over 3 days. This was then used to inoculate 30 ml fresh M9 medium, with same supplements as before, by a dilution of 1/50. The cultures were grown to an OD_{600} 0.4-0.6 and then treated in one of the following ways:

"Base" – cells were recovered by centrifugation at 7000 rcf for 5 minutes at room temperature. The pellet was suspended in 1 ml PBS, and glycerol 10% (v/v), aliquots of 200 μl prepared and stored on ice until use the same day or stored at -80ºC.

"Lyso" – cells were recovered by centrifugation at 7000 rcf for 5 minutes at room temperature, then suspended in 1 ml of 10 mg/ml lysozyme in PBS and incubated at 37ºC for 10 minutes. The suspension was centrifuged to pellet the cells and washed three times and stored as for Base cells.

"Gly" – once the correct OD_{600} was reached the culture was supplemented with filter-sterilised glycine to a final concentration of 2% (w/v) and incubated for a further 3 hours. After this point the cells were pelleted and treated as for Base cells.

"Pen-G" – once the correct OD_{600} was reached the culture was supplemented with 10 μg/ml filter-sterilised penicillin-G and incubated for a further 3 hours. After this point the cells were treated as for Base cells.

"Tween" – once the correct OD_{600} was reached the culture was supplemented with Tween-80 to a final concentration of 0.1% (v/v) and incubated for a further 3 hours. After this point the cells were pelleted and treated as for Base cells.

Prior to electroporation the cells were pelleted at 7000 rcf for 5 minutes at room temperature and washed by suspension in 1 ml ice-cold sterile water and pelleted again. This was repeated three times. The washed cells were then suspended in 200 μl ice-cold water and then kept on ice.

2.4.2.1.2 Transformation

Electrocuvettes with a 0.1 cm gap (BioRad) were chilled on ice. A cell aliquot of 200 µl was halved with one half going to a cuvette (DNA negative control), and to the other 5 µl of plasmid DNA was added, mixed by gentle pipetting and then transferred to a cuvette. These cuvettes were then kept on ice for 10 minutes. The cells were then pulsed using a BioRad MicroPulser at 2.5 kV. Immediately after pulsing 400 µl of pre-warmed (40ºC) nutrient broth was added and the cells thoroughly suspended and transferred to a fresh 1.5 ml micro-centrifuge tube. The cells were then incubated at 40ºC for two minutes and then returned to ice for 90 seconds, then incubated for 3 hours at 37ºC to recover and allow for expression antibiotic resistance proteins. The full amount of cells was then plated to nutrient agar plates containing 50 µg/ml kanamycin and appropriate plasmid antibiotic (table 2.2). The plates were incubated for 4-7 days at 37ºC and then the colonies present were recorded.

2.4.2.2 *Protoplasts*

2.4.2.2.1 Protoplast Transformation

Cultures of *C. fimi* were grown as above to mid-log phase, supplemented with penicillin-G and incubated for a further 3 hours before being pelleted. Protoplasts were formed following one of three protocols:

That of Pulido-Vega who followed the protoplast formation and regeneration procedure of Chang and Cohen, with some alterations [103, 104]:

Harvested cells were suspended in one tenth the original culture volume of SMMP (2x M9 minimal medium, 10 mM cellobiose, 0.5 M sucrose, 0.02 M maleate, 0.02 M $MgCl_2$, pH 6.5 adjusted with NaOH) plus lysozyme prepared in PBS to a final concentration of 2 mg/ml. Cells were incubated with shaking for 2 hours at 37ºC before cells were recovered by centrifugation at 2600 rcf for 15 minutes at room temperature. The supernatant was discarded and the pellet suspended in the same volume of SMMP. This pelleting and suspension was again repeated. One hundred microlitres of cell suspension were spread on nutrient agar plates to test for

osmotically stable survivors, and a further 100 µl plated to recovery medium (per litre, pre-sterilised before mixing: 200 ml 4% agar, 500 ml 1 M sodium succinate, pH 7.3, 100 ml 5% Casamino acids, 50 ml 10% YE, 100 ml 3.5% K_2HPO_4 and 1.5% KH_2PO_4, 25 ml 20% glucose, 20 ml 1 M $MgCl_2$, 5 ml filter-sterilised 2% BSA) to test for protoplast recovery.

PEG Mediated Transformation procedure of Pulido-Vega et al:
Five microlitres of plasmid DNA was mixed with 95 ul 2x SMM (SMM: 0.5 M sucrose, 0.02 M maleate, 0.02 M $MgCl_2$, pH 6.5 adjusted with NaOH) followed by the addition of 0.5 ml protoplast suspension and 1.5 ml 40% (w/v) PEG 6000 in SMMP. This mixture was incubated at room temperature with gentle mixing for 5 minutes. Twenty millilitres SMMP was added to dilute the PEG and centrifuged at 1240 rcf for 20 minutes to pellet the cells. The pellet was suspended in 0.5 ml SMMP and incubated for 3 hours at 37ºC with gentle shaking. Aliquots of 400 µl and 100 µl were spread on recovery media plus 50 µg/ml kanamycin and the appropriate antibiotic.

The following procedure is that of Thierbach et al [105] who used media MMYE, RCG agar and TSMC are described by Katsumata et al [106]:
Pelleted cells were suspended in one fifth the original culture volume of LF (MMYE diluted twice and supplemented with 0.41 M sucrose, 0.01 M $MgSO_4$ and 10 mg/ml lysozyme) and incubated for 2 hours at 37ºC with mild shaking. The suspension was centrifuged at 2600 rcf for 15 minutes at room temperature and the pellet suspended again in the same volume of LF and incubated for a further 1.5 hours. After this incubation 4/5 the original culture volume of TSMC (50 mM Tris-HCl, pH 7.5, 0.5 M sucrose, 10 mM $MgCl_2$, 20 mM $CaCl_2$) was added. The protoplasts were pelleted at 3000 rcf for 10 minutes at room temperature. The supernatant was discarded and the pellet suspended in 5 ml TSMC, centrifuged at 1000 rcf for 10 minutes at room temperature and the pellet suspended in 3 ml TSMC. A fifty microlitre aliquot was spread on nutrient agar to test for osmotically stable survivors, and 50 µl plated on RCG/E for protoplast regeneration (per litre: 5 g glucose, 5 g casamino acids, 2.5 g YE, 3.5 g K_2HPO_4, 0.4 g $MgCl_2.6H_2O$, 30 µg biotin, 2 mg thiamine.HCl, 30 g

polyvinylpyrolidone (MW 10000), 135 g sodium succinate, 1 ml trace elements, 16 g agar, pH 7.6, overlay: 3 ml 0.3% low melting point agarose).

MMYE (per litre):

20 g glucose, 10 g $(NH_2)SO_4$, 3 g urea, 1 g YE, 1 g K_2HPO_4, 0.4 g $MgSO_4.7H_2O$, 2 mg $FeSO_4.7H_2O$, 2 mg $MnSO_4.4.6H_2O$, 50 mg NaCl, 50 μg biotin, 200 μg thiamine, pH 7.2

Trace elements (per litre):

88 mg $Na_2B_4O_7.7H_2O$, 37 mg $(NH_4)6Mo_7O_{27}.4H_2O$, 8.8 mg $ZnSO_4.4H_2O$, 270 mg $CuSO_4.5H_2O$, 7.2 mg $MnCl_2.4H_2O$, 970 mg $FeCl_3.6H_2O$

PEG Mediated Transformation Procedure of Thierbach et al:

Five microlitres of plasmid were added to 100 μl of protoplast suspension followed immediately by 0.7 ml of P1 (2.5 g PEG (6000) dissolved in 7.5 ml TSMC). This was incubated at room temperature for 3 minutes then 7.2 ml of pre-warmed TSMC_2 (TSMC supplemented with 0.1% (w/v) glucose, 0.1% (w/v) YE, 0.1% (w/v) casamino acids) was added and the mixture incubated at 37ºC for 3 hours. The protoplasts were pelleted at 3000 rcf for 10 minutes and suspended in 0.5 ml of TSMC. Aliquots of 400 and 100 μl were spread on RCG/E plates with appropriate antibiotic.

2.4.2.2.2 Protoplast Fusions

The following describes the protocol as followed by Rodriguez et al [107]:

The pelleted cells were washed twice in 5 ml 0.9% (w/v) NaCl and finally suspended in 5 ml HM (0.5 M sodium succinate, 20 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.4, 0.01 M EDTA) plus 10 mg/ml lysozyme. This suspension was incubated for 30 minutes at 37ºC without shaking. EDTA was then added slowly over a 20 minute period from a 0.1 M stock to a final concentration of 10 mM, and incubated for a further 20 minutes. The cells were then centrifuged at 2000 rcf for 10 minutes to pellet and suspended in 5 ml of HM. Aliquots as above were taken and plated on nutrient agar for osmotically stable survivor count and on RM (nutrient agar lower layer, upper layer of 0.5 M sucrose, 20 mM $MgCl_2$, 0.8% (w/v) agar) for protoplast regeneration count.

E. coli Protoplast Formation

The following describes the protocol as followed by Dai et al [108]:

An overnight culture of *E. coli* JM109 transformed with pSB1A2 expressing either LacZα or CenA was used to inoculate 30 ml of LB with ampicillin and grown to and OD_{600} of ~1. Cells were recovered by centrifugation at 3000 rcf for 10 minutes and the cells suspended in 5 ml of 10 mM Tris-HCl, pH 8.0. This step was repeated three times in total and finally suspended in 30 ml of Tris-HCl, pH 8.0, 0.5 M sucrose. EDTA was slowly added over a 20 minute period to a final concentration of 10 mM while being incubated with gentle shaking at 37ºC. After the final EDTA addition the cells were incubated for a further 20 minutes. The cells were recovered by centrifugation at 3000 rcf for 10 minutes, washed twice in 5 ml SMM and finally suspended in 30 ml SMM plus 1 mg/ml lysozyme. This cell suspension was then incubated for 1 hour at 37ºC, the cells pelleted by centrifugation at 2000 rcf for 20 minutes and finally suspended in 5 ml SMM. Aliquots were spread on Luria agar with ampicillin for osmotically stable survivor count and onto EcR plates (Luria broth, 8% (w/v) agar, 0.5 M sucrose) for protoplast regeneration count.

2.4.2.2.3 Electrotransformation of Protoplasts

From the *C. fimi* protoplast cell suspension, 0.5 ml was transferred to a fresh centrifuge tube and centrifuged at 7000 rcf for 10 minutes to pellet the cells and then suspended in 50 µl HM. Five microlitres of plasmid DNA was added and the mix and incubated on ice for 10 minutes. The cells were transferred to pre-chilled electrocuvettes, pulsed at 0.5 kV and 450 µl nutrient broth (with 0.5 M sucrose, 20 mM $MgCl_2$) was added immediately after. The suspension was heated for 90 seconds at 40ºC then on ice for 90 seconds before incubation at 37ºC for 3 hours. The cells were then plated to RM with plasmid appropriate antibiotic.

2.4.2.2.4 PEG Mediated Protoplast Fusions

Following the protocol of Rodriguez et al [107]:

As for the electrofusion protocol, 0.5 ml of *C. fimi* and *E. coli* protoplasts were mixed together, centrifuged to pellet and suspended in 0.3 ml of HM. PEG 6000 (55% w/v) in 2x HM was added to a final volume of 3 ml and the suspension mixed

by inversion. The cell suspension was then incubated at 37ºC for 10 minutes, diluted by the addition of 7 ml of nutrient broth, 1 M sucrose, 40 mM $MgCl_2$, and centrifuged to pellet the cells at 2000 rcf for 10 minutes. The protoplasts were then suspended in 0.5 ml HM and plated on RM with ampicillin.

2.4.3 *Colony PCR*

A small quantity of cells was removed from a colony using a sterile 200 µl pipette tip and the cells suspended in 50 µl of sterile water. This was used as the DNA template by addition of 1 µl of the cell suspension to the PCR reaction mixture. SpecF3 and SpecR1 primers were used for amplification by KoD polymerase following the protocol previously described (section 2.2.2).

2.4.4 *Congo Red Assay*

Single colonies from transformant plates were patched to agar plates containing 50 µg/ml kanamycin, 100 µg/ml spectinomycin and 0.2% (w/v) CMC and incubated for 5 days at 37ºC. Staining was performed by covering the plate with 1 mg/ml Congo red dissolved in water and incubating at room temperature for 15 minutes. Excess dye was removed and the plate covered with 5 M NaCl for 15 minutes at room temperature. The NaCl was removed and the plates were visualised under white light.

CHAPTER 3: GENOMIC ANALYSIS OF *Cellulomonas fimi*

There have been fourteen cellulolytic enzymes cloned from *Cellulomonas fimi*, expressed in *Escherichia coli* and characterised. The first were an endoglucanase and an exoglucanase, CenA and Cex (XynB), discovered in 1986 by Wong et al and O'Neill et al, respectively [109, 110]. These were followed largely in the 1990s by the discovery of CenB and CenC by Meinke et al and Coutihno et al, respectively, in 1991 [70, 111]; CenD by Meinke et al in 1993 [112]; CbhA and XynD in 1994 by Meinke et al and Millward-Sadler et al, respectively [76, 78]; CbhB and XynC were described in 1995 by Shen et al and Clarke et al, respectively [77, 79, 113]; Man26A (ManD) and Man2A (ManA) by Stoll et al in 1999 [82]; Cfx (XynE) by Hekmat et al in 2005 [80]. Besides these cellulolytic enzymes Mayer et al latterly cloned and characterised in 2006 a β-$_N$-acetylglucosaminidase Hex20 (NagB) and a β-glucosidase Nag3 (NagA) [114]. Since then there have been no further publications describing novel lignocellulolytic enzymes from *C. fimi*.

The genomes of other lignocellulolytic organisms have been sequenced and their cellulolytic genes predicted. One of the first to be sequenced was the genome of the cellulolytic fungus *Trichoderma reesei* [115]. This organism has been the source of commercial cellulases and hemicellulases, has been studied for decades and is the archetypal cell-free cellulase system. In the cited paper they compare the cellulolytic genes found within *T. reesei* with other cellulolytic fungi that have been sequenced. They identified hundreds of glycoside hydrolase (GH) genes, 36 of which contained a carbohydrate binding module (CBM), but these numbers were much lower than those found in other fungal species, such as *Aspergillus fumigatus* which reportedly contained 263 GH enzymes and 55 CBM containing enzymes. Obviously not all were associated with cellulolytic activities, but the authors identify 16 hemicellulase genes and 10 cellulolytic genes specifically, again at the lower scale of those compared fungal species.

Clostridium thermocellum is an anaerobic cellulolytic bacterium that has been the paradigm of anaerobic cellulose degradation – the cellulosome. As such it too has been sequenced and studied with enzymes being of interest due to their thermostability and activity levels [116, 117]. The genome was sequenced revealing 60 ORFs capable of binding to scaffoldin, the protein by which the components of the cellulosome are anchored to the cell surface [116]. Gold et al reported that of these 60 potential cellulolytic genes, only about a third have been identified through experimental procedures such as Western blots and standard enzymatic assays. They also report that the cellulosome differs depending on the carbon source – avicel or cellobiose, reporting a core of 29 common proteins, but 6 specific to avicel induction and 5 specific to cellobiose, through mass spectrometry.

More recently the genome of *Cellulomonas flavigena*, a species relatively closely related to *C. fimi*, was sequenced and analysed by Abt et al [118]. This again revealed a wealth of putative cellulosic biomass degrading genes, with 14 potential endoxylanases, 5 β-xylosidases, 3 cellobiohydrolases, 3 β-glucosidases and 2 endoglucanases. However, the full extent of putative cellulosic biomass degrading genes was not reported, with the paper only just touching on cellulose degradation.

From looking at the raw numbers of putative cellulose and hemicellulose hyrdolysing genes in these organisms, it would suggest that *C. fimi* holds the potential for discovery of further lignocellulolytic enzymes. *T. reesei* contains at least 26 genes, *C. thermocellum* a possible 60 genes and *C. flavigena* at least 27 genes with putative cellulosic biomass degrading activity. In *C. fimi*, as stated above, there have been fourteen such enzymes identified. This suggests that there are many potential genes and enzymes still to be identified. The paucity of enzymes so far identified lies with the methods used so far. The first enzymes identified, CenA, CenB and Cex were discovered by screening of a lambda genomic library expressed in *E. coli* under IPTG induction and screening for endo-cellulolytic activity on avicel or CMC [109, 119, 120]. Similarly XynC and XynD were identified by screening libraries using oat-spelt xylan [79, 121], with XynD then adsorbed to avicel to purify from *E. coli* cell extract [78]. Stoll et al screened their library using azo-carob galactomannan

(ACG) to identify mannanase activity and 4-methylumbelliferyl-β-$_D$-mannoside to identify mannosidase activity [82]. Finally, Mayer et al grew *C. fimi* on chitin, identified intracellular chitinase activity and screened a library using 4-methylumbelliferyl β-$_N$-acetyl-$_D$-glucosaminide (MU-GlcNAc), identifying NagA and NagB [114].

Attention was then turned to the secreted cellulases with the culture supernatants of *C. fimi* being screened. CenC was identified after the supernatant was passed through a sephadex resin and eluted fractions tested for CMC-cellulase activity [111, 120]. CenD was identified when a 10 day culture of *C. fimi* was grown with CMC, and the enzymes adsorbed to avicel, BMCC (bacterial microcrystalline cellulose) or sephadex, and analysed by SDS-PAGE [112]. This method identified five cellulose binding polypeptides (cbps); CenA and CenB were identified by N-terminal sequencing as 2 of them. The other three were novel enzymes. One at 75 kDa, was termed CenD and characterised [112]. A second cbp at 95 kDa was later characterised and called CbhA [76], and the final cbp at 120 kDa was later still characterised and named CbhB [77]. Latterly mass spectrometry was used to identify Cfx (XynE) in the supernatant of *C. fimi* when grown on xylan for 3 days using retaining endo-glycosidase specific inactivating dyes [80].

It's these very methods that are probably the reason for the modest selection of enzymes identified, and that no more have been identified in the last 8 years. The use of a lambda genomic library requires the digestion of genomic DNA, the ligation of the fragments into plasmids and the transformation of the host organism, *E. coli* for the above papers, and then expression of these genomic inserts. Large fragments may not be able to persist in the host, or induced expression may not include distal genes, missing their activities. Conversely, genes may be missed as fragments are too small, or dissect the gene, so no functional enzyme is expressed. If there is a gene present there is no guarantee that the gene may be expressed. One possible reason for poor expression is rare codon usage in the heterologous gene, highly likely as *C. fimi* is so GC-rich in comparison to the *E. coli* genome. Nor is it possible to be sure that the whole genome is present within the library.

Methods that use a resin such as sephadex or crystalline cellulose to isolate extracellular enzymes rely on the ability of the enzymes to actually bind the substrate. Not all cellulolytic genes have a CBM that could bind to cellulose or the dextran in sephadex. Of those characterised from *C. fimi* ManA, NagA and NagB have no CBM and EgIII of *T. reesei* also lacks a CBM [122]. Mass spectrometry is a powerful tool for the identification of enzymes in complex mixtures and has been successfully used to identify cellulolytic enzymes [116, 123-126]. Hekmat et al have purely focussed on enzymes that bind their inactivating dyes, specific for retaining enzymes only. This therefore ignores inverting enzymes that may be present in the mixture. These latter techniques are also just focussing on extracellular enzymes that are expressed when either cellulose or xylan are the carbon sources. Any possible intracellular enzymes, or those not expressed in these conditions are therefore going to go undetected. Differential expression dependent on carbon source has been demonstrated in several organisms and in *C. fimi* [127, 128]. The screening methods outlined above use either cellulose or xylan to identify activity in the library or sephadex/avicel fractions. The use of ACG and MU-β-Man by Stoll et al identified a mannanase and mannosidase. Any enzyme not active specifically on cellulose or xylan therefore will go, and probably have gone, undetected with these as substrate, as will any that act purely as exoglucanases.

3.1 Analysis of the Cellulolytic Ability of *C. fimi*

The complete genome was released by the US DOE and uploaded to the NCBI database at the start of 2012. This has allowed a fuller analysis of the cellulolytic capabilities of *C. fimi*, its metabolic pathways and potential promoter DNA motifs that could be of use for heterologous gene expression. Table 3.1 shows the major predicted cellulolytic and hemicellulolytic genes present within the *C. fimi* genome. By using specific keyword searches against the NCBI database, such as "cellulase", "glycoside", "cellulose" and "xylan", I have been able to identify over 100 putative polysaccharide degrading genes. The count of genes identified as having a particular glycoside hydrolase (GH family), carbohydrate binding module (CBM) and cellular localisation are given also. The full list including gene positions is given in table A1

(appendix), with names given by our lab derived from putative annotations and BLAST analysis.

Gene Product	GH Families	CBM Family	Cellular Localisation
Endo-1,4-β glucanase/cellulase	GH_5 – 3 (CenD) GH_6 – 2 (CenA) GH_9 – 5 (CenB, CenC) GH_16 – 2 GH_64 – 1 GH_81 – 2 Endoglucanase -E type – 1	CBM_2 – 5 CBM_2 and CBM_3 – 1 CBM_4_9 - 2 CMB_6 – 1 None – 3 Ricin-B-lectin – 4	Extracellular – 11 Cytoplasmic – 1 Unknown – 1 Cytoplasmic/ membrane – 1 Cell wall/ extracellular – 1
1,4-β cellobiohydrolase	GH 6 – 2 (CbhA) GH 48 – 1 (CbhB)	CMB_2 – 2 None – 1	Extracellular – 2 Unknown – 1
Endo-1,4-β xylanase	GH_10 – 5 (XynB, XynC, XynE) GH_11 – 1 (XynD) GH_43 – 1	CBM_2 – 3 CBM_4_9 – 1 Big_3_4 – 1 Ricin-B-lectin – 1 None – 1	Extracellular – 7
Xylan-1,4-β xylosidase	GH_39 – 1 GH_43 – 4	α/β hydrolase_3 – 1 Ricin-B-lectin – 1 None – 3	Extracellular – 1 Cytoplasmic – 1 Unknown – 1
α-$_L$-arabinofuranosidase	GH_43 – 1 GH_62 – 1 None – 3	Ricin-B-lectin – 1 None – 4	Extracellular – 1 Cytoplasmic – 4
β Mannanase/β Mannosidase	GH_2 – 1 (ManA) GH_26 – 1 (ManD) None – 2	CBM_2 – 1 None – 3	Cell wall/ Extracellular – 1 Cytoplasmic – 1 Unknown – 1
Pectate Lyase	None – 4	Ricin-B-lectin – 2 None – 2	Extracellular – 3 Cell wall/ Extracellular – 1
Chitinase/Endo-β-N-	GH_3 – 1	CBM_2 – 1	Extracellular – 1

acetylhexoaminidase	GH_18 – 1 GH_20 – 1 None – 1	None – 3	Cytoplasmic – 3

Table 3.1. Breakdown of putative cellulolytic and hemicellulolytic genes and their products from the *Cellulomonas fimi* ATCC 484 genome. Gene products based on sequence similarities through BLAST. Glycoside hydrolase (GH) and carbohydrate binding module (CBM) families from sequence homology to other family members, predicted by InterProScan [97]. Cellular localisation of the gene product predicted using Psortb-V3 [98].

By searching the genome for common enzymatic activities associated with cellulose hydrolysis I have collated a list of genes coding for putative polysaccharide degrading enzymes. The genome was annotated automatically using Prodigal to identify open reading frames, followed by protein function prediction using protein domains in Pfam, Swiss-Prot, TIGRFAMs, KEGG, InterPro and COG [129]. Any identified genes of interest, those with putative cellulolytic or hemicellulolytic activity, were translated to amino acid sequences (www.expasy.org) then BLAST searched and named according to the most common hits. In total 103 putative polysaccharide degrading genes were identified including those for all previously known enzymes. Before the genome was sequenced only four endoglucanases were known for *C. fimi* and incorporated only GH families 5 (CenD), 6 (CenA) and 9 (CenB and CenC). After analysis this number has grown to 16 in total including genes annotated as "cellulases", which contain a "cellulase" domain, of GH 5. The scope of GH families has also increased with GH 16, 64 and 81 present. A novel GH 6 cellobiohydrolase was also identified, along with a further 3 endoxylanases, 2 of GH family 10 and 1 of GH family 43. Previously there were 2 reported mannanase/mannosidase enzymes (ManA and ManD, GH 2 and 26, respectively) but this list now incorporates a further 2 enzymes both belonging to GH 26. The only other reported *C. fimi* enzymes are NagA and NagB, a GH 3 multifunctional β-glucosidase and GH 20 acetylhexoaminidase, respectively, with a further 10 GH 3 putative β-glucosidases, a GH 18 chitinase, and another acetylhexosaminidase of no predicted GH family.

No β-xylosidase, arabinofuranosidase or pectate lyase enzymes have been reported previously from *C. fimi* despite these being essential for efficient degradation of lignocellulosic biomass. Genome analysis however identifies six putative genes

annotated as β-xylosidase, 5 GH 43 and 1 GH 39, 5 annotated as arabinofuranosidase genes, 1 GH 43, 1 GH 62 and 3 with no apparent GH family annotation, and 4 annotated as pectate lyase genes. Further to these *C. fimi* also putatively encodes genes for 2 GH 43 arabinanases, 3 acetyl xylan esterases, a large raft of amylases, galactosidases, rhamnogalacturonan lyases, as well as four extracellular catalytic domain free cellulose binding peptides (CBMII).

Christopherson et al have compared the genomes of *C. fimi* ATCC 484 and *Cellvibrio gilvus* ATCC 13127 in their recent paper [129]. The numbers and general features of their analysis are largely consistent with what I found independently. When compared to other cellulolytic actinomycete bacteria *C. fimi* has potentially slightly higher numbers of putative cellulases and xylanases, 19 and 12, respectively. Abt et al for *C. flavigena* report a mere 5 cellulases (2 endo-glucanases, 3 cellobiohydrolases) and 14 xylanases, but Anderson et al increase the cellulase count to 10 [118]. *Thermobifida fusca*, another soil actinomycete, has also been sequenced and analysed [130]. Lykidis et al have compiled a comprehensive list of the lignocellulolytic genes found revealing 9 cellulases and 6 xylanases. The higher abundance in *C. fimi* may be down to the criteria "cellulases" and "xylanases" were counted on. Here I have included any potential endo/exo-glucanases and xylanases/xylosidases in my counts regardless of GH family. Christopherson et al restricted their counts in *C. fimi* and the other genomes they compared to include only GH5, 6, 9, 16, 48 and 94 for the cellulases and GH 10, 11 and 30 for the xylanases, which would restrict the numbers reported. What is clear however, is that soil actinomycetes are a rich source of cellulolytic and hemicellulolytic biomass-degrading enzymes, with *C. fimi* being one of the richest.

This is attested to by the GH families with which genes have been annotated, which cover a wide range. In *C. fimi* the predominant GH families for cellulases are GH 6 and 9, and for xylanases GH 10 and 43, but also present are members of GH 5, 11 which are fairly common families and GH 16, 39, 64 and 81, which are less common. GH families 6 and 9 are both composed of inverting enzymes, and while characterised GH 6 members are exclusively either endo- or exoglucanases, members

of GH 9 also incorporate β-glucosidases and exo-β-glucosaminidases [131]. Characterised GH 10 members exclusively contain retaining endoxylanases, while GH 43 members are inverting hydrolases incorporating xylosidase, arabinofuranosidase, arabinananse and β-galactosidase activities. The presence of GH families with multiple functions may confer a broad specificity to the enzymes, giving *C. fimi* adaptability to different substrates. XynB and XynE in *C. fimi* have been shown to be multifunctional, acting on cellulose and xylan [80]. The diversity of GH families is a conserved feature across cellulose degrading soil bacteria [130, 132]. *Cytophaga hutchinsonii*, an aerobic *Bacteroidetes* cellulolytic bacterium also has a range of cellulase and xylanase GH families (5, 8, 9, 10, 11, 43) as do the fungus *T. reesei* (cellulases – GH 5, 6, 7, 12, 61) and the anaerobic *C. thermocellum* (GH 5, 8, 9, 10, 11, 48), as reported in their corresponding genome papers [115, 116, 133].

Binding of enzymes to their substrate is an important feature of cellulases as described in chapter 1. Traditional carbohydrate binding modules (CBMs) are putatively found in 19 of the predicted cellulase and xylanase enzymes including all known and characterised enzymes except XynE. Putative cellulases and xylanases CelF, CelL, XynG and XynH all lack any form of detectable carbohydrate binding module with CelF and XynG being predicted as extracellular, and CelL and XynH as unknown localisation. This tallies with *T. fusca* where almost all its putative cellulases contain a CBM [130]. There are four annotated CBM_2 modules coded for in *C. fimi* with no apparent catalytic domains (CbpA-D), each predicted to be extracellular with CbpD also containing a chitin binding module. Similar catalytic free CBMs have been found in *T. fusca* [130], and *Saccharophagus degredans* [126]. Presumably they work by physically disrupting the cellulosic structure allowing increased enzymatic attack to occur, or by interacting with other extracellular enzymes to increase their concentration on the substrate [85, 86, 134].

Other putative genes in *C. fimi* have less traditional binding domains, the most common being Ricin-B-lectin domains designated CBM13 in CAZy, found in putatively named CelG, CelN, EglB, EglD, XynE as well as putative α-galactosidase

AgaD, pectate lyases PelB, PelC, AfsE, BxyG and a lone Ricin-B-lectin domain with no detected catalytic domain. Lykidis et al report one intracellular enzyme of unknown function with this CBM family whereas here it is found in predicted extracellular and intracellular enzymes. It has been found in a range of enzymes from bacterial species including xylanases, pectate lyases and feruloyl esterases [131, 135]. Ricin-B-lectin domains are known to bind to galactoside residues [136], and so presumably it anchors these enzymes to the side chains of galactose containing polysaccharides.

Several other putative enzymes also have unusual binding domains. CelA, CenC and CelE, all contain immunoglobulin-like domains which have also been found in *C. thermocellum*, and reported in an endoglucanase from *Penicillium decumbens* [137] and *Alicyclobacillus acidocaldarius* [138], conferring sugar binding properties. XynJ and AbgA, a putative xylanase and arabinogalactan-endo-β-galactosidase, contain bacterial immunoglobulin-like (Big_) domains and according to Pfam these domains are found in a range of surface proteins [139]. This is particularly interesting as the *C. fimi* genome encodes 10 putative proteins with LPXTG domains, characteristic of cell wall anchored proteins [140]. Seven of these genes have no further domains other than a signal peptide (celf_0006, 0651, 0718, 0720, 0735, 1024 and 3189). celf_3434 has a GH 53 domain and a Big_4 domain and is predicted as an arabinogalactan (AbgA), celf_3497 has only a lectin domain, and celf_3746 a phytase-like domain only which hydrolyses phytate (storage form of phosphorous in plants) to release inorganic phosphate (Pfam, 135). LPXTG proteins are found in *C. flavigena* (direct genomic search of annotations, not published) and in *C. gilvus*, and according to Christopherson et al are not found in any other *Cellulomonas* genome [129] but have been reported in a β-mannanase of *Bifidobacterium adolescentis* [140].

3.2 Metabolic Pathways of *Cellulomonas fimi*

3.2.1 *Sugar Metabolism*

C. fimi has been shown to be able to grow on a range of substrates, including simple sugars, crystalline cellulose, filter paper, and chitin [79, 114, 120]. Analysis of the genome reveals complete pathways for not only the degradation of cellulose and hemicellulose but also the full metabolism of the resultant sugars. Glucose is broken down by the Embden-Mayerhof pathway to pyruvate before conversion to lactate by lactate dehydrogenase; this feeds into the mixed acid fermentation pathway for which there is also the full complement of enzymes. There was no pyruvate decarboxylase identified, but this is rare in bacterial species [50, 51]. Xylose, mannose, arabinose, fructose, rhamnose and lactose are all fully metabolised and assimilated.

E. coli is a common expression host for heterologous cellulases, as is *Bacillus subtilis,* and both are a likely candidate for a consolidated bioprocessor host [141-143]. By comparison of the pathways of *E. coli* K-12 substr. MG1655, and *B. subtilis subtilis* 168 in metacyc the sugar and polysaccharide degrading capabilities of the three organisms can be compared although the pathway annotations are far from perfect. The *C. fimi* annotation is automatic without manual curation raising the possibility of missed or erroneous pathways. The *E. coli* metabolic annotation was manually curated and has been regularly updated from literature, and *B subtilis* has some manual curation attached [144, 145].

C. fimi is able to degrade a range of polysaccharides including cellulose, xylan, and chitin as previously stated, with metabolism of arabinan and glycogen predicted in metacyc. Conversely *E. coli* is annotated as only able to degrade glycogen. However, the ability of *E. coli* to utilise simple sugars is seemingly greater than the ability of *C. fimi*. *C. fimi* is not annotated as being able to degrade trehalose, $_{D}$-arabinose or $_{L}$-fucose. *B. subtilis* is also annotated as being able to degrade sucrose and maltose pathways which *E. coli* hasn't been annotated in metacyc as containing.

To test the predicted capability of *C.* fimi to grow on simple sugars, *C. fimi* was grown in Dubos Salt minimal medium with different sugar carbon sources. Growth experiments in our lab using minimal salt medium, shows that *C. fimi* fails to grow without the presence of yeast extract (figure 3.1, DSM only and DSM+Cello). By looking at the biosynthetic pathways in the organism it would appear that amino acids and biotin are necessary for growth to occur (3.2.2). However, the yeast extract can act as a carbon source, and growth has been detected in cultures with no other carbon source added (figure 3.1 DSM+YE). By being able to define the amino acids, cofactors or vitamins that can't be synthesised fully or at all, a true minimal medium can be constructed with no additional carbon other than the polysaccharide substrate of interest. This may be important for industrial degradation of plant material should *C. fimi* become an important strain.

Figure 3.1 depicts the growth of *C. fimi* on a suite of sugars including maltose and sucrose. As can be seen growth is clearly visible and on a par with the other sugars predicted to be metabolised by metacyc, if not better. This suggests that there are mis-annotated genes in the *C. fimi* genome so it appears as though pathways are missing when in fact they are not, or that some genes code for proteins with multiple closely related functions. It should be noted that the sugar concentrations for the growth curves in figure 3.1 were in fact ten-fold lower than was usually used. Sugar concentrations of 0.1 mM were used rather than 10 mM, so although it looks like DSM+YE has a significant amount of growth, under normal conditions the growth from YE would be insignificant, with absorbance for cultures with sugar levelling at ~OD_{600} 1.5. However, it is still evident that *C. fimi* is capable of assimilating all the sugars tested, with no contamination evident when cultures were streaked on to agar plates (data not shown) and no repeats were performed for this reason.

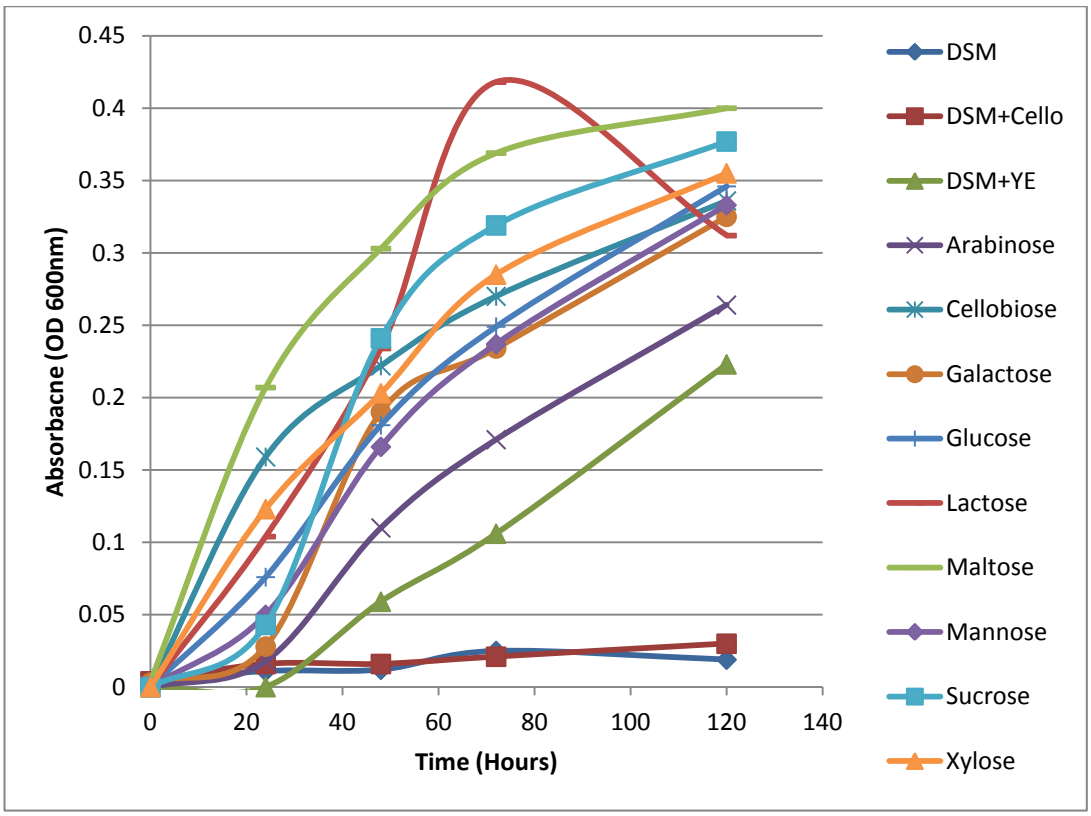


Figure 3.1. Growth of *C. fimi* in Dubos Salt Medium (DSM) with differing carbon sources (0.2 mM). DSM Only and DSM+Cellobiose have no yeast extract (YE), the rest do (0.5 g/L). DSM+YE has no other carbon source. Grown at 37ºC with shaking.

3.2.2 *Amino Acid and Vitamin Synthesis*

As well as being fully capable of metabolising the sugars from hydrolysis of plant cell walls, *C. fimi* is capable of synthesising the majority of amino acids and common vitamins for growth. There are complete pathways for the biosynthesis of all common amino acids except for aspartic acid, appearing to lack aspartate transaminase. For the synthesis of these amino acids, other amino acids are generally used as precursors such as $_{L}$-serine being converted to $_{L}$-cysteine in cysteine biosynthesis pathway I, $_{L}$-valine is converted to $_{L}$-alanine in alanine biosynthesis pathway I, or glutamate being used as the precursor for biosynthesis of arginine, proline and tryptophan (metacyc). This isn't the case for all amino acids and several can be synthesised from products of other metabolic pathways, such as histidine which is biosynthesised from 5-phospho-α-$_{D}$-ribose-1-disphosphate, a product from pentose metabolism, or valine biosynthesis from pyruvate.

C. fimi possesses the pathways necessary to synthesise almost all the major cofactors and vitamins necessary for growth. Biosynthesis pathways were predicted for NAD, Co-A, Co F420, purine and pyrimidine, folic acid, thiamine, lipoate, palmitate, FAD, vitamin-B5, vitamin-B12, vitamin B-6, and vitamin-K. It putatively lacks however the pathway necessary to synthesise biotin which is necessary for the production of fatty acids and gluconeogenesis. There is also a pathway for the production of trehalose which protects against desiccation [146].

3.3 Gene Clustering and Regulation

Stoll et al mapped the positions of the known cellulolytic genes of *C. fimi* on the genome by restriction endonuclease digestion and Southern blotting [147]. They concluded that the known cellulolytic genes are not located in clusters or operons. Looking at the genome and incorporating all the putative genes it becomes clear that this is not the case, and gene clusters are evident and quite frequent. By looking at the surrounding genes of the putative polysaccharide hydrolases, 30 of these genes are flanked with other genes of related function often less than 100 nucleotides from either the start or the end of the gene. Indeed some of the flanking genes actually overlap these putative genes by 4 nucleotides based on predicted start and stop codons. The *xynH-bgaF* cluster in table 3.2 is a prime example of this with the *xynH* start codon beginning one nucleotide before the stop codon of *bgaF*, and *gt36* starting doing the same with the end of *xynH*. This would suggest some of these clusters of genes are expressed at the same time, some as one unit in an operon [148-150]. Table 3.2 depicts the gene clusters involving cellulolytic and hemicellulolytic genes.

Genomic Region	Putative Genes	Gene Orientation
989955 – 997542	*bxyC-bpdt-bpdt-f1esb-afsB-araB*	<=<=<=<=<=>=>
1923428 – 1934555	*bxyF-bxyD-afsD-liftr-f1esb-bpdt-bpdt*	<=<=<=<=<=>=>=>=>
2128016 – 2134275	*cenD-cbhA-liftr*	<=<=<=
3074256 – 3087722	*rok-f1esb-bpdt-bpdt-hyp-agaA-manA-pbpli-aamH*	<=>=>=>=>=><===>=>=>
3533569 – 3538832	*afsE-xynE-bxyG*	<=<=<=
3649422 – 3667427	*liftr-mr-aglA-gt36-xynH-bgaF-f1esb-bpdt-bpdt-bglF-liftr-axeB*	<==>=><=<=<=<=<=<=<=>=>

Table 3.2. Cellulolytic and hemicellulolytic polysaccharide-degrading gene clusters
Transcriptional regulators: *liftr* – LacI family transcriptional regulator; ROK – ROK family protein; *pbpli* – Periplasmic binding protein/LacI transcriptional regulator
Transport: *f1esb* – family 1 extracellular solute binding protein; BPDT – binding-protein-dependent transport systems inner membrane component
Putative genes: *aam** – α-amylase; *afs** – Arabinofuranosidase; *aga** – α-galactosidase; *ara** – Arabinanase; *axe** - Acetyl xylan esterase; *bga** – β-galactosidase; *bgl** – β-glucosidase; *bxy** – β-xylosidase; *xyn** - Xylanase
Other: *gt36* – Glycosyl transferase family 36; *hyp* – Hypothetical protein; *mr* – Mandelate racemase/muconate

As can be seen in table 3.2, 13 of the previously mentioned major enzymes involved in the degradation of lignocellulose form a cluster or potential operon with other associated genes or potential sugar transporters. From an evolutionary stand point this should be expected as the clustering would allow faster production of proteins that, in this case, work together to hydrolyse cellulose. Clustering of cellulolytic genes with other cellulolytic/polysaccharide degrading genes and sugar transporters has also been identified in other cellulolytic genomes including *C. flavigena* (data not shown) and appears to be a common feature [129, 130, 132, 151]. Often these clusters are in the vicinity of a transcriptional regulatory gene, usually of the LacI family. This type of clustering can also be found in the *C. fimi* genome, and are shown in table 3.3 if not already depicted in table 3.2.

Genomic Region	Genes	Gene Orientation
578938 – 588738	*bgaC-f1esb-bpdt-bpdt-liftr-hyp-hyp-abct-tetr*	<=>=>=>=>=>=><=
967746 – 981696	*liftr-f1esb-bpdt-bpdt-bgaA-bgaB-liftr-abct-imt-imt*	=>=>=>=>=>=>=>=>=>=>
1385288 – 1391674	*liftr-aamD-f1esb-bpdt-bpdt*	=>=>=>=>=>=>
2520319 – 2526564	*bpdt-bpdt-f1esb-aamC-rok*	<=<=<=<=<=>
2549244 – 2560545	*anat-aamI-bpdt-bpdt-f1esb-liftr-aamB-aamG*	<=<=<=<=<=<=>=><=
3010411 – 3018856	*aamA-bpdt-bpdt-f1esb-agaB-rok*	<=<=<=<=<=<=>
3026808 – 3033512	*bglG-bpdt-bpdt-f1esb-l1ftr*	<=<=<=<=<=>
3111286 – 3122916	*bgaK-aglB-bgaE-liftr-afu-bpdt-bpdt-f1esb-bglH-nagC-hyp-hyp-tetr*	<=<=<=<=><=<=<=<=<=><=<=<=<=
3629416 – 3637135	*afsC-bpdt-bpdt-f1esb-luxr*	<=<=<=<=<=>
3879031 – 3883795	*bglA-mfs1-l1ftr*	<=<=<=>
4061326 – 4076123	*bgaL-rok-bgaG-bpdt-bpdt-f1esb-abct-rbl*	=><=<=<=<=<=<=>

Table 3.3. Polysaccharide hydrolysing genes clustered with macromolecule transport proteins.
Transcriptional regulators: *liftr* – LacI family transcriptional regulator; *rok* – ROK family protein; *luxr* – LuxR family; *tetr* – Transcriptional regulator TetR
Transport: *f1esb* – family 1 extracellular solute binding protein; *bpdt* – binding-protein-dependent transport systems inner membrane component; *abct* – ABC transporter-like; *imt* – inner-membrane translocator.
Putative genes: *bga** – β-galactosidase; *aam** - α-amylase; *afs** - arabinofuranosidase; *aga** - α-galactosidase; *afu* – α-fucosidase
Other: *anat* – aminoglycoside N(3`)-acetyltransferase; *rbl* – Ricin-B-lectin; *hyp* – Hypothetical protein

Of the clusters observed in table 3.2, five terminate with or incorporate a LacI family transcriptional regulator (*liftr*), one even being annotated as a direct LacI homologue (*pbpli*). From table 3.3 there are 6 gene clusters seemingly incorporating LacI family transcriptional regulators. LacI is a transcriptional repressor, and members of the

same family tend to have the same repressor function [152]. In *Thermobifida fusca* (formerly *Thermonospora*) and *C. flavigena*, CelR, a LacI family transcriptional regulator, has been found with binding sites upstream of cellulolytic genes [132, 153]. Cellulases A-F of *T. fusca* have been sequenced and each had an inverted repeat sequence TGGGAGCGCTCCCA in their 5` upstream regions [153]. This same sequence was found in upstream regions of several *Streptomyces* spp. cellulases and a xylanase from *Thermonospora alba* [153]. Spiridinov and Wilson characterised the CelR transcriptional regulator which binds a DNA motif repressing transcription of cellulase genes, but is itself inhibited by the presence of cellobiose which binds the CelR protein, preventing DNA binding and turning on the previously repressed gene [153]. Anderson et al, in reviewing the cellulolytic systems of aerobic *Actinomycetes*, searched the genomes for CelR binding sites allowing for up to one mismatch and discovered potential sites upstream of putative cellulase genes in all the reviewed organisms bar *Jonesia denitrificans*, including 6 potential sites in *C. flavigena* [132].

It therefore seems probable that *C. fimi* has a CelR transcriptional regulator, or one very similar and potential binding sites upstream of its cellulases. A BLAST search of the genome using the sequence from *T. fusca* CelR protein identified three putative gene products with strong homology: celf_0530 (47% similarity, $1e^{-91}$), celf_1287 (49% similarity, $5e^{-100}$), celf_1704 (49% similarity, $3e^{-92}$). The positions on the genome of these CelR genes largely follow the positional pattern observed in *C. flavigena* and other species, lying near ABC type sugar transporters (which *f1esbp* and *bpdt* are) and glycosyl hydrolases. In *C. fimi celf_0530* is clustered with *bgaC* (table 3.3), downstream of sugar transporters (*celf_1287*<=*bpdt*<=*bpdt*<=*f1esb*), and upstream on the opposite strand of *celE* (*celf_1704*<==>*celE*); in *C. flavigena* DSM 20109 there are two strong matches to CelR, Cfla_2402 – *f1esb*=>*bpdt*=>*bpdt*=>*celR*(2402), and Cfla_3030 – *celR*(3030)<==>*gh9* (putative endoglucanase); in *Cellvibrio gilvus* ATCC 13127 there are again two strong matches to CelR, Celgi_0169 - <=*major facilitator superfamily mfs_1*=>*gcn5-related-N-acetyltransferase*=>*celR*(0169)<=*phosphoglucomutase*, and Celgi_1052 – *celR*(1052)<=*bpdt*<=*bpdt*<=*f1esb*. Using the CelR inverted repeat sequence above

and allowing for 1 mismatch, the *C. fimi* genome was searched using Virtual Footprint [154]. This yielded five hits, one upstream of *cenC*, 2 upstream of hypothetical proteins, one hit within a hypothetical protein and one hit within an isochorismatase hydrolase gene. Either the CelR binding site in *C. fimi* is more degenerate than in other species, simply different or cellulase expression is under the control of a different regulator.

With biosynthesis of cellulases regulated by not only cellobiose but other easily metabolised sugars in *T. fusca*, the same may be true in *C. fimi*, especially with the apparent presence of potential CelR encoding genes [153, 155]. The presence of such regulation would be characterised by motifs upstream of the major cellulolytic genes. I therefore have grouped the 200 bp upstream of genes based on their putative functions. Genes with less than 100 bp between the predicted start codon and the preceding gene have been omitted. There are many programmes available for motif finding. A comprehensive review of different algorithms was performed by Das and Dai [156]. Ensemble algorithms performed best under their analysis. I have used the popular programmes AlignAce, BioProspector, MDScan, MEME and MotifSampler which have been compiled into the user friendly programme Tmod [157]. The defaults for the algorithms were used except for AlignAce (GC background changed from 0.38 to 0.75), BioProspector (only search the forward strand, 25 MonteCarlo simulations), MEME (search forward strand only, return top 5 motifs) and MotifSampler (forward strand only, top 5 hits, repeat 3 times). Motifs found by more than one algorithm or algorithm type, are most likely to be genuine. To assess the strengths of the motifs found, a null set of DNA was also run. This was comprised of 10 randomly selected upstream DNA sequences from those run, combined to form one continuous DNA sequence and randomised 3 times using the sample function in R. The resultant DNA sequence was then divided into 10 separate "sequences". These were run through the algorithms using the same settings as for the actual sequences.

The results from the five algorithms were then compared and motifs that were obviously the same – having the same palindromic sequence found in the same

region of DNA – were combined to form a consensus sequence (figure 3.2). The unique motifs that were identified were compared to known motifs in the Prodoric and RegTransBase databases using TOMTOM of the MEME suite [158]. Using the aligned motif sequences from the algorithm outputs, a position weight matrix was constructed and used to search the *C. fimi* genome for potential hits.

Ideally only genes proven to be co-expressed would be grouped together for motif discovery as it would be most likely that they are under the control of the same regulon. Having no expression data for *C. fimi*, it was assumed that genes putatively of the same or co-acting function are likely to be expressed together, as was often the case in *Thermotoga maritima* [159]. All putative cellulases and glucanases were grouped in a file, and separately all the predicted xylanases, xylosidases, arabinofuranosidases, arabinanases and acetyl xylan esterases were grouped in another. Using Tmod and the previously mentioned algorithm settings, each file was run once through the 5 different programmes.

3.3.2 *Glucanase Motifs*

The glucanase runs only produced one strong looking motif, which was predicted by all the algorithms except MotifSampler. It was consistently found upstream of 9/17 of the glucanase sequences: *cenA*, *cenB*, *cenC*, *celD*, *celA*, *celN*, *celN*, *cbhA* and *eglD*. For AlignAce and MDScan, the higher the score the more significant the motif; for BioProspector and MEME, the lower the p value or E value, the less likely is that the motif has occurred by random based on background frequencies. [ ] denotes that the nucleotides are as easily likely.

AATCGCTTCGG	Align Ace, MAP score 23.4
AAACGATTCG	BioProspector, p value = $2.6e^{-13}$
GAAGCGATTC	BioProspector, p value = $2.4e^{-12}$
GTGGAAGCGC	BioProspector, p value = $1.5e^{-1}$
GAAGCGCTTC	MDScan, Score 3.074
AATCGCTTCG	MDScan, Score 2.992
GAAGCGATTC	MDScan, Score 2.974
CGAATCGCTT	MDScan, Score 2.938
AAGCGATTCG	MDScan, Score 2.863
CGAA[1]CGC[2]TC	MEME, E-value = $3.5e^{-6}$; 1 = A or T, 2 = T or G
CGAAGCGCTTCG	Consensus

Figure 3.2. Combination of identified glucanase motifs to form a consensus sequence

TOMTOM from the MEME suite (http://meme.nbcr.net/meme/intro.html) searches a user motif sequence against available regulatory motif databases. There are two available prokaryotic databases, Prodoric (release 8.9) and RegTransBase (v4) and both were searched against each giving different results. The top hits against Prodoric were for FruR, a LacI family fructose repressor, binding site E = 0.42, and SigI, a RNA polymerase sigma factor regulator E = 0.86. In RegTransBase the top hit was RbsR, a LacI family ribose operon regulator E = $5e^{-3}$, and AlgR, a LacI family regulator of binding protein dependent transport proteins for α-glucosides, E = 0.023. The alignment logos are given in figure 3.3.

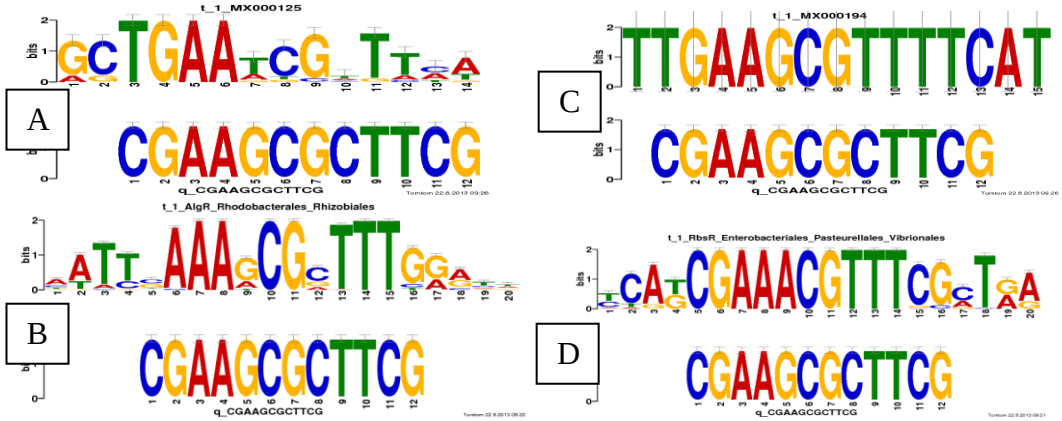


Figure 3.3. TOMTOM output of motif database searches. A) Prodoric 8.9: MX000125 – FruR; C) Prodoric 8.9: MX000194 – SigI; B+D) RegTransBase v4. Top Sequence is that from the database, the bottom is the searched motif (from *C. fimi*). Letter heights are representative of the frequency at which bases arise at each point in the motif (tall letters - no other base found, same height – equal frequency)

Interestingly, by removing the first C and last G from the motif the top hits for the RegTansBase database actually become CelR (E = $9.3e^{-2}$) and FruR (E = $1.1e^{-1}$).

The percentage ratios of each nucleotide in the consensus sequence were calculated and used to make a position weight matrix. Using my own script and this PWM the motif was searched for within the full *C. fimi* genome. An exact match generates the maximum possible score for the motif and an arbitrary cut-off point was used to filter the hits. The cut-off used was 90% of the maximal score, giving rise to only very close matches to the motif, with one or two mismatches. For the glucanase consensus sequence, 54 hits were generated with scores above the cut-off value. Seven of the top 20 hits were found in upstream regions of glycosyl hydrolases – 4 glucanases used to derive the motif, *bxyG* (31 bp downstream of the predicted start codon however), *bglJ*, *bglH* and *bgaE*, and a further four of the hits were upstream of macromolecule/sugar transport genes. The majority of the other hits were actually well within other predicted genes, including 2 hypothetical proteins, a phenylalanyl-tRNA synthetase gene, cytochrome-C oxidase gene and an FG-GAP repeat domain containing protein gene.

3.3.3 *Xylanase Motifs*

The xylanase DNA file searched contained the upstream regions of 23 eligible genes. The files were run and analysed in the same way as the glucanase sequences. Only one likely motif was identified by AlignAce, BioProspector, MDScan and MEME. It is longer than that identified from the glucanase sequences, but bears some similarity around its core.

TTC[1]GCAACGTTGTAGA	AlignAce, MAP 56.691; 1 = T or G
TC[2]CCGACGATGACGA	AlignAce, MAP 40.6; 2 = C or G
TCGACAACGA	BioProspector, p value = $1.7e^{-7}$
CCAAAACGTT	BioProspector, p value = $2.4e^{-7}$
TCGACAACGA	BioProspector, p value = $1.4e^{-5}$
TACAACGATG	BioProspector, p value = $1.5e^{-19}$
AAACGTTGTA	BioProspector, p value = $2.1e^{-19}$
TACAACGTTG	BioProspector, p value = $1.0e^{-18}$
TACAACGTTG	BioProspector, p value = $2.4e^{-15}$
AACGTTGCAC	BioProspector, p value = $8.8e^{-12}$
GAAACGTTTC	MDScan, score 2.942
GAAACGTTTC	MDScan, score 2.942
CGAAACGCTT	MDScan, score 2.821
TTCTACAACGATGTAGAAAAGTGCGA	MEME, E = $1.5e^{-12}$
TTCGACAACGACATCGA	MEME, E = $8.0e^{-8}$
TTCTACAACGTTGTAGAA	Consensus

Figure 3.4. Combination of identified xylanase motifs to form a consensus sequence

This motif was derived from twenty of the sequences occurring at least once in one of the algorithms, but was found consistently upstream of a core of 6 genes found by all 4 algorithms – *bxyG, bglF/bxyA, xfeA, xynD, xynF* and *xynG*. The only Prodoric hit with an E value below 1 in TOMTOM was a match for the MalI binding site, LacI family maltose repressor E = 0.8 (figure 3.5 A). Using RegTransBase (figure 3.5 B+C) the top hits were for the AglR binding site, E = 0.061, CMS2952, a predicted *Streptomyces* spp LacI family regulator of secreted proteins, E = 0.078 and RbsR, a LacI family regulator of ribose transport and metabolism, E = 0.45. Despite the hit for AglR and RbsR, which were also possible matches for the glucanase derived motif, it would seem this could be a separate and distinct motif.

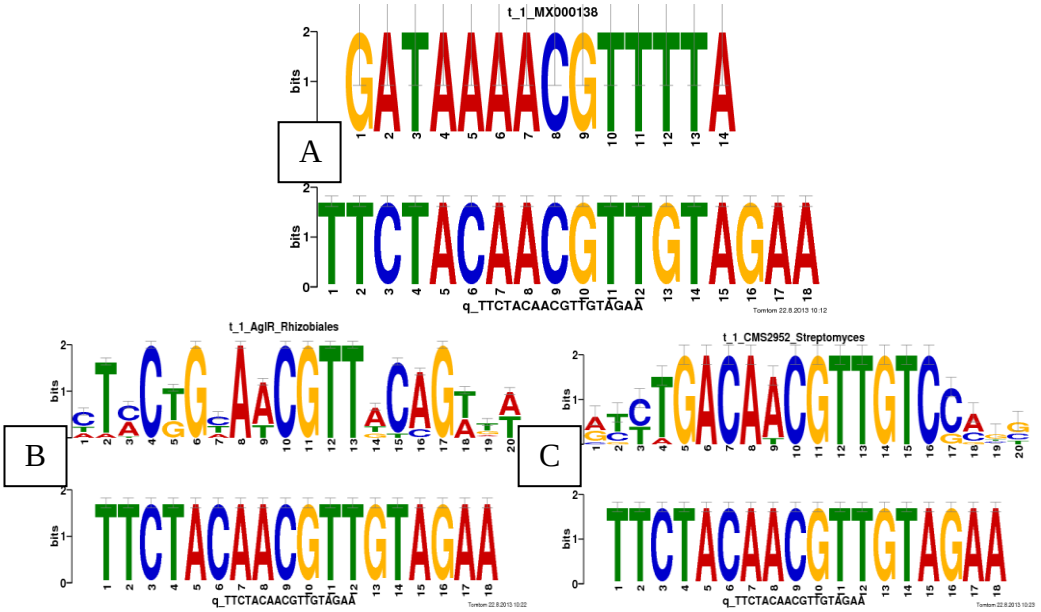


Figure 3.5. TOMTOM alignments of putative xylanase motif to motifs in A) Prodoric 8.9: Mx000136 - MalI, and B+C) RegTransBase. Top and bottom sequences and letter heights as for figure 3.3

As for the glucanase derived motif there appears to be a LacI family transcriptional regulator repressing expression of genes in the absence of simple sugars. When searched against the genome as before for the glucanase motif there were no hits. This means there were no exact matches or matches with 1 or 2 suboptimal nucleotide positions, indicating that the derived consensus is not real, with the peripheral end nucleotides being erroneous and weighted too strongly. At the 80% cut-off, allowing for a more degenerate match to the consensus only 6 hits were found (table 3.4).

Start Position	Strand	Matching Sequence	Nearest Gene, Strand
99364	+	TGCTAAAACGTTGTCGAG	*F1ESB*, +
994751	+	ATCTACAACGTTGCACAA	*F1ESB*, -
994883	+	GTCTACAACGATGTAGAA	*afsB*, +
1614515	+	TCCCACAACGTTGTAGAT	*araA*, +
3717093	+	TTCTACAACGTTGCACGA	*afsA*, +
1038087	-	CTCGAAGACGACGTAGAA	Within NADH-quinone oxidoreducatse

Table 3.4. Putative hits in the *C. fimi* genome of the xylanase derived DNA motif

Shortening the motif by removing the more dubious first 3 and last 3 nucleotides from the sequence (motif – TACAACGTTGTA) and running at a 90% cut-off produced 4 matches, the top being the same as in table 3.4, the second is upstream of a LacI family transcriptional regulator, third *afsB* as above and fourth *araA* as above. At the 80% cut-off with the shorter motif there were 51 possible hits on the genome. The top seven are all described above. In the top 20 there are six which are within a predicted gene, two in the same genic region (just over halfway) of separate 23S rRNA genes, within an integral membrane protein, a BPDT gene, a cyclophilin-type isomerase and a hypothetical protein. One hit was upstream of *celL*, and another two were upstream of sugar transport genes *f1esb* and an ABC transporter. None of these hits share the same regions as those of the glucanase identified motif, indicating that the motifs are distinct from one another.

When the null set of "upstream" DNA was run with these algorithms, motifs were still identified. However, unlike for those identified above, they were of low complexity (GC rich), and were not palindromes as the motifs above were. The top hit for AlignAce was TCCCC[g/c]CCCCGTAAGCG, with a MAP score of 6.5. Potentially it was also "identified" in MDScan (CCCCCCCCCC, E = 1.70). These scores are well below what were given to the glucanase and xylanase derived motifs, increasing the likelihood of the previous motifs being real and over-represented in the DNA sequences given. The second hit on AlignAce (MAP = 4.08) matched to those motifs found by BioProspector (E = 0.14, 0.15 and 0.2) to give the consensus sequence G[g/c]GCTTACGG. When compared to the predicted motifs above, it seems less likely to be genuine, with low complexity, being non-palindromic and having relatively high E values. However, running the motif through TOMTOM as before still produces hits with the top hit in Prodoric to CcpA, a LacI family transcriptional regulator (E = 0.47), and in RegTransBase to GntR (E = 0.02) another member of the LacI family within the E value range given for the glucanase and xylanase motifs.

Discussion

By interrogating the genome I have identified over 100 genes coding for putative enzymes involved in the hydrolysis of cellulose, hemicellulose and the associated sugars including the degradation of pectin (pectate lyases) which are not produced in *T. reesei* the main cellulolytic enzyme mine [115]. Traditional methods of identifying cellulolytic enzymes have failed to reveal the true extent of the arsenal within *C. fimi*, with several methods relying on expression of cellulases when grown on a specific substrate. It becomes clear from reading the literature that cellulolytic cocktails released by organisms can vary hugely dependent on sugar source, with some gene expression only occurring in the presence of simple sugars, under the control of a LacI family regulator. It appears as though the same is true for *C. fimi* as well, with potential repressor binding motifs located upstream of the putative glucanase and xylanases.

These putative enzymes range in their GH families from the function and substrate specific (GH 11 or 81 for example), to the substrate variable (GH 43). This range of predicted GH families and the scale of polysaccharide hydrolysing enzymes appears conserved across species. I feel this is important as it is likely this is what gives the organisms the flexibility to efficiently degrade lignocellulose in all its forms. The production of consolidated bioprocessors has so far met with limited success (see Introduction) and this must in part be due to the relatively small selection of enzymes introduced to the prospective host, and the lack of variety therein.

The grouping of these enzymes often falls into operon-like clusters comprising of sugar transport enzymes and LacI family transcriptional regulators. The release of sugars from biomass is only the first step in bioprocessing. These sugars need to be taken up by the organism and this appears to need specific sugar transporters that are expressed at the same time as the hydrolysing enzymes. Therefore it may well be that inclusion of some of these sugar transporters into the host organism along with the cellulolytic enzymes is needed to confer efficient biomass degradation and utilisation in the host.

The motif CGAAGCGCTTCG was found upstream of the putative glucanase genes, and the motif TACAACGTTGTA was found upstream of the putative xylanase genes. Through homology searching of known regulatory motifs in the databases Prodoric and RegTransBase, they were identified as potentially being members of the LacI family of transcriptional repressors. Members of this family typically mediate their gene expression control with proteins that bind to DNA in the absence of effecter molecules, usually phosphorylated sugars, blocking RNA polymerase from binding to the promoter region, with the operons they control coding for transport and utilisation of the sugar effecter molecule [160]. Members of the family are structurally similar and have been shown to have similar half-site DNA binding sequences with mostly conserved nucleic acids ---AA-C very similar to those predicted in *C. fimi*.

CelR, the cellobiose binding repressor protein, also belongs to the LacI family of transcriptional regulators and has been shown to have effects on gene expression patterns [153, 161]. Deng and Fong created a CelR deletion strain of *T. fusca* and measured growth characteristics of cultures of the mutant versus wild type for glucose and cellobiose as carbon source, and then measured specific gene transcription levels by real-time PCR [161]. They found that growth was severely hampered for glucose as well as cellobiose, exhibiting a very long lag phase of up to 34 hours compared to the wild type of typically 9 hours. At stationary phase for these cultures, cell dry mass was also considerably reduced. When they compared wild type and mutant total cellulase activity, they found the strains were not significantly different however. Taking into account the relative differences in cell mass, they calculated that the mutant strain had a much larger specific activity of 8.84 U/mg (dry weight) than the wild type of 0.44 U/mg (dry weight). They next measured specific gene transcription levels, having identified 9 genes with the reported CelR binding site upstream, 6 having the perfect palindromic sequence. Three of the genes with the perfect biding motif were significantly up regulated in the mutant, 2 were significantly down regulated, and the other to be slightly up regulated but to a much lower, almost insignificant degree. The gene for CelR itself was found in an operon controlled with a single promoter with a perfect CelR

binding sequence upstream, containing genes for an ABC transporter, 2 genes whose function are not mentioned, β-glucosidase then CelR, and this operon was significantly expressed lower in the mutant than wild type. CelR may therefore act as a repressor and an activator, or simply a repressor, and a transcription factor which is repressed usually by CelR is expressed in the mutant to initiate transcription of the observed up regulated genes.

Other experiments investigating the differential expression of genes on a range of substrates has given insight into the control and necessities of natural cellulolytic enzyme systems. There have not been many studies on the differential expression of cellulolytic genes when the organism is grown on different sugar substrates, and of those that have been performed the majority focus on a very small subset of the organism's genes. As far as I am aware there is only one paper that has specifically looked at differing expression in *C. fimi* on different substrates and they focussed on glucanase and xylanase activity. Khanna and Gauri grew *C. fimi* in the presence of CMC, xylan, cellobiose, galactose, glucose, fructose, maltose and xylose [162]. They demonstrated that the specific activity of the 2 groups of enzymes differed depending on the carbon source *C. fimi* was cultured with. Both xylanase and glucanase levels were the highest when grown with CMC or xylan, with glucanases showing highest activity for CMC and xylanase activity highest for xylan. The mono/disaccharides all had a huge repressive effect on both enzyme types, with activity not detected for xylanases on galactose, and not detected for glucanases on xylose, maltose, fructose or cellobiose. They also observed that as CMC was hydrolysed the specific activity levels of both the glucanases and xylanases continued to climb. This would suggest that the released sugars from CMC hydrolysis was a positive inducer of these activities, however as mono and disaccharides had a repressive or inhibitory effect, then the inducers must be longer oligosaccharides. They also note that the presence of xylanase activity in the absence of xylan suggests constitutive expression of (some) xylanases.

In *Cellulomonas flavigena*, Perez-Avalos et al looked only at the specific activity levels of xylanases and xylosidases after growth on avicel, solka floc (highly purified

insoluble cellulose), pre-treated sugar-cane bagasse, xylan, and simper sugars of cellobiose, glycerol and xylose [163]. Maximal activities for both enzymes were detected with sugar-cane bagasse followed by xylan, but low levels of activity were still seen in the presence of avicel and solka floc and even glycerol and xylose. The lowest levels of activity (nearly 50 times lower than the specific activity observed for sugar-cane or xylan) were observed for the mono/disaccharide cultures with cellobiose giving almost no detectable activity. This suggests that the assayed enzymes are induced by both cellulose and xylan, but repressed or inhibited in the presence of simpler mono/disaccharides, similar to what was seen for *C. fimi*, including the apparent constitutive activity.

A more recent study by Rajoka et al, with a larger set of cellulosic, mono/di and oligosaccharides largely found the same pattern as Perez-Avalos et al [164]. Monomeric saccharides all had a negative effect on xylanase and xylosidase activities when compared to xylose or more complex substrates (CMC, kallar grass straw, xylan) with xylan followed by CMC being the best polysaccharides for activity induction, and xylose followed by cellobiose being the best mono/disaccharide inducers. Glucose has the greatest effect on repression, but activity was never suppressed outright. Sanchez-Herrera et al studied the effect of sugar-cane bagasse, solka floc, xylan and glucose again focussing on glucanase and xylanase/xylosidase activities using 2D gel electrophoresis as well as activity assays [128]. As for Perez-Avalos et al bagasse was the greatest inducer of these enzymes, followed by solka floc, then xylan (where the ability to degrade filter paper was not detected at all suggesting suppression of enzymes able to hydrolyse crystalline cellulose). Glucose actively repressed all activity assayed for. By using 2D gel electrophoresis they were able to more accurately detect what was actually happening in the enzyme profile, finding 6 enzymes expressed under all 4 conditions, various levels of overlap between subsets of the carbon sources (eg 1 enzyme was common between glucose and xylan conditions, and 28 enzymes common between bagasse and solka floc conditions). They also detected 11 enzymes specific to induction with bagasse, 10 for xylan, 27 for glucose and 0 for solka floc.

Rajoka and Malik looked at the production of filter paper cellulases (FPases), endo-β-glucanases, and β-glucosidases by *Cellulomonas biazotea* using various pre-treated cellulosic ground plant (grasses) materials, agricultural wastes (cotton stalks, wheat stalks), CMC, α-cellulose, cellobiose and xylan [165]. For this organism FPase activity was most strongly induced by kallar grass and α-cellulose with cellobiose giving less than 20% of the maximal. CMCase activity was largely the same for all substrates. β-glucosidase activity was also fairly steady for all substrates tested, but the highest levels were detected for cellobiose, and the lowest for xylan at 20 times less activity.

The most comprehensive studies have been carried out in bacteria for *Clostridium thermocellum* and *Thermotoga maritima*. The cellulosomal composition of *C. thermocellum* was investigated by Raman et al when grown on amorphous cellulose (Z-Trim) and industrial forms of cellulose (pre-treated switchgrass), cellobiose and combinations of cellulose mixed with xylan and pectin, compared to cultures grown on crystalline cellulose (avicel) using mass spectroscopy [117]. By doing so they identified over 16 novel cellulosomal components mostly of unknown function but included putative carbohydrate esterases (avicel+xylan; avicel+pectin; avicel+xylan+pectin; cellobiose), glycosidases (avicel+xylan; avicel+xylan+pectin; cellobiose; Z-Trim) and pectinases (avicel+pectin; avicel+xylan+pectin). Many enzymes were identified that were common to all conditions, comprising a range of GH families (5, 8, 9 and 48) mostly endoglucanases but also a GH11 xylanase. This is perhaps not surprising as each condition contained cellulose, except for the cellobiose cultures – the major product of cellulose hydrolysis that has been reported many times as a strong inducer of cellulases and xylanases. Of note is the observation that GH5 and GH9 family endoglucanases were differentially expressed, with GH9 actively down regulated in the absence of crystalline cellulose, and GH5 down regulated in the absence of any form of cellulose, and both decreased in the presence of cellobiose. They suggest therefore that GH9 endoglucanases are specifically useful for the hydrolysis of crystalline cellulose, and if true, is not the same as observed for *C. fimi* where the GH9 enzymes CenB and CenC both favour soluble to crystalline cellulose [70, 111].

Another large scale study of *C. thermocellum* was performed by Wilson et al who looked at the global transcriptome when grown on pre-treated *Populus* and switchgrass [166]. The paper was primarily concerned with the comparison of transcriptome sequencing platforms and methods of normalisation, but some salient information pertaining to cellulolytic properties was also reported. For these substrates they discovered glycoside hydrolases (*Cthe*_2089 and 0271) were the most abundant transcripts but also expression of a sugar transport system in an apparent operon (0391-0393) which was specific for cellotriose. Over all there were >200 genes found to be differentially expressed by at least 2 fold (up or down) compared to the housekeeping gene for both substrates. There were differences to be seen between them though with genes 1256+1257 (glycoside hydrolase and cellulose binding protein, respectively) showing higher expression on *Populus* to switchgrass. There were more similarities observed than differences between the 2 substrates, which they postulate is due to the relative homogeneity of the substrates after pre-treatment.

Chhabra et al studied the expression profiles of 269 known genes of *T. maritima* by microarray when grown on galactomannan, barley glucan, laminarin, starch, CMC, glucomannan, β-xylan, glucose, mannose and xylose [159]. These genes included those of glycoside utilisation or modification (65), proteolysis (40), stress response, proteolytic fermentation, sugar transport (21), and transcriptional regulation (69). Under these growth conditions RNA was harvested at early to mid log phase and compared to the expression profiles of each other carbon source in a closed loop design. What was observed was a complex mix of expression profiles but many of the genes could be effectively clustered based on similar expression patterns. Glucose was generally found to repress all glycoside hydrolase genes measured, but genes were up or down regulated in a sugar dependent manner for the other carbon sources assayed.

Cluster 4.1 as they designated it comprises genes strongly up regulated when grown on CMC and include 3 endoglucanases (TM0305, 1524 and 1525), cellobiose phosphorylase (TM1848) and a host of ABC transporter related genes preceded by a

LacI family transcriptional regulator in an apparent operon (TM0303, 0304, 1218-1223). This cluster was also up regulated when grown on barley glucan and glucomannan and generally down regulated on all other substrates. Another cluster (4.4) was strongly up regulated in the presence of galactomannans, but also generally up regulated for CMC, glucomannans and mannose as well. These genes included an α-galactosidase and two β-galactosidases (TM1192, 1193, 1195) whose expression profiles are matched closely by a LacI family transcriptional regulator so may be under its control (TM1200), 2 β-mannosidases (TM1227, 1624) and 2 endoglucomannases (1751, 1752) that appear to be within an operon or gene cluster of very familiar structure similar to those observed in *C. fimi* (1747-1752) incorporating 2 ABC permease proteins and 2 ATP binding proteins but no regulatory gene was clustered close by suggesting a potentially distal regulator.

Clusters 4.5 and 4.6 relate to genes up regulated by xylan with 4.5 also being up regulated by xylose, and 4.6 by glucomannan. A more complicated mix of sugar hydrolysing enzymes is observed when compared to those up regulated on CMC. Cluster 4.5 includes an α-glucuronidase, β-galactosidase, 2 endoxylanases, xylosidase, acetyl xylan esterase, and ATP binding ABC transporter, LacI family transcriptional regulator, crp family transcriptional regulator and a mix of others. Cluster 4.6 is similarly varied including several clusters of potential operons (TM0064-0069) with uronate isomerase, mannoate hydrolase, mannoate oxidoreductase, IcIR family transcriptional regulator and (TM0430-0433) with 3 sugar ABC transporters and a pectate lyase. This greater diversity of genes for xylan opposed to cellulose reflects the complexity of xylan and the range of sugars that may be found in hemicellulose. By looking at the composition of the clusters formed the authors were able to construct 3 separate putative pathways for the extracellular hydrolysis of complex sugars followed by the internalisation of shorter oligosaccharides by specialised transporter proteins and the internal hydrolysis of these oligosaccharides to their constitutive monomer sugars.

These studies cover a small range of cellulolytic bacteria found in differing environments. What is common between them is the apparent control of the genes

by specific sugar inducers and repression in the presence some monomers, primarily glucose. This would suggest that the controlled expression of these enzymes only when in the presence of their main substrate is evolutionarily beneficial, and priority is given to the utilisation of simple high energy sources (eg glucose) over more complex energy sources. From looking at the latter 2 papers mentioned above it is also clear that a very wide range of enzymes is necessary for the efficient and complete hydrolysis of cellulose and hemicellulose. A cell to produce all the necessary enzymes constitutively would have far too great a metabolic burden, hampering its ability to grow and survive. So some enzymes are constitutively expressed at low levels, and when the correct substrate is encountered and no other sugar is available, the released oligo/di/monosaccharides induce expression of a larger host of related enzymes, including sugar specific transporters to increase the internalisation rate of the extracellularly hydrolysed substrate.

In the case of *C. fimi* it therefore seems highly likely that the same sort of systems are also in place. The potential operons of ABC transporters, transcriptional regulators and polysaccharide hydrolysing enzymes resemble motifs identified in several other organisms and in the expression analysis of *T. maritima*. This postulation is also given some credence by the identification of potential transcription factor DNA binding sites upstream of the hydrolysing genes, corresponding to repressors that are alleviated in the presence of sugars (*a la* CelR). Indeed, if the vast majority of the putative polysaccharide hydrolysing genes indentified in the genome are repressed under certain conditions then it is little wonder that the studies performed so far that have focussed on growing cells on cellulose or xylan have only found a small sub set of the potential host. And I feel that any artificial cellulolytic network must in some way incorporate this complexity, by including sugar specific transporters at least. Putting a small number of inducible cellulases or xylanases into a host has not had great success in the production of a consolidated bioprocessor, as discussed in the introduction chapter.

An alternative to expressing cellulolytic systems in a host would be to introduce the enzymes necessary for production of the desired product (usually ethanol) into the

cellulolytic organism instead. *C. fimi* lends itself as a good potential host for this, with the raft of polysaccharide degrading enzymes already available, the ability to utilise almost every sugar released from the biomass, the sugar transport system necessary in place, and the presence of the mixed acid pathway which can be modified for ethanol fermentation [50, 51]. With most amino acid and vitamin biosynthesis pathways present, comparable to those found in *E. coli* and *B. subtilis*, common bioprocessor hosts, *C. fimi* should be as cheap to grow industrially as these organisms at least from a medium perspective.

More work needs to be done to understand how these cellulolytic systems work synergistically on the different substrates they are naturally faced with. We also need to understand how they are controlled so as to be able to replicate this in a synthetic network within a consolidated bioprocessor. With an ever increasing number of genomes available I feel it is important to properly characterise one diverse system, and *C. fimi* could be that model.

CHAPTER 4: CLONING, EXPRESSION AND SCREENING OF PUTATIVE *C. fimi* GENES

Through the analysis of the *Cellulomonas fimi* genome a range of polysaccharide degrading genes have been putatively identified based on sequence homology. A selection of these genes have been cloned as BioBricks, expressed in *Escherichia coli* and *Citrobacter freundii*, and screened for common cellulolytic functionality. BioBricks are the standard synthetic biology unit (see Introduction). I am using the plasmid pSB1A2, an enteric bacteria vector that confers ampicillin resistance, containing the low copy number origin of replication ColE1. To this the BioBrick BBa_J33207 has been inserted which consists of the P_{lac} promoter and the first 76 amino acids of the N-terminus of LacZ, reinstating β-galactosidase activity to *lacZ*ΔM15 bacteria (http://parts.igem.org/Part:BBa_J33207). Downstream of this a synthetic ribosome binding site (RBS), BBa_B0034 which is generally a strong RBS, has been ligated. A gene encoding sequence cloned as a BioBrick can be easily inserted downstream of the promoter/RBS BioBrick and used as an expression cassette for the analysis of gene function as described in Materials and Methods. By adding IPTG to the growth medium expression can be induced by P_{lac} producing functional β-galactosidase and the inserted downstream gene of interest. The *lac* promoter (P_{lac}) and RBS have been extensively used in our laboratory and are known to work well in *E. coli* and are the same for all screening constructs. Using this method we therefore know that any lack of activity or failure of expression is due to reasons other than a poorly performing promoter. This also has allowed for rapid screening of activities in our expression hosts of choice, the *Enterobacteria E. coli* JM109 and *C. freundii* NCIMB 11490. The putatively identified glucanase genes of *C. fimi* have been cloned as BioBricks and screened for common activities. However, as will become evident, *in silico* predicted function and what is actually detected *in vivo* are not necessarily guaranteed to marry. In this chapter I describe the cloning, expression, functional assaying and expression analysis of a range of genes from *C. fimi* on the substrates PNPA, ONPC, ONPX, OPNGal, RBB-CMC and RBB-Xylan.

These synthetic substrates were used due to the specificity conferred by their specific sugars and linkages to the nitrophenol residue, the cleavage of which produces an easily detectable and quantifiable colour shift from clear supernatant to yellow, and requires no specialist equipment to measure, simply a spectrophotometer. Endo-acting enzymes require long chains of sugars as substrate. The RBB dyed CMC and xylan release easily detectable RBB into the reaction supernatant if cleavage of these long chains has occurred, with background sugars simply removed by precipitation as described in Materials and Methods. These assays have been used for decades and are in keeping with contemporaneous literature.

4.1 The Cloned Genes

The genes listed in table 4.1 were all cloned from *Cellulomonas fimi* using the BioBrick method outlined in Materials and Methods. The BioBrick was used as a means to screen the activity of the genes when expressed in *Escherichia coli* under the control of an inducible promoter. The plasmid pSB1A2 has been used in our lab extensively and the P_{lac} inducible promoter has been characterised and works well in our host. Gene names were derived from our own sequencing information based on predicted function through BLAST homology to other proteins in the NCBI database and will be elucidated on later in the chapter. Below are the protein products as annotated by the US DOE when uploaded to NCBI (reference number NC_015514.1) and which tended to match our own annotations.

Some of the genes were cloned before the release of the *C. fimi* genome into the NCBI database so these gene clones are based on our own sequencing data (unpublished) and our predicted start sites. Therefore not all of the predicted gene open reading frames match those in the NCBI database, largely due to incomplete sequences. Where this is the case it has been noted and the difference between ORF start as I predicted it and where the DOE entry predicted it is stated.

Gene Name	DOE Annotation	Genome Position and Gene Length	Difference from DOE Entry	Predicted Function by BLAST
afsA	α-$_L$-arabinofuranosidase domain containing protein	3717150-3718670, 1521 bp		arabinofuranosidase
afsB	α-$_L$-arabinofuranosidase domain containing protein	994923-996431, 1509 bp		arabinofuranosidase
afsC	α-$_L$-arabinofuranosidase domain containing protein	3629416-3630957, 1542 bp		arabinofuranosidase
bxyC	Xylan 1,4-β xylosidase	4081997-4083481, 1485 bp		xylosidase
bxyD	Glycoside Hydrolase Family 43	1926115-1927749, 1635 bp		xylosidase
bxyE	Glycoside Hydrolase Family 43	989955-991277, 1323 bp		xylosidase
bxyF	Glycoside Hydrolase Family 39	1923428-1926118, 2691		xylosidase
celA	Cellulase	418857-420560, 1704 bp		endoglucanase
celD	Glycoside Hydrolase Family 9	52023-54761, 2739 bp		endoglucanase
celE	Glycoside Hydrolase Family 9	1874747-1878067, 3321 bp		endoglucanase
celF	Glycoside Hydrolase Family 9	1612769-1614379, 1611		endoglucanase
celN	Ricin-B-lectin	2665432-2666949, 1518 bp	DOE start at 2665522, 1428 bp	cellulase
xynF	Endo-1,4-β-xylanase	103525-104919, 1395		xylanase
xynG	Glycoside Hydrolase Family 10	1900401-1901690, 1290 bp		xylanase

xynH	Xylan 1,4-β-xylosidase	3655116-3656792, 1677 bp		xylosidase
xynI	Pseudo Gene	4059449-4060426, 977 bp	DOE end at 4061268, 1820 bp	xylanase (catalytic domain)

Table 4.1. Putative *Cellulomonas fimi* genes cloned as BioBricks. *xynI* was deemed a pseudo gene as in the complete sequence defined by US DOE the ORF appeared to extend beyond the stop codon encountered at 4060426, indicating a truncated gene.

The above genes were all cloned as described in the materials and methods section using extracted genomic DNA as PCR template. This is because *C. fimi* colony PCRs were not a reliable method of cloning, often failing to produce a product. When using KOD Hot Start Polymerase, the high GC content of the genome and genes necessitated the use of 5% (v/v) DMSO in the PCR reaction mixture to weaken the hydrogen bonds between DNA strands. When using KOD Xtreme the addition of DMSO was not necessary as the GC PCR buffer is optimised for high GC template already. The other deviations from the standard cycling conditions were the initial denaturation and polymerase activation incubation which was extended from 2 minutes to 5 minutes, and the denaturation step in each cycle was extended from 20 seconds to 1 minute. Without these alterations the PCR reactions would often fail or yield low quantities of DNA. However, these alterations also reduced the specificity of the primers and contaminating products were sometimes observed. In these cases it was necessary to purify the PCR products from agarose gel and then proceed to the next stage of cloning. The gel purifications were performed using Qiagen kits as described in their manual, and purifications from solution were performed using silica beads as described in the methods section.

The PCR products and Edinbrick1 plasmid with ampicillin resistance marker were digested with *Eco*RI-HF and *Spe*I. The GC levels of the genes led to a high prevalence of *Pst*I sites in the genes meaning this enzyme could not be used. The digestion removed BioBrick BBa_J33207 from Edinbrick1 and replaced it with the PCR product, leading to white colonies when *E. coli* JM109 was transformed and plated with ampicillin, IPTG and X-gal. These white colonies were selected, patched, and the plasmids extracted by DNA miniprep. The purified plasmids were linearised and those of the correct molecular weight had their gene insert starts and

ends sequenced using the primers described in table 2.1. A plasmid (pSB1C3) with chloramphenicol resistance marker containing BBa_J33207+RBS was used to add the promoter and RBS upstream of the *C. fimi* gene insert. The BBa_J33207+RBS plasmid was digested with *Eco*RI-HF and *Spe*I, and the gene insert plasmid digested with *Eco*RI-HF and *Xba*I. The digestion products were ligated and plated on ampicillin plates with IPTG and X-gal. Correctly ligated plasmids resulted in blue ampicillin resistant colonies when used to transform *E. coli* JM109. These were sequenced to confirm the presence of promoter, RBS and gene insert using the primer LacZ_F2 (table 2.1). The *E. coli* colonies containing the correctly constructed and sequencing confirmed plasmids were kept as glycerol stocks at -80ºC, and the plasmids stored at -20ºC.

4.2 Screening of Constructs for Activity

For the screening of enzyme activity 15 ml of LB was inoculated with a loop of cells from the glycerol stocks. The LB was supplemented with ampicillin and IPTG from the start and the inoculate cultures were grown at 37ºC with shaking overnight. These cultures were transferred to 50 ml centrifuge tubes and centrifuged to pellet the cells at room temperature for 5 minutes. The supernatant was discarded and the pellet suspended in 1 ml 50% (v/v) PBS and 25% (v/v) glycerol and transferred to 1.5 ml centrifuge tubes and kept on ice. The cells were lysed by sonication as described. The lysates were centrifuged at 13 krpm at 4ºC for 10 minutes to pellet whole cells and debris. The supernatant was kept on ice and stored at -20ºC, and the pellet discarded. This lysate was used for the subsequent assays.

Lysates were screened for endoglucanase (EC 3.2.1.6, Congo red or RBB-CMC), endoxylanase (EC 3.2.1.8, RBB-xylan), β-glucosidase (EC 3.2.1.21, MUG), cellobiohydrolase (EC 3.2.1.91, ONPC), β-galactosidase (EC 3.2.1.23, ONPGal), α-$_L$-arabinofuranosidase (EC 3.2.1.55, PNPA), and β-xylosidase (EC 3.2.1.37, ONPX) activities. The assays were performed as described in the materials and method section using 5 µl of lysate, 5 µl of substrate and 15 µl of sodium acetate buffer, pH 5.5 when performing nitrophenol assays, and a 50 µl assay volume

containing 15 µl lysate, 0.2% (w/v) RBB substrate and sodium acetate buffer, pH 5.5 for the endo- assays. Lysate protein concentration was estimated by Bradford assay. Negative controls for all assays were performed using either *E. coli* or *C. freundii* lysates expressing Edinbrick1 plasmid with no insert, grown, harvested and lysed under the same conditions. Where possible a positive control was used from lysates of *E. coli* expressing *cenA* for endoglucanase activity, *xynB* for endoxylanase, ONPC and ONPX activity and *bglX* for β-glucosidase activity.

4.3 Activity Assays of Expressed Cloned Genes

4.3.1 *Nitrophenol Assays*

These assays detect the release of a sugar derivative that is bound to nitrophenol. If the enzyme of interest is able to cleave the sugar-nitrophenol bond then the nitrophenol is released and when protonated under high pH values it produces a yellow colouration that is detectable at 405 nm. The release of nitrophenol is quantifiable and absorption is directly proportional to the molar concentration present. From standard curves for 2-nitrophenol (ONP) and 4-nitrophenol (PNP) at pH 5.5, the molar extinction coefficients were calculated to be 1192.6 M^{-1} cm^{-1} and 9594.4 M^{-1} cm^{-1}, respectively. Using these values and the Beer-Lambert law:

absorbance = molar coefficient (M^{-1} cm^{-1}) x concentration (M) x pathway length (cm)

the µmol concentration of nitrophenol released per minute was calculated, adjusted for the dilution factor in the assay, and used to define the units (U) of active enzyme present. Units are defined as the amount of enzyme required to release 1 µmol of nitrophenol per minute per ml. The units were then divided by the estimated amount of protein present per ml in mg to give U/mg. All nitrophenol screening data in figures 4.1-4.4 are presented as U/mg to give the clearest indication of activity over background.

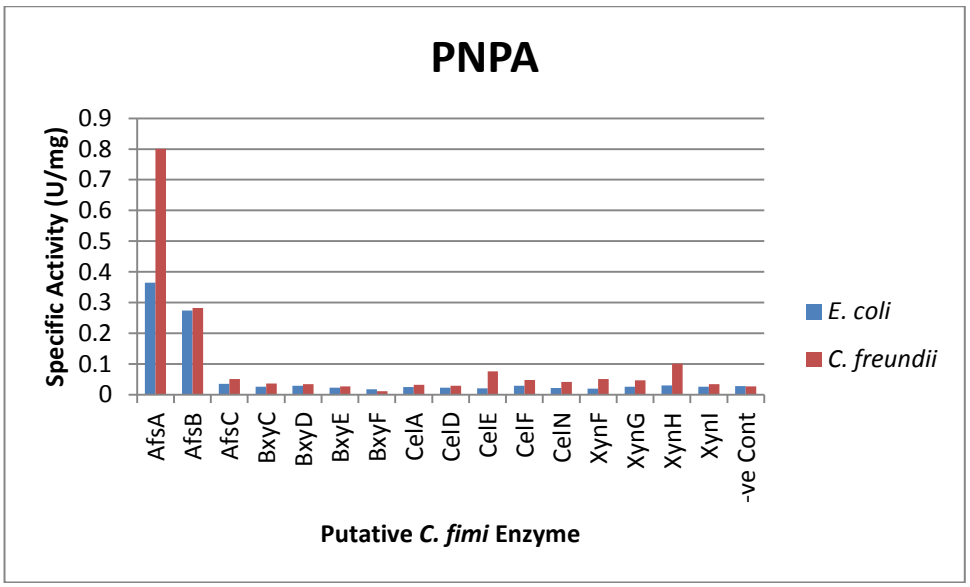


Figure 4.1. Cell lysate screen for arabinofuranosidase activity of *C. fimi* genes using PNPA. Absorbance measured at 405 nm, 1 hour incubation at 37ºC, *E. coli* host is blue and *C. freundii* host is red, -ve Cont is *E. coli* lysate with expressed LacZ

The figures 4.1-4.4 show the relative activities of the gene inserts with IPTG induced expression in both *E. coli* (blue) and *C. freundii* (red) hosts. For PNPA as substrate (figures 4.1) there was no available positive control, however, this in the end was not necessary. Only AfsA and B showed any activity in either host for this substrate, with AfsA being arguably the more active of the genes. Interestingly, despite being sequentially very alike to AfsA and AfsB, AfsC shows roughly background levels of specific activity. None of the other constructs shows any convincing level of activity on this substrate. Of note also is that the specific activity of AfsA is roughly twice as high when expressed in *C. freundii* than in *E. coli*, whereas AfsB is more or less the same.

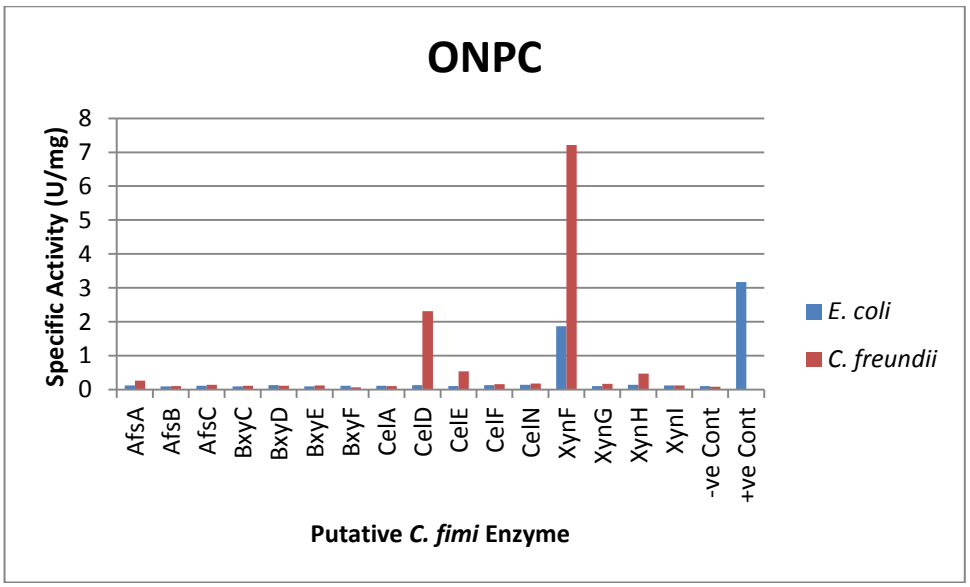


Figure 4.2. Cell lysate screen for cellobiohydrolase activity of *C. fimi* genes using ONPC. Absorbance measured at 405 nm, 1 hour incubation at 37ºC, *E. coli* host is blue, *C. freundii* host is red, -ve Cont and +ve Cont are *E. coli* lysates with expressed LacZ or XynB, respectively

Much like for PNPA, activities towards ONPC were restricted to a small subset of the genes cloned. When expressed in *E. coli* (figure 4.2) the only product with activity is that of XynF with a definite peak well above background akin to that of XynB. The XynF U/mg in this initial screen is over half that recorded for XynB, the most studied of xylanases so far identified in *C. fimi*. In *C. freundii*, activity is evident in CelD and XynF lysates, with the U/mg of CelD at a level similar to that of XynF expressed in *E. coli*. As was seen with AfsA above, in *C. freundii*, the U/mg value of XynF is 4 fold higher than in *E. coli*.

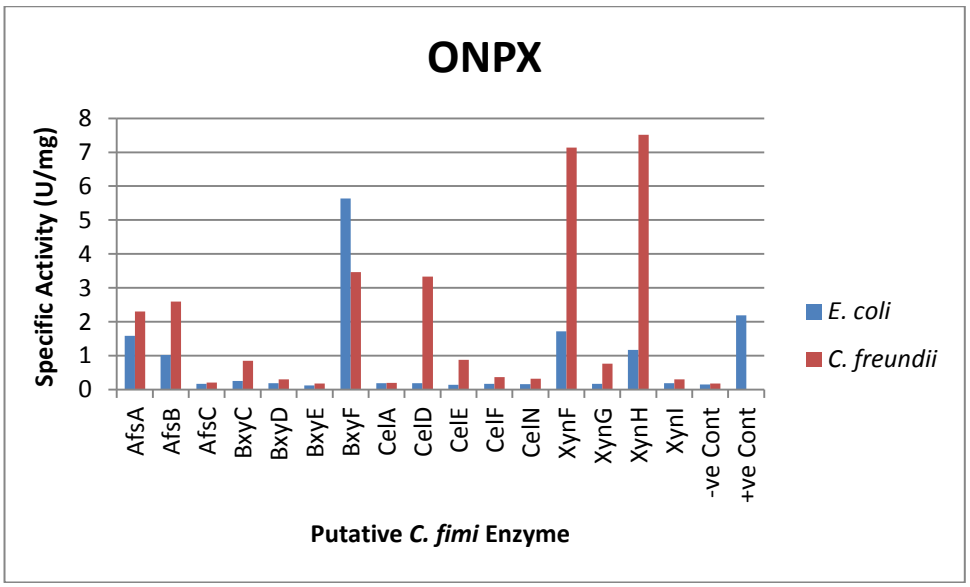


Figure 4.3. Cell lysate screen for β-xylosidase activity of *C. fimi* genes using ONPX. Absorbance measured at 405 nm, 1 hour incubation at 37ºC, *E. coli* host is blue, *C. freundii* host is red, -ve Cont and +ve Cont are *E. coli* lysates with expressed LacZ or XynB, respectively

The hydrolysis of β-xyloside from nitrophenol is the most common function observed for these putative enzymes. When expressed in *E. coli* (figure 4.3), five of the sixteen putative enzymes show activity above background levels. The most active of all is BxyF which shows specific activity of nearly 3 times that of XynB. AfsA, AfsB and XynF seemingly demonstrate multifunctional behaviour with activity clear from the assay, with AfsA and XynF showing specific activity levels near that for XynB, and AfsB roughly half. XynH also displays clear activity on a level with that of AfsB. For the *C. freundii* expression data the results look very different. The same enzymes are still active, but the observed activity with ONPX is altered. XynH lysate is clearly the most active with XynF close behind with nearly 4 times the specific activity seen for XynB expressed in *E. coli*. BxyF is still highly active but nowhere near as much as XynF or XynH apparently are. Along with the arabinofuranosidases, CelD now shows activity against ONPX, and like for ONPC, this is not evident in *E. coli*. By comparing the activities of the constructs in *E. coli* and *C. freundii*, yet again we can observe that there is an increase in specific activity between the former and the latter hosts. Nearly a six fold increase for XynH, 4 fold increase for XynF, 1.5 and 2.5 fold increase for AfsA and AfsB, respectively, but a 0.6 fold decrease for BxyF.

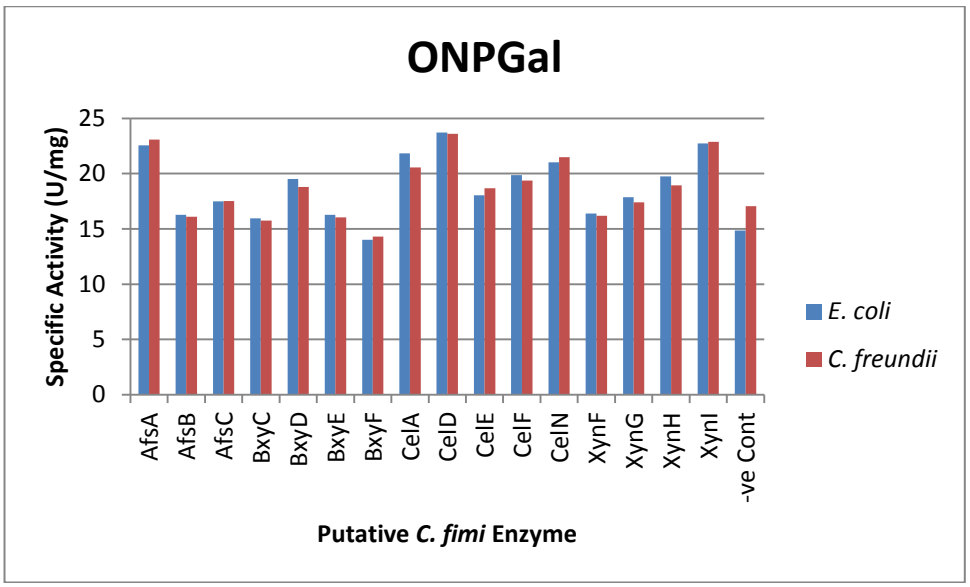


Figure 4.4. Cell lysate screen for β-galactosidase activity of *C. fimi* genes using ONPGal. Absorbance measured at 405 nm, 10 minute incubation at 37ºC, *E. coli* host is blue, *C. freundii* host is red, -ve Cont are *E. coli* lysates with expressed LacZ only

The activity of enzymes against ONPGal is problematic as both *E. coli* JM109 when harbouring a plasmid containing the *lacZα* gene fragment and *C. freundii* express native β-galactosidases and will be induced with the addition of IPTG. This makes the screen for this potential activity in unpurified lysates difficult due to the high background activity. Enzymes were screened for 10 minutes with ONPGal, rather than 1 hour as seen for the other substrates, as long periods of incubation seemed to effectively hydrolyse nearly all the ONPGal present, masking any potential differences in activity. All constructs display a high level of activity as expected (figure 4.4). All of the constructs bar BxyF appear to have higher activity than the Ed1 controls with CelD, AfsA and XynI apparently having the highest activities of all for both *E. coli* and *C. freundii* hosts. Unlike the other substrate screens, the activity levels for the constructs are nearly identical in the two bacterial strains. It is difficult to draw any conclusions from this data, it being extremely unlikely that all the constructs have β-galactosidase activity, however it may be that one of the top 3 enzymes mentioned above has such activity, CelD in particular.

4.3.2 *Endoglucanase and Endoxylanase Assays*

The activity of endo- acting enzymes can be assayed in a number of ways: plate assays using Congo-Red stain to dye long chain polysaccharides with zones of clearing denoting endoglucanase hydrolysis which produces shorter chains which the dye cannot bind to [167]; measuring the increase of sugar reducing-ends by DNS or Anthrone reagents; release of dye bound to the substrate. Each has their advantages and drawbacks. The Congo red assay gives a very clear, easily interpretable positive or negative result but it doesn't stain xylan only soluble cellulose (e.g. CMC). The DNS or Anthrone assays are quantifiable, but quite insensitive to low levels of activity and don't work well with pentose sugars (xylose, arabinose) as substrate, in my experience. For the screening I therefore decided to use polysaccharides dyed with remazol brilliant blue (RBB), which is released when subjected to endo-hydrolysis, and can be measured spectrophotometrically at 590 nm. RBB-dyed CMC and birch wood xylan were used as substrates, the polysaccharides precipitated using three volumes of ethanol and the supernatant absorbance measured. The figures 4.5 and 4.6 depict the absorbance at 590 nm to give a qualitative assessment of activity.

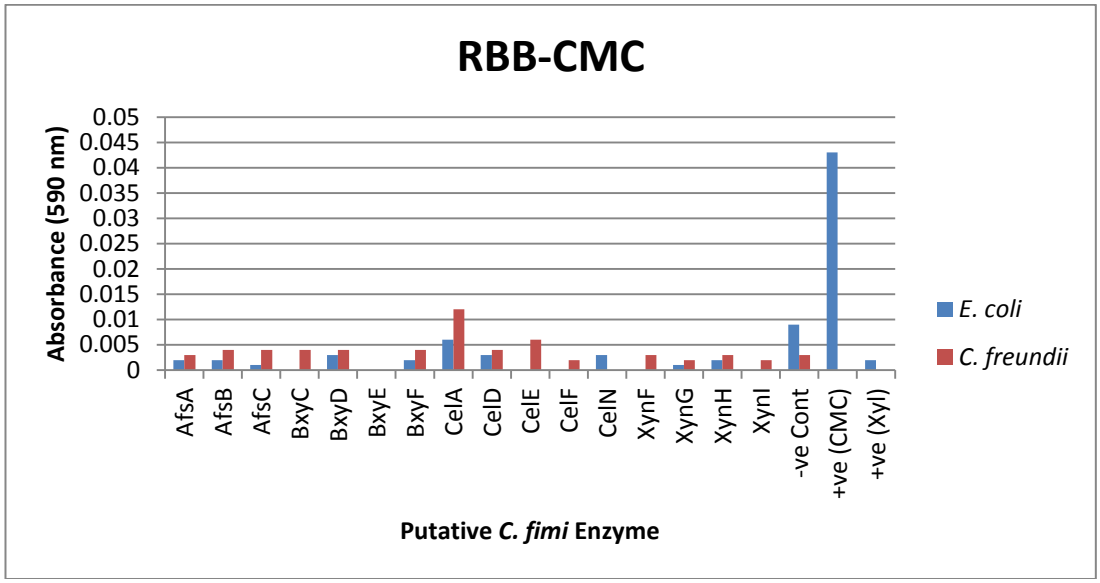


Figure 4.5. Cell lysate screen for endoglucanase activity of *C. fimi* genes using RBB-CMC. Absorbance measured at 590 nm, 3 hour incubation at 37ºC, *E. coli* host is blue, *C. freundii* host is red, -ve Cont, +ve Cont CMC and +ve Cont Xyl are *E. coli* lysates with expressed LacZ only, CenA and XynB, respectively

For the constructs expressed in *E. coli* and assayed using RBB-dyed CMC, there was no endoglucanase activity evident (figure 4.5). All activity fell below that of the Ed1 lysate and well below that recorded for CenA, a known and characterised endoglucanase from *C. fimi*. With *C. freundii* as host (figure 4.5) CelA exhibits an absorbance above background levels, nearly 3 times that of the other constructs, but again well below that of CenA.

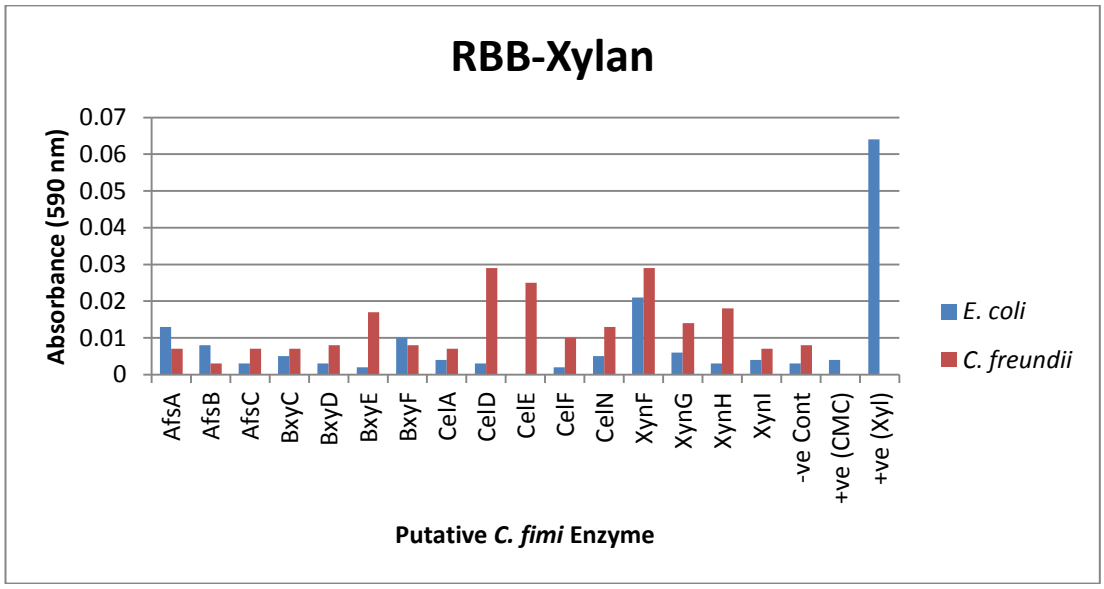


Figure 4.6. Cell lysate screen for endoxylanase activity of *C. fimi* genes using RBB-xylan. Absorbance measured at 590 nm, 3 hour incubation at 37ºC, *E. coli* host is blue, *C. freundii* host is red, -ve Cont, +ve Cont CMC and +ve Cont Xyl are *E. coli* lysates with expressed LacZ only, CenA and XynB, respectively

For the endoxylanase assay several of the enzymes are active showing absorbance values above the background and negative control. When expressed in *E. coli* (figure 4.6) XynF displays the highest level of activity. Other, lesser activity, appears to be present for AfsA and BxyF. When expressed in *C. freundii* (figure 4.6) the AfsA and BxyF activity seemingly abates to background levels. The XynF positive result is repeated, and new activity is observed for CelD, CelE, XynH and BxyE. XynG and CelN appear to also be positive, but to a much lesser degree than the aforementioned enzymes, akin to those levels observed for AfsA and BxyF in *E. coli* and are therefore arguably not true positive results, but noise.

4.3.3 *MUG Assay*

All constructs were screened for MUG activity. Figures 4.7 and 4.8 show tubes containing celll lysates that were assayed as described in the materials and methods. Tubes were visualised under UV light (excitation at 360 nm).

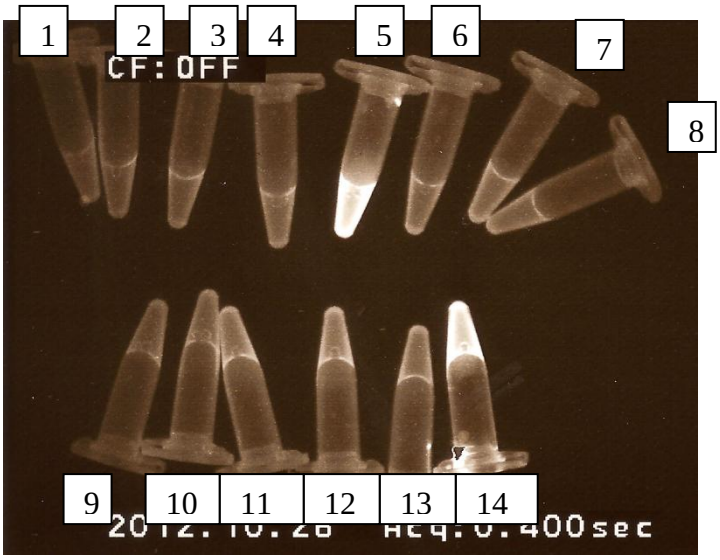


Figure 4.7. *E. coli* cell lysates assyed for β-glucosidase activity using MUG. Reaction mixtures visualised under long wave UV light. 1-14 - AfsA, AfsB, AfsC, BxyF, CelA, CelD, CelE, CelF, CelN, XynF, XynH, XynI, ED1 only, BglX *E. coli* β-glucosidase

Image 4.7 shows *E. coli* lysates and there are 2 tubes with high levels of fluorescence CelA and positive control *E. coli* β-glucosidase BglX (supplied by Damian Barnard). Two further biological repeats of CelA both proved to be negative, but were not photographed, suggesting its positive result was an anomaly most likey caused by a sample tube mix up when performing this particular experiment.

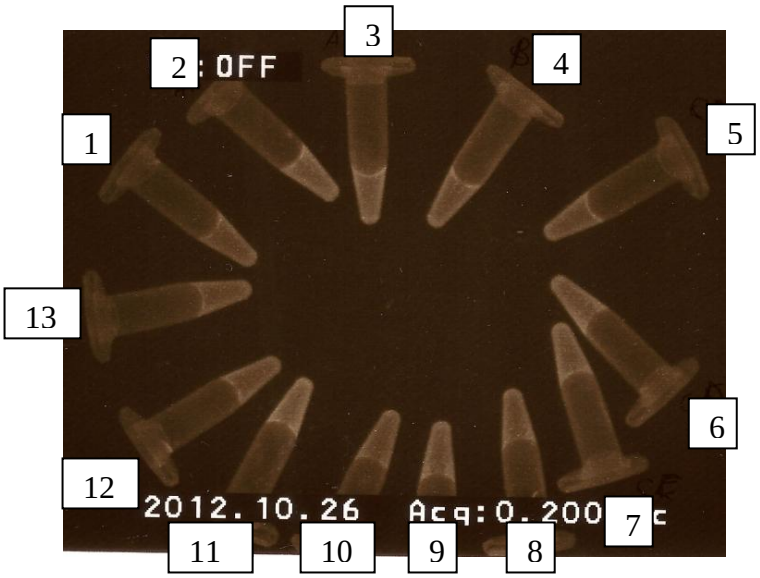


Figure 4.8. *C. freundii* cell lysates assyed for β-glucosidase activity using MUG. Reaction mixtures visualised under long wave UV light. 1-13 - AfsA, AfsB, AfsC, BxyF, CelA, CelD, CelE, CelF, CelN, XynF, XynH, XynI, ED1 only

Figure 4.8 shows the results of the *C. freundii* lysate screens. BglX was not expressed in *C. freundii* but it is clear that all the lysates are negative for glucosidase activity and were not repeated. CelA proved negative once more adding to the likelihood that the positive result of figure 4.7 was erroneous. BxyC, BxyD, and BxyE were assayed on a different day but again they all proved negative for activity; they were not photographed.

4.4 Assay Repeats

To confirm enzymatic action and show positive results were not an anomalous occurrence, biological and technical repeats of positive results were performed, but not for negative results. This was due to the β-galactosidase data showing that the promoter was functional, so a negative result was not due to non-induction of expression but some other factor as will be discussed later in the chapter. Therefore repeating these assays seemed superfluous and this methodology was primarily conceived as a fairly quick way of screening genes of interest easily expressed in *E. coli*. The specific activity for these samples was calculated and the results were averaged. As can be, seen the repeats tend to concur with the initial screens.

4.4.1 *Nitrophenol Active Enzymes*

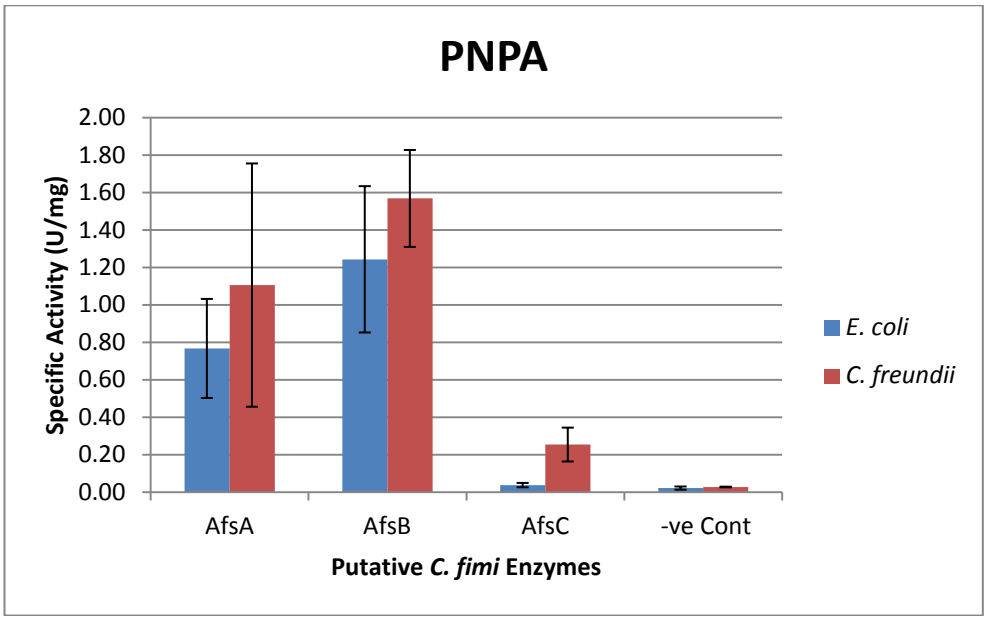


Figure 4.9. Averaged repeats of specific activity of cell lysates expressing constructs with potential enzymatic action against PNPA, -ve Cont is *E. coli* lysates expressing LacZ only. Error bars are +/- standard error. *E. coli* (blue) – AfsA n=5, AfsB n=4, AfsC n=4, -ve Cont n=2; *C. freundii* (red) – AfsA, AfsB and AfsC n=3

As can be seen in figures 4.9-4.11, the initial screens are confirmed. AfsA and AfsB showed high activity with PNPA and retained their ONPX activities (figures 4.9 and 4.11), confirming them to be bifunctional α-$_L$-arabinofuranosidases/β-xylosidases. The repeat assays also confirmed the minimal activity of AfsC with PNPA as substrate. It appears as though specific activity levels of AfsA, AfsB and AfsC are higher when using *C. freundii* as host too, although there is a large amount of variability as evidenced by the standard error (SE) values. SE was used as n was low, so would give a good measure of sample mean precision, with smaller SE denoting more trustworthy measurements.

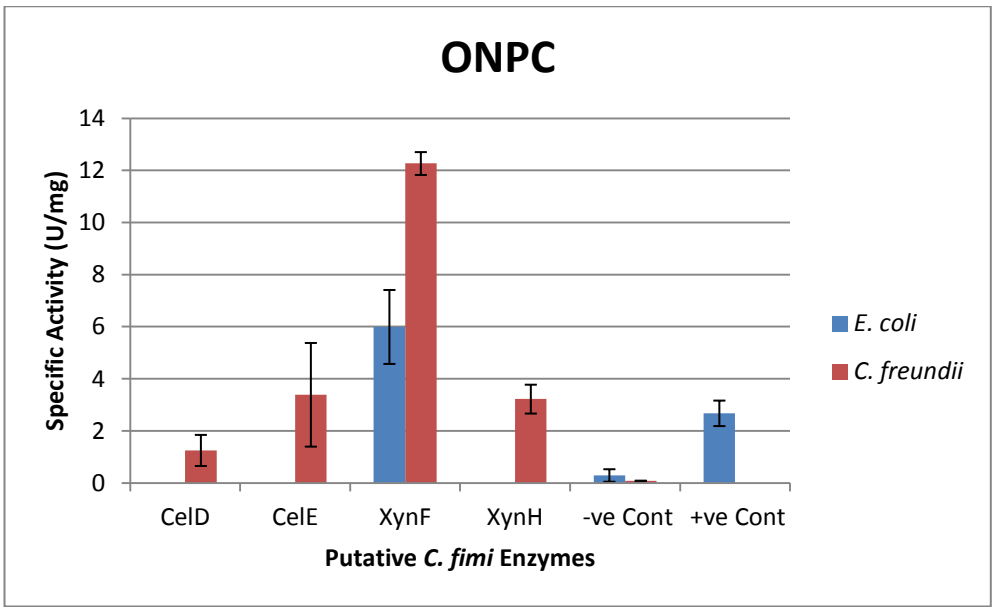


Figure 4.10. Averaged repeats of specific activity of cell lysates expressing constructs of potential enzymatic action against ONPC, -ve Cont and +ve Cont are *E. coli* lysates with expressed LacZ only and XynB, respectively. Error bars are +/- standard error. *E. coli* (blue) – XynF n=4, -ve Cont n=3, +ve Cont n=2; *C. freundii* (red) – CelD n=3, CelE n=2, XynF n=2, XynH n=2, -ve Cont n=3

XynF being the only *E. coli* lysate positive against ONPC was confirmed as having cellobiohydrolase activity (figures 4.10). Its negative result for MUG activity rules out a processively-acting β-glucosidase which may also have released nitrophenol by successive removal of glucose residues from the substrate. XynH when expressed in *C. freundii* appeared to have a small amount of ONPC activity (figure 4.2). This was confirmed during the re-screening of replicate lysates (figure 4.10). A detectable level was seen nearing that of XynB expressed in *E. coli*, well above background of the Ed1 negative control. Genes with detectable activities when expressed in *Citrobacter* and not detected when expressed in *E. coli* were also confirmed. CelD displayed both ONPC and ONPX activity not observed in *E. coli* (figures 4.10 and 4.11). CelE was another enzyme confirmed to be detectably active when expressed in *C. freundii* and not in *E. coli* (figures 4.10 and 4.11).

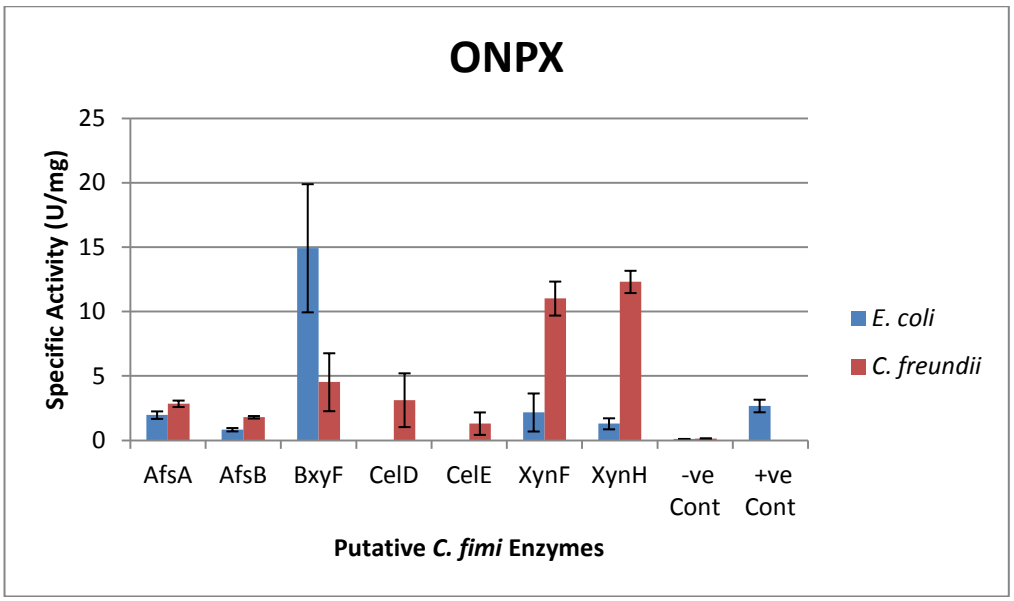


Figure 4.11. Averaged repeats measuring specific activity of cell lysates expressing constructs of potential enzymatic action against ONPX, -ve Cont and +ve Cont are *E. coli* lysates with expressed LacZ only and XynB, respectively. Error bars are +/- standard error. *E. coli* (blue) – AfsA n=3, AfsB n=2, BxyF n=2, XynF n=2, XynH n=3, -ve Cont n=4, +ve Cont n=2; *C. freundii* (red) – AfsA, AfsB, BxyF, CelD, CelE, XynF, XynH, -ve Cont n=2

ONPX hydrolysis by XynF was also confirmed indicating another bifunctional enzyme: a cellobiohydrolase/β-xylosidase (figures 4.11). Of note also are the levels of activity observed when ONPC or ONPX is the substrate in either expression host. The levels observed are greater than those seen for XynB, an archetypal form of this enzyme group. When ONPX is the substrate, the activity of XynF is most marked in *C. freundii*, being almost as high as XynH.

4.4.2 *Endo- Acting Enzymes*

Endoglucanase and endoxylanase activities were assayed using RBB-dyed CMC or xylan, respectively. The following graphs present the mean average absorbances from these biological repeats. As no endoglucanase activity was recorded for *E. coli* there were obviously no repeats performed for this host.

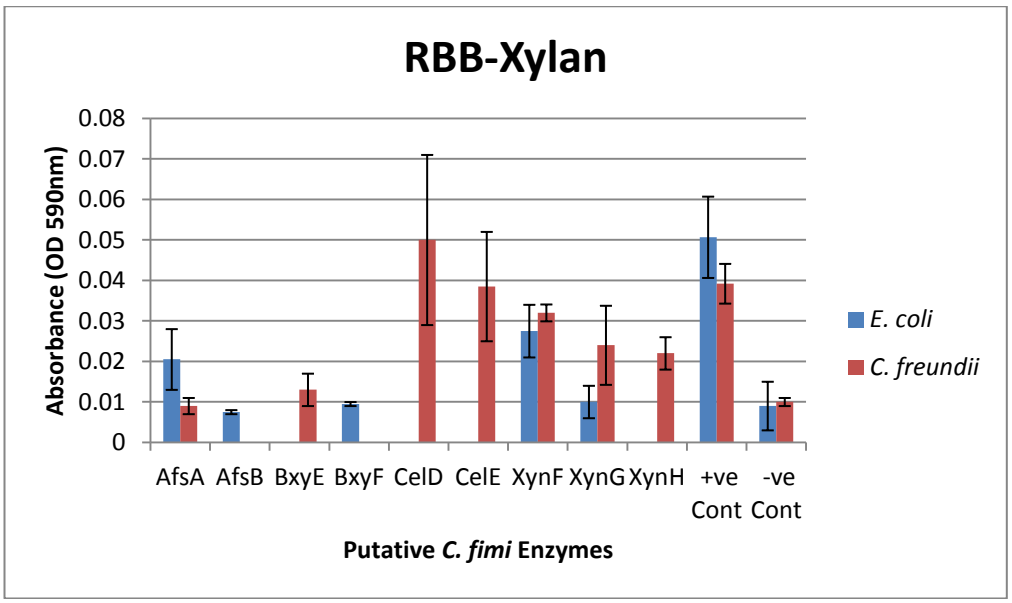


Figure 4.12. Mean absorbance at 590nm of cell lysates expressing constructs of potential enzymatic action against RBB-xylan, -ve Cont and +ve Cont are lysates expressing LacZ only and XynB, respectively. Averaged absorbance, error bars are +/- standard error. *E. coli* (blue) – all samples have n=2 apart from +ve Cont n=3; *C. freundii* (red) – AfsA, BxyE, CelD, CelE, and XynH n=2, XynF n=3, XynG n=4, +ve Cont n=5, -ve Cont n=4

From the above data (figure 4.12) it is clear that XynF has strong endoxylanase activity in both *E. coli* and *C. freundii*. This confirms XynF a multifunctional XynB-like enzyme with endo- and exoxylanase properties as well as exocellulase activity.

In *E. coli* AfsA shows a release of dye beyond that observed for the negative control, but this is not seen when expressed in *C. freundii* with levels almost identical to those observed for the negative control. For AfsB the previous xylanase activity observed in *E. coli* is not confirmed with repeat levels averaging at the same level as the negative control, the same also being true of BxyF and XynG (figure 4.12).

Under *C. freundii* expression, detectable endoxylanase activity not observed for *E. coli* expression lysates is confirmed for CelD, CelE, and XynH (figures 4.12). XynG originally proved negative for xylanase activity in *E. coli*, but positive in *C. freundii*, and this divergent activity is again confirmed with the repeats. All five positive for endoxylanase activity exhibited levels around those seen for the positive control XynB.

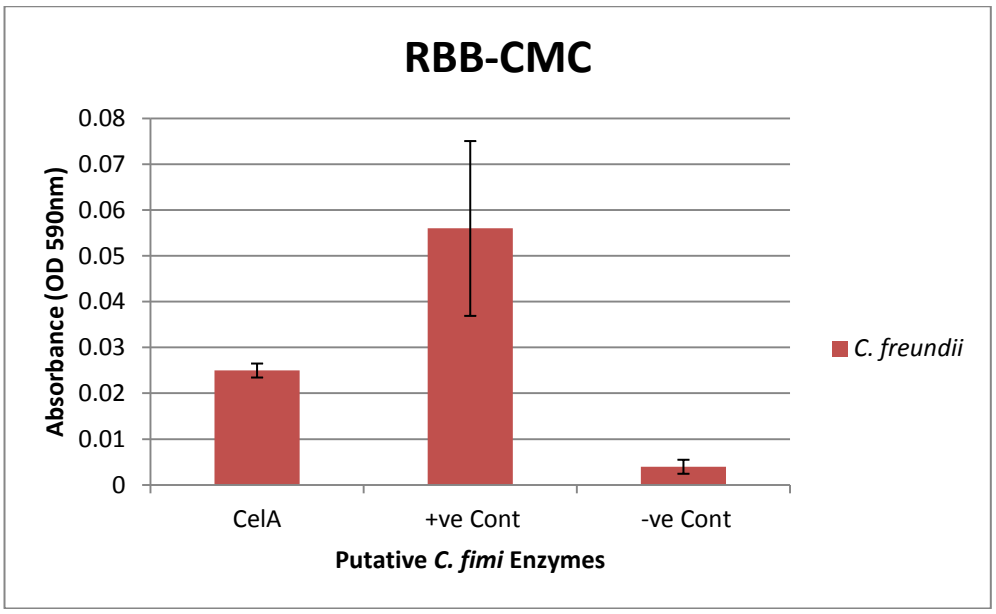


Figure 4.13. Repeats of cell lysates expressing constructs of potential enzymatic action against RBB-CMC, *C. freundii* host, -ve Cont and +ve Cont are lysates expressing LacZ only and CenA, respectively. Mean absorbance, error bars are +/- standard error. CelA, +ve Cont and –ve Cont all n=3

Under *C. freundii* expression, CelA again shows endoglucanase action above that of the Ed1 negative control (figure 4.13). Of the genes cloned this just leaves CelF, CelN, BxyC, BxyD and XynI with no detectable activity against the substrates assayed for. AfsC although very weak, has detectable activity for PNPA.

It is evident too that for all constructs assayed, their specific activity levels were generally higher when expressed in *C. freundii* than with expression in *E. coli* although variance in the repeats is quite large for some. It is hard to draw conclusions on activity levels as a comparison between hosts with this data, but they do show a definite trend to be further explored.

4.5 *E. coli* Expression vs. *C. freundii* Expression

To assess more accurately if the differences in activity levels were genuine, direct same day comparisons of biological replicates were performed on a subset of the active enzymes.

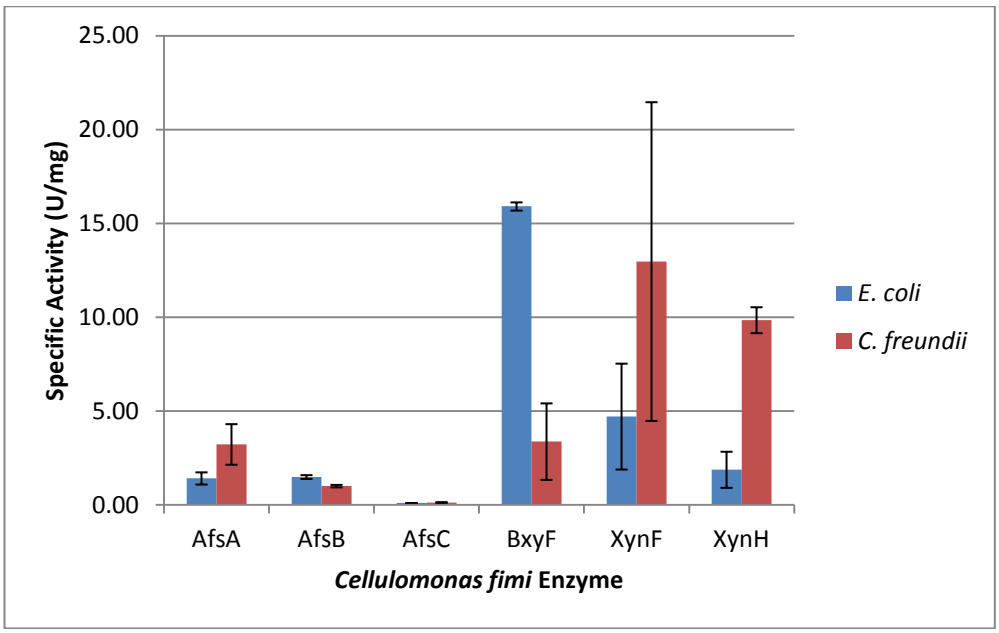


Figure 4.14. Mean repeats of cell lysate specific activity against PNPA (AfsA, AfsB and AfsC), ONPC (XynF) and ONPX (BxyF and XynH) with constructs expressed in either *E. coli* (blue) or *C. freundii* (red). Error bars are +/- standard error. For all enzymes n = 2.

As can be seen in figure 4.14, there is a lot of variation in the samples assayed. To reduce possible variance as much as possible, the above replicates were repeated simultaneously. Despite the general expression noise levels, there appears to be a definite qualitative difference in activity between enzyme activity in *C. freundii* and *E. coli*. For the enzymes assayed above the specific acitivity levels are generally higher for *C. freundii* expression than for *E. coli*. This is the same general trend as was seen for the original activity screens in the two different hosts. Possible reasons for this difference will be discussed later in the chapter.

4.6 Expression Analysis

To determine if the lack of detected functionality for some of the enzymes was due to either no protein present or incorrect substrate, several enzymes had an N-terminal 6xHis tag added to their sequence immediately after the ATG start codon.

The constructs were expressed in *E. coli* under IPTG induction, lysed by sonication in either native binding buffer or denaturing binding buffer as described in Materials and Methods. An Ni-NTA column with a bead volume of 1 ml was used for binding

of potential protein, so as to maximise competition between proteins for binding and reduce the possibility of contaminating untagged proteins. Fractions were analysed by SDS-PAGE as described previously. AfsA was His-tagged as a positive control for purification. BxyD, CelF and CelN were tagged to assess the presence of protein. *E. coli* expressing LacZα only under native conditions, and untagged AfsA also under native conditions were used as negative controls to determine what, if anything, would bind when no tag was present. All gels were silver stained, but if that resulted in no more banding becoming visible, then just the coomassie stain gel is included.

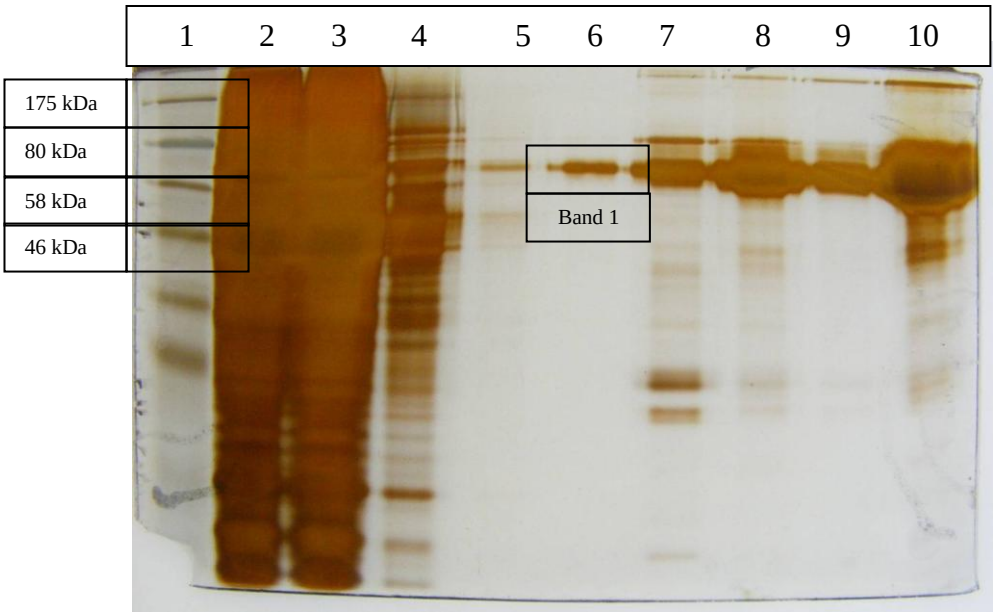


Figure 4.15. SDS-PAGE of His-AfsA purification fractions under native conditions, silver stain. Lane 1 – marker, lane 2 – His-AfsA native cell lysate, lane 3 – flow through, lane 4 – wash 1, lane 5 – wash 2, lane 6 – wash 3, lane 7 – 0.5 M imidazole elution 1, lane 8 – 0.5 M imidazole elution 2, lane 9 – 0.5 M elution imidazole 3, lane 10 – beads

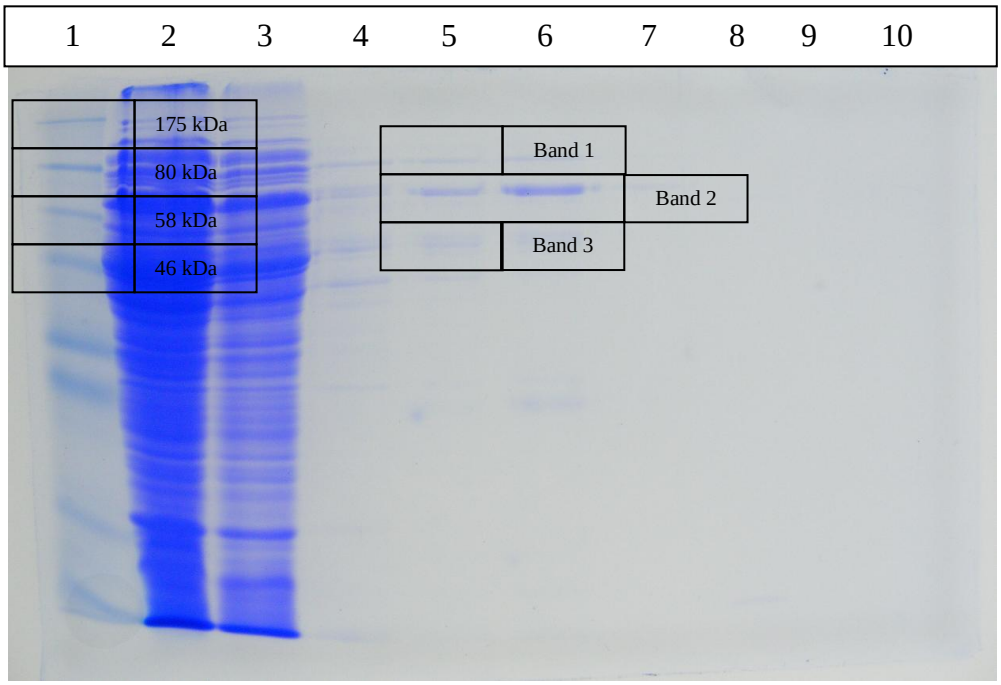


Figure 4.16. SDS-PAGE of His-AfsA purification under denaturing conditions, coomassie stain. Lane 1 – marker, lane 2 – His-AfsA denatured cell lysate, lane 3 – wash 1, lane 4 – wash 3, lane 5 – 0.5 M elution 1, lane 6 – 0.5 M elution 2, lane 7 – 0.5 M elution 3, lane 8 – 1 M elution 1, lane 9 – 1 M elution 2, lane 10 – 1 M elution 3

Figures 4.15 and 4.16 show the purification fractions of His-tagged AfsA in both native and denaturing buffers. There is a clear band in both of roughly the same estimated molecular weight, ~73 kDa (band 1 and band 2 in figure 4.15 and figure 4.16, respectively). Based on predicted sequence, untagged AfsA is 55 kDa and tagged is 56 kDa. The large discrepancy could be partly attributed to MW calculation errors, leading to an over estimation of the true molecular weight. The difference may also be due to modifications such as glycosylation adding to migration retardation, or changes in polypeptide charge through lack of SDS binding [168, 169]. Despite the MW discrepancy, I feel confident that in both cases what we are seeing is tagged AfsA, due to the intensity of the bands and the presence in later stage elutions. These bands are not seen in the negative control of LacZα purification (figure 4.22), untagged AfsA (figure 4.23), or in the fractions of the other tagged proteins (figures 4.17-4.21).

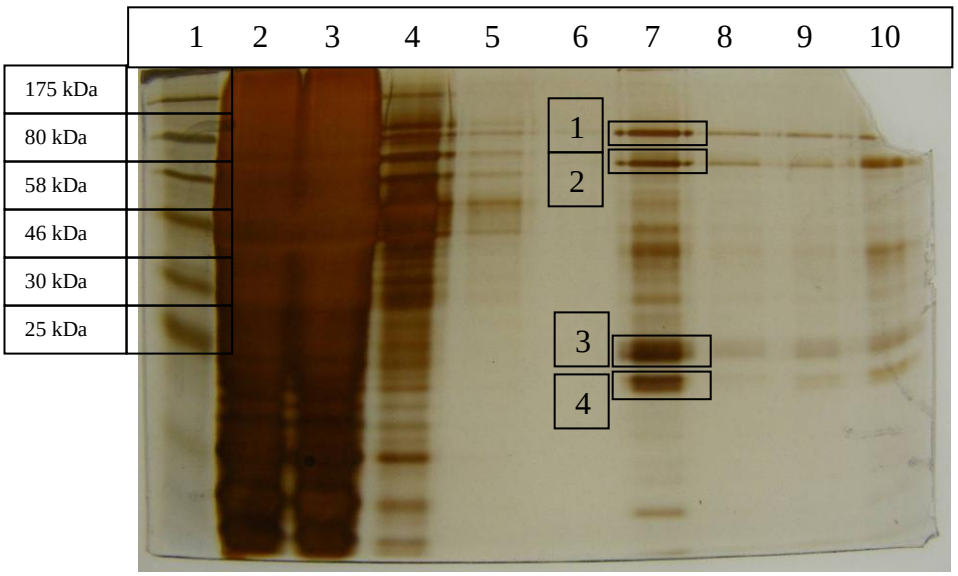


Figure 4.17. SDS-PAGE of His-CelN purification fractions under native conditions, silver stain. Lane 1 – marker, lane 2 – His-CelN native cell lysate, lane 3 – flow through, lane 4 – wash 1, lane 5 – wash 2, lane 6 – wash 3, lane 7 – 0.5 M imidazole elution 1, lane 8 – 0.5 M imidazole elution 2, lane 9 – 0.5 M elution imidazole 3, lane 10 – beads

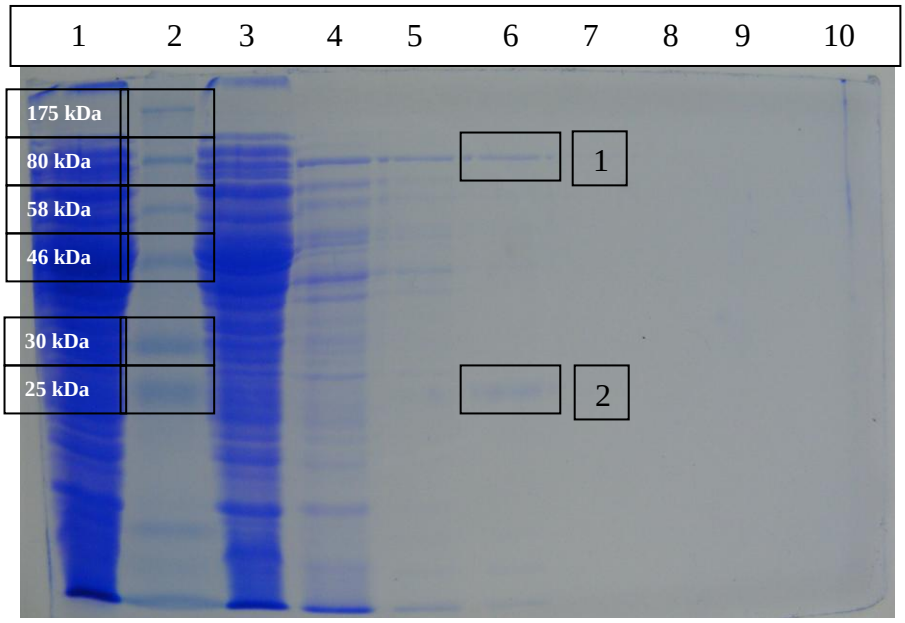


Figure 4.18. SDS-PAGE of His-CelN purification fractions under denaturing conditions, coomassie stain. Lane 1 – His-CelN denatured cell lysate, lane 2 – marker, lane 3 – wash 1, lane 4 – wash 3, lane 5 – 0.5 M elution 1, lane 6 – 0.5 M elution 2, lane 7 – 0.5 M elution 3, lane 8 – 1 M elution 1, lane 9 – 1 M elution 2, lane 10 – 1 M elution 3

Under native or denaturing conditions, His-CelN produces no one band that appears more abundant than another (figures 4.17 and 4.18) and they all lessen equally as the imidazole concentration of the elutions increases, in contrast to the increasing band intensity observed for His-AfsA (figure 4.15). Under native conditions there are a lot

of protein bands not visible when His-AfsA was purified (figures 4.15 and 4.16), suggesting that there was no competition for binding as would be expected in the presence of a tagged peptide, as also seen in the 2 His-negative gels (figures 4.22 and 4.23). Tagged CelN is 54 kDa, making it smaller than His-AfsA. The major bands of figure 4.17 are estimated to be 92 kDa, 75 kDa, for bands 1 and 2, respectively, and 21 kDa and 19 kDa, for bands 3 and 4, respectively. These are either larger than the band assumed to be His-AfsA (bands 1 and 2), or might be expected for His-CelN even allowing for size estimation errors or modifications, or far too small (bands 3 and 4).

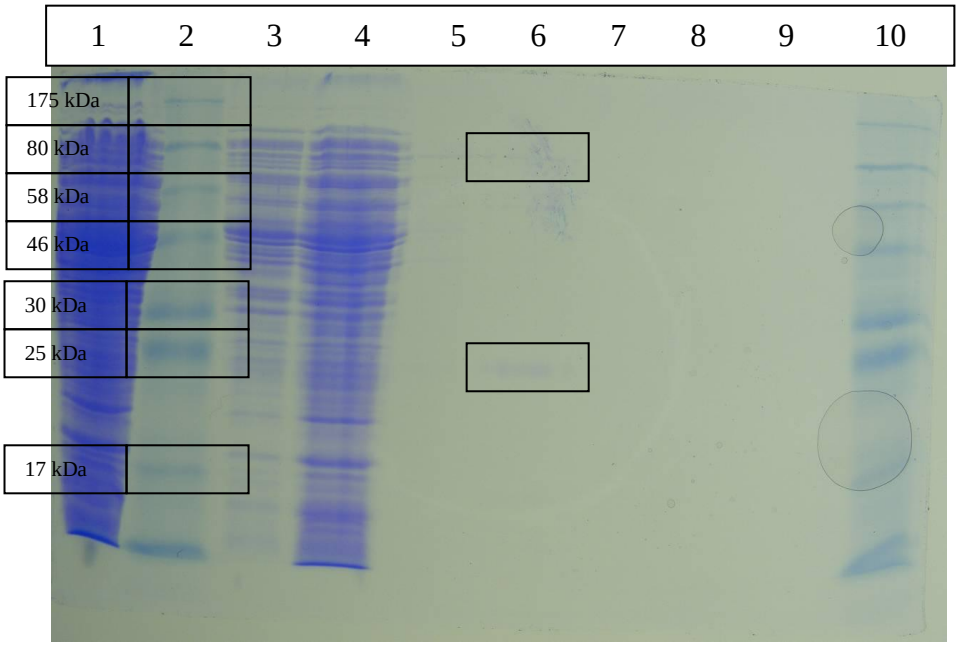


Figure 4.19. SDS-PAGE of His-BxyD purification under denaturing conditions, coomassie stain. Lane 1 – His-BxyD denatured cell lysate, lane 2 – marker, lane 3 – wash 1, lane 4 – wash 3, lane 5 – 0.5 M elution 1, lane 6 – 0.5 M elution 2, lane 7 – 0.5 M elution 3, lane 8 – 1 M elution 1, lane 9 – 1 M elution 2, lane 10 – marker

Under denaturing conditions (figure 4.18), the picture is much simplified when compared to the native gel (figure 4.17). The two bands observed are estimated to be 81 kDa and 24 kDa, bands 1 and 2, respectively, and appear to be the same bands as those purified for His-BxyD under denaturing conditions (figure 4.19), and are also evident in the His-AfsA gels (figures 4.15 and 4.16). Given the near identical gel profiles between denatured His-BxyD and His-CelN, it would seem there is no tagged protein present, especially when compared to the gels of His-AfsA.

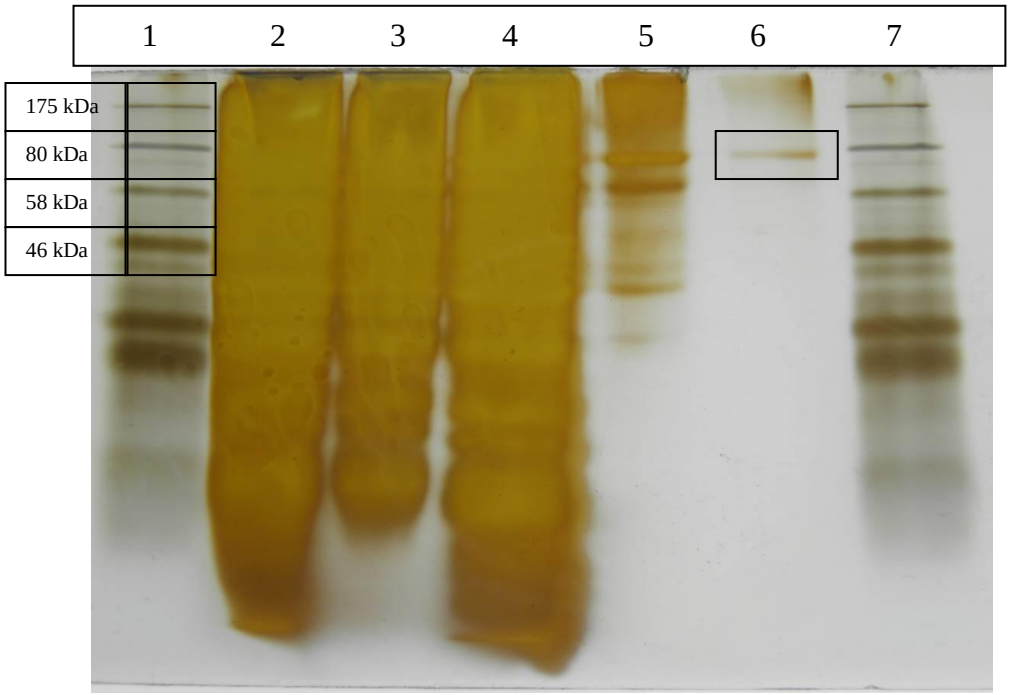


Figure 4.20. SDS-PAGE of His-CelF purification under denaturing conditions, silver stain. Lane 1 – marker, lane 2 – His-CelF denatured cell lysate, lane 3 – flow through, lane 4 – wash 1, lane 5 – wash 2, lane 6 – wash 3, lane 7 – marker

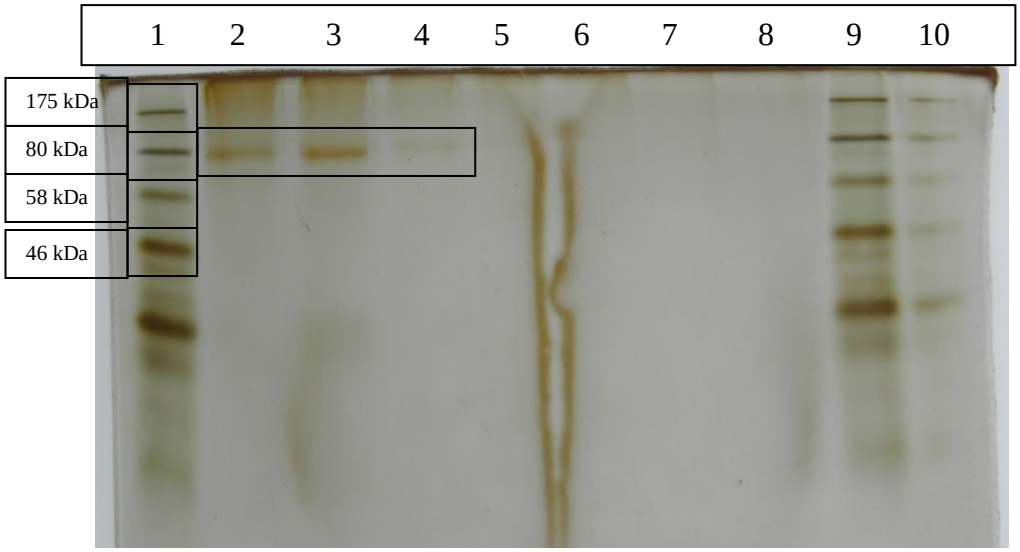


Figure 4.21. SDS-PAGE of His-CelF purification under denaturing conditions, silver stain. Lane 1 – marker, lane 2 – His-CelF denatured cell lysate, lane 3 – flow through, lane 4 – wash 1, lane 5 – wash 2, lane 6 – wash 3, lane 7 – marker

The purification of His-CelF was split over two gels, with the elution fractions on the second gel. The only band visible however, was that observed in lane 6 figure 4.21, the third wash step as seen in figure 4.21. This was estimated to be 79 kDa, the same size as the consistently present bands observed for all gels including the His-AfsA gels. Tagged His-CelF is 57 kDa.

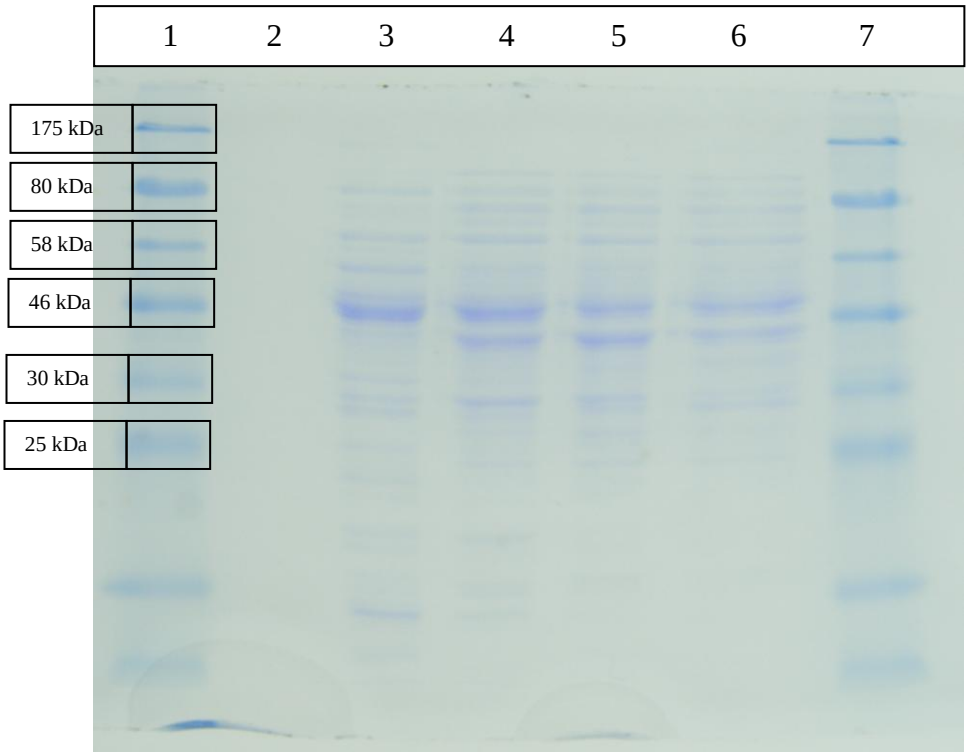


Figure 4.22. SDS-PAGE EdinBrick 1 purification under native conditions, coomassie stain. Lane 1 – marker, lane 2 – Ed1 native cell lysate flow through, lane 3 – wash 1, lane 4 – wash 2, lane 5 – 0.5 M imidazole elution, lane 6 – 1 M imidazole elution, lane 7 – marker

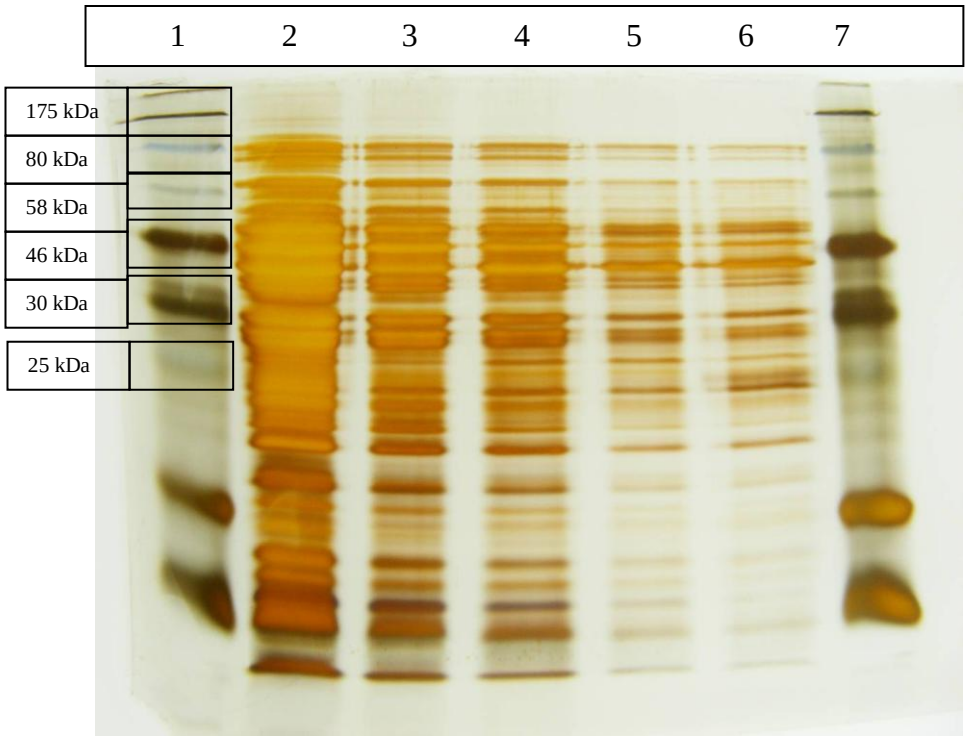


Figure 4.23. SDS-PAGE untagged AfsA purification under native conditions, silver stain. Lane 1 – marker, lane 2 – AfsA native cell lysate, lane 3 – flow through, lane 4 – wash 1, lane 5 – wash 2, lane 6 – 0.5 M imidazole elution, lane 7 - marker

Figures 4.22 and 4.23 are Edinbrick1 only expressing LacZα and untagged AfsA under native conditions, respectively. There is no evidence of the strong banding observed for His-tagged AfsA in either gel. It is evident from these also that without competitive binding to the Ni-NTA column by tagged overexpressed proteins, that a number of proteins native to *E. coli* are able to stick to the column and be eluted, most likely histidine rich or possessing 2^+ metal ion binding ability, including the bands observed throughout at ~80 kDa and <25 kDa. Bolanos-Garcia et al, performed a review of the commonly co-purified proteins that they observed finding them often to be stress response proteins ranging in sizes from 10 kDa to 74 kDa, but it is not an exhaustive list and will vary dependent on host strain used [170, 171].

Discussion

4.7 Function and Expression

4.7.1 *Annotation*

Eleven out of the sixteen cloned genes proved positive for functionality leaving only BxyC, BxyD, CelF, CelN and XynI with no detected activity in either *E. coli* or *C. freundii*. This may be due to one of several reasons: incorrect assumption of function based on BLAST and domains, not screened for the correct substrate or no expression in either host.

The majority of the enzymes matched closely to the predicted putative functions based on the top 20 hits from BLAST. There are discrepancies however. Besides the enzymes with no detected function, CelE and BxyE displayed activities not predicted through sequence homology. Others displayed activities not evidenced from characterised enzymes of the same glycoside hydrolase families reported in the CAZy database [131]. Table 4.2 gives a comparison between observed enzymatic activities, predicted function based on the top 20 hits from BLAST searches of amino acid sequence, protein domains and glycoside hydrolase family and any activities observed that are not in CAZy.

The problem with using BLAST to annotate a genome is that very few actually characterised enzymes are present when compared to the number of automatically annotated proteins from genomes. This is evidenced when looking at the top 10 hits for the *C. fimi* proteins above. Eleven of the 16 proteins had no characterised enzyme hits in the top 10. All were computationally annotated genomes or conceptual translations of genes then annotated automatically. AfsB's sixth hit was a characterised arabinofuranosidase hit from *Rhodanobacter ginsenosidimutan*. CelA's eighth hit was a characterised cellulase from *Paenibacillus xylanilyticus*. CelD and CelE matched to characterised endoglucanase enzymes within their top hits. XynF had two characterised enzymes within its top hits, one being a match to XynE from *C. fimi* itself, the top hit in fact, and the other hit to a bifunctional endoxylanase from *Demequina sp.* JK4. What we see in these cases is that the observed gene product function actually matches with the predicted. The five genes that show no activity have no characterised enzymes in their top hits, casting some doubt over their annotations.

It is entirely possible that these genes have been incorrectly annotated adding to the problem with sequence homology annotation. In 1999 Kyrpides et al note the dubious annotation of genes for *Chlamydia trachomatis* [172] and Gilkes et al [173, 174] describe an "error percolation" in databases through the annotation of an enzyme based on the erroneous annotation of another by sequence homology which greatly increases the probability of erroneous annotations as databases grow ever larger. The use of protein domains can help matters, but the same problems still persist [175]. The GH families are grouped based on sequence homology and have associated functions based on gene annotation. There are far more uncharacterised proteins than characterised proteins available, a lot of which must also be erroneously annotated as above. Simply looking at the characterised enzymes listed also does not give the full picture, but is a more accurate way to annotate a protein with only missed functionality a possibility as seen below in table 4.2. Arabinofuranosidase GH families were inferred by hierarchical clustering with characterised enzymes from each of the 5 potential GH families which exhibit arabinofuranosidase activity using T-coffee [176]. Figure 4.24 is the cladogram

generated by T-coffee. The GH family is stated and abbreviated organism name is given. Bacterial enzymes were used where possible, those that are not eukaryotic are denoted with "euk" suffix. AfsA, B, C and D had no predicted GH family by InterProScan, but in the below cladogram they clearly group with GH51 enzymes. As the representative characterised members of each family also group with other members of the same family suggests the arrangement by T-coffee is likely accurate and the 4 arabinofuranosidases belong to GH family 51. AfsE was predicted by InterProScan to be GH family 62 and this also groups with other GH62 members. Glycoside hydrolase family 54 is made up exclusively of eukaryotic enzymes and has not been found in any bacterial species to date.

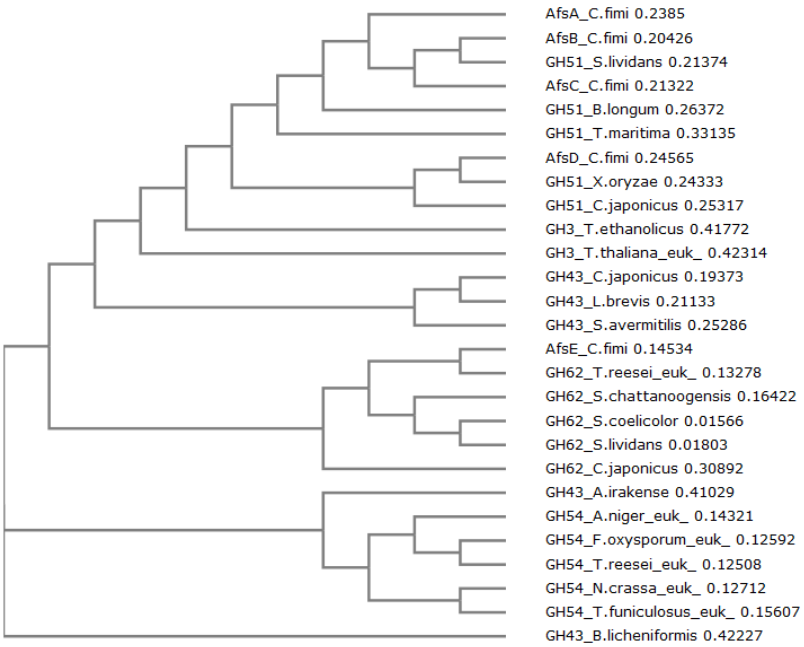


Figure 4.24. Cladogram generated using T-coffee of representative characterised arabinofuranosidase amino acid sequences as listed in CAZy with default settings.

Protein	Observed Enzymatic Action	Predicted Function	Glycoside Hydrolase Domains	Activity Not in CAZy database
AfsA	Arabinofuranosidase, xylosidase	Arabinofuranosidase	arabinofuranosidase C terminus, family independent GH catalytic core	
AfsB	Arabinofuranosidase, xylosidase	Arabinofuranosidase	arabinofuranosidase C terminus, family independent GH catalytic core	
AfsC	Arabinofuranosidase	Arabinofuranosidase	arabinofuranosidase C-terminus, family independent GH catalytic core	
BxyC	None	Xylosidase/arabinofuranosidase	GH 43	No activity
BxyD	None	12/20 GH 43 protein (2/20 xylosidase)	GH 43	No activity
BxyE	Endoxylanase	10/20 GH 43 protein (7/20 xylosidase)	GH 43	
BxyF	Xylosidase	8/20 GH 39 (7/20 xylosidase)	GH 39	
CelA	Endoglucanase	Endoglucanase	GH 5	Xylosidase, endoxylanase
CelD	Cellobiosidase, xylosidase, endoxylanase	8/20 GH family proteins (7/20 endoglucanase)	GH 9	Xylosidase, endoxylanase
CelE	Endoxylanase	4/20 GH 9, 4/20 GH 5, 3 cellobiosidase, 2 endoglucanase	GH 9 N-terminus	Endoxylanase
CelF	None	Endoglucanase	GH 9	No activity
CelN	None	Cellulase	GH 5	No activity
XynF	Cellobiosidase, xylosidase, endoxylanase	Endoxylanase	GH 10	Xylosidase
XynG	Endoxylanase	12/20 GH 10 protein (6/20 endoxylanase)	GH 10	
XynH	Cellobiosidase, xylosidase, endoxylanase	Xylosidase	GH 39	Cellobiohydrolase, endoxylanase
XynI	None	Endoxylanase	None	No activity

Table 4.2. Comparison of observed enzyme functionality and predicted function. Amino acid sequences were run through BLAST and the top hits recorded. Where a clear majority for a function was seen this was recorded. Counts of hits are given where this was not the case. Glycoside hydrolase family domains were predicted using InterProScan [97, 177]. The CAZy database (www.cazy.org, [131]) has characterised enzymes listed under each glycoside hydrolase (GH) family. Functionality not mentioned in the characterised enzymes has been noted.

4.7.2 *Expression*

While BLAST annotations are a possible problem and reason for no detected activity, the lack of expressed protein is the most likely culprit for negative assays. The difference in enzyme specific units between *E. coli* and *C. freundii* and the lack of detectable His-tagged proteins point towards gene expression as a problem.

His-tag purification can fail for a number of reasons. The tags here were N-terminal tags meaning that the tag could be cleaved if the protein contained a signal peptide. Of the genes tagged only CelN had a potential signal peptide (as predicted by SignalP v 3.0, [178, 179]). AfsA, BxyD and CelF have no predicted signal peptides, with AfsA predicted as cytoplasmic by Psortb [98], BxyD and CelF as unknown and CelN as extracellular. The signal may explain why CelN does not appear to be present beyond lack of expression, however I would expect to see some purified protein as it couldn't all be processed, this is not the case and as for the other 2 proteins no definitive bands are seen. Inaccessibility of the tag may also be a reason for lack of purification. This should be negated by the denaturing conditions, certainly enough to be detected as a band. For these reasons I feel it is safe to say that these genes are not expressed, or not properly expressed, leading to no detectable functionality.

4.7.2.1 *Codon Bias*

By comparing the percent composition of codons within *C. fimi* genes expressed and potentially unexpressed in *E. coli* with the percent codon composition of all *E. coli* genes, I hoped to identify potentially problematic codons. Codon bias generally reflects tRNA population, therefore the less used the codon in *E. coli*, the fewer tRNA molecules available for translation and protein synthesis, so therefore limited amounts or no protein produced [180-182]. Figure 4.25 shows the mean percentage counts for each codon triplet for assay negative genes (potentially unexpressed) and assay positive genes. Also displayed are the codon frequencies of the entire *E. coli* genome and the codon frequencies of *C. freundii* based on 207 CDS's (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=546, [183]).

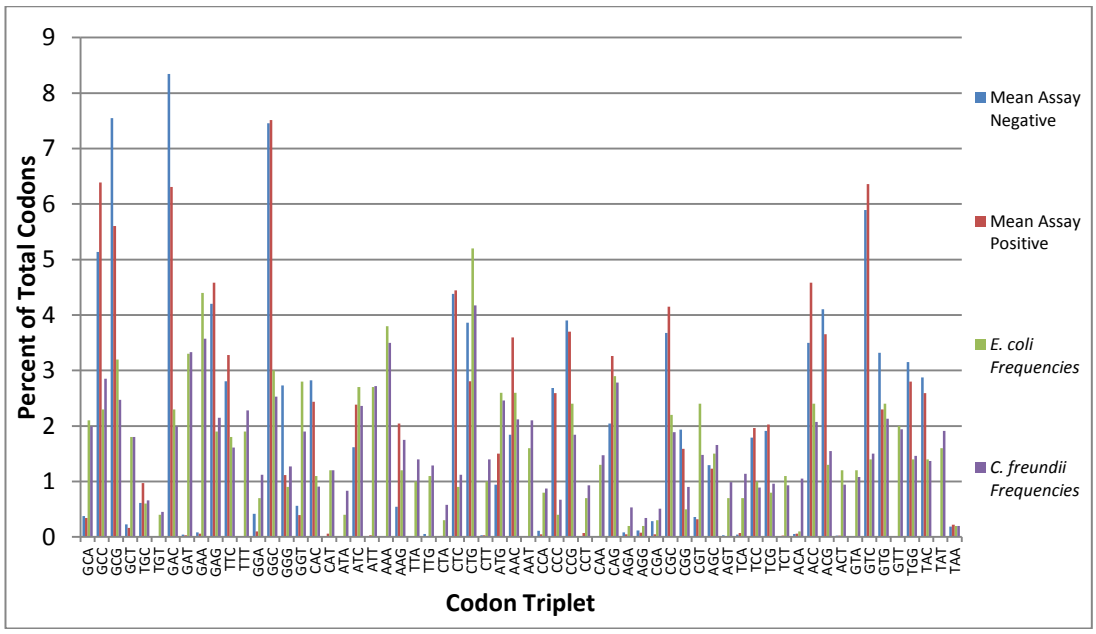


Figure 4.25. Mean frequencies of codons as a percentage of total codons.

By looking at figure 4.25 it appears that there are codons more frequently used in *C. fimi* than in *E. coli* or *C. freundii* and vice versa. Codons GCC, GCG, GAC, GGC and CTC for example are all considerable more frequent for both the expressed and assay negative genes when compared to either host species. However, as many of the assay negative codon frequencies above *E. coli* frequencies are found in the expressed genes as well, this alone cannot be the determining factor in possible expression differences. By comparing the codon frequencies of assay negative and expressed genes using the students t-test, 2 tailed, unequal variance, a mere 5 codons are significantly different at the 5% confidence level – GCG (alanine, assay negative 7.5%, expressed 5.6%), GAC (aspartic acid, assay negative 8.3%, expressed 6.3%), GGA (glycine, assay negative 0.4%, expressed 0.1%), GGG (glycine, assay negative 2.7%, expressed 1.1%) and AAG (lysine, assay negative 0.5%, expressed 2.0%), with differences in GGA, GGG and AAG frequencies significant at the 1% confidence level. Of these five significantly different codons, AAG is of lower frequency in assay negative genes than expressed, but the reverse is true for the other 4. GGA frequencies for both sets of genes are lower than the frequencies of *E. coli* (0.7%) and *C. freundii* (1.12%); GCG, GAC and GGG are above the frequencies of *E. coli* (3.2%, 2.3%, 0.9%, respectively) and *C. freundii* (2.47%, 1.99%, 1.27%, respectively) for both sets of genes. Only GGG is universally higher than the *E. coli*

and *C. freundii* frequencies for the assay negative genes, but not universally higher for the expressed genes.

This may again suggest that GGG is a limiting factor in expression of *C. fimi* genes in *E. coli*. The codon bias for *C. freundii* is quite different to that of *E. coli*. Thirty four of the sixty four codons are more common in the former than for the latter host. In the expressed genes there were 13 codons where all the genes were above the level of *E. coli*. Of these GCC, GAG, CTC, CCC, GTC and TGG were more abundant in *C. freundii* than *E. coli*. This may account for the apparent increase in U/mg observed when *C. freundii* is the host – more abundant heavily used tRNA pools leading to more gene translation and therefore more enzyme present. These differences in codon bias were however generally quite small between the 2 organisms. Only the proportion of ACA was greatly higher in *C. freundii* making up 1.05% of the codons, to 0.1% in *E. coli*. A difference in expression between *E. coli* and *C. freundii* has been documented once before but no possible explanation was proffered as to why [184].

Figures 4.26 and 4.27 show the percentages of a given codon from the total of codons synonymous for an amino acid. This may show a bias in the assay negative enzyme genes towards a particular set of amino acids that could be putting a metabolic burden on amino acid availability not obvious by comparing individual codons.

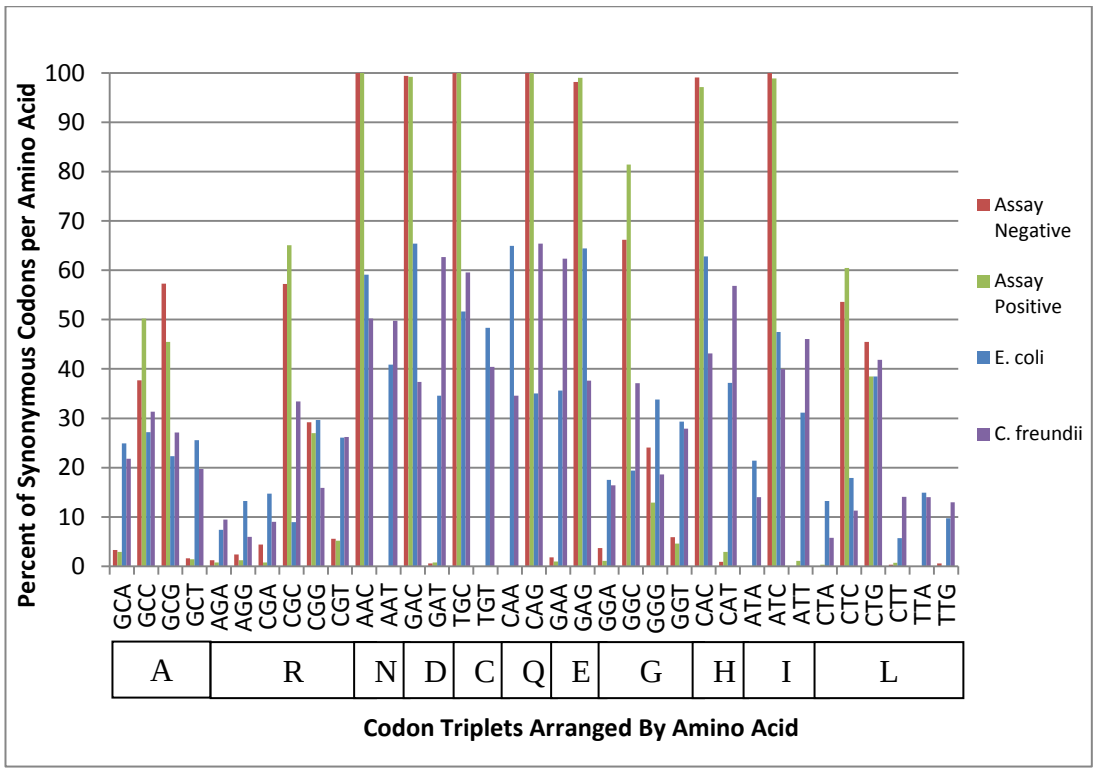


Figure 4.26. Codon frequencies as a percentage of total synonymous codons per amino acid alanine to leucine in alphabetical order, single letter amino acid code is given beneath the relevant codons. Codon counts as a total from all assay negative genes and all expressed genes generated using CUSP from EMBOSS [185]

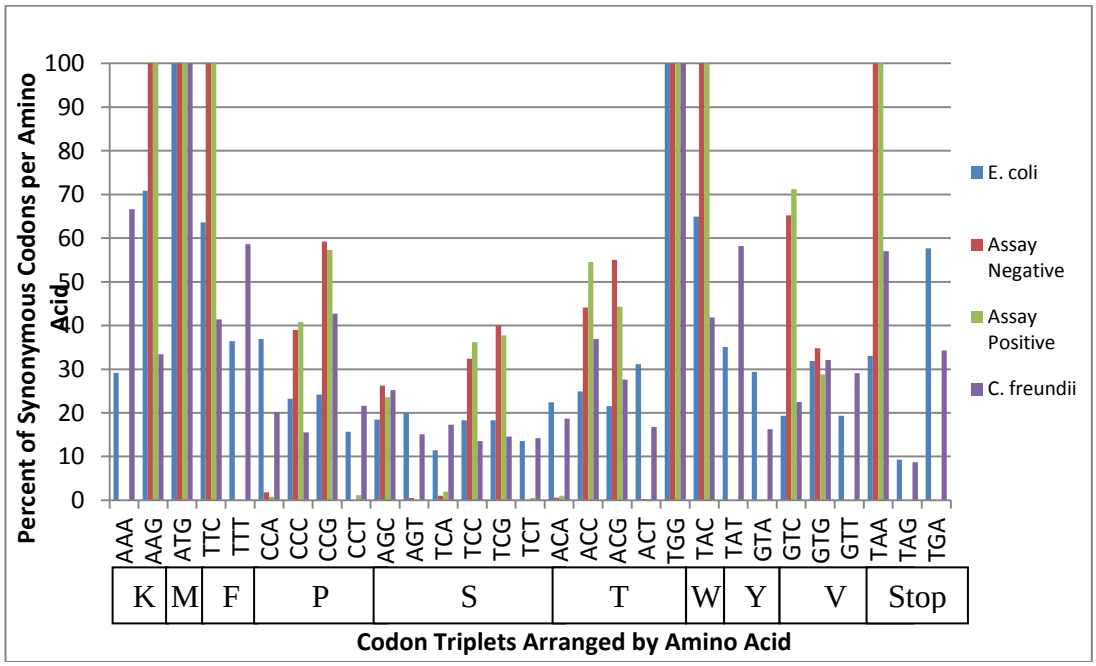


Figure 4.27. Codon frequencies as a percentage of total synonymous codons per amino acid lysine to the 3 stop codons in alphabetical order. The *C. fimi* genes only have TAA as they were artificially given that stop codon during the cloning process. Codon counts as a total from all assay negative genes and all expressed genes generated using CUSP [185]

There is an obvious bias in the *C. fimi* genes towards codons primarily comprising GC. This leads to a heavy bias for some of the amino acids to be supplied in synthesis by one or two tRNAs. For alanine (figure 4.26), almost 100% of the synonymous codons are GCC or GCG for the *C. fimi* genes where as in both *E. coli* and *C. freundii* the codon usage is far more evenly spread. For arginine codons CGC and CGG are the major triplets in *C. fimi* comprising 65% and 27% for expressed genes, respectively and 57% and 29% for assay negative, respectively. In *E. coli* and *C. freundii* these triplet proportions are different with 9% and 33%, respectively for CGC and 30% and 15% for CGG, respectively. Some codons are not recorded at all for some amino acids such as AAT for asparagine, TGT for cysteine, CAA for glutamine, ATA for isoleucine, TTA and TTG for leucine, AAA for lysine, TTT for phenylalanine, AGT for serine, TAT for tyrosine and GTA and GTT for valine. Others are just very low with synonymous codon proportions below 5%. These codon proportions may not accurately represent the true bias of the *C. fimi* genome, but it is clear that there is a huge bias in almost all amino acids for specific codons in the genes cloned. What there isn't however, is a clear bias between the synonymous codons of the expressed genes versus those of the assay negative genes which was somewhat to be expected given the lack of significant differences between the overall individual codon frequencies (figure 4.25).

There is no apparent bias in synonymous codon proportions of the assay negative genes compared to expressed genes. The proportion of synonymous codons is largely the same with only a handful of synonymous codons being significantly different. By comparing the proportions of synonymous codons using a 2-tailed z-test at the 5% confidence level, GCC and GCG coding for alanine, CGA for arginine, GGA, GGC and GGG for glycine, and ACC and ACG for threonine were the only ones significantly different. Of these GCG, CGA, GGA, GGG and ACG were of a higher proportion in the assay negative genes to assay positive. CGA and GGA are recorded as 2 of the 8 lowest abundance codons in *E. coli* and known to have an effect on gene expression levels [180]. Comparing the synonymous codon usage of *E. coli* to *C. freundii* revealed that nearly all synonymous codons are significantly different between the species bar CGT (arginine), GGA and GGT

(glycine), TTA (leucine), ATG (methionine), TCT (serine), TGG (tryptophan), and GTG (valine). Of the 5 codons that are proportionally more abundant in assay negative genes, CGA and GGG are significantly higher in *C. freundii* than *E. coli* while GCG and ACG are significantly lower (GGA being not significantly different).

4.7.2.2 *Translation Initiation Rates*

Recently it has been shown that codon bias is not the only affecter of gene expression. The ribosome binding site (RBS) and its flanking DNA sequence can also have a large effect [186-188]. The Salis laboratory has developed the RBS calculator, an online tool, so any sequence can be assessed to determine its relative translation initiation levels [187]. It does this by calculating the total energy change in Gibbs free energy when the DNA is complexed to the 30S complex of the ribosome, which is then related to a relative change in translation initiation rate for that sequence and the specified 16S rRNA sequence. The calculator then generates a translation initiation score (TU), indicating the relative strength of the RBS, where a score of 10 or 1000 is say 10 fold weaker than a score of 100 or 10,000 respectively.

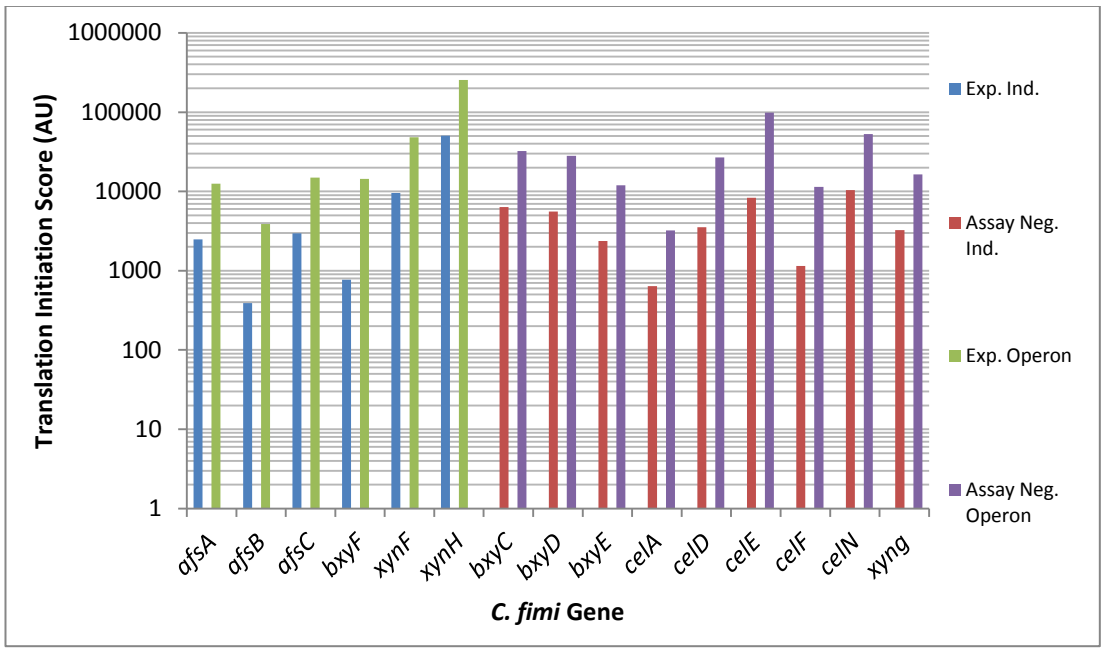


Figure 4.28. Translation initiation rates for *E. coli* (16S rRNA sequence of ACCTCCTTA) as calculated by Salis group RBS calculator (site). Genes *afsA-xynH* are assay positive (exp.) in *E. coli*, and *bxyC-xynG* are assay negative (assay neg.). Translation initiation scores were generated using either the "reverse engineer RBS" tool (Ind.) or "predict operon expression" tool (operon)

The "reverse engineer RBS" tool was initially used which takes a DNA sequence only. The input sequence of all genes was -300 bases before the start codon and as much of the gene as the tool input would allow, with "*Escherichia coli* str. K-12 substr. MG1655 (ACCTCCTTA)" 16S rRNA sequence and free energy model 1.1. Figure 4.28 shows the arbitrary translation initiation scores generated for both expressed and assay negative genes (blue and red bars, respectively). There is no discernible or significant difference between the two sets of genes with a 2-tailed t-test of the difference between means giving a *p*-value of 0.46. The assay negative genes do tend to more consistently have higher TU values though with a range from 637 for *celA* to 10,435 for *xynG*, compared to a low of 390 for *afsB* and high of 50,417 for *xynH* (the next highest being 9536 for *bxyD*).

Figure 4.28 also depicts the results for the operon calculator. As the genes are in fact following the *lacZα* gene fragment, with no promoter of their own they are in an operon arrangement. Input is the promoter region (start of P_{lac} to end of the operator region), then all the DNA sequence immediately after up to 100 bp after the end of the *C. fimi* gene, and the same 16S rRNA sequence as above. A t-test again reveals no significant difference between the sets (*p*-value = 0.54).

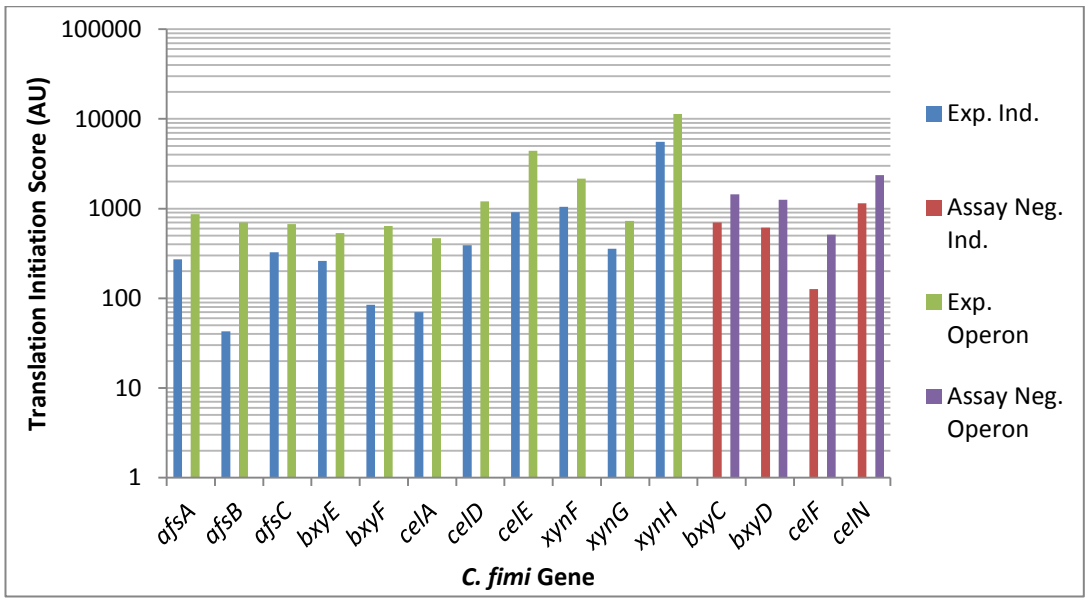


Figure 4.29. Translation initiation rates for *C. freundii* (16S rRNA sequence of AACCGCCCC) as calculated by Salis group RBS calculator (site). Genes *afsA-xynH* are assay positive (exp.) in *E. coli*, and *bxyC-celN* are assay negative (assay neg.). Translation initiation scores were generated using either the "reverse engineer RBS" tool (Ind.) or "predict operon expression" tool (operon)

The JCM (Japanese Collection of Microorganisms) has *C. freundii* NCIMB11490 as the same as JCM 1657, with the 16S rRNA sequence recorded with accession number AJ233408. *C. freundii* is not one of the organisms listed in the RBS calculator, however the instructions say to input the last 9 bases of the 16S rRNA gene when this is the case. The last nine bases of the gene AJ233408 are AACCGCCCC and so this sequence was used. Figure 4.29 shows the translation initiation values for the expressed (blue and green bars) and the assay negative (red and purple bars) with *C. freundii* as host, with the same parameters and sequences as described previously. And again we see little to no difference between the genes sets (t-test *p*-value = 0.7 and 0.48 for ind. and operon, respectively).

There is however a stark contrast in TU values for the same gene between species. For the "individual" TU values, genes in *E. coli* are predicted to have a 9-fold higher translation initiation rate than in *C. freundii* and for the operon scores that increases to 22-fold higher expression in *E. coli* compared to *C. freundii* (except for *afsA, afsB,* and *celA* which have fold differences of 15, 6, and 7, respectively). This indicates that when expression is induced the *E. coli* host cells are most likely under a greater metabolic burden than *C. freundii*, and this may be leading to toxicity for the cells

and potential selection for loss of plasmid or mutated genes/enzymes, which might account for some of the genes not being produced at a detectable level in *E. coli* while they are in *C. freundii*, but there is no obvious evidence of this (appendix figure A2).

Conclusion

From the genes I have cloned here, I have identified eleven novel proteins with a range of activities across different substrates. Of these there were three α-$_L$-arabinofuranosidases, three endoxylanases, a raft of multifunctional enzymes with β-xylosidase and cellobiohydrolase activities and one endoglucanase. The genes *bxyC*, *celA*, *celD*, *celE* and *xynG* were found to be either exclusively expressed in *Citrobacter freundii* NCIMB 11490 or their expression is so low in *Escherichia coli* JM109 that activity is undetectable. The BioBrick is demonstrated here as a useful tool for functional gene screening and that *C. freundii* is a potentially better heterologous host than *E. coli*, at least for *C. fimi* or high GC content genes. This could be due to an amalgamation of factors including the heavy bias of *C. fimi* towards GC rich codons placing a burden on dwindling tRNA pools in *E. coli* which are generally more abundant in *C. freundii*, and the predicted expression levels in the 2 differing hosts giving a lessened metabolic burden in *C. freundii* than *E. coli* and leading to higher levels of heterologous protein whose activity can be assayed.

The above may contribute to why genes have activity exclusively in *C. freundii*, and why activity levels are generally higher in this host also. However, environment can have a large effect on enzyme activity, as is explored in chapter 5. The temperature, pH, salt and metal ions found in the enzymes environment can all increase or decrease activity. As temperature and pH were (to the best of my knowledge) constant in all assays performed, these are unlikely to be potential sources of activity differences. However, as crude cell lysates were used for the assays performed in this chapter, salt and metal ion concentrations are likely to vary between the 2 host species. It may in fact be that as well as the potential genetic component favouring *C. freundii* expression, this host's cytoplasm has a more favourable salt and metal

composition, increasing the activity of the enzymes observed. To my knowledge no studies have been performed to determine the cytoplasmic composition of *C. freundii*, so this is merely supposition.

CHAPTER 5: PURIFICATION AND CHARACTERISATION OF *Cellulomonas fimi* PROTEINS

After expression in *Escherichia coli* and *Citrobacter freundii*, several of the genes showed activities of interest. These genes were selected for further study and characterisation. The genes were expressed at a higher cell culture volume and purified in a 2 step process to homogeneity or near homogeneity and assayed for their physical characteristics.

Putatively named arabinofuranosidase B (AfsB), β-xylosidase F (BxyF), and xylanases F and H (XynF/XynH) were assayed using nitrophenol-sugar derivatives. Their activities at varying pH and temperature levels were measured for each enzymatic substrate they appeared to hydrolyse. Once the temperature and pH optima were identified their activities in the presence of common metal ions and potential sugar inhibitors were tested for positive or negative effects. Surprisingly, for a mesophilic bacterium, the range of optimum pH and temperatures exhibited by this small selection of proteins was wide. Even for the same protein, on different substrates the results are very different.

One of the main tenets of synthetic biology is the rational design of non-naturally occurring gene networks. To be able to do this in a predictable and controlled way the networks need to be realistically mathematically modelled. Performed in this chapter are a series of assays to describe how the enzymes perform under different conditions. This is important as a complex network will require careful balancing of environmental factors to maximise enzyme activities, as not all enzymes work optimally under the same umbrella conditions. I have assayed for the major factors that are known to affect functionally related enzyme's performance (pH, temperature, metal ions and monomer sugars) and mirror characterisation studies that have been previously published on related enzymes, as for modelling purposes consistency between studies will ease this process. Determining the basic kinetics of the enzymes will also contribute towards this, and allow for easier comparison between

enzymes of the same function from different sources to select for those that perform the best.

In this chapter I aim to demonstrate that taking a genomics approach to gene and protein identification can yield new and industrially relevant proteins that would have been overlooked by previous methods when studying *Cellulomonas fimi*. I will show the main physical properties of the purified enzymes in their native, un-tagged state and discuss how they compare to the previously identified *Cellulomonas fimi* genes where appropriate, or to commercially available proteins where *C. fimi* exemplars are not available.

5.1 Purification

5.1.1 *Step 1*

E. coli was grown to a culture volume of 50 ml with induced expression of the proteins of interest. Cells were lysed by sonication in a Tris-HCl/glycerol buffer (pH 8) and passed through the first purification column for binding, washing and elution which was DEAE-sepharose. As an anion exchange resin was being used the protein of interest needed to be negatively charged for binding. For this to happen the buffer had to be of a pH at least 2 values above the protein pI value. The protein pI values were calculated using Compute pI/MW (web.expasy.org/compute_pi). As can be seen in table 5.1 a buffer of pH 8 should have been adequate for the binding of AfsB, BxyF and XynH but not XynF. Being close to neutral, pH 8 seemed like the best pH to allow binding to occur, without damage to the protein which may occur at higher pH levels. This was tested on a small scale to confirm and was found to be the case. XynF was the only protein not to bind to DEAE-sepharose at pH 8. Binding for this particular protein only occurred at pH 11 and above, in keeping with its predicted pI value.

Protein	Amino Acids	MW (kDa)	pI
AfsB	502	55.191	5.01
BxyF	896	98.304	4.98
XynF	464	49.412	8.97
XynH	558	60.577	5.38

Table 5.1. Predicted molecular weights and pIs for the cloned *C. fimi* proteins based on predicted amino acid sequences using *Compute pI/MW* (web.expasy.org/compute_pi)

Therefore for XynF the strategy was reversed. The cell lysate containing the soluble protein was still buffered at pH 8 to bind unwanted *E. coli* proteins but leave the XynF in the flow through with a far reduced complexity of protein. This flow through was then used for the second stage of purification.

Glycerol was used in the buffers as a preservative to protect against damage from freeze/thaw cycles. All fractions and eluted proteins were stored at -20ºC until no longer needed.

5.1.2 *Step 2*

For AfsB, BxyF and XynH the second purification step was by hydrophobic interaction using phenyl-agarose resin. As this relies on a high salt content the enzyme active fractions were made up to 1 M NaCl to match that of the equilibrating buffer. If activity was present in more than one fraction they were combined first then made up to 1 M NaCl. Elutions were performed in a stepwise manner with multiple 1 ml volumes (0.5 CV) per elution concentration to maximise concentration of active enzyme recovered.

XynF was purified secondly on CM-sepharose, a cation exchange resin. As this works in the opposite fashion to the anion exchange resin, then in this instance XynF would bind to the resin and could be eluted to maximise concentration and minimise contamination from other proteins. Again this was done in a stepwise manner with multiple volumes of 1 ml per elution concentration.

5.2 Purity

The purification tables for each protein are given in tables 5.2-5.5. They describe the estimated protein concentrations of each fraction and the enzyme activities exhibited in relation to the original un-purified cell lysate. The concentration of nitrophenol released from the appropriate substrate was calculated from standard curves of 2-nitrophenol for ONPC and ONPX substrates or 4-nitrophenol for PNPA substrate at pH 5.5 and read at 405 nm using a NanoDrop 2000. Where enzyme activity was seen in multiple fractions at the end of the DEAE step, the purification and activity of each fraction is shown individually prior to combination for the second step. Where multiple fractions contained the active enzyme after the second purification step, only the one used for characterisation is shown. This was the fraction that exhibited the highest activity:protein ratio.

Fraction	Volume (ml)	Total Protein (mg)	Activity Units (Units/min)	Total Activity (Ac. Units*ml)	Specific Activity (Tot. Act/mg)	Fold Purification	Yield (%)
Cell Lysate	1	20	0.89	35.65	1.78	1.00	100.00
DEAE 0.3 M (made to 1 M)	4	2	0.33	13.20	26.39	14.81	37.02
HIC Wash 1	2	0.335	0.17	6.65	39.70	22.27	18.65
HIC 0.5 M 3	1	0.335	0.14	5.48	16.37	9.18	15.38
HIC 0.1 M 1	1	0.26	0.11	4.44	17.08	9.58	12.46

Table 5.2. AfsB Purification

Fraction	Volume (ml)	Total Protein (mg)	Activity Units (Units/min)	Total Activity (Ac. Units*ml)	Specific Activity (Tot. Act/mg)	Fold Purification	Yield (%)
Cell Lysate	1	15	5.18	207.28	13.82	1.00	100.00
DEAE 0.4M	6	7.5	0.70	28.17	22.54	1.63	13.59
DEAE 0.4 M (made to 1 M)	6	7.5	1.19	47.63	38.10	2.76	22.98
HIC 20 mM 3	1	0.075	0.31	12.58	167.70	12.14	6.07

Table 5.3. BxyF Purification

Fraction	Volume (ml)	Total Protein (mg)	Activity Units (Units/min)	Total Activity (Ac. Units*ml)	Specific Activity (Tot. Act/mg)	Fold Purification	Yield (%)
Cell Lysate	1	20	1.35	53.89	2.69	1.00	100.00
DEAE FT	4	3	0.18	7.32	9.76	3.62	13.59
DEAE Wash 1	4	0.24	0.07	2.74	45.65	16.94	5.08
CM 0.5 M	1	0.04	0.27	10.79	269.72	100.10	20.02

Table 5. 4. XynF Purification

Fraction	Volume (ml)	Total Protein (mg)	Activity Units (Units/min)	Total Activity (Ac. Units*ml)	Specific Activity (Tot. Act/mg)	Fold Purification	Yield (%)
Cell Lysate	1	13	0.25	10.06	0.77	1.00	100.00
DEAE 0.3 M	4	5	0.04	1.79	1.43	1.85	17.78
DEAE 0.5 M	4	1	0.04	1.51	6.04	7.80	15.00
DEAE 0.5+0.3 (made to 1 M)	8	2.6	0.04	1.43	4.39	5.67	14.17
HIC 0.1 M	2	0.15	0.06	2.38	31.68	40.93	23.61

Table 5.5. XynH Purification

One unit (U) is defined as the amount of enzyme that releases 1 μmol of nitrophenol per min. From standard curves of 2-nitrophenol (ONP) and 4-nitrophenol (PNP) the molar absorbance coefficients (ε) were estimated to be 1192.9 $M^{-1}cm^{-1}$ and 9954.4 $M^{-1}cm^{-1}$, respectively. Using these coefficients the $M^{-1}cm^{-1}$ concentrations of nitrophenol released were estimated by $Au_{(405\ nm)}/(\varepsilon*0.1(\text{pathway length}))$ and to get the mM concentration, and thus the μmol of nitrophenol released, this was multiplied by 1000. The μmol was divided by the minutes the assay ran, and multiplied by 5, the dilution factor of the enzyme in the assay volume, and multiplied by 40 to get the equivalent units in 1 ml (25 μl reaction volume), to give total activity (U/ml). Total activity was multiplied by the total ml volume of the fraction and divided by total protein present in mg to give the specific activity.

The proteins were first tested for binding and elution parameters using DEAE-sepharose on a smaller scale (culture volumes of 20 ml) before scaling up to 50 ml culture volumes. This was to ensure that binding would occur at the given pH and that the elution fraction containing the protein could be identified to maximise yield but minimise volume. This strategy worked reasonably well for AfsB and BxyF. AfsB cleanly eluted in the 0.3 M NaCl fraction only, but BxyF was observed to elute in 0.3 M and 0.4 M fractions. Due to this observation when scaled up the 0.3 M fraction was skipped and the 0.4 M fraction volume increased to 6 ml, which subsequently contained the vast majority of BxyF. XynH on the other hand once scaled up, did not exhibit exactly the same behaviour as previously observed. It ought to have been entirely eluted in the 0.3 M fraction. Once scaled up, a 6 ml volume was again used to elute the protein. The 0.4 M step was omitted as the remainder of the protein could all be eluted in the 0.5 M fraction without a need for an extra separation step between. However, a larger degree of activity was actually observed in the 0.5 M fraction where none should have been apparent from previous experience.

Despite these anomalies the vast majority of the desired proteins were bound and eluted in simplified protein mixtures with little loss to yield. When the fractions were made to 1 M NaCl for binding to the HIC resin, percent yields were actually observed to increase for AfsB and BxyF. The only parameter that may account for this is that NaCl has a positive effect on the activity of AfsB and BxyF. To assess the effect of NaCl on enzyme activity, cell lysates of the enzymes were assayed with and without the addition of 1 M NaCl. The results are given in table 5.6 and are averages of duplicate measurements.

Protein	Mean Specific Activity in 0 M NaCl	Mean Specific Activity in 1 M NaCl	Fold Difference	T-Test p value
AfsB	2.94	3.11	1.06	0.04
BxyF	6.54	8.25	1.26	0.10
XynF	2.92	3.28	1.12	0.06
XynH	1.13	1.18	1.04	0.48

Table 5. 6. Effect of 1 M NaCl on the ability of *C. fimi* proteins to release nitrophenol from sugar substrates in crude *E. coli* cell lysates. Mean of 2 replicates, t-test, 2 tailed, independent

From the data in table 5.6 it would appear that NaCl has little to no effect with regard to enzyme function. The highest difference is with BxyF where the nitrophenol released in the presence of 1 M NaCl increases by only 1.3 fold compared to 0 M NaCl. Only AfsB shows a significant difference at the 5% confidence level between the 2 environments, however a strong change was observed for the other proteins during the purifications. This may be because the large volume of other proteins is effectively diluting the heterologous enzymes, decreasing the chances of ions interacting with our protein of interest. After the first purification step the NaCl concentration is increased to 1 M in a far simplified protein mixture. There are fewer proteins to interact with the sodium ions, or interfere with the ions interacting with AfsB, thereby leading to more interaction with AfsB and the sodium resulting in a far larger observable effect.

The two step process used here has given fold purifications for AfsB, BxyF, XynF and XynH of 9.6, 12.1, 100.1 and 40.9, respectively. Due to the possible effect of NaCl the true fold purification for these enzymes could in fact be lower, but are likely close to the truth. By looking at the SDS gels it also becomes clear that the vast majority of contaminating proteins have been removed in just these 2 steps. For the assays that followed this was a good enough purity, as *E. coli* lacks the natural ability to interact with the substrates. The main problem with the purifications appeared to be with the protein's binding affinity to phenyl-agarose. XynF bound strongly and eluted cleanly from the CM-sepharose column resulting in a 100 fold purification and a single band visible by SDS-PAGE.

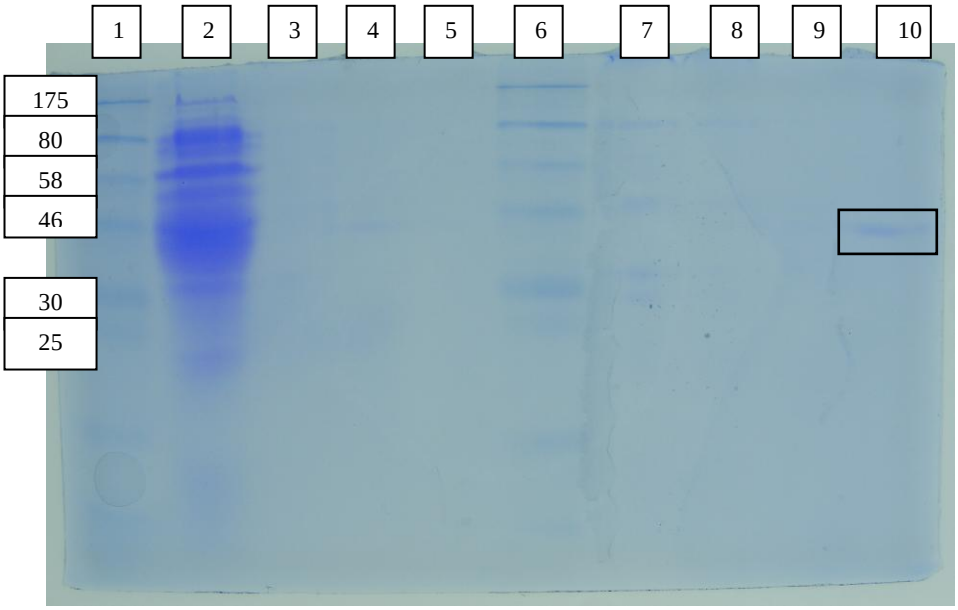


Figure 5.1. SDS-PAGE image of XynF fractions after purification, coomassie stain. Lanes 1 and 6 – Pre-stained protein marker; lane 2 – XynF cell lysate; lane 3 – CM1 DEAE_FT FT; lane 4 – CM1 DEAE_Wash 1 FT; lane 5 – CM1 Wash 1; lane 7 – CM2 DEAE_FT_FT FT; lane 8 – CM2 DEAE_Wash 1_FT FT; lane 9 – CM2 CM1_Wash 1 FT; lane 10 – CM2 0.05 M NaCl elution. Lane numbers are above the gel, marker MW in kDa is displayed beside the gel. Silver staining revealed no extra bands and obscured what was present.

As can be seen from figure 5.1, even just after the DEAE column, the sample complexity is hugely reduced but also very dilute. The first CM-sepharose column packed failed to bind any protein (CM1). This was repacked (CM2) and the enzyme active flow throughs (FTs) were passed through the new column. Binding occurred and XynF was eluted in a single fraction, figure 5.1 lane 10, with the only visible protein boxed. The visible band is estimated at 42 kDa. The predicted full protein is estimated computationally at 49 kDa, but minus the predicted TAT signal peptide this is closer to 46 kDa. The gel based estimate is slightly lower than would be expected, but it seems unlikely to me to be anything else, given its intensity and the fold purification observed.

XynH bound reasonably well to the phenyl-agarose and eluted fairly cleanly at 0.1 M NaCl, but activity was still observed in the succeeding fractions for 3 times 80 mM, and even in the first 60 mM NaCl elution fraction. The protein being spread across several fractions in this fashion would account for the lower fold purification observed. SDS-PAGE analysis reveals several co-purified contaminating proteins,

but large bands of the approximate molecular weight expected for XynH can be observed (figure 5.2).

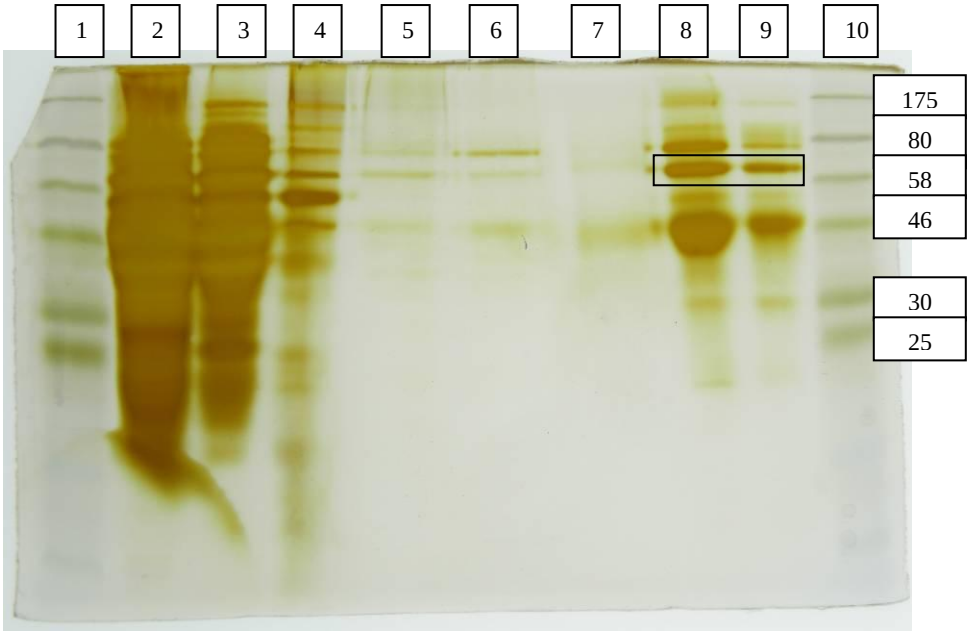


Figure 5.2. Silver stained SDS-PAGE image of XynH fractions after purification, silver stain. Lanes 1 and 10 – Pre-stained protein marker; lane 2 – XynH cell lysate; lane 3 – DEAE 0.3 M NaCl elution; lane 4 – DEAE 0.5 M NaCl elution; lane 5 – HIC FT; lane 6 – HIC 0.5 M NaCl elution; lane 7 – HIC 0.1 M NaCl elution; lane 8 – HIC 0.08 M NaCl elution 1; lane 9 – HIC 0.08 M NaCl elution 2

The predicted molecular weight of XynH is 60 kDa. XynH has no predicted signal sequence so this is probably the size we are truly looking for. The bands of the gel corresponding to that size are highlighted across the 3 lanes they are visible in, in figure 5.2. The SDS-PAGE estimated sizes for the bands are 64.5 kDa, slightly larger than expected.

For AfsB and BxyF however, phenyl-agarose proved to be a less suitable resin. Binding to the resin was still observed for these two proteins, the problem seemingly being how well they bound. BxyF bound tightly to the resin and next to none was lost in the wash steps. Eluting the protein became the problem as BxyF did not elute cleanly. Activity was detected in every elution fraction from 0.1 M NaCl all the way through to the final 0 M NaCl elution. The activity recorded was of a similar level for each of these 16 fractions, a sum total of 17 ml elution volume.

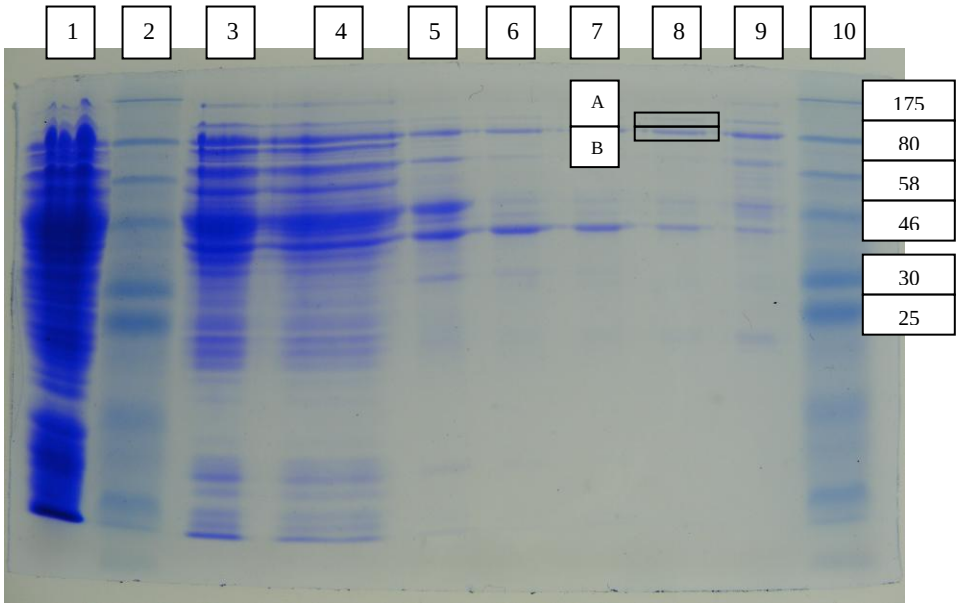


Figure 5.3. SDS-PAGE image of BxyF fractions after purification, coomassie stain. Lanes 2 and 10 – Pre-stained protein marker; lane 1 – BxyF cell lysate; lane 3 – DEAE 0.4 M NaCl elution; lane 4 – DEAE 0.4 M NaCl elution adjusted to 1 M NaCl; lane 5 – HIC 0.08 M NaCl elution 3; lane 6 – HIC 0.06 M NaCl elution 3; lane 7 – HIC 0.04 M NaCl elution 3; lane 8 – HIC 0.02 M NaCl elution 3; lane 9 – HIC 0 M NaCl elution 3

The calculated size of BxyF is 98 kDa. Therefore bands A or B in figure 5.3 are the most likely to be BxyF. There is no predicted signal peptide so this size should be close to the observed. Band A is estimated at 100 kDa and band B as 79 kDa. The other bands from the co-purified proteins are too small to be BxyF, and the HIC 0.02 M NaCl elution 3 fraction appears to be one of the cleaner elutes with seemingly fainter and fewer contaminating proteins visible.

AfsB on the other hand appeared to bind quite poorly to the resin with activity present in the flow through and subsequent wash steps. The highest activity observed being that of the first wash with a specific activity of 39.7 U/mg nitrophenol being released. Some binding must have occurred though for an increase in enzyme activity between the final wash and those enzyme activities recorded for the third 0.5 M NaCl and the first 0.1 M NaCl elutions was clear to see. Activity was still detectable in the fractions up to the second 0.08 M NaCl elution.

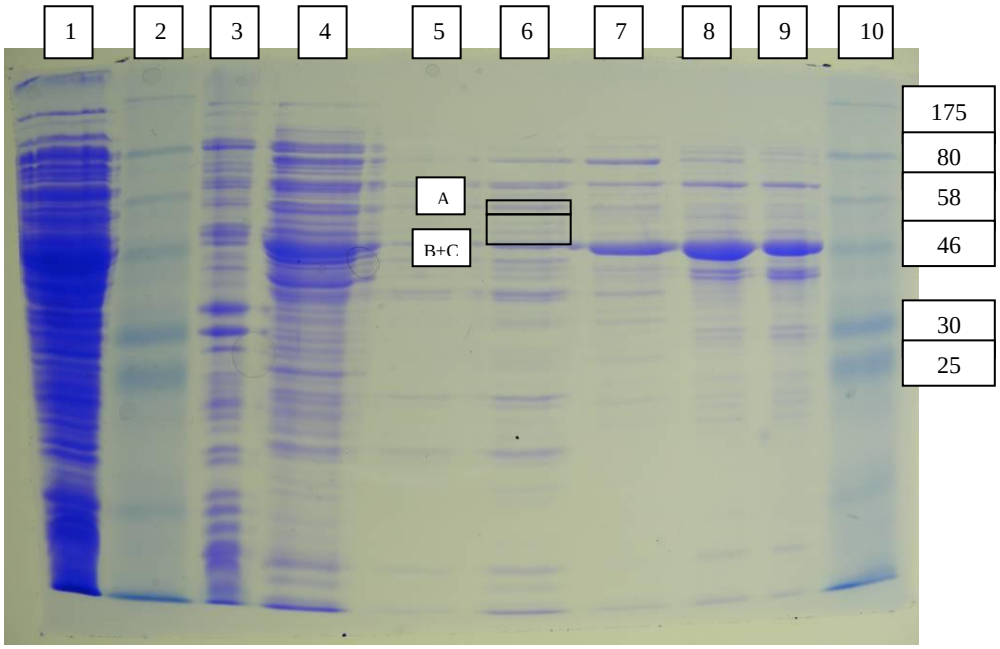


Figure 5.4. SDS-PAGE image of AfsB fractions after purification, coomassie stain. Lane 2 and 10 – Pre-stained protein marker; lane 1 – AfsB cell lysate; lane 3 – DEAE FT; lane 4 – DEAE 0.3 M elution; lane 5 – HIC FT; lane 6 – HIC wash 1; lane 7 – HIC 0.5 M elution 3; lane 8 – HIC 0.1 M elution 3; lane 9 – HIC 0.08 M elution 1

The elution of AfsB is relatively messy when compared to the gels of the other proteins as seen in figure 5.4. What we do see however is that the complexity of the elutions are still greatly reduced. The HIC wash 1 fraction was run in lane 6 of figure 5.4. This fraction had the highest fold purification of AfsB and the most activity. The number of bands in the lane makes it difficult to identify fully which band may be AfsB. There are three main possibilities that lie between the 58 and 46 kDa markers which are highlighted. AfsB is approximated at 55 kDa based on amino acid sequence and has no predicted signal sequence. Band A is estimated at 55 kDa, band B at 51 kDa and band C as 49 kDa. This suggests that band A is that of AfsB and, due to its intensity, that there are other proteins more prominent in the fraction.

In the cases of both BxyF and AfsB, although for seemingly different reasons, the target protein has been spread across a number of fractions accounting for the poor fold recovery after purification (tables 5.2 and 5.3). Regardless of this the proteins recovered are of a relatively simplified mix and have a high enough yield for characterisation experiments.

Despite the contaminating proteins from *E. coli*, there are none within the bacterium that can account for the activities assayed for with the nitrophenol substrates. Therefore the characterisations that follow are of the *Cellulomonas fimi* genes that are being expressed. The presence of proteins can have a stabilising effect on other proteins. The co-purified proteins could therefore make clearer the physical properties observed rather than obfuscate the protein's nature.

5.3 Characterisation

5.3.1 *pH Optima of the Purified Proteins*

The proteins were assayed under standard conditions as previously reported. The only difference being the reaction buffer used was of increasing pH. For the assays fifteen µl of buffer was dispensed to a thin-wall PCR tube before 5 µl of appropriate substrate was added. Just prior to incubation five µl of purified protein was added to the tube and the reaction mixture mixed well before being incubated in a DNA Engine PCR machine for 10 minutes at 37ºC. The reactions were stopped by placing the samples on ice. One volume (25 µl) of 1 M Na_2CO_3 was added to a tube and mixed thoroughly and then immediately the absorbance at 405 nm was measured using a NanoDrop 2000. Each pH point is the average of duplicate runs. The following buffers were used:

- pH 4 – 5.5, 50 mM sodium acetate
- pH 6 – 7.5, 50 mM potassium phosphate
- pH 8 – 9, 20 mM Tris-HCl
- pH 9.5 – 10.5, 50 mM glycine

The data is plotted as a percentage (1^{st} y-axis) of the maximal specific activity (point labels, U/mg) seen based on absorbance at 405 nm adjusted for background which is also shown (2^{nd} y-axis). The standard error as a percentage of activity is shown for each data point. All incubations were for 10 minutes.

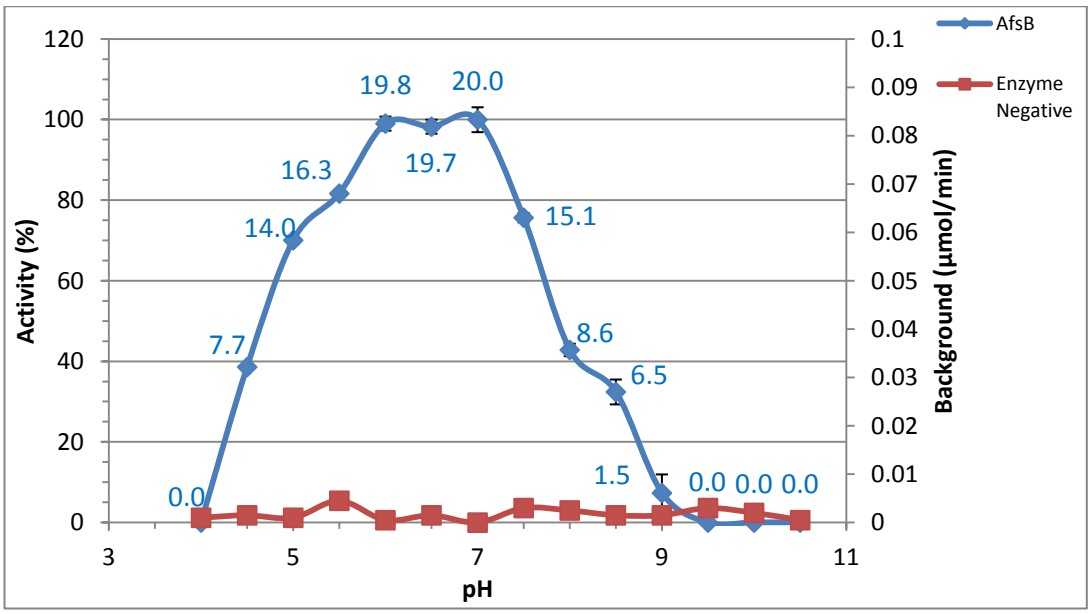


Figure 5.5. Effect of increasing pH on AfsB activity (percentage maximal observed) towards PNPA, adjusted for background. Floating numbers are specific activity of AfsB at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error as percentage of activity shown as bars.

Given that the genes are all from the same organism, they exhibit activity at a range of pH values from as low as 4 for XynF with 2-nitrophenyl β-$_D$-cellobioside (ONPC) to pH 9 for XynH on 2-nitrophenyl β-$_D$-xylopyranoside (ONPX). Neutral pH values however appear to be the most effective. AfsB exhibits maximal activity between pH 6 and 7 (figure 5.5) with little to no difference in activity between pH 6, 6.5 or 7. Either side of that optimal range however, and its activity quickly dies off. At just pH 5.5 or 7.5, activity is reduced by 15% and 22% respectively. An up or down change of pH of 0.5, around a five-fold decrease or increase in H^+ ions, respectively, and the activity drops to 70% and 45%, respectively. It would appear that AfsB is more tolerant of moderately acidic conditions than of moderately alkaline. When getting closer to the extremes of pH there is little difference between activities in acidic or alkaline environments. All activity is essentially non-existent at pH 4, and presumably below although this was not tested, and at pH 9 and above, with AfsB exhibiting just 8% activity at pH 9. As such the optimum pH for AfsB was determined to be pH 6.5 so as to be well protected from slight changes in pH in subsequent experiments.

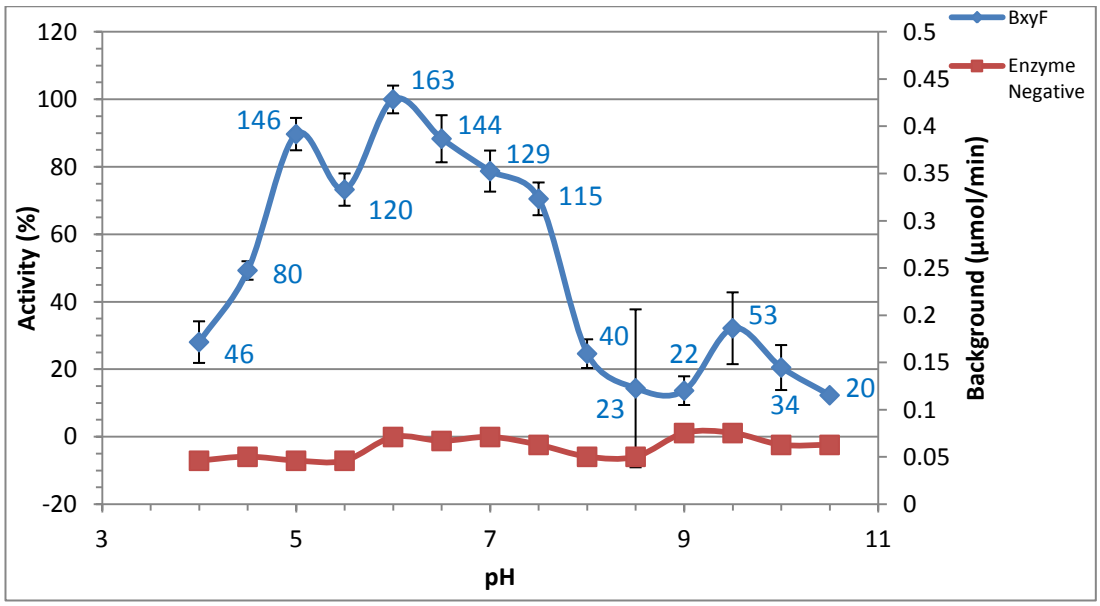
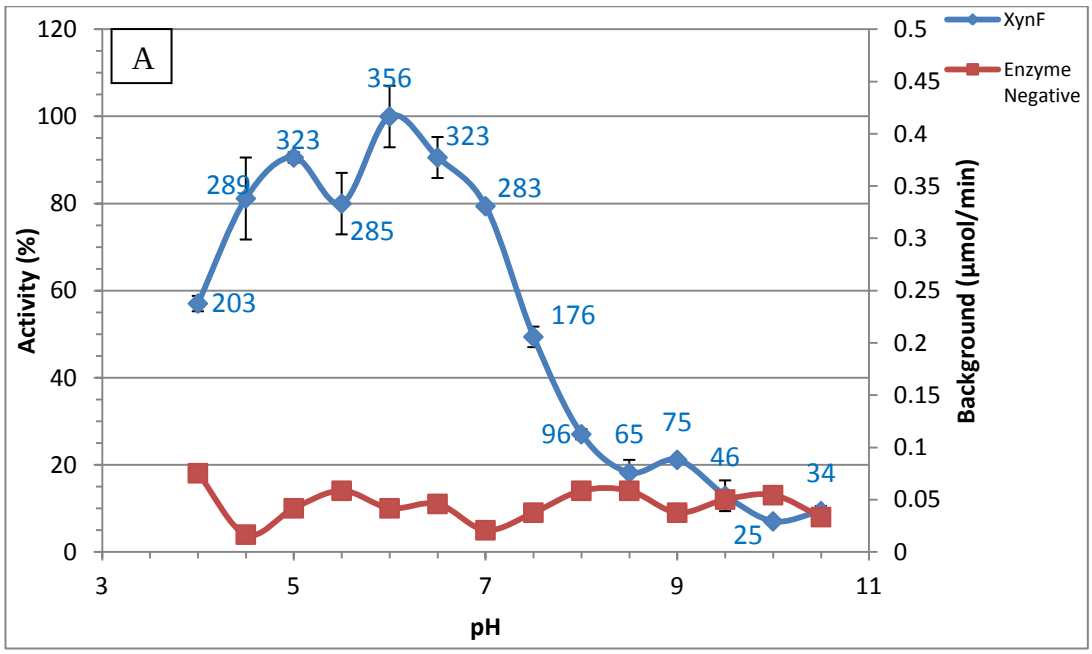
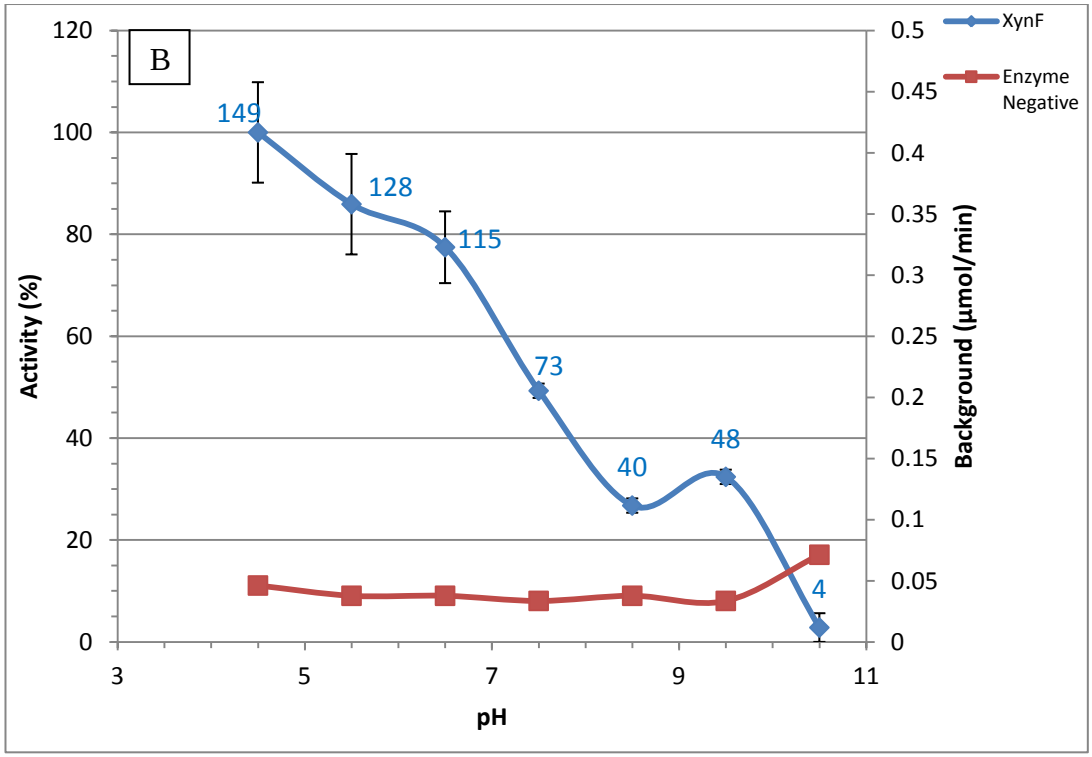


Figure 5.6. Effect of increasing pH on BxyF activity (percentage maximal observed) towards ONPX, adjusted for background. Floating numbers are specific activity of BxyF at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error as percentage of activity shown as bars.

The tolerance shown by AfsB over pH 6-7 does set it slightly apart from the other proteins assayed. For the remaining three proteins, distinct peaks in activity were observed rather than the plateau as in AfsB. With ONPC as substrate BxyF and XynF have distinct activity peaks at pH 6.0 (figures 5.6 and 5.7 A, respectively). XynF activity quickly drops off as pH increases, with 79% activity at pH 7, 27% at pH 8 and 21% at pH 9 and a gradual tailing off after that to below 10% at pH 10.5. BxyF however is more stable in mildly alkaline conditions dropping to 71% activity at pH 7.5, but with a rapid decline thereafter to a low of 13% at pH 9, before a mild resurgence at pH 9.5 increasing to 32%.

At lower pH values the two enzymes have similar responses. BxyF (figure 5.6) steadily increases in activity from 28% at pH 4 to 90% at pH 5, but then a sudden decline is seen at pH 5.5, dropping activity to 73% before recovery at pH 6.0. XynF (figure 5.7 A) also exhibits an increase from 57% at pH 4.0 to 91% at pH 5.0, before a decline at pH 5.5 to 80% and recovery to 100% at pH 6.0. It therefore appears that XynF on ONPC is more stable at lower pH values than BxyF but again, like AfsB, they work optimally at near neutral pH.


A
XynF
Enzyme Negative
Activity (%)
Background (µmol/min)
pH
120
100
80
60
40
20
0
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
3
5
7
9
11
203
289
323
285
356
323
283
176
96
65
75
46
25
34


B
XynF
Enzyme Negative
Activity (%)
Background (µmol/min)
pH
120
100
80
60
40
20
0
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
3
5
7
9
11
149
128
115
73
40
48
4

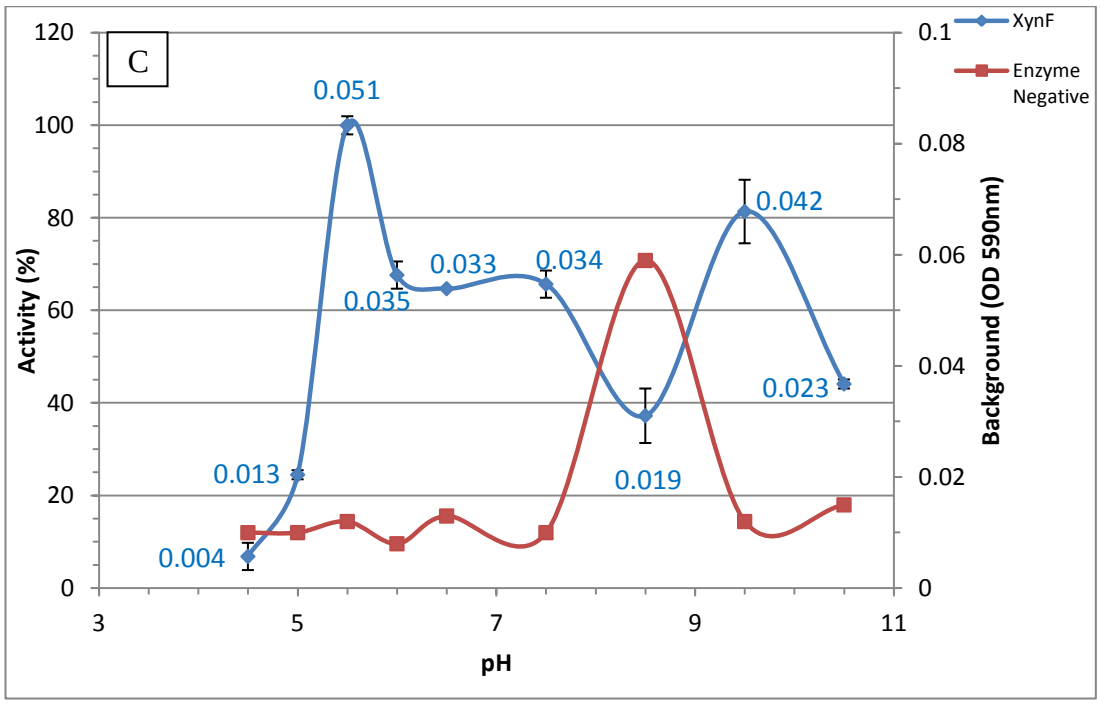


Figure 5.7. Plots showing the effect increasing pH has on purified XynF as a percentage of the maximal activity observed on A) ONPC, B) ONPX, and C) RBB-xylan. Floating numbers are specific activity of XynF at each point (A and B), or OD minus enzyme negative (C). The red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis, A and B), or OD of enzyme negative assays (C). n = 2 for activity screens, +/- standard error as percentage of maximal activity shown as bars.

XynF when assayed on different substrates shows different optimal pH conditions (figure 5.7). Optimum pH is very low at pH 4.5 with ONPX as substrate (figure 5.7 B), if not lower, and as pH increases the activity rapidly decreases. Activity at pH 6.5 is only 77% of the maximum observed, dropping to a low of 3% at pH 10.5. When RBB-Xylan is used as the substrate (figure 5.7 C) this optimum changes again to pH 5.5, with another strong peak at pH 9.5. Unlike on the other two substrates the activity of XynF remains reasonable even in extreme alkaline conditions with >40% activity observable at pH 10.5.

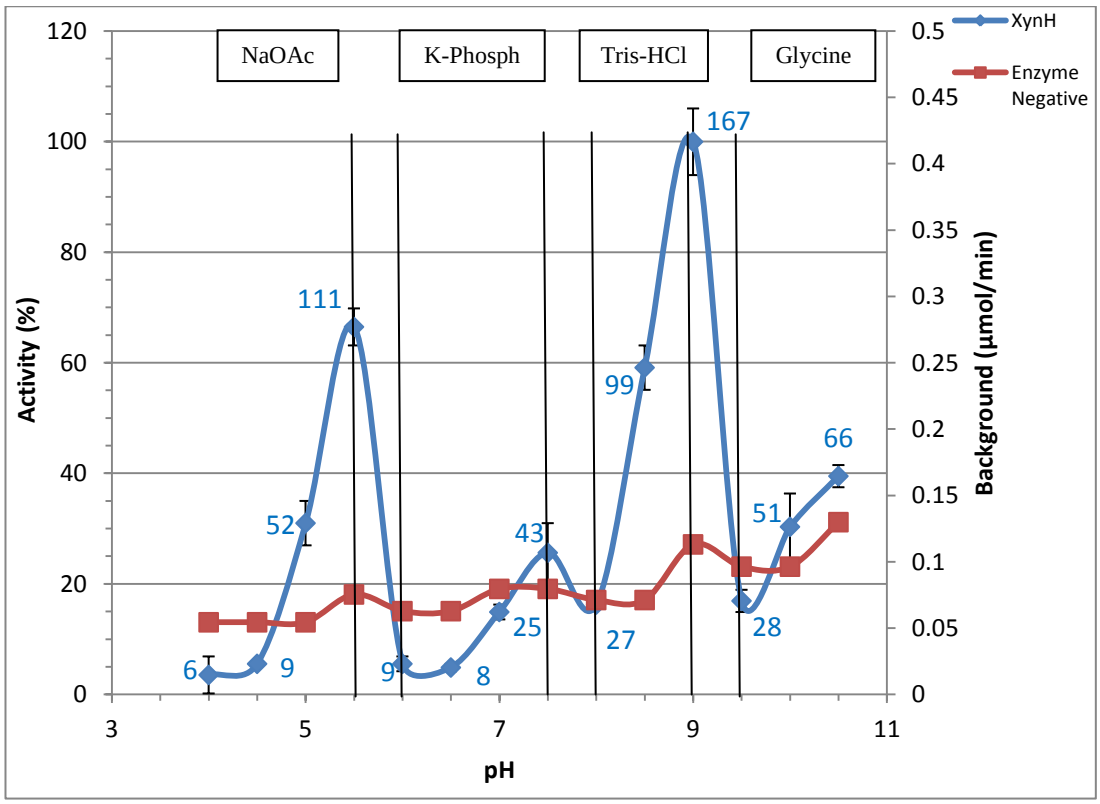


Figure 5.8. A plot showing the effect increasing pH has on purified XynH as a percentage of the maximal activity observed on ONPX. Floating numbers are specific activity of XynH at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error as percentage of maximal activity shown as bars. Buffer ranges are indicated by black lines and labels

XynH exhibits a very different pH profile from the other proteins (figure 5.8). Here we see a possible double peak of activity. The first activity peak is seen at pH 5.5 where XynH has 67% of maximal activity. Either side of this point and the activity drops off dramatically to 31% and 6% at pH 5 and pH 6, respectively. For neutral pH, where optimal activity is seen for AfsB, BxyF and XynF (ONPC), there is a very clear lull in activity with the high between pH 6 and pH 8 being only 25%. The main peak of activity, its apparent optimum, is at pH 9, a point at which the other proteins have lost the majority of their activities. There is no real tolerance shown for variability in pH for XynH at either activity peak. pH 8.5 and pH 9.5 give activities of 59% and 17%, respectively. However, it seems as though alkaline conditions are preferable to acidic. At pH 4, the lowest pH tested, activity was at 4%, whereas at pH 10.5, the highest pH tested, the activity was beginning to climb back up and reached 40%.

For each buffer used, as the pH increases to the max of its buffering range, the activity of XynH increases, before falling again when the next buffer is introduced. The change in activity of XynH is most marked for sodium acetate and Tris-HCl buffers, and more muted when potassium phosphate and glycine buffers are employed (figure 5.8). It would appear that the buffers used have a large and varying effect on the activity of XynH.

The buffer effects are only seen for XynH despite the same substrate used for other enzymes tested and the fact that BxyF is also a GH39 enzyme. The effects observed are therefore unlikely to be related to buffer effects on the substrate (such as altering of charge state or binding of ions released from the buffer interfering with hydrolysis) as effects would also have be seen for BxyF to some extent. Nor is it probably due to the effect of buffer on the catalytic residue of GH39 enzymes which by virtue of their family should be the same – glutamic acid [189], otherwise again it would also be likely some effect would be observed for BxyF. It would appear that XynH itself, its structure, may be being affected in different ways by the different buffers used producing conformational instability [190].

5.3.2 *Temperature Optima of the Purified Enzymes*

Using the pH optima derived above I assayed for the temperature optimum for each protein. Reaction volumes were the same as before. The buffer remained the same per incubation; the temperature was changed using 5ºC intervals between 20ºC and 70ºC. The incubations were performed in a DNA Engine PCR machine. After each incubation the reaction was stopped by placing the samples on ice. One volume of 1 M Na_2CO_3 was added to the sample, mixed thoroughly and the absorbance at 405 nm was immediately measured using a NanoDrop 2000. Each data point is the average of duplicate runs and plotted as the percentage of maximal activity observed based on absorbance. The standard error between replicates is displayed as a percentage of maximal activity as bars at each point. All incubations were for 10 minutes.

The figures 5.9-5.12 show the effect temperature has on each of the proteins. The optimal temperatures occur over a range of temperatures, from as low as 40ºC up to

80ºC. These results seem at odds with the bacterium from which they were cloned, being of the mesophilic range for growth with an optimal temp of 30ºC.

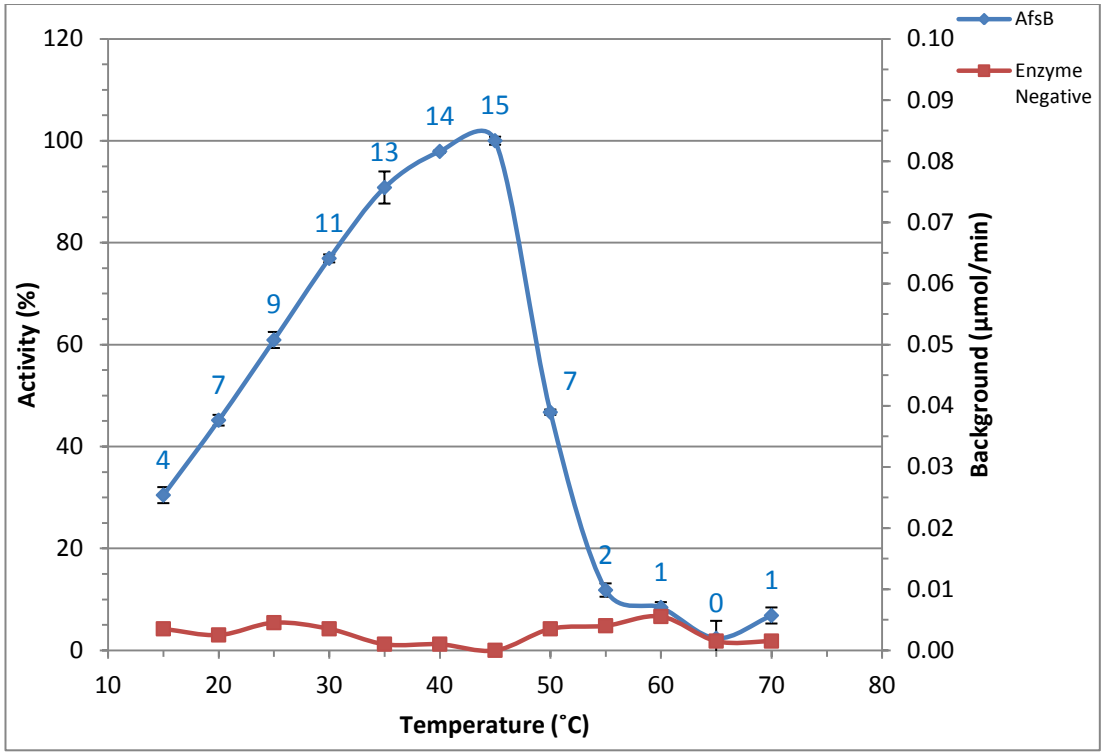


Figure 5.9. Effect of increasing temperature on AfsB activity (percentage maximal observed) towards PNPA at pH 6.5, adjusted for background. Floating numbers are specific activity of AfsB at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error, as percentage of activity, shown as bars.

AfsB has an optimum temperature of 45ºC when assayed at pH 6.0 (figure 5.9). Just below this at 40ºC AfsB exhibits 98% activity however, so either could be considered as optimal. Up to this point there is a steady, almost linear, increase in activity from 30% at 15ºC to 91% at 35ºC just short of the optimal 5 degree range. This suggests that at even relatively low temperatures AfsB is still able to function to a reasonable degree. Indeed, activity is still detectable when incubated at 4ºC (data not shown). Conversely however, activity drops drastically after 45ºC, to a mere 47% activity at 50ºC, then down to 12% of its maximum activity at 55ºC. The ability to release nitrophenol from its sugar substrate is all but wiped out above this point with a low of 2% activity at 65ºC. This suggests that denaturation is occurring at temperatures above 45ºC, and that the protein itself is not of a heat-tolerant nature, working primarily within the mesophilic range.

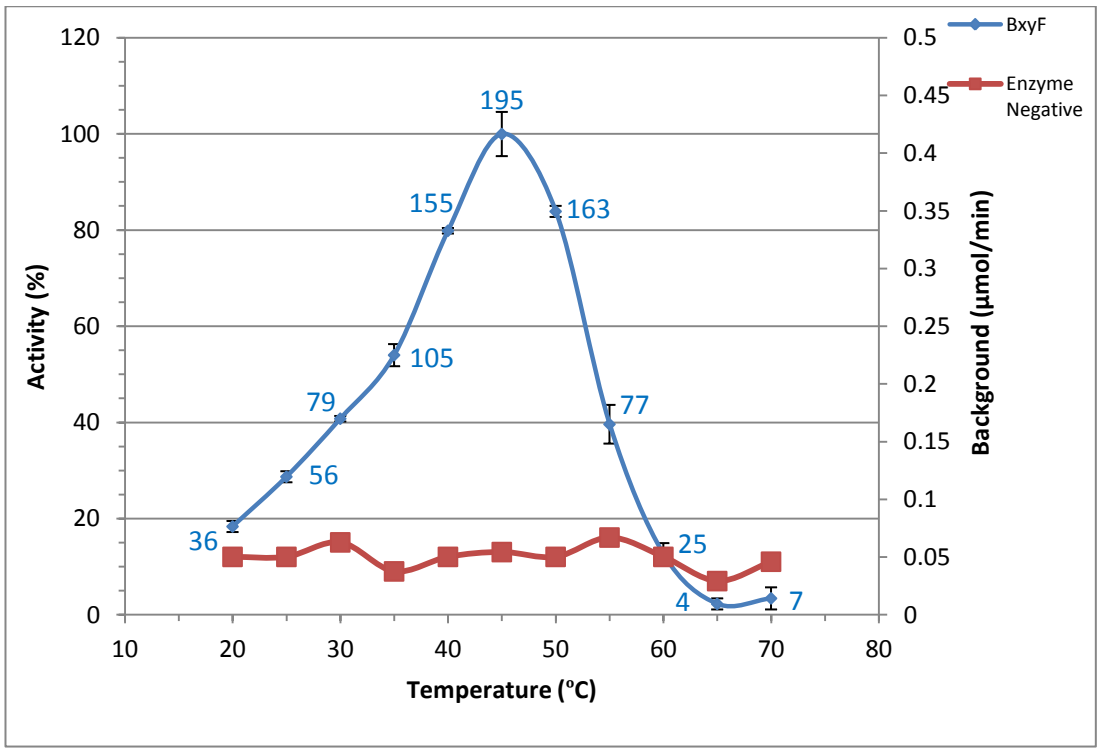
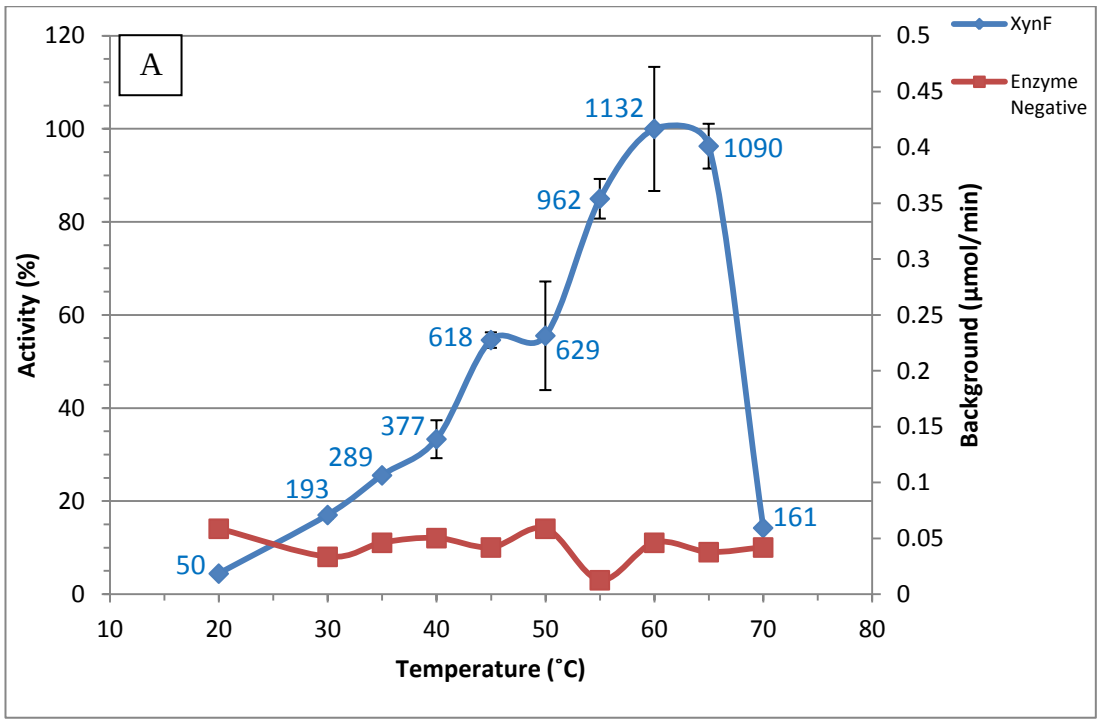
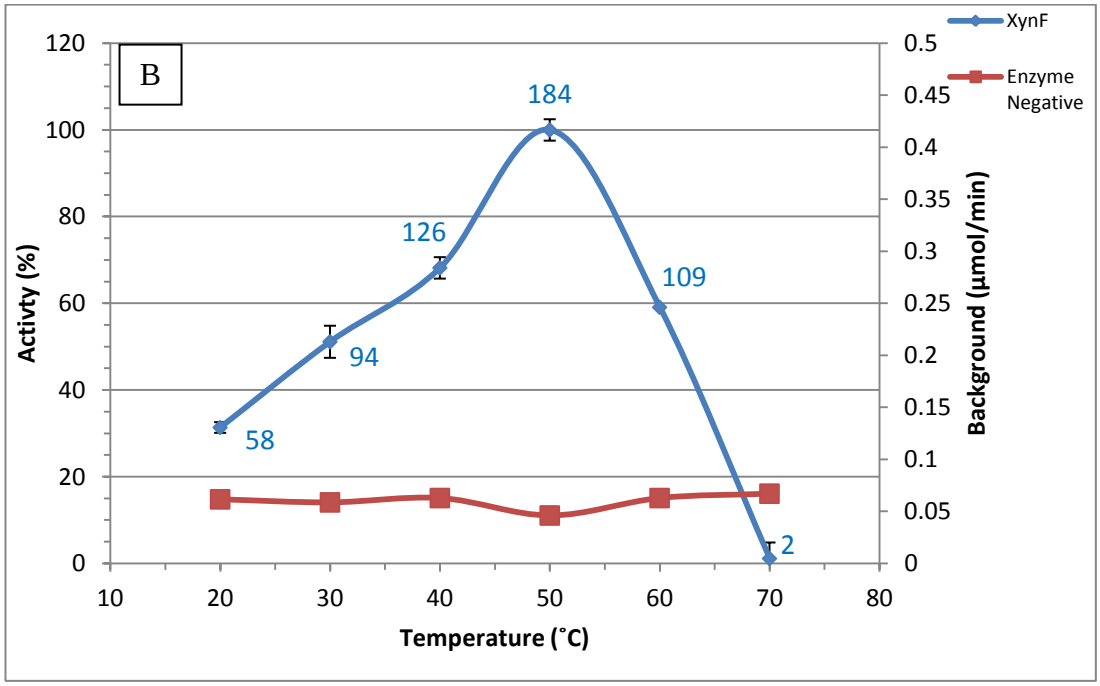


Figure 5.10. Effect of increasing temperature on BxyF activity (percentage maximal observed) towards ONPX at pH 6, adjusted for background. Floating numbers are specific activity of BxyF at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error, as percentage of activity, shown as bars.

BxyF exhibits a similar style plot (figure 5.10). At 20ºC the measured activity is 18% of the maximum detected. This activity then steadily increases to 54% at 35ºC before reaching its peak activity at 45ºC. As for AfsB, activity is seemingly restricted to the mesophilic range, but the BxyF protein has a greater tolerance of super-optimal temperatures. For AfsB in a 5 degree step up from optimal temperature half of the total activity was lost. For BxyF however, at 50ºC 83% activity is still detectable. A higher activity level than is measured at 40ºC which was 80%. Even at 55ºC there is still 40% activity. Nearly all activity is lost however once above 60ºC where activity drops to a low of 2% at 65ºC suggesting denaturation.


A
XynF
Enzyme Negative
Activity (%)
Background (µmol/min)
Temperature (˚C)
1132
1090
962
618
629
377
289
193
161
50


B
XynF
Enzyme Negative
Activty (%)
Background (µmol/min)
Temperature (˚C)
184
126
109
94
58
2

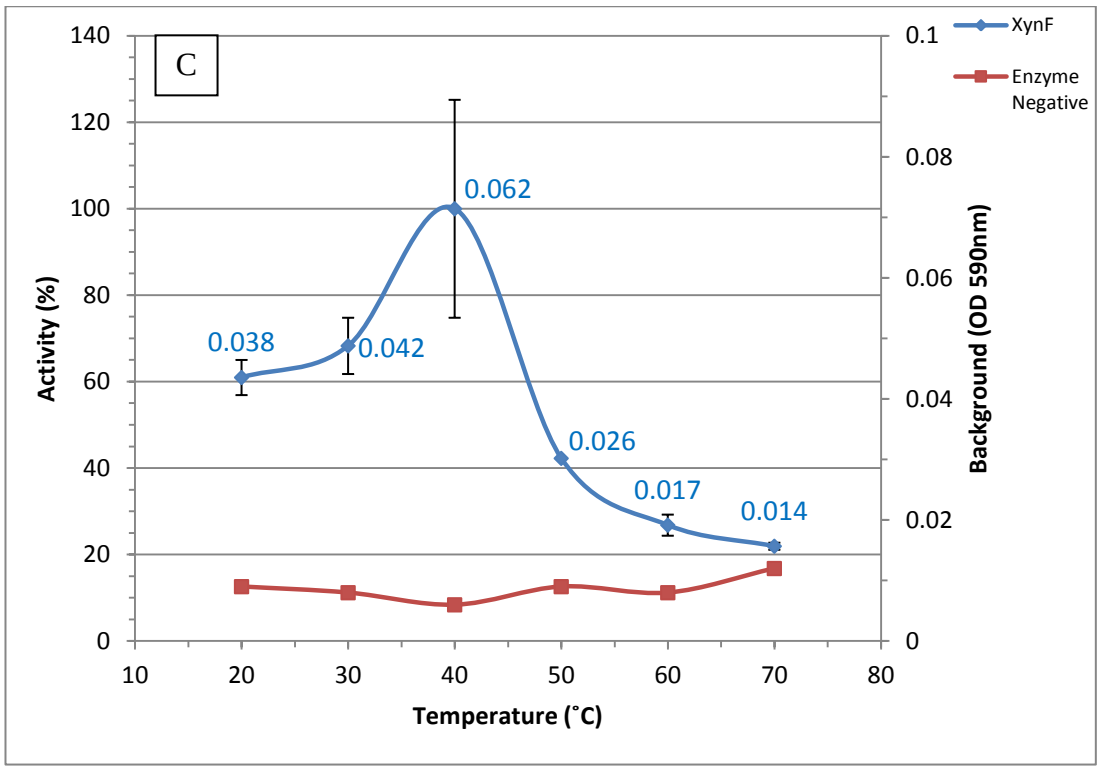


Figure 5.11. Plots showing the effect increasing temperature has on purified XynF as a percentage of the maximal activity observed on A) ONPC at pH 6, B) ONPX at pH 4.5, and C) RBB-xylan at pH 5.5. Floating numbers are specific activity of XynF at each point (A and B), or OD minus enzyme negative (C). The red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis, A and B), or OD of enzyme negative assays (2nd y-axis, C). n = 2 for activity screens, +/- standard error as percentage of maximal activity shown as bars

The plot of XynF activities is yet again different for its three substrates but have roughly similar characteristics. On ONPC at pH 6.0 (figure 5.11, A), activity starts at a low of 5% at 20ºC. It steadily climbs to a small plateau at 45-50ºC where activities are measured at 55% and 56%, respectively, but given the standard error is almost certainly due to an anomalous reading. However, unlike for AfsB and BxyF activity continues to climb as the temperature was increased to 60ºC. Activity remained fairly stable as XynF still retained 96% of its activity even at 65ºC, temperatures at which AfsB and BxyF were apparently almost entirely denatured. The high-end temperature stability was overcome at 70ºC where activity quickly plummeted to 14%, which is still on a par with the level of activity observed at 30ºC.

For XynF assayed using ONPX at pH 4.5 however (figure 5.11, B), the optimal temperature drops. The general shape and characteristics of the graph though are

similar to those of ONPC, with a steady increase in released nitrophenol from 20ºC up to the optimal temperature of 50ºC, before a decline thereafter. However, activities at lower temperatures were higher when ONPX was the substrate than those observed for ONPC. Highest activity was observed with RBB-xylan (figure 5.11, C) as the substrate, resulting in the lowest optimal temperature of just 40ºC. At 20ºC with RBB-xylan, XynF activity was 61% compared to 31% and 4% activity with ONPX and ONPC, respectively. At ten degrees below optimal temperatures, for ONPX activity was much higher, measured at 68% whereas on ONPC, at 10 degrees below the optimum the activity was ~55%. It appears as though tolerance of colder temperatures is much lower when ONPC is the substrate compared to ONPX. Activities on RBB-xylan and ONPX are retained at above optimal temperatures, much more so than for ONPC. This seemingly greater retention of activity is most likely due to the fact the XynF protein is not being denatured. If it were then 60ºC could not be the optimal temperature for ONPC substrate as activity would be lost there also. Protein denaturation appears to occur at 70ºC, as activities for all 3 substrates falls sharply.

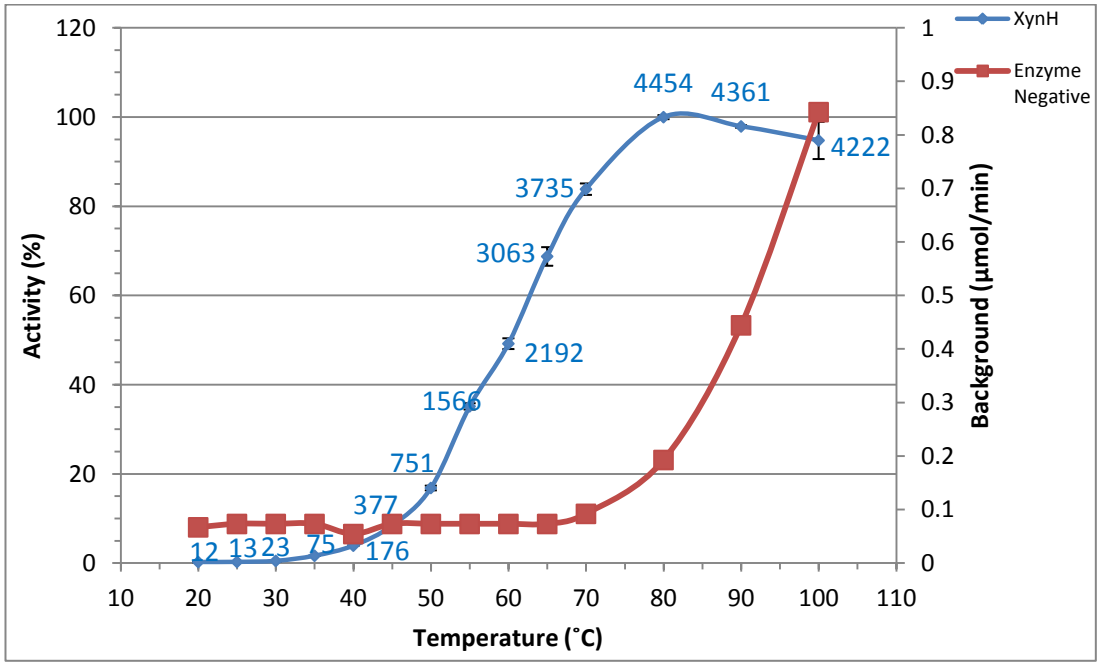


Figure 5.12. Effect of increasing temperature on XynH activity (percentage maximal observed) towards ONPX at pH 9, adjusted for background. Floating numbers are specific activity of XynH at each point, the red line indicates µmol per min of nitrophenol released in enzyme negative assays (2nd y axis). n = 2 for activity screen, +/- standard error, as percentage of activity, shown as bars.

Despite the higher optimal temperatures observed for XynF activity, its general behaviour follows that seen for AfsB and BxyF. The plot for XynH however differs drastically to the others (figure 5.12). Between 20ºC and 30ºC the release of nitrophenol is minimal with less than 0.5% activity observed for these 3 temperatures. The lowest observed activity at 20ºC for the other proteins assayed was 4% for XynF on ONPC substrate. As temperature increases however activity quickly increases again but easily surpasses the optimal temperatures already exhibited. At 45ºC, optima for AfsB and BxyF, there is only 8% activity for XynH. At 50ºC and 60ºC, optima for XynF on substrates ONPX and ONPC, respectively, XynH has 17% and 49% activity, respectively. The optimal temperature of XynH was found to be 80ºC, beyond the temperature range the other proteins were assayed at. XynH exhibits almost a complete reversal of character to those proteins previously assayed where suboptimal temperatures were preferable to super-optimal. At 10 degrees below optimal temperature, activity for XynH is 84%. At ten degrees above optimal, at 90ºC, XynH activity is 98%, dropping only to 95% at 100ºC, 20 degrees above optimal. At much less than twenty degrees above the observed optimal temperature the other 3 proteins assayed were completely denatured exhibiting next to no activity. XynH demonstrates an extremely high tolerance to temperatures generally reserved for extremophiles, but this stability comes at a cost to activity at mesophilic temperatures where the enzyme is relatively inactive.

5.3.3 *Effect of Supplements on Enzyme Activity*

Many enzymes require co-factors such as metal ions to function. Trace amounts are probably present from the buffer systems used and the salt used to elute the protein after purification. To determine whether an increased concentration of metal ions had an effect on enzyme activity, the assay reaction mixtures were supplemented with common metals. Each metal supplemented was of a 1 mM final concentration in the assay mixture. Common sugars may also have an effect through competitive inhibition. These were also tested in the same manner as the metal ions at 1 mM concentrations. The figures 5.13-5.16 are averages of duplicate runs showing percent activity relative to the enzyme-only fraction, with the standard error between

repeats given as a percentage of the maximum activity observed. Each protein was assayed under optimal pH and temperature conditions for 10 minutes and treated as previously described to halt reactions and measure absorbance. The significance of differences was calculated using the t-test in Excel to assess H0: there is no difference in population means vs H1: there is a difference in population means. Enzymes with supplements were compared to the enzyme-only fractions and analysed assuming the samples were independent (the enzyme mixture between runs could not be identical) to each other and of unequal variance. Samples with a p-value less than 0.05 were accepted as being significantly different.

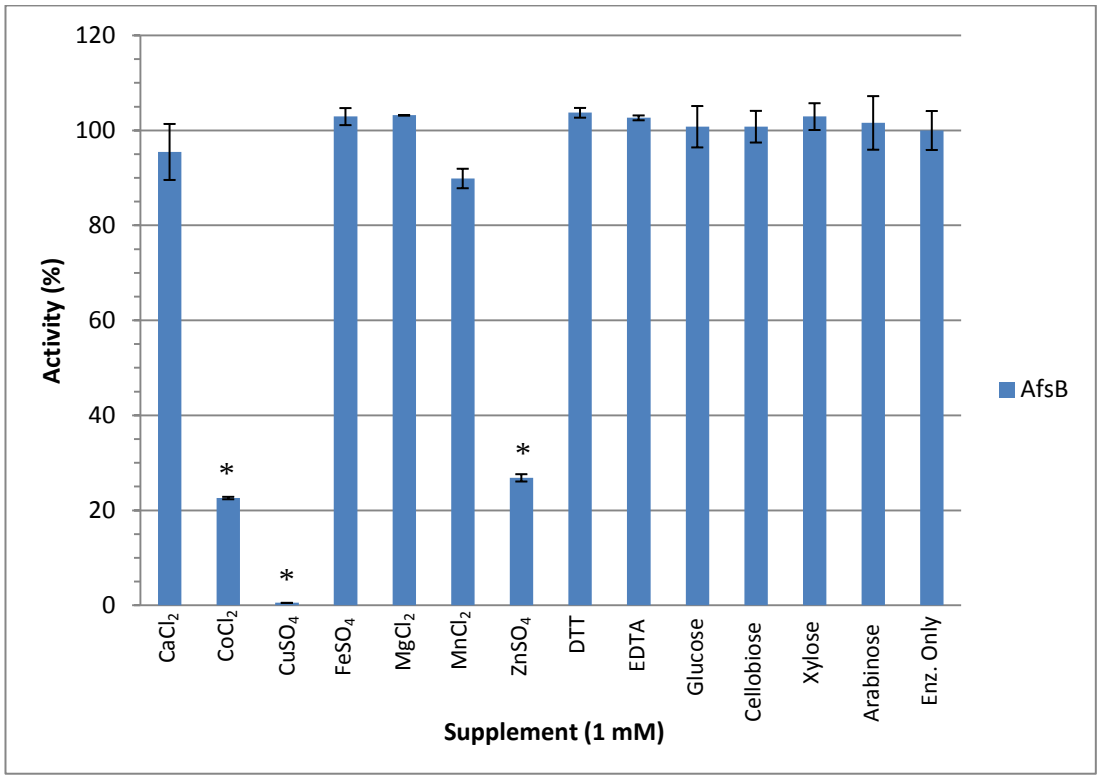


Figure 5.13. Bar plot showing the effect of metal ions and sugar supplements on the activity of AfsB at pH 6.5 and 45ºC on PNPA. Stars denote a difference at the 5% significance level to "Enz only" - AfsB with no supplement.

The most interesting ions would be those that enhance the activity of all the enzymes. Rather than being of benefit to the enzymes, the majority of the metal ions were in fact detrimental to their performance. Copper ions from $CuSO_4$ were universally disadvantageous with all four enzymes recording significant drops in activity. The largest effects were recorded for AfsB and BxyF (figures 5.13 and 5.14, respectively). In both instances the addition of Cu^{2+} ions resulted in 100% loss of

enzymatic function. The activity of XynH was also reduced by a large margin to 55% (figure 5.16). This metal also had a detrimental effect on XynF (figure 5.15), however the effects were most marked when ONPC was the substrate where a 49% loss of function was observed, but only 26% drop when ONPX was the substrate.

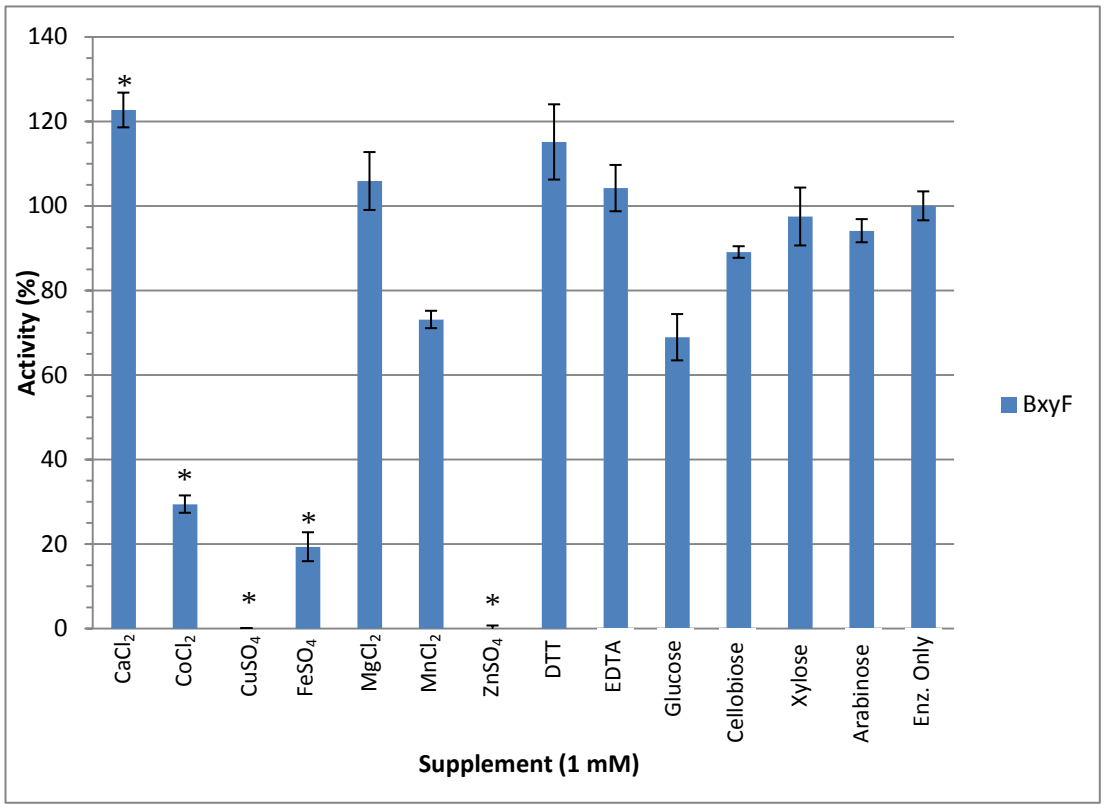


Figure 5.14. Bar plot showing the effect of metal ions and sugar supplements on the activity of BxyF at pH 6.0 and 45ºC on ONPX. Stars denote a difference at the 5% significance level to "Enz only" - BxyF with no supplement.

The only other metal ions to exhibit the same levels of enzyme inhibition as copper were zinc and cobalt. The addition of Zn^{2+} ions resulted in a significant drop in activity for AfsB, BxyF and XynH (figures 5.13, 5.14 and 5.16, respectively). The largest reduction in activity was recorded for BxyF where a total loss of function was seen. AfsB exhibited a loss of 73% of its functionality. XynH (figure 5.16) was more tolerant to Zn^{2+} ions and suffered a loss of only 36% to its activity. XynF displayed little or no change in activity from enzyme only mixtures in the presence of this ion (figure 5.15). Cobalt was found to affect the same set of enzymes again in fairly similar ways to zinc. Addition of Co^{2+} to the enzymes generated a 77% loss in activity for AfsB, a 71% drop for BxyF and a 36% drop for XynH.

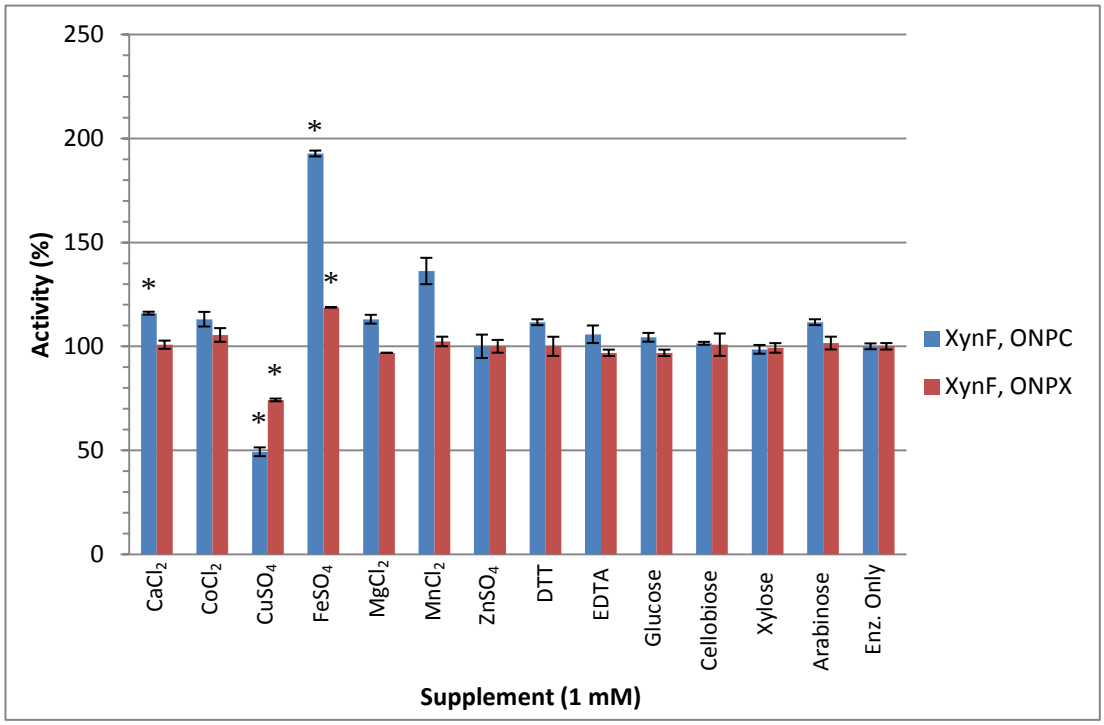


Figure 5.15. Bar plot showing the effect of metal ions and sugar supplements on the activity of XynF at XynF at pH 6.0, 60ºC on ONPC (blue) and pH 4.5, 50ºC on ONPX (red). Stars denote a difference at the 5% significance level to "Enz only" - XynF with no supplement.

The metal ion that most frequently had a positive impact on the enzymes was Fe^{2+}. For XynF on both substrates assayed (figure 5.15), addition of the metal salt resulted in a significant increase in activity compared to the enzyme only mixtures. When XynF was supplemented with 1 mM $FeSO_4$ there was a 19% increase in activity observed on ONPX activity, but a 93% increase in activity when ONPC was the substrate.

Ca^{2+} ions also had a largely positive effect on enzyme activity with enhancements to nitrophenol release observed to a significant level for BxyF and XynF (figures 5.14 and 5.15, respectively). The effect on XynF was isolated to just the ONPC substrate where a 16% increase in activity was detected. The influence of $CaCl_2$ was slightly more marked for BxyF where activity was elevated by 23%.

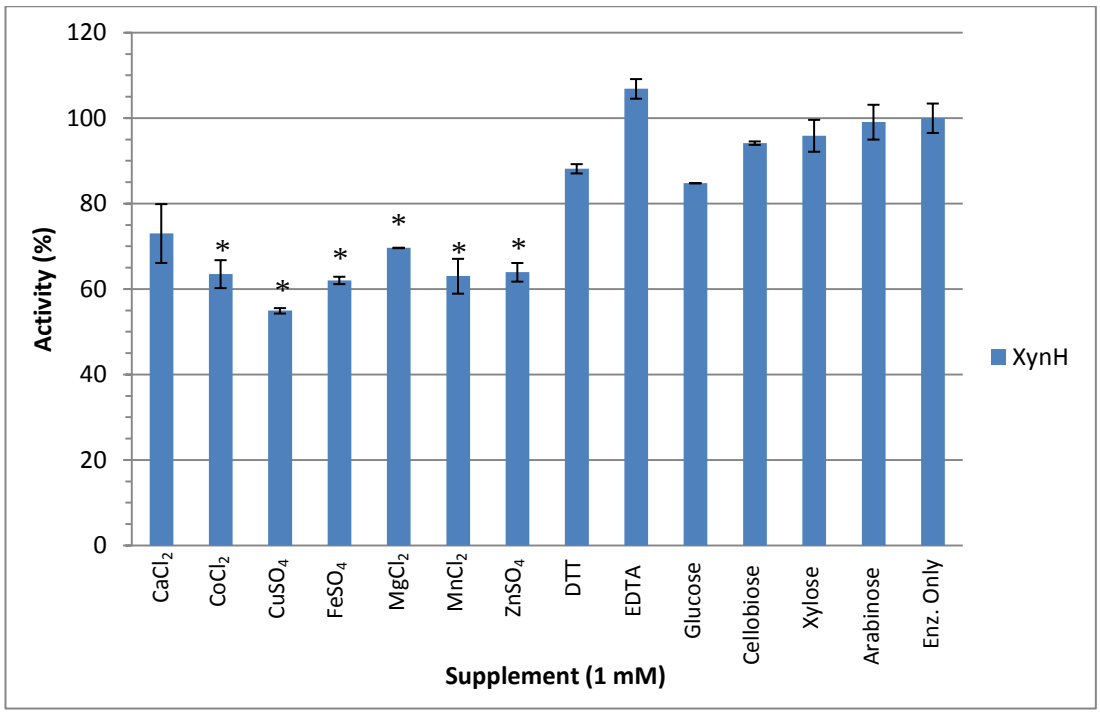


Figure 5.16. Bar plot showing the effect of metal ions and sugar supplements on the activity of XynH at pH 9.0 and 80 ºC on ONPX. Stars denote a difference at the 5% significance level to "Enz only" - XynH with no supplement.

$CaCl_2$ is the only metal supplement that doesn't change enzyme activity for the worse to a significant degree. Even Fe^{2+} ions, which imbue a positive effect on XynF has detrimental effects for BxyF and XynH. For BxyF there was an 81% drop in recorded activity and a decrease of 37% for XynH. XynH is singularly sensitive to metal ions, or at least 2+ ions, as nearly every metal tested resulted in a significant drop in functionality. It is the sole enzyme affected by $MnCl_2$ whose addition to the reaction buffer resulted in a decrease in activity of 28%.

The addition of EDTA, a metal chelator, gave no significant negative or positive effects to the enzymes. Nor did the presence of the reducing agent DTT which prevents the interaction of molecules by inhibiting the formation of disulfide bonds. The addition of simple sugars was to assay for competitive inhibition of the enzymes. Arabinose is released by AfsB, cellobiose (2 β-bound glucose molecules) by XynF and xylose by XynF and XynH. These enzymes may have been able to bind the free sugars leading to a drop in enzyme active sites. There was no evidence of this at the 1 mM concentrations used. Only glucose appeared to have a detrimental effect,

observed for BxyF but this was not shown to be of statistical significance (figure 5.14).

5.3.4 *Vmax and Km Values*

Using diluted samples of purified enzyme, the affect of increasing concentration of substrate was measured. Enzymes were diluted in buffer to a level where activity was still detectable to ensure an excess of substrate. The enzymes were incubated at optimal pH and temperature for 10 minutes with increasing concentrations of their appropriate substrate. Reactions were stopped and the absorbance at 405 nm was recorded as previously described. Using standard curves for 2-nitrophenol (ONP) and 4-nitrophenol (PNP), the release of nitrophenol from the substrate was calculated using the absorbance coefficients for each nitrophenol variant as in previous calculations of specific activity (U/mg). The velocity (V) of hydrolysis is given as U/mg (specific activity). All concentrations were duplicated and velocities were the mean of these replicates. To estimate Vmax and Km values, Eadie-Hofstee plots were used [191]. Any anomalous data that greatly varied from the trendline was removed, the trendline re-applied and the Vmax and Km values estimated when R^2 was at its highest value.

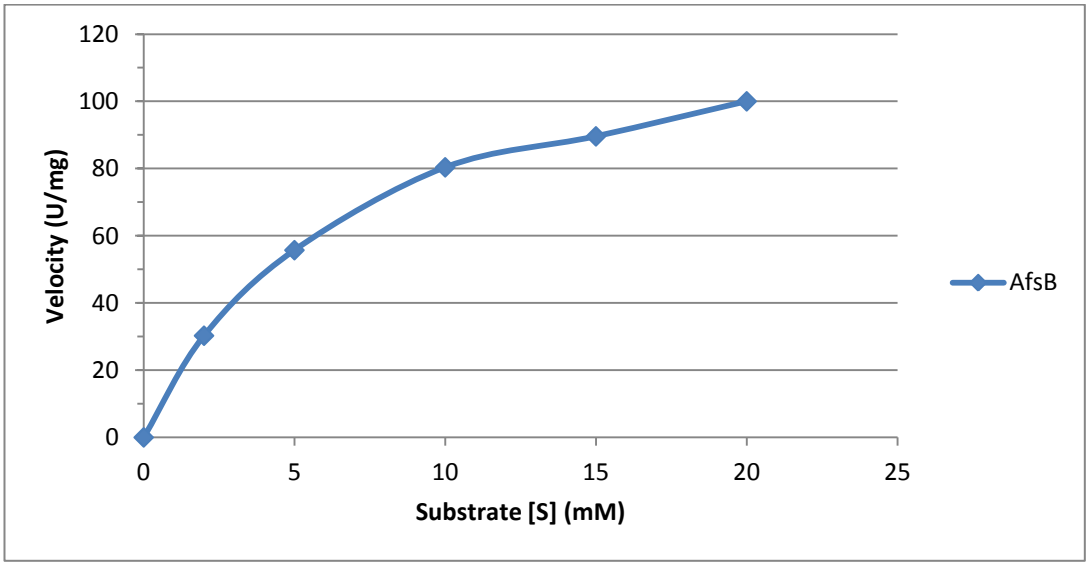


Figure 5.17. Plot of velocity of nitrophenol released from PNPA (S) at increasing concentrations at optimum pH and temperature by AfsB.

AfsB was diluted 1/5 and the velocity vs substrate concentration [S] was plotted (figure 5.17). From this we can estimate that Vmax is in the vicinity of 110 U/mg, with a Km therefore of roughly 5 mM. The Eadie-Hofstee plot of the same data is shown in figure 5.18. For this type of plot Vmax is the y-intercept and –Km is the line gradient.

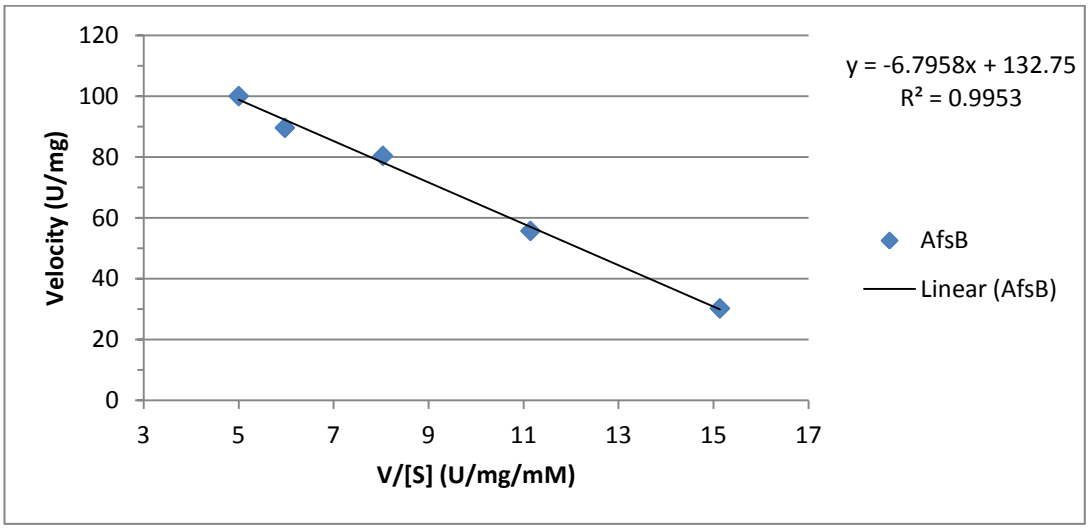


Figure 5.18. Eadie-Hofstee plot of AfsB velocity as a dependent of substrate concentration (mM). The trendline equation is shown as is the R^2 value.

Therefore, using the trendline equation in figure 5.18, the Vmax of AfsB at optimum conditions can be calculated as 133 U/mg, and the Km to be 6.8 mM. This marries up quite well with the values that were expected from the original plot.

Figure 5.19 shows the data for BxyF, which was also diluted by 1/5. There appears to be a levelling off and then an increase of enzyme velocity at the end. This is likely due to background levels of ONPX which is not soluble at high concentrations in aqueous solution. However, the data reasonably suggests a Vmax of 350 U/mg and a Km of 2 mM.

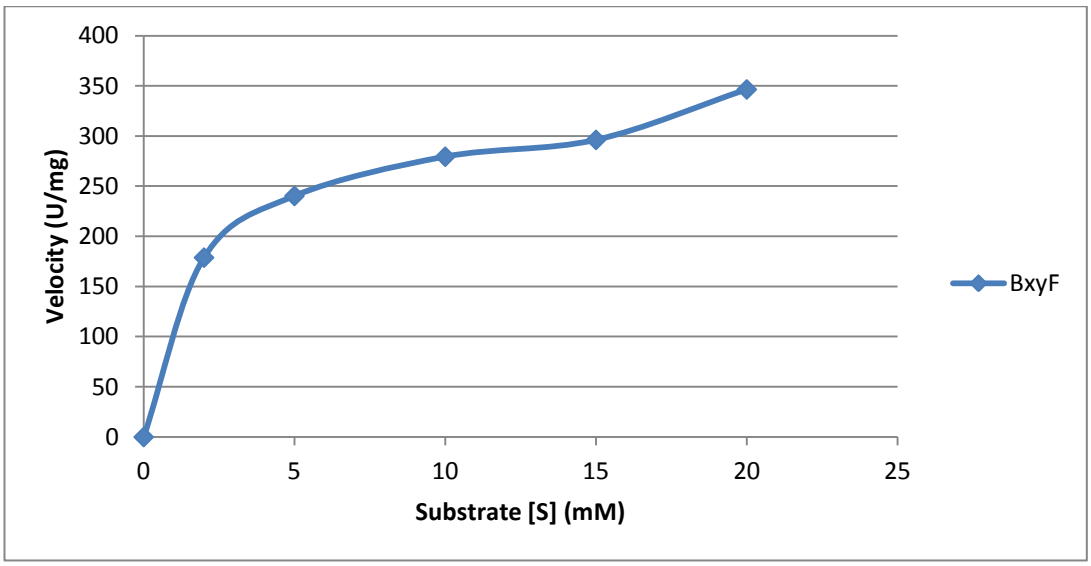


Figure 5.19. Plot of velocity of nitrophenol released from ONPX (S) at increasing concentrations at optimum pH and temperature for BxyF.

The Eadie-Hofstee plot in figure 20 has had the final 20 mM data point omitted which increased the trendline fit to 0.99. Subsequently the Vmax and Km were calculated to be 326 U/mg and 1.7 mM, respectively.

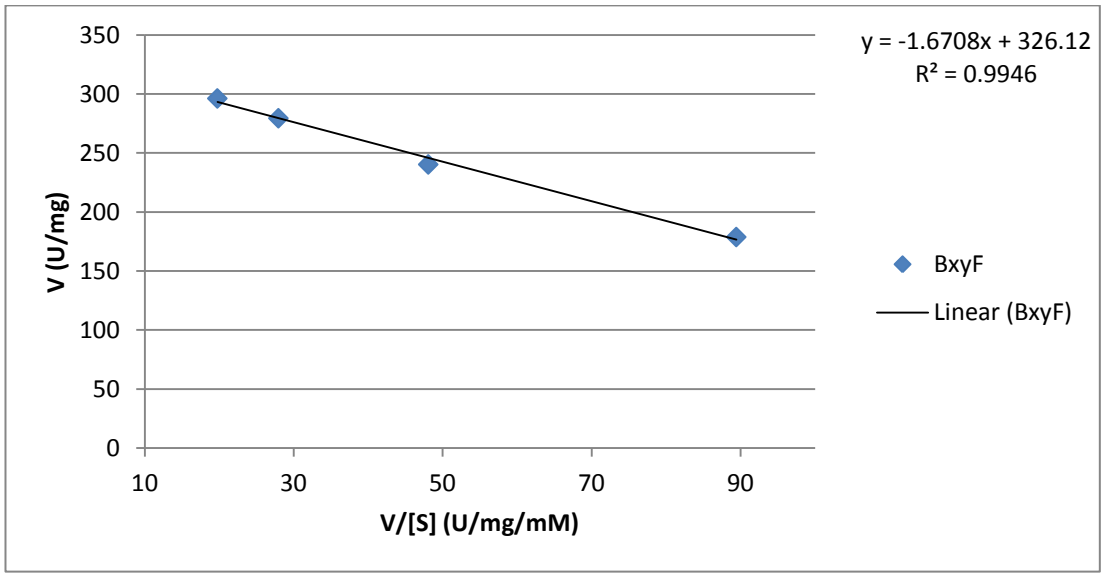


Figure 5.20. Eadie-Hofstee plot of BxyF velocity as a dependent of substrate concentration (mM). The trendline equation is shown as is the R^2 value.

XynF was diluted 1/20 and assayed using ONPC as the substrate. From figure 5.21 we could estimate that Vmax and Km to be 7000 U/mg and 50 mM, respectively. Removing the corresponding 10 mM point from the Eadie-Hofstee plot, which was clearly an outlier, greatly improved the R^2 value from 0.75 to 0.96 (figure 5.22). The

subsequent estimate of Vmax and Km values for XynF with ONPC as substrate are 8942 U/mg and 58 mM, respectively, greatly different values from the predicted values from figure 5.21. If the 10 mM data point is not omitted these estimates increase to a Vmax of 10151 U/mg and Km of 87 mM.

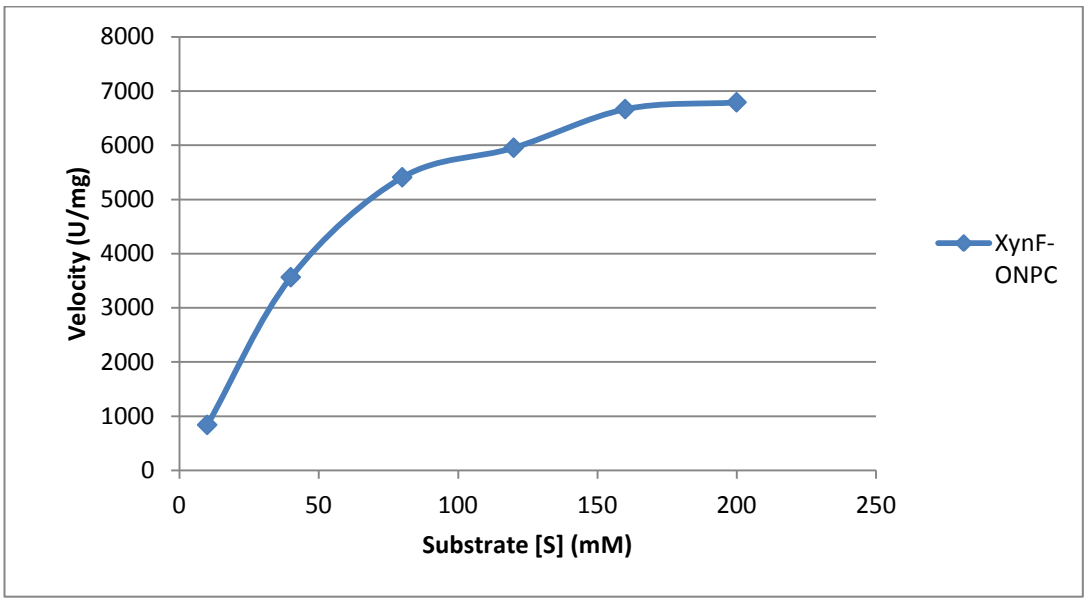


Figure 5.21. Plot of velocity of nitrophenol released from ONPC (S) at increasing concentrations at optimum pH and temperature for XynF.

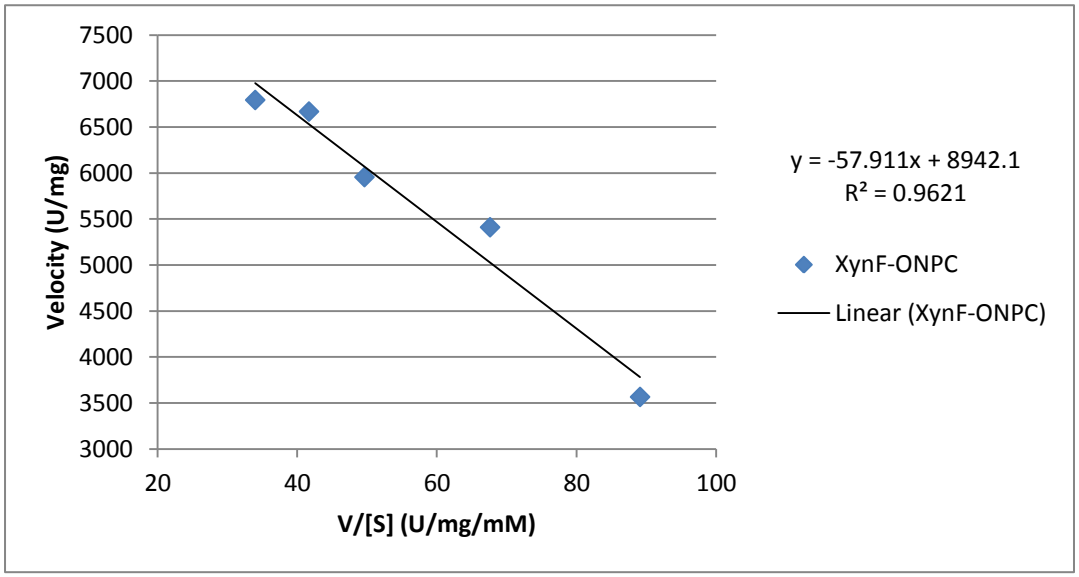


Figure 5.22. Eadie-Hofstee plot of XynF velocity as a dependent of substrate concentration (mM). The trendline equation is shown as is the R^2 value.

XynF on ONPX substrate was diluted by 1/3 and assayed at pH 4.5 at 50ºC. Figure 5.23 shows the specific activity at the increasing substrate concentrations, minus the

405 nm readings for enzyme negative reaction mixtures. It was observed that at higher concentrations of ONPX in the reaction mixture a precipitate was forming. This has produced a high background that has obscured the true activity of XynF. It is therefore not possible to tell if a levelling off of activity has occurred and so Vmax and Km cannot be estimated or calculated.

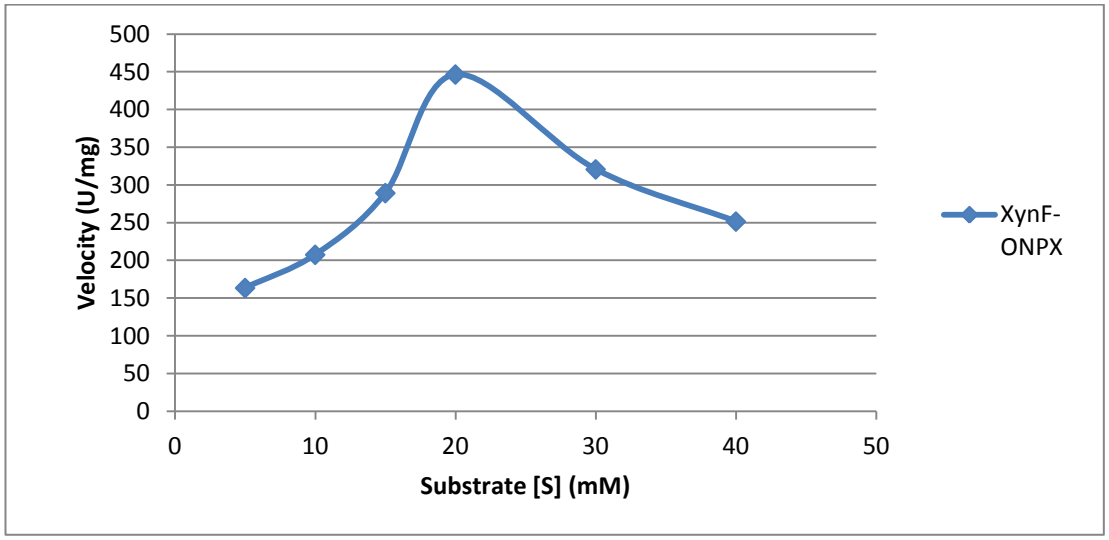


Figure 5.23. Plot of velocity of nitrophenol released from ONPX (S) at increasing concentrations at optimum pH and temperature for XynF.

The same problem for XynH is encountered as for XynF-ONPX. From figure 5.24 we can see that the velocity is not levelling off within the useable substrate concentration range of 0-20 mM. At higher concentrations of ONPX the activity appears to increase rapidly, after enzyme negative background readings are removed. While XynF-ONPX may have levelled before or around the 20 mM concentration, hence the dip after that point (figure 5.23), it appears as though XynH is still increasing in activity.

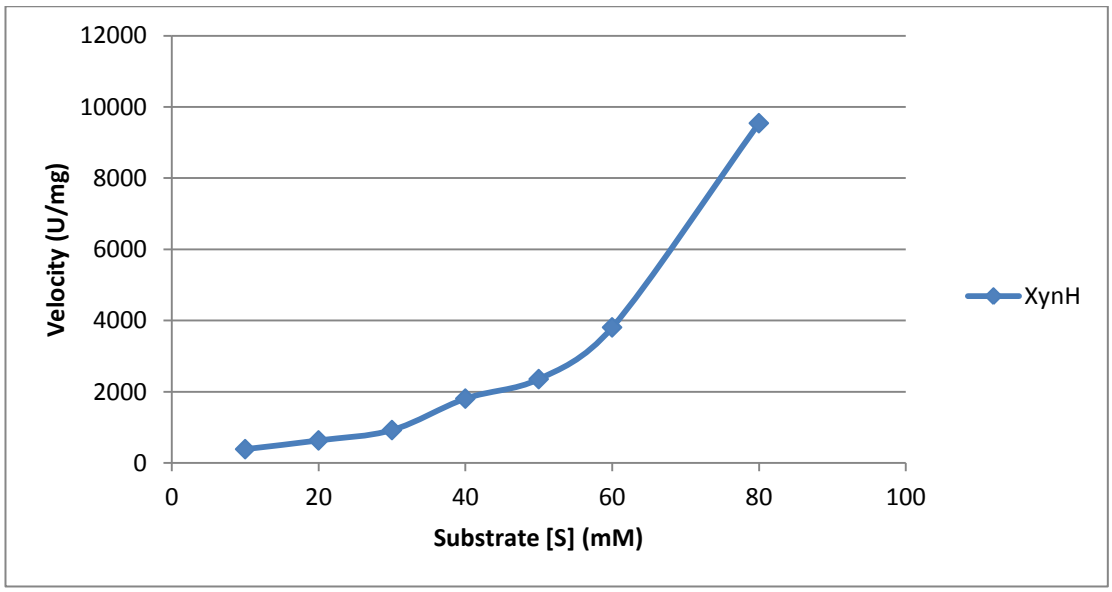


Figure 5.24. Plot of velocity of nitrophenol released from ONPX (S) at increasing concentrations at optimum pH and temperature for XynH.

Discussion

5.4 pH and Temperature Optima

The range of temperature and pH optima for the characterised enzymes is diverse but in keeping with those of other *Cellulomonas fimi* enzymes. CbhA and CbhB have optimal activity at pH 7 and an assumed temperature of 37ºC [76, 113]. CenC has a pH optimum of 5, CenA of 7.5 and CenB of 8.5-9, and temperatures of 45ºC, 37ºC and 37ºC, respectively [73]. Again the 37ºC temperatures are assumed as it is not explicitly stated what temperature is optimum, but this is the temperature the enzymes are assayed at.

XynF is a multifunctional exoglucanase/endoxylanase/xylosidase of GH family 10. As such it is the only characterised enzyme to resemble the activity of known *Cellulomonas fimi* enzymes, namely XynB and XynE which exhibit catalytic activity on both cellulose and xylose [16, 74, 80]. XynF has activity in the range of pH 4.5-6 and 40-60ºC dependent on substrate. XynB and XynE have similar ranges of pH 6 and 7 respectively, and optimum temperatures of 40ºC [16, 80]. These were however determined on one substrate only, in the case of XynE they were determined using

$PNPX_2$ [80], and for XynB using RBB-xylan [16]. It seems reasonable therefore that on other substrates were they to be tested that they too may have differing optima. Different substrates have different structures so will fit the catalytic domain differently leading to more efficient interactions and hydrolysis for some carbon sources than others. This is borne out by the multifunctional enzyme MFCase from *Ampullarea crossean* expressed in the edible fungus *Volvariella volvacea* which has differing pH and temp optima, albeit slight, of 5.5 and 55ºC and 6.0 and 50ºC, respectively on xylan and CMC [192]. A second with similar characteristics is the GH10 XynBE18 from *Paenibacillus sp* which with glucan as substrate has an optimal pH of 6.5 and on xylan of 7.5-9, but the same optimal temperature of 50ºC [193]. Ghatge et al also have described a multifunctional enzyme with differing optima for different substrates [194]. They describe a GH5 bifunctional endoglucanase with an optimal temperature and pH of 65ºC and 6.5, respectively, when CMC was the substrate and 55ºC and 5.5, respectively, when xylan was the substrate. The flexibility of the catalytic domain has been shown in silica to play an important role in the docking of the substrates, which is likely altered by different pH and temperatures [194, 195].

The two β-xylosidases, BxyF and XynH, as well as the α-$_L$-arabinofuranosidase, AfsB, are the first to be cloned and characterised from *C. fimi*. As such there are no direct comparisons to be made from the same organism. BxyF and XynH both belong to GH39 however, and are therefore comparable to others of the same family. BxyF most closely resembles those that have been published, of which there are few; only 10 GH39 enzymes from bacterial sources have been characterised according to CAZy. Its optima of pH 6.0 and 45ºC are not only consistent with the majority of *C. fimi* enzymes, but also the GH39 CcXynB2 enzyme of *Caulobacter crescentus* and the GH39 enzyme of *Bacillus halodurans* which have pH and optima of 6 and 7.5, respectively and a temperature optimum of 55ºC [196, 197]. From related species *Cellulomonas uda*, a β-xylosidase was cloned and characterised of unknown GH family with pH and temperature optima of 5.4-6.1 and 43-45ºC, respectively [198].

Perhaps of more interest is XynH with its temperature optimum of 80ºC and pH optimum of 9, with a secondary peak at 5.5, making it seemingly tolerant of both acidic and basic conditions. XynH displays a radically different activity curve when assayed at different pH levels than the other enzymes here. For each buffer used, as the pH increases to the max of its buffering range, the activity of XynH increases, before falling again when the next buffer is introduced. The change in activity of XynH is most marked for sodium acetate and Tris-HCl buffers, and more muted when potassium phosphate and glycine buffers are employed (figure 5.8). It would appear that the buffers used have a large and varying effect on the activity of XynH.

BxyF appears to be unaffected by these buffer associated differences, despite being the same GH family and therefore having the same catalytic domain and residues as previously stated. BxyF is however larger than XynH with a predicted α/β hydrolase fold domain, absent from XynH, upstream of the catalytic domain. This extra domain may help shield the buffer effects or add extra structural stability to BxyF, protecting against buffer mediated effects, where as XynH is more susceptible to the buffers. The structurally stabilising effect α/β folds were demonstrated by Ugwu and Apte for phosphate buffers at least [190].

Different buffers have been shown to have dramatic effects on a wide range of enzymes. A β-amylase from wheat was found to have differing pH optima at constant temperature depending on the buffer used [199]. These ranged from pH 4.5 for sodium citrate or potassium phosphate buffers, up to pH 5.3 observed when using potassium hydrogen phthalate buffer which, they state, may have been caused by differing rates of enzyme inactivation. More recently Chen et al reported that the optimum pH of protease MCP-01 was markedly different when altering buffers with pH 6-7 observed for phosphate buffer, pH 7-8 for borate buffer, pH 8.5 for Tris buffer and pH 9 for carbonate buffer [200]. They postulated that the differing characteristics seen in each buffer was due to flexibility in the enzyme's structure modulated by the buffer charge and ionic species.

From the experiments testing the effect of metal ions on the activity of the enzymes, it is clear that metal ions can have a profound positive or negative effect on XynH. The largest effects in activity are seen for sodium acetate (releasing Na^+ upon dissociation) and for Tris-HCl (releasing Cl^- upon dissociation). This could indicate that these ions have a positive effect on activity, or that the potassium released by the phosphate buffer, or glycine has either an inert effect on activity or a negative one. It has also been reported that buffer ions can have chaotropic (protein destabilising) or kosmotropic (protein stabilising) effects [190]. In one experiment researchers observed a continual decay in activity of organophosphorus hydrolase when CHES buffer was used, with the activity fully restored when PBS was used (both buffers at pH 6.9), leading them to conclude that the sulfonate group was more chaotropic for this enzyme than the phosphate anion [201]. The half life of enzyme $_L$-amino acid oxidase of *Rhodococcus opacus* was 938 minutes when incubated with glycine-NaOH as buffer compared to 44 minutes when using potassium phosphate or 35 minutes when using TEA-HCl, all at pH 8.6 [202]. For phosphoenolpyruvate decarboxylase at pH 6.2 in MES buffer, which was then supplemented with additives, showed that anions such as HPO_4^{-2} (phosphate) and OAc^- (acetate) were stabilising for the enzyme, Cl^- and Br^- were destabilising, while Na^+ and K^+ were largely inert [203]. An endoxylanase from *B. subtilis* was made more stable and had its activity increased by both K^+ ions and the anion HPO_4^{-2} [204].

The effect a buffer can have on an enzyme can be profound, but is the product of the interplay between a range of competing or complimentary factors affecting protein structure and interactions, so is unpredictable and not necessarily applicable from one protein in a family to another. Whatever the reason for the fluctuating pattern of activity observed for XynH, activity always increases as pH increases for each buffer. This suggests that the optimum pH of the enzyme is more alkaline than acidic, and may in fact be over pH 9, given the continued activity increase when glycine buffer was used.

Its thermo-optimal properties would make it of particular interest for industrial applications. At 80ºC, with over 90% activity seen as high as 100ºC, it is the highest

recorded for a *C. fimi* enzyme to date. These values are higher than those reported for the GH39 xylosidases previously mentioned, and is more akin to temperatures for β-xylosidases from thermophilic bacteria such as *Thermatoga maritima* [205], 90ºC, *Bacillus thermanarcticus* [206], 70ºC and *Geobacillus pallidus* [207], 70ºC to name but a small number. There are several factors that could account for this including amino acid composition and side chain-side chain interactions [208]. Thermo-optimal enzymes could confer an advantage to *C. fimi* or other organisms despite it being a mesophile. Soil temperatures can vary greatly and in areas where there is concerted decay of material, such as composts, both pH and temperature can fluctuate significantly [209]. It may not be that uncommon for organisms such as *C. fimi* to harbour enzymes that function at extremes of temperature and pH. Verma et al, while looking for xylanases in a compost meta-genome identified a xylanase of GH11 with a 75% identity to xylanases in *C. fimi* (although they don't say which) that exhibited optima of 80ºC and pH 9 [210]. Raj et al isolated from the mesophile *Stenotrophomonas maltophilia* a xylanase with optima of pH 8 (with strong activity at pH 9) and temperature of 80ºC [211]. Mitsuishi et al were able to isolate 3 xylanases from a mesophilic fungus, strain Y-94, with unorthodox properties [212]. The discovered the 3 enzymes (xylanases A, B and C) had the same optima, of pH 4.9 and 80ºC, but their stabilities differed with A and B being stable between pH 2.5 and 9 and at 70ºC, whereas C was unstable above pH 5.5 and at 70ºC. Most published alkilophilic or thermostable optimal enzymes are reported from extremophiles, but the above shows that it may not be particularly uncommon for mesophiles to have extremophile-like enzymes to give extra adaptability to a changing environment. Each new genome uploaded to the NCBI database shows just how few enzymes have actually been characterised so more examples may yet be found.

AfsB is a typical bacterial arabinofuranosidase in that it is optimally active at neutral pH 6, as opposed to a more acidic pH as is characteristic of fungal arabinofuranosidases [213]. AfsB has not been identified as belonging to any GH family by domain searching, however its sequence most closely groups with GH51 (see chapter 4). Its optimal values are consistent with those of bacterial AbfB, a

GH51 arabinofuranosidase from *Bifidobacterium longum*, pH 6, 45ºC, GH51 family AbfAC26Sari from *Anoxybacillus kestanbolensis*, pH 5.5, 65ºC, GH51 Abf51A from *Pseudomonas cellulosa*, pH 5.5, 55ºC and the GH51 TpAraF from *Thermotoga petrophila*, pH 6, 65ºC [213-216]. To date I can find no arabinofuranosidases from *Cellulomonas* species with which to compare.

5.5 Effect of Metal Ions on Activity

It is clear from the data that no metal ion is universally beneficial to these enzymes, in fact, just the opposite is true. It is known that certain metal ions have a propensity for inhibition of hemicellulases (Hg^{+2}, Cu^{+2}, Co^{+2}, Fe^{+2}, Zn^{+2} and reagents like DMSO and EDTA), and others have an enhancing effect (Ca^{+2}, Na^{+2}, K^{+} and Mg^{+2}) all at 1 mM [217]. Cu^{2+} ions inhibit the action of all the enzymes, completely in the case of AfsB and BxyF with Zn^{2+} ions having a similar effect on these 2 enzymes. XynF appears to be the most tolerant of metals exhibiting little loss of function even in the presence of copper, and is mildly enhanced by the presence of Fe^{2+}. XynH could be argued to be the most susceptible of all the enzymes to metal interference as it displays a noted drop in activity in the presence of all metals added. None of the enzymes apparently require metals as a cofactor as the addition of EDTA, a chelator of some metals, has no ill effects, nor does the presence of DTT suggesting di-sulphide bonds are not necessary for proper function. As with heavy metal poisoning 2^{+} metal ions can bind to the SH group of amino acids. A catalytic residue with an SH group therefore may explain the inhibitory effect. Typically however, for xylosidases and glucanases the catalytic residues are mainly glutamic acid but with asparagine, glutamine, histidine, tryptophan and tyrosine also found [218]. The ions are not likely to be directly interacting with the catalytic residues, but could be interacting with residues in the catalytic domain, blocking substrate binding, or destabilising the enzyme structure by interacting with stabilising cysteine residues.

The effect of metal ions is not always assayed for when characterising proteins, therefore comparisons will be limited. The arabinofuranosidases AbfAC26Sari and AbfB were both highly susceptible to Cu^{2+} and Hg^{2+-} ions, completely removing

activity, with Zn^{2+} conferring a 50% loss of activity to AbfB, but a slight increase in activity was observed for AbfAC26Sari [213, 219]. These ions are known inhibitors of enzymes from GH51 [213, 219].

Inhibition of β-xylosidases by metals is more usual than enhancement. A β-xylosidase from *Bacillus thermantarcticus* displayed strong inhibition in the presence of Fe^{2+} and Cu^{2+}, but no enhancement, where as a xylanase from the same bacterium showed inhibition by these two ions but also slight enhancement in the presence of Mg^{2+} [206]. Only the presence of KCl improved the activity of CcXynB2, and all other metals, including those tested here resulted in inhibition [196].

5.6 Vmax and Km

The velocity of an enzyme is important to know to determine its effectiveness. The faster it works obviously being the better. Km is a measure of an enzyme's affinity for its substrate, so the lower the Km, the less substrate is needed to reach Vmax. XynF with ONPC as substrate displayed the greatest Vmax, calculated as 8942 U/mg and a Km of 58 mM. The Vmax and Km for AfsB were 133 U/mg and 6.8 mM, respectively, and the Vmax and Km of BxyF were 326 U/mg and 1.7 mM, respectively. XynF-ONPX and XynH not levelling off in the substrate range assayed would suggest that they have a high Km value and so low affinity for the substrates.

When compared with commercially available enzymes these from *C. fimi* compare very favourably. One of the main suppliers of commercial polysaccharide degrading enzymes is Megazyme (www.megazyme.com). Table 5.7 shows the range of arabinofuranosidases, xylosidases and xylanases they produce along with any kinetic data supplied, presumably at optimum temperature and optimal substrate concentrations.

Enzyme and Catalogue Number	Species	Glycosyl Hydrolase Family	Optimal pH and Temperature	Specific Activity (U/mg)	Notes
α-$_L$-arabinofuranosidase (E-AFASE, E-ABFCJ, E-BFCT)	*Aspergillus niger*	51	40ºC, pH 4.0	19	PNPA, 5 mM
	Cellvibrio japonicus	51	50ºC, pH 5.5	21	PNPA, 2.5 mM
	Clostridium thermocellum	51	60ºC, pH 5-6	155	PNPA, 2.5 mM
Cellobiohydrolase (E-CBHI)	*Trichoderma longibrachiatum*	Not stated	70ºC, pH 4.5-5	0.0074	PNPC, 2.5 mM, pH 4, 40ºC
β-$_D$-xylosidase (E-BXSEBP, E-BXSR)	*Bacillus pumilus*	43	35ºC, pH 7.5	21.2	PNPX, 5 mM
	Selenomonas ruminantium	43	40ºC, pH 5.3	64	PNPX, 5 mM

Table 5.7. Commercial enzymes produced by Megazyme

For AfsB at 2.5 mM and 5 mM concentrations of PNPA, specific activity is 38 U/mg and 56 U/mg, respectively. These values are at least double those given for the enzymes from *A. niger* and *C. japonicus*. The thermophilic *C. thermocellum* arabinofuranosidase displays an activity nearly three times that seen for AfsB at a 5 mM substrate concentration and nearly 5 times that of the 2.5 mM substrate concentration. The levels displayed by AfsB however were much lower than arabinofuranosidases reported in the literature. Margolles et al describe AbfB with Vmax of 1019 U/mg and a Km 0.139 mM from *Bifidobacterium longum* and Canakci et al AbfAC26Sari from *Geobacillus caldoxylolyticus* with a Vmax of 417 U/mg and a Km of 0.295 mM for PNPA [213, 214]. Afs(GH54) of *Aerobasidium pullans* is closer to that observed for AfsB with a Vmax of 34.8 U/mg and a Km of 3.7 [220].

The commercial β-xylosidases range in pH values from 5-7.5, and have optimum temperatures around 40ºC. BxyF, XynF and XynH all have optimal pH values within this range but their optimal temperatures are generally higher. The *C. fimi*

enzymes all display higher specific activity, with 240 U/mg for BxyF, 165 U/mg for XynF and 400 U/mg for XynH.

XynF displays considerable activity as a cellobiohydrolase. Megazyme produce only one cellobiohydrolase, that of *T. reesei* which exhibits a very low specific activity at a substrate concentration of 2 mM *p*NPC (table 5.7). At such a low substrate concentration XynF would be far from its maximal velocity, but would, by looking at figure 5.21, still exhibit an activity far beyond that stated for the commercial enzyme. The low specific activity may suggest that the production of this enzyme by Megazyme is quite problematic resulting in only low yields and much contamination by other proteins. Sigma-Aldich also produce a cellobiohydrolase enzyme CBHI from *Hypocrea jecorina* (cat. number E6412), with an optimum temperature and pH of 45ºC and 5-6, respectively. They ship the enzyme in a concentration of 0.13 U/mg on MUC, but do not state its activity at a given substrate concentration or any kinetic data.

Conclusion

I have described in this chapter the purification and characterisation of four novel enzymes from *Cellulomonas fimi*. AfsB a GH51 α-$_L$-arabinofuranosidase, BxyF and XynH, two GH39 β-$_D$-xylosidases and XynF, a multifunctional GH10 enzyme exhibiting endoxylanase, xylosidase and cellobiohydrolase activities on nitrophenyl substrates. Given the affinity and activity displayed by XynF this may be better annotated as a cellobiohydrolase, the third reported in *C. fimi*. They are active over a wide range of pH values (4.5 – 9) and temperatures (40ºC - 80ºC). This all further supports the screening of genomes to identify active enzymes that may be missed by traditional methods.

CHAPTER 6: TRANSFORMATION OF *Cellulomonas fimi*

Cellulomonas fimi has the complete toolset to degrade and utilise cellulose, hemicellulose and its component sugars both hexoses and pentoses. This would make it an almost ideal organism for a consolidated bioprocessor. For this *C. fimi* would need to be transformed and relevant genes introduced to allow the fermentation of the sugars to produce biofuel. Fermentation occurs during anaerobic growth which *C. fimi* is capable of. Ethanol is a natural by-product of this respiration process, one of the main products of bioprocessors so far. One of the ways ethanol is produced is through the mixed-acid fermentation pathway which is present in *Cellulomonas fimi* (see Genome Analysis).

The mixed acid fermentation process is also found in *Escherichia coli* and the introduction of the production of ethanol (PET) plasmid has been successful in transforming *E. coli* to produce elevated levels of ethanol [50]. The plasmid expresses pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase II (*adhII*) from *Zymomonas mobilis*, the two essential enzymes needed to convert pyruvate to acetaldehyde then to ethanol, respectively.

The introduction of these enzymes into Gram-positive bacteria has met with limited success for ethanol production compared to Gram-negative bacteria [221]. The only instance of an *Actinomycete* being transformed to produce ethanol is *Streptomyces sp.* by Hardter et al [222]. They successfully introduced PDC and AdhB enzymes into different *Streptomyces* species and were able to produce 2.5 g/L of ethanol when grown anaerobically with $_{D}$-glucose over 6 days. Possibly more problematic however is that the PDC enzyme is quite rare in bacteria [221]. Talarico et al, by altering the source organism of PDC, were able to greatly improve the production of ethanol in Gram-positive hosts. Using *Bacillus megaterium* as expression host, the *pdc* genes of Gram positive *Sarcina ventriculi*, yeast *Saccharomyces cerevisiae*, Gram negative *Acetobacter pasteurianus* and that of *Z. mobilis* were screened for activity. The *pdc* gene of *S. ventriculi* when expressed displayed the highest specific activity assayed by measuring ADH coupled oxidation of NADH of 5.29, with that

of *Z. mobilis* achieving 1.11 specific units, 0.53 for *S. cerevisiae* and 0.14 for *A. pasteurians* [221]. The authors believe this variation in specific activity is primarily due to codon usage differences between the host species and the gene donor species, with codon usage in Gram-positive bacteria significantly different from that of Gram-negative. This makes the production of ethanol from Gram-positive bacteria a better prospect than it once was, especially after codon optimisation is also instigated.

Transforming *C. fimi* therefore could be very useful for the production of a consolidated bioprocessor. Adding two genes for the production of ethanol to one organism already optimised for the use of lignocellulosic biomass is easier than adding a minimum of 4 genes, the number required for the hydrolysis of cellulose alone (introduction), necessary for converting, for example, *E. coli* into a lignocellulolytic consolidated bioprocessor.

6.1 Electrotransformation

Transformation of *Cellulomonas fimi* with heterologous DNA has not been reported. In related species, the only member of the *Cellulomonas* genus transformed is *Cellulomonas flavigena*. Pulido-Vega et al created protoplasts and then introduced plasmid pC194 isolated from *Bacillus subtilis* SB234 via PEG mediated transfer and cell regeneration, resulting in chloramphenicol-resistant cells [103]. Alemohammad et al successfully used electroporation to transform *C. flavigena* with an *E. coli/Brevibacterium lactofermentum* shuttle vector pJ85 conferring chloramphenicol and kanamycin resistances [223]. Montes-Horcasitas et al latterly used conjugation and electroporation as their transformation methods, successfully introducing plasmid pHT73 from *Bacillus thuringiensis* generating erythromycin resistant cells [224]. As such these are the methods I also adopted when attempting to transform *C. fimi*.

Transformations were attempted using the plasmids pVK168, pTG262 and pCJW93 (table 2.2). pVK168 is an *E. coli/Bacillus subtilis* shuttle vector previously used in

our laboratory that confers spectinomycin resistance. pTG262 is a multiple host shuttle vector available in our laboratory that confers resistance to chloramphenicol, and has been used to transform *E. coli*, *B. subtilis*, *Lactobacillus* sp and *Lactococcus* sp. As both have been used previously to transform Gram-positive bacteria they were considered suitable to see if transformation of *C. fimi* was possible. The final plasmid used was pCJW93, an *E. coli/Actinomycete* shuttle vector which has been used to transform both *E. coli* and a range of high G+C content *Streptomyces* sp., kindly donated by Professor Peter Leadlay [102]. It confers resistance to apramycin as its selectable marker, but also confers resistance to thiostrepton by producing protein Tsr under the control of a constitutive promoter, as the plasmid contains a thiostrepton inducible promoter, *tipA*, and the antibiotic would be lethal to the cell without the presence of Tsr [102, 225]. Antibiotic screens by addition of 5 µl of antibiotic stocks to a plate with *C. fimi* spread upon it indicated that 100 mg/ml spectinomycin (spec), 50 mg/ml chloramphenicol (cml), 50 mg/ml thiostrepton (tsp) and 100 mg/ml ampicillin (amp) are lethal to the bacterium, each creating a large zone of clearing. *Cellulomonas fimi* however proved to be naturally resistant to 100 mg/ml kanamycin (kan) where no zone of clearing was evident. Liquid cultures incorporating final concentrations up to 500 µg/ml erythromycin and up to 500 µg/ml apramycin exhibited little to no inhibition of growth. Liquid culture screens of those antibiotics initially tested on plates confirmed the results. As such for transformations using pCJW93 thiostrepton was used as the selective antibiotic.

The most straightforward method for transformation lies with electroporation. Following the protocol of Montes-Horcasitas, their electrocompetent cell preparation method was derived from that of Bone and Ellar [226]. *Cellulomonas fimi* cells were grown to mid-log phase, cells were recovered by centrifugation and the pellet washed and stored in HG medium (1 mM HEPES pH 7, 10% glycerol). Minimal medium containing cellobiose and kanamycin was used in the cultures to ensure only *C. fimi* would be growing, since its slow growth rate made contaminants a high risk.

Electroporation was then carried out with a 2.5 kV charge using BioRad MicroPulser II apparatus, of unknown capacitance. Using the buffers as described by Bone and

Ellar caused arcing to occur, most likely due to the salt content of the buffer. For subsequent transformation attempts the cells were therefore washed thoroughly in deionised sterile water to remove salt traces. This successfully prevented arcing from occurring.

The survivability of *C. fimi* electroporation was tested. The cfu/ml of thawed *C. fimi* competent cells before and after electroporation was measured. On average, per aliquot of cells, there were $1x10^{13}$ cfu/ml compared to $1.9x10^{9}$ cfu/ml after the procedure. This equates to 0.02% of cells surviving the process, but confirms *C. fimi* is still viable. This survival rate is very low but equates to a large proportion of viable cells still present. Survival rates were not reported for *C. flavigena* in the papers mentioned previously. With optimisation of parameters this rate could be increased with electroporation at optimal conditions for *Fusobacterium nucleatum*, a Gram-negative bacterium, reaching 20% [227].

Plasmids pVK168, pTG262 and pCJW93 were then introduced to the cells as described in the materials and methods section, and plated to selective media containing appropriate antibiotics. Cells transformed with pTG262 or pCJW93 failed to generate any colonies on their respective plates. Those transformed with pVK168 however generated similar numbers of colonies for both plasmid positive and negative transformations, between 30 and 50 colonies in total per plate. Colony size however varied greatly, with some being larger in diameter than others, suggesting they were able to grow slightly faster. There were more of these colonies observed on the DNA positive transformant plates than the negative. This indicated that *C. fimi* was able to generate spec resistant colonies in the absence of plasmid, obscuring whether or not any transformation had taken place especially with negative results for plasmids pTG262 and pCJW93.

The main barrier to transformation of Gram-positive cells is the peptidoglycan layer encasing the cell membrane. It's much thicker than that of Gram-negative cells so adds another layer of impedance to DNA uptake. It has been demonstrated that the use of agents that interrupt the structure of this layer, thereby weakening it, can have

a strong effect on transformation efficiencies [228]. One of the more common additives to growing cells is glycine [228-233]. This works as the glycine replaces $_{L}$- and $_{D}$-alanine in the peptidoglycan layer which impedes synthesis and assembly of the cell wall [230]. The *C. fimi* cell wall contains $_{D}$-alanine so glycine is a suitable addition [234]. Similarly, penicillin-G acts by interfering with cell wall synthesis by disabling transpeptidases. The mode of action of Tween-80 is unclear but it may alter the mycolic acid composition of the cell wall [232]. Lysozyme degrades the peptidoglycan layer generating gaps and weaknesses in the cell wall for DNA to pass through. These supplements can be toxic to the cells, so addition during mid-log phase should still result in many cells being affected but ensuring growth and a large yield of cells.

Cultures of *C. fimi* were therefore treated with lysozyme digestion ("lyso"), glycine ("gly"), penicillin-G ("pen-G"), Tween-80 ("tween") or prepared with no treatment ("base") as described in the materials and methods section. Cells were still viable after wall weakening with on average $7x10^9$ cfu/ml for base, $2.3x10^8$ cfu/ml for lyso, $5.5x10^9$ cfu/ml for gly, $1.5x10^{10}$ cfu/ml for pen-G and $1x10^8$ cfu/ml for tween cell preparations. Transformations of these prepared cells with pTG262 and pCJW93 again failed to generate colonies after electroporation with or without plasmid.

Transformations with pVK168 once again generated colonies for both plasmid DNA positive and negative experiments. Cells treated with lysozyme however, did not produce colonies for either DNA positive or negative transformations. For the other treatments the main differences between positive and negative plates lay in the size of observed colonies. Cells treated with pen-G generated the highest number of colonies, with 4 large colonies observable for DNA positive, with hundreds of very small colonies covering the plate. No observable large colonies were seen on the DNA negative plates, but still hundreds of small colonies, although clearly a lesser number to those on the DNA positive plate. The other treatments appeared to have little effect on resistant colony generation with comparable numbers visible on positive and negative plates. The DNA positive plates for Gly and Tween cell preparations however also had 1 large colony each, with 2 large colonies observed

for base cells, with none observed for negative plates. A selection of large and small colonies from the negative and positive plates were patched to CMC containing plates and assayed for endoglucanase activity using Congo-red (figure 6.1). For this the cells proved positive, except for those patched from the base plate (figure 6.1. A – 1+2), strongly indicating the cells were *C. fimi* and not contaminants.

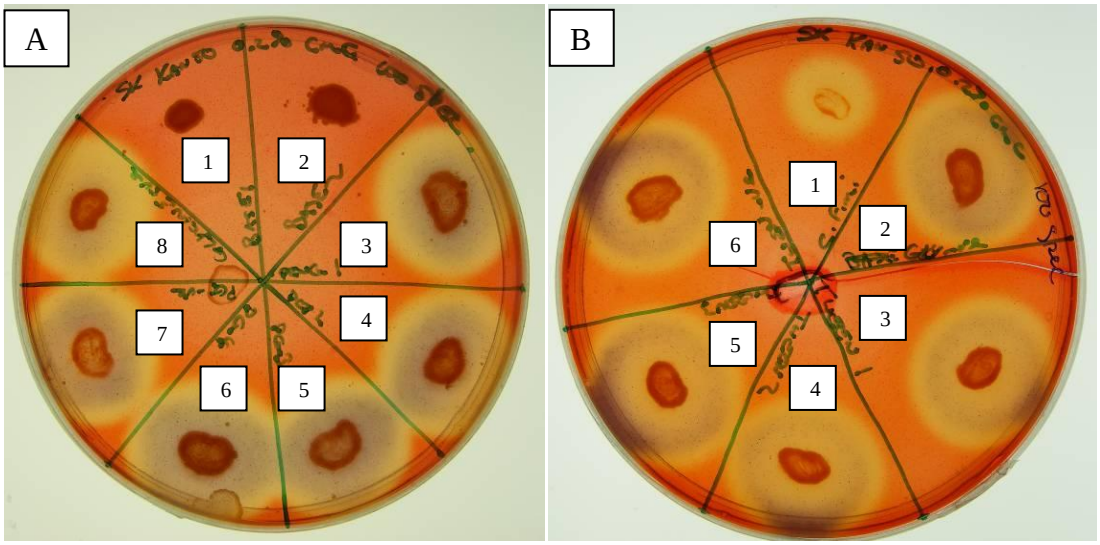


Figure 6.1. CMC plates of colonies selected from transformant plates, stained with Congo red. A) 1+2: Transformants of base plates, DNA +ve; 3-6: Transformants of pen-G plates, DNA +ve; 7: Transformant of pen-G plate, DNA –ve; 8: Transformant of gly plate, DNA +ve. B) 1: *C. fimi*; 2: Transformant of gly plate, DNA –ve; 3-5: Transformant of tween plate, DNA +ve; 6: Transformant of tween plate, DNA -ve

From these experiments it was difficult to draw any conclusions other than *C. fimi* could become naturally spec resistant, and that by comparing the colonies generated for each treatment that pen-G appeared to have the greatest potential positive effect on transformation, but that was being obfuscated by the spontaneously resistant colonies.

A secondary marker of the mCherry gene coding for red fluorescent protein was introduced into pVK168 under control of the *lacZ* promoter (P_{lac}). It was hoped that this would separate any true transformants from background resistance. Electrotransformation of *C. fimi* was again repeated using only base and pen-G cells. A 100 µl aliquot of base untransformed *C. fimi* cells with $2x10^9$ cfu/ml was spread onto 2 spec plates and the resultant colonies counted. On average 650 naturally resistant colonies were generated. Therefore, assuming that 0.02% of cells survive

the electroporation, then we could expect 100 µl of a $1x10^9$ aliquot to generate a negligible number of resistant colonies.

What we actually see however is a countable number of naturally-resistant colonies, but still far more for DNA positive transformants. Figure 6.2 displays the colony counts for base and pen-G cells electroporated with and without plasmid DNA and there is a clear swing in the favour of with DNA. This would suggest that the cells are in fact being transformed. However, no fluorescent colonies were observed and colony PCR to amplify pVK168 spec or mCherry genes all failed.

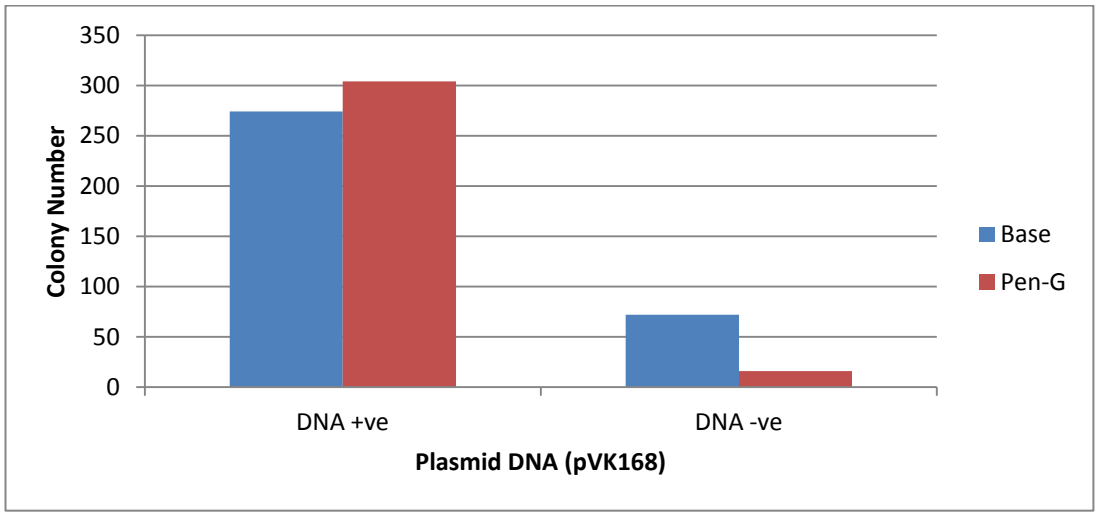


Figure 6.2. Bar chart displaying the number of colonies generated on spec plates after electroporation for either base (blue) or pen-G (red) cell preparation methods. DNA +ve: with pVK168, DNA –ve: without pVK168

6. 2 Protoplasts

The use of protoplasts would allow the employment of chemical transformation procedures which would open the way for expressionless transposome transformation methods. Transposase transformation usually requires the passing of plasmids containing the transposon and the gene coding for the transposase itself under the control of an inducible promoter [235-239]. The transposome consists of transposase complexed to the transposon, requiring no expression of the transposase *in vitro* for transposition of DNA to occur (e.g. EZ-Tn*5* kits produced by Epicentre). The transposase is stored and used in a high salt buffer so electroporation does not

work well as a means of getting the DNA and enzymes into the cell (based on experiments with *E. coli*, DNA, transposase and protocol kindly donated by Maryia Trubitsyna, personal communications) necessitating chemical methods.

Protoplasts of *C. fimi* were generated using the method described by Rodriguez et al [107]. Pulido-Vega et al used a method from Cheng and Cohen to produce protoplasts of *C. flavigena*, and Thierbach et al used a slightly different method to produce *Corynebacter glutamicum* protoplasts, both species closely related to *C. fimi* [103-105]. The methods are quite similar and produce comparable numbers of protoplasts but the Rodriguez protocol has fewer steps, buffers and much simpler regeneration medium so this was adopted as the method of choice.

Two different regeneration methods were used in the Rodriguez paper so both were tested for their regenerative abilities by plating serial dilutions of a protoplast preparation on both. The first regeneration medium was nutrient agar (NA) with a soft agar (0.8% w/v), 0.5 M sucrose and 20 mM $MgCl_2$ top layer poured on top of the protoplasts after spreading. The second medium used the top layer of the first as the bottom layer with no additional top layer. Using the same protoplast suspension, plating on NA plus top layer resulted in $5.5x10^7$ cfu/ml being generated, and for the soft agar bottom layer $7.65x10^7$ cfu/ml was formed. Of that initial protoplast preparation $6.5x10^6$ cfu/ml were osmotically stable cells, cells which still retain a large enough proportion of their cell wall, able to survive without an osmotic stabiliser, meaning a protoplast formation efficiency of over 90%. So soft agar bottom layer was adopted for the regeneration medium during experiments. The top layer also made it nearly impossible to pick single colonies as the cells were embedded into the soft agar which easily came away in large lumps.

Protoplast transformations by direct DNA transfer were carried out following the Pulido-Vega protocol. As *C. fimi* is naturally resistant to spectinomycin, pTG262 was used for the transformations. Once again however, no growth was observed on any plates for either PEG mediated or electro transformation of protoplasts. It could still be possible though that the plasmid DNA is not getting into the cells. Another

method of introducing DNA is through protoplast fusions between different organisms and can result in genomic DNA transfer conferring new attributes to an organism [108, 240-243]. Here I tried to fuse *E. coli* JM109 protoplasts containing either Edinbrick1 or pSB1A2_*cenA*. The presence of *cenA* could lead to homologous recombination at the *cenA* locus, removing the problem of plasmid replication not occurring.

For fusions involving Edinbrick1 no growth was observed on amp regeneration plates. For those fusions with *cenA* in the plasmids 158 colonies were generated in the presence of ampicillin. It was possible that these were *E. coli*. Colonies were off-white, and Gram stained as small negative rods, both of which *C. fimi* can look like (personal experience). These colonies being *E. coli* however does not account for the lack of growth observed for Edinbrick1 cells on the same plate medium, or the slow growth rate. The patched colonies were also able to grow on kan containing plates. This would suggest that either the plasmids are being expressed in *C. fimi* or homologous recombination may have occurred. Colony PCR using amp^r primers was not carried out to confirm this however.

Discussion

From looking at related species, it would appear that transformation of *C. fimi* should be possible. *C. flavigena* has been transformed by Alemohammad and Montes-Horcasitas using plasmids from *Brevibacterium lactofermentum* and *Bacillus thuringiensis* var *kurstaki*, respectively [223, 224]. *Corynebacterium glutamicum*, *C. pseudotuberculosis* and *C. diptheriae* were transformed by electroporation using plasmids derived from *Corynebacteria* [231-233, 244, 245]. *Gordonia* species have been transformed by conjugation and electroporation using plasmids derived from *Rhodococcus* [229, 246] and *G. westfalica* [247] which was also shown to transform and replicate in *Mycobacterium* species and other *Gordonia* species. Other *Actinomycete* species which have been transformed include *Streptomyces* (protoplast, electroporation, conjugation), *Arthrobacter* (electroporation) and *Brevibacterium* (electroporation), but this is not an exhaustive list [102, 248-252].

From these experiments it cannot be determined conclusively that *C. fimi* can or cannot be transformed with heterologous DNA. In keeping with the related species, they do however seem to suggest that it may be possible. Plasmids pTG262 and pCJW93 failed to generate any cml or tsp resistant cells, respectively. Plasmid pVK168 transformations repeatedly generated more colonies on spec plates than were observed for plasmid negative transformation plates. Although *C. fimi* obviously can produce naturally resistant colonies, if no DNA was entering the cells, or no gene expression of the *spec*^{r} gene occurring, the numbers would be more or less the same for DNA positive and negative plates. This suggests that DNA is entering the cells conferring spec resistance.

The lack of pVK168 *spec*^{r} gene amplification by colony PCR is therefore puzzling. This may be due to selection of naturally resistant *C. fimi* colonies from the plate instead of transformants, but for them all to fail would mean not a single transformant colony was selected, which is unlikely. The chances of selecting colonies that may be naturally occurring would be 1/19 for pen-G cells and ~1/4 for base cell preparations. So by selecting 10 colonies, only one or 2 should therefore be natural *C. fimi*. The primers used however are specific for pVK168 backbone and amplify the region containing the *spec*^{r} gene. If the gene was incorporated into the genome then DNA amplification would likely fail. PCRs were not repeated with *spec*^{r} specific primers as the transformant plates were contaminated during storage and discarded prior to this realisation. Another possibility is that the plasmid allows cells enough time to become naturally resistant, then through lack of selective pressure the plasmid is lost.

If transformation is occurring then we have to think that the failure of pTG262 or pCJW93 to produce transformants is due to reasons other than transformation failure. It could be that the plasmids don't replicate or the resistance genes are not expressed in *C. fimi*. Wilkinson et al have observed that heterologous promoters work to varying degrees even within the same genus [102]. Therefore it could be that the heterologous promoters do not work in the host organism. The best remedy for this would be to use a native plasmid from *Cellulomonas fimi* or another member of the

genus, and therefore known to replicate, coupled with native promoters. Indeed the transformations mentioned previously almost all utilise plasmids found within the target species genus. However, *C. fimi* is not known to contain any plasmid DNA [147], nor has any been published for *C. flavigena* or any other *Cellulomonas* spp.

Another reason why the transformations may have failed could be due to the restriction system of *Cellulomonas fimi*. DNA is modified by methylation of specific bases and if the methylation sites of the introduced DNA are different to those of the transformation host then the DNA will be digested. A search of the genome in the same manner of that used to identify the putative polysaccharide hydrolysing genes present in *C. fimi* (chapter 3), searching for the words "restriction", "modification" or "endonuclease", revealed 8 potential genes coding for restriction-modification (RM) subunits, which were then translated into an amino acid sequence and BLAST searched to determine the most frequent hits and either add weight to or throw doubt on the initial US DOE annotation.

Celf_750 was identified as N-6 DNA methylase (with several hits for type I RM modification subunit); Celf_0039 was annotated as a pseudo gene, but matches type II RM modification subunit; Celf_0040, _3430 and _3765 were identified as type III RM restriction subunit, helicase (but several high ranking hits to type III restriction enzyme, subunit R), and DNA methylase (with type III methylase subunit appearing in the top hits), respectively; Celf_0041 was identified as an *Eco*57I restriction endonuclease (a type IV restriction enzyme) homologue; Celf_2537 and _3765 were identified as a restriction endonuclease and adenine specific DNA methyltransferase, respectively, neither having any identified type. All these systems look for specific DNA sequences and restrict those that are not methylated on at least one strand [253-259].

All the plasmid DNA was produced in and harvested from *E. coli* JM109. These strains are K12 restriction negative but modification positive. This means that all the plasmid DNA will be modified in the following ways AAC[N_6]GTGC or

GCAC[N_6]GTT, so if the host restriction system does not recognise this pattern, which is highly likely to be the case, then the introduced DNA will be degraded.

A strategy to combat this would be to clone and express the putative modification enzyme subunits and treat the DNA with these enzymes before transforming them into *C. fimi*. This strategy has been used previously and yielded highly positive results with an efficieny increase of up to 4 orders of magnitude when transforming *Bacillus amyloliquefaciens* and *B. cereus*, and the authors report the first successful transformation of *Nitrobacter hamburgensis* X14 using this technique [260].

Electroporation is not always the most effective method for transforming bacteria. The use of protoplasts has been demonstrated to be more efficient, producing more transformants. Kirchner and Tauch report that groups were able to generate 10^6 transformants per μg of DNA using protoplasts of *C. glutamicum*, but only 10^3-10^5 transformants through initial electroporation experiments, but this was later improved to 10^6 once cell competency optimisation was carried out [261]. Hwang et al also saw a 100 fold drop in transformation efficiency between the use of protoplasts and electroporation for the transformation of *Streptomyces avermitilis* [249] and others have noted similarly [262]. Protoplasts offer a realistic alternative to electroporation and may be a way of performing metabolic engineering in *C. fimi* with intra and interspecies protoplast fusions transferring traits between bacteria through chromosomal crossover [241, 242].

Transformation with pTG262 by two methods resulted in 0 colonies observed on plates. It would seem unlikely that two well established methods of introducing DNA to a cell would fail, so this would seem to confirm that pTG262 is either not replicated or the *cml*^{r} gene is not expressed in *C. fimi*. Fusing protoplasts of *C. fimi* to other bacteria does appear to be possible, with plasmid DNA at least being transferred from *E. coli*. The complete lack of growth of protoplast fusions involving Edinbrick1 only would suggest that the *E. coli* were all dead or unable to regenerate and that *C. fimi* does not grow in the presence of ampicillin. Therefore the cells growing on the amp plates when fused in the presence of pSB1A2_*cenA* were most

likely to be *C. fimi*, which is supported by their ability to grow in the presence of kanamycin which JM109 cells cannot. This also suggests that the pSB1A2 plasmid is unable to replicate in *C. fimi* and homologous recombination has most likely occurred.

The limited evidence from these experiments suggests that *C. fimi* may be susceptible to transformation by either electroporation or PEG mediated transformation of protoplasts. The high G+C content of the *C. fimi* genome would mean that plasmids and antibiotic resistance genes from lower G+C bacteria may not be replicated or expressed. The lack of endogenous plasmids available hampers efforts somewhat, but plasmids from related high G+C bacteria such as *Rhodococcus* or *Brevibacterium* are the most likely to work.

CHAPTER 7: SUMMARY AND FUTURE WORK

The premise of this research was to explore the genome of *Cellulomonas fimi*, a cellulolytic soil bacterium, and identify potential novel genes that could code for unknown polysaccharide hydrolysing enzymes. This was done for the following reasons: dwindling oil reserves are forcing us to look for renewable alternatives to meet our petrochemical needs for transport, fuel and production of many oil by-products. Biofuels, particularly bioethanol, have been touted previously as a viable source but current production is in direct conflict of food production. Ethanol is currently produced on a wide scale from the fermentation of corn starch, taking food crops and diverting their products and land from human consumption to bioethanol production, leading to increasing food prices and the need to find new land for crop growth. Lignocellulosic biomass, a by-product of agriculture, land management and several industries including paper manufacture and the textile industry could be a rich source of sugar for the production of bioethanol, through the sugars trapped in the cell walls of plants, namely cellulose and hemicellulose. Plant material, and particularly cellulose, is recalcitrant to hydrolysis often requiring harsh treatments adding to the expense of biofuel production.

Cellulomonas fimi ATCC484 has been known and studied for decades and has already been the source of several cellulolytic enzymes including endoglucanases, cellobiohydrolases and endoxylanases (introduction). Damian Barnard, a previous PhD. student, has investigated these characterised enzymes, created BioBricks of them and looked for synergistic behaviour between them. There was some observed synergy between these enzymes, but not to such a degree as to fully explain the cellulolytic ability of *C. fimi*. It was decided therefore that further study of the genome of *C. fimi* was necessary, to identify potential new cellulolytic enzymes not yet identified. The aim of my work therefore was through study of the genome of *C. fimi* to identify putative novel genes of polysaccharide degrading enzymes, to clone these as BioBricks, assay for common cellulolytic functionality and then to characterise those identified further.

Within this thesis:

Using the methodology described in section 2.1, a host of over 90 putative polysaccharide and sugar hydrolysing genes were identified. Previously, enzymes from *C. fimi* had been identified through screening of genetic libraries and looking specifically for endoglucanase [109, 119, 120], or endoxylanase [79, 121] activity. Growth of *C. fimi* on chitin was also used to enrich for and identify potential chitinase enzymes [114]. Adsorption to a specific sugar substrate of galactomannan was also used as an enrichment procedure generating a further 2 enzymes with mannanase and mannosidase activity [78]. Extracellular enzymes were looked for after growing *C. fimi* in liquid before passing the culture supernatant through a sephadex column, eluting bound proteins and looking specifically for endoglucanase activity [111, 120]; or those that bound to avicel, BMCC or sephadex and analysing by SDS-PAGE [76, 77, 112]. Lastly dyes were used to specifically bind to and enrich for retaining endoglucanase enzymes which were then identified using mass spectrometry [80].

These methods therefore would only ever identify enzymes that are easily and highly expressed in *E. coli* (genomic library screens), that can bind to the sugar membranes used (sephadex, avicel, BMCC, galactomannan), that are expressed to a high enough level to be detected under a small range of sugar substrates (chitin, CMC, avicel) or those enzymes that are specifically retaining endoglucanases able to bind the specific dye. Only a small subset of enzymes ever would be identified using these methods. Gold et al and others have clearly shown that the expression profile of a cellulolytic organism, aerobic or anaerobic, is highly changeable depending on the growth substrate [116, 124, 130, 263]. The level of genes identified in other cellulolytic organisms of scientific interest has also shown these genomes always contain the potential for the identification of a large variety of cellulolytic enzymes [115, 116, 118, 129, 130].

Results from this study would suggest that these approaches, although successful, are missing large swathes of potentially useful and commercially viable enzymes. Using the genome released by the US DOE, I was able to identify genes that may code for

further cellulolytic enzymes that have not been identified using the methods outlined above. These include 12 further putative endoglucanases, a cellobiohydrolase, 3 endoxylanases, 2 mannosidases, 10 β-glucosidases, a chitinase, 6 β-xylosidases, 5 arabinofuranosidases, 4 pectate lyases and other enzymes necessary for the complete hydrolysis of cellulosic biomass. These cover a range of glycoside hydrolase families, which may act synergistically, but many of which will only be expressed under certain conditions and specific substrates. It also became clear that the catalytic enzymes were not acting alone but were often located in potential operons with sugar binding and transporting membrane proteins, similar to gene clusters identified in other species. This suggests that sugar transport plays a vital role in the efficient utilisation of the sugars released by biomass hydrolysis, which in the construction of a viable consolidated bioprocessor would be necessary.

The preliminary genome annotations were however, based on sequence similarity to other genes in the NCBI database which are predominantly automatically annotated genome sequencing projects, so are likely rife with mis-annotations. As such the next stage was the cloning and functional characterising of some of these identified putative enzymes in *E. coli* JM109 and *C. freundii* NCIMB 11490 (chosen for its relatedness to *E. coli* and native ability to utilise cellobiose). A selection of the genes were cloned by PCR as BioBricks and inserted into pSB1A2 under P_{lac} control complimented with a strong synthetic RBS (BBa_B0034), for inducible expression of the recombinant gene of interest. This allowed me to screen these genes for any that were functional in *E. coli* without the need for further modification of the sequence (eg. codon optimisation, removal/adding of domains). This strategy for rapid screening of genes resulted in the identification of several novel enzymes not previously identified within *C. fimi*.

Fifteen putative genes were cloned by myself in this study, covering novel putative cellulases (*celA, celE, celD, celF, celN*), xylanases (*xynF, xynG, xynH*), xylosidases (*bxyC, bxyD, bxyE, bxyF*) and arabinofuranosidases (*afsA, afsB, afsC*); the latter functionality being previously unreported in *C. fimi*. This study reports that of these, using the methods outlined, in *E. coli* functionality was demonstrated for AfsA and

AfsB for arabinofuranosidase activity, XynF was positive for cellobiohydrolase activity, AfsA, AfsB, BxyF, XynF and XynH were all positive for β-xylosidase activity, and XynF proved positive for endoglucanase activity. In *C. freundii* these activities were confirmed, but additional activity was seen for CelD as a cellobiohydrolase, β-xylosidase and endoxylanase, CelA displayed endoglucanase activity, and BxyE, CelE and XynH also showed some endoxylanase activity.

This study is the first time these genes have been expressed and screened for specific activities. The results, that functional enzymes are able to be expressed as BioBricks from *C. fimi* in *E. coli* is indicative that this method of rapid screening could be applied to other organisms as a quick, first round of screening and discerning the genuine activities of targeted genes of interest once identified by genome interrogation. Two hosts were used and typically *C. freundii* proved to be the better of the 2 for these particular enzymes, displaying a higher specific activity for assay active enzymes and displaying activity for some of the constructs not evident in *E. coli* expression. It is not clear as to why this may be with no literature available on the subject. The obvious reason of codon bias does not fully explain this observation although there is more tolerance of high-GC codons in *C. freundii* than *E. coli* (chapter 4). Translation initiation rates, as calculated by the Salis RBS calculator [187] do show a difference between species however, with *E. coli* having the higher initiation rate by factors of 10, suggesting that *E. coli* is under a greater metabolic burden although toxicity after induction was not obvious. Further studies ought to be carried out to properly understand the differences between *C. freundii* NCIMB11490 and *E.coli* JM109 with regard to these genes functionality if it is to be a viable host for a consolidated bioprocessor.

Each of the cloned genes described throughout this thesis pertain to a glycoside hydrolase family [63, 81] with each family having known functions attached. Here I present enzymes displaying activity not predicted by their GH family as outlined in the CAZy database [131]. CelA is predicted as a GH 5 family member, but exhibits xylosidase and endoxylanase activity; CelD and CelE are GH 9 yet exhibit xylosidase (CelD) and endoxylanase (both) activity; XynF is GH 10 but exhibits

xylosidase and cellobiohydrolase activity. None of the activities listed above are currently described as behaviours of members of the respective families.

On the other hand, some genes displayed no activity in either host (*bxyC, bxyD, celF, celN)* but with no clear reason as to why, but may relate to inaccurate annotation of genomes within the NCBI database [172-174]. This inaccurate annotation is propagated by the ever increasing number of sequenced organisms that are then automatically annotated by comparison to other automatically annotated genomes. There are, in proportion to putative enzyme encoding genes, very few that have actually been characterised and confirmed as annotated. The method used in this thesis could help with this regard.

For these identified, cloned and assayed enzymes to be of use either industrially or as part of a synthetic cellulolytic network, some basic parameters are useful to be known. To this end the optimum temperature and pH of AfsB, BxyF, XynF and XynH are reported, as well as the effects of potential buffer or growth medium components (metal ions, sugars) on enzyme activities. This study shows that enzymes from the same organism have the potential to be starkly and surprisingly different from the expected.

The enzymes purified fairly cleanly with little background and fold purifications of between 9 (AfsB) and 100 fold (XynF). These purified fractions exhibited some anomalous behaviour in fractions with high NaCl concentration however, unpurified crude cell lysates containing the respective enzymes showed little difference when assayed with or without NaCl present but it could not be ruled out entirely from having a potentially positive effect on activity. The enzymes were only purified so far and not to homogeneity as these enzymes' functionalities are not found in *E. coli* itself so native enzymes would not interfere with activity estimates greatly, and should if anything lead to an underestimate of the enzyme activity. It has however meant that the remaining contaminating proteins are obscuring MW estimation.

The purified enzymes were assayed first with increasing pH from 4-10.5 using different buffers but constant temperature of 37ºC, then once the optimal pH was established the optimum temperature was ascertained using the optimal pH from 20-70ºC (up to 100ºC for XynH). The results of these experiments showed interesting characteristics. AfsB showed optima of pH 6.5 and 45ºC. BxyF showed optima of pH 6 and 45ºC. XynF showed optima of pH 6 and 60ºC on ONPC, pH 4 and 50ºC on ONPX and pH 5.5 and 40ºC on RBB-xylan. The pH ranges and temperatures exhibited by these enzymes are slightly varied but in the general range of the known *C. fimi* enzymes with reported (but not necessarily optimal) pH preferences of 6-7 and temps of 37ºC.

XynH exhibited wildly different optima. Despite the apparent effects of the buffering agents on its activity, the optimal pH observed was 9, and it also had a temperature optimum of 80ºC. This is well above what might be expected from *C. fimi* given the other enzymes characterised. Only CenB has a similar predilection for alkaline conditions with a pH optimum of 8.5-9 [73]. Very little, almost no activity in fact, was observed at the lower spectrum of temperature where the initial screens for activity were tested at (37ºC). That anything was seen at all is a good indication that the assays using nitrophenol derivatives are sensitive otherwise activity may have been missed here. The optimal values of XynH does however beg the question of other enzymes within the genome, if they too are similarly inclined to the extremes of pH and temperature, or even more so, then they may appear negative in the screens. An improved strategy might be to assay for activity at alkaline, neutral and acidic pH levels and at low (20ºC), medium (45ºC) and high (70ºC) temperatures to ensure the best chance of identifying active clones.

It was demonstrated through this study that the presence of metal ions can have a dramatic effect on enzyme activity. When supplemented with $CoCl_2$, $CuSO_4$ and $ZnSO_4$ the enzymes exhibited a reduction in observed activity to nothing as was the case for AfsB and BxyF, whereas other metal ions exhibited a more muted negative effect. XynH for example was only ever reduced to 55% activity by a supplement, $CuSO_4$, which completely destroys AfsB and BxyF activity. $FeSO_4$ when added to

reaction mixtures containing XynF had a significantly positive effect on activity, as did $CaCl_2$ when added to BxyF and XynF reaction mixtures. Largely however the addition of 2^+ metal ions had a negative impact on activity, and the addition of sugar monomers, thought to be inhibitory, had no real effect at 1 mM concentrations, but obviously may at higher levels.

Vmax and Km of an enzyme are measures of their maximum theoretical activity and of their affinity for their substrates. These are useful parameters for modelling the behaviour of an enzyme in a system and for comparing different enzymes to one another. Using Eadie-Hofstee plots [191] to linearise the velocity versus substrate concentration data, the Vmax and Km of AfsB, BxyF and XynF (with ONPC as substrate) were determined. It was impossible to determine these parameters for XynH or XynF (ONPX as substrate) due to the low solubility of ONPX in the buffer solution, which led to precipitation at higher concentration levels, below those necessary to determine the kinetics. AfsB was determined to have a Vmax of 133 U/mg and Km of 6.8 mM. BxyF was found to have a Vmax of 326 U/mg and Km of 1.7 mM. XynF with ONPC as substrate was found to have a Vmax of 8942 U/mg and a Km of 58 mM.

Future work:

These figures suggest that the cloned *C. fimi* enzymes are of potential commercial use, especially when compared to those available (chapter 5). Each enzyme should work synergistically as part of larger network to help fully hydrolyse cellulosic biomass, but they have industrial uses in their own right. Xylanases have been used for improving paper quality, plant cell protoplasting, coffee mucilage liquefaction, bread dough making, and extraction of oil from subterranean mines [264]. β-xylosidases along with their function in the total saccharification of plant material, being shown to alleviate end product inhibition of xylanases, have been used in the food industry for the thinning and clarification of fruit juices, as well as in the pulping process of paper [265, 266]. Arabinofuranosidases have been used much like xylosidases as aides to saccharification of cellulosic biomass, but have also been used to fractionate pectin from beet pulp, increases the digestibility of grasses for

livestock, delignification of pulp, thinning and clarification of fruit juices and enhancing the aroma of wine through the hydrolysis of grape monoterpenyl α-$_L$-arabinofuranosylglucosides [267]. Their uses cover a wide range of environments and activities so an ever diverse industrial selection will always be favourable. Pre-treatments of plant material are typically at extremes of pH and temperature making XynH of particular industrial interest.

Further characterisation of these enzymes would be useful if turned to a commercial use. Through extending the purification process by use of ultrafiltraion, salting out or HPLC, the enzymes could be purified to homogeneity allowing analysis by native and denaturing PAGE to determine molecular weight, isoelectric point or if they act as monomers or homocomplexes. To ease purification thereafter it would be beneficial to attach a tag, for example the 6xHis tag used during this study, but the effects of adding such a tag would need to be determined. It is obvious by looking at XynH activity and the effect of different salts and metal ions that the buffer composition is of great importance. Discerning the optimal buffer/media composition for effective hydrolysis would also need to be carried out before commercialisation. The enzymes reported in this thesis have all been characterised on artificial substrates. To be demonstrably of industrial use they must be also assayed for activity on more natural complex biomass substrates. The enzymes could be incubated with their complex substrate, raw plant material, cellulose or xylan from different sources and the released sugars determined using thin layer chromatography or HPLC and reducing sugar assays such as the DNS assay.

Future work must also continue the cloning and screening of putative genes, not just in *C. fimi* but many other organisms. Once the function of a gene has been identified it will go to improve the annotation of all genomes in which a homologue is found. The presented characterised genes of this thesis can now be used in combination with other characterised lignocellulose degrading enzymes to search for synergies, and to begin to build networks capable of being expressed by a heterologous host. Damian Barnard has already demonstrated the synergisitic capabilities of XynF. By including in these networks enzymes with multiple functions, such as XynF, it may

be possible to create small efficient networks where one enzyme performs multiple roles complimentary but not overlapping with the functions of other enzymes in that network. It is from this growing assortment of cellulolytic BioBricks that a workable consolidated bioprocessor will be born.

CHAPTER 8: REFERENCES

1. Kerr, R.A., *Oil resources splitting the difference between oil pessimists and optimists.* Science, 2009. **326**(5956): p. 1048-1048.
2. Talebnia, F., D. Karakashev, and I. Angelidaki, *Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation.* Bioresource Technology. **101**(13): p. 4744-4753.
3. Maheshwari, R., *The science behind the biofuel controversy.* Current Science, 2008. **95**(5): p. 594-602.
4. Service, R.F., *Green energy another biofuels drawback: The demand for irrigation.* Science, 2009. **326**(5952): p. 516-517.
5. Searchinger, T., et al., *Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change.* Science, 2008. **319**(5867): p. 1238-1240.
6. von Blottnitz, H. and M.A. Curran, *A review of assessments conducted on bio-ethanol as a transportation fuel from a net energy, greenhouse gas, and environmental life cycle perspective.* Journal of Cleaner Production, 2007. **15**(7): p. 607-619.
7. Tilman, D., et al., *Beneficial biofuels-the food, energy, and environment trilemma.* Science, 2009. **325**(5938): p. 270-271.
8. Melillo, J.M., et al., *Indirect emissions from biofuels: how important?* Science, 2009. **326**(5958): p. 1397-1399.
9. Glazer, A.N. and H. Nikaido, eds. *Chapter 10: Biomass*. Microbial Biotechnology: Fundamentals of Applied Microbiology. 1995, W.H. Freeman and Company: New York. 327-358.
10. Himmel, M. and S. Pictaggio, eds. *Chapter 1: Our Challenge is to Aquire Deeper Understanding of Biomass Recalcitrance and Conversion*. Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy, ed. M. Himmel. 2008, Blackwell Publishing. 1-6.
11. Wackett, L.P., *Biomass to fuels via microbial transformations.* Current Opinion in Chemical Biology, 2008. **12**(2): p. 187-193.
12. Harris, P. and B. Stone, eds. *Chapter 4: Chemistry and Molecular Organisation of Plant Cell Walls*. Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy, ed. M. Himmel. 2008, Blackwell Publishing. 61-93.
13. Alberts, B., et al., eds. *Chapter 19: Cell Junctions, Cell Adhesion and the Extracellular Matrix*. Fifth ed. Molecular Biology of the Cell. 2008, Garland Science. 1195-1202.
14. Girio, F.M., et al., *Hemicelluloses for fuel ethanol: A review.* Bioresource Technology, 2010. **101**(13): p. 4775-4800.
15. Honda, Y., et al., *An investigation of the pH-activity relationships of cex, a family 10 xylanase from* Cellulomonas fimi*: Xylan inhibition and the influence of nitro-substituted aryl-beta-D-xylobiosides on xylanase activity.* Journal of Bioscience and Bioengineering, 2002. **93**(3): p. 313-317.
16. Wang, Q. and T. Xia, *Enhancement of the thermostability and hydrolytic activity of GH10 xylanase by module shuffling between* Cellulomonas fimi

Cex and Thermomonospora alba *XylA.* World Journal of Microbiology & Biotechnology, 2007. **23**(8): p. 1047-1055.
17. Carapito, R., et al., *Efficient hydrolysis of hemicellulose by a* Fusarium graminearum *xylanase blend produced at high levels in* Escherichia coli. Bioresource Technology, 2009. **100**(2): p. 845-850.
18. Carapito, R.l., et al., *Molecular basis of arabinobio-hydrolase activity in phytopathogenic fungi.* Journal of Biological Chemistry, 2009. **284**(18): p. 12285-12296.
19. Johnson, D. and R. Elander, eds. *Chapter 14: Pretreatments for Enhanced Digestability of Feedstocks*. Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy, ed. M. Himmel. 2008, Blackwell Publishing. 436-453.
20. Sousa, L.D., et al., *'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies.* Current Opinion in Biotechnology, 2009. **20**(3): p. 339-347.
21. Wyman, C.E., et al., *Coordinated development of leading biomass pretreatment technologies.* Bioresource Technology, 2005. **96**(18): p. 1959-1966.
22. Klinke, H.B., A.B. Thomsen, and B.K. Ahring, *Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.* Applied Microbiology and Biotechnology, 2004. **66**(1): p. 10-26.
23. Merino, S.T. and J. Cherry, *Progress and challenges in enzyme development for biomass utilization, in biofuels*. 2007, Springer-Verlag Berlin: Berlin. p. 95-120.
24. Klyosov, A., ed. *Cellulases of the Third Generatio*. Biochemistry and Genetics of Cellulose Degradation, ed. J. Aubert, P. Beguin, and J. Millet. 1988, Academic Press. 87-100.
25. Herpoel-Gimbert, I., et al., *Comparative secretome analyses of two* Trichoderma reesei *RUT-C30 and CL847 hypersecretory strains.* Biotechnology for Biofuels, 2008. **1**: p. 12.
26. Durrand, H., et al., eds. *Classical and molecular genetics applied to* Trichoderma reesei *for the selection of improved cellulolytic industrial strains*. Biochmeistry and Genetics of Cellulose Degradation, ed. J. Aubert, P. Beguin, and J. Millet. 1988, Acaademic Press. 135-152.
27. Lynd, L.R., et al., *Microbial cellulose utilization: Fundamentals and biotechnology.* Microbiology and Molecular Biology Reviews, 2002. **66**(3): p. 506-+.
28. Huang, C.M., et al., *Molecular-cloning and expression of multiple cellulase genes of* Ruminococcus flavefaciens *strain-186 in* Escherichia coli. Applied Microbiology and Biotechnology, 1989. **31**(3): p. 265-271.
29. Lam, T.L., R.S.C. Wong, and W.K.R. Wong, *Enhancement of extracellular production of a* Cellulomonas fimi *exoglucanase in* Escherichia coli *by the reduction of promoter strength.* Enzyme and Microbial Technology, 1997. **20**(7): p. 482-488.
30. Wong, W.K.R., et al., *The cloning, expression and characterization of a cellobiase gene encoding a secretory enzyme from* Cellulomonas biazotea. Gene, 1998. **207**(1): p. 79-86.

31. Lau, A.T.Y. and W.K.R. Wong, *Purification and characterization of a major secretory cellobiase, Cba2, from* Cellulomonas biazotea. Protein Expression and Purification, 2001. **23**(1): p. 159-166.
32. Fu, Z.B., et al., *A two-stage refinement approach for the enhancement of excretory production of an exoglucanase from* Escherichia coli. Protein Expression and Purification, 2006. **48**(2): p. 205-214.
33. Wulff, N.A., H. Carrer, and S.F. Pascholati, *Expression and purification of cellulase Xf818 from* Xylella fastidiosa *in* Escherichia coli. Current Microbiology, 2006. **53**(3): p. 198-203.
34. Andric, P., et al., *Effect and modeling of glucose inhibition and in situ glucose removal during enzymatic hydrolysis of pretreated wheat straw*. Applied Biochemistry and Biotechnology, 2010. **160**(1): p. 280-297.
35. Gomathi, D., et al., *Production of bio-ethanol from pretreated agricultural byproduct using enzymatic hydrolysis and simultaneous saccharification*. Microbiology, 2012. **81**(2): p. 201-207.
36. Zhang, W., et al., *Optimisation of simultaneous saccharification and fermentation of wheat straw for ethanol production.* Fuel, 2013. **112**: p. 331-337.
37. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update*. Current Opinion in Biotechnology, 2005. **16**(5): p. 577-583.
38. French, C.E., *Synthetic biology and biomass conversion: a match made in heaven?* Journal of the Royal Society Interface, 2009. **6**: p. 12.
39. Edwards, M.C. and J. Doran-Peterson, *Pectin-rich biomass as feedstock for fuel ethanol production.* Applied Microbiology and Biotechnology, 2012. **95**(3): p. 565-575.
40. McBride, J.E., et al., *Utilization of cellobiose by recombinant beta-glucosidase-expressing strains of* Saccharomyces cerevisiae*: characterization and evaluation of the sufficiency of expression.* Enzyme and Microbial Technology, 2005. **37**(1): p. 93-101.
41. Lee, J.H., et al., *Constitutive coexpression of* Bacillus *endoxylanase and* Trichoderma *endoglucanase genes in* Saccharomyces cerevisiae. Journal of Microbiology and Biotechnology, 2007. **17**(12): p. 2076-2080.
42. Lejeune, A., D.E. Eveleigh, and C. Colson, *Expression of an endoglucanase gene of* Pseudomonas fluorescens *var cellulosa in* Zymomonas mobilis. FEMS Microbiology Letters, 1988. **49**(3): p. 363-366.
43. Yoon, K.H., S.H. Park, and M.Y. Pack, *Transfer of* Bacillus subtilis *endo-beta-1,4-glucanase gene into* Zymomonas anaerobia. Biotechnology Letters, 1988. **10**(3): p. 213-216.
44. Brestic-Goachet, N., et al., *Transfer and expression of an* Erwinia chrysanthemi *cellulase gene in* Zymomonas mobilis. Journal of General Microbiology, 1989. **135**(4): p. 893-902.
45. Yanase, H., K. Nozaki, and K. Okamoto, *Ethanol production from cellulosic materials by genetically engineered* Zymomonas mobilis. Biotechnology Letters, 2005. **27**(4): p. 259-263.
46. Saha, B.C., *Hemicellulose bioconversion.* Journal of Industrial Microbiology & Biotechnology, 2003. **30**(5): p. 279-291.

47. van Zyl, W.H., et al., *Next-generation cellulosic ethanol technologies and their contribution to a sustainable Africa.* Interface Focus, 2011. **1**(2): p. 196-211.
48. la Grange, D.C., R. den Haan, and W.H. van Zyl, *Engineering cellulolytic ability into bioprocessing organisms.* Applied Microbiology and Biotechnology, 2010. **87**(4): p. 1195-1208.
49. Lee, W.J., et al., *Kinetic studies on glucose and xylose transport in* Saccharomyces cerevisiae. Applied Microbiology and Biotechnology, 2002. **60**(1-2): p. 186-191.
50. Ingram, L.O., et al., *Genetic-engineering of ethanol-production in* Escherichia coli. Applied and Environmental Microbiology, 1987. **53**(10): p. 2420-2425.
51. Ohta, K., et al., *Metabolic engineering of* Klebsiella oxytoca *m5a1 for ethanol-production from xylose and glucose.* Applied and Environmental Microbiology, 1991. **57**(10): p. 2810-2815.
52. Zhou, S.D. and L.O. Ingram, *Simultaneous saccharification and fermentation of amorphous cellulose to ethanol by recombinant* Klebsiella oxytoca *SZ21 without supplemental cellulase.* Biotechnology Letters, 2001. **23**(18): p. 1455-1462.
53. Zhou, S.D., F.C. Davis, and L.O. Ingram, *Gene integration and expression and extracellular secretion of* Erwinia chrysanthemi *endoglucanase CelY (celY) and CelZ (celZ) in ethanologenic* Klebsiella oxytoca *P2.* Applied and Environmental Microbiology, 2001. **67**(1): p. 6-14.
54. Liu, S.Q. and N. Qureshi, *How microbes tolerate ethanol and butanol.* New Biotechnology, 2009. **26**(3-4): p. 117-121.
55. Knight, T.F.J., *Indempotent vector design for standard assembly of biobricks.* http://hdl.handle.net/1721.1/21168, 2003.
56. Joshi, N., et al., *Novel approaches to biosensors for detection of arsenic in drinking water.* Desalination, 2009. **248**(1-3): p. 517-523.
57. Lohmueller, J., et al., *Progress toward construction and modelling of a tri-stable toggle switch in* E. coli. Synthetic Biology, IET, 2007. **1**(1.2): p. 25-28.
58. Haynes, K., et al., *Engineering bacteria to solve the Burnt Pancake Problem.* Journal of Biological Engineering, 2008. **2**(1): p. 8.
59. Shetty, R., D. Endy, and T. Knight, *Engineering BioBrick vectors from BioBrick parts.* Journal of Biological Engineering, 2008. **2**(1): p. 5.
60. Bayer, E.A., et al., *Cellulose, cellulases and cellulosomes.* Current Opinion in Structural Biology, 1998. **8**(5): p. 548-557.
61. Mitsuzawa, S., et al., *The rosettazyme: A synthetic cellulosome.* Journal of Biotechnology, 2009. **143**(2): p. 139-144.
62. Maki, M., K.T. Leung, and W.S. Qin, *The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass.* International Journal of Biological Sciences, 2009. **5**(5): p. 500-516.
63. Henrissat, B., *A classification of glycosyl hydrolases based on amino-acid-sequence similarities.* Biochemical Journal, 1991. **280**: p. 309-316.
64. Mansfield, S.D. and R. Meder, *Cellulose hydrolysis - the role of monocomponent cellulases in crystalline cellulose degradation.* Cellulose, 2003. **10**(2): p. 159-169.

65. Meyer, A.S., L. Rosgaard, and H.R. Sørensen, *The minimal enzyme cocktail concept for biomass processing.* Journal of Cereal Science, 2009. **50**(3): p. 337-344.
66. Balow, A., ed. *Chapter 59*. Second ed. The Prokaryotes, ed. A. Balow, et al. Vol. 2. 1992, Springer Verlag. 1323-1329.
67. Poulsen, O.M. and L.W. Petersen, *Growth and enzyme-production of* Cellulomonas sp *atcc-21399 on microcrystalline cellulose - effects of increasing concentration of a mineral medium.* Applied Microbiology and Biotechnology, 1989. **30**(5): p. 535-539.
68. Shen, H., et al., *Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase-a (CenA) of* Cellulomonas fimi *alters its conformation and catalytic activity.* Journal of Biological Chemistry, 1991. **266**(17): p. 11335-11340.
69. Chaudhary, P., N.N. Kumar, and D.N. Deobagkar, *The glucanases of* Cellulomonas. Biotechnology Advances, 1997. **15**(2): p. 315-331.
70. Meinke, A., et al., *Unusual sequence organization in CenB, an inverting endoglucanase from* Cellulomonas fimi. Journal of Bacteriology, 1991. **173**(1): p. 308-314.
71. Johnson, P.E., et al., *Structure of the N-terminal cellulose-binding domain of* Cellulomonas fimi *CenC determined by nuclear magnetic resonance spectroscopy.* Biochemistry, 1996. **35**(45): p. 14381-14394.
72. Tomme, P., et al., *Interaction of polysaccharides with the N-terminal cellulose-binding domain of* Cellulomonas fimi *CenC .1. Binding specificity and calorimetric analysis.* Biochemistry, 1996. **35**(44): p. 13885-13894.
73. Tomme, P., et al., *Characterization of CenC, an enzyme from* Cellulomonas fimi *with both endo- and exoglucanase activities.* Journal of Bacteriology, 1996. **178**(14): p. 4216-4223.
74. Chaudhary, P. and D.N. Deobagkar, *Purification and characterization of xylanases from* Cellulomonas sp *NCIM 2353.* Biotechnology and Applied Biochemistry, 1997. **25**: p. 127-133.
75. Notenboom, V., et al., *Exploring the cellulose/xylan specificity of the beta-1,4-glycanase Cex from* Cellulomonas fimi *through crystallography and mutation.* Biochemistry, 1998. **37**(14): p. 4751-4758.
76. Meinke, A., et al., *Cellobiohydrolase-A (CbhA) from the cellulolytic bacterium* Cellulomonas fimi *is a beta-1,4-exocellobiohydrolase analogous to* Trichoderma reesei *Cbh-II.* Molecular Microbiology, 1994. **12**(3): p. 413-422.
77. Shen, H., et al., *Cellobiohydrolase-B, a 2nd exo-cellobiohydrolase from the cellulolytic bacterium* Cellulomonas fimi. Biochemical Journal, 1995. **311**: p. 67-74.
78. Millwardsadler, S.J., et al., *Evidence for a general role for high-affinity noncatalytic cellulose-binding domains in microbial plant-cell wall hydrolases.* Molecular Microbiology, 1994. **11**(2): p. 375-382.
79. Clarke, J.H., et al., *A modular xylanase from mesophilic* Cellulomonas fimi *contains the same cellulose-binding and thermostabilizing domains as xylanases from thermophilic bacteria.* FEMS Microbiology Letters, 1996. **139**(1): p. 27-35.

80. Hekmat, O., et al., *Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry - Discovery of a novel retaining beta-1,4-glycanase in* Cellulomonas fimi. Journal of Biological Chemistry, 2005. **280**(42): p. 35126-35135.
81. Henrissat, B. and G. Davies, *Structural and sequence-based classification of glycoside hydrolases.* Current Opinion in Structural Biology, 1997. **7**(5): p. 637-644.
82. Stoll, D., H. Stalbrand, and R.A.J. Warren, *Mannan-degrading enzymes from* Cellulomonas fimi. Applied and Environmental Microbiology, 1999. **65**(6): p. 2598-2605.
83. Stoll, D., et al., *Mannanase Man26A from* Cellulomonas fimi *has a mannan-binding module.* FEMS Microbiology Letters, 2000. **183**(2): p. 265-269.
84. Le Nours, K., et al., *The structure and characterization of a modular endo-beta-1,4-mannanase from* Cellulomonas fimi. Biochemistry, 2005. **44**(38): p. 12700-12708.
85. Jing, H.Q., et al., *Production and purification of the isolated family 2a carbohydrate-binding module from* Cellulomonas fimi. Protein Expression and Purification, 2009. **64**(1): p. 63-68.
86. Carrard, G., et al., *Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(19): p. 10342-10347.
87. Jervis, E.J., C.A. Haynes, and D.G. Kilburn, *Surface diffusion of cellulases and their isolated binding domains on cellulose.* Journal of Biological Chemistry, 1997. **272**(38): p. 24016-24023.
88. Valjamae, P., G. Pettersson, and G. Johansson, *Mechanism of substrate inhibition in cellulose synergistic degradation.* European Journal of Biochemistry, 2001. **268**(16): p. 4520-4526.
89. Poon, D.K.Y., S.G. Withers, and L.P. McIntosh, *Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the* Cellulomonas fimi *xylanase Cex through NMR spectroscopic analysis.* Journal of Biological Chemistry, 2007. **282**(3): p. 2091-2100.
90. Macleod, A.M., et al., Streptomyces lividans *glycosylates an exoglucanase (Cex) from* Cellulomonas fimi. Gene, 1992. **121**(1): p. 143-147.
91. Sandercock, L.E., et al., *Degradation of cellulases in cultures of* Cellulomonas fimi. FEMS Microbiology Letters, 1996. **143**(1): p. 7-12.
92. Miller, R.C., Jr, et al., eds. Cellulomonas fimi *Cellulases and their Genes*. Biochemistry and Genetics of Cellulose Degradation, ed. J. Aubert, P. Beguin, and J. Millet. 1988, Academic Press. 235-248.
93. Langsford, M.L., et al., *The cellulase system of* Cellulomonas fimi. J Gen Microbiol, 1984. **130**(6): p. 1367-1376.
94. Fu, Z.B., et al., *Cell death caused by hyper-expression of a secretory exoglucanase in* Escherichia coli. Protein Expression and Purification, 2005. **42**(1): p. 67-77.
95. J.G. Holt, D.H.B., ed. *Bergey's Manual of Systematic Bacteriology*. 9th ed. Genome, ed. W.R. Hensyl. Vol. 2. 2000, Lippincott, Williams and Wilkins: Philadelphia.
96. Aziz, R.K., et al., *The RAST server: Rapid annotations using subsystems technology.* Bmc Genomics, 2008. **9**: p. 15.

97. Quevillon, E., et al., *InterProScan: protein domains identifier.* Nucleic Acids Research, 2005. **33**: p. W116-W120.
98. Yu, N.Y., et al., *PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes.* Bioinformatics, 2010. **26**(13): p. 1608-1615.
99. Chung, C.T., S.L. Niemela, and R.H. Miller, *One-step preparation of competent* Escherichia coli *- transformation and storage of bacterial-cells in the same solution.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(7): p. 2172-2175.
100. Gromova, I., Celis, J.E., *Protein Detection in Gels by Silver Staining: A Procedure Compatible with Mass-Spectrometry.* 3rd Edition ed. Cell Biology: A Laboratory Handbook., ed. J.E. Celis, Carter, N., Hunter, T., Simons, K., Small, J.V. and Shotton, D. 2006, New York, NY, USA: Elsevier Science. 219-223.
101. Shearman, C.A., K.L. Jury, and M.J. Gasson, *Controlled expression and structural organization of a* Lactococcus lactis *bacteriophage lysin encoded by 2 overlapping genes.* Applied and Environmental Microbiology, 1994. **60**(9): p. 3063-3073.
102. Wilkinson, C.J., et al., *Increasing the efficiency of heterologous promoters in Actinomycetes.* Journal of Molecular Microbiology and Biotechnology, 2002. **4**(4): p. 417-426.
103. Pulidovega, B., et al., *Transformation of protoplasts of* Cellulomonas flavigena. Journal of Industrial Microbiology, 1991. **8**(2): p. 137-140.
104. Chang, S. and S.N. Cohen, *High-frequency transformation of* Bacillus subtilis *protoplasts by plasmid DNA.* Molecular & General Genetics, 1979. **168**(1): p. 111-115.
105. Thierbach, G., A. Schwarzer, and A. Puhler, *Transformation of spheroplasts and protoplasts of* Corynebacterium glutamicum. Applied Microbiology and Biotechnology, 1988. **29**(4): p. 356-362.
106. Katsumata, R., et al., *Protoplast transformation of glutamate-producing bacteria with plasmid DNA.* Journal of Bacteriology, 1984. **159**(1): p. 306-311.
107. Rodriguez, H., et al., *Formation, regeneration, and fusion of protoplasts in a* Cellulomonas *strain.* Current Microbiology, 1991. **23**(5): p. 265-269.
108. Dai, M.H., et al., *Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of* Escherichia coli. Metabolic Engineering, 2005. **7**(1): p. 45-52.
109. Wong, W.K.R., et al., *Characterization and structure of an endoglucanase gene CenA of* Cellulomonas fimi. Gene, 1986. **44**(2-3): p. 315-324.
110. Oneill, G., et al., *Structure of the gene encoding the exoglucanase of* Cellulomonas fimi. Gene, 1986. **44**(2-3): p. 325-330.
111. Coutinho, J.B., et al., *Nucleotide-sequence of the endoglucanase-C gene (CenC) of* Cellulomonas fimi*, its high-level expression in* Escherichia coli*, and characterization of its products.* Molecular Microbiology, 1991. **5**(5): p. 1221-1233.
112. Meinke, A., et al., *Cellulose-binding polypeptides from* Cellulomonas fimi *- endoglucanase-D (CenD), a family a beta-1,4-glucanase.* Journal of Bacteriology, 1993. **175**(7): p. 1910-1918.

113. Shen, H., *Gene Cloning and Characterization of an Exocellobiohydrolase (CbhB) from Cellulomonas fimi.* Thesis, 1995.
114. Mayer, C., et al., *Characterization of a beta-N-acetylhexosaminidase and a beta-N-acetylglucosaminidase/beta-glucosidase from* Cellulomonas fimi. Febs Journal, 2006. **273**(13): p. 2929-2941.
115. Martinez, D., et al., *Genome sequencing and analysis of the biomass-degrading fungus* Trichoderma reesei *(syn.* Hypocrea jecorina*).* Nature Biotechnology, 2008. **26**(5): p. 553-560.
116. Gold, N.D. and V.J.J. Martin, *Global view of the* Clostridium thermocellum *cellulosome revealed by quantitative proteomic analysis.* Journal of Bacteriology, 2007. **189**(19): p. 6787-6795.
117. Raman B, P.C., Hurst GB, Rodriguez M Jr, McKeown CK, Lankford PK, Samatova NF, Mielenz JR., *Impact of pretreated Switchgrass and biomass carbohydrates on* Clostridium thermocellum *ATCC 27405 cellulosome composition: a quantitative proteomic analysis.* PLOS ONE, 2009. **4**.
118. Abt, B., et al., *Complete genome sequence of* Cellulomonas flavigena *type strain (134(T)).* Standards in Genomic Sciences, 2010. **3**(1): p. 15-25.
119. Gilkes, N.R., et al., *Mode of action and substrate specificities of cellulases from cloned bacterial genes.* Journal of Biological Chemistry, 1984. **259**(16): p. 455-459.
120. Moser, B., et al., *Purification and characterization of endoglucanase-C of* Cellulomonas fimi*, cloning of the gene, and analysis of* in vivo *transcripts of the gene.* Applied and Environmental Microbiology, 1989. **55**(10): p. 2480-2487.
121. Clarke, J.H., et al., *Multiple xylanases of* Cellulomonas fimi *are encoded by distinct genes.* FEMS Microbiology Letters, 1991. **83**(3): p. 305-310.
122. Rabinovich, M.L., M.S. Melnick, and A.V. Bolobova, *The structure and mechanism of action of cellulolytic enzymes.* Biochemistry-Moscow, 2002. **67**(8): p. 850-871.
123. Blume, L.R., et al., *Characterization of* Clostridium thermocellum *isolates grown on cellulose and sugarcane bagasse.* Bioenergy Research, 2013. **6**(2): p. 763-775.
124. Adav, S.S., et al., *Quantitative iTRAQ secretome analysis of cellulolytic* Thermobifida fusca. Journal of Proteome Research, 2009. **9**(6): p. 3016-3024.
125. Chundawat, S.P.S., et al., *Proteomics-based compositional analysis of complex cellulase-hemicellulase mixtures.* Journal of Proteome Research, 2011. **10**(10): p. 4365-4372.
126. Taylor, L.E., et al., *Complete cellulase system in the marine bacterium* Saccharophagus degradans *strain 2-40(T).* Journal of Bacteriology, 2006. **188**(11): p. 3849-3861.
127. Hekmat, O., et al., *Specificity fingerprinting of retaining beta-1,4-glycanases in the* Cellulomonas fimi *secretome using two fluorescent mechanism-based probes.* Chembiochem, 2007. **8**(17): p. 2125-2132.
128. Sanchez-Herrera, L.M., et al., *Differential expression of cellulases and xylanases by* Cellulomonas flavigena *grown on different carbon sources.* Applied Microbiology and Biotechnology, 2007. **77**(3): p. 589-595.
129. Christopherson, M.R., et al., *The genome sequences of* Cellulomonas fimi *and "*Cellvibrio gilvus*" reveal the cellulolytic strategies of two facultative*

*anaerobes, transfer of "*Cellvibrio gilvus*" to the genus* Cellulomonas, *and proposal of* Cellulomonas gilvus *sp nov.* PLOS One, 2013. **8**(1).
130. Lykidis, A., et al., *Genome sequence and analysis of the soil cellulolytic actinomycete* Thermobifida fusca *YX.* Journal of Bacteriology, 2007. **189**(6): p. 2477-2486.
131. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.* Nucleic Acids Research, 2009. **37**: p. D233-D238.
132. Anderson, I., et al., *Genomics of aerobic cellulose utilization systems in Actinobacteria.* PLOS ONE, 2012. **7**(6).
133. Xie, G., et al., *Genome sequence of the cellulolytic gliding bacterium* Cytophaga hutchinsonii. Applied and Environmental Microbiology, 2007. **73**(11): p. 3536-3546.
134. Levy, I. and O. Shoseyov, *Cellulose-binding domains biotechnological applications.* Biotechnology Advances, 2002. **20**(3-4): p. 191-213.
135. Boraston, A.B., et al., *A novel mechanism of xylan binding by a lectin-like module from* Streptomyces lividans *xylanase 10A.* Biochemical Journal, 2000. **350**: p. 933-941.
136. Rutenber, E., M. Ready, and J.D. Robertus, *Structure and evolution of ricin B-chain.* Nature, 1987. **326**(6113): p. 624-626.
137. Liu, G.D., et al., *An endo-1,4-beta-glucanase PdCel5C from cellulolytic fungus* Penicillium decumbens *with distinctive domain composition and hydrolysis product profile.* Enzyme and Microbial Technology, 2013. **52**(3): p. 190-195.
138. Eckert, K., et al., *Gene cloning, sequencing, and characterization of a family 9 endoglucanase (CeIA) with an unusual pattern of activity from the thermoacidophile* Alicyclobacillus acidocaldarius *ATCC27009.* Applied Microbiology and Biotechnology, 2002. **60**(4): p. 428-436.
139. Punta, M., et al., *The Pfam protein families database.* Nucleic Acids Research, 2012. **40**(D1): p. D290-D301.
140. Kulcinskaja, E., et al., *Expression and characterization of a* Bifidobacterium adolescentis *beta-mannanase carrying mannan-binding and cell association motifs.* Applied and Environmental Microbiology, 2013. **79**(1): p. 133-140.
141. Kim, J.Y., *Overproduction and secretion of* Bacillus circulans *endo-beta-1,3-1,4-glucanase gene (bglBC1) in* B. subtilis *and* B. megaterium. Biotechnology Letters, 2003. **25**(17): p. 1445-1449.
142. Tang, X.J., et al., *Medium optimization for the production of thermal stable beta-glucanase by* Bacillus subtilis *ZJF-1A5 using response surface methodology.* Bioresource Technology, 2004. **93**(2): p. 175-181.
143. Yang, M.J., et al., *Expression of a* Bacillus subtilis *endoglucanase in protease-deficient* Bacillus subtilis *strains.* Journal of Microbiology and Biotechnology, 2004. **14**(2): p. 430-434.
144. Keseler, I.M., et al., *EcoCyc: fusing model organism databases with systems biology.* Nucleic Acids Research, 2013. **41**(D1): p. D605-D612.
145. Barbe, V., et al., *From a consortium sequence to a unified sequence: the* Bacillus subtilis *168 reference genome a decade later.* Microbiology-Sgm, 2009. **155**: p. 1758-1775.

146. Avonce, N., et al., *Insights on the evolution of trehalose biosynthesis.* BMC Evolutionary Biology, 2006. **6**(1): p. 109.
147. Stoll, D., *Mapping of genes encoding glycoside hydrolases on the chromosome of* Cellulomonas fimi. Canadian Journal of Microbiology, 2001. **47**(12): p. 1063-1067.
148. Das, A. and C. Yanofsky, *Restoration of a translational stop-start overlap reinstates translational coupling in a mutant trpB'-trpA gene pair of the* Escherichia coli *tryptophan operon.* Nucleic Acids Research, 1989. **17**(22): p. 9333-9340.
149. Scherbakov, D.V. and M.B. Garber, *Overlapping genes in bacterial and phage genomes.* Molecular Biology, 2000. **34**(4): p. 485-495.
150. Lee, S.J., et al., *Coupled expression of MhpE aldolase and MhpF dehydrogenase in* Escherichia coli. Biochemical and Biophysical Research Communications, 2006. **346**(3): p. 1009-1015.
151. Miller, M.E.B., et al., *diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of* Ruminococcus flavefaciens *FD-1.* PLOS One, 2009. **4**(8).
152. Lewis, M., *The lac repressor.* Comptes Rendus Biologies, 2005. **328**(6): p. 521-548.
153. Spiridonov, N.A. and D.B. Wilson, *Characterization and cloning of CelR, a transcriptional regulator of cellulase genes from* Thermomonospora fusca. Journal of Biological Chemistry, 1999. **274**(19): p. 13127-13132.
154. Munch, R., et al., *Virtual footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes.* Bioinformatics, 2005. **21**(22): p. 4187-4189.
155. Lin, E. and D.B. Wilson, *Regulation of beta-1,4-endoglucanase synthesis in* Thermomonospora fusca. Applied and Environmental Microbiology, 1987. **53**(6): p. 1352-1357.
156. Das, M.K. and H.-K. Dai, *A survey of DNA motif finding algorithms.* BMC Bioinformatics, 2007. **8 Suppl 7**: p. S21.
157. Sun, H.C., et al., *Tmod: toolbox of motif discovery.* Bioinformatics, 2010. **26**(3): p. 405-407.
158. Gupta, S., et al., *Quantifying similarity between motifs.* Genome Biology, 2007. **8**(2).
159. Chhabra, S.R., et al., *Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium* Thermotoga maritima. Journal of Biological Chemistry, 2003. **278**(9): p. 7540-7552.
160. Weickert, M.J. and S. Adhya, *A family of bacterial regulators homologous to gal and lac repressors.* Journal of Biological Chemistry, 1992. **267**(22): p. 15869-15874.
161. Deng, Y. and S.S. Fong, *Development and application of a PCR-targeted gene disruption method for studying CelR function in* Thermobifida fusca. Applied and Environmental Microbiology, 2010. **76**(7): p. 2098-2106.
162. Khanna, S. and Gauri, *Regulation, purification, and properties of xylanase from* Cellulomonas fimi. Enzyme and Microbial Technology, 1993. **15**(11): p. 990-995.

163. PerezAvalos, O., et al., *Induction of xylanase and beta-xylosidase in* Cellulomonas flavigena *growing on different carbon sources.* Applied Microbiology and Biotechnology, 1996. **46**(4): p. 405-409.
164. Rajoka, M.I., *Regulation of synthesis of endo-xylanase and beta-xylosidase in* Cellulomonas flavigena*: a kinetic study.* World Journal of Microbiology & Biotechnology, 2005. **21**(4): p. 463-469.
165. Rajoka, M.I. and K.A. Malik, *Cellulase production by* Cellulomonas biazotea *cultured in media containing different cellulosic substrates.* Bioresource Technology, 1997. **59**(1): p. 21-27.
166. Wilson, C.M., et al., *Global transcriptome analysis of* Clostridium thermocellum *ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass.* Biotechnology for Biofuels, 2013. **6**(179).
167. Teather, R.M. and P.J. Wood, *Use of Congo red polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen.* Applied and Environmental Microbiology, 1982. **43**(4): p. 777-780.
168. Rath, A., et al., *Detergent binding explains anomalous SDS-PAGE migration of membrane proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(6): p. 1760-1765.
169. Denzinger, T., et al., *Isolation, primary structure characterization and identification of the glycosylation pattern of recombinant goldfish neurolin, a neuronal cell adhesion protein.* Journal of Mass Spectrometry, 1999. **34**(4): p. 435-446.
170. Bolanos-Garcia, V.M. and O.R. Davies, *Structural analysis and classification of native proteins from* E. coli *commonly co-purified by immobilised metal affinity chromatography.* Biochimica Et Biophysica Acta-General Subjects, 2006. **1760**(9): p. 1304-1313.
171. Andersen, K.R., N.C. Leksa, and T.U. Schwartz, *Optimized* E. coli *expression strain LOBSTR eliminates common contaminants from His-tag purification.* Proteins-Structure Function and Bioinformatics, 2013. **81**(11): p. 1857-1861.
172. Kyrpides, N.C. and C.A. Ouzounis, *Whole-genome sequence annotation: 'Going wrong with confidence'.* Molecular Microbiology, 1999. **32**(4): p. 886-887.
173. Gilks, W.R., et al., *Modeling the percolation of annotation errors in a database of protein sequences.* Bioinformatics, 2002. **18**(12): p. 1641-1649.
174. Gilks, W.R., et al., *Percolation of annotation errors through hierarchically structured protein sequence databases.* Mathematical Biosciences, 2005. **193**(2): p. 223-234.
175. Kunin, V., et al., *A bioinformatician's guide to metagenomics.* Microbiology and Molecular Biology Reviews, 2008. **72**(4): p. 557-578.
176. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* Journal of Molecular Biology, 2000. **302**(1): p. 205-217.
177. Zdobnov, E.M. and R. Apweiler, *InterProScan - an integration platform for the signature-recognition methods in InterPro.* Bioinformatics, 2001. **17**(9): p. 847-848.
178. Bendtsen, J.D., et al., *Non-classical protein secretion in bacteria.* Bmc Microbiology, 2005. **5**: p. 13.

179. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* Journal of Molecular Biology, 2004. **340**(4): p. 783-795.
180. Kane, J.F., *Effects of rare codon clusters on high-level expression of heterologous proteins in* Escherichia coli. Current Opinion in Biotechnology, 1995. **6**(5): p. 494-500.
181. Zdanovsky, A.G. and M.V. Zdanovskaia, *Simple and efficient method for heterologous expression of Clostridial proteins.* Applied and Environmental Microbiology, 2000. **66**(8): p. 3166-3173.
182. Rocha, E.P.C., *Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization (vol 14, pg 2279, 2004).* Genome Research, 2004. **14**(12): p. 2510-2510.
183. Nakamura, Y., T. Gojobori, and T. Ikemura, *Codon usage tabulated from international DNA sequence databases: status for the year 2000.* Nucleic Acids Research, 2000. **28**(1): p. 292-292.
184. Jiang, P.X., et al., *Reconstruction of the violacein biosynthetic pathway from* Duganella sp *B2 in different heterologous hosts.* Applied Microbiology and Biotechnology, 2010. **86**(4): p. 1077-1088.
185. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: The European molecular biology open software suite.* Trends in Genetics, 2000. **16**(6): p. 276-277.
186. Kudla, G., et al., *Coding-Sequence Determinants of Gene Expression in* Escherichia coli. Science, 2009. **324**(5924): p. 255-258.
187. Salis, H.M., E.A. Mirsky, and C.A. Voigt, *Automated design of synthetic ribosome binding sites to control protein expression.* Nature Biotechnology, 2009. **27**(10): p. 946-U112.
188. Farasat, I., et al., *Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria.* Molecular Systems Biology, 2014. **10**(6).
189. Lombard, V., et al., *The Carbohydrate-active Enzymes database (CAZy) in 2013.* Nucleic Acids Research, 2014. **42**(D1): p. D490-D495.
190. Sydney O. Ugwu, S.P.A., *The effect of buffers on protein conformational stability.* Pharmaceutical Technology, 2004: p. 16.
191. Eadie, G.S., *The inhibition of cholinesterase by physostigmine and prostigmine.* Journal of Biological Chemistry, 1942. **146**(1): p. 85-93.
192. Zhao, L.C., et al., *Purification and Characterization of the Recombinant Multifunctional Cellulase from* Volvariella volvacea. Food Biotechnology, 2012. **26**(2): p. 164-179.
193. Shi, P.J., et al., Paenibacillus sp *Strain E18 bifunctional xylanase-glucanase with a single catalytic domain.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3620-3624.
194. Ghatge, S.S., et al., *Characterization of modular bifunctional processive endoglucanase Cel5 from* Hahella chejuensis *KCTC 2396.* Applied Microbiology and Biotechnology, 2014. **98**(10): p. 4421-4435.
195. Fang, S., et al., *Cloning and characterization of a new broadspecific beta-glucosidase from* Lactococcus sp. *FSJ4.* World Journal of Microbiology & Biotechnology, 2014. **30**(1): p. 213-223.
196. Correa, J.M., et al., *Expression and characterization of a GH39 beta-xylosidase II from* Caulobacter crescentus. Applied Biochemistry and Biotechnology, 2012. **168**(8): p. 2218-2229.

197. Smaali, I., C. Remond, and M.J. O'Donohue, *Expression in* Escherichia coli *and characterization of beta-xylosidases GH39 and GH-43 from* Bacillus halodurans *C-125.* Applied Microbiology and Biotechnology, 2006. **73**(3): p. 582-590.
198. Rapp, P. and F. Wagner, *Production and properties of xylan-degrading enzymes from* Cellulomonas uda. Applied and Environmental Microbiology, 1986. **51**(4): p. 746-752.
199. Ballou, G.A. and J.M. Luck, *The effects of different buffers on the activity of [beta]-amylase.* Jour Biol Chem, 1941. **139**((1)): p. 233-240.
200. Chen, X.L., et al., *Effects of different buffers on the thermostability and autolysis of a cold-adapted protease MCP-01.* Journal of Protein Chemistry, 2002. **21**(8): p. 523-527.
201. Wu, C.F., et al., *GFP-visualized immobilized enzymes: Degradation of paraoxon via organophosphorus hydrolase in a packed column.* Biotechnology and Bioengineering, 2002. **77**(2): p. 212-218.
202. Geueke, B. and W. Hummel, *A new bacterial L-amino acid oxidase with a broad substrate specificity: purification and characterization.* Enzyme and Microbial Technology, 2002. **31**(1-2): p. 77-87.
203. Jensen, W.A., et al., *Stability studies on maize leaf phosphoenolpyruvate carboxylase - the effect of salts.* Biochemistry, 1995. **34**(2): p. 472-480.
204. Park, M.T., et al., *Orthophosphate anion enhances the stability and activity of endoxylanase from* Bacillus sp. Biotechnology and Bioengineering, 2001. **72**(4): p. 434-440.
205. Xue, Y.M. and W.L. Shao, *Expression and characterization of a thermostable beta-xylosidase from the hyperthermophile,* Thermotoga maritima. Biotechnology Letters, 2004. **26**(19): p. 1511-1515.
206. Lama, L., et al., *Purification and characterization of thermostable xylanase and beta-xylosidase by the thermophilic bacterium* Bacillus thermantarcticus. Research in Microbiology, 2004. **155**(4): p. 283-289.
207. Quintero, D., et al., *Isolation and characterization of a thermostable beta-xylosidase in the thermophilic bacterium* Geobacillus pallidus. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2007. **1774**(4): p. 510-518.
208. Kumar, S., C.J. Tsai, and R. Nussinov, *Factors enhancing protein thermostability.* Protein Engineering, 2000. **13**(3): p. 179-191.
209. Atchley, S.H. and J.B. Clark, *Variability of temperature, pH, and moisture in an aerobic composting process.* Applied and Environmental Microbiology, 1979. **38**(6): p. 1040-1044.
210. Verma, D., et al., *Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (mxyL) retrieved from compost-soil metagenome.* PLOS One, 2013. **8**(1).
211. Raj, A., S. Kumar, and S.K. Singh, *A highly thermostable xylanase from* Stenotrophomonas maltophilia*: purification and partial characterization.* Enzyme Research, 2013. **2013**: p. 429305.
212. Mitsuishi, Y., et al., *Purification and properties of thermostable xylanases from mesophilic fungus strain Y-94.* Agricultural and Biological Chemistry, 1987. **51**(12): p. 3207-3213.
213. Margolles, A. and C.G. de los Reyes-Gavilan, *Purification and functional characterization of a novel alpha-L-arabinofuranosidase from*

Bifidobacterium longum *B667.* Applied and Environmental Microbiology, 2003. **69**(9): p. 5096-5103.
214. Canakci, S., et al., *Purification and characterization of a highly thermostable alpha-L-Arabinofuranosidase from* Geobacillus caldoxylolyticus *TK4.* Applied Microbiology and Biotechnology, 2007. **75**(4): p. 813-820.
215. Beylot, M.H., et al., *The* Pseudomonas cellulosa *glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity.* Biochemical Journal, 2001. **358**: p. 607-614.
216. Souza, T., et al., *Structure of a novel thermostable GH51 alpha-L-arabinofuranosidase from* Thermotoga petrophila *RKU-1.* Protein Science, 2011. **20**(9): p. 1632-1637.
217. Juturu, V. and J.C. Wu, *Insight into microbial hemicellulases other than xylanases: a review.* Journal of Chemical Technology and Biotechnology, 2012. **88**(3): p. 353-363.
218. Czjzek, M., et al., *Enzyme-substrate complex structures of a GH39 beta-xylosidase from* Geobacillus stearothermophilus. Journal of Molecular Biology, 2005. **353**(4): p. 838-846.
219. Canakci, S., et al., *Cloning, purification, and characterization of a thermostable alpha-L-arabinofuranosidase from* Anoxybacillus kestanbolensis *AC26Sari.* Applied Microbiology and Biotechnology, 2008. **81**(1): p. 61-68.
220. Saha, B.C. and R.J. Bothast, *Purification and characterization of a novel thermostable alpha-L-arabinofuranosidase from a color-variant strain of* Aureobasidium pullulans. Applied and Environmental Microbiology, 1998. **64**(1): p. 216-220.
221. Talarico, L.A., et al., *Construction and expression of an ethanol production operon in Gram-positive bacteria.* Microbiology-Sgm, 2005. **151**: p. 4023-4031.
222. Uwe Hardter, M.L., Sandra Ebeling, Andreas Bechthold *Ethanol production in* Actinomycetes *after expression of synthetic adhB and pdc.* Open Journal of Biotechnology, 2012. **6**: p. 13-16.
223. Alemohammad S.J., P.J.T., *Transformation of the coryneform bacterium* Cellulomonas flavigena *with plasmid DNA via electroporation.* Biotechnology Techniques, 1990. **4**(2): p. 147-148.
224. Montes-Horcasitas, C., et al., *Efficient transformation of* Cellulomonas flavigena *by electroporation and conjugation with* Bacillus thuringiensis. Current Microbiology, 2004. **49**(6): p. 428-432.
225. Murakami, T., T.G. Holt, and C.J. Thompson, *Thiostrepton-induced gene-expression in* Streptomyces lividans. Journal of Bacteriology, 1989. **171**(3): p. 1459-1466.
226. Bone, E.J. and D.J. Ellar, *Transformation of* Bacillus thuringiensis *by electroporation.* FEMS Microbiol Lett, 1989. **49**(2-3): p. 171-7.
227. Han, Y.W., et al., *Sonoporation is an efficient tool for intracellular fluorescent dextran delivery and one-step double-crossover mutant construction in* Fusobactetium nucleatum. Applied and Environmental Microbiology, 2007. **73**(11): p. 3677-3683.

228. Aune, T.E.V. and F.L. Aachmann, *Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed.* Applied Microbiology and Biotechnology, 2010. **85**(5): p. 1301-1313.
229. Arenskotter, M., et al., *Identification and application of plasmids suitable for transfer of foreign DNA to members of the genus* Gordonia. Applied and Environmental Microbiology, 2003. **69**(8): p. 4971-4974.
230. Buckley, N.D., et al., *An effective strategy, applicable to* Streptococcus salivarius *and related bacteria, to enhance or confer electroporation competence.* Applied and Environmental Microbiology, 1999. **65**(12): p. 5642-5642.
231. Dorella, F.A., et al., *An improved protocol for electrotransformation of* Corynebacterium pseudotuberculosis. Veterinary Microbiology, 2006. **114**(3-4): p. 298-303.
232. Haynes, J.A. and M.L. Britz, *Electrotransformation of* Brevibacterium lactofermentum *and* Corynebacterium glutamicum - *growth in Tween 80 increases transformation frequencies.* FEMS Microbiology Letters, 1989. **61**(3): p. 329-334.
233. Haynes, J.A. and M.L. Britz, *the effect of growth-conditions of* Corynebacterium glutamicum *on the transformation frequency obtained by electroporation.* Journal of General Microbiology, 1990. **136**: p. 255-263.
234. Schleife.Kh and O. Kandler, *Peptidoglycan types of bacterial cell-walls and their taxonomic implications.* Bacteriological Reviews, 1972. **36**(4): p. 407-477.
235. Ashour, J. and M.K. Hondalus, *Phenotypic mutants of the intracellular Actinomycete* Rhodococcus equi *created by* in vivo *Himar1 transposon mutagenesis.* Journal of Bacteriology, 2003. **185**(8): p. 2644-2652.
236. Felsheim, R.F., et al., *Transformation of* Anaplasma marginale. Veterinary Parasitology, 2010. **167**(2-4): p. 167-174.
237. Koledin, T., G.L. Newton, and R.C. Fahey, *Identification of the mycothiol synthase gene (mshD) encoding the acetyltransferase producing mycothiol in Actinomycetes.* Archives of Microbiology, 2002. **178**(5): p. 331-337.
238. Petzke, L. and A. Luzhetskyy, In vivo *Tn5-based transposon mutagenesis of Streptomycetes.* Applied Microbiology and Biotechnology, 2009. **83**(5): p. 979-986.
239. Siegl, T. and A. Luzhetskyy, *Actinomycetes genome engineering approaches.* Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology, 2012. **102**(3): p. 503-516.
240. Imada, C., M. Okanishi, and Y. Okami, *Intergenus protoplast fusion between* Streptomyces *and* Micromonospora *with reference to the distribution of parental characteristics in the fusants.* Journal of Bioscience and Bioengineering, 1999. **88**(2): p. 143-147.
241. Michael Tyurin, M.K., Vel Berzin, *Electrofusion of cells of* Acetogen Clostridium sp. *MT 351 with erm(B) or cat in the chromosome.* Journal of Biotech Research, 2012. **4**: p. 1-12.
242. Winkler, J., M. Rehmann, and K.C. Kao, *Novel* Escherichia coli *hybrids with enhanced butanol tolerance.* Biotechnology Letters, 2010. **32**(7): p. 915-920.

243. Zhang, E., et al., *Improved method of protoplast fusion between marine luminescent bacterium and* Escherichia coli *cultivated with fatty acid.* Fisheries Science, 2008. **74**(6): p. 1290-1296.
244. Tauch, A., et al., *Efficient electrotransformation of* Corynebacterium diphtheriae *with a mini-replicon derived from the* Corynebacterium glutamicum *plasmid pGA1.* Current Microbiology, 2002. **45**(5): p. 362-367.
245. Tsuchida, Y., et al., *Characterization of a new 2.4-kb plasmid of* Corynebacterium casei *and development of stable corynebacterial cloning vector.* Applied Microbiology and Biotechnology, 2009. **81**(6): p. 1107-1115.
246. Shavandi, M., et al., *Biodesulfurization of dibenzothiophene by recombinant* Gordonia alkanivorans *RIPI90A.* Bioresource Technology, 2009. **100**(1): p. 475-479.
247. Broker, D., M. Arenskotter, and A. Steinbuchel, *Transfer of megaplasmid pKB1 from the rubber-degrading bacterium* Gordonia westfalica *strain Kb1 to related bacteria and its modification.* Applied Microbiology and Biotechnology, 2008. **77**(6): p. 1317-1327.
248. Neesen, K. and G. Volckaert, *Construction and shuttling of novel bifunctional vectors for* Streptomyces spp *and* Escherichia coli. Journal of Bacteriology, 1989. **171**(3): p. 1569-1573.
249. Hwang, Y.S., et al., *Optimization of transformation procedures in avermectin high-producing* Streptomyces avermitilis. Biotechnology Letters, 2001. **23**(6): p. 457-462.
250. Miteva, V., S. Lantz, and J. Brenchley, *Characterization of a cryptic plasmid from a Greenland ice core* Arthrobacter *isolate and construction of a shuttle vector that replicates in psychrophilic high G+C Gram-positive recipients.* Extremophiles, 2008. **12**(3): p. 441-449.
251. Amador, E., J.F. Martin, and J.M. Castro, *A* Brevibacterium lactofermentum *16S rRNA gene used as target site for homologous recombination.* FEMS Microbiology Letters, 2000. **185**(2): p. 199-204.
252. Veiga-Crespo, P., et al., *Proposal of a method for the genetic transformation of* Gordonia jacobaea. Journal of Applied Microbiology, 2006. **100**(3): p. 608-614.
253. Janulaitis, A., et al., *Purification and properties of the Eco57I restriction endonuclease and methylase - prototypes of a new class (type-IV).* Nucleic Acids Research, 1992. **20**(22): p. 6043-6049.
254. Murray, N.E., *Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle).* Microbiology and Molecular Biology Reviews, 2000. **64**(2): p. 412-+.
255. Xu, Q., et al., *Identification of type II restriction and modification systems in* Helicobacter pylori *reveals their substantial diversity among strains.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(17): p. 9671-9676.
256. Lepikhov, K., et al., *Characterization of the type IV restriction modification system BspLU11III from* Bacillus sp *LU11.* Nucleic Acids Research, 2001. **29**(22): p. 4691-4698.
257. Tock, M.R. and D.T.F. Dryden, *The biology of restriction and anti-restriction.* Current Opinion in Microbiology, 2005. **8**(4): p. 466-472.

258. Rao, D.N., D.T.F. Dryden, and S. Bheemanaik, *Type III restriction-modification enzymes: a historical perspective.* Nucleic Acids Research, 2014. **42**(1): p. 45-55.
259. Oliveira, P.H., M. Touchon, and E.P.C. Rocha, *The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts.* Nucleic Acids Research, 2014. **42**(16): p. 10618-U803.
260. Zhang, G., et al., *A Mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria.* PLOS Genetics, 2012. **8**(9).
261. Kirchner, O. and A. Tauch, *Tools for genetic engineering in the amino acid-producing bacterium* Corynebacterium glutamicum. Journal of Biotechnology, 2003. **104**(1-3): p. 287-299.
262. Yoshida, N. and M. Sato, *Plasmid uptake by bacteria: a comparison of methods and efficiencies.* Applied Microbiology and Biotechnology, 2009. **83**(5): p. 791-798.
263. Zhang, Y.H.P. and L.R. Lynd, *Regulation of cellulase synthesis in batch and continuous cultures of* Clostridium thermocellum. Journal of Bacteriology, 2005. **187**(1): p. 99-106.
264. Juturu, V. and J.C. Wu, *Microbial xylanases: Engineering, production and industrial applications.* Biotechnology Advances, 2012. **30**(6): p. 1219-1227.
265. Yang, X., et al., *Two xylose-tolerant GH43 bifunctional beta-xylosidase/alpha-arabinosidases and one GH11 xylanase from* Humicola insolens *and their synergy in the degradation of xylan.* Food Chemistry, 2014. **148**: p. 381-387.
266. Jordan, D.B. and K. Wagschal, *Properties and applications of microbial beta-D-xylosidases featuring the catalytically efficient enzyme from* Selenomonas ruminantium. Applied Microbiology and Biotechnology, 2010. **86**(6): p. 1647-1658.
267. Saha, B.C., *Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology.* Biotechnology Advances, 2000. **18**(5): p. 403-423.

pSortB	gene	designation	start	stop	length	strand	DOE product
Extracellular	cenB	Celf_0019	19822	22959	3138	-	GH9
Extracellular	celD	Celf_0045	52023	54761	2739	-	GH9
Cytoplasmic	bglK	Celf_0080	96472	97689	1218	+	GH3
Extracellular	xynF	Celf_0088	103525	104919	1395	+	endo-1,4-beta xylanase
Cytoplasmic	eglA	Celf_0089	105034	105786	753	+	GH16
Cytoplasmic	bglJ	Celf_0140	155684	157960	2277	-	GH3
Unknown	celG	Celf_0233	257060	258901	1842	-	Cellulase
Extracellular	cbpD	Celf_0270	302635	303702	1068	-	chitin binding 3 and CBM_2
Unknown	gamB	Celf_0293	326384	328348	1965	-	GH15 related protein
Extracellular	rglA/celK	Celf_0345	387514	389727	2214	+	cellulose binding family 2
Extracellular	xynD	Celf_0374	416010	417947	1938	-	GH11
Unknown	*celA*	Celf_0376	418857	420560	1704	-	Cellulase
Extracellular	*pelB*	Celf_0398	437717	438988	1272	-	pectate lyase
Extracellular	*cbpB/celH*	Celf_0403	442658	443869	1212	-	cellulose binding family 2
Cyto memb	*xfeA/celI*	Celf_0404	444115	445416	1302	+	cellulose binding family 2
Extracellular	*agaD*	Celf_0438	476010	477695	1686	-	ricin b lectin
Unknown	*bglE*	Celf_0468	505436	507964	2529	+	GH3
Cytoplasmic	*bgaC*	Celf_0526	578937	580982	2046	-	beta galactosidase
Extracellular	*xynC*	Celf_0574	633960	638012	4053	+	endo-1,4-beta xylanase
Extracellular	*bglD*	Celf_0583	647221	652458	5238	+	GH3
Cyto memb	*gamA*	Celf_0635	699847	701397	1551	-	GH15 related protein
Extracellular	*xynJ*	Celf_0788	865594	869244	3651	+	GH43
Extracellular	*pelD*	Celf_0807	885253	887370	2118	+	pectate lyase
Unknown	*manD*	Celf_0862	945722	948754	3033	-	mannan endo-1,4-beta mannosidase

Unknown	*bgaA*	Celf_0855	972316	974592	2277	+	beta galactosidase
Unknown	*bgaB*	Celf_0856	972316	976649	4334	+	beta galactosidase
Unknown	*bxyE*	Celf_0899	989955	991277	1323	-	GH43
Cytoplasmic	*afsB*	Celf_0903	994923	996431	1509	+	alpha-L-arabinofuranosidase domain protein
Unknown	*araB*	Celf_0904	996433	997542	1110	+	GH43
Extracellular	*pulA*	Celf_1126	1222768	1228689	5922	+	alpha-1,6-glucosidase pullanase type
Unknown	*celL*	Celf_1230	1339431	1340732	1302	-	1,4-beta cellobiohydrolase
Extracellular	*Cex*	Celf_1271	1381472	1383004	1533	-	GH10
Cytoplasmic	*aamD*	Celf_1276	1386521	1388170	1650	+	Alpha-amylase catalytic region
Unknown	*axeD*	Celf_1329	1447824	1448933	1110	+	cellulose binding family 2
Extracellular	*pelC*	Celf_1340	1463945	1465264	1320	-	pectate lyase
Extracellular	*celF*	Celf_1481	1612769	1614379	1611	+	GH9
Unknown	*araA*	Celf_1482	1614607	1616976	2370	+	GH43
Extracellular	*cenC*	Celf_1537	1678243	1681548	3306	+	GH9
Extracellular	*celE*	Celf_1705	1874747	1878067	3321	+	GH9
Extracellular	*xynG*	Celf_1729	1900401	1901690	1290	+	GH10
Cytoplasmic	*bxyF*	Celf_1744	1923428	1926118	2691	-	GH39
Unknown	*bxyD*	Celf_1745	1926115	1927749	1635	-	GH43
Cytoplasmic	*afsD*	Celf_1746	1927746	1929290	1545	-	alpha-N-arabinofuranosidase
Extracellular	*celJ/axeA*	Celf_1754	1938086	1939582	1497	-	cellulose binding family 2
Extracellular	*cbpA/celB*	Celf_1913	2116913	2119669	2757	-	cellulose binding family 2
Extracellular	*cenD*	Celf_1924	2128016	2130259	2244	-	GH5
Extracellular	*cbhA*	Celf_1925	2130389	2133007	2619	-	1,4-beta cellobiohydrolase
Unknown	*bglC*	Celf_2053	2270790	2273537	2748	-	GH3
Unknown	*manB*	Celf_2091	2316225	2318174	1950	+	PKD domain containing protein

Cytoplasmic	*aamE*	Celf_2232	2471592	2473304	1713	-	alpha amylase
Extracellular	*aamC*	Celf_2283	2523568	2525229	1662	-	alpha amylase
Cytoplasmic	*aamI*	Celf_2311	2550032	2551873	1842	-	alpha amylase catalytic domain
Cytoplasmic	*aamB*	Celf_2316	2557214	2558959	1746	+	alpha amylase
Extracellular	*aamG*	Celf_2317	2559046	2560545	1500	-	alpha amylase
Extracellular	*celN*	Celf_2403	2665432	2666949	1518	+	ricin b lectin, pfam:GH5
Cytoplasmic	*bgaI*	Celf_2624	2921643	2922428	786	+	GH16
Cytoplasmic	*aamA*	Celf_2714	3010411	3012129	1719	-	alpha amylase
Cytoplasmic	*agaB*	Celf_2718	3015311	3017533	2223	-	extracellular solute binding protein 1, alpha galactosidase
Cytoplasmic	*bglG/bxyB*	Celf_2726	3026808	3029090	2283	-	GH3
Cytoplasmic	*aamF*	Celf_2761	3068809	3071085	2277	+	alpha amylase
Cytoplasmic	*agaA*	Celf_2769	3079662	3081848	2187	-	alpha galactosidase
Cytoplasmic	*manA*	Celf_2770	3082006	3084534	2529	+	GH2 immunoglobulin, 2nd domain periplasmic binding/LacI trascriptional regulation
Cyto/membrane	*aamH*	Celf_2772	3085530	3087722	2193	+	1-4, alpha glucan branching enzyme
Cytoplasmic	*bgaK*	Celf_2783	3105074	3106528	1455	-	beta galactosidase
Unknown	*aglB*	Celf_2784	3106752	3109034	2283	-	GH31
Cytoplasmic	*bgaE*	Celf_2785	3109031	3111049	2019	-	beta galactosidase, GH31
Unknown	*bglH*	Celf_2791	3118765	3121623	2859	+	GH3
Cytoplasmic	*nagC*	Celf_2792	3121654	3122916	1263	-	N-acetylglucosamine-6-phosphate deacetylase
Extracellular	*cbpC/celM*	Celf_2856	3189989	3192286	2298	-	cellulose binding family 2
Cytoplasmic	*axeC*	Celf_2961	3303494	3304525	1032	-	Cephalosporin-C deacetylase (pfam: acetyl xylan esterase)
Cytoplasmic	*nagA*	Celf_2983	3328711	3330405	1695	-	beta phosphoglucomutase, GH3
Cytoplasmic	*nagB*	Celf_3081	3453602	3455092	1491	-	GH20

Extracellular	*eglD*	Celf_3113	3490179	3491453	1275	-	GH16
Extracellular	*afsE*	Celf_3155	3533569	3535071	1503	-	Alpha-N-arabinofuranosidase, ricin be lectin, GH62?
Extracellular	*cfx(xynE)*	Celf_3156	3535136	3536617	1482	-	endo-1,4-beta xylanase
Extracellular	*bxyG*	Celf_3157	3536721	3538832	2112	-	GH43, pfam: ricin b lectin
Extracellular	*chiA*	Celf_3161	3540963	3542621	1659	-	GH18
Unknown	*cenA*	Celf_3184	3561504	3562853	1350	-	1,4-beta cellobiohydrolase
Cytoplasmic	*afsC*	Celf_3249	3629416	3630957	1542	-	alpha-L-arabinofuranosidase domain protein
Unknown	*aglA*	Celf_3268	3652236	3654392	2157	+	alpha glucuronidase
Unknown	*xynH*	Celf_3270	3655116	3656792	1677	-	xylan-1,4-beta xylosidase
Unknown	*bgaF*	Celf_3271	3656789	3658885	2097	-	beta galactosidase, xylan 1,4-beta xylosidase
Cytoplasmic	*bglF/bxyA*	Celf_3275	3662699	3665035	2337	-	GH3
unknown	*axeB*	Celf_3277	3666435	3667427	993	+	Cephalosporin-C deacetylase (pfam: acetyl xylan esterase)
Unknown	*bgaD*	Celf_3133	3706317	3708326	2010	+	beta galactosidase
Cytoplasmic	*afsA*	Celf_3321	3717150	3718670	1521	+	alpha-L-arabinofuranosidase domain protein
Extracellular	*eglB*	Celf_3330	3727780	3729399	1620	-	glucan-1,3-beta D glucosidase
Unknown	*bglB*	Celf_3372	3772939	3775200	2262	-	GH3, KEGG: beta glucosidase like
Extracellular	*cbhB*	Celf_3400	3798755	3802027	3273	-	GH48
Extracellular	*abgA*	Celf_3434	3834673	3837402	2730	+	LPXTG-motif cell wall anchor domain (KEGG: arabinogalactan..)/GH53
Unknown	*rglB*	Celf_3440	3844437	3847022	2586	-	FG-gap repeat protein
Cyto/membrane	*eglE*	Celf_3454	3858671	3860815	2145	-	GH81
Unknown	*bglA*	Celf_3472	3879031	3881346	2316	-	GH3
Unknown	*bgaH*	Celf_3551	3977992	3980373	2382	-	GH35, pfam: beta galacrosidase
Cell wall/Ext	*manC*	Celf_3612	4044282	4045808	1527	-	GH26
Cytoplasmic	*agaC*	Celf_3623	4056186	4057523	1338	+	putative alpha galactosidase

Cytoplasmic	*bgaJ*	Celf_3624	4057701	4059452	1752	+	beta galactosidase
Unknown	*xynI*	Celf_3625	4059449	4060426	977	+	Adjusted Start
Cytoplasmic	*bgaL*	Celf_3626	4061326	4064412	3086	+	GH 2 TIM barrel
Unknown	*bgaG*	Celf_3628	4065955	4068600	2646	-	GH35, pfam: beta galacrosidase
Unknown	*bxyC*	Celf_3643	4081997	4083481	1485	+	Xylan-1,4-beta xylosidase
Cell wall/Ext	*eglC/celC*	Celf_3773	4216184	4219540	3357	-	endo-1,3(4)-beta glucanase
Cell wall/Ext	*pelA*	Celf_3775	4221739	4223307	1569	-	pectate lyase/Amb allergen

Table A1. Predicted *C. fimi* genes and putative function (aam – α-amylase, abg – arabinogalactan-endo-β-galactosidase, afs – arabinofuranosidase, aga – α-galactosidase, agl – α-glucuronidase, ara – arabinanase, axe – acetyl xylan esterase, bga – β-galactosidase, bgl – β-glucosidase, bxy – β-xylosidase, cel – cellulase, cbp – cellulose binding protein, chi – chitinase, egl – endoglucanase, gam – glucoamylase, man – mannosidase, pul – α-1,6-glucosidase pullanase type, pel – pectate lyase, rgl – rhamnogalacturonan lyase, xyn - xylanase

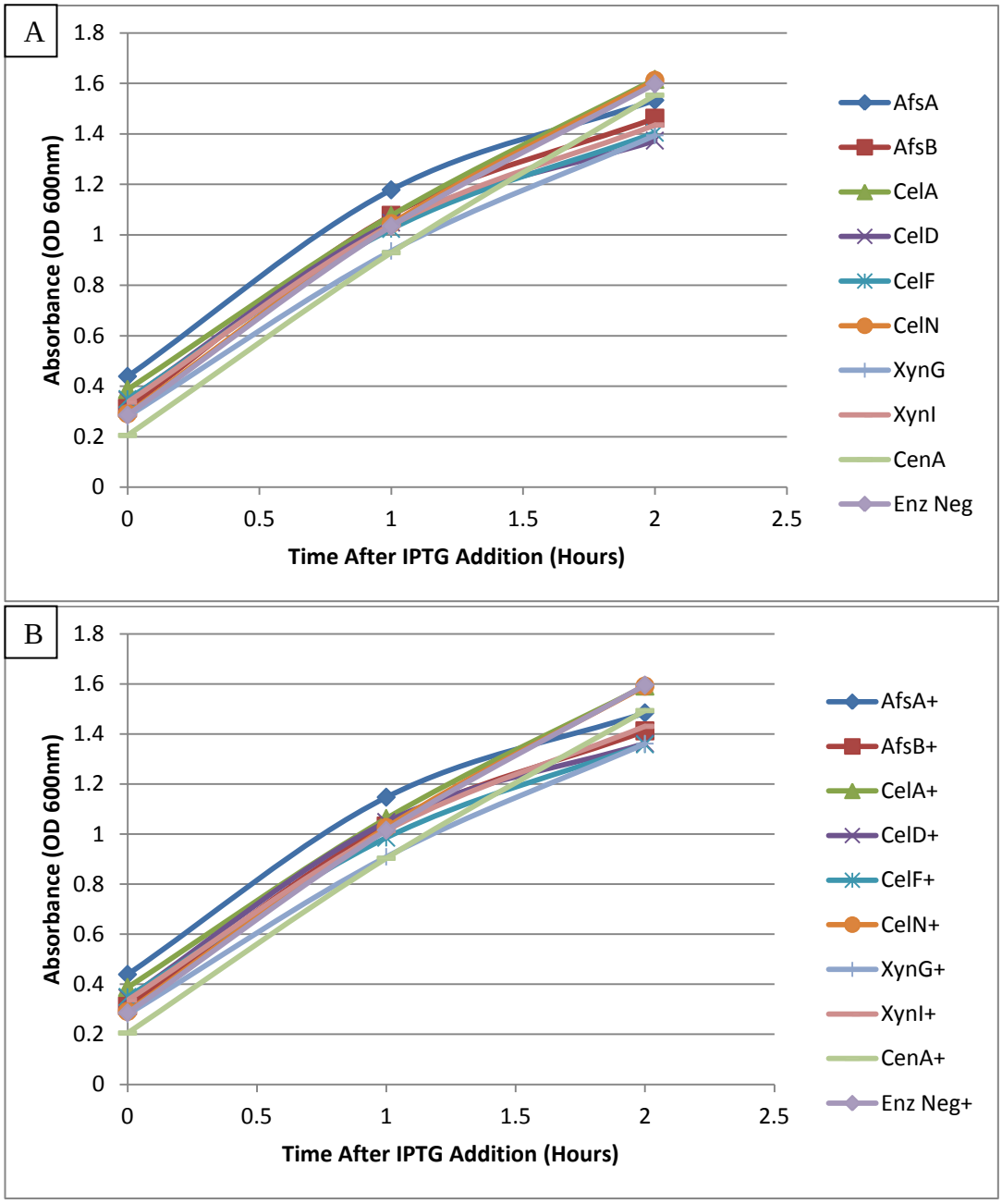


Figure A2. Effect of IPTG induced gene expression on *E. coli* growth rate. A – IPTG negative; B – IPTG positive