
THE UNIVERSITY
*of* EDINBURGH

# Development of novel systems for bioconversion of cellulosic biomass to useful products

**Kwabena Obeng Duedu**


Doctor of Philosophy (PhD)

**The University of Edinburgh**

**2015**

# Declaration

I hereby declare that the research reported in this thesis is original and was carried out by me and supervised by Prof. Christopher E. French. Work from other authors where cited have been duly acknowledged. This work has not been submitted at any institution wholly or partially for the award of any degree.

Signature:_____ Date: 24th November, 2015

**Kwabena Obeng Duedu**
BSc (Medical Lab. Science), MPhil (Microbiology), MRSB, CBiol, AFHEA

# Lay Summary

There is increasing concern regarding sustainable energy sources such as biofuels, to replace declining oil reserves. Bioethanol (ethanol produced biologically) and biodiesel (diesel produced biologically) are two major forms of fuels that are sustainable and environmentally friendly. The production of bioethanol currently is done using food crops like corn and sugarcane. This leads to increased prices and shortage of such food crops. An alternative way of producing bioethanol is using sugar that is obtained from plants. Around us, plants and plant products are common. For example, paper, wood and many other things are produced from plants and these contain the sugar needed for making bioethanol. The abundance of these raw materials makes them the only imaginable resource that can potentially substitute a substantial portion of the fossil fuels we use today, but current methods for producing biofuels from non-food crops are cost intensive and not economically viable. There is therefore the need to develop effective ways of producing the sugar needed from the paper and other plant based resources around us. This thesis explored ways of converting paper (and potentially other plant materials) into sugar for subsequent production into biofuels using engineered bacteria. It was found that, engineering bacteria with some combinations of enzymes could facilitate the conversion of paper into sugar. There is however the need to carefully screen and select which enzyme combinations should be used to engineer the bacteria.

# Abstract

There is increasing concern regarding alternative, sustainable energy sources, such as biofuels, to replace declining oil reserves. The abundance of lignocellulosic biomass makes it the only imaginable resource that can potentially substitute a substantial portion of the fossil fuels we use today, but current methods for producing biofuels from non-food crops are cost intensive and not economically viable. Synthetic biology provides several potential approaches for developing biologically mediated processes for the conversion of lignocellulosic biomass into biofuels. Such systems are based on engineered microbes that produce enzymes for catalysing the conversion of cellulose into fermentable sugars and subsequently into high value products. Effective degradation of cellulose requires multiple classes of enzyme working together. In naturally occurring cellulose degrading microbes, bioconversion is catalysed by a battery of enzymes with different catalytic properties. However, naturally occurring cellulases with multiple catalytic domains seem to be rather rare in known cellulose-degrading organisms.

Using synthetic biology approaches, seven cellulases with multiple catalytic domains were engineered and tested to determine the usefulness of such chimeric enzymes to replace cloning of multiple enzymes for biomass conversion. Catalytic domains were taken from *Cellulomonas fimi* endoglucanases CenA, CenB and CenD, exoglucanase Cex, and β-glucosidase, Cfbglu as well as *Cytophaga hutchinsonii* cellodextrinase CHU2268. All fusions retained both catalytic activities of the parental enzymes. To investigate the benefits of fusion, *Citrobacter freundii* NCIMB11490 was transformed with either fused or non-fused enzymes and cultured with cellulose blotting papers as main carbon source.

Cells expressing fusions of Cex with CenA or CenD reproducibly showed higher growth than cells expressing non-fused versions, as well as more rapid physical destruction of paper. The opposite was observed for the other combinations. Comparing two different Cex and CenA fusions, CxnA2, which contains two carbohydrate binding modules (CBMs), degraded filter paper faster and led to better growth than CxnA1, which contains only one CBM. It was observed that CxnA1 was exported to the supernatant of *E. coli* and *C. freundii* cultures, as also seen for Cex and CenA, although there is no clear biological mechanism for this.

Monitoring of growth using colony counts is laborious, but the use of optical density is not possible for cellulose-based cultures as it is affected by the insoluble cellulose particles. The SYBR Green I/propidium iodide live/dead staining protocol was therefore evaluated for growth measurements and was found to allow rapid measurements of large numbers of samples.

In conclusion, these studies have demonstrated a simple and useful method for making chimeric proteins from libraries of multiple parts. The results demonstrate that use of fusion proteins can improve biomass conversion *in vivo*, and could potentially reduce the necessity for cloning of multiple enzymes and improve product yields. A simple and effective method for monitoring growth of bacteria in turbid cultures using a fluorimeter has also been developed.

# Dedication

*Nhyira nka Awurade!*

*(To God be the Glory!)*

I dedicate this work to my beloved daughter, Nana Aba Intsifua Duedu who sacrificed her precious first father-daughter days for me to complete this work.

# Acknowledgements

I am highly indebted my supervisor Prof. Chris E. French for his guidance and mentorship throughout this journey. I also appreciate the efforts of members of my PhD committee (Dr. Louise Horsfall, Prof. Steve E. Fry, and Prof. Alistair Elfick) whose feedback has been very instrumental in keeping me on track.

Special thanks to the late Professors Sir Kenneth and Lady Noreen Murray for establishing the Darwin Trust of Edinburgh as well as its board of trustees for awarding my studentship. Without it, I would have missed this Edinburgh experience.

I also thank all previous and current lab members who in diverse ways contributed to the success of my work.

Finally, I thank my wife and daughter and son for their love and support given to me even when I was absent from home. To my parents Christian and Susanna, my uncle Charles as well as the entire family, I say thank you for all the love and encouragement.

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| | | |
|---|---|---|
| Amp | - | Ampicillin |
| ATCC | - | American Type Culture Collection |
| BSA | - | Bovine serum albumin |
| BMCC | - | Bacterial microcrystalline cellulose |
| CAZy | - | Carbohydrate-Active Enzymes Database |
| CBM | - | Carbohydrate binding module |
| CBP | - | Consolidated bioprocessing |
| CD | - | Catalytic domain |
| CMC | - | Carboxymethyl cellulose |
| Cml | - | Chloramphenicol |
| DMSO | - | Dimethyl sulphoxide |
| E.C. | - | Enzyme commission |
| EFD | - | Equivalent fluorescent DNA |
| GFP | - | Green fluorescent protein |
| GH | - | Glycosyl hydrolase |
| iGEM | - | International Genetically Engineered Machine Competition |
| IPTG | - | Isopropyl β-D-1-thiogalactopyranoside |
| Kan | - | Kanamycin |
| LA | - | Luria agar |
| LB | - | Luria broth |
| MUC | - | 4-methylumbelliferyl β-D-cellobioside |
| MUG | - | 4-methylumbelliferyl β-D-glucopyranoside |
| NCBI | - | National Centre for Biotechnology Information |

| | | |
|---|---|---|
| NCIMB | - | National Collections of Industrial, Marine and Food Bacteria |
| ORI | - | Origin of replication |
| PCR | - | Polymerase chain reaction |
| RBS | - | Ribosome binding site |
| RFP | - | Red fluorescent protein |
| SBS | - | School of Biological Sciences |
| SEM | - | Scanning electron microscopy |
| TSS | - | Transformation and storage solution |
| UK | - | United Kingdom |
| X-gal | - | 5-bromo-4-chloro-3-indoyl-β-D-galactoside |

# Chapter 1

# Introduction

## 1.1 Background and overview of this thesis

Having received all my pre-doctoral training in infectious disease related fields, biomass conversion was not an area I envisaged for my PhD. However, after reading Chris French's paper titled "Synthetic biology and biomass conversion: A match made in heaven?" (French, 2009), I developed an interest in this field not for the biomass conversion per se but the synthetic biology. Although the exact meaning of the term 'synthetic biology' is unclear (Schmidt, 2015), it is regarded by many practitioners as the application of engineering principles to biology (French *et al.*, 2015).

The ultimate goal of biomass conversion is to sustainably generate useful high value chemicals and materials such as biofuels and plastics from non-food crops. Current methods for the conversion processes are however not cost-effective (Lynd *et al.*, 2005). Synthetic biology on the other hand seems well placed to provide a solution to this by enabling the engineering of an ideal biofuel producing micro-organism (IBPM) (French, 2009). This thesis investigates the use of multicatalytic cellulases to improve the catalysis of enzymatic cellulose hydrolysis. The second component of this thesis explores the potential of adopting live-dead staining of bacteria in flow cytometry for investigating growth of bacteria in media containing insoluble cellulosic substrates.

## 1.2 The need for biofuels

There has been an ever-increasing concern for alternative and sustainable energy sources (Nuffield Council on Bioethics, 2011; Creutzig *et al.*, 2015; Fulton *et al.*, 2015). This is largely due to high $CO_2$ emissions and the negative impact fossil fuels have on the environment as well as their declining levels. One such alternative is nuclear energy which has gained considerable industrial ground and application in daily use. Dangers however associated with this source emphasize the need for safer, environmentally friendly and sustainable energy sources. Hydropower (hydroelectric, tidal and ocean thermal power) and geothermal appear to be safer alternatives to nuclear energy. These sources however are not practically useful in the long distance transport sector, thus emphasizing the need for liquid fuels. The only sustainable source of liquid fuels appears to be plant derived biofuels.

The use of biofuels is not new. In 1900, Rudolf Diesel demonstrated an engine he constructed running on groundnut oil (Knothe, 2001). Biofuels were also reported during the 1920s and 1930s as well as during World War II (Nuffield Council on Bioethics, 2011). Henry Ford, founder of Ford Motor Company is also quoted to have said in 1925 that “The fuel of the future is going to come from fruit like that sumac out by the road, or from apples, weeds, sawdust - almost anything. There is fuel in every bit of vegetable matter that can be fermented. There’s enough alcohol in one year’s yield of an acre of potatoes to drive the machinery necessary to cultivate the fields for a hundred years”. Years later, major advances and applications of plant-derived fuels have been seen with bioethanol and biodiesel (Antoni *et al.*, 2007; Dale, 2008).

The United States (US) and the European Union (EU) have individually set targets to expand the production and use of biofuels. The Energy Independence and Security Act (EISA) of 2007 establishes that blending of renewable fuels (Biomass diesel and cellulosic biofuels) into transportation fuels in the US should increase from 9 billion

gallons in 2008 to 36 billion gallons by 2022 (Van Dyk and Pletschke, 2012). Similarly, the EU has also projected biofuel supply in transportation fuels to reach 25% by 2030 (Himmel *et al.*, 2007; US EPA, 2013). Despite the potential of plant derived biofuels to meet these targets, large scale applications are currently problematic due to the difficulties in converting plant biomass (mostly made of cellulose, hemicellulose, lignin, pectin, and protein) into bioalcoholic derivatives and biodiesel. Existing systems for ethanol production involve the use of sugarcane and maize (corn) starch whereas that for biodiesel involves esterification of fatty acids derived from vegetable oils (Hill *et al.*, 2006; Granda *et al.*, 2007). The use of staple foods and their products is considered to be a major contributing factor to the rise in food prices (Sachs, 2008). Rapidly growing non-food plants such as scrub willow, switchgrass and miscanthus can be cultivated in industrial quantities on land unsuited for food production, thus providing an alternative to food crops being used as feedstocks (Heaton *et al.*, 2008; French, 2009).

In order to overcome challenges associated with biomass conversion, the need for efficient and robust systems to convert plant biomass is emphasized. Current understanding and technologies in synthetic biology and its applications provide useful insights into possibilities of developing the requisite systems for efficient biomass conversion.

### 1.3 Cellulose structure and accessibility to enzymatic hydrolysis

The plant cell wall is made up of various polysaccharides such as cellulose, callose, hemicellulose (and pectin) and lignin in addition to proteins, enzymes and water as well as other substances (cutin, suberin and inorganic compounds) (Albersheim *et al.*, 1973; Showalter, 1993; Gibeaut and Carpita, 1994; Showalter, 2001; Lerouxel *et al.*, 2006). Cellulose is a linear polymer of D-glucose linked together with β-(1→4)

linkages of repeating cellobiose units and forms the microfibrillar foundation of the plant cell wall (Haworth, 1932; Hon, 1994). Cellulose and callose (a polysaccharide of glucose linked by β-1,3-linkages) are synthesized at the plasma membrane whereas the other non-cellulosic components are secreted to the cell surface and forms a porous matrix which surrounds the cellulose microfibrils (Gibeaut and Carpita, 1994). The cellulose microfibrils are made up of about 30 to 100 molecules of cellulose forming a para-crystalline structure stabilized by intra- and intermolecular hydrogen bonds oriented in a parallel or antiparallel manner as well as Van der Waals forces (Heredia *et al.*, 1995; Harris *et al.*, 2010). Studies on α-cellulose, the fraction obtained from separating hemicelluloses, pectic substances and lignin, have shown that the crystalline structure is surrounded by another less organized structure in which other polysaccharides (eg. mannans and xylans) are linked to cellulose by covalent bonds (Brillouet and Mercier, 1981; Düsterhöft *et al.*, 1991; Koller *et al.*, 1991). These bonds are extremely resistant and cannot be broken by even treatment with acids (Taiz, 1984; Heredia *et al.*, 1995). Cellulose microfibrils have been found to be about 3 to 5 nm in diameter using atomic force microscopy (AFM) (Ding and Himmel, 2006). The study also found that the microfibrils bundle together to form macrofibrils which split to form smaller bundles at the ends and eventually cellulose elementary fibrils (Figure 1-1).


**Figure 1-1: Structural and chemical complexity of cell-wall biomass.**
(A) Example of high-density bales of corn stover harvested on the eastern plains of northern Colorado. (B) An atomic force micrograph of the maize parenchyma cell-wall surface. The diameter of individual microfibril is about 3 to 5 nm. Scale bar, 50 nm. The macrofibrils in the upper-most layer that can also split at the end (indicated by the arrow). (Ding and Himmel, 2006; Himmel *et al.*, 2007)

The complex interaction of cellulose in plant cell walls with other polysaccharides (Figure 1-2) makes it difficult for cellulases to access and degrade it. Himmel et al (2007) identified (in a review) the factors that contribute to recalcitrance of plant biomass enzymatic (as well as chemical) degradation include:

(i) the composition of the epidermal tissue of the plant body (eg. the cuticle and epicuticular waxes)

(ii) the density and arrangement of the vascular bundles

(iii) the sclerenchymatous (thick wall) tissue composition

(iv) the amount of lignin

(v) the structural make-up (homogeneity and complexity) of cell-wall constituents (eg. microfibrils and matrix polymers)

(vi) the effectiveness of enzymes acting on an insoluble substrate and

(vii) the presence of inhibitors naturally existing in cell walls or those generated during conversion processes affecting subsequent fermentation.

The cellulose of cell wall microfibrils consists of precisely arranged cellodextrins resulting in a crystalline structure extremely resistant to enzymatic hydrolysis. Although cellulose is a homopolysaccharide, it occurs in plants intertwined with other polymers forming a complex structure (figure 1-2). Such polymers include lignin and hemicellulose which fill spaces between cellulose microfibrils (Kumar *et al.*, 2009). Lignin is a complex polymer of phenylpropane units whereas hemicellulose is a mixture of polysaccharides (Verardi *et al.*, 2012).

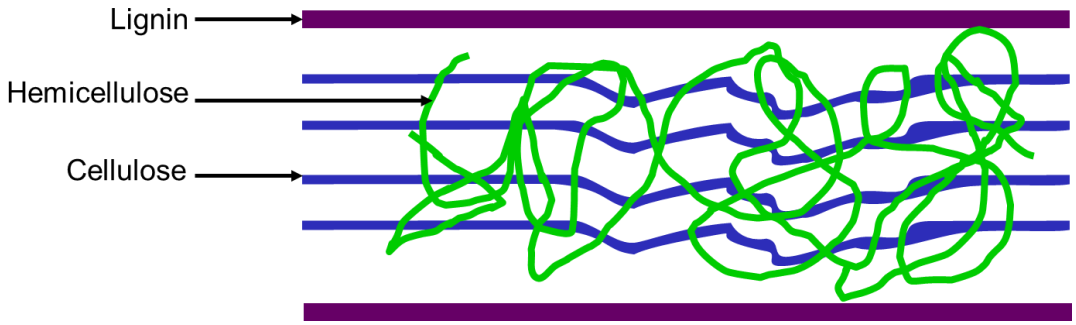



**Figure 1-2: A simplified representation of the major components of lignocellulose** (Adapted from Kumar et al (2009)

Although hydrolysis of cellulose could be achieved by cellulases, the complex interactions that result in the formation of plant feedstock make it difficult for cellulose to be accessed. Cellulases belong to the family of glycoside hydrolases (EC 3.2.1.-), a family of enzymes that hydrolyse the glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (http://www.cazy.org/Glycoside-Hydrolases.html). The hydrolysis of cellulose is achieved by the combined activity of three classes of cellulases (Figure 1-3); endo-β-

1,4-glucanases (EC 3.2.1.4), exo- β-1,4-glucanases (or cellobiohydrolases) and β-glucosidases. Endoglucanases hydrolyse the glycosidic bonds within the amorphous regions of the cellulose chain. The selection of surface sites by endoglucanases is probably random. A water molecule is inserted into the β-(1,4) bond generating long chain oligomers (cellodextrins) with reducing and non-reducing ends. Water molecules occupy the space under the non-reducing chain end preventing it from reannealing into the cellulose crystal. There are two forms of exoglucanases and these act unidirectionally on the cellodextrins either from the reducing (EC 3.2.1.176) or non-reducing ends (EC 3.2.1.91) in a processive (or progressive) manner. This causes the removal of cellodextrins from the microfibril and is considered the rate-limiting step in cellulase action (Himmel *et al.*, 2007). Exoglucanases are thus referred to as “processive” cellulases. The activity of endo- and exo- glucanases results in the liberation of cellobiose which is hydrolysed by β-glucosidases (EC 3.2.1.21) to release glucose (Schülein, 1988).

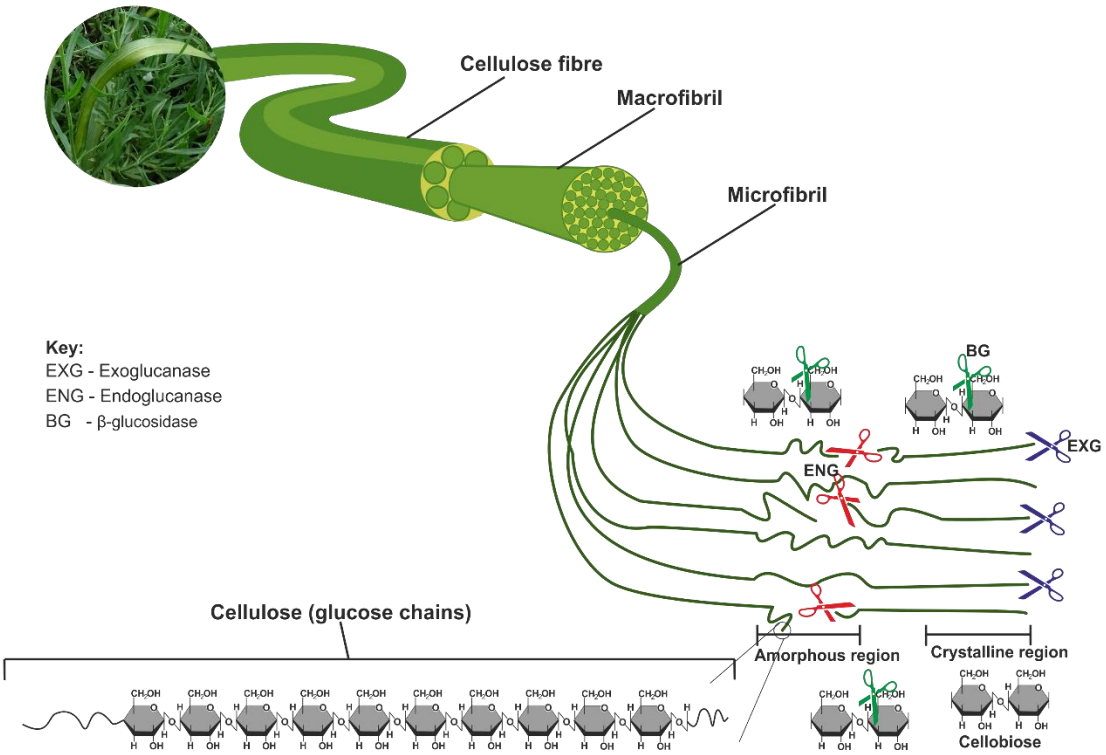



**Figure 1-3: Cellulose structure and enzymatic hydrolysis by free enzyme systems**

The endoglucanases (endo-β-(1,4)-glucanases) or β-(1,4)-D-glucan-4-glucanohydrolases (EC 3.2.1.4) act randomly on soluble and insoluble β-(1,4)-glucan substrates and cleave within the cellulose chains. The exoglucanases (exo-β-(1,4)-D-glucanases), including both the β-(1,4)-D-glucan glucohydrolases (EC 3.2.1.74) liberate D-cellobiose from β-(1,4)-D-glucans. The β-D-glucosidases or β-D-glucoside glucohydrolases (EC 3.2.1.21) act to release D-glucose units from D-cellobiose and soluble cellodextrins (Himmel *et al.*, 2010).

In order to overcome biomass recalcitrance, various strategies are being pursued by researchers. Advanced technologies (eg. crystallography, atomic force and electron microscopy, etc.) have enabled studies that have provided new understanding into plant cell wall structure, biosynthesis and chemistry (Cosgrove, 2005; Ding and Himmel, 2006; Caffall and Mohnen, 2009; Keegstra, 2010; Pettolino *et al.*, 2012). Although the level of understanding is still basic, the body of information available has shown that removal of various non-cellulose cell wall polymers and/or phenolic esters

(eg. linkages between lignin and hemicellulases) results in a decrease in the enzyme concentrations required to convert cellulose to cellobiose or glucose (Torget *et al.*, 1992; Iiyama *et al.*, 1994; Yang and Wyman, 2004). This has led to the theory that plants could be engineered with altered cell wall compositions to enhance enzymatic and chemical degradation. For example, new barley varieties with reduced lignin compositions are being engineered for biofuel production from post-harvest waste (Halpin, 2013, personal communication). Another approach to overcome biomass recalcitrance is engineering catalysts and bioconversion systems (Himmel *et al.*, 2007). Two approaches to improving the performance of cellulases are:

(i) enzyme mining to find new highly active enzymes

(ii) knowledge-based protein engineering

The latter approach requires understanding of the mechanisms of action of the enzymes to be engineered. A third plausible approach will be a combination of the two approaches where highly acting individual enzymes are re-engineered to enhance their processivity and catalytic output, for example, by making fusion proteins to enhance synergy.

### 1.4 Microbial cellulases and biomass conversion

Microbial enzyme systems for biomass conversion may occur in one of two forms; free or complexed systems. Free enzyme systems are present mostly in aerobic bacteria and fungi whereas the complexed systems are present mostly in anaerobic bacteria (Bayer *et al.*, 1998a; Grabber *et al.*, 2008; Hendriks and Zeeman, 2009; Park *et al.*, 2010). Examples of bacteria and fungi with free enzyme systems include *Cellulomonas fimi, Cytophaga hutchinsonii, Trichoderma reesei* and *Aspergillus niger* (Beguin and Aubert, 1994; Zhang and Lynd, 2006). Those with complexed systems

produce multienzyme complexes called cellulosomes. Examples of such microorganisms include *Acetivibrio cellulolyticus, Bacteroides cellulosolvens*, and some clostridia. *Clostridium thermocellum* for example has been and is currently being used as a model organism for cellulose degradation (Beguin *et al.*, 1992; Doi, 2008). Cellulose degrading microbes utilize mainly carbohydrates and are usually unable to use proteins or lipids as energy sources for their growth (Lynd *et al.*, 2002).

Cellulases may be produced as secreted cellulases or cell associated cellulases (Henrissat *et al.*, 1998; Henrissat and Davies, 2000). To utilize insoluble cellulosic substrates, the cellulolytic organism must produce extracellular cellulases. Most aerobic bacteria and fungi that produce non-complexed cellulases secrete them. Typical examples of these are *C. fimi* and *T. reesei*. The expression of the cellulases is often induced by the presence of the substrate and is repressed when utilizable sugars are present (Sukumaran *et al.*, 2005). Various cellulolytic organisms have adopted various strategies to effectively hydrolyse cellulose whether free insoluble cellulose particles or cellulose embedded within the other cell wall polymers. For example, cellulolytic filamentous fungi and actinobacteria use hyphal extensions to penetrate the cellulosic substrate presenting their cellulose systems within the cavities produced (Tomme *et al.*, 1995; Lynd *et al.*, 2002). This strategy enables these organisms to produce free (non-complexed) cellulases. Anaerobic bacteria lack the hyphal property and are unable to effectively penetrate cellulosic material. This might explain their adaptation to produce cellulases as complexed systems which positions them at the site of hydrolysis. The cellulosomes occur as protrusions on the cell walls and their composition varies with respect to the organism (Lynd *et al.*, 2002).

Following lessons from how enzymes occur in natural organisms, efforts in exploring these enzymes have focused on synergy. Thus, to obtain an ideal biomass producing microorganism (IBPM), it is necessary to create an enzyme combination capable of

effectively accessing, hydrolysing and liberating sugars from lignocellulosic material. Additionally, the IBPM should be able to utilize some of these sugars for its continuous growth. Another important component of this IBPM is that it should be able to tolerate the desired product or other by-products that may be released into the medium in the course of its culture in an industrial context (French, 2009). This therefore requires assembly of an array of small parts into a concerted and coordinated pathway to act on lignocellulosic materials.

Whereas some studies have focused on developing and testing individual enzymes, others have combined these enzymes with an ultimate goal of achieving synergy in their activities (US EPA, 2013). Boisset et al (2000) cloned two cellobiohydrolases from *Humicola insolens* and tested them either alone or in a mixture which also included *β*-glucosidase. They observed significant synergy when the two were combined in a mixture. It was realised that one of the cellobiohydrolases, Cel6A, which was less active when used singly, had a partial endoglucanase character and this together with Cel7A produced more reducing ends to be acted upon (Boisset *et al.*, 2000). Various degrees of synergy were also found when various cellulases from *Fibrobacter succinogenes* were cloned and combined differently. Overall, having all the cellulases gave the highest synergy reflecting their complementarity and specificities for the various substrates (Qi *et al.*, 2007). Other studies (Zhou and Ingram, 2000; Park *et al.*, 2010; Zhang *et al.*, 2010) have also found synergy between various cellulose hydrolases, suggesting that the natural situation where a battery of enzymes in a cellulolytic microorganism complement each other to convert biomass can be replicated synthetically.

### *1.4.1 Cellulomonas fimi cellulases*

*C. fimi* is a Gram-positive non-spore forming soil facultative anaerobe which produces an array of extracellular enzymes that hydrolyse various polymers such as cellulose, starch, xylan, mannan and chitin (Beguin *et al.*, 1977; Langsford *et al.*, 1984; Stalbrand *et al.*, 1998; Stoll *et al.*, 1999; Le Nours *et al.*, 2005). It grows best at 30°C and is one of the best-studied cellulolytic bacteria. *C. fimi* and other cellulomonads as well as *Actinotalea fermentas* (formerly *Cellulomonas fermentas*) are the only known facultative anaerobes known to degrade cellulose aerobically or anaerobically (Lynd *et al.*, 2002). Anaerobic degradation of cellulose by other anaerobes such as the Clostridia involves the use of cell surface enzymes (cellulosomes). No evidence of typical cellulosome structures or homologs of scaffoldins, dockerins or cohesins were found when three cellulomonad genomes were analysed (Christopherson *et al.*, 2013). *Cellulomonas flavigena* on the other hand has been reported to have cellulosome-like structures on its cell surface involved in cellulose degradation (Vladut-Talor *et al.*, 1986; Lamed *et al.*, 1987; Kenyon *et al.*, 2005). Additionally, *C. uda* has also been reported to use a mixture of cell-free and cell-associated cellulases (Stoppok *et al.*, 1982). The genome of *C. fimi* has been sequenced recently and showed no evidence of cellulosome-like components suggesting that its cellulase system is of the secreted type (Christopherson *et al.*, 2013).

To date, three classes of β-1,4-glucanases have been identified biochemically in *C. fimi* namely endo- and exo- glucanases and β-glucosidases (Table 1-1). They include four endoglucanases (CenA, CenB, CenC and CenD), one exoglucanase/xylanase (Cex), two exocellobiohydrolases (CbhA and CbhB) and two β-glucosidases (Cfbglu and Nag3A). The catalytic domains of the cellulases they encode come from six different families of glycoside hydrolases: family 3 (Cfbglu and Nag3A); family 5 (CenD); family 6 (CenA and CbhA); family 9 (CenB and CenC); family 10 (Cex); family

48 (CbhB). All except the β-glucosidases contain additional carbohydrate binding modules (CBM). The CBMs belong to family 2 (CenA, CenD, Cex, CbhA and CbhB) and family 4 (CenC, two tandem CBMs). CenB has two CBMs, a CBM3 and a CBM2 separated by three fibronectin III-like (Fn3) domains. The properties of CenA, CenB, CenD, Cex and Cfbglu (used in this study) are described further below.

**Table 1-1: Previously reported cellulases of *Cellulomonas fimi***

| Gene name | Locus tag | Protein Family | Encoded protein function | Reference |
|---|---|---|---|---|
| *cenA (cel6A)* | Celf_3184 | GH6 CB2 | endo-β-1,4-glucanase | (Wong *et al.*, 1986; Gilkes *et al.*, 1989; Gilkes *et al.*, 1991a) |
| *cenB (cel9A)* | Celf_0019 | GH9 CBM3 CBM2 | endo-β-1,4-glucanase | (Greenberg *et al.*, 1987b; Meinke *et al.*, 1991a) |
| *cenC (cel9B)* | Celf_1537 | GH9 CBM4 CBM4 | endo-β-1,4-glucanase | (Coutinho *et al.*, 1991) |
| *cenD (cel5A)* | Celf_1924 | GH5 CBM2 | endo-β-1,4-glucanase | (Meinke *et al.*, 1993) |
| *cex (xyn10A)* | Celf_1271 | GH10 CBM2 | exo-1,4-β-glucanase<br>endo-1,4-β-xylanase | (O'Neill *et al.*, 1986a; O'Neill *et al.*, 1986b) |
| *cbhA (cel6B)* | Celf_1925 | GH6 CBM2 | exo-β-1,4-cellobiohydrolase | (Meinke *et al.*, 1994) |
| *cbhB (cel48A)* | Celf_3400 | GH48 CBM2 | exo-β-1,4-cellobiohydrolase | (Shen *et al.*, 1995) |
| *cfbglu* | Celf_2783 | GH3 | β-glucosidase | (Wakarchuk *et al.*, 1984; Kim and Pack, 1989) |
| *nag3A* | Celf_2983 | GH3 | β-glucosidase<br>β-N-acetylglucosaminidase | (Mayer *et al.*, 2006) |

#### *1.4.1.1 Endoglucanase A (CenA)*

The endoglucanase CenA has a molecular weight of 47 kDa. It is made up of an N-terminal 32 amino acid (aa) signal peptide, 110 aa CBM and a 284 aa catalytic domain (CD). The CBM and CD are joined together by a 23 aa proline-threonine rich rigid linker. The catalytic activity of CenA takes place within the amorphous regions of the cellulose chain. Apart from cellulose, CenA has also been shown to hydrolyse other forms of glucose polymers. It hydrolyses soluble cellotetraose and to a lesser extent cellotriose and cellobiose (Din *et al.*, 1994; Kleman-Leyer *et al.*, 1994; Damude *et al.*, 1996). By the use of Northern blotting, it has been shown that the transcription of the genes *cenA* and c*ex* is regulated by the carbon source present. CenA expression could thus be induced by glycerol, cellobiose and cellulose but was repressed by the presence of glucose (Greenberg *et al.*, 1987b). Din *et al.* (1994) found a unique characteristic of CenA which has not been reported in other endoglucanases. They observed that the CBM of CenA disrupts the structure of cellulose fibres resulting in the release of fine cellulose particles without any detectable hydrolytic activity. These could then be hydrolysed by the catalytic domain suggesting intramolecular synergism between the CBM and catalytic domain.

#### *1.4.1.2 Endoglucanase B*

The second endoglucanase from *C. fimi* is 1,012 aa long CenB. Mature CenB is a 110 kDa protein with a catalytic domain and two CBMs belonging to families CBM3 and CBM2. It also has a 33 aa signal peptide. The last 100 aa at the C-terminus of CenB has more than 50% similarity to the CBMs of CenA and Cex. The N-terminus of each CBM is made up of about 18 aa rich in proline and threonine residues analogous to the proline-threonine linkers of Cex and CenA (Meinke *et al.*, 1991a). Another feature

of CenB is that its CBMs flank three fibronectin type-III (Fn3) like repeat sequences of 98 aa with >60% identity. The repeats in CenB are related to similar repeats of 95 aa in chitinase A1 from *Bacillus circulans* which also share about 70% identity (Meinke *et al.*, 1991b). Fibronectin in eukaryotes is a multifunctional plasma protein typically involved in protein-protein interactions in muscle and neural tissues. The presence of the Fn3-like domains suggest possible interaction of these cellulases or proteins (including cellulases) exhibiting similar Fn-like repeats. Purified CenB has been reported to show a high affinity for binding to and hydrolysis of crystalline cellulose. It hydrolysed up to 87% of bacteria microcrystalline cellulose (BMCC) and released 50 times more reducing sugar than CenA (Kleman-Leyer *et al.*, 1994). Unlike the other cellulases in *C. fimi*, CenB is not induced by cellulose or the other polysaccharides. It is constitutively expressed irrespective of the polysaccharide present. It has two promoters controlling its expression; a strong inducible and a weak constitutive promoter. The inducible promoter is distal to the transcription start site and is induced only in the presence of cellulose. The presence of glycerol, cellobiose or glucose does not affect this promoter (Greenberg *et al.*, 1987a).

#### *1.4.1.3 Endoglucanase D*

The endoglucanase CenD was first identified as a 75 kDa cellulose-binding polypeptide (Cbp) from culture supernatants of *C. fimi* by adsorption to cellulose or Sephadex. This was the first evidence of the identification of a glycoside hydrolase family 5 (previously cellulase family A) enzyme in *C. fimi* (Gilkes *et al.*, 1991b; Meinke *et al.*, 1993). It has four domains; a 39 aa signal peptide followed by a CD sequence 405 aa long, two repeats of 98 aa Fn3-like proteins and a 105 aa long CBM. The CD is connected to the Fn3-like domain by a short linker of three Gly residues. Similarly, the two repeat Fn3-like domains are joined by another short linker of three Gly

residues. Joining the Fn3-like domain and the CBM is a short Pro-Thr-Thr linker. The CBM of CenD belongs to family 2 similar to those of CenA, CenB and Cex (Meinke *et al.*, 1993). The rate of hydrolysis of avicel (microcrystalline cellulose) by CenD was found to be similar to those of CenA and CenB. On BMCC, hydrolysis by CenD was similar to that of CenB, but that of CenA was about 20 fold lower (Meinke *et al.*, 1993; Sandercock *et al.*, 1996).

#### *1.4.1.4 Exoglucanase (Cex)*

To date, Cex is the only known β-1,4-exoglucanase of *C. fimi* apart from CbhA and CbhB. Cex has a coding region of 1452 bp (484 codons) (O'Neill *et al.*, 1986a). It hydrolyses cellulose and cellotetraose to release cellobiose from non-reducing chain ends (Kleman-Leyer *et al.*, 1994; White *et al.*, 1994). It has also been shown to possess a weak endoglucanase activity (http://www.uniprot.org/uniprot/P07986) and also activity against xylan. This activity has been reported to be about 40 times greater than that on cellulose (Notenboom *et al.*, 1998). It has an N-terminal catalytic domain which is separated by a proline-threonine linker from the non-catalytic CBM which binds to amorphous, semi-crystalline and crystalline cellulose chains (Gilkes *et al.*, 1991b; Ong *et al.*, 1993a; Creagh *et al.*, 1996; McLean *et al.*, 2000). The CBM is attached to the catalytic domain. It migrates along the surface of crystalline cellulose thereby facilitating the diffusion of the catalytic domain (Jervis *et al.*, 1997a).

#### *1.4.1.5 C. fimi β-glucosidase (cfbglu)*

Initial studies on the glycanases of *C. fimi* found that cell lysates contained at least two β-glucosidases. One of these was found to constitutive and could hydrolyse *p*-nitrophenyl-β-D-glucoside (PNPG) but not cellobiose. The second (inducible)

however hydrolysed both PNPG and cellobiose. The former was found to be an aryl-β-D-glucosidase and the latter a “true” cellobiase (β-D-glucoside glucohydrolase) (Wakarchuk *et al.*, 1984). The locus tags for these genes are Celf_2983 (Nag3A) and Celf_2783 (Cfbglu) respectively. The expression of Cfbglu was induced about seven fold in the presence of cellobiose (Wakarchuk *et al.*, 1984) and the enzyme hydrolyses cellobiose into two glucose monomers (Kim and Pack, 1989). Like other members of the GH3 family, Cfbglu has no CBM. Cell-free culture supernatants of *C. fimi* were found to be devoid of β-D-glucosidase activity suggesting this type of activity in *C. fimi* is strictly intracellular (Wakarchuk *et al.*, 1984). This is further supported by the absence of a signal peptide in Cfbglu.

### *1.4.2 Cytophaga hutchinsonii cellodextrinase CHU_2268*

*Cytophaga hutchinsonii* is an aerobic Gram negative gliding bacterium found in soil and decaying plant matter. It was initially isolated as *Spirochaeta cytophaga* (Hutchinson and Clayton, 1919). Very little has been studied about the mechanisms of cellulose degradation in members of the *Cytophaga* genus. The genome of *C. hutchinsonii* has been sequenced (Xie *et al.*, 2007). Additionally, cloning, heterologous expression and characterization of endoglucanases, CHU_1965/Cel9A (GH9 family) (Louime *et al.*, 2007) and CHU_1107/ChCel5A (GH5 family) (Zhu *et al.*, 2013) have been reported. The genome sequence shows a notable absence of genes encoding exoglucanases. Very little is known about the GH family 3 β-glucosidase CHU_2268 of *C. hutchinsonii* (Liu, 2012). Studies on this enzyme in our lab have shown that it is able to degrade cellooligonucleotidesaccharides up to at least cellohexose. Glucose is released from this degradation rather than cellobiose suggesting it is probably a cellodextrinase (glucan-1,4-β-glucosidase, EC 3.2.1.74).

CHU2268 thus appears to have characteristics of both β-glucosidases (EC 3.2.1.21) and EC 3.2.1.74 (C. K. Liu *et al.*, *In preparation*).

### *1.4.3 Protein domain organization and opportunities for protein engineering*

Non-complexed cellulases are often comprised of modular proteins (figure 1-4). The arrangement of these individual modules contributes to the function of the entire enzyme. The modular nature of proteins offers benefits such as increased stability, cooperative functions and protection of intermediates and enhanced protein stability (Ostermeier and Benkovic, 2000). Schulz and Schirmer (1979) define a protein domain as ‘‘a subregion of the polypeptide chain that is autonomous in the sense that it possesses all the characteristics of a complete globular protein”. Domains can further be considered as structural (geometrically separate entities) or functional (functionally autonomous regions). Functional domains consist of one or more structural domains (Schulz and Schirmer, 1979).

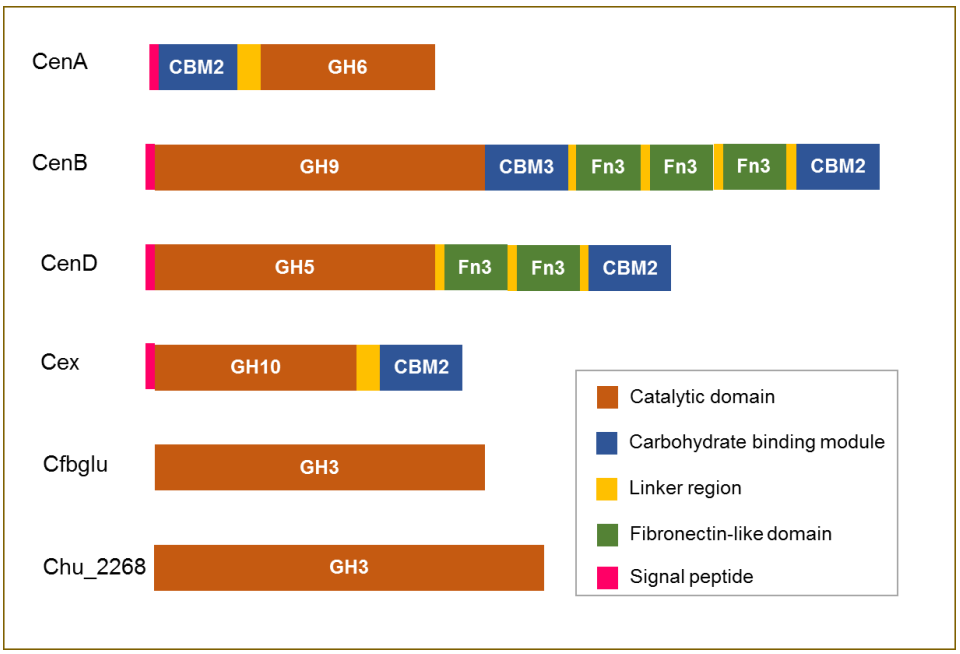



**Figure 1-4: Modular composition of some cellulases from *C. fimi* and *C. hutchinsonii*.**

Endo- and exo- glucanases are often characterized by a CD and a CBM. Other domains such as the fibronectin-like domains in CenB and CenD may also be present. Domains are linked to each other by short or long linker regions. On the other hand, β-glucosidases typically do not have CBMs. A common feature of the linkers of *C. fimi* cellulases is that they are rich in Pro and Thr residues whether long or short.

Current advances in protein engineering have resulted from creating multi-functional fusion proteins containing domains from various proteins (Nixon *et al.*, 1997; Zhang *et al.*, 2009; Elleuche, 2015; Yu *et al.*, 2015). Chimeric enzymes catalysing the same multistep sequential reaction have been shown to demonstrate superiority over the individual native enzymes (Bulow and Mosbach, 1991). Making a protein bi- or multi-functional potentially facilitates enzymatic performances and expands the range of substrate specificities. This feature is described as playing a central role in evolution (Elleuche, 2015). In nature, evolution of multifunctional chimeric proteins has been suggested. For example, *Aspergillus nidulans* encodes in a single open reading frame (ORF), a large protein with five distinct catalytic activities essential for tryptophan biosynthesis (Charles *et al.*, 1986; Conrado *et al.*, 2008). Another example is that of an endoglucanase-exoglucanase fusion protein from *Anaerocellum thermophilum*

(Zverlov *et al.*, 1998). Artificially fused enzymes such as multicatalytic cellulases have also been reported (Boisset *et al.*, 2000; Qi *et al.*, 2007; Park *et al.*, 2010; Zhang *et al.*, 2010). Fusion proteins can catalyze divergent consecutive (Fig. 1-5a), non-consecutive (Fig. 1-5b), or opposing reactions (Fig. 1-5c) (Khandeparker and Numan, 2008; Elleuche, 2015).

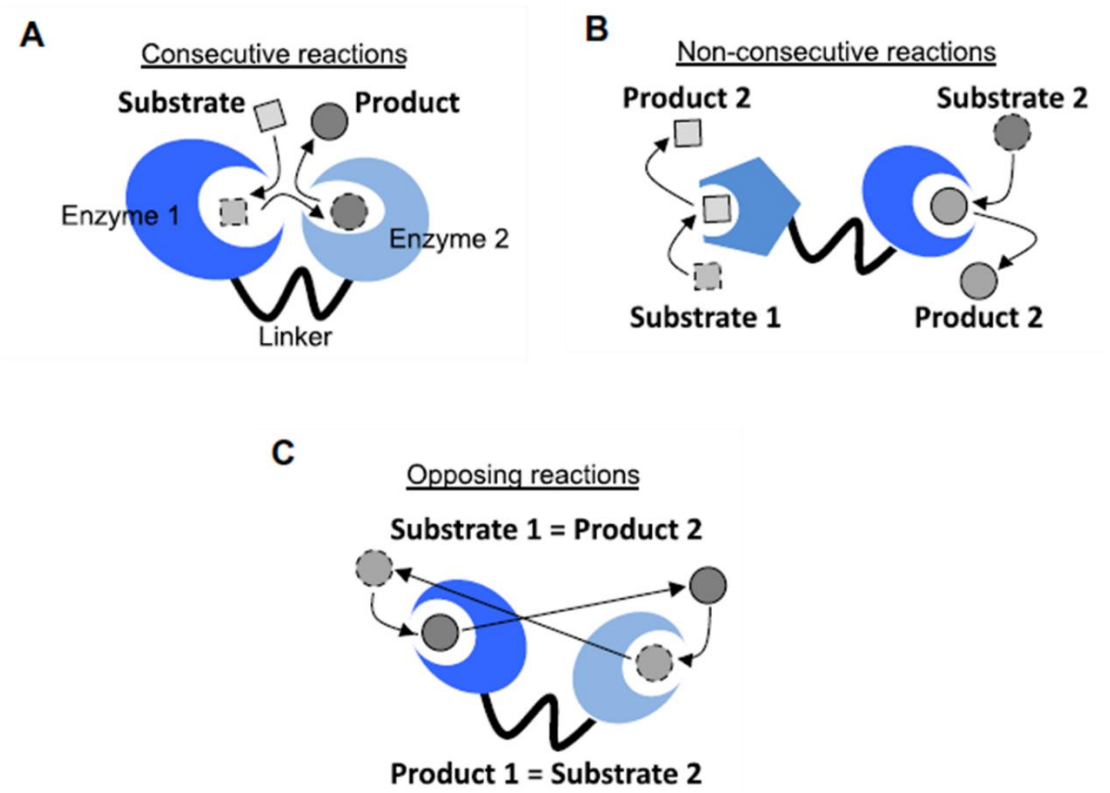



**Figure 1-5: Bicatalytic enzyme designs.**
A: Two enzymes catalyze consecutive reactions. B: Non-consecutive reactions are catalysed. C: Catalysis of opposing reactions. Enzymes are indicated in light and dark blue, and substrates and products are given in light and dark grey. Arrows highlight substrate to product conversion and propagation (Elleuche, 2015).

## 1.5 DNA assembly from standard parts

The backbone of synthetic biology is the rapid, efficient and versatile DNA assembly methods that facilitate the building of new genetic devices and pathways (Arkin, 2008;

Ellis *et al.*, 2011). As distinct from traditional cloning, synthetic biology cloning adopts the general engineering principle of assembling standard parts. This makes the approach dynamic in that parts can be combined quickly in any defined order. In addition to building genetic modules and circuits, synthetic biology offers the potential of combining various modules to form larger systems such as chromosomes and organisms. One such application is the synthetic yeast 2.0 project aiming to build the world's first synthetic eukaryotic genome (Annaluru *et al.*, 2014; Schindler and Waldminghaus, 2015).

Several standards for assembling DNA have been developed (Knight, 2003; Li and Elledge, 2007; Engler *et al.*, 2009; Gibson *et al.*, 2009; Quan and Tian, 2009; Sarrion-Perdigones *et al.*, 2011; Zhang *et al.*, 2012; Sarrion-Perdigones *et al.*, 2013; Colloms *et al.*, 2014; Li *et al.*, 2014; Trubitsyna *et al.*, 2014; Storch *et al.*, 2015). The BioBrick standard is a restriction-ligation-based method that has in one way or another influenced the development of all other assembly standards. It is based on the principle of idempotency, where the prefix and suffix are maintained in all assembled parts enabling consecutive rounds of hierarchical cloning (Knight, 2003). A major disadvantage is that each part must not have forbidden restriction sites; the restriction sites in the prefix (EcoRI & XbaI) and suffix (SpeI & PstI). In addition, this standard does not allow the assembly of more than two parts in a single reaction. With the need for building more complex systems growing, focus has shifted to methods that allow assembly of multiple parts in a single reaction, otherwise called the one-pot methods (Kahl and Endy, 2013).

One pot methods can be categorized by the principles used in the assembly method. The sequence and ligase independent cloning (SLIC) (Li and Elledge, 2007), Gibson isothermal assembly (Gibson *et al.*, 2009), circular polymerase extension cloning (CPEC) (Quan and Tian, 2009) and Seamless Ligation Cloning Extract (SLiCE)

(Zhang *et al.*, 2012) are based on end-homology between the ends of neighbouring parts. Prior to amplification, the homology ends are added by PCR, hence for each part, new primers are required. Furthermore, each part needs to be sequenced to confirm the correctness of the parts. Other methods such as the original Golden Gate (Engler *et al.*, 2009), modular cloning (MoClo) (Weber *et al.*, 2011) and GoldenBraid (Sarrion-Perdigones *et al.*, 2011; Sarrion-Perdigones *et al.*, 2013) are based on the use of type IIS restriction enzymes generating user-defined overhangs. They are quick but like the BioBrick standard, requires the removal of forbidden restriction sites within DNA parts. Furthermore, parts are required to be cloned into specific plasmids to generate the necessary restriction overhangs. Other methods combine the use of type IIS restriction enzymes and linkers. These include GenBrick (BBF RFC98) (Gasiūnaitė *et al.*, 2013), modular overlap-directed assembly with linkers (MODAL) (Casini *et al.*, 2014) and the Biopart Assembly Standard for Idempotent Cloning (BASIC) (Storch *et al.*, 2015). The use of linkers is very useful for making fusion proteins in that the linkers can serve as fusion linkers between protein domains. The BioBrick and GenBrick assembly standards were used this work.

### *1.5.1 BioBrick assembly*

In order to accord flexibility in combining DNA parts, the BioBrick® assembly standard was introduced (Knight, 2003). The BioBrick® parts contain standard prefix and suffix sequences with a combination of restriction sites. The prefix contains EcoRI-NotI-XbaI whereas the suffix contains SpeI-NotI-PstI. This allows very flexible pairwise assembly of DNA parts in the BioBrick® format. There are two main types of BioBrick assembly (http://parts.igem.org/About_Assembly); standard assembly (figure 1-6a) and 3A (three antibiotic) assembly (figure 1-6b). Both assembly types involves cloning two parts together. Standard assembly involves two plasmids containing the two parts

to be joined. The upstream part is digested with EcoRI and SpeI and the downstream part is digested with EcoRI and XbaI. The digests are separated by agarose gel electrophoresis after which the smaller band from the upstream digest and the larger band from the downstream digest are excised, purified and ligated. The two initial plasmids may contain the same or different antibiotic selection cassettes, however the resulting plasmid will have the antibiotic selection cassette of the downstream part. In 3A assembly, three plasmids each with a different antibiotic cassette are used. The two plasmids containing the parts to be combined serve as donor plasmids and the third as the acceptor plasmid that receives the two parts. The plasmids are digested with EcoRI/SpeI (donor 1), XbaI/PstI (donor 2) and EcoRI/PstI (acceptor). This method does not require gel purification and has higher success rates compared to standard assembly.

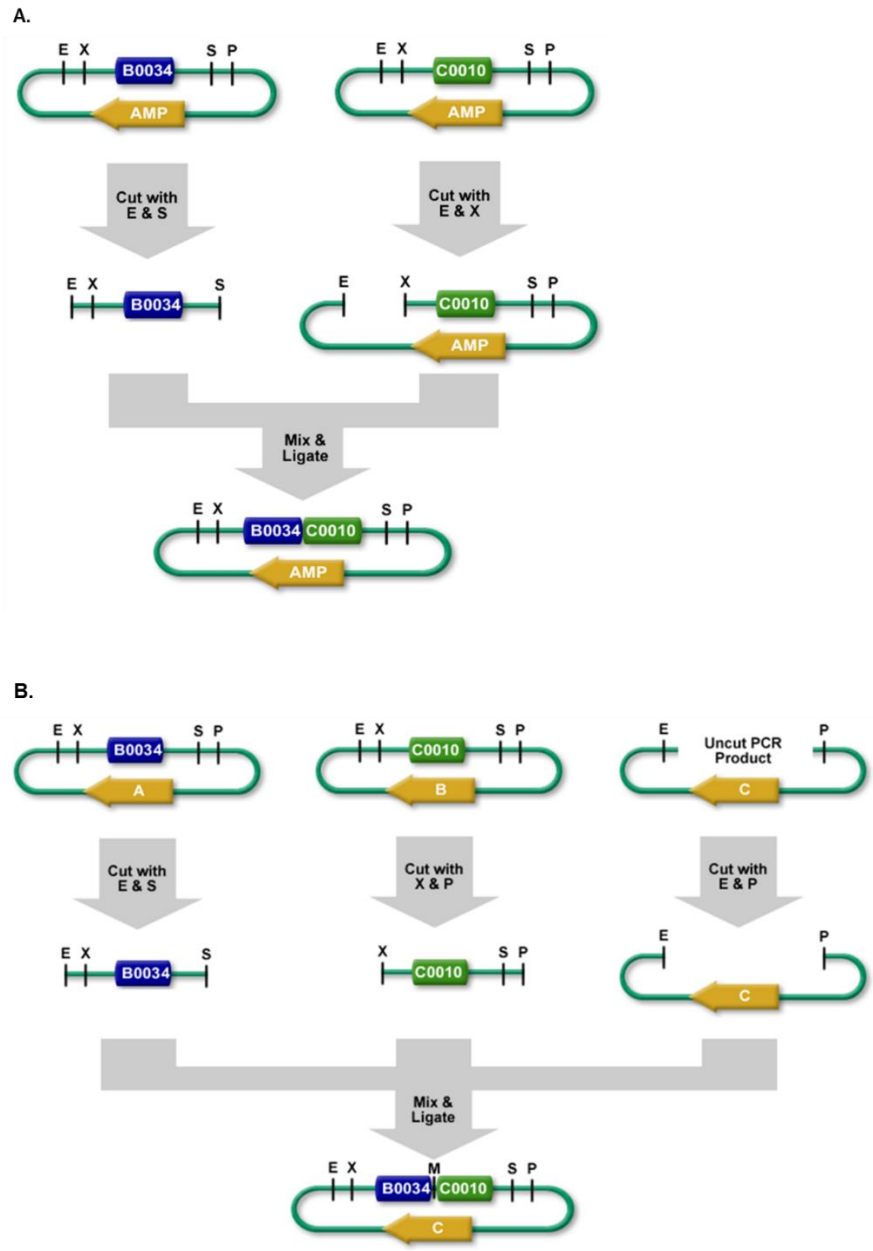



**Figure 1-6: Overview of BioBrick standard and 3A assembly methods**
(A) The BioBrick Standard Assembly and (B) the 3A BioBrick Assembly are illustrated. The BioBrick is flanked by restriction sites forming the prefix (EcoRI and XbaI) on the left and the suffix (SpeI and PstI) on the right. For the Standard Assembly the blue part is cut out of its plasmid with EcoRI/SpeI and inserted into the second plasmid cut by EcoRI/XbaI. Gel purification of the digested parts is necessary whereas in the 3A assembly, gel purification is not necessary. In the 3A assembly, the two parts to be combined are cut out with EcoRI/SpeI and XbaI/PstI respectively. They are then inserted into a third plasmid cut with EcoRI/PstI. Source: Registry of Standard Biological Parts.

Although BioBrick assembly is highly flexible it is relatively slow for assembly of large number of parts. To overcome this problem, ‘one-pot’ assembly methods have been introduced. Examples of these include the Gibson and Golden-Gate assembly methods among others (Engler *et al.*, 2009; Gibson *et al.*, 2009). A common feature of these is that assembly is based on end-homology. Hence, in order to use these methods, homology ends must be added to each part usually by PCR which also introduces possibilities of PCR-induced mutations. To ensure fidelity of each assembly, sequencing must be performed. All these processes do not make these assembly methods flexible to handle.

### *1.5.2 GenBrick assembly*

The GenBrick method is a novel technique that incorporates the flexibility of the BioBrick method with the rapidity of the ‘one-pot’ assembly methods. A key feature is that it can be fully automated and does not require PCR or gel-electrophoresis. GenBrick is based on the proprietary Genabler technology developed for Scottish Enterprise by Ginkgo BioWorks (Boston, USA). Parts developed in GenBrick format are fully BioBrick compatible and can be assembled by normal BioBrick assembly, but they can also be assembled using Genabler assembly, which allows rapid 'single pot' assembly of up to ten parts in a single reaction. GenBrick uses type IIS restriction enzymes which cut outside their recognition sequence. The advantage of this type of cutting is that the user can define which sticky ends are generated. GenBrick uses EarI which recognizes the sequence 5’-CTCTTC-3’. It cuts one base downstream of the recognition site and leaves three base overhangs determined by the downstream sequence. In addition to EarI, SapI, which only differs from EarI by a 5’ G in the recognition sequence (5’-GCTCTTC-3’), may be used. The advantage of SapI is that its recognition sequence is encountered less frequently than EarI. SapI however tends

to settle in the tube and this affects the efficiency of cutting (https://www.neb.com/faqs/1/01/01/why-isn-t-sapi-cutting). Extra mixing of the tube during incubation is thus needed.

GenBricks have two components; Bricks and Linkers. Linkers are long oligonucleotides up to about 80 base pairs. A Brick could be a coding sequence, a promoter sequence or origin of replication, etc. These are usually more than 80 base pairs and do not contain start or stop codons, which makes it easier to add N- or C-terminal tags as well as to generate fusion proteins. Each linker consists of four oligonucleotides and can be referred to (following Genabler terminology) as Eye Long, Eye Short, Hook Long, and Hook Short (EL, ES, HL and HS). The complete linker looks like this (figure 1-7):

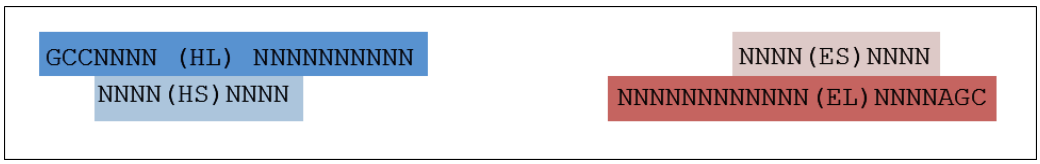



**Figure 1-7: GenBrick Linkers**

Linkers are assembled with parts in the pre-assembly step. The long homology arms of the linkers ensure the correct positioning of parts during the assembly step.

The long single strand 3'- overlaps should be about 15 to 18 bases in length and complementary whereas the double stranded regions should be at least 10 bases in length. The 5'- overhang of the HL begins with GCC whereas that of the EL begins with CGA. These will correspond to GGC and TCG which will be generated on the brick (or vector) when digested with EarI. These 5'- overhangs are defined by a standard prefix (GAATTC G CGGCCG CT TCTAGA TGGC CTCTTC T **TCG**) and suffix (CTGCAG CGGCC GCT ACTAGT A CTCTTC A **GGC**) which are used in designing primers for making GenBrick parts. These are the standard RFC 10

BioBrick prefix and suffix for a coding sequence, with the addition of short sequence including the EarI sites plus a start and stop codon. Coding Bricks are cloned in a Donor Vector flanked by two inward-facing EarI sites. These generate the same 5'-overhangs TCG at the upstream end, and GGC at the downstream end.

To perform a GenBrick assembly, two stages are involved; the pre-assembly and final assembly. Each pre-assembly consist of one eye, one brick and one hook. The brick in its Donor Vector is placed with the linkers (four oligonucleotides) in a tube with some EarI and T4 DNA ligase. The tube is cycled at temperatures between 16˚C and 37˚C using a thermal cycler. The resulting products are purified to remove unligated oligonucleotides. For final assembly, the pre-assemblies are simply mixed together in a tube, then used to transform *E. coli*. Ligation is not necessary and may even reduce efficiency. The transformants are plated on LA with an appropriate antibiotic and IPTG. X-Gal may be added if the Acceptor vector contains *lacZ'*$\alpha$. Colonies bearing correct assemblies are selected based on selectable markers (e.g. blue/white selection) and can be tested, for example by plasmid DNA miniprep and restriction analysis, or colony PCR using the vector eye long and linker1 hook long oligonucleotides as primers to amplify the entire assembly, if its size permits this. An example of a four-part GenBrick assembly is shown in figure 1-8.

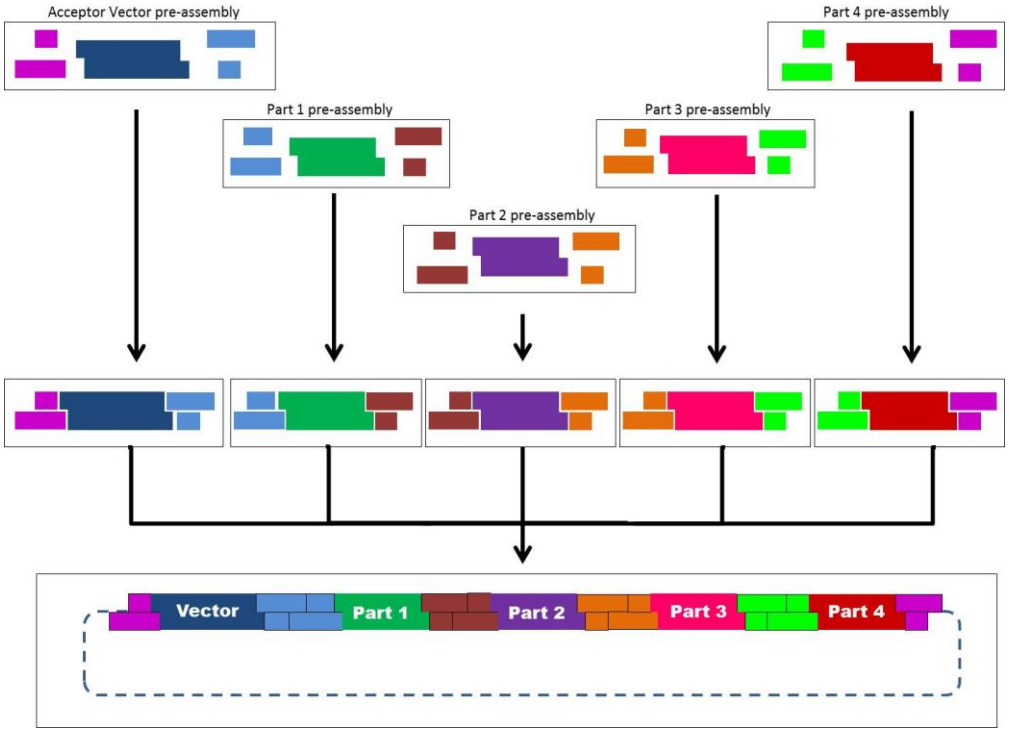



**Figure 1-8: Overview of GenBrick assembly**

The GenBrick assembly involves the use of EarI restriction enzyme to cut all parts giving overhangs with CGG and TCG at the 5′ and 3′ ends respectively. Linkers are ordered as two pairs of synthesized oligonucleotides with short overhangs GCC or AGT at the 5′ ends and longer overhands of 15 to 18 bases at the 3′ ends. Each pair of linker is annealed to the corresponding part in the pre-assembly step followed by a combination of all part/linker components in the final assembly step.

## 1.6 Previous results in this laboratory

Studies in this laboratory (Lakhundi, 2011; Barnard, 2012; Liu, 2012) have explored the applications of synthetic biology using BioBricks in designing and testing modular parts of cellulases from the gram-positive rod *Cellulomonas fimi* and gram-negative rod *Cytophaga hutchinsonii*. Among these cellulases are the endoglucanases CenA and CenB from *C. fimi* and the exoglucanases Cex and Chu_2268 from *C. fimi* and *C. hutchinsonii* respectively. Some combinations of some of these parts have also been made and tested. One of these is the Cex-CenA part which has been functionally

expressed in *Escherichia coli*, *Citrobacter freundii* and *Bacillus subtilis*. Results showed that *C. freundii* bearing this construct can grow on filter paper (figure 1-5).

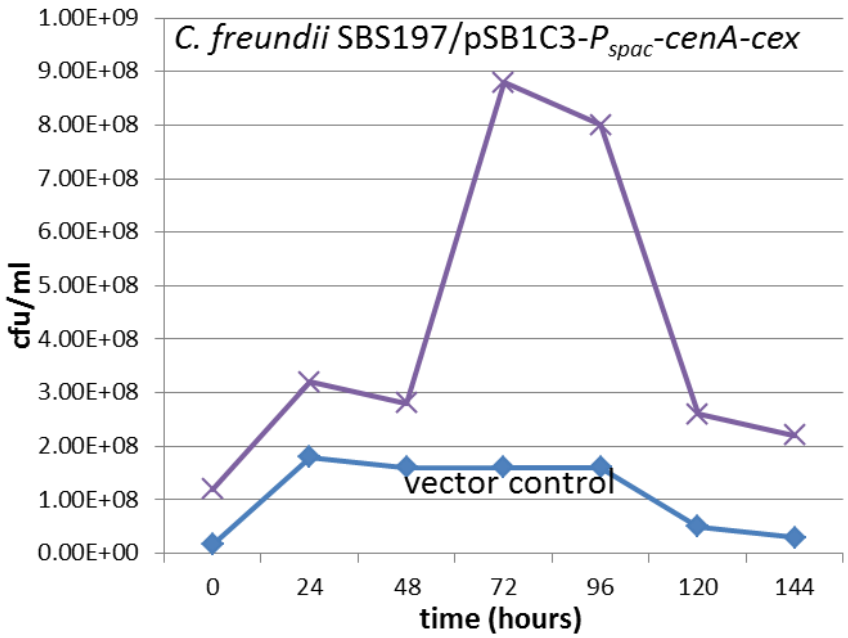



**Figure 1-9: Growth on filter paper by *C. freundii*** (Lakhundi, 2011).
*C. freundii* strain SBS197 expressing the *cex-cenA* construct grows in minimal media containing filter paper whereas a similar strain without the construct does not.

These findings establish a foundation for further work exploiting the flexibility of assembling parts and devices from standard parts to form effective systems for biomass conversion.

### 1.7 Secretion of CenA and Cex in *E. coli*

Gilkes *et al.* (1984a) showed that *C. fimi* cellulases CenA and Cex expressed in *E. coli* were exported to the periplasm. Though their activities were found to be complementary, their export to the periplasm was at different efficiencies. This finding is particularly interesting as it could mean that the signal recognition peptide of CenA

and Cex are also recognized by *E. coli. C. freundii* and *E. coli* being Gram-negative organisms contain an outer membrane which could prevent extracellular secretion leading to the trapping and preventing the release of the expressed recombinant CenA and Cex. Its however been reported that both CenA and Cex are able to escape from the periplasm into the medium (Guo *et al.*, 1988).

By co-expressing the *out* operon of a type II secretion system and recombinant cellulases, the problem of impeded extracellular secretion could be avoided. Endoglucanase genes *celY* and *celZ* from *Erwinia chrysanthemi* were functionally integrated into the chromosome P2 of *Klebsiella oxytoca* P2 using surrogate promoters from *Zymomonas mobilis* (Zhou *et al.*, 2001).

## 1.8 Aim and Objectives

The project aims was to use synthetic biology tools to investigate ways of improving microbial degradation of cellulosic biomass.

**Specific Objectives:**

1. Investigate synergy between cellulases through fusion proteins
2. Develop a reproducible assay to monitor growth of bacteria in cultures containing insoluble substrates

# Chapter 2

# Materials and Methods

## 2.1 Bacteria strains

*Escherichia coli* and *Citrobacter freundii* were used in this study. *E. coli* JM109 (New England Biolabs, Inc.), derivative of the non-pathogenic K12 strain was used for cloning and protein expression experiments. Its genotype is *F´traD36 proA$^{+}$B$^{+}$ lacI$^{q}$ Δ(lacZ)M15/Δ(lac-proAB) glnV44 e14$^{-}$ gyrA96 recA1 relA1endA1 thi hsdR17* (Yanisch-Perron *et al.*, 1985). *E. coli* BL21 strain (Agilent Technologies, Inc.) was occasionally used for protein expression as described. It is a derivative of the *E. coli* B strain which is also non-pathogenic and has genotype *F- dcm ompT hsdS($r_B$- $m_B$-) gal [malB+]$_{K-12}$(λS)*. For cellulose degradation experiments *C. freundii* NCIMB11490 (ATCC 8090) was used.

## 2.2 Overview of molecular biology procedures

### *2.2.1 DNA extraction*

DNA extraction was performed using commercial DNA extraction kits from QIAGEN GmbH, Germany. For genomic DNA from bacteria cultures, the DNeasy Blood and Tissue kit was used. Plasmid DNA was extracted using the QIAprep spin miniprep kit. The manufacturer's spin protocols were followed for each of these kits.

### *2.2.2 Polymerase chain reaction (PCR)*

PCR was performed using Q5 DNA polymerase (New England Biolabs, Inc. USA) according to the manufacturer's protocol with modifications. PCR was performed in a

final reaction volume of 50 µl (Table 2-1). The Q5 high GC enhancer buffer was included in all reactions. For templates that did not amplify using the standard reaction conditions (Table 2-2), the initial denaturation step was extended to 1 minute. PCR was performed in either the MJ Research PTC 200 DNA engine (Bio-Rad Laboratories, Inc. USA) fitted with dual 48 x 0.2 ml sample blocks or the Techne 3PrimeG gradient thermal cycler (Bibby Scientific Ltd., UK) fitted with a single 48 × 0.2 ml sample block.

**Table 2-1: Components of PCR reaction mixes**

| Component | 50 µl Reaction | Final Concentration |
|---|---|---|
| 5X Q5 Reaction Buffer | 10 µl | 1× |
| 10 mM dNTP mixture | 1 µl | 200 µM |
| 10 µM Forward Primer | 2.5 µl | 0.5 µM |
| 10 µM Reverse Primer | 2.5 µl | 0.5 µM |
| Template DNA | 1 µl | |
| Q5 High-Fidelity DNA Polymerase | 0.5 µl | 0.02 U/µl |
| 5X Q5 High GC Enhancer | 10 µl | 1× |
| Nuclease-Free Water | to 50 µl | |

Source - https://www.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491

**Table 2-2: Thermocycling Conditions for routine PCR**

| STEP | TEMP | TIME |
|---|---|---|
| Initial Denaturation | 98°C | 30 seconds |
| 25–35 Cycles | 98°C | 5–10 seconds |
| | *50–72°C | 10–30 seconds |
| | 72°C | 20–30 seconds/kb |
| Final Extension | 72°C | 2 minutes |
| Hold | 4–10°C | |

Source - https://www.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491

PCR products were purified for downstream applications (restriction digests, cloning, etc.) using the QIAquick PCR purification kit (QIAGEN GmbH, Germany) according to manufacturer's instructions.

### *2.2.3 Primers and other oligonucleotides*

Primers here refer to oligonucleotides used for PCR. Other oligonucleotides as used here refer to those oligonucleotides that were annealed in pairs or multiples to form double stranded DNA fragments or linkers for cloning.

#### *2.2.3.1 Primers*

Three main types of primers were designed and used in the various studies reported in this thesis. These were

- Cloning primers
- Mutagenic primers
- Other modification primers

- Sequencing primers

Cloning primers were designed for amplifying genes or templates for subsequent cloning. For these primers, the standard GenBrick prefix (5'-C GAATTC CT TCTAG ATG GC CT CTTCT TCG-3') and suffix (5'-ATC CTGCAG CT ACTAGTA CTCTTCA GGC-3') were included before and after the forward and reverse primers respectively (Figure 2-1a). The final primer length was limited to a maximum of 50 nucleotides. Mutagenic primers were designed for performing single point mutations to remove unwanted restriction sites in plasmid constructs. The forbidden restriction recognition sites were EcoRI (5'-GAATTC-3'), XbaI (5'-TCTAGA-3'), EarI (5'-CTCTTC-3'), SpeI (5'-ACTAGT-3') and PstI (5'-CTGCAG-3'). Non-overlapping divergent primers were designed centred on the unwanted restriction site (Figure 2-1b). A nucleotide change was determined in order not to change the amino acid sequence. The nucleotide to be changed was positioned at the 5' end of the forward primer. Other modification primers were designed mainly for deleting or inserting DNA fragments. To delete a fragment of DNA from a plasmid construct, divergent non-overlapping primers were designed to exclude the unwanted region. Primers meant for inserting a piece of DNA were tagged with the DNA sequence to be inserted (Figure 2-1c). General BioBrick sequencing primers (pSBNX3insf2 and pSBNX3insr2) were used for sequencing inserts in BioBrick or GenBrick constructs in pSB1C3, pSB1K3 or pSB1T3. In addition, sequence or construct specific primers were also designed to sequence internal regions. Primers used in this study are listed in table 2-3.

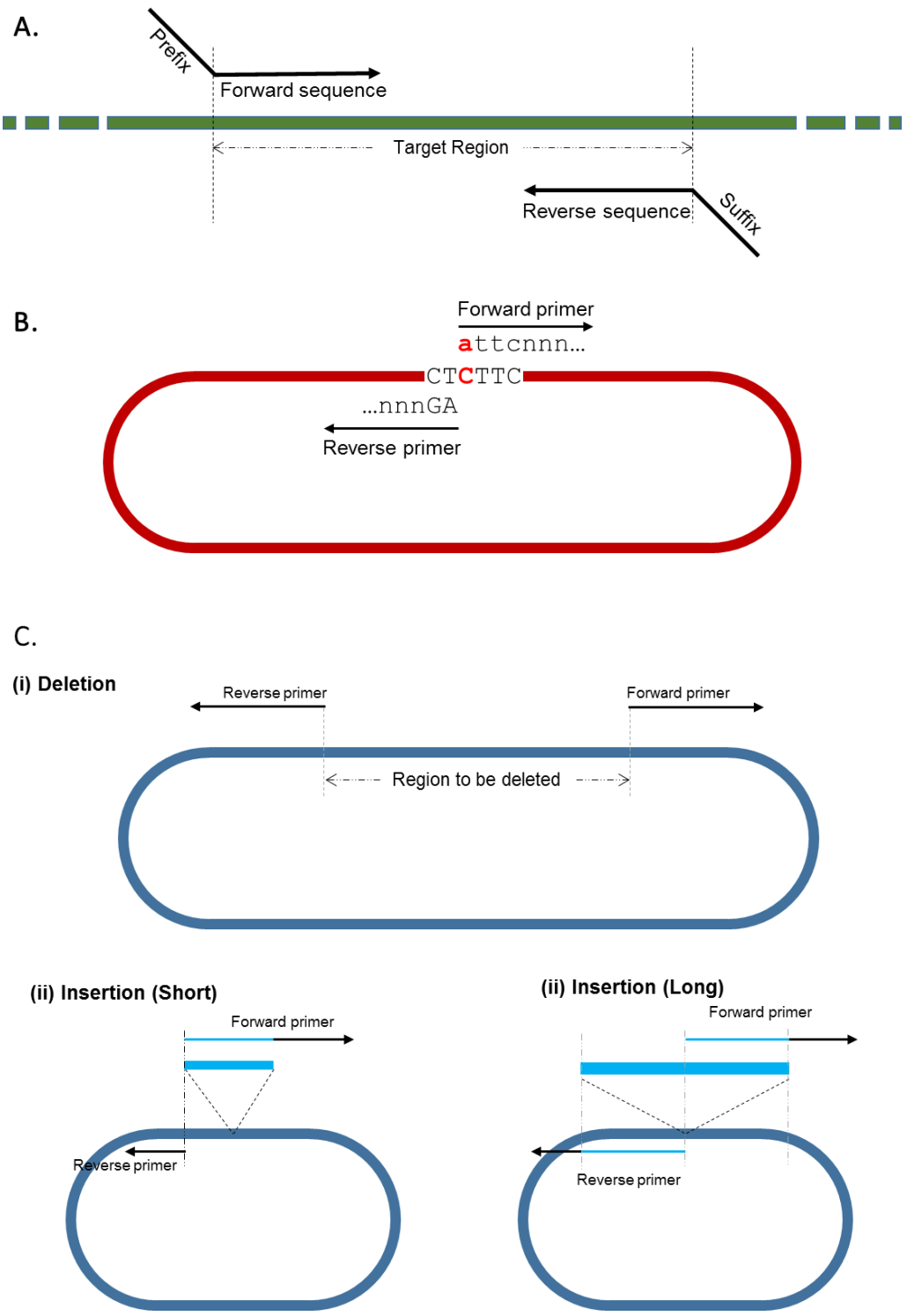



**Figure 2-1: Primer design principles and strategies**

(A) Primer design principle for amplification of parts for synthetic biology cloning. The standard prefix and suffix are included in the forward and the reverse primers respectively; (B) Primer design for removing unwanted restriction sites. In the

example shown the forward primer is designed such that the EarI recognition site is removed. Assuming the sequence of the recognition site is in frame, the third 'C' (CTC - Leu) is replaced with an 'A' maintaining it as a codon for Leu (CTA); (C) i. Primers for deletion of a region within a plasmid are designed such that they are divergent and exclude the unwanted region. ii. To insert a short sequence (eg. his-tag), the additional sequence is included as a tail to just one primer. iii. To insert longer sequences (e.g. an 18 amino acid long linker to fuse two proteins), the additional sequence is split into two with each half tailing the forward and reverse primers.

*2.2.3.2 Other oligonucleotides*

To add signal peptide coding sequences (about 96 bp or more) to a protein coding sequence, oligonucleotides were designed and annealed (see annealing protocol below - 2.2.4) with predefined overhangs compatible with the downstream cloning method. Each oligonucleotide was limited to a maximum length of 50 nucleotides. A schematic representation of the design principle for oligonucleotides is given in figure 2-2. A list of oligonucleotides used in this study is given in table 2-3.

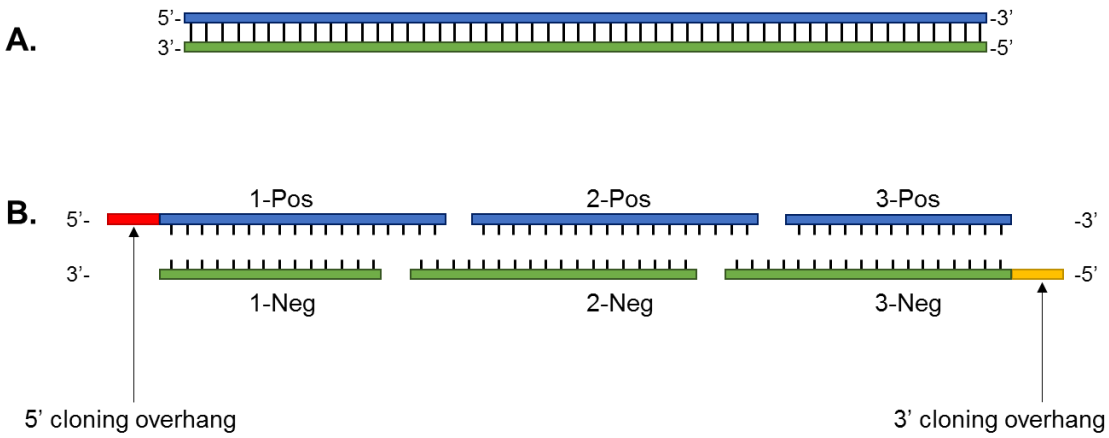



**Figure 2-2: Design of oligonucleotides for making DNA parts**

The DNA to be made (A) is designed as oligonucleotides (B). Internal overhangs of about 5 bp are left to facilitate annealing and correct positioning of each piece.

**Table 2-3: Sequences of oligonucleotides used**

| Name | Sequence |
|---|---|
| **A. Primers for making parts for cloning** | |
| T1SScf_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG gac gct acc ggc ggt aac g |
| T1SScf_r1 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC ggg cac aat gtg att ttg ttg c |
| GBchu2268_r2 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC ctc att aaa ata tat ttc tgt c |
| GBchu2268_f2 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG gag aaa aaa aca gaa gcg |
| GBcex_r1 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC gcc gac cgt gca ggg cgt gc |
| GBcex_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG ctg ccc gcc cag gcc gcg |
| GBcenA_r2 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC cca cct ggc gtt gcg cgc |
| GBcenA_f2 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG ggc tgc cgc gtc gac tac gc |
| gbCBD$_{Cex}$_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG tcc ggt ccg gcc ggg tgc |
| gbCBD$_{Cex}$_r1 | ATC CTGCAG CT AC TAG TA CTCTTC A GGC gc cga ccg tgc agg gcg tgc |
| gbCBD$_{CenA}$_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG ccc ggc tgc cgc gtc gac tac |
| gbCBD$_{CenA}$_r1 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC gc tgg tcg tcg gca cgg tgc |
| GBcfbglu_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG ggc gac cgg ttc cag cag gc |
| GBcfbglu_r1 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC ggg ctg gta ggt cgc ggc |
| GBcex_f2 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG cct agg acc acg ccc gca c |
| GBcenA_f3 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC tcc acc cgc aga acc gccg |
| TISS.TolC_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG aga aat gtc gcg cctg |
| TISS.HylD_r1 | ATC CTGCAG CT AC TAG TAC TCT TCA GGC gcg ctc acgcagcgcttc |
| T1SS.ABC_f1 | C GAATTC CT TCTAG ATG GCC TCT TCT TCG agcaacctcgattcacatgc |
| TISS.TolC_f2 | CGAATTCCTTCTAGATGG GCTCTTC TTCG aga aat gtc gcg cctg |
| TISS.HylD_r2 | ATCCTGCAGCTACTAGT GCTCTTC AGGC gcg ctc acg cag cgc ttc |
| T1SS.ABC_f2 | CGAATTCCTTCTAGATGG GCTCTTC TTCG agc aac ctc gat tca cat gc |
| | |
| **B. Primers for removing unwanted restriction sites** | |
| MABELcex_r1 | tt cgc gtt gat gcc ctc gac g |
| MABELcex_f1 | A agc aac tcg ctc tac gac ctc g |
| MABELcfbglu_f1 | a agc ctc ggc ctc cag gc |

| | |
|---|---|
| MABELcfbglu_r1 | tt cat gat ctc gac gtc ctc ggg |
| tolCEarI-mut_f1 | Aca gctc accg cctt aaa cggc attgg |
| tolCEarI-mut_r1 | tct tcgg cggc gtcc acca tttgc |
| abcEarI-mut_f1 | Aaa gag gcca gcgt aaca ggtta acg |
| abcEarI-mut_r1 | ggc gatg atga aaac atcta acg |
| abcEarI-mut_f2 | Act tcga ccat gtcga ccatt gtcg |
| abcEarI-mut_r2 | ggt gatc atctc gcga act gttcc agc |
| | |
| **C. Primers for deletions** | |
| CexΔSP_f2 | Atg ctgcc cgcc cagg ccgc gacc acg |
| CenAΔSP_f2 | Atg ggctg ccgcg tcgac tacg ccg |
| RBS-PlacZ_r1 | cta gta cct cct tga gct cta gta t |
| | |
| **D. Primers for insertions** | |
| GBsuffixF1 | gcctg aagag tacta gtagc ggccg ctgc |
| GBprefixR1 | cgaag aagag ccc at ctaga agcgg ccgcg aa |
| GBAccEF1 | tacta gagct cttct tcg ta ctagt agcgg ccgct gcag |
| GBAccER1 | ctcta gtact cttca ggc ct ctaga agcgg ccgca agttc |
| J04450f1 | caatacgcaaaccgcctctc |
| J04450r1 | tataaacgcagaaaggcccacc |
| RBS-B0034rev | CGGCGTAGGCTCTTCACGA catctagtatttctcctcttt |
| mRFP-E1010fwd | ACA CCC ACT GCTCTTC T ACA gcttcctccgaagacgttatc |
| GFP-E0040fwd | ACA CCC ACT GCTCTTC T ACA cgtaaaggagaagaacttttc |
| mRFP-E1010rev | CGG CGT AGG CTCTTCA TGT gcgatctacactagcactatc |
| GFP-E0040rev | CGG CGT AGG CTCTTC A TGT tttgtatagttcatccatg |
| Term-B0015fwd | ACA CCC ACT GCTCTTC T GCC taataatactagagccaggc |
| His.RBS.Plac_r1 | GTG ATG GTG GTG GTG ATG catatggta cctcct tgagct ctagta t |
| His.Cex_f1 | ctg cccg ccca ggcc gcga ccacg |
| His.CenA_f1 | ggc tgccg cgtc gactac gccg |
| morB-SP.RBS.PlacZ_r1 | ttgctaaagctggtatccggcat cta gta cct cct tga gct cta gta t |
| morB-SPcex_f2 | cccgggcctgtttaccccgctgcag ctgcccgcccaggccgcgaccacg |
| morB-SPcenA_f2 | cccgggcctgtttaccccgctgcag cccggctgccgcgtcgactacgc |
| PTcenA_r1 | GGG AGT TGG AGT TGT AGG TGT CGG ccacctggcgttgcgcgccatc |
| PTcexΔSP_f3 | ACA CCT CCC ACG CCT ACA CCC ACC ctgcccgcccaggccgcgaccacg |
| cexΔCBD_r3 | gctcgcgccgaaggcctccatc |
| pSB1X3_f1 | TAA TAA tactagtagcggccgctgc |
| PTcenB_r1 | GGG AGT TGG AGT TGT AGG TGT CGG atcagcggctcagccgcagacctc |

| | |
|---|---|
| PTcenC_r1 | GGG AGT TGG AGT TGT AGG TGT CGG gctgcgcggacgctgcacggcg |
| PTcenD_r1 | GGG AGT TGG AGT TGT AGG TGT CGG gcgcgtcgtgcaggtcgcgccg |
| cenA.PT_f1 | cccacgccgaccccgacgccgac |
| PTcbhA_r1 | GGG AGT TGG AGT TGT AGG TGT CGG GCCGAGCGTGCAGGCGACGCCG |
| PTcbhB_f1 | ACA CCT CCC ACG CCT ACA CCC ACC TCGTCAACGACCCGCCGGCGATC |
| CbhBΔCBD_r1 | GGTCTGCGGCGTCTTCGTGGTG |
| PTcfbglu_r1 | GGG AGT TGG AGT TGT AGG TGT CGG gggctggtaggtcgcggcgtc |
| PT.CBD(cenA)_f2 | ACA CCT CCC ACG CCT ACA CCC ACC cccggctgccgcgtcgactacg |
| PTchu2268_r1 | GGG AGT TGG AGT TGT AGG TGT CGG ctcattaaaatatatttctgtc |
| PT.CBD(cenB)_f2 | ACA CCT CCC ACG CCT ACA CCC ACC gcgcccacctacaactacgccg |
| PTcfbglu_f1 | ACA CCT CCC ACG CCT ACA CCC ACC ggcgaccggttccagcaggc |
| | |
| **E. Oligonucleotides for making parts or linkers** | |
| Pro-ThrCex HL | GCC ccgac gccga cgccc accac gccga ccccg acgccc acga |
| Pro-ThrCex HS | gtg gtggg cgtcg gcgt cgg |
| Pro-ThrCex EL | CGA cgtcg gggtc ggcgtc ggcgt cgtggg cgtcgg ggtc ggc |
| Pro-ThrCex ES | cgc cga cgcc gaccc cgacg |
| Pro-ThrCenA HL | GCC cccacg ccgacccc gacgccg acga ccccca cgccg acgccg acc |
| Pro-ThrCenA HS | gtcg gcgtcg gggtcgg cgtggg |
| Pro-ThrCenA EL | CGA cgtcggc gtgggg gtggggg tcgggg tcggc gtcg gcgtg ggggtc |
| Pro-ThrCenA ES | ccga ccccc accccc acgcc gacg |
| Pro-ThrsynKD1 HL | GCC ccgacacct acaactcc aactccc acacctc ccacgc ctacac |
| Pro-ThrsynKD1 HS | ggagtt ggagttg taggtgt cgg |
| Pro-ThrsynKD1 EL | CGA tgtgggagttgg ggtaggggtg ggtgtagg cgtggga ggtgtg |
| Pro-ThrsynKD1 ES | ccacc cctaccc caactcc caca |
| Pro-ThrsynKD2 HL | GCC ccaactc cgactccta ctccaa cccccactacg ccaacgccg |
| Pro-ThrsynKD2 HS | ggttg gagtaggagt cggagttgg |
| Pro-ThrsynKD2 EL | CGA cgtggtgg gtgtaggt gtgggagtc ggcgttggcg tagtggg |
| Pro-ThrsynKD2 ES | actcc cacaccta cacccac cacg |
| RBS linker1 HL | gccc accaccatca ccatcattga tactagag aaag |
| RBS linker1-2 HS | tgatgg tgatggtg gtg |
| RBS linker1 EL | CGA catctagta tttctcctc tttctctagt atcaa |
| RBS linker1 ES | agga gaaatactagatg |
| RBS linker2 HL | gcc caccac catcac catca ttgatactagagaact |
| RBS linker2 EL | CGA catctag tacctcc ttgagttc tctagta tcaa |
| RBS linker2 ES | caa ggagg tacta gatg |

| | |
|---|---|
| RBS-SPcenA.1pos | GCC ctcaa ggaggt actagatg tccaccc gcagaa ccgccg cag |
| RBS-SPcenA.1neg | cag cagcgc tgcggcgg ttctgcgg gtggacat ctagtacctc cttgag |
| RBS-SPcenA.2pos | Cgct gctggc ggccgc ggccgtc gccgtcgg cggtc |
| RBS-SPcenA.2neg | Ggtca gaccg ccgacgg cgacg gccgcg gccgc |
| RBS-SPcenA.3pos | tgac cgccctcac caccacc gccgcgc aggcg gct |
| RBS-SPcenA.3neg | CGA agccgc ctgcgcgg cggtggtgg tgagggc |
| RBS-SPcex.1pos | GCC ctcaag gaggtact agatgccta ggaccac gcccgc |
| RBS-SPcex.1neg | gccg ggtgcgg gcgtggt cctagg catcta gtacct ccttgag |
| RBS-SPcex.2pos | Acc cggccacc cggcccgc ggcgc ccgcacc gctctgc |
| RBS-SPcex.2neg | gtc gtgcg cagagc ggtgcg ggcgcc gcgg gccgg gtg |
| RBS-SPcex.3pos | gca cgacg ctcgcc gccg cggc ggcga cgctc |
| RBS-SPcex.3neg | cga cgacg agcgt cgccg ccgcgg cggcg agc |
| RBS-SPcex.4pos | gtc gtcggc gccac ggtcgt gctgcc cgccc aggcc gcg |
| RBS-SPcex.4neg | CGA cgcg gcctgg gcgg gcag cacg acc gtgg cgc |
| SPcenA.1v2pos | TCG tcca cccgca gaaccg ccgcag |
| SPcenA.1v2neg | cag cagcg ctgcg gcgg ttctgc gggt gga |
| SPcex.1v2pos | TCG ccta ggacca cgcc cgc |
| SPcex.1v2neg | gccg ggtgcg ggcgtg gtcct agg |
| RBSlinkerRev2 | cattctagtacctccttgagttctct |
| Pro-Thr KD3 HL | ACA ccg acacct acaactc caactcc cacacct cccacgc ctacac |
| Pro-Thr KD3 EL | TGT tgtgg gagttgg ggtaggg gtgggt gtaggc gtggga ggtgtg |
| SPcenA.3v2neg | TGT agccg cctgcg cggcgg tggtg gtga gggc |
| SPcex.4v2neg | TGT cgcgg cctggg cggg cagca cgaccg tggcgc |
| | |
| **F. Sequencing primers** | |
| pSBNX3insf2 | aaat aggcg tatca cgagg c |
| pSBNX3insr2 | cagt gagcg aggaa gcctg c |
| fusion.cex101_r | gacg aggtt gaact cgctg |
| fusion.CBDcenA_r | ggtg taggt ccagt cgagc |
| fusion.chu-Nterm_f | cgttt acga tcaa agcag |
| fusion.cenA-Nterm_f | ctgga cgcgc tcctg tgg |

### *2.2.3 Plasmids and plasmid design*

BioBrick high copy number plasmids pSB1C3 (2070 bp), pSB1T3 (2461 bp), and pSB1K3 (2204 bp) (Figure 2-3, Registry of Standard Biological Parts) were used for cloning. In addition blunt ended cloning where specified was performed with the Zero Blunt® PCR Cloning Kit (Life Technologies Ltd., UK) and the CloneJET PCR Cloning Kit (Thermo Fisher Scientific Inc., Lithuania). The origin of replication for the BioBrick plasmids is a pUC19-derived pMB1 which gives copy numbers from 100-300 per cell (Yanisch-Perron *et al.*, 1985).

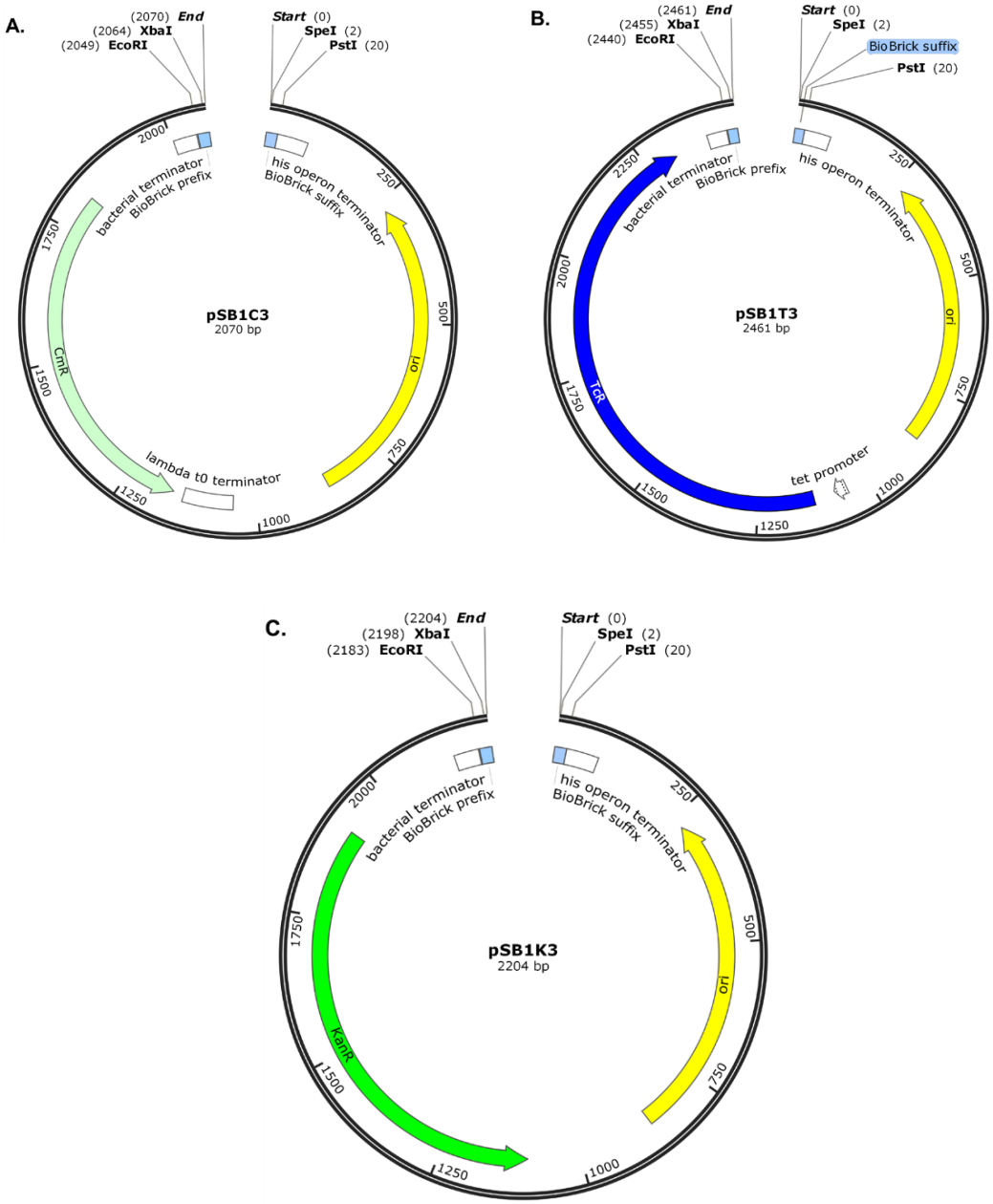



**Figure 2-3: Plasmid maps for pSB1C3, pSB1T3 and pSB1K3 showing major features**

The plasmids harbour antibiotic resistance genes for chloramphenicol ($Cm^R$), tetracycline ($Tc^R$) and kanamycin ($Kan^R$). Plasmid maps constructed with Snapgene Viewer Software (GSL Biotech LLC, USA).

To enable simultaneous cloning of multiple parts in BioBrick plasmids based on the Genabler technology attempts were made to modify the standard BioBrick prefix and suffix for pSB1C3 and pSB1T3 to make them Genabler compatible (thereinafter referred to as GenBrick plasmids). Genabler technology is a proprietary technology developed for Scottish Enterprise by Ginko Bioworks. The modification involved introduction of EarI sites into the prefix and suffix to make donor (pSB1C3-EarI) and acceptor (pSB1T3-EarI) plasmids. Each plasmid to be modified had a red fluorescent protein (RFP) coding device (BBa_J04450) cloned into its BioBrick cloning site. Two steps of mutagenesis with blunt end ligation (MABEL) (See section 2.2.5) were performed to insert the EarI sites. Primer pairs for the donor vector PCRs were GBsuffixF1/J04450r1 and GBprefixR1/J04450f1. Similarly, primer pairs for the acceptor vector PCRs were GBAccEF1/Jo4450r1 and GBAccER1/J04450f1. In each instance (donor and acceptor), one EarI site was successfully inserted into either the prefix or the suffix. None of the efforts to insert the second EarI site gave desired results irrespective of which site (prefix or suffix) was inserted first.

Following failure to insert the second EarI site into the BioBrick plasmids pSB1T3 and pSB1C3 to serve as acceptor and donor vectors respectively, a synthetic piece of double stranded DNA encoding a truncated *E. coli lacZ* gene (*lacZ'α*) was designed to contain the GenBrick prefix and suffix. The wild type *LacZ'α* had an internal EarI site and this was modified during the design of the synthetic DNA. The synthetic double stranded DNA was ordered from GENEART (Life Technologies, UK). The sequence of the *LacZ'α* is given in figure 2-4.

> **Synthetic LacZ'alpha for GenBrick Cloning**

cagaaatcatccttagcgaaGAATTCGCGGCCGCTTCTAGATGGGCTCTTCT**TCG**gcag
tgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacact
ttatgcttccggctcgtatgttgtgtgaaattgtgagcggataacaatttcacacagga
aacagctatgaccatgattacggattcactggccgtcgttttacaacgtcgtgactggg
aaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctgg
cgtaatagcgaagaagcccgcaccgatcgcccttcccaacagttgcgcagcctgaatgg
cgaatggcgctttgcctggtttccggcaccagaagcggtgccggaaagctggctggagt
ga**GCC**TGAAGAGTACTAGTAGCGGCCGCTGCAGtccggcaaaaaagggcaagg

**Figure 2-4: Sequence of the synthetic *lacZ′* for engineering GenBrick donor a vector**

The prefix sequence contained a forward facing EarI site which leaves a 3' overhang (5'-CGA-3') on the bottom strand. The suffix contained a reverse facing EarI site which leaves a 5'-GCC-3' overhang on the 3'-end of the top strand.

The synthesized DNA was digested with EcoRI/PstI. Similarly, pSB1C3-BBa_J04450 was digested with same enzymes. The digested products were cleaned using the QIAquick PCR purification kit and ligated overnight at 16°C. The ligation was used to transform *E. coli* JM109 competent cells. The transformed cells were plated onto LA with chloramphenicol, IPTG and X-Gal. Blue colonies were picked and screened for the correct insert and restriction sites. To do this, selected colonies were inoculated into 5 ml LB and grown overnight at 37°C with shaking at 200 rpm. Plasmid DNA was extracted using the QIAprep spin miniprep kit. The extracted DNA was digested with EarI and electrophoresed on 0.8% agarose gel in 1X TAE (Tris-acetate-EDTA buffer) buffer. Minipreps that gave the correct size bands were confirmed by sequencing. Subsequent work on developing a wide range of donor and acceptor plasmids for GenBrick assembly was performed by Lina Gasiūnaitė, an MSc student who worked

with me during the initial phases. In addition to these plasmids, other modifications to the pSB1C3 plasmid with BBa_J04450 or a green fluorescent protein coding device BBa_K741002 were made. These modifications (described in detail in methods, Chapter 3) were to enable the fusion of other proteins to the N- or C- termini of the RFP or GFP proteins.

A list of plasmids used in this study has been given in the table 2-4.

**Table 2-4: Plasmids used in this study**

| Name | Description | Source |
|---|---|---|
| pSB1C3-BBa_K523015 | $P_{lac}$+*lacZ'α*+RBS+*cenA* | This study |
| pSB1C3-BBa_K523016 | $P_{lac}$+*lacZ'α*+RBS+*cex* | This study |
| pSB1C3-BBa_K523025 | $P_{lac}$+*lacZ'α*+RBS+*cxnA* | C.E. French |
| pSB1C3-BBa_J15509 | $P_{lac}$+*lacZ'α*+RBS+*cex*+RBS+*cenA* | This study |
| pSB1C3-BBa_J33207 | $P_{lac}$+RBS+*lacZ'α* | C.E. French |
| pSB1C3-BBa_J04450 | pSB1C3 with mRFP cassette | Registry of standard Biological Parts |
| pDBM2091 | pSB1C3 with $P_{lac}$+*lacZ'α*+RBS+*cenA*+RBS+*cex* | (Barnard, 2012) |
| pKDCenACex | pDBM2091 with fused *cenA/cex* | This study |
| pDBM2093 | pSB1C3 with $P_{lac}$+*lacZ'α*+RBS+*cenB*+RBS+*cex* | (Barnard, 2012) |
| pKDCenBCex | pDBM2093 with fused *cenB/cex* | This study |
| pDBM2095 | pSB1C3 with $P_{lac}$+*lacZ'α*+RBS+*cenD*+RBS+*cex* | (Barnard, 2012) |
| pKDCenDCex | pDBM2095 with fused *cenD/cex* | This study |
| pKDCxnA(NF)Cfbglu | pSB1C3 with $P_{lac}$+*lacZ'α*+RBS+*cxnA*+RBS+*cfbglu* | This study |
| pKDCxnACfbglu | pSB1C3 with fused *cxnA/cfbglu* | This study |
| pDBM1350 | pSB1K3 with $P_{lac}$+*lacZ'α*+*cfbglu*+RBS+*cenA* | (Barnard, 2012) |
| pKDCfbgluCenA | pSB1K3 with fused *cfbglu/cenA* | This study |
| pASChu2268(NF)CenA | pSB1C3 with $P_{lac}$+*lacZ'α*+RBS+*chu2268*+RBS+*cenA* | A.A. Salinas-Vaccaro |
| pKDChu2268CenA | pSB1C3 with fused *chu2268/cenA* | This study |

All plasmids have chloramphenicol resistance except pDBM1350 and pKDCfbgluCenA which have kanamycin resistance

### *2.2.4 Restriction digests, ligations and transformations*

*2.2.4.1 Restriction digests and ligations*

Restriction digests were conducted in 30 µl reaction volume. For a single digest, the reaction mixture was made up of 21 µl deionized water, 3 µl restriction enzyme buffer, 1 µl restriction enzyme and 5 µl DNA. Initial experiments using 3 µl DNA often resulted in faint or non-visible bands for small inserts although the plasmid backbone was clear. The 5 µl volume of DNA was therefore chosen as the amount of plasmid DNA to include in digests in order to give visible bands for small inserts. For double digests, the water volume was reduced to 20 µl and 1 µl of the second enzyme was added. Modifications to this general restriction digest protocol were made where necessary. Such modifications are described into detail where mentioned.

Generally, ligation reactions were set up in 20 µl reaction mixtures unless otherwise stated. This included 14 µl deionized water, 2 µl ligase buffer, 1 µl ligase and 3 µl DNA. For blunt end ligation reactions, 1 µl polynucleotide kinase (PNK) was added and water adjusted accordingly. Ligation reactions were incubated at 16°C overnight unless otherwise stated. To ligate oligonucleotides into double stranded DNA, the oligonucleotides were first phosphorylated as follows:

- 20 µl of 100 µM sense oligonucleotide
- 20 µl of 100 µM antisense oligonucleotide
- 0.5 µl of T4 PNK
- 0.5 µl of 10 mM ATP
- 5 µl of 10X PNK buffer

The reaction mixture was incubated at 37°C for 30 minutes after which 5 µl 5M NaCl was added. To anneal the oligonucleotides, the mixture was placed in a heat block

pre-heated to 98°C. The block was switched off for gradual cooling overnight. To make up linkers for GenBrick assembly, the reaction mixture was diluted to 0.5 µM (1.38 µl of annealed oligonucleotides added to 98.62 µl of elution buffer, EB) in order to add to reactions. To anneal multiple pairs to form a DNA part, the annealed oligo pairs were ligated overnight using T4 DNA ligase. The T4 DNA ligase and T4 PNK were from New England Biolabs whereas the ATP was obtained from Sigma Aldrich.

*2.2.4.2 Competent cell preparation and transformation*

Competent cells were prepared using a one-step protocol (Chung *et al.*, 1989). The chosen strain was grown in 5 ml LB overnight to make up the starting culture. Overnight culture (250 µl) was used to inoculate 50 ml of LB in a 250 ml conical flask without antibiotics. The culture was grown at 37°C, 200 rpm until the cells reached optical density (OD 600 nm) between 0.3 and 0.5. The culture was transferred immediately to a 50 ml centrifuge tube and chilled on ice. The chilled tubes were centrifuged at 4,000 x g for 5 minutes and the supernatant discarded. The pellets were resupended in 5 ml transformation and storage solution (TSS) and left on ice for 30 minutes. Cells (100 µl) were pipetted into 1.5 ml tubes and either transformed immediately or stored at -80°C. The TSS consists of 17 ml LB, 5 ml 40% w/v polyethylene glycol (PEG 3350), 1 ml of 1 M $MgCl_2$ and 1 ml DMSO. TSS components (except DMSO) were autoclaved prior to mixing. TSS was stored at 4°C.

To transform cells, 2 µl of plasmid DNA or 5 µl of ligation was added to fresh or thawed (on ice) competent cells. The tube was mixed by flicking three times and placed on ice for 45 minutes. The cells were heat-shocked in a water bath at 42°C for 60 seconds and returned to ice for a further 60 seconds. LB (900 µl) was added to each tube and mixed by inversion. The tubes were incubated (positioned sideways) at 37°C, 200

rpm for one hour. Cells (100 µl) were plated on LA with antibiotics, IPTG or X-Gal as appropriate and incubated at 37°C overnight. Colonies were then selected for screening.

### *2.2.5 Mutagenesis with blunt end ligation (MABEL)*

The MABEL protocol was developed by Prof. Christopher E. French (University of Edinburgh). The protocol applied to plasmids (Figure 2-5) and was used as follows:

- removing unwanted restriction sites by changing a base
- single or multiple (continuous) nucleotide deletions
- insertion of up to 64 bp of DNA. eg. as linkers for making fusion proteins

Two divergent non-overlapping primers were designed as described (section 2.2.3). PCR was performed with the primers using Q5 DNA polymerase according to the manufacturer's instructions. An aliquot (10 µl) of the PCR product was checked for successful amplification by agarose gel electrophoresis. The remainder of the PCR product was purified from the reaction mixture and used to set up a 20 µl ligation reaction as described (section 2.2.4). Prior to ligation, the reaction mixture was treated with T4 polynucleotide kinase. The ligation mixture was used to transform *E. coli* JM109 competent cells and plated on LA with appropriate antibiotics. About five colonies were selected and grown overnight in 5 ml of LB. Plasmid minipreps were prepared and subjected to restriction digest to check for the change. Digests that appeared correct were then sequenced for confirmation.

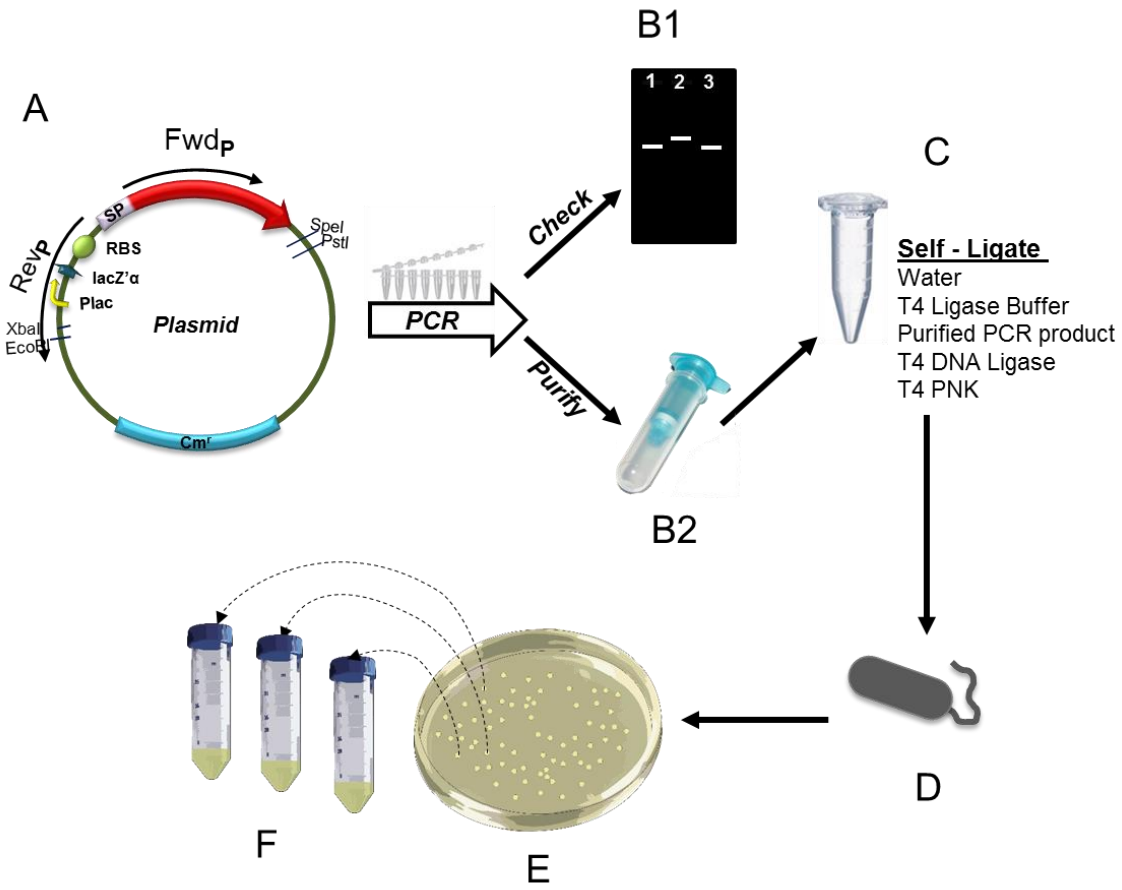



**Figure 2-5: Illustration of the mutagenesis with blunt end ligation protocol**

The MABEL protocol is used to illustrate a deletion of piece of DNA (SP) in a plasmid. Divergent non-overlapping primers ($Fwd_p$ and $Rev_p$) are used for PCR. An aliquot of the PCR is checked on a gel (B1) and the remainder purified (B2). The purified product is then self-ligated (C) and used to transform *E. coli* competent cells (D). Colonies are selected on agar plates, minipreped (F) and confirmed by restriction digest and sequencing.

### *2.2.6 Agarose gel electrophoresis of DNA*

Products from restriction digests and PCR as well as other forms of DNA (synthesised and ligated oligonucleotides) were analysed by agarose gel electrophoresis on gels containing GelGreen™ nucleic acid gel stain (Biotium Inc., USA). To incorporate stain into agarose gels, 10 µl of the 10,000X stock solution was added to 100 ml of molten agarose according to the manufacturer's protocol. Agarose gels were prepared to a

concentration of 0.8% (w/v) agarose in 1X TAE buffer. A higher percentage of agarose (1.5% (w/v)) was used when products of shorter lengths (< 800 bp) were expected. To load DNA onto agarose gels, 10 µl of PCR product was mixed with 2 µl of 6x purple gel loading dye (New England Biolabs Inc., USA) on a piece of parafilm. The mixture was loaded directly into wells within the agarose gel. A stock solution of 20X TAE buffer (96.8 g/l Tris base; 7.44 g/l sodium-EDTA; 22.8 ml glacial acetic acid) was diluted to 1X using distilled deionised water. The 0.8% agarose (w/v) gels were electrophoresed at 70 V for 40 to 60 minutes. Lower or higher percentages were electrophoresed at lower voltage/longer time or higher voltage/shorter time respectively.

### *2.2.7 DNA sequencing & bioinformatics*

Plasmid DNA preparations or clean PCR products were sequenced using the Sanger DNA sequencing method. The Big-Dye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, UK) was used to perform cycle sequencing reactions in a 10 µl reaction volume (2 µl 5X sequencing buffer, 1 µl 3.2 pmol forward or reverse primer, 1 µl big-dye reagent 2 to 6 µl DNA). Nuclease free water was used to adjust the reaction volume to 10 µl. Cycling conditions and the amount of DNA was determined according to the manufacturer's instructions. The sequencing reaction was then sent to Edinburgh Genomics, University of Edinburgh for capillary electrophoresis. Chromatograms from the sequencing were analysed and trimmed manually. The sequence was then aligned with the expected sequence using blastn from Basic Local Alignment Search Tool (BLAST) on the National Centre for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

## 2.3 Cultivation and enumeration of bacteria

### *2.3.1 Media and growth conditions*

Luria agar (LA) and broth (LB) (Bertani, 2004), and M9 minimal medium (Cold Spring Harb., 2010) with various supplements were used in the cultivation and selection of bacteria.

Luria broth and agar were obtained from the media suite of the School of Biological Sciences, Darwin Building, University of Edinburgh. The broth was used directly whereas the agar was melted in a microwave (Sharp R-372M, 25 L, 900 W, 2450 MHz). For 250 ml of LA the microwave was set to 30% power for 20 minutes. The melted agar was placed in a water bath at 55°C for about 15 minutes to cool. Appropriate antibiotics and other supplements (IPTG or X-Gal) (table 2-5) were added and mixed just before pouring into petri dishes. The concentrations and sources of antibiotics are given in table 2-5. The recipe for the LB and LA are given in table 2-6.

M9 minimal medium was prepared (table 2-6) and used for cellulose utilization, protein localization and qualitative (plate) activity experiments. For *E. coli* cultures, M9 medium was prepared and supplemented with 0.2 g/l yeast extract and 4 g/l glucose. M9 was chosen over LB for protein localization studies because LB often produced very high background fluorescence measurements.

**Table 2-5: Antibiotic and Chemical Supplements**

| Antibiotic/Chemical | Final Concentration | Supplier |
|---|---|---|
| Chloramphenicol | 40 µg/ml † or 15 µg/ml‡ | Duchefa Biochemie, The Netherlands |
| Ampicillin | 100 µg/ml | Melford Laboratories Ltd, UK |
| Carbenicillin | 80 µg/ml | Melford Laboratories Ltd, UK |
| Kanamycin | 50 µg/ml | AppliChem GmbH, Germany |
| Tetracycline | 5 µg/ml | Duchefa Biochemie, The Netherlands |
| IPTG | 90 µg/ml | Sigma Aldrich Inc., Germany |
| X-Gal | 90 µg/ml | Duchefa Biochemie, The Netherlands |

†Concentration used for *E. coli;* ‡Concentration used for *C. freundii*.

**Table 2-6: Recipes for media preparation**

| Media/Reagent | Components | Quantities |
|---|---|---|
| Luria Broth (LB) and Luria Agar (LA)* | Bacto Tryptone<br>Oxoid Yeast Extract<br>NaCl<br>NaOH<br>Oxoid Agar No. 3*<br>Distilled Water | 10 g<br>5 g<br>10 g<br>0.2 g<br>15 g*<br>1 L |
| M9 Minimal Medium (Broth and Agar*) | 4x M9 Salts<br>1M $CaCl_2$<br>1M $MgSO_4$<br>Carbon Source<br>Yeast extract<br>Agar*<br>Distilled water | 25 ml<br>10 µl<br>200 µl<br>Varying†<br>Varying‡<br>15 g/L*<br>Up to 100ml |
| 4x M9 Salts | $Na_2HPO_4$<br>$KH_2PO_4$<br>NaCl<br>$NH_4Cl$<br>Water | 28 g<br>12 g<br>2 g<br>4 g<br>1 L |
| 1x PBS | $KH_2PO_4$<br>NaCl<br>$Na_2HPO_4$<br>Water | 0.21 g<br>9 g<br>0.726 g<br>1 L |

*Agar was only added for solid plate media (LA and M9 agar plates). †Carbon source provided was either glucose (positive controls) or cellulose filter paper. No carbon source was added to negative controls. For expression of proteins and activity testing on plates, 4 g/L glucose was added. ‡The amount of yeast extract (YE) depended on the preparation; 0.2 g/L for expression of proteins and plate activity tests, 1 g/L for cellulose utilization experiments. All media were autoclaved at 121°C, for 15 minutes. LB, M9 salts and PBS were made by staff at the Darwin Media Suite (Level 8), University of Edinburgh.

### *2.3.2 Monitoring and quantification of growth*

Cell growth was monitored by measuring the OD of cells at 600 nm in cuvettes (FisherBrand, Cat. No. FB55147). To quantify cells, additional measurements were

included. However, in cultures containing insoluble cellulosic particles, the OD600 was not considered to give a reliable indication of cell growth.

*2.3.2.1 Optical density measurement*

To perform the OD600 measurements, 100 µl of the culture was pipetted into a cuvette. Phosphate buffered saline (900 µl) (PBS, table 2-6) was added. The cuvette was covered with a piece of parafilm and mixed by inversion. 1 ml of PBS was used as blank against which the OD of cultures were measured. Measurements were performed using the 600 nm absorbance module with the Modulus™ Single Tube Multimode Reader (Turner BioSystems, Inc.).

*2.3.2.2 Viable cell counts (cfu)*

Viable cells were quantified as colony forming units (cfu) per millilitre of culture sample. To perform viable counts, cultures were serially diluted in 1.5 ml tubes and spotted onto LA plates. 10 µl of each dilution was spotted in quadruplicate on plates using the map in figure 2-6.

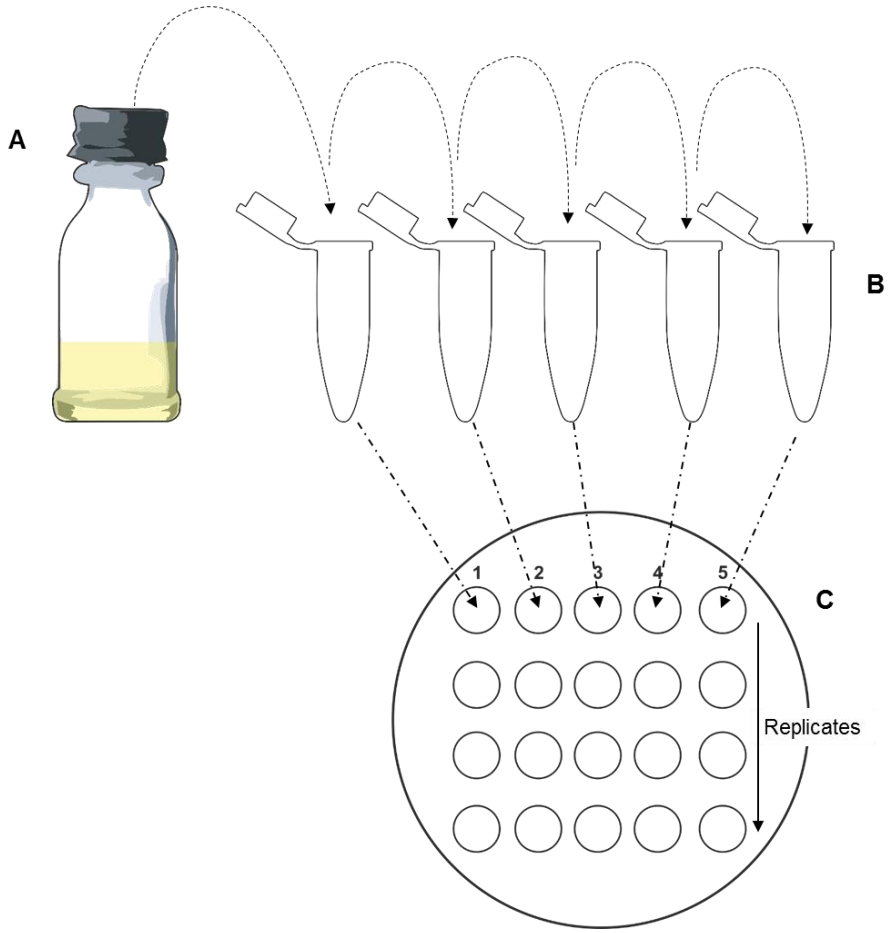



**Figure 2-6: Scheme for colony count and replicates**

Serial dilutions of the culture (A) are made in 1.5 ml tubes (B) and plated by spotting 10 µl in replicates.

*2.3.2.3 Total protein*

Total protein content of cultures was used as measurement of growth according to a protocol developed by C. French (manuscript in preparation). The protein concentration was determined using the Pierce™ Coomassie Plus (Bradford) Assay Kit (Thermo Scientific Inc., Cat. 23236). A sample of well mixed culture (100 µl) was mixed with 900 µl reagent in a cuvette. The mixture was incubated at 65°C for 1 hour to lyse cells. The cuvettes were allowed to cool down for 1 hour after which the OD was measured at 600 nm using the Modulus system with water used as blank. Protein

concentration of dilutions of bovine serum albumin were prepared as a standard to estimate the actual protein concentration of samples.

*2.3.2.4 Live/Dead cell staining*

Live/dead staining was performed using SYBR Green I nucleic acid gel stain (SYBR-I) and propidium iodide (PI). A sample of culture (100 µl) was diluted with 900 µl 1x PBS. The mixture was then stained with 10 µl each of SYBR-I (1:30 dilution of 10,000X stock) and PI (1 mg/ml) and incubated at room temperature for 15 minutes in the dark. Fluorescence (green) of live cells were measured using the blue filter on the Modulus™ Single Tube Multimode Reader. SYBR-I and propidium iodide were obtained from Life Technologies, UK.

## 2.4 Protein expression and purification

### *2.4.1 Growth and induction conditions*

All constructs were under the regulation of the *lac* promoter BBa_J33207 (Registry of Standard Biological Parts). To express proteins, cultures were made by inoculating fresh LB (with appropriate antibiotics) with an overnight culture (1:200 dilution). A vessel of volume at least five times the volume of culture was used. Cells were grown at 37°C with shaking at 250 rpm until reaching OD600 between 0.4 and 0.6 at which point they were induced with IPTG to a final concentration of 0.38 mM. Induction was continued for 16 hours after which cells were harvested.

### *2.4.2 Preparation of lysates and fractions*

The harvested cells were centrifuged at 16,000 x g for 15 minutes. The supernatant was collected and analysed for secreted enzymes. To prepare total cell lysates, the

pellets were resuspended in 1X PBS and sonicated for 3 pulses at 10 µm (amplitude) for 30 seconds on ice. Samples were cooled for one minute in between pulses on ice. The sonicated samples were centrifuged at 16,000 x g for 10 minutes and the supernatant collected as the total cell lysate. To separate cytoplasmic and periplasmic fractions, cell pellets from 50 ml of cell culture were harvested by centrifugation at 16,000 x g for 15 minutes (Rutter *et al.*, 2013). The cell pellets were resuspended in 5 mL of shock buffer (0.5 mM sodium ethylenediaminetetraacetic acid (EDTA), 100 mM Tris–HCl, pH 8.0, phenylmethylsulfonyl fluoride (PMSF) (20 µg/ml) and 0.5 M sucrose). The suspension was incubated on ice for 5 min and then centrifuged at 16,000g for 5 min. Pellets were warmed to room temperature and resuspended in 5 mL of ice-cold water. After 1 min on ice, 283 µl of 20 mM $MgCl_2$ was added. The osmotically shocked cells were centrifuged at 16,000g for 5 min and the supernatant was collected as the periplasmic fraction. The cell pellet was lysed by ultrasonication and centrifuged at 16,000g for 5 min. The supernatant was collected as the cytoplasmic fraction.

### 2.5 Enzyme activity assays

Qualitative determination of endo- and exo- glucanase activities was performed using the Congo Red carboxymethyl cellulose (CMC) assay and the 4-methylumbelliferyl β-D-cellobioside (MUC) assay respectively. Quantitative exoglucanase activity assays were performed with MUC whereas endoglucanase activity assays were performed with the Azo CM-Cellulose assay (Megazyme, Ireland).

Activity assays were performed either on live cells, cell lysate or cell culture supernatant. To obtain cell lysate or supernatant, a fresh LA plate was inoculated with glycerol stocks of the sample and incubated at 37°C overnight. A colony from the plate

was then transferred into 5ml LB supplemented with appropriate antibiotics and grown overnight at 37°C on a shaker. The following morning, 100ml of fresh LB was warmed at 37°C for 30 minutes in a conical flask. 1 ml of the overnight culture was then used to inoculate the pre-warmed LB and allowed to grow at 37°C for 2 hours. The absorbance at 600nm ($A_{600}$) was determined against a blank (LB) and the culture was continued if the $A_{600}$ was less than 0.6. Otherwise, 50 ml was transferred into each of two new flasks (also pre-warmed at 37°C). To one flask, 0.38 mM final concentration of IPTG was added and marked induced. Both cultures were grown for a further 4 hours. After growth, cultures were transferred into 50 ml centrifuge tubes and centrifuged at 13,000 × g for 10 minutes. About 10 ml of the supernatant was transferred into fresh sterile tubes and the rest discarded. The pellets were re-suspended in 5 ml of 50% (v/v) glycerol-PBS solution. Lysis of cells was performed by sonication as described above.

### *2.5.1 Exoglucanase activity*

For qualitative determination of exoglucanase activity in live cell cultures, LA was prepared with incorporation of an appropriate antibiotic as indicated earlier. MUC stock solution was prepared in 80% dimethyl sulfoxide (DMSO) to a concentration of 5 mg/ml and 100 µl was pipetted onto the agar plate and spread. The plate was allowed to dry in a laminar flow high-efficiency particulate air (HEPA) filtered cabinet. Cells to be tested were streaked onto the plates and incubated at 37°C overnight. Release of 4-methylumbeferone (4-MU) occurs when the exoglucanase hydrolyses the 4-MUC. Fluorescence of the released 4-MU was detected under long wavelength (365nm) blue light.

Quantitative assays were performed on cell lysate or supernatant by adding 50 µl of cell lysate to a UV grade cuvette (Fisher Scientific, CXA-110-005J). PBS (1.450 ml) containing MUC (final concentration 0.2mM) was added. The cuvette was covered with parafilm and inverted about 6 times to mix. Fluorescence was measured using the UV filter in the Modulus™ Single Tube Multimode Reader, after which the reaction was incubated at 37°C for 60 minutes and fluorescence measured again. A blank solution was prepared like samples but with distilled water. The value of blanks was subtracted from samples. Dilutions of 4-MU were prepared and measured as a standard curve to determine the concentration of 4-MU released in samples.

### *2.5.2 Endoglucanase activity*

For qualitative determination of endoglucanase activity, carboxymethyl cellulose (CMC) was added to LA to a final concentration 0.2% (w/v) CMC and autoclaved. Prior to pouring into plates, appropriate antibiotics and IPTG were added as indicated earlier. Cells to be tested were streaked onto the plates and grown at 37°C overnight. Plates were flooded with 5 ml of 0.1% Congo Red followed by 5 ml of 1M NaCl. Endoglucanase activity was detected as zones of clearing around the colonies (Chaudhary *et al.*, 1997).

For quantitative determination of endoglucanase activity, the Azo CM-cellulose assay (Megazyme, Ireland) was performed according to manufacturer's protocol with modifications. Cell lysate (0.4 ml) or supernatant (0.4ml) to be tested was added to 0.4 ml of Azo-CM cellulose reagent. The mixture was mixed by vortexing after which 300 µl was collected into a 1.5 ml tube (time 0) and the reaction stopped by the addition of 750 µl of absolute ethanol. The remainder was incubated at room temperature for 60 minutes. The stopped reaction was vortexed and centrifuged at

15,000xg for 10 minutes. The supernatant containing depolymerized low molecular weight substrate dyed fragment was collected into a cuvette and the absorbance determined at 600 nm using distilled water as blank. Dilutions of Remazol Brilliant Blue R were prepared and measured as a standard curve to determine the amount of dye released.

### *2.5.3 Assay standards*

Assay standards were prepared to convert fluorescence and optical density readings to enzyme units as well as determine the detection ranges for reliable quantification with each assay. Standard curves were prepared for MUC and CMC activity assays using 4-Methylumbelliferone (4-MU) (Sigma Aldrich) and Remazol Brilliant Blue R (RBB) (Sigma Aldrich). An initial stock of 100 mM 4-MU was prepared in 80% v/v DMSO. The stock was then diluted in concentrations from 10 mM to 0 using 1x PBS and the fluorescence measured as done for samples. The RBB standards were prepared in similar concentrations using water as diluent and blank. Absorbance was measured at 590 nm.

## 2.6 Statistical analysis

Data was entered into Microsoft Excel 2013 and exported into IBM SPSS version 21 for analysis. Data was summarized as graphs (box plots, frequency distributions and scattergrams) and where necessary inferential statistics performed. Appropriate measures of centrality (mean, median) and of dispersion (standard deviation, standard error) were calculated. Prior to statistical hypothesis testing, the data was checked for normality using the Shapiro–Wilk test and Q-Q plot in order to determine the appropriate statistical tests to employ. Statistically significant differences and/or

associations between categorical data were analysed using student t-tests or their non-parametric analogues. Regressions and correlations were used to determine independent associations.

# Chapter 3

## Characterization of a *Cellulomonas fimi* exoglucanase-endoglucanase fusion protein for degradation of cellulosic biomass

### 3.1 Summary

Effective degradation of cellulose requires multiple classes of enzyme working together. However, naturally occurring cellulases with multiple catalytic domains seem to be rather rare in known cellulose-degrading organisms. A fusion protein made from *Cellulomonas fimi* exo- and endo- glucanases, Cex and CenA which improves breakdown of cellulose is described. A homologous carbohydrate binding module (CBM-2) present in both glucanases was fused to give a fusion protein CxnA. CxnA or unfused constructs (Cex+CenA, Cex, or CenA) were expressed in *Escherichia coli* and *Citrobacter freundii*. The latter recombinant strains were cultured at the expense of cellulose filter paper. The expressed CxnA had both exo- and endo-glucanase activities. It was also exported to the supernatant as were the non-fused proteins. In addition, the hybrid CBM from the fusion could bind to microcrystalline cellulose. Growth of *C. freundii* expressing CxnA was superior to that of cells expressing the unfused proteins. Physical degradation of filter paper was also faster with the cells expressing fusion protein than the other constructs. Our results show that fusion proteins with multiple catalytic domains can improve the efficiency of cellulose degradation. Such fusion proteins could potentially substitute cloning of multiple enzymes as well as improving product yields.

## 3.2 Background

About 75% of the total biomass of lignocellulosic materials is made of cellulose and hemicellulose, making cellulose the most suitable feedstock for production of biofuels and renewable feedstock chemicals (Bayer *et al.*, 2007; Gomez *et al.*, 2008; Carroll and Somerville, 2009). Cellulose can be degraded naturally and used as a source of energy by various bacteria and fungi. Such organisms produce endo-β-1,4-glucanases (EC 3.2.1.4), exo-β-1,4-glucanases or cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) (Knowles *et al.*, 1987; Van Dyk and Pletschke, 2012). The combined action of these enzymes results in saccharification of cellulose. Endoglucanases and cellobiohydrolases break the long cellulose into short-chain units (cellobiose and other short oligonucleotidesaccharides) which are hydrolysed by β-glucosidases into glucose (Xiao *et al.*, 2004; Van Dyk and Pletschke, 2012). The endoglucanases and cellobiohydrolases usually consist of two functional units, the catalytic domain (CD) which hydrolyses the β-1,4-glycosidic bonds in cellulose and a non-catalytic CBM which binds to cellulose. In addition to their substrate binding function, CBMs have been shown to improve the enzymatic activity of their cellulases. CBMs may loosen individual cellulose chains from the cellulose surface prior to hydrolysis or enhance the solubilisation of individual glucan chains of the cellulose surface. In addition, some CBMs also serve as thermostabilizing domains (Teeri *et al.*, 1992a; Teeri *et al.*, 1992b; Linder and Teeri, 1996; Hall *et al.*, 2011).

The cellulases of *Cellulomonas fimi* are well studied, and the genome has been sequenced (Christopherson *et al.*, 2013). Two of the best studied cellulases of *C. fimi* are the endoglucanase CenA and the bifunctional exoglucanase/xylanase Cex. The N-terminal family-2 CBM of CenA, unlike those of other endoglucanases, has been shown to disrupt the structure of cellulose fibres, resulting in the release of fine

cellulose particles without any detectable hydrolytic activity (Din *et al.*, 1994). Cex possesses a homologous CBM at its C-terminus. The N-terminal GH10 catalytic domain of Cex hydrolyses cellulose and cellotetraose to release cellobiose from non-reducing chain ends (Kleman-Leyer *et al.*, 1994; White *et al.*, 1994), and has also been shown to possess endoglucanase activity as well as activity against xylan (MacLeod *et al.*, 1994; Notenboom *et al.*, 1998; Notenboom *et al.*, 2000). In both CenA and Cex, the CBM is linked to the respective catalytic domain by a proline-threonine (PT) linker. Whist the PT linker acts as a flexible tether, the CBM anchors the catalytic domains onto amorphous, semi-crystalline and crystalline cellulose chains enabling the catalytic domains to hydrolyse nearby cellulose chains (Gilkes *et al.*, 1991b; Ong *et al.*, 1993b; Creagh *et al.*, 1996; McLean *et al.*, 2000; Poon *et al.*, 2007). In addition the CBM initiates mobility and facilitates the diffusion of the catalytic domain (Jervis *et al.*, 1997b).

Gene fusion is an essential tool in systems and synthetic biology, and has been widely used in the production of synthetic bifunctional enzymes (Bulow, 1987; Ljungcrantz *et al.*, 1989; Tamada *et al.*, 1994; Seo *et al.*, 2000). In cellulose hydrolysis, gene fusion has been used in studying functions and effectiveness of various CBMs (Murashima *et al.*, 2003; Thongekkaew *et al.*, 2013; Zhao *et al.*, 2013) as well as to create enzymes with multiple activities (An *et al.*, 2005; Bower *et al.*, 2006; Lu *et al.*, 2006). Warren and colleagues described a fusion of Cex and CenA made by the use of restriction enzymes. Although the resulting fusion had both endoglucanase and exoglucanase activities, it lacked the flexible PT linkers of both enzymes and all of the CBM of Cex, and most of the CBM of CenA. It did not bind to cellulose, but, like native CenA and Cex, was translocated to the periplasm when expressed *in E. coli* (Warren *et al.*, 1987). A xylanase-endoglucanase gene fusion of *Clostridium thermocellum xynX*

gene and an enhanced *Erwinia chrysanthemi PY35 cel5Z*::Ω gene has been constructed using restriction independent overlapping (An *et al.*, 2005).

Although various fusions of cellulases have been reported, their effectiveness for cellulose degradation *in vivo* has not been examined. It has been previously demonstrated that recombinant *Citrobacter freundii* (a close relative of *E. coli* possessing the native ability to assimilate cellobiose) expressing CenA and Cex is able to grow at the expense of crystalline cellulose (French *et al.*, 2015; S. S. Lakhundi *et al.*, *In preparation*). Here, the preparation of a Cex-CenA fusion protein retaining all the essential properties for effective cellulose degradation is described.

## 3.3 Methods

### *3.3.1 Construction of the fusion protein*

The gene encoding the fusion protein was constructed from previously prepared pSB1C3 plasmids containing BBa_K523016 ($P_{lac}$+*lacZ'α*+RBS+*cex*) and BBa_K523015 ($P_{lac}$ + *lacZ'α* + RBS + *cenA*). The *lac* promoter was BBa_J33207 and ribosome binding site was BBa_J15001 (Lakhundi, 2011). An NcoI restriction site was introduced into the CBM of the two constructs by PCR such that the amino acid sequences were not changed. The *cenA* construct was amplified using the forward primer cenAfNcoI and reverse primer pSBNX3insr2. Likewise the *cex* construct was amplified using the forward primer pSBNX3insf2 and reverse primer cexrNcoI. Primer sequences are provided in table 2-3. PCR amplifications were performed using KOD hot start DNA polymerase (Novagen) using the MJ Research PTC-200 DNA Engine Thermal Cycler. The amplified PCR products were each purified using the QIAquick PCR purification kit (QIAGEN GmbH, Germany). The purified DNA products were

then digested with NcoI (New England Biolabs, Inc.) for one hour and purified. The purified products were ligated overnight at 16°C in a 30 µL reaction volume containing 10 µL of each digested product, 3 µL T4 DNA ligase buffer (with ATP), 1 µL T4 ligase, 1 µL T4 polynucleotide kinase (PNK) and 5 µL nuclease free water. The ligation product was used as a template for PCR with primers pSBNX3insf2 and pSBNX3insr2 using KOD polymerase. PCR products were electrophoresed on 0.8% agarose gels in 1X TAE buffer. Electrophoresis was performed at 100 V, 50 mA for 40 minutes using the Bio-Rad mini-sub cell horizontal electrophoresis system. The gel was post-stained with SYBR-Safe nucleic acid stain (1 µL 10 000 x stock in 50 ml deionized water) for 20 minutes and visualized on the Life Technologies Safe Imager 2.0. Bands were excised and purified using the QIAquick gel extraction kit (QIAGEN GmbH, Germany). The purified products were digested with EcoRI and PstI at 37°C for one hour. The plasmid pSB1C3 was also digested with EcoRI/PstI. Plasmid digests and PCR product digests were ligated together as described above and cloned into *E. coli* JM109. White clones were tested for Cex and CenA activities as described below. Positive clones were grown overnight in 5ml Luria broth supplemented with 40 µg/ml chloramphenicol. Plasmid DNA was isolated using the QIAprep spin miniprep kit (QIAGEN GmbH, Germany) and tested for insert size. Constructs with the right sizes were confirmed by Sanger sequencing performed by Edinburgh Genomics, University of Edinburgh, UK. All procedures were performed according the respective manufacturers' protocols.

### *3.3.2 Enzyme activity assays*

Enzyme activities were performed as described in Chapter 2.5 above.

### *3.3.3 Substrate binding assays*

Binding of enzymes in crude lysates to cellulosic substrate was evaluated as described by Warren *et al.,* (1987) with slight modifications. Crude lysate (50 µl) was added to 1 ml PBS containing 20 mg or 200 mg cellulose (avicel) (20µm microcrystalline powder, Aldrich Chemistry, Cat. No. 310697). The mixture was mixed by gentle vortexing and incubated at 0°C for 1 hour with 1 minute of vortexing every 15 minutes. The avicel was pelleted by centrifugation at 17 000 x g at 4°C. The supernatants were collected into fresh tubes and assayed quantitatively for Cex and CenA enzyme activity as described for culture supernatants. The pellet was resuspended in 400 µl fresh PBS. To the mixture 1.1 ml of PBS containing MUC (final concentration 0.2mM) was added and incubated at 37°C for 1 hour. Enzyme activity as 4-MU fluorescence was measured. For comparison and also as a positive control, assays without avicel were included alongside and treated in the same way.

### *3.3.4 Protein expression and electrophoresis*

*E. coli* JM109/pSB1C3-BBa_K523015, *E. coli* JM109/pSB1C3-BBa_K523016, *E. coli* JM109/pSB1C3-BBa_K523025, *E. coli* JM109/pSB1C3-BBa_J15509 and *E. coli* JM109/pSB1C3-BBa_J33207 were grown overnight in 5ml LB cultures with 40 µg/ml chloramphenicol. Following incubation, the cultures were used to inoculate 50 ml fresh LB pre-warmed to 37°C in a 1:100 dilution. Cultures were grown to $OD_{600}$ of about 0.6 and induced by adding IPTG to a final concentration of 0.38 mM. Cells were harvested after 16 hours by centrifugation (11 300 x g, 20 min) and washed twice with phosphate buffered saline (PBS). The culture supernatant was stored at -80°C and assayed for enzyme activity. The cell pellet was washed, resuspended in 5ml PBS and sonicated for 5 pulses at 10 µm (amplitude) for 20 seconds on ice. The sonicated samples were

centrifuged as before and the supernatant was collected as cell-free lysates. Samples were either analysed immediately or stored at -80°C overnight.

The cell extracts were analysed by electrophoresis using Mini-PROTEAN® TGX™ precast gels under non-denaturing conditions. The gels were run in 1x Tris-Glycine buffer (0.25M Tris, 1.92M Glycine; pH 8.3) at 40 V for 3 hours. For endoglucanase activity, gels were hand-prepared according to the manufacturer's instructions for Mini-PROTEAN system with modifications. CMC (0.2% w/v) was incorporated into the gels. After electrophoresis, the gels were rinsed once in distilled water. They were then flooded with 0.1% w/v Congo red solution and incubated at room temperature on a rotary shaker (80 rpm, 45 min). This was followed by washing in 1M NaCl with shaking. Gels to be assayed for exoglucanase activity were immersed in 10 mM MUC and incubated at 37°C for 1 hour. Due to poor migration in native-PAGE gels (discussed further in results) lysates were also analysed by agarose gel electrophoresis on 5% w/v agarose gels using Tris-borate buffer (90 mM Tris base, 90 mM Boric acid, pH 8.44).

### *3.3.5 Preparation of periplasmic fractions*

To investigate the location of expressed proteins, cell fractions (cytoplasmic and periplasmic) were prepared (Rutter *et al.*, 2013). Expression of the recombinant proteins from *E. coli* JM109 cells was performed as described above in M9 media (Cold Spring Harb., 2010) supplemented with 0.02% yeast extract and 40 µg/ml chloramphenicol.

### *3.3.6 Cellulose degradation experiments*

*C. fimi* NCIMB11490 (ATCC8090) was transformed with pSB1C3-BBa_K523025 or pSB1C3-BBa_J15509 and selected on plates containing chloramphenicol. Initial experiments showed that 40 µg/ml chloramphenicol slowed growth, hence 15 µg/ml was used. Enzyme activity was confirmed for all constructs on agar plates. M9 minimal medium supplemented with 1 g/l yeast extract and a main carbon source was used for cellulose degradation experiments. Cellulose degradation experiments were conducted using M9 minimal medium (Cold Spring Harb., 2010) supplemented with filter/blotting paper (Whatman® GB003, 0.8 mm, 300 gsm or Ford 428 mill 0.2 mm, 148gsm) as the main carbon source. For each construct, 7 ml cultures were set up in 30 ml McCartney bottles with either two 1.5 cm squares of Ford 428 paper (approx. 66 mg) or 1 cm squares of Whatman's paper (approx.. 60 mg). Inocula were prepared from overnight cultures grown in LB. The cultures were centrifuged at 6000 g for 10 minutes at 4°C. The supernatant was discarded and pellets were re-suspended in M9 medium to make a 10x concentrate. Each tube was inoculated to a density of 0.1 $OD_{600}$ after which chloramphenicol (15 µg/ml) and IPTG (0.38 mM) were added. In a separate experiment using conical flasks, four 1.5 cm squares of Ford 428 paper were added to 20 ml culture. All other culture constituents were proportional to those prepared for the bottles. The bottles and flasks were grown at 37°C on a rotary shaker (200 rpm). Positive (with glucose) and negative (no carbon) controls were also set up. In addition, a vector control (*C. freundii* NCIMB11490/pSB1C3-BBa_J033207) was included in all experiments. Cultures containing filter paper were agitated vigorously (maximum speed) for 1 minute every 24 hours on the Genie 2 vortex mixer (Scientific Industries, Inc.). Colony counts were prepared by serial dilution in sterile phosphate-buffered saline (PBS) and subsequent plating on LA without antibiotic.

### *3.3.7 Thin layer chromatography*

One millilitre aliquots were collected from the cellulose utilization cultures and analysed by thin layer chromatography (TLC) (Fry, 1988). The aliquots were concentrated using a DNA Speed Vac (DNA 110, Savant Instruments, Inc. USA) using the low drying rate overnight. The concentrate was resuspended in 100 µl 5% v/v chlorobutanol. To assess the sugar, three microliters of the resuspended samples were loaded as spots onto 20 cm x 20 cm TLC Silica Gel 60 plates (Merck KGaA, Germany) alongside a master mix standard (MMS) containing the seven main cell wall monosaccharides (galacturonic acid (GalA), galactose glucose, mannose, arabinose xylose and rhamnose). The plates were allowed to dry at room temperature and run in a TLC tank with 5:2:2:1:1 EPPAW buffer (ethlyacetate, propanol, pyridine, acetic acid and water). The TLC was run for 3 hours and air dried overnight in a fume hood. The plates were developed by dipping into thymol solution (5 mg thymol, 95 ml ethanol, 5 ml $H_2SO_4$). Colour development was achieved by heating plates in an oven (105°C, 5 min).

## 3.4 Results

### *3.4.1 Generation of the fusion protein.*

By analysis of the CBM sequences of Cex and CenA, a consensus sequence PWNGSIP was found. It was determined that an NcoI site could be introduced into both without changing the amino acid sequence (Figure 3-1a). The two plasmids containing each of the fusion partners were amplified by the respective primers, digested with NcoI and ligated together. The ligation product was used as a template for PCR which yielded two products; a larger around 3.1 kb and a smaller around 2.1 kb. The larger was presumed to be correct and was purified from gel. The smaller was

presumed to be a mixture of Plac-lacZα-cenA and Plac-lacZα-cex from intact template which had made it through the purification. The 3.1 kb fragment was gel purified and digested with EcoRI/PstI and ligated into pSB1C3 (replacing mRFP, BBa_J04450) pre-digested with the same enzymes (Figure 3-1b). *E. coli* JM109 competent cells were transformed and white colonies selected from plates containing chloramphenicol. The resulting fusion protein had the catalytic domains of Cex at the N-terminus and CenA at the C-terminus. In between these was a new chimeric CBM generated as a result of the fusion (Figure 3-2).

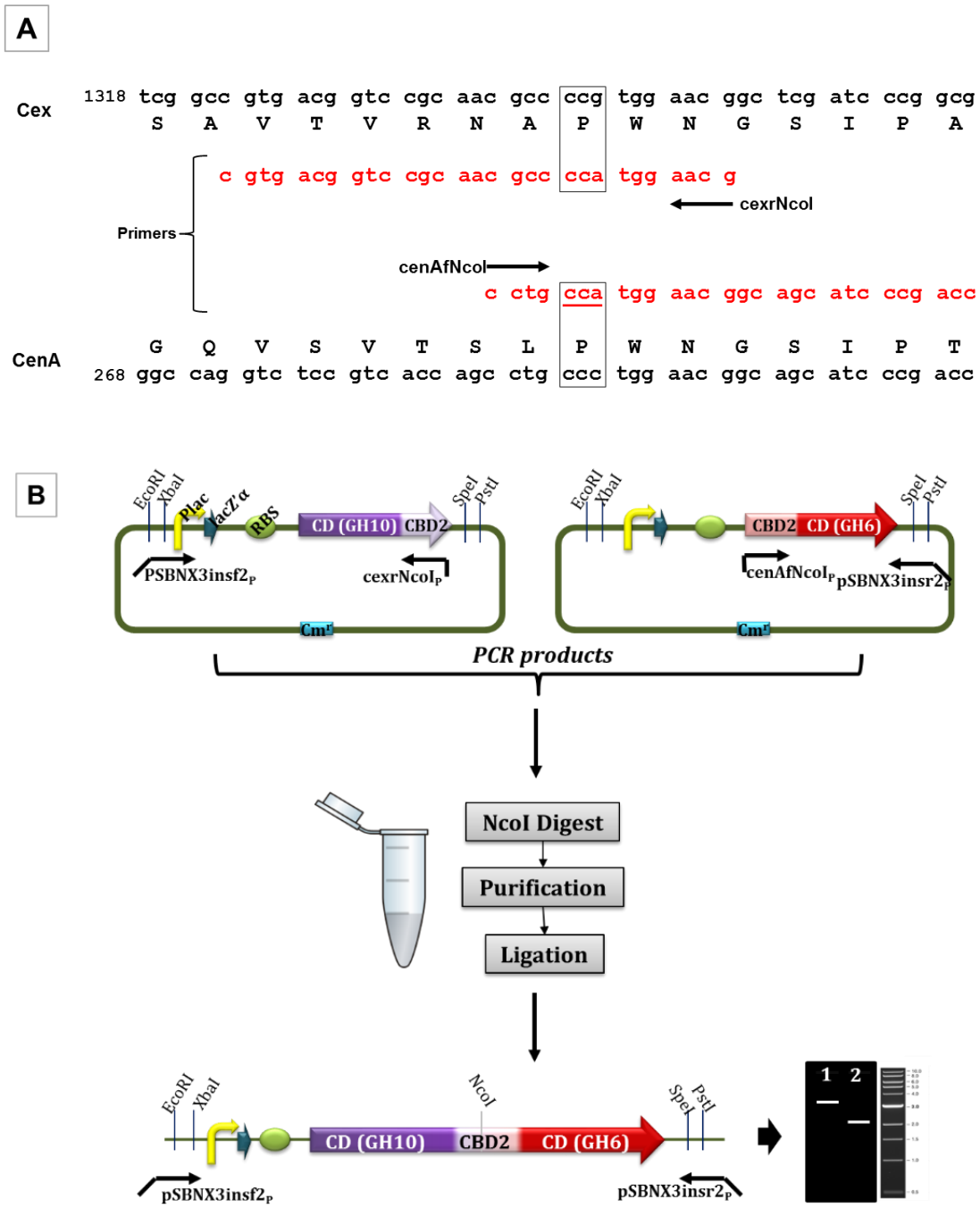



**Figure 3-1: Illustration of the design and construction of the exoglucanase-endoglucanase fusion**

(A) Part of the sequence of *cex* (from nt 1318) and *cenA* (from nt 268) showing the position where the point mutation was created as well as the position of the primers. (B) Steps that were involved in the generation of the fusion protein.

> CBM Cex

SGPAG CQVLW GVNQW NTGFT ANVTV KNTSS APVDG WTLTF SFPSG QQVTQ AWSST 432

VTQSG SAVTV RNA PWNGSIP AGGTA QFGFN GSHTG TNAAP TAFSL NGTPC TVG 485

> CBM CenA

PGCRV DYAVT NQWPG GFGAN VTITN LGDPV SSWKL DWTYT AGQRI QQLWN GTAST 87

NGGQV SVTSL PWNGSIP TGGTA SFGFN GSWAG SNPTP ASFSL NGTTC TGTVP TTS 142

> CBM CxnA

SGPAG CQVLW GVNQW NTGFT ANVTV KNTSS APVDG WTLTF SFPSG QQVTQ AWSST 432

VTQSG SAVTV RNA PWNGSIP TGGTA SFGFN GSWAG SNPTP ASFSL NGTTC TGTVP TTS 490

**Figure 3-2: Amino acid sequences of Cex, CenA and CxnA indicating a PWBGSIP consensus in all three sequences**

### *3.4.2 Enzyme activity assay standards*

The detection ranges for reliable quantification were determined from dilutions of 4-MU and RBB. The lowest detection limit was defined as the lowest concentration of analyte that produces a reading equivalent to half of the reading obtained from a sample with twice its concentration. That is, 'x' is said to be the lowest detection limit when

$$y_x = ½\ a$$

where 'y' is the reading obtained for 'x' concentration and 'a' is the reading obtained for 2x.

Similarly, the highest detection limit was defined as the concentration of analyte whose reading is equivalent to twice the reading of half of its concentration. That is, 'n' is said to be the highest detection limit when

$$w_n = 2b$$

where ‘w’ is the reading obtained for ‘n’ concentration and ‘b’ is the reading obtained for ½ n.

With these definitions, a linear curve was able to be drawn with 0,0 as the intercept. The regression equation from the curves (figure 3-3) was used to convert the fluorescence and optical density readings from the MUC and CMC assays respectively to enzyme units. All samples that gave readings above the highest detection limit were diluted and re-measured.

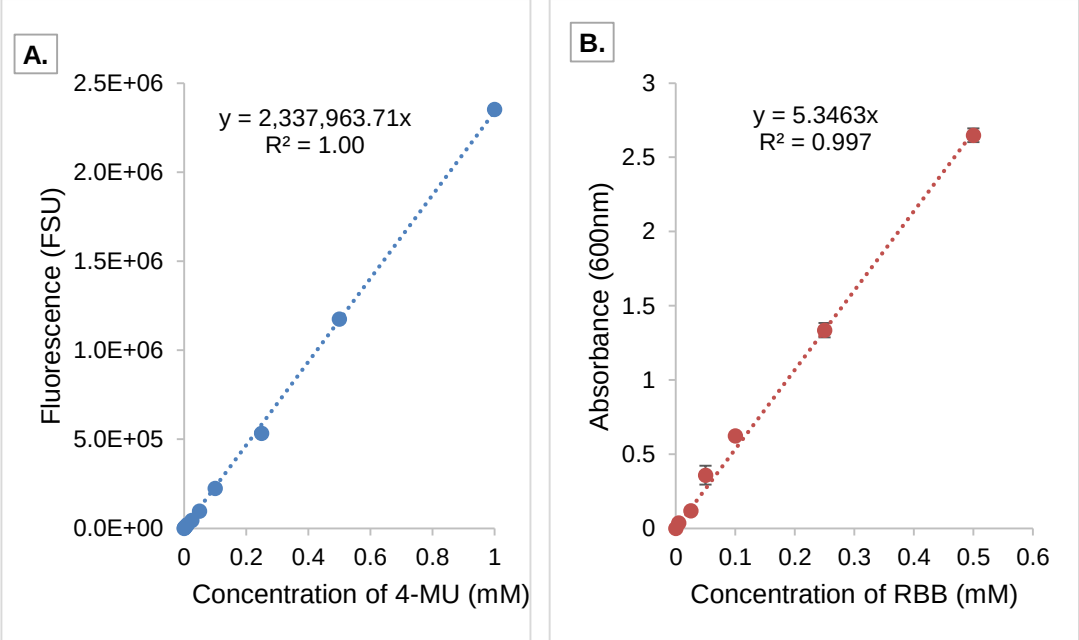



**Figure 3-3: The detection ranges of 4-MU and RBB standards**

Detection ranges of 4-MU and Remazol Brilliant Blue (RBB) were determined by measuring the fluorescence and absorbance respectively of serially diluted suspensions. (A) MUC detection range was 1 nM to 1 mM corresponding to fluorescence values of 2,456 FSU to 2,351,032 FSU. (B) RBB detection range was 5 nM to 0.5 mM corresponding to OD values of 0.035 to 2.648

### *3.4.3 Cellulase activity of the fusion protein*

*E. coli* JM109/pSB1C3-BBa_K523025, showed fluorescence on MUC plates as well as zones of clearing on CMC plates indicating Cex and CenA activities respectively (figure 3-4; A & B). Both activities were also detected in cell lysates on agar plates (figure 3-4; C & D) and also by electrophoresis of cell lysates on non-denaturing (native) PAGE gels. On native-PAGE gels (4% stacking, 12% resolving) however, the proteins did not migrate through the gels (precast or in-house prepared), but remained at the junction between the stacking and resolving gels. Similarly, no migration was observed in 4 - 15 % precast gradient gels with proteins remaining in the wells. Electrophoresis was therefore repeated using 5% agarose gels in Tris-borate buffer (pH 8.44) (Figure 3-4; E & F). Unexpectedly, the active proteins migrated towards the cathode, rather than the anode, although the predicted isoelectric points (pI) of the proteins (Cex, pI 6.18; CenA, 7.56; CxnA 6.19) were lower than the pH of the buffer. Activity spots due to CenA and Cex migrated further than those due to CxnA, indicating that both activities were associated with a protein of a different size from either CenA or Cex, as expected (figure 3-4; E& F).

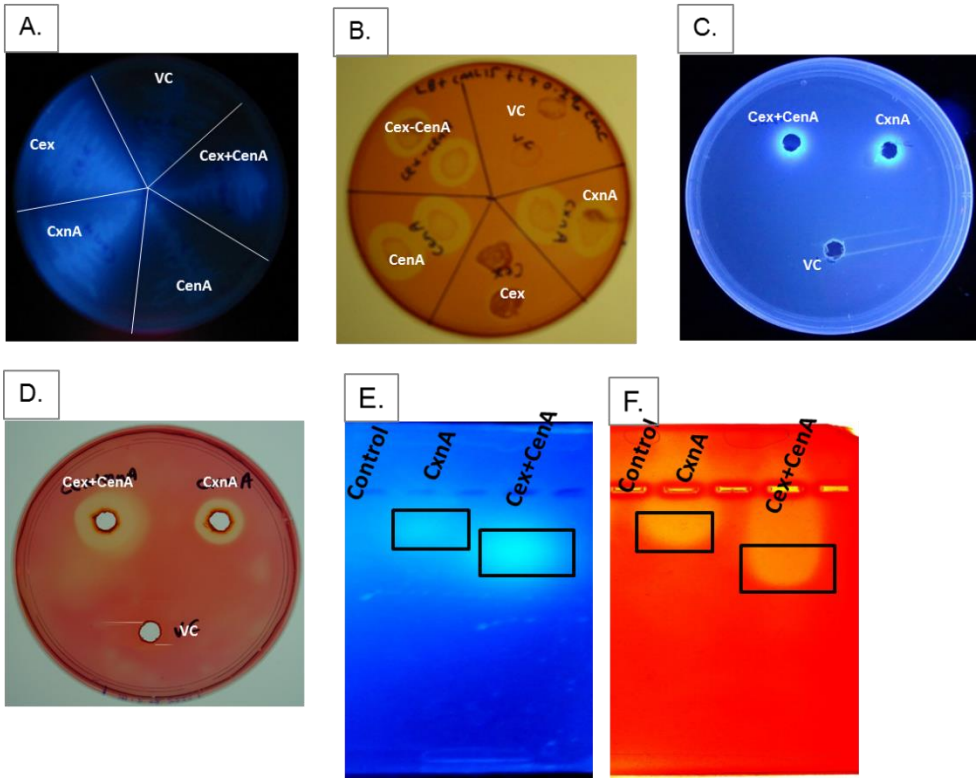



**Figure 3-4: Enzyme activity in cells expressing cex, cenA, cxnA and non-fused cex+cenA**

Exoglucanase activity (A, C & E) is detected by fluorescence of 4-MU following incubation of samples (whole cells or lysates) with MUC. Endoglucanase activity (B, D & F) is detected by the formation of a halo following incubation of samples (whole cells or lysates) with CMC followed by Congo Red stain and destaining with NaCl. A & B shows activity in whole cells; C & D shows activity in cell lysates on agar plates; E & F shows activity from cell lysates following electrophoresis on agarose gels. Enzymes were expressed by cells growing on plates and could also be detected from crude cell lysates run on a 5% horizontal agarose gel. Exoglucanase and endoglucanase activities determined by MUC and Congo Red assays respectively. Vector control (VC) is pSB1C3-BBa_J33207.

### *3.4.4 Investigation of the ability of the fused CBM to bind to cellulose*

To determine whether the fused CBM of CxnA could bind to cellulosic substrate, the crude protein extracts were incubated with and without microcrystalline cellulose (avicel) and the reduction in activity due to binding to cellulose was determined.

Binding of the enzymes to 20 mg/ml and 200 mg/ml of avicel in PBS was also tested. There was no significant difference between the enzyme bound to 20 mg/ml and 200 mg/ml of avicel. Unbound enzyme activity was 4.5% and 38% for non-fused and fused (CxnA) enzymes (figure 3-4). The remaining activity was recovered from the avicel pellet.

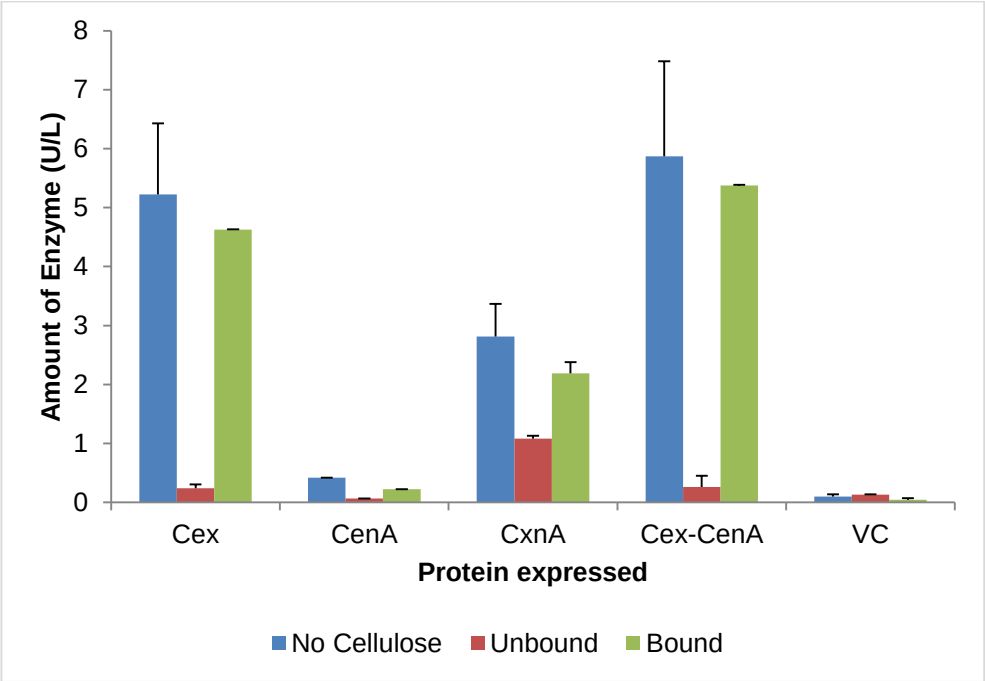



**Figure 3-5: The gene fusion retained its property of binding to cellulose**

The ability of the chimeric CBM to bind to cellulose was tested by incubating 50 µl of crude lysate with 20 mg/ml avicel at 0°C for 1 hour. The avicel was separated by centrifugation. MUC assays were performed aliquots of samples taken before addition of avicel (no avicel), supernatant after binding test and avicel is separated (unbound) and then in a resuspended pellet (bound). The Azo-CMC endoglucanase activity does not work in the presence of avicel. No significant difference ($p>0.33$) was found between the activity obtained before addition of avicel and the activity recovered from the resuspended avicel. There was however significant differences ($p<0.01$) between the activity recovered from the supernatant after binding and that obtained prior to addition of avicel. No statistically significant difference ($p>0.1$) was obtained for the vector control. Three biological replicated were performed.

### *3.4.5 Investigation of extracellular secretion/leakage of expressed proteins*

Secretion of cellulase is important for accessibility and hydrolysis of substrate. To determine whether CxnA was secreted into the culture supernatant, the enzyme activity levels in culture medium (extracellular), periplasmic and cytoplasmic protein extracts were compared. About 20% of the exoglucanase activity was found in the extracellular fraction of both CxnA and non-fused enzymes (figure 3-5). The majority of enzyme activity was however located in the periplasmic fraction.

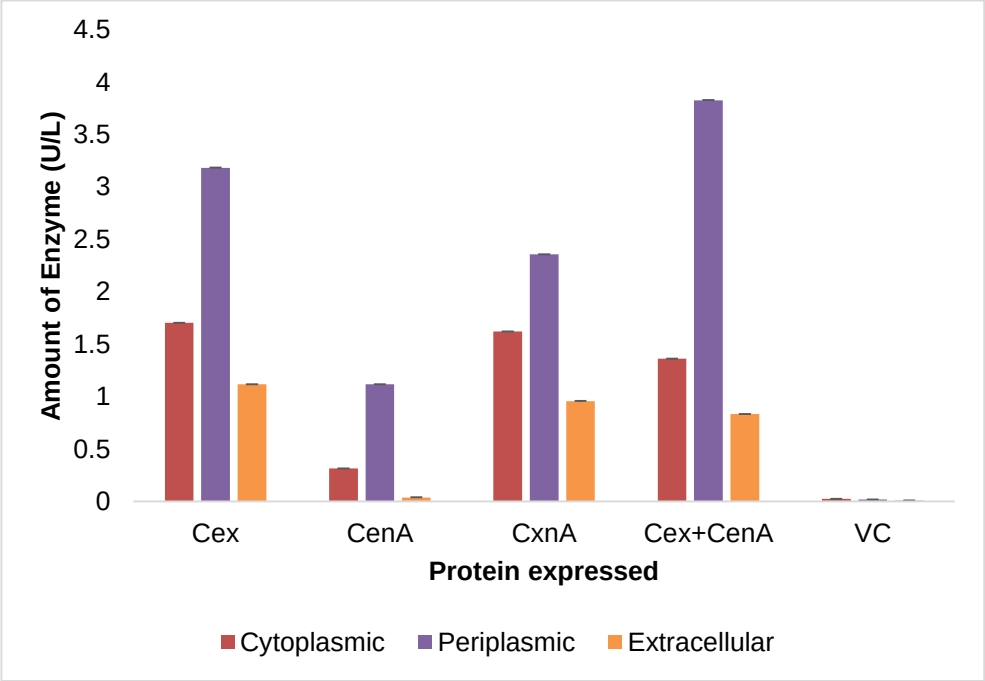



**Figure 3-6: Enzymes were secreted/leaked into the culture supernatant**

Secretion/leakage of proteins from the periplasm into the supernatant was investigated by comparing enzyme activities obtained from culture supernatant, cytoplasm and periplasmic fractions. Cytoplasmic and periplasmic fractions were resuspended in equal volumes and diluted to 1× equivalent of extracellular fraction. The activities from all fractions were statistically different ($p<0.01$) for all of the constructs except the vector control where no statistical difference ($p>0.1$) was observed. Three biological replicates were used.

### *3.4.6 Growth of* Citrobacter freundii *utilizing cellulosic substrates*

Growth of *C. freundii* NCIMB11490/pSB1C3-BBa_K523025 (fused) was higher than *C. freundii* NCIMB11490/pSB1C3-BBa_J15509 (non-fused) (figure 3-6; a, b & c). This observation was consistent irrespective of the filter paper used. Physical destruction of filter paper was faster with cells expressing the fused enzyme as opposed to cells expressing the non-fused. In shake flasks (100 ml conical flasks, 62 mm base diameter, foam stopper), significant destruction of filter paper (Ford 428) was seen at day 7 (Figure 3-6d) for the cells expressing the fused construct. In McCartney bottles (30 ml capacity, 28 mm base diameter, plastic screw cap) however physical destruction was observed earlier (day 3; figure 3-6e). Destruction of filter paper was faster for cells expressing the fused enzyme than cells expressing the non-fused enzymes (figure 3-6e). By day 5 however (McCartney bottles), filter paper was broken down in cells expressing both fused and non-fused enzymes whereas those of the vector control were not destroyed.

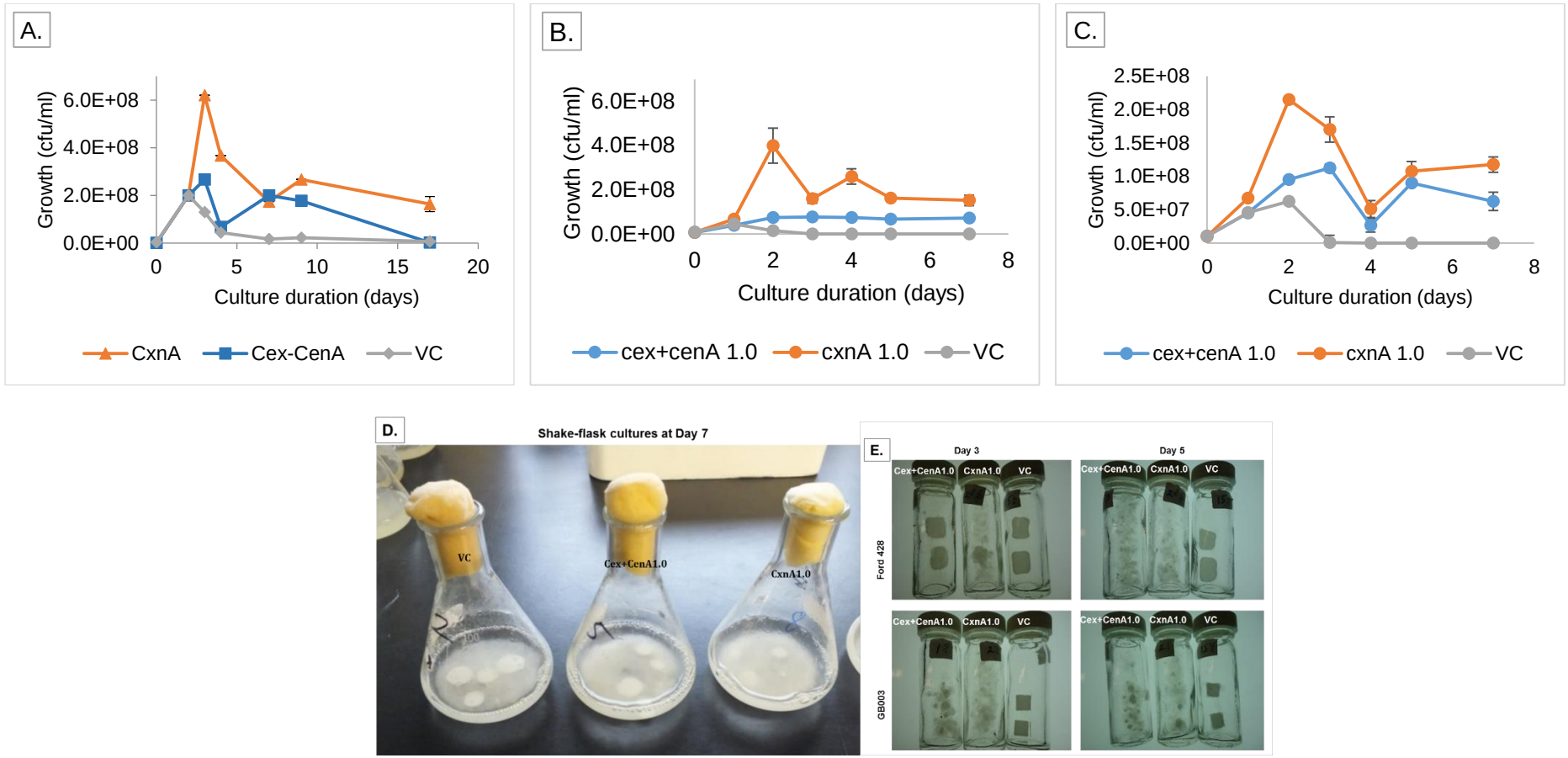



**Figure 3-7: *C. freundii* expressing the gene fusion grew better and physical destroyed filter paper faster than the non-fused genes**

*C. freundii* with pSB1C3-BBa_J15509 (Cex+CenA1.0) or pSB1C3-BBa_K2523025 (CxnA1.0) or pSB1C3-BBa_J33207 (VC) cultured with the Ford 428 (A: conical flasks & B: McCartney bottles) and the GB003 (C) blotting papers. The cultures were grown with shaking at 200 rpm and monitored daily. In addition, cultures were vortexed for one minute (highest speed on Genie 2 Vortex Mixer, (SLS, UK)) daily. Growth was monitored by performing viable counts on serial dilutions. Physical destruction of filter paper observed in both conical flasks (D) and McCartney bottles (E). Three biological replicates were used.

### *3.4.7 Determination of the release of sugars hydrolysed cellulose filter paper*

To determine whether the filter paper was being hydrolysed by cells expressing the recombinant cellulases, aliquots of the cultures were taken at various time points and analysed using thin layer chromatography. Spots that were suspected to be cellobiose and cellotriose were seen in samples from filter paper cultures (fused and non-fused) on days 2, 4 and 7 (figure 3-7 b, c & d). The spots were absent in the day 0 samples (figure 3-7a). On the day 17 plates however, the spots for cellobiose and cellotriose was only seen in cultures expressing the fusion protein but not the non-fused (figure 3-7e).

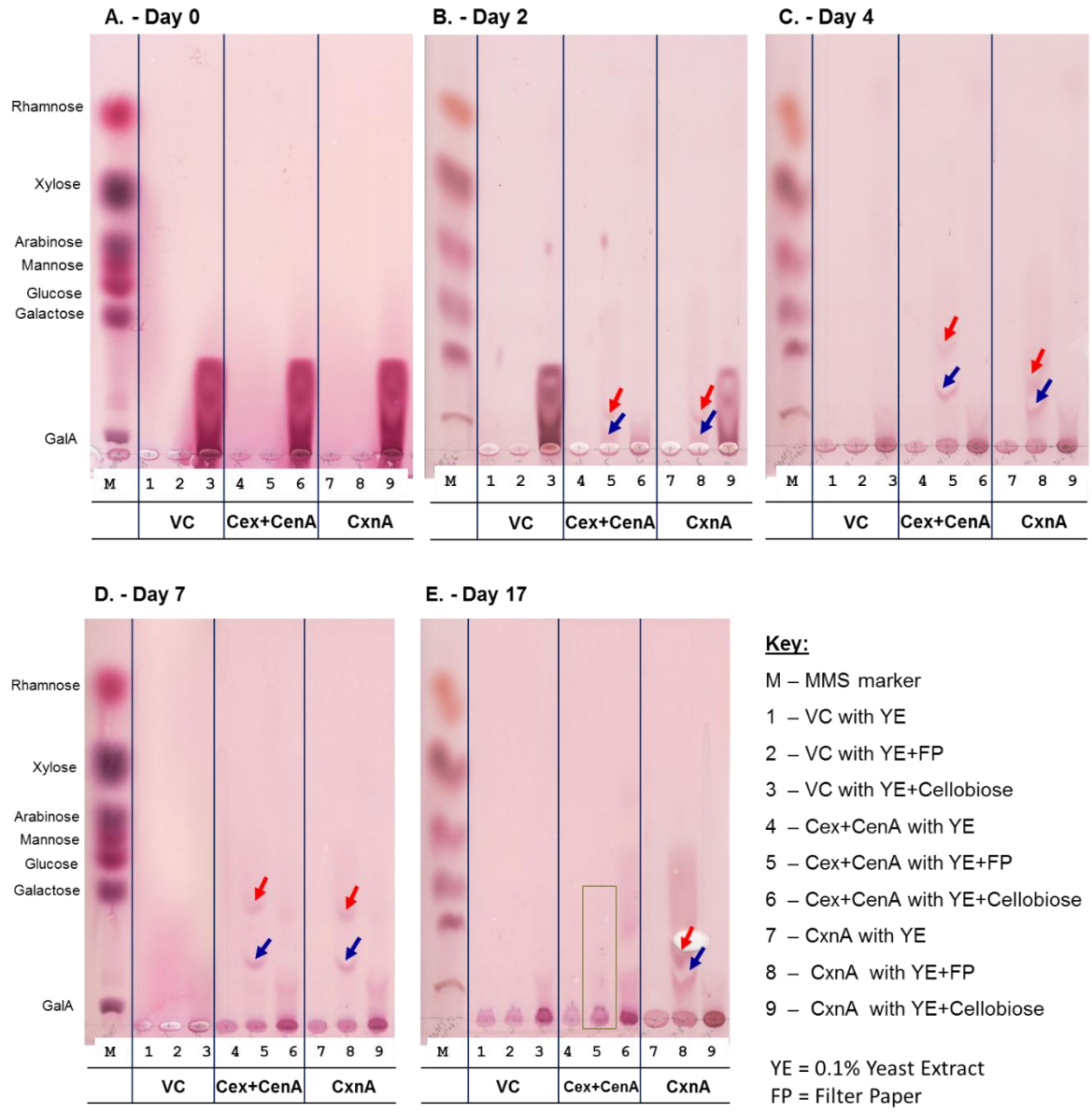



**Figure 3-8: Hydrolysed sugars suspected to be cellobiose and cellotriose were seen on thin layer chromatographic analysis**

Aliquots (1 ml) of the growing cultures were taken at various time intervals and concentrated to about 100 µl in a speedvac. The concentrate was resuspended in 100 µl chlorobutanol after which 5 µl was analysed on silica gel TLC plates.

To confirm that the spots seen were cellobiose and cellotriose, the samples from filter paper cultures were re-run (figure 3-8) alongside a marker made up of cellotetraose, cellotriose, cellobiose and glucose. The spots were present in all the samples as seen

earlier except the suspected cellotriose spot from the day 17 sample. The missing spot could be due to hydrolysis that occurred during storage (three months). The spots were confirmed as cellobiose and cellotriose.

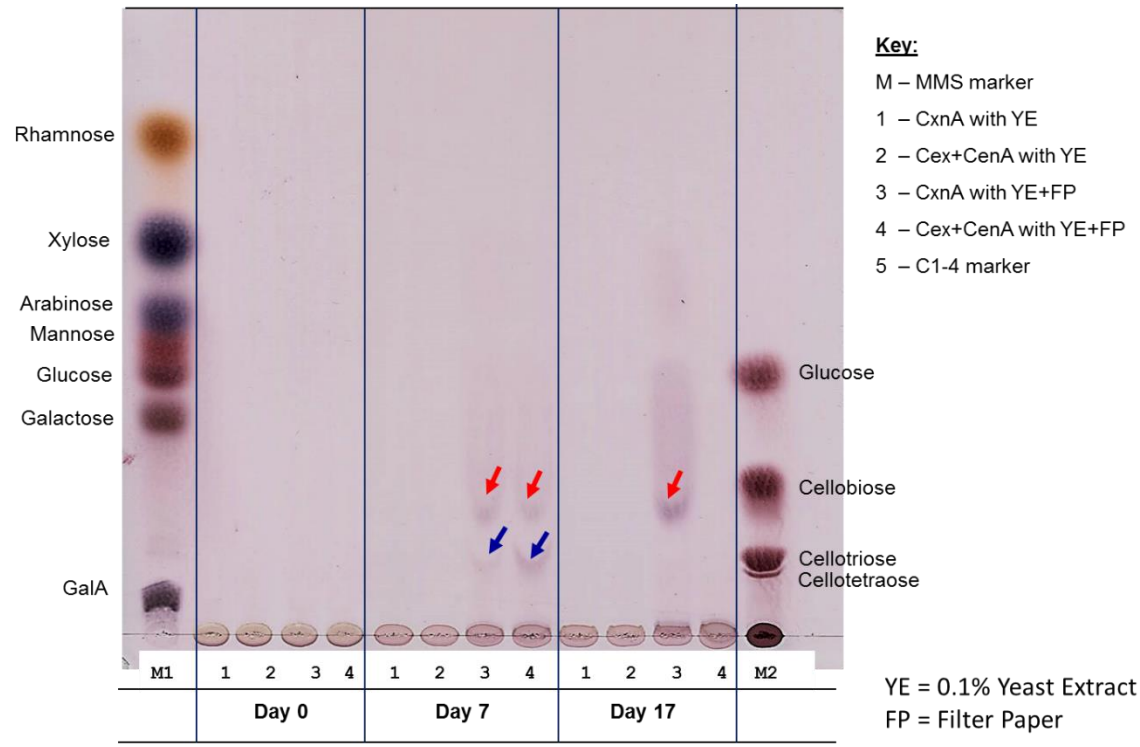



**Figure 3-9: Spots of hydrolysed sugars correspond to cellobiose and cellotriose**

Aliquots (1 ml) of the growing cultures were taken at various time intervals and concentrated to about 100 µl in a speedvac. The concentrate was resuspended in 100 µl chlorobutanol after which 5 µl was analysed on silica gel TLC plates.

## 3.5 Discussion

In order to test synergistic effects of co-localization of different enzyme domains for microbial cellulose hydrolysis, there was a need to design and test simple enzyme complexes with multiple enzyme activities. Several cellulases have been isolated and characterized. However, finding a combination of cellulases that will enable strains to utilize crystalline cellulose as a carbon source is challenging. The expression of *C.*

*fimi* exoglucanase/xylanase Cex and endoglucanase CenA in *C. freundii* which allows the recombinants to grow on cellulosic substrates has been described earlier (Lakhundi, 2011; French *et al.*, 2015; S. S. Lakhundi *et al.*, *In preparation*). Fusion enzymes catalysing the same multistep sequential reaction have been shown to demonstrate superiority over the individual native enzymes (Bulow and Mosbach, 1991). Cellulases with co-localized domains are exemplified by the multienzyme machines known as cellulosomes. Co-localization improves synergy between domains resulting in effective degradation of cellulose (Bayer *et al.*, 1998a; Bayer *et al.*, 1998b; Bayer *et al.*, 2004).

The fusion protein CxnA was found to bind cellulose. Binding however was not as great as that of the native CBM. The reason for this is unclear but could be due some small changes in the conformation of the chimeric CBM. Further engineering of the CBM could potentially improve binding as well as help to understand requirements and process of binding. In addition to binding, it has been suggested that CBMs enhance the cellulose degradation by loosening individual chains from the cellulose surface (Teeri *et al.*, 1992b). The ability of the chimeric CBM of CxnA to do this has not been determined.

Enzyme activity was detected in all fractions (cytoplasmic, periplasmic and culture media). The mechanism by which *E. coli* and *C. freundii* secrete these proteins from the periplasm is not clear. Secretion however indicates that the signal peptides of the cellulases are recognized by the *E. coli* and *C. freundii* Sec pathway. The ability of cellulolytic organisms to effectively secrete cellulases into the medium is necessary to enhance degradation of cellulosic substrates.

Synergy in other cellulase fusions has been reported (Bulow, 1987; An *et al.*, 2005; Thongekkaew *et al.*, 2013). Fusion of Cex and CenA was reported earlier (Warren *et*

*al.*, 1987). Though this fusion contained both exoglucanase and endoglucanase activities, it was unable to bind to cellulose because it lacked an intact CBM. Furthermore, it was not tested whether the fusion could hydrolyse cellulosic substrates. To check whether the new fusion protein could hydrolyse cellulose better than non-fused ones, *Citrobacter freundii* was transformed with the fusion as well as with the individual cellulases. The ability of the transformants to utilize filter paper for growth was tested. *C. freundii* was chosen because it is a close relative of *E. coli* with native ability to assimilate cellobiose. Furthermore, it is not considered pathogenic (Advisory Committee on Dangerous Pathogens, UK) although it is capable of causing nosocomial infections in the immunocompromised. By using *C. freundii,* there was no need to include a β-glucosidase to the cultures or clone it into the constructs. From previous experiments, it was determined that *C. freundii* required a small amount (0.1% w/v) of yeast extract to help it grow and reach sufficient numbers to produce the initial enzymes to break down cellulose. It was observed that all the cultures (including negative controls) grew similarly within the first 48 hours. However, numbers of live cells within the negative control cultures dropped from this point forward whereas the number of live cells within the cultures expressing the cellulases continued to increase. Cultures expressing the fused enzyme increased about 50% more than the cultures expressing the non-fused enzymes. This was confirmed when the total growth was expressed as the total area covered by the growth curves.

In order to assess the hydrolysis of the filter paper and production of cellobiose, aliquots of the cultures were analysed over time. It was observed that the cellulases enabled cells expressing them to produce cellobiose and cellotriose in culture from day 2. Production of these products was observed for the cultures until day 17 when the cultures expressing the non-fused proteins no longer showed free cellobiose. Furthermore, the colony counts of the non-fused cultures had also reduced by about

90%. Physical degradation of filter paper by the fusion constructs started earlier than the non-fused constructs. These results are evidence that the fused protein improves degradation of the filter paper releasing more cellobiose such that there is still a detectable quantity to support growth at a time when the non-fused constructs no longer show detectable cellobiose. The fused constructs may perhaps be more stable than the unfused enzymes making it stay active for a longer period. The stability of the enzymes was not determined.

## 3.6 Conclusions

This work has demonstrated that fusing homologous domains of two different enzymes can give a new gene fusion which when expressed is functional and may be more effective than the native proteins. Co-localization of the catalytic domains of cellulases may enhance the overall catalytic activity of both enzymes. Such fusions could potentially replace cloning of multiple enzymes for improved biomass conversion and downstream processing. The apparent leakage of the *C. fimi* cellulases from *E. coli* is an interesting finding. Further investigations to understand the mechanism of secretion will provide useful insights into how to enhance extracellular secretion of cellulases and other proteins.

# Chapter 4

## Design and application of novel multicatalytic cellulases for improved biocatalysis of the degradation of cellulosic biomass

### 4.1 Summary

In naturally occurring cellulose degrading microbes, bioconversion is catalyzed by a battery of enzymes with different catalytic properties. The construction and characterization of an endoglucanase-exoglucanase fusion protein (CxnA1.0) which improves microbial degradation of cellulosic biomass was earlier described. In order to investigate the capabilities of recombinant multimeric fusion proteins in improving cellulosic biomass conversion, we constructed six additional fusions of cellulases from *Cellulomonas fimi* (endoglucanases CenA, CenB and CenD, exoglucanase Cex, and β-glucosidase) and *Cytophaga hutchinsonii* (cellodextrinase CHU2268). The fusions were made using rigid proline-threonine linkers. Enzyme activity was tested by expressing fused and non-fused devices in *Escherichia coli*. All fusions retained catalytic activity of both parental enzymes. To investigate the benefits of fusion, *Citrobacter freundii* NCIMB11490 was transformed with either fused or non-fused enzymes and cultured with cellulose blotting papers (Whatman® GB003, 0.8 mm, 300 gsm or Ford 428 mill 0.2 mm, 148gsm) as main carbon source. Cells expressing the fusions of Cex with CenA or CenD reproducibly showed higher growth than cells expressing non-fused versions, as well as more rapid physical destruction of paper. The opposite was observed for the other combinations. A newly constructed Cex/CenA fusion (CxnA2) which contains two CBMs degraded the GB003 paper faster than the previously reported Cex/CenA fusion (CxnA1) which contains only one

CBM and led to improved cell growth. Our results demonstrate that use of fusion proteins can improve biomass conversion in vivo, and could potentially substitute cloning of multiple enzymes and improve product yield. Our procedure for generating fusion proteins from existing libraries of constructs is simple and may be useful in other applications.

## 4.2 Background

Cellulose hydrolysis requires three classes of cellulases: endo-β-1,4-glucanases (EC 3.2.1.4), exo-β-1,4-glucanases (or cellobiohydrolases) and β-glucosidases (EC 3.2.1.21) (Lynd *et al.*, 2002; Lynd *et al.*, 2005; French, 2009; Horn *et al.*, 2012). Cellobiohydrolases hydrolyse the β-1-4 glycosidic linkages in cellulose and from the reducing end (EC 3.2.1.176) or non-reducing end (EC 3.2.1.91). Endoglucanases hydrolyse internal β-1-4 glycosidic linkages forming new reducing and non-reducing ends, and together with the cellobiohydrolases release cellobiose from cellulose chains. The cellobiose is then hydrolysed to glucose by the β-glucosidases. In recombinant cellulose degradation systems, co-expression of cellulases from all three classes is essential for the conversion of cellulosic biomass into sugars (French, 2009). Our understanding of cellulases and natural microbial biomass degradation systems has been improved by the many studies that have been performed (Sternberg, 1976; Leschine, 1995; Lynd *et al.*, 2002; Lynd *et al.*, 2005; Wilson, 2011; Gupta *et al.*, 2012; Du *et al.*, 2015; Tan *et al.*, 2015). Despite this wealth of knowledge, efforts to transfer natural cellulose degradation systems to heterologous hosts have had limited success (Lynd *et al.*, 2002; Lynd *et al.*, 2005). Synthetic biology offers the opportunity of evaluating many combinations of cellulases in order to identify the most efficient for the development of recombinant systems. Additionally, it also offers options and the flexibility for creating a single recombinant microorganism for

simultaneous degradation of biomass and biofuel production (Lynd *et al.*, 2005; French, 2009; French *et al.*, 2013; French *et al.*, 2015). Synthetic biology uses two genetic engineering perspectives on how to create the ideal microbial biofuel production system. The two are: (1) Engineer a natural cellulose degrading organism to generate useful products and (2) Engineer an organism which can produce useful products to degrade cellulose. With synthetic biology, a host that performs neither of the two (degrading cellulose or producing fuels; eg. *Escherichia coli* and *Citrobacter freundii*) is engineered to perform them both.

One of the reasons that have been suggested as a cause for the general failure to achieve effective degradation of crystalline cellulose in recombinant strains is the absence of synergy between the different classes of enzymes or between enzymes with overlapping catalytic activity (Lynd *et al.*, 2002; French *et al.*, 2015). Although studies on synergy are complicated due to the different interactions that may be present, an empirical synthetic biology approach (combinatorial generic engineering) has been proposed (French, 2009). The combinatorial approach involves parallel screening of a large library of different combinations of cellulases from different classes. This screening will help determine empirically which combinations of enzymes are most effective for degradation of the various types of cellulases. Efforts in this direction have been made by the C. French Group in the University of Edinburgh (Lakhundi, 2011; Barnard, 2012; Liu, 2012). These studies employed the use of BioBrick cloning to construct various genetic devices encoding cellulases from *Cellulomonas fimi* and *Cytophaga hutchinsonii*. Expression of cellulases from these devices was tested in *E. coli, C. freundii* and *Bacillus subtilis*. The results from these were then used to construct modular devices encoding multiple cellulases and subsequently tested for activity against amorphous cellulose, crystalline cellulose, mannan and xylan. From these experiments, the activities of *E. coli* and *C. freundii*

expressing *C. fimi* cellulases CenA, CenB, CenC, CenD (endoglucanases), Cex, CbhA, CbhB (cellobiohydrolases) and Cfbglu (β-glucosidase) as well as *C. hutchinsonii* cellodextrinase CHU_2268 were elucidated. *E. coli* expressing CenA best utilized carboxymethyl cellulose (CMC) and filter paper. Synergistic effects were seen from co-cultures of cells expressing CenB and CenD. Co-expressing CbhA and CbhB with CenA improved utilization of CMC in *E. coli* but not *C. freundii.* Also, *C. freundii* expressing Cex and CenA was able to grow on microcrystalline cellulose as the main carbon source although it best grows with filter paper (figure 4-1). These experiments also revealed that *C. freundii* NCIMB11490 expressing the heterologous cellulases was unable to utilize CMC. Cells expressing CenA-Cex exhibited higher CenA activity compared to CenA-Cex-CbhA-CbhB leading to better deconstruction of filter paper (Barnard, 2012). Of the *C. hutchinsonii* cellulases studied, CHU_2268 was found to be the most promising cellulase for recombinant expression (Liu, 2012).

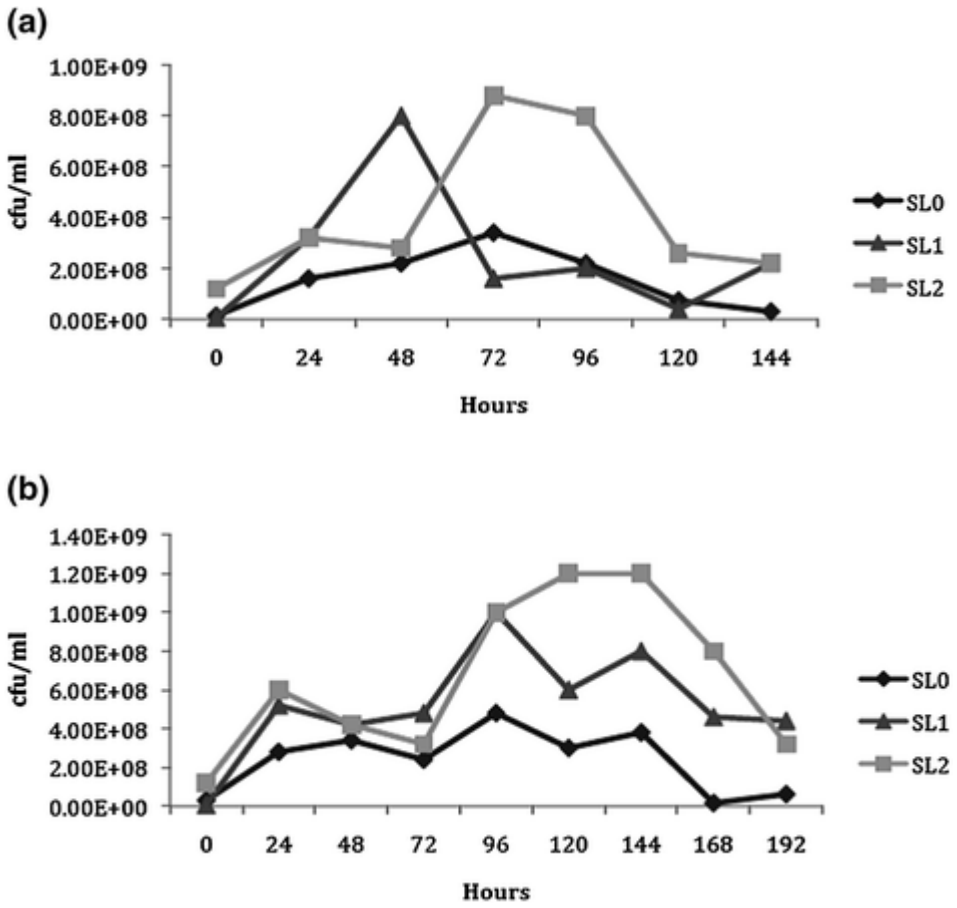



**Figure 4-1: Growth of *Citrobacter freundii* SBS197 expressing CenA and Cex at the expense of cellulosic substrates**

Aliquotes of cultures containing filter paper and *C. freundii* expressing cellulase genes were taken at various time intervals and the colony counts determined. SL0 *C. freundii* SBS 197 vector control; SL1 *C. freundii* SBS197 expressing cenA and cex from *lac* promoter; SL2 *C. freundii* SBS197 expressing *cenA* and *cex* from *spac* promoter. (a) Growth with 20 g/l filter paper +1 g/l yeast extract; (b) growth with 20 g/l Avicel +1 g/l yeast extract (Lakhundi, 2011; French *et al.*, 2015)

Apart from co-expression, fusion proteins have been demonstrated to improve biocatalysis (Bulow, 1987; Bulow and Mosbach, 1991; An *et al.*, 2005; Elleuche, 2015). In chapter 3 of this thesis it was demonstrated that a fusion protein made from CenA and Cex improved the hydrolysis and utilization of filter paper as a main carbon source by recombinant *C. freundii*. To further investigate the effects of such fusion proteins, six additional fusion proteins were made from existing modular constructs

(Barnard, 2012) using a simple straightforward protocol, mutagenesis with blunt-end ligation (MABEL). *C. freundii* ATCC8090 (NCIMB11490) transformed with the fused or unfused proteins was grown at the expense of filter paper to assess the catalytic ability of the fused proteins.

## 4.3 Methods

### *4.3.1 Bacteria strains, media and cellulose degradation experiments*

*E. coli* JM109 was used as a host strain for cloning whereas *C. freundii* ATCC8090 was used a host strain for cellulose degradation experiments. Media and cellulose degradation experiments were carried out as described earlier in Chapter 3 of this thesis. Cellulose degradation experiments were conducted repeatedly in 30 ml McCartney bottles, 50 ml falcon tubes and 100 ml conical flasks.

### *4.3.2 Construction of the fusion proteins*

Fusion proteins were generated using the MABEL protocol described earlier (Chapter 2) in this thesis. A selection of modular devices constructed earlier in the lab were selected from a library for modification into fusion proteins. Primers with a fusion linker tag at the 5'-end (forward) or 3'-end (reverse) were designed and used for PCR. The PCR was performed using Q5 DNA polymerase (NEB) with the GC enhancer buffer according to the manufacturer's instructions. Forward primers were designed to match the first 18 to 21 nucleotides of the C-terminal fusion partner without the ATG start codon and signal peptide if present. Similarly, reverse primers were designed to flank the last 18 to 21 nucleotides of the N-terminus fusion partner without the stop codon(s). Together the forward and reverse primers inserted a 16 amino acid linker rich in proline and threonine residues. This type of linker was used because similar

linkers occur naturally in *C. fimi* cellulases linking the catalytic domains (CDs) to the carbohydrate binding modules (CBMs). Figure 4-2 is an illustration of the primer structure and the fusion construction process. A list of primers and constructs is given in table 4-1.

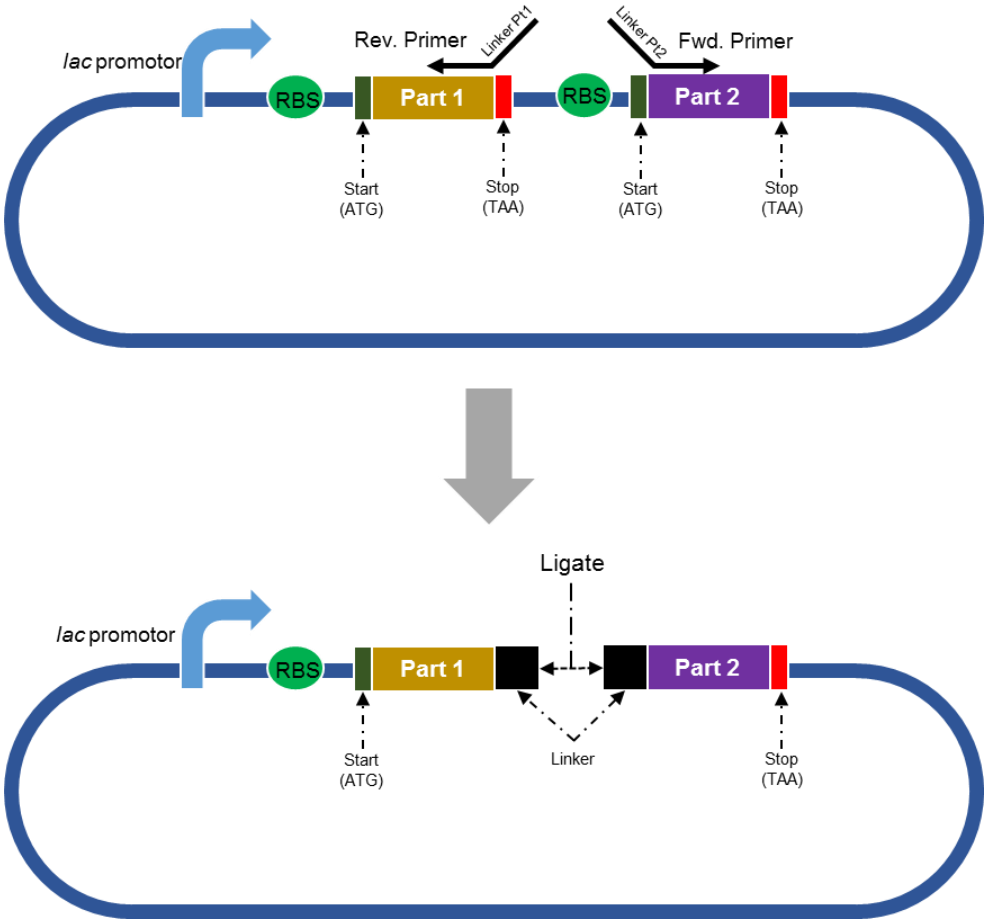



**Figure 4-2: Construction design of the fusion proteins using MABEL**

Primers tagged with the fusion linkers were used for PCR to amplify the entire plasmid (pSB1C3) containing the cellulases. The resulting linear fragment which also contains the added linker was self-ligated using T4 DNA ligase and T4 polynucleotide kinase (T4 PNK). *E. coli* was transformed with the ligated construct and selected on plates with appropriate antibiotics. Fusion was confirmed by DNA sequencing.

**Table 4-1: Constructs and primer pairs used to make fusions**

| Construct | Primer | Reference |
|---|---|---|
| *PlacZ-cenA-cex* | PTcexΔSP_f3 / PTcenA_r1 | (Barnard, 2012) |
| *PlacZ-cenB-cex* | PTcexΔSP_f3 / PTcenB_r1 | (Barnard, 2012) |
| *PlacZ-cenC-cex* | PTcexΔSP_f3 / PTcenC_r1 | (Barnard, 2012) |
| *PlacZ-cenD-cex* | PTcexΔSP_f3 / PTcenD_r1 | (Barnard, 2012) |
| *PlacZ-cfbglu-cenA* | PT.CBD(cenA)_f2 / PTcfbglu_r1 | (Barnard, 2012) |
| *chu2268-cenA* | PT.CBD(cenA)_f2 / PTchu2268_r1 | AS |
| *chu2268-cenB* | PT.CBD(cenB)_f2 / PTchu2268_r1 | AS |
| *cxnA-cfbglu* | PTcfbglu_f1 / PTcenA_r1 | This study |

Modular devices containing *chu2268* and *cenA* or *cenB* were made by Alejandro Salinas-Vaccaro (AS). Primer sequences listed in Table 2-3.

### *4.3.3 Enzyme activity assays*

Protein expression and extraction was performed as described in Chapter 2 of this thesis. Endoglucanase, exoglucanase and β-glucosidase activities were tested using carboxymethyl cellulose (CMC), 4-Methylumbelliferyl β-D-cellobioside (MUC) and 4-Methylumbelliferyl β-D-glucopyranoside (MUG) respectively as substrates. The CMC was incorporated into the agar prior to pouring to a final concentration of 0.2%. For MUC and MUG, 100 µl stock (5 mg/ml) solution was spread onto the plates and allowed to dry in a laminar flow hood just before inoculation. Enzyme activity in cells expressing the enzymes was performed as described in chapter 3. In addition to these, activity testing of crude cell lysates was also performed on the M9 plates. Using the reverse side of a sterile 200 µl pipette, wells were made in the agar for each sample to be tested. Fifty microliters of crude cell lysate was pipetted into the

respective wells and incubated at 37°C for 1 hour. MUC and MUG plates were examined right away for fluorescence using long wave UV light. The CMC plates were flooded with 5 ml Congo red solution and de-stained with 1M NaCl as described earlier (Chapter 2).

### *4.3.4 Protein gel electrophoresis*

The crude cell lysates were analysed on polyacrylamide gels for their sizes. Protein concentration of the cell lysates was determined using the Pierce Coomassie plus (Bradford) assay kit (ThermoScientific, UK). About 40 µg of cell lysates were mixed with sample buffer and loaded onto the wells. Mini-PROTEAN® TGX® precast gels (4-15% gradient; Bio-Rad laboratories) were used to run the electrophoresis on the Mini-PROTEAN® Tetra cell according to the manufacturer's instructions. Electrophoresed gels were stained with Coomassie brilliant blue R-250 stain (1 g/L) for 4 hours and de-stained. The stain was prepared using the de-stain solution (45% (v/v) ethanol, 45% (v/v) water and 10% (v/v) glacial acetic acid) and filtered prior to use.

## 4.4 Results

### *4.4.1 Construction of the fusion enzymes*

Constructs were first checked by digesting the resulting plasmids with EcoRI. Plasmids that appeared to be the right size (figure 4-3) were then submitted for sequencing at Edinburgh Genomics. All the CenC+Cex fusions (CxnC) were found to contain CenB instead of CenC. The PCR was repeated and re-sequenced but the result was the same. Sequencing repeatedly failed (no reads) for the

CHU_2268+CenB fusions (i.e ChuB). The failed constructs (*cxnC* and *chuB*) were therefore excluded from further testing.

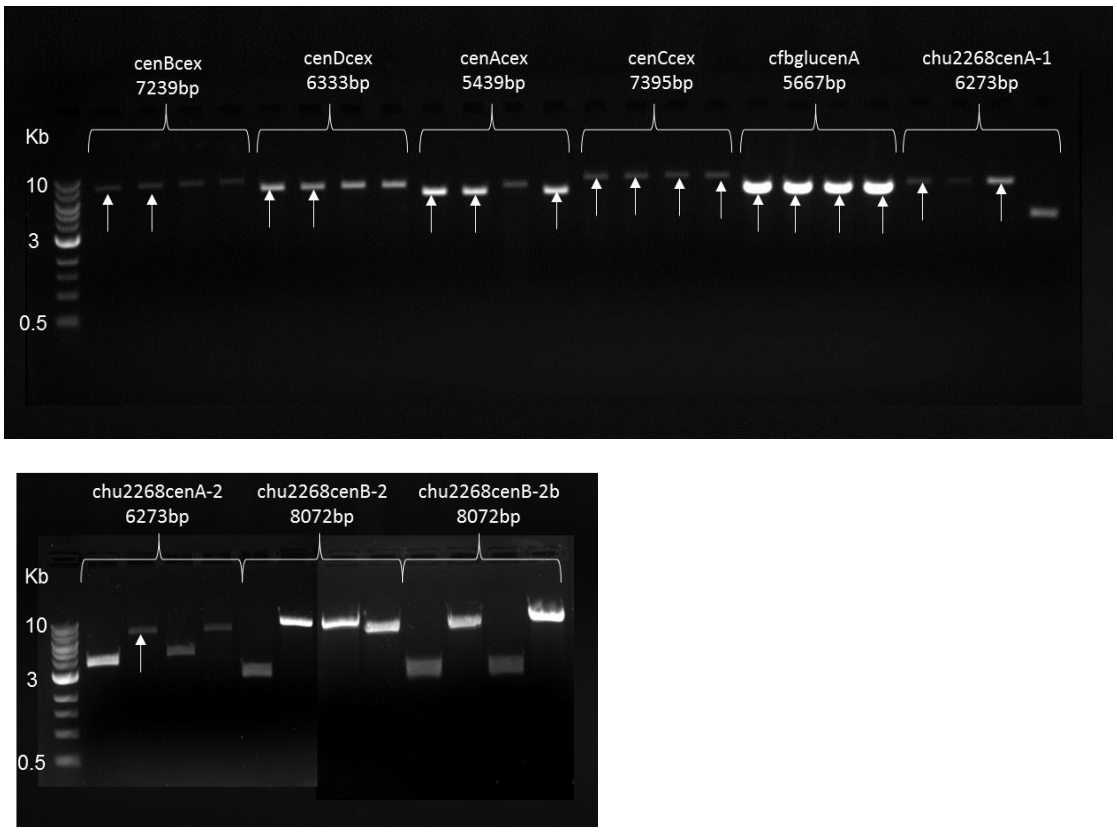



**Figure 4-3: EcoRI restriction digests of plasmid minipreps arising from MABEL**

Digests were performed for 1 hour at 37°C with 2 µl of plasmid DNA in a 30 µl reaction volume. Lane one shows a 1 kb ladder (NEB) run alongside plasmids from four selected colonies on the transformation plates. Expected band sizes indicated below names of constructs and actual bands (also confirmed by sequencing) shown by arrows.

### *4.4.2 Amino acid sequence, domain composition and predicted homology models*

Amino acid sequences of the generated fusion proteins and the predicted domain composition are given in table 4-2 and figure 4-4. Each new fusion had at least two

catalytic domains and one carbohydrate binding module. The fusion of CxnA1.0 with Cfbglu resulted in a fusion (CxnACfbglu) with three catalytic domains (exoglucanase, endoglucanase and β-glucosidase) and one CBM fulfilling the three classical enzyme requirement for the hydrolysis of cellulose. CxnA 2.0, CenBCex and CenDCex contained two catalytic domains each. In addition CxnA 2.0 and CenDCex contained two CBM2s. The CBMs of CxnA 2.0 were located at the ends whereas in the case of CenDCex, one CBM was in the middle and the other at the C-terminus. CenBCex had three CBMs, CBM3 and CBM2 from CenB and a third CBM2 from the Cex. Both CfbgluCenA and Chu2268CenA had two catalytic domains (β-glucosidase and endoglucanase) and the CBM2 from CenA in the middle.

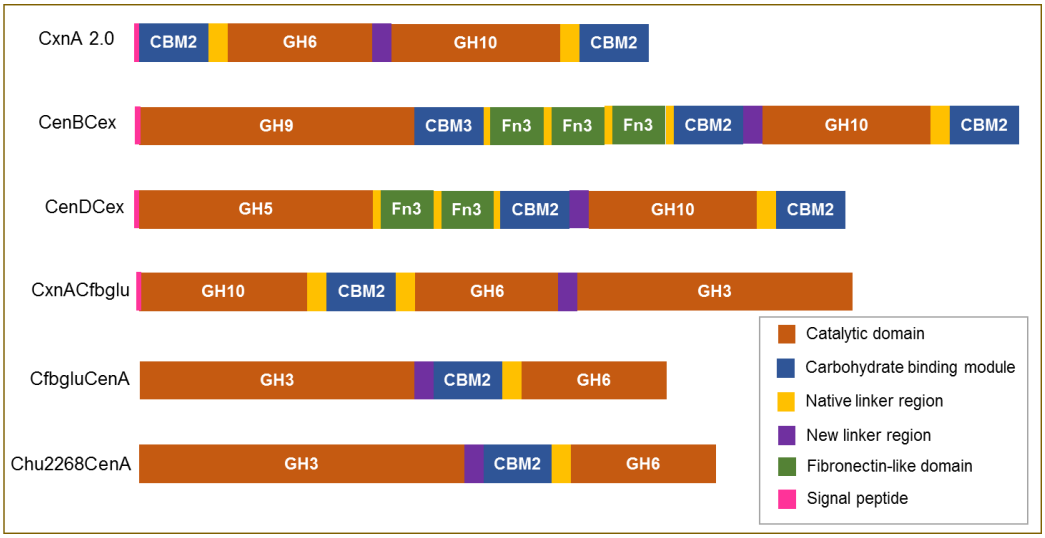



**Figure 4-4: Illustrative diagram of the domain organization/constitution of the fusion proteins**

No changes to the native structure or composition of the fusion partners were made except deletions of the stop and start codons from the N-terminal and C-terminal fusion partners respectively.

**Table 4-2: Amino acid sequences of fusion proteins indicating fusion linker regions**

| Name | Sequence |
|---|---|
| CxnA 2.0 | MSTRRTAAALLAAAAVAVGGLTALTTTAAQAAPGCRVDYAVTNQWPGGFGANVTITNLGDPVSSWKLDWTYTAGQRIQQLWNGTASTNGGQVSVTSLPWNGSIP<br>TGGTASFGFNGSWAGSNPTPASFSLNGTTCTGTVPTTSPTPTPTPTTPTPTPTPTPTPTPTVTPQPTSGFYVDPTTQGYRAWQAASGTDKALLEKIALTPQAYW<br>VGNWADASHAQAEVADYTGRAVAAGKTPMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNPIVILEPDALAQLGDCSGQGDRVGFLKYAAKSLTLKG<br>ARVYIDAGHAKWLSVDTPVNRLNQVGFEYAVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSRNGNGSNGEWCNPRGRALGERPVAVNDGSGLDALLWV<br>KLPGESDGACNGGPAAGQWWQEIALEMARNARW**PTPTTPTPTPPTPTPT**LPAQAATTLKEAADGAGRDFGFALDPNRLSEAQYKAIADSEFNLVVAENAMKWDA<br>TEPSQNSFSFGAGDRVASYAADTGKELYGHTLVWHSQLPDWAKNLNGSAFESAMVNHVTKVADHFEGKVASWDVVNEAFADGGGRRQDSAFQQKLGNGYIETAF<br>RAARAADPTAKLCINDYNVEGINAKSNSLYDLVKDFKARGVPLDCVGFQSHLIVGQVPGDFRQNLQRFADLGVDVRITELDIRMRTPSDATKLATQAADYKKVV<br>QACMQVTRCQGVTVWGITDKYSWVPDVFPGEGAALVWDASYAKKPAYAAVMEAFGASPTPTPTTPTPTPTTPTPTPTSGPAGCQVLWGVNQWNTGFTANVTVKN<br>TSSAPVDGWTLTFSFPSGQQVTQAWSSTVTQSGSAVTVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPTAFSLNGTPCTVG* |
| CenBCex | MLRQVPRTLVAGGSALAVAVGVLVAPLATGAAAAPTYNYAEALQKSMFFYQAQRSGDLPADFPVSWRGDSGLTDGADVGKDLTGGWYDAGDHVKFGFPMAFSAT<br>MLAWGAIESPTGYSKAGSLDELKDNLRFVSDYFVKAHTAPNELYVQVGDGEADHKWWGPAEVMTMARPSHKISASCPGSDVAAETAAALASSAIVLKGDDPAYA<br>ATLVSHAKQLYTFADTYRGAYSDCVTAASAYYKSWSGYQDELVWGAYWLYKATGDATYLAKAEAEYDKLGTENQSTTRSYKWTIAWDNKQFGTYALLAMETGKQ<br>KYVDDANRWLDYWTVGVNGQKVPYSPGGQAVLDSWGALRYAANTSFVALVYSDWMTDATRKARYHDFGVRQINYALGDNPRSSSYVVGFGANPPTAPHHRTAHG<br>SWLDSITTPAQSRHVLYGALVGGPGSPNDAYTDSRQDYVANEVATDYNAGFTSALARLVEEYGGTPLASFPTPEQPDGDQLFVEAMLNQPPSGTFTEVKAMIRN<br>QSAFPARSLKNAKVRYWFTTDGFAASDVTLSANYSECGAQSGKGVSAGGTLGYVELSCVGQDIHPGGQSQHRREIQFRLTGPAGWNPANDPSYTGLTQTALAKA<br>SAITLYDGSTLVWGKEPTGTTTDTTPPTTPGTPVATGVTTVGASLSWAASTDAGSGVAGYELYRVQGTTQTLVGTTTAAAYILRDLTPGTAYSYVVKAKDVAGN<br>VSAASAAVTFTTDTTGETEPPTTPGTPVASAVTSTGATLAWAPSTGDPAVSGYDVLRVQGTTTTVVAQTTVPTVTLSGLTPSTAYTYAVRAKNVAGDVSALSAP<br>VTFTTAAPPVDTVAPTVPGTPVASNVATTGATLTWTASTDSGGSGLAGYEVLRVSGTTQTLVASPTTATVALAGLTPATAYSYVVRAKDGAGNVSAVSSPVTFT<br>TLPVTSTPSCTVVYSTNSWNVGFTGSVKITNTGTTPLTWTLGFAFPSGQQVTQGWSATWSQTGTTVTATGLSWNATLQPGQSTDIGFNGSHPGTNTNPASFTVN<br>GEVCG**PTPTTPTPTPPTPTPT**LPAQAATTLKEAADGAGRDFGFALDPNRLSEAQYKAIADSEFNLVVAENAMKWDATEPSQNSFSFGAGDRVASYAADTGKELY<br>GHTLVWHSQLPDWAKNLNGSAFESAMVNHVTKVADHFEGKVASWDVVNEAFADGGGRRQDSAFQQKLGNGYIETAFRAARAADPTAKLCINDYNVEGINAKSNS<br>LYDLVKDFKARGVPLDCVGFQSHLIVGQVPGDFRQNLQRFADLGVDVRITELDIRMRTPSDATKLATQAADYKKVVQACMQVTRCQGVTVWGITDKYSWVPDVF<br>PGEGAALVWDASYAKKPAYAAVMEAFGASPTPTPTTPTPTPTTPTPTPTSGPAGCQVLWGVNQWNTGFTANVTVKNTSSAPVDGWTLTFSFPSGQQVTQAWSST<br>VTQSGSAVTVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPTAFSLNGTPCTVG* |

| | |
|---|---|
| CenDCex | MHSASRTRARTRVRTAVSGLLAATVLAAPLTLVAAPAQAATGDDWLHVEGNTIVDSTGKEAILSGVNWFGFNASERVFHGLWSGNITQITQQMAQRGINVVRVP<br>VSTQLLLEWKAGTFLKPNVNTYANPELEGKNSLQIFEYWLTLCQKYGIKVFLDVHSAEADNSGHVYNMWWKGDITTEDVYEGWEWAATRWKDDDTIVGADIKNE<br>PHGTQGSTERAKWDGTTDKDNFKHFAETASKKILAINPNWLVFVEGVEIYPKPGVPWTSTGLTDYYGTWWGGNLRGVRDHPIDLGAHQDQLVYSPHDYGPLVFD<br>QKWFQKDFDKASLTADVWGPNWLFIHDEDIAPLLIGEWGGRLGQDPRQDKWMAALRDLVAERRLSQTFWVLNPNSGDTGGLLLDDWKTWDEVKYSTMLEPTLWK<br>HGGKYVGLDHQVPLGGVGSTTGTSISQVGGGTPDTTAPTAPTGLRAGTPTASTVPLTWSASTDTGGSGVAGYEVYRGTTLVGTTTATSYTVTGLAADSAYTFSV<br>RAKDGAGNTSAASAAVTARTAAGGGDVTAPSVPTGLTAGTPTATSVPLTWTASTDTGGSGVTGYEVYRGSTLVARPTGTSHTVTGLSAATAYTFTVRAVDAAGN<br>VSAASAPVGVTTAPDPTTGSCAVTYTANGWSGGFTAAVTLTNTGTTALSGWTLGFAFPSGQTLTQGWSARWAQSGSSVTATNEAWNAVLAPGASVEIGFSGTHT<br>GTNTAPATFTVGGATCTTR**PTPTTPTPTPPTPTPT**LPAQAATTLKEAADGAGRDFGFALDPNRLSEAQYKAIADSEFNLVVAENAMKWDATEPSQNSFSFGAGD<br>RVASYAADTGKELYGHTLVWHSQLPDWAKNLNGSAFESAMVNHVTKVADHFEGKVASWDVVNEAFADGGGRRQDSAFQQKLGNGYIETAFRAARAADPTAKLCI<br>NDYNVEGINAKSNSLYDLVKDFKARGVPLDCVGFQSHLIVGQVPGDFRQNLQRFADLGVDVRITELDIRMRTPSDATKLATQAADYKKVVQACMQVTRCQGVTV<br>WGITDKYSWVPDVFPGEGAALVWDASYAKKPAYAAVMEAFGASPTPTPTTPTPTPTTPTPTPTSGPAGCQVLWGVNQWNTGFTANVTVKNTSSAPVDGWTLTFS<br>FPSGQQVTQAWSSTVTQSGSAVTVRNAPWNGSIPAGGTAQFGFNGSHTGTNAAPTAFSLNGTPCTVG* |
| CxnACfbglu | MPRTTPAPGHPARGARTALRTTLAAAAATLVVGATVVLPAQAATTLKEAADGAGRDFGFALDPNRLSEAQYKAIADSEFNLVVAENAMKWDATEPSQNSFSFGA<br>GDRVASYAADTGKELYGHTLVWHSQLPDWAKNLNGSAFESAMVNHVTKVADHFEGKVASWDVVNEAFADGGGRRQDSAFQQKLGNGYIETAFRAARAADPTAKL<br>CINDYNVEGINAKSNSLYDLVKDFKARGVPLDCVGFQSHLIVGQVPGDFRQNLQRFADLGVDVRITELDIRMRTPSDATKLATQAADYKKVVQACMQVTRCQGV<br>TVWGITDKYSWVPDVFPGEGAALVWDASYAKKPAYAAVMEAFGASPTPTPTTPTPTPTTPTPTPTSGPAGCQVLWGVNQWNTGFTANVTVKNTSSAPVDGWTLT<br>FSFPSGQQVTQAWSSTVTQSGSAVTVRNAPWNGSIPTGGTASFGFNGSWAGSNPTPASFSLNGTTCTGTVPTTSPTPTPTPTTPTPTPTPTPTPTPTVTPQPTS<br>GFYVDPTTQGYRAWQAASGTDKALLEKIALTPQAYWVGNWADASHAQAEVADYTGRAVAAGKTPMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNP<br>IVILEPDALAQLGDCSGQGDRVGFLKYAAKSLTLKGARVYIDAGHAKWLSVDTPVNRLNQVGFEYAVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSR<br>NGNGSNGEWCNPRGRALGERPVAVNDGSGLDALLWVKLPGESDGACNGGPAAGQWWQEIALEMARNARW**PTPTTPTPTPPTPTPT**MGDRFQQAGRPRRRGPARA<br>VNRTGQEDASTGSRERSHACIVSFRHRHPATGHAPGTQQSPPRPPPVTETRRTLMTTTRPSGRQFSDDFLWGSATASYQIEGAHDEGGRGPSIWDTFSRTPGKV<br>LNGDTGDVAVDHYHRVPEDVEIMKSLGLQAYRFSIAWPRIQPTGSGEFNQAGLDFYSDLVDRLIAAGIKPVATLYHWDLPQPLEDEGGWANRATAYRFVEYARK<br>LAEVLGKRVDLWTTLNEPWCSAFLGYASGVHAPASPTRSRPCAPSTTSTSRTASRAARSARSSARTRPSRSREPARDARRRRLRRLRRGQAPHRHHRERGLPRP<br>AARRRVPQGGLRGHRAPHRLVLRRAGRPRAHPHPAHRARRQLLLDRPREEGQRAGADPGKPGPDGHRASEHSSWVGADEVEWLPQPGPHTAMGWNIEPDGLVDL<br>LLELRDRYPSQPLAITENGAAFYDTVSEDGRVHDPERVGYLHDHVDAVGEAIDKGADVRAYFVWSLLDNFEWRYGYDRRFGIVRVDYDTHERIVKDSGLWYREL<br>VRTRTIAPAEDAATYQP* |

| | |
|---|---|
| CfbgluCenA | MGDRFQQAGRPRRRGPARAVNRTGQEDASTGSRERSHACIVSFRHRHPATGHAPGTQQSPPRPPPVTETRRTLMTTTRPSGRQFSDDFLWGSATASYQIEGAHD<br>EGGRGPSIWDTFSRTPGKVLNGDTGDVAVDHYHRVPEDVEIMKSLGLQAYRFSIAWPRIQPTGSGEFNQAGLDFYSDLVDRLIAAGIKPVATLYHWDLPQPLED<br>EGGWANRATAYRFVEYARKLAEVLGKRVDLWTTLNEPWCSAFLGYASGVHAPASPTRSRPCAPSTTSTSRTASRAARSARSSARTRPSRSREPARDARRRRLRR<br>LRRGQAPHRHHRERGLPRPAARRRVPQGGLRGHRAPHRLVLRRAGRPRAHPHPAHRARRQLLLDRPREEGQRAGADPGKPGPDGHRASEHSSWVGADEVEWLPQ<br>PGPHTAMGWNIEPDGLVDLLLELRDRYPSQPLAITENGAAFYDTVSEDGRVHDPERVGYLHDHVDAVGEAIDKGADVRAYFVWSLLDNFEWRYGYDRRFGIVRV<br>DYDTHERIVKDSGLWYRELVRTRTIAPAEDAATYQP**PTPTTPTPTPPTPTPT**PGCRVDYAVTNQWPGGFGANVTITNLGDPVSSWKLDWTYTAGQRIQQLWNGT<br>ASTNGGQVSVTSLPWNGSIPTGGTASFGFNGSWAGSNPTPASFSLNGTTCTGTVPTTSPTPTPTPTTPTPTPTPTPTPTPTVTPQPTSGFYVDPTTQGYRAWQA<br>ASGTDKALLEKIALTPQAYWVGNWADASHAQAEVADYTGRAVAAGKTPMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNPIVILEPDALAQLGDCS<br>GQGDRVGFLKYAAKSLTLKGARVYIDAGHAKWLSVDTPVNRLNQVGFEYAVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSRNGNGSNGEWCNPRGRA<br>LGERPVAVNDGSGLDALLWVKLPGESDGACNGGPAAGQWWQEIALEMARNARW* |
| Chu2268CenA | MKKITVLISIWLSAAAFFSCEKKTEAGSKPQAFDKEVQDLLKNMSLEEKAGQMTQIDIRNLLNNGYGNTDEKLDTARLKEAIQTYHVGSILNCIQAYTPEKWVE<br>LISQIQNEALQSPNKIPVLYGTDAMHGVGFIKDAVLFPHNIGMAASRNDQLVSQAAQVTSTEARSVGLTWNFAPVLDVGREPYWSRFEETFGEDVYITTQMGSA<br>AVQMMEGSDLTSKTNIASCLKHFIGYSAPKNGIDRTQSHIPEIVLREYYLPPFREAINKGASSIMINSAEINGIPCHGNKWLLTDLLRTELGFTGMVVSDWEDV<br>IRLHTWHKVAATPKEAVMMAVNAGVDMSMVPNDYSFPKYLVELVKEGKVSMARIDEAVGRILTLKIKLGLMKNPLPSIADVGVVGSDAHQQIALNAARESITLL<br>KNDKNILPLAKDKKILLVGPAANSLSALHSSWSYTWQGSNESLYPETTKTIREALEASGNKANIRTNATTGFDDAANYDVSFIQKNTAGVDVIIVCVGEAAYAE<br>QPGVIKDLNLPEAQKQLIVAAKKTGKPVIVCLVEGRPRLFPEEEALADAVIMCYRPGSKGADAFAEILYGDINPSGKLPFTYPRYDGDITTYDYKFKETEQQLK<br>PGVSEFVAFNPQWPFGHGLSYTTFAYSNLNVNKSNFTKNDSVLVTVDISNTGARTGKIAVELYSRDHFASITPSERRLRKYTKIELKAGEKQTVSFTIKAADLQ<br>FVNKDLKTVTEAGAFDLMIGNLQTEIYFNE**PTPTTPTPTPPTPTPT**PGCRVDYAVTNQWPGGFGANVTITNLGDPVSSWKLDWTYTAGQRIQQLWNGTASTNGG<br>QVSVTSLPWNGSIPTGGTASFGFNGSWAGSNPTPASFSLNGTTCTGTVPTTSPTPTPTPTTPTPTPTPTPTPTPTVTPQPTSGFYVDPTTQGYRAWQAASGTDK<br>ALLEKIALTPQAYWVGNWADASHAQAEVADYTGRAVAAGKTPMLVVYAIPGRDCGSHSGGGVSESEYARWVDTVAQGIKGNPIVILEPDALAQLGDCSGQGDRV<br>GFLKYAAKSLTLKGARVYIDAGHAKWLSVDTPVNRLNQVGFEYAVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSRNGNGSNGEWCNPRGRALGERPV<br>AVNDGSGLDALLWVKLPGESDGACNGGPAAGQWWQEIALEMARNARW* |

Homology models of fusion proteins were generated (figure 4-5) by submitting the amino acid sequences to the Phyre2 server (Kelley and Sternberg, 2009; Kelley *et al.*, 2015). The sequence for the CenBCex exceeded the 1500 sequence length allowed for intensive mode. The last nine amino acids were therefore deleted and re-submitted. Models were scored with a 0 to 9 confidence key with 9 being the highest and 0 the lowest. The confidence for the CBM domain models were 0 whereas those of the CDs were 9. The models however provide an idea of the sort of interactions between the domains and how mobile the catalytic domain will be. Modelling for CenBCex and CxnACfbglu was not successful although models for the specific catalytic domains were generated.

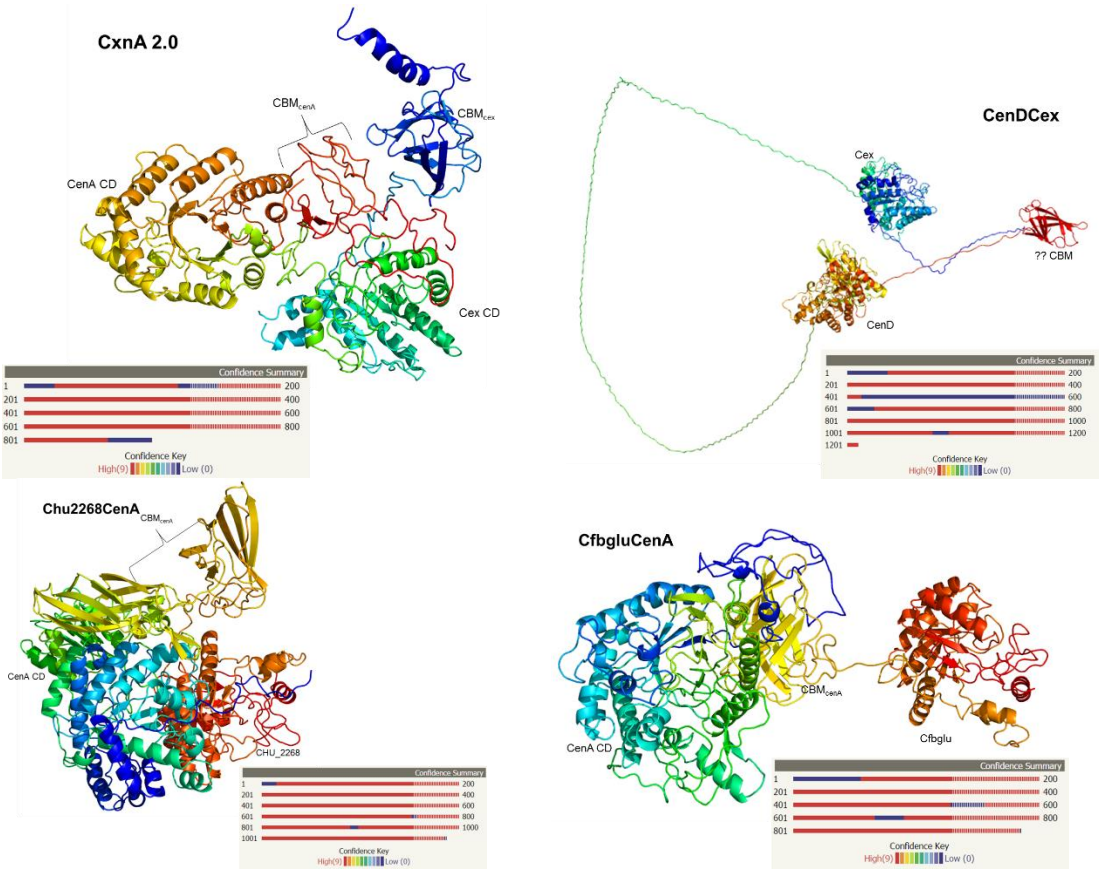



**Figure 4-5: Homology models predicted from sequences of the gene fusions**
Models were generated on the Phyre2 server using the intensive mode with the default parameters on the website (Kelley *et al.*, 2015). The DNA sequence for the fusion genes were constructed manually using the DNA sequences of the individual cellulases and the fusion linkers. Protein sequences modelled were translated from the constructed DNA sequences using the translate tool on the Sequence Manipulation Suite (Stothard, 2000). Models are coloured rainbow (N->C).

### *4.4.3 Enzyme activity*

*E. coli* JM109 transformed with the fused and non-fused constructs showed activities in both live cells and cell lysates except in the case of CenBCex and CHU2268+CenA. CenBCex didn't show exoglucanase activity in the live cells on plates whereas the non-fused CHU2268+CenA didn't show β-glucosidase activity on the plates (figure 4-6). Within cell lysates however all the proteins (fused and non-fused) showed their respective enzyme activities (figure 4-7).

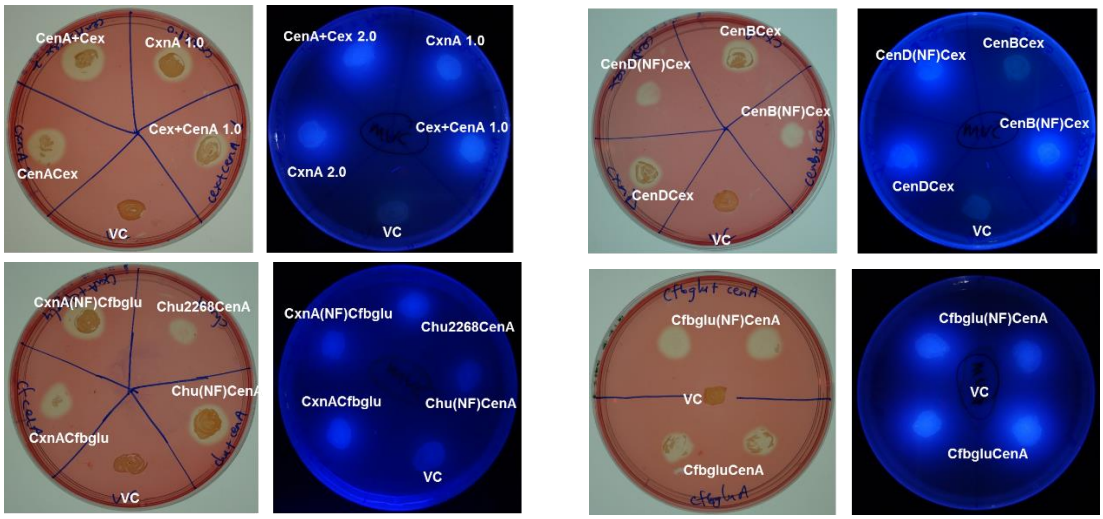



**Figure 4-6: *E. coli* JM109 cells expressing the gene fusions constructs showed enzyme activity of the parent (non-fused) enzymes**

Activity was tested from *E. coli* JM109 cells expressing the gene fusions and the non-fused constructs. Endoglucanase activity was tested on M9-agar plates supplemented with 0.2% CMC. The endoglucanases hydrolyse the CMC and as a result the Congo-Red dye is not picked up leaving zones of clearing. Exoglucanase activity was tested with MUC whereas β-glucosidase activity was tested with MUG. Exoglucanases and β-glucosidases hydrolyse MUC and MUG to release 4-MU which is fluorescent under long wavelength UV light. Cex+CenA1.0 and CxnA1.0 are constructs described in Chapter 3.

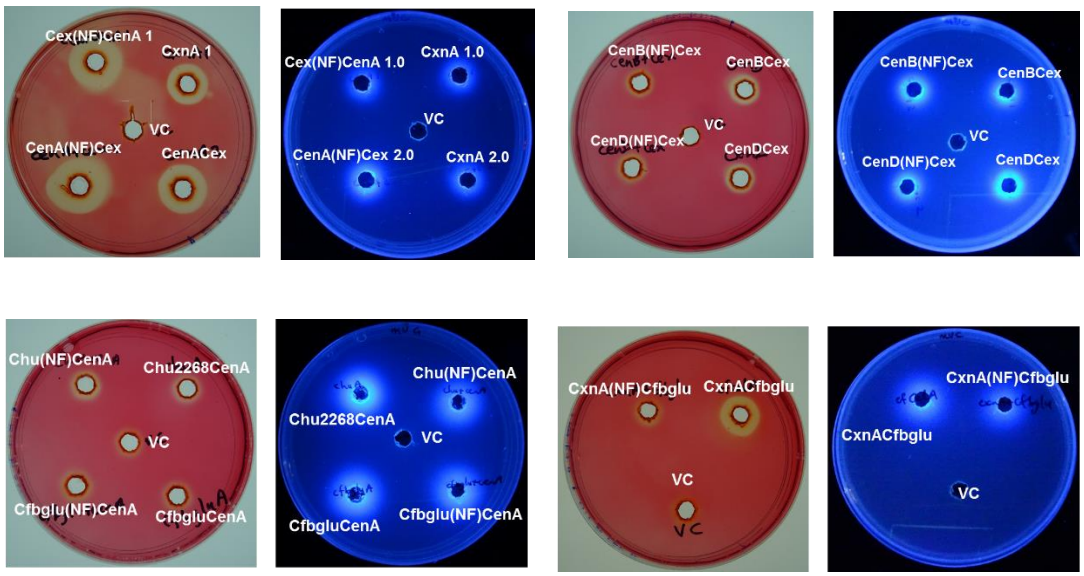
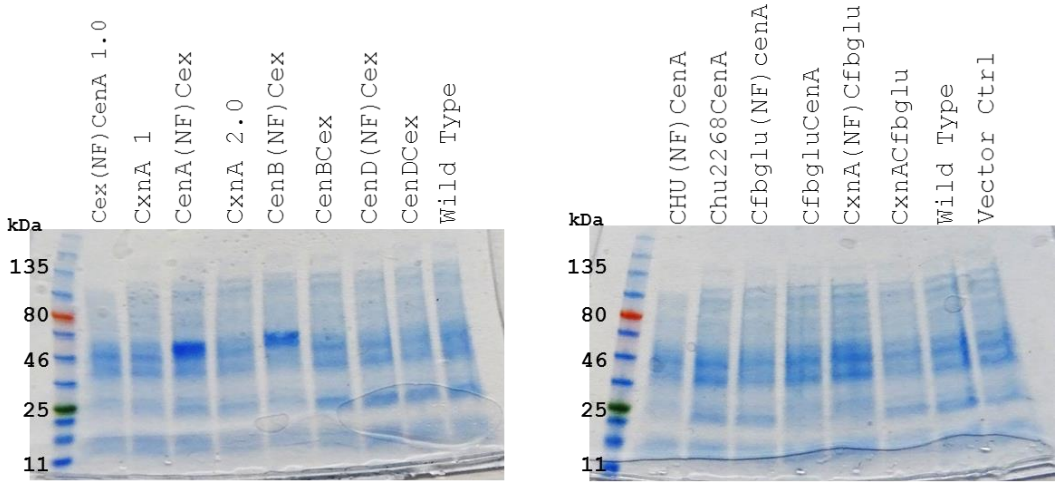



**Figure 4-7: Crude cell lysates from *E. coli* JM109 cells expressing the gene fusions showed enzyme activities of the parent (non-fused) enzymes**

Whole cell lysates were made by sonicating pellets (re-suspended in 1× PBS) from cultures induced for the expression of the gene fusions and non-fused versions. Tests were performed with 50 µl cell lysate as described for the live cells (figure 4-6). Cex+CenA1.0 and CxnA1.0 are constructs described in Chapter 3.

### *4.4.4 Protein gel electrophoresis*

On SDS-PAGE gels, additional visible bands were only seen for Cex from the non-fused Cex+CenA 1.0 (pSB1C3-BBa_J15509, table 2-4) and CenA+Cex 2.0 (pDBM2091, table 2-4) samples.




| Protein | Molecular Weight (kDa) |
|---|---|
| Cex | 51.22 |
| CenA | 46.74 |
| CenB | 109.01 |
| CenD | 78.95 |
| Cfbglu | 62.54 |
| CHU2268 | 83.67 |
| CxnA 1.0 | 84.21 |
| CxnA 2.0 | 95.92 |
| CenBCex | 158.18 |
| CenDCex | 128.12 |
| Chu2268CenA | 129.02 |
| CfbgluCenA | 107.89 |
| CxnACfbglu | 148.31 |

**Figure 4-8: SDS-PAGE of fused and non-fused proteins**

Proteins were expressed in *E. coli* BL21 cells under the *lac* promotor. Cell extracts were analysed on Bio-Rad Mini PROTEAN® TGX® Precast gels (4-15% gradient). Extra bands were only seen for non-fused Cex+CenA1.0 and CenA+Cex2.0. The bands correspond approximately to Cex which is 51.22 kDa. Wells were loaded with 40 ng of cell lysate

### *4.4.5 Growth on cellobiose by C. freundii expressing the chimeric cellulases*

To determine whether growth is related to the cellulosic substrates provided a control experiment was performed using cellobiose as a main carbon source. All constructs including the vector control grew well on cellobiose (figure 4-9) although there were some small differences in the overall level of growth for each construct. Growth of cells expressing the various combinations Cfbglu and CenA (*C. freundii* NCIMB11490/pDBM1350 and *C. freundii* NCIMB11490/pKDCfbgluCenA) lagged in the first three days before rising sharply on day 5 (figure 4-9f). Another notable change in pattern was that of the cells expressing the non-fused Chu2268(NF)CenA (*C. freundii* NCIMB11490/pASChu2268(NF)CenA). Unlike the cells expressing the fused version (*C. freundii* NCIMB11490/pKDChu2268CenA) or vector control, growth lagged during the first day and peaked at day 3 (figure 4-9e).

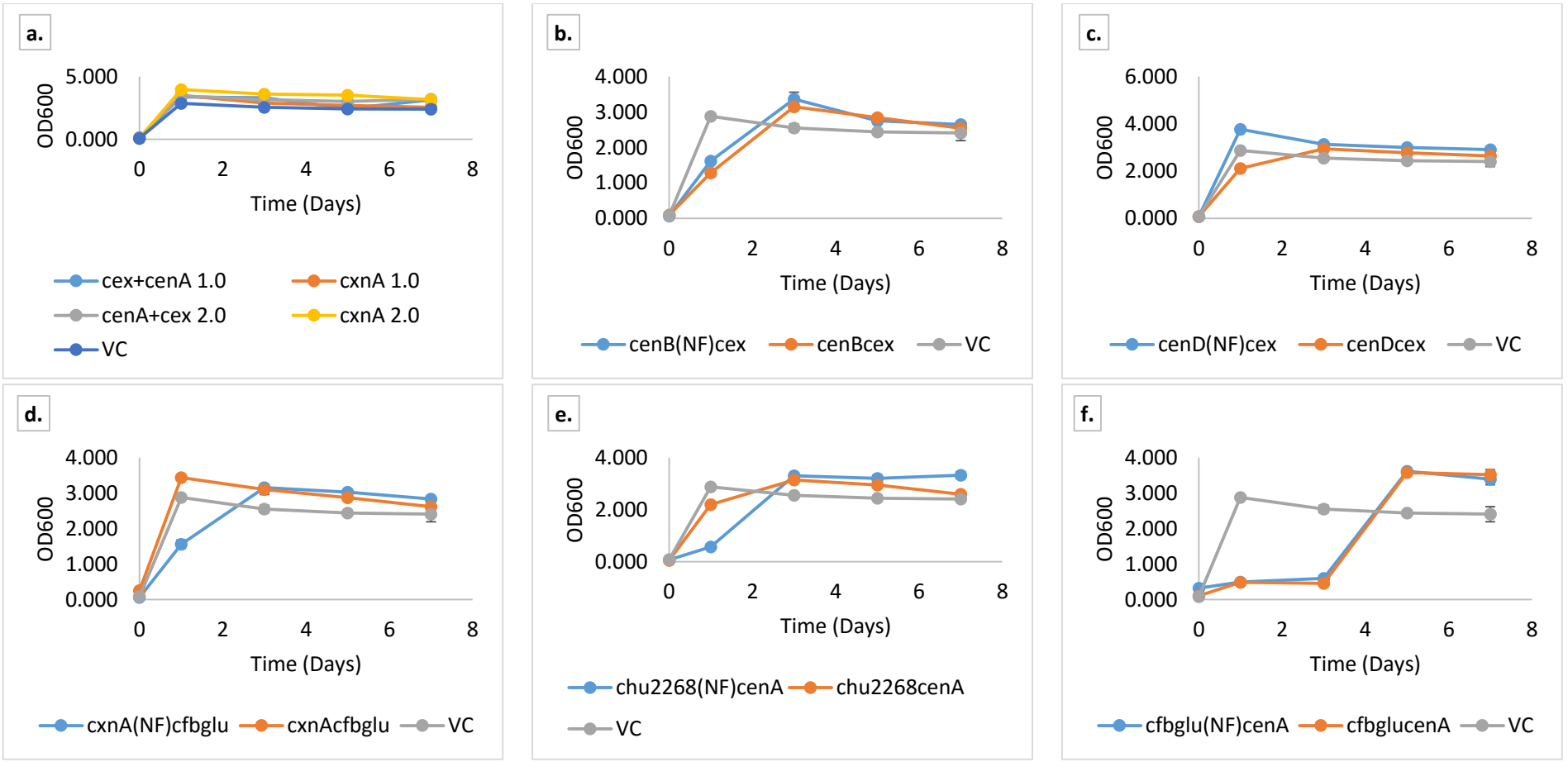



**Figure 4-9: *C. freundii* NCIMB11490 expressing cellulases grows well on cellobiose as well as those not expressing cellulases – monitoring with OD600**

Recombinant constructs ((a) CenA/Cex, (b) CenB/Cex, (c) CenD/Cex, (d) CxnA1.0/Cfbglu, (e) CHU2268/CenA and (f) Cfbglu/CenA) cultured in M9 media supplemented with 4 g/L cellobiose as main carbon. Cultures were shaken at 200 rpm and monitored by checking their turbidity. Plots represent means of two sets of experiments conducted side-by side (each in duplicate), one with conical flasks and the other with 50 ml falcon tubes. The same initial culture was used to inoculate each respective flask/falcon tube. Conical flasks contained 25 ml of culture supplemented with 6 (1.5 x 1.5 cm) squares of Ford 428 blotting paper (about 198 mg in total) whereas falcon tubes contained 7 ml of culture supplemented with 2 pieces of the paper (about 66 mg in total). Cultures were vortexed for one minute every 24 hours. All cultures contained 1 g/l yeast extract.

### *4.4.6 Growth and degradation of filter paper by recombinant C. freundii expressing the chimeric cellulases*

Preliminary investigation of growth of *C. freundii* ATTC8090 expressing the recombinant proteins was monitored using the total protein assay as described in Chapter 5. Total protein assay was used since OD does not give reliable results in the presence of insoluble substrate material. The Ford's 428 Gold Medal blotting paper was used as the main carbon source. The fusion protein CxnA 1.0 and its non-fused version was included in the experiments for comparison with CxnA 2.0 and its non-fused version. In general, total protein increased sharply within 24 hours and declined slowly from day 3 onwards (figure 4-10). Growth of cells expressing the constructs CxnA 1.0 (Chapter 3), CxnA 2.0, CenDCex and Chu2268CenA was higher than cells expressing the non-fused versions. There was no significant difference between growth of cells expressing the other fused and non-fused proteins. The total protein results were inconclusive because growth of the vector (negative) control was similar to those expressing the recombinant cellulases. The experiments were repeated and monitored using plate counts.

Experiments were conducted using either the Ford's 428 or the Whatman GB003 blotting paper. Plate count results for fused and non-fused CenA/Cex and CenD/Cex were similar to the total protein results (figure 4-10 a & c; 4-11 a, b, e & f). Growth of cells expressing the fusion proteins CxnA 2.0 and CenDCex were higher than those expressing the non-fused versions of the protein. Additionally, growth of cells expressing the CxnA 2.0 was also higher than the CxnA 1.0 (figure 4-11 a & b).

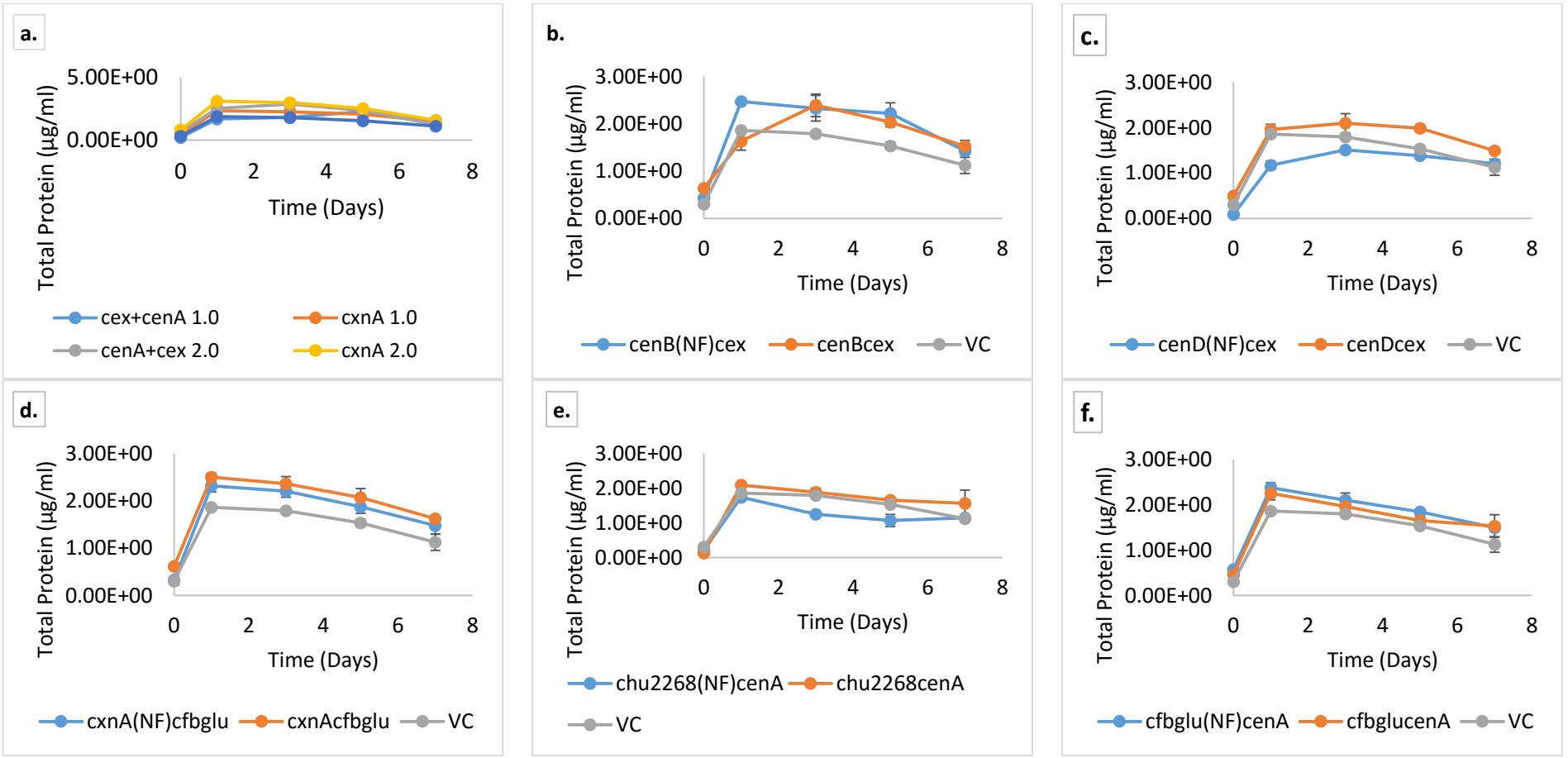
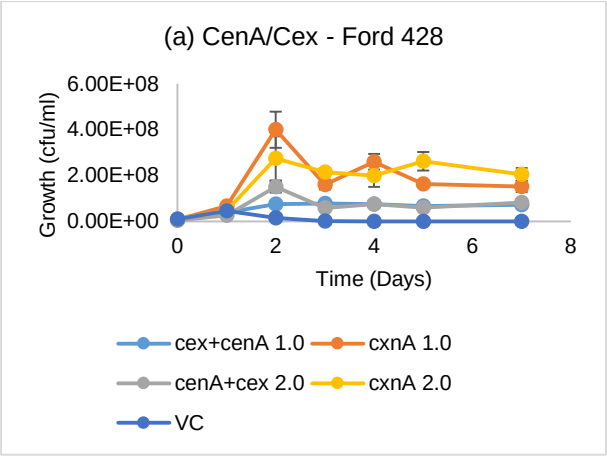
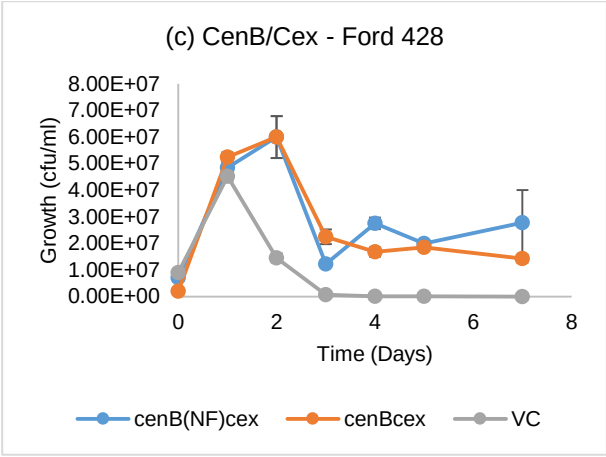
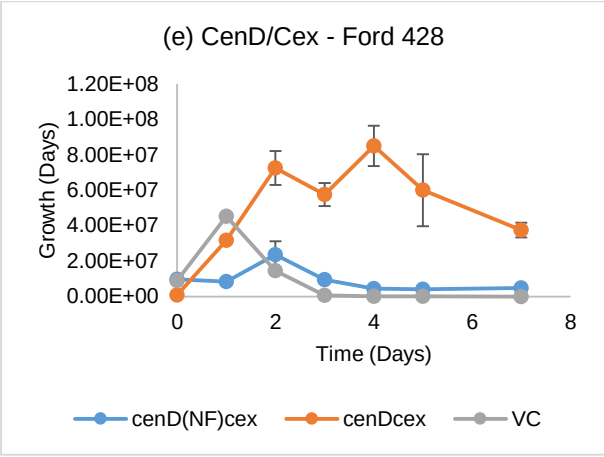


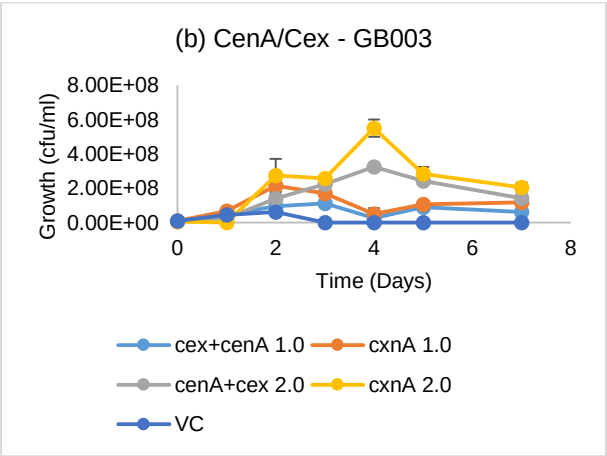

**Figure 4-10: *C. freundii* NCIMB11490 expressing cellulases grows well on cellobiose as well as those not expressing cellulases – monitoring with Bradford total protein**

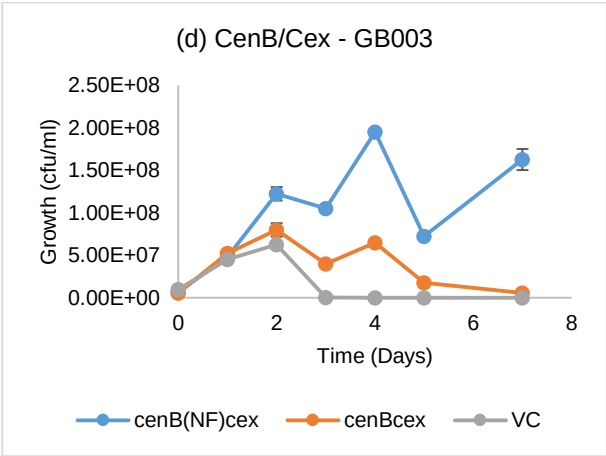
Recombinant constructs ((a) CenA/Cex, (b) CenB/Cex, (c) CenD/Cex, (d) CxnA1.0/Cfbglu, (e) CHU2268/CenA and (f) Cfbglu/CenA) cultured in M9 media supplemented with 4 g/L cellobiose as main carbon. Plots represent means of two sets of experiments conducted side-by side (each in duplicate), one with conical flasks and the other with 50 ml falcon tubes. The same initial culture was used to inoculate each respective flask/falcon tube. Conical flasks contained 25 ml of culture supplemented with 6 (1.5 x 1.5 cm) squares of Ford 428 blotting paper (about 198 mg in total) whereas falcon tubes contained 7 ml of culture supplemented with 2 pieces of the paper (about 66 mg in total). One hundred microliters of culture sample was collected, lysed and analysed for the total protein content. Cultures were vortexed for one minute every 24 hours. All cultures contained 1 g/l yeast extract.

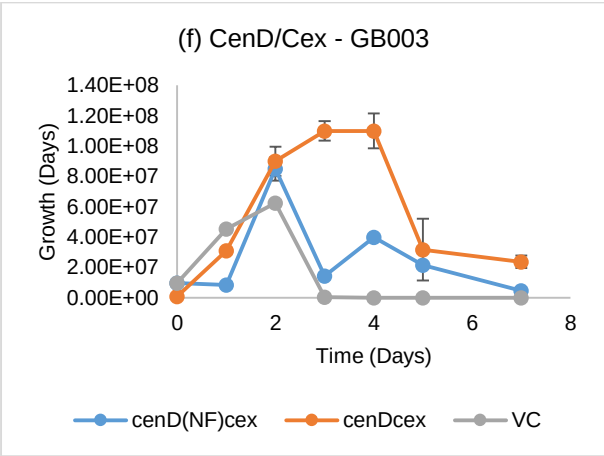

(a) CenA/Cex - Ford 428
Growth (cfu/ml)
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
2
4
6
8
Time (Days)
cex+cenA 1.0
cxnA 1.0
cenA+cex 2.0
cxnA 2.0
VC


(c) CenB/Cex - Ford 428
Growth (cfu/ml)
8.00E+07
7.00E+07
6.00E+07
5.00E+07
4.00E+07
3.00E+07
2.00E+07
1.00E+07
0.00E+00
0
2
4
6
8
Time (Days)
cenB(NF)cex
cenBcex
VC


(e) CenD/Cex - Ford 428
Growth (Days)
1.20E+08
1.00E+08
8.00E+07
6.00E+07
4.00E+07
2.00E+07
0.00E+00
0
2
4
6
8
Time (Days)
cenD(NF)cex
cenDcex
VC


(b) CenA/Cex - GB003
Growth (cfu/ml)
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
2
4
6
8
Time (Days)
cex+cenA 1.0
cxnA 1.0
cenA+cex 2.0
cxnA 2.0
VC


(d) CenB/Cex - GB003
Growth (cfu/ml)
2.50E+08
2.00E+08
1.50E+08
1.00E+08
5.00E+07
0.00E+00
0
2
4
6
8
Time (Days)
cenB(NF)cex
cenBcex
VC


(f) CenD/Cex - GB003
Growth (Days)
1.40E+08
1.20E+08
1.00E+08
8.00E+07
6.00E+07
4.00E+07
2.00E+07
0.00E+00
0
2
4
6
8
Time (Days)
cenD(NF)cex
cenDcex
VC

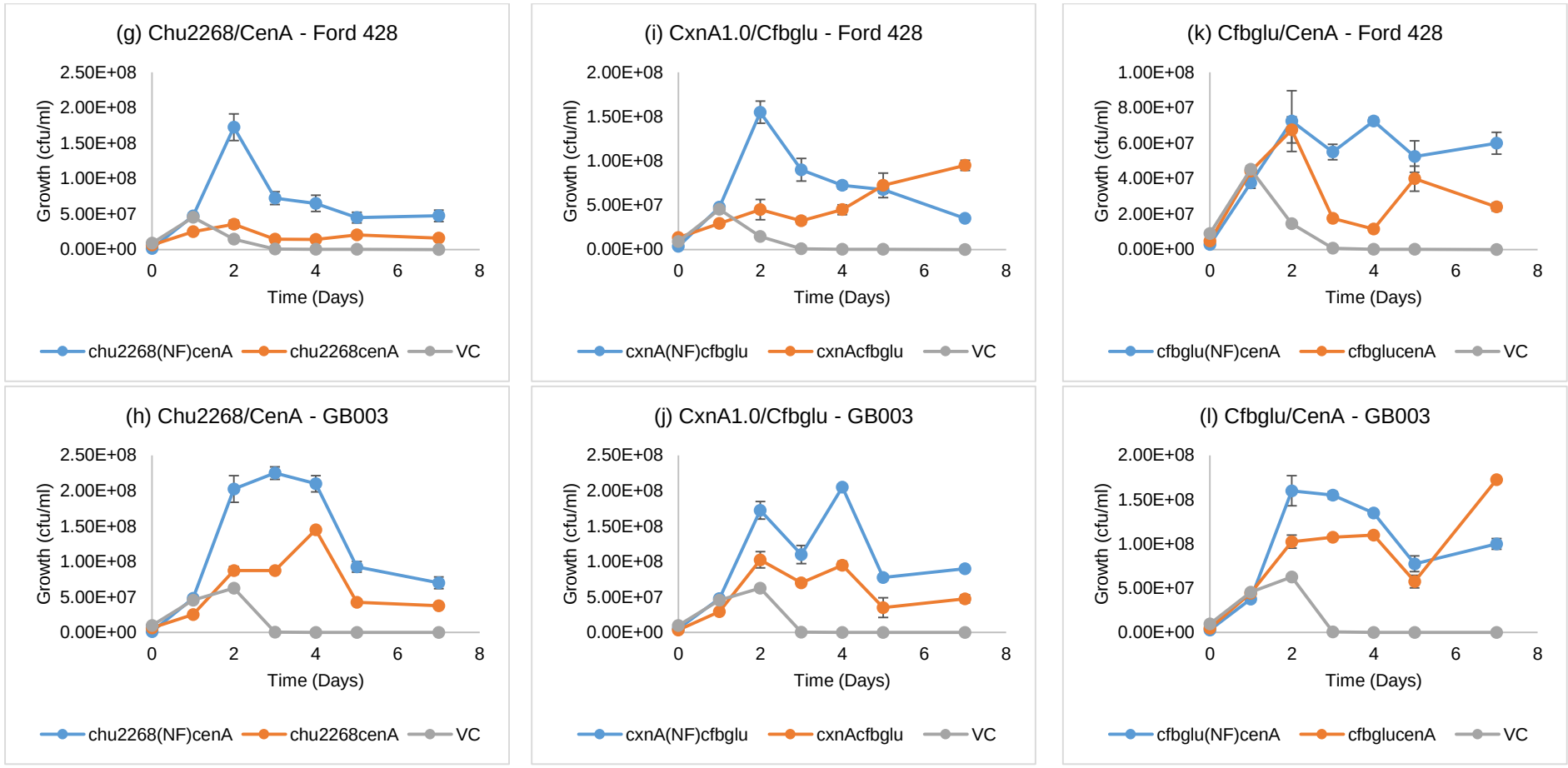



**Figure 4-11: *C. freundii* expressing the recombinant cellulases were able to grow on filter paper**

Cultures were conducted in conical flasks containing 25 ml of M9 culture supplemented with 6 (1.5 x 1.5 cm) squares of Ford 428 blotting paper (about 198 mg in total) and falcon tubes containing 7 ml of culture supplemented with 2 pieces of the paper (about 66 mg in total). Cultures were incubated at 37°C on a shaker (200 rpm) and vortexed for one minute every 24 hours. All cultures contained 1 g/l yeast extract. Plots represent the average of two sets of three biological replicates (a set in falcon tubes and another in conical flasks).

Physical degradation of the paper was not observed in the shake flasks after 7 days of culturing. Degradation however occurred in the falcon tubes and the McCartney bottles between 3 and 7 days. The cells expressing fused CenA/Cex and CenD/Cex degraded filter paper faster than the non-fused versions (figure 4-12 a & c). There was no significant difference in degradation observed between cells expressing fused and non-fused CenB/Cex (figure 4-12b) and CxnA1.0/Cfbglu (figure 4-12e). Cells expressing non-fused CHU2268/CenA (figure 4-12d) and Cfbglu/CenA (figure 4-12f) degraded filter paper faster than those expressing the fused versions.

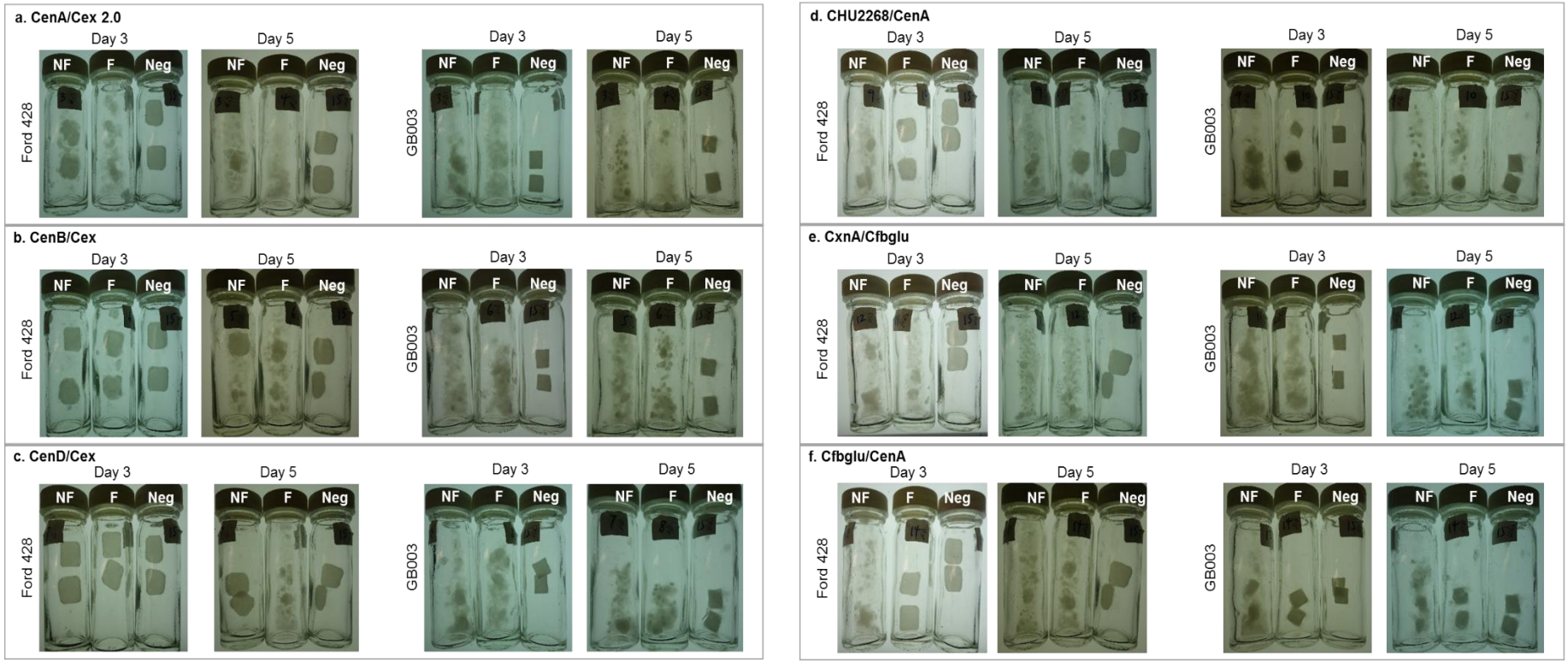



**Figure 4-12: Physical degradation of filter paper by *C. freundii* expressing the recombinant cellulases**

Cellulose utilization and filter paper degradation experiments were conducted in McCartney bottles with two 1.5 x 1.5 cm squares of Ford 428 or two 1 x 1 cm squares of Whatman GB003 blotting papers as main carbon source. Cultures were vortexed for one minute every 24 hours. NF=Non-fused; F=Fused. Images taken on a white light box.

## 4.5 Discussion

It was earlier observed that a fusion of *C. fimi* exoglucanase Cex and endoglucanase CenA enabled *C. freundii* to degrade and utilize cellulose filter paper as a main carbon source for growth better than *C. freundii* with the non-fused version. In order to test synergistic effects of different combinations of cellulases additional fusions were made from a library of modular parts. The MABEL procedure employed for the fusion of was quick and efficient for making fusions. The protocol is similar to the QuickChange site-directed mutagenesis kit protocol (Agilent Technologies, Inc. USA) but simpler and cheaper. There were initial attempts to make fusions using the GenBrick DNA assembly method to make the fusion proteins (Gasiūnaitė *et al.*, 2013). Although this method is quick, it requires considerable amount of time to make all the parts needed for the fusions. With an already existing library of combination parts, the MABEL protocol was therefore thought to be the fastest and easiest way to make fusions for quick testing of synergy between cellulases.

To ensure constructs retained catalytic activity of the parent enzymes, enzyme activity testing was performed for both live cells and cell lysates. All constructs showed catalytic activities within live cells and cell lysates except the CenBCex and CHU2268+CenA which didn't show activity for Cex (exoglucanase) and CHU2268 (β-glucosidase) in live cells respectively. In order to detect enzyme activity, cells must secrete some of the expressed enzymes into the medium to catalyze the substrate. It therefore suggest that CenBCex and CHU2268 may not be secreted. This however is unlikely to be cause of the absence of Cex activity in the CenBCex construct because catalytic activity for endoglucanase (CenB) part was detected in the live cells. In other experiments (data not shown), deletion of the signal peptide of Cex resulted in loss of catalytic activity. However, catalytic activity was restored when Cex without its signal peptide is fused with other enzymes like CenA (CxnA2.0) and CenD

(CenDCex). Further investigation into the loss of Cex activity in live cells is required to make conclusive deductions. The CHU2268 on the other hand not showing β-glucosidase activity is not unusual based on results from other members of the lab. Though the CHU2268+CenA construct did not show activity, a His-tagged version of it shows activity in live cells (A. Salinas-Vaccaro 2014, Personal communication).

The idea of making fusion proteins is to bring their catalytic domains together which leads to synergy (Bayer *et al.*, 1998a; Bayer *et al.*, 1998b; Bayer *et al.*, 2004). Synergism based on colocation was described based on dilution experiments from the fusion of CelY (exoglucanase) and CelZ (endoglucanase) from *Clostridium stercorarium* (Riedel and Bronnenmeier, 1998). Homology models for the fusion proteins were generated to provide an idea about the geometry of the various domains of the fusion protein. Thus this may provide insights into whether when the CBMs bind to cellulose, the catalytic domains will be free to 'move' and catalyze hydrolysis. The use of long rigid linkers in the design of the fusion proteins is to enable this characteristic. Phyre2 uses a template-based modelling (TBM) approach to build protein models, predict ligand binding sites and analyse effects of amino acid variants (Kelley *et al.*, 2015). The result therefore is not a true representation of the crystal structure of the proteins. Phyre2 was chosen over other TBM platforms because it has an intensive mode which models the different fusion partners together. For example when SWISS-MODEL (Arnold *et al.*, 2006; Bordoli *et al.*, 2009; Biasini *et al.*, 2014) was used, it only gave segregated models of the catalytic domains and sometimes the CBM. The models of CxnA1.0 (Chapter 3 of this thesis) and CxnA2.0 were different and representative. The CxnA2.0 showed both catalytic domains and what was presumed to be the CBMs. The CBMs were not distinct as seen in CxnA1.0 or the Cex or CenA models. The CenDCex models on the other hand showed three domains, possibly two CDs and one CBM based on the other models. The model

however had very long loops that may include residues which ought to be part of one or both of the catalytic domains. Since the modelling platform uses already existing structures as a template to model new ones, it might be the case that there is no good structure for a CenD like protein to serve as the template for modelling. The Chu2268CenA model showed the two catalytic domains close together unlike the CfbgluCenA model in which the catalytic domains were separate.

The absence of additional bands on the protein gel requires a brief comment. The result is not unusual as the *lac* promoter is not particularly strong and hence it is not surprising that proteins might not be expressed at high levels. The two extra bands corresponding to Cex from the non-fused CenA/Cex and CenB/Cex shows Cex in these constructs may be expressed in high quantities. Further investigation is needed to characterize this expression. This common feature between the two constructs did not have any correlation with degradation and utilization of cellulose filter paper.

Cellulases that contain multiple separate catalytic domains are rare. Examples have been reported from two extremely thermophilic anaerobes, *Caldocellum saccharolyticum* (Saul et al., 1989; Saul et al., 1990; Gibbs et al., 1992; Te'o et al., 1995) and *Anaerocellum thermophilum* (Bolshakova *et al.*, 1994; Zverlov *et al.*, 1998; Yang *et al.*, 2009). The *celB* gene from *C. saccharolyticum* codes for a protein with molecular weight of about 120 kDa consisting of a central domain with no enzymatic activity linked by pro-thr rich linkers to two enzymatic domains at its ends (Saul *et al.*, 1989; Saul *et al.*, 1990). The first is a bifunctional exoglucanase/xylanase domain whereas the second is an endoglucanase domain. With Cex being a bifunctional exoglucanase/xylanase linked with Pro-Thr linkers to other domains, our fusion

proteins CxnA1.0 and 2.0, CenBCex and CenDCex are synonymous. There have been no reports of CelB being used for cellulose degradation experiments and such data could help in fine tuning the fusion proteins reported here. CelA from *C. saccharolyticum* is also a multidomain enzyme with N-terminal endoglucanase and C-terminal enzyme which has sequence homology with *Clostridium thermocellum* cellulase CelS (Te'o *et al.*, 1995). Another multidomain enzyme from *C. saccharolyticum* has also been reported with N-terminal β-mannanase and C-terminal endoglucanase linked by a Pro-Thr rich linker (Gibbs *et al.*, 1992). The occurrence of Pro-Thr rich linkers in these multidomain cellulases suggests they are most likely the ideal type of linkers for re-engineering nature's toolbox as was done in this study.

A key conclusion from the results of this study is that not all fusions or co-localizations are likely to improve catalytic activity. The CxnA 1.0 fused with the β-glucosidase (Cfbglu) resulted in a fusion which could not enable *C. freundii* to utilize filter paper better than those expressing CxnA and Cfbglu separately. Furthermore, lower viable cell counts were seen with cells expressing CxnA1.0/Cfbglu (fused or non-fused) compared to CxnA1.0 alone. It was expected that as the CxnA1.0/Cfbglu degraded filter paper at a similar rate as CxnA1.0, the recombinant *C. freundii* will grow as much but that wasn't the case. Since *C. freundii* produces β-glucosidase naturally the extra β-glucosidase may be negatively affecting the entire reaction. It will thus be interesting to investigate the performance of the CxnA1.0/Cfbglu constructs in *E. coli* or a chassis that does not produce β-glucosidases. It has been reported that arrangement of the β-glucosidase and cellulase is critical for domain-domain interaction and optimal performance.

## 4.6 Conclusions

This study has revealed that not all fusions or co-localizations of cellulases could potentially enhance the overall catalytic activity of enzymes. The *C. fimi* cellulases CenA and Cex have been found to be the best candidates for gene fusion that leads to increased catalytic activity. In addition to CenA, CenD has also been demonstrated as a good fusion partner to Cex that improves cellulose utilization. The CenA/Cex fusions could potentially have industrial applications.

## 4.7 Limitations and further experiments

A limitation to the conclusions drawn is that there are no quantitative data on the expression levels and activity of the proteins. Without this data, it cannot be concluded as to whether the ‘success’ or ‘failure’ of fusion proteins in giving better results than unfused proteins is due to (i) higher/lower expression levels, (ii) reduced levels of activity of enzymes due to e.g. steric hindrance or improper folding or (iii) increased/decreased burden on cells due to protein expression and/or secretion. Further experiments should be performed to determine quantitatively the expression levels of the fusion proteins. In addition, the fusion proteins need to be purified and analysed to determine their crystal structures and stability.

# Chapter 5

## Two-colour fluorescence fluorimetric analysis for direct quantification of bacteria in cultures with insoluble cellulosic substrates

### 5.1 Summary

Engineering microbial systems for cellulose degradation requires testing and characterization of large libraries of enzyme combinations and their performance in a suitable chassis. This requires enzyme activity testing and *in vivo* monitoring of cell growth on cellulose as a main carbon source. Estimation of bacterial growth on cellulose is however complicated and cannot be performed by rapid traditional methods such as spectrophometric measurements at 600 nm (OD600) due to the presence of insoluble cellulosic particles in the growth media. Colony counts are the only suitable alternative but these are laborious and not high-throughput. Enumeration of bacteria by flow cytometry is another alternative but the presence of cellulosic particles must be dealt with. The method of enumerating bacteria stained with SYBR I Green nucleic acid gel stain and Propidium Iodide (SYBR-I/PI) by flow cytometry was adapted for a fluorimeter and evaluated. The fluorimetry adapted SYBR-I/PI enumeration of bacteria in turbid growth media had direct correlations with OD600 ($p>0.001$). When tested in media containing insoluble particles, the method gave reproducible results which OD600 could not give. An alternative method based on the assessment of total protein using the Pierce Coomassie Plus (Bradford) Assay was also evaluated and compared. In all, the SYBR-I/PI method was found to be the quickest and most reliable. The protocol is thus potentially useful for assessing *in vivo*

activity of cellulose degrading enzyme systems. To enable comparison of fluorescence results across labs and instruments, a fluorescence intensity standard unit, the equivalent fluorescent DNA (EFD) was proposed and evaluated.

## 5.2 Background

Monitoring bacterial growth is essential for assessing many microbial applications. Determination of bacterial cell numbers can be done by direct and indirect methods. The commonest and oldest methods for this are turbidimetric measurements (optical density at 600 nm, OD600) and viable (plate/colony) counts (Breed and Dotterrer, 1916; Koch, 1970). Turbidimetric measurements are indirect methods which are fast and usually preferred when a large number of cultures are to be counted. The readings obtained from these measurements are a representation of the cell numbers (Koch, 1970). However to obtain definitive numbers, the readings must be correlated initially with cell number determined by other means (eg. plate counts). Plate counts on the other hand gives a direct measure of viable cells within the sample. Direct enumeration by microscopy using Petroff-Hausser counting chambers can also be performed (Treuer and Haydel, 2011). The aforementioned methods are however not universally applicable due to various limitations. Turbidimetric methods are unreliable for direct enumeration of bacteria cells in media containing insoluble substances such as food and environmental samples. Unavailability of suitable culture media and low concentrations of viable bacteria are major limitations to plate counting. To get around these challenges, different approaches based on the use of fluorochromes have been devised to investigate microbial viability and density (Diaper *et al.*, 1992; Kaprelyants and Kell, 1992; Caron and Badley, 1995; Porter *et al.*, 1996; Barbesti *et al.*, 2000; Foladori *et al.*, 2010; Tamburini *et al.*, 2014).

Fluorochromes used in staining and enumerating bacteria cells by flow cytometry are based on membrane integrity, DNA binding and energy transfer between the fluorochromes (Humphreys *et al.*, 1994; Sgorbati *et al.*, 1996; Barbesti *et al.*, 2000; Gregori *et al.*, 2001). Barbesti et al (2000) demonstrated that when DNA is simultaneously stained by SYBR-I (membrane permeant) and PI (non-membrane permeant) there is a decrease in the fluorescence of SYBR-I and an increase in the fluorescence of PI. This is due to a strong energy transfer between the two fluorochromes which facilitates discrimination between living and dead bacteria. This transfer is due to the extremely high quantum yield of DNA bound SYBR-I complex (~0.8, Molecular Probes Inc., USA) and the overlapping of its emission spectrum (figure 5-1) with the absorption spectrum of PI. As a result, the fluorescence of SYBR-I is 'quenched' by PI when stained with both. 'Dead cells' are regarded as cells with compromised membranes. The compromised membrane integrity allows both PI and SYBR-I to permeate the cells at which point such cells will fluoresce red (PI) when excited. 'Live cells' on the other hand allows only SYBR-I to permeate and when excited, fluoresce green.

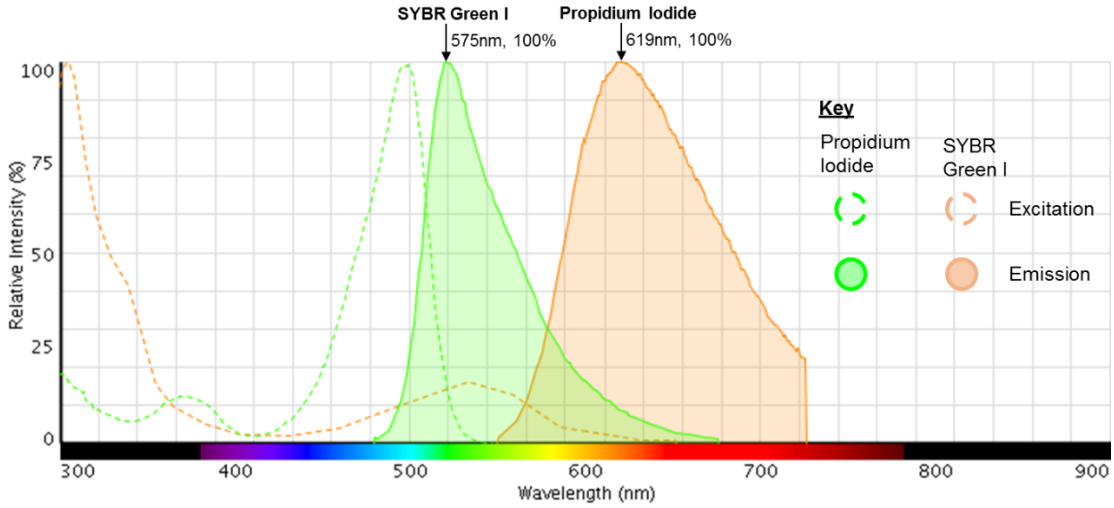



**Figure 5-1: Excitation and emission spectra of SYBR green I and propidium iodide**

(Data from Fluorescence SpectraViewer, Life Technologies)

Fluorescence has been used in quantitation for a long time. Although the technique has improved over time, a major challenge it faces is standardization and references for fluorescent measurements. A special issue on "Quantitative Fluorescence Cytometry: An Emerging Consensus" published by the journal *Cytometry* identified some of these challenges (Lenkei *et al.*, 1998b). Among them were **(1)** inter-laboratory comparisons (Purvis and Stelzer, 1998; Waxdal *et al.*, 1998; Zenger *et al.*, 1998), **(2)** instrumentation (Purvis and Stelzer, 1998; Wood, 1998; Wood and Hoffman, 1998), and **(3)** reagent and calibration standards (Gratama *et al.*, 1998; Lenkei *et al.*, 1998a; Purvis and Stelzer, 1998; Schwartz *et al.*, 1998; Shapiro *et al.*, 1998; Wood and Hoffman, 1998; Zhang *et al.*, 1998). A common response to these challenges has been the development of a fluorescence intensity standard (FIS), the MESF (molecules of equivalent soluble fluorochrome) for use in flow cytometry (Gaigalas *et*

*al.*, 2001; Schwartz *et al.*, 2002; Wang *et al.*, 2002; Schwartz *et al.*, 2004). The MESF is based on a comparison between the number of fluorophores in two solutions, where one solution is a standard with known values. The standard is often a suspension of labelled microbeads.

Naturally occurring cellulose degrading microbes use a battery of multiple catalytic enzymes to hydrolyse cellulose. We have described applications of synthetic biology that expand the technical capabilities of engineering efficient cellulose degrading enzyme systems (French *et al.*, 2015) making it potentially easier to develop an ideal biofuel producing microorganism (IBPM) (French, 2009). The IBPM must be able to (1) produce enzymes required for the hydrolysis of cellulose, (2) take up the hydrolysed product and (3) produce high value products without poisoning itself (French *et al.*, 2015). Each of these stages requires tuning to get the best output. Characterization and fine tuning of microbial cellulose degrading systems requires reliable methods for monitoring growth of cells on cellulose as a main carbon source. Substrates used for such experiments (eg. avicel, paper and pre-treated plant materials) are generally insoluble. This makes the use of turbidimetric methods (eg. OD600), which are fast and suitable for screening large number of samples, unreliable. Colony counting on the other hand is an arduous process. Dual staining of samples with SYBR-I and PI has been demonstrated as a useful method for quantifying bacteria in environmental and food samples using flow cytometry. The application of this method for fluorimetry has not been reported. Here, SYBR-I/PI dual staining was adopted and used to evaluate the cell content of cultures containing microcrystalline cellulose (avicel). A major challenge for fluorescence measurement is the inability to make comparable fluorescence intensity measurements across laboratories and between different instruments. In response to this, a FIS based on

DNA stained with SYBR-I and PI was also developed. This standard, like the MESF used in flow cytometry, is based on equivalency between the intensity of fluorophores in two solutions, one standard (known concentration(s) of DNA) and the other the unknown sample. An alternative method for semi-quantification of bacteria using the total protein content of the sample was also evaluated and compared with the two colour fluorescence method.

## 5.3 Methods

### *5.3.1 Reagents, equipment and sample preparation*

The SYBR-I/PI protocol was evaluated using DNA and cell suspensions of *Escherichia coli* strains JM109 and DH5α and *Citrobacter freundii* strains ATCC8090 and SBS197. The SBS197 strain was obtained from the Biology Teaching Laboratory of the School of Biological Sciences (SBS), University of Edinburgh. *E. coli* and *C. freundii* were chosen because they are the hosts we use in our biomass degradation experiments. *E. coli* and *C. freundii* were grown in 100 ml Luria Broth (LB) overnight at 37°C, 200 RPM. The cells were centrifuged at 5000 x g for 10 minutes, the supernatant discarded and the pellet resuspended in 5 ml 1 x PBS. Human genomic DNA (Cat. No. BIO-35025, 200 ng/µl) was obtained from Bioline, UK and used as a calibration standard. Additionally, plasmid DNA was extracted from *E. coli* JM109 harbouring the BioBrick plasmid, pSB1A2-BBa_K523025 using the QIAGEN plasmid midi kit (QIAGEN GmbH, Germany) according to the manufacturer's instructions. Dilutions of the DNA were made in nuclease free water and stained with the fluorochromes. Propidium iodide (Cat. No. P3566, 1 mg/ml stock solution) and SYBR-I (Cat. No. S-7563, 10,000 X concentrate) were obtained from Life Technologies, UK. SYBR-I dilutions were made using Pierce™ Dimethylsulfoxide (DMSO) (Cat. No.

20688, Thermo Scientific). Fluorescence measurements were taken using the Modulus™ Single Tube multimode reader (P/N 998-9203, Turner BioSystems, USA).

### *5.3.2 Fluorescence staining of cells and DNA*

One ml of diluted DNA or cell suspension was stained with 10 µl each of SYBR-I (1:30 dilution of commercial stock) and PI in a cuvette (FisherBrand FB55147, Fisher Scientific, UK). The samples were covered with parafilm, mixed by inverting about 6 times and incubated in the dark for 15 minutes at room temperature. The polarity of PI allows it to penetrate only cells with compromised membranes whereas SYBR-I is able to permeate all cells. In free DNA or dead cells, staining with the two fluorochromes activates an energy transfer from SYBR-I to PI (Barbesti *et al.*, 2000). As a result, 'dead' cells or DNA fluoresce red whereas live cells emit green fluorescence.

### *5.3.3 Fluorescence measurement and energy transfer*

Fluorescence was measured using the blue (P/N 9200-040, $\lambda_{ex}$ = 460 nm, $\lambda_{em}$ = 515 - 570 nm) or green (P/N 9200-042, $\lambda_{ex}$ = 525 nm, $\lambda_{em}$ = 580 - 640) Modulus™ fluorescence kits (Turner BioSystems Inc., USA). Measurements obtained from the instrument were given in fluorescence units (FSU). To convert the FSU to a FIS which can be compared between instruments and across labs, a standard was generated using dilutions of DNA stained with either SYBR-I, PI or both. Each calibration standard was diluted to a volume of 10 ml. Aliquots of 1 ml were transferred to labelled cuvettes in triplicates for SYBR-I, PI and SYBR-I/PI staining. All samples were measured individually in both the red and green fluorescence channels. Energy transfer from SYBR-I to PI was evaluated.

### *5.3.4 Discrimination and semi-quantification of cells*

To assess the discriminatory capacity of the protocol, various degrees of damage were induced in the *E. coli* and *C. freundii* cells. To induce damage, 500 µl of cell suspension was ultrasonicated at 10 µm (amplitude) for various intervals. Cell damage was assessed by performing total protein quantification on 100 µl of each sample using the Pierce Coomassie plus (Bradford) assay kit (ThermoScientific, UK). The ultrasonicated samples were then stained with SYBR-I, PI or both. Measurements were taken for all samples in both the green and red fluorescence channels.

### *5.3.5 Determination of how presence of cellulose particles affects fluorescence*

To assess the effects of addition of cellulose on fluorescence, 20 mg of cellulose powder (20 µm microcrystalline powder, Aldrich Chemistry, Cat. No. 310697) was added to cell suspensions. Two sets of cell suspensions were prepared at different cell densities. The turbidity of cells was determined by measuring the absorbance at 600 nm using the absorbance module (Model E6076, GLOMAX MultiJR, Promega, UK) with the Modulus™ reader. To one set of cells, cellulose powder was added and the turbidity re-measured. Both sets of samples were stained in parallel with SYBR-I/PI and the fluorescence determined as described above.

### *5.3.6 Total protein assay*

The Pierce Coomassie plus (Bradford) assay was used to determine the total protein concentration in cells. Reactions were set up in 1 ml volumes containing 100 µl of sample and 900 µl of Bradford reagent and mixed. To determine protein content of

ultrasonicated cell suspensions, the reaction mixture was incubated at room temperature for 10 minutes according to the manufacturer's instructions after which absorbance was measured at 600 nm using the Modulus™ reader. For other cell suspensions, the reaction mixture was incubated at 65°C for one hour to lyse the cells. The samples were allowed to cool for one hour at room temperature during which time cellulose particles settle at the bottom that so they do not interfere with the absorbance measurements. Absorbance values were converted to protein concentration by comparing with a standard curve prepared with dilutions of the 2 mg/ml bovine serum albumin (BSA) included in the kit.

### *5.3.7 Statistical analysis*

Microsoft Excel was used for data entry, organization and generation of graph plots. Statistical tests were conducted in IMB SPSS version 21. Correlations and tests for linearity between fluorescence and DNA concentration or cell density were analysed using Pearson's product-moment correlation tests. To determine whether fluorescence depended on cell density or degree of damage, Kendall's rank correlation (tau b) was used. All experiments were performed in at least three replicates and the means plotted with standard error values.

## 5.4 Results

### *5.4.1 Staining characteristics and energy transfer*

Both stains fluoresced in either channel due to the overlap between their absorption and emission spectra (figure 5-2). Furthermore, fluorescence values recorded for SYBR-I were higher than those of PI for the same sample. SYBR-I fluorescence in the red channel was 10.06 % (SD = ± 2.16) of the fluorescence recorded from the red

channel. Fluorescence of propidium iodide in the green channel was 0.34 % (SD = ±0.25) of that in the red channel. Background fluorescence of PI is not likely to affect fluorescence of SYBR-I in the green channel but that of SYBR-I will potentially affect the fluorescence of PI in the red channel.

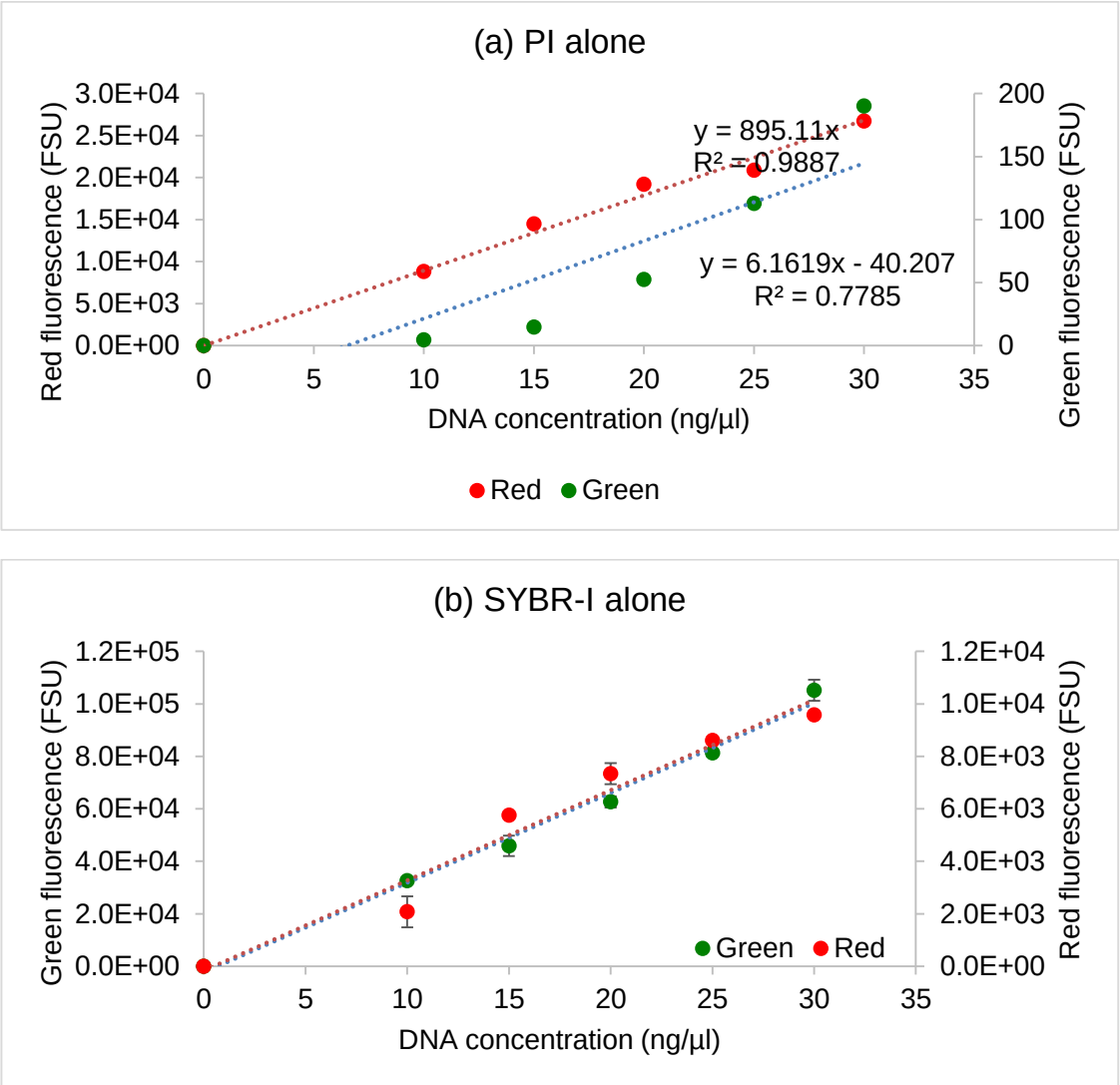



**Figure 5-2: Red and green fluorescence when DNA was stained with SYBR-I or PI**

DNA was serially diluted and stained with SYBR-I and PI. Fluorescence measurements were measured in both the green and red fluorescence channels. Plots represent means of three biological replicates tested side by side.

Energy transfer was assessed after compensation for SYBR-I and PI emission in the red and green channels respectively as described by Barbesti et al (2000). The energy transfer from SYBR-I to PI was evaluated as a decrease in green SYBR-I fluorescence and increase in PI fluorescence of DNA samples stained with both fluorochromes as compared to the same samples stained with only one of the two fluorochromes. Reduction of SYBR-I fluorescence was 98.6 % (SD = ±0.18) whereas increase in PI was 120.2 % (SD = ±15.5) indicating a strong discriminatory capability (table 5-1 and figure 5-3).

**Table 5-1: Energy transfer from SYBR-I to PI**

| Amount of DNA (ng/µl) | SYBR-I decrease (%) | Standard Deviation (±) | PI increase (%) | Standard Deviation (±) |
|---|---|---|---|---|
| 10 | -99.0 | 0.27 | 146.7 | 23.31 |
| 15 | -99.8 | 0.07 | 121.9 | 17.84 |
| 20 | -99.4 | 0.14 | 98.4 | 31.13 |
| 25 | -99.7 | 0.11 | 115.6 | 31.60 |
| 30 | -99.4 | 0.13 | 118.6 | 37.61 |

Values are means of three measurements. SYBR-I fluorescence was almost quenched when DNA was stained with both stains.

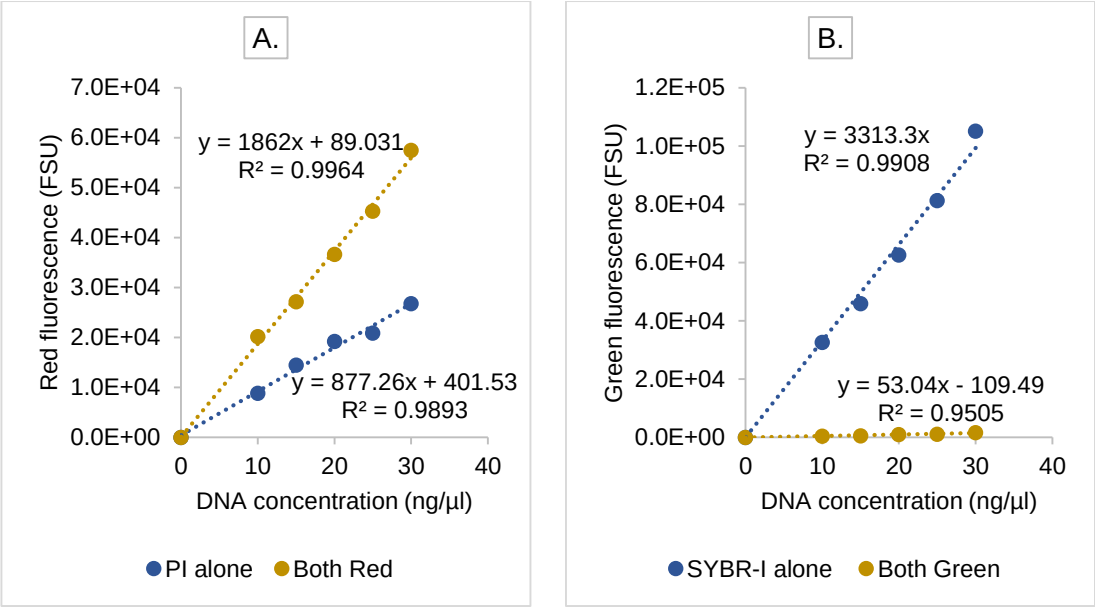



**Figure 5-3: Differences between single and dual staining of DNA**

DNA was serially diluted and stained with either SYBR-I or PI alone and fluorescence determined. Fluorescence determination was performed in the red fluorescence channel for PI alone (a) and in the green fluorescence channel for SYBR-I alone (b). Dual stained DNA was measured in both the red and green fluorescence channels. Six replicates were performed.

### *5.4.2 Definition of a equivalent fluorescent DNA (EFD) unit*

A fluorescence intensity standard was developed based on the fluorescence of DNA stained with either of the two fluorochromes to enable comparison of fluorescence across laboratories and instruments. An equivalent fluorescent DNA (EFD) unit was defined as the amount of fluorescence obtained from staining 1 ng/µl of DNA for 15 minutes at room temperature (25°C). A best fit line obtained for each stain was used to determine the EFD unit for that particular stain after compensating for fluorescence of SYBR-I and PI in the red and green channels (figure 5-4).

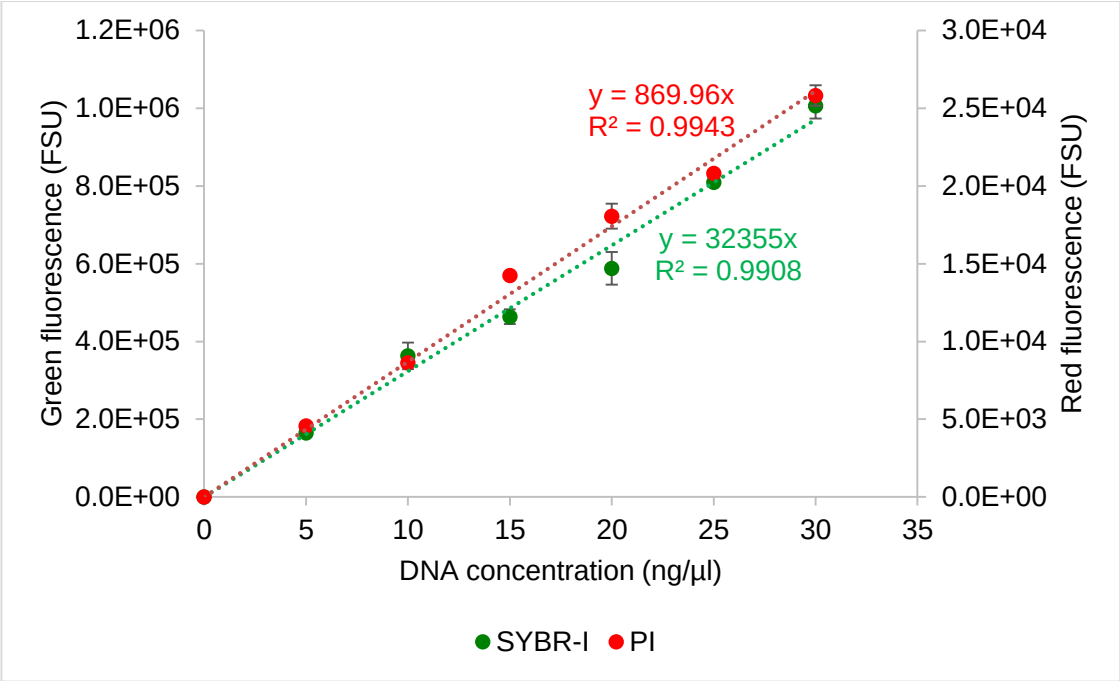



**Figure 5-4: Fluorescent DNA standard curve for SYBR-I and PI**

DNA was serially diluted and the concentrations determined using the nanodrop. Diluted DNA was stained with either SYBR-I or PI and the fluorescence measured in the green or red fluorescence channels respectively. Plots represent six replicates of each dilution that was prepared and measured.

### *5.4.3 Assessment of single fluorescent staining of bacteria cultures*

Quantification of bacteria cells was assessed first by staining dilutions of *E. coli* suspensions with single fluorochromes. There was no correlation between cell density (OD600) and the fluorescence obtained when PI alone was used to stain cells (figure 5-5). On the other hand, a strong correlation ($R^2 = 0.988$) between cell density and fluorescence was obtained with SYBR-I alone. Although SYBR green I is a green fluorescent stain, red fluorescence was detected from cultures stained with SYBR green I. This fluorescence was however 3 to 6 fold lower than the green fluorescence detected. The direct correlation between cell density and fluorescence demonstrates that SYBR-I permeates all cells.

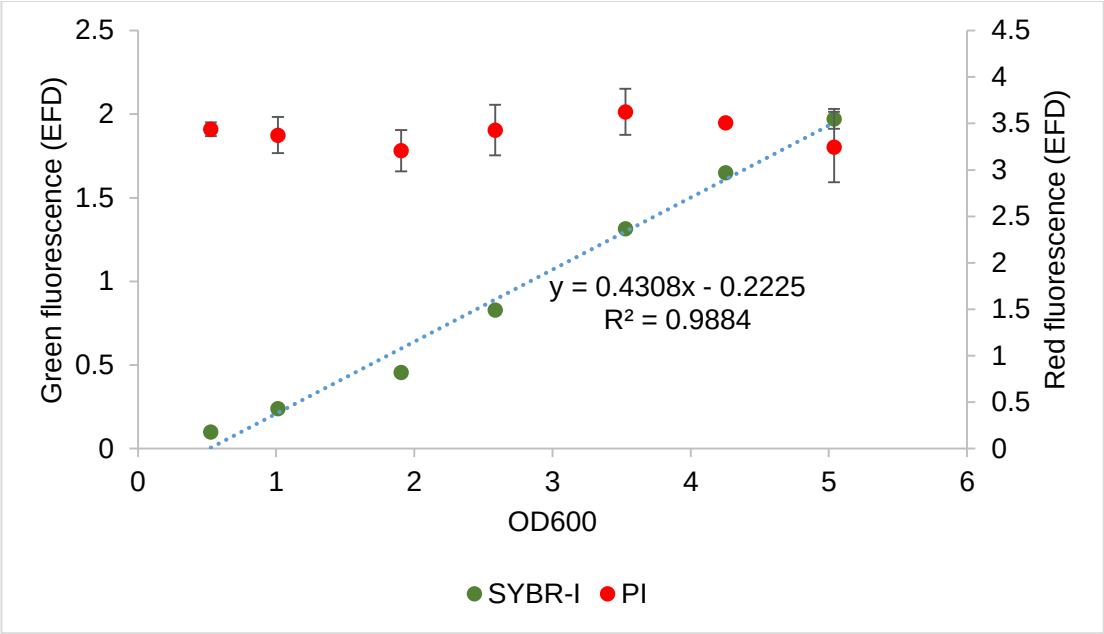



**Figure 5-5: Fluorescence of cell cultures stained with single fluorochromes**

Cell suspensions were prepared by resuspending pellets from overnight cell cultures. The pellets were diluted and stained with either SYBR-I or PI. Two biological replicates (each made up of three technical replicated) were performed.

To investigate the discriminatory abilities of the two stains, varying levels of cell damage was used as a marker. The cell damage was induced by subjecting aliquots of the same cell suspension to different durations of ultra-sonication. Cell pellets of *E. coli* JM109 obtained from an overnight culture were resuspended in PBS and sonicated. Increasing duration of sonication gave higher extracellular protein concentrations indicating release of protein from cells whereas the cell density measured by absorbance at 600 nm reduced (figure 5-6) confirming that some cells have been damaged.

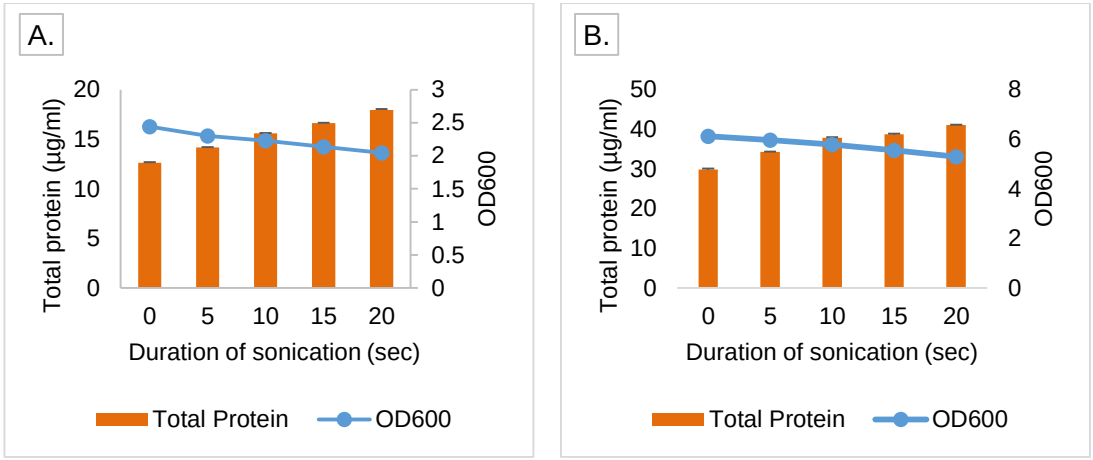



**Figure 5-6: Effects of sonication on cells determined by total protein and cell density**

Cell suspensions were ultrasonicated to induce various degrees of cell damage. The total protein of sonicated cells was determined using the Bradford assay.

The correlation between cell damage or membrane permeability and staining properties of the two stains was investigated by comparing fluorescence measurements obtained from the sonicated cell suspensions. Figure 5-7 demonstrates that both stains give higher fluorescence with higher degrees of cell damage indicating that they bind extracellular readily.

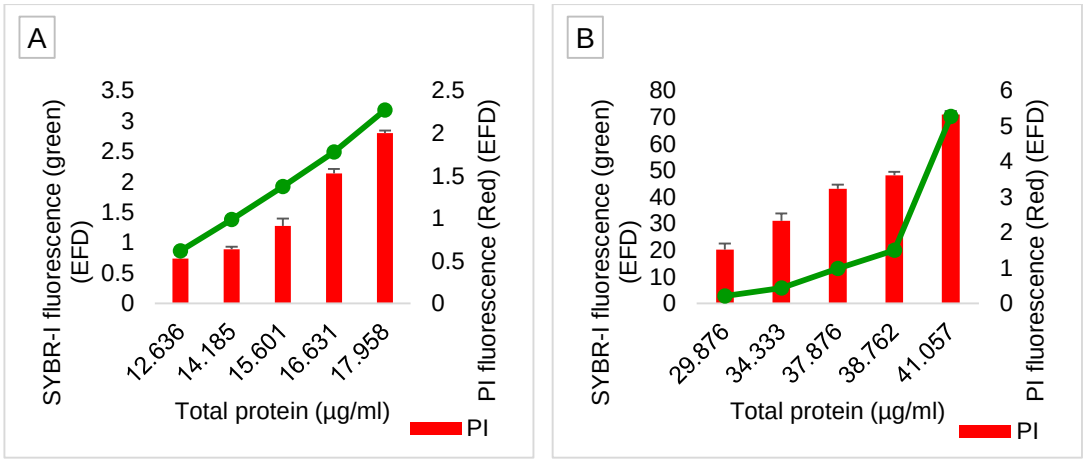



**Figure 5-7: Effects of sonication on staining properties of SYBR-I and PI**

Fluorescence and total protein were determined for cell suspensions sonicated to induce various degrees of membrane damage. Fluorescence of PI as a result of cell damage determined by total protein was compared.

### *5.4.4 Discriminating between live and dead cells by dual staining*

Simultaneous staining of cells with both SYBR-I and PI allows detection and quantification of viable bacteria cells by flow cytometry. To demonstrate this with fluorimetry, cell suspensions of *E. coli* JM109 and *C. freundii* NCIMB11409 were sonicated to induce various degrees of cell damage and stained with both SYBR-I and PI. The extracellular protein in the cell suspension increased with the duration of sonication (representing increasing cell lysis and release of protein). However, when the same samples were totally lysed by incubation at 65°C (1 hour), the total protein was not significantly different from each other (Figures 5-8; a & b).

Dual staining of bacteria cells allows differentiation of live and dead cells, where green fluorescence represents the live cell population stained with SYBR-I and red fluorescence for dead cell population stained with PI. The population of live cells as determined by dual staining decreased with increasing sonication duration whilst the dead cells increased (Figures 5-8; c and d). There was a strong direct correlation between duration of sonication and the dead cell population for both *E. coli* ($R^2$ = 0.99; figure 5-8c) and *C. freundii* ($R^2$ = 0.99; figure 5-8d). Although there was an inverse correlation between the duration of sonication and the live cell population, the correlation was weaker than that of the dead cell population ($R^2$ = 0.87, *E. coli*; $R^2$ = 0.76, *C. freundii*). Furthermore there were strong correlations between the dead cell measurements and the extracellular protein released as a result of sonication for both *E. coli* ($R^2$=0.98) and *C. freundii* ($R^2$=0.99) (Figure 5-8; e & f)

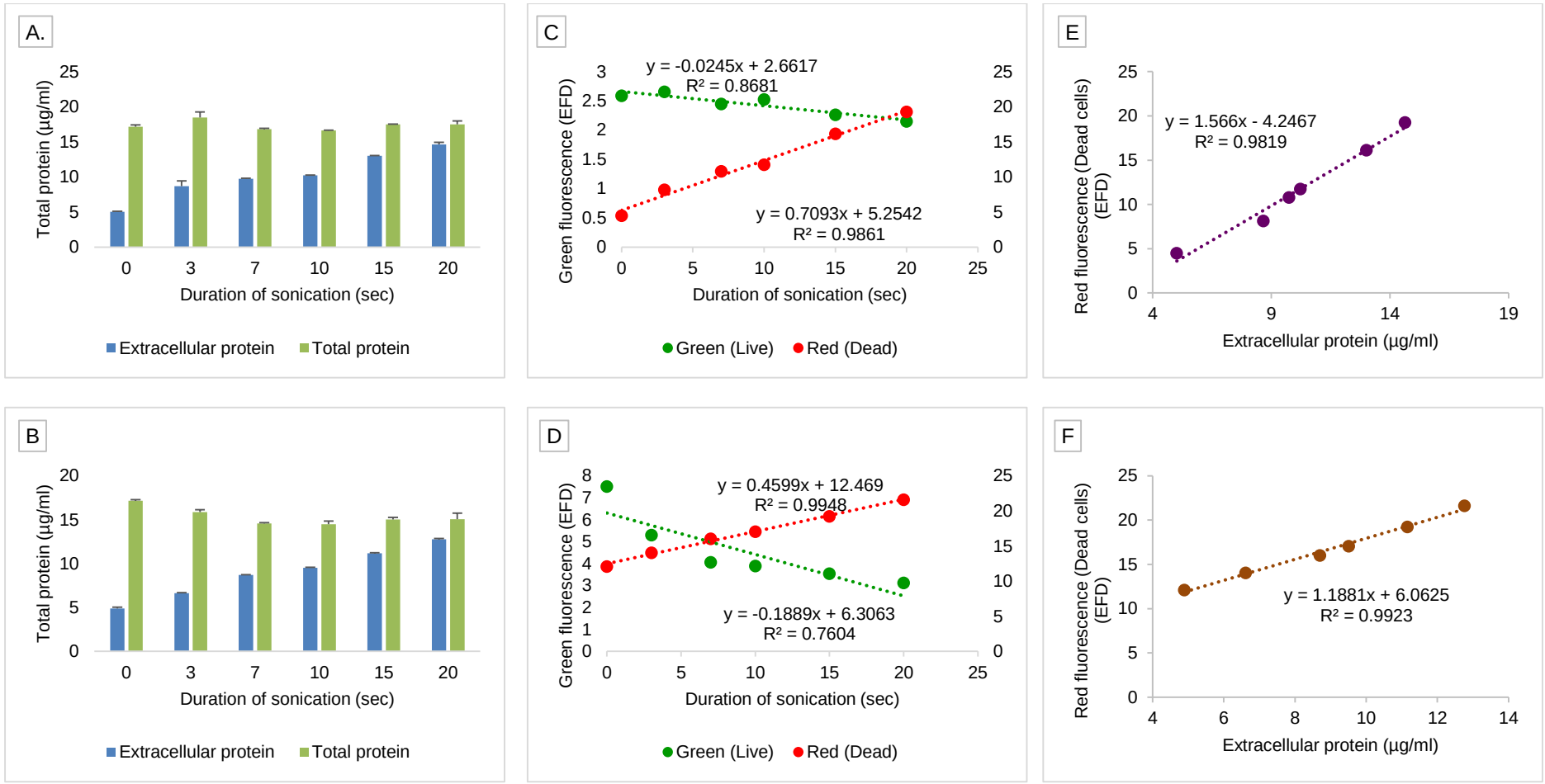



**Figure 5-8: Dual staining with SYBR-I and PI could discriminate live and dead *E. coli* and *C. freundii* cells**

Extracellular protein concentration (a & b) and dead cells (c & d) were estimated from cells that had been sonicated to induce varying degrees of cell damage. Correlations between cell damage and protein concentration as well as the live cells and protein concentration were determined and compared. Plots a, c and e are for *E. coli* whereas plots b, d and f are for *C. freundii*. Three biological replicated performed.

### *5.4.5 Effects of the presence of insoluble cellulose on turbidity of cell suspensions*

To determine what effects addition of an insoluble substance will have on the cell density (OD600), cellulose powder was added to cell suspensions after initial OD600 had been measured. Figures 5-9 and 5-10 show results from two sets of experiments involving two different strains of *E. coli* and *C. freundii.* As expected, the addition of cellulose resulted in high OD600 values. The values however did not correlate with the values obtained before the addition of cellulose. To investigate whether leaving cuvettes to stand for some time to allow avicel to settle will improve the correlation between OD600 without and with avicel, OD600 measurements were taken at 5 and 15 minutes after addition of avicel. The OD600 was did not significantly change with longer standing time although the values for 15 minutes wait were lower than the values for 5 minutes. The decrease was similar for all the samples tested. To determine whether there was a fixed relationship between increases in OD600 with addition of cellulose, the OD600 of blank cuvettes with added cellulose was compared with the cell suspensions with added cellulose. Additionally, increases in the OD600 from cultures with different cell densities were also compared. The results showed original OD600 of the cells could not be deduced reliably when the measurement was made in the presence of avicel. Furthermore, addition of cellulose to suspensions with lower cell densities made them seem comparable to suspensions with cell densities 10 fold higher (Figure 5-10).

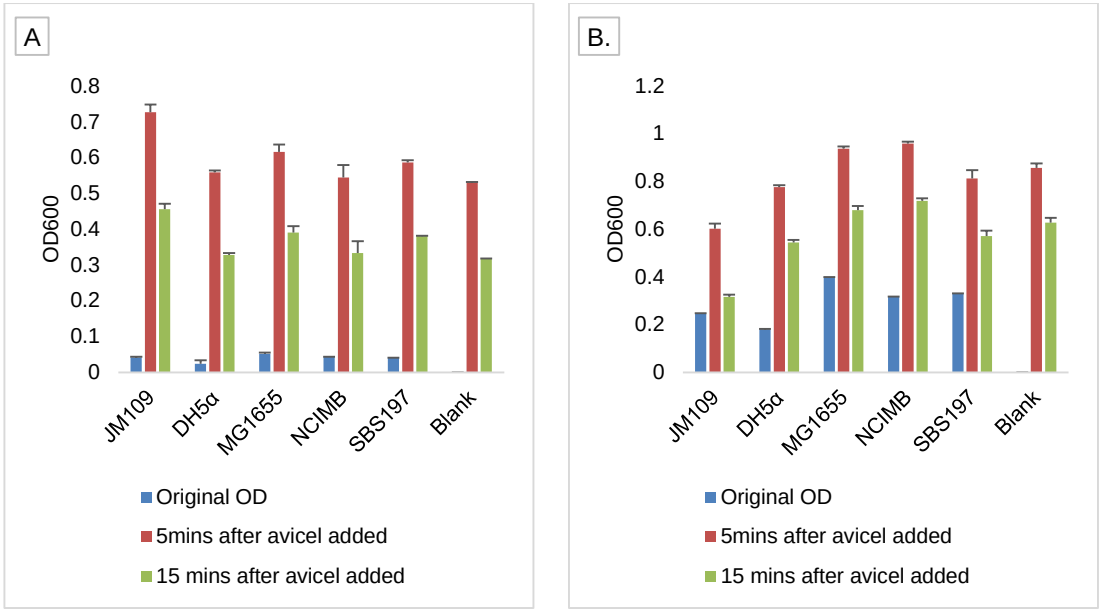



**Figure 5-9: Changes in the turbidity of cell suspensions with the addition of cellulose powder**

Cell densities (OD600) measured in cultures of different strains of *E. coli* and *C. freundii*. The strains of *E. coli* are JM109, DH5*α* and MG1655 whereas those of *C. freundii* are NCIMB11490 and SBS197. 'A' is results from cells with densities less than those in 'B'. Error bars are standard errors of three biological replicates.

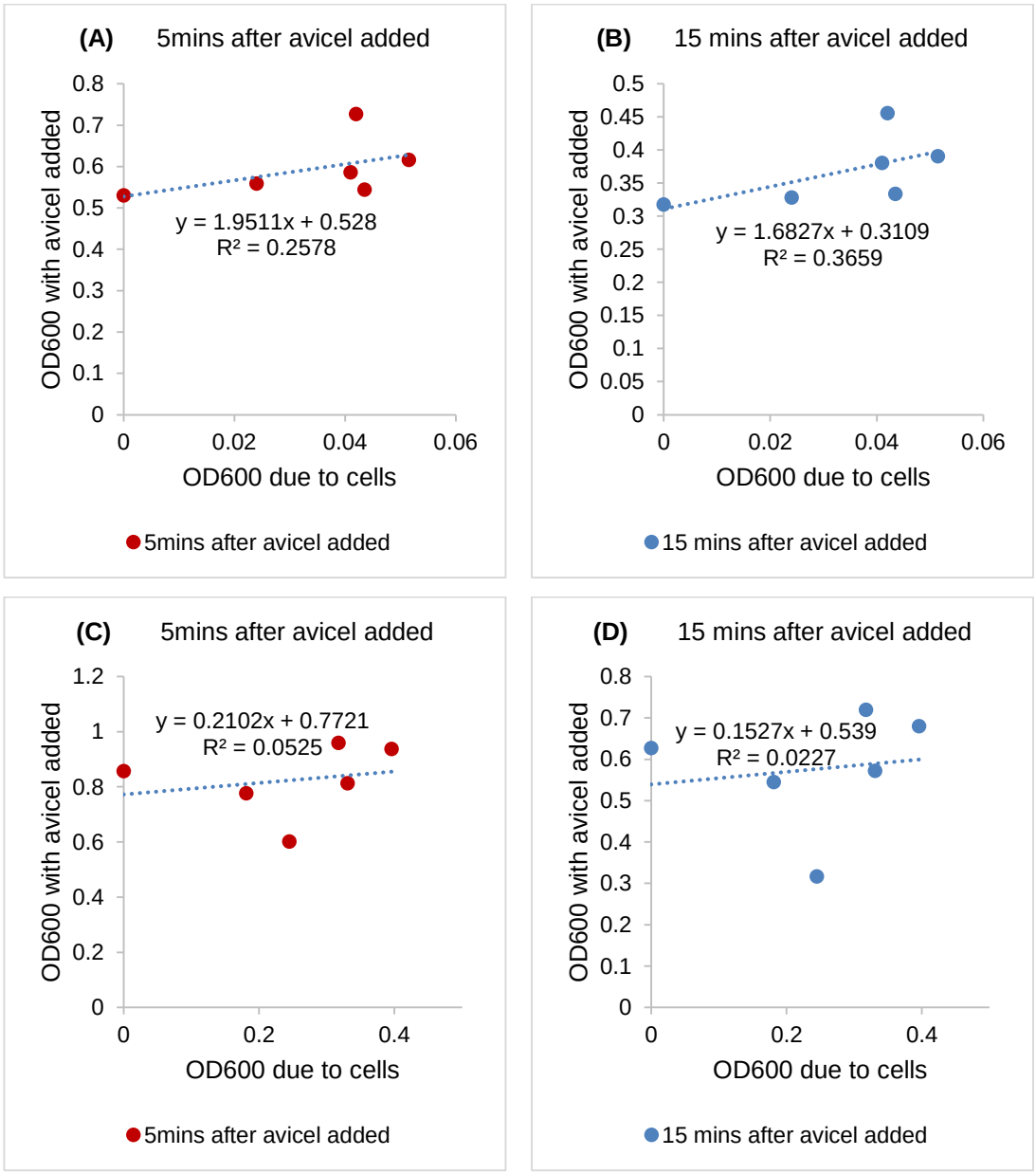



**Figure 5-10: Correlations between the turbidity before and after addition of cellulose**

Turbidity of cell suspensions were determined prior to and after the addition of avicel. Cell suspensions were made to settle and turbidity determined by measuring absorbance at 600 nm. Three biological replicates were performed.

### *5.4.6 Detecting live and dead cells in the presence of cellulose*

To investigate the ability of the SYBR-I and PI dual staining to estimate the amount of cells present in suspensions with cellulose, correlations between OD600 and the fluorescence obtained were determined. *E. coli* and *C. freundii* cell suspensions with or without cellulose were investigated. The addition of avicel did not significantly affect the fluorescence measurements (Figure 5-11).

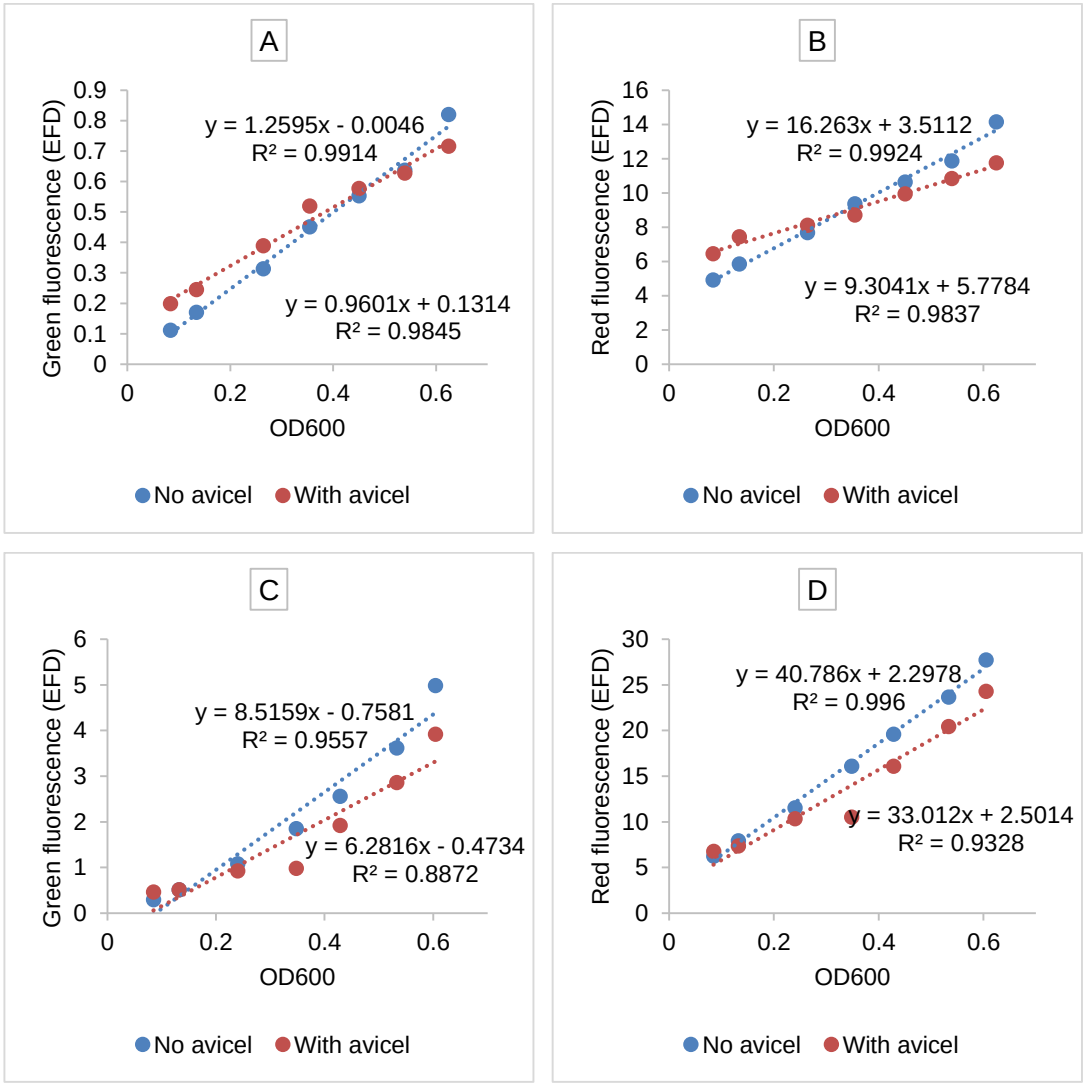



**Figure 5-11: Effects of avicel on the fluorescence measurements**

Fluorescence measurements were obtained for cells prior to and after addition of avicel. Cell suspensions of *E. coli* (a and b) and *C. freundii* (c and d) were used. Three biological repeats were performed.

Pearson's correlation was used to test whether there was any correlation between OD600 before addition of avicel and the live cells (green fluorescence) before and after addition of avicel. Strong correlations were obtained for all cells whether avicel was present or not (Table 5-2). To determine whether the fluorescence measured depended on the OD600, Kendall tau correlation test was performed (Table 5-2). The results showed that irrespective of the type of bacteria used or whether there was avicel or not, fluorescence depended on the OD600 before addition of avicel.

**Table 5-2: Significance of correlations between OD600 and fluorescence in the presence or absence of cellulose**

| | | Pearson Correlation | Kendall's tau_b Correlation |
|---|---|---|---|
| **JM109** | Correlation Co-efficient | 0.995** | 1.000** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| **JM109+avicel** | Correlation Co-efficient | 0.990** | 1.000** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| **NCIMB** | Correlation Co-efficient | 0.978** | 1.000** |
| | Sig. (2-tailed) | 0.000 | 0.000 |
| **NCIMB+avicel** | Correlation Co-efficient | 0.942** | 1.000** |
| | Sig. (2-tailed) | 0.002 | 0.000 |

Correlations were tested for linearity (Pearson) and dependence (Kendall's tau) between the OD600 values and the fluorescence measurements. **Correlation is significant at the 0.01 level (2-tailed).

### *5.4.7 Effects of addition of cellulose on total protein estimation*

The effects of addition of cellulose on total protein was determined as was done for fluorescence above. Total protein data like fluorescence was normally distributed. There was no significant difference between total protein for cells with or without avicel

(p = 0.117 (*E. coli*) and 0.600 (*C. freundii*); figure 5-12). Significant correlations were also observed between OD600 and the total protein of *E. coli* without cellulose (p = 0.001, $R^2$ = 0.960) and with cellulose (p = 0.003, $R^2$ = 0.923) as well as *C. freundii* without cellulose (p < 0.001, $R^2$ = 0.990) and with cellulose (p < 0.001, $R^2$ = 0.984). As to whether total protein depended on OD600 of cells, all were found to strongly depend (tau_b = 1, p <0.001) except in the case of *E. coli* with cellulose for which dependence was weaker (tau_b = 0.810) but significant (p = 0.011).

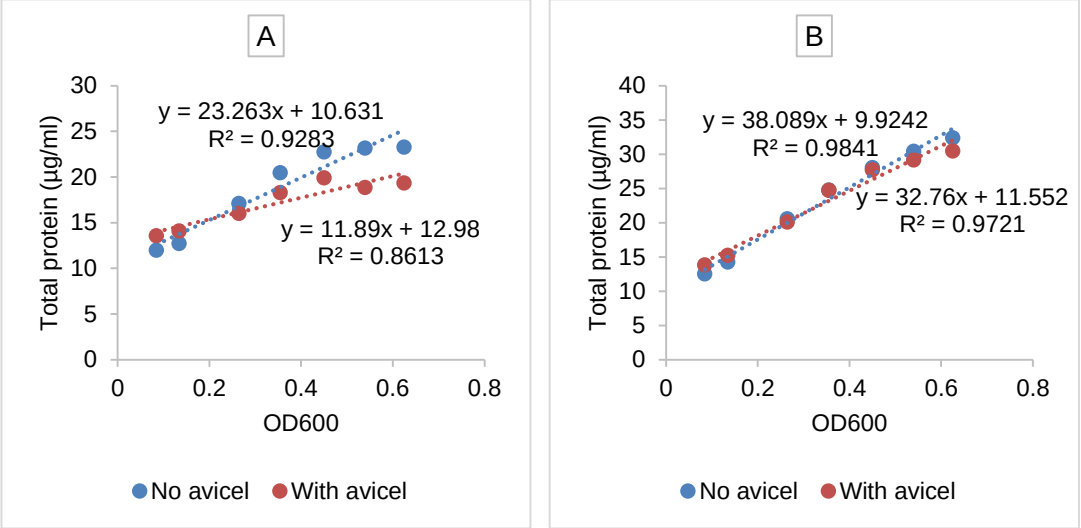



**Figure 5-12: Effects of addition of avicel on total protein**

Total protein were obtained for cell suspensions prior to and after addition of avicel. (a) *E. coli* and (b) *C. freundii* cell suspensions were used. The scatterplots show the original OD600 (i.e. before addition of avicel) versus the total protein with and without avicel. Three biological replicates were performed.

## 5.5 Discussion

There have been tremendous advances in fluorescence cytometry (instruments, fluorophores and methods) over the years which enable direct analysis and quantification of bacteria and other cells in different environments (Melamed *et al.*, 1972; Lebaron *et al.*, 1998; Lenkei *et al.*, 1998b; Gregori *et al.*, 2001; Nunez *et al.*, 2004; Foladori *et al.*, 2010; Tamburini *et al.*, 2014). In this study, it has been

demonstrated that a simple two colour fluorescence fluorimetric technique can effectively be used to monitor growth of bacteria in turbid growth cultures containing insoluble cellulosic substrates. This is essential for assessing *in vivo* activity of cellulose degrading enzyme systems for biofuel production. The total protein assay evaluated in this study was also found to be suitable for this purpose. The fluorescence technique however has some advantages over the total protein technique. Whilst the fluorescence method can be completed in about 20 minutes, the total protein will be completed in not less than two hours. Additionally, the stains used in the fluorescence technique are not regarded as harmful by European Union regulations whereas the protein reagent is harmful. Another advantage of using the SYBR-I/PI method over the total protein is that it allows direct estimation of live and dead cells making it suitable for experiments where cell lysis is suspected. With an appropriate standard curve (eg. plate count or flow cytometry), the green fluorescence values can be converted directly to cell numbers.

Barbesti et al (2000) demonstrated that there is energy transfer from SYBR-I to PI in bacterial cells stained with both fluorochromes. This study has demonstrated that the energy transfer also occurs with DNA in solution but at a slightly higher percentage. This is expected as the fluorochromes readily bind DNA and do not need to cross barriers (eg. membranes) as occurs for cells (Melamed *et al.*, 1972; Lebaron *et al.*, 1998). The strong energy transfer facilitates discrimination of live and dead bacteria and it has been shown that it is not affected by the metabolic state (eg. stationary or exponential) or the type of bacteria (gram positive or negative) (Barbesti *et al.*, 2000). The discriminatory properties of the SYBR-I/PI dual staining have been shown to be very effective, with strong correlations demonstrated between dead cells (red fluorescence) and the degree of induced damage (sonication). This also means that

the fluorimetric application of the SYBR-I/PI method can account not only for cells with compromised cell membranes as happens in flow cytometry but also for completely lysed cells. Flow cytometry will count cells meaning completely lysed cells will not be accounted for. This makes the application of two colour fluorescence staining with fluorimetry applicable in a wider context.

Obtaining conditions necessary for the ideal measurement of fluorescence intensity is in practice very difficult (Gaigalas *et al.*, 2001). The development of the MESF unit is a practical approach that seeks to take the application of fluorescence cytometry from just enumeration to actual quantitation. With this unit, quantitative fluorescence data are no longer dependent on the instrument or the environment within which cells are present (eg. media) but are standardized and comparable with others (Schwartz *et al.*, 2004). Although the unit has not been evaluated on other fluorescence platforms such as fluorimeters, there are some foreseeable challenges. Fluorimeters do not count cells but rather give a value for the total fluorescence obtained. The fluorescence obtained is also dependent on the amount of DNA present which these stains (i.e. SYBR-I and PI) bind to. Thus, a recombinant organism harbouring a large piece of foreign DNA will likely produce higher fluorescence than the wild type. The equivalent fluorescent DNA (EFD) unit developed in this study leverages the principles of the MESF and the foreseeable challenges of its application in fluorimetry to best serve its purpose. The EFD value obtained for a sample can be directly compared to another sample of the same constitution provided the DNA standard and samples were measured on the same instrument. Rather than using a known quantity of standard beads, the EFD can be converted to a quantitation standard when cell quantities are standardized with plate counts or another appropriate measure.

## 5.6 Conclusions

Analysis of bacterial cells stained with SYBR-I and PI using a fluorimeter has been evaluated. The application of this dual fluorescence staining technique in fluorimetry is simple, fast and can be easily adopted for automation or large screening applications. With fluorimetry, dual staining of samples with SYBR-I and PI will detect not only cells with compromised membranes but completely lysed cells as well. This property can be an advantage or disadvantage depending on the application. The use of the EFD as a unit for comparing fluorescence intensity in fluorimetry across laboratories and instrument platforms has also been demonstrated. This standard is cheap and can easily be made, yet is suitable for comparison with other samples or applications.

## 5.7 Limitations

Despite the advantages highlighted, the method does have some limitations. On its own, dual staining followed by fluorescence measurement does not give absolute quantification of cells. A standard curve must be generated, for example from plate counts or other suitable method, and used to estimate the number of cells from the fluorescence reading. Furthermore, the method is not suitable for directly comparing quantitation obtained from different bacteria, due to its dependence on the amount of DNA present and inability to separate brightly and dimly fluorescing cells.

# Chapter 6

## Conclusions and further work

### 6.1 Summary of key contributions

The abundance and sustainable characteristics of lignocellulosic biomass makes them an ideal choice for biofuel production. As highlighted in various sections of this thesis, biofuels from lignocellulosic biomass is yet to become a key part of the fuels we use today due to challenges in its processing. This work focused on one aspect of the biomass to biofuel processing chain; cellulose degradation to sugars. It has been demonstrated that engineering multicatalytic enzymes could potentially substitute cloning of multiple enzymes. By investigating this further, it was found that making fusions does not automatically mean those fusions will be better than the non-fused versions. Although multicatalytic enzymes have shown promise, careful selection of enzymes is required. There is a need to design and screen various combinations of cellulases in order to choose the best fusion partners. Here it has been demonstrated that a large library of fusion proteins can be made using the MABEL method. The process is efficient and fast and can be employed to make fusions, deletions and substitutions without any expensive commercial kit.

Although some of the fusion proteins show promise, the data generated in these studies cannot be solely used to draw firm conclusions. As discussed earlier, it is not known whether the ‘success’ or ‘failure’ of the fusion proteins is due to improper folding or instability as a result of pH or temperature changes in the medium.

Furthermore, it is also not known whether expressing the fusion or non-fused proteins overburden the cells. These factors should be investigated in order to guide further research on how to improve upon these fusion proteins. Experiments to monitor the burden of the expression of the heterologous proteins on cells could be performed as reported (Ceroni *et al.*, 2015).

The combinatorial approach to engineering cellulose degradation systems must go hand in hand with high throughput screening methods. Plate counting provides a direct way of monitoring cellulose utilization which is a primary requirement for further engineering to produce high value chemicals. As discussed earlier, this process is laborious and not ideal for combinatorial engineering, hence, the development and characterization of the SYBR-I/PI dual staining method. This study has demonstrated that the two colour fluorescence staining could be applied to fluorimetry, for which instrumentation is cheaper than flow cytometry. Furthermore, the development of a fluorescence intensity standard, the equivalent fluorescent DNA to go with fluorimetry provides a cheaper way of standardizing fluorescent measurements for easy comparison across different experiments, laboratories and instrument platforms.

## 6.2 Future work

The most critical experiments will be to purify the enzymes, measure their specific activities compared to unfused versions, and use this information to determine the absolute expression and secretion levels of each fused and unfused enzyme. This would give strong clues as to the reasons for improved or worsened performance in each case.

Other studies could focus on determining the crystal structures as well as the stability of the expressed proteins (fusion compared to native enzymes) in the various types of media used. Effects of pH, temperature, age of culture as well as cell lysis should be investigated. Additionally, since secretion helps in getting the expressed enzymes to substrates, investigations should be performed to determine whether the proteins are secreted well or not. Other studies on the morphology of cells expressing the recombinant enzymes in the presence of cellulosic substrates and glucose should also be performed. It will be interesting to know whether cells expressing the recombinant proteins adopt new characteristics like clustering together or forming protrusions or extensions. As discussed above some cellulosic organisms without secretory systems form hyphal extensions to help get the cellulases to the substrates. Though not a general rule this information together with that on secretion will be useful in drawing strong conclusions for both the fusion proteins and the *C. freundii* as a host for expressing heterologous cellulases. These studies will ultimately help to understand why the fusion or non-fused proteins performs better or otherwise.

It is also recommended that the fusion proteins be tested in other hosts (e.g. *E. coli* and *Saccharomyces cerevisiae*). For example it will be interesting to know whether the fusion of endoglucanase-exoglucanase-β-glucosidase will be better in a host that does not have native ability to assimilate cellobiose (e.g. *E. coli*) other than in *C. freundii*. Results from these studies will help understand whether the host plays a role in determining how the fusion protein will perform.

The fluorescence method developed in this work needs to be further evaluated to determine its applicability to other cellulosic substrates such as pre-treated biomass. Further investigation into the possibility of adopting the methods for smaller volumes (e.g. in 96 well microplates) should also be investigated.

# References


Albersheim, P., Bauer, W. D., Keestra, K. & Talmadge, K. W. 1973. The structure of the wall of suspension-cultured sycamore cells. *In:* Loewus, F. (ed.) *Biogenesis of Plant Cell Wall Polysaccharides.* Academic Press.

An, J. M., Kim, Y. K., Lim, W. J., Hong, S. Y., An, C. L., Shin, E. C., Cho, K. M., Choi, B. R., Kang, J. M., Lee, S. M., et al. 2005. Evaluation of a novel bifunctional xylanase–cellulase constructed by gene fusion. *Enzyme Microb Technol,* 36**,** 989-995.

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., et al. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science,* 344**,** 55-8.

Antoni, D., Zverlov, V. V. & Schwarz, W. H. 2007. Biofuels from microbes. *Appl Microbiol Biotechnol,* 77**,** 23-35.

Arkin, A. 2008. Setting the standard in synthetic biology. *Nat Biotechnol,* 26**,** 771-4.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics,* 22**,** 195-201.

Barbesti, S., Citterio, S., Labra, M., Baroni, M. D., Neri, M. G. & Sgorbati, S. 2000. Two and three-color fluorescence flow cytometric analysis of immunoidentified viable bacteria. *Cytometry,* 40**,** 214-8.

Barnard, D. K. 2012. Design and Construction of Modular Genetic Devices encoding the Enzymatic Hydrolysis of Lignocellulosic Biomass.*[PhD Thesis], University of Edinburgh.*

Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol,* 58**,** 521-54.

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. 1998a. Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol,* 8**,** 548-57.

Bayer, E. A., Lamed, R. & Himmel, M. E. 2007. The potential of cellulases and cellulosomes for cellulosic waste management. *Current Opinion in Biotechnology,* 18**,** 237-245.

Bayer, E. A., Shimon, L. J., Shoham, Y. & Lamed, R. 1998b. Cellulosomes-structure and ultrastructure. *J Struct Biol,* 124**,** 221-34.

Beguin, P. & Aubert, J. P. 1994. The biological degradation of cellulose. *FEMS Microbiol Rev,* 13**,** 25-58.

Beguin, P., Eisen, H. & Roupas, A. 1977. Free and cellulose-bound cellulases in a *Cellulomonas* species. *J Gen Microbiol,* 101**,** 191-196.

Beguin, P., Millet, J. & Aubert, J. P. 1992. Cellulose degradation by *Clostridium thermocellum*: from manure to molecular biology. *FEMS Microbiol Lett,* 79**,** 523-8.

Bertani, G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol,* 186**,** 595-600.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res,* 42**,** W252-8.

Boisset, C., Fraschini, C., Schulein, M., Henrissat, B. & Chanzy, H. 2000. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Appl Environ Microbiol,* 66**,** 1444-52.

Bolshakova, E. V., Ponomariev, A. A., Novikov, A. A., Svetlichnyi, V. A. & Velikodvorskaya, G. A. 1994. Cloning and expression of genes coding for carbohydrate degrading enzymes of *Anaerocellum thermophilum* in *E.coli*. *Biochem Biophys Res Commun,* 202**,** 1076-80.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc,* 4**,** 1-13.

Bower, B., Larenas, E. & Mitchinson, C. 2006. Exo-endo cellulase fusion protein. Google Patents.

Breed, R. S. & Dotterrer, W. D. 1916. The number of colonies allowable on satisfactory agar plates. *J Bacteriol,* 1**,** 321-31.

Brillouet, J. M. & Mercier, C. 1981. Fractionation of wheat bran carbohydrates. *Journal of the Science of Food and Agriculture,* 32**,** 243-251.

Bulow, L. 1987. Characterization of an artificial bifunctional enzyme, beta-galactosidase/galactokinase, prepared by gene fusion. *Eur J Biochem,* 163**,** 443-8.

Bulow, L. & Mosbach, K. 1991. Multienzyme systems obtained by gene fusion. *Trends Biotechnol,* 9**,** 226-31.

C. K. Liu, Salinas-Vaccaro, A., Banard, D. K., Lakhundi, S., Cain, N. & French, C. E. *In preparation*. CHU2268 may be the 'missing' exoglucanase of *Cytophaga hutchinsonii*.

Caffall, K. H. & Mohnen, D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res,* 344**,** 1879-900.

Caron, G. N.-V. & Badley, R. A. 1995. Viability assessment of bacteria in mixed populations using flow cytometry. *Journal of Microscopy,* 179**,** 55-66.

Carroll, A. & Somerville, C. 2009. Cellulosic biofuels. *Annu Rev Plant Biol,* 60**,** 165-82.

Casini, A., MacDonald, J. T., De Jonghe, J., Christodoulou, G., Freemont, P. S., Baldwin, G. S. & Ellis, T. 2014. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res,* 42**,** e7.

Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. 2015. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Meth,* 12**,** 415-418.

Charles, I. G., Keyte, J. W., Brammar, W. J., Smith, M. & Hawkins, A. R. 1986. The isolation and nucleotide sequence of the complex AROM locus of *Aspergillus nidulans*. *Nucleic Acids Res,* 14**,** 2201-13.

Chaudhary, P., Kumar, N. N. & Deobagkar, D. N. 1997. The glucanases of *Cellulomonas*. *Biotechnology Advances,* 15**,** 315-331.

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

Chung, C. T., Niemela, S. L. & Miller, R. H. 1989. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A,* 86**,** 2172-5.

Cold Spring Harb. 2010. Recipe: M9 minimal medium (modified). *Cold Spring Harb Protoc.,* doi:10.1101/pdb.rec12296.

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D. & Rosser, S. J. 2014. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res,* 42**,** e23.

Conrado, R. J., Varner, J. D. & DeLisa, M. P. 2008. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr Opin Biotechnol,* 19**,** 492-9.

Cosgrove, D. J. 2005. Growth of the plant cell wall. *Nat Rev Mol Cell Biol,* 6**,** 850-61.

Coutinho, J. B., Moser, B., Kilburn, D. G., Warren, R. A. & Miller, R. C., Jr. 1991. Nucleotide sequence of the endoglucanase C gene (cenC) of *Cellulomonas fimi*, its high-level expression in *Escherichia coli*, and characterization of its products. *Mol Microbiol,* 5**,** 1221-33.

Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. & Haynes, C. A. 1996. Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc Natl Acad Sci U S A,* 93**,** 12229-34.

Creutzig, F., Ravindranath, N. H., Berndes, G., Bolwig, S., Bright, R., Cherubini, F., Chum, H., Corbera, E., Delucchi, M., Faaij, A., et al. 2015. Bioenergy and climate change mitigation: an assessment. *GCB Bioenergy,* 7**,** 916-944.

Dale, B. 2008. Biofuels: thinking clearly about the issues. *J Agric Food Chem,* 56**,** 3885-91.

Damude, H. G., Ferro, V., Withers, S. G. & Warren, R. A. 1996. Substrate specificity of endoglucanase A from *Cellulomonas fimi*: fundamental differences between endoglucanases and exoglucanases from family 6. *Biochem J,* 315 ( Pt 2)**,** 467-72.

Diaper, J. P., Tither, K. & Edwards, C. 1992. Rapid assessment of bacterial viability by flow cytometry. *Applied Microbiology and Biotechnology,* 38**,** 268-272.

Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. 1994. C1-Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci U S A,* 91**,** 11383-7.

Ding, S. Y. & Himmel, M. E. 2006. The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry,* 54**,** 597-606.

Doi, R. H. 2008. Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers. *Ann N Y Acad Sci,* 1125**,** 267-79.

Du, R., Yan, J., Li, S., Zhang, L., Zhang, S., Li, J., Zhao, G. & Qi, P. 2015. Cellulosic ethanol production by natural bacterial consortia is enhanced by *Pseudoxanthomonas taiwanensis*. *Biotechnol Biofuels,* 8**,** 10.

Düsterhöft, E.-M., Voragen, A. G. J. & Engels, F. M. 1991. Non-starch polysaccharides from sunflower (*Helianthus annuus*) meal and palm kernel (*Elaeis guineenis*) meal—preparation of cell wall material and extraction of

polysaccharide fractions. *Journal of the Science of Food and Agriculture,* 55**,** 411-422.

Elleuche, S. 2015. Bringing functions together with fusion enzymes - from nature's inventions to biotechnological applications. *Appl Microbiol Biotechnol,* 99**,** 1545-56.

Ellis, T., Adie, T. & Baldwin, G. S. 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol (Camb),* 3**,** 109-18.

Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One,* 4**,** e5553.

Foladori, P., Bruni, L., Tamburini, S. & Ziglio, G. 2010. Direct quantification of bacterial biomass in influent, effluent and activated sludge of wastewater treatment plants by using flow cytometry. *Water Research,* 44**,** 3807-3818.

French, C. E. 2009. Synthetic biology and biomass conversion: a match made in heaven? *J R Soc Interface,* 6 Suppl 4**,** S547-58.

French, C. E., Barnard, D. K., Fletcher, E., Kane, S. D., Lakhundi, S. S., Liu, C.-K. & Elfick, A. 2013. Synthetic Biology for Biomass Conversion. *In:* Suib, S. L. (ed.) *New and Future Developments in Catalysis.* Amsterdam: Elsevier.

French, C. E., Horsfall, L., Barnard, D. K., Duedu, K., Fletcher, E., Joshi, N., Kane, S. D., Lakhundi, S. S., Liu, C.-K., Oltmanns, J., et al. 2015. Beyond Genetic Engineering: Technical Capabilities in the Application Fields of Biocatalysis and Biosensors. *In:* Giese, B., Pade, C., Wigger, H. & von Gleich, A. (eds.) *Synthetic Biology.* Springer International Publishing.

Fry, S. C. 1988. *The growing plant cell wall: chemical and metabolic analysis,* Harlow, Essex, England, New York, Longman Scientific & Technical; Wiley.

Fulton, L. M., Lynd, L. R., Körner, A., Greene, N. & Tonachel, L. R. 2015. The need for biofuels as part of a low carbon energy future. *Biofuels, Bioproducts and Biorefining,* 9**,** 476-483.

Gaigalas, A. K., Li, L., Marti, G., Henderson, O., Vogt, R. & Barr, J. 2001. The development of fluorescence intensity standards. *J Res Natl Inst Stand Technol,* 106**,** 381-389.

Gasiūnaitė, L., Lewicka, A., Pashkuleva, H., Villanueva, H., Thornton, H., Trubitsyna, M. & French, C. 2013. GenBrick – a rapid multi-part assembly method for bioBricks. *http://dspace.mit.edu/handle/1721.1/81331*.

Gibbs, M. D., Saul, D. J., Luthi, E. & Bergquist, P. L. 1992. The beta-mannanase from "*Caldocellum saccharolyticum*" is part of a multidomain enzyme. *Appl Environ Microbiol,* 58**,** 3864-7.

Gibeaut, D. M. & Carpita, N. C. 1994. Biosynthesis of plant cell wall polysaccharides. *FASEB J,* 8**,** 904-15.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd & Smith, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods,* 6**,** 343-5.

Gilkes, N. R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H. D., Kilburn, D. G., Warren, R. A. & Miller, R. C., Jr. 1991a. Structural and functional relationships in two families of beta-1,4-glycanases. *Eur J Biochem,* 202**,** 367-77.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1991b. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev,* 55**,** 303-15.

Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1989. Structural and functional analysis of a bacterial cellulase by proteolysis. *J Biol Chem,* 264**,** 17802-8.

Gomez, L. D., Steele-King, C. G. & McQueen-Mason, S. J. 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist,* 178**,** 473-485.

Grabber, J., H, , Mertens, D., R, , Kim, H., Funk, C., Lu, F. & Ralph, J. 2008. Cell wall fermentation kinetics are impacted more by lignin content and ferulate cross-linking than by lignin composition. *J Sci Food Agric,* 89**,** 122–9.

Granda, C. B., Zhu, L. & Holtzapple, M. T. 2007. Sustainable liquid biofuels and their environmental impact. *Environmental Progress,* 26**,** 233-250.

Gratama, J. W., D'Hautcourt, J.-L., Mandy, F., Rothe, G., Barnett, D., Janossy, G., Papa, S., Schmitz, G., Lenkei, R. & The European Working Group on Clinical Cell, A. 1998. Flow cytometric quantitation of immunofluorescence intensity: Problems and perspectives. *Cytometry,* 33**,** 166-178.

Greenberg, N. M., Warren, R. A., Kilburn, D. G. & Miller, R. C., Jr. 1987a. Regulation and initiation of cenB transcripts of *Cellulomonas fimi*. *J Bacteriol,* 169**,** 4674-7.

Greenberg, N. M., Warren, R. A., Kilburn, D. G. & Miller, R. C., Jr. 1987b. Regulation, initiation, and termination of the *cenA* and *cex* transcripts of *Cellulomonas fimi*. *J Bacteriol,* 169**,** 646-53.

Gregori, G., Citterio, S., Ghiani, A., Labra, M., Sgorbati, S., Brown, S. & Denis, M. 2001. Resolution of viable and membrane-compromised bacteria in freshwater and marine waters based on analytical flow cytometry and nucleic acid double staining. *Appl Environ Microbiol,* 67**,** 4662-70.

Guo, Z., Arfman, N., Ong, E., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. & Miller, R. C. 1988. Leakage of *Cellulomonas fimi* cellulases from *Escherichia coli*. *FEMS Microbiology Letters,* 49**,** 279-283.

Gupta, P., Samant, K. & Sahu, A. 2012. Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential. *Int J Microbiol,* 2012**,** 578925.

Hall, M., Rubin, J., Behrens, S. H. & Bommarius, A. S. 2011. The cellulose-binding domain of cellobiohydrolase Cel7A from *Trichoderma reesei* is also a thermostabilizing domain. *J Biotechnol,* 155**,** 370-6.

Halpin, C. 2013, personal communication. Controlling lignin biosynthesis in barley. 16th November, 2013**,** Edinburgh, UK.

Harris, D., Bulone, V., Ding, S. Y. & DeBolt, S. 2010. Tools for cellulose analysis in plant cell walls. *Plant Physiology,* 153**,** 420-426.

Haworth, W. N. 1932. Molecular structure of cellulose and of amylose. *Nature,* 129**,** 365.

Heaton, E. A., Dohleman, F. G. & Long, S. P. 2008. Meeting US biofuel goals with less land: the potential of Miscanthus. *Global Change Biology,* 14**,** 2000-2014.

Hendriks, A. T. W. M. & Zeeman, G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology,* 100**,** 10-18.

Henrissat, B. & Davies, G. J. 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol,* 124**,** 1515-9.

Henrissat, B., Teeri, T. T. & Warren, R. A. 1998. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett,* 425**,** 352-4.

Heredia, A., Jiménez, A. & Guillén, R. 1995. Composition of plant cell walls. *Zeitschrift für Lebensmittel-Untersuchung und Forschung,* 200**,** 24-31.

Hill, J., Nelson, E., Tilman, D., Polasky, S. & Tiffany, D. 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 11206-11210.

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. & Foust, T. D. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science,* 315**,** 804-807.

Himmel, M. E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R. & Bayer, E. A. 2010. Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels,* 1**,** 323-341.

Hon, D. N. S. 1994. Cellulose: a random walk along its historical path. *Cellulose,* 1**,** 1-25.

Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. 2012. Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels,* 5**,** 45.

Humphreys, M. J., Allman, R. & Lloyd, D. 1994. Determination of the viability of trichomonas vaginalis using flow cytometry. *Cytometry,* 15**,** 343-348.

Hutchinson, H. B. & Clayton, J. 1919. On the decomposition of cellulose by an aerobic organism ( Spirochaeta cytophaga, n. sp.). *The Journal of Agricultural Science,* 9**,** 143-172.

Iiyama, K., Lam, T. & Stone, B. A. 1994. Covalent cross-links in the cell wall. *Plant Physiol,* 104**,** 315-320.

Jervis, E. J., Haynes, C. A. & Kilburn, D. G. 1997a. Surface diffusion of cellulases and their isolated binding domains on cellulose. *J Biol Chem,* 272**,** 24016-23.

Jervis, E. J., Haynes, C. A. & Kilburn, D. G. 1997b. Surface diffusion of cellulases and their isolated binding domains on cellulose. *Journal of Biological Chemistry,* 272**,** 24016-23.

Kahl, L. J. & Endy, D. 2013. A survey of enabling technologies in synthetic biology. *J Biol Eng,* 7**,** 13.

Kaprelyants, A. S. & Kell, D. B. 1992. Rapid assessment of bacterial viability and vitality by rhodamine 123 and flow cytometry. *Journal of Applied Bacteriology,* 72**,** 410-422.

Keegstra, K. 2010. Plant cell walls. *Plant Physiol,* 154**,** 483-6.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc,* 10**,** 845-58.

Kelley, L. A. & Sternberg, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

Kenyon, W. J., Esch, S. W. & Buller, C. S. 2005. The curdlan-type exopolysaccharide produced by *Cellulomonas flavigena* KU forms part of an extracellular glycocalyx involved in cellulose degradation. *Antonie Van Leeuwenhoek,* 87**,** 143-8.

Khandeparker, R. & Numan, M. T. 2008. Bifunctional xylanases and their potential use in biotechnology. *J Ind Microbiol Biotechnol,* 35**,** 635-44.

Kim, H. K. & Pack, M. Y. 1989. Cloning and expression of *Cellulomonas fimi* β-glucosidase genes in *Escherichia coli*. *Enzyme Microb Technol,* 11**,** 313-316.

Kleman-Leyer, K. M., Gilkes, N. R., Miller, R. C., Jr. & Kirk, T. K. 1994. Changes in the molecular-size distribution of insoluble celluloses by the action of recombinant *Cellulomonas fimi* cellulases. *Biochem J,* 302 ( Pt 2)**,** 463-9.

Knight, T. 2003. Idempotent vector design for standard assembly of BioBricks. Available: http://hdl.handle.net/1721.1/21168.

Knothe, G. 2001. Historical perspectives on vegetable oil-based diesel fuels. *Inform,* 12**,** 1103-1107.

Knowles, J. K. C., Lehtovaara, P. & Teeri, T. T. 1987. Cellulase families and their genes. *TIBTECH,* 5**,** 255–261.

Koch, A. L. 1970. Turbidity measurements of bacterial cultures in some available commercial instruments. *Analytical Biochemistry,* 38**,** 252-259.

Koller, A., O'Neill, M. A., Darvill, A. G. & Albersheim, P. 1991. A comparison of the polysaccharides extracted from dried and non-dried walls of suspension-cultured sycamore cells. *Phytochemistry,* 30**,** 3903-3908.

Kumar, P., Barrett, D. M., Delwiche, M. J. & Stroeve, P. 2009. Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Industrial & Engineering Chemistry Research,* 48**,** 3713-3729.

Lakhundi, S. 2011. A synthetic biology approach to cellulose degradation.*[PhD Thesis], University of Edinburgh.*

Lamed, R., Naimark, J., Morgenstern, E. & Bayer, E. A. 1987. Specialized cell surface structures in cellulolytic bacteria. *J Bacteriol,* 169**,** 3792-800.

Langsford, M. L., Gilkes, N. R., Wakarchuk, W. W., Kilburn, D. G., Miller Jr, R. C. & Warren, R. A. J. 1984. The cellulase system of *Cellulomonas fimi J Gen Microbiol,* 130**,** 1367-1376.

Le Nours, J., Anderson, L., Stoll, D., Stalbrand, H. & Lo Leggio, L. 2005. The structure and characterization of a modular endo-β-1,4-mannanase from *Cellulomonas fimi*. *Biochemistry,* 44**,** 12700-8.

Lebaron, P., Parthuisot, N. & Catala, P. 1998. Comparison of blue nucleic acid dyes for flow cytometric enumeration of bacteria in aquatic systems. *Appl Environ Microbiol,* 64**,** 1725-30.

Lenkei, R., Gratama, J. W., Rothe, G., Schmitz, G., D'Hautcourt, J. L., Årekrans, A., Mandy, F. & Marti, G. 1998a. Performance of calibration standards for antigen quantitation with flow cytometry. *Cytometry,* 33**,** 188-196.

Lenkei, R., Mandy, F., Marti, G. & Vogt, R. (eds.) 1998b. *Quantitative fluorescence cytometry: an emerging consensus*: John Wiley & Sons, Inc.

Lerouxel, O., Cavalier, D. M., Liepman, A. H. & Keegstra, K. 2006. Biosynthesis of plant cell wall polysaccharides - a complex process. *Curr Opin Plant Biol,* 9**,** 621-30.

Leschine, S. B. 1995. Cellulose degradation in anaerobic environments. *Annu Rev Microbiol,* 49**,** 399-426.

Li, M. V., Shukla, D., Rhodes, B. H., Lall, A., Shu, J., Moriarity, B. S. & Largaespada, D. A. 2014. HomeRun vector assembly system: a flexible and standardized cloning system for assembly of multi-modular DNA constructs. *PLoS One,* 9**,** e100948.

Li, M. Z. & Elledge, S. J. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods,* 4**,** 251-6.

Linder, M. & Teeri, T. T. 1996. The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. *Proc Natl Acad Sci U S A,* 93**,** 12251-5.

Liu, C. K. 2012. The cellulose degradation system of *Cytophaga hutchinsonii.[PhD Thesis], University of Edinburgh.*

Ljungcrantz, P., Carlsson, H., Mansson, M. O., Buckel, P., Mosbach, K. & Bulow, L. 1989. Construction of an artificial bifunctional enzyme, beta-

galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry,* 28**,** 8786-92.

Louime, C., Abazinge, M., Johnson, E., Latinwo, L., Ikediobi, C. & Clark, A. M. 2007. Molecular cloning and biochemical characterization of a family-9 endoglucanase with an unusual structure from the gliding bacteria *Cytophaga hutchinsonii*. *Appl Biochem Biotechnol,* 141**,** 127-38.

Lu, P., Feng, M. G., Li, W. F. & Hu, C. X. 2006. Construction and characterization of a bifunctional fusion enzyme of *Bacillus*-sourced beta-glucanase and xylanase expressed in Escherichia coli. *FEMS Microbiol Lett,* 261**,** 224-30.

Lynd, L. R., van Zyl, W. H., McBride, J. E. & Laser, M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol,* 16**,** 577-83.

Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev,* 66**,** 506-77, table of contents.

MacLeod, A. M., Lindhorst, T., Withers, S. G. & Warren, R. A. 1994. The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry,* 33**,** 6371-6.

Mayer, C., Vocadlo, D. J., Mah, M., Rupitz, K., Stoll, D., Warren, R. A. & Withers, S. G. 2006. Characterization of a beta-N-acetylhexosaminidase and a beta-N-acetylglucosaminidase/beta-glucosidase from *Cellulomonas fimi*. *FEBS J,* 273**,** 2929-41.

McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A. & Kilburn, D. G. 2000. Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng,* 13**,** 801-9.

Meinke, A., Braun, C., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1991a. Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. *J Bacteriol,* 173**,** 308-14.

Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1991b. Multiple domains in endoglucanase B (CenB) from *Cellulomonas fimi*: functions and relatedness to domains in other polypeptides. *J Bacteriol,* 173**,** 7126-35.

Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1993. Cellulose-binding polypeptides from *Cellulomonas fimi*: endoglucanase D (CenD), a family A beta-1,4-glucanase. *J Bacteriol,* 175**,** 1910-8.

Meinke, A., Gilkes, N. R., Kwan, E., Kilburn, D. G., Warren, R. A. & Miller, R. C., Jr. 1994. Cellobiohydrolase A (CbhA) from the cellulolytic bacterium *Cellulomonas fimi* is a beta-1,4-exocellobiohydrolase analogous to *Trichoderma reesei* CBH II. *Mol Microbiol,* 12**,** 413-22.

Melamed, M. R., Adams, L. R., Zimring, A., Murnick, J. G. & Mayer, K. 1972. Preliminary evaluation of acridine orange as a vital stain for automatic differential leukocyte counts. *Am J Clin Pathol,* 57**,** 95-102.

Murashima, K., Kosugi, A. & Doi, R. H. 2003. Solubilization of cellulosomal cellulases by fusion with cellulose-binding domain of noncellulosomal cellulase engd from *Clostridium cellulovorans*. *Proteins,* 50**,** 620-8.

Nixon, A. E., Warren, M. S. & Benkovic, S. J. 1997. Assembly of an active enzyme by the linkage of two protein modules. *Proc Natl Acad Sci U S A,* 94**,** 1069-73.

Notenboom, V., Birsan, C., Warren, R. A., Withers, S. G. & Rose, D. R. 1998. Exploring the cellulose/xylan specificity of the β-1,4-glycanase cex from *Cellulomonas fimi* through crystallography and mutation. *Biochemistry,* 37**,** 4751-8.

Notenboom, V., Williams, S. J., Hoos, R., Withers, S. G. & Rose, D. R. 2000. Detailed structural analysis of glycosidase/inhibitor interactions: complexes of Cex from *Cellulomonas fimi* with xylobiose-derived aza-sugars. *Biochemistry,* 39**,** 11553-63.

Nuffield Council on Bioethics 2011. Biofuels: ethical issues. Abingdon: Nuffield Council on Bioethics.

Nunez, R., Garay, N., Villafane, C., Bruno, A. & Lindgren, V. 2004. Description of a flow cytometry approach based on SYBR-14 staining for the assessment of

DNA content, cell cycle analysis, and sorting of living normal and neoplastic cells. *Exp Mol Pathol,* 76**,** 29-36.

O'Neill, G., Goh, S. H., Warren, R. A., Kilburn, D. G. & Miller, R. C., Jr. 1986a. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. *Gene,* 44**,** 325-30.

O'Neill, G. P., Warren, R. A., Kilburn, D. G. & Miller, R. C., Jr. 1986b. Secretion of *Cellulomonas fimi* exoglucanase by *Escherichia coli*. *Gene,* 44**,** 331-6.

Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. 1993a. The cellulose-binding domain (CBD(Cex)) of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. *Biotechnol Bioeng,* 42**,** 401-9.

Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. 1993b. The cellulose-binding domain ($CBD_{Cex}$) of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. *Biotechnol Bioeng,* 42**,** 401-9.

Ostermeier, M. & Benkovic, S. J. 2000. Evolution of protein function by domain swapping. *Adv Protein Chem,* 55**,** 29-77.

Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A. & Johnson, D. K. 2010. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnology for Biofuels,* 3.

Pettolino, F. A., Walsh, C., Fincher, G. B. & Bacic, A. 2012. Determining the polysaccharide composition of plant cell walls. *Nat Protoc,* 7**,** 1590-607.

Poon, D. K. Y., Withers, S. G. & McIntosh, L. P. 2007. 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,* 282**,** 2091-2100.

Porter, J., Deere, D., Pickup, R. & Edwards, C. 1996. Fluorescent probes and flow cytometry: new insights into environmental bacteriology. *Cytometry,* 23**,** 91-6.

Purvis, N. & Stelzer, G. 1998. Multi-platform, multi-site instrumentation and reagent standardization. *Cytometry,* 33**,** 156-165.

Qi, M., Jun, H. S. & Forsberg, C. W. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. *Appl Environ Microbiol,* 73**,** 6098-105.

Quan, J. & Tian, J. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One,* 4**,** e6441.

Riedel, K. & Bronnenmeier, K. 1998. Intramolecular synergism in an engineered exo-endo-1,4-beta-glucanase fusion protein. *Mol Microbiol,* 28**,** 767-75.

Rutter, C., Mao, Z. & Chen, R. 2013. Periplasmic expression of a *Saccharophagus cellodextrinase* enables *Escherichia coli* to ferment cellodextrin. *Appl Microbiol Biotechnol,* 97**,** 8129-38.

S. S. Lakhundi, Barnard, D. K., Duedu, K. O., Cain, N., Nagy, R., Krakowiak, J. & French, C. E. *In preparation*. *Citrobacter freundii* as a test platform for recombinant cellulose degradation systems.

Sachs, J. D. 2008. Surging food prices and global stability. *Scientific American,* 298**,** 40-40.

Sandercock, L. E., Meinke, A., Gilkes, N. R., Kilburn, D. G. & Warren, R. A. J. 1996. Degradation of cellulases in cultures of *Cellulomonas fimi*. *FEMS Microbiology Letters,* 143**,** 7-12.

Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juarez, P., Fernandez-del-Carmen, A., Granell, A. & Orzaez, D. 2011. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One,* 6**,** e21622.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A. & Orzaez, D. 2013. GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol,* 162**,** 1618-31.

Saul, D. J., Williams, L. C., Grayling, R. A., Chamley, L. W., Love, D. R. & Bergquist, P. L. 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*". *Appl Environ Microbiol,* 56**,** 3117-24.

Saul, D. J., Williams, L. C., Love, D. R., Chamley, L. W. & Bergquist, P. L. 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. *Nucleic Acids Res,* 17**,** 439.

Schindler, D. & Waldminghaus, T. 2015. Synthetic chromosomes. *FEMS Microbiol Rev,* 39**,** 871-891.

Schmidt, J. C. 2015. Synthetic Biology as late-modern technology. *In:* Giese, B., Pade, C., Wigger, H. & von Gleich, A. (eds.) *Synthetic Biology.* Springer International Publishing.

Schülein, M. 1988. Cellulases of *Trichoderma reesei*. *In:* Willis A. Wood, S. T. K. (ed.) *Methods in Enzymology.* Academic Press.

Schulz, G. & Schirmer, R. H. 1979. The covalent structure of proteins. *Principles of Protein Structure.* Springer New York.

Schwartz, A., Gaigalas, A. K., Wang, L., Marti, G. E., Vogt, R. F. & Fernandez-Repollet, E. 2004. Formalization of the MESF unit of fluorescence intensity. *Cytometry Part B: Clinical Cytometry,* 57B**,** 1-6.

Schwartz, A., Marti, G. E., Poon, R., Gratama, J. W. & Fernández-Repollet, E. 1998. Standardizing flow cytometry: A classification system of fluorescence standards used for flow cytometry. *Cytometry,* 33**,** 106-114.

Schwartz, A., Wang, L., Early, E. A., Gaigalas, A. K., Zhang, A., Marti, G. E. & Vogt, R. F. 2002. Quantitating Fluorescence Intensity From Fluorophore: The Definition of MESF Assignment. *J Res Natl Inst Stand Technol,* 107**,** 83-91.

Seo, H. S., Koo, Y. J., Lim, J. Y., Song, J. T., Kim, C. H., Kim, J. K., Lee, J. S. & Choi, Y. D. 2000. Characterization of a bifunctional enzyme fusion of trehalose-6-phosphate synthetase and trehalose-6-phosphate phosphatase of *Escherichia coli*. *Appl Environ Microbiol,* 66**,** 2484-90.

Sgorbati, S., Barbesti, S., Citterio, S., Bestetti, G. & De Veechi, R. 1996. Characterization of number, DNA content, viability and cell size of bacteria from natural environments using DAPI/PI dual staining and flow cytometry. *Minerva Biotech,* 8**,** 9-15.

Shapiro, H. M., Perlmutter, N. G. & Stein, P. G. 1998. A flow cytometer designed for fluorescence calibration. *Cytometry,* 33**,** 280-287.

Shen, H., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. 1995. Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium *Cellulomonas fimi*. *Biochem J,* 311 ( Pt 1)**,** 67-74.

Showalter, A. M. 1993. Structure and function of plant cell wall proteins. *Plant Cell,* 5**,** 9-23.

Showalter, A. M. 2001. Introduction: plant cell wall proteins. *Cell Mol Life Sci,* 58**,** 1361-2.

Stalbrand, H., Mansfield, S. D., Saddler, J. N., Kilburn, D. G., Warren, R. A. & Gilkes, N. R. 1998. Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β-1,4-glucanases. *Appl Environ Microbiol,* 64**,** 2374-9.

Sternberg, D. 1976. Beta-glucosidase of *Trichoderma*: its biosynthesis and role in saccharification of cellulose. *Appl Environ Microbiol,* 31**,** 648-54.

Stoll, D., Stalbrand, H. & Warren, R. A. 1999. Mannan-degrading enzymes from *Cellulomonas fimi*. *Appl Environ Microbiol,* 65**,** 2598-605.

Stoppok, W., Rapp, P. & Wagner, F. 1982. Formation, location, and legulation of endo-1,4-beta-glucanases and beta-glucosidases from *Cellulomonas uda*. *Appl Environ Microbiol,* 44**,** 44-53.

Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T. & Baldwin, G. S. 2015. BASIC: A new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth Biol,* 4**,** 781-7.

Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques,* 28**,** 1102, 1104.

Sukumaran, R. K., Singhania, R. R. & Pandey, A. 2005. Microbial cellulases - Production, applications and challenges. *Journal of Scientific & Industrial Research,* 64**,** 832-844.

Taiz, L. 1984. Plant cell expansion: regulation of cell wall mechanical properties. *Annu Rev Plant Physiol,* 35**,** 585-657.

Tamada, Y., Swanson, B. A., Arabshahi, A. & Frey, P. A. 1994. Preparation and characterization of a bifunctional fusion enzyme composed of UDP-galactose 4-epimerase and galactose-1-P uridylyltransferase. *Bioconjug Chem,* 5**,** 660-5.

Tamburini, S., Foladori, P., Ferrentino, G., Spilimbergo, S. & Jousson, O. 2014. Accurate flow cytometric monitoring of *Escherichia coli* subpopulations on solid food treated with high pressure carbon dioxide. *J Appl Microbiol,* 117**,** 440-50.

Tan, T. C., Kracher, D., Gandini, R., Sygmund, C., Kittl, R., Haltrich, D., Hallberg, B. M., Ludwig, R. & Divne, C. 2015. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nat Commun,* 6**,** 7542.

Te'o, V. S., Saul, D. J. & Bergquist, P. L. 1995. celA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Appl Microbiol Biotechnol,* 43**,** 291-6.

Teeri, T. T., Penttila, M., Keranen, S., Nevalainen, H. & Knowles, J. K. 1992a. Structure, function, and genetics of cellulases. *Biotechnology,* 21**,** 417-45.

Teeri, T. T., Reinikainen, T., Ruohonen, L., Jones, T. A. & Knowles, J. K. C. 1992b. Domain function in *Trichoderma reesei* cellobiohydrolases. *J Biotechnol,* 24**,** 169-176.

Thongekkaew, J., Ikeda, H., Masaki, K. & Iefuji, H. 2013. Fusion of cellulose binding domain from *Trichoderma reesei* CBHI to *Cryptococcus sp.* S-2 cellulase enhances its binding affinity and its cellulolytic activity to insoluble cellulosic substrates. *Enzyme Microb Technol,* 52**,** 241-6.

Tomme, P., Warren, R. A. & Gilkes, N. R. 1995. Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol,* 37**,** 1-81.

Torget, R., Himmel, M. & Grohmann, K. 1992. Dilute-acid pretreatment of two short-rotation herbaceous crops. *Applied Biochemistry and Biotechnology,* 34-35**,** 115-123.

Treuer, R. & Haydel, S. E. 2011. Acid-fast staining and Petroff-Hausser chamber counting of mycobacterial cells in liquid suspension. *Curr Protoc Microbiol,* Chapter 10**,** Unit 10A 6.

Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A. & French, C. E. 2014. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res,* 42**,** e154.

US EPA. 2013. *Renewable fuel standards (RFS)* [Online]. Available: http://www.epa.gov/otaq/fuels/renewablefuels/index.htm [Accessed 6th March, 2013.

Van Dyk, J. S. & Pletschke, B. I. 2012. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes--factors affecting enzymes, conversion and synergy. *Biotechnol Adv,* 30**,** 1458-80.

Verardi, A., De Bari, I., Ricca, E. & Calabrò, V. 2012. Hydrolysis of lignocellulosic biomass: current status of processes and technologies and future perspectives. *In:* Lima, M. A. P. (ed.) *Bioethanol.* InTech.

Vladut-Talor, M., Kauri, T. & Kushner, D. 1986. Effects of cellulose on growth, enzyme production, and ultrastructure of a *Cellulomonas species*. *Archives of Microbiology,* 144**,** 191-195.

Wakarchuk, W. W., Kilburn, D. G., Miller Jr., R. C. & Warren, R. A. J. 1984. The preliminary characterization of the β-glucosidases of *Cellulomonas fimi*. *J Gen Microbiol,* 130**,** 1385-1389.

Wang, L. L., Gaigalas, A. K., Schwartz, A., Abbasi, F. & Vogt, R. F. 2002. Quantitating fluorescence intensity from fluorophores: Practical use of MESF values. *J Res Natl Inst Stand Technol,* 107**,** 339-353.

Warren, R. A., Gerhard, B., Gilkes, N. R., Owolabi, J. B., Kilburn, D. G. & Miller, R. C., Jr. 1987. A bifunctional exoglucanase-endoglucanase fusion protein. *Gene,* 61**,** 421-7.

Waxdal, M. J., Monical, M. C. & Palini, A. G. 1998. Inter-laboratory relative fluorescence intensity measurements using FlowCal 575 calibration beads: A baseline study. *Cytometry,* 33**,** 213-218.

Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS One,* 6**,** e16765.

White, A., Withers, S. G., Gilkes, N. R. & Rose, D. R. 1994. Crystal structure of the catalytic domain of the β-1,4-glycanase Cex from *Cellulomonas fimi*. *Biochemistry,* 33**,** 12546-52.

Wilson, D. B. 2011. Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol,* 14**,** 259-63.

Wong, W. K., Gerhard, B., Guo, Z. M., Kilburn, D. G., Warren, A. J. & Miller, R. C., Jr. 1986. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. *Gene,* 44**,** 315-24.

Wood, J. C. S. 1998. Fundamental flow cytometer properties governing sensitivity and resolution. *Cytometry,* 33**,** 260-266.

Wood, J. C. S. & Hoffman, R. A. 1998. Evaluating fluorescence sensitivity on flow cytometers: An overview. *Cytometry,* 33**,** 256-259.

Xiao, Z., Zhang, X., Gregg, D. J. & Saddler, J. N. 2004. Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. *Appl Biochem Biotechnol,* 113-116**,** 1115-26.

Xie, G., Bruce, D. C., Challacombe, J. F., Chertkov, O., Detter, J. C., Gilna, P., Han, C. S., Lucas, S., Misra, M., Myers, G. L., et al. 2007. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol,* 73**,** 3536-46.

Yang, B. & Wyman, C. E. 2004. Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. *Biotechnol Bioeng,* 86**,** 88-95.

Yang, S. J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., Davis, M., Westpheling, J. & Adams, M. W. 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. *Appl Environ Microbiol,* 75**,** 4762-9.

Yanisch-Perron, C., Vieira, J. & Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* 33**,** 103-19.

Yu, K., Liu, C., Kim, B. G. & Lee, D. Y. 2015. Synthetic fusion protein design and applications. *Biotechnol Adv,* 33**,** 155-64.

Zenger, V. E., Vogt, R., Mandy, F., Schwartz, A. & Marti, G. E. 1998. Quantitative flow cytometry: Inter-laboratory variation. *Cytometry,* 33**,** 138-145.

Zhang, J., Yun, J., Shang, Z., Zhang, X. & Pan, B. 2009. Design and optimization of a linker for fusion protein construction. *Progress in Natural Science,* 19**,** 1197-1200.

Zhang, X. Z., Sathitsuksanoh, N. & Zhang, Y. H. 2010. Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans*: Heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. *Bioresour Technol,* 101**,** 5534-8.

Zhang, Y.-z., Kemper, C., Bakke, A. & Haugland, R. P. 1998. Novel flow cytometry compensation standards: Internally stained fluorescent microspheres with matched emission spectra and long-term stability. *Cytometry,* 33**,** 244-248.

Zhang, Y., Werling, U. & Edelmann, W. 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res,* 40**,** e55.

Zhang, Y. H. & Lynd, L. R. 2006. A functionally based model for hydrolysis of cellulose by fungal cellulase. *Biotechnol Bioeng,* 94**,** 888-98.

Zhao, L., Pang, Q., Xie, J., Pei, J., Wang, F. & Fan, S. 2013. Enzymatic properties of *Thermoanaerobacterium thermosaccharolyticum* beta-glucosidase fused to *Clostridium cellulovorans* cellulose binding domain and its application in hydrolysis of microcrystalline cellulose. *BMC Biotechnol,* 13**,** 101.

Zhou, S., Davis, F. C. & Ingram, L. O. 2001. Gene integration and expression and extracellular secretion of *Erwinia chrysanthemi* endoglucanase CelY (*celY*) and CelZ (*celZ*) in ethanologenic *Klebsiella oxytoca* P2. *Appl Environ Microbiol,* 67**,** 6-14.

Zhou, S. & Ingram, L. O. 2000. Synergistic hydrolysis of carboxymethyl cellulose and acid-swollen cellulose by two endoglucanases (CelZ and CelY) from *Erwinia chrysanthemi*. *J Bacteriol,* 182**,** 5676-82.

Zhu, Y., Zhou, H., Bi, Y., Zhang, W., Chen, G. & Liu, W. 2013. Characterization of a family 5 glycoside hydrolase isolated from the outer membrane of cellulolytic *Cytophaga hutchinsonii*. *Appl Microbiol Biotechnol,* 97**,** 3925-37.

Zverlov, V., Mahr, S., Riedel, K. & Bronnenmeier, K. 1998. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology,* 144**,** 457-465.