Structure and function characterisation of a bacterial copper dependent mono-oxygenase from *C. cellulovorans*, with particular insight into biofuel production.

By

James William Christopher Robinson

MSc by Research

University of York

Chemistry

January 2014

Abstract

Lytic polysaccharide mono-oxygenases (LPMOs) are a recently discovered, and rapidly expanding, family of enzymes capable of oxidizing complex polysaccharide polymers. They are attracting interest in biotechnology as potential factors in the efficient treatment of recalcitrant biomass with a particular focus on biofuel production. Previous work has identified key physiological features in LPMOs that are archetypal in the activity of these enzymes. Here we report that an enzyme from *Caldibacillus cellulovorans*(*C. cellulovorans)* with high sequence conservation with other LPMOs*,* displays high affinity copper binding characteristics typical of a CBM33 molecule. By using EPR, the active site of *C. cellulovorans* CBM33 is revealed as an axial type 2 copper centre in a histidine brace conformation, emblematic of the LMPO clan. Oxidative degradation of chitin and cellulose was proven with mass spectrometry studies, product identity suggested a propensity for C1 oxidation; however, potential C4 site oxidation by-products were observed. The broad spectrum activity of *C. cellulovorans* CBM33 could potentially stimulate a more diverse selection of efficient biofuel treatments. Further to this, oxidative activity at 80°C creates a new paradigm for high temperature biotechnology treatments for the degradation of recalcitrant polysaccharides.

Table of Contents

List of Figures

Acknowledgements

I would like to thank Dr Paul Walton and Dr Gideon Davies for giving me the opportunity to carry out research in their groups for the past year and for the support that they have provided. Their influence has fuelled the interest that I have in the field and for science in general.

My thanks go to all the member of the Walton and Davies groups for their support and friendship this past year; Rebecca Gregory, Sherry Chan, Dan Wright, Andrew Thompson, Luisa Ciano, Lorna Clarke, Emma Dux, and in particular Glyn Hemsworth for his support in every area. I also thank Simon Grist and Sally Lewis for maintaining the wet lab. YSBL is a comfortable working environment that provides personal support and a rich intelligent stimulation. I am grateful for the chance to work with everyone and hope to return in the future.

Finally, I would like to thank my amazing family for all their support, financial and otherwise, and my fiancée Elinka for tolerating me.

Author's Declaration

EPR data collection at The University of York was carried out by Rebecca Gregory in my company and under the supervisor of Dr Paul Walton.

MALDI/TOF mass spectrometry measurements and PCR of the Champion pET-SUMO vector were carried out by Glyn Hemsworth in my presence.

All other work described in this thesis was performed by me, except where stated.

Chapter 1 – Introduction

Biofuels

Future energy security is an increasing global issue that is polarising opinions in science, economics and politics. [1] Remaining fossil fuel deposits are fast being depleted for the production of electricity, transport, heating and oil based polymers. [2] Increased international demand from both industrially established and developing nations has led to forecasts of negative net fuel availability within 50 years, especially for states dependant on the importation of fuels.[3] Decreasing fossil fuel stocks coupled with exponentially increasing demand and market diversity have driven fuel costs to the highest ever point. Fossil fuels are rapidly becoming an unviable option, economically and environmentally, for many nations. [2,4] Further to demand rapidly eclipsing demand, certain forecasts predict that there will be a dramatic, and potential fatal, global temperature shift in the near future if over 40% of available fossil fuels are burnt. [5] Even if these worst case scenario forecasts are false it is increasingly obvious that energy is rapidly becoming one of the most prominent global issues.

Accumulation of global pressure has triggered an increase in the interest and financial support for alternative fuel sources. [6] Scientists have hailed biofuel as an energy source with both the production potential and environmental sustainability to replace fossil fuels in a balanced energy market. [4] Biofuels are currently estimated as 10% of the global energy production (50 EJ per annum) and the field as a whole contributes as the most prolific renewable energy source. [7]

Biofuels as a field of research and industry is incredibly diverse. Two major biofuel fields exist: the production of fuel biomass feedstock and harvesting oil like metabolic products from algal species and the formation of short chain alcohols from biomass. Though related, the two are distinct in the production process, production capabilities, economic outlook and scientific niche; this work will focus on the production of ethanol and short chain hydrocarbons (generally 4C or fewer) from glucose. Glucose is a potential feedstock with low cost, water solubility, non-toxic and abundant in plant biomass

There are currently mature biofuel markets around the world: ethanol produced from processed sugar cane in Brazil and corn ethanol produced in the United States of America supporting beliefs that biomass fuels can realistically enhance the global energy portfolio. [8]

These first generation biofuels produced from plant based simple sugars have factors limiting their successful integration into a global energy market:

1. Decreased food security caused by limitations in available arable land. The financial benefits of bio-mass for the production of fuel over agricultural crops is leading to decreasing global food security and higher food prices.
2. Regional water shortages, affecting the agricultural industry and water availability for human consumption. Depletion of fresh water sinks is a realistic danger due to the large volumes of water required in existing biofuel production protocols the irrigation of crops and downstream reaction steps and product washing.
3. The financial input required to reduce CO_2 emission by one tonne has been calculated according to the existing market. The cost of reducing CO_2 emissions outweighs the positive environmental effect of 1st generation biofuels leading to their dismissal as a long term energy solution due to the incompatibility with the current economic environment. [8] The positive global reach of 1st generation biofuels is regarded as low considering existing infrastructure and production protocols.
4. Limitations in market infrastructure including the transportation, processing facilities, refining facilities and experts and fuel compatibility has led to high company start-up costs and a nullification of positive environmental impacts.

International bodies, governments and companies have begun to emphasise the production of biofuel from sources that do not have a detrimental impact on food, water and land security globally, 2nd generation biofuels. [9] The focus of bio-technology has shifted towards the step-wise liberation of sugars from agricultural waste and forest residues rich in lignocellulose and then conversion to fuels.

Biomass structure

The superfluity and high potential energy of lignocellulose as a feedstock for biofuel production has made it an impossible avenue for companies to ignore when searching for greener, renewable technologies to contribute to the complex global energy network.

Cellulose is a polymer formed by the condensation of D-glucose monomers through β–1, 4–glycosidic bonds to form linear chains (Figure 1.1A). Single cellulose chains aggregate laterally through hydrogen bonding and Van de waals interactions to form rod like structures called microfibrils. [10] Glucan chains do not exist as single entities in nature and therefore in the process of real time synthesis multiple chains are produced simultaneously and bind to form dense microfibrils in a crystalline matrix (Figure 1.1B). [11] Cellulose is the base component of plant cell walls and provides protection, structure and support for the individual cell and synergistic rigidity with tissue.

During cell wall synthesis cellulose is loaded with hemicellulose (branching polysaccharide polymers, such as mannan and xylan[12]) and pectins to form a voluminous web of branching carbohydrate side chains. These spaces are occupied by

bound lignin (phenylpropanoid polymers), [13] the effect is complete encapsulation of cellulose leading to biomass recalcitrance (Figure 1.1C). Recent work aimed at characterising the molecular make up of lignocellulose has suggested that lignin physically inhibits degradation molecule access to cellulose.

Cellulase action

For cellulosic biofuel to be a major component of the global energy production network, saccharification of the complex cellulose crystal must be achieved. [9] Existing biotechnology protocols utilise step-wise degradation of the crystalline structure.

Firstly the general recalcitrance of the lignocellulose complex is reduced by disruption of the hemicellulose or lignin network to liberate cellulose. A variety of techniques exist including dilute sulphuric acid pre-treatment, alkaline explosion and organic solvent extractions. [14]
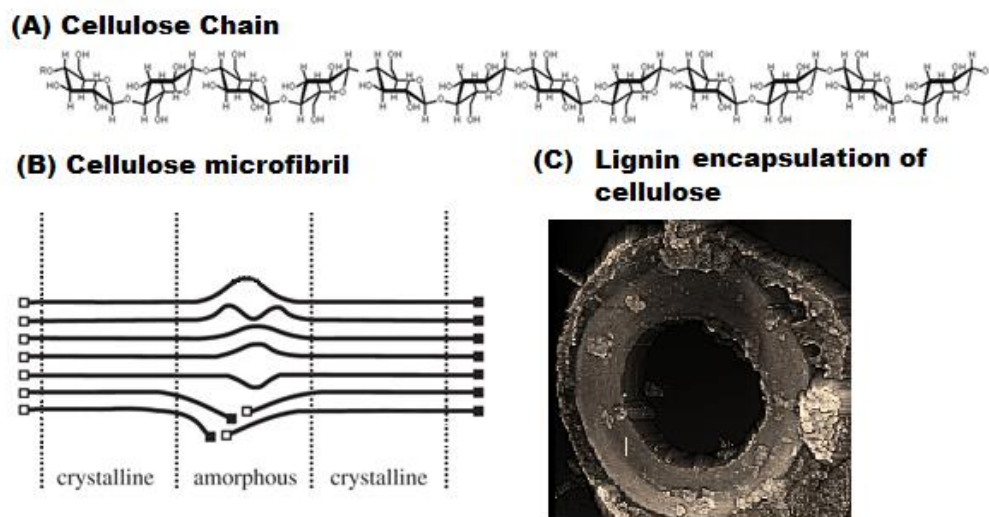
Figure 1.1 The complex structure of cellulose.
Figure 1.1 shows the complex structure of lignocellulose. (Figure from Nguyen (2010) [15]).
(A) Shows the linear structure of the cellulose polymer, single molecule polyglucan like this does not form readily in nature, instead the chain has a propensity to polymerise in rod like components to create dense, crystalline structures.
(B) Cartoon representation of a cellulose microfibril displays the inaccessibility of individual cellulose chains due to the steric packing in crystalline regions. Not shown are the stabilising molecules bound into the complex 3D structure which further stabilise the crystal structure.
(C) Image displaying a cross section of a plant cell wall displaying the complete encapsulation of cellulose by lignin. (Figure adapted from Lawrence Livermore National Laboratory website).

Non-cellulose cell wall component removal greatly increases the accessibility of the cellulose, enabling the second step, enzymatic degradation of the cellulose matrix. [13] Existing enzyme

treatments contain a wide spectrum of carbohydrate active molecules. A combination of endo-acting and exo-acting cellulases will cleave cellulose into glucose monomers (Figure 1.2A). [16,17]

These enzymes are greatly inhibited by the recalcitrant structure of cellulose, even after the pre-treatment process. As evidenced in Figure 1.2, Parts B and C, cellulases have active site envelopes buried with the protein core. These pockets are created by aligned hydrophobic residues that hinder the entrance of water molecules protecting an open cavity within the protein core containing the catalytically dependant residues.[18]

The globular structure of cellulases means that the enzyme is only able to cleave glycosidic bond when a single cellulose strand is able to enter the enclosed groove (endoglucanase) or tunnel (exoglucanase) active sites. [19] Because of this, cellulase enzymes are only able to degrade cellulose in amorphous regions of cellulose and a large protein load is required to completely degrade the matrix. The inaccessibility of the active site to truly crystalline polysaccharides inhibits the production of short oligosaccharides and reduces the viability of lignocellulosic biofuels without serious steps to improve recalcitrant polysaccharide breakdown. Despite a rapidly expanding repertoire of techniques, high enzyme load and the initial energy input required have made lignocellulosic fuels increasingly unviable economically for the field of biotechnology.
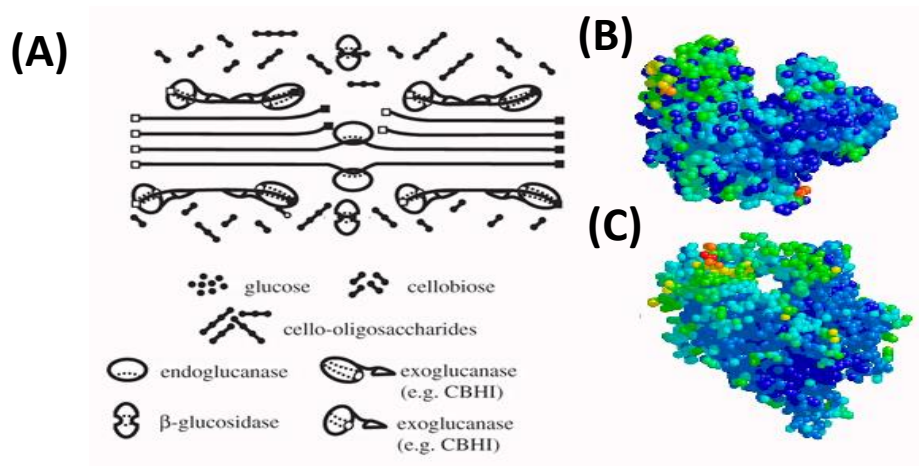
Figure 1.2 Cellulase action on crystalline cellulose.
Figure 1.2 shows the structural features of cellulase enzymes and the synergistic breakdown of cellulose. (Figure adapted from Nguyen (2010) [15]).
(A) Cartoon of cellulose activity on a microfibril of cellulose with particular emphasis on different site of attack for the endo and exo cellulase enzymes
(B) Space filling model of an endo-glucanase active anywhere along an exposed cellulose chain enabled by the groove like binding site. The different coloured residues show the amino acid residue polarity of charge.
(C) Space filling model of an exo-glucanase active and the cellulose chain ends. The different coloured residues show the amino acid residue polarity of charge.

Lytic Polysaccharide Mono-oxygenases

The inestimable potential of cellulase treatments in biotechnology has driven research into improving the poor activity of cellulose degrading cocktails. An extended database of cellulase enhancing conditions exists; citing enzymes, small molecules, buffer components and thermal conditions. [20,21]

In one experiment, a fungal protein, of unknown function, was shown to amplify hydrolysis of crystalline cellulose two fold and in the process greatly decreasing the required enzyme load. Due to the acceleration in hydrolytic activity observed when these proteins were present they were named Glycoside Hydrolases, GH61 (See Table 1.1). [22-26]

	Bacterial	Fungal
First discovered as	Chitin binding protein 21. Reports that CBP21 enhances chitinase activity.	*Cel* Glycoside hydrolase 61 GH61 enhances the activity of classical hydrolases.
Reclassified by affinity	Carbohydrate binding module 33 Recognise as being bacterial homologues of the copper dependant mono oxygenase proteins from fungi.	Glycoside hydrolase 61 A copper dependant molecule that exhibited oxidative activity and no hydrolase activity.
Reclassified by activity	Auxiliary Activity 9 Bacterial LPMOs characterised as oxidative molecules acting by an unknown mechanism but encompassing CBM33 enzymes.	Auxiliary Activity 10 Fungal LPMOs were reclassified as no hydrolase activity was observed and enzymes generated oxidative reaction products.

Table 1.1 **Changes in LPMO classification in the CAZy database.**
Table 1.1 shows how bacterial and fungal LPMOs have been classified since their discovery finally leading to their classification as Auxiliary Activity enzymes with oxidative action.
The CAZy database categorises enzymes active on carbohydrates according to the Henrissat classification, divided into families by the structural and activity similarities. [27] Families AA9 and AA10 in the CAZy database are LPMOs from all kingdoms of life. [28]

Attempts were made to characterise LPMO enzymes as a potential augmentation of existing biofuel protocols, a crystal structure of CelGH61B from *Hypocrea jecorina* was published allowing for structural elucidation. [29] CelGH61 had an immunoglobulin-like β-sandwich fold, with a di-valent metal active site situated on a flat binding face (Figure 1.3).
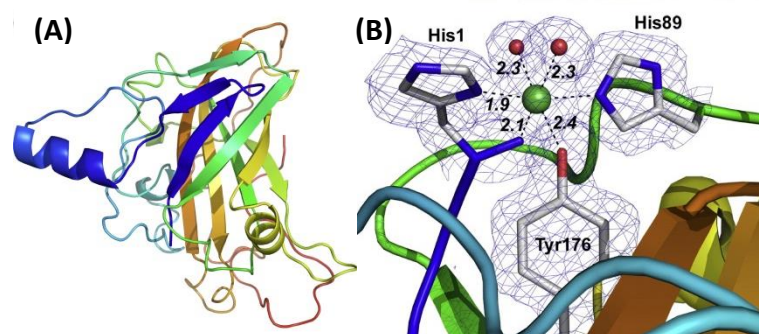
Figure 1.3 The structure of GH61.
GH61 structure both in a whole molecule, line and ribbon representation and an electron density model of the active site.
(A) Cartoon representation of CelGH61 enzyme showing the β-barrel structure of the immunoglobin like fold where the different colours represent different domains of the 3D protein.
(B) Showing a di-valent metal active site on a flat binding face with electron density showing the two histidine residues coordinating the copper ion (Green) and the ligand bond lengths.
Figure adapted from Quinlan et al.[24]

Nickel was modelled into the site; the metal ion coordinated by a histidine brace and a tyrosine residue. The N-terminal histidine residue of the amino acid chain coordinates the metal ion and was therefore forecast as being essential for activity.

The structural aspects of GH61 are atypical of classical GHs and suggestions were made that GH61 enzymes were not classical hydrolases. No archetypal binding groove or tunnel was observed and importantly the key acid/base catalytic residue pairs found in hydrolytic enzymes were absent from the structure.

Active site

Work on GH61 enzymes began to focus on the characteristics of the active site and the identity of the metal ion.

Ambiguity surrounding the metal identity was partially alleviated when a GH61 structure was published; [24] Quinlan et al identified the metal as being a Cu(II) ion. The structure of GH61 was recognised as similar to a chitin binding protein (CBP21, renamed CBM33) from *Serratia marcescens* which had been published previously. [30] This molecule was shown to have oxidative action on crystalline chitin though the mechanism and metal dependency were missed. Utilising this information Quinlan et al predicted GH61 and CBM33 as being copper dependent monooxygenases. Electro paramagnetic resonance (EPR) studies of the lone electron in GH61 allowed for characterisation of the copper centre. The G value obtained for GH61 of 2.27, as shown in (Figure 1.4), is representative of Type (II) copper centres and has become indicative of some LPMOs. [24] This type (II) copper centre can be characterised by a single copper molecule bound by nitrogen or oxygen ligands in a square planar geometry.

	gII	gI	AII Gauss
Copper (II) in aqueous acetate buffer, pH 5	2.42	2.08	118
Copper (II) -TaGH61A*	2.27	2.06	153
met-dopamine beta-mono-oxygenase	2.29	2.06	155
met-peptidylglycine alpha-hydroxylating mono-oxygenase	2.25	2.05	162
met-copper amine oxidase	2.32		153
met-lysyl oxidase	2.28		153
Type 2 site in copper methane mono-oxygenase	2.24	2.04	185
Copper (II)-BaCBM33	2.25		125
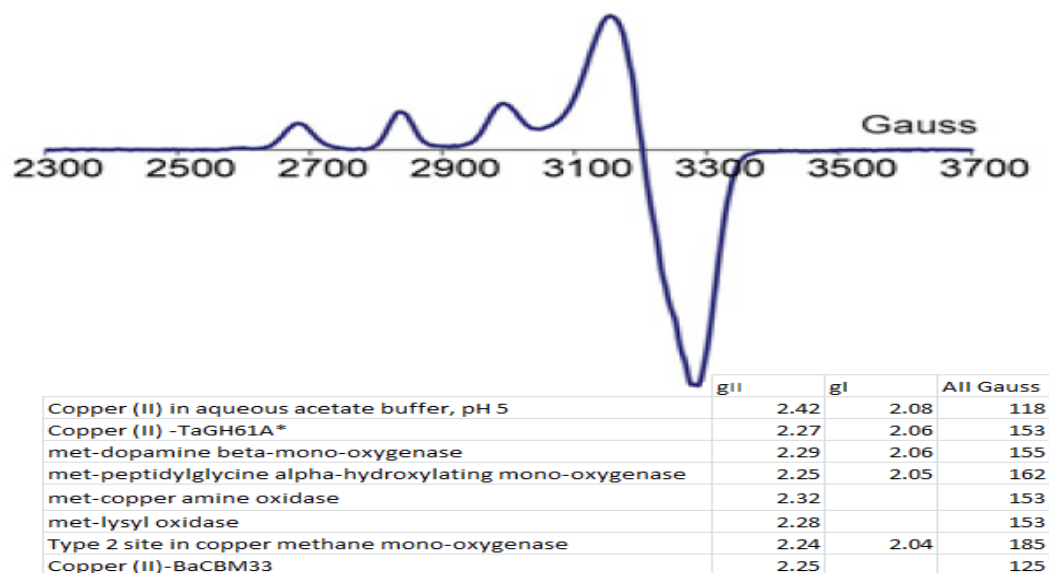
Figure 1.4 Type 2 copper centre analysed by EPR.
Figure 1.4 showing EPR spectrum of a copper centre coordinated by O/N ligands in a square planar geometry (Type 2 copper centre)
The Four peaks are indicative of a type 2 copper centre and have been seen in the EPR spectra of LPMO enzymes analysed by EPR. Adapted from Quinlan *et al.* [24]

Both GH61 and CBM 33 have been shown to contain similar, type (II) copper active sites, and overall β-barrel structure with a flat binding face. They are structurally and mechanistically related; [31] both have a metal ion active site situated on a flat binding face essential for substrate orientation. Within the active site; two histidine residues, including the N-terminal residue coordinate the copper. In GH61 a tyrosine residue also coordinates the copper and the N-terminal histidine was found to be methylated; neither trait has been identified in CBM33 which gives some indication to the different substrate activities (Figure 1.5).
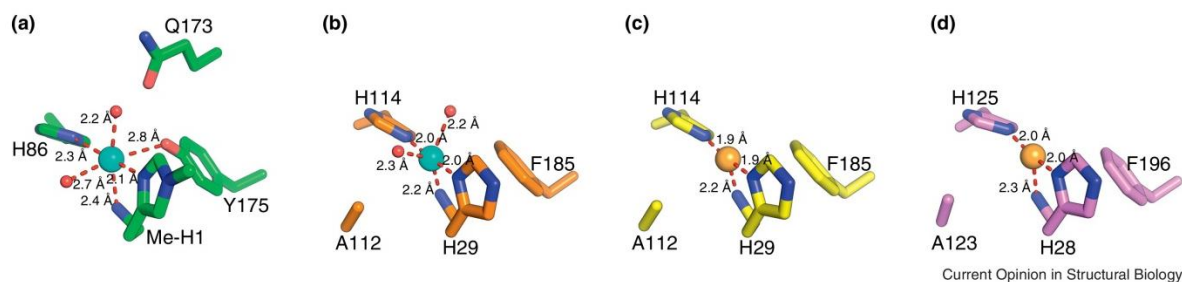
Figure 1.5 Comparison of a GH61 active site and two species of CBM33s.
Figure shows a carton comparative analysis of different copper active sites in a variety of LPMO enzymes.
Cu(II) centre in *Thermoascus auranticus GH61* (PDB:2yet). Blue sphere represents a copper ion in the 2+ state.
Cu(II) centre in *Enterococcus faecalis CBM33* (PDB:4alr)). Blue sphere represents a copper ion in the 2+ state.
Cu(I) centre in *Enterococcus faecalis CBM33 (*PDB:4alt). The orange sphere represents a copper ion in the 1+ state, though the coordination of the copper indicated that x-ray reducing of the copper has returned the active site to Copper 2+.
Cu(I) centre in *Bacillus amyloliquefaciens CBM33 (*PDB:2yox) The orange sphere represents a copper ion in the 1+ state though the coordination of the copper indicated that x-ray reducing of the copper has returned the active site to Copper 2+, though the coordination of the copper indicated that x-ray reducing of the copper has returned the active site to Copper 2+..
The glutamine residue in GH61 active site is replaced by Phenylalanine in both the species of CBM33s and there is an additional alanine. These differences are thought to dictate the different activities and binding characteristics. It is clear to see the conserved geometry of the coordinating residues about the copper active site.
Figure adapted from Hemsworth *et al* (2013) [32]

Both CBM33 and GH61 act on the polysaccharide matrix in a non-hydrolytic manner; a mechanism involving the copper dependent insertion of atmospheric oxygen into the lignocellulose chain has been shown. [33-35] Activity in LPMOs is dependent on the redox activation of the enzyme by reducing agents such as celliobiose dehydrogenase enzymes (CDH1) or small molecules. [36] This activation is thought to allow binding of atmospheric oxygen for the formation of a catalytic reactive oxygen species. Evidence has been provided that GH61 and CBM33 molecules can create random length chain breaks in the polysaccharide chain by oxidative action to cleave a glycosidic bond and produce a charged carboxylic acid chain break. [35,37] Recent theoretical work has suggested that they break glycosidic bonds using a copper-oxyl mediated, oxygen rebound mechanism. They suggested that previous theories underestimated the reactive oxygen species 'power' required to break the glycosidic bond. [38] The mechanism by which the cellulose chain is broken remains unresolved. Questions remain about the specificity of the reaction, due to the observation of C_1, C_4, and C_6 oxidation products, as well as further doubts about the mechanism for substrate binding.

These enzyme families are now regarded as lytic polysaccharide mono-oxygenases (LPMOs) and their classification in the carbohydrate-active enzymes database (CAZy; http://www.cazy.org) was altered accordingly. [39] In 2013, the GH61 and CBM33 families were reclassified by function to the Auxiliary Activity families 9 and 10 respectively. [40]

LPMO enzymes have an important role in future cellulosic biofuel protocols due to the potential increase in yield, rate and required enzyme load. Existing treatments are limited by the enzymes that are available and characterised. Biotechnology must advance as a field and find alternative enzymes with broader range of temperature and pH tolerance. Most importantly, more enzymes must be characterised that have a broad activity spectrum that will allow for complete conversion of lignocellulose material to glucose, including the non cellulose polysaccharides in the complex.

C. cellulovorans CBM33

C. cellulovorans is a thermophillic, cellulolytic, bacterium that obtains nutrients through the degradation of wood and plant matter rich in lignocellulose. [41] *C. cellulovorans* releases an extracellular 'cellulosome' cocktail containing enzymes for the degradation of plant polysaccharides. A component of the *C. cellulovorans* cellulosome is a multi-domain β-1, 4-mannanase; [42] a GH5 mannanase enzyme, which is bracketed by CBM3 modules and linked to an N-terminal protein thought to be an AA10 enzyme. The gene is thought to be a LPMO due to the high sequence conservation with known CBM33s (Uniprot: Q9RFX5 - mannan endo-1, 4-β-mannosidase). Discovering a LPMO from *C. cellulovorans* CBM33 enzyme is of particular interest for 2 reasons:

1. It is the only LPMO present in an organism that is known to release extracellular enzyme cocktails for the breakdown of material rich in lignocellulose such as woody materials. Therefore, it is a possibility that the protein exhibits activity on a variety of carbohydrate substrates to accelerate the release of polysaccharide chains from the complex matrix. This protein is of particular interest due to the presence of the encoding gene on an operon active on mannan polymers, it is possible that this protein is active on the complex crystalline substrate mannan to allow nutrient uptake by the cell. Activity of an LPMO on mannan has not yet been shown. This novel activity would therefore be very important in biotechnology for both biofuels and for other fields that are dependent on the degradation of complex polysaccharides.

2. *Calibacillus Cellulovorans* is a bacterial strain that grows in high temperatures (50-80°C). This means that the enzymes of the cellulosome are likely to be stable at those temperatures. The thermo stability of these enzymes is desirable as it allows for biotechnological applications at higher temperatures to produce higher yields. The increased stability can give a larger range of bio-industrial treatment conditions.

LPMO enzymes will have an important role in future cellulosic biofuel production pipelines due to the potential increase in yield, rate and reduced enzyme load that they induce. Existing treatments are limited by the enzymes that are available and characterised. Biotechnology must advance as a field and find alternative enzymes with broader range of temperature and pH tolerance. Most importantly, more enzymes must be characterised that have a broad activity spectrum that will allow for complete

conversion of lignocellulose material to glucose, including the non cellulose polysaccharides in the complex.

Aim

The project aimed to characterise *C. cellulovorans* CBM33 structure and function with a view to biotechnological applications. Work will focus on the purification of *C. cellulovorans* CBM33 from E. *coli*. Experimentally the main aims are to confirm the identity of *C. cellulovorans* CBM33 as a LPMO and elucidate the activity and mechanism.

If successful the works stands to make a significant impact with wide reaching implications in biotechnology. *C. cellulovorans* CBM33 could improve the yield and cost efficiency of existing biofuel treatments and enable expansion to other biological feedstock such as hemicellulose. The enzyme could also impact bacterial washing powders, water treatment

Chapter 2 – Purification of *C. cellulovorans* CBM33

2.1 Expression vector

Expression system optimisation has given rise to a class of vectors derived from the *E.coli* strain pBR322 called pET vectors [43]. These vectors have been manipulated to alter the available leader sequences, expression signals, relevant restriction sites, and other features to optimise recombinant gene expression (see Figure 2.1).
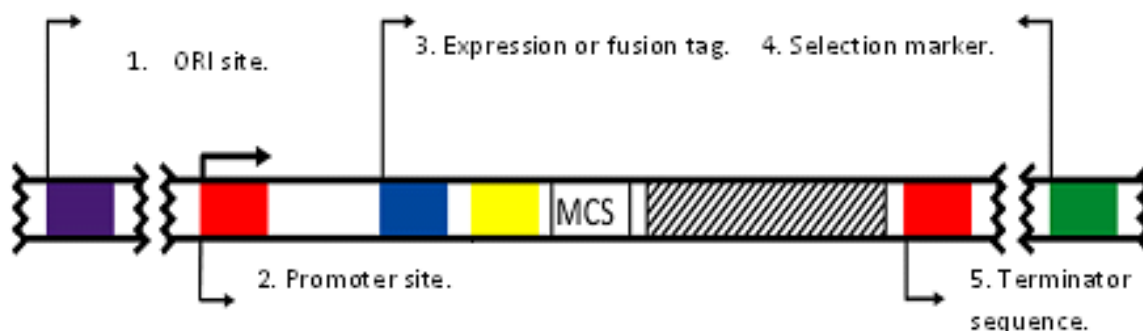
Figure 2.1 Components of an expression vector.

ORI site is a short sequence of DNA where replication initiates. Recombinant plasmids contain an ORI site separate from the host replication site. Most expression strains have a genetically modified ORI site that has replication inhibitory mechanisms for hyper replication of the plasmids.

Promoter sites are regions of DNA that allow for specific recognition of a gene transcription start site by RNA polymerase. Expression tags are transcribed regions of the gene that improve the downstream purification process. These can include regions that will help with cell localisation (PelB leader sequence), affinity purification (poly-histidine sequence) or the solubility of the intracellular recombinant protein (SUMO).

Selection marker gene, encoding for the translation of a protein to provide a specific anti-biotic resistance to the strain to allow for selective cloning. Common antibiotics used in recombinant protein expression are ampicillin (AmpR – β lactamse gene) or kanamycin (kanR – nptII gene).

Terminator region of DNA to stop the transcription of the recombinant genes.

A former member of the Davies' group, Ed Taylor, designed and ordered a codon optimised vector for the CCCBM33 (Figure 2.2). The C.cellulovorans gene was cloned into a pET-11a vector for optimal expression in competent E.coli strains:

1. The selection marker (see Figure 2.1) in the pET-11a vector is an ampicillin resistance gene. The gene encodes for a β-lactamase enzyme, the enzyme has the ability to catabolise ampicillin like antibiotics to give resistance for selected growth

2. *C.cellulovorans* CBM33 expressed from pET-11a has an expression tag for periplasmic targeted expression. The N-terminal sequence of 22 amino acid residues called a PelB leader promotes periplasmic translocation of the recombinant protein. The presence of an expression tag is designed to allow for more efficient purification to yield higher concentrations of recombinant protein.

```
                                  CAT
 ATGAAGTATCTGCTGCCGACGGCTGCTGCTGGTCTGCTGCTGCTGGCTGCCCAACCGGCTATGGCGCATGGTGGT
 ATGGTGTTCCCGGCGACCCGTACGTATGCCTGCTACGTCGATGGCAAAGTGCATGGTAATGGCGGTGACCTGAAC
 ATGATTAATCCGGCATGTCTGGATGCGCTGGCCATCAGTGGCAACTATCAGTTTTGGAACTGGTTCGGTAATCTG
 ATTTCAAACGCTGGCGGTCGTCACCGCGAAATTATCCCGGATGGCAAGCTGTGCGGTCCGACCGCAAGCTTTGAT
 GGCATGAATCAGGCTCGTACGGACTGGTGGACCACGCGTCTGCAACCGGGTGCAACCATCACGGTGCGTGTTAAC
 GCATGGGCTCCGCATCCGGGCACCTGGTATCTGTACGTGACGCGTGATGGTTGGGACCCGACCCAGCCGCTGAAA
 TGGAGCGATCTGGAACCGACCCCGTTCTCTCAAGTTACGAATCCGCCGATTAACAGCTCTGGCCCGGACGGTGCC
 GAATATTCCTGGCAGGTCCAACTGCCGAATAAGCAGGGCCGTCACATTATCTACATGATCTGGCAACGCAGTGAT
 TCCCCGGAAGCGTTTTATAACTGTTCAGACGTTTACTTCGGCTCGGGTCCGATTGCCTACGAATTTGGTGATCCG
                          CGCGAAGGCGGTACCATGATC
                              TAAGGATCC
```

Figure 2.2 Sequence of the ordered C. cellulovorans CBM33 gene.

Figure 2.2shows the codon optimised expression gene for *C.cellulovorans* CBM33 with an attached PelB sequence for periplasmic targeted expression. (Optimized Sequence Length:708, GC%:56.54)

3. Pet-11a has a T7 promoter controlling the expression of the recombinant gene. The vector does not however have a readily expressed T7 RNA polymerase; instead, polymerase expression is regulated through the Lac operon. Lac genes are conditionally expressed in the presence of lactose which allows for inducible expression in vitro by addition of a lactose analogue, IPTG.

Recombinant gene expression involves the experimental production of a protein from cloned DNA. High protein yields are dependent on the ability to control expression.

Protein expression is dependent on the RNA polymerase catalysed gene transcription to messenger (m) RNA. Regulation of transcription comes from promoters which are regions of DNA that allow the binding of polymerase molecules which allows for transcription of the genes. An effective mechanism for the control of gene expression is the insertion of an inducible promoter into the expression vector to allow control of the expression. Inducible promoters only activate expression in the presence of physical or chemical stimulus.

The Lac promoter in expression strains

The Lac operon in its natural form is a set of genes encoding the uptake and metabolism of lactose in *E.coli* and selected bacterial strains. Three structural genes (*lacZ*, *lacY*, and *lacA*) control the uptake and breakdown of lactose in sugar monomers and their expression is tightly regulated by a repressor molecule.

The strong regulation of the Lac operon has made inducible expression of recombinant protein a powerful tool in microbiology. In laboratory experiments, as in natural *E.coli* the expression of genes in the Lac operon are completely switched off the absence of a substrate. The Lac I gene encodes for the translation of a repressor protein (Lac repressor) that is constantly expressed in live *E.coli* cells. Lac repressor regulates the expression levels of the Lac operon in *E. coli* in order to prevent the genes being transcribed when a useful substrate is present. The repressor molecule forms a tetramer complex that binds to the DNA that encodes for the Lac operon.

The repressor molecule forms a dimer readily with itself and will then form a complex made up from four individual subunits of the repressor protein. Each monomer has a domain at the N terminal of the subunit, residues 1-49, that when unbound is flexible relative to the protein but takes a rigid shape when able to bind to the major groove of DNA cis of the promoter site. [44]This binding domain has exposed helix-turn-helix motifs that allows for sequence recognition of DNA bases in the operator region. Binding of the helix-turn-helix domain allows for hydroxyl groups of the deoxyribose backbone to anchor the repressor to the operator site.

The binding of Lac repressor to this major operator site (Om) and allows for a second operator region, regarded as the auxiliary operator site (Oa), which is distal from the major operator region to bind to a second helix-turn-helix region of the repressor molecule. This multiple site binding of the operator regions creates a complex folded conformer of the *E. coli* which prevents binding of the transcription machinery. The repressor molecule directly hinders RNA polymerase binding and therefore negatively regulates the expression of the Lac genes.

Lactose and lactose mimics have a higher affinity for Lac repressor than the operator sites and, when present in the cell, bind to the repressor tetramer. Lactose binding to the repressor molecule causes an allosteric alteration the protein. The allolactose binding site was shown to be in the hinged section of the complex protein distal to the DNA binding domain. Binding of IPTG to LacI is mediated by a water mediated, hydroxyl conformation about IPTG which alters the structure of the protein, especially the dynamic DNA binding domains. [45] This alteration in the 3D structure alters the orientation of the helix-turn-helix binding domains relative to the DNA which causes dissociation of the repressor from the operator site. This dissociation allows RNA polymerase access to the promoter and transcribes the gene in the multiple cloning sites.

In vitro, a lactose mimic, IPTG, is used to selectively induce the genes controlled by the Lac promoter, IPTG directly mimics allolactose and is transported across the membrane of the cell and can interact with the transcription machinery. The pET vectors, a selection of expression strains derived from BL21 have utilised this powerful induction tool to control recombinant gene expression in order to optimise translated protein yield. The pET vector has a gene attached to the Lac promoter which transcribes a T7 RNA polymerase. This means that when IPTG is added into the culture it is taken into the *E.coli* cell and promotes the expression of the Lac operon genes, in this case a T7 RNA polymerase. The transcribed mRNA is translated and a copy of T7 polymerase is produced. T7 polymerase then binds to a different promoter site, in a residue specific manner, and catalyses the transcription of the genes controlled by the T7 promoter. Normally, the recombinant gene is cloned into the vector downstream of the T7 promoter site so that the gene of interest is expressed when IPTG is added to the system. Is a powerful inducer for in vitro work as there is no expression prior to addition

but after the IPTG is induced the molecule binds to the repressor and, as the cell cannot break down the synthetic version, transcription is constantly turned on to optimise recombinant protein yield (See Figure A1).

2.2 Expression strain

Once the recombinant gene has been cloned into a circular DNA vector it must be inserted into a host cell for expression of the gene to synthesise the desired protein of interest. There are a varied of selection of cell strains which have each been genetically altered to help increase to amount of recombinant protein produced. In order to test the expression of *C.cellulovorans* CBM33, from the pET-11a vector, an expression strain matrix experiment using different *E.coli* strains. These strains are all derivatives of BL21 *E. coli* cells that have been genetically altered to control aspects of recombinant protein production. Several strains of *E. coli* expression cells were tested in a small scale expression test.

Figure 2.3 is an SDS-PAGE gel showing the isolated soluble protein fraction from the tested strains. To isolation methods were used, analysing the whole cell isolate and the fraction obtained from the periplasm, utilising the PelB leader sequence periplasmic expression to separate C. cellulovorans CBM33 from the cytosolic fraction.
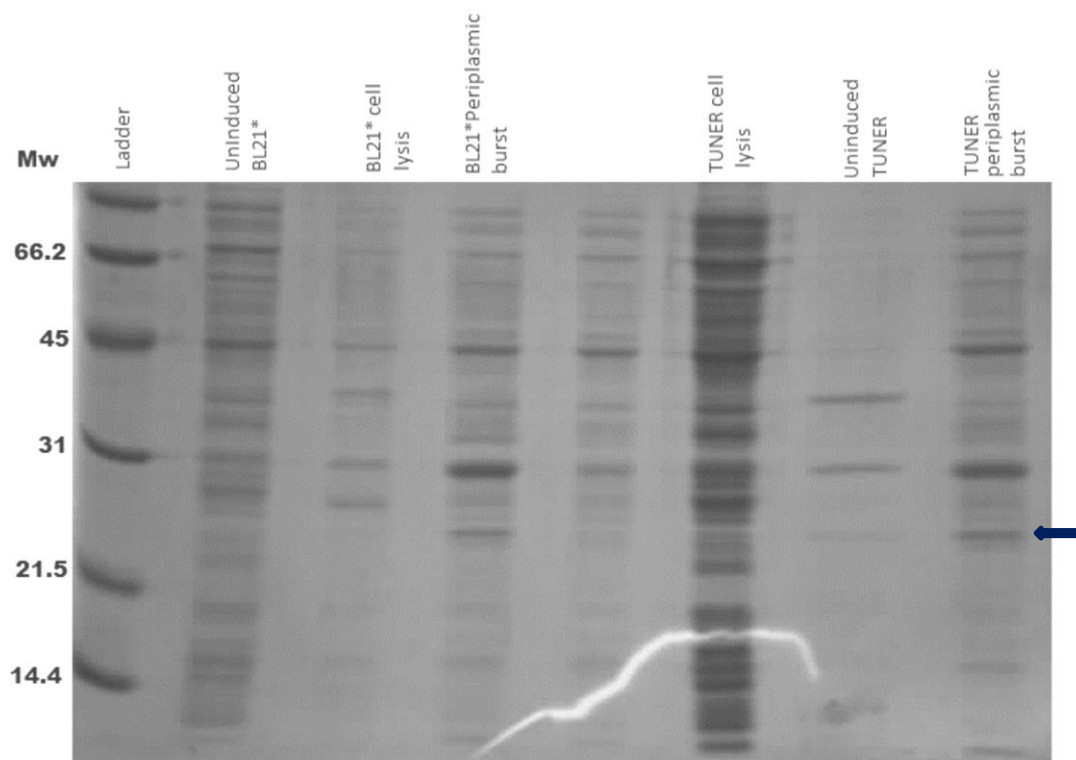
Recombinant expression of *C. cellulovorans* CBM33 from TUNER and BL21* cultures gave clear soluble bands showing that purification was a possibility from these cells. The basal expression levels within the BL21* cell cultures is lower than that in TUNER, and therefore it is reasonable to assume that purification of recombinant protein from this system would be easier. Predicted translocation of *C. cellulovorans* CBM33 to the periplasm appears successful due to the presence of the recombinant protein in the periplasmic isolated fraction. Successful translocation mechanism will improve the purification protocol by reducing contamination and should leave an active protein which runs at the correct weight on an SDS-PAGE gel. In an attempt to improve the expression yields of *C. cellulovorans* CBM33 we carried out a sequence of expression optimisation studies.

BL21* is a commonly used cell strain for the expression of recombinant proteins. Genetic manipulation leads to hyper transcription of DNA to complete saturate the translation machinery. The physiology of the cells have been altered to maximise the efficiency of polymerase activity, mRNA stability and minimise RNA breakdown [46]. A further deletion of the LON and ompT protease genes reduces protein recycling. [47-49] In order to maximise recombinant protein expression the transcription levels of T7 promoter modules is much higher than unaltered cells which means that upon induction of the Lac operon genes by IPTG the expression of T7 polymerase will be higher and therefore the transcription of the recombinant gene will be more efficient and the final concentration of the protein will be higher relative to the cell content. Modifications to catabolic genes of BL21* greatly reduces the levels of basal protein expression. Further truncation of the RNAase gene reduces the rate of mRNA cleavage and therefore saturates the ribosomes for optimal translation. [49] As a further protection to recombinant protein yields the competent cells have a resistance to phage T1 (*fhu*A2), which ensures healthy cell growth.

Induction optimisation of PelB *C. cellulovorans* CBM33

The initial protein yield from BL21* was low and in order to optimize the expression of the *C. cellulovorans* CBM33 gene, an experiment was devised to study how the time of induction with respect to culture cell density affects the soluble protein yield. IPTG induction can be affected by different factors from uptake rates, culture cell density, growth rate and the concentration maintenance inside the cell. By tuning the time of expression induction the expression could be increased to optimize the recombinant

protein yield. Small scale expression cultures were prepared identically and grown at 37°C to OD_{600} 0.4 before the temperature was reduced to 16°C. One set of cultures was inoculated with IPTG to induce recombinant gene expression at an optical density (OD_{600}) 0.8 (early) and another at 1.5 (late). In order to compare the yields of the two sample the cells were harvested and the protein isolate analysed by SDS-PAGE gel.
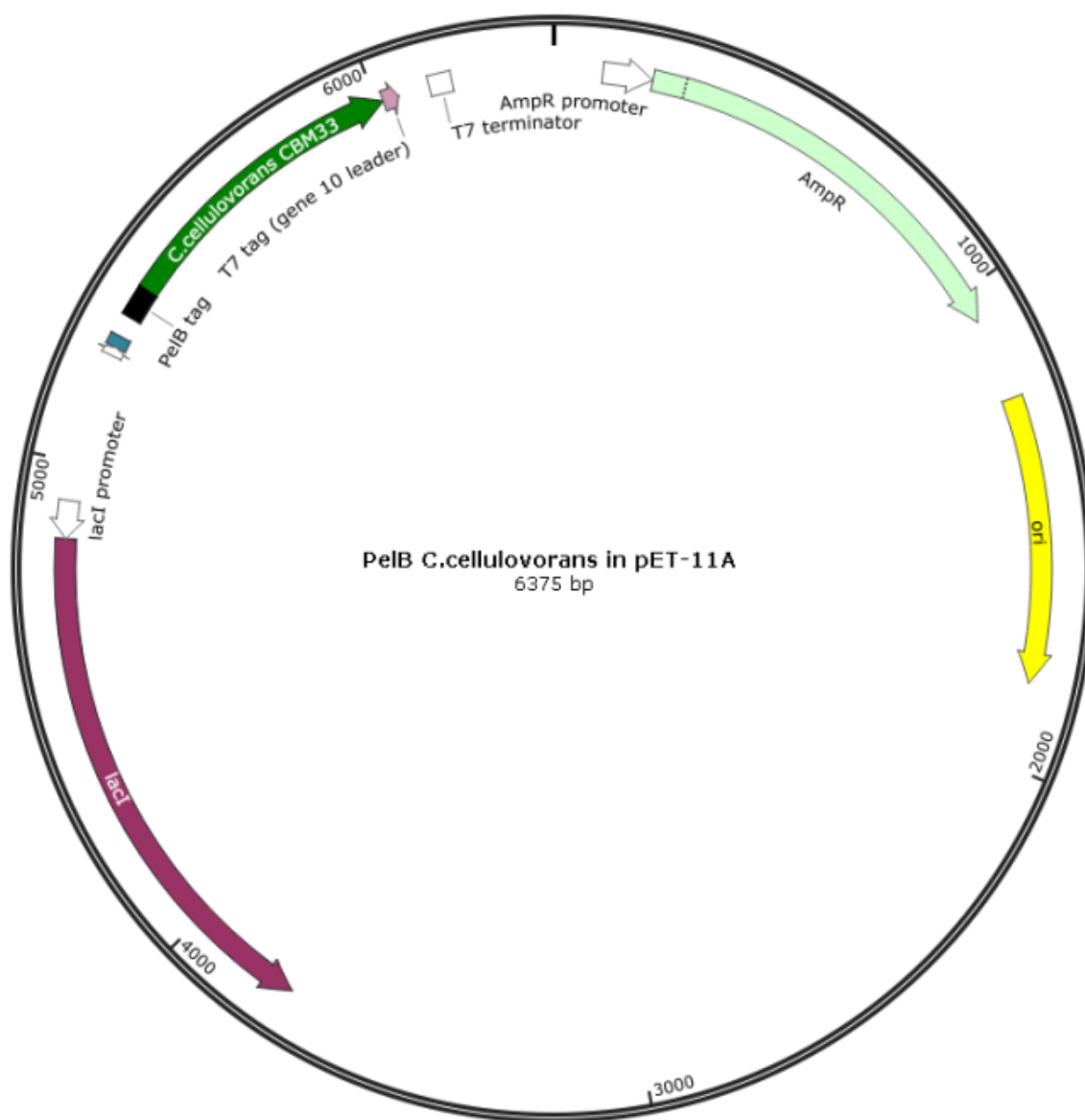
Figure 2.4 A vector map of pET-11A containing a PelB tagged C. cellulovorans.
Figure 2.4 shows a vector map of *C. cellulovorans* CBM33 in a pET-11A vector designed for expression from *E. coli.* Components of the cell that are essential to the optimal expression of recombinant protein, including the origin of replication (ori), Ampicillin resistance gene (AmpR), Lac I promoter, T7 promoter and the recombinant gene, are marked.
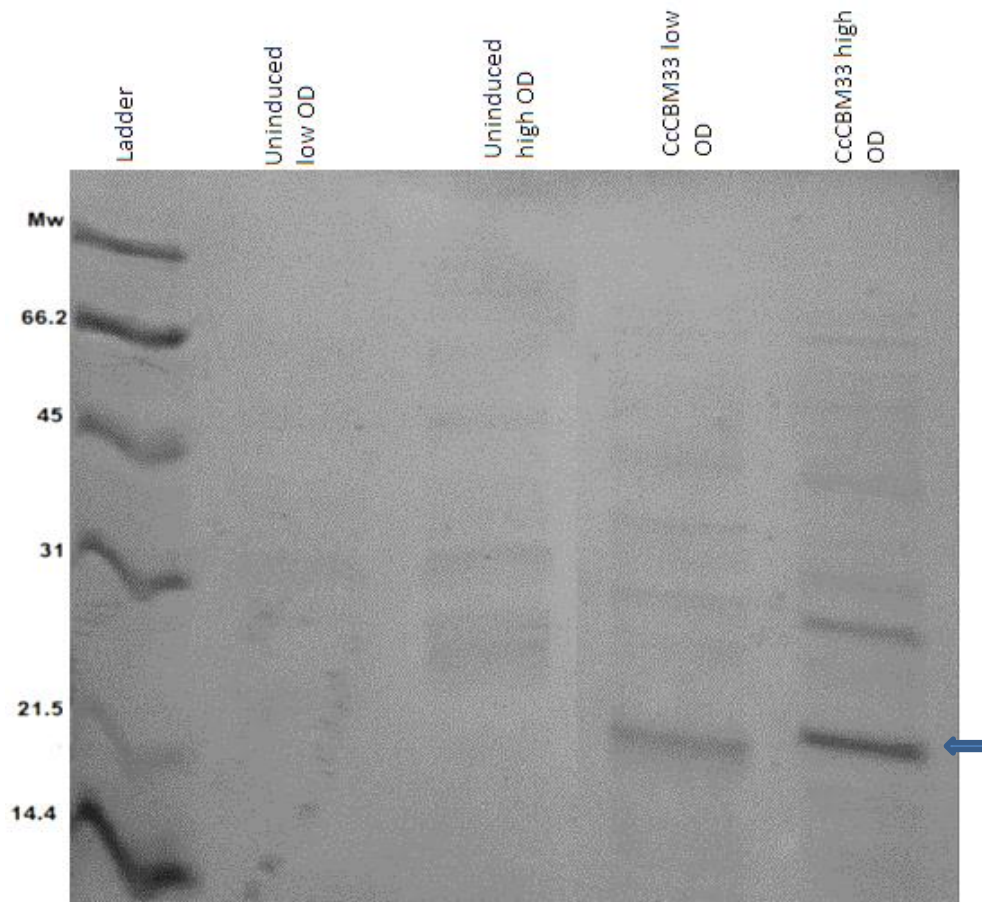
Figure 2.5 Induction optimisation
Figure showing an SDS-PAGE gel analysing the isolates from BL21* competent cell inoculated with *C. cellulovorans* CBM33 and gene expression induced using IPTG at different time points. OD_{600} 0.8 was labelled early induction and OD_{600} 1.5 late induction. *C. cellulovorans* CBM33 is marked on the gel with an arrow at the expected molecular weight.

Figure 2.5 shows that when the cultures were allowed to reach higher cell density before IPTG was added to the culture to induce recombinant gene expression the protein band increased in density. Considering the increase in protein yield seen in small scale expression tests cells were induced at OD_{600} 1.5 for all future experiments.

Purification of *C. cellulovorans* CBM33

The periplasmic isolate from the BL21* cells was impure and required purification to allow for successful characterisation. If it were possible to design an efficient purification protocol to isolate pure recombinant *C. cellulovorans* CBM33 then characterization of the protein would be possible and problems caused by poor expression would be negated.

27

Certain physiological properties of *C. cellulovorans* CBM33 restrict typical purification methods. Usual purification techniques, using Histidine tag purification by nickel affinity is not possible as the poly-histidine region would bind copper molecules and inhibit any binding studies. Alternative N-terminal purification tags have been disregarded due to the structural and functional importance of the N-terminal histidine residue. The majority of proteases available for the cleavage of a purification tag exhibit non-specific cleavage and possible expression artefacts, which, attached to the N-terminal histidine are predicted as disrupting the folding and activity of *C. cellulovorans* CBM33 making the construct useless. C-terminal tags were considered, however, again the inability to cleave the tag either at all or in a site specific manner could not be ensured.

A selection of purification techniques was tested in an attempt to purify *C. cellulovorans* CBM33.
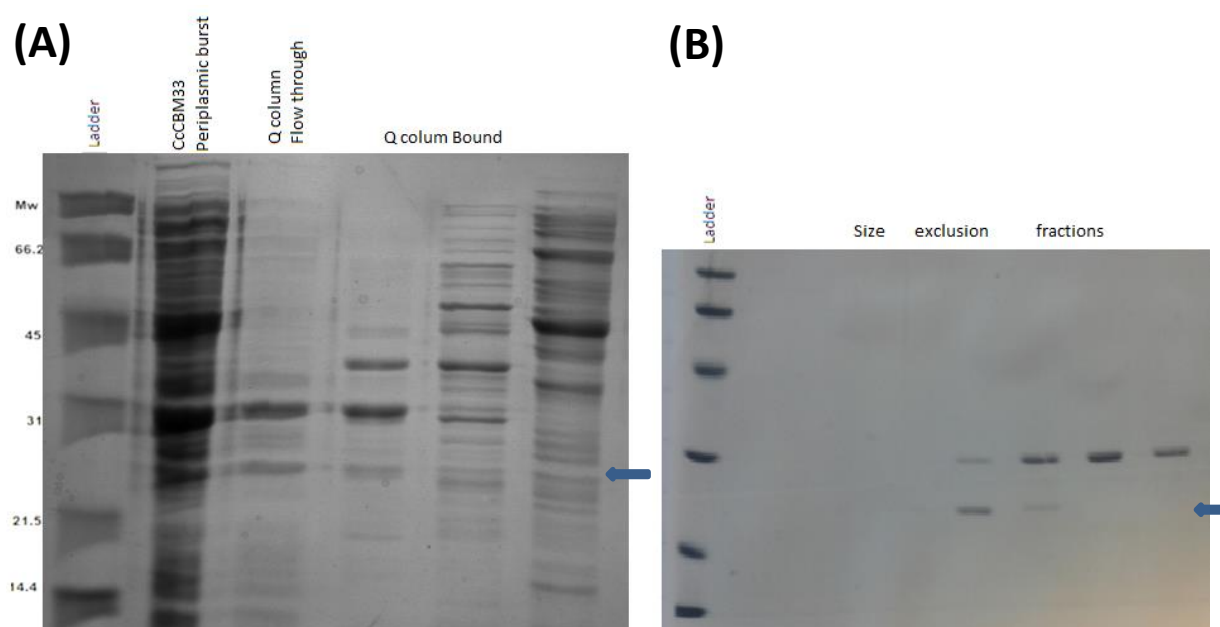
Figure 2.6 Sequential *C. cellulovorans* purification.
Figure showing SDS-PAGE gels analysing Sequential purification of *C. cellulovorans* CBM33 by (A) Q column – anion exchange and (B) Gel filtration (S75). *C. cellulovorans* CBM33 did not bind with the Q column and was therefore present in the flow through (Blue arrow on both gels). Size exclusion chromatography purified the recombinant protein apart from a contaminant band.

Figure 2.6 shows results of the sequential purification of *C. cellulovorans* CBM33 by anion exchange and gel filtration influenced by the existing purification protocol of *Bacillus amyloliquefaciens* CBM33. [50] *C. cellulovorans* CBM33 did not bind to the Q column and was present in the, very dilute, flow through. At pH 8 *C. cellulovorans* CBM33 should have bound to an anion exchange column according to the Expasy predicted pI, the failure of this method made the successful purification of *C. cellulovorans* CBM33 in good yields unlikely. The results in Figure 2.6 (B) shows the gel filtration results, the protein was present, but in unmeasurably low concentrations. The

possibility of increasing the volume of production cultures to increase recombinant protein yields was dismissed as a possibility as there was a contaminant protein at around 31 kDa. The presence of this contaminant prevented successful purification and characterization of *C. cellulovorans* CBM33 and other purification methods were explored.

Purification protocol optimisation

Ammonium sulphate precipitation

In an attempt to improve the yield of *C. cellulovorans* CBM33 from purification techniques we attempted to precipitate contaminants from the solution by addition of ammonium sulphate. The ammonium sulphate concentration in the protein solution was raised step-wise and the precipitant analysed. Recombinant *C. cellulovorans* CBM33 protein precipitated between 2.0 M and 2.5 M final concentration of ammonium sulphate. (Results not shown due to gel distortion in the presence of ammonium sulphate) Whilst this technique did purify the recombinant *C. cellulovorans* CBM33 from some contaminant proteins that precipitated in lower ammonium sulphate concentrations, the sample was not pure. Furthermore the large contaminant band observed in earlier purification steps remained.

Excess ammonium sulphate, an artefact from the purification method, inhibited downstream purification steps and analysis by SDS-PAGE gel. Repeat washing steps to remove the excess decreased yields of protein. Precipitation of recombinant *C. cellulovorans* CBM33 by ammonium sulfate precipitation was dismissed as a purification step.

Hydrophobic interaction chromatography

The hydrophobic interactions of proteins are usually mutually exclusive from the anionic interaction characteristics, therefore as an alternative or further purification method Hydrophobic Interaction Chromatography was tested.

The flow through obtained from a Q column, containing *C. cellulovorans* CBM33, was loaded onto a Phenyl sepharose HIC column equilibrated in a 50 mM Tris pH 8.0, 1 M ammonium sulphate buffer.
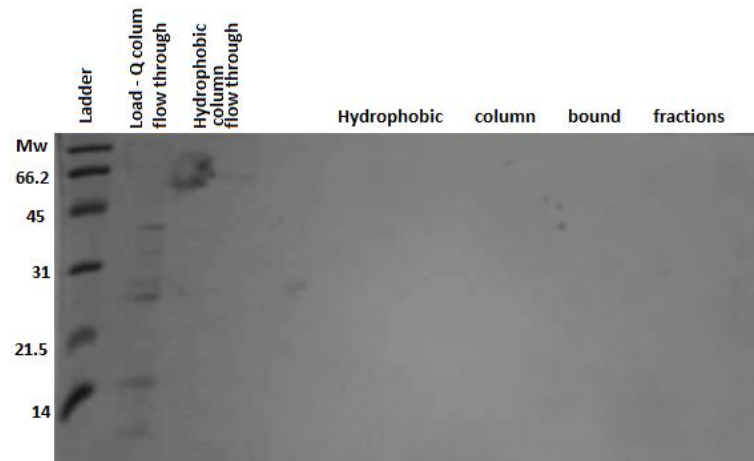
Figure 2.7 *Caldibacillus Cellulovorans HIC purification*

Figure showing an SDS-PAGE gel analysing fractions from a Hydrophobic Interaction Chromatography column. The protein load was a dilute and impure sample of *C. cellulovorans* CBM33, purification by HIC did not yield a workable sample of *recombinant protein.* There is no visible band on the gel showing *C. cellulovorans* CBM33 either due to the low concentration of protein or the high concentration of ammonium sulphate in the elution buffer.

There were no visible protein bands on the SDS-PAGE gel analyzing fractions of the hydrophobic gel, either due to ammonium sulphate interference with the gel or lack of protein yield (Figure 2.7).

The elution buffer at this stage of purification contained a large concentration of ammonium sulphate, attempted removal by buffer exchange led to loss of the entire *C. cellulovorans* CBM33 enzyme. Analysis of hydrophobic column products suggested that the HIC purification step had a negative effect on the final protein yield and therefore was disregarded as a purification possibility.

Protein identification of the purified protein bands

Two proteins co-purified by the designed purification protocol; (Figure 2.6-B) the protein with Mw= 23377 Dalton is hypothesised as being *C. cellulovorans* CBM33 but the identity of the contaminant protein is unknown. It is difficult to understand the co-purification with *C. cellulovorans* CBM33 despite the difference in size and without further structural and functional understanding it is challenging to design a purification process to separate the two. Both proteins were sent to the technology facility at the University of York for protein identification by standard mass spectrometry techniques.

A mascot database search against the NCBInr database showed that the protein with Mw= 23377 Dalton had sequence identity with a multi-domain beta-1,4-mannanase precursor from *Caldibacillus cellulovorans* and the contaminant band is β-lactamase.

These results showed that we had co-purified *C. cellulovorans* CBM33 and β- lactamase, a purification artefact that was incorporated into the PelB *C. cellulovorans* CBM33 (pET-11a) vector as resistance selection. β- lactamase selectively deactivates certain penicillin

based antibiotics [51] and was present in our system due to induction by the antibiotic ampicillin.

The contaminant protein was identified as a protein essential for cell growth in the PelB *C. cellulovorans* CBM33 gene construct. Resistance selection is essential in molecular biology to ensure the growth of only the correct construct and therefore it will be impossible to remove β- lactamase from this expression system.

Attempted purification results for PelB *C. cellulovorans* CBM33 were unsuccessful and other constructs were considered. In order to increase the likelihood of successful purification, multiple plasmids containing the *C. cellulovorans* CBM33 gene were designed.

1. PelB tagged *C. cellulovorans* CBM33 gene in an YSBL-LIC vector. This construct has the same periplasmic expression mechanism as the previous construct as results showed that this acted as a successful, natural, purification step. The YSBL-LIC vector, a pET-28(non His tagged) derivative, does not contain the gene encoding for β-lactamase and therefore utilizing the same purification protocol it should be possible to purify *C. cellulovorans* CBM33.

2. Intracellular *C. cellulovorans* CBM33 gene in an YSBL-LIC vector. This construct was designed to address the issue of poor expression. It is possible that inefficient translocation or tag cleavage during PelB *C. cellulovorans* CBM33 expression was reducing the yield of soluble protein. If this were the case an Intracellular expression system would produce a greater yield of *C. cellulovorans* CBM33.

3. SUMO *C. cellulovorans* CBM33 is a gene which encodes for the recombinant protein with an attached ubiquitin like protein that will greatly increase the solubility of the expressed protein. Within the SUMO domain is a poly-histidine tag that will allow efficient purification and better protein yields.

PelB *C. cellulovorans* CBM33 in an YSBL-LIC vector

Recombinant *C. cellulovorans* CBM33 was partially purified from pET---11a vector described above. The protein that co-purified was an expression artefact that is present to breakdown the antibiotic that we were using to select for cells with our construct. In an attempt to utilise the successful purification technique that we designed above but to by-pass the contaminant co-purification *C. cellulovorans* CBM33 was successfully cloned into a YSBL—LIC vector with kanamycin resistance gene encoding an aminoglycoside 3'-phosphotransferase (Uniprot: P00551).

We sent the synthesized vector for sequence analysis and the results confirmed that we successfully made the construct in Figure 2.8. The *C. cellulovorans* CBM33 construct with kanamycin resistance in the standard YSBL—LIC vector had the same expression mechanism as in the PelB *C. cellulovorans* CBM33 pET 11-A vector. A 22 amino acid N-terminal PelB tag for periplasmic targeted expression; using the periplasmic

recombinant protein purification protocol described above this change in vector should by pass the issue of having co-purification of *C. cellulovorans* CBM33 with β- lactamase.

Expression

The soluble periplasmic protein fraction obtained from the cell harvest was loaded onto a gel for analysis of the proteins by size. The expression levels of *C. cellulovorans* CBM33 were very similar to those observed in the PelB pET-11A construct with lower recombinant protein yields and major contaminant bands on an SDS-PAGE gel. (Results not shown due to indiscernible protein band) *C. cellulovorans* CBM33 runs at the predicted Mw on an SDS-PAGE gel which shows that the PelB leader sequence was cleaved upon translocation to the periplasm as the hypothetical mechanism predicted.
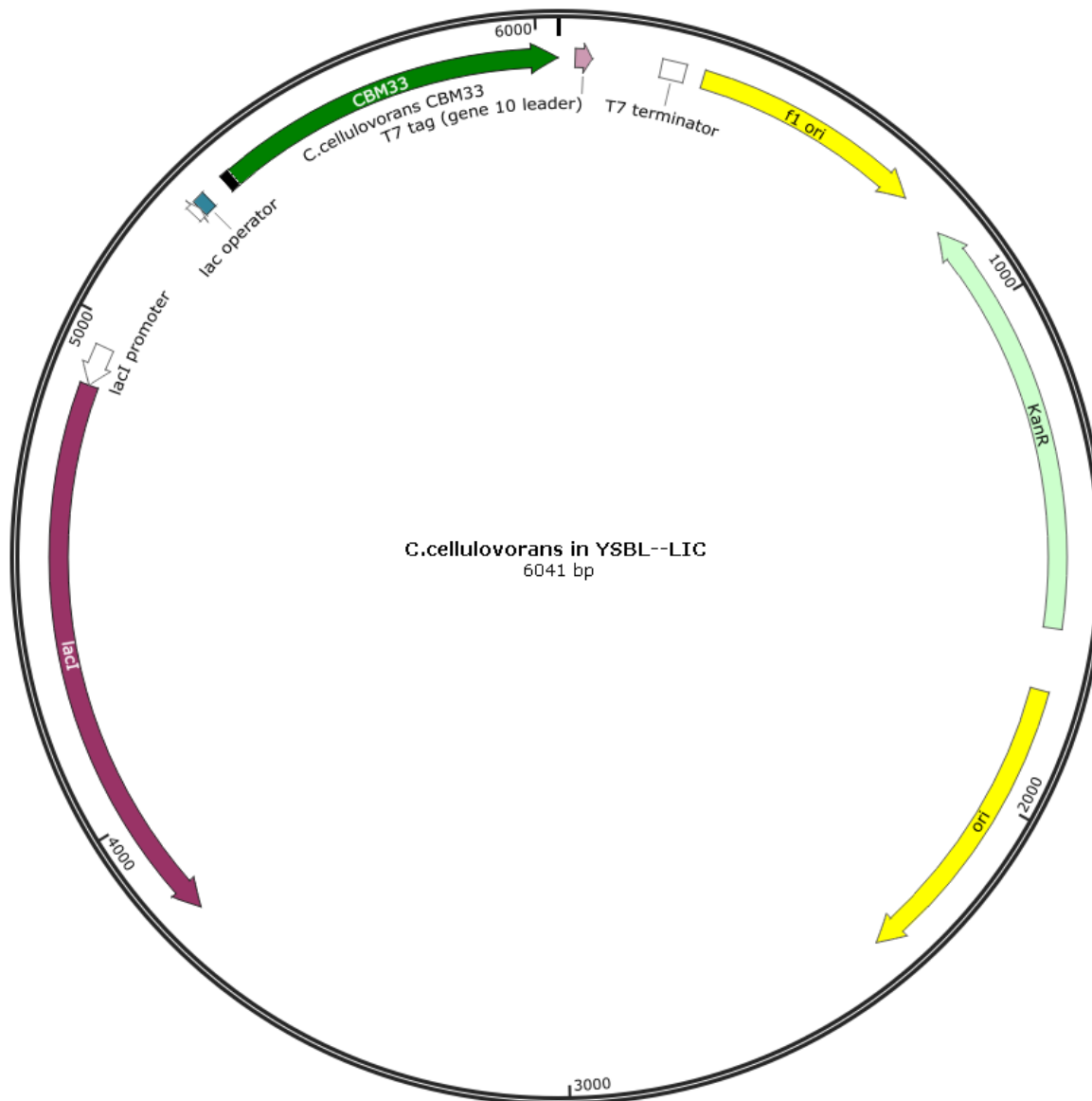
Figure 2.8 vector map PelB *C. cellulovorans* **CBM33 YSBL-LIC**
Figure showing a vector map for the PelB *C. cellulovorans* CBM33 gene. The vector (pET-28 derivative) was chosen to alter the resistance selection gene from ampicillin resistance to kanamycin. This step was taken in order to remove the contaminant protein seen in previous purification experiments.

Purification of *C. cellulovorans* CBM33 from YSBL-LIC

Based on the purification results of *C. cellulovorans* CBM33 in pET-11A, the protein was purified using the same protocol as the contaminant protein that co-purified previously would not be present.

The protein isolate was purified by sequential purification by Q-column and gel filtration both attached to a Fast Performance Liquid Chromatography AKTA and the fractions analysed by SDS-PAGE gel (Figure 2.9). The yield of purified *C. cellulovorans* CBM33 was negligible, protein bands seen in Figure 2.9 (B) were concentrated by acetone precipitation before they were visible on an SDS-PAGE gel.

Design of the construct to remove β-lactamase as a co-purification artefact was successful and a pure *C. cellulovorans* CBM33 sample was obtained from this construct. However, continued poor expression meant that the construct was not useful for purification.
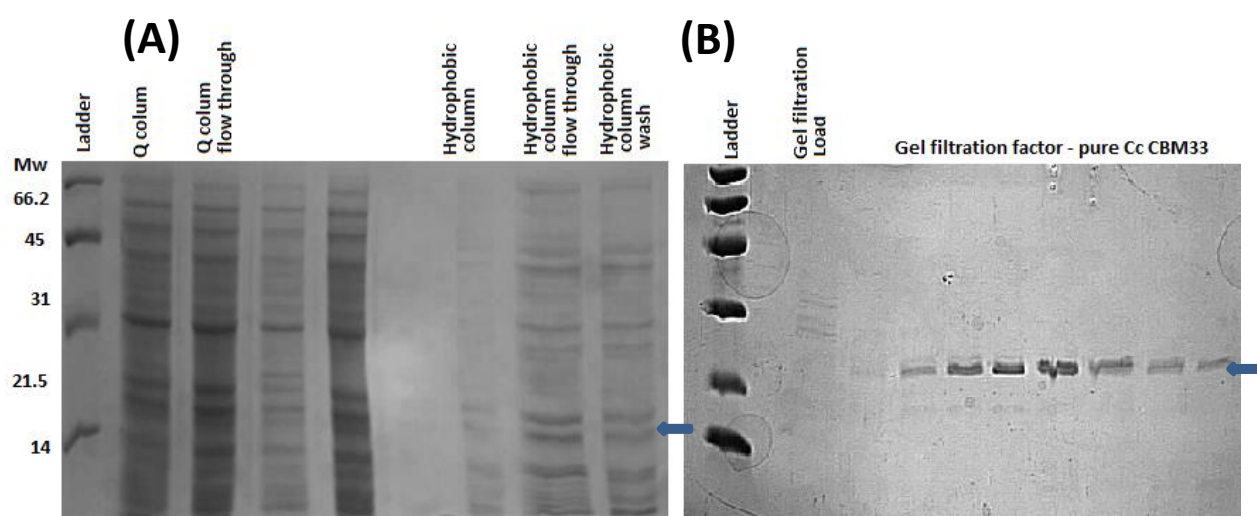
Figure 2.9 Sequential C. cellulovorans purification
Figure showing an SDS-PAGE gel showing the sequential purification of PelB *C. cellulovorans* CBM33 by (A) Q column anion exchange chromatography (Lanes 2-6), HIC (Lanes 7-9) and (B) Gel filtration.

It was interesting to note that results suggested the expression of PelB *C. cellulovorans* CBM33 constructs in small scale expression tests was higher than in the scaled up cultures (comparison of small scale expression in Figure 2.4 and large scale expression in Figure 2.6). This phenomenon was observed for *C. cellulovorans* CBM33 expressed in both the YSBL-LIC and pET-11a vectors. The original small scale (5 ml) expression tests of *C. cellulovorans* CBM33 carried out by Glyn Hemsworth, from Davies group, suggested the periplasmic expression would be strong with a clearly visible band on SDS-PAGE gel however the larger scale cultures (500 mls) did not give workable yields. These results suggested that the PelB tag attached to the N-terminal residue of *C. cellulovorans* CBM33 could be inhibiting the soluble expression as suggested previously, decreasing protein yields.

Intracellular *C. cellulovorans* CBM33 in an YSBL—LIC vector

In an experiment designed to test whether a *C. cellulovorans* CBM33 construct without an N-terminal expression tag would display higher expression; primers were designed for a *C. cellulovorans* CBM33 construct without a tag which would be targeted for cytoplasmic expression (Figure 2.10).

Expression of Intracellular *C. cellulovorans* CBM33 vector

When deciding which expression strain would provide the optimal environment for expression of intracellular *C. cellulovorans* CBM33; the protein identification data from PelB *C. cellulovorans* CBM33 in pET-11A construct were considered. The Mw of this protein was 2 Da lower than the predicted showing the assumed formation of one di-sulfide bond in the folded protein.
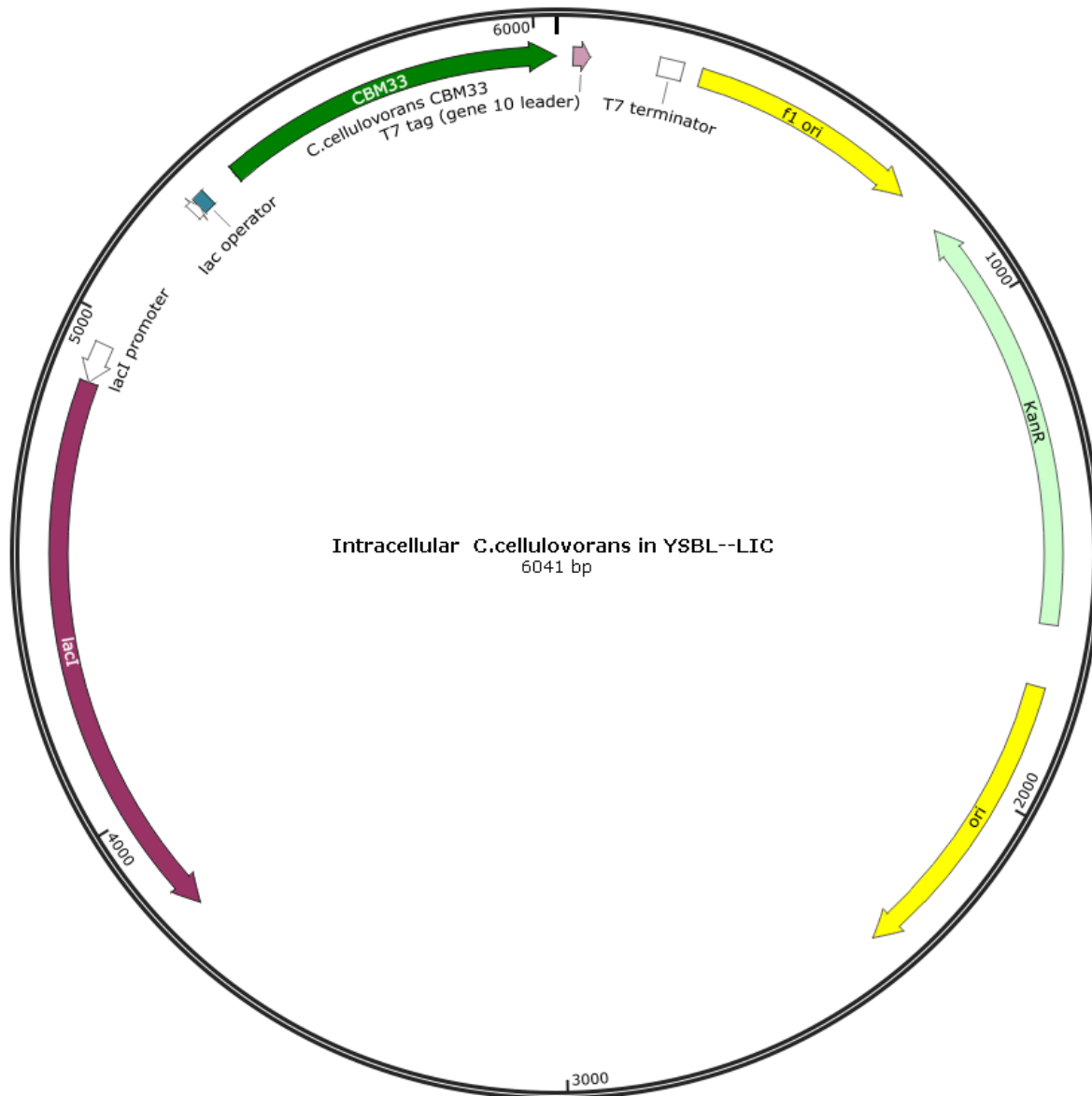
Figure 2.10 Vector map Intracellular *C. cellulovorans CBM33 YSBL-LIC*
Figure showing the vector map for the Intracellular *C. cellulovorans* CBM33 gene in YSBL-LIC. The gene is designed for expression of recombinant in the cystosol of an *E. coli* expression strain so the PelB leader sequence was removed in cloning.

Di-sulphide bridges are bonds formed between two cysteine residues that are distal in the primary sequence of the protein. These intra-molecular bonds are important and in some cases essential for protein structure, folding, stability and therefore activity. SHuffle *E.coli* cells from New England Biolabs promote the formation of di-sulphide bonds within the cytoplasm and were therefore selected as the expression system for this vector to increase stability and solubility of intracellular *C. cellulovorans* CBM33.

The cell components were isolated by centrifugation and both the soluble protein and insoluble fractions were analyzed by SDS-PAGE gel (Figure 2.11). There is clearly a

protein that is present in the soluble fraction that is induced by the addition of IPTG of an Mw that corresponds with *C. cellulovorans* CBM33. The expression of the intracellular *C. cellulovorans* CBM33 protein in these small scale cultures is poor.

Scaling up the growth cultures to 500 ml flasks produced a negligible band of recombinant protein, near impossible to reach a reasonable concentration for downstream purification. When considering how low protein purification yields have been observed in other constructs this construct was not considered as a sensible avenue for experimentation. The intensity of the *C. cellulovorans* CBM33 protein band on SDS gel (Figure 2.11) was similar to the levels observed in periplasmic preparations. The hypothesis that expression would be increased in comparison to the PelB constructs was disproven.
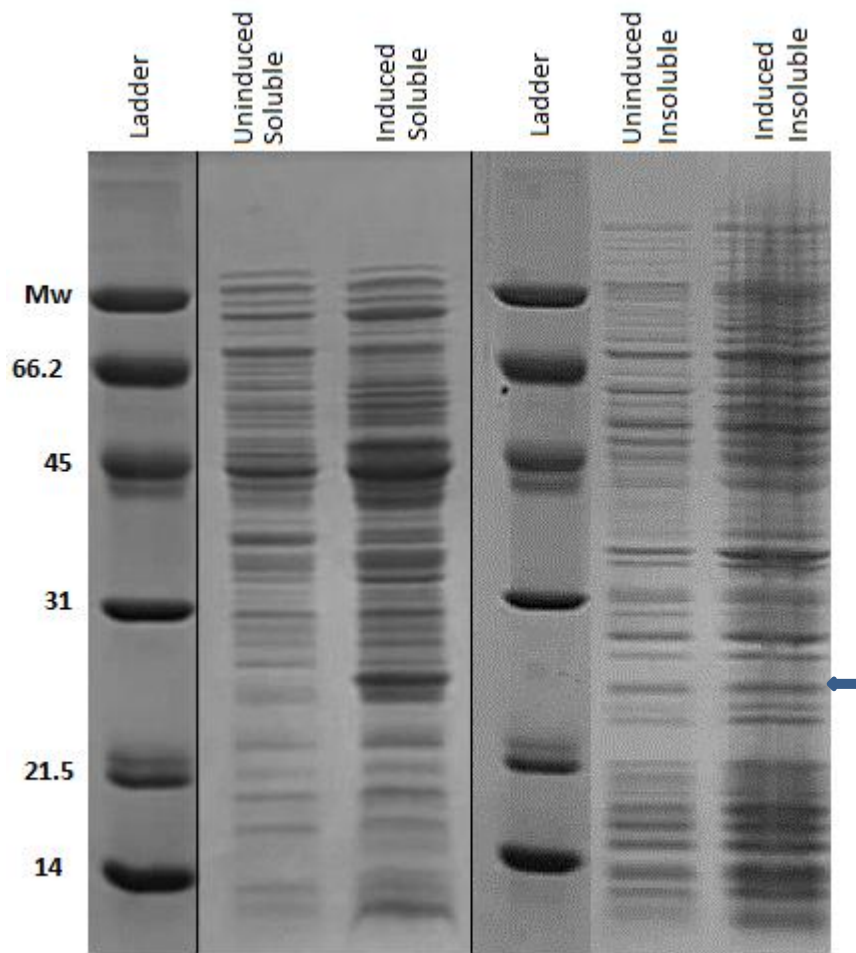
Figure 2.11 Intracellular *C. cellulovorans* expression test
Figure showing an SDS-PAGE gel analysing the insoluble and soluble expression of the Intracellular *C. cellulovorans* CBM33 gene from Shuffle cells.

2.3 SUMO *C. cellulovorans* CBM33 in a Champion pET—SUMO vector

Purification of *C. cellulovorans* CBM33, from both periplasmic isolation and native intracellular isolation, produced unworkably low yields of recombinant protein; at this stage the largest barrier to efficient purification is the low yield of soluble *C. cellulovorans* CBM33. In order to meet the aims of the project and characterize *C. cellulovorans* CBM33, soluble expression must be increased.

The field of molecular biology has utilised the attachment of small molecule tags to recombinant proteins in order to increase soluble expression. An example is the cloning of small ubiquitin-like modifier (SUMO) proteins into recombinant expression vectors to increase the solubility of the attached protein domain. SUMO proteins contain an ubiquitin like fold which plays a role in a variety of natural expression systems but also

38

creates a very stable fusion construct with recombinant proteins which increases the expression stability and therefore the soluble protein yield. [52]

We cloned *C. cellulovorans* CBM33 into an expression system with a SUMO construct attached to the N-terminus in an attempt to increase soluble expression. Primer design allowed for exact design of *C. cellulovorans* CBM33 vector with physiological traits improving expression (see Figure 5.4). Further to this, the SUMO domain attached to the enzyme contained al poly-Histidine region at the N-terminal. These positively charged residues would allow for efficient purification by allowing the recombinant protein to bind to a nickel column and purify from the cell isolate.

Whilst successful purification of recombinant protein was currently the inhibiting factor in our characterisation of *C. cellulovorans* CBM33 and the opportunities to purify by nickel affinity chromatography appeared to solve this; it was still imperative that any future studies be carried out on *C. cellulovorans* CBM33 in the native form. Therefore, it was essential that it was possible to guarantee specific cleavage of the SUMO domain to leave the active conformer of *C. cellulovorans* CBM33. SUMO *C. cellulovorans* CBM33 gene was encoded so that the N-terminal histidine forms a peptide bond, linking the SUMO tag to *C. cellulovorans* CBM33 but also preventing copper binding.

SUMO protease, a commercially available cysteine protease which identifies the proteolytic site through structural recognition, allows for specific cleavage to leave to leave the N-terminus of the protein as it would be in the native system. The protease has been shown to recognise the cleavage site in a site specific manner dependant on the formation of salt bridges between the protease molecule and the SUMO domain attached to the recombinant protein. [53]

Small scale expression tests of SUMO *C. cellulovorans* CBM33

The protein isolated from small scale cultures was analysed by SDS-PAGE gel (Figure 2.15). A band of protein appears on an SDS-PAGE gel at Molecular weight (Mw= 36778 Da) which is the correct weight for the predicted SUMO tagged moiety. The band is not present in the un-induced cells showing that the protein is induced by the IPTG treatment during growth and the majority of the induced protein is present in the soluble fraction. Expression of *C. cellulovorans* CBM33 from the SUMO construct was strong and the bound SUMO domain made the protein soluble in solution for downstream purification. Results show that intelligent design of the vector allowed for an increase in soluble protein.
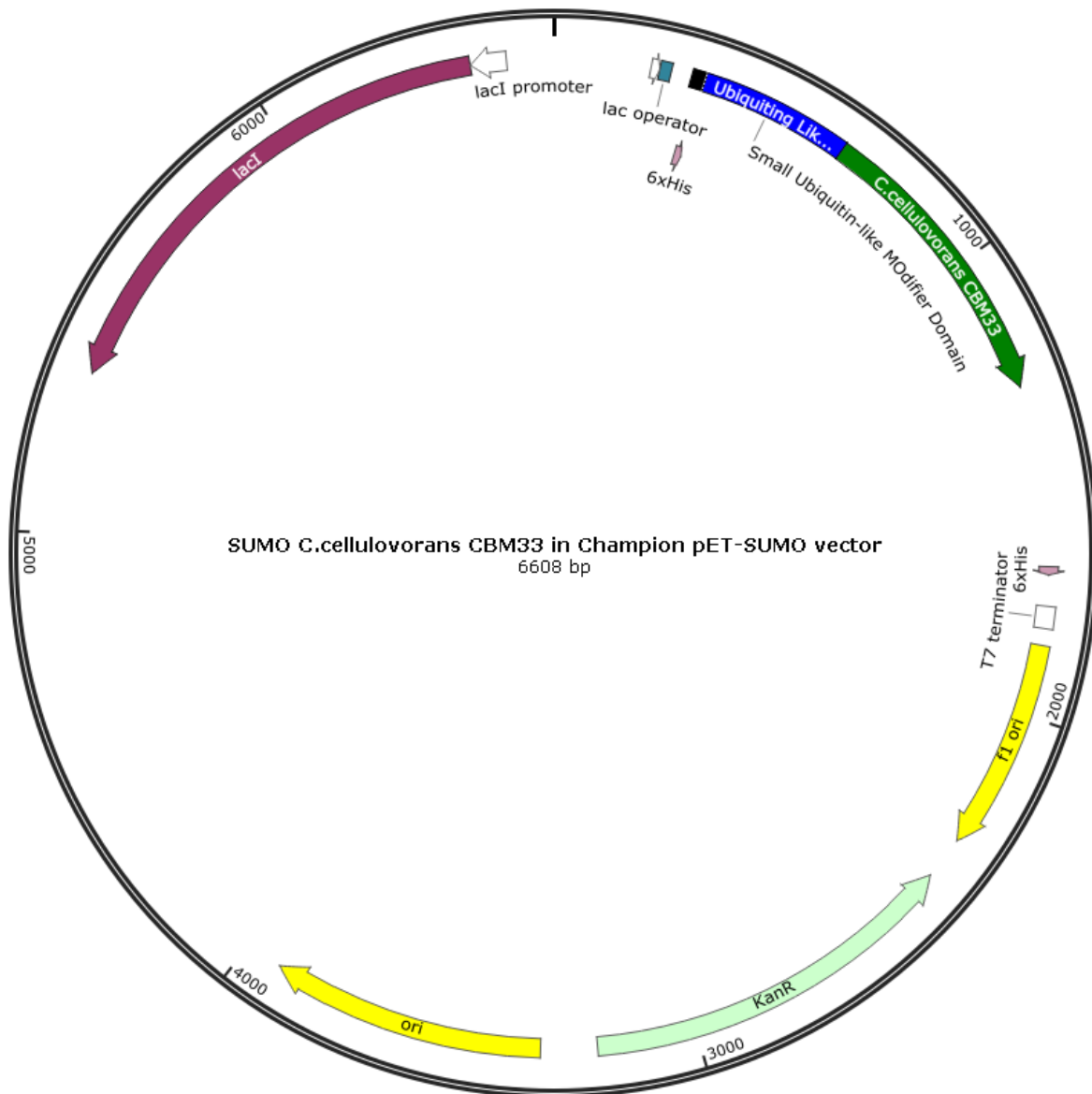
Figure 2.12 Vector map SUMO *C. cellulovorans* CBM33 Champion pET-SUMO
Figure showing the vector map for the SUMO *C. cellulovorans* CBM33 gene in Champion pET-SUMO. In order to increase the ease of downstream purification and the solubility of expressed recombinant protein led to an insertion of a SUMO moiety onto the N-terminal of the recombinant protein. The protein, solubility tag, has an attached poly-histidine tag for optimization of purification.
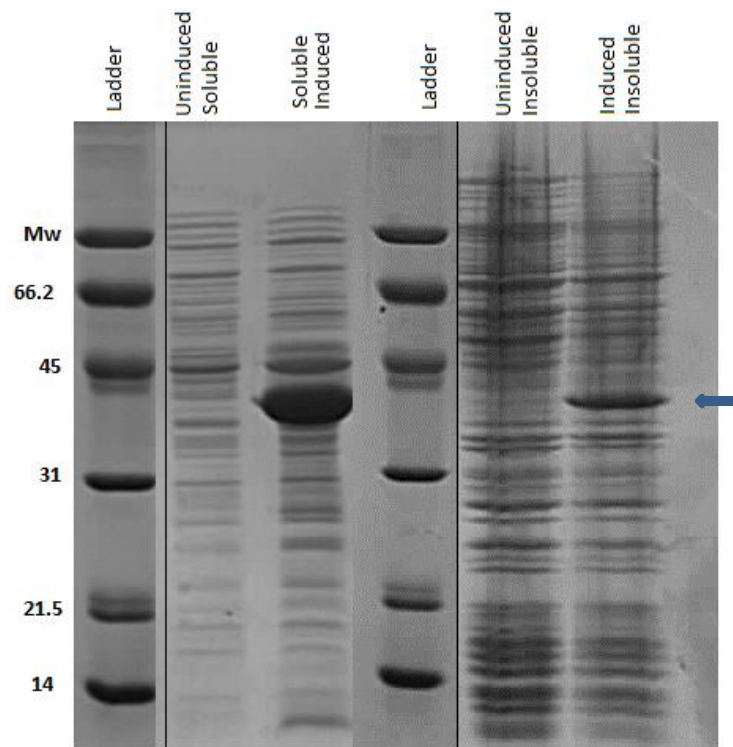
Figure 2.13 SUMO C. cellulovorans CBM33 expression test
SDS-PAGE gel analysis of SUMO *C. cellulovorans* CBM33 expression from SHuffle cells. The SUMO tagged recombinant protein has a Mw of 36KDa and is present in good yields in the soluble expression test.

Purification of SUMO *C. cellulovorans* CBM33

The first stage for purification of *C. cellulovorans* CBM33 from the SUMO construct involved loading onto a Nickel resin column and the fractions collected for analysis by SDS-PAGE gel. SUMO *C. cellulovorans* CBM33 bound to the nickel column due to the poly-histidine region in the SUMO domain and only eluted in a gradient of Imidazole (Figure 2.14). Fractions A6-C6 contained pure *C. cellulovorans* CBM33 SUMO fusion, (Mw= 36778 Da) the protein was very concentrated and this suggested that the yield had the potential to be high.

41
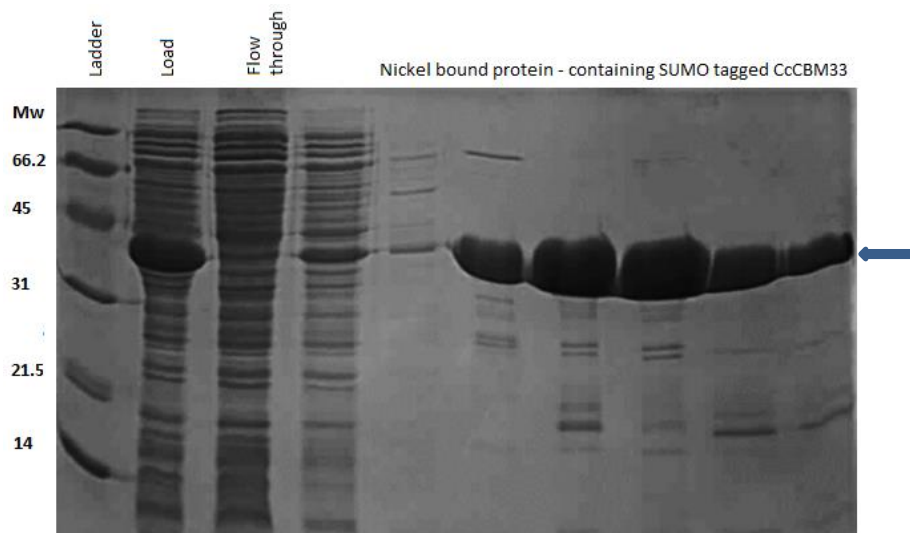
Figure 2.14 SUMO C. cellulovorans CBM33 1st Nickel column
SDS-PAGE gel analysis of SUMO *C. cellulovorans* CBM33 purification fractions by Nickel affinity chromatography. The soluble expression of the protein from the YSBL-LIC vector coupled with the strong affinity of the SUMO tag for the nickel column has led to efficient isolation of a concentrated sample of recombinant protein. Due to high concentrations of the protein in the SDS-PAGE gel load the gel has begun to bow, a sign that too much protein was present in the load.
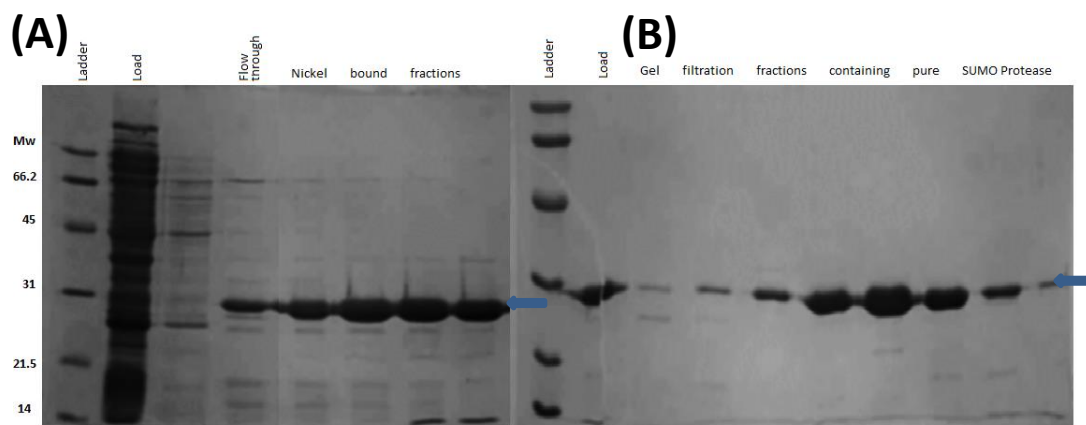
Figure 2.15 SUMO protease purification
SDS-PAGE gel analysis of SUMO Protease purification by (A) Nickel column and (B) Gel filtration.

Whilst obtaining the pure protein was a significant step forward, it was still necessary to cleave the SUMO moiety from the protein using SUMO protease in order to allow characterisation of SUMO protease.

Purification of SUMO protease

The SUMO protease vector was purchased and inserted into a pLysS expression strain, Figure 2.15 shows that two step purification of the protease by Nickel column and gel filtration gave pure protein. Protein identification of the SUMO protease sample by mass

spectrometry came back positive and the recombinant protein concentration was calculated at 4.1 mg/mL.

SUMO domain cleavage from SUMO *C. cellulovorans* CBM33

Imidazole, a known inhibitor of SUMO protease activity, [54] was removed from the solution by buffer exchange to maximise the efficiency of the cleavage reaction. 125 mg of SUMO *C. cellulovorans* CBM33 was incubated with 1.25 mg of SUMO protease overnight for 16 hours with 0.1 mM of 2-Mercaptoethanol as a reducing agent.

After the incubation of SUMO protease with *C. cellulovorans* CBM33 the reaction solution contained several products; SUMO protease, the cleaved SUMO domain, un-cleaved SUMO *C. cellulovorans* CBM33 and native *C. cellulovorans* CBM33. *C. cellulovorans* CBM33 was the only constituent of the mixture which does not contain a poly-histidine region and therefore was the only molecule that would not bind to a nickel resin column and would therefore be separated into the flow through. Results in (Figure 2.16) confirm that we successfully cleaved the SUMO domain from *C. cellulovorans* CBM33 and the removal of the His-tag meant that the native enzyme eluted in the flow through.
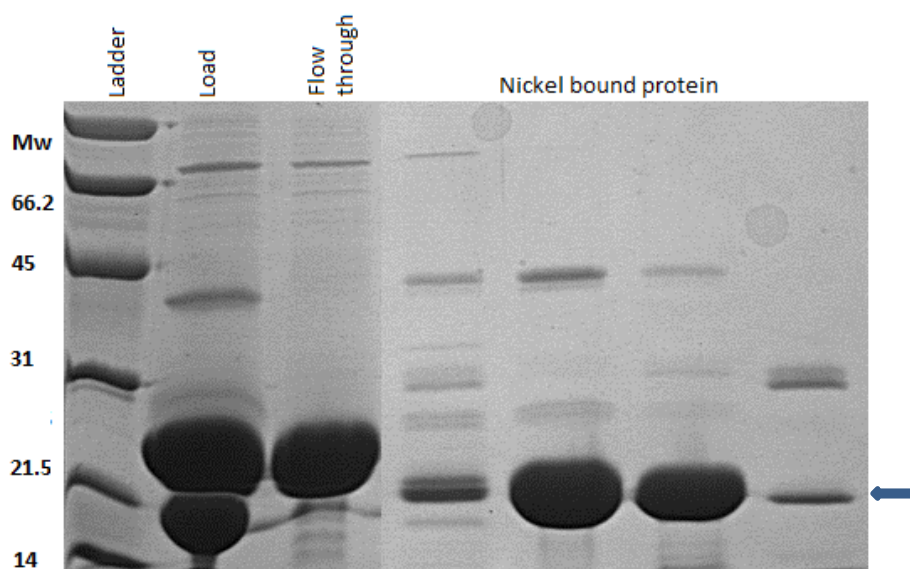
Figure 2.16 SUMO *C. cellulovorans* CBM33 2nd Nickel column
SDS-PAGE gel analysis of Nickel column separation of the cleaved SUMO moiety and SUMO protease (Lanes 4-7) from the native *C. cellulovorans* CBM33 (Lane 3). The protein load on the gel is so high that it has completely bowed but the yield and purity seem to confirm this process as an effective purification protocol.

Purification of native *C. cellulovorans* CBM33

To further purify the *C. cellulovorans* CBM33 protein the solution was concentrated by centrifuge to less than 1 ml using a VivaSpin column with a 10 KDa filter to remove buffer. The concentrated recombinant protein solution was loaded onto a Superdex S75 column for purification by gel filtration. SDS analysis showed that *C. cellulovorans* CBM33 was running pure on the gel at the expected molecular weight. Analysis of the S75 fractions showed the protein was very concentrated and pure. Due to the high concentration of protein the SDS-PAGE gel was completely overloaded and was therefore illegible and was not shown.

To confirm the identity and purity of *C. cellulovorans* CBM33 the protein was analysed by Electrospray Time of Flight mass spectrometry (TOF MS ES+) and the results show a protein present in the sample Mw=23377.9 Da (Figure 2.17). This value is 2 Da less than the predicted molecular weight calculated through ExpasyTranslate, this result reaffirms that *C. cellulovorans* CBM33 contains one disulphide bridge in the secondary structure. The peak also shows that the cleavage of the SUMO domain the cysteine protease, successfully expressed and purified above, worked efficiently and that the cleavage was specific at the N-terminal histidine.

These results showed that cloning a SUMO *C. cellulovorans* CBM33 construct into a Champion pET-SUMO vector allowed for successful purification. The intelligent vector design addressed each of the problems with purification of *C. cellulovorans* CBM33 experienced previously, optimisation of the expression and purification of *C. cellulovorans* CBM33 yield of 152 mg of pure protein.
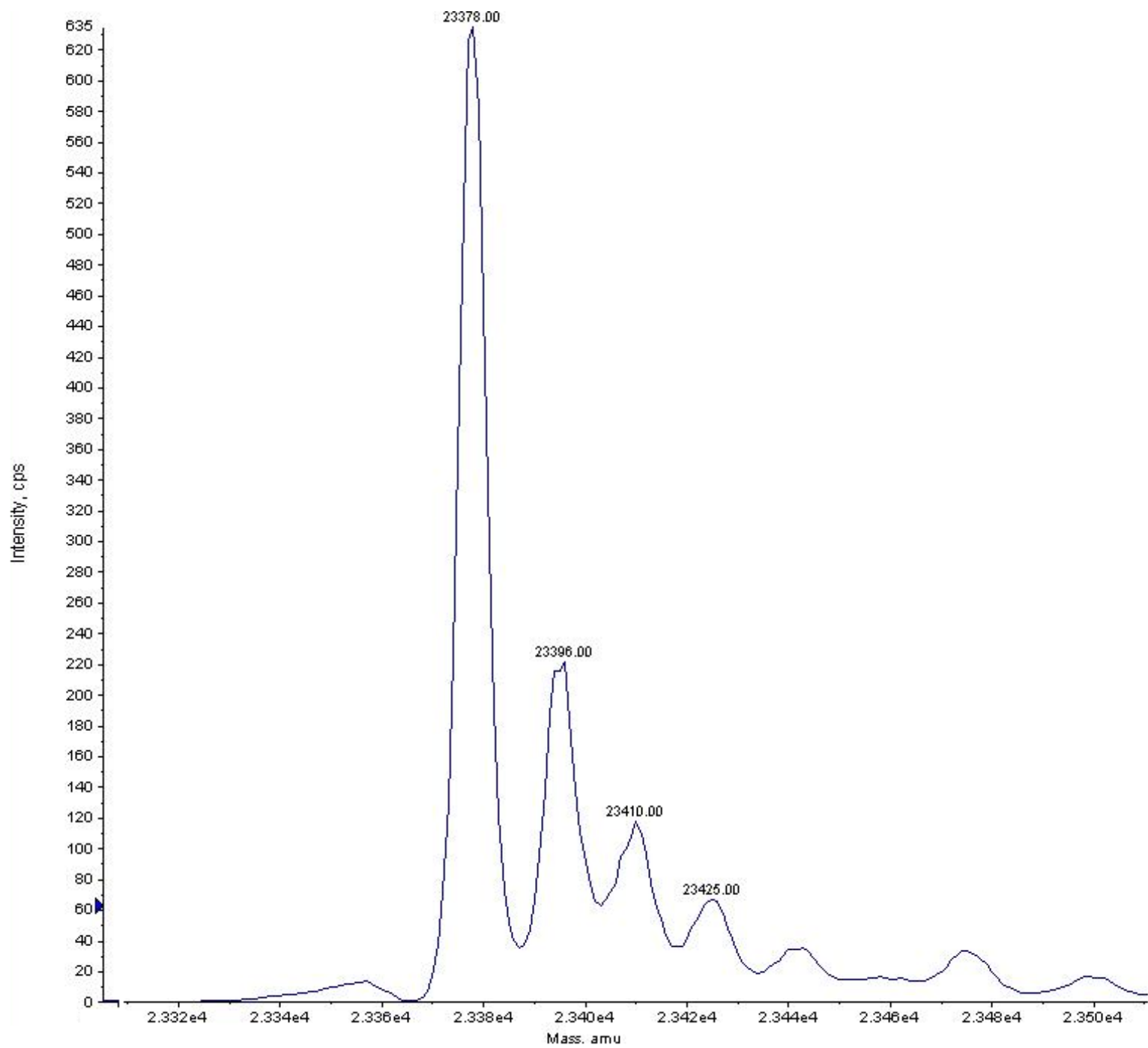
Figure 2.17 Protein ID by mass spectrometry.
Mass reconstruction using Bruker Daltronics software from Electrospray TOF analysis of pure *C. cellulovorans* CBM33. The main protein peak is at 23370 Da which matches the predicted weight of *C. cellulovorans* CBM33 minus two. This two Da decrease most likely represents the formation of a di-sulphide bond (loss of two hydrogen atoms).

Chapter 3 – characterisation of *C. cellulovorans* CBM33

C. cellulovorans CBM33 was successfully purified from Shuffle *E.coli* cells which had been transformed with and expression vector for a construct of *C. cellulovorans* CBM33 with a SUMO like domain attached to the N-terminal. The acquisition of pure protein enabled us to begin experiments to characterise the protein. Initially we attempted to show *C. cellulovorans* CBM33 is an LPMO molecule before going on to elucidate the mechanism of action.

Isothermal Calorimetry (ITC)

ITC measures the thermal transfer within a system when "ligand" molecules in solution are titrated into a solution containing the molecule being characterised. Previous LPMO characterisation experiments have indicated the requirement for copper bound in the active site for enzymatic activity. Therefore, carrying out ITC copper binding experiments on *C. cellulovorans* CBM33 by measurement of the thermal shift evolved when a high concentration of copper is titrated into a protein solution. The temperature change can be used to calculate the reaction stoichiometry and the protein binding affinity for copper in this buffer and temperature.

The experimental data for *C. cellulovorans* CBM33 gave a K_D of 4.31×10^{27} M calculated from $1/K_a$. The value obtained was below the lower detection limit of ITC and therefore regarded as unreliable and the background measurements were noisy. However, using the known concentration of protein and ligand it is possible to calculate that the stoichiometric binding ratio of *C. cellulovorans* CBM33 with copper was 0.604 at 298 K (Figure 3.1). Due to the detection limits of the ITC machine, there were no titration peaks within the range of copper saturation; therefore, assumptions based on the results were limited to the statement that *C. cellulovorans* CBM33 binds copper very tightly.

Despite purification of *C. cellulovorans* CBM33 focusing on obtaining pure protein in the apo form, without metal bound in the active site, the experimental stoichiometric ratio of copper bound to *C. cellulovorans* CBM33 is lower than the predicted value. Other LPMO characterisation experiments have obtained a similarly low stoichiometric value; from existing LPMO copper binding data, *C. cellulovorans* CBM33 should bind to copper in an equi-molar ratio. Sub stoichiometric binding ratios were also observed in ITC characterisation of CBM33 from *Bacillus amyloliquefaciens.* [50] *Both C. cellulovorans* CBM33 and *Bacillus amyloliquefaciens CBM33* have a high affinity for copper, shown by the tight binding characteristics in ITC experiments, and therefore, if the binding ratio were 1:1 then it should have been reflected in the ITC results. X-ray crystallography and EPR studies on *Bacillus amyloliquefaciens CBM33* went on to confirm a mono-copper binding site.

Having considered the evidence, it was assumed that *C. cellulovorans* CBM33 was a mono-copper centre. The low stoichiometry could have been explained by suggested copper artefact binding from the purification or experimental equipment. If this copper was bound to *C. cellulovorans* CBM33 tightly then it may not have been removed by treatment with EDTA and therefore this could have altered the ITC results.
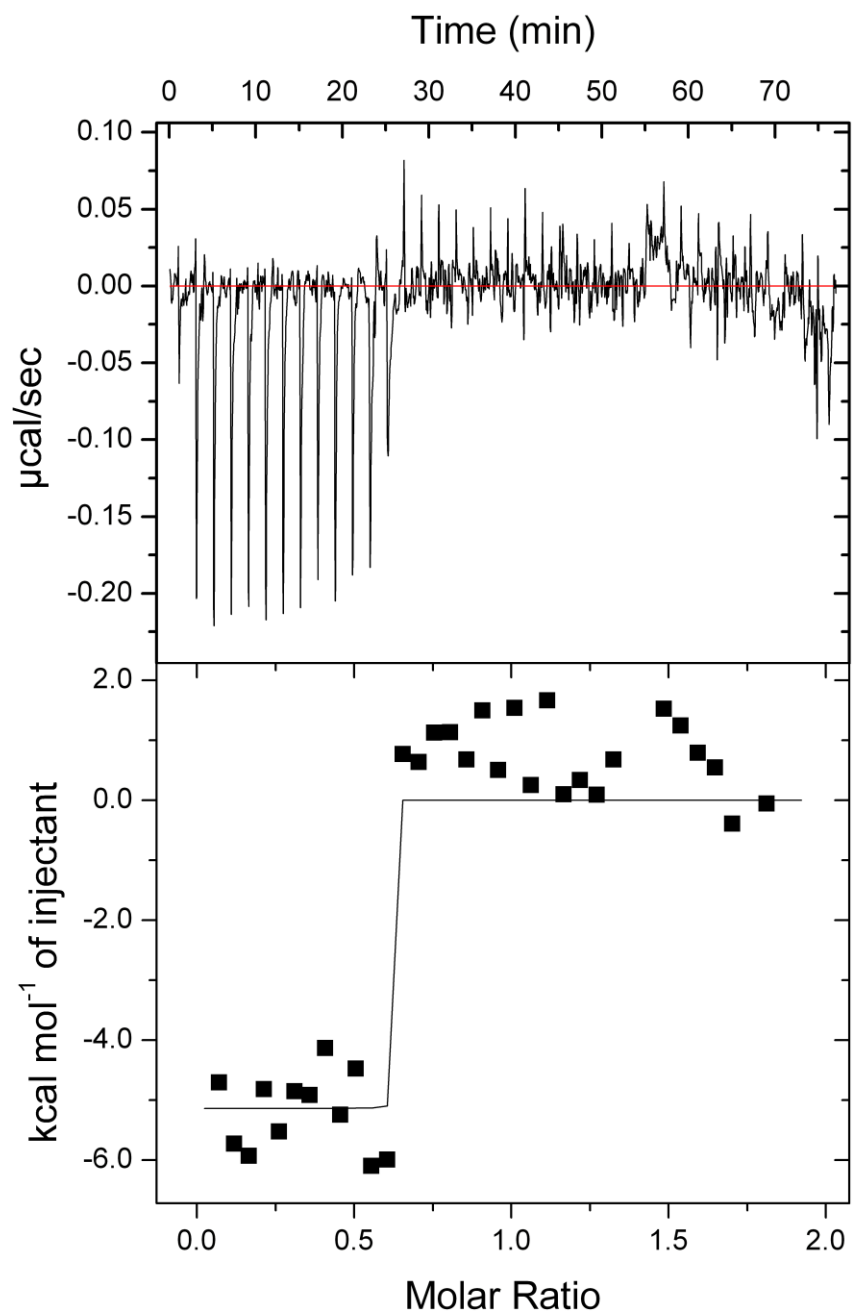
Crystallisation screening *C. cellulovorans* CBM33

In an attempt to investigate the structure of *C. cellulovorans* CBM33 further crystal screens were set up with *C. cellulovorans* CBM33 purified to 6.7 mg/mL from the SUMO construct. This experiment was designed to elucidate the three dimensional structure of the protein and resolve the copper stoichiometry. INDEX and PACT plates were set up in copper loaded and apo conditions and crystals allowed time to grow. No protein crystals grew in the conditions tested.

Electro paramagnetic resonance experiments on *C. cellulovorans* CBM33
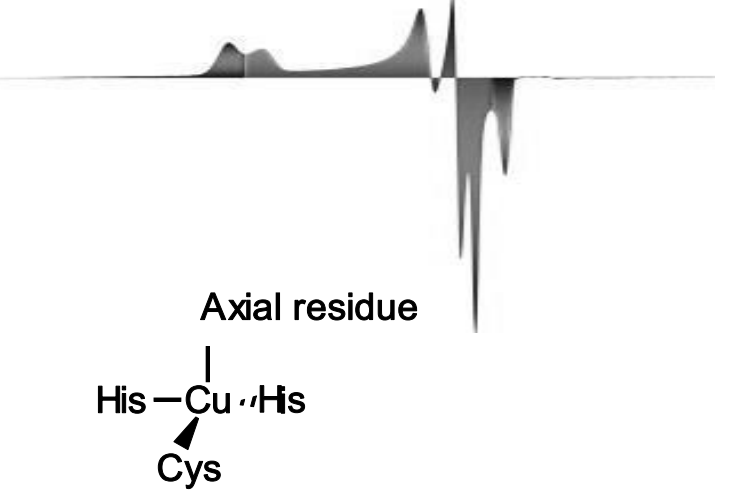
Attempts to structurally characterize *C. cellulovorans* CBM33 through traditional crystallography methods failed within the project timeframe. Therefore alternative methods to characterise *C. cellulovorans* CBM33 structure in a way that can link structure and function were investigated. Electron paramagnetic resonance (EPR) was used to try and "infer the structure of the close-lying ligands of the" copper ion. [55]

It is possible to study the ligand characteristics in Cu(II) as there is one free electron in the copper molecule. EPR measures the spin of the lone electron in the copper ion, this phrase refers to the electron spin about the copper nucleus giving the electron angular momentum in the magnetic field. The electron can transition between, energetically similar, empty electron orbitals and there is a large angular momentum shift when the free electron transitions between orbitals. A free Cu (II) has two regions in the copper trace, the parallel and the adjacent. Protein residues binding the d orbital as coordinating ligands drastically alters the angular momentum of the free electron and represents the symmetry of the coordination. The ligand binding creates a separation in the parallel region of the EPR trace to give individual peaks.

The position of these separated peaks can be used as a fingerprinting method for the exact symmetry of the parallel region of the trace and from this information the identity of the coordinating residues can be assigned.

Copper centres display conserved coordination residues according to the type of centre and each type of copper centre visible using EPR gives a unique trace (see Table 3.1). The individual characteristics of the trace which are brought about by the d orbital energies

altering the splitting in the parallel and adjacent directions can act like a fingerprinting process to identify the coordination environment of the copper ion. The key parameters in this regard are the g tensor values (g factors are the EPR parameter incorporating "chemical shift" information) and the hyperfine coupling value (represented as a - the difference in magnetic field caused by the presence of nuclei magnetic fields).[56]

	Geometry and residues	Active site and EPR structure
Type 1 copper centre	Single copper molecule coordinated in a trigonal planar geometry by two histidine residues, a Cysteine and a variable ligand in the axial position.[57]	Axial residue His—Cu···His Cys

Type 2 copper centre	Square planar, single copper, active site with coordination from two histidine residues in a T shaped brace.	

Table 3.1 Comparative figure showing different characteristics of single copper, active sites.
Table showing an example of single copper centres that give a visible EPR spectrum. Table describes the geometry and ligands about the active site and gives a representation of an active site.

The EPR spectrum measured from a sample of *C. cellulovorans* CBM33 characterised by continuous X-band EPR was adjusted using a measured background and plotted in Figure 3.2. *C. cellulovorans* CBM33 EPR measurements gave values of g1=2.28 (2920 Gauss), g2=2.08 (3200 Gauss), g3=2.06 (3230 Gauss), A1=152.39 in mKaisers. The different g tensor values represent the intensity of interaction in a direction relative to the copper centre and represent an axial geometry given by the angular motion alteration of the lone electron in the copper centre by the coordinating residues in the x and y plane.
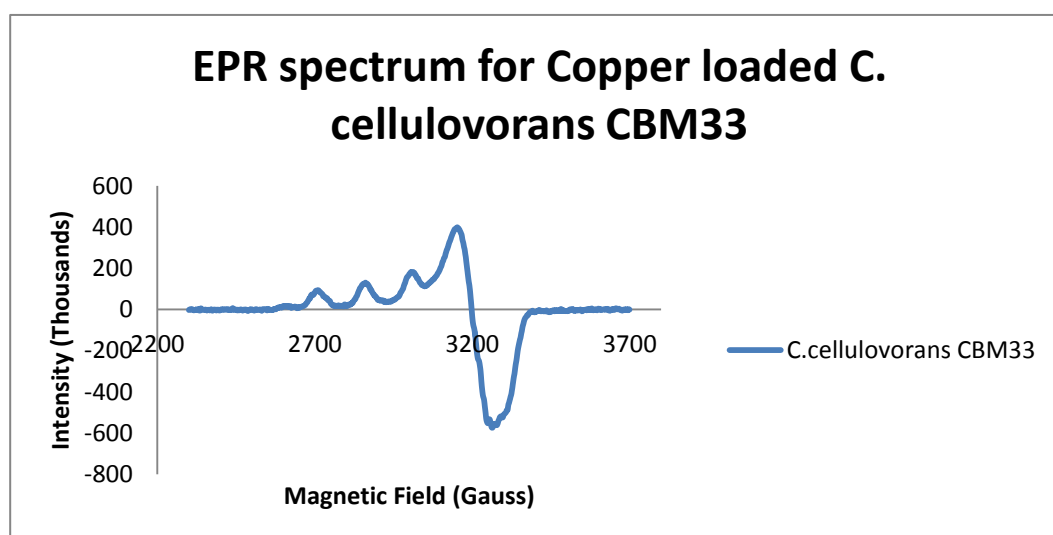
Figure 3.2 EPR spectrum for copper loaded *C. cellulovorans* CBM33
Figure showing the EPR trace of *C. cellulovorans* CBM33 and allowing for analyses of EPR characteristics such as g tensor values and splitting patterns.

The pattern of splitting observed in the parallel region (calculated by observing the gII and gI values from experimental data and displayed in Figure 3.2) of *C. cellulovorans* CBM33 is indicative of a type 2 copper centre. The EPR peaks display an axial geometry with the ligands in the X and Y planes (relative to the copper ion) are symmetrical and the Z plane ligands are at a different distance.

The axial nature of the copper active site suggests that *C. cellulovorans* CBM33 has LPMO characteristics (See Table 3.1). Comparative analysis of parallel gII and the shows that the value is indicative of a type (II) copper centre.

	gII	gI	AII (Gauss)
Copper (II) in aqueous acetate buffer, pH 5	2.42	2.08	118
Copper (II) -*Ta*GH61A*[24]	2.27	2.06	153
met-dopamine beta-mono-oxygenase [58]	2.29	2.06	155
met-peptidylglycine alpha-hydroxylating mono-oxygenase [59]	2.25	2.05	162
met-copper amine oxidase [60]	2.32		153
met-lysyl oxidase [61]	2.28		153
Type 2 site in copper methane mono-oxygenase [62]	2.24	2.04	185
Copper (II)-*Ba*CBM33 [50]	2.25		125
Copper (II)- *C. cellulovorans* CBM33	2.28	2.07	152

Table 3.2 EPR characteristics of characterised type (II) copper centres in LPMO molecules.
Figure showing EPR characteristics for a selection of different LPMO molecules. The gII represents *g* tensor values whilst A is the distance between the parallel region peaks and represents the hyperfine splitting of the copper ion free electron in the protein.

As the characteristics of the *C. cellulovorans* CBM33 EPR spectrum match those of a type 2 copper centre and we can, therefore, assume that the copper active site of *C. cellulovorans* CBM33 has axial symmetry with nitrogen or oxygen coordinating the copper ion. These findings provide more evidence to confirm the hypothesis that *C. cellulovorans* CBM33 is a member of the LPMO enzyme family. From this result we are able to can infer oxidative activity on polysaccharide molecules. The A value of *C. cellulovorans* CBM33 (152 guass), as a representation of the peak splitting is very different from what would be expected from a CBM33 which usually lie closer to a value of A=140. The larger distance between gI and gII suggests that the *C. cellulovorans* CBM33 copper centre is far more similar in structure to an archetypal GH61 than to a CBM33 active site. It is interesting that despite *C. cellulovorans* CBM33 being classified as a CBM33 enzyme it has a structure and function closer to that of a GH61 LPMO.

EPR analysis with Cellulose

In an attempt to elucidate the mechanism and substrate binding characteristics of *C. cellulovorans* CBM33, potential substrates were added to an EPR cell containing enzyme, as above, and the effect of substrate proximity on the copper ion with the ligand environment observed. The angular momentum of the unpaired electron in copper can be altered by interaction with atoms in close proximity which will alter the g value.

For a *C. cellulovorans* CBM33 sample incubated with Avicel (cellulose) the values of g1= 2.28, g2= 2.06, g3=2.01, A1=160 mKaisers were calculated. Upon addition of cellulose into the EPR reaction mixture containing *C. cellulovorans* CBM33 the g values observed in the perpendicular region of the copper centre were reduced and generated some splitting patterns around 3300 G. Small fluctuations in the spectra around this area suggest that the characteristics of the *C. cellulovorans* CBM33 active site have changed and further study on this phenomenon could elucidate to the mechanism of the enzyme. The spectra as a whole looked less like an archetypal copper 2 centre, alterations in the spectra showed distortion of the active site.

These results suggested that *C. cellulovorans* CBM33 was interacting with cellulose. As the fluctuations in the g values are at sites specific to a copper 2 centre we can suggest that the alterations of the copper centre are caused by a specific residue, in close proximity to the copper ion, with the cellulose substrate.
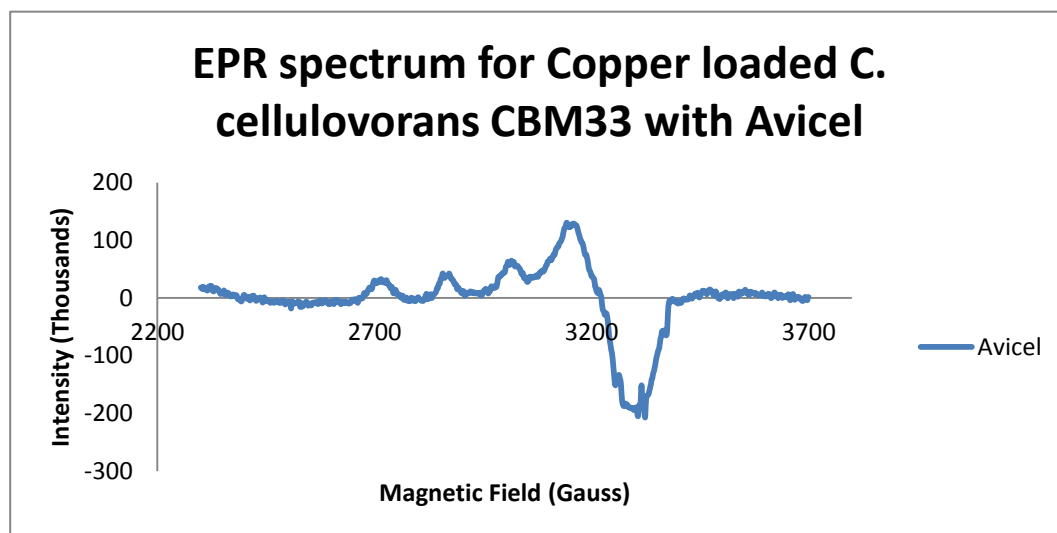
Figure 3.3 EPR trace of *C. cellulovorans* CBM33 sample with Avicel.

EPR analysis with Squid pen chitin

Squid Pen Chitin was added to the *C. cellulovorans* CBM33 EPR cell and the copper trace from this sample was studied. For a *C. cellulovorans* CBM33 sample incubated with Squid Pen Chitin values of g1=2.27, g2=2.06, g3=2.02 and A1=156 mKaisers were calculated.

The g values observed for both the perpendicular and parallel regions of the copper centre were reduced upon addition of chitin. Like in the cellulose EPR experiment there seemed to be greater resolution of the perpendicular region and this generated fluctuations between 3220 G and 3400 G. The most noticeable of these evolved peaks was at 3380 G; this characteristic was visible in neither the cellulose or apo *C. cellulovorans* CBM33 samples. More work is needed to be able to account for the alterations in the copper active site observed with chitin; however, the trace suggested that *C. cellulovorans* CBM33 may exhibit residue specific binding and activity on chitin.
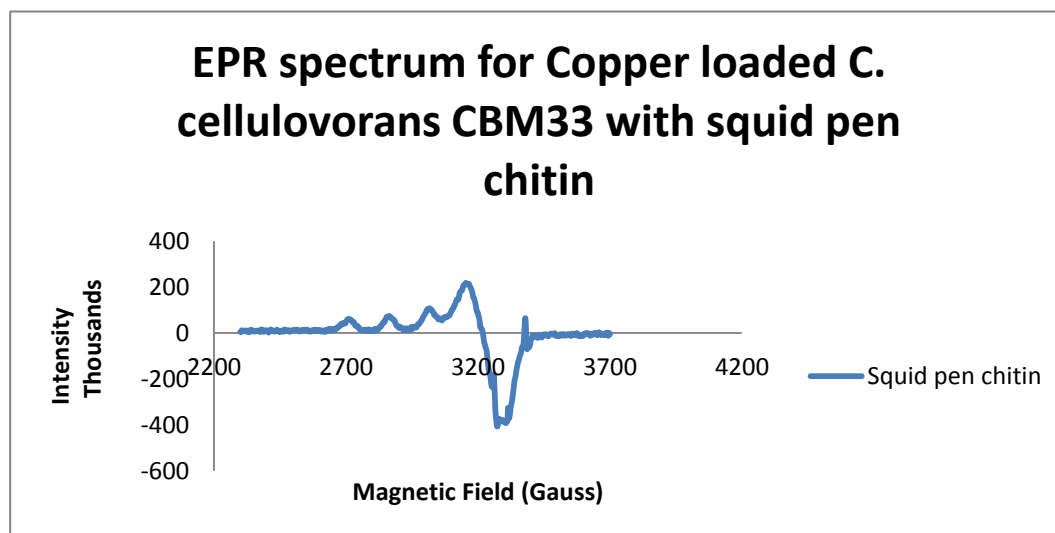
Figure 3.4 EPR spectrum of *C. cellulovorans* CBM33 with Squid Pen Chitin.

EPR analysis with Ivory Nut Mannan

Ivory nut mannan was added to the *C. cellulovorans* CBM33 EPR cell and studied the copper trace from this sample.
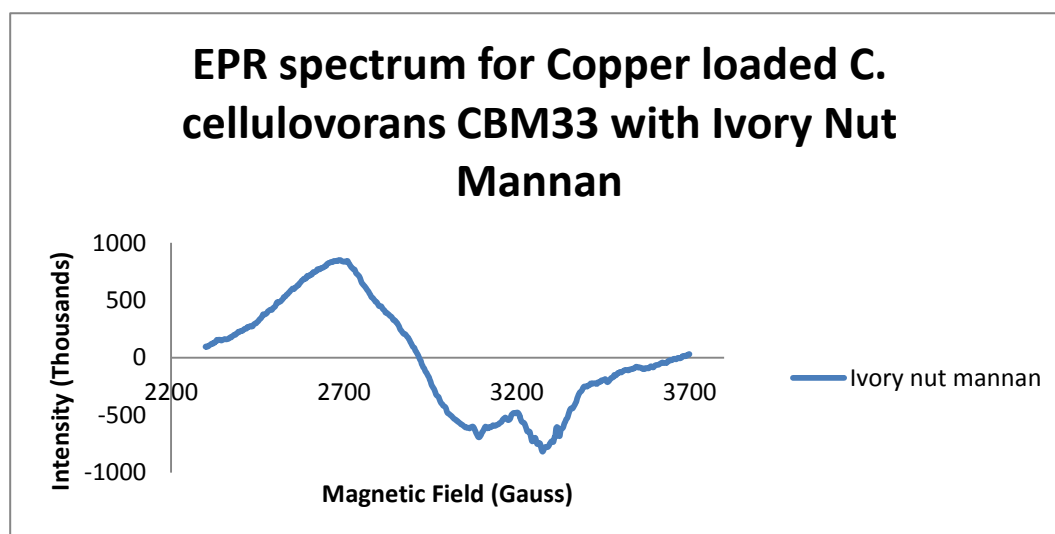
The EPR spectrum for the copper centre in *C. cellulovorans* CBM33 in the presence of Ivory Nut Mannan (Figure 3.5) is completely distorted from the characterised, axial anisotropic, type 2 copper centre characterised in *C. cellulovorans* CBM33. Addition of Ivory Nut Mannan into the EPR cell drastically altered the original *C. cellulovorans* CBM33 copper spectrum.
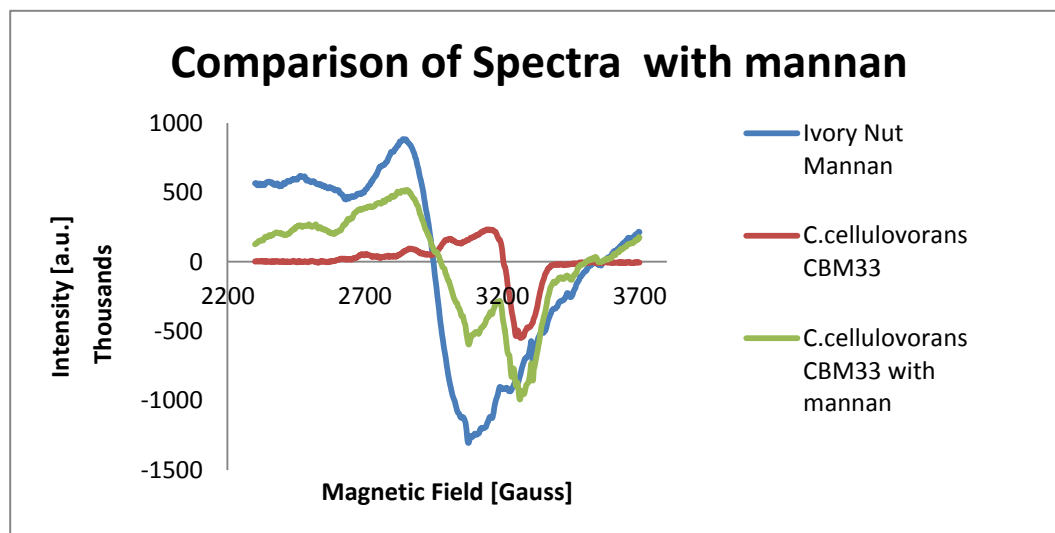
Figure 3.6 EPR spectra comparison of *C. cellulovorans* CBM33 with Ivory Nut Mannan with two blank samples.
Figure shows an EPR trace for *C. cellulovorans* CBM33 in the presence of ivory nut mannan; and for comparison two controls EPR spectra showing only C*. cellulovorans* CBM33and Ivory Nut Mannan respectively in an attempt to clarify the irregular peak observed in Figure 3.5.

In an attempt to clarify the strange copper profile observed in Figure 3.5, a further experiment was carried out where the EPR spectrum of a cell containing only Ivory Nut Mannan was measured.

Figure 3.6 shows that the EPR spectrum measured for a sample containing only Ivory Nut Mannan exhibited a similar distorted shape as the spectrum with *C. cellulovorans* CBM33. An attempt to normalize the spectra for *C. cellulovorans* CBM33 in the presence of Ivory Nut Mannan using the mannan sample as a background was unsuccessful.

It was possible to assign sensible g values to the *C. cellulovorans* CBM33 copper centre, in the presence of mannan, if the anisotropy were axial and $g_x=g_y>g_z$. We assigned the peak at 3100 G as the g parallel values of g1=1.97, g2=2.29 and g3=2.14. When the g parallel values for *C. cellulovorans* CBM33 incubated with mannan were compared to other copper centre in Table 3.1 would place the copper centre, in this reaction, outside the general classification range for LPMO molecules. The data made it difficult to judge

whether this was the case however the similar peaks observed in both the mannan blank spectrum and the enzyme cell suggested that this was incorrect.

The distortion of the spectrum makes analysis of the copper centre using this method, using g and A values impossible. The spectrum of *C. cellulovorans* CBM33 with Ivory Nut Mannan was indicative of a very strong interaction between copper and the mannan molecule. However, from the distortion we could suggest that the interaction is non-specific.

Mass spectrometry product analysis

The activity and substrate specificity of *C. cellulovorans* CBM33 is still undefined. To evaluate *C. cellulovorans* CBM33 activity on polysaccharide polymers, mass spectrometry analysis was carried out on the reaction products evolved from *C. cellulovorans* CBM33 incubation with potential substrates.

Mechanistic details of polysaccharide breakdown by LPMOs remain absent; however, we can begin to make assumptions about key steps and intermediates by studying the products evolved in the reaction. *C. cellulovorans* CBM33 was incubated with polysaccharide substrate and the solution analysed by Matrix-assisted laser desorption/ionization (MALDI). The oligosaccharide product's mass allows for identification of the lytic oxidation site.

Product analysis - *C. cellulovorans* CBM33 incubated with Avicel

C. cellulovorans CBM33 was incubated with crystalline Avicel in conditions that would prompt enzymatic activity. Analysis of the evolved species suggested that *C. cellulovorans* CBM33 exhibits celluolytic activity. Aldonic acid oligosaccharide cleavage products were observed, in the *C. cellulovorans* CBM33 samples, at dimer unit lengths (DP) = 4, 5, 6, 7, 8, 9. (See Table 3.2) There was also evidence of unmodified oligosaccharides; (DP = 5, 6, 7, 8, 9) though these were also seen in the Avicel blanks the intensity of the peak increased when *C. cellulovorans* CBM33 is present as was the case in similar GH61 experiments. [24] A minor product was observed at DP_n-2 Da, previous work has suggested unopened lactones or C4 oxidation products as the cause for this peak (See figure 3.7). [35]

Product analysis - *C. cellulovorans* CBM33 incubated with Squid Pen Chitin

C. cellulovorans CBM33 was incubated with crystalline squid pen chitin and the reaction products analysed using mass spectrometry. *C. cellulovorans* CBM33 displayed chinolytic activity on crystalline Squid Pen Chitin (β-chitin). The predominant reaction product was aldonic acid oligosaccharides with DP6 and DP8 (See Figure 3.8). These findings corroborated with data collected for other AA10 enzymes which have been described as creating lytic chain breaks in chitin structure to form even polymer unit oligosaccharides.

Product analysis - *C. cellulovorans* CBM33 incubated Mannan polymers

Attempting to confirm, for the first time, the activity of an LPMO on crystalline mannan, a variety of mannan polymers were incubated with *C. cellulovorans* CBM33 and the products analysed by mass spectrometry. None of the mannan samples contained lytic products when analysed by mass spectrometry. The mannan solutions were contaminated with free oligosaccharides which prevented the solution from crystallising on the MALDI-TOF analysis plate. When it was possible to crystallise a reaction solution the mass spectrometry reading was noisy and inconclusive (Data not shown as there was not reading). Attempts were made to optimise the analysis protocol by increasing the crystallisation agents for the matrix and decontaminating washes of the mannan prior to the reaction however neither variation yielded data showing activity on mannan. Further work is needed to characterise the activity of *C. cellulovorans* CBM33 on other crystalline polysaccharides. The linking of the *C. cellulovorans* CBM33 domain to a mannanase when the protein is produced from the parent organism implies heavily that activity on mannan be the primary purpose of *C. cellulovorans* CBM33.
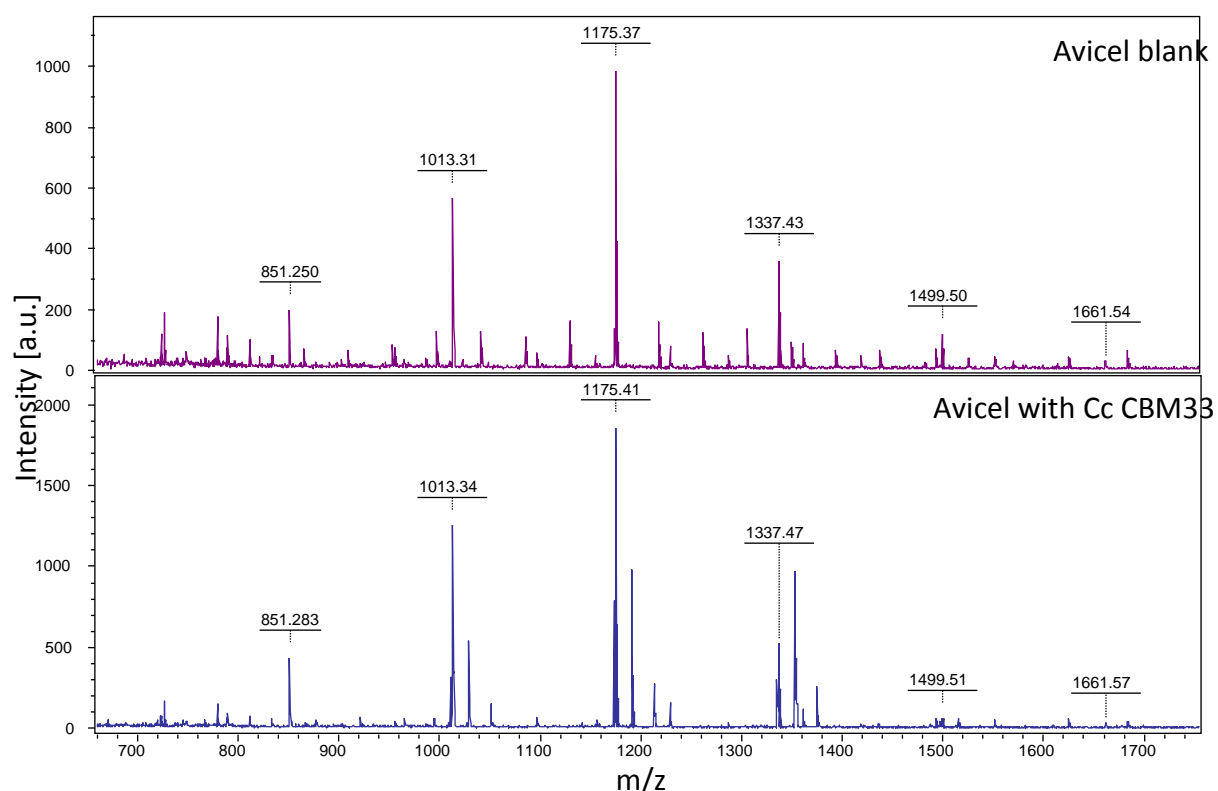
Reaction Product		DP5	DP6	DP7	DP8	DP9
Un-modified oligosaccharide	m/z	851	1013	1175	1337	1499
	Integrated area	40	130	201	63	15
Lactone or	m/z	N/A	1011	1173	1335	N/A

hypothetical C4 oxidation	Integrated area		29	82	33	
Aldonic acid	m/z	867	1029	1191	1353	1515
	Integrated area	3	53	100	110	7
Aldonic acid + 2Na	m/z	N/A	1051	1213	1375	N/A
	Integrated area		15	26	24	

Figure 3.6 Mass Spectrometry analyses of the reaction products evolved when *C. cellulovorans* CBM33 is incubated with Avicel. The peaks represent the m/z of individual saccharide reaction products. Table displays the saccharide product characteristics.
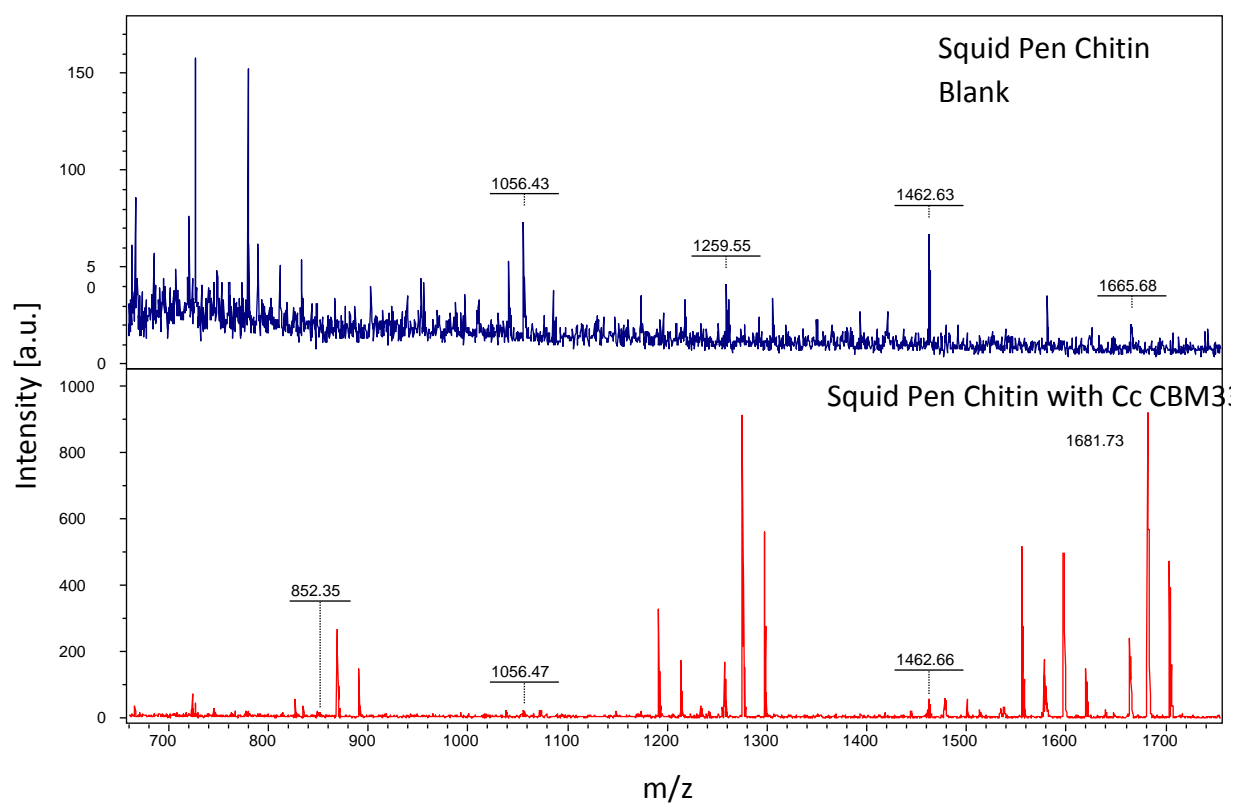
Reaction Product		DP6	DP8
Un-modified oligosaccharide	m/z	1259	1665
	Integrated area	5	12
Lactone or hypothetical C4 oxidation	m/z	1257	1663
	Integrated	16	26

	area		
Aldonic acid	m/z	1275	1681
	Integrated area	86	114
Aldonic acid + 2Na	m/z	1297	1703
	Integrated area	49	58

Figure 3.7 Mass Spectrometry analyses of the reaction products evolved when *C. cellulovorans* CBM33 is incubated with squid pen chitin.

Analysis of *C. cellulovorans* CBM33 activity on chitin under thermophillic conditions

C. cellulovorans is a thermophillic bacterium that can grow in conditions up to 80°C. Identifying enzymes that are able to catalyse reactions efficiently at high temperatures is a goal in biotechnology to allow treatments at a wider range of temperatures.

Figure 3.8 clearly shows that oxidative products were produced in the reaction mixture which confirms that activity of *C. cellulovorans* CBM33 on chitin in thermophillic conditions. The characteristics of the mass spectrometry analysis were altered between the samples carried out at 30°C and those at 80°C. The variation in reaction products increases at 80°C which is most likely due to the decreased stability of crystalline cellulose at higher temperatures. The activity of an oxidative enzyme at thermophillic temperatures is important to increase the efficiency of enzyme treatments to reduce the enzyme load required. Results would suggest that increasing the temperature of a synergistic reaction between *C. cellulovorans* CBM33 and classical hydrolases would require a far smaller enzyme load that existing biofuel treatments and would therefore reduce cost and production potential.
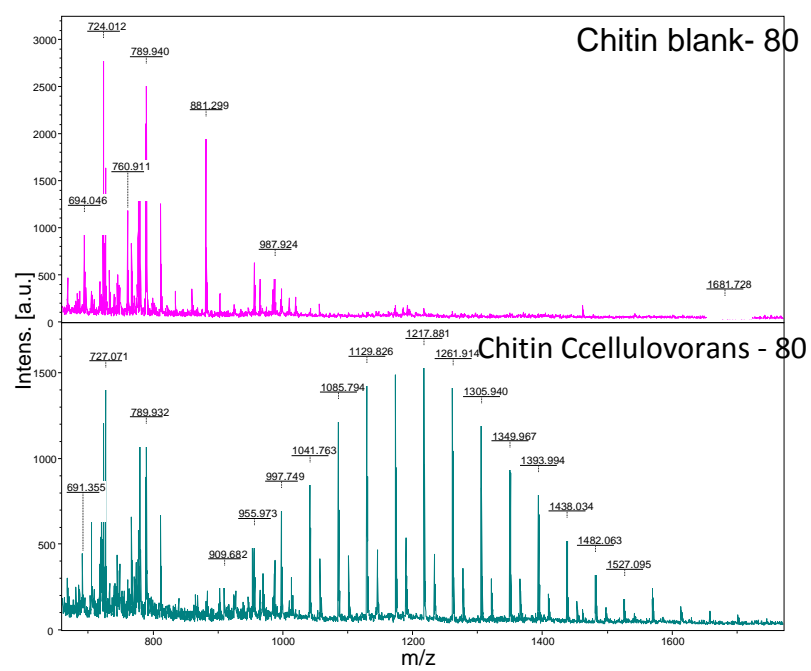
Figure 3.8 Mass Spectrometry analyses of the reaction products evolved when C. *cellulovorans* CBM33 is incubated with Squid Pen Chitin in a reaction vessel at 80°C

Chapter 4 - Discussion

The results of our research show that we have successfully purified a protein from *Caldibacillus cellulovorans* which is hypothesised to be a lytic polysaccharide mono-oxygenase in accordance with current literature.

Bio-informatics searches suggest that the *C. cellulovorans* CBM33 enzyme has sequence conservation with existing CBM33 molecules. *C. cellulovorans* CBM33 with an ubiquitin like protein attached at the N-terminal has been purified by sequential Nickel column and gel filtration. This technique could provide a novel paradigm for efficient purification of recombinant proteins whilst conserving the delicate native structure.

Protein identification results confirmed successful expression, site specific cleavage of the SUMO domain and re-affirmed the presence of one di-sulphide bridge in *C. cellulovorans* CBM33.

The purification protocol utilizing the poly-histidine tag provided a final yield of 150 mg/L in a pure sample. We experimentally showed that *C. cellulovorans* CBM33 bound copper in a stoichiometric ratio 0.6. This value is lower than expected when examining the metal binding site in comparison to other CBM33 molecules. We would expect a binding ratio of 1.0. Lower ratios have been observed in related experiments involving related AA10 proteins and other single copper dependent proteins. [30,50] This feature may suggest copper binding by LPMO molecules tight enough to prevent stripping of the copper ion by EDTA. Otherwise there could be a sequence based physiological reason for why the protein retains copper such as steric hindrance of EDTA.

EPR experiments allowed examination of the binding features within copper center of *C. cellulovorans* CBM33. EPR results confirmed that *C. cellulovorans* CBM33 had an active site with typical type 2 copper binding geometry values, characteristics indicative of an LPMO protein. G value comparisons with LPMO molecules from other sources appear to confirm that the square planar ligands coordinating the coper ion will be nitrogen or oxygen atoms which most likely come from histidine molecules as in other LPMO molecules.

These finding suggest that the *C. cellulovorans* CBM33 active site is not coordinated in an archetypal square planar orientation; distortion of the active site has been described in other CBM33 molecules.[63,64] However, the g values obtained for *C. cellulovorans* CBM33 are closer to the characteristics seen in GH61 molecules rather than CBM33s.

The hyperfine splitting displayed in EPR analysis of *C. cellulovorans* CBM33 reaction with both chitin (squid pen) and cellulose (Avicel) matches those seen in other LPMOs and supports the idea of reactive oxygen species formation in the mechanism of degradation. These results suggest that *C. cellulovorans* CBM33 is active on chitin and cellulose in an oxidative manner.

In previous work, AA9 and AA10 were thought to have exclusive activities, on chitin and cellulose respectively, although some more recent studies have shown activity on both. Interestingly, in the case of *C. cellulovorans* CBM33, the active site is somewhere between that of typical AA9 and AA10 enzymes previously classified. This raises questions about the link between cooper ion ligands in the active site and the activity of the protein. No accepted hypothesis currently exists regarding the substrate specificity of LPMOs and differences in the active site residues could provide the answer; for example changes in the steric properties, hydrophobicity, or charge around the active site altering the specific binding of a substrate.

The EPR spectrum for *C. cellulovorans* CBM33 incubated with Ivory Nut Mannan is less definitive. The trace shows a complete loss of any feature resembling a type 2 copper center, the distortion is so complete that it is difficult to make any assumptions about the interaction. The distortion of g values suggests an interaction of *C. cellulovorans* CBM33 with mannan, with the mannan in a close proximity to the copper ion; however, the results do not give enough evidence to suggest or support novel activity of a CBM33 protein on complex mannan polymers.

The hypothesized activity of *C. cellulovorans* CBM33 on crystalline polysaccharides was further tested by analysis of reaction products. Mass spectrometry analysis of reaction products showed that there were mass peaks at several Mws that matched oxidation products when *C. cellulovorans* CBM33 was incubated with both cellulose and chitin. These results suggest that we have confirmed the activity of a LPMO isolated from a *Caldibacillus cellulovorans* on both chitin and cellulose. Importantly oxidative activity on chitin was also proven at 80°C which provides potential for more thermally stable enzyme treatments in bio-technology.

Study of the reaction products from the reaction with cellulose where the products have degrees of polymerization between 4 and 9 with the production of a 7 monomer chain favored. In comparison, chitin experiments produced products at every other DP-unit, for example, 6 and 8 sugar unit products; though, the octomer product appears to be favored over the hexamer. These reaction profiles suggest a different mechanism of activity on chitin and cellulose and supports EPR characterization that suggests *C. cellulovorans* CBM33 falls between the typical parameters of AA10 and AA9 characteristics.

Accounting for the different activity of *C. cellulovorans* CBM33 of chitin and cellulose, the hypothetical activity on mannan polymers and the ability to catalyze reactions at a wide range of temperatures makes the enzyme a fantastic candidate for biotechnological treatments due to its broad spectrum of activity. Further to this the native enzyme complex released from *C. cellulovorans* involving two carbohydrate binding domains and a classical hydrolase will allow for more complex treatments allowing for enzymatic cooperation and increasing of rates as well as diversification of substrates.

Structure and function characterization of *C. cellulovorans* CBM33 has facilitated a wide range of new possibilities in biotechnology as well as enhancing existing areas. Despite this, there remains a gap in our knowledge of how *C. cellulovorans* CBM33 acts and biotechnological treatments in general, and therefore, further study is required

Future work

Initial steps will focus to resolve the irregular results obtained in copper-protein binding experiments. An elegant experiment, involving sequential competition titration of a protein with zinc then copper, was developed in the characterization of newly discovered AA11. Experimental stoichiometry values obtained for *Ba*CBM33 rose to around 0.8 per enzyme molecule which is closer to the expected value and reflective of the structural results obtained by x-ray crystallisation. Further to more complex copper binding experiments, binding to other metal captions should be investigated to give a representation of the specific affinity of the enzyme active site for copper in comparison to the less specific binding of other metals. Some CBM33 proteins have been reported as exhibiting limited activity in the presence of di-valent metals other than copper. [35]

The affinity, stability and melting temperature of *C. cellulovorans* CBM33 could be investigated through utilization of differential scanning fluorimetry. [65] These experiments would give further indication as to copper binding and would confirm the LPMO characteristics of *C. cellulovorans* CBM33.

The most important piece of future work will be to try and confirm the hypothetical activity of *C. cellulovorans* CBM33 on complex mannan polymers. This is a process that has not been discovered and the novel activity would open a wider field open bio-technological treatments. Attempts to visualize mannan polymer, *C. cellulovorans* CBM33 catalyzed, reaction products using MALDI-TOF were unsuccessful. The crystal matrix created on the Bruker mass spectrometer imaging plate designed to absorb the sample and allow laser excitation to allow ionization and flight on the instrument. However, the mannan samples did not exhibit normal crystallization characteristics and therefore the sample components could not be imaged using this method. The samples gave incredibly noisy reading and this coupled with unusual crystallization it would appear that there are a plethora of mannan oligosaccharide units present (both before and after incubation) which are inhibiting the correct crystallization causing the samples to be unable to fly. To overcome this barrier, and to give a clearer image of the reaction solutions, future work could focus on liquid state analysis of the reaction components which would remove the requirement for volatilization. For instance the reaction mixture could be analyzed by High Performance Liquid Chromatography (HPLC) or High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). This process would allow for more accurate mass analysis of reaction products and allow assigned activity of *C. cellulovorans* CBM33 on mannan polymers.

The understanding of other LPMO molecules increased greatly when it was possible to visualize and study the three dimensional structure. Influenced by this, it will be important to experimentally find conditions in which both Apo and copper bound crystals of *C. cellulovorans* CBM33 can be grown and diffraction measured. A structure resolved by this method will elucidate details about the protein structure to function relationship such as active site coordination, coordinating bond lengths (variable in LPMOs with a potential link to activity of product length) and protein surface residues quoted as playing an essential role in substrate orientation and binding and therefore activity.

The reason for increasing interest in LPMO molecules is the potential role that they could play in bio-technological treatments to break down the highly complex and recalcitrant structure of polysaccharides. For these proteins to be useful in bio-technology there is a requisite that they function efficiently in standard industry processes. As mentioned above, the first LPMO molecules were identified through their synergistic acceleration of a standard hydrolase treatment activity on crystalline cellulose. Repeating this experiment to study the activity of classical hydrolases and *C. cellulovorans* CBM33 in concert will demonstrate the activity on cellulose, chitin and hypnotically mannan polymers.

Application

LPMOs are attracting such a large amount of attention globally due to the wide range of bio-technological niches they could fill. The primary research motive is the formulations of biofuel in a renewable manner from sustainable sources in a way that does not negatively impact on global food, water and land security. This work has shown that *C. cellulovorans* CBM33 exhibits activity on both chitin and cellulose polymers; this makes *C. cellulovorans* CBM33 an industrial enzyme with the potential to diversify treatments due to the fact that *C. cellulovorans* as a host organism is thermophillic and therefore could potentiate reactions at higher temperatures and temperature controlled single reaction chambers.

If the activity of *C. cellulovorans* CBM33 on mannan substrates could be proven, bio-fuel feedstock would increase greatly. Using current techniques utilizing the complete global biomass harvest would only produce around 150 EJ/year. [19] Discovery of an enzyme which could facilitate the recovery of monomeric sugars from complex mannans and glucomannans greatly expands the available feedstock for bio-fuel production as mannan is a recalcitrant bio-polymer present in the cell wall of plants.

Further to the potential *C. cellulovorans* CBM33 offers to the field of bio-fuels, there are other potential uses. Glucomannans are present in many foods; it is added to liquids as a

thickening agent but is known to stain fabric fibers when they come into contact. When glucomannan, in food, contacts fibers of clothing it stains the fabric. Removal of mannan stains has been an issue for detergent industries for many years and it is only recently that enzymes have been developed for bio-tech treatments able to remove these stains. *C. cellulovorans* CBM33 could provide a realistic supplement to these treatments due to the high natural thermo-stability suited to high temperature clothes washers and stain removal.

Enzymes active in the degradation of complex polysaccharides have been regarded with interest for some time in the field of nutritional science. The human digestion system is able to digest only a small percentage of plant material, due to the high percentage of recalcitrant polysaccharides in the cell walls, and this therefore decreases the energy and nutrition that can be obtained from plant based foods with respect to the land area, volume of water, time, money and man power essential for their production. Bio-technological techniques have been suggested that involved the supplementation of symbiotic fauna to the human body to enhance uptake of simple sugars isolated from previously indigestible materials. *C. cellulovorans* CBM33 is a molecule that is predicted to disrupt the complex matrix of polysaccharides. A predicted wide range of activity and working temperature would make *C. cellulovorans* CBM33 a valuable addition to nutrient optimization treatments.

Chapter 5 - Experimental

Standard methods

DNA sequencing

DNA was sequenced using Sanger sequencing (Sanger ABI 3730xl) provided by GATC. Samples were sent using research group, pre ordered, barcodes and results were made available for download and delivered by email.

Polymerase Chain Reaction (PCR)

Gene amplification was achieved by utilising a polymerase chain reaction using commercially available polymerase. [66] A typical reaction mixture is shown in Table 5.1 with Q5 High-Fidelity DNA polymerase (New England Biolabs) usually used, though other polymerases (specified in individual protocols) were sometimes chosen to improve products yields or as an alternative when PCR of longer sequences was unsuccessful.

PCR reaction component	Quantity
Template DNA (100ng/µl)	0. 5 µl
Forward primer DNA	1 µl
Reverse Primer DNA	1 µl
dNTPs (2mM)	5 µl
Polymerase buffer (10 x High Fidelity buffer)	10 µl
Polymerase enzyme	0.5 µl
MilliQ ddH$_2$O	32 µl

Table 5.1 Standard polymerase chain reaction
Table displaying the standard mixture components for the replication of sample DNA by polymerase chain reaction.

If the PCR reaction was unsuccessful using Q5 polymerase then multiple reactions were set up using Q5 Hot start polymerase (New England Biolabs), Phusion polymerase (Thermo Scientific), KOD polymerase KOD Hot Start polymerase (Novagen).

The reaction was controlled thermally in a PCR cycler (See Table 5.2), the 50 µl reaction mixture was contained in a thin walled PCR tube to ensure accurate heating.

Temperature (°C)	Duration (Seconds)	Repeats	Notes
98	120		Denature the DNA initially to single stranded DNA. This allows binding of the primers by complementary Watson–Crick base pairing.
98	30	X 30	Denature strands.
60	30		Anneal the primers to the template strand
72	20		Extension of the copy strand by the polymerase, incorporating the nucleotides to duplicate the strands.
4	∞		Cooling to 4 °C kills the polymerase reaction and stores the products without breakdown

Table 5.2 PCR cycling protocol for the standard PCR reaction used.

Pouring an agarose gels for analysis by electrophoresis

0.5 g of Agarose powder was added to the TRIS base, acetic acid and EDTA buffer (TAE - 40mM Tris, 20mM acetic acid, and 1mM EDTA) at pH 8. The mixture was heated for 80 seconds in the microwave at 800 w; the solution was stirred after 40 seconds to avoid evaporation. Once the solution has cooled to room temperature, 1 μL SYBR Safe (purchased as 10,000X concentrate in DMSO) is added and mixed by agitation before pouring the gel into the casting block and inserting the 6 well medium comb for 50 minutes before running.

Preparing the samples for agarose gel electrophoresis

The set gel is placed in a running tank filled above the level of the gel with TAE buffer. The products from the PCR reaction need to be mixed with a dye before they are added to the gel. The SYBR Safe added to the gel chelates with the DNA. The fluorescence, of the complex formed, was imaged under UV irradiation. A ladder consisting stand sized oligonucleotides was loaded onto the gel to be used as a marker for comparative measurement of the DNA passage through the gel, New England Bioloabs 1 kb ladder was used.

Running an electrophoresis gel

The lid is applied to the tank which ensures that the electrical circuit is closed. A power pack is used to apply electrical current (200 v) through the gel for one hour. The TAE, acid/base buffer ensure the DNA is negatively charged and therefore is attracted to the

positive electrode when a current is applied. Components of the solution are separated by size; the larger the chain the slower it travels allowing accurate resolution by size. The agarose gel is imaged; the stains used in the loading buffer bind to the DNA and are visible under UV light.

Transformation of plasmid containing recombinant gene into competent cells

Competent *E.coli* cell aliquots were defrosted on ice and 1 µl DNA added to the solution. DNA was gently mixed by pipette before heat shocked for 30 seconds at 40°C before returning back to ice for a further 2 minutes. Upon addition of 250 µl of SOC growth media the reaction mixture is incubated at 37°C for 1 hour in an Eppendorf Thermoblock shake. [67]

Preparation of LB agar media for the growth of cells

LB agar solution is made up in a ratio of 5 g NaCl: 5 g Tryptone: 2.5 g Yeast Extract: 7.5 g Agar per 500 ml of MilliQ H_2O. 200 mLs of LB agar is melted by microwave, care must be taken to avoid boiling of the liquid and agitation should be used to mix and distribute the heat. Allow the contents to cool to near room temperature and add antibiotic(s) for resistance selection in the vector to a final concentration between 30-50 µg/mL in LB agar mixture (Ampicillin -35 µg/mL, Kanamycin – 50 µg/mL, Chloramphenicol – 35 µg/mL). Under abiotic conditions pour the agar into empty plates, ensuring the surface is smooth and level, before being allowed to dry for two hours; place in a drying oven 40 minutes before use.

Plating out cells onto LB agar plates

250 µL of cells of the growth culture is pipetted onto the agar plate under aseptic conditions. Spread cells evenly with a glass spreader that has been flamed with ethanol and a roaring blue flame. Allow the cells to dry onto the plate and then place in the 37°C oven to grow overnight.

Small Scale Expression Tests

LB solution is made up in a ratio of 10 g sodium chloride; 10 g Tryptone, 5 g yeast extract powder per litre. Small, 5 ml cell cultures were grown from a single colony at 37°C and agitation of 180 rpm. The growth temperature is reduced to 16°C once the OD_{600} reaches 0.4. Once the optical density reaches 1.5 (0.8 in early experiments) IPTG is added to a final concentration of 1 mM. The inoculated culture was left to grow over night at 16°C.

Glycerol stock of cells containing the desired recombinant gene constructs

0.5 mls of culture was mixed with 0.5 mls of glycerol (80%) and mix gently in an Eppendorf tube. Drop the tube into liquid nitrogen to freeze and place in a -80°C freezer.

Large Scale Protein Production for Purification

500 ml cell cultures were grown from a single colony at 37°C and agitation of 180 rpm. The growth temperature is reduced to 16°C once the OD_{600} reaches 0.4. Once the optical density reaches 1.5 (0.8 in early experiments) IPTG is added to a final concentration of 1 mM. The inoculated culture was left to grow over night at 16°C.

Isolation of the periplasmic fraction from cells

The cell media was transferred into centrifuge tubes and balanced before centrifugation with a Sorvall F10 rotor, 8000 rpm for 30 minutes to obtain a pellet of cells.

1. Cell pellets obtained were weighed before being suspended in 3 volumes (with respect to weight in grams) of iced 50 mM Tris pH 8.0, 200 mM NaCl and 20% w/v Sucrose.
2. 40 μl of 10 mg/ml Hen Egg White Lysozyme was added for every gram of cell paste and solution incubated on ice for 1 hour with occasional agitation.
3. Add 60 μl of 1 M $MgSO_4$ for every gram of cell paste and incubate on ice for a further 20 minutes.
4. The protoplasts were spun down in a F13S rotor at 10000 rpm for 20 minutes and the supernatant isolated into a fresh tube and stored on ice.
5. Pellet is suspended again, this time in ice cold milliQ water (same volume as used in step 1) and incubated on ice for a further hour.
6. Repeat Step 4.
7. 4 sonication pulses were applied to the solution at maximum amplitude to decrease the viscosity of the solution
8. Sample was diluted in 5 volumes of 50 mM TRIS pH8, 50 mM NaCl buffer ready for purification.

Cell Lysis following intracellular protein expression

1. Harvest cells were above (Sorvall F10 rotor, 8000 rpm, and 30 minutes) and the pellets obtained weighed.
2. Suspend in 3 volumes (with respect to weight in grams) of iced 50mM TRIS pH8, 250 mM NaCl, and 30 mM Imidazole buffer.
3. 10 sonication pulses were applied to the cell suspension in intermediates before being returned to ice to cool. The lysed solution was centrifuged in the F13S rotor at 10000 rpm for 20 minutes.
4. The cell lysate was then used for downstream purification.

Protein analysis by polyacrylamide gel electrophoresis

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) involves the binding of sodium dodecylsulfate to protein back bone which forces the unfolding of protein to create a charged molecule that can be resolved through a polymer gel by applied current.

12% (w/v) SDS-PAGE gels were poured according to the following recipe:

Resolving Gel	Stacking Gel
2.5 ml Resolving gel Buffer	1.3 ml Stacking gel buffer
4.2 ml Acrylamide Stock	500 µl Acrylamide stock
3.2 ml ddH$_2$O	3.2 ml ddH$_2$O
16 µl TEMED	8 µl TEMED
100 µl 10% APS	100 µl 10% APS

Table 5.3 Shows the standard recipe for making up a 15% SDS-PAGE gel using Hoefer gel kit.

Samples were mixed with a quarter volume sample buffer, mixed, and then heated to 94°C to boil the protein. The sample was loaded into a Hoefer gel kit filled with running buffer. The gel was run at 100 volts for 50 minutes with a standard protein ladder into one lane. Stain the gel with coomassie blue before de-staining and analysing.

Specific methods

Cloning *C. cellulovorans* CBM33

Ed Taylor ordered a plasmid containing *C. cellulovorans CBM33* from GenScript in a pET-11a vector (Ampicillin resistance selection) with an attached PelB tag for targeting the protein to the periplasm. All further cloning was carried out using this DNA template.

In-fusion cloning was used to insert the *C. cellulovorans CBM33 gene into various expression vectors* following the standard procedures described by Clontech. [68]

Primer design

Primers were designed using serial cloner 2.6 (http://serialbasics.free.fr/Serial_Cloner-Download.html), oligonucleotide sequences, with the standard primers pre-requisites, were ordered from Eurofins MWG Operon. Primers were ordered with overlapping sticky ends so that annealing to the template strand would be specific and efficient so as to give a good yield in PCR reactions. [69] Special attention was paid to the GC content and the melting temperature to ensure that PCR artefacts were avoided in the cloning process and the DNA product produced had the correct sequence.

C. cellulovorans CBM33 YSBL-LIC	
Forward primer 5' to 3'	AGGAGATATACCATGCATGGTGGTATGGTGTTCCC
Reverse primer 5' to 3'	TGAGGAGAAGGCGCGTTAGATCATGGTACCGCCTTC
SUMO C. cellulovorans CBM33 Champion PET Sumo	
Forward primer 5' to 3'	GAACAGATTGGTGGTCATGGTGGTATGGTGTTCCC
Reverse primer 5' to 3'	TACCTAAGCTTGTCTTTAGATCATGGTACCGCCTTC

Table 5.4 DNA primers for *C. cellulovorans* CBM33.
Table showing the sequence of the DNA primers designed and ordered to clone recombinant genes into expression vectors.

DNA purification

The DNA product from the initial PCR reactions is contaminated with used PCR products, artefacts and impurities that must be removed from the solution before any downstream reactions can take place. This is done by DNA purification techniques; DNA produced from our PCR reactions was purified using the QIAquick PCR Purification Kit from Qiagen. All centrifuge reactions for the following protocol were carried out at ≥10,000 x g (13,000 rpm) in a lab, tabletop, micro centrifuge. The kit has all buffers made up to working concentration and the standard protocol was followed. [70]

DPN-1 Digest reaction

DNA synthesized from E.coli has undergone N6 methylation of adenine due to methyltransferases in the E.coli. Dpn1 digestion is carried out to remove the template DNA which bears the methylation. The reaction produces non-methlyated sample which makes downstream processes cleaner

In fusion reaction

In fusion allows for the cloning of DNA fragments into a linear expression vector to form a complete, circular, vector without the need for restriction and ligation. The DNA insert containing *C. cellulovorans* CBM33 was combined with linear vectors using the CLonetech InFusion Kit and following the standard protocol. [71]

Transformation of *C. cellulovorans* CBM33 vector into NovaBlueGiga cells

An aliquot of NovaBlueGiga cells were transformed with 1 µl of *C. cellulovorans* CBM33 vector DNA following the protocol above. NovaBlueGiga cells are genetically engineered contain *recap endA* mutations, which facilitate high yields of excellent quality plasmid DNA. [72] Cells are plated out following the above protocol and grown overnight at 37°C.

Testing the DNA constructs

The plated cells were left to grow overnight into a selection of single colonies; within each colony the cells contained the same genetic sequence. All colonies that grew on the antibiotic plates have been selected due to the presence of resistance genes; tests were carried out to confirm that the correct gene had been cloned into the plasmid in the InFusion reaction and that no contaminant colonies were present.

Colony PCR

Using a pipette tip, a single colony was scraped from the agar plate and dipped into a Colony PCR solution carry out gel electrophoresis as described above and analyze by agarose gel (Dream Taq Buffer acts as the loading dye). The size of the DNA fragment showed whether an insert of the correct size has been incorporated into the vector.

PCR reaction component	Quantity
Milli Q ddH$_2$O	9. 5 µl
Forward primer DNA T7 protease forward primer (20 µM)	1.5 µl
Reverse Primer DNA YSBL-LIC (- H) reverse primer (20 µM)	1.5 µl
dNTPs (10 mM)	5 µl
Dream Taq Green Buffer	1.5 µl
Dream Taq Polymerase	0.5 µl

Table 5.5 Colony PCR recipe.
Showing the standard recipe for making up the Colony PCR reaction mixture to carry out DNA accumulation from a colon selected for the presence of a recombinant gene.

Colony PCR grow up

The colony picked for colony PCR is then added into a 5 ml LB culture with specific antibiotic(s). The overnight culture was set up. After overnight incubation the cell suspension was harvested by centrifuge at 13.3k rpm and the supernatant removed. Overnight cultures were used to increase the cell density in the suspension to allow extraction of DNA from the cells or for inoculation of larger growth

DNA isolation from cells using SpinPrep DNA kit from Novagen

The harvested cell pellet is suspended in buffer P1 from Novagen according to existing protocols. [73]

1. Suspend cell pellet in 250 µl of Buffer P1 (containing LyseBlue Reagent and RNAase) in a Eppendorf tube and mix.
2. Add 250 µl of Buffer P2 and invert 12 times and incubate for 5 minutes on ice (solution should turn blue).
3. Add 350 µl of Buffer N3 and invert 12 times to quench the lysis (should turn colourless).
4. Centrifuge for 10 minutes in a desktop centrifuge at 13k rpm.
5. Using Gilson, apply the supernatant to QIAprep tube and centrifuge for 60 seconds.
6. Wash tube with 500 µl of PB Buffer and spin for 60 seconds.
7. Add 750 µl of PE Buffer and centrifuge twice for 60 seconds each, allowing the sample to stand for 60 seconds in-between.
8. Elute the membrane bound DNA by washing the tube with 60 µl of EB Buffer into a clean Eppendorf, allow the sample to stand for 60 seconds before centrifuging for 60 seconds.

Expression of recombinant *C. cellulovorans* CBM33

Transformation of expression cells

The correct DNA sample is transformed into the competent cells following the protocol in general methods. Expression of PelB *C. cellulovorans* CBM33 in pET-11A was tested in Bl21, BL21*, Rossetta2, C41 and TUNER cells (See page 52).

Expression of recombinant *C. cellulovorans* CBM33

Cultures of LB media were inoculated with cells containing the *C. cellulovorans* CBM33 plasmid and grown following the protocol in general methods (see page 52).

Purification of periplasmic *C. cellulovorans* CBM33

Q column anion exchange purification

C. cellulovorans CBM33 and other proteins isolated from *E. coli* cells were suspended in 50 mM TRIS pH8, 50 mM NaCl. At pH 8 *C. cellulovorans* CBM33 should bind to an annion exchange column according to the Expasy predicted pI. The protein solution was passed through a Q-column in the 50 mM TRIS pH8, 50 mM NaCl buffer column attached to a Fast Performance Liquid Chromatography AKTA. [50] The separated fractions were collected by automatic fractionation and the UV of each fraction measured. UV absorbance represents protein presence and indicated the fraction should be analysed further by SDS PAGE gel.

Hydrophobic Interaction Chromatography column purification

Protein fraction is concentrated by centrifugation to below 10 mls as described above. Sequential buffer exchange washes are carried out into 50 mM TRIS pH8, 50 mM NaCl, 1 M Ammonium Sulphate buffer. The protein solution was loaded onto a Phenyl sepharose Hydrophobic Interaction Chromatography column attached to a Fast Performance Liquid Chromatography AKTA. Protein that interacts strongly with the column was eluted by a decreasing concentration gradient of Ammonium sulphate from 1 M across 20 column volumes. The separated fractions were collected by automatic fractionation and the UV of each fraction measured to indicate protein presence.

Analytical ammonium sulphate cut

Proteins generally have a propensity to precipitate in larger concentrations of ammonium sulphate. The concentration of ammonium sulphate in solution was sequentially increased to analyse precipitation.

Take six Eppendorf tubes and label:

Name	Reaction components
1.5 M Ammonium sulphate pellet	0.3 mls of 4 M ammonium sulphate solution
2.0 M Ammonium sulphate pellet	0.2 mls mls of 4 M ammonium sulphate solution
2.5 M Ammonium sulphate pellet	66 mg of Amonium sulphate powder
3.0 M Ammonium sulphate pellet	66 mg of Amonium sulphate powder
3.5 M Ammonium sulphate pellet	66 mg of Amonium sulphate powder
Soluble fraction	Empty

Table 5.6 Shows the recipe for a standard Ammonium sulphate precipitation experiment.

0.5 mls of cold crude protein fraction was added to the first tube and mixed. The sample was incubated on ice for 10 minutes and then spun down on a bench top centrifuge at 13000 rpm. The supernatant was transferred into the next tube by pipette.The process was repeated for each of the following tubes and the supernatant from the 3.5 M Ammonium sulphate fraction placed into the empty tube. The pellets was dissolved in 50 mM TRIS pH8 and the protein concentration in each sample analysed by A280 and SDS-PAGE gel. [50]

Gel filtration

C. cellulovorans CBM33 protein fraction was concentrated to less than 5 ml and applied to a HiLoad 26/60 Superdex 75 column equilibrated in GF buffer (20 mM sodium acetate pH 5.0, 250 mM NaCl). The column was run at low pressure, attached to a Fast Performance Liquid Chromatography AKTA. Protein presence in automated fraction collection allowed for combination of *C. cellulovorans* CBM33 fractions.

Purification of intracellular *C. cellulovorans* CBM33

Hi-Trap crude Nickel column purification

The protein fraction was diluted in four volumes of 50 mM TRIS pH8, 50 mM NaCl, 30 mM imidazole. The solution was loaded onto a Hi-Trap crude Nickel column in the buffer column attached to a Fast Performance Liquid Chromatography AKTA equilibrated in the same buffer. [74] Solution components that bound to the Nickel resin were eluted by an increased gradient of Imidazole; increased from 30 mM to 300 mM over 20 column volumes. Elution fractions were collected by automatic fractionation and the UV of each fraction measured.

Gel filtration

C. cellulovorans CBM33 protein fraction was concentrated to less than 5 ml and applied to a HiLoad 26/60 Superdex 75 column equilibrated in GF buffer (20 mM sodium acetate pH 5.0, 250 mM NaCl). The column was run at low pressure, attached to a Fast Performance Liquid Chromatography AKTA. Protein presence in automated fraction collection allowed for combination of *C. cellulovorans* CBM33 fractions.

Purification of SUMO protease

Hi-Trap crude Nickel column purification

The protein fraction was diluted in four volumes of 50 mM TRIS pH8, 50 mM NaCl, 30 mM imidazole, 1 mM β mercapto-ethanol. The solution was loaded onto a Hi-Trap crude Nickel column in the buffer column attached to a Fast Performance Liquid Chromatography AKTA equilibrated in the same buffer. [74] Solution components that bound to the Nickel resin were eluted by an increased gradient of Imidazole; increased

from 30 mM to 300 mM over 20 column volumes. Elution fractions were collected by automatic fractionation and the UV of each fraction measured.

Gel filtration

C. cellulovorans CBM33 protein fraction was concentrated to less than 5 ml and applied to a HiLoad 26/60 Superdex 75 column equilibrated in GF buffer (20 mM sodium acetate pH 5.0, 250 mM NaCl, 1 mM dithiothreitol). The column was run at low pressure, attached to a Fast Performance Liquid Chromatography AKTA. Protein presence in automated fraction collection allowed for combination of *C. cellulovorans* CBM33 fractions.

Protein characterisation

Protein Mass Determination by Electrospray Ionisation Mass Spectrometry

The protein was buffer exchanged into 2mM TRIS on a centrifugal concentrator by repeated dilution and concentration cycles. This buffer exchange proved an essential step to remove trace salt which inhibited sample ionisation on the Mass spectrometer. The mass spectrometry and mass reconstruction were carried out by the TechFacility at the University of York.

Isothermal Titration Calorimetery

Isothermal titration calorimetry was performed using a Micro2000 -ITC calorimeter (MicroCal). Protein solution, in sodium acetate buffer, pH5, was placed in the cell at a concentration of 52.5 μM and a 10-fold more concentrated solution of $CuCl_2$ in the syringe was titrated in. Titrations were performed at 298 K. A 2 μL injection was conducted initially however was disregarded in data analysis due to spike in noise and thermal transfer. 10 μL injections were used during the titration with a 5 min interval between each injection. The $CuCl_2$ solution was made up in identical gel filtration buffer from *C. cellulovorans* CBM33 purification process. Data was analysed using the Origin 7 software package (MicroCal). Heats of dilution were subtracted from the data, but previous analysis of related LPMO proteins routinely returned sub-stoichiometric binding values with Cu. Utilising evidence from previous work the Calorimeter injection system was modified to remove Cu-containing brass components.

Electro Paramagnetic Resonance (EPR) analysis of the copper active site in *C. cellulovorans* CBM33

Copper EPR spectrum of *C. cellulovorans* CBM33

Continuous wave X-band, frozen solution, EPR spectra of 0.5 mM solution of *C. cellulovorans* CBM33 (15% v/v glycerol) at pH 5 (acetate buffer) and 155 K were acquired on a Bruker ESP 300 spectrometer operating at 9.35 GHz, with a modulation amplitude of 4G and microwave power of 5 mW.

Spectra were normalised by background subtraction for each reading and the g value calclulated using the following equation, ($g = \frac{hv}{\beta Bo}$) where h=Plank's constant, v=Frequency, β=Bohr magnetron and B_o=field [Gauss].

Copper EPR spectrum of *C. cellulovorans* CBM33 with substrate

2 mg of substrate was added to an EPR tube containing the above reaction and re-measured.

Crystallisation Trials

Attempts were made to crystallise *C. cellulovorans* CBM33 to resolve the 3 Dimensional structures by X-ray crystallography. Trials were set up using *C. cellulovorans* CBM33 concentrated to 6.4mg/ml in pH 5 sodium acetate buffers. Parallel trials were attempted to study Apo and Cu-bound *C. cellulovorans* CBM33 protein structure, $Cu(NO_3)_2\cdot 3H_2O$ was added to the copper study to a final concentration of 1 mM. Using a Mosquito robot (TTP Labtech) separate INDEX and PACT screens were set up and allowed to crystallise at 289 °K.

Mass spectrometry of *C. cellulovorans* CBM33 reaction products

Reaction of *C. cellulovorans* CBM33 with crystalline substrates

0.2% w/v solid substrate was suspended in 10 mM ammonium acetate, pH 5.0, 1 mM ascorbic acid. *C. cellulovorans* CBM33 was loaded with equi-molar amounts of copper and added to the reaction mixture to a final concentration of 1 µM. Reactions were incubated at 30 °C in a rotary incubator. Substrate was separated by centrifugation at 16,000*g* at 4 °C for 5 min and the supernatant used for the analysis.

Matrix-assisted laser desorption ionization–time of flight/time of flight Spectrometry

1 µl of protein sample was mixed with an equal volume of 10 mg/ml 2,5-dihydroxybenzoic acid in 50% acetonitrile, 0.1% trifluoroacetic acid on a SCOUT-MTP 384 target plate (Bruker). The spotted samples were then dried in a vacuum desiccator before being analyzed by mass spectrometry on an Ultraflex III matrix-assisted laser

desorption ionization–time of flight/time of flight (MALDI-TOF/TOF) instrument (Bruker), as described in Vaaje-Kolstad *et al*. _ENREF_67 Results were analysed using the Bruker Daltronics Flex Analysis software.
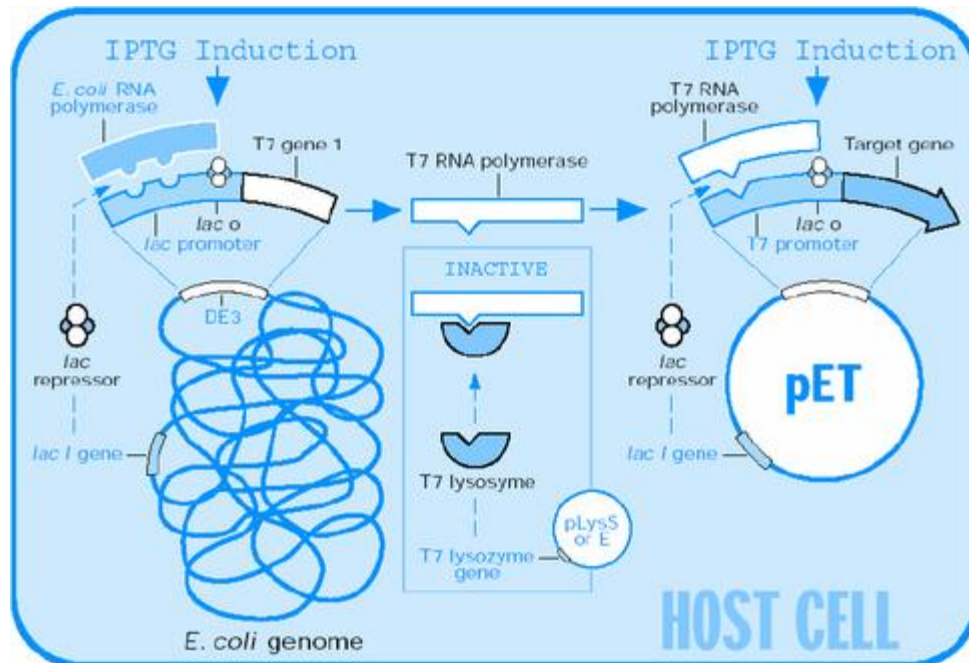
Appendix

Figure A1 Recombinant gene expression from *E. coli* expression cells.
Figure displayes the mechanism by which the addition of IPTG to the growth culture overcomes the permanent down regulation of the Lac operon genes in expression strains of *E.coli.* Lac repressor, produced from LacI gene, binds to an operator site which sterically hinders the binding and activity of RNA polymerase. IPTG binds to the LacI repressor protein allostercally changing the 3D shape and causing it to dissociate from the DNA, leaving the promoter open. *E. coli* RNA polymerase binds and transcribes large volumes of gene of interest mRNA (in this case T7 polymerase inserted into the multiple cloning site).T7 polymerase then goes on to catalyse the expression of the recombinant gene controlled by the T7 promoter. IPTG cannot be metabolised by *E. coli so once it is bound to the repressor molecule the operon genes remain active.(Figure from Novagen recombinant gene expression catalogue).*

List of abbreviations

AA9 – Auxiliary Activities 9

AA10 - Auxiliary Activities 10

APS – Ammonium Persulfate

CAZy - Carbohydrate-Active enZYmes

CBM33 – Carbohydrate binding module 33CBP21 – Chitin binding protein 21

CDH 1 – cellobiose dehydrogenase 1

CO_2 – Carbon di-oxideEJ - Exajoule

EPR – Electron paramagnetic resonance

GH61 - Glycoside Hydrolase Family 61

HIC – Hydrophobic interaction Chromatography

IPTG – Isopropyl β---D---1---thiogalactopyranoside

ITC – Isothermal calorimetry

LB – Lysogeny Broth

LPMO – Lytic polysaccharide mono-oxygenase

MALDI/TOF – Matrix-assisted laser desorption/ionization time of flight mass spectrometry

OD_{600} – Optical density 600 nm

PCR – Polymerase Chain Reaction

SDS-PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis

SUMO - Small Ubiquitin-related Modifier

TEMED TetraMethylEthyleneDiamine

Tris - Tris (hydroxmethyl) aminomethane

ULP – Ubiquitin like protein

Bibliography

Referencing in the style of JACS (Journal of the American Chemical Society) using EndNote X5.

(1)   Umbach, F. *Energy Policy* **2010**, *38*, 1229.

(2)   IEA *OECD/International Energy Agency* **2007**.

(3)   Jansen, J. C.; Seebregts, A. J. *Energy Policy* **2010**, *38*, 1654.

(4)   IEA *OECD/International Energy Agency* **2008**.

(5)   McKibben, B. In *Rolling Stone* 2012.

(6)   Brown, M. A. a. S., Benjamin K. *The MIT Press* **2011**.

(7)   Horn, S.; Vaaje-Kolstad, G.; Westereng, B.; Eijsink, V. G. *Biotechnology for Biofuels* **2012**, *5*, 45.

(8)   IEA *OECD/International Energy Agency,* **2008**.

(9)   Sathitsuksanoh, N.; George, A.; Zhang, Y. H. P. *Journal of Chemical Technology & Biotechnology* **2013**, *88*, 169.

(10)   Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H. *Science* **2004**, *306*, 2206.

(11)   Brown Jr, R. M. *Food Hydrocolloids* **1987**, *1*, 345.

(12)   Carroll, A.; Somerville, C. *Annual Review of Plant Biology* **2009**, *60*, 165.

(13)   Ding, S.-Y.; Liu, Y.-S.; Zeng, Y.; Himmel, M. E.; Baker, J. O.; Bayer, E. A. *Science* **2012**, *338*, 1055.

(14)   Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804.

(15)   Nguyen, C. ENZYMATIC CONVERSION OF CELLULOSIC BIOMASS TO CELLULOSIC ETHANOL - A REVIEW. [Online Early Access]. Published Online: 2010.

(16)   Merino, S.; Cherry, J. *Adv Biochem Eng Biotechnol* **2007**, *108*, 95

(17)   Bhutto, A. W.; Qureshi, K.; Harijan, K.; Zahedi, G.; Bahadori, A. *RSC Advances* **2014**, *4*, 3392.

(18)   Sandgren, M.; Wu, M.; Karkehabadi, S.; Mitchinson, C.; Kelemen, B. R.; Larenas, E. A.; Ståhlberg, J.; Hansson, H. *Journal of Molecular Biology* **2013**, *425*, 622.

(19)   Li, B.-Z.; Balan, V.; Yuan, Y.-J.; Dale, B. E. *Bioresource Technology* **2010**, *101*, 1285.

(20)   Reese, E. *Appl Microbiol* **1956**, *4*, 39

(21)   Reese, E.; Siu, R.; Levinson, H. *J Bacteriol* **1950**, *59*, 485

(22)   Vaaje-Kolstad, G.; Westereng, B.; Horn, S. J.; Liu, Z.; Zhai, H.; Sørlie, M.; Eijsink, V. G. H. *Science* **2010**, *330*, 219.

(23)   Harris, P.; Welner, D.; McFarland, K.; Re, E.; Poulsen, J.; Brown, K.; Salbo, R.; Ding, H.; Vlasenko, E.; Merino, S. *Biochemistry* **2010**, *49*, 3305

(24)   Quinlan, R.; Sweeney, M.; Lo Leggio, L.; Otten, H.; Poulsen, J.; Johansen, K.; Krogh, K.; Jorgensen, C.; Tovborg, M.; Anthonsen, A. *Proc Natl Acad Sci U S A* **2011**, *108*, 15079

(25)   Beeson, W.; Phillips, C.; Cate, J.; Marletta, M. *J Am Chem Soc* **2012**, *134*, 890

(26)   Langston, J.; Shaghasi, T.; Abbate, E.; Xu, F.; Vlasenko, E.; Sweeney, M. *Appl Environ Microbiol* **2011**, *77*, 7007

(27)   Park, B. H.; Karpinets, T. V.; Syed, M. H.; Leuze, M. R.; Uberbacher, E. C. *Glycobiology* **2010**, *20*, 1574.

(28)   Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic acids research* **2009**, *37*, D233.

(29)   Karkehabadi, S.; Hansson, H.; Kim, S.; Piens, K.; Mitchinson, C.; Sandgren, M. *J Mol Biol* **2008**, *383*, 144.

(30)   Vaaje-Kolstad, G.; Houston, D.; Riemen, A.; Eijsink, V.; van Aalten, D. *J Biol Chem* **2005**, *280*, 11313

(31)   Vaaje-Kolstad, G.; Bohle, L. A.; Gaseidnes, S.; Dalhus, B.; Bjoras, M.; Mathiesen, G.; Eijsink, V. G. *Journal of molecular biology* **2012**, *416*, 239.

(32)   Hemsworth, G. R.; Davies, G. J.; Walton, P. H. *Current Opinion in Structural Biology* **2013**, *23*, 660.

(33)   Vaaje-Kolstad, G.; Houston, D. R.; Riemen, A. H.; Eijsink, V. G.; van Aalten, D. M. *The Journal of biological chemistry* **2005**, *280*, 11313.

(34)   Bayer, E. A.; Belaich, J.-P.; Shoham, Y.; Lamed, R. *Annual Review of Microbiology* **2004**, *58*, 521.

(35)   Beeson, W. T.; Phillips, C. M.; Cate, J. H.; Marletta, M. A. *J Am Chem Soc* **2012**, *134*, 890.

(36)   Westereng, B.; Ishida, T.; Vaaje-Kolstad, G.; Wu, M.; Eijsink, V. G. H.; Igarashi, K.; Samejima, M.; Ståhlberg, J.; Horn, S. J.; Sandgren, M. *PloS one* **2011**, *6*, e27807.

(37)   Quinlan, R. J.; Sweeney, M. D.; Lo Leggio, L.; Otten, H.; Poulsen, J. C.; Johansen, K. S.; Krogh, K. B.; Jorgensen, C. I.; Tovborg, M.; Anthonsen, A.; Tryfona, T.; Walter, C. P.; Dupree, P.; Xu, F.; Davies, G. J.; Walton, P. H. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 15079.

(38)   Kim, S.; Ståhlberg, J.; Sandgren, M.; Paton, R. S.; Beckham, G. T. *Proceedings of the National Academy of Sciences* **2013**.

(39)   Zoia, L.; Orlandi, M.; Argyropoulos, D. *J Agric Food Chem* **2008**, *56*, 10115

(40)   Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P. M.; Henrissat, B. *Biotechnology for biofuels* **2013**, *6*, 41.

(41)   Huang, X. P.; Hudson, J.A., Rainey, F.A., Nichols, P.D., and Morgan,H.W. .

(42)   Sunna, A.; Gibbs, M. D.; Chin, C. W.; Nelson, P. J.; Bergquist, P. L. *Appl Environ Microbiol* **2000**, *66*, 664.

(43)   Watson, N. *Gene* **1988**, *70*, 399.

(44)   Charles, E. B.; Mitchell, L. *Nature Structural & Molecular Biology* **2000**, *7*, 209.

(45)   Daber, R.; Stayrook, S.; Rosenberg, A.; Lewis, M. *Journal of Molecular Biology* **2007**, *370*, 609.

(46)   Studier, F. W.; Moffatt, B. A. *J Mol Biol* **1986**, *189*, 113.

(47)   F. William Studier, S. B., New York   (US); Parichehre Davanloo, Basel (CH); Alan H. Rosenberg, Setauket, New York   (US); Barbara A. Moffatt, East Lansing, Michigan   (US); and John J. Dunn, Bellport, New York   (US) 1990.

(48)   Donahue Jr, R. A.; Bebee, R. L. *Focus* **1999**, *21*, 49.

(49)   Novagen, Ed. 2008.

(50)   Hemsworth, G. R.; Taylor, E. J.; Kim, R. Q.; Gregory, R. C.; Lewis, S. J.; Turkenburg, J. P.; Parkin, A.; Davies, G. J.; Walton, P. H. *J Am Chem Soc* **2013**, *135*, 6069.

(51)   Ogawara, H. *Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme* **1975**, *20*, 1214.

(52)   Wang, Y.; Ladunga, I.; Miller, A. R.; Horken, K. M.; Plucinak, T.; Weeks, D. P.; Bailey, C. P. *Genetics* **2008**, *179*, 177.

(53)   Elmore, Z. C.; Donaher, M.; Matson, B. C.; Murphy, H.; Westerbeck, J. W.; Kerscher, O. *BMC biology* **2011**, *9*, 74.

(54)   Mossessova, E.; Lima, C. D. *Molecular cell* **2000**, *5*, 865.

(55)   Peisach, J.; Blumberg, W. E. *Archives of Biochemistry and Biophysics* **1974**, *165*, 691.

(56)   Bunce, N. J. *Journal of Chemical Education* **1987**, *64*, 907.

(57)   Hall, J. F.; Kanbi, L. D.; Strange, R. W.; Hasnain, S. S. *Biochemistry* **1999**, *38*, 12675.

(58)   Katterle, B.; Gvozdev, R. I.; Abudu, N.; Ljones, T.; Andersson, K. K. *Biochem J* **2002**, *363*, 677.

(59)   Boswell JS, R. B., Kulathila R, Merkler D, and Blackburn NJ   *Biochemistry* **1996**, *35*.

(60)   Medda, R.; Padiglia, A.; Pedersen, J. Z.; Rotilio, G.; Finazzi Agro, A.; Floris, G. *Biochemistry* **1995**, *34*, 16375.

(61)   Gacheru, S. N.; Trackman, P. C.; Shah, M. A.; O'Gara, C. Y.; Spacciapoli, P.; Greenaway, F. T.; Kagan, H. M. *J Biol Chem* **1990**, *265*, 19022.

(62)   Chan SI, C. K., Yu SS, Chen CL, Kuo SS *Biochemistry* **2004**, *43*, 4421.

(63)   Rotilio, G.; Morpurgo, L.; Giovagnoli, C.; Calabrese, L.; Mondovi, B. *Biochemistry* **1972**, *11*, 2187.

(64)   Vaaje-Kolstad, G.; Westereng, B.; Horn, S. J.; Liu, Z.; Zhai, H.; Sorlie, M.; Eijsink, V. G. *Science* **2010**, *330*, 219.

(65)   Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem J* **2004**, *382*, 769.

(66)   Gyllensten, U. B. *BioTechniques* **1989**, *7*, 700.

(67)   Hanahan, D. *Journal of Molecular Biology* **1983**, *166*, 557.

(68)   Clonetech: 2013.

(69)   Abd-Elsalam, K. A. *african Journal of biotechnology* **2003**, *2*, 91.

(70)   QIAGEN; QIAGEN, Ed. 2010.

(71)   Casadaban, M. J.; Cohen, S. N. *Journal of Molecular Biology* **1980**, *138*, 179.

(72)   Murphy, M. B.; Fuller, S. T.; Richardson, P. M.; Doyle, S. A. *Nucleic acids research* **2003**, *31*, e110.

(73)   Siebenkotten, G.; Leyendeckers, H.; Christine, R.; Radbruch, A. *Qiagen News* **1995**, *2*, 11.

(74)   Hengen, P. N. *Trends in Biochemical Sciences* **1995**, *20*, 285.