Cloning, Expression, Purification and Characterization of Lytic Polysaccharide Monooxygenases from Streptomyces coelicolor (A3)2 and Jonesia denitrificans

sophanit mekasha

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Science
Master Thesis 60 credits 2013

Cloning, expression, purification and characterization of lytic polysaccharide monooxygenases from *Streptomyces coelicolor* (A3)2 and *Jonesia denitrificans*

Masters Thesis

By Sophanit Mekasha

Protein Engineering and Proteomics Group

Department of Chemistry, Biotechnology and Food Science

The Norwegian University of Life Sciences

2013

AKNOWLEDGEMENS

This dissertation would not have been possible without the help and contribution of many great individuals and institutions. It is produced as a result of great collaboration with the Protein Engineering and Proteomic group at UMB and Schmidt-Dannert lab at the University of Minnesota.

Primarily my utmost gratitude goes to my advisor, Dr. Gustav Vaaje-Kolstad for his countless advise, patience, encouragement and support. This research would not have been possible without his invaluable guidance. I would also like to thank him for reading and commenting on this manuscript. I would also like to thank my supervisor, Zarah Forsberg for providing me with all the support and direction.

I am deeply indebted to Professor Vincent Eijsink for his encouragement, consistent support in carefully reading and commenting on revisions of this manuscript and for giving me the chance to visit the University of Minnesota. I am extremely grateful to Professor Claudia Schmidt-Dannert for opening her laboratory and collaborating in most of the molecular biology work, and for sharing her knowledge to my success. Her invaluable discussions, encouragement and guidance have helped me structure the technical details of my work. I am also thankful to members of Schmid-Dannert's lab, who created warm working environment during my visit and for all the assistance they extended to me.

My deepest gratitude extends to all fellow lab mates who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever. I also sincerely appreciate the financial support from NOCC that funded my visit to Minnesota where I conducted parts of the research discussed in this dissertation.

Last but not least, none of this would have been possible without the love and support of my immediate family, to whom this dissertation is dedicated. I would like to express my heart-felt gratitude for their constant source of love, concern, support and strength all these years. My extended family and friends have also aided me throughout this endeavor.

Ås, March 15th 2013

Sophanit Mekasha

ABSTRACT

Enzymatic conversion of the abundant and recalcitrant polysaccharides cellulose and chitin is an important step in the conversion of biomass to valuable products. Hence, development of enzyme technology for biomass saccharification is important. It is well known that degradation of cellulose and chitin requires synergistic action of hydrolytic enzymes (cellulases and chitinases). Recently, it has been discovered that another class enzymes makes an important contribution to the degradation process, namely the Lytic Polysaccharide Monooxiginases or LPMOs. These enzymes have the ability to cleave crystalline cellulose or chitin in an oxidative manner. While LPMO-encoding genes are abundant, only few enzymes have been characterized. The catalytic mechanism of these enzymes remains elusive. There is thus a clear need for further characterization and comparative studies of diverse LPMOs.

The current study focuses on characterization of two LPMO-domains from different Gram-positive bacteria with different modular structure and substrate specificities. These are: Cels2, a cellulose active enzyme from *Streptomyces coelicolor* comprising an LPMO and a CBM2 domain and Jden1381 a putatively chitin-active multidomain protein from *Jonesia denitrificans* consisting of an LPMO, a CBM5/12 and a GH18 (chitinase) domain.

Mutational characterization of CelS2 function by probing conserved residues (Arg212, Ser215, and Phe219) in the predicted catalytic site showed the requirement of these residues for enzymatic catalysis. Analysis of activity on phosphoric-acid swollen cellulose using both mass spectrometry and HPLC for detection of soluble products, showed that the R212A and F219Y mutations inactivated the enzyme, whereas the S215A and F219A mutations reduced activity by approximately 50 and 85 %, respectively.

So far, the existence of active LPMOs attached with chitinases in a single protein has not been reported. The existence of such combinations suggest a highly chitinolytic potential for the chitinase. To assess this potential a codon-optimized gene encoding full length *Jden*1381 was cloned in *E.coli* using both pET32b and pUCBB-eGFP expression vectors. For analysis of individual effects of the domains, a variety truncated versions of Jden1381 were expressed in pUCBB-eGFP. Of the five Jden1381 variants expressed in this study, five yielded soluble protein and four were purified and characterized. For characterization of the full length Jden1381, a crude extract was used. Analysis of product profiles using UHPLC and MALDI-TOF mass spectrometry showed chitinolytic activity on α-chitin, β-chitin and collidal chitin exerted by both the N-terminal LPMO domain (generating oxidized products) and the C-terminal GH18 domain (generating native products).

In conclusion, this study provides novel insights into the catalytic mechanism of LPMOs, from the CelS2 work, while the studies on Jden1381 show, for the first time, that nature has developed multi-modular enzymes comprising both LPMO and GH domains acting on the same substrate.

SAMMENDRAG

Enzymatisk nedbrytning av polysakkaridene cellulose og kitin er en essensielt trinn i prosessen hvor biomasse konverteres til verdifulle produkter. Utvikling og forbedring av enzymteknologi for omdanning av biomasse er derfor viktig for å forbedre prosessene, samt for å få en bedre forståelse for hvordan enzymene virker. Det er velkjent at effektiv nedbrytning av cellulose og kitin krever samspill mellom komplementerende hydrolytiske enzymer (glykosid hydrolaser; cellulaser og kitinaser). Nylig har også en ny klasse enzymer viktig for biomassenedbrytning blitt oppdaget nemlig lytisk polysakkarid monooxyginaser (LPMOer). Disse enzymer spalter krystallinsk cellulose eller kitin ved hjelp av en oksidativ mekanisme. I kombinasjon med vanlige glykosid hydrolaser øker LPMOene nedbrytningshastigheten av biomasse, hvilket gjør disse enzymene meget interessante for biomas relatert enzymteknologi. Siden oppdagelsen av LPMOer ble gjort for kun 3 år siden har bare noen få enzymer blitt karakterisert. Samtidig er den katalytiske mekanismen fortsatt ukjent. Det er derfor et stort behov for ytterligere karakterisering av ulike LPMOs.

Denne studien fokuserer på karakterisering av to LPMO-moduler fra ulike Gram-positive bakterier. Det første enzymet er Cels2, en cellulose aktiv LPMO fra *Streptomyces coelicolor* bestående av en LPMO katalytisk module og en CBM2 cellulose-bindene modul. Det andre enzymet er Jden1381 som er en ukarakterisert kitin-aktivt multidomene enzyme fra *Jonesia denitrificans* bestående av en LPMO katalytisk module, en CBM5/12 kitin-bindenen modul og en GH18 katalytisk modul (kitinase).

Funksjonen til tre konserverte aminosyrer forbundet med det aktive setet til CelS2 (Arg212, Ser215, og Phe219) ble karakterisert ved mutagenes og analyse av aktivitet av Aktivitet mot PASC (phosphoric-acid swollen cellulose) ble målt med både massespektrometri og HPLC og det viste seg at mutasjonene R212A og F219Y inaktiveres enzymet, mens mutasjonene S215A og F219A redusert aktivitet med henholdsvis 50 og 85% sammenliknet med villtype enzymet.

Det andre enzymet analysert i dette studiet, Jden1381, representerer et hittil ukarakterisert kombinasjon av kitin-aktive katalytiske moduler, nemlig en LPMO kombinert med en kitinase (GH18) . Siden det er kjent at LPMOer og kitinaser viser sterk synergi sammen, bør det kitinolytiske potensiale til et slikt multidomeneprotein være stort.

ABBREVIATIONS

CAZy	Carbohydrate Active Enzyme database
CBM	Carbohydrate Binding Module
CBP	Chitin binding protein
CelS2	LPMO coding gene from *Streptomyces coelicolor*
DEAE	Diethylminoethyl cellulose
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DP	Degree of Polymerization
GH	Glycoside Hydrolases
His tag	Hexa-histidine tag
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl-β-D-thio-galactoside
Jden1381fl	Chitinase coding gene from *Jonesia denitrificans*, full length
Jden1381fl_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, full length, C-terminally His tagged
Jden1381-LPMO	Chitinase coding gene from *Jonesia denitrificans*, N-terminal single domain
Jden1381-LPMO_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, N-terminal single domain, C-terminally His tagged
Jden1381-LPMO-CBM5/12	Chitinase coding gene from *Jonesia denitrificans*, N-terminal multi- domain
Jden1381-LPMO-CBM5/12_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, N-terminal single domain, C-terminally His tagged
Jden1381-CBM5/12	Chitinase coding gene from *Jonesia denitrificans*, single middle domain
Jden1381-CBM5/12_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, single middle domain, C-terminally His tagged
Jden1381-CBM5-GH18	Chitinase coding gene from *Jonesia denitrificans*, C-terminal multi- domain
Jden1381-CBM5-GH18_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, C-terminal multi- domain, C-terminally His tagged
Jden1381-GH18	Chitinase coding gene from *Jonesia denitrificans*, C-terminal single domain
Jden1381-GH18_C-His_6	Chitinase coding gene from *Jonesia denitrificans*, C-terminal single domain, C-terminally His tagged
LPMO	Lytic Polysaccharide Monooxiginase
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization-Time of Flight
OD_{600}	Optical Density at 600 nano meter
PCR	Polymerase Chain Reaction
PDB	Protein structure database
PMSF	Phenylmethylsulfonyl fluoride
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TAE	Tris-acetate-EDTA
Tris	Tris(hydroxymethyl)aminomethane
TRX tag	Thioredoxin tag

SI UNITS

µg	Micro gram
µl	Micro liter
µM	Micro molar
Kb	Kilo-base
kDa	Kilo-Dalton
ng	Nano gram
psi	Pounds per second inch pressure unit
v/v	volume/volume
w/v	weight/volume

Table of Contents

1 INTRODUCTION 1
1.1 Polysaccharides 1
1.1.1 Structure and classification of polysaccharides 1
1.1.2 Degradation of polysaccharides in nature 2
1.2 Chitin and cellulose – structure and classification 2
1.2.1 Structure and classification of cellulose 2
1.2.2 Structure and classification of chitin 5
1.3 Glycoside hydrolases (EC 3.2.1.x) – classification 6
1.4 Enzymatic degradation of cellulose and chitin 7
1.4.1 Cellulose degradation 8
1.4.2 Chitin degradation 8
1.5 The modularity of glycoside hydrolases 9
1.6 Carbohydrate Binding Modules (CBMs)- Function, classification and structure 10
1.7 Lytic polysaccharide monooxygenases (LPMOs) 11
1.7.1 Structure and function 13
1.8 Carbohydrate degradation by *Streptomyces coelicolor* 17
1.9 Carbohydrate degradation by *Jonesia denitrificans* 19
1.10 Protein Expression 21
1.10.1 Gene optimization and modification 21
1.10.2 Expression vectors 22
1.10.3 Host strains 22
1.10.4 Protein secretion 23
1.11 Protein purification 24
1.11.1 Ion-exchange chromatography (IEC) 24
1.11.2 Size exclusion chromatography (SEC) 25
1.11.3 Affinity-chromatography 26
1.12 Protein characterization 27
1.12.1 Experimental techniques 27
1.12.2 Bioinformatics techniques 28
1.13 Aim of this project 30
2 MATERIALS 31

2.1 Laboratory equipment 31
2.2 Chemicals 32
2.3 Proteins and enzymes 33
2.4 DNA 34
2.5 Carbohydrate substrates 34
2.6 Kits 34
2.7 Primers 36
2.8 Bacterial Strains 38
2.9 Plasmids 39
3 METHODS 40
3.1 Microbiology methods 40
3.1.1 Cultivation media 40
3.1.2 Cultivation of bacterial strains 43
3.1.3 Long-term storage of bacterial strains 44
3.2 Molecular biology methods 45
3.2.1 Plasmid isolation using the NucleoSpin® Plasmid kit 45
3.2.2 Plasmid purification from E.coli using the Wizard® Plus SV miniprep DNA purification system 46
3.2.3 Polymerase Chain Reaction-based methods 48
3.2.4 Agarose gel electrophoresis 54
3.2.5 Extraction and purification of DNA fragments from agarose gels 55
3.2.6 Restriction digestion 56
3.2.7 Ethanol/EDTA Precipitation of DNA 58
3.2.8 Cloning 59
3.2.9 Competent cells for transformation 63
3.2.10 DNA sequencing 65
3.3 Protein Expression 67
3.3.1 Cultivation of BL21 (DE3) cells for optimal expression 67
3.3.2 Stimulation of transcription by IPTG induction 67
3.4 Protein extraction 68
3.4.1 Periplasmic extracts of *E. coli* 68
3.4.2 Cytoplasmic extract of E. coli 69
3.5 Protein purification 70
3.5.1 Ion-Exchange Chromatography 70

3.5.2 Size Exclusion Chromatography (SEC) 71
3.5.3 Protein purification by immobilized metal ion affinity chromatography 72
3.6 Protein concentration and concentration measurement 74
3.6.1 Protein concentration 74
3.6.2 Concentration measurement 75
3.7 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 76
3.8 Analysis of enzyme activity 77
3.8.1 Matrix-Assisted Laser Desorption and Ionization Time of Flight- mass spectroscopy (MALDI-TOF MS) 77
3.8.2 High Performance Liquid Chromatography (HPLC) 79
3.9 Bioinformatics methods 83
4 RESULTS 85
4.1 Bioinformatic analysis of CelS2 and Jden1381 85
4.1.1 Domain structure and physiochemical properties 85
CelS2-N from Streptomyces coelicolor 86
4.1.2 Multiple sequence alignment (MSA) 86
4.1.3 Homology modeling 87
Jden1381 from *Jonesia denitrificans* 89
4.1.4 Sequence alignment 89
4.1.5 Gene optimization 91
4.2 Mutagenesis, molecular cloning and transformation 91
4.2.1 Site Directed Mutagenesis of CelS2 91
4.2.2 Gene cloning of Jden1381 and gene truncation 92
4.3 Protein expression and purification 95
4.3.1 Protein expression of CelS2-N mutants 95
4.3.2 Expression of Jden1381 96
4.3.3 Protein purification 98
4.4 Enzyme characterization - CelS2-N 102
4.4.1 Choice of substrate 102
4.4.2 Mapping of enzymatic activity of CelS2- R212A, S215A, F219A and F219Y mutants by HPAEC 103
4.4.3 Analysis of initial rates of $CelS2_{WT}$ and $CelS2_{S215A}$ 105
4.5 Enzyme characterization - Jden1381 106
4.5.1 Analysis of chitooligosaccharides released by full length Jden1381 107

4.5.2 MALDI-TOF MS analysis of oligosaccharides released by the Jden1381 LPMO domain. 108
5 DISCUSSION **114**
6 REFERENCES **126**
APPENDICES **133**

1 INTRODUCTION

1.1 Polysaccharides

Carbohydrates are the most abundant biological molecules in nature. These molecules have a general formula $(CH_2O)_n$ where "n" represents the number of carbons present in the sugar molecule. Some carbohydrates contain sulfur (e.g. glycosaminoglycans or GAGs) or nitrogen (e.g. chitin). In their simplest form carbohydrates are monosaccharides but often they occur as oligo- or polysaccharides. Polysaccharides include insoluble crystalline compounds such as cellulose and chitin, which serve as structural components in the cell walls of plants, microorganisms and crustaceans. In addition, polysaccharides serve as building materials for microbial capsules, providing resistance to stressful conditions such as dehydration and playing a role in environmental interactions (Davies *et al.* 1997, Senni *et al.* 2011, Zaragoza *et al.* 2009 & Zeltins *et al.* 1995).

1.1.1 Structure and classification of polysaccharides

Polysaccharides are built through glycosidic linkage of monosaccharides leading to formation of either branched or linear sugar chains. The large assortment of monosaccharides combined with the many potential coupling and branching points provides an enormous diversity. Polysaccharides are divided into two classes; homo-polysaccharides (composed of identical monosaccharide units) and hetero polysaccharides (composed of two or more types of monosaccharides). Homo-polysaccharides can be further classified by the type of linkage joining the monosaccharide units. The linkage can be either an α- or a β- depending on the configuration of the hydroxyl group joining the monomers (Lindhorst *et al.*, 2007, Robyt, J. F. (1997) & Varki *et al*., 1999). The boundary between oligosaccharide and polysaccharide is vague. Usually, oligosaccharides are considered to contain up to ~20 monosaccharide units while polysaccharides contain 20 or more monosaccharide units (www.newworldencyclopidea.org).

1.1.2 Degradation of polysaccharides in nature

In nature, polysaccharides are degraded through enzymatic cleavage of the glycosidic bonds that connect the monomers. The enzyme reactions are mostly hydrolytic, but can also involve beta-elimination (polysaccharide lyases) or oxidation (lytic polysaccharide monooxygenases). Enzymes responsible for hydrolyzing polysaccharides are called glycoside hydrolases. These enzymes are produced by different prokaryotes and eukaryotes for different reasons. Polysaccharides like cellulose, chitin and various hemicelluloses are among the main polysaccharides that are subjected to enzymatic degradations. The diversity of polysaccharides and their complex structures have resulted in the existence of huge diversity of enzymes for the degradation of these bio-molecules. Glycoside hydrolases are named after their target substrates; for example, hydrolytic enzymes for cellulose and chitin are called cellulases and chitinases, respectively. Organisms normally produce multiple enzymes for efficient degradation of their target polysaccharide substrates (Béguin *et al.,* 1994, Min *et al.* 2012 & Warren 1996). In this study, we focus on the enzymatic conversion of two robust and abundant polysaccharides, cellulose and chitin.

1.2 Chitin and cellulose – structure and classification

Cellulose and chitin are the most abundant, insoluble bio-polymers found in nature with closely related chemical structure, function and mode of polymerization. Cellulose serves as structural component of plant cell walls (Fig 1.2), while chitin is the major constituent of fungal cell walls, cuticles of insects, exo-skeletons of crustaceans and the cell walls of zoo-plankton and some algae (Wilson 2009, Saito *et al*., 2000 & Raabe *et al.*, 2006). While cellulose is composed of glucose, chitin is composed of *N*-acetylglucosamine, i.e. a glucose variant where the C2 sugar is substituted with an acetamido group.

1.2.1 Structure and classification of cellulose

Cellulose is composed of glucopyranose units linked by β-1, 4-glycosidic bonds in a linear chain. Cellulose synthesis starts from polymerization of two β-glucopyranose units where the polymerized unit is called cellobiose. In this reaction, the β-glucopyranose is linked with a hydroxyl at the 4^{th} position of another glucose residue producing cellobiose (Haworth 1937). Further polymerization reactions lead to formation of cellulose. The orientation of the

monomeric units in cellulose is in a "flip-flop" manner, where each monomer is rotated 180^{o} relative to the following residue. The polymerization produces long rigid chains with a reducing end and a non-reducing end (Fig. 1.1) (Filipponen 2009 & Teeri 1997).

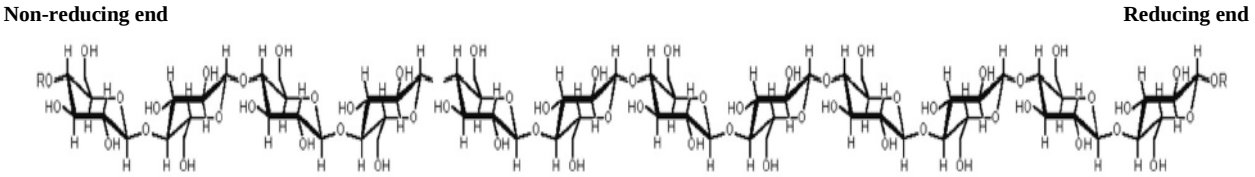


Figure 1.1. Structure of a cellulose chain. Cellulose is a linear polysaccharide containing hundreds to thousands of β1→4 linked glucose units. Each glucose unit is inverted 180 ^{o} relative to the preceding glucose unit (Picture source: Horn *et al.*, 2012).

Cellulose is found in plants and is also present in fungi and algae. Cellulose is also synthesized by certain prokaryotes (e.g. *Acetobacter*, *Rhizobium,* and *Agrobacterium*) (Dewick 2009, Brown R.M., JR. 2003 & O'Sullivan 1997). Cellulose produced by living organisms is called native cellulose. The number of glucose units in native cellulose is diverse depending on the source, such as primary or secondary cell walls. Cellulose polymers from primary cell walls contain about 8000 glucose units while secondary cell walls contain about 15000 glucose units per chain. The number of glucose units required to form an insoluble product is approximately eight. Polymers that contain more than eight glucose units have greater affinity to one another than to water (Brown R.M., JR., 2003).

Plant fibers are composed of cellulose chains that are highly ordered into nanostructures known as microfibrils. Aggregates of these microfibrils produce macro-fibrils. The quantity of microfibrils in macrofibrils is diverse depending on the source. Bundles of these macrofibrils produce cellulose fibers (Fig 1.2b) (Brown, JR. 2003 & Donaldson, 2007).

The microfibrils of cellulose are embedded in a matrix of complex heteropolymeric network that contains hemicellulose and pectin. Hemicelluloses are branched polysaccharides containing mainly pentoses (like xylose and arabinose) and hexoses (like mannose, glucose and galactose) that can form hydrogen bonds to the surface of cellulose fibrils. Pectin is composed primarily of uric acids as galacturonic acid units. Pectin glue cells together creating layer known as middle lamella that connects two plant cells (Fig 1.2a) (Cosgrove 2005, Somerville *et al.,* 2004 & Taherzadeh *et al.,* 2008).

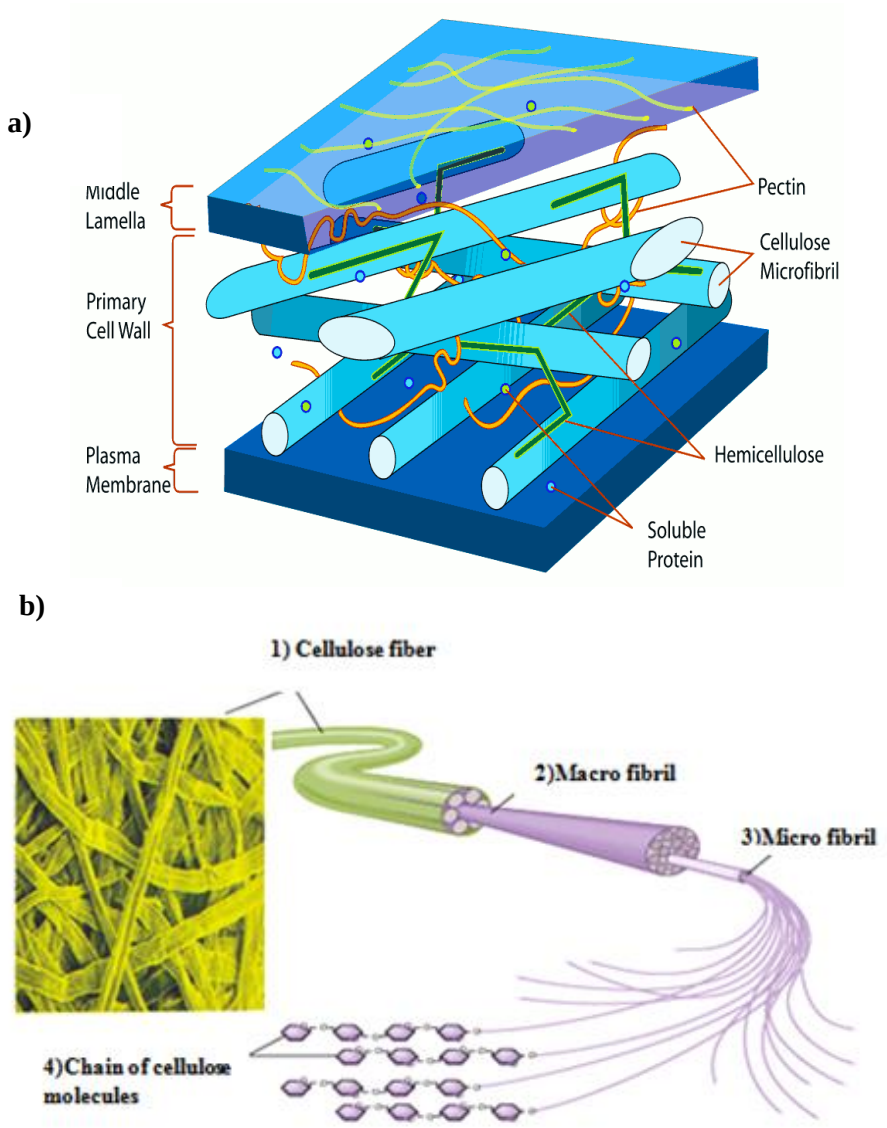


Fig 1.2. Organization of cellulose in plant cell walls. a**)** Structure of cell wall. The figure shows the components of primary plant cell wall. The cell wall is enclosed with plasma membrane and a layer called middle lamella. Cellulose microfibrils are embedded with complex heteropolymers such as pectins and hemicellulose. The cellulose-heterosaccharide complex is surrounded with proteins. b**) 1)** Cellulose fiber from ponderosa pine. **2)** Cellulose fibers contain macro-fibrils which comprise bundles of micro-fibrils (**3**). **4)** Micro-fibrils consist of bundle of glucose polymers or cellulose chains. Picture (source: a) http://www.wpclipart.com/plants/diagrams/Plant_cell_wall_diagram.png.html b) Modified picture from http://nutrition.jbpub.com/resources/chemistryreview9.cfm.

Cellulose occurs in different crystalline forms: highly ordered crystalline regions, pseudo-ordered regions (para-crystalline) and so called amorphous regions representing disordered or non-crystalline cellulose chains (Boraston *et al.*, 2002). A simplified illustration of cellulose chain arrangements in amorphous and crystalline regions is shown in Figure 1.3.

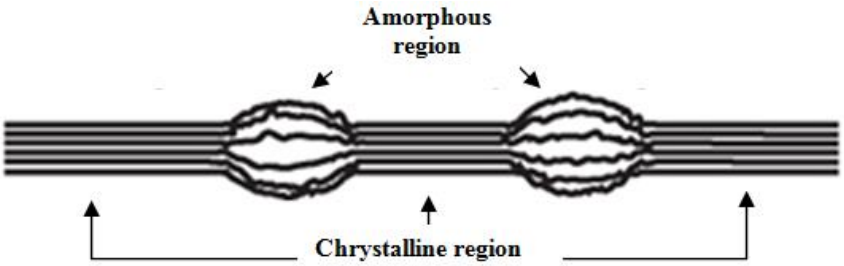


Fig 1.3. Simplified structural view of cellulose showing the difference between amorphous and crystalline regions of cellulose (source: Modified figure from Oke 2010)

Native cellulose is classified into several classes according to the alignment of cellulose chains. Most commonly, cellulose is classified into cellulose I and cellulose II. Cellulose I, (which is the most common type) consists of glucose chains that are aligned in parallel while cellulose II consists of anti-parallel chains (Brown, R. M., JR 2003).

1.2.2 Structure and classification of chitin

Following cellulose, chitin is the second most abundant insoluble bio-molecule on earth and is widely distributed among unlike organisms. This molecule is a result of polymerization of *N*-acetyl-D-glucosamine or GlcNAc via β-1,4- glycosidic bonds. The length of chitin polymers varies from 100-8000 depending on the organism. As in cellulose, each chitin monomer is rotated 180^{o} relative to its preceding residue (Fig 1.4) (Carlstrom 1957, & Mulisch 1993).

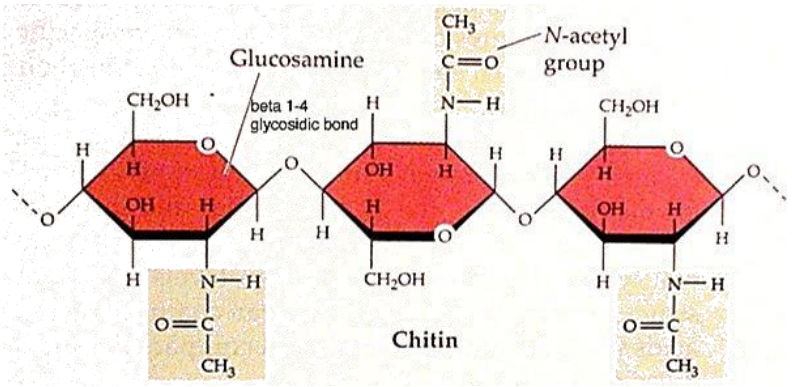


Fig 1.4. Chemical structure of chitin (Source: Academic Brooklyn cuny http://academic.brooklyn.cuny.edu/biology/bio4fv/page/chitin.jpg)

Chitin chains tend to be arranged in similar was as in cellulose. Chitin fibers from cuticle of lobster contain chitin chains that are associated with protein, forming nanofibrils that are 2-5

nm in diameter. Cluster of these nanofibrils form chitin-protein fibers (Fig 1.5) (Raabe *et al.*, 2006).

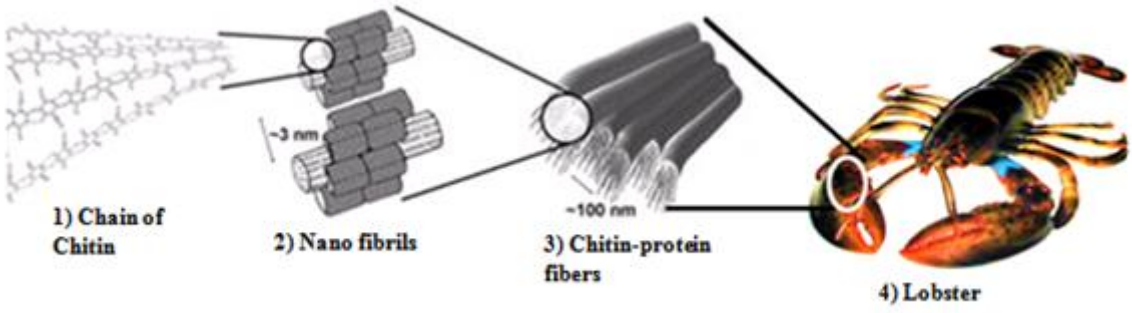


Fig 1.5. Organization of chitin in lobster cuticles. 1) Chains of N-acetyl-glucose amine form chitin chains **2**) Chitin chains are wrapped with proteins, forming nano-fibrils which are ~ 3nm in diameter **3**) These nano-fibrils cluster to form chitin-protein fibers which are ~100 nm in diameter. **4)** Lobster *Homarus americanus* (Picture source: Modified picture from D. Raabe *et al.*, 2006).

In resemblance to its analogue, cellulose, chitin chains are organized in highly ordered manner creating the crystalline structure of the bio-polymer. Crystallinity and morphology of chitin may vary. According to the orientation of chains, chitin is classified into α-chitin (anti-parallel), β-chitin (parallel) and γ-chitin (mixture of parallel and anti-parallel). Among these classes α-chitin is the most common and rigid type (Mulisch 1993). A more soluble derivative of chitin, chitosan, is obtained when acetyl groups are removed, converting the GlcNAc units to glucosamine. Although the definition of chitin vs. chitosan based on the degree of acetylation (DA) is vague, some researchers state that the DA of chitin is greater than 50 % while chitosan has DA value lower than 50 %. Chitosan is less crystalline than chitin, more accessible to aqueous solvents, and under some conditions, water soluble. The latter is due to the fact that the amino group of glucosamine has a pKa of ~ 6.5, which gives these sugars a positive charge at mildly acidic pH. Hence, chitosan is soluble at mildly acidic pH. In nature chitin has tendency to be covalently or non-covalently bound to other macromolecules like protein, carotenoids and glucans (Zhang *et al.*, 2012, Tharanathan *et al.*, 2003).

1.3 Glycoside hydrolases (EC 3.2.1.x) – classification

Enzymes that have polysaccharide hydrolytic activity are called glycoside hydrolases (GHs). The International Union of Biochemistry and Molecular Biology (IUBMB) Nomenclature Committee gave these enzymes Enzyme Commission number EC 3.2.1.x where “x”

represents the substrate. However, this nomenclature has its limitations in representing enzymes with broader specificities and does not take into account the concepts of evolutionary heritage leading to sequence and structural similarities. These limitations led to development of an alternative classification of these enzymes into different families based on their amino acid sequences (Henrissat 1991).

Enzymes that participate in carbohydrate degradation and/or modification are classified in the carbohydrate active enzymes (CAZy) database (Cantarel *et al.* 2009) where enzymes are organized in families based on their amino acid similarity i.e. evolutionary relationship. Members of one family have similar overall structures and catalytic machineries, but may have very different binding and/or catalytic activities. The information of this database is growing due to continuous discovery and characterization of relevant enzymes. Per February 2013, the CAZy database listed131 glycoside hydrolase (GH) families. Chitinases belong to families 18 and 19, cellulases fall into families 1, 3, 5, 6, 7, 8, 9, 12, 16, 44, 45 and 48 while chitosanases occur in families 5, 8 and 46 (http://cazy.org, Warren (1996) & Thu *et al.*, 2010).

The CAZy database classifies several other enzymes types and non-enzymatic carbohydrate-binding domains, and several of these are needed for efficient polysaccharide degradation. This is discussed in more detail, below.

1.4 Enzymatic degradation of cellulose and chitin

Enzymatic degradation of cellulose and chitin is challenging due to the insolubility and crystallinity of these bio-polymers (Schwarz 2001). Therefore, cellulolytic and chitinolytic bacteria and fungi often produce an array of enzymes that work together. Previous studies have shown that the degradation of chitin and cellulose involves the same four types of hydrolytic enzymes. Three of these enzymes act on the polymers and are classified as endo- and exo- types according to their modes of attack on the polysaccharide chain (Fig 1.6) (Horn *et al.,* 2006 & Teeri 1997). The fourth enzyme-type is a beta-glycosidase, that converts the oligomeric (mainly dimeric) products from the other three enzymes into monomeric sugars.

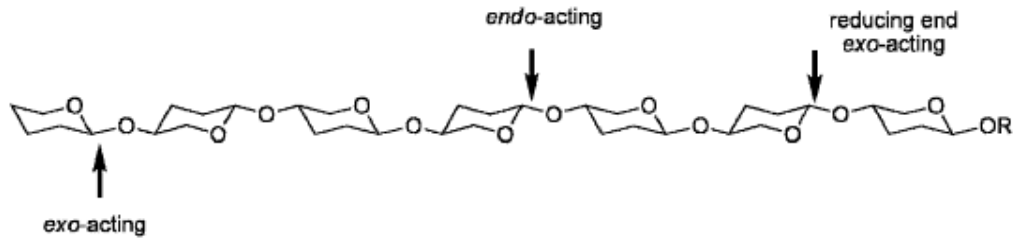


Fig 1.6. Structural scheme of binding and catalytic sites for endo- and exo- acting glycoside hydrolases (Picture source: http://www.cazypedia.org/index.php/Glycoside_hydrolases)

1.4.1 Cellulose degradation

Cellulases are sub-classified and named after their mode of action. The main enzyme classes are (1) endo-β-(1,4)-glucanases, which cleave randomly within the chain producing new chain ends, and (2) exo- β-(1,4)-D-glucanases, also known as cellobiohydrolases (CBH), which release cellobiose units from chain ends, in a processive manner. There are two types of CBH, known as CBHI and CBHII. CBHI acts on reducing ends of the chain while CBHII acts on the non-reducing ends (Fig. 1.6). A β-D-glucosidase, also known as cellobiase converts the cellobiose produced by the other glucanases to glucose. These enzymes work synergistically in the degradation of cellulose. Studies have shown that of the three polymer-active enzyme types, the exo-glucanases catalyze most of the bond cleavages during the saccharification of crystalline cellulose (Brown R.M., JR. 2003).

Among the families of GHs listed on CAZy database, families 6, 7, 8, 9, 44, 45 and 48 mainly or only comprise cellulases. Cellulase are also found in families 1, 3, 5, 12 and 16 that contain enzymes with a wide range of substrate specificity.

1.4.2 Chitin degradation

Chitin degrading enzymes belong to families 18 and 19 of the glycoside hydrolases. The role of family 19 enzymes in biomass conversions is rather unclear, whereas family 18 enzymes occur in chitinolytic machineries that resemble cellulolytic machineries. As cellulases, these family 18 chitinases include exo- and endo acting enzymes. *Serratia marcescens,* one of the best studied chitinolytic microorganisms, produces three GH18 enzymes when grown on

chitin, namely ChiA (exo-chitinase working from the reducing end), ChiB (exo-chitinase working from the non-reducing end) and ChiC (endo chitinase) (Fig 1.7).

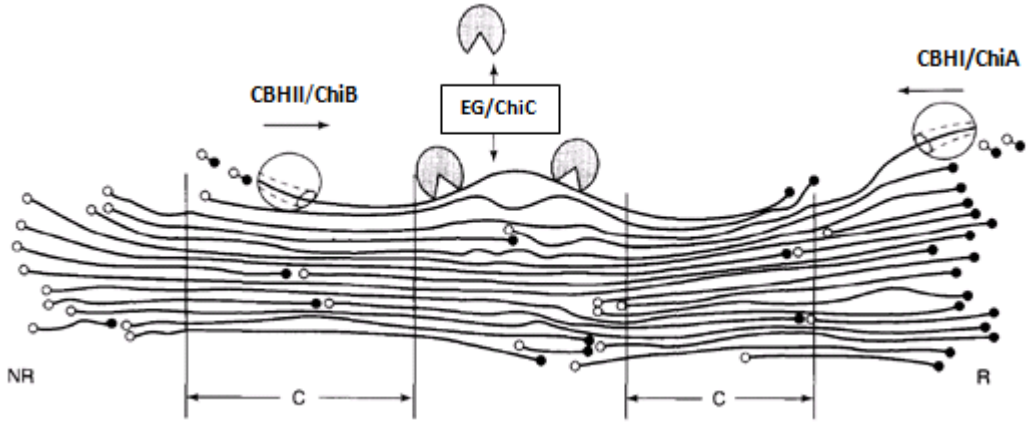


Fig 1.7. Schematic representation of the attack points of cellulases/chitinases during cellulose and chitin hydrolysis. Endo-glucanases (EG) and ChiC hydrolyze within the cellulose and chitin chains respectively. Endo acting enzymes preferably attack amorphous regions, producing new chain ends. CBHI and ChiA attack from the reducing end and degrade processively towards the non-reducing end producing cellobiose and chitobiose, respectively. CBHII and ChiB and attack from the non-reducing end of their respective substrate and degrade processively towards the reducing end. C indicates the most rigid or crystalline section of the cellulose/chitin. NR = non-reducing end R = Reducing end. White open circles represent non-reducing ends. Black circles represent reducing ends. The dimeric products, which may reach large concentrations, are inhibitory for the cellulases/chitinases and their conversion to monomeric sugars by β-glucosidases is thus important (not shown in figure) (Source: Modified figure from Teeri 1997).

1.5 The modularity of glycoside hydrolases

Catalytic modules of GHs are often linked with one or more carbohydrate binding modules (CBMs). CBMs are mostly involved in mediating contact between the catalytic module and the substrate (see section 1.6 for more details) Some GHs may also be linked with modules having other functions than carbohydrate/substrate binding in addition to CBMs. Esterase, fibronectin type III-like and dockerin are examples of non-CBM domains attached to GHs (Fig. 1.8) (Henrissat *et al.*, 2000, Forsberg *et al.* 2011 & Vaaje-Kolstad *et al.* 2005a).

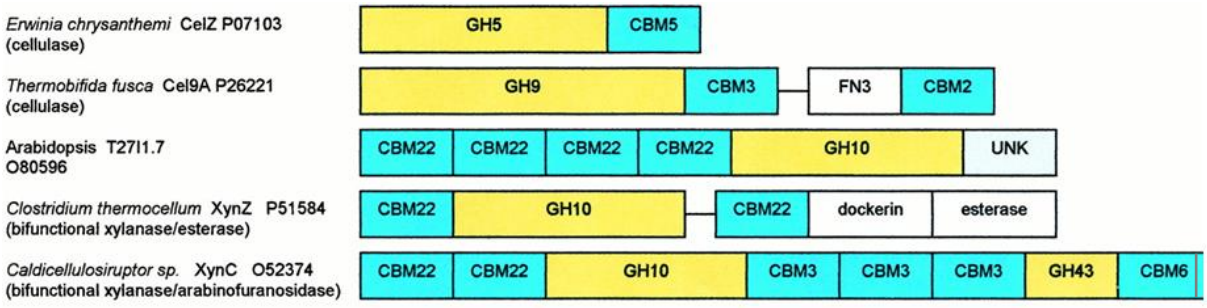


Fig 1.8. Example of the modular structures of glycoside hydrolases. Yellow boxes represent the catalytic domain of glycoside hydrolases (GHs) from different families. The blue boxes represent carbohydrate binding modules of diverse families (CBMs). Light gray blue box labeled "UNK" represents domains of unknown function. Fn3 (Fibronectin type III – like) domains, an esterase domain and a dockerin (possibly involved in cellulosome formation) are represented as white boxes. Defined linker peptides that separate domains are represented by black lines (Source: Modified figure from Herrisat *et al.*, 2000).

1.6 Carbohydrate Binding Modules (CBMs)- Function, classification and structure

An important factor that promotes efficient hydrolysis of chitin and cellulose by glycoside hydrolases is the ability of the enzymes to bind and attach to the substrate. The crystalline structure and insoluble nature of chitin and cellulose makes this a challenge for the enzymes. Once the enzymes have attached to the substrate it may be beneficial for them to remain attached. Most catalytic modules are thus equipped with CBMs that mediate substrate binding and optimize contact between enzyme and substrate. For many years, the functions of all CBMs were thought to relate to binding only. However, recent studies show that some of these domains may have extended functions that provide a vital contribution to efficient carbohydrate hydrolysis (discussed below) (Boraston *et al.* 2004, Forsberg *et al.* 2011 & Vaaje-Kolstad *et al.* 2005a).

CBMs are currently (by February 2013) classified into 66 families according to their sequence similarities and are listed in the CAZy (Carbohydrate-Active Enzymes) database. CBMs with the ability to bind cellulose are found in families 1,2, 3, 4, 6, 8, 10, 16, 17, 30, 33, 37, 44, 46, 49, 59, 63 and 64, whereas CBMs that bind chitin are found in families 5, 12, 14, 18, 19, 33, 37, 50, 54 and 55. Recent studies have showed that family 33 CBMs in fact are enzymes with a lytic polysaccharide monooxygenase activity (Vaaje-Kolstad *et al.* 2010, Forsberg *et al.* 2011; Aachmann et al., 2012). This enzymatic activity was shown to boost the efficiency of GHs (Vaaje-Kolstad et al., 2005a, 2010), indicating that these CBM33s are crucial for efficient carbohydrate hydrolysis. In fact, since the year 2010, family 33 CBMs has been proposed to be regrouped and named **L**ytic **P**olysaccharide **M**ono-**O**xygenases (LPMOs). (Discussed in section 1.7)) (Horn *et al*., 2012-BfB review). Hereafter, CBM33s are referred as LPMOs.

The structure and function of CBMs have been studied intensely for several decades. A general trend is the presence of aromatic amino acids on the binding surface/ site that interacts with the targeted carbohydrate. CBMs targeting single chained polysaccharides (like e.g.

xylan) usually have binding clefts, whereas CBMs targeting insoluble, crystalline substrates like cellulose and chitin usually have flat binding surfaces (Boraston *et al*., 2004). Structural studies of CBMs have revealed high structural diversity of cellulose and chitin targeting CBMs (Fig 1.9). Since chitin and cellulose are relatively similar crystalline polysaccharides it is not uncommon that CBMs targeting chitin also bind cellulose and vice versa.

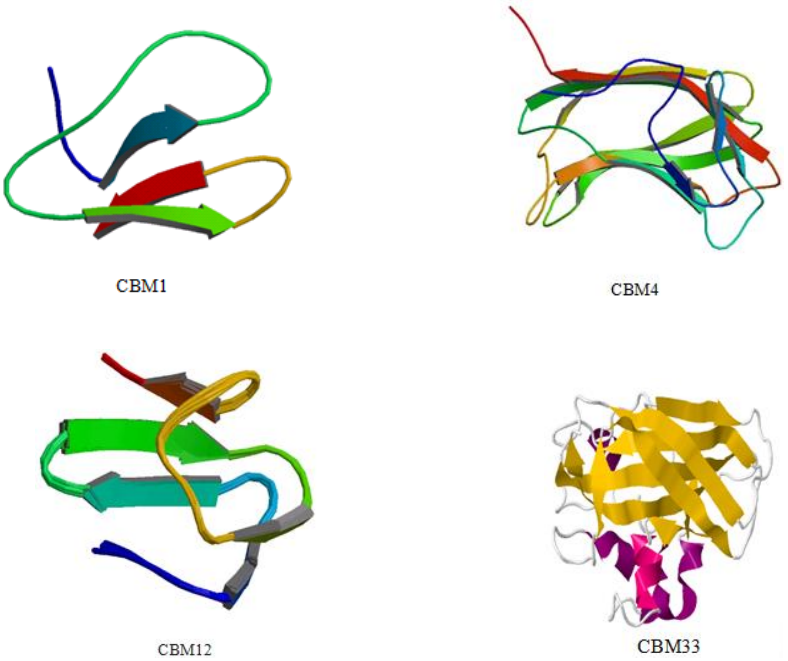


Fig 1.9. Examples of structure of carbohydrate binding modules. Family 1: CBM1 from Cellulase 7A (*Trichoderma reesei*) (pdb 1CBH) (Kraulis *et al.* 1989). Family 4: CBM4 from Cellulase 9B (*Cellulomonas fimi*) (pdb 1ULO) (Johnson et al. 1996). Family 12: CBM12 from Chitinase ChiA1 (*Bacillus circulans*). (pdb 1ED7) (Ikegami et al. 2000). Family 33: CBM33 (known as CBP21) a one-domain protein from *Serratia marcescens* (pdb 2BEM) (Vaaje-Kolstad *et. al* 2005b).

1.7 Lytic polysaccharide monooxygenases (LPMOs)

For decades the concept of cellulose/chitin degradation has been based on the synergistic action of endo-enzymes, processive exo-enzymes and β-glucosidases. However, recent studies demonstrate the involvement of a novel enzyme type, namely bacterial and fungal enzymes listed in family 33 carbohydrate binding modules (CBM33s) and family 61 glycoside hydrolases (GH61s). It is now clear that these enzymes cleave cellulose and chitin chains in their crystalline context and by doing so, they increase the efficiency of classical GHs. Interestingly, CBM33s and GH61s use an oxidative mechanism for cleavage. Leaving one of the newly generated chain ends oxidized. In the case of CBM33s this usually is the C1

sugar, meaning that the enzyme generates aldonic acids. Therefore, it has been proposed to reclassify and rename this group of enzymes as lytic polysaccharide monooxiginases or LPMOs as this name represents their lytic and monooxidizing function (Horn *et al*., 2012).

The first LPMO that was characterized and structurally solved was isolated from the Gram negative soil bacterium *Serratia marcescens*. This organism produces at least five chitin degrading enzymes and is known to be one of the most efficient chitin degrading bacteria. One of the most abundant enzymes produced by *S. marcescens* is a single-domain CBM33-type LPMO known as CBP21. Like all other well-studied LPMOs, CBP21 is a cooper-dependent monooxygenase (Vaaje-Kolstad *et al*., 2010; Aachmann *et al*., 2012), the activity of which depends on an electron donor (or "reducing agent") such as ascorbic acid.

LPMOs are wide spread in bacteria, viruses and fungi. Fungal LPMOs (GH61s) share low sequence similarity with bacterial LPMOs (usually less than 10%), but have a common fold and catalytic motif. Previous studies have shown that the action of LPMOs the CBM33 and GH61 families on cellulose is similar (Westereng *et al*., 2011, Horn *et al*., 2012). However, while all known CBM33 oxidize C1, some GH61s from *Neurospora crassa* and *Thermoascus aurantiacus* (*Ta*GH61) generate C4 and/or C6 oxidized products, in addition to C1 oxidized products (Horn *et al*. 2012, Phillips *et al.* 2011, Quinlan *et al.* 2011 & Westereng *et al.* 2011). C1 oxidizing GH61s generate non-modified non-reducing end thus may benefit cellobiohydrolases that attack from non-reducing ends. On the other hand C4 oxidizing GH61s may benefit cellobiohydrolases attacking from the reducing ends. Such anticipated very specific synergies have, however, not yet been demonstrated by experiment. The action of C1 and C4 oxidizing GH61s, as well as our current understanding of complete cellulolytic enzyme machineries, are illustrated in Figure 1.10.

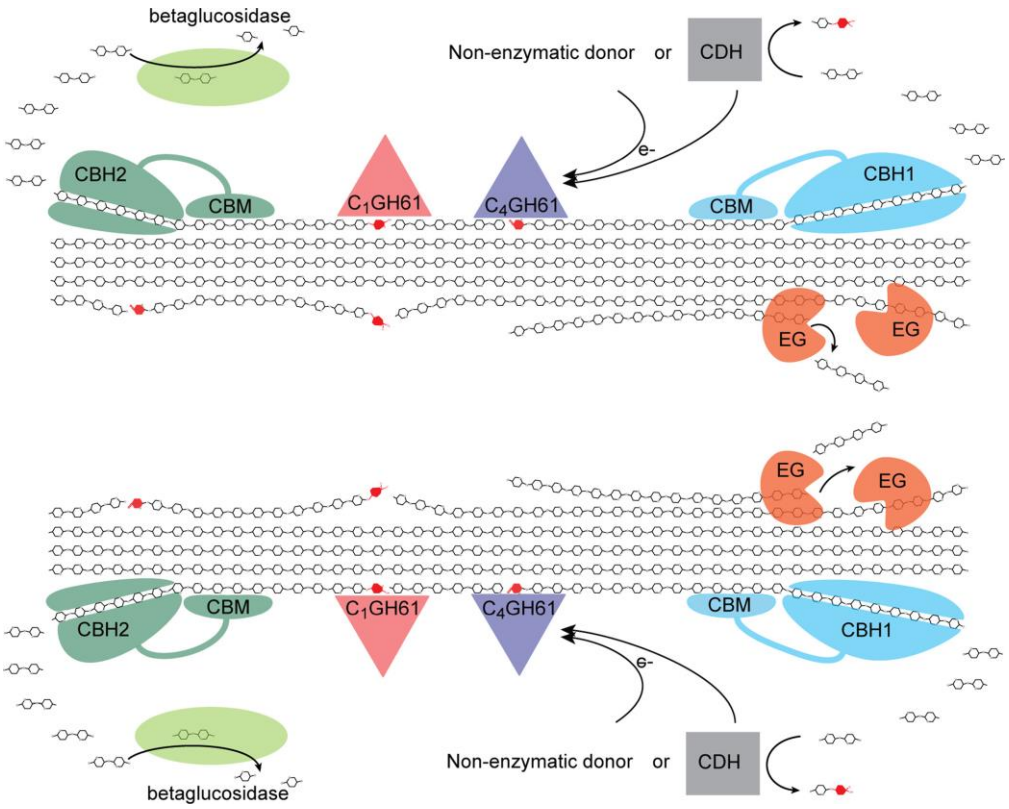


Fig 1.10. Schematic illustration showing fungal enzymatic depolymerization cellulose. The figure shows enzymatic cleavage of crystalline cellulose facilitated by C1 and C4 oxidizing GH61s (named C_1GH61 and C_4G61 respectively). C_1GH61s are indicated with red triangles while C_4GH61s are indicated with blue triangles. Oxidized ends are colored red. These enzymes generate non-oxidized reducing or non-reducing ends where cellobiohydrolases (CBH1 and CBH2) may attack. The catalytic sites of CBHs are attached to cellulose chains with the help of cellulose binding domains (CBMs). It has to be noted that many cellulolytic enzyme systems contain several CBHs and endoglucanases (EGs) that may act on various parts of the substrate. Products of CBHs are cellobiose and are further degraded to glucose by beta-glucosidases (shown in light green). Cellobiose-dehydrogenase (CDH) may provide GH61s with electrons. However, previous studies show that organisms that do not contain genes encoding for CDHs may be provided with electrons from other non-enzymatic reductants such as ascorbic acid and reduced glutathione. (Source: Horn *et al*. 2012)

1.7.1 Structure and function

Previous structural studies of GHs show that these enzymes share common substrate binding site architectures, which can be roughly described as groove, cleft or tunnel. The substrate-interacting surfaces are often lined with aromatic residues, which are important in substrate binding and, in the case of processive enzymes, displacement (Zakariassen *et al*., 2009). Interestingly, LPMOs do not show a groove, tunnel or cleft, but have flat substrate-binding surfaces (Vaaje-Kolstad *et al*., 2005b, Aachmann *et al*., 2012, Karehabadi *et al*., 2008) explaining their preference for binding and cleaving polysaccharides organized in flat, crystalline arrangements like chitin and cellulose. Furthermore, the majority of solvent exposed residues of the LPMO binding surfaces are polar residues (discussed below). A

schematic illustration of the interaction between an LPMO (CBP21 in this case) and a crystalline surface is shown in Fig. 1.11.

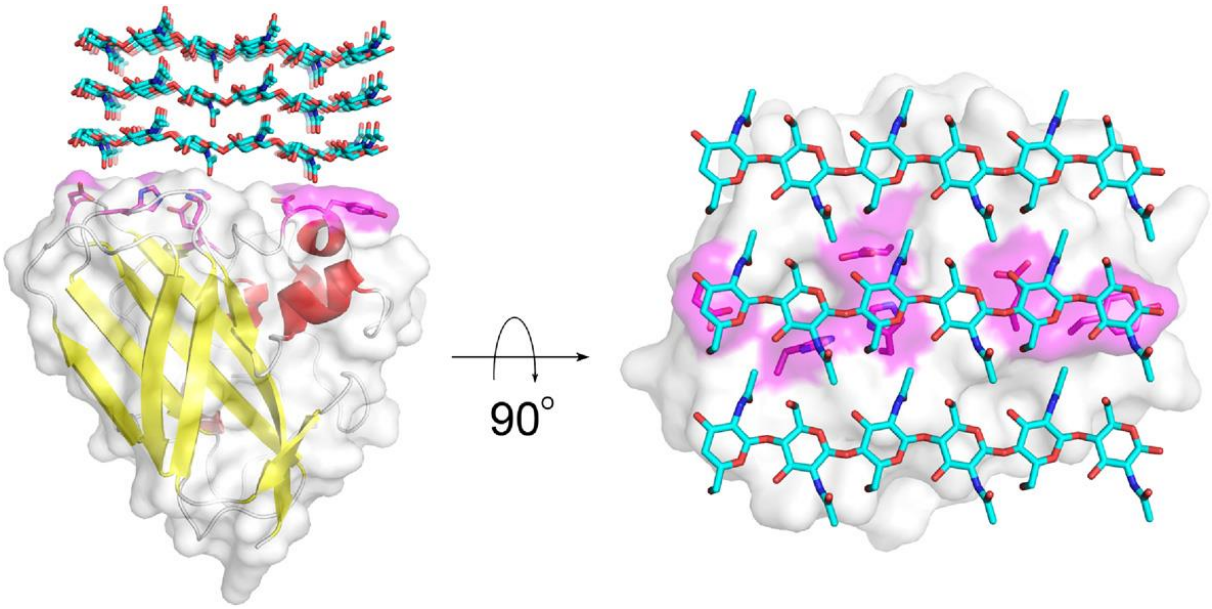


Fig. 1.11 Schematic illustration of interaction between CBP21 and chitin. The figure shows interaction between the flat surfaces of CBP21 and β-chitin. The left figure shows a side view of the interaction. The right figure represents a 90° rotated view showing the interaction from a top view. The side chain of residues known to interact with chitin is shown as stick and their surfaces are colored magenta. The top view shows all known interacting residues as some of these are hidden in the side view. Note that this orientation is hypothetical and the actual orientation of interaction is unknown (Source: Horn *et al*., 2012).

As the architecture of LPMO suggests, the catalytic site of LPMOs is exposed to solvent. In both CBM33-type and GH61-type LPMOs the active site comprises two conserved histidine residues that bind a metal ion (Fig. 1.12, panel d; the metal is copper; see below). Mutational probing of conserved residues on the putative binding surface of LPMOs has shown that they are important for both binding and the catalytic function of the enzyme (Vaaje-Kolstad *et al*., 2005a). It is not yet known which residues determine substrate specificity, but comparison of sequences and structure does provide some hints. The latter is illustrated by Fig. 1.12, which shows a comparison of CBP21 from *Serratia marcescens* and chitin-active *Ef*CBM33A from *Enterococcus faecalis* V583 and which highlights residues in the binding surface that are different in the cellulose active CBM33 CelS2 from *Streptomyces coelicolor* (Vaaje-Kolstad *et al*., 2012).

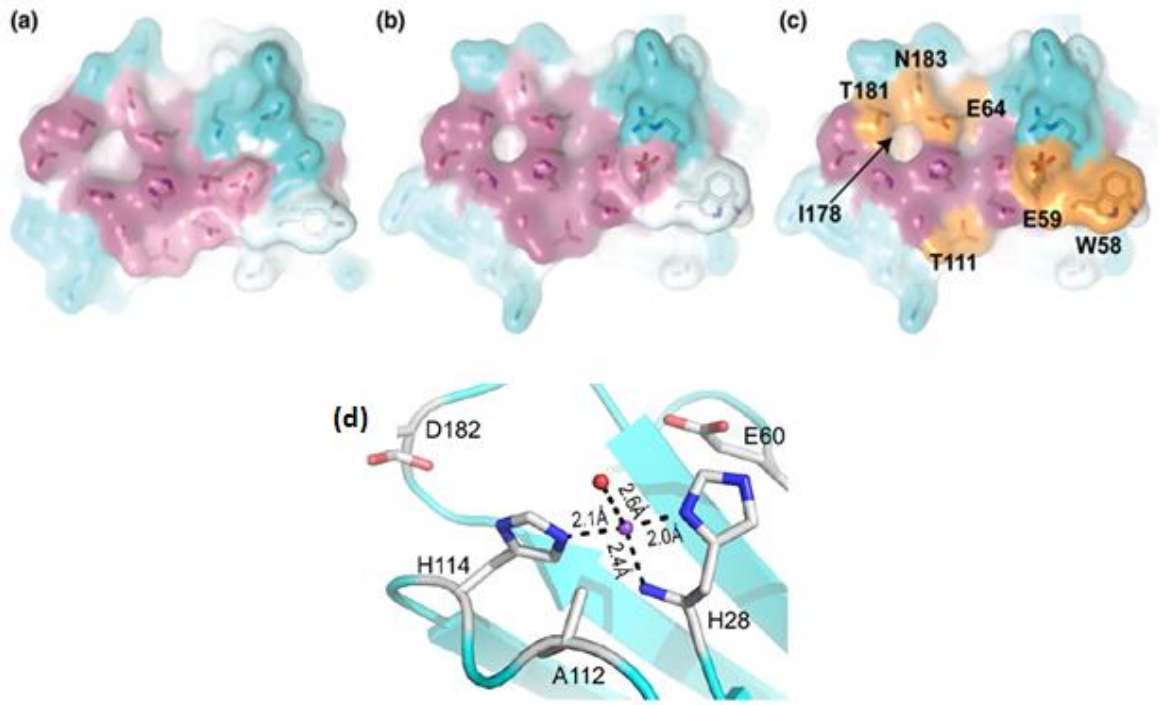


Fig 1.12. Catalytic sites of chitin-active CBM33 type LPMOs from *Serratia marcescens* and *Enterococcus faecalis* V583. The amino acid side chains of surface exposed residues of CBP21 from *Serratia marcescens* (**a**) and CBM33A from *Enterococcus faecalis* V583 (**b** and **c**) are shown as sticks. The surface exposed residues of both CBM33s contain two histidines. These histidines bind metals and this is shown for CBM33A (**d**). The color coding in panels a, b and c represents degree of conservation of residues. Blue; represents non-conserved residues, white; residues conserved to certain degree and magenta; highly conserved residues. In panel c, residues that are not conserved in cellulose active CBM33 CelS2 are coloured orange (Picture Source: Vaaje-Kolstad *et al.,* 2012)

After some initial confusion (Harris *et al*., 2010, Vaaje-Kolstad *et al*., 2010), it is now clear that LPMOs are copper dependent-enzymes (Quinlan *et al*., 2011, Phillips *et al*., 2011,Vaaje-Kolstad *et al*., 2012 Aachmann *et al*.,2012) where the metal is coordinated by the two conserved histidines (Fig 1.12d). The reaction mechanism is currently a subject of discussion and research, but from recent work on CBP21 it seems clear that copper is reduced on the enzyme after which it can transfer an electron to molecular oxygen, as shown in Fig 1.13. This activated oxygen (a superoxo intermediate) can then initiate the reaction by abstracting a proton from the substrate. Phillips *et al*., 2012 have suggested a complete putative mechanism for LPMOs in the GH61 family, but most steps in this scheme are putative (Fig 1.14).

Notably, several studies have shown that the necessary electron may be donated by cellobiose dehydrogenase, as indicated in both Fig. 1.14 and 1.10.

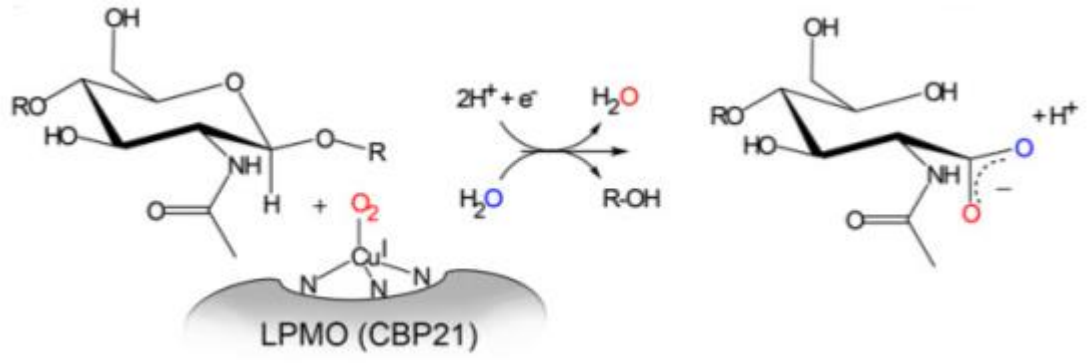


Fig 1.13. **Schematic overview of the proposed reaction mechanism catalyzed by LPMOs.** The figure summarizes experimental evidence obtained for CBP21. CBP21 coordinates reduced copper ion (Cu(I)) through the conserved histidines, then activates molecular oxygen through electron transfer (and copper oxidation), which then through an unknown mechanism results in glycosidic bond cleavage. The C1 carbon is oxidized by single oxygen (red oxygen) and then hydrolyzed (blue oxygen), generating the end product, an aldonic acid. The copper binding involves three nitrogen atoms, one from the N-terminal amino group and two imidazole nitrogens from two conserved histidines.
A more detailed full and speculative description of the reaction mechanism is shown in figure 1.14 (Source for this figure: Aachmann *et al.*, 2012).

Note that Phillip *et al.* propose a mechanism that may lead to formation of either C1 or C4 oxidized sugars (Fig. 1.14). The idea is that the superoxo intermediate extracts a hydrogen atom either from C1 or C4 leading to formation of copper hyperoxo intermediate and a substrate radical. Initiation of O-O bond cleavage is performed by a second electron from e.g. CDH, leading to release of water and formation of a copper oxo radical which couples with the substrate radical, thereby hydroxylating the sugar chain at C1 or C4. The glycosidic bond is destabilized due to addition of the oxygen atom, leading to elimination of the adjacent glucan and formation of a lactone or ketoaldose. Phillips et al further suggested that the elimination of glucan may be facilitated by a general acid, possibly a third highly conserved histidine located near the metal binding histidines that is found in most GH61s (Phillips *et al*., 2011).

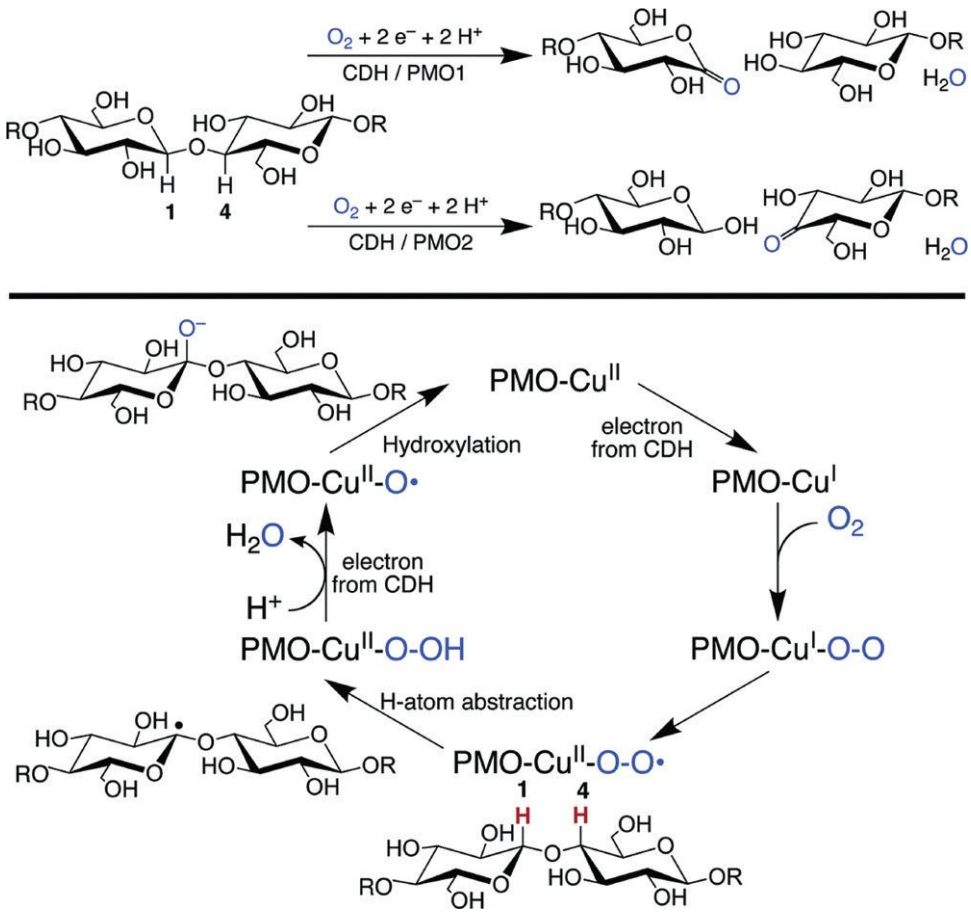


Fig 1.14. Proposed mechanism for LPMOs. C1 and C4 oxidizing GH61s are named PMO1 and PMO2 respectively. Top panel: PMO1 abstracts a hydrogen from C1 generating sugar lactones. PMO2 abstracts a hydrogen atom from C4 generating ketoaldoses. The bottom panel shows a postulated reaction mechanism for LPMOs: Cu(II) is reduced to Cu(I) by the heme domain of CDH and oxygen binds. As a result, a copper superoxo intermediate is formed through internal electron transfer, which abstracts a H• from C1 or C4. A second electron from CDH leads to homolytic cleavage (single bond cleavage) of the Cu-bound hydroperoxide. The Copper oxo species (Cu-O•) then couples with the substrate radical, hydroxylating the substrate. Addition of oxygen destabilizes the glycosidic bond leading to elimination of the adjacent glucan. (Source: Phillips *et al.*, 2011)

1.8 Carbohydrate degradation by *Streptomyces coelicolor*

Streptomyces coelicolor A3 (2) is filamentous, gram positive (Gram+) and ubiquitous soil-dwelling bacterium. The genomic DNA of this species is known to have high GC (guanine/cytosine) content (72.12%) and contains a large number of chromosomal genes coding for most natural antibiotics used today. This bacterium is essential for its environment because of its ability to process insoluble biomass such as lignocellulose and chitin (Bentley *et al.*, 2002).

S. coelicolor A3 (2*)* degrades both cellulose and chitin in addition to other polysaccharides. According to the CAZy database, the genome of *Streptomyces coelicolor* A3(2) contains genes encoding for 148 GHs, 61 glycosyl transferases, 9 polysaccharide lyases, 26

carbohydrate esterases and 83 carbohydrate binding modules. These include at least 22 cellulases, 7 LPMOs (CBM33-type), and 13 chitinases. The carbohydrate binding enzymes are listed in Appendix A.

As many other carbohydrate degrading organisms, *Streptomyces coelicolor* A3(2) expresses both single domain GHs and multi modular enzymes. As an example, the domain architectures for all the seven LPMOs putatively encoded in *S. coelicolor* are shown in Fig. 1.15.

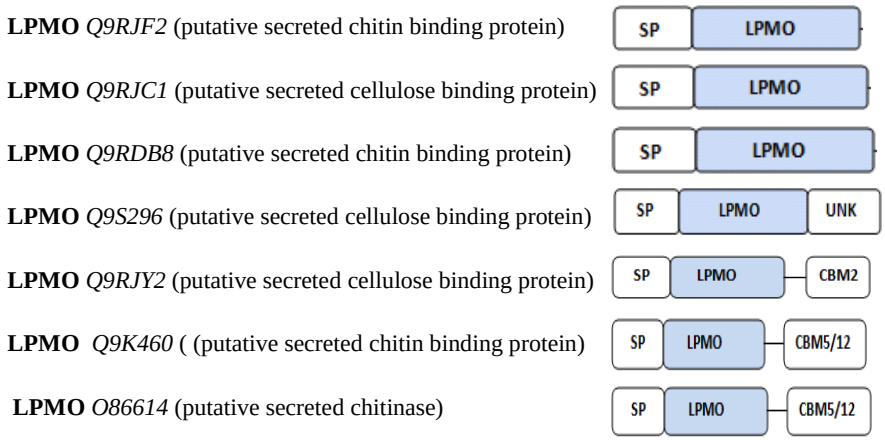


Fig 1.15. Modular structure of putative LPMOs from *S. coelicolor A3(2).* Abbreviations: SP, signal peptides; LPMO, Lytic polysaccharide monooxygenase; CBM, carbohydrate binding module; UNK, region with unknown function. All LPMO domains are colored light blue. SP and other domains are labeled. The putative protein functions are predicted base on sequence comparison with similar and are generally not based on experimental data and likeyl to be wrong in some cases. CBM2 domains tend to bind to cellulose, whereas CBM5/12 domains are normally associated with chitin-binding. The accession numbers for the seven putative LPMOs were obtained from CAZy database and all modular figures were constructed based on the annotated module structures on pfam database.

Of the seven LPMOs putatively produced by *Streptomyces coelicolor* A3(2), only one has been cloned and characterized, namely CelS2 (UniProt code Q9RJY2). CelS2 is a two-domain LPMO containing an N-terminal LPMO and a CBM2 domain (see also Fig 1.15). CelS2 was in 2011 shown to be a cellulose active LPMO that boosts the degradation of cellulose by cellulases. The genome information of *S. coelicolor* shows that CelS2 is located 248 bp apart from a gene encoding for a putative cellulase (CelB, a cellulose in the GH12 family; accession number: Q9RJY3) suggesting the co-expression of these enzymes. Co-regulation of these two enzymes was experimentally studied using the closely related strain *Streptomyces halstedii* JM8 showing a gene encoding for Cel2 (a cellulase in the GH12 family) is clustered with another gene encoding for cellulose binding protein referred as p40

(which is a highly similar domain with CelS2). The open reading frame (ORF) for p40 is located 216 bp downstream from the ORF encoding Cel2. This type of transcriptional co-regulation of CelS2 with genes coding for enzymes involved in cellulose degradation indicates the involvement of CelS2 in cellulose degradation (Garda *et al.*, 1997, Forsberg *et al.*, 2011 & http://www.genedb.org/gene/SCO1187#SCO1187).

1.9 Carbohydrate degradation by *Jonesia denitrificans*

Jonesia denitrificans is a coryneform bacterium (aerobically growing and asporogenous gram positive rod) originally isolated from cooked ox blood. The organism has an irregular rod-like shape with a dimension of 0.3-0.5 µm in diameter and 2-3µm in length (Fig 1.16). The genome of this organism is 2,749,646 base pairs in size and contains 2,558 protein coding genes. *J. denitrificans* utilizes sugar derivatives such as cellobiose (glucose dimers), D-sorbitol and D-galactose. In addition, the cell wall of this organism contains amine containing sugars including galactoseamine and glucoseamine (Rüdiger *et al*., 2009, Funke *et al.*, 1997).

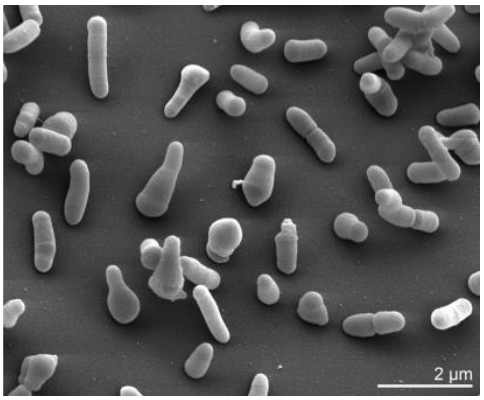


Fig 1.16. Electron micrograph of *Jonesia denitrificans*. The picture is taken from Pukall *et al*. 2009; photo credit: Dr. Manfred Rohde at Helmholtz Centre for Infection Research, Braunschweig.

Reference materials for *J. denitrificans* are scarce. However, due to the organism's phylogenetic position, the complete genome of this organism was sequenced and published by Pukall *et al.* in 2009. The bioinformatics resource known as Kyoto encyclopedia of genes and genomes (*http//genome.jp/KEGG*) has recently included the *J. denitrificans* "metabolic genome" in the database, which allows prediction of the putative strategies that this bacterium has for carbohydrate degradation.

Jonesia denitrificans is predicted to contain several genes encoding for carbohydrate active enzymes. Per November 2012, the CAZy database listed 64 GHs, 21 glycosyl transferases, 2 polysaccharide lyases, 7 carbohydrate esterases and 32 carbohydrate binding modules for this organism. These include at least 6 cellulases, 2 LPMOs, and 1 chitinase (http://www.cazy.org/b1048.html). The carbohydrate binding enzymes are listed in Appendix B.

According to the KEGG pathway database, *J. denitrificans* is capable of degrading chitin to *N*-acetyl glucosamine which then enters into other metabolic pathways (Fig 1.17). The conversion of chitin to *N*-acetylglucoseamine is performed by two hydrolytic enzymes. The first step of chitin degradation is performed by a GH18 chitinase (EC 3.2.1.14), converting chitin to chitobiose units. Chitobiose units are further converted to *N*-acetyl-glucosamine by a GH20 beta-*N*-acetylhexoseaminidase (Fig 1.17). The GlcNAc is then taken up by the bacterium and enters the amino sugar metabolic pathway.

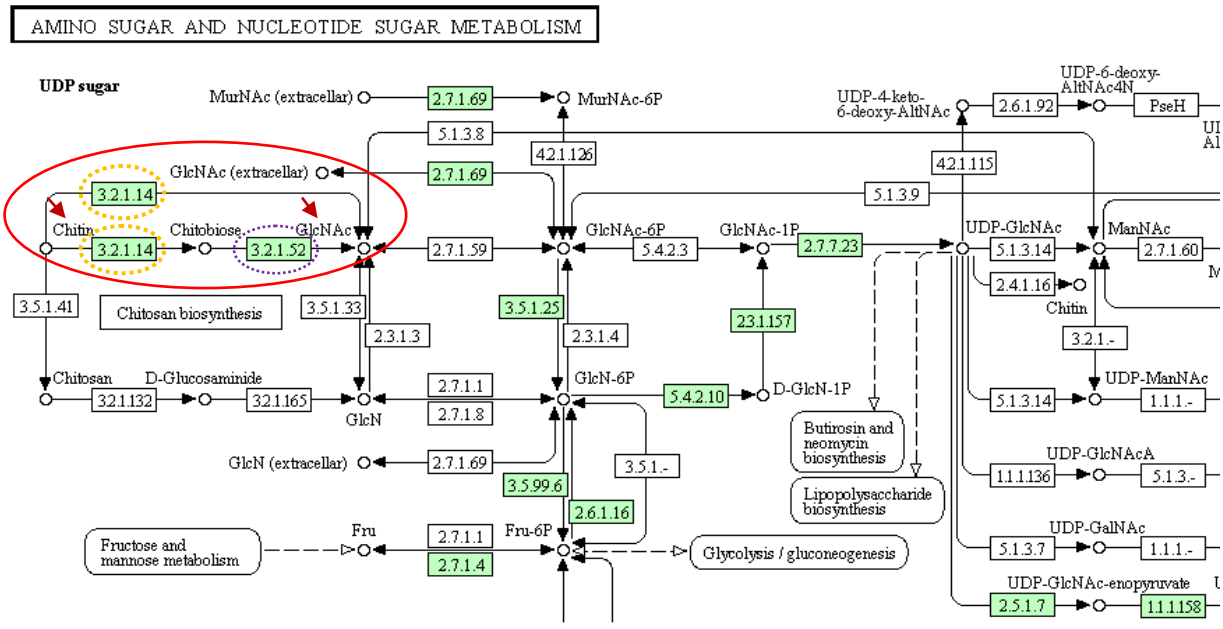


Fig 1.17. Amino-sugar metabolism of *J. denitrificans*. The predicted pathways for degradation of chitin are shown in the figure. The enzymes involved in these processes are indicated with their EC- numbers. Chitin degradation is indicated by the red circle. Chitin is degraded to chitobiose by EC 3.2.1.14 (chitinase) (circled orange with broken line; one gene). Chitobiose is further degraded to *N*-acetyl-glucosamine by EC 3.2.1.52 (beta-*N*-acetylhexoseaminidase) (circled purple with broken line) (Source: http://www.genome.jp/kegg-bin/show_pathway?jde00520).

Analysis of *J. denitrificans* GHs in the CAZy database revealed the presence of one GH18 chitinase that contains 651 amino acids. The modular structure of this chitinase is unique compared all other chitinases of this family. Judged by its sequence, the enzyme is predicted to contain three chitin active domains namely; an N-terminal CBM33 type LPMO, CBM5/12 chitin binding domain and c-terminal GH18. N-terminally, the chitinases has a signal peptide for secretion. The three domains are predicted to be linked by two linker peptides that are 22-25 amino acids in length. The first linker peptide is rich in aspartic acids, asparagines, glycines and threonines, while the second linker peptide is rich in prolines, aspartic acids and glycines. The modular arrangement of this chitinase is shown Fig. 4.1.

1.10 Protein Expression

The goal of the present study was to produce and characterize LPMOs (see below for more details) and to do so the genes encoding these proteins need to be cloned. Factors determining the success of heterologous protein expression include optimization of the gene of interest, selection of suitable expression vectors and use of optimal host strains (Waltson *et al.* 2007).

1.10.1 Gene optimization and modification

There exists a variety of codons representing the 20 amino acids used in protein synthesis. There is a substantial difference in the preference of codon usage amongst microorganisms, meaning that regularly used codons in one bacterium will be rarely used in another. Cloned genes may contain codons that are rarely used in the production strain (e.g. *Escherichia coli*). Thus, expression of the recombinant gene may be slow or even absent. In addition, there are variations in the guanine and cytosine content (GC-content) that may interfere with expression levels. For instance, a gene with GC content may be suitable for expression in host organisms having low GC-content but not for expression in organisms with high GC-content. It may therefore be advantageous to optimize the DNA sequence of the to-be-expressed gene. This can be done by synthesizing the target gene where the GC content has been adapted and rare codons have been substituted with codons that are used by the production strain (Ermolaeva *et al.,* 2001, Maertens *et al.,* 2010 & Sastalla *et al.,* 2009).

1.10.2 Expression vectors

Expression vectors are plasmids that can be used to express a foreign gene in a host cell, often at high levels. These types of plasmids are constructed by combination of defined DNA fragments. Expression vectors are different depending on the components they are assembled from. One of these components is the promoter which drives the transcription of inserted genes (Baneyx 1999).

Known promoters for expression in *E.coli* include the lac promoter (*plac*) derived from the lactose utilization system of *E.coli* and the T7 promoter, which is derived from a bacteriophage. The efficiency of protein expression depends on the strength of the promoter. The T7 promoter is a strong promoter that promotes high levels of protein expression comparing to e.g. *plac*. Some promoters require to be induced in order to be activated while others are constitutive. The T7 promoter is an inducible promoter that can be turned on by the non-hydrolyzable lactose analogue isopropyl-β-D-1-thiogalactopyranoside (IPTG). In some cell types T7 promoters may be active even in the absence of inducer (Vaaje-Kolstad *et al.*, 2005). This phenomenon is known as promoter leakage (Baneyx 1999). It is important to note that promoter strength not necessarily is a success criterion. Too fast production or too much protein may be one of the reasons for the formation of so-called inclusion bodies, i.e. non-soluble and denatured protein (see below).

1.10.3 Host strains

The Gram negative bacterium *E. coli* is the most commonly used and preferable strain for protein expression, for various reasons. The growth rate of *E. coli* is very higher compared to the growth rates of other possible host strains (Sørensen *et al.*, 2004).

A problem often encountered in heterologous protein overexpression is low yields of soluble protein due to the physiological response of the host strain. Excess production of protein may be toxic to host organisms and prevent the proteins from being folded properly into functional protein. As a result, excessively expressed proteins may accumulate in the host cell forming non-functional protein aggregates known as inclusion bodies. *In vitro* re-folding of inclusion bodies into soluble proteins is possible and is widely used for recovery of complex and toxic

proteins like membrane proteins. However, re-folding may be difficult and unsuccessful (Wagner *et al.* 2007).

In general, host strains must stably maintain expression plasmids and confer the genetic material relevant to the expression system. Many host strains are selected or genetically modified to control expression levels based on either the type of expression vector they carry or the type of protein they express. For instance, host strains that are designed to express recombinant genes cloned on T7 promoter based expression vectors contain gene for T7 RNA polymerase (DE3) (Casali 2003).

1.10.4 Protein secretion

Microbes secrete part of their proteins. Secretion systems vary, between eukaryotes and prokaryotes and among prokaryotes. Gram negative bacteria contain double membranes where each has its own function and composition. Therefore, extracellular protein translocation in Gram-negative bacteria requires the passage of the protein through these double membranes. Gram negative bacteria secrete some of their proteins in the space between these membranes, known as periplasmic space (Mergulhão *et al.* 2005). Gram-positive bacteria have only one membrane and are generally considered as a better organism for secretion. Clearly, if one wants to have an overexpressed protein secreted, it is required to have knowledge of the secretion processes applicable for the host stain (Cleland *et al.* 1996 & Mergulhão *et al.* 2005).

Recombinant protein secretion within the periplasmic space or into the extracellular environment reduces the risk for cytoplasmic protein"overloading" in the host cell. It will also provide comparatively easier protein recovery due to the simplicity of protein extraction. For instance, recovery of extracellularily secreted proteins does not require cell disruption. In the case of export to the periplasmic space, proteins can be recovered from periplasmic extracts that can be obtained from simple osmotic shocking procedures (Mergulhão *et al.* 2005).

Proteins intended to be translocated to external environments or the periplasmic space contain N-terminal amino acid sequences (usually 18-30 amino acids in length) called signal peptides (or leader peptides). Signal peptides are cleaved off during the translocation process by signal

peptide peptidases that are associated with the innermembrane, where their active sites face the periplasm (Baneyx, 1999).

1.11 Protein purification

Protein purification may require a single or multiple steps depending on the nature of the protein and the complexity of the starting material. There exists a range of different purification techniques, varying from "rough" techniques aimed at simplifying or concentrating the starting mixture (e.g. ultrafiltration, ammonium sulfate precipitation) to high resolution chromatographic techniques for "final" purification steps. Chromatographic techniques exploit protein properties such as net charge, hydrophobicity, or size. Alternatively, many current techniques for purification of recombinant proteins are based on adding amino acid tags that convey affinity to a specific target molecule. Below, some key chromatographic methods used in this study are discussed.

1.11.1 Ion-exchange chromatography (IEC)

Ion-exchange chromatography separates proteins according to their charge. IEC can be carried out at physiological conditions, hence it is a robust technique for separation of biomolecules like proteins that are intended to maintain their native structure and function. IEC is divided into anion- and cation-exchange chromatography depending on the charge of the column material (Yang *et al.* 1996).

Anion-exchange columns are made from positively charged materials. The most commonly used anion column is made of ethylaminoethyl-cellulose (DEAE cellulose).
During protein purification, positively charged amine groups that are attached to a resin will interact with negatively charged residues on the proteins (i.e. aspartic and glutamic acid side chains; pKa~4.4) and retain sufficiently charged proteins on the column (Fig 1.18). The pH of the buffer is adjusted to a value greater than the pKa value of the anionic groups in order to maintain these in the deprotonated state. Retention of proteins on the column will be determined by the number of (exposed) anionic side chains, the charge on the column, the

actual pKa's of the anionic side chains, and the pH of the buffer (Guetta 2006, Berg *et al.* 2002 & Vydac product manual).

In contrary, cation exchange columns are made from negatively charged materials and interact with positively charged amino acids (arginine, pKa~12.48, histidine, pKa~6.5 and lysine, pKa~10). The basic side chains of these residues will be maintained in a protonated state by adjusting the pH of the running buffer to below 6 or 7.

Proteins that are bound both to ion exchange columns are usually eluted with salt gradients, this salt normally being sodium chloride. It is recommended to increase the salt concentration gradually in order to separate proteins according to their interaction strength (Vydac product manual).

1.11.2 Size exclusion chromatography (SEC)

Size exclusion chromatography, also known as gel filtration, is a separation method where proteins are separated according to size. Unlike other chromatographic methods, SEC does not require the interaction of molecules with the stationary phase. The whole separation process happens in one and the same buffer, which is an advantage for sensitive biomolecules that do not tolerate chemical shift. The SEC column is made of an inert porous matrix of particles that contain pores of different sizes. The column has to be equilibrated with buffer with buffer prior to purification, so that the pores as well as the space between the particles is filled with buffer. A SEC column contains void volume (V_o) (volume of buffer outside of the pores) and total column volume (V_t) (volume of buffer outside and inside the pores). During separation, the larger molecules that cannot enter into the particle pores will elute in the void volume at the buffer flow rate. Medium sized molecules that have partial access to the particle pours will subsequently elute according to size. The smallest molecules and salt have full access to all pores and elute together at V_t (Fig 1.18-2) (Amersham Biosciences).

1.11.3 Affinity-chromatography

In affinity chromatography proteins are separated according to specific and reversible interactions with specific immobilized ligands attached to the stationary phase. (Fig 1.18-3). This is a robust biomolecule purification method due to its high specificity and the use of natural conditions that maintain of protein functionality. Protein recovery from affinity chromatography can be achieved either by using competitive ligands or by changing the pH, ionic strength or polarity (Amersham Biosciences). For instance, chitinases and poly-Histidine (Poly-His) tagged recombinant proteins can be purified by such affinity chromatography, using chitin-beads and colums with immobilized nickel, respectively (Anderson *et al.* 1999 & Taira *et al.* 2004).

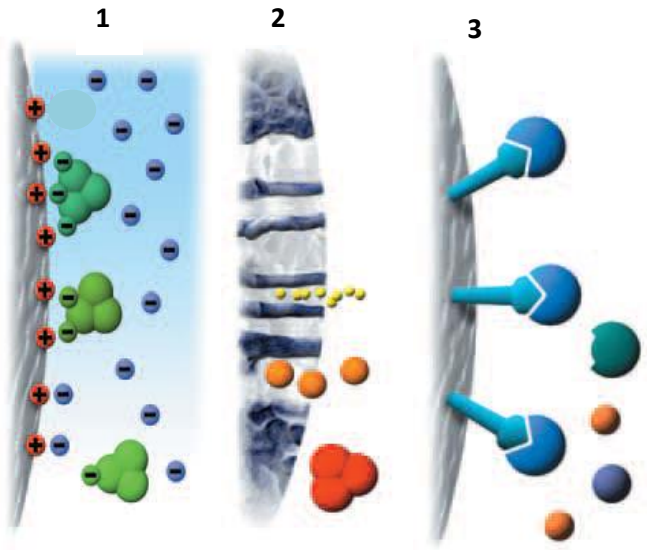


Fig 1.18. **Schematic illustration of chromatographic methods 1)** Anion-exchange chromatography. Proteins (green) with a higher number of residues with negative side chains interact more strongly with the positively charged stationary phase and are thud retained. These electrostatic interactions are shielded by salts, which can thus be used to elute the proteins. Proteins with few negatively charged residues elute prior to proteins with high number of negatively charged residues. **2)** Size exclusion chromatography. Proteins with higher molecular weight do not fit into the column particle pores and elute first. Medium sized proteins that have partial accessibility to the particle pores elute later, whereas small proteins fitting into all pores elute last. **3)** Affinity chromatography. Proteins that can interact with specific ligands will be retained, while other proteins will pass through. This interaction can be reversed by changing environmental factors such as pH. (Figure Source: Amersham pharmacia biotech Handbook "Affinity chromatography: principles and methods" http://www.bio-sun.com.cn/download/1.pdf)

1.12 Protein characterization

Once a protein has been cloned and successful strategies for expressed and purified have been established, the physical characterization of the protein properties can start. In modern protein chemistry, bioinformatics techniques are also an essential toolbox for helping planning, guiding and understanding experiments in the lab. Studies of protein including its biological function, identification of 3-dimentional (3D) structure, evolutionary relations, optimal functional conditions, and kinetics fall under protein characterization. Most of all identification of the 3D structure of a protein is crucial step since it may contain valuable information for understanding its function. Therefore, traditional protein characterization often starts with searching for sequence similarities which in turn identify proteins of resemble structures.

1.12.1 Experimental techniques

1.12.1.1 Enzyme assays

Enzyme assays are experimental methods used to measure enzyme activities and its kinetics. Study of enzyme activity is key step for further characterization of the particular enzyme. The activity of enzymes on its particular substrate may be studied for instance through quantification of end products using quantitative analytical methods such as colorimetric analysis. The activity and kinetic of an enzyme correlates to factors like pH, temperature and substrate concentrations. Optimization of factors is essential for visualization of the enzymes character and all these can be studied with enzyme assays.

1.12.1.2 Site directed mutagenesis

Site directed mutagenesis is a useful technique for studying relations between structure and function of protein. Specific nucleotide substitution within a gene sequence leads to alteration of protein the gene encodes. This type of nucleotide alteration is known as site directed mutagenesis. Amino acids for substitution are selected by homologous protein sequence comparisons. And the introduction of site specific mutations can be achieved through different methods. Yet, polymerase chain reaction (PCR) based mutation is a common

method. PCR based mutation is obtained through amplification of the entire plasmid bearing the gene of interest by using complementary synthetic oligonucleotides containing the desired mutation (Hogrefe *et. al* 2002). Overview of site directed mutagenesis is shown in figure 1.19

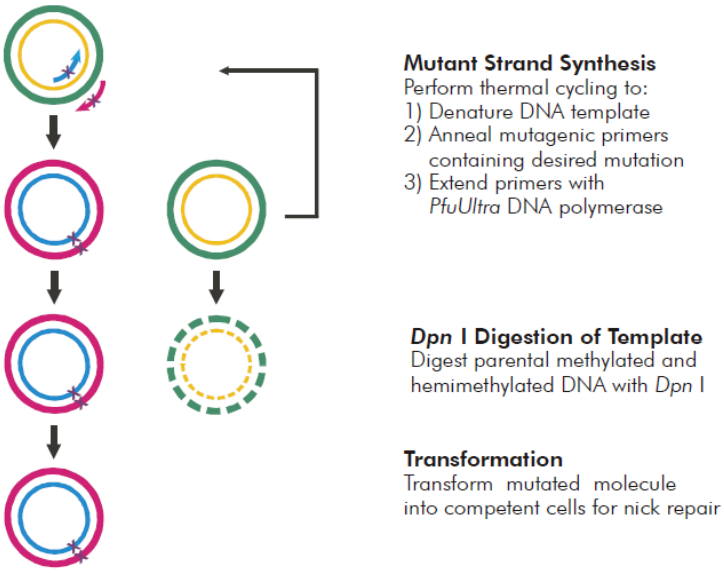


Fig. 1.19 Overview of quick change site directed mutagenesis.

1.12.2 Bioinformatics techniques

1.12.2.1 Homology modeling

Homology modeling is simply means mapping of protein of known sequence but not structure with two or more evolutionary related proteins with known structure. Two proteins of same origin are expected to have similar structure and function. Therefore, proteins with known structure can serve as template for structural prediction of structurally unknown proteins (Wallner *et al.* 2005).

1.12.2.2 Sequence alignment

Multiple-sequence alignment is a useful technique for detecting possibly important residues in a protein. Amino acid residues that are critical for stability, structure or function are anticipated to be conserved throughout evolution and may be detected by critical inspection of sequence alignments (Schueler-Furman et al. 2003). Their essentiality is often predicted according to their degree of conservation, i.e., the more a residue is conserved throughout

evolution the more important it is for the protein. It is important to note that there are many different reasons that may lead to a residue being conserved: activity, stability, foldability, flexibility. It is thus not straightforward to extract functional information from sequence alignments.

1.12.2.3 General database tools : CAZy, Uniprot and Pfam

Biological databases contain massive amounts of useful information, including protein and DNA sequences, protein structures and all sorts of protein classifications. Most of these databases provide freely accessible information. In this study CAZy, Uniprot and Pfam have been of particular importance.

The CAZy database (http://www.cazy.org/) contains families carbohydrate-active enzymes and binding-modules. Members of a family are structurally related and have similar catalytic mechanism, but may vary in terms of substrate-specificity. Protein classes accessible in CAZy include the Glycoside Hydrolases (GH) and Carbohydrate Binding Modules (CBM).

Uniprot archives protein sequence and functional information.

Pfam is a searchable collection of multiple sequence alignments, each summarized as a Hidden Markov Models. By searching Pfam, one can easily and reliably identify known protein domains in a protein sequence.

1.13 Aim of this project

There is accumulating evidence that LPMOs play a vital role in hydrolysis of recalcitrant polysaccharides and, thus, biomass conversion. However, many aspects of these enzymes require extensive additional experimental work, addressing issues such as the catalytic mechanism and (possible) functional diversity among the many family members. LPMOs show great diversity in source and sequence, which indicates variation and substrate specificity. This has created huge interest in performing comparative studies of these enzymes. Accordingly, the aim of this study was to characterize two LPMOs from different Gram-positive bacteria with different modular structures and likely to have different substrate specificities.

The first part of this study addresses the characterization of CelS2 from *Streptomyces coelicolor*, two-domain CBM33-type cellulose-active LPMO containing an N-terminal LPMO domain following by a C-terminal CBM2 domain. The goal of this part of the study was to evaluate the role of five highly conserved surface exposed residues located near the catalytic center, using site-directed mutagenesis. Accordingly, six mutations were planned: R212A, D214A, S215A, E217A, F219A and F219Y. To simplify analysis, the mutations were made in truncated CelS2 variant lacking the CBM2 domain, which expresses better than the full length protein. The functional properties of these mutants were compared with those of the truncated wild type enzyme.

The second part of this study addresses the cloning, expression and characterization of novel protein called Jden1381 from *Jonesia denitrificans*, which comprises an N-terminal CBM33-type LPMO followed by a putatively chitin-binding CBM5/12 and, C-terminally a putative chitinase from family GH18. The aim was to clone, express and purify full length and truncated versions of this enzyme and to study the functional properties of the various constructs. Jden1381 is rather special because the protein combines an LPMO with a glycoside hydrolase, suggesting that the synergy between LPMOs and hydrolases may also be accomplished by having these catalytic modules in one and the same protein.

2 MATERIALS

2.1 Laboratory equipment

Equipments	Supplier
Automated pipettes	Labsystems
Pipette tips	Thermo scientific
Eppendorf tubes	Avant
Cryo tubes	Sarstedt
pH meter	Metrohm
Cuvettes	Eppendorf
Electrophoresis power supply	
PowerPac 300	BioRad
SDS-Page gel electrophoresis equipment	BioRad
SDS-PAGE Gels	BioRad
DNA gel electrophoresis power supply	
PowerPac Basic	BioRad
Mini sub cell GT (chamber)	BioRad
Electroporator	BioRad
MicroPulser Electroporator 100-120V/220-240V	
Protein Gel shaker	Ika®
Stirrer	Ika®
Spectrophotometer	
Eppendorf Biophotometer	Eppendorf
Ultrospec 3300 pro	GE Health care
UV/Visible Spectrophotometer NanoDrop 2000	Thermo Scientific
Centrifuges	
Table centrifuge	Eppendorf
Centrifuge 5430R	Eppendorf
Cooling centrifuge	Beckman Coulter
Thermo cycler (pcr) machine	
Master cycler	Eppendorf
Pcr tubes (micro tubes)	Axygen
ThermoMixer	
Protein purification equipment	Akta purifier
Columns- HisTrap/DEAE/SEC	GE Healthcare
Software – Unicorn 5.20	
Sterilization hood	Teistar AV-100
HPLC equipment	
ICS-300	Dionex
Column- Carbo-Pack PA1	Dionex
UHPLC-Ultimate 3000	Dionex
Column- HELIC	
Software – Chromeleon® 7	Dionex
Agilent 1100 series	Agilent
Column- Tsk Gel™ Column (Amide 80)	Tosoh Bioscience
Software – Chem Station version B.04.01	Agilent
Culture incubator	Infors
Water bath	Julabo

2.2 Chemicals

Chemical	Supplier
2,5-Dihydroxybenzoic acid (DHB)	Bruker Daltonics
Acetic acid 99.8 %	VWR
Acetonitrile (CH_3CN)	Fulltime
Acrylamide	National Diagnostics ProtoGel
Agar bacteriological (Agar No. 1)	Oxoid
Agarose, SeaKem®	Lonza
Albumin, bovine serum (BSA), Fraction V	Sigma-Aldrich
Ammonium Persulfate	Bio-Rad
Ammonium sulfate $(NH_4)_2SO_4$	Merck
Ampicillin	Sigma-Aldrich
Ascorbic acid	Sigma-Aldrich
Bacto™ Peptone	Becton, Dickinson and Company
Bacto™ yeast extract	Becton, Dickinson and Company
Bacto™Tryptone	Becton, Dickinson and Company
Biotin	Sigma-Aldrich
Bis-Tris ($C_8H_{19}NO_5$)	Sigma-Aldrich
Brain heart infusion (BHI)	Oxoid
Calcium chloride ($CaCl_2$)	Sigma-Aldrich
Calcium sulfate ($CaSO_4$)	Sigma-Aldrich
Coomassie Brilliant Blue R250	Merck
D(+)-Glucose monohydrate	VWR
DL-Dithiothreitol (DTT)	Sigma-Aldrich
D-Sorbitol	Sigma-Aldrich
Ethanol 96 % (v/v)	Arcus
Ethidium bomide, ultrapure Bioreagent	J.T. Baker
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA-Na_2)	Sigma-Aldrich
Gallic acid	Sigma-Aldrich
Glycerol 85 % (w/v)	Merck
Hydrochloric acid (HCl)	Merck
Imidazole	Sigma-Aldrich
Isopropyl β-D-1-thiogalactopyranoside	Sigma-Aldrich
LE-Agarose	Gene Mate
L-Gluthation, reduced	Sigma-Aldrich
Magnesium chloride ($MgCl_2$)	Qiagen
Magnesium sulphate ($MgSO_4$)	Merck
Magnesium sulphate heptahydrate ($MgSO_4$) x $7H_2O$	Merck
MES ($C_6H_{13}NO_4S$) hydrate	Sigma-Aldrich
Methanol, HPLC grade	LAB-SCAN
N-Acetyl-D-glucosamine	Sigma-Aldrich

Phenylmethanesulfonylfluoride (PMSF)	Sigma-Aldrich
Phosphoric acid (KOH) 85 % (w/v)	Merck
Potassium chloride (KCl)	Merck
Potassium dihydrogen phosphate (KH_2PO_4)	Merck
Potassium phosphate dibasic (K_2HPO_4)	Sigma-Aldrich
Potassium sulphate (K_2SO_4)	Sigma-Aldrich
Sodium acetate ($C_2H_3NaO_2$)	Sigma-Aldrich
Sodium chloride (NaCl)	Sigma-Aldrich
Sodium hydroxide (NaOH) 50 % (w/v)	J.T. Baker
Sodium sulphate (Na_2SO_4)	Sigma-Aldrich
Sodiumdodecylsulfate (SDS)	Bie & Berntsen
Sulphuric acid (H_2SO_4)	Sigma-Aldrich
Tetramethylethylenediamine (TEMED)	Bio-Rad
Thiamine	Sigma-Aldrich
Tris(hydroxymethyl)aminomethan (Tris-HCl)	Sigma-Aldrich
Yeast extract	Remel
α-cyano-hydroxy-cinnamic acid (CHCA), 97 % (w/v)	Aldrich
B-lactoglobulin from Bovine Milk, 90 % (w/v)	Sigma-Aldrich

2.3 Proteins and enzymes

Protein/enzyme	Suppliers
Bovine Serum Albumine	New England Biolabs (NEB)
Celluclast	Novozymes
Cel5A from *Thermobifida fusca*	Gift from David Wilson (Cornell University)
DNA polymerase	
Vent Polymerase (with 10x reaction buffer)	New England Biolabs (NEB)
Phusion polymerase (with 5x isothermal buffer)	New England Biolabs (NEB)
Pfu polymerase (with 10x reaction buffer)	New England Biolabs (NEB)
Protein standard	
Benchmark Protein Ladder	Bio-Rad
Restriction buffers	
NEBuffer 3 (10x)	NEB
NEBuffer 4 (10x)	NEB
Restriction enzymes	
BamHI (20U/µl)	NEB
BamHI-HF (20U/µl)	NEB
NotI (20U/µl)	NEB
NotI-HF(20U/µl)	NEB
T4 DNA ligase (with 10x reaction mix)	Promega
T5 Exonuclease (with 5x isothermal buffer)	NEB
Taq DNA ligase (with 10x reaction buffer)	NEB

2.4 DNA

DNA	Suppliers
dNTP-mix	Invitrogen
DNA standards	
1Kb DNA ladder	Fermentas
GeneRuler 1Kb Plus DNA ladder	Thermo Scientific

2.5 Carbohydrate substrates

Table 2.1 Carbohydrate substrates used for degradation experiment.

Substrates	Source	Specifications provided by the supplier	Supplier
α-chitin	Shrimp shell	Dried and milled chitin (~400 µm particle size, ash 1.7 %, 4.7 % moisture)	Sea garden
β-chitin	Squid pen	Dried and milled (~400 µm particle size)	France chitin, Marseille, France
Collidal chitin	Shrimp shell	Hydrochloric acid swollen chitin	Prepared in-house (Shimahara *et al.*, 1988)
Avicel	Cellulose	~50 µm particle size	PH101 Sigma-Aldrich
PASC	Shrimp shell	Phosphoric acid swollen cellulose	K. Igarashi

2.6 Kits

Kit	Suppliers
BigDye® Terminator v3.1 Cycle Sequencing Kit Ready reaction mix pGEM®.3Zf(+) double-stranded DNA control Template -21M13 Control Primer (forward) BigDye Terminator v1.1/3.1 Sequencing Buffer (5x)	Perkin Elmer/Applied Biosystems
dNTP set (100mM) PCR Grade	Invitrogen
GoTaq® Green Master Mix GoTaq DNA polymerase $MgCl_2$ (3 mM) dNTP (400µM) Green GoTaq® reaction buffer (pH 8.5) (2x)	Promega
Illustra™ GFX™ PCR DNA and Gel Band Purification Kit GFX columns	GE Healthcare

Collection tubes Color-coded bottles of capture buffer Wash buffer Elution buffer[1] (sterile water) Elution buffer[2] (Tris-HCl)	
Phusion® High Fidelity DNA Polymerase High Fidelity pfu (2,000 U/ml) DMSO (100%) $MgCl_2$ solution (50mM) Phusion GC buffer (5x) Phusion HF reaction buffer (5x)	New England Biolabs (NEB)
QuickChange II Site-Directed Mutagenesis kit Pfu High-Fidelity DNA polymerase (2.5 U/µl) 10x reaction buffer Dpn I restriction enzyme (10U/µl) Oligonucleotide control primer #1[34-mer (100ng/µl)] 5'CCA TGA TTA CGC CAA GCG CGC AAT TAA CCC TCA C 3' Oligonucleotide control primer #2[34-mer (100ng/µl)] 5'GTG AGG GTT AAT TGC GCG CTT GGC GTA ATC ATG G 3' pWhitescript 4.5-kb control plasmid (5 ng/µl) dNTP Mix XL1-Blue supercompetent cells pUC18 control plasmid (0.1 mg/µl in TE buffer)	Agilent Technologies
T4 Ligase T4 DNA ligase T4 ligase buffer (5x) 250 mM Tris-HCl (pH 7.6) 50 mM $MgCl_2$ 5 mM ATP 5 mM DTT 25% (w/v) polyethylene glycol-8000	Invitrogen™
T5 Exonuclease T5 DNA ligase (10,000U/m) NEBuffer 4 (10x)	New England Biolabs (NEB)
Taq DNA ligase Taq DNA ligase (40,000U/ml) λ DNA-BSTEII digest	New England Biolabs (NEB)
Wizard® Plus SV Minipreps DNA purification System 20ml Cell Resuspension Solution (CRA) 20ml Cell Lysis Solution (CLA) 30ml Neutralization Solution (NSB) 20ml Column Wash Solution (CWA) 50 Wizard® SV Minicolumns 50 Collection Tubes (2ml) 550µl Alkaline Protease Solution 13ml Nuclease-Free Water 20 Vacuum Adapters	Promega
$Vent_R$® DNA polymerase $Vent_R$® DNA polymerase (2,000U/ml) $MgSO_4$ (100 mM) ThermoPol reaction buffer (10x)	New England Biolabs (NEB)

2.7 Primers

Table 2.2. Primers by name and sequence

Primer name	Primer Sequence (5' → 3')
*CelS*2-R212A-F	TCATGCAGTGGGTGGCTTCGGACAGCCAGG
*CelS*2-R212A-R	CCTGGCTGTCCGAAGCCACCCACTGCATGA
*CelS*2-D214A-F	TGGGTGCGTTCGGCCAGCCAGGAGAAC
*CelS*2-D214A-R	GTTCTCCTGGCTGGCCGAACGCACCCA
*CelS*2-S215A-F	GTGGGTGCGTTCGGACGCCCAGGAGAACTTCTTC
*CelS*2-S215A-R	GAAGAAGTTCTCCTGGGCGTCCGAACGCACCCAC
*CelS*2-E217A-F	GTTCGGACAGCCAGGCGAACTTCTTCTCCTG
*CelS*2-E217A-R	CAGGAGAAGAAGTTCGCCTGGCTGTCCGAAC
*CelS*2-F219A-F	GCCAGGAGAACGCCTTCTCCTGCTC
*CelS*2-F219A-R	GAGCAGGAGAAGGCGTTCTCCTGGC
*CelS*2-F219Y-F	CAGCCAGGAGAACTACTTCTCCTGCTCGG
*CelS*2-F219Y-R	CCGAGCAGGAGAAGTAGTTCTCCTGGCTG
BamHI_*Jden*1381-F	CCGTAGCAATGGATCCATGAAGAAGAGAAAGTTGAGAGCGTCAGC
*Jden*1381_NotI-R	GGCATTACGAGCGGCCGCTCATTGTAAACCAGTAGCAATCGCTGTAATCAAGTCG
*Jden*1381_NotI_His-R	GTAATGCCGCGGCCGCTCAATGATGATGATGATGATGTTGTAAACCAGTAGCAATCGCTG
BamHl_*Jden*1381-LPMO-F	CCGTAGCAATGGATCCATGAAGAAGAGAAAGTTGAGAGCGTCAGC
*Jden*1381-LPMO_Notl-R	TCGTAATGCCGCGGCCGCTCATGAGACCACAACATCCATACAGTTG
*Jden*1381-LPMO_His_NotI-R	TCGTAATGCCGCGGCCGCTCAATGATGATGATGATGATGTGAGACCACAACATCCATAC
*Jden*1381-CBM5/12-F	GAGGAGATCTGGATCCATGAATAACGGTGGCAATACAGGTGGC
*Jden*1381-CBM5/12-R	GATGCTCGAGGCGGCCGCTCAAGGTGGAGTGCCACCTTCACCTGGG
*Jden*1381-CBM5/12_HIS-R	GATGCTCGAGGCGGCCGCTCGATGATGATGATGATGATGAGGTGGAGTGCCACCTTCACC
*Jden*1381-LPMO-CBM5/12-His _NotI-R	GATGCTCGAGGCGGCCGCTCAAGGTGGAGTGCCACCTTCACCTGGG
*Jden*1381-GH18-F	CCGTAGCAATGGATCCATGCCTGATACGCCTGGTACCGGC
*Jden*Et-F	CCCAGATCTGGGTACCGACGACGACGACAAGCATGGTTGGGTGACAGATCCACCGTCCAG

*Jden*Et-R	GTGGTGGTGGTGGTGCTCGAGTGCGGCCGCTCATTGTA AACCAGTAGCAATCGCTG
pET32b-F	CAGCGATTGCTACTGGTTTACAATGAGCGGCCGCACTC GAGCACCACCACCACCAC
pBBinF	CATCCTGAACTTATCTAGACC
pBBinR	GCAGGTCCTGAAGTTAACTAG
T7 Fwd	TAATACGACTCACTATAGGG
T7 Rev	CCCTATAGTGAGTCGTATTA
pRSET-B SeqF	GATCTCGATCCCGCGAAATT
pRSET-B SeqR	TGTTAGCAGCCGGATCAAGC

Table 2.3. Primers by name and description

Primer name	Primer description
*CelS*2-R212A-F	*CelS*2-R212A, forward mutational primer
*CelS*2-R212A-R	*CelS*2-R212A, reverse mutational primer
*CelS*2-D214A-F	*CelS*2-D214A, forward mutational primer
*CelS*2-D214A-R	*CelS*2-D214A, reverse mutational primer
*CelS*2-S215A-F	*CelS*2-S215A, forward mutational primer
*CelS*2-S215A-R	*CelS*2-S215A, reverse mutational primer
*CelS*2-E217A-F	*CelS*2-E217A, forward mutational primer
*CelS*2-E217A-R	*CelS*2-E217A, reverse mutational primer
*CelS*2-F219A-F	*CelS*2-F219A, forward mutational primer
*CelS*2-F219A-R	*CelS*2-F219A, reverse mutational primer
*CelS*2-F219Y-F	*CelS*2-F219Y, forward mutational primer
*CelS*2-F219Y-R	*CelS*2-F219Y, reverse mutational primer
BamHl_*Jden*1381-F	*Jden*1381, forward cloning primer, full length
*Jden*1381_NotI-R	*Jden*1381, reverse cloning primer, full length
BamHl_*Jden*1381_His-F	*Jden*1381, forward cloning primer, full length with C-terminal His-tag
*Jden*1381_NotI_His-R	*Jden*1381, reverse cloning primer, full length with C-terminal His-tag
BamHl_*Jden*1381-LPMO-F	*Jden*1381-LPMO, forward cloning primer, N-terminal domain
*Jden*1381-LPMO_NotI-R	*Jden*1381-LPMO, reverse cloning primer, N-terminal domain
BamHl_*Jden*1381-LPMO_His-F	*Jden*1381-LPMO, forward cloning primer, N-terminal domain with C-terminal is tag
*Jden*1381-LPMO_NotI_His-R	*Jden*1381-LPMO, reverse cloning primer, N-terminal domain with

	C-terminal His tag
*Jden*1381-*CBM*5/12-F	*Jden*1381-*CBM*5/12, forward cloning primer
*Jden*1381-*CBM*5/12-R	*Jden*1381-*CBM*5/12-F, reverse cloning primer
*Jden*1381-LPMO-CBM5/12-F	*Jden*1381-LPMO-CBM5/12, forward cloning primer, N-terminal multi domain
*Jden*1381-LPMO-CBM5/12-R	*Jden*1381-CBM33-CBM5/12, reverse cloning primer, N-terminal multi domain
*Jden*1381-LPMO-CBM5/12-F	*Jden*1381-LPMO-CBM5/12, forward cloning primer, N-terminal multi domain with C-terminal His-tag
*Jden*1381-LPMO-CBM5/12-R	*Jden*1381-CBM33-CBM5/12, reverse cloning primer, N-terminal multi domain with C-terminal His tag
*Jden*Et-F	*Jden*1381, forward cloning primer, with N-terminal Enterokinase cleavage site
*Jden*Et-R	*Jden*1381, reverse cloning primer, with N-terminal Enterokinase cleavage site
pET32b-F	pET32b, forward vector cloning primer
pET32b-R	pET32b, reverse vector cloning primer
pBBinF	pUCBB, forward sequencing primer
pBBinR	pUCBB, Reverse sequencing primer
T7 Fwd	pET32b, forward sequencing primer
T7 Rev	pET32b, Reverse sequencing primer
pRSET-B SeqF	pRSET-B, forward sequencing primer
pRSET-B SeqR	pRSET-B, Reverse sequencing primer

2.8 Bacterial Strains

Strain	Source
Escherchia coli BL21Star™ (DE3)	Invitrogen
T7 express competent *Escherchia coli* (C2566H)	NEB
Escherchia cloli JM109	Stratagene
Escherchia cloli Rosetta ™	Merck Millipore
Escherchia cloli TOP10	Invitrogen
Escherchia cloli XL1-blue	Agilent

2.9 Plasmids

Plasmid	Source and reference
pET32b	Novagene (Appendix C)
pRSETB/CelS2	(Forsberg *et al.*, 2011) (pRSETb/CelS2-N map retrieved from Addgene is shown in Appendix D)
pUCBB-eGFP	(Vick *et al.*, 2011) (pUCBB-eGFP map retrieved from Addgene is shown in Appendix 5 Fig E)

3 METHODS

3.1 Microbiology methods

3.1.1 Cultivation media

All media were prepared using MilliQ water filtered through a 0.22 µm Millipore filter and autoclaved at 15 psi (1 bar) and 121 ^{o}C for 20 minutes.

3.1.1.1 Agar plates

Agar was added to the LB medium to an end concentration of 15 g/L (w/v) before volume adjustment and autoclaving. For agar plates intended for positive selection of transformants, 100 µg/mL ampicillin was added after cooling to ~50^{o}C, just before pouring the plates. The medium (1 L) was distributed over approximately 20 Petri dishes. The agar plates were left in the laminar flow cabinet to cool and solidify for 20 minutes and stored in plastic bags at +4^{o}C.

3.1.1.2 Antibiotics

Ampicillin

Ampicillin (Sigma-Aldrich) is a beta-lactam antibiotic that inhibits bacterial growth through inhibition of peptidoglycan cross linking (cell wall synthesis) by inhibiting transpeptidases on the inner surface of the bacterial membrane. Ampicillin is active against both Gram-positive and Gram-negative bacteria. Expression plasmids harboring the β-lactamase (*bla*) promoter and gene express the ampicillin resistance protein, β-lactamase, which allows selection of *Escherichia coli* transformants harboring these plasmids. Ampicillin is commonly used as a selective marker in biotechnology.

In this study, ampicillin was used in agar and liquid media (Table 3.1). A stock solution of 50 mg/ml ampicillin was used in all cases.

3.1.1.3 Luria Bertani (LB)

Liquid medium

- 10 g Bacto Trypton
- 10 g NaCl (5 g for "low salt LB")
- 5 g Bacto yeast extract

All solid ingredients were mixed in a 1L flask and dissolved in 800 mL dH_2O by stirring with a magnet stirrer. After the solid components were dissolved, the volume was adjusted to 1 liter with dH_2O and the solution was autoclaved. For media intended for growing transformants, 50 µg/ml of ampicillin was added prior to cultivation.

3.1.1.4 M9 minimal medium

M9 minimal media supplemented with either 0.2% (w/v) glucose or 0.2 % (v/v) glycerol was prepared. The ingredients include different salts and vitamin (thiamine).

For making 10X M9 salt, the salts listed below were dissolved in 900 ml dH_2O. After all salts were dissolved, the volume was adjusted to 1000 ml and autoclaved.

10 X M9 salts

- Na_2HPO_4 60 g/L
- KH_2PO_4 30 g/L
- NH_4Cl 10 g/L
- NaCl 5 g/L

The ingredients listed below were made and autoclaved (except for thiamine) separately.

- 1 M thiamine HCl (sterile filtered, storage 4 ^oC)
- 100 mM $CaCl_2$
- 1M $MgSO_4$
- 20 % (w/v) glucose
- 20% (v/v) glycerol

A 887 ml autoclaved dH_2O was prepared and the following amounts of the ingredients were added.

- 100 ml 10x M9 salts (see above)
- 1 ml 1 M thiamine HCl (sterile filtered, storage 4 ^oC)
- 1 ml 100 mM $CaCl_2$
- 1 ml 1M $MgSO_4$
- 10 ml 20% (w/v) glucose or 20% (v/v) glycerol

3.1.1.5 Terrific broth (TB)

- 10 x TB salts:
  - 23.12 g KH_2PO_4
  - 125.41 g K_2HPO_4

The chemicals were carefully mixed, and dissolved in dH_2O to a final volume of 1 liter and autoclaved.

Liquid medium

- 12 g Bacto Trypton
- 24 g Bacto yeast extract
- 5 ml of 97% (v/v) Glycerol (Final concentration was ~ 0.5 % (v/v)

All ingredients were carefully mixed and dissolved in dH_2O to a final volume of 897 ml before autoclaving. After autoclaving, 100 ml 10x TB salts were added.

3.1.1.6 Brain Heart Infusion Broth (BHI)

37 g BHI was completely dissolved in dH_2O to a final volume of 1 liter and autoclaved.

3.1.1.7 2 x TY

2xTY medium

- 16g Bacto Tryptone
- 10g Yeast extract
- 5g NaCl

All ingredients were carefully mixed and dissolved in dH_2O to a final volume of 1000 ml before autoclaving.

3.1.2 Cultivation of bacterial strains

All media and reagents used for cultivation were sterilized either by autoclaving or sterile-filtration (0.45 µm pore size). All culturing work was performed in sterile conditions. To start new cultures, a single colony from an agar plate or a piece of glycerol stock (see section 3.4) was inoculated in 5 ml medium of choice in culture tubes and incubated overnight at 30°C or 37°C with shaking at 220 rpm as described in Table 3.1. For media recipes, see section 3.1.1. The growth of microorganisms was monitored by measuring optical density at a wavelength of 600 nm (OD_{600}).

Table 3.1. Culturing conditions of strains used.

Strains	Culturing conditions
E. coli C2566H (NEB)	Strains harboring plasmids encoding ampicillin resistance were cultivated in liquid LB-medium supplemented with 50 µg/ml ampicillin, with shaking at either 18 or 30 ^{o}C or on LB-agar plates supplemented with 100 µg/ml ampicillin at 37 ^{o}C. For optimal expression of *Jden*1381, the bacteria were cultured in either M9 or LB media supplemented with 50 µg/ml ampicillin at 30 ^{o}C for 3-5 days, with shaking.
E. coli Bl21 Star™ (DE3) cells (Invitrogen)	Strains harboring plasmids encoding ampicillin resistance were cultivated in liquid LB-medium supplemented with 50 µg/ml

	ampicillin, with shaking at 30 ^{o}C or on LB-agar plates supplemented with 100 µg/ml ampicillin at 37 ^{o}C. For optimal expression of CelS2 mutants, the cells were cultured in either BHI or TB-medium supplemented with 50 µg/ml ampicillin at 30 ^{o}C for 16 h, with shaking. For optimal expression of *Jden*1381 variants, the bacteria were cultured in either M9, 2xTY or LB media supplemented with 50 µg/ml ampicillin at 30 ^{o}C for 3-5 days, with shaking.
E. coli JM109 cells (Invitrogen)	Cells harboring plasmids encoding ampicillin resistance were cultivated in liquid LB-medium supplemented with 50 µg/ml ampicillin, with shaking, at 37 ^{o}C, or on LB-agar plates supplemented with 100 µg/ml ampicillin, at 37 ^{o}C.
E. coli Rosetta ^{TM} cells (Merck Millipore)	Transformants carrying plasmids encoding ampicillin resistance were cultivated in liquid LB-medium supplemented with 50 µg/ml ampicillin with shaking, at 30 ^{o}C, or on LB-agar plates supplemented with 100 µg/ml ampicillin, at 37 ^{o}C.
E. coli TOP10 cells (Invitrogen)	Transformants carrying plasmids encoding ampicillin resistance were cultivated in liquid LB-medium supplemented with 50 µg/ml ampicillin with shaking, at 37 ^{o}C, or on LB-agar plates supplemented with 100 µg/ml ampicillin, at 37 ^{o}C.

3.1.3 Long-term storage of bacterial strains

Glycerol enables bacterial strains to be stored frozen for a long time without harming the cells. Thus, bacterial strains harboring different constructs were preserved as follows:

- 1 ml overnight bacterial culture
- 300 µl glycerol (85 % (v/v), sterile)

Bacterial stains were carefully mixed with glycerol in cryo-tubes and stored at-80^{o}C.

3.2 Molecular biology methods

3.2.1 Plasmid isolation using the NucleoSpin® Plasmid kit

The NucleoSpin® Plasmid kit was used to purify plasmids from *E. coli* TOP10 cells in the molecular biological work related to mutagenesis of the CelS2 gene.

Materials

- Overnight grown bacterial culture containing the plasmid of interest

NucleoSpin® Plasmid/Plasmid (NoLid) kit (Macherey-Nagel)

- NucleoSpin® Plasmid/Plasmid (NoLid) Column
- Collection tubes (2 ml)
- Resuspension buffer A1
- Lysis buffer A2
- Neutralization buffer A3
- Wash buffer A4
- Wash buffer AW
- Elution buffer AE

Procedure

Plasmids were isolated from *E.* coli TOP10 cells following the protocol for plasmid DNA preparation supplied with the NucleoSpin® Plasmid/Plasmid (NoLid) kit. All reaction steps were carried out at room temperature and all centrifugations were done at 11,000 x g using a bench top centrifuge (Eppendorf 5415R).

An overnight culture of *E. coli* TOP10 harboring the plasmid of interest was prepared starting with a glycerol stock. The overnight culture (1 ml) was transferred to 1.5 ml eppendorf tube and cells harvested by centrifugation for 30 seconds. The medium was completely removed before resuspending the pellet in 250 µl resuspension buffer A1. For cell lysis, 250 µl lysis buffer A2 (SDS/alkaline lysis) was mixed into the sample by inverting the tube 6-8 times. The sample was then incubated until the lysate appeared clear. Maximum incubation time was 5 minutes. The lysis method used here is based on the alkaline SDS method developed by Birnboim and Doly (Birnboim *et al*., 1979). The method separates plasmid DNA from other cellular components such as protein and chromosomal DNA by taking advantage of the differences in size and nature of these components. Plasmid DNA is relatively small,

supercoiled while bacterial chromosomal DNA is much larger and less supercoiled. This topological variation allows selective precipitation of chromosomal DNA and proteins from plasmid DNA. When the cells are lysed under alkaline conditions, nucleic acids and proteins will denature.

The lysis reaction was stopped and the lysate neutralized by adding 300 µl neutralization buffer A3 and inverting the tube 6-8 times before the lysate was centrifuged for 5 minutes. During this step, proteins and chromosomal DNA will remain denatured and precipitated while plasmids will renature and stabilize in the solution. Buffer A3 creates appropriate conditions for binding of plasmid DNA to the silica membrane of the NucleoSpin® Plasmid/Plasmid (NoLid) column.

Precipitated proteins, cell debris and genomic DNA were then pelleted by a centrifugation step. The clear supernatant (750 µl) was loaded onto a column placed in a 2 ml collection tube and spun down for 1 minute. The flow-through was discarded. Contaminations like salts, metabolites and soluble cellular components were removed by washing the column with 600 µl wash buffer A4 containing ethanol followed by 1 min centrifugation.

The flow-through was discarded. For a better clean up and for obtaining good sequencing results, the column was washed with an additional 600 µl of wash buffer AW (pre heated at 50^{o}C) followed by centrifugation for 1 min.

The flow-through was discarded and remaining liquid was removed by an additional 2-minute centrifugation. The pure plasmid DNA was then eluted from the column under low ionic strength conditions with 50 µl of the slightly alkaline Buffer AE (5 mM Tris/HCl, pH 8.5) using a 1 minute centrifugation step. The plasmid was stored at -20 ^{o}C.

3.2.2 Plasmid purification from E.coli using the Wizard® Plus SV miniprep DNA purification system

The Wizard® Plus SV miniprep DNA purification system was used to isolate and purify plasmids from chemically competent *E. coli* JM109 cells and electro competent *E.coli* C2566H cells in the molecular biological work related to cloning of the *Jden*1381 gene.

Materials

- Overnight culture of JM109 cells containing the plasmid of interest
- Cell Resuspension Solution (CRA)
- Cell Lysis Solution (CLA)
- Neutralization Solution (NSB)
- Column Wash Solution (CWA)
- Wizard® SV Mini-columns
- Collection Tubes (2ml)
- Alkaline Protease Solution
- Nuclease-Free Water

Procedure

All reaction steps were carried out at room temperature and all centrifugations were done using an Eppendorf 5415R centrifuge operated at maximum speed.

An overnight culture of *E. coli* JM109 harboring the expression vectors (1 ml) was transferred to 1.5 ml eppendorf tubes and cells harvested by centrifugation for 5 minutes. The medium was poured off and excess media was completely removed by blotting the inverted tube on a paper towel. The pellet was thoroughly resuspended in 250 µl Cell Resuspension Solution by vortexing or pipetting. For cell lysis, 250 µl Cell Lysis Solution was mixed into the sample by inverting the tube 4 times and incubated until the cell suspension cleared (1-5 minutes). In order to inactivate endonucleases or other proteins and improve the quality of isolated DNA, 10 µl of Alkaline Protease Solution was added and mixed by inverting the tube 4 times. The mixture was incubated for 5 minutes. It is important not to exceed 5 minutes of incubation as nicking of the plasmid may occur. For neutralization of the reaction, 350µl of Neutralization Solution was added and mixed immediately by inverting the tube 4 times. The bacterial lysate was spun down for 10 minutes at maximum speed (14,000 x g). The cleared lysate (approximately 750 µl) was transferred to the prepared Spin Column inserted into a Collection Tube by decanting and spun down for 1 minute. The Spin Column was removed from the tube and the flow-through was discarded from the Collection Tube. The Spin Column was then reinserted into the Collection Tube and the plasmid was washed by adding 750 µl of Column Wash Solution (which was pre-diluted with 95% ethanol) followed by a 1 minute centrifugation. After discarding the flow-through from the collection tube, the washing step was repeated by adding an additional 250µl of Column Wash Solution. For complete

removal of the washing solution, the column was spun down for 2 minutes. For elution of the plasmid, the spin column containing the plasmid was transferred to a new, sterile 1.5 ml eppendorf tube and 100µl of Nuclease-Free Water was added and incubated for 2-5 minutes in order to let the Nuclease-Free Water distribute through the column. Finally the plasmids were eluted by centrifugation for 1 min. After eluting the plasmid, the assembly was removed from the 1.5ml eppendorf tube and the Spin Column was discarded. The isolated plasmid was stored at -20^{o}C until further use.

3.2.3 Polymerase Chain Reaction-based methods

Polymerase chain reaction (PCR) is an *in vitro* technique used for amplification of specific nucleotide sequences. The reactions require both sequence specific oligonucleotide primers complementary to the sequence of interest or template, dNTPs, and a thermo-stable DNA polymerase. In this study PCR was used for four purposes:

- Gene amplification
- Gene truncation
- Site directed mutagenesis
- Verification of transformants

3.2.3.1 Gene amplification

PCR was used for amplifying the *Jden*1381 gene from *Jonesia denitrificans* DSM 20603 as well as the pET32 b vector in which *Jden*1381 was inserted using Gibson assembly (for more details, see section 3.2.8.2) The *Jden*1381 gene used in this study was codon optimized to express in *E. coli* (Appendix F).

Materials

- Template DNA
- Phusion® High-Fidelity DNA Polymerase kit (NEB):
  - dNTP mix, 10 mM
  - 5x Phusion® GC Buffer
  - Phusion® DNA polymerase (2 U/µl)

- DMSO (100%)

- Primers (See Table 2.2 & 2.3)
- Nuclease-free dH_2O

Procedure

PCR reaction mixes (50µl) were set up on ice in 0.2 ml PCR tubes according to Table 3.2. Reaction mixes were placed in a Master cycler gradient 120V (Eppendorf) and amplification was carried out using the cycling parameters in Table 3.3. Amplification and linearization of the pET32b expression vector was carried out using the cycling parameters in Table 3.4.

Table 3.2. Reaction setup for PCR using Phusion® High-Fidelity DNA Polymerase

Reagents	Volume (Final concentration)
dNTPs 10 mM	1 µl (2 mM)
5x GC buffer	10 µl
DMSO	0.5 µl
DNA Template	2 µl (approximately 80 ng)
Forward primer	0.5 µl (1 pmol)
Reverse primer	0.5 µl (1 pmol)
dH_2O	35.5 µl
Phusion® DNA Polymerase	0.5 µl (1 U)

Table 3.3. Cycling parameters for PCR-amplification of *Jden*1381 using Phusion ® High-Fidelity DNA Polymerase.

Step	Temperature	Time (minutes:seconds)	Number of cycles
Initial denaturation	98°C	1:30	1
Denaturation	98°C	0:10	25
Annealing	60°C	0:30	
Elongation	72°C	3:00	
Final elongation	72°C	5:00	1

Table 3.4. Cycling parameter used for PCR-amplification of the pET32b vector

Step	Temperature	Time (minutes:seconds)	Number of cycles
Initial denaturation	98°C	5:0	1
Denaturation	98°C	0:30	25
Annealing	60°C	1:0	
Elongation	72°C	7:0	
Final elongation	72°C	10:00	1

After amplification, the PCR products were analyzed by agarose gel electrophoresis, described in section 3.2.4. The appropriate DNA band was cut from the gel and the DNA fragment purified as described in section 3.2.5.

3.2.3.2 Gene truncation

Gene truncation is a method used to trim genes at defined regions. A gene can be truncated from its 5'- , 3'- or both ends simultaneously. Deletion of N- or C-terminal regions from proteins is used to determine the functionality of these regions, e.g. in substrate-binding or protein-protein interactions. In this study, truncations were performed on the *Jden*1381 gene (see Fig. 4.1 for domain description) from *Jonesia denitrificans* for four purposes: *i*) to assess relative contribution of the individual domains (the N-terminal LPMO and CBM5/12) on both substrate binding and catalytic activity of the chitinase (C-terminal GH18 domain), *ii*) for individual characterization of the LPMO and GH18 domains, *iii*) to simplify purification procedures (for C-His_6 versions) and *iv*) to solve encountered complications on expression of the full length Jden1381.

The mature full length *Jden*1381 contains 620 residues divided into three domains (see section 1.9). The first residue of mature LPMOs is histidine (His32 in this case) (Westereng *et al.*, 2011). Residue numbering in this thesis starts from these histidine. This means residue number one is His 32 in all Jden1381 variants. Table 3.5 shows overview over all constructs made for expression of Jden1381 variants. The table contains information of number of amino acids starting from His32, features of each construct (i.e., presence or absence of signal peptide and 6xHis tags) and target proteins from each construct.

Table 3.5. Overview of constructs for expression of Jden1381 variants.

Constructs	Amino acids	Tags	SP	Expected soluble protein
pET32b/*Jden*1381fl	1 - 626	His_6, N-terminal	No	LPMO-CBM5/12-GH18
pUCBB/*Jden*1381 fl	1-620	No	Yes	LPMO-CBM5/12-GH18
pUCBB/*Jden*1381fl_C-His_6	1-626	His_6, C-terminal	Yes	LPMO-CBM5/12-GH18
pUCBB/*Jden*1381-LPMO	1-142	No	Yes	LPMO
pUCBB/*Jden*1381-LPMO_C-His_6	1-148	His_6, C-terminal	Yes	LPMO
pUCBB/*Jden*1381-LPMO-CBM5/12	1-245	No	Yes	LPMO-CBM5/12
pUCBB/*Jden*1381-LPMO-CBM5/12_C-His_6	1-251	His_6, C-terminal	Yes	LPMO-CBM5/12
pUCBB/Jden1381-CBM5/12	1-103	No	No	CBM5/12
pUCBB/Jden1381-	1-106	His_6, C-	No	CBM5/12

CBM5/12_C-His_6		terminal		
pUCBB/*Jden*1381-CBM5/12-GH18	1-478	No	No	CBM5/12-GH18
pUCBB/*Jden*1381-CBM5/12-GH18_C-His_6	1-484	His_6, C-terminal	No	CBM5/12-GH18
pUCBB/Jden1381-GH18	1-375	No	No	GH18
pUCBB/Jden1381-GH18_C-His_6	1-381	His_6, C-terminal	No	GH18

Procedure

Truncation of genes was performed by PCR based amplification. Accordingly, region specific primers (see Table 2.2 & 2.3) were designed for all truncations. Gene fragments encoding truncated proteins were amplified by Phusion® High Fidelity DNA polymerase (NEB) using the same procedure described above (section 3.2.3.1; Tables 3.2 and 3.3).

3.2.3.3 Site Directed Mutagenesis

To introduce a specific amino acid substitution in a gene, the QuickChange™II Site- Directed mutagenesis kit (Agilent) was used. The kit is based on a method that uses PfuUltra HF DNA polymerase and a PCR cycler to introduce point mutations or insertions/deletions in DNA. The mutagenic primers are complementary to opposite strands of the vector and they are extended by pfuUltra DNA polymerase by thermal cycling thereby incorporating the mutation. The PCR product is then treated with *DpnI*, which is an endonuclease that specifically digests dam-methylated DNA. Since DNA from *E.coli* is dam-methylated, the enzyme will digest the parental DNA leaving the mutated PCR product intact. See section 1.12.1.2 and Fig 1.19 for more details.

Material

- QuickChangeII Site-Directed Mutagenesis Kit
  Pfu Ultra HF DNA polymerase (2.5 U/µl)
  10x reaction buffer
  Dpn I restriction enzyme (10U/µl)
  dNTP Mix
  XL1-Blue supercompetent cells

- DNA template:
    - N-terminal domain from CelS2 (CelS2-N) (Fig.1.15)
- *E.coli* Top10 competent cells
- Primers (See Table 2.2 & 2.3)

Procedure

The parental DNA harboring the *celS2* gene (nucleotide sequence of *celS2*is shown in Appendix G) was prepared from an overnight culture of transformed *E.coli* Top10 cells (section 3.2.1). The concentration of plasmid DNA was determined using a Biophotometer (at 260nm) (Eppendorf). The PCR reactions were set up by following the instructions provided by the supplier.

The PCR reactions (50 µl) were set up in 0.2ml PCR tubes on ice by mixing the reagents listed in Table 3.6 and amplification was achieved using the cycling parameters listed in Table 3.7. After the PCR reaction, the parental DNA was digested by adding 1 µl *DpnI* and incubating at 37 ^{o} C for 1 hour. After digestion, the DNA was transformed into chemically competent cells; either *E. coli* Top10 or XL1-Blue super-competent cells (see section 3.2.9.1). 150 µl and 25µl of the transformation mixtures were spread on separate pre-warmed LB-agar plates supplemented with 100 µg/ml Ampicillin and the plates were incubated overnight at 37^{o}C. The introduction of mutation was evaluated by DNA sequencing (see section 3.2.10)

Table 3.6. PCR mix setup for QuickChangeII Site-Directed Mutagenesis

Reagents	Amount
10x reaction buffer	5 µl (1x)
DNA template	25 ng
Forward primer*	125 ng
Reverse primer*	125 ng
dNTP mix	1 µl
dH_2O	to 50 µl
Pfu Ultra HF DNA polymerase	1µl (2.5U)

*See section 2.7; Table 2.2 and 2.3

Table 3.7. Cycling parameters used in the QuickChange procedure

Step	Temperature	Time	Number of cycles
Initial denaturation	95°C	30 seconds	1
Denaturation	95°C	30 seconds	16
Annealing	55°C	1 minute	
Elongation	68°C	3.5 minutes	

3.2.3.4 Transformant verification

GoTaq® Green DNA Polymerase (Promega) is a premixed ready-to-use solution containing bacterially derived Taq DNA polymerase, dNTPs, $MgCl_2$ and reaction buffers at optimal concentrations for efficient amplification of DNA templates by PCR. The GoTaq® Green Master Mix contains two dyes (blue and yellow) that allow real-time monitoring of progress during electrophoresis. Reactions assembled with GoTaq® Green Master Mix can be directly loaded onto agarose gels. In 1 % agarose gels, the blue dye migrates at the same rate as 3-5 Kb DNA fragments, and the yellow dye migrates at a rate faster than the primers (<50bp). In this study, the GoTaq® Green Master Mix was used to verify transformed *E. coli* BL21 (DE3), *E. coli* C2566H, *E. coli* JM109 harboring various plasmids with (parts of) the *Jden*1381 gene (see section 3.2.3.2). Template DNA for these reactions was prepared by suspending a small amount of transformant colony in sterile water.

<u>Materials</u>

- GoTaq® Green Master Mix (2x) (Promega)
  - *Taq* DNA polymerase
  - $MgCl_2$ (3 mM)
  - dNTP (400µM)
  - Green GoTaq® reaction buffer (pH 8.5) (2x)
- DNA template
- Primers (See Table 2.2 & 2.3)

Procedure

PCR reactions (approximate volume = 10µl) were set up on ice in 0.2 ml PCR tubes according to Table 3.8. Reaction mixes were placed in a Master cycler gradient 120V (Eppendorf) and amplification was carried out using the settings shown in Table 3.9.

Table 3.8. PCR reactions set up using GoTaq® Green Master Mix.

Reaction Component	Volume (final concentration)
Forward primer	0.3 µl (3 pmol)
Reverse primer	0.3 µl (3 pmol)
Template DNA	2 µl of culture suspension
dH_2O	2.4 µl
GoTaq® Green Master Mix	5 µl
Final volume	10 µl

Table 3.9. Program for the thermal cycler when using GoTaq® Green DNA polymerase Master Mix.

Step	Temperature	Time (minutes:seconds)	Number of cycles
Initial denaturation	95°C	5:00	1
Denaturation	95°C	0:45	25
Annealing	60°C	0:45	
Elongation	72°C	2:10	
Final elongation	72°C	7:00	1

3.2.4 Agarose gel electrophoresis

PCR products were analyzed by running electrophoresis on 1% agarose gels. Linear DNA molecules are negatively charged and, when subject to an electric field in a gel matrix they are separated according to size. For DNA visualization, ethidium bromide was added to the gel. Ethidium bromide intercalates between the stacked nucleotide bases and because of its fluorescent properties, DNA binding ethidium bromide can be visualized with UV light. The size of the DNA molecules is determined by comparison to a DNA ladder comprised of DNA fragments of known size.

Materials

- Agarose
- 1x TAE-buffer:
  - 4.85 g Tris-base
  - 1.14 ml Acetic acid, 99.8% (v/v)
  - 2 ml 0.5 M EDTA, pH 8.0

- Dissolved and mixed in 1 liter dH_2O, yielding a final pH of 8.5.
- Ethidium bromide, 10 mg/ml
- 10x loading buffer (Takara)
- 1 kb DNA ladder (Fermentas) or GeneRuler™ 1Kb plus DNA ladder (Thermoscientific) (Appendix H)

Procedure

1 % (w/v) agarose gels were made by dissolving 0.5 g agarose in 50 ml 1xTAE-buffer by careful heating in a microwave-oven. The solution was cooled to below 60°C before 1 µl Ethidium bromide solution was added. The gel solution was poured into the gel-chamber of a Mini-Sub Cell GT cell with a UV-transparent gel casting tray (Bio-Rad) with a 15-wells comb, and left to solidify. After 30 minute, the comb spacer was removed and the gel transferred to an electrophoresis chamber. 1xTAE- buffer was added to the chamber covering the gel completely. After adding 0.1 volume 10x loading dye to both the 1 kb ladder and each of the samples, the sample were applied to the gel. The gel was run for 40 minutes at 90 volt using a PowerPac Basic™ power supply (Bio-Rad) and DNA-bands were visualized by UV light.

3.2.5 Extraction and purification of DNA fragments from agarose gels

DNA extraction from agarose gel has at least two advantages; i) it is useful to recover DNA generated by restriction enzymes digestion from agarose gel, and ii) for purification of PCR products (i.e., removing of enzymes, primers or undesirable products resulting from the thermocycling). The method which was used in this study is based on the use of silica matrix to bind DNA in the presence of chaotropic agent (a substance with ability to denature proteins and breakdown the polymeric structure of agarose). The matrix bound DNA is washed with an ethanolic buffer to remove the impurities, salts and other contaminations, and the purified DNA are then eluted in a low ionic strength buffer.

Materials

- 1% Agarose gel containing DNA fragments of interest
- Illustra™ GFX™ PCR DNA and Gel Band Purification Kit
  - illustra™ GFX™ MicroSpin™ GFX columns

- 2 ml collection tubes
- Color-coded bottles of capture buffer
- Wash buffer
- Elution buffer (sterile water)

- Scalpel

Procedure

All centrifugations were carried out at room temperature for 1 minute at 11,000 x g using a Centrifuge 5415R (Eppendorf).

DNA fragments or pcr products were excised from the agarose gel by cutting carefully around the band with a clean scalpel. The weight of the gel piece was determined and for every 100 mg of agarose gel, 500 µl Color-coded capture buffer containing chaotropic agent were added, followed by incubation at 65°C while shaking at 550 rpm until the gel slice was completely dissolved. After the gels were dissolved, approximately 750 µl of the samples were transferred to Illustra™ GFX™ MicroSpin™ collection columns which were placed into 2 ml collection tubes and incubated for 1 minute to let the DNA bind to the membrane. After 1 minute of incubation, the bound DNA was separated from the dissolved gel and other impurities by centrifugation for 10 minutes. This let the impurities to pass through the membrane. The flow-through was discarded and column containing the bound DNA was washed with 750 µl wash buffer containing ethanol, using centrifugation to collect the flow through. After removing the wash buffer, the column was centrifuged for a second time for removing residual wash buffer from the column. The dry column was then placed in a sterile 1.5 ml eppendorf tube and the PCR product was eluted with 40 µl Elution buffer (or sterile water). The concentration of extracted DNA was measured at 260 nm using a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific).

3.2.6 Restriction digestion

Restriction endonucleases are sequence-specific enzymes that cleave DNA molecules at specific sites, producing a double-stranded break in the DNA strand. Cleavage may form blunt ends, meaning that both strands are cut at the same position, or overhanging (cohesive) ends, meaning that the two strands are cut at slightly different positions. Different restriction

enzymes may require different reaction conditions for optimal activity. Under non- standard conditions, restriction enzymes may cut at sites similar to their defined recognition sequences. This phenomenon is known as star activity. In order to avoid exhibition of such activities, some enzymes are designed by mutational modification, for instance *BamHI*^{HF} and *NotI*^{HF} (HF =high fidelity), to cleave with higher fidelity. Such HF versions of endonucleases have the same recognition sequence as their corresponding wild types.

3.2.6.1 Double restriction digestion of Jden1381 fragments and pUCBB-eGFP

The expression vector pUCBB-eGFP (Appendix E) and *Jden*1381 (full length and truncated genes; see Table 3.5 for more details on the truncated domains) were double digested using *BamHI* and *NotI* or *BamHI*^{HF} and *NotI*^{HF} (NEB). For the *Jden*1381 variants, the recognition sequences for these enzymes were inserted at the 5' and 3' ends of by PCR amplification (see section 3.2.3.2).

The optimal NEBuffer for both *BamHI* and *NotI* is NEBuffer 3. The optimal buffer for both and is NEBuffer 4.

Recognition sites:

BamHI: 5' ….G▼GATCC….3'

3'….CCTAG▲G….5'

NotI: 5' ….GC▼GGCCGC….3'

3'….CGCCGG▲CG….5'

Materials

- pUCBB-eGFP (50µg) and DNA fragments containing *Jden*1381 variants (50 µg)
- Restriction enzymes

  BamHI, 20U/µl (NEB) or *BamHI*^{HF}, 20U/µl (NEB)

  NotI, 20U/µl (NEB) or *NotI*^{HF}, 20U/µl (NEB)
- NEBuffer 3 (10x)

- NEBuffer 4 (10x)
- Sterile water

Procedure

A 50 µl reaction mixture was set up on ice with 5 µl 10x buffer mixed with 2.5 µl *BamHI* or BamHI^{HF} and 2.5µl *NotI* or *NotI*^{HF}. A total of 1 µg DNA was diluted with 39 µl of sterile water. The reaction was incubated at 37 ^{O}C for 16 h for optimal digestion. The digested fragments were verified by agarose gel electrophoresis (see section 3.2.4) and purified by gel extraction (see section 3.2.5)

3.2.7 Ethanol/EDTA Precipitation of DNA

Ethanol/EDTA precipitation is a method used to purify and concentrate DNA or RNA. Ethanol precipitation allows contaminants to be removed along with excess liquid while the DNA will precipitate and form a solid pellet at the bottom of the tube. EDTA helps to stabilize DNA during precipitation. Ethanol precipitation is a recommended DNA cleaning method used to obtain material that has sufficient quality for further manipulations and that gives high quality and consistent signals in sequencing (see section 3.2.10).

Materials

- 125 mM EDTA
- 96% (v/v) ethanol
- 70 % (v/v) ethanol

Procedure

For precipitation, 20 µl of DNA solution were transferred to sterile 1.5 ml eppendorf tubes. Precipitation of DNA was accomplished by adding 2µl 125 mM EDTA and 62.5 µl 96 % ethanol, followed by incubation at room temperature for 15 minutes. After incubation, the precipitate was pelleted by centrifugation at 16,000 rpm, 4 ^{o}C for 30 minutes. The ethanol and EDTA mixture was then removed by vacuum suction. In order to wash the DNA pellet, 60 µl

70 % ethanol was added followed by centrifugation at 16,000 rpm, 4 ^{o}C for 15 minutes. After centrifugation, the ethanol was immediately removed. To dry the DNA, the eppendorf tubes were left open in an activated laminar flow cabinet for 30 minutes.

3.2.8 Cloning

3.2.8.1 Cloning of Jden1381-gene fragments into pUCBB-eGFP

Double-restriction digests of *Jden*1381 variants and the pUCBB-eGFP expression vector were performed in order to obtain fragments with cohesive ends (see section 3.2.6). For ligating the *Jden1381* variants into pUCBB-eGFP vectors a T4 DNA ligase kit (Invitrogen) was used. T4 DNA ligase is an enzyme that catalyzes formation of phosphodiester bonds between double-stranded DNA ends with 3'-hydroxyl and 5'-phosphate termini. This reaction requires the presence of ATP. The T4 DNA ligase kit contains T4 DNA ligase reaction buffer that supplies ATP and stabilizes the ligation reaction. To achieve optimal ligation, the molar ratio of insert vs. vector in the reactions was approximately 3:1. The cloning of *Jden*1381variants in pUCBB-eGFP vector was designed so that they are incorporated right after *BamHI* recognition site that is located 39 bp downstream from the lac_promoter. Schematic illustration of the cloning steps of Jden1381 into pUCBB-eGFP vector is shown in Appendix I.

Materials

- pUCBB-eGFP restriction digested vector
- *Jden*1381 restriction digested insert
- T4 DNA ligase kit (Invitrogen)
  - T4 DNA ligase (5U/µl)
  - 5 x T4 DNA ligase reaction buffer
- Sterile dH_2O

Procedure

Reactions, 20 µl were set up according to Table 3.10 and incubated for 24 hours at 16°C. The DNA ligase was subsequently inactivated by incubating the reactions at 65°C for 10 minutes.

Table 3.10. Reaction set up for cloning of *Jden*1381 variants into pUCBB-eGFP expression vector.

Reaction Component	Cloning Reaction	Negative Control
Restriction digested *Jden*1381 variants	5.5 µl (248ng of *Jden*1381fl, 67ng of *Jden*1381-LPMO, 95ng of *Jden*1381-LPMO-CBM5/12, 41 ng of *Jden*1381-CBM5/12, 184ng of Jden1381-CBM5/12-GH18 and 145.5 ng of *Jden*1381- GH18)	-
Linearized vector, pUCBB-eGFP	4 µl (100 ng)	4 µl
5 x T4 DNA ligase reaction buffer	2 µl	5 µl
T4 DNA ligase	1.5 µl	2 µl
Nuclease-free dH_2O	7 µl	13,5 µl

3.2.8.2 Cloning of Jden1381-full length into pET32b

The pET-32b expression vector is designed for high level protein expression. The vector allows constructs where the protein of interest is expressed with a 109-residue N-terminal theoredoxin tag (Trx.Tag^{TM}) followed by a His-tag. The Trx.Tag coding sequence is located 89bp downstream from a T7 promoter and is following by a 6-residue His tag starting 21bp down stream from the TRX.Tag. Following the His tag (93 bp downstream) there is an enterokinase cleavage site. The intention for cloning *Jden*1381 in pET-32b was to utilize the beneficial effect of the TRX.Tag on protein expression and solubility, the N-terminal His tag for one step protein purification and the enterokinase cleavage site for cleavage of the enzyme to obtain mature protein. The pET-32b vector map is shown in Appendix C.

Mature full-length Jden1381 has an N-terminal LPMO. Since residue number 1 of the mature protein is a catalytically crucial conserved histidine (Westereng *et al.*, 2011) the cloning of *Jden*1381full length in pET-32b was designed so that cleavage of the enterokinase site would lead to the mature protein starting with this histidine, i.e.,His32 (this means the signal peptide of the protein was not included). The sequence around the cleavage site looks like this (arrow indicates cleavage site):

….. GACGACGACGACAAG▼CATGGTTGGGTG……
Asp Asp Asp Asp Lys His Gly Trp Val

Cloning of *Jden*1381 into pET32b expression vector was done using the Gibson Assembly strategy. Gibson Assembly is a one step enzyme dependent DNA assembly developed by Dr. Daniel Gibson in 2009 (Gibson, D. *et al.* 2009 & 2011). Gibson assembly requires the presence of three enzymes, a 5' exonuclease, a DNA ligase and a DNA polymerase. The principle is shown in Figure 3.1. In this study, the Gibson method was slightly modified based on previous experiences; i.e, in order to increase amount of DNA, the ratio of DNA to Gibson reaction mix (see below) used in this study was 1:1 whereas the original Gibson protocol suggests 1:3 ratio (Dr. Soojin Yeom, personal communication). Both pET32b and the *Jden*1381 fragment had previously been linearized by running a PCR amplification and overlapping sequence had been incorporated on both the 3' and 5' ends of the vector and the gene, using specifically designed primers (see Table 2.2 & 2.3 & section 3.2.3.1). The PCR products were treated with *DpnI* in order to remove the parental DNA (discussed in section 3.2.6) and purified by DNA Gel extraction (see section 3.2.5).

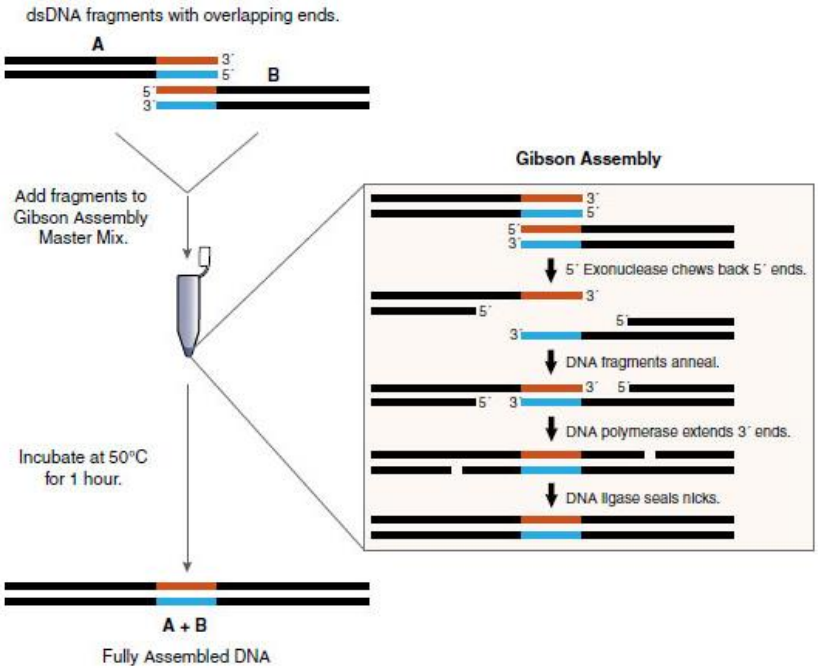


Fig 3.1. Principle of Gibson assembly. The gene and the vector are amplified with primers containing 25-27 nucleotides of overlapping sequence introduced at the 5' ends of the primers. The amplified and linearized dsDNA will be digested from their 5' ends by 5' exonucleases, thereby enhancing annealing of overlapping sequences of the vector and the gene. A DNA polymerase extends the sequences and closes the gaps while DNA ligase seals nicks and complete the cloning process. Note that construction of a vector involves two such

ligations; both fragments have overlapping extensions at both ends. (Source: New England Biolabs Japan; http://www.nebj.jp/products/detail/1238)

Materials

- Isothermal buffer (5x)
  - 1M Tris-HCl pH 7.5 (3 ml)
  - 2M $MgCl_2$ (150 µl)
  - 100 mM dGTP (60 µl)
  - 100 mM dATP (60 µl)
  - 100 mM dTTP (60 µl)
  - 100 mM dCTP (60 µl)
  - 1M DTT (300 µl)
  - PEG-8000 (1.5 g)
  - 100 mM NAD (300µl)

Gibson Master Mix

The composition of the Gibson Master Mix is shown in Table 3.11.

Table 3.11. Components and amount of Gibson Assembly master mix

Component	Amount
5x Isothermal Buffer	2µl
0.1U/µl T5 Exonuclease	0.4 µl
Taq DNA Ligase	1µl
Phusion Taq polymerase	0.125µl

Procedure

After preparation of the appropriate solutions (Table 3.11 & 3.12) for Gibson assembly they were stored in aliquots of 3.5 µl at -80 ^{o}C in PCR tubes. One reaction tube was thawed on ice and a total volume of 6.5µl containing equimolar amounts of the to-be ligated DNA fragments

(10 ng for pET32b and 3.27 ng for Jden1381) was added. After mixing, the tubes were immediately placed in thermocycler and incubated at 37 ^{O}C for 5 minutes followed by incubation at 50 ^{o}C for 1 hour. Subsequently, the reaction mixture was immediately transformed to *Escherichia coli* T7 express C2566H electro-competent cells (NEB) (see section 3.2.13.2).

3.2.9 Competent cells for transformation

Competent cells have the ability to take up foreign DNA. There are two main types of competent cells, chemically competent and electro-competent cells. These two types of cells differ in their transformation efficiency, which is expressed as the number of transformants per microgram of plasmid DNA. The most commonly used strain for transformation is *E. coli*. Chemically competent cells are treated with $CaCl_2$ in order to facilitate attachment of plasmid DNA to the cell membrane. To open the cell membrane and allow entrance of plasmid DNA, the cells are given a heat-shock.

Rapidly growing cells (those on log phase) are made competent easier than cells in other growth stage. Electro competent cells are prepared from cells which are brought to the log phase. Transformation into electro-competent cells is facilitated by creating pores in the cells with an electrical pulse. Both transformation techniques include a cell recovery step after the actual transformation. A common medium used for post transformational recovery is Super Optimal Broth (SOC) medium.

3.2.9.1 Chemical transformation in E. coli competent cells

<u>Materials</u>

- Super Optimal Broth (SOC) medium

    - 20g Bacto Tryptone
    - 5g Bacto Yeast Extract
    - 2ml of 5M NaCl.
    - 2.5ml of 1M KCl.

- 10ml of 1M $MgCl_2$
- 10ml of 1M $MgSO_4$

All chemicals were dissolved in or mixed with dH_2O to a total volume of 900 ml before autoclaving. After cooling to room temperature 10 ml of 2 M glucose (sterile filtered) was added and the medium was stored at 4°C.

- OneShot® TOP10 chemically competent *E. coli* cells (Invitrogen)
- OneShot® BL21 Star™ (DE3) chemically competent *E. coli* cells (Invitrogen)
- JM109 chemically competent *E. coli* cells (Stratagene)
- XL1-blue chemically competent *E. coli* cells (Agilent)
- Rosetta ™ chemically competent *E. coli* cells (Merck Millipore)

JM109 and OneShot® TOP10 chemically competent *E. coli* cells were used for plasmid amplification of pUCBB-eGFP and pRSET-B respectively. OneShot® BL21 Star (DE3) chemically competent *E. coli* cells were used for periplasmic expression of *Jden*1381 variants and the R212A, D214A, S215A and E217A mutants of CelS2, while *E. coli* XL1- blue cells were used for expressing the F219A and F219Y mutants of CelS2.

For transformation, a 2.5 µl ligation-mixture was added to 25 µl competent cells in sterile cell culturing tubes followed by incubation on ice for 30 minutes. After a heat-shock at 42°C for 30 seconds the cells were cooled on ice for 1 minute. For cell recovery, 250 µl SOC medium heated to room temperature was added and the transformation mixture was incubated for 1 hour at 37°C with shaking at 220 rpm. Following centrifugation at 4000 rpm for 5 minutes, the cells were resuspended in 100 µl SOC medium and spread out on preheated LB-agar plates supplied with ampicillin. After overnight incubation of the plates at 37°C, pre-cultures of clones were made by inoculation of small liquid with transformants. Glycerol stocks of the transformed cells were made according to section 3.1.3.

To verify that the transformation was successful, PCR reactions using GoTaq Green DNA Polymerase Master Mix and appropriate primer sets were set up according to section 3.2.3.4.

3.2.9.2 Plasmid transformation by electroporation

Material

- SOC medium (see 3.11.1)
- T7 express (C2566H) electro competent *E. coli* (discussed above)

Procedure

T7 express (C2566H) electro competent *E. coli* cells were used for maintaining pET32/*Jden*1381 constructs. (3.2.3.2) For transformation of the Gibson-assembled DNA, 10 µl of cloned gene was mixed with 25 µl T7 express competent *E. coli* cells in a sterile 2 mm gap electroporation cuvette (Bio-Rad) and incubated for 5 minutes. Transformation was enhanced by using a Micopulser (Bio-Rad) giving 2kV for 5.4 milliseconds. Control cells, without added plasmid, were treated in the same way. After the pulse, 250 µl SOC medium was gently mixed and the cells followed by incubation at 37 ^{o}C for 1 hour with shaking at 220 rpm. Following centrifugation at 4000 rpm for 5 minutes, the cells were resuspended in 100 µl SOC medium and spread out on preheated LB-agar plates supplied with ampicillin. After overnight incubation at 37°C, pre-cultures of clones were made by inoculation of overnight liquid cultures with transformants. Glycerol stocks of the transformed cells were made according to section 3.1.3.

To verify that the transformations were successful, PCR reactions were set up using GoTaq Green DNA Polymerase Master Mix and appropriate primers, according to section 3.2.3.4. The primers used are described in Table 2.2 &2.3.

3.2.10 DNA sequencing

The success of the mutagenesis reactions according to section 3.2.3.3 was verified by sequencing the complete gene for every CelS2 mutant. In addition, all *Jden*1381 fragments prepared for cloning according to section 3.2.3.1 and 3.2.3.2 were sequenced to check for the introduction of undesirable mutations. For all mutants of CelS2 the in-house sequencing service was used. Sequencing reactions were prepared using the BigDye® Terminator v3.2

Cycle Sequencing Kit (Applied Biosystems) and especially designed sequencing primers (see Table 2.2). For all Jden1381 variants, the plasmids were sent to ACGT (including sequencing primers) INC (www.acgtinc.com) for sequencing. All data analysis of sequences was performed using the GENtle software suite for sequence analysis (Manske 2003).

BigDye® Terminator v3.1 Sequencing PCR

During DNA sequencing by chain-termination, the same principle as for PCR is used, only with the addition of dideoxynucleotides (ddNTPs). ddNTPs lack the 3' hydroxyl group and addition of a ddNTP instead of a dNTP to the growing DNA strand results in termination of the elongation. The four different ddNTPs have different fluorescent tags, absorbing light at different wave-lengths. After amplification, every base from the DNA-strand can be visualized by separating all possible sequence lengths and measuring the fluorescence for each amplicon. An ABI Prism 3100- genetic analyzer and the BigDye® Terminator Cycle Sequencing Kit v3.1were used for dye terminator sequencing.

Materials

- pRSET-B plasmids harboring CelS2 mutants
- BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems)
- Ready Reaction Premix
- BigDye Terminator v1.1/3.1 Sequencing Buffer (5x)
- Sequencing primers (See Table 2.2 & 2.3)
- dH_2O

For each plasmid, two reactions were set up for sequencing, one with the forward pRSET-B-primer and one with the reverse pRSET-B-primer. Both reactions were set up on ice in 0.2 ml PCR tubes according to Table 3.12. The reaction tubes were placed in a Master cycler gradient 120V (Eppendorf) and the PCR reactions were carried out using settings shown in Table 3.13.

Table 3.12. Reaction mixture for DNA sequencing.

Reagent	Quantity
Template	250 ng
Primer	3.2 pmol
BigDye Sequencing Buffer	2 µl
dH_2O	To a final reaction volume of 20 µl
Ready Reaction Premix	1 µl

Table 3.13. Thermal cycling conditions for DNA sequencing.

Reaction step	Temperature	Time (minutes:seconds)	Number of cycles
Initial Denaturation	96°C	1:00	1
Denaturation	96°C	0:10	25
Annealing	50°C	0:05	
Elongation	60°C	4:00	

3.3 Protein Expression

3.3.1 Cultivation of BL21 (DE3) cells for optimal expression

For optimal expression of *E. coli* BL21 Star™ (DE3) strains encoding each of the CelS2-N mutants and each of the *Jden*1381 variants, were cultivated in different media (LB, TB, BHI and M9 section 3.1.1 and Table 3.1), and different cultivation conditions (temperature and length of cultivation) were tested.

3.3.2 Stimulation of transcription by IPTG induction

Transcription of the gene of interest in pET32b constructs is controlled by the strong phage T7 promoter that drives transcription of downstream genes (see map in Appendix C). The T7 promoter is transcribed by T7 RNA polymerase and the transcription of this polymerase is controlled by an isopropyl β-D-thiogalactoside (IPTG)-inducible promoter. Therefore, for optimal expression of the inserted gene, T7 RNA polymerase production was induced by addition of 0.4 mM IPTG.

Materials

1 M IPTG, sterile filtered

Procedure

5 ml overnight culture was transferred to a culture flask containing 300 ml medium supplemented with 50 µg/ml ampicillin, followed by incubation at 37°C with shaking at 220 rpm until the OD_{600} reached 0.6. OD_{600} was measured using Biophotometer (Eppendorf).

Then, gene expression was induced by adding IPTG to a final concentration of 0.4 mM. After cultivation for 4 more hours, the culture was harvested and a cytoplasmic extract was made according to section 3.4.2

In the pRSET-B vector gene expression is also supposed to be controlled by the T7 RNA polymerase promoter (see map in Appendix D). However, due to the leakiness of the promoter, the pRSET-B vectors did not require IPTG induction for expressing the CelS2 mutants.

3.4 Protein extraction

3.4.1 Periplasmic extracts of *E. coli*

The CelS2-encoding gene was previously cloned into pRSET-B where it was linked to the signal peptide of CBP21 in order to enhance translocation of the LPMO to the periplasmic space. (Forsberg *et al*., 2011) The *Jden*1381 variants were cloned into pUCBB-eGFP vector as described in section 3.2.8.1 These cloned *Jden*1381 variants carry their native signal peptide. These signal peptide drive the translocation of all expressed LPMOs to the periplasmic space in *E. coli* and the LPMOs can thus be extracted by lysing the cells' outer membranes using cold osmotic shock, as described below. (Nb. This applies to all proteins in this study, except the full length *Jden*1381 cloned in pET32b that lack signal peptide (see section 3.2.8.2)

Materials

- Spheroplast buffer:
    - 50 µl 0.5 M EDTA, pH 8.0
    - ml 1 M Tris, pH 8.0
    - 8.55 g sucrose
    - 125 µl 50 mM PMSF
    - All ingredients were carefully mixed and dissolved in dH_2O to a final volume of 50 ml, and kept on ice.
- Sterile dH_2O
- 20 mM $MgCl_2$

Procedure

150 ml overnight culture was transferred to 250 ml centrifuge bottles and incubated on ice for 20 minutes before centrifuging at 5500 x g for 10 minutes, using a Beckman coulter Avanti J-25 centrifuge with a JA-14 rotor, at 4°C. After centrifugation the pellet was resuspended in 15 ml ice-cold spheroplast buffer and kept on ice for 5 minutes. After another centrifugation, this time at 10,000 x g, the pellet was resuspended in 12.5 ml ice-cold dH_2O and left on ice for 45 seconds. 625 µl 20 mM $MgCl_2$ was then added to the suspension, followed by another centrifugation as above. The final supernatant, the periplasmic extract, was filtered through a 0.22 µm sterile filter and stored at 4°C.

3.4.2 Cytoplasmic extract of E. coli

The full length *Jden*1381 gene was cloned into the pET32b expression vector without its signal peptide and transformed to *E. coli* BL21 (DE3) cells for expression (see section 3.2.9.1). Due to the lack of signal peptide, the expressed protein will be located in the cytoplasm of *E. coli* BL21 (DE3) cells and can be extracted by disrupting the cells. In this study BugBaster®Protein Extraction Reagent (Novagene®) was used for preparing cytoplasmic extracts.

Material

- BugBuster® Protein Extraction Reagent (Novagen®)

Procedure

50 ml overnight culture was transferred to 250 ml a centrifuge bottle and incubated on ice for 20 minutes before centrifuging at 8000 x g for 10 minutes, using a Beckman Coulter Avanti J-25 centrifuge with a JA-14 rotor, at 4°C. For cell lysis, the pellet was resuspended in 5 ml of BugBuster® Protein Extraction Reagent and incubated on ice for 10 minutes. For checking protein expression in the whole cell lysate, an aliquot was kept aside. The remaining cell lysate was spun down as before and the supernatant that contains the released proteins was filtered through a 0.22µl µm sterile filter and stored at 4°C.

3.5 Protein purification

For protein characterization it is important to work with purified proteins. In this study, different techniques for protein separation and purification were used, depending on the protein. The successfully expressed, CelS2 mutants (R212A, S215A, F219A and F219Y) and, the *Jden*1381 variants (the N-terminal single domain *Jden*1381-LPMO and the N-terminal double domain *Jden*1381-LPMO-CBM5/12 were purified by a combination of ion exchange chromatography and gel-filtration. The 6xHis-tagged variants; the C-terminal 6xHis tagged N-terminal single domain *Jden*1381-LPMO and the C-terminal 6xHis-tagged N-terminal double domain *Jden*1381-LPMO-CBM5/12_C-His_6 were purified by the use of a His-trap column. See Table 3.5, below, for a closer description of the various Jden1381 variants.

3.5.1 Ion-Exchange Chromatography

Ionic exchange chromatography (see section 1.11.1) is frequently used for purification of proteins, polypeptides, nucleic acids and other charged biomolecules. The technique separates proteins based on their surface ionic charge using beads modified with positively or negatively charged groups. Proteins with low binding-affinity to the column material are washed off the column by using low-salt buffer. To elute the proteins that bind strongly to the column, a high-salt buffer is used. The salt masks the charged groups on the column material allowing the protein to be eluted. Since each protein has a different charge on the surface, proteins are eluted at varying salt concentrations and may thus be separated in different elution fractions. This method was used for purification of CelS2 mutants; R212A (pI = 4.41), S215A (pI = 4.50), F219A (pI = 4.50) and F219Y (pI = 4.50) and the Jden1381 variants; Jden1381-LPMO (pI = 4.57) and Jden1381-LPMO-CBM5/12 (pI = 4.12).

Materials

- Periplasmic extract containing CelS2 mutants and Jden1381 variants (12.5 ml)
  - Binding buffer: 50 mM Tris- HCl pH 7.5
  - Elution buffer: 1 M NaCl in 50 mM Tris- HCl pH 7.5

  All buffers were filtrated through a 0.45 µm sterile- filter (Millipore)
- 10 mM Tris-HCl pH 7.5

Procedure

The ion exchange column, HiTrap™ DEAE FF, 5 ml (GE Healthcare), was connected to an Äkta purifier chromatographic system (GE Healthcare), a fully automated liquid

chromatography system, and washed with elution buffer, to remove any contamination on the column, and then equilibrated using the binding buffer. The purification steps used in this study varied slightly from time to time depending on the amount of protein prepared for purification or on the gradient of elution buffer.

The pH of the sample was adjusted to 7.5 by adding 37.5 ml of 50 mM Tris-HCl and the sample (50 ml) was loaded onto the column. Unbound proteins were eluted over approximately 25 minutes (2 column volumes) at 4ml/min flow rate. The protein of interest was eluted by a gradient from 0 % to 50 % elution buffer, over approximately 10-15 minutes at 4 ml/min flow rate. Eluted proteins were detected by online monitoring of absorption at 280 nm and collected using a fraction collector (approximately 3 ml per fraction). Fractions were analyzed using SDS-PAGE (see section 3.7). Fractions containing the (partially purified) protein of interest were pooled and concentrated to 1 ml using Amicon Ultra-15 Centrifugal Filter Units with a 3 kDa cut-off cellulose membrane (Millipore), while at the same time changing the buffer to 10 mM Tris-HCl pH 7.5 (see section 3.6.1). Further purification was done using size exclusion chromatography (see section 3.5.2). See Fig. 18a & Appendix J; Fig. J-1 & J-2 for examples of obtained chromatograms.

3.5.2 Size Exclusion Chromatography (SEC)

Size exclusion chromatography separates proteins on the basis of size and shape. The column matrix is composed of beads with pores in different sizes, where small proteins enter all the pores, thereby using longer time passing through the column and eluting after the larger proteins. See section 1.11.2 for more details.

Materials

- Running buffer (50 mM Tris-HCl, pH 7.5; 200 mM NaCl):
  50 ml 1 M Tris-HCl, pH 8.0
  11.6 g NaCl
  Dissolved in 900 ml dH_2O, regulated to pH 7.5 with 6 M HCl before the volume was brought up to 1 liter by addition of dH_2O.

Procedure

A HiLoad 16/60 Superdex G-75 column (GE Healthcare) was connected to an Äkta purifier chromatographic system (GE Healthcare). Concentrated protein (1 ml) obtained from the first step of purification (ion exchange chromatography; see section 3.5.1) was applied through a 2 ml loading loop at 0.3 ml/min flow rate, followed by application of 3 column volumes of running buffer. Protein elution was monitored by recording absorbance at 280 nm. The eluate was collected in fractions of 5 ml. Fractions putatively containing the protein of interest were analyzed by SDS-PAGE according to section 3.7. Since the over-expressed proteins yielded a highly visible band in the starting material, we considered it sufficient to determine protein identity in the chromatographic fractions by comparison with the periplasmic extract. See Fig. 4.18b & Appendix J; Fig. J-3 for examples of obtained chromatograms.

3.5.3 Protein purification by immobilized metal ion affinity chromatography

HisTrap™ HP is a ready-to-use column for preparative purification of His-tagged recombinant proteins by immobilized metal ion affinity chromatography (IMAC). The column is pre-packed with charged Ni Sepharose™ High Performance beads. Histidines form complexes with nickel ions and both *Jden*1381-LPMO_C-His_6 and *Jden*1381-LPMO-CBM5_C-His_6 will bind strongly to the column material.

Materials

- Binding buffer (20 mM Tris-HCl, 5 mM imidazole, 250 mM NaCl):
  2.4 g Tris
  0.34g Imidazole
  14.6 g NaCl
  All chemicals were dissolved in dH_2O. The pH was adjusted to 7.5 with 6 M HCl, the total volume was adjusted to 1 l and the buffer was sterile-filtered through a 0.22 µm membrane and degassed by sonication for 30 minutes.
- Elution buffer (20 mM Tris, 250 mM Imidazole, 250 mM NaCl,):
  17.02 g Imidazole
  14.6 g NaCl
  2.4g Tris

All chemicals were dissolved in dH_2O. The pH was adjusted to 7.5 with 6 M HCl, the total volume was adjusted to 1 l, and the buffer was sterile-filtered through a 0.22 µm membrane and de-gassed by sonication for 30 minutes.

- Stripping buffer (0,02 mM sodium phosphate, 500 mM NaCl, 50 mM EDTA):
  3.28 g sodium phosphate
  29.2 g NaCl
  100 ml 0.5 M EDTA, pH 8.0
  All ingredients were mixed and dissolved in 500 ml dH_2O and the pH was adjusted to 7.4 with 6 M HCl before dH_2O was added to a final volume of 1 liter. The buffer was sterile-filtered through a 0.22 µm membrane and degassed by sonication for 30 minutes.
- dH_2O
- 1 M NaCl
- 20 % Ethanol

Procedure

The pH of the sample (typically 30 ml of binding buffer) was adjusted to 7.5 through a buffer exchange with 3 x the sample volume of binding buffer using an Amicon Ultra-15 Centrifugal Filter Unit with a 10 kDa cut-off (Millipore). Prior to loading, the sample was filtered through a 0.22 µm filter, to remove any particles that could clog the column.

A His-Trap™ HP 1 ml (GE Healthcare) pre-packed column was connected to an Äkta purifier chromatographic system (GE Healthcare). The column was prepared for binding by washing with 5 column volumes of distilled water before equilibration using 5 column volumes of binding buffer. The sample (approximately 30 ml) was then loaded onto the column and binding buffer was run at 4ml/min flow rate. The binding buffer contained imidazole at a low concentration to prevent unspecific binding. The protein of interest was eluted by applying a linear gradient of 20 column volumes from 0% to 100% elution buffer. His-tagged protein was eluted at approximately 10 % of elution buffer and collected in fractions of 1 ml. For letting all non-binding proteins had passed through the column, elution and fraction collection were continued until the absorbance at 280 nm had reached the baseline. After elution, the column was washed with 20 column volumes of 1 M NaCl to remove all bound proteins, followed by washing with 20 column volumes of dH_2O and application of 20 % ethanol for storage. The purified protein was analyzed by SDS-PAGE according to section 3.7 and stored at 4°C. See Appendix J; Fig. I-4 for examples of obtained chromatograms.

If the column was to be used for purification of other proteins, a "stripping" step was performed to remove all bound molecules. The column was washed with 10 column volumes of stripping buffer, followed by 10 column volumes of binding buffer and, finally, 10 column volumes of dH_2O. The column was recharged using 0.5 ml 0.1 M $NiSO_4$ solution followed by a 5 column volume wash with dH_2O and application of 20 % ethanol before storage.

3.6 Protein concentration and concentration measurement

3.6.1 Protein concentration

Amicon® Ultra-15centrifugal Filter Unit is used for fast ultra-filtration and protein concentration. The device has a capacity for high concentration factors and easy concentrate recovery from both dilute and complex protein samples. The device is composed of a cup, a filter unit containing cellulose membrane with varied Molecular Weight Cutoff, MWCO and a centrifuge tube (for collection of filtrates). In this study, Amicon® Ultra-15 was used for concentrating purified protein and buffer change.

Materials

- Amicon®-15 Centrifugal Filter Unit, 10, 000 MWCO (Millipore)
- Purified protein
  - CelS2 mutants: $CelS2_{R212A,}$ $CelS2_{S215A}$, $CelS2_{F219A}$ $CelS2_{F219Y}$,
  - Jden1381 variants: Jden1381-LPMO, Jden1381-LPMO_C-His_6, Jden1381-LPMO-CBM5/12, Jden1381-LPMO-CBM5/12_C-His_6
- Micro-centrifuge - centrifuge 5430R (eppendorf)
- 10 mM Tris-HCl buffer

Procedure

Purified protein was added to Amicon®-15 Unit and the device was spun down at 4500 rcf at 4^oC. When buffer change is needed, the filtrate was discarded and 10 mM Tris-HCl was added to the purified protein, 3 times every 10 minutes. The processing time was 15-60 minutes depending on the volume of the sample.

3.6.2 Concentration measurement

Quick Start ™ Bradford Protein Assay is a method for determining the concentration of proteins in a sample, which is based on the binding of Coomassie Brilliant Blue G-250 dye to proteins (Bradford 1976). When the dye binds to protein it is converted from the cationic form (A_{max} = 470 nm) to the stable unprotonated blue form (A_{max} = 595 nm), and this change can be detected using a spectrophotometer. Protein concentration is calculated by using a standard curve.

Materials

- 5x Dye Reagent, Protein Assay (BioRad)
    - Commasie Blue G250 (100 mg)
    - 95% ethanol (95%)
    - 85 % phosphoric acid (85%)
    - Distilled water (up to 1 L)
- Sample buffer:
    - 10 mM Tris-HCl pH pH 8.0
- BSA Standard protein (2 mg/ml)
- Polystyrene cuvettes, 1 ml (Brand)
- Bovine serum albumin (BSA) for calibration
- Spectrophotometer (eppendorf)

Procedure

2 µl of sample was diluted in 798µl sample buffer and mixed with 200 µl 5x Dye Reagent. After 5 minutes incubation at room temperature, the sample absorbance at 595 nm was measured using spectrophotometer.

The spectrophotometer, a BioPhotometer from (Eppendorf), was calibrated with a standard curve for bovine serum albumin (BSA) with the concentrations 2, 1.5, 1, 0.75, 0.5, 0.25 and 0.125 mg/ml. Triplicates of the standards incubated in 5x Dye Reagent for 5 minutes were analyzed.

When measuring samples, a reference containing 800 µl sample buffer was mixed with 200 µl 5x Dye Reagent and incubated for 5 minutes. After the photometer was zeroed using the

reference, samples were analyzed in triplicates. Protein concentrations were calculated as the mean value of three replicates.

3.7 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a widely used technique for separation of proteins according to their mass. The addition of anionic detergents like SDS or LDS (lithium dodecyl sulphate) and a reducing agent like β-mercaptoethanol (for reducing disulphide bridges) to the sample denatures the protein and provides each protein with a uniform negative charge. The proteins are then separated based on size only by using electrophoresis and can be visualized by protein specific staining, e.g. with Coomassie brilliant blue. A molecular marker containing proteins with known masses is used to estimate the mass of the sample proteins.

Materials

- NuPage® LDS sample buffer 4x (Invitrogen)
- NuPage Sample reducing agent 10x (Invitrogen)
- 20x MES-SDS Running buffer pH 7.3:
  97.6 g MES
  60.6 g Tris Base
  10 g SDS
  3.0 g EDTA
  All chemicals were dissolved in dH_2O and the final volume was adjusted to 500 ml. Before use, the buffer was diluted 20 times in dH_2O.
- 1x wash solution:
  23.5 ml 85% (w/v) Phosphoric acid
  Diluted in 976.5 ml dH_2O
- 1x fixer solution:
  300 ml 96% (v/v) Ethanol
  23.5 ml 85% w/v) Phosphoric acid
  Diluted in dH_2O to a final volume of 1 liter
- Coomassie staining solution:

200 ml 96% (v/v) Ethanol

117.6 ml 85% (w/v) Phosphoric acid

100 g Ammonium sulphate

1.2 g Coomassie Brilliant Blue R 250

All ingredients were dissolved in dH_2O to a final volume of 1 L.

- Bench Mark™ Protein Ladder (Invitrogen) (Appendix K)

Procedure

To 18 µl protein sample, 2 µl NuPage® LDS sample buffer was added and the sample was boiled for 10 minutes. A NuPage® polyacrylamide gel, Bis-Tris 10 % (Invitrogen) was installed in the XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen) and the chamber was filled with 1x MES-SDS running buffer. In the first well, 5 µl of the protein ladder was applied, followed by application of 15 µl of each sample in the remaining wells. After running the gel using a Power supply (BioRad) for 35 minutes at 200V, the gel was released from the plastic plates and fixed in 1x fixer solution for 1 hour and washed in 1x washing solution for 2 x 10 min. For visualization of proteins, the gel was treated with the Coomassie Brilliant Blue R-250 staining solution for 2 hours. Coomassie blue binds non-specifically to all proteins. Destaining of the gel was done by washing in dH_2O for 3-5 hours.

3.8 Analysis of enzyme activity

3.8.1 Matrix-Assisted Laser Desorption and Ionization Time of Flight- mass spectroscopy (MALDI-TOF MS)

MALDI-TOF is non-quantitative technique used for determination of the masses of small molecules and proteins. MALDI-TOF MS is based on soft ionization and transportation of analytes in vacuum by using a laser. Ionized molecules are accelerated by a strong magnet field through an analyzer tube which has a mass analyzer in the end; ions are being separated based on their mass to charge ratios (*m/z)*. The Time-of- flight (TOF) analyser measures the time the ions use to reach the detector, reflecting the molecular mass of the ions. Smaller ions travel at the highest speed (Walsh 2002).

The activities of Jden1381-LPMO, Jden1381-LPMO_C-His_6, Jden1381-LPMO-CBM5/12 and Jden1381-LPMO-CBM5/12-C_His_6 towards β-chitin were analyzed (qualitatively) using MALDI-TOF MS for product analysis. As control CBP21was used as control. The activity of Jden1381 variants boosted when copper and ascorbic acid were added, therefore, 1µM copper and 1mM ascorbic acid were included in all reactions containing Jden1381 variants

Materials

- Matrix:
    - 4.5 mg DHB (2,5-Dihydroxybenzoic acid)
    - 150 µl acetonitrile
    - 350 µl dH_2O

DHB and acetonitrile were mixed thoroughly by vortexing until the DHB was completely dissolved; then dH_2O was added. The DHB solution was stored at 4°C for no more than a week.

- MALDI target- plate
- Enzymes; Jden1381-LPMO, Jden1381-LPMO_C-His_6, Jden1381-LPMO-CBM5/12, Jden1381-LPMO-CBM5/12-C_His_6 and CBP21
- Substrates: β-chitin
- Buffer: 50 mM Tris-HCl pH 8
- 0.5 mM copper chloride
- 0.1M ascorbic acid
- MALDI-TOF Mass spectrometer, Ultraflex

Procedure

All reactions (200 µl), were set up with 1 mM reducing agent, 20 mM buffer, 1µM enzyme and 5mg/ml β-chitin and 1µM $CuCl_2$ in cryo-tubes. The reaction mixtures were incubated horizontally in a shaker for 16 hours but at 37°C and 900 rpm. All reactions were run in triplicate.

A negative control, i.e. a reaction without added enzyme was set-up to account for the presence of chitooligosaccharides in the substrate solution. For a positive control, β-chitin was treated with purified CBP21 from *Serratia marcescens*, in a reaction with 1 mM ascorbic acid as reducing agent and 1µM $CuCl_2$ in Tris buffer pH 8 at 37 ^{o} C.

Samples (1 µl) were taken out from each sample and applied to a spot on the pre-cleaned MALDI plate and 2 µl of DHB-matrix was applied on top of the sample. The sample/DHB-matrix- spot was completely dried by a warm stream of air before the MALDI plate was inserted into the MALDI-TOF analyzer. The software FlexControl version 3.3 was used for controlling the system and FlexAnalysis version 3.3 was used for analysis and processing of data. Most samples were analyzed using laser beam intensities between 20% - 30% of maximum capacity.

3.8.2 High Performance Liquid Chromatography (HPLC)

HPLC is a chromatographic technique used for separation, identification, quantification and purification of individual compounds from sample mixtures. An HPLC set-up consists of a column containing the stationary phase, one or more pumps to drive the mobile phase and the samples through the column and a detector.

The column can contain different stationary phases. The retention time is calculated as the time it takes from the sample is injected until the compound is detected. HPLC delivers high performance (high resolution) because the systems run at high pressure, which again allows the use of stationary phases with very low particle sizes.

3.8.2.1 Analysis of oligomeric products from chitin by UHPLC

The activity of full length Jden1381 on different chitin substrates was analyzed by examining the production of both oxidized and native chito-oligosaccharides.

<u>Materials</u>

- Reaction buffer
  - 100 mM Tris-HCl pH 8
- Reducing agent

  - 0.1 M ascorbic acid
- Substrate
  - β- chitin, squid pen
  - α-chitin, shrimp
  - Colloidal chitin (produced in house by G. Vaaje-Kolstad according to Shimahara *et al.,* 1988)
- 1 mM $CuCl_2$
- Enzyme: Periplasmic extract from *E. coli* BL21(DE3) containing full length Jden1381
- 100% (v/v) 5mM H_2SO_4
- dH_2O
- Standard solutions containing 200 µM of native DP 1 and 6 (Megazyme).
- Column: Rezex RFQ-Fast Acid H+ (8%) column, size: 100 x 7.80 mm

Procedure

Reactions (200 µl) were set up in triplicates in cryo-tubes with 5 mg/ml substrate, 10 µl periplasmic extract in 20 mM reaction buffer, with 1 mM ascorbic acid and 1 µM $CuCl_2$. All samples were incubated for 16 hrs at 37°C with shaking at 990 rpm. After 16 hrs of incubation, the insoluble products from all the samples were separated from the soluble products by centrifugation at maximum speed and 20µl of the soluble products was transferred to a HPLC vial containing 20 µl of 50mM H_2SO4 to stop the reaction. A negative control was generated by incubating 5 mg/ml all chitin substrates in separate reactions with 10 µl of periplasmic extract from a BL21 (DE3) transformant harboring the pUCBB/empty construct, in a reaction with 1 mM ascorbic acid as reducing agent and 1µM $CuCl_2$ in Tris-HCl buffer pH 8 at 37 ^{o} C.

The samples were analyzed using an UHPLC system equipped with Rezex RFQ-Fast Acid H+ (8%) column. The samples were applied to the column by injecting 8 µl and the oligosaccharides were separated using 85 ^{O}C column temperature and a flow rate of 1ml/min. A mix of native chito-oligosaccharides with a degree of polymerization (DP) between 1 and 6 (200 µM of each) was used as standard. The recorded chromatograms were analyzed using with Chromeleon 7.0 software.

3.8.2.2 Analysis of oxidized oligomeric products from cellulose by HPAEC

For determination of optimal substrate for CelS2, the activity of $CelS2_{WT}$ was tested on Avicel and PASC (see section 2.5). The soluble products, i.e. released oxidized and native cello-oligomers were analyzed.

The activity of CelS2 mutants R212A, S215A, F219A and F219Y on PASC was analyzed by examining soluble products. For in depth analysis, the formation of oxidized products generated over time by $CelS2_{S215A}$ and $CelS2_{WT}$ on PASC were assessed (See Fig. 4.19 bellow). For monitoring the time course of product formation from the action $CelS2_{S215A}$ and $CelS2_{WT}$ on PASC, soluble products (native and oxidized oligomers), those released to the solvent and remain trapped in the solid fraction, were analyzed (for simplicity, here and after, the later are referred as "trapped" products). To permit HPAEC analysis and quantification of the products, both products generated by the CelS2 variants were further degraded with Cel5A, an endoglucanase from *Thermobifida fusca* (Jung *et al.*, 2001)

The addition of copper to the reaction containing CelS2 variants, decreased activity. Therefore, further analyses of CelS2 were made without addition of copper.

Materials

- Reaction buffer
  - 50 mM Bis-Tris pH 6.5
- Reducing agents
  - 0.1 mM ascorbic acid
- 30 mM EDTA
- Other reducing agents used for optimization
  - 0.1 M L-Reduced glutathione
  - 0.1 M Gallic acid
  - 0.1 M Gluconic acid
  - 0.1 M D-Glucose amine
  - 0.1 M Tartaric acid
- Purified enzymes
  - $CelS2_{WT}$
  - $CelS2_{R212A}$

- $CelS2_{S215A}$
- $CelS2_{F219A}$
- $CelS2_{F219Y}$

- Substrates (see section 2.5)
- Column: CarboPac PA1, 2x250 mm
  - Pre column: CarboPac^{TM} PA1, 2x50 mm
- HPLC Buffer:
  - Buffer A 0.1 M NaOH (Degassed)
  - Buffer B 0.1M NaOH, 1M NaOAc (Degassed)
- dH_2O

- Standards: Mixture of (Glc4, Glc5, GlcGlc-ox, Glc_2Glc-ox, Glc_3Glc-ox), varied concentration
  - Native standards were purchased from Megazyme
  - Oxidized standards were prepared by enzymatic (LPMO) degradation.
- Purified Cel5A from *Thermobifida fusca* (2.5mg/ml)

Procedure

For determination of the substrate preference of wild-type CelS2, reactions (200 µl) were set up in triplicates in cryo-tubes with 5 mg/ml Avicel or 2 mg/ml PASC, 1 µM enzyme in 20 mM reaction buffer, with 1 mM ascorbic acid. All samples were incubated at 50°C with shaking at 990 rpm. The products generated after 16 h were analyzed by using a Dionex-Bio-LC equipped with a CarboPac PA1 column as described below.

For analyzing the time course of product formation by $CelS2_{WT}$ and the $CelS2_{S215A}$ mutant, reactions (600 µl) were set up in triplicates in cryo-tubes with 2 mg/ml PASC, 1 µM enzyme in 20 mM buffer, with 1 mM ascorbic acid in 20 mM Bis-Tris PH6.5. All samples were incubated at 50°C with shaking at 990 rpm. At appropriate times, 30 µl samples were taken by transferring 30 µl of the reaction to sample tubes that were centrifuged at maximum speed. The supernatant (containing the soluble products) was separated from the pellet (containing the trapped products). Before HPAEC, Cel5A was added to both the supernatant and the pellet (resuspended in 30 µl of 20 mM buffer containing 1µl of 30 mM EDTA) for

depolymerization of all poly- and oligosaccharides present. The final concentration of Cel5A used for soluble and trapped products was 8µMand 16 µM respectively. The reaction mixtures for the soluble products were incubated for 30 minutes at 50°C with shaking at 990 rpm, while reactions for the trapped products were incubated at similar conditions for 16 hrs. The reaction of Cel5A was stopped by adding 41µl 0.1M NaOH and the products were analyzed by using a Dionex-Bio-LC system equipped with a CarboPac PA1 column. The amount of sample applied to the column was 2.0µl and application was done using a flow rate of 0.25 ml/min with 0.1 M NaOH at a column temperature of 30 ^{o}C. Cello-oligosaccharides were eluted at flow rate of 0.25 ml/min at a column temperature of 30 ^{o}C by applying a stepwise linear gradient with increasing amounts of sodium acetate (NaOAc) by changing from 0.1M NaOH to 0.1M NaOH/0.1M NaOAc in 10 minutes, then to 0.1 M NaOH/0.3MNaOAc in 35 minutes following by 0.1M NaOH and 1.0M NaOAc in 40 minutes. Before application of the next sample, the column was reconditioned with 0.1 M NaOH for 9 minutes. Eluted cello-oligosaccharides were monitored using an electrochemical detector (ED 1) equipped with a disposable gold (Au) working electrode, silver /silver chloride (Ag/AgCl) reference electrode and a titanium auxiliary electrode (http://www.dionex.com). Both running and elution buffers were degassed for 15 minutes and kept under pressure (N_2, approximately 5 psi) to prevent adsorption of atmospheric carbondioxide. The recorded chromatograms were analyzed using Chromeleon 7.0.

A negative control, i.e. a reaction without added enzyme was set-up to account for the presence of cellulose oligomers in the substrates. For a positive control, PASC (2mg/ml) or Avicel (5mg/ml) were treated with purified $CelS2_{WT}$ (1µM) with 1 mM ascorbic acid as reducing agent in 20 mM Bis-Tris pH 6.5. Both the positive and negative controls were incubated at 50°C with shaking at 990 rpm. Both control samples were incubated for corresponding incubation times as the main samples. For accuracy of data analysis, reaction buffer was run as blank and any noise resulted from the buffer were subtracted from the main peaks (those generated from chitooligosaccharides).

3.9 Bioinformatics methods

All genomic and proteomic computational analyses were done by using publicly available bioinformatics tools. Table 3.10 show a list of bioinformatics tools used in this study.

Table 3.14. Overview of bioinformatic tool used in this study**.**

Bioinformatics tools	Purpose	Web address
Expasy ProtParam	To compute the physical and chemical parameters of CelS2 and *Jden*1381	http://web.expasy.org/protparam/
CAZy	For annotation of carbohydrate active proteins	http://www.uniprot.org/
pfam	For annotation of domain structures of Cels2 and Jden1381	http://pfam.sanger.ac.uk//
Expasy SwissModel	For structural prediction of CelS2	http://swissmodel.expasy.org/
T-coffee server	For sequence alignment of CelS2 with homologous LPMOs	http://www.ebi.ac.uk/Tools/msa/tcoffee
ClustalW	For sequence alignment of CBP21 and the LPMO domain of Jden1381	http://www.ebi.ac.uk/Tools/msa/clustalw

4 RESULTS

4.1 Bioinformatic analysis of CelS2 and Jden1381

4.1.1 Domain structure and physiochemical properties

CelS2 is one of the seven putative CBM33-type LPMOs encoded in the genome of *S. coelicolor*. Analysis of the domain structure of the enzyme using the Pfam database (www.pfam.org) showed that it consists of a signal peptide, an N-terminal CBM33-type LPMO domain and a C-terminal CBM2 cellulose binding domain (Fig 4.1A). Jden1381 encoded in the genome of *J. denitrificans* contains signal peptide followed by a CBM33-type LPMO domain, a CBM5/12 chitin binding domain and a GH18 chitinase domain (Fig 4.1B). The amino acid sequences of CelS2 and Jden1381 are shown in Appendix L and M respectively.

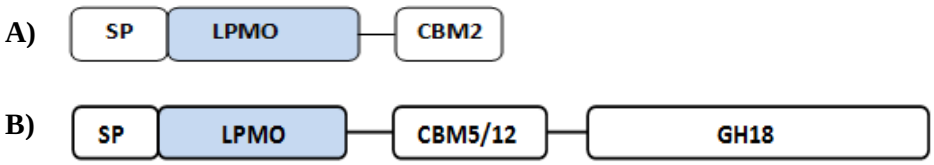


Fig 4.1. Modular structure of CelS2 (A) and Jden1381 (B) Abbreviations: SP, signal peptide; LPMO, Lytic polysaccharide monooxygenases (CBM33-type in both cases); CBM2, family 2 carbohydrate binding domain; CBM5/12, family 5 or 12 carbohydrate binding domain and GH18, family 18 glycoside hydrolase. Nb. The family 5 and 12 CBMs are separated in CAZy but are one class in Pfam.

Using the Expasy ProtParam tool the physiochemical properties of the N-terminal LPMO domain of CelS2 (CelS2-N), the full length proteins and the truncated variants of Jden1381 were calculated (Table 4.1).

Table 4.1. Physico-chemical properties of CelS2-N and of Jden1381 variants. The physical and chemical parameters of CelS2-N and Jden1381 were computed using the Expasy ProtParam tool. Calculations do not include the signal peptide (except for bp), nor possible tags, but only the mature proteins and their truncated forms. Note that all work with CelS2 was done with a truncated form of CelS2 comprising only its N-terminal LPMO domain and referred to as CelS2-N (see Fig. 4.1)

Protein	Uniprot	bp	Amino acids	Mw (kDa)	pI
CelS2-N	Q9RJY2	684	1 - 93	21.9	4.48
Jden1381, full length	C7R4I0	1956	1 - 620	66.3	4.26
Truncated variants					
Jden1381-LPMO	-	522	1 - 142	15.5	4.57
Jden1381-LPMO-CBM5/12	-	831	1 - 245	25.8	4.12

CelS2-N from Streptomyces coelicolor

4.1.2 Multiple sequence alignment (MSA)

The amino acid sequence of 13 bacterial LPMOs were compared with the N-terminal domain of CelS2 (Uniprot accession Q9RJY2) in a multiple sequence alignment made using the T-coffee server (http://www.ebi.ac.uk/Tools/msa/tcoffee) (Fig. 4.2) in order to identify highly conserved residues. The sequence alignment shows considerable sequence conservation including the histidines responsible for copper binding (Figs. 4.2 and 4.3). Conserved amino acids targeted in this study were selected by combination of results from this multiple sequence alignment and inspection of a model of the active site of CelS2 obtained using homology modeling (see section 4.1.3).

```
tr|A4X9B3|A4X9B3_SALTO    MSSYRART-AALLTAATTLLAA-AAV-LTVRSEPAAAHGAAMVPGSRTFL 46
tr|A8LVV7|A8LVV7_SALAI    MSY-RSRT-AALLTAAATLLVA-VVA-LTARSEPAAAHGAAMVPGSRTFL 45
tr|B1VNK5|B1VNK5_STRGG    MARR-KKLLTSLVAVLATLLGG-IGL-TLMGQDNAQAHGVTMTPGSRTYL 47
tr|C4RK97|C4RK97_9ACTO    -----------MLIAAATLAAG-AAT-LVASPNPAAAHGAAMTPGSRTYL 37
tr|D2BFT3|D2BFT3_STRRD    MSKWR-SI--AVAAA-VALMST-LLA-VVLVPGQASAHGAMMVPGSRTFF 44
tr|D5ZX22|D5ZX22_9ACTO    MTLR-SRF-VSLAAVLATLLGG-LGL-SFLWQNNAQAHGVAMVPGSRTYL 46
tr|D6ESG8|D6ESG8_STRLI    MVRR-TRL-LTLAAVLATLLGS-LGVTLLLGQGRAEAHGVAMMPGSRTYL 47
tr|D6Y7U3|D6Y7U3_THEBD    MGRLRRSTSKTTVSISAAALAA-LVA-SVLPSSPAAAHGAMMMPGSRTYL 48
tr|D8I1M0|D8I1M0_AMYMU    MTR-R-RS--TILAAVTVLLAS-LTA-ILLNTGTAEAHGAMMKPGSRTFL 44
tr|D9T4V8|D9T4V8_MICAI    MHR--SRT-AALLTAAATLALG-AFA-LATNSGPAAAHGAAMTPGARTYL 45
tr|E4N0M0|E4N0M0_KITSK    MTRR-T-T-PTLLAALATAVAT-LCA-LLVGQPPAQAHGVAVVPGSRTYL 46
tr|E8W5Y2|E8W5Y2_STRFA    MARG-KRLLVSLTAVFATLLGG-IAL-TLFGQGNAQAHGVTMTPGSRTYL 48
tr|F4FFD8|F4FFD8_VERMA    MA--------AMFFAAVTLAASAVVA-VTASPDPAAAHGAAMTPGARTYL 42
tr|Q9RJY2|Q9RJY2_STRCO    MVRR-TRL-LTLAAVLATLLGS-LGVTLLLGQGRAEAHGVAMMPGSRTYL 47
                                        ..               * ***. : **:**::

tr|A4X9B3|A4X9B3_SALTO    CWQDGLSPTGEIQPYNPACSAAVDQSGANSLYNWFSVLRSDADGRTVGFI 96
tr|A8LVV7|A8LVV7_SALAI    CWKDGLTPSGEIQPNNPACSAAVAQSGANSLYNWFSVLRSDADGRTVGFI 95
tr|B1VNK5|B1VNK5_STRGG    CMLDARTGTGALDPTNPACKAALDESGANALYNWFAVLDSNAGGRGAGYV 97
tr|C4RK97|C4RK97_9ACTO    CWKDGLAPTGEIKPNNPACSAAVAQNGPNSLYNWFSVLRSDAGGRTVGFI 87
tr|D2BFT3|D2BFT3_STRRD    CWQDGLSSTGQIIPINPACGAAVAQSGPNSLYNWFSVLRSDGAGRTRGFI 94
tr|D5ZX22|D5ZX22_9ACTO    CQLDAITGTGALNPTNPACRDALNKSGSSALYNWFAVLDSRAAGRGPGYV 96
tr|D6ESG8|D6ESG8_STRLI    CQLDAKTGTGALDPTNPACQAALDQSGATALYNWFAVLDSNAGGRGAGYV 97
tr|D6Y7U3|D6Y7U3_THEBD    CWKDGLTPQGNIVPNNPACAAAVAQSGTNALYNWFAVLRSDGAGRTRGYI 98
tr|D8I1M0|D8I1M0_AMYMU    CWQDGLSSTGEIKPINPACAAAVGVSGANSLYNWFAVLRSDGAGRTRGFI 94
tr|D9T4V8|D9T4V8_MICAI    CWKDGLTGTGEIRPNNPACSSAVAANGANSLYNWFSVLRSDAGGRTVGFI 95
tr|E4N0M0|E4N0M0_KITSK    CYQDGRTSTGALDPTNPACRAALAQSGTTPLYNWFAVLDSNAGGRGQGYV 96
tr|E8W5Y2|E8W5Y2_STRFA    CWLDAKTSTGSLDPTNPACKAALSESGSNALYNWFAVLDSNAGGRGAGYV 98
tr|F4FFD8|F4FFD8_VERMA    CWRDGLSPTGEIRPQNPACSAAVAQSGANSLYNWFSVLRSDAGGRTTGFI 92
tr|Q9RJY2|Q9RJY2_STRCO    CQLDAKTGTGALDPTNPACQAALDQSGATALYNWFAVLDSNAGGRGAGYV 97
                          *  *. :  * : * ****  *:  .*...*****:** * . **  *::

tr|A4X9B3|A4X9B3_SALTO    PDGQLCSGGNP---GFLGYDLARIDWPLTHLTAGQNIEFRYSNWAHHPGT 143
tr|A8LVV7|A8LVV7_SALAI    PDGQLCSGGNP---GFLGYDLARTDWPLTHLTAGRTMEFRYSNWAHHPGT 142
tr|B1VNK5|B1VNK5_STRGG    PDGTLCSAGDRSPYDFTGYNAPRSDWPRTHLTAGKTIQVKHSNWAAHPGS 147
tr|C4RK97|C4RK97_9ACTO    PDGKLCSGGNP---GFSGYDAARTDWPLTHLTAGARFDFKYSNWAHHPGT 134
tr|D2BFT3|D2BFT3_STRRD    PDGQACSGGNP---GYSGFDLPRADWPVTHLTAGAGIQFKYNKWAAHPGW 141
tr|D5ZX22|D5ZX22_9ACTO    PDGTLCSAGDRSPYDFSAYNAARADWPRTHLTSGATVKVQYSNWAAHPGD 146
tr|D6ESG8|D6ESG8_STRLI    PDGTLCSAGDRSPYDFSAYNAARSDWPRTHLTSGATIPVEYSNWAAHPGD 147
tr|D6Y7U3|D6Y7U3_THEBD    PDGKLCSADAK-VYDFSGFDLARDDWPVTHLTAGATIQIRYNMWAHHPGT 148
tr|D8I1M0|D8I1M0_AMYMU    PDGKLCSGGNP---NYAGFDGV-GAWPLTHLTSGAQFDFSYNAWAAHPGW 140
tr|D9T4V8|D9T4V8_MICAI    PDGKLCSGGNP---GFSGYDAARNDWPITHLTAGRSMEFRYSNWAHHPGT 142
tr|E4N0M0|E4N0M0_KITSK    PDGTLCSAGNKSPYDFSAYNAPRDDWPRTHLTAGAAIEVDYSNWAAHPGE 146
tr|E8W5Y2|E8W5Y2_STRFA    PDGKLCSAGDRSPYNFTGYNAARSDWPRTHLTAGRTIQVKHSNWAAHPGS 148
tr|F4FFD8|F4FFD8_VERMA    PDGQLCSGGAT---GFRGFDLARDDWPLTHLTAGRTMEFRYSNWAHHPGT 139
tr|Q9RJY2|Q9RJY2_STRCO    PDGTLCSAGDRSPYDFSAYNAARSDWPRTHLTSGATIPVEYSNWAAHPGD 147
                          ***  **..     .: .::     ** ****:*  . . :. ** ***

tr|A4X9B3|A4X9B3_SALTO    FYFYVTKDSWSPTRPLAWSDLEEQPFLTVTNPPQRGGPGTDDGHYYFAGT 193
tr|A8LVV7|A8LVV7_SALAI    FSFYITKDSWSPTRPLAWSDLEEQPFLTVTNPPQRGAVGTNDGHYYFTGT 192
tr|B1VNK5|B1VNK5_STRGG    FRVYLSKPGYSPSTELGWDDLEL--IETVTDPPQSGGPGTDGGHYYWNLD 195
tr|C4RK97|C4RK97_9ACTO    FYFYVTKDSWSPTRALAWSDLETQPFLTVTNPPQNGPVGTNEGHYYFSGN 184
tr|D2BFT3|D2BFT3_STRRD    FYLYVTKDGWNPNQALTWDDLESQPFHTADHPQSVGSPGTNDAHYYWNAT 191
```

Fig 4.2. Multiple sequence alignment of the N-terminal domain of CelS2 with 13 bacterial CBM33-type LPMO domains. The degree of residue conservation is indicated under the sequences with: and *, based on standard settings in T-coffee. Residues marked with * are fully conserved. Two conserved metal binding histidines are highlighted grey. The five conserved amino acid that are predicted to be located in the catalytic site of CelS2 (see Fig. 4.4) and that were targeted in the present study are highlighted blue. The red rectangle indicates CelS2. Note that the first of the two conserved histidines is residue number 1 in the mature, secreted proteins. According to normal practice, residue numbering in this thesis is according to this sequence alignment; this means that the mature protein starts with His35.

4.1.3 Homology modeling

The structure of CelS2-N was predicted by using Swiss-model, an automated homology modeling server that is accessible via the Expasy web server (http://swissmodel.expasy.org/). The CelS2-N structure was predicted using CBP21 (pdb code 2BEM), a chitin-active LPMO from *Serratia marcescens*, as template. A sequence alignment of CelS2 and CBP21 is shown in figure 4.3. The overall sequence identity between the two proteins (excluding their signal peptides) is 22 %. Because of the low sequence identity and the presence of various large insertions and deletions in the sequence alignment (Fig. 4.3) the model of CelS2-N is expected to contain major errors. However, the model did produce a plausible active site (with several conserved residues), and this part of the model was considered sufficiently reliable for further work.

```
CBP21    MNKTSRTLLSLGLLSAAMFG--------VSQQANAHGYVESPASRAYQCK 42
CelS2-N  MVRRTRL-LTLAAVLATLLGSLGVTLLLGQGRAEAHGVAMMPGSRTYLCQ 49
         * : :*  *:*. : *:::*         . :*:*** .  *.**:* *:

CBP21    LQLNTQCGSVQYEPQSVEGLKGFPQAGP---------------------- 70
CelS2-N  LDAKTGTGAL--DPTNPACQAALDQSGATALYNWFAVLDSNAGGRGAGYV 97
         *: :*  *::  :* .     .: *:*.

CBP21    ADGHIASADKSTFFELD--QQTPTRWNKLNLKTGPNS-FTWKLTARHSTT 147
CelS2-N  PDGTLCSAGDRSPYDFSAYNAARSDWPRTHLTSGATIPVEYSNWAAHP-G 116
         .** :.**.. : :::.  : : : * : :*.:*..  . :.  * *.

CBP21    SWRYFITKPNWDASQPLTRASFDLTPFCQFNDGGAIPAAQV---THQCNI 164
CelS2-N  DFRVYLTKPGWSPTSELGWDDLELIQTVTNPPQQGSPGTDGGHYYWDLAL 196
         .:* ::***.*..:. *   .::*          . *.::      :  :

CBP21    PADRSGSHVILAVWDIADTANAFYQAIDVNLSK----------------- 197
CelS2-N  PSGRSGDALIFMQWVRSDSQENFFSCSDVVF------------------- 227
         *:.***. :*:  *  :*: : *:.. ** :
```

Fig 4.3. Pairwise alignment of CelS2-N and CBP21. The conserved metal binding histidines are highlighted grey. Among the five highly conserved residues of CelS2 that were selected for mutation (Fig 4.2) the two that are identical to residues in the catalytic site of CBP21 are highlighted green (see Fig. 4.4), whereas the other three residues are highlighted red.

Based on the MSA (Fig 4.2), the pairwise alignment of CelS2-N and CBP21 (Fig 4.3) and analysis of the predicted CelS2-N structure, the conserved region of the metal binding site was mapped for potentially important residues (Fig 4.4). Five conserved residues (Fig. 4.3) located in close vicinity to the metal binding histidines (R212, D214, S215, E217 and F219) were selected for site-directed mutagenesis in order to probe their function. Three of these otherwise conserved residues differ between chitin-active CBP21 and cellulose active CelS2-N. Of these, Arg212, which corresponds to Ile180 in CBP21 caught special attention. In CBP21 this residue defines the wall of a small “hole” close to the metal binding site (Fig. 5.2) The arginine in CelS2-N seems to fill this hole (see Ch. 5 for further Discussion).

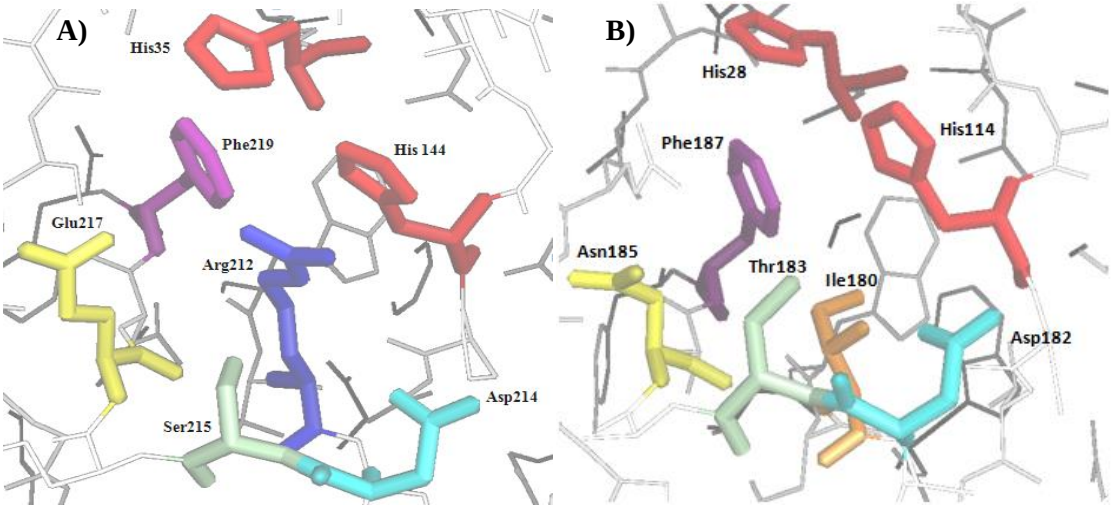


Fig 4.4. Comparison of CBP21 and CelS2-N active sites. (A) Model of CelS2-N obtained by homology-modelling using CBP21 (PDB code 2BEM) as template. (B) Active site of CBP21. The metal binding histidines are labeled and shown as red sticks, i.e His35 and His144 for CelS2-N and His28 and His114 for CBP21. The five amino acids selected for mutagenesis in CelS2-N and their equivalents in CBP21 are also labeled and shown as sticks with varying colors: Arg 212= blue; Asp214 = cyan; Ser215 = pale-green; Glu217 = yellow and Phe219 = purple. The corresponding residues in CBP21 are: Ile180 = orange; Asp182 = cyan; Thr183 = pale-green; Asn185 = yellow and Phe187 = purple) The figures were made using PyMOL (DeLano, W. L. *et al.*, 2005)

Jden1381 from *Jonesia denitrificans*

4.1.4 Sequence alignment

The two catalytic domains of Jden1381 were aligned with relative enzymes (CBP21 from *S. marcescens* for the LPMO domain and ChiC from *S. coelicolor* for the GH18 domain; Fig. 4.5 and 4.6). Jden1381-LPMO has 32 % sequence identity with CBP21 while Jden1381-GH18 has 69 % sequence identity with ChiC. The presence of conserved catalytic motifs essential for activity (metal binding histidines for the LPMOs and the DXDXE motif containing the catalytic glutamate for the GH18 chitinases) was detected in the respective Jden1381 domains.

```
Jden1381-LPMO    HGWVTDPPSRQALCASGETSFDCGQISYEPQSVEAPKG--------ATTC 73
CBP21            HGYVESPASRAYQCKL-QLNTQCGSVQYEPQSVEGLKGFPQAGPADGHIA 76
                 **:* .*.**   *   : . :**.:.*******. **        .  .

Jden1381-LPMO    SGGNEAFAILDDNSKP-WPTTEIAS-TVDLTWKLTAPHNTSTWEYFVDG- 120
CBP21            SADKSTFFELDQQTPTRWNKLNLKTGPNSFTWKLTARHSTTSWRYFITKP 126
                 *..:.:*  **::: . * . :: : . .:****** *.*::*.**:

Jden1381-LPMO    ------QLHQT---------FDQKGQQPPTSLTHTLTDLP--TGEHTILA 153
CBP21            NWDASQPLTRASFDLTPFCQFNDGGAIPAAQVTHQCNIPADRSGSHVILA 176
                       * ::         *:: *  *.:.:**  .  .  :*.*.***

Jden1381-LPMO    RWNVSNTNNAFYNCMDVVVS----------------------------- 173
CBP21            VWDIADTANAFYQAIDVNLSK---------------------------- 197
                  *::::* ****:.:** :*
```

Fig. 4.5. Sequence alignment of the Jden1381 LPMO domain and CBP21. The Jden1381 LPMO domain contains several large inserts relative to CBP21, but shows strong sequence conservation near the conserved crucial histidines (yellow). Signal peptides are not included in the alignment. The alignment was generated by ClustalW2.

```
Jden1381-GH18    DTPGTGDERIVGYFTNWGVYGRDYHVKNIKTSGAADHLTHIMYAFGNVQGGKCTIGDAYA 336
ChiC             -NPGTGAEVKMGYFTNWGVYGRNYHVKNLVTSGSADKITHINYAFGNVQGGKCTIGDSYA 292
                  .**** *  :***********:*****: ***:**::*** ***************:**

Jden1381-GH18    DYDKAYTAAQSVDGVADTWDQPLRGNFNQLRKLKAEYPHIKVVWSFGGWTWSGGFGQAAQ 396
ChiC             DYDKAYTADQSVDGVADTWDQPLRGNFNQLRKLKAKYPNIKILYSFGGWTWSGGFPDAVK 352
                 ******** *************************:**:**:::*********** :*.:

Jden1381-GH18    NPEAFAQSCRDLVEDPRWADVFDGIDIDWEYPNACGATCD-TSGRDAYRDLLAALRTEFG 455
ChiC             NPAAFAKSCHDLVEDPRWADVFDGIDLDWEYPNACGLSCDETSAPNAFSSMMKAMRAEFG 412
                 ** ***:**:****************:********* :** **. :*: .:: *:*:***

Jden1381-GH18    DD-LVTSAIPADATDGGKIDAANYAGGAEYLDWIMPMSYDYFGAWDKNGPTAPHSPLTSY 514
ChiC             QDYLITAAVTADGSDGGKIDAADYGEASKYIDWYNVMTYDFFGAWAKNGPTAPHSPLTAY 472
                 :* *:*:*:.**.:********:*. .::*:**   *:**:**** ************:*

Jden1381-GH18    QGIPIQGYDTTSTINKLTGLGIPADKILLGIGFYGRGWTGVTDPTPGSSATGAAPGTYEA 574
ChiC             DGIPQQGFNTADAMAKFKSKGVPADKLLIGIGFYGRGWTGVTQSAPGGTATGPATGTYEA 532
                 :*** **::*:.:: *:.. *:****:*:*************:.:**.:***.*.*****

Jden1381-GH18    GIEDYKVLAQRCPATGQVAGTSYGFCDGQWWSYDTPQDIIHKMNYANTENLGGAFFWELS 634
ChiC             GIEDYKVLKNSCPATGTIAGTAYAHCGSNWWSYDTPATIKSKMDWAEQQGLGGAFFWEFS 592
                 ******** : ***** :***:*..*..:*******  *  **::*: :.********:*

Jden1381-GH18    GDTADGDLITAIATGLQ---------------------------------------- 651
ChiC             GDTTNGELVSAIDSGLK---------------------------------------- 609
                 ***::*:*::** :**:
```

Fig 4.6. Sequence alignment of the Jden1381-GH18 domain (residues 277 – 651) and ChiC domain (residues 234 - 609) from *Streptomyces coelicolor*. The catalytic DXDXE motif diagnostic for GH18 chitinases is shaded yellow. ChiC from *S. coelicolor* has been shown to be highly up-regulated when *S. coelicolor* is induced by chitobiose ([1]Saito 2000) Signal peptides and other domains are not included in the alignment.

4.1.5 Gene optimization

The full length *Jden*1381 gene was codon optimized for expression in *E. coli* using the OptimumGene^{TM} Gene Design software (http://www.genscript.com/codon_opt.html) prior to gene synthesis. The sequence identity of the codon optimized gene with the original nucleotide sequence is 76 %. Furthermore, the GC- content (Guanine-Cytosine content) of the codon optimized gene decreased by 7 % (from 58% to 51 %) compared to the original nucleotide sequence (Appendix F). The synthetic gene displayed in Appendix F was used as starting point for all *Jden*1381-constructs.

4.2 Mutagenesis, molecular cloning and transformation

4.2.1 Site Directed Mutagenesis of CelS2

PCR products for all mutations (see section 3.2.3.3 for details) were analyzed by agarose gel before transformation (see Fig 4.7 for examples). Identification of correct transformants and introduction of correct mutations were verified by DNA sequencing. Plasmids were isolated from four transformants, on average, per mutant for gene sequence verification according to section 3.2.1. Gene sequencing confirmed all mutants to be correctly inserted without any PCR-generated errors in the sequence.

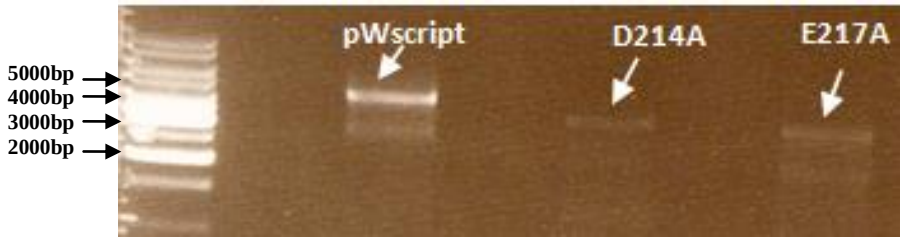


Fig 4.7. Agarose gel showing results of PCR amplifications for site-directed mutagenesis of CelS2. The left lane shows a 1 Kb DNA ladder. The sizes (bp) of some markers are indicated. The lanes containing the PCR genes are labeled by the type of mutant that was made; the lane labeled pWscript shows the result of a control reaction using ingredients supplied with the Quick-Change kit. The expected fragment sizes are 4500 bp for pWscript and 3500 bp for D214A and E217A.

4.2.2 Gene cloning of Jden1381 and gene truncation

Expression of Jden1381 variants caused several problems and, therefore, a large series of approaches was tested, using different truncations, different tags, different expression systems and different host strains. An overview over all successfully constructed expression vectors is provided in Table 4.2. A schematic overview over truncated versions of Jden1381 is also provided in Fig. 4.8

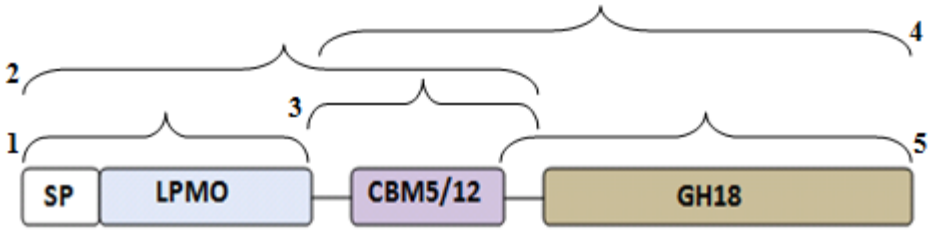


Fig 4.8. Simplified illustration showing the truncated forms of *Jden*1381 that were generated in this study. The constructs labeled 1 – 5 are referred in the main text and in Table 4.2 as:1, Jden1381-LPMO; 2, Jden1381-LPMO-CBM5/12; 3, Jden1381-CBM5/12; 4, Jden1381-CBM5/12-GH18; 5, Jden1381-GH18. Note that there are linker peptides connecting the CBM5/12 domain to the rest of the protein and that these linker peptide, referred to as N-terminal or C-terminal, respectively, are included in some of the constructs.

Table 4.2. Overview of constructs for expression of Jden1381 variants. For expression all constructs were transformed to chemically competent *E.coli* BL21(DE3). Of the 13 constructs, six expressed as soluble proteins. See also Fig. 4.8 for a graphical representation of the truncated Jden1381 variants and Table 3.5 for overview of all expected soluble proteins.

Constructs	Amino acids	Tags	SP	Soluble protein
pET32b/*Jden*1381fl	1 - 626	His_6, N-terminal	No	LPMO-CBM5/12-GH18
pUCBB/*Jden*1381 fl	1-620	No	Yes	LPMO-CBM5/12-GH18
pUCBB/*Jden*1381fl_C-His_6	1-626	His_6, C-terminal	Yes	-
pUCBB/*Jden*1381-LPMO	1-142	No	Yes	LPMO
pUCBB/*Jden*1381-LPMO_C-His_6	1-148	His_6, C-terminal	Yes	LPMO
pUCBB/*Jden*1381-LPMO-CBM5/12	1-245	No	Yes	LPMO-CBM5/12
pUCBB/*Jden*1381-LPMO-CBM5/12_C-His_6	1-251	His_6, C-terminal	Yes	LPMO-CBM5/12
pUCBB/Jden1381-CBM5/12	1-103	No	No	-
pUCBB/Jden1381-CBM5/12_C-His_6	1-106	His_6, C-terminal	No	-

pUCBB/*Jden*1381-CBM5/12-GH18	1-478	No	No	-
pUCBB/*Jden*1381-CBM5/12-GH18_C-His_6	1-484	His_6, C-terminal	No	-
pUCBB/Jden1381-GH18	1-375	No	No	-
pUCBB/Jden1381-GH18_C-His_6	1-381	His_6, C-terminal	No	-

Since we expected expression and/or purification problems two variants of each truncated protein were generated, one with no tag and one with a C-terminal 6xHis tag (indicated by "-C-His_6" in the construct name). To minimize the chances of misplacing domain borders, linker peptides (Fig. 4.8) were generally included in the truncated versions (except for Jden1381-LPMO). All methods and strategies used for cloning are described in detail in section 3.2.8. Progress in cloning steps was monitored using agarose gel electrophoresis and, where appropriate, DNA sequencing. Examples of agarose gel-based verifications of cloning steps are shown in Figures 4.9, 4.10 and 4.11. Transformants were usually verified by colony PCR (see section 3.2.3.4) and an example of colony PCR verification is shown in Fig. 12.

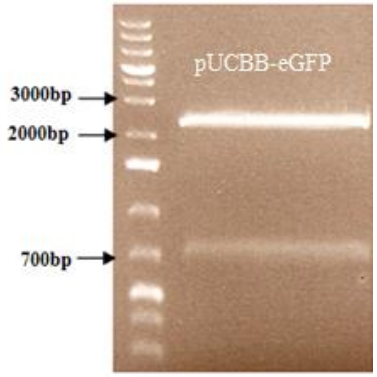


Fig 4.9. Agarose gel electrophoresis of the double restriction digested expression vector pUCBB-eGFP. The left lane shows the 1 Kb DNA ladder. The sizes (bp) of some marker fragments are indicated. The vector was digested with *Bam*HI and *Not*I restriction enzymes. The expected sizes of the generated fragments are 2374 bp and 700 bp where the 2374 bp fragment represents the vector backbone used for generation of pUCBB/Jden1381fl, pUCBB/Jden1381_C-His_6, pUCBB/*Jden*1381-LPMO, pUCBB/*Jden*1381-LPMO_C-His_6, pUCBB/*Jden*1381-LPMO-CBM5/12, pUCBB/*Jden*1381-LPMO-CBM5/12_C-His_6, pUCBB/*Jden*1381-CBM5/12 pUCBB/*Jden*1381-CBM5/12_C-His_6, pUCBB/*Jden*1381-CBM5/12-GH18, pUCBB/*Jden*1381- CBM5/12-GH18_C-His_6, pUCBB/*Jden*1381-GH18 and pUCBB/*Jden*1381-GH18_C-His_6

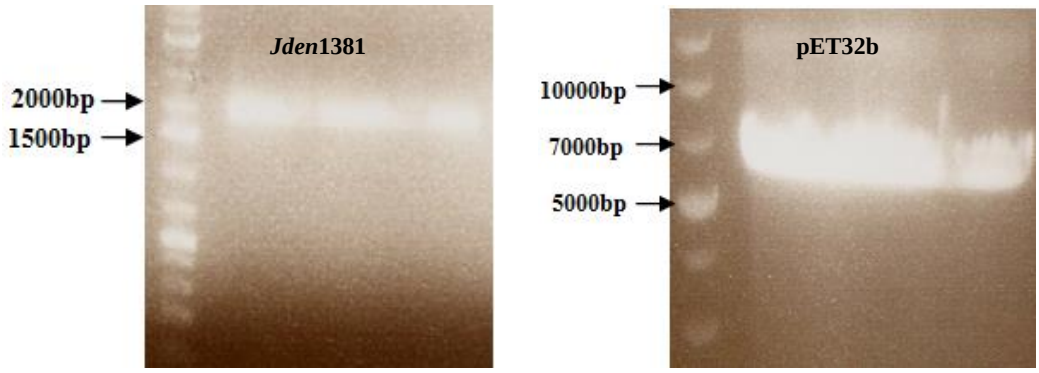


Fig 4.10. PCR amplification of the *Jden*1381gene and pET32b visualized on an 1% agarose gel. These amplifications were done for the generation of pET32b/*Jden*1381 by Gibson assembly (section 3.2.8.2). The left lanes show a 1 kb ladder. The sizes (bp) of some marker-fragments are indicated. The left and right gels show PCR products for *Jden*1381 and pET32b, respectively. The expected PCR product sizes are 1956 bp for *Jden*1381 and 5955 bp for pET32b.

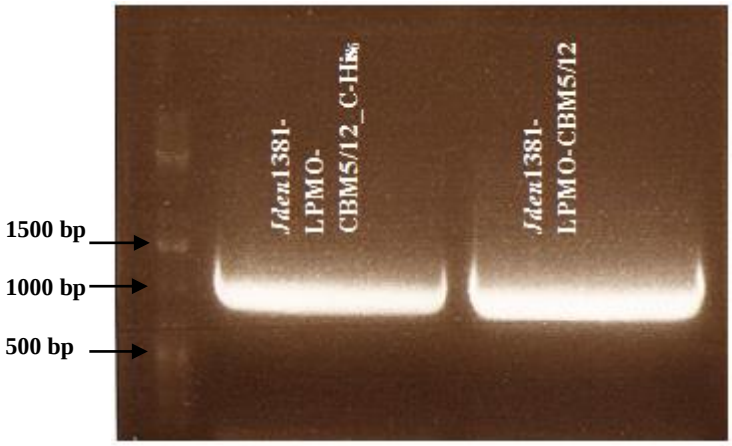


Fig 4.11. Agarose (1 %) gel electrophoresis PCR-generated fragments for generation of *Jden*1381-LPMO-CBM5/12. The left lane shows a 1 Kb DNA ladder and the size (bp) of some marker fragments is indicated. The gel shows two fragments, one with (left) and one without (right) C- terminal His tag, labeled *Jden*1381-LPMO-CBM5/12_C-His_6 and *Jden*1381-LPMO-CBM5/12, respectively. The expected sizes for these fragments are 831 bp and 849 bp respectively.

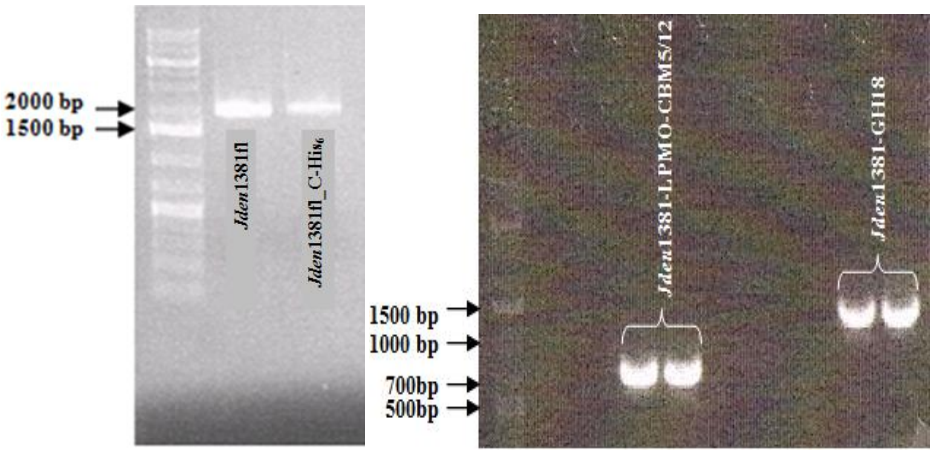


Fig. 4.12. Agarose (1 %) electrophoresis of products from colony PCR. The pictures show PCR fragments genereated according section 3.2.3.4. directly fom selected *E.coli* transformants putatively containing expression constructs. The left lanes show a 1 kb ladder and the sizes (bp) of some marker-fragments are indicated. The left gel demonstrates successful construction of the full length *Jden*1381 (size = 1956 bp; pUCBB/*Jden*1381) and full length *Jden*1381 and C-terminally His tagged full length *Jden*1381 (size = 1974 bp; pUCBB/*Jden*1381_C-His_6). The right gel demonstrates construction of truncated versions: the left two lanes show two PCR amplifications of the insert in *Jden*1381-LPMO-CBM5/12 (expected size = 831bp) and the right two lanes show PCR amplifications of *Jden*1381-CBM5/12-GH18 (expected size =1131 bp).

4.3 Protein expression and purification

4.3.1 Protein expression of CelS2-N mutants

All constructs encoding CelS2-N mutants (R212A, D214A, S215A, E217A, F219A and F219Y) were transformed into chemically competent *E. coli* BL21 (DE3) for expression. SDS PAGE analysis of periplasmic extracts revealed that *E. coli* BL21 (DE3) expressing the CelS2-N mutants: R212A, S215A, F219A and F219Y produced considerable amounts of protein, and that a substantial fraction was soluble (Fig. 4.13a, & 4.15). Furthermore, $CelS2_{D214A}$ and $CelS2_{E217A}$, a product with a ~5 kDa larger size than expected could be observed in cell pellets fractions, which most likely is the recombinant protein with an non-cleaved signal peptide (Fig. 4.13b). The SDS analysis for D214A and E217A (Fig 4.13b) revealed that these mutants were expressed as inclusion bodies in the cytoplasm. For comparison, the CelS2-N wild type was expressed following the same standard procedure and analyzed on SDS PAGE gel. Comparisons showed that all expressed soluble mutant proteins had similar production yields and migration distances as the wild type. As an example, the comparison between periplasmic extracts from a strain producing CelS2-N wild type and a strain producing $CelS2_{R212A}$ is shown in figure 4.13a.

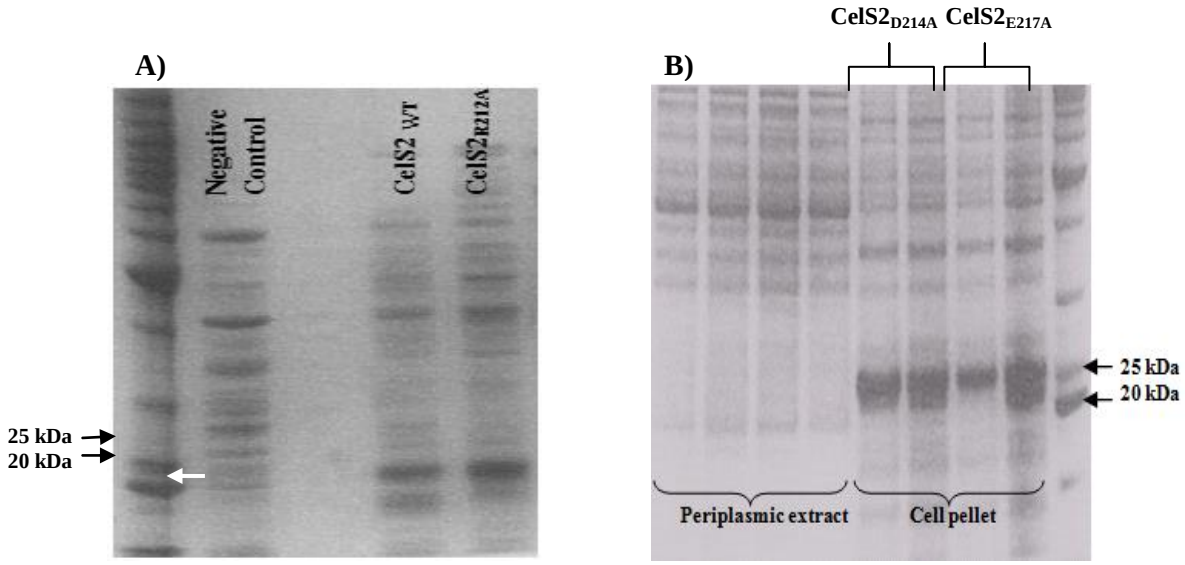


Fig 4.13. SDS-PAGE gels illustrating production of CelS2 variants. A): Coomassie stained SDS-PAGE gel with a protein standard (left lane), a negative control [periplasmic extract of non-transformed BL21 (DE3)] and periplasmic extracts from cells producing $CelS2_{WT}$ and $CelS2_{R212A}$. Some of the markers are labeled by their masses indicated in kDa. The band representing the CelS2-N enzymes is indicated by white arrow. B) Coomassie stained SDS-PAGE gel with a protein standard (right lane), The left four lanes show periplasmic extracts from BL21(DE3) transformed with pRSET-B/$CelS2_{D214A}$ (the first two lanes)and pRSET-B/$CelS2_{E217A}$ (the next two lanes). The right four lanes show proteins in the cell pellets obtained when preparing these periplasmic extracts for $CelS2_{D214A}$ (the first two lanes) and $CelS2_{E217A}$ (the next two lanes) The gel picture demonstrates that these two mutants expressed in enormous amount and aggregated as insolubel proteins. The expected size for unprocessed CelS2 is 25 kDa; the expected size for correctly processed CelS2 is 21.9kDa.

4.3.2 Expression of Jden1381

All thirteen plasmid constructs containing the *Jden*1381 full length and truncated versions were successfully transformed into *E.coli* BL21 (DE3) cells for expression (see Table 3.5 or 4.2 for detail). For one of the constructs, pET32b/*Jden*1381fl, the protein was expected to be expressed in the cytoplasm and protein production was analyzed in cell-free extracts and whole cell lysate (section 3.4.2). For twelve of the constructs, proteins were expected to be exported to the periplasm and protein production was analyzed in the same way as for the CelS2-N mutants.

The result, illustrated for pET32b/*Jden*1381fl in Fig. 4.14, showed that soluble intracellular protein of the correct size was obtained for Jden1381fl. Furthermore, soluble protein with the correct size was observed in periplasmic extracts for these variants: Jden1381fl, Jden1381-LPMO, Jden1381-LPMO_C-His_6, Jden1381-LPMO-CBM5/12 and Jden1381-LPMO-CBM5/12_C-His_6 (Fig. 4.15 & 4.17). For the following variants no protein was obtained:

Jden1381fl_C-His_6, Jden1381-CBM5/12, Jden1381-CBM5/12_C-His_6, Jden1381-CBM5/12-GH18, Jden1381-cbm5/12-GH18_C-His_6, Jden1381-GH18 and Jden1381-GH18_C-His_6. This is summarized in Table 4.2. As illustrated by Fig 4.14, IPTG induction was not always needed to obtain reasonable protein amounts (compare lanes 3 and 4 in Fig, 4.14). Furthermore, for pET32b/*Jden*1381fl, expressing longer Jden1381 variants, we observed proteolytic processing of the protein as demonstrated by the presence of truncated protein forms in the abstracts (see Fig. 4.14).

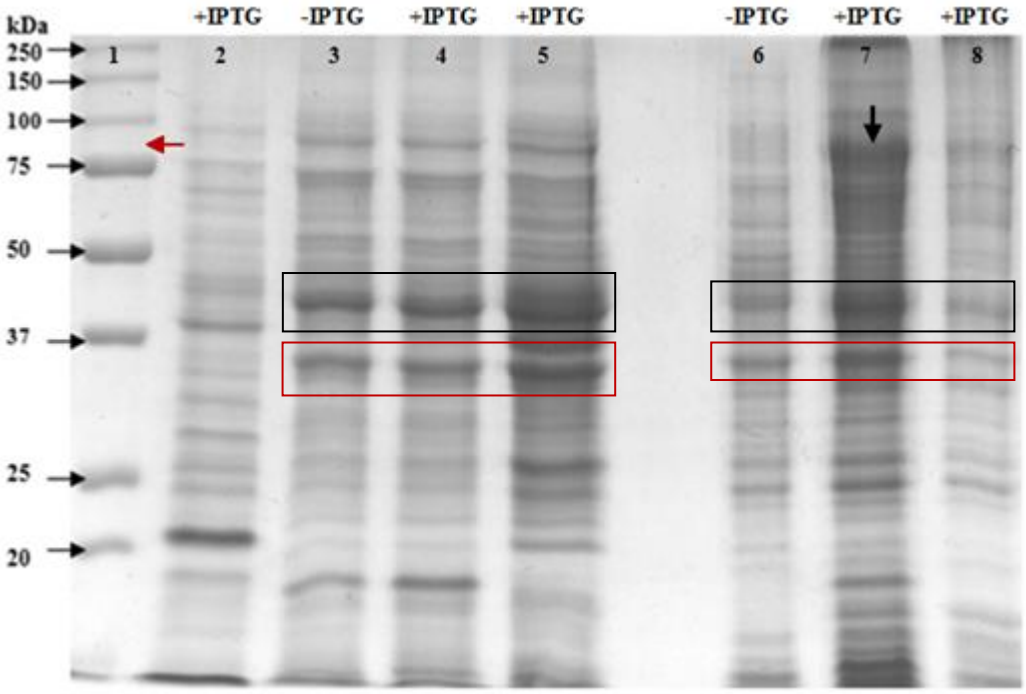


Fig 4.14. SDS-PAGE analysis of expression of Jden1381fl from pET32b/*Jden*1381. Cell-free extracts obtained from cultures with and without IPTG induction for construct pET32b/*Jden*1381. The size of the full length protein (including the Trx, 6xHis and Enterokinase tags) is approximately 83 kDa (indicated by a red arrow). Lanes 2-5 show cell free extracts, lanes 6-8 show whole cell lysates. Lane1, protein standard with masses indicated in kDa. Lane 2, pET32b/empty induced with 0.4 mM IPTG. Lane 3, pET32-b/*Jden*1381 without IPTG induction. Lane 4, pET-32b/ *Jden*1381induced with 0.4 mM IPTG; cell free extract made three hours after induction (expected protein size 83 kDa). Lane 5, pET32b/*Jden*1381induced with 0.4 mM IPTG; cell free extracted made 4 hours after induction. Lane 6, Cell lysate from culture harboring pET32-b/*Jden*1381 without IPTG induction. Lane 7, cell lysate from culture harboring pET-32b/ *Jden*1381 induced with 0.4 mM IPTG; cell free extract made 3 hours after induction. Lane 8, cell lysate from culture harboring pET32b/*Jden*1381induced with 0.4 mM IPTG; cell free extract made 4 hours after induction. Note that all lanes, apart from showing a faint band possible corresponding to the full-length protein (red arrow) show at least two more prominent bands that are absent in the negative control, including bands at approximately 33kDa (marked with red rectangle) and 43kDa (marked with black rectangle). The sizes of these bands match the calculated molecular weights of N-terminally truncated variants of the full length tagged protein, containing only the LPMO domain (33 kDa) or the LPMO and the CBM5/12 domain (43 kDa). In addition, lane 7 reveals a prominent band (indicated by arrow) corresponding the expected size of the full length protein indicating a possible aggregation of the protein in the cytoplasm.

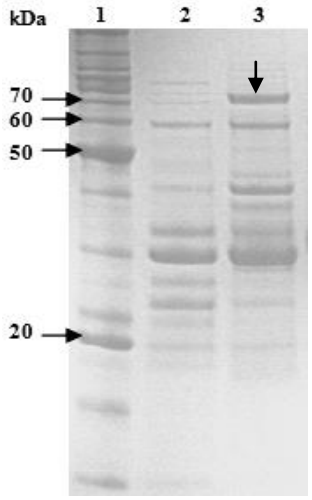


Fig 4.15. SDS-PAGE analysis of expression of Jden1381fl from pUCBB/*Jden*1381fl. The expected size of Jden1381fl is 66 kDa. Lane 1, protein standard with masses indicated in kDa. The SDS-gel shows periplasmic extract from cultures for constructs pUCBB/empty (Lane 2) and pUCBB/*Jden*1381fl (Lane 3). Lane 3 shows a prominent band of approximate size of 70 kDa (indicated by arrow). The size of this band matches the expected size of unprocessed Jden1381fl (69kDa)

4.3.3 Protein purification

Both the CelS2-N mutants (R212A, S215A, F219A and F219Y) and the non-His tagged *Jden*1381 variants (Jden1381-LPMO and Jden1381-LPMO-CBM5/12) were purified to approximately 90% purity using a two step purification procedure comprising an ion exchange (DEAE) (section 3.5.1) and a size exclusion step (section 3.5.2). Example of chromatograms illustrating these two types of purification steps is provided in Fig. 4.18 (showing purification steps for $CelS2_{R212A}$). SDS-PAGE gels illustrating these purification steps are shown in Fig 4.16a-b, respectively. The C-terminally His- tagged variants Jden1381- LPMO_C-His_6 and Jden1381-LPMO-CBM5/12_C-His_6 were purified to approximately 97 % purity by His-Trap affinity chromatography (section 3.5.3). SDS-PAGE gels showing the final purified proteins used for further work are shown in figures. 4.16b-d & 4.17. Chromatograms of ion exchange, size exclusion and His-Trap affinity chromatography obtained from purification steps of Jden1381-LPMO, Jden1381-LPMO_C-His_6 and Jden1381-LPMO-CBM5/12 are shown in Appendix J; DEAE, (Fig. J-1 & J-2), Size exclusion chromatography (Fig. J-3) and His-Trap affinity chromatography (Fig. I-4)

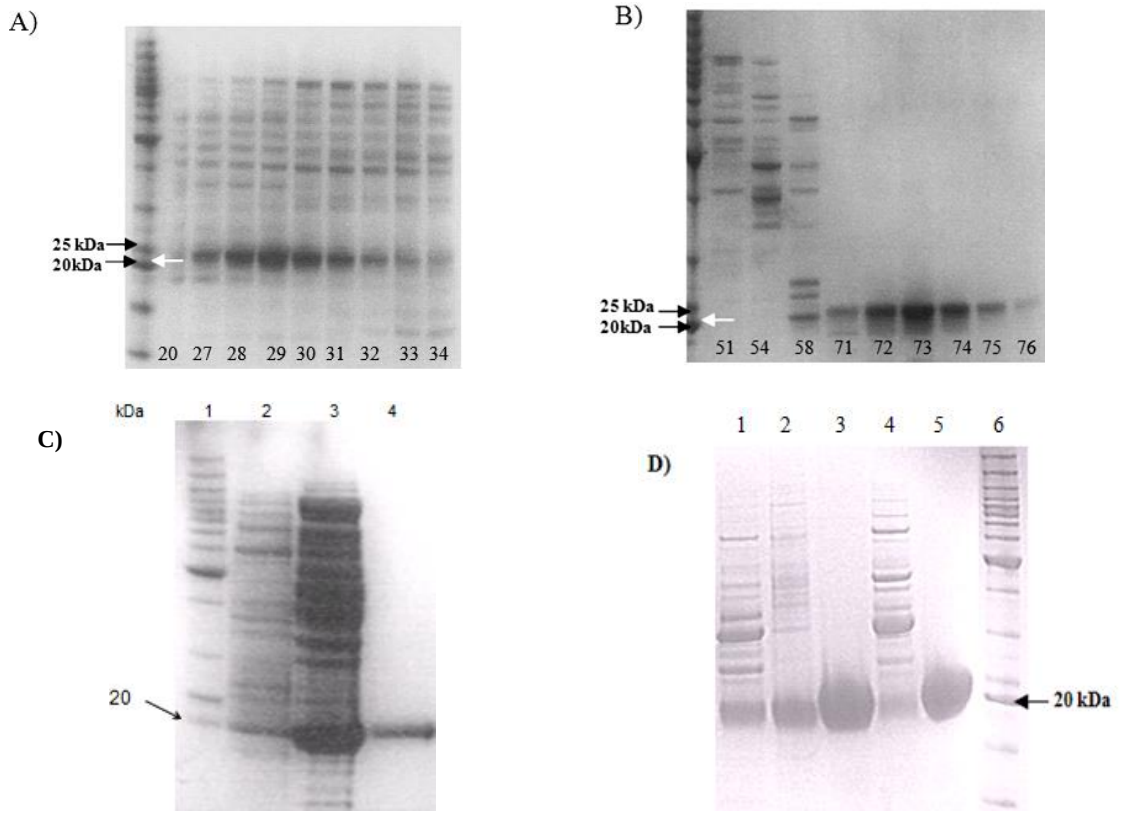


Fig 4.16. SDS-PAGE analysis of protein purification steps of CelS2 mutants. The left lanes of all (except for D) SDS gels contain protein standard with masses indicated in kDa and the approximate protein sizes are indicated by arrows. The expected size for CelS2-N is 21.9kDa. **A**) Coomassie stained SDS gel showing fractions from DEAE purification step of $CelS2_{R212A}$ mutant protein. Lanes containing fractions of the purification step are labeled with fraction number. Fractions 27-34 contain considerable amount of $CelS2_{R212A}$ thus collected for further purification step. **B**) Protein fractions from size exclusion chromatography of $CelS2_{R212A}$. The fraction numbers are labeled. The SDS gel reveals that fraction 71-76 contain pure $CelS2_{R212A}$ thus collected for further use. Chromatograms from DEAE and size exclusion chromatography of $CelS2_{R212A}$ are shown in Fig. 4.18. **C**) Purification steps for $CelS2_{S215A}$ mutant Lane 2, DEAE purified $CelS2_{S215A}$ Lane 3, periplasmic extracts from BL21(DE3) transformed with pRSET-B/*CelS2*$_{S215A}$. Lane 4, $CelS2_{S215A}$ purified with size exclusion chromatography. **D**) Purification steps for $CelS2_{F219A}$ and $CelS2_{F219Y}$. Lane1, Periplasmic extract from BL21(DE3) transformed with pRSETB/*CelS2*$_{F219A}$. Lane 2, DEAE purified $CelS2_{F219A}$. Lane 3, SEC purified $CelS2_{F219A}$. Lane 4, Periplasmic extract from pRSETB/*CelS2*$_{F219Y}$. Lane 5, SEC purified $CelS2_{F219Y}$. Lane 6, protein standard with masses indicated in kDa.

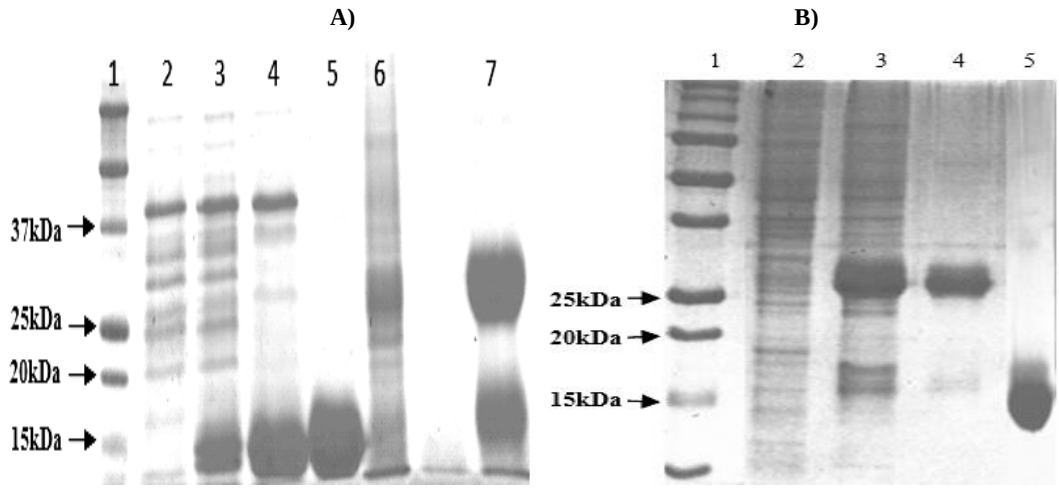


Fig 4.17. SDS-PAGE analysis of protein purification steps of Jden1381 variants. The left lanes of all (except for D) SDS gels contain protein standard with masses indicated in kDa. A) SDS gel analysis for step wise purification steps for Jden1381-LPMO (15 kDa) and Jden1381-LPMO-CBM5/12_C-His_6 (27 kDa). Lane 2, periplasmic extract from cultures for constructs pUCBB/empty. Lane 3, periplasmic extract from cultures for constructs pUCBB/*Jden*1381-LMPO. Lane 4, Jden1381-LPMO purified with anion exchange chromatography (DEAE). Lane 5) Jden1381-LPMO purified with size exclusion chromatography. Lane 6: periplasmic extract from cultures for construct pUCBB/Jden1381-LPMO-CBM5/12_C-His_6 (DE3). Lane 7: Jden1381-LPMO-CBM5/12_C-His_6 purified with HisTrap affinity chromatography. Note that a prominent band of approximate size of 15 kDa is observed, suggesting the separation of a certain amount of the LPMO and CBM5/12 domains. B) SDS gel image revealing purified Jden1381-LPMO-CBM5/12 and Jden1381-LPMO_C-His_6. Lane 2: Periplasmic extract from cultures for pUCBB/Jden1381-LPMO-CBM5/12. Lane 3, DEAE purified Jden1381-LPMO-CBM5/12. Lane 4: SEC purified Jden1381-LPMO-CBM5/12. Lane 5: HisTrap affinity chromatography purified Jden1381-LPMO_C-His_6. Examples of chromatograms for purification steps of Jden1381-LPMO, Jden1381-LPMO-CBM5/12 and Jden1381-LPMO-CBM5/12_C-His_6 are shown in Appendix J.

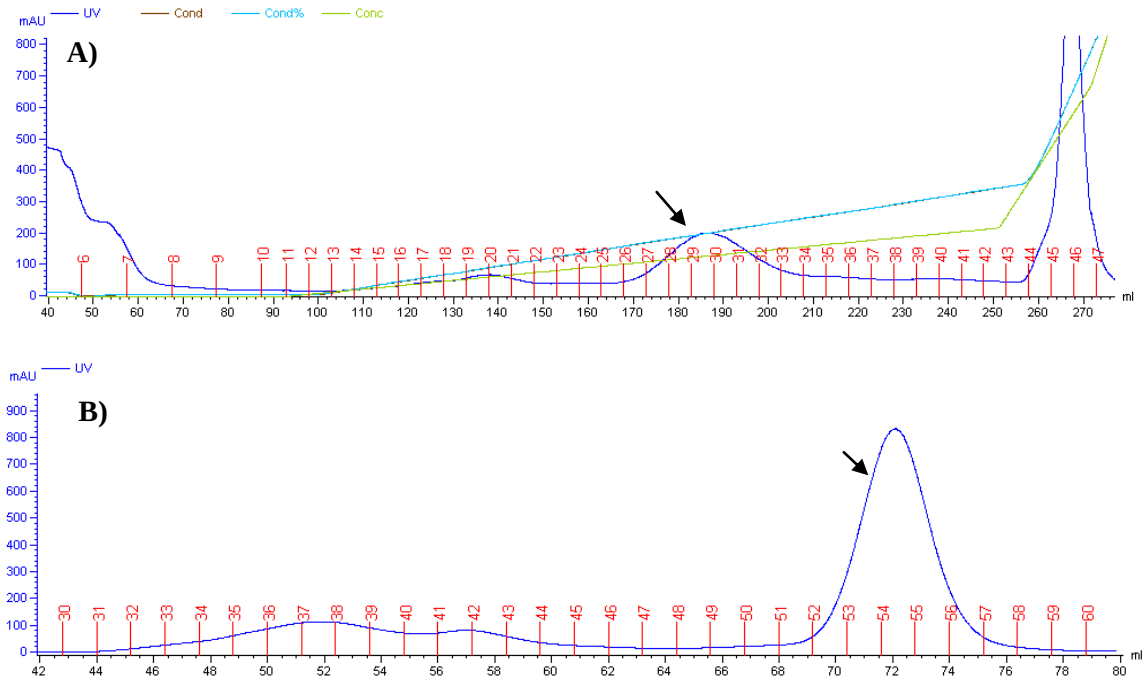


Fig 4.18. Chromatogram for ion exchange (A) and size exclusion (B) of CelS2$_{R212A}$. Fraction numbers are shown in red. The x-axes show volume (ml) of buffer passed through the column. The Y-axes show measurement of UV absorbance at 280nm (blue). The chromatogram show eluted proteins detected by online monitoring of absorption at 280 nm and collected in 5 ml fractions. In panel A; light blue line, measurement of conductivity used to follow column equilibration and salt gradient formation; green line, concentration measurement of elution buffer. Peaks containing CelS2$_{R212A}$ are indicated by arrows.

Protein concentrations were measured using the Bradford protein assay as described in section 3.6.2. Protein concentrations in the samples of purified protein were typically 1 to 5 mg/ml. When necessary, protein solutions were concentrated using Amicon® Ultra centrifugal filter device, 10,000 MWCO (Millipore) as described in section 3.6.1. Purified proteins were stored at 4 ^{o}C. To get an impression of yields, Table 4.3 shows an overview over of the yield of purified protein expressed in mg per liter culture.

Table 4.3. Overview of protein yields expressed as mg purified protein per liter culture medium. The other protein variants targeted in this study were not purified because they were not produced in sufficient amounts (Jden1381fl, Jden1381fl_C-$_{His6}$, Jden1381-CBM5/12, Jden1381-CBM5/12_C-His_6, Jden1381-CBM5/12-GH18, Jden1381-CBM5/12-GH18_C-His_6, Jden1381-GH18 and Jden1381-GH18_C-His_6) or they were produced as insoluble protein ($CelS2_{D214A}$ and $CelS2_{E217A}$).

Protein	Yield (mg/L)
$CelS2_{R212A}$	5.2
$CelS2_{S215A}$	1.3
$CelS2_{F219A}$	5.0
$CelS2_{F219Y}$	3.2
Jden1381-LPMO	1.3
Jden1381-LPMO_C-His_6	1.0
Jden1381-LPMO-CBM5/12	0.9
Jden1381-LPMO-CBM5/12_C-His_6	1.2

4.4 Enzyme characterization - CelS2-N

Table 4.4. Concentration of enzyme and substrate in reaction set up to screen the effect of CelS2 mutations

Substrate	Enzyme concentration						
	Cel5A	CelS2-WT	CelS2-R212A	CelS2-S215A	CelS2-F219A	CelS2-F219Y	CelS2-H144A
Avicel 5 mg/ml	-	1µM	1µM	1µM	-	-	1µM
Pasc 2mg/ml		1µM	1µM	1µM	1µM	1µM	1µM
	8µM	-	-	-	-	-	-
	16µM	-	-	-	-	-	-

4.4.1 Choice of substrate

For determination of the optimal substrate for CelS2-N, the wild type enzyme activity was evaluated on both Avicel-PH101 and phosphoric acid swollen cellulose (PASC). For evaluation of substrate preferences, the amounts of soluble sugar products released from both substrates were compared using HPAEC for product analysis and quantification as described in section 3.8.2.2. The results (Fig. 4.19) indicated that CelS2-N is more active on PASC compared to Avicel, hence PASC was chosen as substrate for determining the activity of CelS2 wild type and CelS2-N mutants.

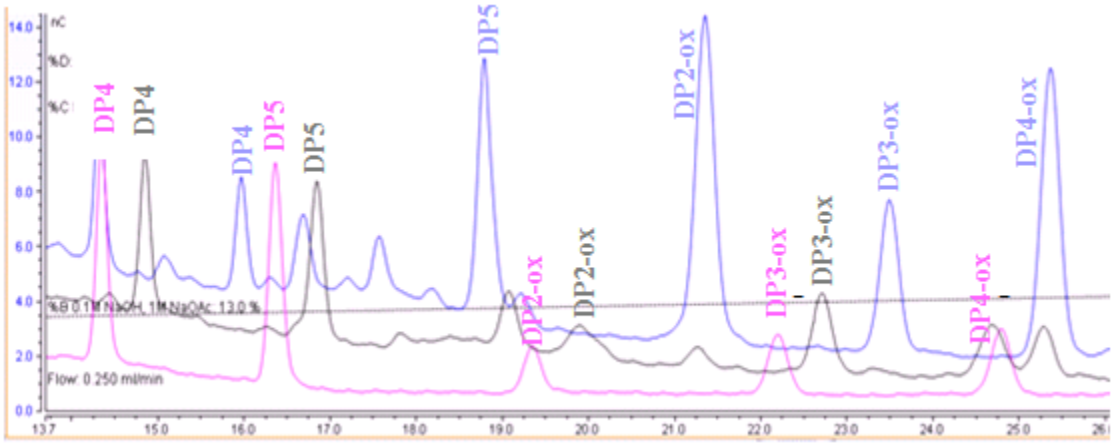


Fig 4.19. Catalytic activity of CelS2 on Avicel and PASC. Soluble products were analyzed with high pressure anion exchange chromatography (HPAEC), using a mix of native (DP4-DP5) and oxidized (DP2ox-DP4ox) cello-oligosaccharides as standards (pink chromatogram). Here the standards were used for qualitative analysis hence varied amount of standards was used. The figure shows chromatograms representing products generated from the action of $CelS2_{WT}$ on Avicel-PH101 (grey line) and PASC (purple line). The reactions (200 µl) were prepared by mixing 1µM CelS2 with 5mg/ml Avicel-PH101 or 2mg/ml PASC in 20 mM BisTris buffer pH 6.5 (at 50 ^{o}C) with 1mM Ascorbic acid. The reactions were incubated 50 ^{o}C over night with shaking at 990 rpm.

4.4.2 Mapping of enzymatic activity of CelS2- R212A, S215A, F219A and F219Y mutants by HPAEC

In order to investigate the effect of the mutations, CelS2 mutants (including an inactive H144A mutant made and purified by Z. Forsberg) were screened for activity on PASC and compared with the wild type (Fig. 4.20). These assays indicated that the S215A and the F219A mutants are active, but less active than the wild type, whereas the R212A and F219Y mutants appeared to be inactive as they did not release oxidized cellooligosaccharides from the substrate (similar to $CelS2_{H144A}$). Tables 4.5 & 4.6 show estimates of the relative amounts of products, which indicate that the activities of S215A and F219A amount to approximately 50 % and 15 % of the wild type activity, respectively. The products generated from both wild type and active mutants were identified as native and oxidized products by comparison of their retention times to those of the corresponding available standards, Glc4, Glc5, GlcGlc-ox, Glc_2Glc-ox, Glc_3Glc-ox and Glc_4Glc-ox.

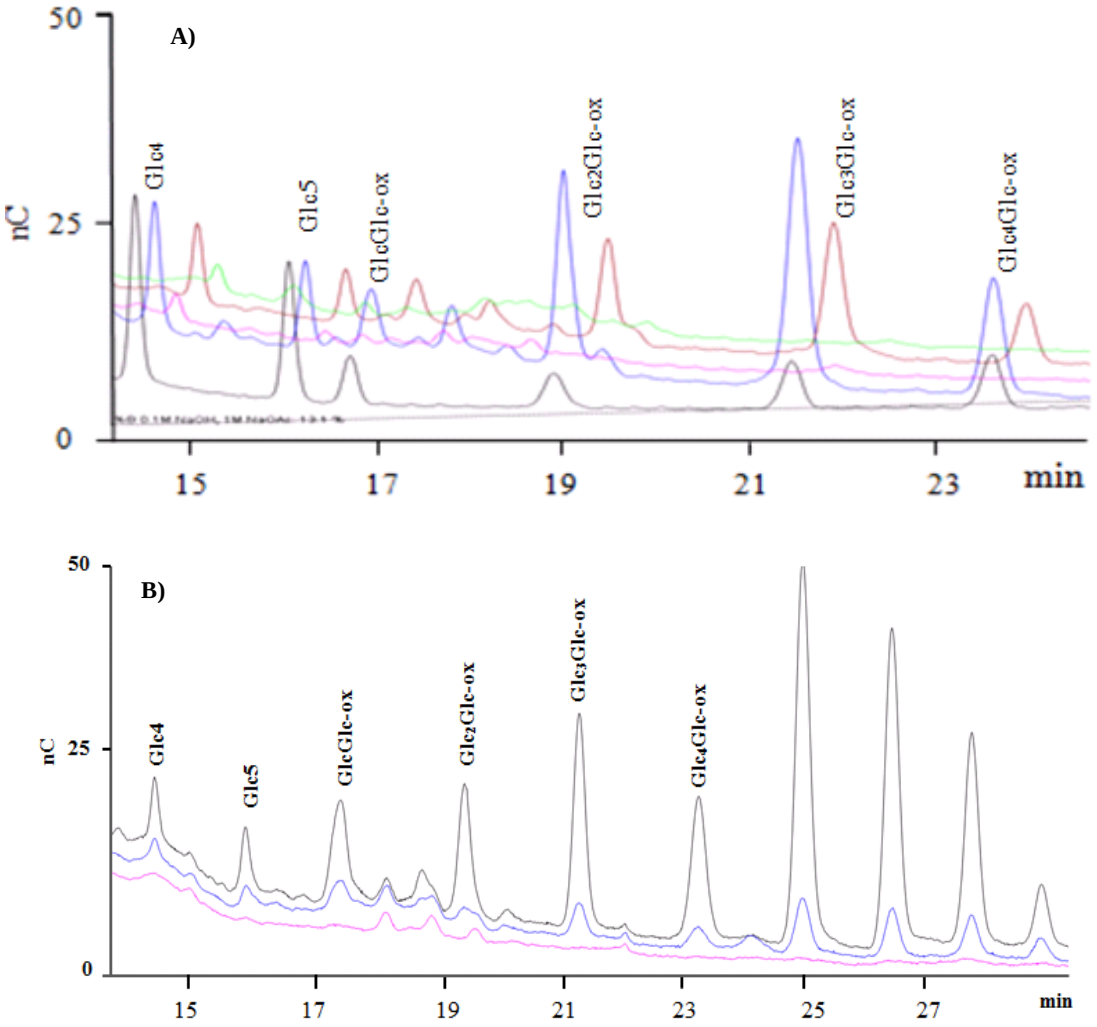


Fig 4.20. HPAEC analysis of products generated by CelS2-wild type and CelS2 mutants. . The reactions (200 µl) were prepared by mixing 1µM enzyme with 2mg/ml PASC in 20 mM BisTris buffer pH 6.5 (at 50 ^{o}C) with 1mM ascorbic acid. The reactions were incubated at 50 ^{o}C for 16 h with shaking at 990 rpm. Chromatograms representing the different enzymes have different colors. A mixture known oligomers (varied amount), containing native (Glc4 and Glc5) and oxidized (GlcGlc-ox - Glc4Glc-ox) sugars was used to identify cello-oligomers produced by the enzymes. Panel A: black, standard mixture; blue, CelS2 $_{WT}$; pink, CelS2$_{R212A}$; brown, CelS2-S215A and green, CelS2$_{H144A}$. Note that active CelS2 variants produce both native and oxidized products. Panel B: grey, CelS2$_{WT}$; blue, CelS2$_{F219A}$; pink CelS2$_{F219Y}$.

Table 4.5. Concentrations of native and oxidized products generated by CelS2$_{WT}$ and CelS2$_{S215A}$, as estimated from HPAEC chromatograms.

Cello-oligomers	Retention time (min)	Peak area from CelS2$_{WT}$/PASC	Peak area from CelS2$_{S215A}$/PASC	Ratio of released sugar; CelS2$_{WT}$:CelS2$_{S215A}$
Glc4	14.3	0.89	0.49	1.8
Glc5	16.01	0.65	0.40	1.6
GlcGlc-ox	16.69	0.96	0.53	1.8
Glc$_2$Glc-ox	18.9	2.21	1.17	1.8
Glc$_3$Glc-ox	21.3	3.37	1.80	1.8
Glc$_4$Glc-ox	23.4	1.4	0.7	2

Table 4.6. HPEAC chromatogram based concentration of native and oxidized products generated by $CelS2_{WT}$ and $CelS2_{F219A}$

Cello-oligomers	Retention time (min)	Peak area from $CelS2_{WT}$/PASC	Peak area from $CelS2_{F219A}$/PASC	Ratio of released sugar from $CelS2_{WT}$:$CelS2_{F219A}$
Glc_3Glc-ox	21.3	1.07	0.16	6.6
Glc_4Glc-ox	23.4	0.88	0.12	7.3

4.4.3 Analysis of initial rates of $CelS2_{WT}$ and $CelS2_{S215A}$

There is currently no good assay for proper determination of the catalytic rates of LPMOs and the activity data presented in the previous section are rough estimates. In attempt to obtain more insight into one of the mutational effects, we used a time series of HPAEC chromatograms (section 3.8.2.2) to compare the initial rates of $CelS2_{WT}$ and $CelS2_{S215A}$ acting on PASC. To do so, we assessed the generation of oxidized reaction products over time, both products released from the substrate into solution and soluble products remaining attached to the substrate. The latter were obtained by spinning down the tubes and removing most of the supernatant. Quantification was simplified by degrading the complex product mixtures (the soluble and the trapped products, separately) to Glc, Glc_2 and Glc-GlcA with a cellulase (Cel5A from *T. fusca)* prior to analysis. Figure 4.21, shows the time courses for production of oxidized chitobiose (Glc-GlcA/ DP2-ox) from PASC. The Figure shows that the data have limited accuracy and that the curves do not look linear. The results do confirm that $CelS2_{S215A}$ yield less products than CelS2 wild-type; at 360 min, it is estimated that wild type CelS2-N produces approximately two-fold more soluble products and four-fold more trapped products compared to $CelS2_{S215A}$ (Fig. 4.18 & Appendix N).

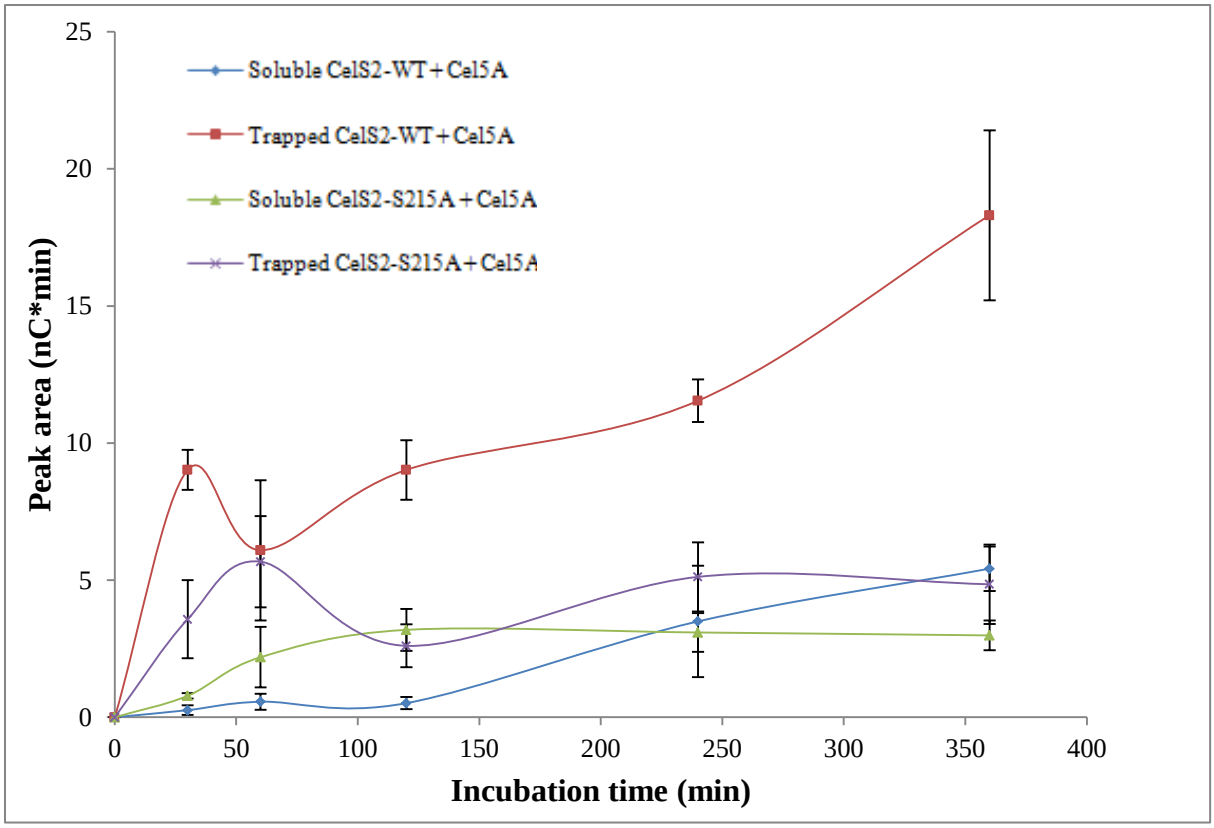


Fig 4.21. Time course for product formation by CelS2$_{WT}$ and CelS2$_{S215A}$. Phosphoric acid swollen cellulose (PASC) was treated with 2 µM CelS2$_{WT}$ or CelS2$_{S215A}$ in a reaction containing 2mg /ml substrate, 1mM Ascorbic acid and 20 mM BisTris buffer (pH 6.5 at 50 ^{o}C). The reactions were incubated at 50 ^{o}C in ThermoMixer while shaking at 900rpm. Both products released from the substrate into solution and soluble products remaining attached to the substrate taken at various time points were further degraded with Cel5A as described in section 3.8.2.2. The figure shows the peak areas for the peak representing oxidized chitobiose (see the text above for more details).

4.5 Enzyme characterization - Jden1381

Table 4.7. Concentration of enzyme and substrate in reactions set up to analyze the activity of Jden1381 variants on chitin.

Substrate	Enzyme concentration					
	Jden1381 (periplasmic extract)	Jden1381-LPMO (purified)	Jden1381-LPMO_C-His$_6$ (purified)	Jden1381-LPMO-CBM5/12 (purified)	Jden1381-LPMO-CBM5/12_C-His$_6$ (purified)	CBP21 (purified)
α-chitin 5mg/ml	10 µl of periplasmic extract	1 µM	-	1µM	-	-
β-chitin 5mg/ml	10 µl of crude extract	1µM	1µM	1µM	1µM	1µM
Collidal chitin 5mg/ml	10 µl of periplasmic extract	1µM	-	1µM	-	-

4.5.1 Analysis of chitooligosaccharides released by full length Jden1381

The activity of full length Jden1381, produced using construct pUCBB/*Jden*1381was tested on α-chitin, β-chitin and colloidal chitin; product formation was analyzed by chromatography using an UHPLC with UV detection, as described in section 3.8.2.1 Reactions were carried out with addition of reductant (ascorbic acid). Adding a reductant was necessary to promote LPMO activity. The activity of Jden1381-LPMO boosted when copper was added, therefore copper was added in further enzymatic assays. The results (Fig. 4.22) showed that Jden1381 produces native chito-oligosaccharides from all chitin variants tested, although product ratios differed. The reaction with α-chitin generated predominantly (GlcNAc)2 and GlcNAc, the reaction with β-chitin gave products ranging from GlcNAc to (GlcNAc)6 and the reaction with colloidal chitin gave products ranging from (GlcNAc)2 to (GlcNAc)6. For control reactions, crude extract from BL21 (DE3) cells containing the pUCBB/empty construct was treated with α-chitin, β-chitin and colloidal chitins and products were analyzed, using the same conditions as Jden1381. All such reactions do not show any detectable product formation. An example of such control reaction (from α-chitin) is included in figure 4.22. Comparison with chromatograms obtained from substrates incubated without enzyme showed that all these products were enzyme-generated (results not shown). Although these data are insufficient to make firm quantitative conclusions about the activity of Jden1381 on the various substrates, they do seem to suggest that full length Jden1381 hydrolyzes α-chitin more efficiently than β- and colloidal chitin.

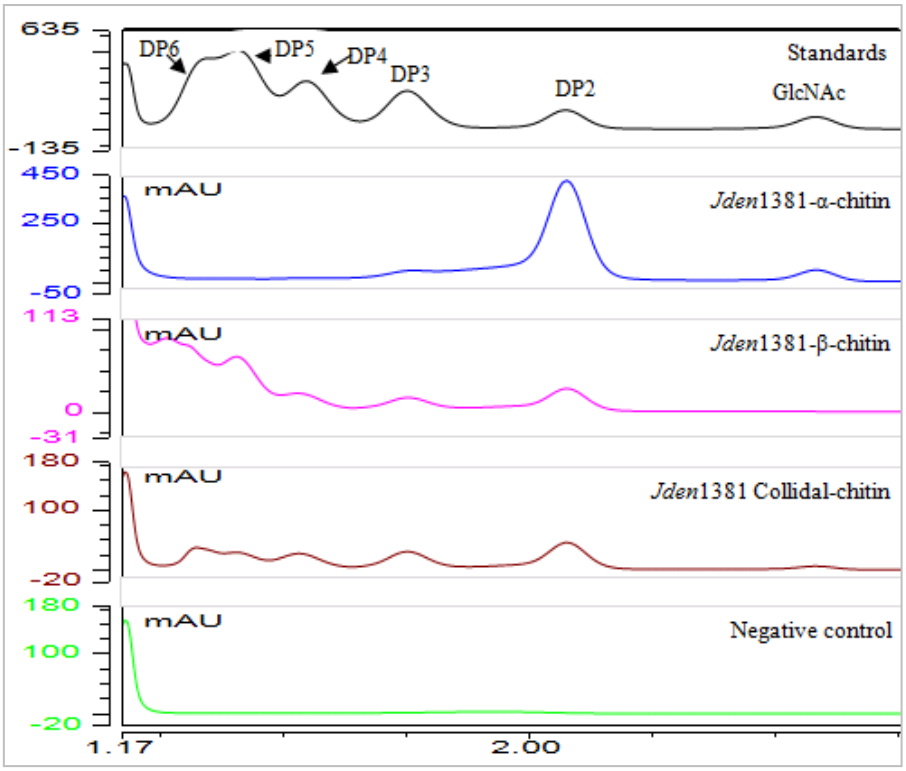
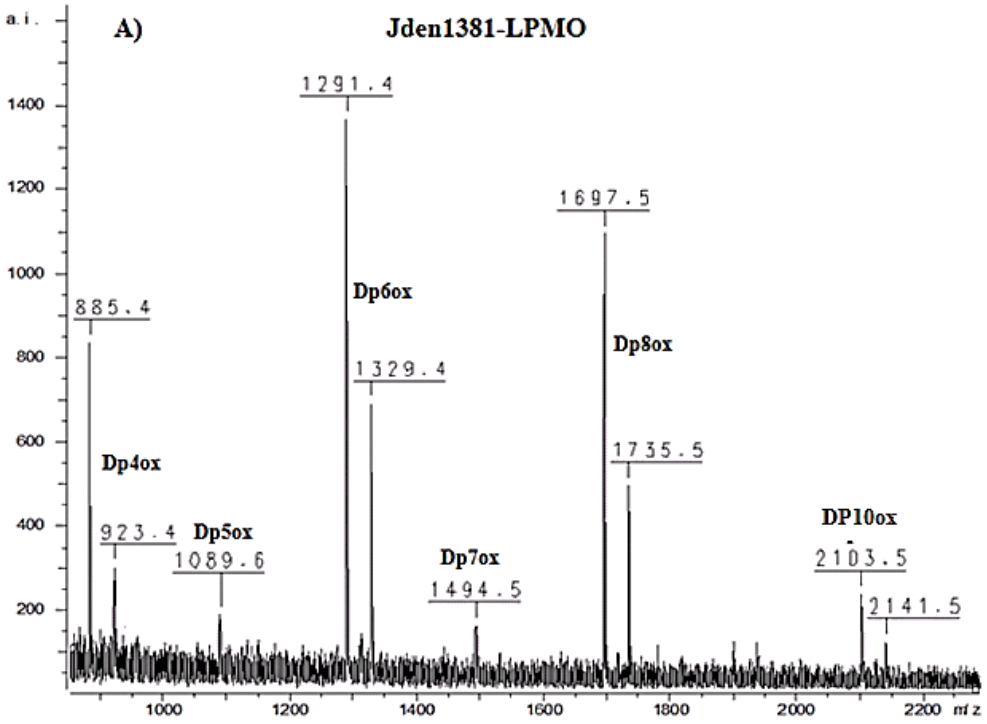


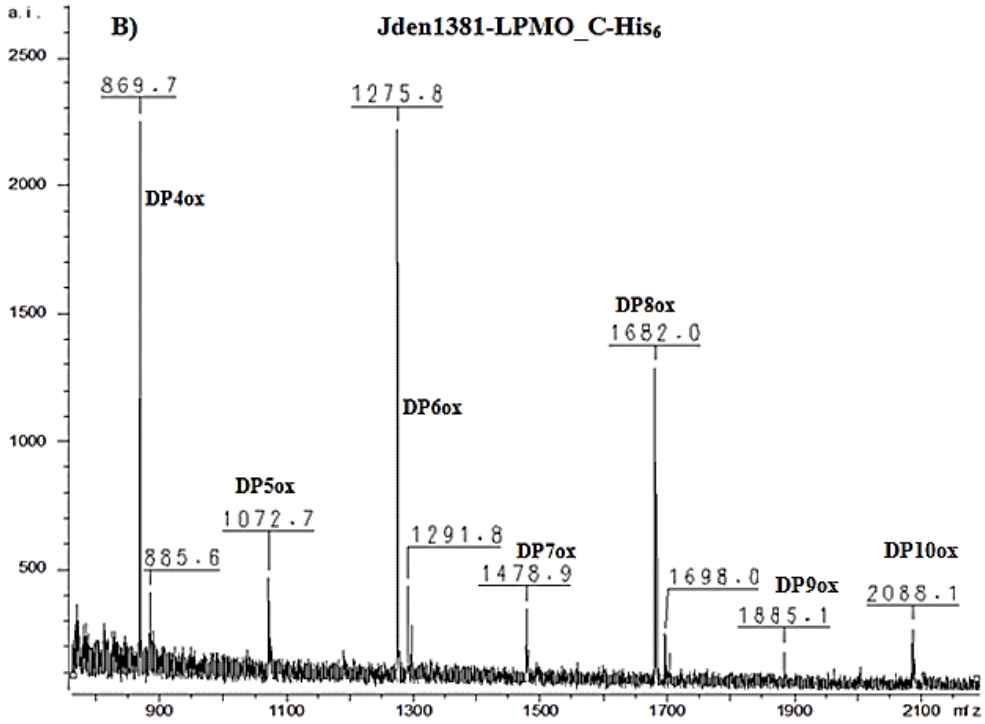
Fig. 4.22. UHPLC analysis of products generated by full length Jden1381 upon incubation with various forms of chitin chitin. Standards (200 µM of GlcNAc & DP2-DP6) are labeled by their degree of polymerization (DP). The figure shows aligned chromatograms with elution time on the X-axis and detector signal on the Y-axis. The chromatograms are labeled with the enzyme and substrate names. A negative control showing the action of a crude extract from BL21 (DE3) cells containing the pUCBB/empty construct on α-chitin is labeled as "negative control".

4.5.2 MALDI-TOF MS analysis of oligosaccharides released by the Jden1381 LPMO domain.

Activities of Jden1381-LPMO, Jden1381-LPMO_C-His_6, Jden1381-LPMO-CBM5/12 and Jden1381-LPMO-CBM5/12 were tested on β-chitin and soluble products were analyzed by MALDI-TOF MS as described in section 3.8.1. For positive control, CBP21was treated with β-chitin under the same condition as Jden1381 variants. These analysis shows that all tested Jden1381-LPMO variants produced oxidized chito-oligosaccharides (Fig. 4.23; Table 4.7). Generally, the mass spectra of products generated by the LPMO-containing Jden1381 variants (Fig. 4.23a-c) looked similar to the mass spectrum generated by CBP21 (Fig. 4.23d). Product generation by Jden1381-LPMO-CBM5/12_C-His_6 was detected on MALDI-TOF MS and did not show any product formation (result not included).

For analysis product type (specifically products length) generated by Jden1381-LPMO variants, β-chitin was treated with Jden1381-LPMO and Jden1381-LPMO-CBM5/12. The reaction setups were prepared as described in section 3.8.1 and the reaction was extended for 72 hours, taking samples at 16 hr and 72hr. The products were analyzed on MALDI-TOF MS. The results for both enzymes show that the maximum size of product generated is an oxidized tetramer, DP4ox (Fig. 4.24). Products from Jden1381-CBM5/12 include at least one glucose amine moiety, GlcN (deacetylated sugar) bearing tetramer (Fig. 4.24, lower panel).


A)
Jden1381-LPMO
a.i.
1291.4
1697.5
885.4
Dp6ox
Dp8ox
1329.4
1735.5
Dp4ox
923.4
Dp5ox
1089.6
Dp7ox
1494.5
DP10ox
2103.5
2141.5
m/z


B)
Jden1381-LPMO_C-His6
a.i.
869.7
1275.8
DP4ox
DP8ox
1682.0
DP6ox
DP5ox
1072.7
1291.8
885.6
DP7ox
1478.9
DP10ox
1698.0
DP9ox
2088.1
1885.1
m/z

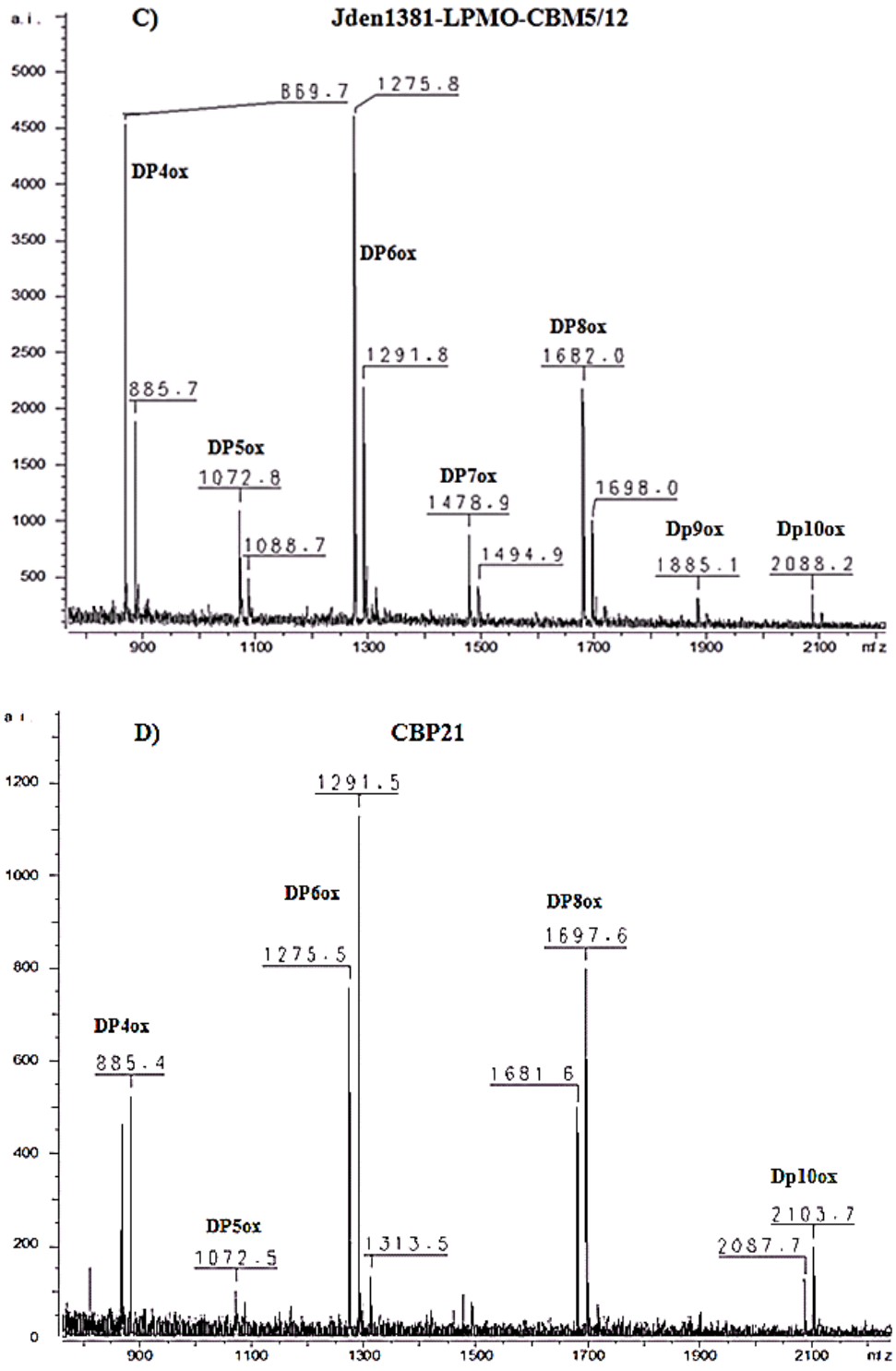


Fig 4.23. MALDI-TOF MS analysis of products release by various Jden1381 variants from β-chitin. Peaks appear in clusters due to the formation of various adducts ($M + Na^+$; $M + K^+$; $M – H + 2K^+$; ($M – H + Na^+ + K^+$); they are labeled by their degree of polymerization, while oxidation is indicated by "ox". Table 4.7 provides a list of masses. A, Jden1381-LPMO; the products are mainly observed as ($M + K^+$) and ($M -H + 2K^+$) adducts; B, Jden1381-LPMO_C-His_6; the products are mainly observed as ($M + K^+$) or ($M + Na^+$). C, Jden1381-LPMO-CBM5/12; the products are mainly observed as ($M + K^+$) or ($M + Na^+$) adducts. D, CBP21 from *Serratia marcescens* (positive control); the products are mainly observed as ($M + K^+$), ($M + Na^+$) and ($M – H + Na^+ + K^+$) adducts.

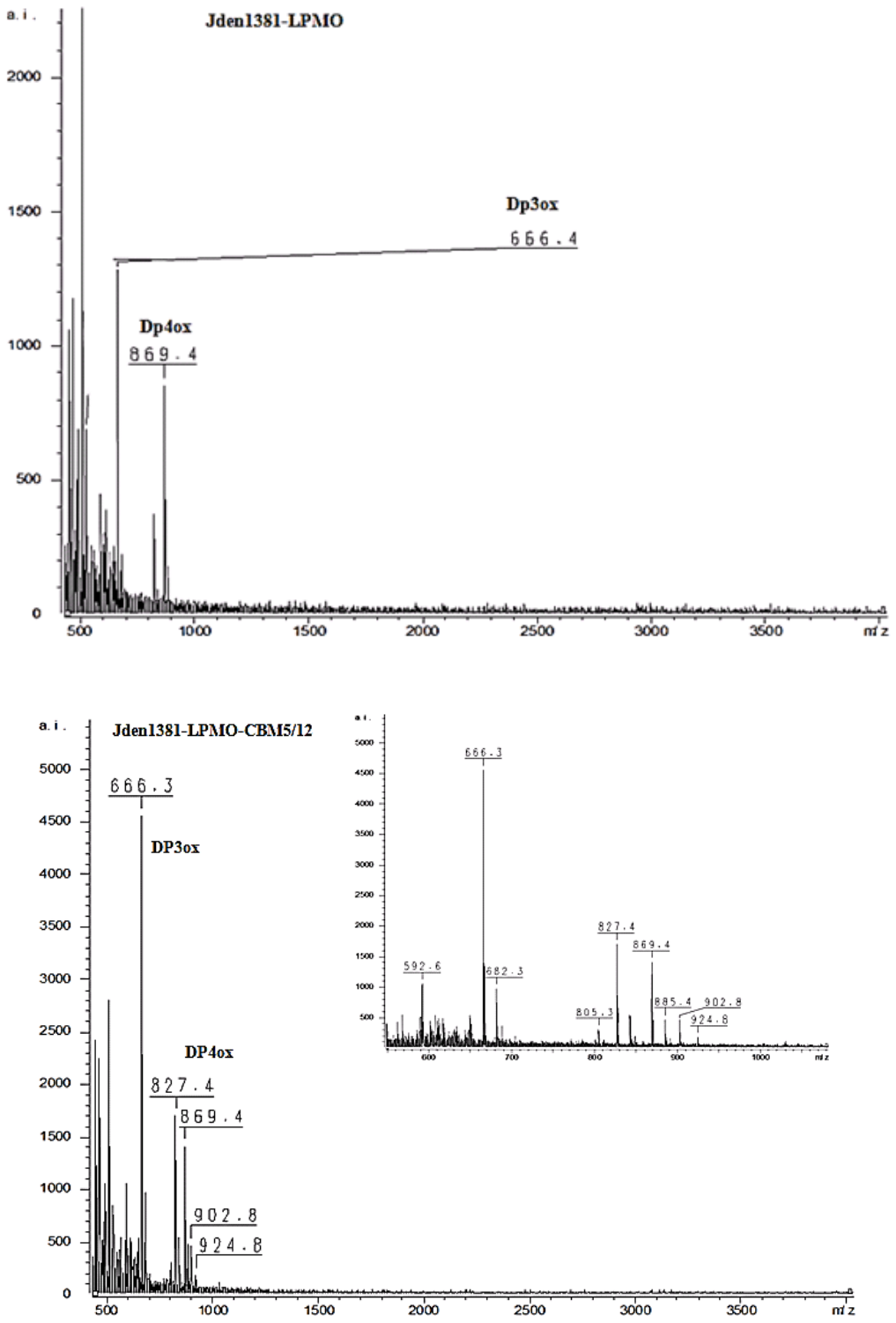


Fig 4.24. MALDI-TOF MS analysis of products release by Jden1381-LPMO and Jden1381-LPMO-CBM5/12 from β-chitin. Peaks appear in clusters due to the formation of various adducts ($M + Na^+$; $M + K^+$; $M –H + 2K^+$); they are labeled by their degree of polymerization, while oxidation is indicated by "ox". Note that only two types of products are observed in both types of Jden1381-LPMO variants. Upper panel, Jden1381-LPMO; the products are mainly observed as ($M + Na^+$) adducts; Lower panel, Jden1381-LPMO-CBM5/12; the products are mainly observed as ($M + Na^+$), ($M + K^+$) and ($M -H + 2K^+$). Note that the lower panel contains a spectra at m/z = 827.4 which may indicate the presence of one deacetylated sugar moiety containing product, DP4ox variant. In the zoom in picture, peaks are labeled with observed atomic mass. Table 4.7 provides a list of masses.

Table 4.7. Overview of m/z values

Products and their adducts	Observed m/z	Theoretical m/z
DP3-ox+ Na	666.3	643.4
DP4-ox+ Na	869.7	846.8
DP4-ox + K	885.4 or 885.7	846.5
DP5-ox + Na	1072.8	1049.9
DP5-ox + K	1088.7	1049.8
DP6-ox + Na	1275.5 or 1275.8	1252.6
DP6-ox + K	1291.5 or 1291.8	1252.6
DP6-ox –H+2K	1329.4	1252.6
DP6-ox –H + K + Na	1313.5	1252.6
DP7-ox + Na	1478.9	1456
DP7-ox + K	1494.5 or 1494.8	1456
DP8-ox + Na	1682.0	1659.1
DP8-ox + K	1698.0	1659.1
Dp8-ox–H + 2K	1735.5	1659.1
DP9-ox + K	1885.1	1846.2
DP10-ox + Na	2087.7 or 2088.2	2064.8
DP10-ox + K	2103.7	2064.8

5 DISCUSSION

Complete degradation of cellulose and chitin requires the synergistic action of glycoside hydrolases (GHs). Today, it is well known that efficient degradation by these enzymes is promoted by LPMOs, which carry out oxidative cleavage of chains in the crystalline regions of the substrate (Horn *et al* 2012). The majority of LPMOs remains uncharacterized and the present study therefore aimed at studying two LPMOs, a cellulose active LPMO from *S. coelicolor* called CelS2 and an LPMO-containing multi-domain protein called Jden1381 from *J. denitrificans*.

CelS2

Forsberg *et al.* (2011) have previously shown that CelS2 cleaves cellulose into mixed population of reducing-end oxidized (aldonic acid bearing end) and native cello-oligomers in a reaction that depend on the presence of a reductant. The reaction mechanism of this oxidative cleavage remains elusive and requires extensive investigation. In order to obtain more insight into catalysis, detailed studies of the roles of individual catalytic site residues are necessary. Accordingly, in this study, attempts were made to assess the significance of conserved surface exposed catalytic site residues for activity. This was done by creating six mutants: R212A, D214A, S215A, E217A, F219A and F219Y.

The six residues were selected for mutation through combination of multiple sequence alignment and prediction of their orientation on the catalytic site using CBP21 as template (section 4.4). Since these residues are surface exposed, it was anticipated that removing solvent reactive side chains would give hints about the function of these residues. Therefore, all of these residues were substituted to alanine.

Judged by its predicted structure, the position of Phe219 of CelS2 is replaced with Tyrosine (Tyr) in many cellulose active LPMOs. This phenomenon is particularly common in GH61 type LPMOs such as, *Thermoscus aurantiacus* (TaGH61A), *Phenerochaete chrysosporium* (pcGH61D), *Hypocrea jecorina* (HjCel61B). Interestingly, Tyr 160, a corresponding residue of Phe219 in CelS2, of pcGH61D interacts with binding metals and it seems that this residue is essential metal coordination (Westereng *et al.,* 2011 and Fig. 5.4). In order to assess the

functional differences between these two residues within these types of LPMOs, F219Y mutant was constructed.

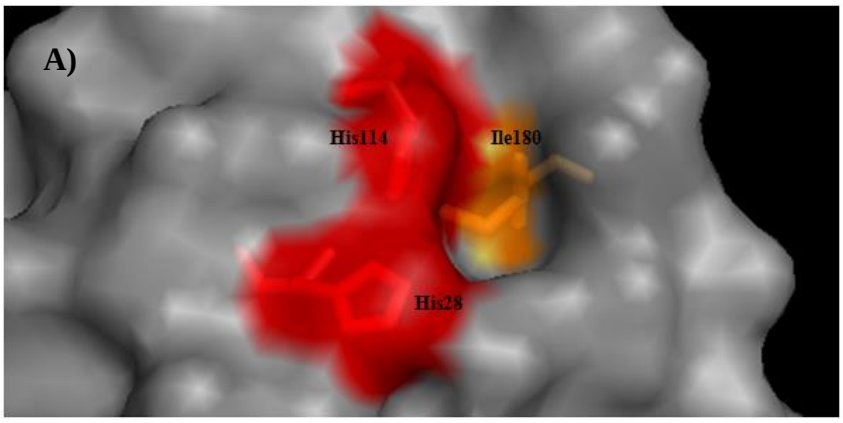
All the six residues selected for mutation were successfully mutated. However, obtaining soluble protein from pRSETB/CelS2$_{D214A}$ and pRSETB/CelS2$_{E217A}$ harboring cells was challenging. Attempts for expression of these mutants resulted in formation of protein aggregates (Fig 4.13). This type of protein aggregate was not observed during expression of either the wild type or the other mutants (R212A, S215A, F219A and F219Y). Formation of protein aggregate in D214A and E217A mutants may be an indication for alteration of protein structure caused by the mutants, which led to formation of misfolded or partially folded proteins. If this is the case, it seems that D214 and E217are essential for protein stability. Interestingly, D182 and N185 of CBP21, the corresponding residues of D214A and E217A of CelS2, have shown minor binding effects when they are substituted by alanine (Vaaje-Kolstad *et al.,* 2005b).

Mutants R212A, S215A, F219A and F219Y could be produced in soluble form and during expression and purification; they showed similar behavior as the wild-type enzyme. Two mutations (R212A and F219Y) led to complete loss of activity, whereas the other two mutations, S215A and F219A, resulted in reduced activity.

The arginine at position 212 was chosen for mutation because it seems to make a major difference between cellulose-active and chitin-active LPMOs. Figure 4.4 show that chitin-active LPMOs, with an Ile at this position (Fig. 5.2a), have a small pocket on their surface that could be adapted to binding an N-acetyl group in chitin. This pocket is expected to be absent if Ile is replaced by Arg. Arginine is large and basic residue and its substitution to a small amino acid like alanine cause changes in the chemical environment of the active site and may also have structural effects. The predicted structure for CelS2 (Fig 4.4) indicates that R212 is situated in close proximity to the metal coordinating site; so, the R212A mutation may also affect metal-binding directly. The model also suggests that several potentially important intramolecular interactions are lost when Arg is replaced by Ala. On the other hand, it is not uncommon for LPMOs to have an Ala in this position, as is the case for e.g. E7 from *Thermobifida fusca* (Fig. 5.1 and Moser *et al*., 2007), which has been shown to be active on both cellulose and chitin (Z. Forsberg, pers. com.). However, mutation of Arg to Ala may have been too drastic for CelS2 and mutation to Ile would in retrospect have been a more conservative choice for assessing the effect of Arg in substrate binding or specificity.

```
E7                1 MHRYSRTGKHRWTVRALAVLFTALLG-----LTQWTAPASAHGSVINPAT     45
                    |.|.:|        :..||.:...|||      |......|.|||..:.|.:
CelS2             1 MVRRTR-------LLTLAAVLATLLGSLGVTLLLGQGRAEAHGVAMMPGS     43

E7               46 RNYGCWL--RWGHDHLNPNMQYEDPMCWQAW-QDNPNAMWNWNGLYRDWV     92
                    |.|.|.|  :.|...|:|.    :|.|..|. |....|::||..:.....
CelS2            44 RTYLCQLDAKTGTGALDPT----NPACQAALDQSGATALYNWFAVLDSNA     89

E7               93 GGNHRAALPDGQLCSGGLTEGGRYRSMDAV-GPWKTTDVNN--TFTIHLY    139
                    ||.....:|||.|||.|......:.:.:|. ..|..|.:.:  |..:...
CelS2            90 GGRGAGYVPDGTLCSAGDRSPYDFSAYNAARSDWPRTHLTSGATIPVEYS    139

E7              140 DQASHGADYFLVYVTKQGFDPTTQPLTWDSLELVHQTGSYPPAQ------    183
                    :.|:|..| |.||:||.|:.||:: |.||.|||: ||.:.||.|
CelS2           140 NWAAHPGD-FRVYLTKPGWSPTSE-LGWDDLELI-QTVTNPPQQGSPGTD    186

E7              184 --NIQFTVHAPN-RSGRHVVFTIWKASHMDQTYYLCSDVNFV........    222
                      :..:.:..|: |||..::|..|..|...:.::.||||.|.
CelS2           187 GGHYYWDLALPSGRSGDALIFMQWVRSDSQENFFSCSDVVFD........    228
```

Fig 5.1. Pairwise alignment of E7 (T. fusca) and Cels2. Signal peptides are highlighted blue. Arg212 in CelS2/Ala206 in E7 are highlighted yellow. Arg212 is thought to eliminate a small pocket in the surface of CelS2, close to the catalytic site (see Fig 5.2).



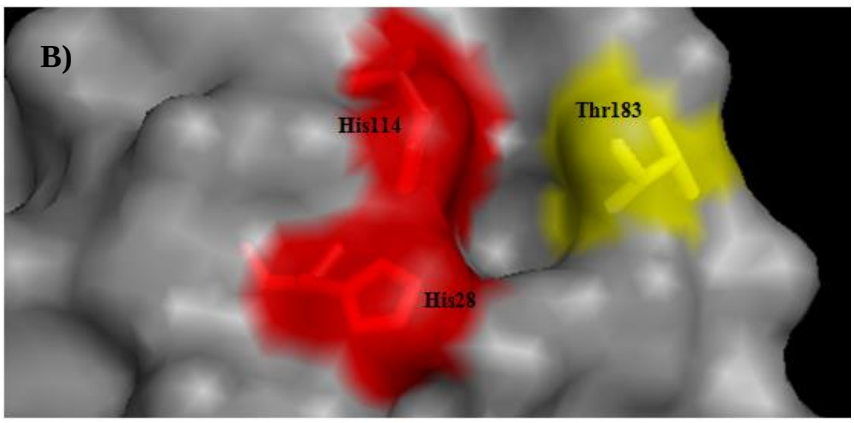


Fig. 5.2. Surface structure of CBP21. The figures show the presence of a small pocket on the surface of CBP21. Ile 180 (A), colored orange and Thr183 (B) colored yellow are two of the residues that make the small pocket. This small pocket is assumed to be adapted to bind the N-acetyl group of chitin. The two conserved histidines are labeled red are labeled. The figure was made by using PyMOL. (DeLano, W. L. *et al*., 2005)

.

Ser215 was chosen for mutation because this residue is a relatively conserved polar side chain close to the active site (Fig. 4.4). The position of Thr183 of CBP21, the corresponding residue of Ser215 of CelS2, is on the small surface pocket that is assumed to have a function in substrate binding (discussed above) (Figure 5.2B). Mutation of Ser215 to Ala only had a modest effect on activity (~ 50 % reduction; Fig. 4.20a & 4.21), suggesting that the hydroxyl group on Ser215 has a role in catalysis, but is not essential. For in depth analysis, the formation of oxidized products generated over time (30-360 minutes) by $CelS2_{S215A}$ and $CelS2_{WT}$ on PASC was assessed (Fig. 4.21). In this analysis both soluble products released to the solvent and those remained trapped in the insoluble fractions were analyzed. The latter were obtained by spinning down the tubes and removing most of the supernatant. For simpler quantification of products, both soluble products (those released to the solvent and remain trapped in the solid fraction) were degraded by endoglucanase Cel5A from *Thermobifida fusca* (section 4.4.3) and the time course of formation of oxidized chitobiose (Glc-GlcA) was analyzed.

The results, which generally showed high standard deviations and uncertainty, showed that the S215A mutant had lower product yields but, perhaps, a slightly higher initial rate. This indicates a possible effect on substrate accessibility (the mutant works better but can access less of the substrate). After 360 minutes, the formation of soluble products released to the solvent by $CelS2_{WT}$ was approximately two fold higher compared with the S215A mutant, while the production of soluble products that were trapped in the insoluble fraction by $CelS2_{WT}$ was approximately four fold higher compared to S215A mutant (Fig 4.21).

The substitution of S215 of CelS2 with Thr183 in CBP21 indicates the importance of having a polar residue at this position for catalytic activities of these LPMOs. Interestingly though, several LPMOs contain hydrophobic residues at this position (Fig 5.3). The CBM33-type putatively cellulose-active LPMO E7 from *Thermobifida fusca* has a Met at this position. The GH61-type fungal LPMO TtGH61E from *Thelavida terrestris* has a Val at this position. This indicates that S215 may not be directly important for activity in CelS2 and that the minor mutational effect observed here is due to (minor) local structural effects that may have an effect on substrate binding.

As described in section 1.7, some cellulose active GH61s generate more than one product type. In addition to, or perhaps instead of, C1 oxidized sugars they may generate C4 or C6 oxidized sugars (Phillips *et al.*, 2011). Sequence and structural comparison between LPMOs demonstrates the presence of conserved hydrophobic residues at the position equivalent to Ser215 in many cellulose active GH61 type LPMOs and in some CBM33-type LPMOs such as E7 (Fig 5.1). Perhaps this residue affects the oxidation pattern.

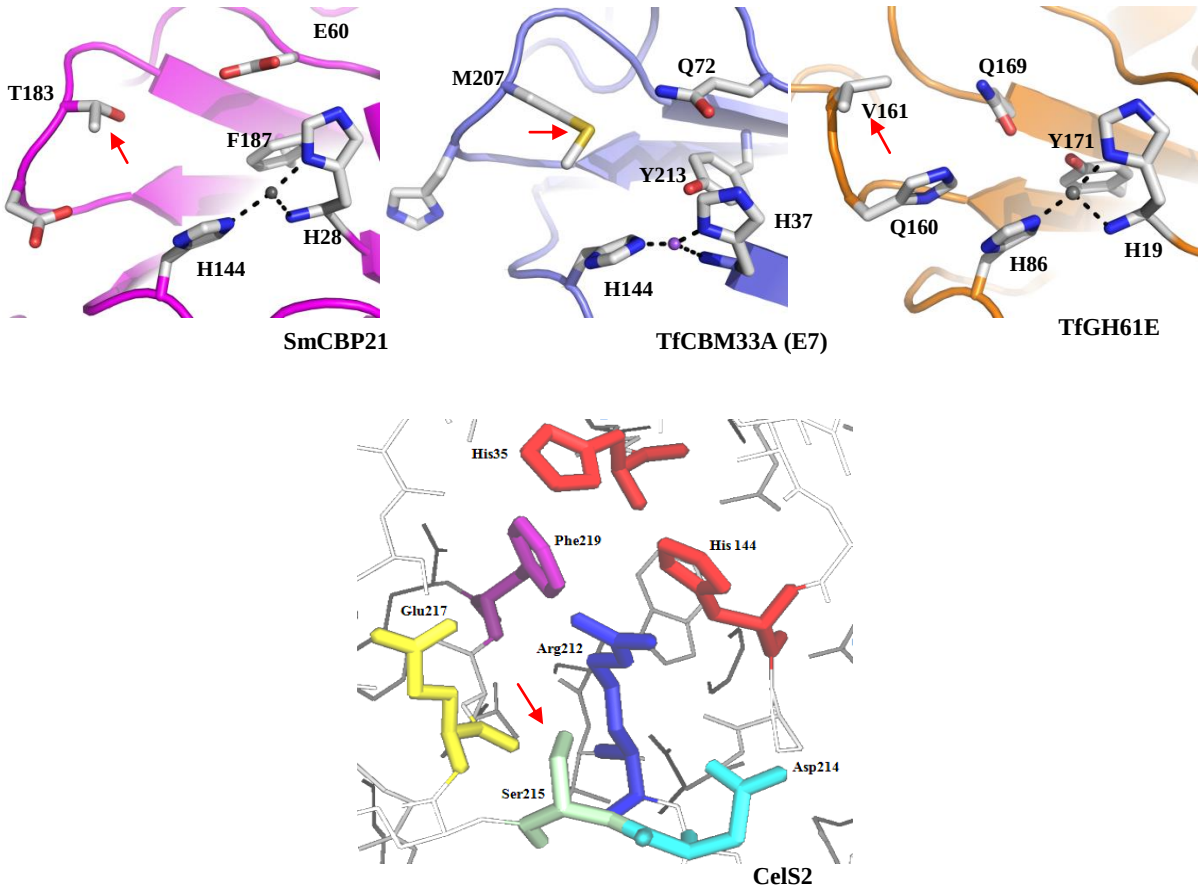


Fig 5.3 Catalytic site comparison between CBM33-type LPMOs from *Serratia marsescens* (CBP21) and *Thermobifida fusca* [*Tf*CBM33A (E7)] and a GH61-type cellulose-active LPMO from *Thelavida terrestris* (*Tt*GH61E). Upper panel: The catalytic sites of CBP21 (A) and TtGH61 (C) are from crystal structures, whereas the catalytic site of TfCBM33A (E7) is from a homology model. Note that *T183 in CBP21 (equivalent to S215 in CelS2)* is replaced by the hydrophobic residues Met and Val in TfCBM33A and TtGH61E, respectively. These residues are highlighted by red arrow. (Source: World Intellectual property Organization; International publication number WO 2012/019151 A1) All figures were made using PyMOL (DeLano, W. L. *et al*., 2005).

The F219A mutation caused partial reduction in enzymatic activity while the F219Y mutation caused complete deactivation. The activity of F219A is reduced to approximately 15% of the activity of $CelS2_{WT}$. It is remarkable that complete removal of the phenyl ring of Phe219 leaves the enzyme active, whereas addition of just a hydroxyl group inactivates the enzyme. It

is not surprising that mutation of Phe219 has effects, since the residue is in close proximity (approximately 4 Å from His144) of the catalytic center (see Fig. 4.4). Looking at the catalytic center one would assume that the F219A mutation has effects beyond not that dramatic reduction in activity. Clearly, a more in-depth enzymatic characterization of this mutant is needed to understand the role of the phenylalanine, which is highly conserved (see below for further discussion).

As mentioned earlier, the F219Y mutation was made because Tyr is observed at this position in several (putatively) cellulose-active LPMOs (see Fig. 5.3; Phe187 in CBP21, which is analogous to Phe219 in CelS2, is replaced by Tyr213 in E7,Tyr171 in TtGH61E and Tyr175 in TaGH61A), where it aids in coordinating the copper ion essential for LPMO activity (Fig. 5.4). The most plausible explanation for the inactivating effect of the F219Y mutation is that the extra hydroxyl group displaces the copper ion instead of aiding in binding the metal as is suggested role for this residue in GH61-type LPMOs (see fig 5.4; Quinlan *et al.*, 2011 and Li *et al.*, 2012). Another explanation for the loss of activity is that the introduction of the hydroxyl group creates steric hindrance at the catalytic site which might cause catalytic site distortion and hinder substrate binding. Alternatively, the addition of a hydroxyl group close to the active site may disturb H-bonding networks or active site electrostatics as eluded to above.

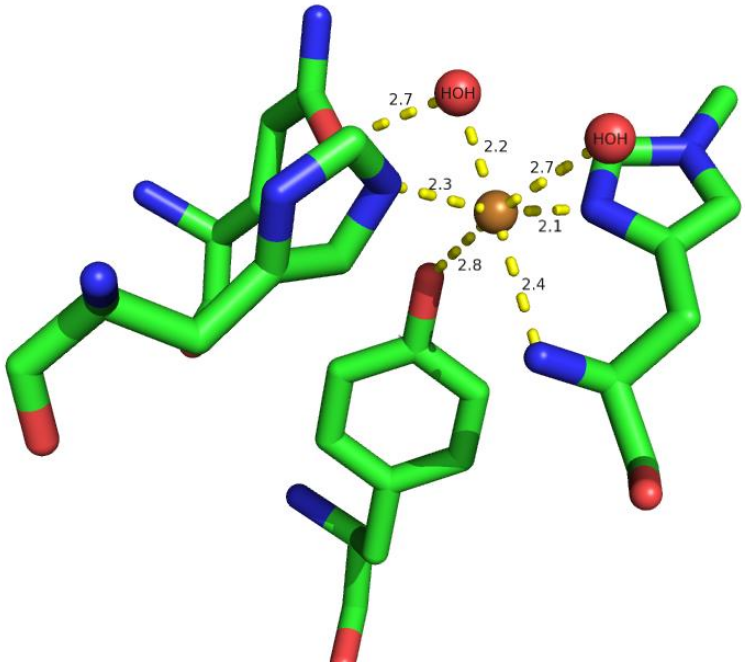


Fig 5.4. Copper ion coordination by TaGH61A (PDB ID: 2YET; Quinlan et al). Residues shown are His86 (left in figure), Gln173 (top in figure), His1 (right in figure) and Tyr175 (bottom in figure). Please note that the signal sequence has not been taken into account when numbering the residues. It should also be noted that His1 is modified by post translational

modification by methylation of a nitrogen atom on the imidazole ring. Two water molecules are shown as red spheres. The copper ion is shown as a gold colored sphere. All distances are in Å. The figure was made by using PyMOL. (DeLano, W. L. *et al*., 2005)

The distance between Phe187 of CBP21 and the two metal binding histidines is 3.8 Å (for His28) and 4.2 Å (for His114). Interestingly, when analyzing the distances between these residues in the other CBM33-type LPMOs with structures that are available, the Phe is equally close to the N-terminal histidine (Fig. 5.5). Thus, changing the Phe to Tyr and thereby increasing the "reach" of the Phe residue by ~1.4Å would indeed require some adjustment of the catalytic site residues, especially the N-terminal histidine (given the CelS2 catalytic site is very similar to that of CBP21, EfCBM33A and GbpA). It is not possible to conclude about these possible explanations in the absence of structural information for both wild-type and mutant. Even if phenylalanine is substituted to tyrosine in many other LPMOs, the result of this study shows that this single substitution is not appropriate for CelS2 and that additional mutations would most likely be required. Hopefully will future structures of CBM33-type LPMOs with tyrosines in the catalytic site be of help in understanding why the CelS2 Phe212Tyr is devoid of activity.

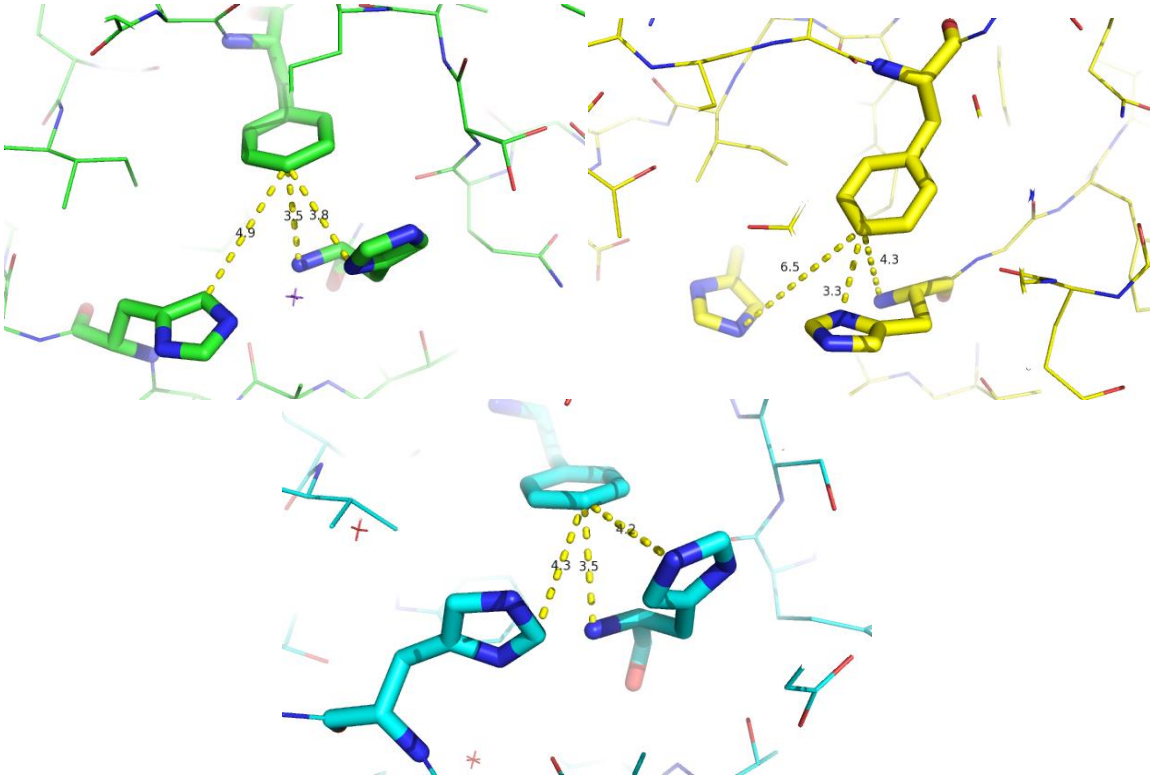


Fig 5.5. Distances between the active site Phe (Phe219 in CelS2) and the closest atom of the catalytic histidines in CBP21 (green colored carbon atoms, His114 on the left, His28 on the right and Phe187 at the top), EfCBM33A (yellow colored

carbon atoms, His114 on the left, His29 on the right and Phe185 at the top) and GbpA (cyan colored carbon atoms; His121 on the left, His24 on the right and Phe193 at the top). The figure was made by using PyMOL. (DeLano, W. L. *et al*., 2005)

It is important to note that the enzyme characterizations carried out so far are rather preliminary and that much more work is needed to fully characterize and understand the mutational effects. The work on CelS2 has many weaknesses which could be improved. For instance, due to lack of quantified cello-oligomer standards, the study lacks a true quantitative assay. Furthermore, the analyses rely on single substrate type (PASC) and limited analytical methods were used. In addition to improving the mentioned weaknesses, additional methods and work cold be included. For instance, pH-activity curves could have been constructed for analysis of the catalytic activity of mutants compared to the wild type at different pHs. Due to its accuracy, it would also be interesting to include oxidized chito-oligomer generation by the mutants and wild type using MALI-TOF MS, however, this method can not be used for quantification of chitooligosaccharides (MALDI-TOF MS is a qualitative method not quantitative).

Jden1381

The unique modular structure of the multi-domain LPMO from *Jonesia denitrificans* encoded by *Jden*1381caught interest when selecting enzymes for characterization in this study (Fig 4.1). The gene was successfully cloned and expressed in full length and truncated versions in pUCBB-eGFP and pET32b expression vectors (section 4.2.2). So far, there has not been published any work on cloning or characterization of multi-domain LPMOs that contain a glycoside hydrolases on the same gene. The current study shows that Jden1381 indeed is an enzyme with multiple catalytic activities, since the full-length enzyme showed both LPMO and GH18 activity.

In this study, a total of 13 constructs were made for expressing Jden1381both in full length and truncated versions because of four reasons: *i*) to assess relative contribution of the individual domains (the N-terminal LPMO and CBM5/12) on both substrate binding and catalytic activity of the chitinase (C-terminal GH18 domain), *ii*) for individual characterization of the LPMO and GH18 domains, *iii*) to simplify purification procedures

(for C-His_6 versions) and *iv*) to solve encountered complications on expression of the full length Jden1381.

Of the 13 constructs, five were able to provide considerable amount of soluble proteins (pUCBB/*Jden*1381fl, pUCBB/*Jden*1381-LPMO, pUCBB/*Jden*1381-LPMO_C-His_6, pUCBB/*Jden*1381-LPMO-CBM5/12 and pUCBB/*Jden*1381-LPMO-CBM5/12_C-His_6) while pET32b/*Jden*1381fl showed weak expression. The rest of the constructs (pUCBB/*Jden*1381fl_C-His_6, pUCBB/*Jden*1381-CBM5/12, pUCBB/*Jden*1381-CBM5/12_C-His_6, pUCBB/*Jden*1381-CBM5-GH18, pUCBB/*Jden*1381-CBM5/12-GH18_C-His_6, pUCBB/*Jden*1381-GH18 and pUCBB/*Jden*1381-GH18_C-His_6) did not show expression of the recombinant protein. A number of reasons and speculations can be discussed concerning the lack of success in expressing these constructs, however, I would like to highlight possibly main reason; since most of these constructs bear truncated versions, there may be error or inaccuracy on the annotated domain prediction (section 4.1.1)

Two different clones were made for expressing the full length protein of which the pUCBB/*Jden*1381fl clone worked. The pET32b/*Jden*1381 did not yield a considerable amount of protein. The success of pUCBB/*Jden*1381 was probably due to, among many tested conditions, a change in the growth media from LB to 2xTY (see Table 3.1). The pET32b/*Jden*1381 construct was generated using the Gibson Assembly cloning method as described in section 3.2.8.2. Combination of pET32b expression vector together with the Gibson Assembly cloning method was anticipated to be beneficial to obtain high level of soluble and mature Jden1381fl (discussed in section 3.2.8.2). In order to achieve this, a restriction free cloning method with a potential of assembling large DNA fragments (as pET32b) was required. The method is efficient and worked well. During expression of full length Jden1381 we repetitively observed conversion of the protein to smaller fragments; possibly representing the protein's individual domains or combinations thereof (an example is shown in figure 4.14). Bands possibly representing full length Jden1381 were generally weak, but in some cases, overnight cultures of BL21(DE3) transformant harboring pUCBB/*Jden*1381fl construct grown in 2xYT media at 30 ^{o}C gave better yield; see Fig.4.15 & Table 3.1. For unclear reasons, attempt to purification of this Jden1381fl was not successful, therefore, the activity of full length Jden1381 was tested using samples from periplasmic extract obtained from a culture of BL21(DE3) harboring pUCBB/*Jden*1381fl.

Periplasmic extracts containing full length Jden1381 showed activity towards collidal chitin, β-chitin and α-chitin from which they produce native chito-oligosaccharides (Fig. 4.22). This clearly shows that Jden1381 contains chitinase activity. The result shows the production of more native chitobiose from *Jden*1381 acting on α-chitin compared to collidal chitin or β-chitin. This may indicate that the enzyme preference for α- chitin, which would be interesting since chitinases generally have lower activity on α-chitin than on other chitin forms. The products observed from collidal and β-chitin are mixtures of native chitooligosaccharides (DP1-DP6). Due to lack of time, no effort to analyze the presence of products from LPMO (oxidized chitooligosaccharides) in this product mixture was made thus, it is not certain whether the activities measured in this assay include a possible "boosting" effect of the LPMO domain. The different product profiles obtained for the different substrates may be simply due to differences in catalytic efficiency, as discussed above and in section 4.5.1. However, it is tempting to speculate that the other two domains are involved in orienting the GH18 module on the substrate and that this orientation varies between substrates.

Truncated and well-purified versions of Jden1381 containing the LPMO domain generated oxidized products when acting on β-chitin in the presence of copper ion and ascorbic acid, similar to what has been observed for other chitin-active LPMOs (Fig 4.23 & 4.24). It would have been interesting to do similar analysis on the other two chitin forms, especially α-chitin, since the full length enzyme seems to be particularly active on this highly recalcitrant chitin form.

Judged by its sequence, the C- terminal GH18 domain resembles the well studied endo-acting chitinases ChiC from *Streptomyces coelicolor* (69 % sequence identity) and endochitinase Chi21702 from *Sanguibacter antarctiucs* (74 % sequence identity) (Park *et al.*, 2009, Saito, *et al.*, 1999), which may imply that Jden1381-GH18 is an endo-acting chitinase that contains an N-terminal LPMO domain.

The results of these two studies are preliminary which gave general insights. The results gave inspiration for further deep investigation of LPMOs. Therefore, further work on Jden1381is planned while the work on CelS2 will continue by Zarah Forsberg (the supervisor for CelS2 work).

In order to study these cellulose and chitin active LPMOs, huge amount of time, effort and work were invested. More characterization work on both CelS2 for Jden1381 could not be done because of the huge amount of effort was spent on cloning and purification. In this study, 13 genes were cloned (Jden1381 variants) and six mutations were made (CelS2 mutants) and the relative catalytic activity/effect of the expressed enzymes were assessed. However, even though some very interesting data was generated, there is room for improvement in our methods which should be marked for advanced further work.

For optimal characterization of Jden1381, sufficient amount of enzyme is required, thus attempts will be made to solve challenges in protein expression. For optimizing expression of Jden1381variants, attempts will be made to clone these genes in expression vectors derived from Gram positive bacteria (*Bacillus subtilis,* for instance). If expression of all Jden1381 variants succeeded, synergetic effects between the N-terminal LPMO domain, the CBM12/5 domain and the C-terminal GH18 domain will be assessed and the LPMO and GH18 domains will be characterized individually. It will be of interesting to investigate the possible effect of the sub-domains in substrate binding on GH18. The current study in particular shed light on substrate preference of Jden138, that α-chitin is favored over β-and colloidal chitin, which is not a common behavior of most chitinases (discussed above). This may indicate that this chitinase have even more unique features beyond its domain structure for its enzyme family.

Outlook

For the past decades, biomass has become significant resource for production of biofuel. Biomass conversion requires biochemical transformations such as fermentation of sugar to ethanol. Enzymatic conversion of biomass into simple sugars is preferable over chemical conversion due to the potential of enzymes in preserving the structure of carbohydrates (Horn *et al.*, 2012).

Chitinases and cellulases are among promising candidates to be utilized for processing continuously growing amount of renewable biomass. The presence of recalcitrance polysaccharides in biomass is the key challenge for process acceleration and efficient degradation by these enzymes. In addition, enzymatic hydrolysis is the primary process bottleneck with respect to cost and yield in biomass conversion; as a result, having a deep understanding of the hydrolytic activities of enzyme candidates is crucial to making biomass

to biofuel transformation cost competitive. Decades of research have been dedicated in developing advanced enzymatic technology addressing this issue. The discoveries of LPMOs and GH61s may open an opportunity to solve this challenge, although we are just starting to understand how these enzymes work. The geographical diversity of these enzymes indicates the presence of many LPMOs with wide variety in substrate specificities, functions and potential.

Many laboratory studies are performed by "artificial" substrates (e.g. purified cellulose/chitin, filter paper etc.) that are convenient for characterizing these enzymes. However, this might not show the whole picture about the enzymes' behavior unless "real" activities are assessed with natural substrates (e.g. intact plant cell walls or crustacean shells). Furthermore, it will be interesting to compare possible behavioral changes by LPMOs when acting on natural substrates rather than "artificial" substrate.

LPMOs are important for bio-economy. Today LPMOs are already being used in multi-enzyme cocktails such as Cellic CTec3, the most cost-efficient enzyme cocktail which decreases the required enzyme dose for biomass conversion by 4/5 (Novozymes) and it is probably only getting better because much potential is still untapped.

6 REFERENCES

Aachmann, F.L., Sørlie, M., Skjåk-Bræk, G., Eijsink, V. G. H. and Vaaje-Kolstad, G. (2012) "NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions" *Proc Natl Acad Sci U S A, 109(46): 18779-18784.*

Anderson, K., Carlsson, M., Gustafsson, C., Heijbel, A. (1999) "Purification of Poly(His)-tagged Recombinant Proteins using HisTrap" *Amersham Pharmacia Biotech UK Limited*

Baneyx, F. (1999) "Recombinant protein expression in *Escherichia coli*" *Current Opinion in Biotechnology 10:411-421*

Béguin P., Aubert J.P. (1994) The biological degradation of cellulose, *FEMS Microbiol Rev.13 (1):25-58*

Bentley, S.D., *et al* (2002), Complete genome sequences of medel actinomycete *Streptomyces colicolor* (A3)2, *Nature (417) 141-147*

Birnboim, H.C. and Doly, J. (1979) "A rapid alkaline extraction procedure for screening recombinant plasmid DNA" *Nucleic Acids Reas.,24;7(6):1513-23*

Boraston, A.B., Kwan, E., Chiu P., Warren, R. A. J. and Kilburn, D. G. (2002), Recognition and hydrolysis of noncrystalline cellulose, *The journal of biological chemistry, Vol 278 No. 8 pp 6120-6127*

Boraston, A.B., Bolam D. N., Gilbert, H. J., Davies, G. J. (2004) "Carbohydrate-binding modules: fine-tuning polysaccharide recognition". *Biochem. J., 2004. 382(Pt 3): p. 769-81*.

Brown, R. M., JR. (2003) "Cellulose Structure and Biosynthesis: What Is in Store for the 21st Century" *Journal of Polymer Science. Part A Polymer Chemistry, Vol 42, 487-495*

Casali, N. (2003) "*Escherichia coli* host strains" *Methods in Molecular Biology™ Volume 235, Page 27-48*

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

Carlstrom, D. (1957) "The Crystal Structure of α-chitin (poly-N-Acetyl-D-Glucosamine)" *J Biophys Biochem Cy* 3(5):669-683.

Cleland, J.L., Craik, C.S. (1996) "Protein Engineering: Principles and Practices" *Wiley-Liss, Inc. page 5-10*

Cosgrove, D.J. (2005) Growth of the plant cell wall" *Nature Review Molecular Cell Biology 6, 850-861*

Davies, G.j., Wilson. K.S. and Herissat, B (1997) "Nomenclature for sugar-binding subsites in glycosyl hydrolases" *Biochem. J.321,557-559*

DeLano, W. L. and Lam, J. W. (2005) "PyMOL: acommunications tool for computational models" *Abstr. Pap. Am. Chem. Soc. 230, 1371–1372*

Garda, A.L., Fernández-Abalos, J.M., Sánchez, P., Ruiz-Arribas, A. and Santamaria, R.I. (1997) "Two genes encoding an endoglucanases and a cellulose-binding protein are clustered and co-regulated by a TTA codon in *Streptomyces hastedii* JM8" *Biochem. J. 324,403-411*

Dewick, M. P. (2009) "Medicinal Natural Products. A Biosynthetic Approach", *3rd edition, page 494*

Eijsink, V. G. H., Vaaje-Kolstad, G., Vårim, K., Horn, S. J. (2008)"Towards new Enzymes for biofuels: lessons from chitinase research" *Trends in biotechnology, Volume 26, issue 5 Page 228-23.*

Ermolaeva, Maria D. (2001) Synonymous Codon Usage in Bacteria. *Curr. Issues Mol. Biol. 3(4): 91-97*

Filipponen. I (2009), "The synthetic strategies for unique properties in cellulose nanocrystal materials" *Wood and paper science, Raleigh, North Carolina (North Carolina state university)*

Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunæs, A.C., Stenstrøm, Y., MacKenzie, A., Sørlie, M., Horn, S.J. and Eijsink, V. G. H. (2011), "Cleavage of cellulose by a CBM33 protein", *Protein science 2011 VOL 20:1479—1483*

Funke, G., Von Graevenitz A., Clarridge JE 3rd, Bernard KA., (1997) « Clinical microbiology of coryform bacteria» *Clin Microbiol Rev. 1997 Jan 10(1) 125-59.*

Gront D., Grabowski M, Zimmerman MD, Raynor J, Tkaczuk KL, Minor W.(2012) "Assessing the accuracy of template-based structure prediction metaservers by comparison with structural genomics structures" *J. Struct Funct Genomics, 13(4):213-25*

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred

kilobases. Nat. Methods 6, 343–345.

Gibson, Daniel G. (2011) "Enzymatic Assembly of Overlapping DNA Fragments" *Methods in Enzymolog, volume 498, Elsevier Inc. ISSN 0076-6879,*

Guetta, Daniel (2006) "Acidity, Basicity and pKa" http://www.columbia.edu/~crg2133/Files/CambridgeIA/Chemistry/AcidityBasicitykPa.pdf

Handbook: Amersham Biosciences "Affinity Chromatography: Principles and methods" *Edition AC 18-1022-29*

Handbook: Amersham Biosciences "Gel filtration: Principles and methods" *Edition Al 18-1022-18*

Haworth, W.N (1937), "The structure of carbohydrate and vitamin C" *Nobel Lecture*

Henrisat, Bernard (1991) "The classification of glycosyl hydrolases based on amino acid sequence similarities" *Biochem. J. 280, 309-316*

Herrisat, B. and Davies, G. J. (2000) 2Glycoside Hydrolases and Glygosyltransferases. Families, Modules and Implications for Genomics" *Plans Physiology vol 124 no. 4, 1515-1519*

Himmel, Michael E., Ding, Shi-You, Johnson, David K., Adney, William S., Nimlos, Mark R., Brady, John W., Foust, Thomas D. (2007) "Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production" *Science 315, 804*

Hogrefe, H. H., Cline, J., Youngblood, G.L., Allen R.M. (2002) "Creating randomized amino acid libraries with QuickChange Multi Site-Directed Mutagenesis Kit" *Biotechniques 33(5):1158-6*0, 1162, 1164-5

Horn, S. J., Sikorski, P., Cederkvist, J.B., Vaaje-Kolstad, G., Sørlie, M., Synstad, B., Vriend, G., Vårum, K.M. and Eijsink, V.G. H. (2006) "Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides" *PNAS Vol. 103 no. 48, 18089-18094*

Horn, S.J., Vaaje-Kolstad, G., Westereng, B. and Eijsink, V. G. H. (2012) "Novel enzymes for the degradation of cellulose" *Biotechnology for Biofuel, 5:45*

http://www.cazy.org/

http://www.dionex.com/en-us/index.html

http://www.emdmillipore.com

http://www.genscript.com/codon_opt.html

http://www.newworldencyclopedia.org/entry/Carbohydrate

http://pfam.sanger.ac.uk

http://www.uniprot.org

http://en.wikipedia.org/wiki/Arginine

http://www.wpclipart.com/plants/diagrams/Plant_cell_wall_diagram.png.html

Jose T., Raphael L., Arthur J.C., Ely M., Edward A.B., Yuval S. and Thomas A.S. (1996) "Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose" *The EMBO Journal vol.15 no.21 pp.5739-5751, 1996*

Jung, H., Wilson, D. B. and Walker, L. P. (2001) "Binding Mechanisms for *Thermobifida fusca* Cel5A, Cel6B and Cel48A Cellulose-Binding Modules on Bacterial Microcrystalline Cellulose" *Biotechnol Bioeng. 20;80(4):380-92*

Kai Zhang, Andreas Geissler, Steffen Fischer, Erica Brendler, and Ernst Baucker (2012) "Solid-State Spectroscopic Characterization of α-Chitins Deacetylated in Homogeneous Solutions" *The journal of physical chemistry B., 116, 4584−4592*

Karkehabadi, S., Hansson, H., Kim, S., Piens, C. M., Sandgren, M. (2008) "The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from Hypocrea jecorina, at 1.6 angstrom Resolution". *J Mol Biol, 2008. 383(1): p. 144-154.*

Li, X ., Beeson, W.T., Phillips, C.M., Marletta, M.A. and Cate, J.H.D. (2012) "Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases" *Structure, 20:1051–1061.*

Liu, Y.S, Baker, J.O., Zeng Y., Himmel, M.E., Haas T. and Ding, Shi-You (2011) " Hydrolyzes crystalline cellulose on hydrophobic faces2, *The journal of biological chemistry, Vol 286, NO 13, 11195-11201*

Lindhorst, T.K. (2007) "Essentials of carbohydrate chemistry and biochemistry", *3rd edition, page 1*

Lloyd D. (2007) "Cellulose microfibril aggregates and their size variation with cell wall type" *Wood Sci Technol (2007) 41:443–460*

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M (1993) " A Thioredoxin Gene Fusion Expression System that Circumvent Inclusion Body Formation in the *E. coli* Cytoplasm" *Nature biotechnology 11, 187-193*

Maertens, B., Spriestersbach, A., Groll,. Uritza von, Roth, U., Kubicek, J., Gerrits, M., Graf, M., Liss, M., Daniela, d., Wagner, R. and Schäfer (2010) "Gene optimization mechanisms: A multi-gene study revels a high success rate of full-length human proteins expression in *Escherichia coli*" *Protein Science Volume 19, Issue 7*

Mulisch, M (1993) "Chitin in Protistan Organisms- Distribution, Synthesis and Deposition" *Europ. J. Protistol. 29, 1-18*

Mergulhão, F.J.M., Summers, D.K., Monteiro, G.A. (2005) "Recombinant protein secretion in *Escherichia coli*" *Biotechnology Advances 23, 177-202*

Min D, Li Q, Jameel H, Chiang V, Chang HM. (2012) "The Cellulase-Mediated Saccharification on Wood Derived from Transgenic Low-Lignin Lines of Black Cottonwood (Populus trichocarpa)" *Appl Biochem Biotechnol 168 (4):947-55*

Moser, F, Irwin, D., Chen, S., and Wilson, D (2007) " Regulation and Characterization of *Thermobifida fusca* Carbohydrate-Binding Module Proteins E7 and E8" *Biotechnology and Bioengineering Volume 100, Issue 6*

New England Biolabs Japan; http://www.nebj.jp/products/detail/1238)

http://www.cazypedia.org/index.php/Glycoside_hydrolases

Nutrition resources-Chemistry review-Carbohydrates
http://nutrition.jbpub.com/resources/chemistryreview9.cfm

Oke, Isdin (2010) "Nanoscience in nature: cellulose nanocrystals" *Studies by undergraduate researchers at Guelph, Vol 3, No 2*
(http://journal.lib.uoguelph.ca/index.php/surg/rt/printerFriendly/1132/1656)

O'Sullivan, A. C. (1997) "Cellulose: the structure slowly unravels" *Cellulose 4, 173-207*

Park, H.J., Kim, D., Kim, I. H., Lee,C. E., Kim, I. C., Kim, Y., Kim, S. J., Lee, H. K. and Yim, J. H. (2009) "Characteristics of cold-adaptive endochitinase from Antarctic bacterium *Sanguibacter antarcticus* KOPRI 21702" *Enzyme and Microbial Technology, Volume 45, Issue 5, Pages 391-396*

Raabe, D., Romano, P., Sachs, C., Fabritius, H., Al-Sawalmih, A., Yi, S.B., Servos, G. and Hartwig, H.G (2006) "Microstructure and crystallographic texture of the chitin–protein network in the

biological composite material of the exoskeleton of the lobster *Homarus americanus" Material Science and Engineering A 421, 143-153*

Robyt. J. F (1997) "Essentials of Carbohydrate Chemistry" *Springer, Page 157*

Rüdiger, P., Gahrich-Schröter, G., Lapidus, A., Nolan, M., Rio. T.G.D., Lucas, S., Chen, F., Tice, H., Pitluck, S., Cheng, J.F., Copeland, A., Saunder, E., Brettin, T., Detter, J.C., Bruce, D., Goodwin, L., Pati, A., Ivanova, N., Mavromatis, K., Ovchinnikova, G., Chen, P., Göker, M., Bristow, J., Elisen, J.A., Markowitz, P., Hugenholtz, P., Kyripides, N.C., Klen, H.P. and Han, C (2009) "Complete genome sequence of Jonesia denitrificans type strain (Prevot 55134^{T})" *Standards in Genomic Sciences* 1: 262-269

Rudrapatnam N. Tharanathan and Farooqahmed S. Kittur (2003)" Chitin - The Undisputed Biomolecule of Great Potential" *Critical Reviews in Food Science and Nutrition Volume 43, Issue 1*

Saito, A., Fujii, T., Yoneyama, T., Resenbach, M., Ohno, T., Watanabe, T. and Miyashita, K. (1999) "High-multiplicity of Chitinase genes in *Streptomyces coelicolor* A3(2)" *Biosci. Biotechnol. Biochem., 63(4), 170-718*

Saito, A., Ishizaka, M., Francisco Jr, P.B., Fujii, T. and Miyashita, K. (2000b) "Translational co-regulation of five chitinase genes scattered on the *Streptomyces coelicolor* A3(2) chromosome" *Microbiology 146,2937-2946*

Saito, Yukie, Okano, Takeshi, Gaill, Françoise, Chanzy, Henri and Putaux, Jean-Luc (2000a) "Structural data on the intra crystalline swelling of β-chitin" *International Journal of Biological Macrobiology 28: 81-88*

Sastalla, I., Chim, K., Cheung. G.Y.C., Pomerantsev. A. and Leppla, A.H " Codon-optimized Fluorescent Proteins Designed for Expression on Low-GC Gram-Positive Bacteria" (2009) *Appl environ Microbiol. 75(7):2099-2110*

Schueler-Furman, O. and Beker, D. (2003) "Conserved Residue Clustering and Protein Structure Prediction" *Proteins: Structure, Function and Genetics 52:225-235*

Schwarz, W.H., (2001) The cellulosome and cellulose degradation by anaerobic bacteria, *Appl Microbiol Biotecnol 56:634-649*

Senni, K., Pereira, J., Gueniche, F., Delbarre-Ladrat, C., Sinquin, C., Ratiskol, J., Godeau, G., Fischer, A.M., Helley, D. and Colliec-Jouault, S. (2011) "Marine Polysaccharides: A Source of Bioactive Molecules for Cell Therapy and Tissue Engineering" *Mar Drugs.9(9):1664-1681*

Shimahara, K. and Takiguchi, Y. (1988) "Preparation of crustacean chitin" *Methods Enzymol, 16:, 417-423*

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Peredez, A., Persson, S., Raab, T., Vorwerk, S and Youngs, H. (2006) "Towards a System Approach to Understanding Plant Cell Walls" *Science Review, Vol. 306*

Sørensen, H. P., Mortensen, K. K. (2004) "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*" *Journal of Biothechnolugy 115; 113-128*

Taherzadeh, M. J. and Karimi, K., "Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review" *Int J Mol Sci. 9(9):1621-1651*

Taira, T., Toma, N. and Ishihara, M. (2004) "Purification, Characterization and Antifungal Activity of Chitinases from Pineapple (Ananas comosus) Leaf" *Biosci. Biotechnol. Biochem., 69(1), 189-196*

Teeri, T.T. (1997), "Crystalline cellulose degradation: new insight into the function of cellobiohydrolases". *Trends Biotechnol 15:160–167*

Thu, V. Vuong and Wilson, David B (2010) "Glycoside Hydrolases: Catalytic Base/Nucleophile Diversity" *Biotechnology and Bioengineering 107: 195-205*

Vaaje-Kolstad G, Horn SJ, van Aalten DMF, Synstad B, Eijsink VGH (2005a) "The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation" *J Biol Chem 280:28492–28497.*

Vaaje-Kolstad, G., Houston, D. R., Riemen, A.H.K., Eijsink, V.G.H. and Aalten, D.M.F. (2005b) "Crystal Structure and Binding Properties of the *Serratia marcescens* Chitin-binding Protein CBP21" *The Journal of Biological Chemistry, Vol. 280, No 12, pp11313-11319*

Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z. L., Zhai, H., Sørlie, M. and Eijsink, V. G. H. (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides, *Science 330, 219-222.*

Vaaje-Kolstad, Liv Anette Bøhle, Sigrid Gåseidnes, Bjørn Dalhus , Magnar Bjørås, Geir Mathiesen and Vincent G. H. Eijsink (2012) "Characterization of the Chitinolytic Machinery of Enterococcus faecalis V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme" *J. Mol. Biol. 416:239-254*

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., Marth, J. (1999) "Essentials of Glycobiology", page 17

Vick, J. E, Johnson, E.T., Choudhary, S., Bloch, S.E., Lopez-Gallego, F., Srivastava, P., Tikh, I., Wawrzyn, G.T. and Scmidt-Dannert, C. (2011) "Optimized compatible set of BioBrick™ vectors for metabolic pathway engineering" *Appl microbial Biotechnology 92:1275-1286*

Vydac product manual "Principles and applications of high performance ion-exchange chromatography for bio-separations" http://www.seaviewsci.com/vydac/vydacpubs/vhpionex.pdf

Wagner S., Baars L.,Ytterberg, A. J., Klussmeier A., Wagner, C. S., Nord, O., Nygren, Per-Åke, Wijk, K. J. V., Gier, Jan-Willem (2007) "Consequences of Membrane Protein Overexpression in *Escherichia coli*" Molecular and *Cellular proteomics 6.9*

Wallner, B., Elofsson (2005) "All are not equal: A benchmark of different homology modeling programs" *Protein Science 14:1315-1327*

Waltson, J. D., Caudy, A. A., Myers, R. M. and Witkowski. J. A. (2007) Recombinant DNA: genes and genomes: a short course. 3rd edition San Fransosco: W.H. Freeman.ISBN 0-7167-2866-4.

Warren, R.A.J. (1996) "Microbial Hydrolysis of Polysaccharides" *Annu. Rev. Microbiol. 50:183-212*

Wilson, D. (2009) "The first evidence that a single cellulase can be essential for cellulose degradation in a cellulolytic microorganism", *Molecular Microbiology 74[6], 1287-1288*

Westereng, B., Ishida, T., Kolstad-Vaaje, G., Wu, Miao, Eisink, V.G.H., Igarashi, K., Samejima, M., Ståhlberg, J., Horn , S. and Sandgren, M (2011) "The putative Endonuclease PcGH61D from *Phanerochaete chrysosporium* is a Metal-Dependent Oxidative Enzyme that Cleaves Cellulose" *PLoS One; 6 (11):e27807*

World Intellectual property Organization; International publication number WO 2012/019151 A1

Yang, Y.-B., Harrison. K., Kindsvater, J. (1996) "Characterization of a novel stationary phase derivative from a hydrophilic polystyrene-based resin for protein cation exchange high-performance liquid chromatography" *Journal of chromatography A 723: 1-10*

Zakariassen, H., Aam, B. B., Horn, S. J., Vårim, K. M., Sørlie, M. and Eijsink, V. G. (2009) "Aromatic residues in the catalytic center of chitinase A from *Serratia marcescens* affect processively, enzyme activity and biomass converting efficiency" *J. Biol Chem 17; 284(16):10610-7*

Zaragoza, O., Rodrigues, M.L., Jesus, M.D., Frases, S., Dadachova, E., Casadeval, A (2009) "The capsule of the fungal pathogen *Cryptococcus neoformans*"*Adv Appl Microbiol. 68:133-216*

Zetins, A., Schrempf H. (1995) "Visualization of α-Chitin with a Specific Chitin-Binding Protein (CBH1) from *Streptomyces olivaceoviridis*" *Analytical Biochemistry 231,287-294*

APPENDICES

Appendix A) Genomic data for glycoside hydrolases and related proteins of *streptomyces coielicolor* A3(2) (data taken from CAZy database)

List Of Proteins

Protein Name	Family	Reference Accession
α-L-arabinofuranosidase (AbfB;SCO5932)	CBM13,GH62	CAA16189.1
Branching enzyme (GlgBI;SC6A11.16c;SCO5440)	CBM48,GH13	CAA58314.1
branching enzyme II (GlgBII;SC4G10.11c;SCO7332)	CBM48,GH13	CAB92878.1
Cel1 (SCO0765 or SCF81.24c)	CBM4,GH9	CAB61539.1
cellulose-active enzyme (SCO1188;SCG11A.19;CelS2)	CBM33,CBM2	CAB61600.1
chitin-binding protein (Chb3;SCO0481;SCF80.02)	CBM33	CAB57190.1
chitinase A (ChiA;Chi18bA;SCK15.05c;SCO5003)	CBM16,GH18	CAB92596.1
chitinase B (ChiB;Chi18bB;SCO5673;SC8B7.05c)	CBM16,GH18	CAA20216.1
chitinase C (ChiC;Chi18aC;2SC6G5.20c;SCO5376)	CBM2,GH18	CAB94547.1
chitinase E (ChiE;Chi18aE;SCO5954;SC7H1.24)	CBM2,GH18	CAA16211.1
chitinase F (ChiF;SC5H1.29c;SCO7263)	CBM12,GH19	CAB42954.1
chitinase L (ChiL;Chi18cL;SCO2799;2SCC13.07c)	CBM2,GH18	CAC10108.1
GlgX (SCO7338;SC4G10.17)	CBM48,GH13	CAB92884.1
SCO0284 or SCF85.12	GH27,CBM51	CAB54169.1
SCO0335	CBM32,CBM32	CAB56140.1
SCO0505 or SCF34.24	CBM35	CAD55445.1
SCO0535 or SCF11.15	CBM3,GH16	CAB59592.1
SCO0545 or SCF11.25	CBM32,CBM32,GH55	CAB59602.1
SCO0554 (ManA;SCF11.34c)	GH5,CBM2	CAD55266.1
SCO0643 or SCF91.03c	CBM33	CAB61160.1
SCO0674 or SCF91.34c (XlnA;XysA)	GH10,CBM2	CAB61191.1
SCO0764 or SCF81.23c	CBM13	CAB61540.1
SCO0787 or 3SCF60.19	GH16,CBM13	CAC14352.1
SCO0829 or SCF43A.19	CBM12	CAB48906.1
SCO1061 / SCG22.07c	GH93,CBM13	CAB95280.1
SCO1187 (CelB;SCG11A.18)	GH12,CBM2	CAB61599.1
SCO1226 or 2SCG1.01c	CBM12	CAD55167.1
SCO1734 (SCI11.23)	CBM33	CAB50949.1
SCO2226 or SC10B7.21c	GH13,CBM41,CBM48,GH13	CAB90874.1
SCO2291 (AxeA)	CE4,CBM2	CAB61737.1
SCO2292 (XlnB)	GH11,CBM2	CAB61738.1
SCO2383 or SC4A7.11	CBM13	CAB62715.1
SCO2833 (Chb)	CBM33	CAB65563.1
SCO3842 or SCH69.12	CBM12	CAB45209.1
SCO4117	CBM51	CAB92370.1
SCO4257 or SCD8A.30	CBM13	CAB77351.1
SCO4258 or SCD8A.31	CBM13	CAB77352.1
SCO4481 or SCD65.24	CBM13	CAD55485.1
SCO4488 or SCD69.08 (putative protein kinase)	CBM13	CAB92109.1

SCO4670 or SCD40A.16c	CBM12	CAB81861.1
SCO5456 or SC3D11.13c	CBM48,GH13	CAB76010.1
SCO5685 or SC5H4.09c	GH3,CBM6	CAB91121.1
SCO5786 or SC4H2.07c	CBM2	CAA18323.1
SCO6078 or SCBAC1A6.02c	CBM48,GH13	CAC33923.1
SCO6082 (GlgX3)	CBM48,GH13	CAC33927.1
SCO6234 (ManA2;SC2H4.16)	GH5,CBM10	CAA20610.1
SCO6345 or SC3A7.13	CBM33,CBM5	CAA20076.1
SCO6348 / SC3A7.16c	GH101,CBM32	CAA20079.1
SCO6377	CBM13	CAA20165.1
SCO6428 or SC1A6.17c	GH92,CBM6	CAA18915.1
SCO6456 or SC9B5.23c	CBM13	CAA22765.1
SCO6546 (SC5C7.31c)	CBM2,GH48	CAA20643.1
SCO6548 (SC5C7.33)	CBM2,GH6	CAA20645.1
SCO6572 or SC3F9.07	CBM6	CAA19630.1
SCO6596 or SC8A6.17	GHnc,CBM32	CAA19789.1
SCO6665 or SC5A7.15C	GH16,CBM6	CAA19944.1
SCO7015 or SC1H10.04c	CBM32,CBM32,CBM32,GH87	CAB88148.1
SCO7019 (Aml)	CBM25,CBM25,GH13	CAB88152.1
SCO7037 or SC4G1.03	CBM32,CBM32,GH87	CAC01535.1
SCO7170 or SC9A4.32	CBM6	CAC01659.1
SCO7211 or SC2H12.10c	CBM13	CAB94634.1
SCO7212 or SC2H12.11c	CBM6,CBM13	CAB94635.1
SCO7225 / SC2H12.24 (ChiM)	CBM33,CBM5	CAB94648.1
SCO7228 or SC2H12.27c	CBM35,PL9	CAB94651.1
SCO7406 or SC6D11.02c	GH30,CBM13	CAB76325.1
SCO7637 or SC10F4.10c	CBM2,GH5	CAC16970.1
SCO7774 or SC5E9.22	CBM13	CAC14502.1
SCO7775 or SC5E9.23	CBM13	CAC14503.1
Xylanase A (XlnA;SCO5931;SC10A5.36c)	GH10,CBM13	CAD55241.1
xyloglucan-specific endo-β-1,4-glucanase (SCO6545)	GH74,CBM2	CAA20642.1

List Of CBM33 type LPMOs

Protein Name	Family	Reference Accession
cellulose-active enzyme (SCO1188; SCG11A.19; CelS2)	CBM33,CBM2	CAB61600.1
chitin-binding protein (Chb3; SCO0481; SCF80.02)	CBM33	CAB57190.1
SCO0643 or SCF91.03c	CBM33	CAB61160.1
SCO1734 (SCI11.23)	CBM33	CAB50949.1
SCO2833 (Chb)	CBM33	CAB65563.1
SCO6345 or SC3A7.13	CBM33,CBM5	CAA20076.1
SCO7225 / SC2H12.24 (ChiM)	CBM33,CBM5	CAB94648.1

Appendix B) Putative Carbohydrate-active enzymes detected in the genome *Jonesia denitrificans* DSM 20603 (data taken from CAZy database)

List Of Proteins

Protein Name	Family	Reference Accession
Jden_0062	CBM42	ACV07738.1
Jden_0324	GH64,CBM13	ACV07995.1
Jden_0329	CBM13,PL3	ACV08000.1
Jden_0385	GH53,CBM61	ACV08054.1
Jden_0538	CBM2,GH6	ACV08202.1
Jden_0731	GH74,CBM2	ACV08395.1
Jden_0732	GH10,CBM2	ACV08396.1
Jden_0733	GH10,CBM2	ACV08397.1
Jden_0734	GH5,CBM2	ACV08398.1
Jden_0735	GH6,CBM2	ACV08399.1
Jden_1134	GH48,CBM2	ACV08790.1
Jden_1381	CBM33,CBM5,GH18	ACV09037.1
Jden_1648	CBM22,CBM22,GH10,CBM9	ACV09296.1
Jden_1726	CBM48,GH13	ACV09372.1
Jden_1729	CBM48,GH13	ACV09375.1
Jden_1926	GH13,CBM25,CBM25	ACV09568.1
Jden_1927	GH13,CBM25,CBM25,CBM25	ACV09569.1
Jden_1928	CBM22,GH43,CBM6	ACV09570.1
Jden_2228 (fragment)	GH5,CBM10,CBM10	ACV09865.1
Jden_2229	GH5,CBM2	ACV09866.1
Jden_2383	GH11,CBM2,CBM2	ACV10017.1
Jden_2447	CBM33,CBM2	ACV10079.1

LPMOs

Jden_1381 CBM33,CBM5,GH18

Jden_2447 CBM33,CBM2

Appendix C) Map and features of expression vector pET32-a(+)

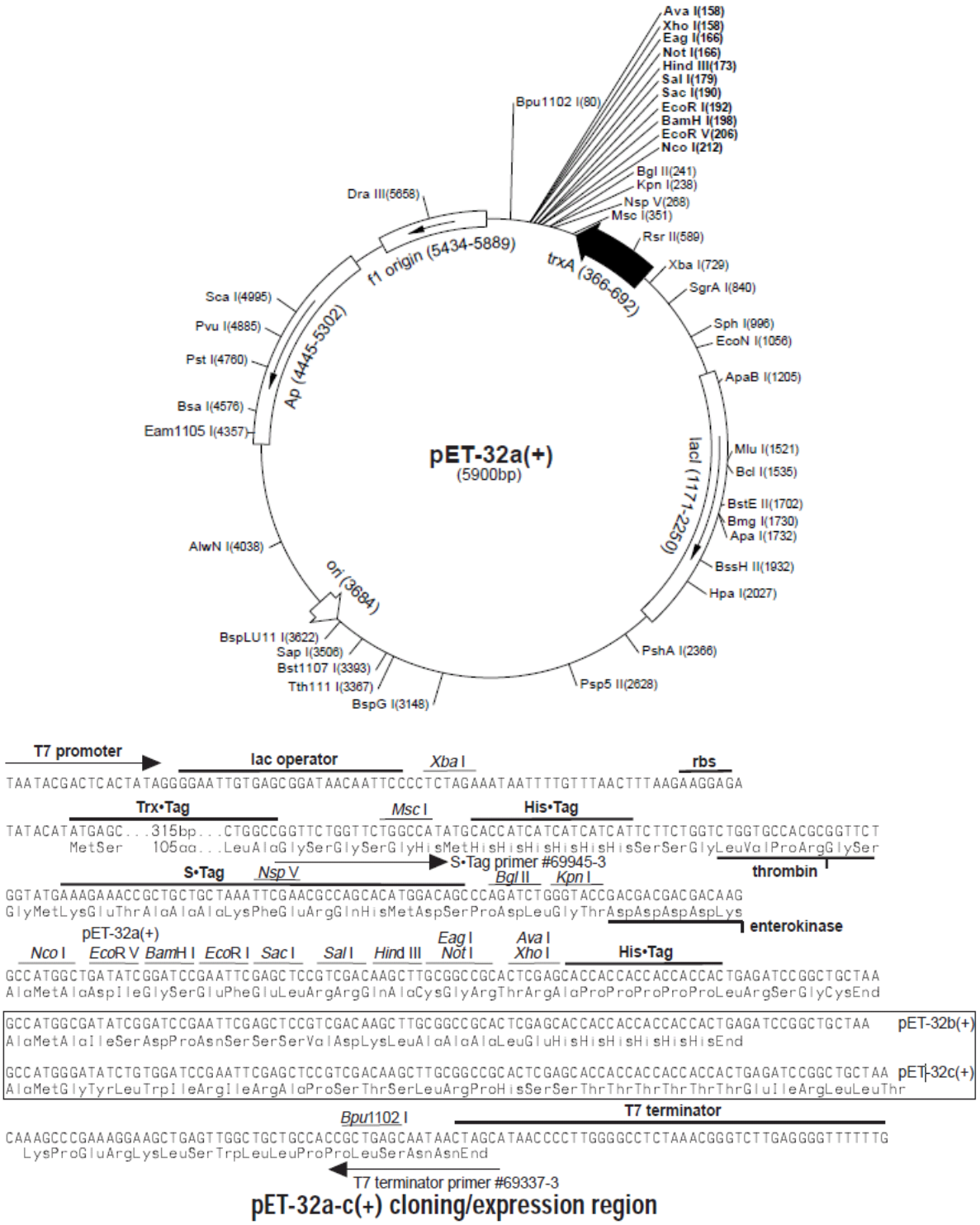
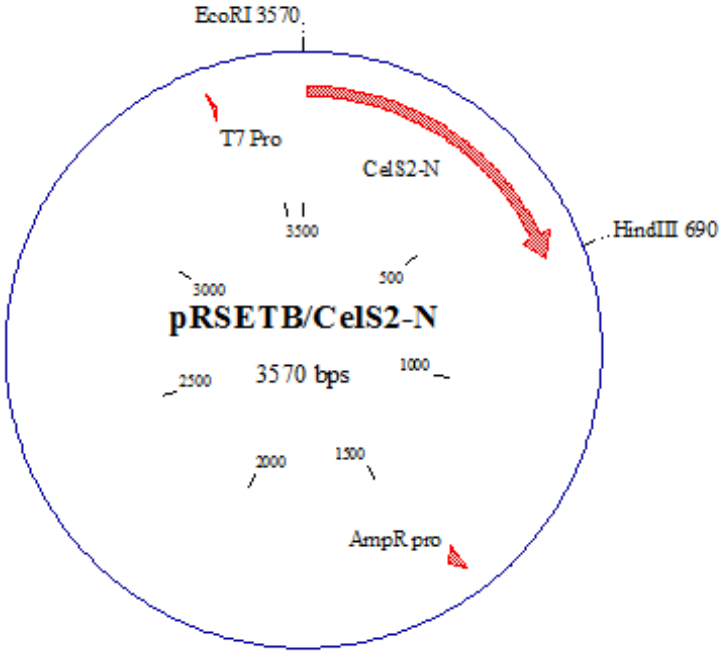


Fig x. Map and features of pET32a+. pET32 type expression vectors contain three versions, pET32a(+), pET32b(+) and pET32c(+). pET32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209. (Novagen, http://www.emdmillipore.com)

Appendix D) Plasmid construct of pRSETB-CelS2-N

(Plasmid map was made using CloneManager6)



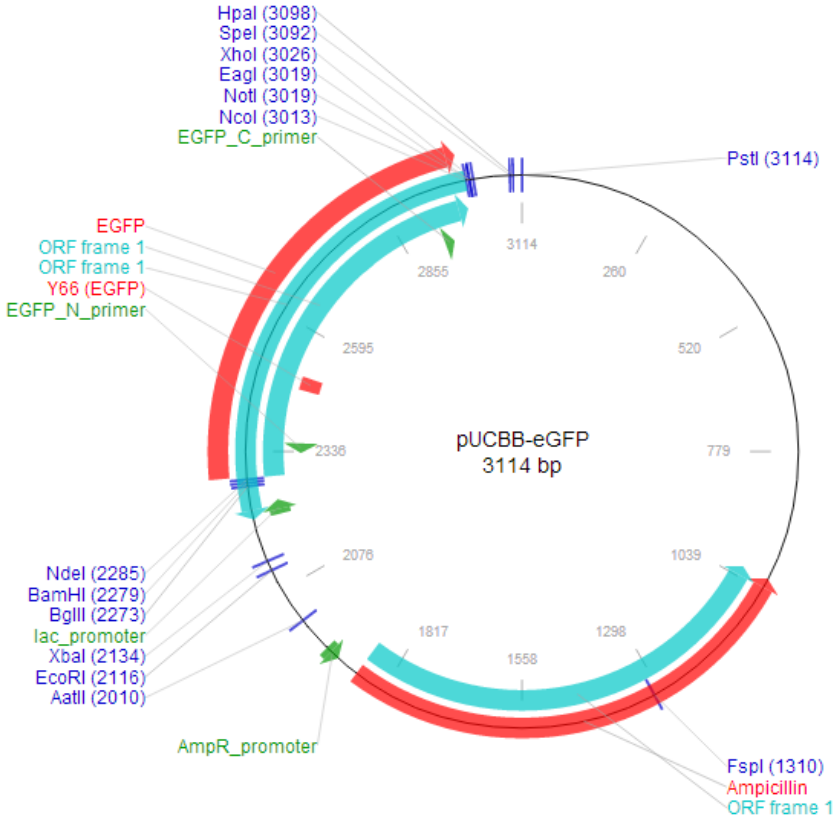
Molecule Features:

Type	Start	End	Name	Description
GENE	6	689	CelS2-N	
GENE	1419	1447	AmpR pro	
GENE	3364	3375	T7 Pro	

Enzymes (2 sites)

HindIII 690, EcoRI 3570,

Appendix E) Map of the pUCBB-eGFP expression vector (data taken from http://www.addgene.org)



Feature Name	Start	End
Ampicillin	1875	1015
AmpR_promoter	1945	1917
lac_promoter	2211	2240
EGFP	2290	3003
EGFP_N_primer	2353	2332
Y66 (EGFP)	2470	2499
EGFP_C_primer	2940	2961

ORF	Start	End
ORF frame 1	1875	1015
ORF frame 1	3012	2212
ORF frame 1	2287	3006

Enzyme Name	Cut
FspI	1310
AatII	2010
EcoRI	2116
XbaI	2134
BglII	2273
BamHI	2279
NdeI	2285
NcoI	3013
NotI	3019
EagI	3019
XhoI	3026
SpeI	3092
HpaI	3098
PstI	3114

Appendix F) Gene optimization for *Jden*1381 from *Jonesia denitrificans* DSM 20603

Original Gene, G-C content 58 %

>Jden_1381

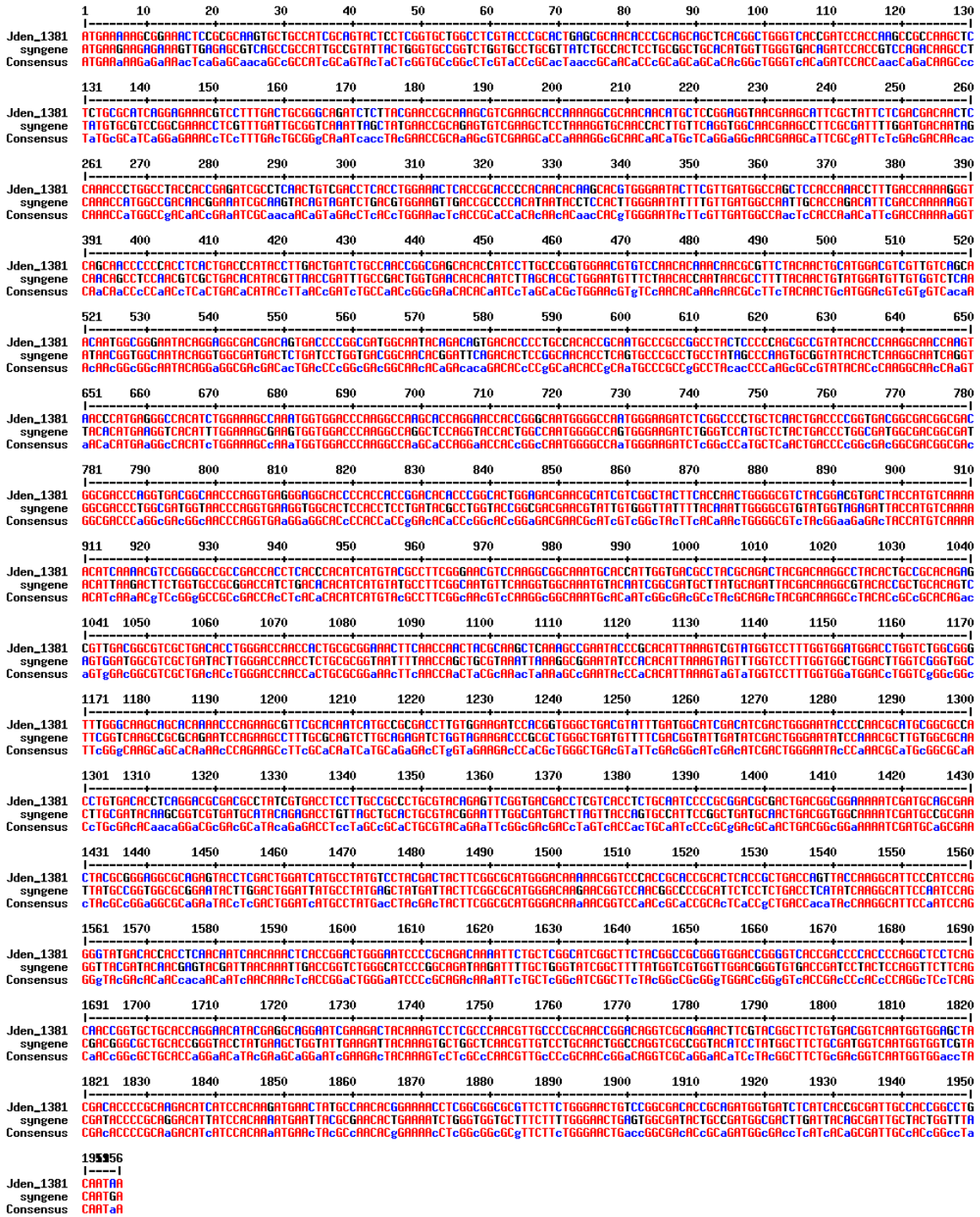
ATGAAAAAGCGGAAACTCCGCGCAAGTGCTGCCATCGCAGTACTCCTCGGTGCTGGCCTCGTACCCGCACTGAGCGCAACACC
CGCAGCAGCTCACGGCTGGGTCACCGATCCACCAAGCCGCCAAGCTCTCTGCGCATCAGGAGAAACGTCCTTTGACTGCGGGC
AGATCTCTTACGAACCGCAAAGCGTCGAAGCACCAAAAGGCGCAACAACATGCTCCGGAGGTAACGAAGCATTCGCTATTCTC
GACGACAACTCCAAACCCTGGCCTACCACCGAGATCGCCTCAACTGTCGACCTCACCTGGAAACTCACCGCACCCCACAACAC
AAGCACGTGGGAATACTTCGTTGATGGCCAGCTCCACCAAACCTTTGACCAAAAGGGTCAGCAACCCCCCACCTCACTGACCC
ATACCTTGACTGATCTGCCAACCGGCGAGCACACCATCCTTGCCCGGTGGAACGTGTCCAACACAAACAACGCGTTCTACAAC
TGCATGGACGTCGTTGTCAGCAACAATGGCGGGAATACAGGAGGCGACGACAGTGACCCCGGCGATGGCAATACAGACAGTGA
CACCCCTGCCACACCGCAATGCCCGCCGGCCTACTCCCCCAGCGCCGTATACACCCAAGGCAACCAAGTAACCCATGAGGGCC
ACATCTGGAAAGCCAAATGGTGGACCCAAGGCCAAGCACCAGGAACCACCGGGCAATGGGGCCAATGGGAAGATCTCGGCCCC
TGCTCAACTGACCCCGGTGACGGCGACGGCGACGGCGACCCAGGTGACGGCAACCCAGGTGAGGGAGGCACCCCACCACCGGA
CACACCCGGCACTGGAGACGAACGCATCGTCGGCTACTTCACCAACTGGGGCGTCTACGGACGTGACTACCATGTCAAAAACA
TCAAAACGTCCGGGGCCGCCGACCACCTCACCCACATCATGTACGCCTTCGGGAACGTCCAAGGCGGCAAATGCACCATTGGT
GACGCCTACGCAGACTACGACAAGGCCTACACTGCCGCACAGAGCGTTGACGGCGTCGCTGACACCTGGGACCAACCACTGCG
CGGAAACTTCAACCAACTACGCAAGCTCAAAGCCGAATACCCGCACATTAAAGTCGTATGGTCCTTTGGTGGATGGACCTGGT
CTGGCGGGTTTGGGCAAGCAGCACAAAACCCAGAAGCGTTCGCACAATCATGCCGCGACCTTGTGGAAGATCCACGGTGGGCT
GACGTATTTGATGGCATCGACATCGACTGGGAATACCCCAACGCATGCGGCGCCACCTGTGACACCTCAGGACGCGACGCCTA
TCGTGACCTCCTTGCCGCCCTGCGTACAGAGTTCGGTGACGACCTCGTCACCTCTGCAATCCCCGCGGACGCGACTGACGGCG
GAAAAATCGATGCAGCGAACTACGCGGGAGGCGCAGAGTACCTCGACTGGATCATGCCTATGTCCTACGACTACTTCGGCGCA
TGGGACAAAAACGGTCCCACCGCACCGCACTCACCGCTGACCAGTTACCAAGGCATTCCCATCCAGGGGTATGACACCACCTC
AACAATCAACAAACTCACCGGACTGGGAATCCCCGCAGACAAAATTCTGCTCGGCATCGGCTTCTACGGCCGCGGGTGGACCG
GGGTCACCGACCCCACCCCAGGCTCCTCAGCAACCGGTGCTGCACCAGGAACATACGAGGCAGGAATCGAAGACTACAAAGTC
CTCGCCCAACGTTGCCCCGCAACCGGACAGGTCGCAGGAACTTCGTACGGCTTCTGTGACGGTCAATGGTGGAGCTACGACAC
CCCGCAAGACATCATCCACAAGATGAACTATGCCAACACGGAAAACCTCGGCGGCGCGTTCTTCTGGGAACTGTCCGGCGACA
CCGCAGATGGTGATCTCATCACCGCGATTGCCACCGGCCTGCAATAA

***Jden*1381 synthetic gene; G-C content 51 %**

>Jden_1381 synthetic gene

ATGAAGAAGAGAAAGTTGAGAGCGTCAGCCGCCATTGCCGTATTACTGGGTGCCGGTCTGGTGCCTGCGTTATCTGCCACTCC
TGCGGCTGCACATGGTTGGGTGACAGATCCACCGTCCAGACAAGCCTTATGTGCGTCCGGCGAAACCTCGTTTGATTGCGGTC
AAATTAGCTATGAACCGCAGAGTGTCGAAGCTCCTAAAGGTGCAACCACTTGTTCAGGTGGCAACGAAGCCTTCGCGATTTTG
GATGACAATAGCAAACCATGGCCGACAACGGAAATCGCAAGTACAGTAGATCTGACGTGGAAGTTGACCGCCCCACATAATAC
CTCCACTTGGGAATATTTTGTTGATGGCCAATTGCACCAGACATTCGACCAAAAAGGTCAACAGCCTCCAACGTCGCTGACAC
ATACGTTAACCGATTTGCCGACTGGTGAACACACAATCTTAGCACGCTGGAATGTTTCTAACACCAATAACGCCTTTTACAAC
TGTATGGATGTTGTGGTCTCAAATAACGGTGGCAATACAGGTGGCGATGACTCTGATCCTGGTGACGGCAACACGGATTCAGA
CACTCCGGCAACACCTCAGTGCCCGCCTGCCTATAGCCCAAGTGCGGTATACACTCAAGGCAATCAGGTTACACATGAAGGTC
ACATTTGGAAAGCGAAGTGGTGGACCCAAGGCCAGGCTCCAGGTACCACTGGCCAATGGGGCCAGTGGGAAGATCTGGGTCCA
TGCTCTACTGACCCTGGCGATGGCGACGGCGATGGCGACCCTGGCGATGGTAACCCAGGTGAAGGTGGCACTCCACCTCCTGA
TACGCCTGGTACCGGCGACGAACGTATTGTGGGTTATTTTACAAATTGGGGCGTGTATGGTAGAGATTACCATGTCAAAAACA
TTAAGACTTCTGGTGCCGCGGACCATCTGACACACATCATGTATGCCTTCGGCAATGTTCAAGGTGGCAAATGTACAATCGGC
GATGCTTATGCAGATTACGACAAGGCGTACACCGCTGCACAGTCAGTGGATGGCGTCGCTGATACTTGGGACCAACCTCTGCG
CGGTAATTTTAACCAGCTGCGTAAATTAAAGGCGGAATATCCACACATTAAAGTAGTTTGGTCCTTTGGTGGCTGGACTTGGT
CGGGTGGCTTCGGTCAAGCCGCGCAGAATCCAGAAGCCTTTGCGCAGTCTTGCAGAGATCTGGTAGAAGACCCGCGCTGGGCT
GATGTTTTCGACGGTATTGATATCGACTGGGAATATCCAAACGCTTGTGGCGCAACTTGCGATACAAGCGGTCGTGATGCATA
CAGAGACCTGTTAGCTGCACTGCGTACGGAATTTGGCGATGACTTAGTTACCAGTGCCATTCCGGCTGATGCAACTGACGGTG
GCAAAATCGATGCCGCGAATTATGCCGGTGGCGCGGAATACTTGGACTGGATTATGCCTATGAGCTATGATTACTTCGGCGCA
TGGGACAAGAACGGTCCAACGGCCCCGCATTCTCCTCTGACCTCATATCAAGGCATTCCAATCCAGGGTTACGATACAACGAG
TACGATTAACAAATTGACCGGTCTGGGCATCCCGGCAGATAAGATTTTGCTGGGTATCGGCTTTTATGGTCGTGGTTGGACGG
GTGTGACCGATCCTACTCCAGGTTCTTCAGCGACGGGCGCTGCACCGGGTACCTATGAAGCTGGTATTGAAGATTACAAAGTG
CTGGCTCAACGTTGTCCTGCAACTGGCCAGGTCGCCGGTACATCCTATGGCTTCTGCGATGGTCAATGGTGGTCGTACGATAC
CCCGCAGGACATTATCCACAAAATGAATTACGCGAACACTGAAAATCTGGGTGGTGCTTTCTTTTGGGAACTGAGTGGCGATA
CTGCCGATGGCGACTTGATTACAGCGATTGCTACTGGTTTACAATGA

Alignment of the two genes; 76 % sequence identity



Appendix G) Nucleotide sequence of *celS2* (data obtained from Uniprot database)

The start and stop codon are underlined. The sequence coding for the signal peptide is black. The red sequence represents the coding sequence for the mature N-terminal LPMO domain of CelS2. The codons for the five conserved residues selected for mutations are colored blue. The sequence coding for the C-terminal CBM2 domain is colored green.

```
   1 atggttcgac gcaccagact cctcaccctc gcggcggtac tggccaccct gctcggctcg
  61 ctcggcgtga cccttctgct cgggcagggg cgggccgagg cgcacggcgt ggcgatgatg
 121 cccggctccc gcacctacct gtgccagctg gacgccaaga ccggcaccgg cgccctcgac
 181 ccgacgaacc ccgcctgcca ggccgccctc gaccagagcg gggcgacggc cctgtacaac
 241 tggttcgccg tgctcgactc caacgcgggg ggccgcggcg ccggttacgt gccggacggc
 301 accctgtgca gcgccggcga ccgttccccg tacgacttct ccgcctacaa cgccgcccgc
 361 tcggactggc cccgcacgca cctgacgtcg ggtgcgacga tcccggtgga atacagcaac
 421 tgggcggccc accccgggga cttccgggtg tacctgacca agccgggctg gtcgcccacg
 481 tccgagctgg gctgggacga cctggagctg atccagacgg tgaccaaccc gccccagcag
 541 ggctcgccgg gcaccgacgg gggccactac tactgggacc tcgcgctgcc ctcgggccgc
 601 tcgggcgacg cgttgatctt catgcagtgg gtgcgttcgg acagccagga gaacttcttc
 661 tcctgctcgg acgtcgtctt cgacggcggc aacggagagg tcaccggcat ccgcggttcc
 721 gggagcaccc cggacccgga cccgacaccg accccgacgg acccgaccac cccgcccacg
 781 cacaccggct cctgcatggc cgtgtactcg gtggagaact cctggagcgg cggcttccag
 841 gggtcggtcg aggtgatgaa ccacggcacc gagccgctga acggctgggc cgtgcagtgg
 901 cagccgggcg gcgggaccac gctcggcggg gtgtggaacg gttcgctgac cagcggctcc
 961 gacggtacgg tcacggtccg caacgtggac cacaaccgcg tcgtaccacc ggacgggagc
1021 gtgaccttcg gcttcaccgc cacttcgacg ggcaatgact tcccggtcga ctcgatcggc
1081 tgcgtggcac cctga
```

Appendix H) DNA ladders

1 Kb DNA ladder (Fermentas)

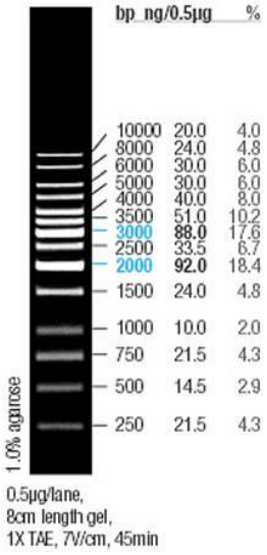


Fig H-1. 1 Kb DNA ladder used to identify DNA fragments. This DNA ladder is useful to determine DNA fragments from of size (250-10000) (Source; Fermentas).

GeneRuler^{TM} 1Kb Plus ladder

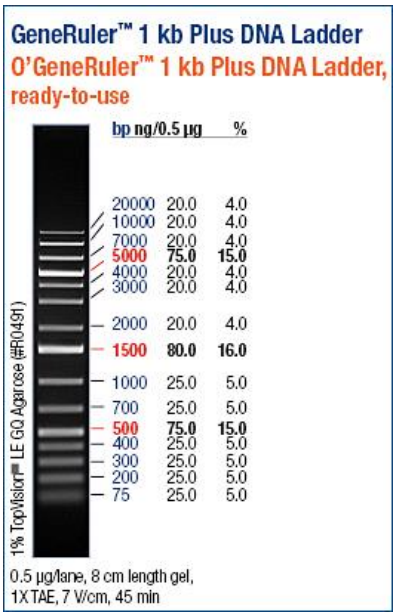
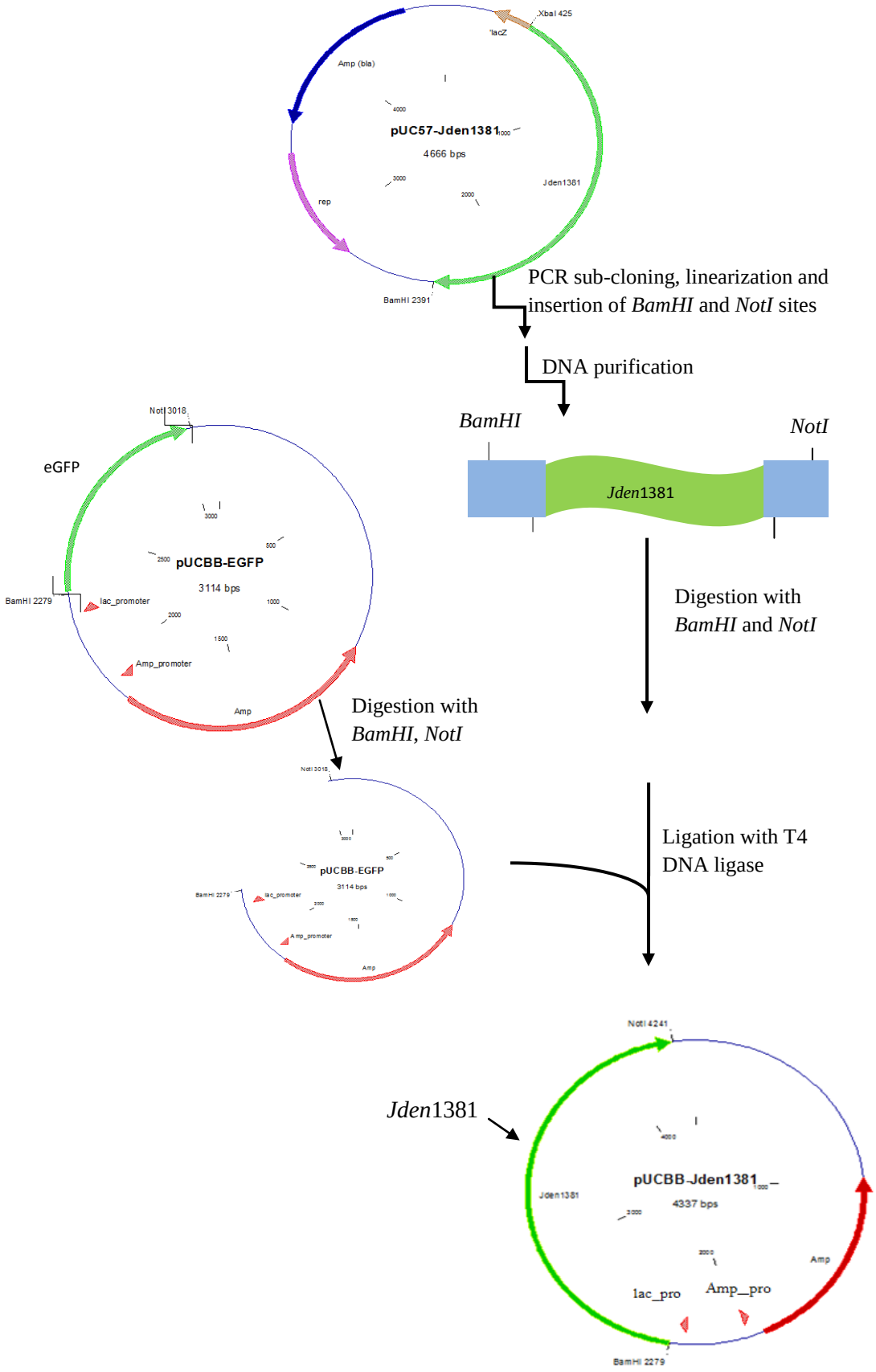


Fig H-2. 1Kb plus DNA ladder used to identify DNA fragments. This DNA contains DNA bands (75-20000bp), hebnce is useful to identify both small and large DNA fragments. (Source; Thermoscientific)

Appendix I) Cloning steps for *Jden*1381 in pUCBB vector (Plasmid map was made using CloneManager6)



Appendix J) Examples of chromatograms obtained during protein purification.

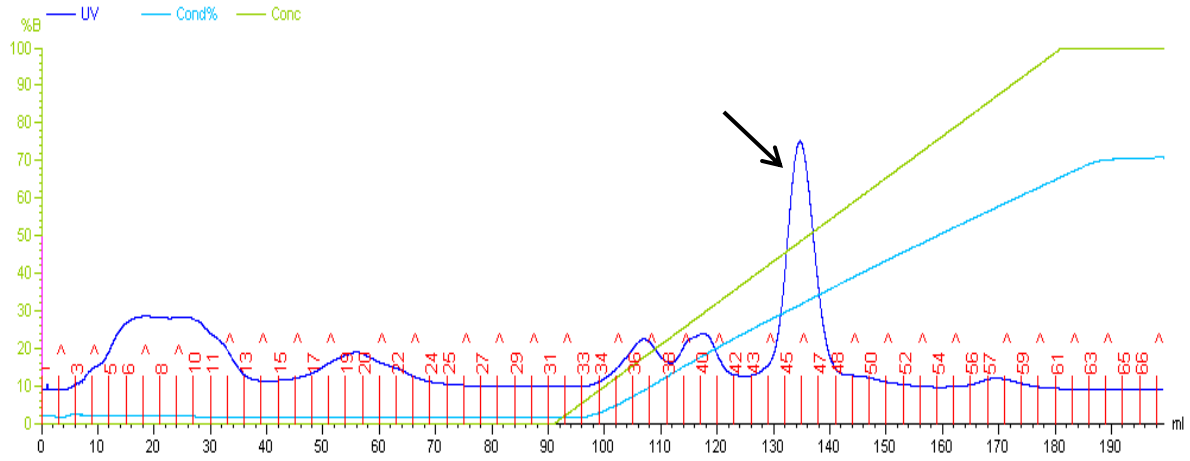


Fig. J-1 Ion exchange chromatography used for purification of Jden1381-LPMO. Fraction numbers are shown in red. The x-axes show volume of buffer passes through the column. The Y-axes show measurement of elution buffer concentration. The chromatogram show eluted proteins detected by online monitoring of absorption at 280 nm and collected in 3 ml fractions. Light blue line; measurement of conductivity used to follow column equilibration and salt gradient formation; green line, concentration measurement of elution buffer. Peak containing Jden1381-LPMO is indicated by arrow.

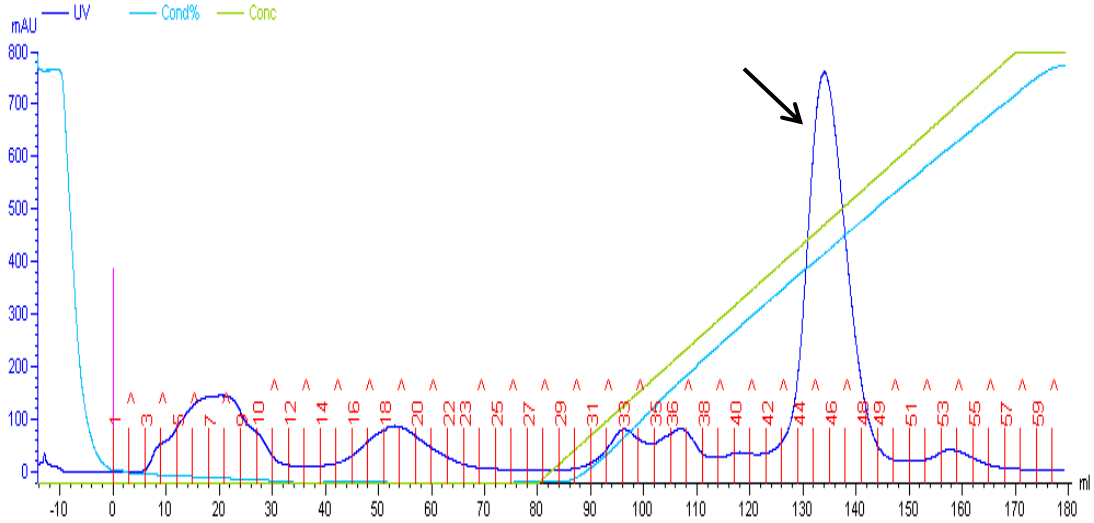


Fig. J-2 Ion exchange chromatography used for purification of Jden1381-LPMO-CBM5/12. Fraction numbers are shown in red. The x-axes show volume of buffer passes through the column. The Y-axes show measurement of UV absorbance at 280nm (blue). The chromatogram show eluted proteins detected by online monitoring of absorption at 280 nm and collected in 3 ml fractions. Light blue line; measurement of conductivity used to follow column equilibration and salt gradient formation; green line, concentration measurement of elution buffer. Peak containing Jden1381-LPMO-CBM5/12 is indicated by arrow.

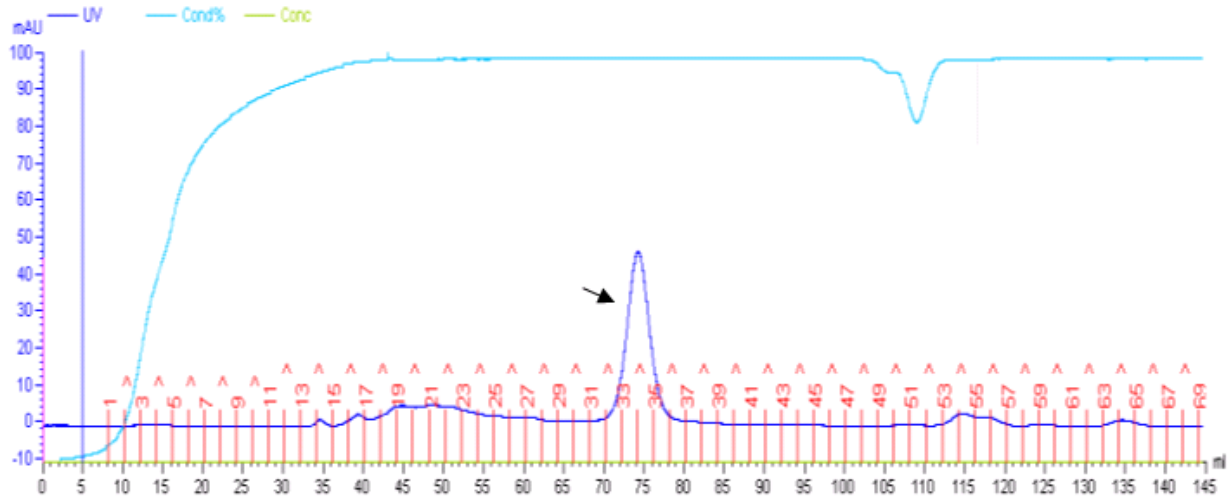


Fig.J-3. Size exclusion purification for Jden1381-LPMO. Fraction numbers are shown in red. The x-axes show volume of buffer passes through the column. The Y-axes show measurement of UV absorbance at 280nm (blue). The chromatogram show eluted proteins detected by online monitoring of absorption at 280 nm and collected in 5 ml fractions. Light blue line; measurement of conductivity used to follow column equilibration and salt gradient formation. Peak containing Jden1381-LPMO is indicated by arrow.

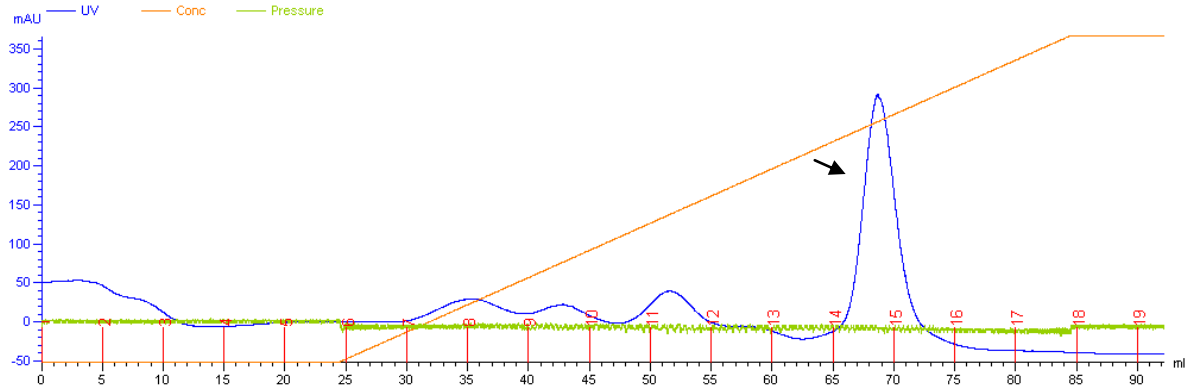
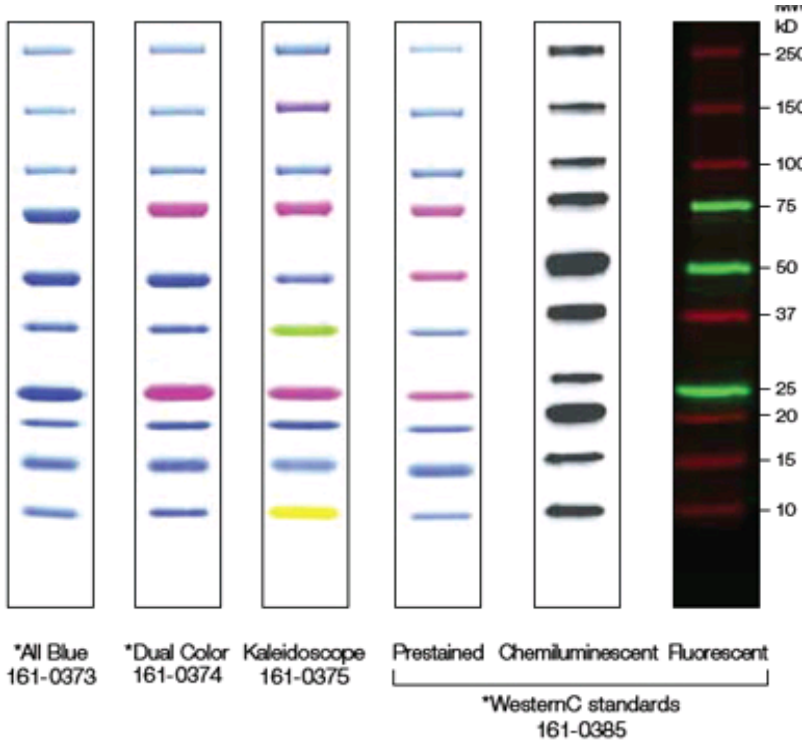


Fig. J-4. Chromatogram from HisTrap purification of Jden1381-LPMO-CBM5_C-His$_6$. Fraction numbers are shown in red. The x-axes show volume of buffer passes through the column. The Y-axes show measurement of UV absorbance at 280nm (blue). The chromatogram show eluted proteins detected by online monitoring of absorption at 280 nm and collected in 5 ml fractions. Brown line, concentration of elution buffer; green line, measurement of pressure. Peak containing Jden1381-LPMO-CBM5/12_C-His$_6$ is indicated by arrow.

Appendix K) Protein marker used in this study.

In this study,, protein identification was performed by SDS-Gel analysis using Banch marker protein ladder from Bio.Rad.

250 kDa protein ladder (Bio-Rad)



Appendix L) Amino acid sequence of CelS2, full length

```
        10         20         30         40         50         60
MVRRTRLLTL AAVLATLLGS LGVTLLLGQG RAEAHGVAMM PGSRTYLCQL DAKTGTGALD

        70         80         90        100        110        120
PTNPACQAAL DQSGATALYN WFAVLDSNAG GRGAGYVPDG TLCSAGDRSP YDFSAYNAAR

       130        140        150        160        170        180
SDWPRTHLTS GATIPVEYSN WAAHPGDFRV YLTKPGWSPT SELGWDDLEL IQTVTNPPQQ

       190        200        210        220        230        240
GSPGTDGGHY YWDLALPSGR SGDALIFMQW VRSDSQENFF SCSDVVFDGG NGEVTGIRGS

       250        260        270        280        290        300
GSTPDPDPTP TPTDPTTPPT HTGSCMAVYS VENSWSGGFQ GSVEVMNHGT EPLNGWAVQW

       310        320        330        340        350        360
QPGGGTTLGG VWNGSLTSGS DGTVTVRNVD HNRVVPPDGS VTFGFTATST GNDFPVDSIG


CVAP
```

Fig x Amino acid sequence of celS2, full length . Residues 1-34 signal peptide; 35-227, mature CBM33 type LPMO and 265-360, CBM2 domain. (Source: uniprot and pfam databases, accession number: Q9RJY2)

Appendix M) Amino acid sequence of Jden1381

```
        10         20         30         40         50         60
MKKRKLRASA AIAVLLGAGL VPALSATPAA AHGWVTDPPS RQALCASGET SFDCGQISYE

        70         80         90        100        110        120
PQSVEAPKGA TTCSGGNEAF AILDDNSKPW PTTEIASTVD LTWKLTAPHN TSTWEYFVDG

       130        140        150        160        170        180
QLHQTFDQKG QQPPTSLTHT LTDLPTGEHT ILARWNVSNT NNAFYNCMDV VVSNNGGNTG

       190        200        210        220        230        240
GDDSDPGDGN TDSDTPATPQ CPPAYSPSAV YTQGNQVTHE GHIWKAKWWT QGQAPGTTGQ

       250        260        270        280        290        300
WGQWEDLGPC STDPGDGDGD GDPGDGNPGE GGTPPPDTPG TGDERIVGYF TNWGVYGRDY

       310        320        330        340        350        360
HVKNIKTSGA ADHLTHIMYA FGNVQGGKCT IGDAYADYDK AYTAAQSVDG VADTWDQPLR

       370        380        390        400        410        420
GNFNQLRKLK AEYPHIKVVW SFGGWTWSGG FGQAAQNPEA FAQSCRDLVE DPRWADVFDG

       430        440        450        460        470        480
IDIDWEYPNA CGATCDTSGR DAYRDLLAAL RTEFGDDLVT SAIPADATDG GKIDAANYAG

       490        500        510        520        530        540
GAEYLDWIMP MSYDYFGAWD KNGPTAPHSP LTSYQGIPIQ GYDTTSTINK LTGLGIPADK

       550        560        570        580        590        600
ILLGIGFYGR GWTGVTDPTP GSSATGAAPG TYEAGIEDYK VLAQRCPATG QVAGTSYGFC

       610        620        630        640        650
DGQWWSYDTP QDIIHKMNYA NTENLGGAFF WELSGDTADG DLITAIATGL Q
```

Fig. x Amino acid sequence of Jden1381, full length. Amino acid residues 1-31 Signal peptides, 32-171 CBM33 type LPMO domain, 202-244 CBM5/12 domain and 284-636 GH18 domain. (Source: pfam database, accession number: C7R4I0)

Appendix N) Raw data from experiment for determination the time course for oxidized product generation by $CelS2_{S215A}$ and $CelS2_{WT}$

Table N-1. The table shows the detected oxidized chitobiose (Glc-GlcA/DP2-ox) generated by $CelS2_{S215A}$ and $CelS2_{WT}$ from three parallel reactions. Both soluble products; those released to the solvent and those remain trapped in the solid phase (referred as "trapped") were quantified. The area of the peak is used for quantification. The mean value for the three parallel reactions was for all products and standard deviation (St.dev) was calculated for all reactions. Fig 21 was constructed based on this data. (See Section 4.4.3 for more detail)

	Soluble	Trapped	Soluble	Soluble	Trapped	
$CelS_{WT}$			Mean	St.dev	Mean	St.dev
Minutes/parallel	Peak area (nC*min)	Peak area (nC*min)	Peak area (nC*min)		Peak area (nC*min)	
0	0	0	0	0	0	0
30-1	0,4608	8,788	0,258533	0,175906	9,0217	0,728815333
30-2	0,1413	8,4384				
30-3	0,1735	9,8387				
60-1	0,8699	5,3152	0,565967	0,290408	6,089833	2,558564149
60-2	0,2913	8,9462				
60-3	0,5367	4,0081				
120-1	0,7352	9,663	0,514233	0,224479	9,020333	1,086652761
120-2	0,2864	7,7657				
120-63	0,5211	9,6323				
240-1	5,7682	10,9173	3,494867	2,032675	11,542	0,774376504
240-2	1,8525	11,3003				
240-3	2,8639	12,4084				
360-1	6,3432	16,518	5,418833	0,810435	18,30947	3,099274691
360-2	5,083	21,8882				
360-3	4,8303	16,5222				
	Soluble	Trapped				
$CelS2_{S215A}$			Mean	Stdev	Mean	Stdev
Minutes/parallel	Peak area (nC*min)	Peak area (nC*min)	Peak area (nC*min)		Peak area (nC*min)	
0	0	0	0	0	0	0
30-1	0,8878	4,0379	0,7813	0,096088	3,573767	1,427940073
30-2	0,7011	1,9715				
30-3	0,755	4,7119				
60-1	3,4624	7,5844	2,188767	1,104726	5,673633	1,66313284
60-2	1,4902	4,5517				
60-3	1,6137	4,8848				
120-1	3,6672	3,363	3,186467	0,765762	2,604467	0,781806129
120-2	2,3034	1,8013				
120-63	3,5888	2,6491				
240-1	3,8993	6,4221	3,091533	0,705473	5,118767	1,261875578
240-2	2,7789	5,0313				
240-3	2,5964	3,9029				
360-1	2,7857	6,1484	2,9849	0,546161	4,848667	1,442985954
360-2	2,5663	5,1017				
360-3	3,6027	3,2959				