Norwegian University
of Life Sciences

Master's Thesis 2017 60 ECTS
Faculty of Chemistry, Biotechnology, and Food Science (KBM)

Rational Engineering of a Family 6 Glycoside Hydrolase

Heidi Østby
Biotechnology

Rational Engineering of a Family 6 Glycoside Hydrolase

Master's Thesis

Heidi Østby

Protein Engineering and Proteomics Group
Department of Chemistry, Biotechnology, and Food Science
Norwegian University of Life Sciences

2017

Table of Contents

Acknowledgements

The research for this thesis was completed at the Faculty of Chemistry, Biotechnology, and Food Science at the Norwegian University of Life Sciences, with Dr. Gustav Vaaje-Kolstad, Marianne S. Jensen, and Dr. Lasse Fredriksen as my advisors.

I am beyond grateful to my main thesis advisor, Dr. Gustav Vaaje-Kolstad, whose dynamic and instructive advising kept me committed to and passionate about this project through all its various stages. His feedback was always full of useful detail, and reflected his inspiring dedication to scientific research. Professor Vincent G. H. Eijsink is the mentor who ignited my passion for protein engineering. I have tremendous respect for him as a researcher and as an advisor – despite his busy schedule, he always made time to check in with me concerning the progression of my project, and invariably inspired my research in new directions.

I have had the great fortune of having both Gustav and Vincent as teachers during my time at NMBU, and owe my interest in protein science to their contagious enthusiasm for the field.

I could not have completed this thesis without the close guidance of my advisor Marianne S. Jensen, who helped me enormously every step of the way, following each phase of my research incredibly closely while simultaneously (miraculously) finding the time to complete critical work on her PhD. When I felt lost, I always knew I could turn to Marianne, and she would take the time to demonstrate a confusing technique or clarify a vexing question – for which I will always be grateful. Her accuracy and work ethic in the lab inspired me to work harder and taught me to strive for precision and excellence in my own work.

Dr. Lasse Fredriksen always kept life in the lab positive and fun, cheering me up when I made seemingly irreversible mistakes. Lasse taught me to see the learning potential in all wrong turns, reminding me that research is first and foremost a series of lessons learned by trial and error. A brilliant researcher, Lasse is extremely knowledgeable and accomplished but humble to a fault, and has explained innumerable difficult methods to me over the past year.

In addition to my main advisors, every member of the PEP group always answered my countless questions and took time out of his or her busy schedule to teach theoretical and

practical aspects of numerous complex laboratory methods. I will forever be grateful for all the knowledge they have given me this year.

Dr. Heidi Rudi, my academic advisor, is in many ways the reason I came to NMBU, and has supported me through several complicated academic processes throughout my time here. She immediately responded to my many frustrated emails about transferring credits across international school systems with patience, clarity, and incredibly helpful expertise. Helping me navigate the difficult process of transferring to NMBU from an American university, she made my life much easier and was always available for advising whenever I needed it.

My fellow master's students, Ingrid Heggenes and Bjørnar Flatin, were a constant source of motivation and kindness throughout this past year. Thank you for listening and for suggesting new ways in which I could better my project, and for the hundreds of moments of laughter shared over iced coffee in our study room.

I am immeasurably thankful for the unconditional love and support from my family. My desire to understand how things work at an extraordinarily detailed level comes from my father, and my recognition that there is beauty and creativity in science comes from my mother. Both my parents have supported me through every single step of my academic career, and have always pushed me to follow my dreams. They taught me that what matters is making a difference in the world, whatever your passion or field. Thank you to my inspirational sisters, Marie and Hanna, who show me every day what it means to stay true to your interests and dreams. Your humor and love have managed to pick me up every time I stumbled.

I am grateful to all my friends at NMBU for making my time here joyful and memorable, especially Bjørnar Flatin and Martine Dall'Osso Teigset. Bjørnar, thank you for your positivity and humor through every study session over the past four years. I have so much admiration for your genuine interest and compassion for those around you. Thank you for never once letting me doubt my intelligence or knowledge. Martine, I will always consider myself enormously fortunate to have randomly ended up in the same apartment as you four years ago – little did I know that I would be making a life-long friend. Thank you for your endless kindness and warmth that have comforted me every time I needed it.

Finally, to my boyfriend, Eddie Shankar: four years of 17,000 km distance has not been easy, but it never would have been possible without your positivity, strength, and belief in our future. Thank you for teaching me to not sweat the small stuff, and for putting up with the ridiculous time difference that would drive any person insane. Our relationship has not been like most others, and I know that your unwavering encouragement in the face of my insecurities surrounding the future have helped me learn to believe in myself and my abilities. I cannot express how much your love and support has meant to me. Thank you from the bottom of my heart.

Ås, May 11th 2017

Heidi Østby

Abstract

Plant-based lignocellulosic biomass represents a significant global reserve of organic renewable energy, and is considered to have high potential to replace substantial portions of fossil-fuel dependency. Sugars derived from the degradation of lignocellulosic biomass can be utilized in the production of more environmentally-friendly products such as biofuels and numerous other value-added compounds. However, the degradation of the cellulose polysaccharide is complicated by its highly recalcitrant nature resulting from strong intermolecular hydrogen bonds and cross-linkages with other components such as lignin.

A variety of bacteria and fungi have evolved the ability to degrade cellulose through the use of cellulases, β-glucosidases, and lytic polysaccharide monooxygenases (LPMOs). Such enzymes are thus the focus of intense research for use in biorefineries that aim to convert woody biomass to soluble sugars. Biorefineries utilize a variety of methods, including enzymatic degradation, to purify and isolate various components from biomass. These industrial processes are often characterized by relatively extreme conditions such as high temperatures and acidic or alkaline pH. There is therefore great interest in improving the properties of cellulases for use under these conditions. One approach is the engineering of cellulases with the aim of altering specific properties such pH stability and/or pH optimum.

The main aim of this study was to use rational protein engineering in an attempt to optimize a family 6 glycoside hydrolase (GH), mgCel6A, for use at pH 5.0 or lower for more than 24 hours at 60°C or higher. The mgCel6A wild-type enzyme was extensively characterized to determine its pH optimum and degree of pH stability on the industrial BALI™ cellulose substrate, as well as on model substrates.

Mutant design aimed at optimizing the pH optimum and stability of mgCel6A was guided by analysis of a homology model of the catalytic domain of the enzyme (mgCel6AΔCBM), a multiple sequence alignment (MSA) of biochemically characterized family 6 GHs, and studies of pH-engineering literature. Mutations were introduced on the surface of the enzyme and in or near the active site. Produced mgCel6AΔCBM enzyme variants (Q152E, L165E, N21D/N46D/A104E/Q152E/L165E/I242E/I271D/Q276E, T44D, D158A, and D158N) were

characterized to investigate their pH optima and stability as compared to the wild-type enzyme.

Of the six successfully produced mutant enzyme variants, three appeared to have identical stabilities and pH optima compared to that of mgCel6AΔCBM. One mutant variant with additional negative surface charge (N21D/N46D/A104E/Q152E/L165E/I242E/I271D/Q276E) showed a significant decrease in thermal stability. Two mutants lacking an aspartate presumed to raise the pK_a of the catalytic acid (D158A and D158N) showed considerable acidic shifts in their pH-activity profiles. Characterization of these mutants indicated that they had obtained pH optima of approximately 5.0 and 5.5, respectively, for 24-hour reactions at 60°C with the BALI™ cellulose substrate. However, the activities of these mgCel6AΔCBM variants were drastically reduced as compared to that of the wild-type enzyme, showing that further optimization of these mutants is most likely needed for optimal performance under industrial conditions.

Future mutagenesis studies of the D158A and D158N mutants may provide a deeper understanding of the reduced catalytic activity, and may also help in designing modified versions of these mutants with preserved reduced pH optima and concomitantly increased activity. A combination of rational engineering and directed evolution may be advantageous in the further development of mgCel6AΔCBM towards industrial use.

Sammendrag

Plantebasert lignocellulosisk biomasse utgjør en betydelig global reserve av organisk fornybar energi, og antas å ha stort potensiale til å erstatte mye av det fossile brennstoffet vi i dag er svært avhengige av. Sukker som utvinnes ved nedbryting av lignocellulosisk biomasse kan brukes i produksjonen av mer miljøvennlige produkter slik som biodrivstoff og flere andre foredlede produkter. Nedbryting av cellulose er imidlertid vanskelig grunnet sterke intermolekylære hydrogenbindinger og kryss-linker med andre komponenter slik som lignin.

En rekke typer bakterier og sopp har utviklet evnen til å bryte ned cellulose ved hjelp av cellulaser, β-glucosidaser og lytiske polysakkaridmonooksygenaser (LPMOer). Det forskes derfor intenst på disse enzymene for bruk i bioraffinerier hvor målet er å konvertere biomasse til løselige sukkere. Bioraffinerier benytter en rekke metoder, inkludert enzymatisk nedbryting, for å rense og isolere ulike komponenter i biomassen. Disse industrielle prosessene karakteriseres ofte av relativt ekstreme forhold, slik som høye temperaturer og sur eller basisk pH. Det er derfor av stor interesse å kunne forbedre cellulasenes egenskaper til bruk under slike forhold. En tilnærmingsmåte er å utvikle cellulaser med den målsetting å endre spesifikke egenskaper som pH-stabilitet og/eller pH-optimum.

Hovedformålet med denne oppgaven har vært å bruke rasjonell protein engineering i et forsøk på å optimalisere en familie 6-glykosid hydrolase (GH), mgCe16A, for bruk ved pH 5.0 eller lavere og ved 60°C eller høyere, i mer enn 24 timer. mgCel6A ble karakterisert for å bestemme villtypeenzymets pH-optimum og grad av pH-stabilitet på det industrielle BALI™ cellulose-substratet og på modellsubstrater.

Mutantdesign med mål om å optimalisere enzymets pH-optimum og stabilitet var basert på analyse av en homologimodell av det katalytiske domenet av enzymet (mgCel6AΔCBM), en multippel sekvenssammenstilling av biokjemiske karakteriserte familie 6-GHer, og pH-engineering litteratur. Mutasjoner ble introdusert både på overflaten av enzymet, samt i og i nærheten av det aktive setet. Produserte mgCel6AΔCBM mutanter (Q152E, L165E, N21D/N46D/A104E/Q152E/L165E/I242E/I271D/Q276E, T44D, D158A, og D158N) ble deretter testet for å undersøke deres pH-optimum og stabilitet sammenliknet med villtypeenzymet.

Tre av seks produserte mutanter viste like stabiliteter og pH-optima som villtypeenzymet. En av mutantvariantene av mgCe16AΔCBM med flere negative ladninger på overflaten (N21D/N46D/A104E/Q152E/L165E/I242E/I271D/Q276E) viste en betydelig reduksjon i termostabilitet. To andre mutanter som manglet en aspartat (antatt å øke pK_a verdien til den katalytiske syren; mutanter D158A og D158N), fikk betydelig lavere pH-optima. Karakterisering av disse mutantene indikerte at de hadde pH-optima på rundt 5.0 og 5.5 i en 24-timers reaksjon ved 60°C på BALI™ substratet. Imidlertid var aktiviteten til disse mutantvariantene av mgCel6AΔCBM betydelig redusert i forhold til villtypeenzymet. Resultatene indikerer at videre optimalisering av disse mutantvariantene er nødvendig for optimal ytelse ved industrielle betingelser.

Fremtidige mutagenesestudier av mutantene D158A og D158N kan gi en dypere forståelse av den reduserte katalytiske aktiviteten, og kan også være til hjelp i design av modifiserte versjoner av disse mutantene med tilsvarende redusert pH-optimum og økt aktivitet. En kombinasjon av rasjonell design og styrt evolusjon (directed evolution) kan være fordelaktig i videre utvikling av mgCel6AΔCBM for industrielle forhold.

Abbreviations

$A_{260}/A_{280}/A_{540}$	Absorbance of ultraviolet light at 260/280/540 nanometers
Å	Ångström (0.1 nanometers)
bp	Base Pair
BSA	Bovine Serum Albumin
CAZy	Carohydrate-Active Enzyme
CBH	Cellobiohydrolase
CBM	Carbohydrate-Binding Module
CBM2	Family 2 Carbohydrate-Binding Module
CMC	Carboxymethyl Cellulose
dGTP	Deoxyguanosine Triphosphate
dH_2O	Milli-Q® Sterile Water
DNA	Deoxyribonucleic Acid
DNS	3,5-Dinitrosalicylic Acid
dNTP	Deoxynucleoside Triphosphates
DSC	Differential Scanning Calorimetry
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
EG	Endoglucanase
GH	Glycoside Hydrolase
GH6	Family 6 Glycoside Hydrolase
GlcX (1-6)	Glucose (Glc1), Cellobiose (Glc2), Cellotriose (Glc3), Cellotetraose (Glc4),

	Cellopentaose (Glc5), and Cellohexaose (Glc6)
HPAEC-PAD	High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection
ICS	Ion Chromatography System
IMAC	Immobilized Metal Ion Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilodalton
LB	Lysogeny Broth
LIC	Ligation-Independent Cloning
LPMO	Lytic Polysaccharide Monooxygenase
mgCel6A	Family 6 Glycoside Hydrolase from a Metagenome (mg)
mgCel6AΔCBM	Truncated Variant of mgCel6A, lacking the Carbohydrate-Binding Module (CBM)
MSA	Multiple Sequence Alignment
MWCO	Molecular Weight Cut-off
Ni-NTA	Nickel (Ni^{2+})-Nitrilotriacetic Acid
nm	Nanometers
PASC	Phosphoric-Acid Swollen Cellulose
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

S.O.C.	Super Optimal Broth with Catabolite Repression
SOE	Splicing by Overlap Extension
TAE	Tris-Acetate-EDTA
TB	Terrific Broth
TGS	Tris-Glycine-SDS
UV	Ultraviolet
v/v	Volume/Volume
w/v	Weight/Volume

1. Introduction

1.1. The Need for Novel Energy Sources

Cellulose is the most abundant organic polymer found on Earth, and can be considered a significant renewable resource in the progressive global move towards generating more environmentally-friendly products (Klemm et al. 2005; Payne et al. 2015). The current heavy reliance on fossil fuels as the principal global energy source for transportation, generation of electricity, heating, and chemical production is a finite solution, and the discovery and development of alternative, sustainable energy sources are therefore of great interest. Currently, oil represents the primary energy source for the transportation industry, and for the production of chemicals, plastics, and synthetic fibers (Cherubini 2010). The detrimental effects of the emission of greenhouse gases resulting from the combustion of fossil fuels are well-documented, and novel energy sources with the potential to reduce or eliminate these emissions are in high demand (Oreskes 2004). Renewable plant-based biomass has been the focus of intense research for several decades due to its potential to reduce industrial fossil fuel dependency. Sugars derived from the degradation of plant-based biomass can be fermented into ethanol or butanol, which can then be distilled and used as biofuels (Cherubini 2010; Horn et al. 2012). In addition, these sugars and other derivatives of the breakdown of biomass can also be used to produce various value-added products with a large variety of applications (Ferreira-Leitão et al. 2010).

1.2. Biomass

Biomass is a general term comprising organic material derived from plant or animal matter. Numerous types of biomass exist, and their compositions vary depending on the starting material from which they are derived. The term lignocellulosic biomass refers to biomass from various plant sources rich in lignin, cellulose, and hemicellulose. The amount of each main constituent varies, but lignocellulosic biomass is generally thought to be made up of 40-50% cellulose, 20-40% hemicellulose, and 20-30% lignin, forming a tightly-packed co-polymeric matrix (***Figure 1.2.1***). In addition to these three main components, lignocellulosic biomass also contains minor amounts of proteins, lipids, soluble sugars, and minerals (Horn et

al. 2012). Cellulose is one of the main constituents of the cell wall in plants, and, in combination with lignin, contributes to the high tensile strength that enables plants to maintain their rigidity. In addition to plants, cellulose can also be found to a lesser extent in other organisms such as some bacteria, fungi, algae, invertebrates, and amoeba (Habibi et al. 2010).

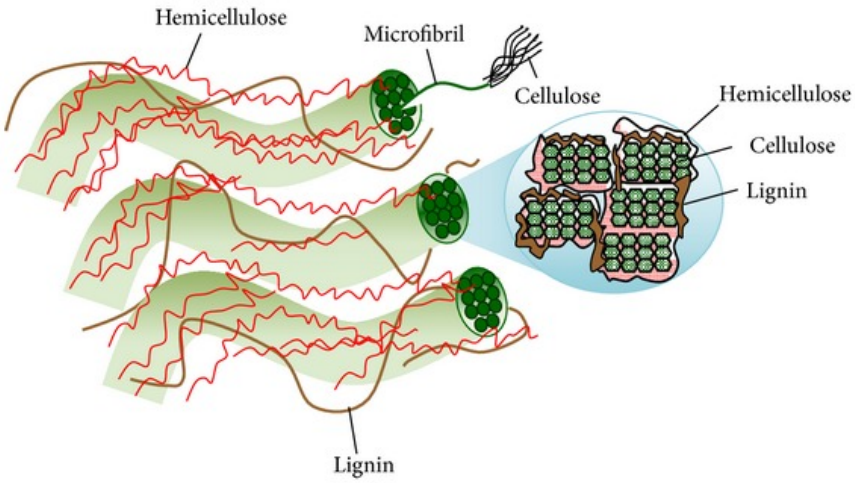


Figure 1.2.1. Major Components of the Plant Cell Wall. The figure shows the three principal components of the plant cell wall: lignin, hemicellulose, and cellulose. The cellulose chains form microfibrils (see Section 1.2.3) localized in a cross-linked matrix with hemicellulose and lignin. The cross-section shows the packing of cellulose microfibril structures and the surrounding lignin and hemicellulose. Figure obtained from Lee et al. (2014).

1.2.1. Lignin

Lignin, thought to make up 20-30% of lignocellulosic biomass, is a highly complex aromatic polymer recognized for its heterogeneity (de Gonzalo et al. 2016; Horn et al. 2012). Lignin consists of three main monolignols, guaiacyl, p-hydroxyphenyl, and syringyl, which are linked together to form a highly branched network. The extensive branching of this polymer makes it exceedingly resistant to both biological and chemical degradation (Abdel-Hamid et al. 2013). It also enables lignin to contribute a significant tensile strength and rigidity to the plant cell wall (de Gonzalo et al. 2016). In plant-based biomass, lignin is usually cross-linked with cellulose and hemicellulose. While certain forms of hemicellulose can be relatively easy to degrade enzymatically (Horn et al. 2012) (see Section 1.2.2), cellulose contains numerous highly crystalline regions which prevent its efficient enzymatic depolymerization (Alberts et

al. 2014) (see Section 1.2.3). The cross-linkages between cellulose and lignin in biomass thus further increase the recalcitrance of this polysaccharide.

1.2.2. Hemicellulose

Hemicelluloses comprise a large group of non-cellulose polysaccharides that show a great amount of variation in their structures. Nevertheless, a common feature amongst all hemicelluloses is the β(1→4) linked units of D-xylose, D-mannose, or D-glucose that comprise the polysaccharide backbone. Examples of hemicellulose include xylan, mannan, glucomannan, and xyloglucan (Horn et al. 2012; Scheller & Ulvskov 2010). The significant variation in hemicellulose structure includes varying amounts of branching between the diverse members of this polysaccharide group. While hemicelluloses are generally considered to be more easily degraded than cellulose (see Section 1.2.3), highly complex branched hemicellulose regions can contribute to biomass recalcitrance due to the interlinking of these components in plant cell walls. For this reason, enzymatic cocktails used in the industrial degradation of lignocellulosic biomass often include hemicellulase enzymes which break down these heterogeneous polysaccharides into fermentable sugars (Gírio et al. 2010; Horn et al. 2012).

1.2.3. Cellulose

Cellulose is a linear polysaccharide consisting of hundreds to thousands of repeating β(1→4)-linked D-glucose units in a linear chain. The repeating glucose units are covalently linked via *O*-glycosidic bonds from the equatorial OH group on C4 of one glucose monomer to the C1 atom of the adjacent glucose monomer (***Figure 1.2.2***). These glucose units exist in the chair conformation, with their substituent hydroxyl groups in equatorial positions and the substituent hydrogen atoms in axial positions (Klemm et al. 2005). The orientation of these groups enables the cellulose molecule to form strong hydrogen bonding networks within single cellulose chains as well as between adjacent chains (Himmel et al. 2007). This extensive hydrogen bonding network allows cellulose chains to form highly ordered crystalline structures known as cellulose microfibrils (Alberts et al. 2014; Chami Khazraji & Robert 2013; Sharma & Yazdani 2016). The extensive interaction patterns make crystalline cellulose microfibrils extremely recalcitrant to industrial enzymatic hydrolysis (Himmel et al.

2007). Every second monomer in the cellulose chain is rotated 180° to allow for the preferred conformation of the bond between the C4 oxygen atom and the C1 carbon atom. Thus, the repeating unit in cellulose is a dimer or disaccharide, also known as cellobiose. One end of the linear cellulose chain contains an OH-group at C4: this is known as the non-reducing end. The other end of the cellulose molecule contains a free OH-group on C1 (***Figure 1.2.2***). This OH-group is in equilibrium with the aldehyde form, and is thus capable of acting as a reducing agent. This end of the cellulose chain is therefore known as the reducing end (Klemm et al. 2005). Cellulose is most commonly isolated from wood pulp, and can be treated thermochemically, biochemically, chemically, and/or mechanically in numerous steps to depolymerize the polysaccharide into its monomers (Section 1.2.4) (Cherubini 2010).

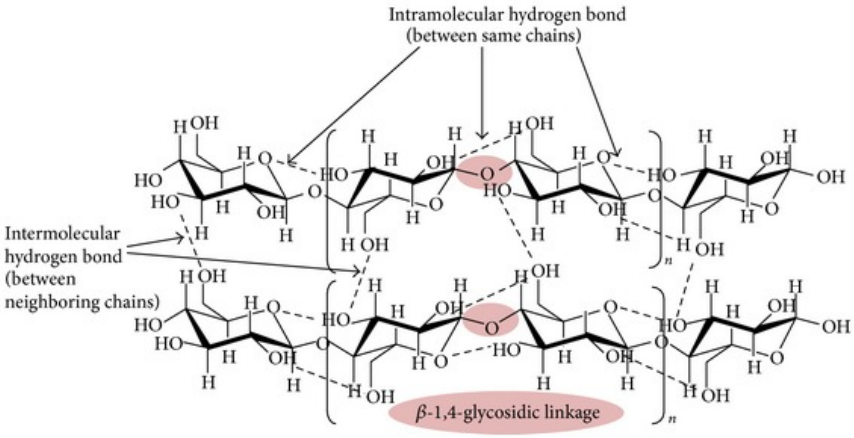


Figure 1.2.2. Chemical Structure of the Cellulose Molecule. The figure shows the β(1→4) glycosidic bonds between individual D-glucose monomers, as well as the extensive hydrogen bonding that occurs both within cellulose chains and between adjacent chains. The reducing (right) and non-reducing (left) ends of the cellulose chains are also shown. Figure obtained from Lee et al. (2014).

1.2.4. Industrial Processing of Biomass

The potential role of lignocellulosic biomass in replacing fossil fuels is highly dependent on efficient and systematic conversion into its fundamental components. Biorefineries specialize in these efficient degradations, and apply a variety of chemical, biological, and mechanical approaches to purify and isolate various components from biomass for use in the production of biofuels and other value-added products (Cherubini 2010; Ferreira-Leitão et al. 2010).

Borregaard AS, a private Norwegian company, represents an example of a modern biorefinery that produces sustainable biochemicals and biofuels from lignocellulosic biomass. As part of this conversion, Borregaard utilizes a BALI™ pre-treatment and separation process to convert the starting material, wood pulp derived from Norway spruce, *Picea abies*, into high-value lignin, ethanol, and other products (Rødsrud et al. 2012). ***Figure 1.2.3*** shows an overview of the BALI™ process.

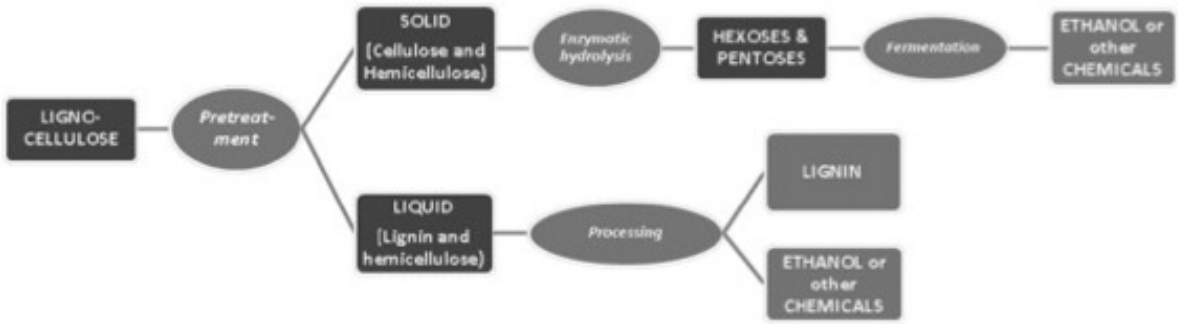


Figure 1.2.3. The Borregaard BALI™ Process. The figure shows a schematic description of the BALI™ process employed at Borregaard AS. The pretreatment step to the left comprises cooking steps and the use of sulfite (to solubilize lignin). The solid cellulose and hemicellulose fraction make up the BALI™ cellulose substrate used in this study. Figure obtained from Rødsrud et al. (2012); reprinted with permission from Elsevier.

The initial steps of the BALI™ process include chemical pretreatment and fractionation, where the lignin in the lignocellulose-based starting material is made water-soluble and cellulose crystallinity is reduced. The two resulting fractions, one liquid and one solid, contain lignin and soluble hemicellulose, and cellulose and non-soluble hemicellulose, respectively. The liquid fraction is processed in order to degrade the hemicellulose to its constituent monosaccharides, and to separate the lignin. The lignin can then be used as a performance chemical in many different Borregaard products, while the hemicellulose-derived monosaccharides are fermented into ethanol or other products. The solid fraction, consisting of the pretreated cellulose and non-soluble hemicellulose, make up the BALI™ substrate. At Borregaard AS, this solid fraction is enzymatically treated, resulting in fermentable sugars (Rødsrud et al. 2012).

1.3. Bacterial and Fungal Degradation of Plant Polysaccharides

Various bacteria and fungi have evolved enzymatic machineries for the efficient degradation of nutrient-rich biomass into its constituent monomers. Here we will focus on enzyme systems for the conversion of cellulose, which is abundant and highly recalcitrant. Cellulolytic enzyme machineries consist of three main enzyme classes which act synergistically to efficiently catalyze the depolymerization of cellulose (***Figure 1.3.1***).

Currently, there are three distinct systems by which cellulolytic organisms enzymatically catalyze the degradation of cellulose. The first involves cellulosomes, large multi-enzymae complexes found on the organism's cell surface (Krause et al. 2003; Payne et al. 2015). The second system by which cellulolytic organisms degrade cellulose is via "free" enzymes, referring to cellulases that are secreted as single catalytic units, which may or may not be attached to one or more carbohydrate-binding modules (CBMs) via linker domains (Section 1.4). The third system involves a so-called polysaccharide-utilization locus (PUL), identified in certain members of the Bacteriodetes phylum. PULs encode proteins that enable bacteria to bind to their substrate, degrade it, and transport enzymatic hydrolysis products into the cell. This system has been shown to function on various carbohydrate substrates, including cellulose (Naas et al. 2014).

Cellulases represent a large, diverse class of enzymes that catalyze the hydrolysis of cellulose molecules into shorter oligosaccharide units. Cellulases can act processively or non-processively, attaching within and/or from the ends of the cellulose chains (Section 1.5.1). Lytic polysaccharide monooxygenases (LPMOs) utilize an oxidative mechanism to disrupt internal, highly crystalline regions of cellulose, thus generating novel chain ends to which processive exo-acting cellulases can bind (Horn et al. 2012; Vaaje-Kolstad et al. 2010). Finally, β-glucosidases convert cellobiose disaccharides to the final hydrolysis product, D-glucose (Section 1.5.1).

These three classes of enzymes are produced by several types of bacteria and fungi that break down cellulose into glucose, which may serve as their main source of carbon.

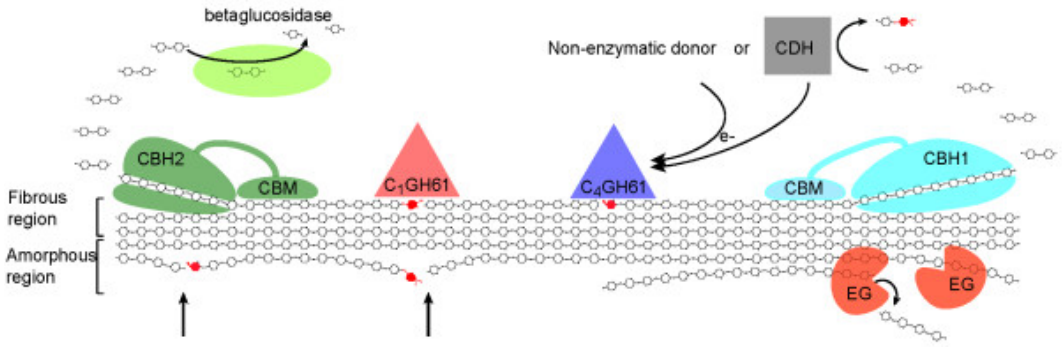


Figure 1.3.1. Representation of the Synergistic Action of cellulases, LPMOs, and β-glucosidases in the Enzymatic Degradation of Cellulose. The figure shows the concomitant action of free cellulase enzymes (cellobiohydrolases (CBHs) and endoglucanases (EGs), see Section 1.5.1), β-glucosidases, and LPMOs (represented by the red and blue triangles). The CBHs are shown with carbohydrate-binding modules (CBMs). The arrows in the lower left portion of the figure represent novel chain ends generated by the action of LPMOs upon which cellulases can act (oxidized sugars are shown in red). Figure obtained from Horn et al. (2012).

1.4. Classification of Carbohydrate-Active Enzymes

The enormous diversity of enzymes involved in the degradation of plant cell wall polysaccharides and other carbohydrates presented the need for a large-scale classification system. In 1999, Henrissat and colleagues constructed the Carbohydrate-Active Enzyme (CAZy) Database, which serves as an up-to-date list of enzymes known to catalyze the hydrolysis, modification, or formation of glycosidic bonds. The current enzymatic classes described in the CAZy database include glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activity enzymes (AAs).

Within the glycoside hydrolase class of carbohydrate-active enzymes, there are currently 135 families listed in the CAZy database (Lombard et al. 2014). Traditional classification systems categorized GH enzymes based on the preferred substrate upon which they act. However, several glycoside hydrolases display relatively broad substrate specificities, making it difficult to classify these based solely on this criterion (Naumoff 2011). Studies in the evolutionary mechanisms of proteins have confirmed that structure tends to be more conserved than sequence (Lesk 2010). The current GH classification system employed by CAZy therefore classifies GHs into families based on both amino acid sequence and structural fold similarity,

allowing for a more systematic understanding of potential evolutionary relationships between glycoside hydrolases from different organisms. In addition, the CAZy classification system is performed "module by module." This means that distinctive domains of the same GH enzyme may be placed into different families, depending on their most probable evolutionary origin. The CAZy database also further groups some families into "clans," which assemble GH families with similar tertiary structures, hydrolysis mechanisms, and catalytic residues. Glycoside hydrolases in the same clan are considered to share a common evolutionary history.

In addition, the database contains a list of known carbohydrate-binding modules (CBMs), which are non-catalytic domains that have substrate-binding properties (Boraston et al. 2004; Lombard et al. 2014). CBMs are generally thought to aid in the correct placement of the catalytic domain on the substrate, and thus increase the overall concentration of productively-bound enzyme on the surface of the substrate to be hydrolyzed. These binding domains appear to have a minimal effect on the catalytic mechanism, and instead appear to simply assist the enzyme catalytic domain in adsorption onto the substrate (Várnai et al. 2013).

1.5. Cellulases

1.5.1. Cellulose Degradation by Cellulases

Cellulases, which catalyze the hydrolysis of glycosidic bonds, include endoglucanases (EGs) and cellobiohydrolases (CBHs). CBHs are processive enzymes that attach to one end of a cellulose molecule and cleave glycosidic bonds along the chain, producing cellobiose. EGs, which can be both processive and non-processive, hydrolyze glycosidic bonds within the cellulose chain (Payne et al. 2015). In addition to the highly ordered crystalline regions of cellulose discussed above, less-ordered amorphous areas have also been described. These typically occur in microfibril regions with high amounts of internal strain (Sharma & Yazdani 2016). Endoglucanases are thought to act on these amorphous regions, cleaving glycosidic bonds to produce chain ends upon which processive CBHs can act to produce cellobiose. Finally, β-1,4-glucosidases cleave the cellobiose glycosidic bond, forming glucose. While β-glucosidases play a critical role in the breakdown of cellulose to its monosaccharide constituent glucose, they do not act directly on the cellulose substrate, and can therefore not

strictly be considered cellulases. These three categories of enzymes act synergistically in order to catalyze the breakdown of cellulose (Horn et al. 2012).

The difference in the mechanism by which EGs and CBHs operate is reflected in the structure of the active sites of these enzymes (***Figure 1.5.1***). Endoglucanases have an open groove-like active site which enables them to hydrolyze *O*-glycosidic bonds in the middle of a cellulose molecule, creating new chain ends. Exo-acting CBHs, however, contain tunnel-like active sites that seem optimized for binding at the end of a cellulose molecule and hydrolyzing glycosidic bonds in the chain dimer by dimer (Sukharnikov et al. 2011).

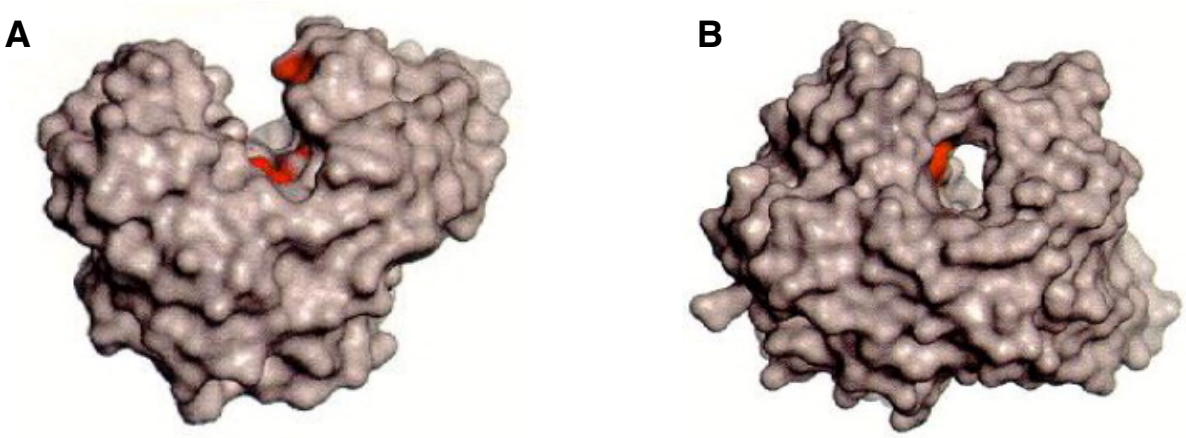


Figure 1.5.1. Active Site Topologies of endoglucanases (EGs) and cellobiohydrolases (CBHs). The figure shows the difference in structure of the active sites of EGs and CBHs, reflected in their mechanism of action (Section 1.5.1). **A:** The open groove-like cleft typical of endoglucanases (Cel6A from *Thermobifida fusca*). **B:** The tunnel active site typical of cellobiohydrolases (Cel6A from *Hypocrea jecorina*). Proposed catalytic amino acids are shown in red (see Section 1.5.3.1 for details). Figure modified from Davies and Henrissat (1995); reprinted with permission from Elsevier.

1.5.2. Catalytic Proficiency of Glycoside Hydrolases

As discussed in Section 1.2.3, the cellulose microfibril structure is extremely recalcitrant to degradation due to the extensive hydrogen bonding networks found between adjacent cellulose chains. In addition to this strong intermolecular bonding, the *O*-glycosidic bonds within the cellulose molecule add another dimension to the polysaccharide's resistance to hydrolysis. *O*-glycosidic bonds have been shown to be extraordinarily stable, 2-4 times more so than the bonds between nucleotides and amino acids in DNA and proteins respectively. A 1998 study by Wolfenden *et al.* estimated that *O*-glycosidic bonds have an uncatalyzed half-

life of nearly 5 million years at 25°C from pH 7 to pH 14 (Wolfenden et al. 1998). This makes GHs astoundingly efficient enzymes, since they can increase the rate of hydrolysis of an *O*-glycosidic bond by a factor of up to 10^{17}, making their catalytic efficiency nearly unparalleled in the enzyme world. In fact, glycoside hydrolases are the most efficient known enzymatic catalysts that do not utilize a cofactor for catalysis (Payne et al. 2015).

1.5.3. Glycoside Hydrolase Catalytic Mechanisms

While much remains to be understood about glycosidic bond hydrolysis, glycoside hydrolase enzymes are thought to function via two main mechanisms, which involve inversion or retention of the configuration at the anomeric carbon (C1), respectively (André et al. 2003). *O*-glycosidic bond hydrolysis occurs via general acid catalysis. This type of reaction requires an acid/proton donor and a base/nucleophile, or, in terms of the amino acids, two carboxylic acids, nearly always in the form of aspartates or glutamates (Davies & Henrissat 1995; Rye & Withers 2000).

Regardless of whether the mechanism proceeds via inversion or retention, the protonated residue must always be located within hydrogen-bonding distance to the oxygen atom in the substrate glycosidic bond. The principal difference between the two hydrolysis mechanisms lies in the location of the residue acting as a nucleophile (retention) or base (inversion) (***Figure 1.5.2***). In an enzyme operating via the inverting catalysis mechanism, the two catalytic residues are approximately 10 Å apart, with one acting as an acid and the other as a base. In retaining enzymes, the catalytic acid residue can act as both an acid and a base, and is located approximately 5 Å away from a nucleophilic residue which forms an enzyme-substrate intermediate as part of the hydrolysis mechanism (Davies & Henrissat 1995; Vuong & Wilson 2010). In inverting GHs, the second carboxylic acid is therefore further away from the substrate than in an enzyme operating via the retaining mechanism. This is due to the fact that inverting enzymes require the activation of a water molecule, which must fit between the carboxylic residue acting as a base and the substrate. After activation by the catalytic base, this water molecule attacks the anomeric carbon while the oxygen in the glycosidic bond becomes protonated by the catalytic acid, resulting in a single-displacement type mechanism. The glycosidic bond is broken, the aglycone departs, and the configuration of the anomeric carbon-oxygen bond is inverted.

In a retaining-type mechanism, however, the second carboxylic acid is located close enough to the substrate to directly attack the anomeric carbon, and the glycosidic oxygen becomes protonated by the catalytic acid residue in the same way as in the inverting mechanism, forming the aglycone (***Figure 1.5.2***). The novel enzyme-substrate bond has the opposite configuration at the substrate anomeric carbon as the starting C1-glycosidic oxygen bond. While the aglycone is free to depart at this stage, the enzyme remains bonded to the substrate. The deprotonated catalytic acid now acts as a base, and activates an incoming water molecule that hydrolyzes the enzyme-substrate bond, while, once again, inverting the anomeric configuration. This mechanism is referred to as a double-displacement, with the net result being a retention of the configuration at the substrate anomeric carbon (Davies & Henrissat 1995; Rye & Withers 2000).

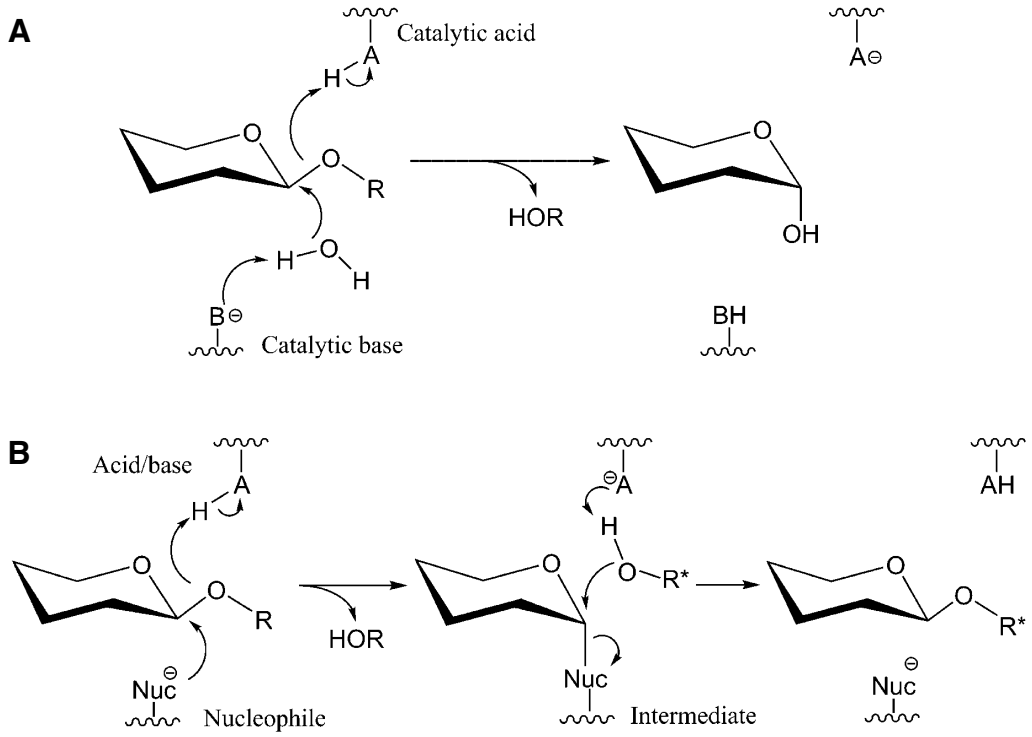


Figure 1.5.2. Inverting and Retaining Mechanisms of Glycoside Hydrolases. Panel A: The inverting mechanism, where the catalytic base (B-) residue is thought to activate a water molecule which attacks the anomeric carbon of the substrate. The departing aglycone (ROH) acquires its proton from the catalytic acid (AH). **Panel B:** The retaining double-displacement mechanism, which occurs in two central steps. The departing aglycone is formed in the first step when the nucleophilic residue attacks the anomeric carbon of the substrate and the glycosidic oxygen acquires a proton from the catalytic acid. In the second step, the deprotonated catalytic acid residue activates a nearby water molecule (R*OH) which attacks the anomeric carbon, breaking the substrate-enzyme bond. Figure modified from Vuong and Wilson (2010); reprinted with permission from John Wiley and Sons.

In addition to these two main mechanisms, novel mechanisms or variations thereof have been found to operate in certain families, including GHs that lack a catalytic acid residue, GHs that depend on a portion of the substrate to act as the catalytic nucleophile, GHs that use residues other than aspartate and glutamate as their catalytic residues, and GHs that require various cofactors to hydrolyze their substrate (Vuong & Wilson 2010; Withers & Williams 2013). There are also a few families of GHs where the mechanism of catalysis is not yet fully understood.

1.5.3.1. Catalytic Mechanism of Family 6 Glycoside Hydrolases (GH6s)

The GH6 family represents an example of a family of GHs whose catalytic mechanism is not fully understood. While these enzymes are known to lead to inversion of the anomeric configuration, and the identity of the catalytic acid is well-established, the catalytic base has proved far more difficult to identify (Piens & Davies 2015). Early studies of GH6s from *Thermobifida fusca*, *Hypocrea jecorina*, and *Cellulomonas fimi* speculated in the roles of two conserved aspartates acting as catalytic bases (Damude et al. 1995; Rouvinen et al. 1990). However, mutation of these residues did not abolish enzymatic activity, signifying that these residues are incapable of serving as the catalytic base in a classical single-displacement inverting mechanism. 1999 and 2002 studies of Cel6A from *Thermobifida fusca* and *Hypocrea jecorina* (previously known as *Trichoderma reesei*), respectively, considered that GH6s may not operate with a traditional catalytic base residue (Koivula et al. 2002; Wolfgang & Wilson 1999). In 2003, André *et al*. proposed the presence of a complex consisting of a water molecule hydrogen bonded to three amino acids and to the substrate glycosidic oxygen atom, which enables the proton transfer from the catalytic acid to the substrate glycosidic oxygen. The catalysis mechanism based on this proposed complex involves a conformational change in the enzyme upon substrate binding, bringing a conserved aspartate in close proximity to the substrate. According to the proposed mechanism, this residue is then able to act as a catalytic base by abstracting a proton from another water molecule, which it then transfers to the catalytic acid (André et al. 2003). More recent studies on GH6 mechanisms, however, have proposed that these enzymes do not require a catalytic base, and that both nucleophilic attack of the substrate and reprotonation of the catalytic acid following substrate hydrolysis can occur via the transfer of a proton through a "water wire" of several water molecules (Mayes et al. 2016).

While it has proven challenging to determine the exact identity of the catalytic base in the GH6 catalytic mechanism, another conserved aspartate has been studied in detail, and has been shown in numerous GH6s to raise the pK_a value of the catalytic acid (Damude et al. 1995; Koivula et al. 1996; Wolfgang & Wilson 1999). The pK_a value represents an acid's dissociation constant, and signifies whether an acid will donate or accept a proton at a given pH value. When the pH of the solution is lower than the acid's pK_a value, the acid will be protonated. The pK_a value of the catalytic acid thus plays a critical role in determining the pH-dependency of the activity of cellulases, including the GH6 family (see Section 1.6.1).

1.6. Protein Engineering

Protein engineering is a tool commonly used in modern biotechnology to modify various characteristics of proteins. The overall goal is often to develop proteins that are more suitable for use under specific pre-determined conditions. To efficiently alter specific physiochemical properties of enzymes through engineering, it is first imperative to define the enzyme's optimal conditions for activity, as well as to understand the ways and mechanisms through which proteins fold and maintain their stability.

1.6.1. Enzyme Optima

Throughout the course of evolution, organisms have adapted to the various environmental conditions of their respective habitats. As a result, the extracellular enzymes produced by these organisms have also evolved to tolerate these circumstances. Such circumstances include specific temperatures, pHs, salt and other ion concentrations, and more. Natural enzymes can therefore be expected to have various "optima" related to these factors, since enzymes need to produce a maximum amount of product, while also retaining stability over a prolonged period of time. In this sense, an enzyme's optimum for a factor is a balance between efficient catalysis and maintaining stability.

In order to understand the mechanisms by which a protein retains its stability over time, it is first necessary to understand the processes by which proteins fold and acquire their natural three-dimensional structures.

1.6.2. Protein Folding and Stability

Protein folding is a complex topic that has intrigued scientists for decades. Research has shown that protein folding appears to be a compromise between a "trial and error" method, and a continuation of "already established" strong interactions between various residues or domains. In short, proteins appear to somewhat randomly search through numerous conformations of the polypeptide chain, allowing even distant residues to interact, and then hierarchically "build upon" the strongest interactions to adopt the most stable overall structure. The protein is only said to be in its final native structure when all interactions, both those within distinct domains, and those between domains, have been formed (Dobson 2003; Loladze et al. 1999).

The mechanisms of protein folding and the ways in which proteins stabilize their three-dimensional structures are processes that have yet to be fully understood. Under physiological conditions, unfolded polypeptide chains with the same amino acid sequence fold into the same three-dimensional conformation: this structure is referred to as the native state. Naturally, this folding goes against the second law of thermodynamics, namely that a system's total entropy will increase over time. In order to compensate for the instability that this loss of entropy confers, folded proteins maintain certain properties that stabilize their ordered native states. These properties include the burial of hydrophobic, non-polar residues in the protein core, the satisfaction of the hydrogen-bonding potential of polar residues, and the compact packing of amino acids in the protein center, all of which encourage the protein to correctly fold into its native state (Lesk 2010).

In addition to these characteristics, folded proteins also have large numbers of defined interactions between specific amino acids that stabilize their native states. These interactions include hydrogen bonds, salt bridges, disulfide bridges, electrostatic interactions, van der Waals interactions, and many more (Jaenicke 1991). Many proteins also bind to cofactors, which stabilize their native states even further. Through these interactions and folding patterns, proteins are able to surmount the loss of entropy conferred by constraints of the polypeptide chain. The resulting conformational stability is only marginal, however, and represents the energetic equivalent of only a few hydrogen bonds. This implies that relatively minor (engineered) changes in proteins may have large effects on stability.

The protein sequence of amino acids is the main determinant in the final native state structure, including the structure of the essential active site, where enzymes perform their unique catalysis mechanisms. In general, the main chain "backbone" of proteins is very flexible, and allows for secondary structure elements, such as α-helices, β-sheets, and loops, to take form during the folding process (Lesk 2010). The twenty common amino acids have varying biochemical properties, including their polarity, length of the side chain, charge, and rigidity. Notably, two particular amino acids stand out in terms of rigidity: proline and glycine. Proline's sidechain is bonded to the backbone chain, forming a pyrrolidine ring that severely restricts the flexibility of the main chain at proline residues. Glycine, on the other hand, is the only amino acid that lacks a side chain: instead, it has only a single hydrogen atom. This makes glycine the most flexible residue, which is why it is commonly found in protein loops (Krieger et al. 2005).

While stability is an exceedingly broad term, in protein science it denotes a protein's ability to maintain its three-dimensional structure in the face of exposure to extreme conditions such as significant deviations in temperature, pH, and salt concentration, or exposure to chemicals such as alcohols, acids, or bases (Jaenicke 1991). Once a protein has been destabilized, it will generally become denatured, and this denaturation process is seldom reversible (Daniel et al. 1996; Lesk 2010). The elements of protein structures that contribute to the stabilization of their three-dimensional native states are therefore of particular interest in the field of protein chemistry.

1.6.3. A Brief History of Protein Engineering

Novel techniques that can be used to investigate folding mechanisms, intramolecular interactions, and stability within proteins are constantly being developed and improved. One such technique, protein engineering, aims to enhance various enzymatic properties such as stability, selectivity, and processivity, or to alter an enzyme's characteristic optima, in order to efficiently modify an enzyme for use in a specific catalytic process. The field of protein engineering first arose in the 1980s and 1990s, and has evolved considerably over the past three decades in tandem with breakthroughs in molecular biology, bioinformatics, biotechnology and structural biology.

In the early days of enzymology, many catalytic mechanisms were already well-understood, and the developing discipline of X-ray crystallography was beginning to generate a plethora of enzymatic three-dimensional structures. Unfortunately, there was no precise or effective method to determine the roles of specific amino acids in an enzyme's proposed catalytic mechanism (Brannigan & Wilkinson 2002). A significant breakthrough in enzymology came with the development of site-directed mutagenesis, which allowed for the modification of specific amino acids in a protein sequence through mutation of single codons in the gene that encodes the protein (Carter 1986). This technique forms the basis of one of the two main approaches to engineering an enzyme with novel properties, namely rational design.

In rational protein design, knowledge of the protein sequence and structure, as well as biophysical properties of the specific enzyme to be engineered, are used to determine candidate amino acids that may be altered. In theory, this knowledge, in combination with a general knowledge of factors affecting protein stability, folding, interaction and substrate-binding, should allow for the generation of novel enzymes with a desired characteristic property (Bornscheuer & Pohl 2001; Kaushik et al. 2016). In practice, however, detailed structural data is not always available, which can pose a significant challenge to the design of rational mutations. In addition, the effect of a specific mutation on protein folding and other properties may not always be predictable, due to difficulties involved in the prediction of certain intra-molecular interactions within a protein, such as long-distance electrostatic interactions (Spector et al. 2000).

The other main approach used in protein engineering is directed evolution, which, in contrast to rational design, relies on the use of random mutagenesis to create large-scale libraries of mutant clones. High-throughput screening can then be used to select mutants with the desired properties. This method is generally considered to be the most effective way of obtaining large numbers of modified enzymes containing a desired trait (Yang et al. 2014). While directed evolution circumvents the need for enzymatic structural information, it requires advanced screening technology and is not as cost-effective as simpler rational design approaches (Kaushik et al. 2016).

Combinations of these main approaches, along with continuous advances in protein engineering tools, have broadened the ways in which novel enzymes can be engineered. In fact, many of the most successful examples of protein engineering have been results of a

fusion of techniques (Bommarius et al. 2011; Chica et al. 2005; Eijsink et al. 2005). Using rational design approaches, scientists can pre-select certain areas of the protein believed to have the highest chance of accommodating a particular mutation and/or being relevant for the property of interest. Then, using directed evolution methods, smaller, more specific libraries can be generated and screened for this property of interest. This type of combination of techniques is often referred to as semi-rational protein engineering (Lutz 2010).

1.6.4. Engineering Stability and pH Optimum

The arrival of site-directed mutagenesis opened new doors for the world of protein engineering, enabling researchers to investigate the mechanisms of protein stability through experimentation, and, in terms of enzymes, to investigate catalytic mechanisms through well-designed mutagenesis studies.

In order to increase stability in a protein, it was recognized early on that any such mutation should not alter the folding mechanism, or the structure of the folded or the unfolded state, of the protein (Fersht et al. 1992). In enzymes, such mutations should also not affect the ability of the active site to perform catalysis. Residues in the active site are highly optimized for catalysis, but often do not contribute significantly to the global stability of the protein (Shoichet et al. 1995). As a consequence, well-designed mutations to these residues may serve to increase overall stability, but this will often be at the cost of the enzyme's functionality (Bloom et al. 2006). A natural question thus evolves: which parts, folds, or domains of proteins do contribute significantly to the overall stability? Early protein engineering studies demonstrated that surface-exposed residues can easily be mutated without substantial negative effect on the protein stability, which seems to indicate that stability stems from the more compact, inflexible regions of proteins (Matthews 1993).

Proteins and enzymes found in extremophiles may be useful in understanding protein stability. In contrast to early conclusions by Matthews, studies of such proteins indicated that surface residues are important for stability. For example, studies of proteins stemming from acidophilic and alkalophilic organisms have shown that these proteins generally tend to contain considerable amounts of charged residues, in particular on their surfaces (Jaenicke 1991). Several site-directed mutagenesis studies have shown that the introduction or removal

of charged residues on an enzyme's surface may cause large-scale changes to the electrostatic interaction network of a protein, thus significantly influencing its stability (Cockburn & Clarke 2011; Grimsley et al. 1999; Spector et al. 2000). Similarly, crystallographic studies of thermophilic organisms showed that enzymes produced by these organisms often contain more salt bridges than enzymes from mesophilic organisms, and that these may contribute to their overall stability at high temperatures (Kumar et al. 2000; Yip et al. 1998). Other studies have had somewhat mixed outcomes: while certain studies have shown that some salt bridges can stabilize the overall structure of proteins, other mutations designed to introduce salt bridges have been shown to be destabilizing (Kumar & Nussinov 1999; Makhatadze et al. 2003). Increasing stability by engineering surface charges is no easy feat, due to the complex electrostatic network mentioned earlier, and this should be kept in mind when designing rational mutations with this aim.

The degree of solvent exposure is another vital parameter to consider when designing mutations aimed to increase protein stability. While early studies of the degree of solvent-exposure of charged residues seemed to indicate that buried salt bridges have a destabilizing effect on proteins, more recent experiments have shown that this is not always the case (Dao-pin et al. 1991; Kumar & Nussinov 1999; Makhatadze et al. 2003).

While there is a large body of literature on stability engineering, the engineering of pH optima seems less well-explored. Nevertheless, there are some successful examples in the literature. Certain enzymatic pH-activity profiles have been modified through rational engineering by introducing charged amino acids in close proximity to the catalytic residues (Hirata et al. 2004; Kim et al. 2006; Nielsen et al. 2001). Such mutations have been shown to function by modifying the pK_a values of the catalytic residues. As discussed in Section 1.5.3.1, the ionization states of these residues are vital in terms of being able to catalyze the enzyme's specific reaction. Unfortunately, these pK_a shifts have often been slight, and mutations have mostly been based on comparative protein engineering, where the mutations selected have been determined by comparison with a homologous enzyme with the desired pH optimum range.

Accumulated data indicate that the introduction of charges is a good way to change enzyme properties, including the pH-activity profile, pH-stability and thermal stability (Cockburn & Clarke 2011). Such mutations will preferably either form salt bridges with naturally charged

residues in the protein, potentially increasing stability, alter the pK_a of one or more catalytic residues, causing a shift in the pH-activity profile, or simply contribute to an increased overall amount of charges in an attempt to "replicate" the large proportion of charged residues found in acidophilic and alkalophilic enzymes.

1.7. Origins of mgCel6A

The research described in this thesis focuses on the characterization and rational engineering of a family 6 glycoside hydrolase, called mgCel6A, with the main aim of optimizing the enzyme for use in industrial conditions. The gene coding for this enzyme was synthesized based on the published metagenome of a thermostable microbial community (Reddy et al. 2013), and full-length and truncated variants of this enzyme were expressed and characterized (Jensen et al. 2017).

1.7.1. NorZymeD

This study and the research preceding it were performed as part of the NorZymeD project at the Protein Engineering and Proteomics Group at the Norwegian University of Life Sciences. NorZymeD is an enzyme development project with numerous partners spread across Norway. The project aims to discover, characterize, and apply novel enzymes for industrial biomass processing. Enzyme sources include biogas reactors, the Svalbard reindeer gut, the Arctic shipworm, hot environments, and the artic mid-ocean ridge.

NorZymeD-generated cellulases considered to have high potential for industrial use are tested at Borregaard AS (Section 1.2.4), as this company is also involved in the NorZymeD project.

1.7.2. Previous Work on mgCel6A

All previous work was performed by Marianne S. Jensen and Dr. Lasse Fredriksen (Jensen et al. 2017).

Prior to beginning this study, the gene coding for the mgCel6A enzyme was synthesized based on publicly available metagenome data published by Reddy et al. in 2013. The metagenome in question was obtained from sequencing of the bacterial community resulting from an incubation at 55°C of high-solids rice straw inoculated with compost. This thermophilic community was found to contain large numbers of Actinobacteria, with a majority of the genomic data predicted to belong to the *Micromonospora* genus. Most of the cellulases analyzed from this community were shown to be cellobiohydrolases containing a characteristic family 2 CBM (CBM2) (Reddy et al. 2013).

1.7.2.1. Bioinformatics

The amino acid sequence of the mgCel6A enzyme was annotated using the InterPro server, which predicts relevant protein domains and families based on predictive models from a variety of databases, including Pfam, PROSITE, and many others (Finn et al. 2016; Finn et al. 2017). The results determined that the enzyme is a family 6 GH consisting of a catalytic domain and a CBM2, connected via a proline- and threonine-rich linker domain. The presence of these domains was also confirmed using the dbCAN server, which allocates signature domains to relevant areas of the protein using family classification data from the CAZy database (Yin et al. 2012). In addition, a BLASTp search using the NCBI Protein Blast Tool indicated that the catalytic domain of the mgCel6A enzyme had an 84% sequence identity with an endoglucanase produced by *Micromonospora rosaria* (Camacho et al. 2009; Zhang & Madden 1997). An additional BLAST search using the CBM of mgCel6A returned a 67% identity with the CBM of an endoglucanase produced by *Micromonospora echinaurantiaca*. It is therefore reasonable to assume that the naturally occurring mgCel6A enzyme is produced by an organism in this genus. Since the enzymes with the highest sequence identity with the catalytic domain and the CBM of mgCel6A are both endoglucanases, it is also natural to assume that mgCel6A is an endoglucanase. This theory is also supported by the fact that structure homology-modeling servers such as SWISS-MODEL found the the homolog with the highest sequence identity to the catalytic domain of mgCel6A for which the structure has been solved to be an endoglucanase from *Thermobifida fusca* (PDB ID: 1TML) (Arnold et al. 2006; Berman et al. 2000; Biasini et al. 2014; Bordoli et al. 2008) (Section 4.1.2).

1.7.2.2. Characterization of mgCel6A

Following the bioinformatics analysis, the full-length enzyme (mgCel6A) and a truncated version consisting only of the catalytic GH6 domain (mgCel6AΔCBM) were expressed in *E. coli*, purified, and characterized (Jensen et al. 2017).

1.7.2.2.1. Apparent Melting Temperature (T_m)

To evaluate the effects of the CBM2 domain on the thermostability of the mgCel6A enzyme, differential scanning calorimetry (DSC) was performed (***Figure 1.7.1***). Both enzymes were found to have apparent melting points at 76°C. The thermograms of both enzymes display a shoulder, peaking at approximately 64°C (mgCel6A) and 68°C (mgCel6AΔCBM), possibly indicating a minor conformational change occurring before the main unfolding event. This pre-unfolding event was present in both enzyme variants, indicating that it was not due to the presence or lack of the CBM2 domain. These results suggest that the CBM does not have a significant effect on the thermostability of mgCel6A (Jensen et al. 2017).

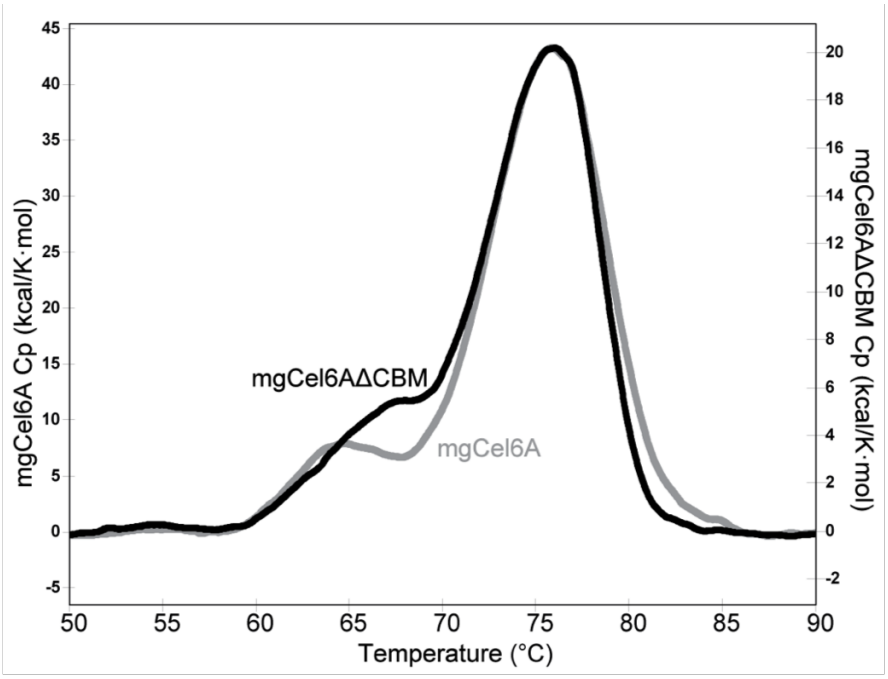


Figure 1.7.1. Thermograms produced during DSC of mgCel6A and mgCel6AΔCBM. The figure shows the molar heat capacity (C_p) of the two proteins over a temperature range from 50°C to 90°C. The gray curve shows the C_p values for the full-length protein, while the black curve shows the C_p values of the catalytic domain (mgCel6AΔCBM). Note that the two curves have individual y-axes. The proteins were heated at a rate of 1°C/minute in 75 mM phosphate-citrate buffer pH 6.0. Figure obtained from Jensen et al. (2017); reprinted with permission from the authors.

1.7.2.2.2. Optimal Enzyme Conditions

Marianne S. Jensen had previously determined the optimal conditions for enzymatic activity of both mgCel6A and mgCel6AΔCBM on carboxymethyl cellulose (CMC) and BALI™ cellulose.

Both variants were found to have an optimum temperature of 85°C on 1% (w/v) CMC in 15-minute reactions. For 15-minute reactions with 1% (w/v) BALI™ cellulose, however, the optimum temperature was found to be 75°C, indicating that substrate characteristics influence the enzyme's optimal conditions. Interestingly, in 24-hour reactions with 1% (w/v) BALI™ cellulose, the optimum temperature was reduced to 60°C, likely due to a decrease in thermostability over longer incubation times (results not shown).

Both mgCel6A and mgCel6AΔCBM showed identical optimum temperatures in all reactions tested, indicating that the CBM domain did not confer any additional thermostability to the full-length enzyme.

1.8. Aim of this Study

Previous work in the NorZymeD project (Sections 1.7.1 and 1.7.2) had yielded a putative *Micromonospora* endoglucanase with considerable thermostability called mgCel6A. Both the full-length enzyme and a truncated variant consisting only of the catalytic domain (mgCel6AΔCBM) were expressed and purified. Various physiochemical properties of the enzyme variants were investigated, including melting temperature, optimal conditions, stability, activity on various substrates, and the effect of the CBM domain on the enzyme (Jensen et al. 2017).

The aim of the present study was to further characterize and rationally engineer this enzyme to reduce its pH optimum and increase its pH stability. The goal was to develop a mutant variant of mgCel6A capable of operating at pH 5.0 or lower, and 60°C or higher, for more than 24 hours. The target substrate was the BALI™ cellulose substrate, a Borregaard AS-derived byproduct from processing of Norway spruce (Section 1.2.4).

The first part of this study involved the characterization of the full-length mgCel6A and the truncated variant, mgCel6AΔCBM. The pH optima and stability of the two enzyme variants were investigated through enzyme activity and stability assays.

The second part of this study concerned the rational design of mutant mgCel6AΔCBM variants. Homology modeling was used to generate a model of the mgCel6AΔCBM enzyme. The model was used in conjunction with studies of pH-engineering literature and a multiple sequence alignment of other GH6s with known pH optima to design a total of twelve rationally-designed mutant variants of mgCel6AΔCBM. All mutations were aimed at affecting stability, activity, or both, and were either on the protein surface or close to the catalytic center.

Six rationally-designed mutants were successfully expressed and purified. In the final part of this study, the wild-type mgCel6AΔCBM enzyme and mutant variants were subjected to similar pH optimum and stability assays, thus enabling comparison of the enzymes.

2. Materials

2.1. Laboratory Equipment and Materials

Table 2.1.1 shows an overview of the laboratory equipment and materials used throughout this study.

Table 2.1.1. Laboratory Equipment and Materials. The equipment and materials used during this study are listed along with the relevant supplier.

Category	Equipment	Supplier
Appliances	Bransonic® Ultrasonic Cleaner, Branson 3510	Sigma-Aldrich
	Certoclav Sterilizer, A-4050	CertoClav
	Freezer, -20°C	Bosch
	Freezer, -80°C	Sanyo
	Incubator, New Brunswick™ Innova® 44, 37°C	Eppendorf
	Incubator, 37°C	Termaks
	Laminar Flow Workbench, Telstar AV-100	Azbil Telstar, S.L.
	Microwave Oven, MD142	Whirlpool
	Milli-Q® Direct Water Purification System, Direct 16	Merck Millipore
	Refrigerator, 4°C	Bosch
	Sonics Vibra-Cell™ Ultrasonic Processor, VC750	Sonics & Materials, Inc.
	Thermomixer™ C	Eppendorf
	Water bath, 42°C	Julabo
	Water bath, 100°C	Stuart

Centrifuges	Centrifuge, Avanti™ J-25	Beckman Coulter
	Centrifuge, Heraeus™ Multifuge™ X1R	Thermo Scientific
	Microcentrifuge, Heraeus™ Pico™ 21	Thermo Scientific
	Microcentrifuge, MiniStar	VWR
Filtration and Ultrafiltration Equipment	Membrane dry vacuum pump/compressor, VCP 80	VWR
	Multiscreen®$_{HTS}$ 96-Well Plates, 0.45 μm Durapore® Membrane	Merck Millipore
	Multiscreen®$_{HTS}$ Vacuum Manifold	Merck Millipore
	Syringe Filtration Unit, Filtropur S 0.45 μm PES Membrane	Sarstedt
	Ultrafiltration Unit Vivaspin® 20, 10,000 MWCO PES	Sartorius
	Vacuum Filtration Systems, 0.2 μm PES Membrane	VWR
Gel Equipment	8-Well Comb	Bio-Rad
	Benchtop UV Transilluminator	UVP
	Electrophoresis Cell, Mini-PROTEAN® Tetra Vertical	Bio-Rad
	Electrophoresis System, Mini-Sub® Cell GT Horizontal	Bio-Rad
	Gel Casting Tray	Bio-Rad

	Gel Imaging System, Gel Doc™ EZ	Bio-Rad
	Power Supply, PowerPac™ 300	Bio-Rad
	Power Supply, PowerPac™ Basic	Bio-Rad
	Precast SDS Gels, Mini-PROTEAN® TGX Stain-Free™	Bio-Rad
	Stain-Free Sample Tray	Bio-Rad
	UV Sample Tray	Bio-Rad
Instruments and Columns	Bioreactor, LEX-48	Epiphyte Three
	HPAEC Column, Dionex™ CarboPac™ PA1, 250x2 mm	Thermo Scientific
	HPAEC System, Dionex™ ICS-3000	Thermo Scientific
	IMAC Column, HisTrap™ HP, 5 mL	GE Healthcare Life Sciences
	ÄKTA Pure	GE Healthcare Life Sciences
	PCR Machine, Mastercycler Gradient	Eppendorf
	pH Meter, 827 pH Lab	Metrohm
	Spectrophotometer, BioPhotometer® D30	Eppendorf
	Spectrophotometer, Multiskan™ FC Microplate Photometer	Thermo Scientific
Miscellaneous Equipment	Automated Pipettes	Thermo Scientific
	Cuvettes, Disposable	Eppendorf
	Glassware	Duran Group

	Inoculation Loops, 10 μL, blue	Sarstedt
	Magnet, Teflon Stirring Bar	SP Scienceware
	Magnetic Stirrer, RCT Basic	IKA
	Petri dishes, 90 mm	Heger
	Pipette Refill Tips	Thermo Scientific and VWR
	Scale, Entris 3202I-1S	Sartorius
	Scalpel, Stainless Steel Surgical Blade No. 10	Swann-Morton
	Thermometer	VWR
	Toothpicks	Playbox
	Vortex, MS2 Minishaker	IKA
Tubes, Vials, and Plates	Cellstar® Centrifuge Tubes, 15 mL and 50 mL	Greiner Bio-One
	Centrifuge Bottles, Nalgene®, Wide-Mouth with Sealing Caps, Style 3141, 500 mL	Sigma-Aldrich
	Centrifuge Tubes, Nalgene®, Oak Ridge, Style 3119, 25 mL	Sigma-Aldrich
	Eppendorf Tubes, 1.5 mL	Axygen
	Microtiter® 96-Well UV Microplates	Thermo Scientific
	PCR Tubes, 0.2 mL	Axygen
	Screw Cap Micro Tubes, 2 mL	Sarstedt
	Sealing Tape, Nunc™	Thermo Scientific

	Snap Ring Micro-Vials, 0.3 mL, and Snap Ring Caps	VWR
	Sarstedt Tubes, 13 mL	Sarstedt

2.2. Chemicals

Chemicals used in this study are shown in ***Table 2.2.1***.

Table 2.2.1. Chemicals. The table shows the chemicals used throughout this study along with the relevant supplier.

Chemical	Supplier
3,5-Dinitrosalicylic acid ($C_7H_4N_2O_7$)	Sigma-Aldrich
Agar Powder	VWR
Antifoam 204	Sigma-Aldrich
Avicel® PH-101	Sigma-Aldrich
Bacto™ Tryptone	Becton, Dickinson and Co.
Bacto™ Yeast Extract	Becton, Dickinson and Co.
Carboxymethyl cellulose (CMC) sodium salt ($C_{28}H_{30}Na_8O_{27}$)	Sigma-Aldrich
Cellobiose ($C_{12}H_{22}O_{11}$) (Glc2)	Fluka Analytical
Cellohexaose ($C_{36}H_{62}O_{31}$) (Glc6)	Megazyme
Cellopentaose ($C_{30}H_{52}O_{26}$) (Glc5)	Megazyme
Cellotetraose ($C_{24}H_{42}O_{21}$) (Glc4)	Megazyme
Cellotriose ($C_{18}H_{32}O_{16}$) (Glc3)	Megazyme
Citric acid ($C_6H_8O_7$)	Sigma-Aldrich
Dipotassium phosphate (K_2HPO_4)	Merck

Dithiothreitol (DTT) ($C_4H_{10}O_2S_2$)	Invitrogen
DNA Gel Loading Dye, 6X	New England Biolabs
Ethanol (C_2H_6O) absolute	VWR
Ethylenediaminetetraacetic acid (EDTA) ($C_{10}H_{16}N_2O_8$)	Merck
Glucose ($C_6H_{12}O_6$) (Glc1)	Fluka Analytical
Glc1 – Glc6 Stock Solution, 1.0 mg/mL	Made in lab
Glycerol ($C_3H_8O_3$), 85%	Merck
Hydrochloric acid (HCl)	Merck
Imidazole ($C_3H_4N_2$)	Sigma-Aldrich
Isopropyl β-D-1-thiogalactopyranoside (IPTG) ($C_9H_{18}O_5S$)	Sigma-Aldrich
Kanamycin ($C_{18}H_{36}N_4O_{11}$)	Sigma-Aldrich
Monopotassium phosphate (KH_2PO_4)	Merck
NEBuffer 2.1	New England Biolabs
NuPAGE® LDS Sample Buffer (4X)	Invitrogen
NuPAGE® Sample Reducing Agent (10X)	Invitrogen
peqGREEN DNA/RNA Dye	Peqlab
Potassium sodium tartrate tetrahydrate ($C_4H_4KNaO_6 \cdot 4H_2O$)	Sigma-Aldrich
SeaKem® LE Agarose	Lonza
Sodium acetate anhydrous (CH_3COONa), BioUltra	Sigma-Aldrich
Sodium chloride (ClNa)	VWR
Sodium hydroxide (HNaO) solution, 50%	Fluka Analytical
Sodium hydroxide (HNaO)	VWR

Sodium phosphate dibasic (HNa_2O_4P)	Sigma-Aldrich
Sucrose ($C_{12}H_{22}O_{11}$)	Sigma-Aldrich
Tris Base, ULTROL® Grade ($C_4H_{11}NO_3$)	Calbiochem
Tris-Acetate-EDTA (TAE) Buffer (50X)	Thermo Scientific
Tris-Glycine-SDS (TGS) Buffer (10X)	Bio-Rad

2.3. Plasmids

Table 2.3.1 shows the plasmids used during this study.

Table 2.3.1. Plasmids. The table shows the plasmids utilized during this study along with the relevant supplier.

Plasmid	Supplier
pNIC-CH	Addgene (Gift from Opher Gileadi) (see Appendix A)
pUC57	GenScript®

2.3.1. pUC57

Genes ordered from GenScript® were codon-optimized for *E. coli* and delivered in the pUC57 vector.

2.3.2. pNIC-CH

The pNIC-CH vector was used to create all recombinant plasmids produced in this study. This vector is a pET cloning and expression vector which utilizes a T7 promoter to transcribe the plasmid-encoded gene of interest via the host organism's T7 RNA Polymerase. The vector contains a *sacb* gene, allowing for negative selection using sucrose (see Section 3.7 for details). This gene is flanked by BfuAI restriction sites which enable the gene to be removed.

A target gene can then be inserted into the vector and the encoded protein can be expressed in an expression host such as One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*. The vector also contains an *aph* gene which confers kanamycin-resistance to the bacteria harboring the plasmid (see Section 2.13), as well as a region coding for a C-terminal 6xHis-tag immediately following the *sacb*/target gene. The addition of a terminal His-tag allows for the target protein encoded by the gene of interest to be isolated from other *E. coli* proteins through immobilized metal ion affinity chromatography (Section 3.14).

Additional data on pNIC-CH, including a vector map and the cloning site for insertion of a target gene, can be found in Appendix A.

2.4. Primers

Table 2.4.1 shows an overview of the primers used in this study. For the sequences of the primers, see ***Table 7.2.1*** in Appendix B.

Table 2.4.1. Primers by name and description. The table lists the primers used in this study and a description of their use. For primer sequences, see **Table 7.2.1** in Appendix B.

Primer	Description
mgCel6A_F LIC	*mgcel6a* Forward Primer with Ligation-Independent Cloning (LIC) Overhang (primer made by Dr. Alasdair Mackenzie)
mgCel6A_R LIC	*mgcel6a* Reverse Primer with LIC Overhang (primer made by Dr. Alasdair Mackenzie)
mgCel6AΔCBM_R LIC	*mgcel6aΔcbm* Reverse Primer with LIC Overhang (primer made by Dr. Alasdair Mackenzie)
M1_F	*mgcel6aΔcbm M1* (N21D) Forward Mutational Primer for Site-Directed Mutagenesis
M1_R	*mgcel6aΔcbm M1* (N21D) Reverse Mutational Primer for Site-Directed Mutagenesis

M2_F	*mgcel6aΔcbm M2* (N46D) Forward Mutational Primer for Site-Directed Mutagenesis
M2_R	*mgcel6aΔcbm M2* (N46D) Reverse Mutational Primer for Site-Directed Mutagenesis
M3_F	*mgcel6aΔcbm M3* (Q152E) Forward Mutational Primer for Site-Directed Mutagenesis
M3_R	*mgcel6aΔcbm M3* (Q152E) Reverse Mutational Primer for Site-Directed Mutagenesis
M4_F	*mgcel6aΔcbm M4* (Q276E) Forward Mutational Primer for Site-Directed Mutagenesis
M4_R	*mgcel6aΔcbm M4* (Q276E) Reverse Mutational Primer for Site-Directed Mutagenesis
M5_F	*mgcel6aΔcbm M5* (A104E) Forward Mutational Primer for Site-Directed Mutagenesis
M5_R	*mgcel6aΔcbm M5* (A104E) Reverse Mutational Primer for Site-Directed Mutagenesis
M6_F	*mgcel6aΔcbm M6* (L165E) Forward Mutational Primer for Site-Directed Mutagenesis
M6_R	*mgcel6aΔcbm M6* (L165E) Reverse Mutational Primer for Site-Directed Mutagenesis
M7_F	*mgcel6aΔcbm M7* (I242E) Forward Mutational Primer for Site-Directed Mutagenesis
M7_R	*mgcel6aΔcbm M7* (I242E) Reverse Mutational Primer for Site-Directed Mutagenesis
M8_F	*mgcel6aΔcbm M8* (I271D) Forward Mutational Primer for Site-Directed Mutagenesis
M8_R	*mgcel6aΔcbm M8* (I271D) Reverse Mutational Primer for Site-Directed Mutagenesis
M10_F	*mgcel6aΔcbm M10* (T44D) Forward Mutational Primer for Generation of Right Fragment in Splicing by Overlap Extension

M10_R	*mgcel6aΔcbm M10* (T44D) Reverse Mutational Primer for Generation of Left Fragment in Splicing by Overlap Extension
M11_F	*mgcel6aΔcbm M11* (D158A) Forward Mutational Primer for Generation of Right Fragment in Splicing by Overlap Extension
M11_R	*mgcel6aΔcbm M11* (D158A) Reverse Mutational Primer for Generation of Left Fragment in Splicing by Overlap Extension
M12_F	*mgcel6aΔcbm M12* (D158N) Forward Mutational Primer for Generation of Right Fragment in Splicing by Overlap Extension
M12_R	*mgcel6aΔcbm M12* (D158N) Reverse Mutational Primer for Generation of Left Fragment in Splicing by Overlap Extension
pLIC-for	pNIC-CH Forward 5' Sequencing Primer (from Addgene)
pLIC-rev	pNIC-CH Reverse 3' Sequencing Primer (from Addgene)

2.5. Proteins and Enzymes

The table below lists proteins and enzymes used in this study.

Table 2.5.1. Proteins and enzymes. The table shows the proteins and enzymes used in this study along with the relevant supplier.

Protein/Enzyme	Supplier
BenchMark™ Protein Ladder	Invitrogen
Bovine Serum Albumin (BSA) (100X)	New England Biolabs
Q5® Hot Start High-Fidelity Master Mix (2X)	New England Biolabs
Red Taq DNA Polymerase Master Mix (2X)	VWR
T4 DNA Polymerase	New England Biolabs

2.6. DNA

Table 2.6.1 shows DNA and nucleotides used in this study.

Table 2.6.1. DNA. The table shows DNA and nucleotides used in this study along with the relevant supplier.

DNA	Supplier
dGTP Mix, 10 mM	New England Biolabs
GeneRuler 1 kb DNA Ladder, ready-to-use	Thermo Scientific

2.7. Kits

Table 2.7.1 gives an overview of the kits used in this study.

Table 2.7.1. Kits. The table lists the kit name, the relevant supplier, and the contents of the kit.

Kit	Supplier	Contents
NucleoSpin® Gel and PCR Clean-up Kit	MACHEREY-NAGEL	Binding Buffer NTI Collection Tubes (2 mL) NucleoSpin® Gel and PCR Clean-Up Columns (yellow rings) Wash Buffer NT3
NucleoSpin® Plasmid Kit	MACHEREY-NAGEL	Collection Tubes (2 mL) Lysis Buffer A2 Neutralization Buffer A3 NucleoSpin® Plasmid Columns (white rings) Resuspension Buffer A1 Wash Buffer A4
QuikChange II Site-Directed Mutagenesis Kit	Agilent Technologies	dNTP Mix DpnI Restriction Enzyme PfuUltra High-Fidelity DNA Polymerase Reaction Buffer (10X)

2.8. Cellulose-based Substrates

The table below shows the cellulose-based substrates used in this study along with the origin of the substrate/the relevant supplier.

Table 2.8.1. Cellulose-based substrates. The substrates used in this study are shown along with the origin of the substrate/relevant supplier.

Substrate	Supplier
BALI™ cellulose	Borregaard AS (see Section 1.2.4)
Carboxymethyl cellulose (CMC)	Sigma-Aldrich
Phosphoric acid-swollen cellulose (PASC)	Produced by Dr. Lasse Fredriksen from Avicel® PH-101

2.9. Bacterial Strains

The bacterial strains used in this study and the relevant supplier are shown in ***Table 2.9.1***.

Table 2.9.1. Bacterial strains. The table shows the bacterial strains used in this study along with the relevant supplier.

Bacterial Strain	Supplier
One Shot® TOP10 Chemically Competent *E. coli*	Thermo Scientific
One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*	Thermo Scientific

2.10. Cultivation Media and Agar

Various types of media were used to cultivate bacterial strains of *E. coli* throughout this study. Lysogeny broth (LB) medium was used in liquid form in overnight cultures of *E. coli*, and in the form of agar plates when plating transformation mixtures to obtain single colonies. Terrific broth (TB) medium was used when cultivating *E. coli* for protein expression. Super optimal broth with Catabolite repression (S.O.C.) medium from Invitrogen was used in *E. coli* transformations.

2.10.1. Lysogeny Broth (LB)

Liquid medium:

- 10 g Bacto™ Tryptone
- 10 g NaCl
- 5 g Bacto™ Yeast Extract

All components were mixed in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L, and the medium was autoclaved. The medium was then stored at room temperature. When using the medium for bacterial cultivation, kanamycin was added prior to inoculation to give a final antibiotic concentration of 50 μg/mL.

Agar plates (with and without sucrose):

- 10 g Bacto™ Tryptone
- 10 g NaCl
- 5 g Bacto™ Yeast Extract
- 15 g Agar Powder

All components were dissolved in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L, and the medium was autoclaved. The solution was allowed to cool to about 50°C. Working at a laminar flow workbench, kanamycin was added to the solution to give a final antibiotic concentration of 50 μg/mL. When making agar plates with sucrose, sucrose was added to the solution prior to autoclaving to give a final concentration of 5% (w/v). The liquid medium was then gently poured into petri dishes and allowed to solidify. Plates were subsequently stored at 4°C.

2.10.2. Terrific Broth (TB)

Liquid medium:

- 6 g Bacto™ Tryptone
- 12 g Bacto™ Yeast Extract
- 2 mL Glycerol (85%)

All components were added to a 1 L flask, and dH_2O was added to a final volume of 450 mL. Aeration bottle caps were attached to the flasks, and the medium was autoclaved. The solution was allowed to cool to room temperature before adding 50 mL phosphate solution (Section 2.11.1) and 150 µL Antifoam 204. Before inoculation with bacterial culture, kanamycin was added to give a final antibiotic concentration of 50 µg/mL.

2.10.3. Super Optimal Broth with Catabolite repression (S.O.C.) Medium

Provided ready-made by Invitrogen.

2.11. Buffers and Other Solutions

2.11.1. Phosphate Solution for Terrific Broth (TB) Medium

- 23.14 g KH_2PO_4
- 125.4 g K_2HPO_4

The components were dissolved in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L, and the solution was autoclaved. The solution was stored at room temperature.

2.11.2. Phosphate-Citrate Buffers

Phosphate-citrate buffers were used in all enzyme activity assays (Section 3.18). The buffers were made by mixing 0.1 M citric acid and 0.2 M sodium phosphate dibasic solutions at various ratios to a total volume of 20 mL to obtain buffers ranging from pH 3.0 to pH 8.0 (see Table 2.11.1). The buffers were then heated to the temperature at which they would be used, and the pH was adjusted by adding 0.1 M citric acid or 0.2 M sodium phosphate dibasic until the correct pH was obtained. The solutions were then allowed to cool to room temperature before they were filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system. The buffers were stored at room temperature.

0.1 M Citric acid:

19.2 g citric acid was dissolved in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L. The solution was then filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system, and stored at room temperature.

0.2 M Sodium phosphate dibasic:

28.4 g sodium phosphate dibasic was dissolved in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L. The solution was then filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system, and stored at room temperature.

Phosphate-Citrate Buffers:

Table 2.11.1. Phosphate-citrate buffers. The table shows the volumetric ratios of 0.2 M sodium phosphate dibasic and 0.1 M citric acid used to make phosphate-citrate buffers of varying pHs.

Final pH	Volume (mL)	
	0.2 M Sodium phosphate dibasic	0.1 M Citric acid
3.0	4.08	15.92
3.5	6.04	13.96
4.0	7.72	12.28
4.5	9.00	11.00
5.0	10.28	9.72
5.5	11.36	8.64
6.0	12.84	7.16
6.5	14.20	5.80
7.0	17.44	2.56
7.5	17.98	2.02
8.0	19.53	0.47

2.11.3. Tris-HCl Buffer pH 8.0

Tris-HCl Buffer pH 8.0, 1 M and 20 mM:

121.14 g Tris base was dissolved in 800 mL dH_2O in a 1 L flask. The pH was adjusted to 8.0 using HCl. dH_2O was then added to a final volume of 1 L, and the solution was filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system, and stored at room temperature. The buffer was subsequently diluted with dH_2O under sterile conditions to make a 20 mM solution.

2.11.4. Ni-NTA Buffers for Ion Metal Affinity Chromatography (IMAC)

Ni-NTA Binding Buffer:

- 0.34 g imidazole
- 29.22 g NaCl
- 50 mL 1 M Tris-HCl buffer pH 8.0

All components were dissolved in 800 mL dH_2O in a 1 L flask. dH_2O was then added to a final volume of 1 L, and the medium was filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system.

Ni-NTA Elution Buffer:

- 17.02 g imidazole
- 14.61 g NaCl
- 25 mL 1 M Tris-HCl buffer pH 8.0

All components were dissolved in 400 mL dH_2O in a 500 mL flask. dH_2O was then added to a final volume of 500 mL, and the medium was filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system.

2.11.5. 3,5-Dinitrosalicylic Acid (DNS) Reagent

- 16 g NaOH
- 300 g Potassium sodium tartrate tetrahydrate
- 10 g 3,5-Dinitrosalicylic acid (DNS)

16 g NaOH was weighed and added to a 1 L volumetric flask. 800 mL dH_2O and a magnet were then added to the flask, which was placed on magnetic stirring at 40°C. 300 g potassium sodium tartrate tetrahydrate and 10 g DNS were weighed and mixed together. This mixture was gradually added to the volumetric flask. Upon complete dissolution, the solution was

cooled to room temperature and transferred to a dark flask. The DNS reagent was then stored at room temperature.

2.11.6. Eluents for High-Performance Anion Exchange Chromatography

Eluent A:

- 10.4 mL 50% (w/v) NaOH

Eluent B:

- 84.02 g CH_3COONa
- 5.2 mL 50% (w/v) NaOH

Eluent C:

- dH_2O

To prepare Eluent A, 2 L dH_2O was added to a 2 L volumetric flask. The eluent was then transferred to an ICS eluent bottle. To prepare Eluent B, 84.02 g sodium acetate (CH3COONa) was weighed and added to a 1 L volumetric flask. 1 L dH_2O was then added to the flask gradually before a magnet was added, and the flask was placed on magnetic stirring. Once the sodium acetate had completely dissolved, the solution was filtered under sterile conditions into an ICS eluent bottle using a 0.2 µm PES membrane vacuum filtration system. To prepare Eluent C, approximately 2 L dH_2O was added directly to an ICS eluent bottle. All three eluents were degassed for 10 minutes using a Sonics Vibra-Cell™ Ultrasonic Processor. Before attaching the ICS eluent bottles to the ICS-3000 system, 10.4 mL and 5.2 mL 50% (w/v) NaOH were added to Eluent A and B, respectively. The solutions were thoroughly mixed before the eluent bottles were attached to the ICS-3000 system.

2.12. Software and Online Resources

Table 2.12.1 gives an overview of the various software and online resources utilized throughout this study.

Table 2.12.1. Software and Online Resources. The software/online resource is shown in the table along with the relevant supplier.

Software/Online Resource	Supplier
BioEdit Sequence Alignment Editor	Accessed via mbio.ncsu.edu/BioEdit/bioedit.html
BioMath Calculator, dsDNA: µg to pmol	Promega
BLAST	NCBI
CAZy	AFMB
Chromeleon® 7 Chromatography Data System	Thermo Scientific
Clustal Omega	EMBL-EBI
dbCAN	Accessed via csbl.bmb.uga.edu/dbCAN
Image Lab™ Software	Bio-Rad
InterPro	EMBL-EBI
pDRAW32	AcaClone
Pfam	EMBL-EBI
PDB	RCSB
ProtParam Tool	ExPASy
PyMOL 1.8	Schrödinger; Accessed via pymol.org
QuikChange T_m Primer Calculator	Agilent Technologies
SWISS-MODEL	Swiss Institute of Bioinformatics
Unicorn™ 6.4.1	GE Healthcare Life Sciences

2.13. Antibiotics

Aminoglycosides are a group of broad-spectrum antibiotics commonly used to treat infections caused by Gram-negative bacteria. Aminoglycoside antibiotics consist of one or more aminated sugars linked via glycosidic bonds to a dibasic cyclitol. These antibiotics act by inhibiting normal protein synthesis in the bacteria they target. This inhibition occurs through binding of the aminoglycosides to the bacterial ribosomal 30S subunit, thus preventing elongation of peptide chains during protein synthesis. The result is most commonly a premature stop of protein synthesis, resulting in the production of impaired proteins. The insertion of these proteins into the bacterial membrane results in increased permeability and eventually the death of the bacterial cell.

In this study, the aminoglycoside antibiotic kanamycin was used to select for *E. coli* bacteria containing the pNIC-CH plasmid (see Appendix A). This plasmid contains the *aph* gene, which codes for aminoglycoside 3'-phosphotransferase (APH), an enzyme that inactivates numerous types of aminoglycosides via phosphorylation. The *aph* gene thus confers kanamycin-resistance to all bacteria that have taken up the pNIC-CH plasmid. Kanamycin was therefore used as a selectable marker in all liquid and agar media throughout this study, at a concentration of 50 μg/mL.

A kanamycin stock solution was made by adding 10 mL dH_2O to 0.5 g kanamycin. Aliquots of 1 mL were made, and these antibiotic solutions were stored at -20°C.

3. Methods

3.1. Cultivation of *Escherichia coli* (*E. coli*) Strains

Escherichia coli is a Gram-negative, facultative anaerobe, naturally found in the large intestine of warm-blooded organisms. *E. coli* is one of the most closely studied model prokaryotic organisms, and is widely used in the fields of biotechnology, molecular biology, biochemistry, and genetics. Different strains of this bacterium are commonly used *in vitro* to express and produce various heterologous proteins. This is done by inserting the gene coding for the protein of interest into a plasmid designed for expression of a target gene, which is then transformed into an *E. coli* expression host. The plasmid-harboring bacteria are cultivated and induced to promote protein expression, allowing for subsequent purification of the protein.

E. coli has numerous characteristics that make it highly suitable for use as a host in cloning and expression methods. Being a model organism, the genome of *E. coli* has been extensively studied, and its metabolic processes are well-understood. In addition, the organism's rapid growth rate, ability to grow with or without the presence of oxygen, and simple transformation procedure, all contribute to its frequent use in *in vitro* protein production.

In this study, two strains of *E. coli* were used for cloning and expression. When cloning the plasmid containing the inserted gene of interest, One Shot® TOP10 Chemically Competent *E. coli* cells were used to generate large numbers of pNIC-CH plasmid. One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells were then used to express the recombinant proteins.

Both *E. coli* strains were grown in liquid and agar LB medium. When using liquid LB medium, *E. coli* colonies harboring the recombinant pNIC-CH plasmid were cultivated overnight at 37°C and 220 rpm. Before inoculation, kanamycin was added to the medium to a concentration of 50 μg/mL. When grown on agar LB medium, various volumes of transformation mixture were plated on LB agar plates supplemented with 50 μg/mL kanamycin followed by cultivation overnight at 37°C.

In addition, One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells were also grown in TB liquid medium during protein expression. Prior to inoculation, the liquid TB medium was supplemented with kanamycin to a final concentration of 50 μg/mL, as well as with antifoam and phosphate solution, as detailed in Section 2.10.2.

All media and reagents used in cultivation were autoclaved and/or filtered under sterile conditions using a 0.2 μm PES membrane vacuum filtration system, and all bacterial cultures were made while working at a laminar flow workbench to prevent contamination. Overnight cultures were made by inoculating 3 or 5 mL of LB liquid medium in a sterile culture tube (supplemented with kanamycin as described above) with a single colony from an LB agar plate, or directly from a glycerol stock (Section 3.2) using an inoculation loop.

3.2. Long-Term Storage of Bacteria – Glycerol Stocks

When storing bacteria over longer periods of time, glycerol stocks were made to preserve the cells carrying the pNIC-CH plasmids with the gene of interest. Glycerol stabilizes the bacterial cells at freezing temperatures, and prevents damage to the cell membranes.

Glycerol stocks were made by adding 1 mL of bacterial overnight culture and 500 μL of sterile 50% (v/v) glycerol to a 2 mL screw cap micro tube, yielding a final glycerol concentration of 17% (v/v). The components were thoroughly mixed, and the tubes were stored at -80°C.

When using glycerol stocks to inoculate new overnight cultures, an inoculation loop was used to slightly touch the surface of the frozen glycerol stock, before inserting the loop into the culture tube containing medium and kanamycin. The inoculation loop was then removed, and the culture tube was left at 37°C and 220 rpm overnight.

3.3. Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) is an essential biotechnological method used to amplify specific DNA sequences. PCRs require a thermostable DNA polymerase, template DNA, forward and reverse primers, and deoxynucleoside triphosphates (dNTPs), which are used to synthesize novel complementary DNA. During a standard qualitative PCR, these four components are added to a PCR tube and placed in a thermocycler PCR machine. This machine promotes the three essential PCR steps in the reaction by changing the temperature in various pre-programmed steps, which last for specific periods of time.

The first step in a conventional PCR is the heating of the reaction to ensure melting of the template DNA. In this melting step, the two template DNA strands separate from one another, resulting in single-stranded template DNA carrying the target sequence of interest. In the second PCR step, the temperature is lowered, enabling the forward and reverse primers to bind to the single-stranded DNA in the regions where they are complementary to the template DNA. The primers are designed so that these complementary regions are adjacent to the target DNA sequence. The binding of the primers to the template DNA strands is called annealing. Finally, the temperature is raised again, allowing the thermostable DNA polymerase to bind to the primers and begin synthesizing the novel DNA containing the target sequence. The polymerase extends the chain of DNA by adding dNTPs to the continually growing strand in a process known as elongation. In every cycle of these three steps, the number of target DNA molecules is doubled.

In conjunction with this study, PCR was used for several purposes:

- Amplification of the wild-type genes, *mgcel6a* and *mgcel6aΔcbm*, and the mutant M9 gene, *M9* (Section 3.3.1) (see Appendix C for DNA sequences of wild-type genes and Appendix E for amino acid sequences of all mutant genes)
- Site-Directed Mutagenesis (Section 3.3.2)
- Splicing by Overlap Extension (Section 3.3.3)
- Colony PCR: Verification of correct transformation of the One Shot® TOP10 Chemically Competent *E. coli* cells following ligation-independent cloning (Section 3.3.4)

3.3.1. Amplification of the full-length (*mgcel6a*) and truncated (*mgcel6a*Δ*cbm*) wild-type genes and the Mutant M9 (*M9*) gene

Prior to beginning this study, the full-length *mgcel6a* gene was ordered from GenScript®. The gene was delivered in the pUC57 vector. The gene had been codon-optimized for expression in *E. coli* and was amplified as described in Jensen et al. (2017). To amplify the truncated *mgcel6a*Δ*cbm* gene, an alternate reverse primer was used. The same was done for the mutant *M9* gene later in the study. This work was performed by Marianne S. Jensen.

Following amplification, the wild-type and mutant M9 PCR products were run on an agarose gel (Section 3.4) and isolated from the gel using the NucleoSpin® Gel and PCR Clean-up Kit (Section 3.5). The dsDNA concentration was then measured using the D30 BioPhotometer® (Section 3.6) before the purified PCR products were subjected to ligation-independent cloning (LIC) (Section 3.7) to insert the genes into the pNIC-CH plasmid. Finally, the pNIC-CH plasmids containing the genes of interest were transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Section 3.8), which were plated on LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1).

3.3.2. Site-Directed Mutagenesis (SDM)

During the course of this study, PCR was used for to create mutants of the *mgcel6a*Δ*cbm* gene. To create mutants M1-M8 (see ***Table 4.1.1*** for an overview of mutants), PCR was used in conjunction with site-directed mutagenesis (SDM), using the QuikChange II Site-Directed Mutagenesis Kit. In this method, the forward and reverse mutant primers are both designed to contain the desired mutation, and are otherwise complementary to opposite strands of the template plasmid. Following the annealing of these primers to the template DNA, PfuUltra high-fidelity DNA Polymerase binds to the primers and begins to synthesize novel DNA. The result of this elongation is a plasmid containing the desired mutation.

The PCR product is then treated with DpnI endonuclease to digest parental non-mutated DNA, which is *E. coli*-derived and therefore methylated. The DpnI enzyme recognizes specific methylation sites in the parental DNA and cuts the plasmid at these sites, thereby promoting the degradation of the DNA lacking the mutation of interest. Finally, the newly

mutated plasmid is usually transformed into XL1-Blue Supercompetent Cells from Agilent Technologies. However, in this study, the transformation was instead done using One Shot® TOP10 Chemically Competent *E. coli* cells, as recommended by Dr. Jennifer Loose.

Materials:

- QuikChange II Site-Directed Mutagenesis Kit
- Mutant Forward Primer (M1_F, M2_F, M3_F, M4_F, M5_F, M6_F, M7_F, M8_F; see ***Table 2.4.1***)
- Mutant Reverse Primer (M1_R, M2_R, M3_R, M4_R, M5_R, M6_R, M7_R, M8_R; see ***Table 2.4.1***)
- Template DNA
- dH_2O

Method:

A glycerol stock of One Shot® TOP10 Chemically Competent *E. coli* cells containing the pNIC-CH plasmid carrying *mgcel6aΔcbm* was used to create an overnight culture (Section 3.1). The plasmid was isolated from the overnight culture using the NucleoSpin® Plasmid Kit, and the dsDNA concentration was measured using the D30 BioPhotometer®. PCRs were subsequently prepared according to ***Table 3.3.1***. All reaction mixtures were made on ice in 0.2 mL PCR tubes, and the PCRs were performed in a Mastercycler Gradient PCR machine using the conditions shown in ***Table 3.3.2***.

Table 3.3.1. PCR components for site-directed mutagenesis to create mutants M1-M8.

Component	Volume (μL)
Reaction Buffer (10X)	5
Mutant Forward Primer	1.1-10.2 (125 ng)
Mutant Reverse Primer	1.1-10.2 (125 ng)
Template DNA (*mgcel6aΔcbm* in pNIC-CH)	1.72 (10 ng)
dNTP Mix	1
dH_2O	To 50 μL
PfuUltra High-Fidelity DNA Polymerase (added last)	1

Table 3.3.2. PCR conditions for site-directed mutagenesis to create mutants M1-M8.

Step	Number of Cycles	Temperature (°C)	Time (mm:ss)
Initial Denaturation	1	95	00:30
Melting	16	95	00:30
Annealing		55	01:00
Elongation		68	07:30

Following amplification, 1 μL DpnI was added to each PCR product. The reactions were thoroughly mixed and then centrifuged in a Heraeus™ Pico™ 21 microcentrifuge at 11000 g for 1 minute, before they were incubated at 37°C for 1 hour. Finally, the mutated plasmids were transformed into One Shot® TOP10 Chemically Competent *E. coli* cells, which were plated on LB agar plates supplemented with kanamycin.

3.3.3. Splicing by Overlap Extension (SOE)

Due to difficulties obtaining correctly mutated genes for certain mutants of the *mgcel6aΔcbm* gene, an alternate PCR method was used to create mutants M10-M12 (see ***Table 4.1.1*** for an overview of mutants). For these mutants, PCR was used in conjunction with splicing by

overlap extension (SOE), followed by ligation-independent cloning to insert the mutated gene into the pNIC-CH plasmid.

Gene SOE is a technique that allows for the introduction of point mutations through PCR. This method generates two fragments of a gene insert that share short regions of complementarity in which the mutation is located. These fragments are then used as template DNA in a second PCR to generate the full-length gene of interest.

In the first PCR, a left and a right fragment are made for each mutant. The mutant fragments both contain a terminal 25-bp region that is complementary to the other mutant fragment. In the second PCR, the strands hybridize according to these regions of complementarity. The Q5 Polymerase enzyme then extends each fragment strand from the 3' end of the 25 bp region, based on the sequence of the other fragment strand. The result is a full-length double-stranded gene containing the desired mutation.

The final PCR products are subjected to ligation-independent cloning (LIC) with T4 DNA Polymerase to digest the mutated gene (Section 3.7), before the newly constructed pNIC-CH plasmid is transformed into One Shot® TOP10 Chemically Competent *E. coli* cells.

3.3.3.1. First PCR

Materials:

- Q5 High-Fidelity Master Mix (2X)
- 10 µM mgCel6A_F LIC and mgCel6AΔCBM_R LIC primers (see ***Table 2.4.1***)
- 10 µM M10_F, M11_F or M12_F primers (see ***Table 2.4.1***)
- 10 µM M10_R, M11_R, M12_R primers (see ***Table 2.4.1***)
- Template DNA
- Nuclease-Free dH_2O

Method:

PCRs were prepared according to ***Table 3.3.3***. All reaction mixtures were made on ice in 0.2 mL PCR tubes, and the PCRs were performed in a Mastercycler Gradient PCR machine using the conditions shown in ***Table 3.3.4***.

To generate the left fragments for each mutant, the mgCel6A_F LIC forward primer and the relevant mutant reverse primer were used. To generate the right fragments for each mutant, the relevant mutant forward primer and the mgCel6AΔCBM_R LIC reverse primer were used.

The sizes of the fragments were estimated using pDRAW32 (AcaClone).

To obtain the template DNA, a glycerol stock of One Shot® TOP10 Chemically Competent *E. coli* cells containing the pNIC-CH plasmid carrying *mgcel6a*Δ*cbm* was used to create an overnight culture. The plasmid was isolated from the overnight culture using the NucleoSpin® Plasmid Kit (Section 3.9), and the dsDNA concentration was measured using the D30 BioPhotometer®.

Table 3.3.3. PCR components for left and right fragment generation to create mutants M10-M12.

Component	Volume (μL)
Q5 High-Fidelity Master Mix (2X)	25
10 μM Forward Primer	2.5
10 μM Reverse Primer	2.5
Template DNA (*mgcel6a*Δ*cbm* in pNIC-CH)	1
Nuclease-Free dH_2O	19

Table 3.3.4. PCR conditions for left and right fragment generation to create mutants M10-M12.

Step	Number of Cycles	Temperature (°C)	Time (mm:ss)
Initial Denaturation	1	98	00:30
Melting	25	98	00:07
Annealing		64	00:20
Elongation		72	00:16
Final Extension	1	72	02:00

The PCR products were then run on an agarose gel to verify the size of the left and right fragments, and the DNA bands were isolated from the gel and purified using the NucleoSpin® Gel and PCR Clean-up Kit. The dsDNA concentration was measured using the D30 BioPhotometer®.

3.3.3.2. Second PCR

Materials:

- Q5 High-Fidelity Master Mix (2X)
- 10 µM mgCel6A_F LIC and mgCel6AΔCBM_R LIC (see ***Table 2.4.1***)
- Mutant Left Fragment Template DNA
- Mutant Right Fragment Template DNA
- Nuclease-Free dH_2O

Method:

The Promega online dsDNA: µg to pmol conversion calculator (promega.com) was used to determine the amount of each mutant fragment needed for the second PCRs, based on measured DNA concentration. The reactions were made using a 1:1 molar ratio of left:right fragment.

To generate the mutant *mgcel6aΔcbm* genes, the mgCel6A_F LIC forward primer and mgCel6AΔCBM_R LIC reverse primers were used.

PCRs were prepared according to ***Table 3.3.5***. All reaction mixtures were made on ice in 0.2 mL PCR tubes, and the PCRs were performed in a Mastercycler Gradient PCR machine using the conditions shown in ***Table 3.3.6***.

Table 3.3.5. PCR components for full-length insert generation to create mutants M10-M12.

Component	Volume (µL)
Q5 High-Fidelity Master Mix (2X)	25
10 µM Forward Primer	2.5
10 µM Reverse Primer	2.5
Left Fragment Template DNA	0.5-1.5
Right Fragment Template DNA	1.0-1.5
Nuclease-Free dH_2O	To 50 µL

Table 3.3.6. PCR conditions for full-length insert generation to create mutants M10-M12.

Step	Number of Cycles	Temperature (°C)	Time (mm:ss)
Initial Denaturation	1	98	00:30
Melting	25	98	00:07
Annealing		64	00:20
Elongation		72	00:20
Final Extension	1	72	02:00

The PCR products were run on an agarose gel to verify the correct size of the newly spliced gene, and then isolated from the gel and purified using the NucleoSpin® Gel and PCR Clean-up Kit. The dsDNA concentration was subsequently measured using the D30 BioPhotometer® before the PCR products were subjected to T4 DNA Polymerase digestion

followed by ligation-independent cloning (LIC) to generate the mutated pNIC-CH plasmid. Finally, the plasmid was transformed into One Shot® TOP10 Chemically Competent *E. coli* cells, which were plated on LB agar plates supplemented with kanamycin and sucrose.

3.3.4. Colony PCR

In order to verify that the mutated gene had been correctly cloned into the pNIC-CH vector during LIC (Section 3.7), colony PCR was performed on the One Shot® TOP10 Chemically Competent *E. coli* colonies grown on LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1).

Colony PCR is an efficient procedure that enables confirmation of the presence of an inserted gene in a plasmid. The method requires an initial heat lysis step, which ruptures the bacterial cell membrane, releasing the plasmid into solution. During PCR, this plasmid is used as the template DNA for amplification, using primers specific to the insert gene. If the insert has been correctly cloned into the plasmid, the PCR will amplify the mutated insert gene. The PCR products can then be visualized on an agarose gel (Section 3.4) to confirm that the insert is of the correct size, thereby ensuring that the LIC and subsequent transformation have been successful.

Materials:

- Red Taq DNA Polymerase Master Mix (2X)
- 10 μM mgCel6A_F LIC Forward Primer and mgCel6AΔCBM_R LIC Reverse Primer (see ***Table 2.4.1***)
- Template DNA
- dH_2O

Method:

PCRs were prepared according to ***Table 3.3.7***. All reaction mixtures were made on ice in 0.2 mL PCR tubes, and the PCRs were performed in a Mastercycler Gradient PCR machine using the conditions shown in ***Table 3.3.8***.

Preparing the template DNA:

Working at a laminar flow workbench, sterile toothpicks were brushed against transformed colonies that had grown on the LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1). The tips of the toothpicks were then brought into contact with the bottom of the PCR tubes, thus transferring a small amount of the colony to the tube. The tubes were then microwaved for 2 minutes to lyse the bacterial cells, before allowing them to cool on ice. The other reaction components were then added to the PCR tubes containing the template DNA.

Table 3.3.7. PCR components for colony PCR to verify transformation One Shot® TOP10 Chemically Competent E. coli cells.

Component	Volume (μL)
Red Taq DNA Polymerase Master Mix (2X)	25
10 μM Forward Primer	1
10 μM Reverse Primer	1
Template DNA (mutant *mgcel6aΔcbm* in pNIC-CH)	-
dH_2O	23

A positive control reaction was also simultaneously performed to verify correct insertion of the mutated gene into the pNIC-CH plasmid. The positive control contained 1 μL non-mutated *mgcel6aΔcbm* in pNIC-CH, and 22 μL dH_2O, as well as the volumes of the other three reaction components listed in ***Table 3.3.7***.

Table 3.3.8. PCR conditions for colony PCR to verify transformation One Shot® TOP10 Chemically Competent E. coli cells.

Step	Number of Cycles	Temperature (°C)	Time (mm:ss)
Melting	25	95	00:25
Annealing		55	00:30
Elongation		72	01:00
Final Extension	1	72	05:00

The PCR products were then run on an agarose gel (Section 3.4) to verify the correct size of the mutated insert (864 bp) (see Appendix C for DNA sequences of the wild-type genes).

3.4. Agarose Gel Electrophoresis

Agarose gel electrophoresis was used throughout this study to separate DNA fragments based on their relative size. Agarose is a linear polysaccharide composed of repeating agarobiose units. Following dissolution in a suitable buffer via heating and subsequent cooling of the solution, agarobiose chains bind non-covalently to form a gel matrix containing pores of a determined size. When DNA samples are run on an agarose gel, smaller molecules will encounter less resistance when travelling through the gel pores, and thus travel farther through the gel. This principle allows for the separation of DNA molecules based on their size.

To enable the DNA molecules to travel through the gel, a current is applied throughout the gel chamber, creating an electric field with an anode and a cathode end. Due to the negative charge of the phosphate backbone, DNA molecules will travel through the agarobiose pores towards the positively charged anode.

In this study, all PCR products were run on 1% agarose gels. To enable visualization of the DNA bands following migration through the gel, peqGREEN dye was added to the gels before solidification. 6X DNA Gel Loading Dye was added to all samples before application on the gel. Note that loading dye was not added when running PCR products from Colony

PCR, as these already contained a red dye present in the Red Taq DNA Polymerase Master Mix (Section 3.3.4). A GeneRuler 1 kb DNA ladder consisting of DNA fragments of known size in bp was run alongside all samples to enable size comparison and determination. All gels were run in a Mini-Sub® Cell GT Horizontal Electrophoresis System using a PowerPac™ Basic power supply.

Materials:

1% Agarose Gel:

- 0.6 g agarose
- 60 mL 1X Tris-Acetate-EDTA (TAE) Buffer
- 3 μL peqGREEN Dye

Loading Buffer:

- DNA Gel Loading Dye (6X)

Running Buffer:

- Tris-Acetate-EDTA (TAE) Buffer (1X)

DNA Ladder:

- GeneRuler 1 kb DNA Ladder, ready-to-use

Method:

Preparing the 1% agarose gel:

The agarose and the TAE buffer were mixed in a 100 mL flask and microwaved until the solution began to boil and the agarose powder was completely dissolved. The solution was then left to cool to about 55°C before 3 μL peqGREEN dye was added. Finally, the solution was poured into a gel casting tray with an 8-well comb and allowed to solidify for 30 minutes.

Preparing the electrophoresis chamber and loading the samples:

Once the gel was solid, the spacer was removed, and the casting tray containing the gel was transferred to a Mini-Sub® Cell GT Horizontal Electrophoresis System. 1X TAE buffer was then added to the chamber. 6X DNA Gel Loading Dye was added to each sample to be run on the gel in a 1:10 volumetric ratio. 10 µL of the ladder and 50 µL of the samples were then loaded into the wells on the gel. Note that 6X DNA Gel Loading Dye was not added to the GeneRuler 1 kb DNA ladder before application onto the gel, since this ladder already contains a loading dye.

Running the gel and visualizing the DNA bands:

The gel was run for 50 minutes at 90V. The current was applied to the chamber using a PowerPac™ Basic power supply. Once the run had been completed, the DNA bands were visualized with a Benchtop UV Transilluminator. Finally, gel imaging was conducted using a UV Sample Tray and a Gel Doc™ EZ Imager. Gel images were generated using the Image Lab™ Software.

3.5. Extraction and Purification of DNA Fragments from Agarose Gels

Following agarose gel electrophoresis, DNA fragments were extracted and purified from the gel using the NucleoSpin® Gel and PCR Clean-up Kit. Protocol 5.2 (Rev. 04) from the user manual, DNA extraction from agarose gels, was followed as written, with a few minor changes (detailed below).

Materials:

- 1% agarose gel with DNA fragments of interest
- Scalpel
- NucleoSpin® Gel and PCR Clean-up Kit
- dH_2O

Method:

Extracting DNA fragments from the 1% agarose gel:

Once the gel had been removed from the electrophoresis chamber (Section 3.4), it was placed on the Benchtop UV Transilluminator. UV light was then applied in short intervals to visualize the bands and gently cut them out of the gel using a scalpel.

Purification of the DNA fragments:

Each gel slice was placed into a pre-weighed 1.5 mL Eppendorf tube. The Eppendorf tubes were weighed to determine the mass of each gel slice. According to the NucleoSpin® Gel and PCR Clean-up Kit Protocol, 200 µL Binding Buffer NTI was added for each 100 mg of gel. The tubes were then incubated at 50°C for 10 minutes to ensure complete dissolution of the gel slice. NuceloSpin® Gel and PCR Clean-Up Columns were placed into 2 mL Collection Tubes, and the dissolved gel samples were loaded onto the columns. The tubes were then centrifuged for 30 seconds at 11000 g, and the flow-through was discarded. The column silica membranes were washed by applying 700 µL Wash Buffer NT3 and centrifuging for 30 seconds at 11000 g. The silica membranes were then dried by centrifuging the tubes for 1 minute at 11000 g. Finally, the columns were transferred to new 1.5 mL Eppendorf tubes. DNA elution was performed by adding 25 µL dH_2O to the columns and incubating at room temperature for 1 minute, before the final centrifugation for 1 minute at 11000 g.

The isolated and purified plasmids were stored at -20°C.

All centrifugation steps were performed in a Heraeus™ Pico™ 21 Microcentrifuge.

3.6. Measuring dsDNA Concentration – A_{260}

To measure the concentration of dsDNA in all samples, a D30 BioPhotometer® was used. The concentration of dsDNA in a sample is determined by measuring its UV absorbance at 260 nm, A_{260}. The Beer-Lambert law is used to determine the sample's DNA concentration in

µg/mL. This law relates the absorbance of a sample to the path length through which the UV light travels and to the concentration of the sample. The law is usually written as $A = \varepsilon * b * c$, where A is the absorbance, ε the extinction coefficient, b the path length, and c the concentration. At 260 nm, the average extinction coefficient of double-stranded DNA is 0.02 $(\mu g/mL)^{-1}$ cm^{-1}. This value relates the measured A_{260} to the concentration of DNA in µg/mL.

Note that in this study, the path length b was always equal to 1 cm.

Materials:

- dsDNA
- dH_2O
- Disposable cuvette

Method:

Diluted samples to be measured in the spectrophotometer were prepared in 0.2 mL PCR tubes. The dsDNA samples were diluted 12X by adding 5 µL sample to 55 µL dH_2O and mixing thoroughly. The whole reaction volume was then transferred to a disposable cuvette which was subsequently placed in the D30 BioPhotometer® instrument. The A_{260} was then measured, and the dsDNA concentration of the 12X dilution was obtained. The final concentration was determined by multiplying the concentration by 12.

3.7. Ligation-Independent Cloning (LIC)

In LIC, the 3'→5' exonuclease activity of T4 DNA Polymerase is used to degrade one of the two DNA strands in a gene insert from the 3' end. By using only dGTPs in the reaction mixture, the polymerase's exonuclease activity will outcompete its polymerase activity until the first guanine (G) base is reached in the strand that is being degraded. At this point, the T4 DNA Polymerase ceases its exonuclease activity. The result is a double-stranded DNA molecule with a single-stranded overhang.

When the corresponding digestion is done on the pNIC-CH vector using only dCTPs, the overhangs will be complementary and will then anneal to one another when the two components are combined in the cloning process. The result is a circular pNIC-CH plasmid carrying the gene fragment of interest, which can then be transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Section 3.8).

The digestion of the pNIC-CH vector removes the *sacb* gene from the plasmid. This gene codes for levansucrase, an enzyme that hydrolyzes sucrose while simultaneously producing fructose polymers called levans. Levans are toxic to *E. coli* bacteria, and any bacteria carrying an intact *sacb* gene will thus be unable to grow in the presence of sucrose. Sucrose is therefore used as a selectable marker following the transformation of the mutated plasmid into One Shot® TOP10 Chemically Competent *E. coli* cells. This ensures that all colonies that grow on the LB agar plates supplemented with kanamycin and sucrose lack the *sacb* gene, and instead contain the inserted gene of interest (see Section 2.3.2 and Appendix A).

Before performing LIC, the Promega online dsDNA: µg to pmol conversion calculator (promega.com) was used to determine the amount of insert needed for the T4 DNA Polymerase digestion, based on the concentrations measured according to Section 3.3.3.2. The reactions were prepared containing 0.2 pmol of mutated insert.

Materials:

- NEBuffer 2.1
- Bovine Serum Albumin (BSA) (20X)
- dGTP Mix
- Dithiothreitol (DTT) (100 mM)
- T4 DNA Polymerase
- Template DNA (mutant *mgcel6aΔcbm* genes)
- T4 DNA Polymerase-Digested pNIC-CH Vector (digestion performed by Dr. Lasse Fredriksen)
- Ethylenediaminetetraacetic acid (EDTA) (25 mM)
- dH_2O

Method:

The *mgcel6aΔcbm* mutant genes to be inserted were digested using T4 DNA Polymerase. The reactions were prepared according to ***Table 3.7.1***. The reaction mixtures were made on ice in 0.2 mL PCR tubes, and the LIC digestion was performed in a Mastercycler Gradient PCR machine using the conditions shown in ***Table 3.7.2***.

Table 3.7.1. Components used in LIC digestion of mutant *mgcel6aΔcbm* genes.

Component	Volume (µL)
NEB Buffer 2	2
BSA (20X)	1
dGTP Mix	2
DTT	1
Template DNA (mutant *mgcel6aΔcbm* genes)	1.1-1.2 (0.2 pmol)
T4 DNA Polymerase (added last)	1
dH_2O	To 20 µL

Table 3.7.2. PCR conditions used in LIC digestion of mutant *mgcel6aΔcbm* genes.

Step	Number of Cycles	Temperature (°C)	Time (mm:ss)
Lic Digestion	1	22	60:00
Heat Deactivation	1	75	21:00

4.5 µL of the T4-digested gene fragments were then mixed with 0.5 µL of the T4-digested pNIC-CH plasmid in 0.2 mL PCR tubes. The reactions were left at room temperature for 1 hour to allow the insert and the vector to coalesce.

A negative control reaction was also performed simultaneously in order to verify correct insertion of the mutated gene into the pNIC-CH plasmid. The negative control contained 0.5 µL T4-digested pNIC-CH plasmid, and 4.5 µL dH_2O instead of T4-digested insert DNA.

2 µL EDTA was then added to each reaction before transformation into One Shot® TOP10 Chemically Competent *E. coli* cells (Section 3.8). The transformed cells were plated on LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1).

3.8. Transformation of One Shot® TOP10 Chemically Competent *E. coli*

Following either SDM or LIC (Section 3.3.2 and Section 3.7, respectively), the pNIC-CH plasmid containing a wild-type or mutated *mgcel6aΔcbm* gene was transformed into One Shot® TOP10 Chemically Competent *E. coli* cells. This *E. coli* strain is designed to enhance cloning of all genomic DNA, and therefore allows for effective amplification of the pNIC-CH plasmid containing the gene of interest.

Materials:

- Plasmid DNA (pNIC-CH containing gene of interest)
- One Shot® TOP10 Chemically Competent *E. coli* cells
- Super Optimal broth with Catabolite repression (S.O.C.) medium
- LB agar plates supplemented with kanamycin to 50 µg/mL (Section 2.10.1)
- Sarstedt 13 mL tubes

Method:

To transform the plasmid into One Shot® TOP10 Chemically Competent *E. coli*, the transformation protocol from Invitrogen was used (Publication Number MAN0000633, Rev. A.0). The Chemical transformation procedure was followed as written, with a few minor changes (detailed below). One vial of One Shot® TOP10 Chemically Competent *E. coli* cells was used for each transformation reaction.

The vials containing the One Shot® TOP10 Chemically Competent *E. coli* cells were gently thawed on ice. The vials were then quickly spun down, before 50 µL of the cells were

transferred to pre-chilled Sarstedt 13 mL tubes. 5 µL of plasmid DNA was then added to the cells in the tubes, and the reactions were gently swirled to mix the components before being incubated on ice for 30 minutes. After the incubation, the reactions were heat-shocked by incubating the tubes for 30 seconds in a 42°C water bath, before placing the tubes back on ice for 2 minutes. Working at a laminar flow workbench, 250 µL S.O.C. medium was added to each tube. The tubes were then incubated at 37°C and 220 rpm for 1 hour.

Following the 1 hour-incubation, 100 µL of the cell suspension was plated on room-temperature LB agar plates supplemented with kanamycin to a final concentration of 50 µg/mL. When transforming LIC products, 100 µL of the reaction mixture was plated on room-temperature LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1). The plates were then incubated at 37°C overnight.

Colonies able to grow on the kanamycin or kanamycin and sucrose plates were then used for Colony PCR for LIC products (Section 3.3.4) and/or inoculation of overnight cultures in LB liquid medium supplemented with kanamycin (Section 2.10.1).

The resulting overnight cultures were used to isolate the amplified pNIC-CH plasmid containing the gene of interest (Section 3.9).

3.9. Plasmid Isolation and Purification

Following cultivation of transformed One Shot® TOP10 Chemically Competent *E. coli* cells in LB liquid medium, the amplified pNIC-CH plasmid containing the gene of interest was isolated from the overnight culture using the NucleoSpin® Plasmid Kit. Protocol 5.1 (Rev. 09) from the User manual, Isolation of high-copy plasmid DNA from *E. coli*, was followed as written, with a few minor changes (detailed below).

Materials:

- Overnight culture of transformed One Shot® TOP10 Chemically Competent *E. coli* cells
- NucleoSpin® Plasmid Kit
- dH_2O

Method:

To harvest the bacterial cells, overnight culture was added to 1.5 mL Eppendorf tubes in 1 mL increments, and centrifuged for 30 seconds at 11000 g. The supernatant was discarded between each centrifugation. After cells from the entire 5 mL One Shot® TOP10 Chemically Competent *E. coli* overnight culture had been collected, 250 µL Resuspension Buffer A1 was added to each tube, and the pellets were resuspended by pipetting up and down. This buffer contains RNase A, which degrades cellular RNA following the subsequent lysis of the harvested bacterial cells. The buffer also contains other components, such as ethylenediaminetetraacetic acid (EDTA), which help stabilize the target plasmid and aid in the disruption of the bacterial cell walls. 250 µL Lysis Buffer A2 was then added to each tube, and the reactions were mixed gently by inverting the tubes 8 times. The NaOH and sodium dodecyl sulfate (SDS) present in this buffer help degrade the bacterial cell walls and cell membranes, effectively lysing the bacterial cells. These components also disturb hydrogen bonds within double-stranded bacterial DNA and denature bacterial proteins. The tubes were then left at room temperature for 5 minutes to ensure complete lysis of the bacterial cells, before adding 300 µL Neutralization Buffer A3 to each tube. The addition of the Neutralization Buffer enables the bacterial DNA to reform its double-stranded arrangement. While this is relatively simple for short plasmid DNA, longer genomic DNA is very unlikely to reanneal during this step. The reactions were again mixed by inverting the tubes 8 times until the reactions were colorless, and the tubes were centrifuged for 5 minutes at 11000 g. The single-stranded genomic DNA will form aggregates with the denatured bacterial proteins and can thus be easily separated from the soluble double-stranded plasmid DNA during this step. NucleoSpin® Plasmid Columns were placed into 2 mL Collection Tubes, and the supernatants from the preceding centrifugation were loaded onto the columns. The tubes were centrifuged for 1 minute at 11000 g, and the flow-through was discarded. 600 µL Wash

Buffer A4 was then added to each column, before centrifuging the tubes for 1 minute at 11000 g. The addition of this buffer enables the excess salts and other contaminants present on the column to be washed away. The column silica membranes were dried by centrifuging the tubes for 2 minutes at 11000 g, and final elution of purified plasmid DNA was performed by adding 50 µL dH_2O to the columns and incubating at room temperature for 1 minute, before the final centrifugation for 1 minute at 11000 g.

Isolated plasmid DNA was subsequently stored at -20°C.

All centrifugation steps were performed in a Heraeus™ Pico™ 21 Microcentrifuge.

Once the plasmid had been isolated, the dsDNA concentration was measured using a D30 BioPhotometer® (Section 3.6). The plasmid DNA was then sequenced (Section 3.10) before subsequent transformation into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells (Section 3.11).

3.10. DNA Sequencing

pNIC-CH plasmids carrying the gene of interest were sequenced following isolation from One Shot® TOP10 Chemically Competent *E. coli*. Sequencing was performed on all mutant pNIC-CH plasmids to verify the correct mutations had been inserted into the *mgcel6aΔcbm* gene (see ***Table 4.1.1*** for an overview of mutants).

To prepare samples for sequencing, plasmid DNA, forward or reverse primers, and dH_2O were combined in 1.5 mL Eppendorf tubes. Sequencing reactions were prepared as shown in ***Table 3.10.1***. Forward and reverse sequencing reactions were prepared for each mutant plasmid using 10 µM pLIC-for and pLIC-rev sequencing primers (see Table ***Table 2.4.1***) that anneal to the plasmid on each side of the gene of interest. These primers are used to ensure the entire gene is sequenced correctly.

The concentration of plasmid DNA measured using the D30 BioPhotometer® (Section 3.6) was used to determine the volume required to add 400-500 ng plasmid DNA to each reaction.

Sequencing was performed by GATC Biotech. Analysis of sequencing results was performed using BioEdit Sequence Alignment Editor.

Table 3.10.1. Components for GATC sequencing of mutant *mgcel6aΔcbm* genes.

Component	Volume (µL)
Plasmid DNA	3.0-8.5 (400-500 ng)
Forward or Reverse Primer	2.5
dH_2O	To 11 µL

After verification of the sequencing results, mutated plasmids were transformed into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells for protein expression (Section 3.11).

3.11. Transformation of One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*

Following DNA sequencing (Section 3.10), the pNIC-CH plasmid containing a wild-type or mutated *mgcel6aΔcbm* gene was transformed into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells. This *E. coli* strain is designed to enable recombinant protein expression following induction by isopropyl β-D-1-thiogalactopyranoside (IPTG).

The transformation into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* was performed in the same manner as the transformation into One Shot® TOP10 Chemically Competent *E. coli* (Section 3.8), with a few changes in reaction component volumes (detailed below).

Materials:

- Plasmid DNA (pNIC-CH containing gene of interest)
- One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells
- Super Optimal broth with Catabolite repression (S.O.C.) medium
- LB agar plates supplemented with kanamycin to 50 μg/mL (Section 2.10.1)
- Sarstedt 13 mL tubes

Method:

To transform the plasmid into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*, the BL21 Star™ (DE3) One Shot® Chemically Competent Cells transformation protocol from Invitrogen was used. The Basic Transformation Procedure (MAN0000662) was followed as written, with a few minor changes (detailed below). One vial of One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells was used for each transformation reaction.

The vials containing the One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells were gently thawed on ice. The vials were then quickly centrifuged in a MiniStar microcentrifuge, before 25 μL of the cells were transferred to pre-chilled Sarstedt 13 mL tubes. 2.5 μL of plasmid DNA was then added to the cells in the tubes, and the reactions were gently swirled to mix the components before being incubated on ice for 30 minutes. After the incubation, the reactions were heat-shocked by incubating the tubes for 30 seconds in a 42°C water bath, before placing the tubes back on ice for 2 minutes. Working at a laminar flow workbench, 75 μL S.O.C. medium was added to each tube. The tubes were then incubated at 37°C and 220 rpm for 1 hour.

Following the 1 hour-incubation, 20 μL of the reaction mixture was plated on room-temperature LB agar plates supplemented with kanamycin to a final concentration of 50 μg/mL. The plates were then incubated at 37°C overnight.

Colonies able to grow on the kanamycin plates were then used for inoculation of overnight cultures in LB liquid medium supplemented with kanamycin (Section 2.10.1).

The resulting overnight cultures were used in the preparation of glycerol stocks (Section 3.2), and for the inoculation of TB liquid medium to be used in bacterial cultivation and protein expression (Section 3.12).

3.12. Protein Expression

Bacterial cultivation in TB liquid medium and subsequent protein expression following isopropyl β-D-1-thiogalactopyranoside (IPTG)-induction was performed using a LEX-48 Bioreactor. This system allows for bacterial cultivation and protein expression with combined aeration and mixing, along with temperature control during growth.

The addition of IPTG promotes protein expression by binding to the repressor molecule produced by transcription of the *lacI* gene in the pNIC-CH vector (Appendix A). Without IPTG-induction, the repressor binds to the *lac* operator, blocking binding of the T7 RNA Polymerase (produced by the host organism) to the T7 promoter. When the repressor is bound to IPTG, however, it can no longer bind the *lac* operator, and the polymerase is therefore free to bind to the T7 promoter, resulting in the transcription of the inserted target gene.

Materials:

- TB liquid medium supplemented with kanamycin, antifoam, and phosphate solution (Section 2.10.2)
- 1 M IPTG
- One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* overnight culture

Method:

To express the wild-type and mutant recombinant proteins, 2 mL of overnight cultures of One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* containing the appropriate plasmid were used to inoculate 500 mL of autoclaved TB liquid medium. Prior to inoculation, the medium was supplemented with kanamycin, antifoam and phosphate solution, as specified in Section 2.10.2.

Cultures were then grown at room temperature with aeration for 24 hours in a LEX-48 Bioreactor. After 24 hours, protein expression was induced by adding 100 µL 1M IPTG, giving a final concentration of 0.2 mM. The cultures were then incubated for another 24 hours before the cells were harvested (Section 3.13).

3.13. Bacterial Harvesting, Cell Lysis, and Protein Extraction

Following protein expression as described in Section 3.12, the bacterial cells were harvested from the culture. The cells were then lysed via sonication to release the overexpressed protein into the solution. Sonication relies on the use of sound energy to disrupt bacterial cell walls and cell membranes, thereby releasing the cell's components into the solution. The protein can be extracted from the remaining cell debris through centrifugation, and purified from the protein extract using immobilized metal ion affinity chromatography (IMAC) (Section 3.14).

Materials:

- Ni-NTA Binding Buffer (Section 2.11.4)
- Bacterial culture containing expressed target protein

Method:

Cell Harvesting:

Harvesting was performed in several centrifugation steps to collect the bacterial cells in a pellet. Following protein expression, the 500 mL One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* bacterial culture was transferred to centrifuge bottles and centrifuged for 15 minutes at 5000 g in an Avanti™ J-25 Centrifuge using the JA-10 rotor. The cell pellet was resuspended in 50 mL of the supernatant; the rest of the supernatant was discarded. The suspension was then transferred to two 25 mL centrifuge tubes and centrifuged for 15 minutes at 5000 g in a Heraeus™ Multifuge™ X1R Centrifuge using the Fiberlite™ F15-8 x 50cy Fixed Angle rotor. This second centrifugation step was performed to transfer the pellet to the centrifuge tubes. The supernatant was discarded.

Cell Lysis:

In order to make the proteins available for purification, the bacterial cells were lysed to release the proteins. The pellet of harvested cells was frozen at -80°C for 1 hour to make the bacterial cell walls more brittle, thus facilitating subsequent cell resuspension and lysis. The cells were then thawed on ice and resuspended in 30 mL Ni-NTA Binding Buffer (Section 2.11.4), before being lysed by sonication using a Sonics Vibra-Cell™ Ultrasonic Processor. The sonication was done at 30% amplitude using pulses of 5 seconds for 3 minutes. The cell suspension was kept on ice throughout the sonication process.

Protein Extraction:

To separate the soluble proteins in solution from the remaining cell debris, the sonicated bacterial suspension was transferred to Cellstar® 50 mL centrifuge tubes and centrifuged for 15 minutes at 15000 g in a Heraeus™ Multifuge™ X1R Centrifuge using the Fiberlite™ F15-8 x 50cy Fixed Angle rotor. The pellet containing the ruptured bacterial cell walls and insoluble matter was stored at -20°C, while the supernatant containing the soluble components was filtered into a Cellstar® 50 mL centrifuge tube using a Filtropur S 0.45 µm PES Membrane syringe filtration unit.

3.14. Protein Purification – Immobilized Metal Ion Affinity Chromatography (IMAC)

Immobilized Metal Ion Affinity Chromatography (IMAC) is a chromatographic method used to isolate proteins with terminal His-tags. In protein purification, a solution containing the specific protein of interest is loaded onto a column prepackaged with a resin containing a metal ion to which the target protein carries an affinity. The protein then binds reversibly to the column resin, while proteins that lack an affinity for the metal ion are washed out. Elution of the protein is achieved by applying a solution containing a compound that can outcompete the binding of the target protein to the column material, in this case a Ni Sepharose matrix in a HisTrap™ High-Performance (HP) column.

In this study, all recombinant proteins were cloned into the pNIC-CH vector (Section 2.3.2 and Appendix A), which contains a region encoding a C-terminal 6xHis-tag separated from the protein of interest by a single alanine. All proteins in this study were therefore expressed with a C-terminal 6xHis-tag. Since histidine contains an imidazole ring which has a very high affinity for divalent metals such as Ni^{2+}, the target proteins were purified using a Ni^{2+} affinity HisTrap™ (HP) column prepacked with a Ni Sepharose matrix. A Ni-NTA Elution Buffer (Section 2.11.4) was used to elute the target proteins. This buffer has a high concentration of imidazole, which outcompetes the 6xHis-tag in Ni^{2+}-binding, thereby enabling the target protein to be eluted.

Chromatography was performed using an ÄKTA Pure chromatography system, which monitors the A_{280} of the column flow-through and visualizes this graphically, enabling detection of the target protein as it is eluted form the column. Chromatography results were visualized using the Unicorn™ 6.4.1 software.

Materials:

- Ni-NTA Binding Buffer (Section 2.11.4)
- Ni-NTA Elution Buffer (Section 2.11.4)
- Filtered recombinant protein in Ni-NTA Binding Buffer solution
- dH_2O

Method:

The HisTrap™ HP 5 mL column was attached to the ÄKTA Pure chromatography system. To remove any residual ethanol (the column storage buffer contains 20% ethanol), the system was first flushed with 5 column volumes (CV) dH_2O using a flow rate of 2 mL/minute. The column was then calibrated by flushing the system with 5 CV Ni-NTA Binding Buffer using the same flow rate. The sample containing the target protein was then loaded onto the column at a flow rate of 1 mL/minute. The slower flow rate employed during this step is presumed to facilitate the binding of all proteins with Ni^{2+} affinity to the column matrix. Following sample loading, additional Ni-NTA Binding Buffer was applied to the column, allowing for the removal of all non-bound or weakly-bound proteins. Once the A_{280} value had stabilized, indicating that all non-target proteins had been eluted, a linear gradient of Ni-NTA Elution

Buffer (0%-100%) was applied over 16 CV with a flow rate of 2 mL/minute. The increased concentration of imidazole in the column leads to the preferential binding of imidazole to Ni^{2+}, freeing the target protein and enabling it to be released from the column resin. Eluted target protein was collected using a fraction collector, collecting 4 mL eluent per fraction.

Fractions containing the highest concentrations of the protein of interest were subsequently analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to verify the identity of the target protein based on molecular weight in kDa (Section 3.15). Fractions putatively containing the target protein were identified based on observation of A_{280} during elution.

3.15. Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Once recombinant proteins had been purified via IMAC (Section 3.14), fractions determined to contain the target protein were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). This commonly-used electrophoresis technique allows for the separation of proteins according to their molecular weight.

While DNA molecules carry a net negative charge, facilitating their separation on a gel through the application of an electric current, the net charge of a protein is determined by its amino acid composition. To separate proteins by size using an electric current, proteins must therefore first acquire a net negative charge, enabling them to travel through a gel towards the positively charged anode. This is accomplished by using sodium dodecyl sulfate (SDS) or lithium dodecyl sulfate (LDS). These compounds disrupt a protein's tertiary structure by dsiturbing non-covalent bonds. The SDS or LDS compound, which carries a net negative charge, binds relatively evenly throughout the linearized protein, thus creating a proportional relationship between the protein's charge and molecular weight.

To facilitate protein denaturation and linearization, protein samples are mixed with SDS or LDS and a reducing agent such as dithiothreitol (DTT), which aids in the breaking of intramolecular disulfide bonds. The mixture is boiled before the protein samples are applied

to a pre-cast SDS-PAGE gel. The gel is then run in a Tris-Glycine-SDS (TGS) buffer. The SDS in the buffer ensures that the target proteins remain denatured and negatively charged throughout the run.

In this study, all IMAC-purified protein fractions of interest were run on Mini-PROTEAN® TGX Stain-Free™ Precast Gels. 4X NuPAGE® LDS Sample Buffer and 10X NuPAGE® Sample Reducing Agent were mixed in a working solution, which was added to all samples before boiling. In addition to LDS, the 4X NuPAGE® LDS Sample Buffer also contains a loading dye, enabling subsequent visualization of the protein bands following migration through the gel. A BenchMark™ Protein Ladder consisting of proteins of known molecular weight in kDa was run alongside all samples to enable size comparison and determination. All gels were run in a Mini-PROTEAN® Tetra Cell using a PowerPac™ 300 power supply.

The cell pellet generated during the final centrifugation step during protein extraction (Section 3.13) was also examined via SDS-PAGE. This was done to confirm the successful lysis of the harvested bacterial cells, as well as to ensure the target proteins had not aggregated within the bacterial cytoplasm, forming inclusion bodies.

Similarly, the flow-through generated during protein purification via IMAC was also analyzed via SDS-PAGE. The flow-through contains all proteins that do not bind or bind weakly to the column during sample application. The flow-through is examined using SDS-PAGE to verify that the target protein was not eluted during the Ni-NTA Binding Buffer wash step.

Materials:

- Protein fractions resulting from IMAC purification (Section 3.14)
- IMAC flow-through (Section 3.14)
- Pellet generated via centrifugation during protein extraction (Section 3.13)
- Tris-Glycine-SDS (TGS) Buffer (1X)
- NuPAGE® LDS Sample Buffer (4X)
- NuPAGE® Sample Reducing Agent (10X)
- BenchMark™ Protein Ladder
- Mini-PROTEAN® TGX Stain-Free™ Precast Gel

Method:

Preparing the LDS/Reducing Agent Working Solution:

To prepare the LDS/Reducing Agent solution to be added to all samples before boiling, 50 µL of the 4X NuPAGE® LDS Sample Buffer was mixed with 10 µL of the 10X NuPAGE® Sample Reducing Agent and 40 µL dH_2O in a 1.5 mL Eppendorf tube. This gave a final concentration of 2X NuPAGE® LDS Sample Buffer and 1X NuPAGE® Sample Reducing Agent. The 4X NuPAGE® LDS Sample Buffer contains Coomassie G250 and phenol red loading dyes.

Preparing the protein samples:

10 µL of each IMAC protein fraction or of the IMAC flow-through was mixed with 10 µL of the LDS/Reducing Agent Working Solution in a 1.5 mL Eppendorf tube. When preparing pellet samples, a pipette tip was brought into contact with the pellet, and then immersed in 10 µL dH_2O in a 1.5 mL Eppendorf tube. 10 µL LDS/Reducing Agent Working Solution was then added to the tube. All tubes were then boiled for 10 minutes before applying the samples to the pre-cast gels.

Preparing the electrophoresis chamber and loading the samples:

A Mini-PROTEAN® TGX Stain-Free™ Precast Gel was inserted into a cassette and placed into the Mini-PROTEAN® Tetra Cell. 1X TGS Buffer was then added to the inner chamber and the outer chamber. 10 µL of the ladder and 15 µL of the samples were then loaded into the wells on the gel. Note that LDS/Reducing Agent Working Solution was not added to the BenchMark™ Protein ladder before application onto the gel, since this ladder already contains a loading dye.

Running the gel and visualizing the protein bands:

The gel was run for 15 minutes at 300V. The current was applied to the chamber using a PowerPac™ 300 power supply. Once the run had been completed, gel imaging was conducted using a Stain-Free Sample Tray and a Gel Doc™ EZ Imager. Gel images were generated using the Image Lab™ Software.

Once the purity of the fractions containing the target protein had been analyzed, and the size of the target protein had been confirmed by visualizing the gel, the relevant IMAC fractions were pooled and concentrated via ultrafiltration (Section 3.16).

3.16. Ultrafiltration and Buffer Exchange

IMAC fractions containing the target protein were pooled and subjected to ultrafiltration to concentrate the diluted protein. A simultaneous buffer exchange was also performed from Ni-NTA Elution Buffer to a 20 mM Tris-HCl buffer, pH 8.0 (Section 2.11.3).

Ultrafiltration is a technique used to separate various macromolecules in a solution. Via centrifugation, the solution is pushed through a semipermeable membrane with a specific molecular weight cut-off (MWCO). This value is indicative of the membrane pore size. Molecules with a lower molecular weight than the membrane MWCO value will pass through the membrane into the permeate. By selecting an ultrafiltration membrane with a MWCO value smaller than the molecular weight of the target protein, a solution containing the protein

of interest can effectively be concentrated by removing superfluous buffer and other molecules via filtration.

In this study, protein ultrafiltration and concentration was performed with a simultaneous buffer exchange. A total of 40 mL 20 mM Tris-HCl buffer, pH 8.0 was applied to the membrane in 5-15 mL increments throughout the ultrafiltration. Between each buffer addition, the protein solution was centrifuged to approximately 2 mL. The buffer exchange process is performed to remove excess imidazole and salts from the protein solutions, which may be undesirable in potential protein crystallization experiments or enzymatic assays.

The molecular weight of all recombinant proteins in this study was approximately 31 kDa (see ***Table 4.1.2*** for an overview of physiochemical properties of all proteins generated during this study). A Vivaspin® 20, 10,000 MWCO Polyethersulfone (PES) ultrafiltration unit was therefore used in the ultrafiltration of all proteins in this study. The centrifugation was performed at 4500 g in a a Heraeus™ Multifuge™ X1R.

Materials:

- Protein fractions resulting from IMAC purification (Section 3.14)
- 20 mM Tris-HCl Buffer, pH 8.0 (Section 2.11.3)
- Vivaspin® 20, 10,000 MWCO PES Ultrafiltration Unit
- dH_2O

Method:

To rinse the membrane before beginning the protein ultrafiltration, 5 mL dH_2O was applied to the Vivaspin® 20, 10,000 MWCO PES Ultrafiltration Unit and the unit was centrifuged for 2 minutes.

The IMAC-purified fractions containing the protein of interest were then pooled in a Cellstar® 50 mL centrifuge tube, and 10 mL of this pooled solution was added to the ultrafiltration unit, along with 5 mL 20 mM Tris-HCl Buffer, pH 8.0. The solution was mixed by pipetting, and the unit was centrifuged until the volume reached approximately 2 mL.

The remainder of the pooled target protein fractions, as well as a further 35 mL 20 mM Tris-HCl Buffer, pH 8.0 were then added to the ultrafiltration unit gradually in 5/10 mL increments during continuous centrifugation, until the final volume was 1.5 mL.

Note that when adding the 20 mM Tris-HCl Buffer, pH 8.0 to the ultrafiltration unit, careful pipetting was performed close to the bottom of the unit to prevent clogging of the membrane and to thoroughly mix the solution. In addition, the protein solution was carefully monitored throughout the ultrafiltration to ensure that no protein aggregation occurred.

Once the target protein had been sufficiently concentrated, the protein concentration was determined using direct photometric measurement of A_{280} (Section 3.17).

3.17. Measuring Protein Concentration – A_{280}

Following ultrafiltration of target proteins, a D30 BioPhotometer® was used to measure protein concentration. The protein concentration of a sample is estimated by measuring its UV absorbance at 280 nm, A_{280}. By subsequently using the Beer-Lambert law, $A = \varepsilon * b * c$, the sample's protein concentration can be determined. The Beer-Lambert law relates the absorbance of a sample to the path length through which the UV light travels and to the sample's concentration. A represents the sample's absorbance, ε is the extinction coefficient, b is the path length through which the light travels, and c is the sample's concentration.

The extinction coefficient ε of each recombinant protein was determined using the ProtParam tool (Gasteiger et al. 2005). This tool allows for the calculation of various physiochemical parameters of a protein, including its theoretical extinction coefficient ε, based on its amino acid sequence. In order for this method to be precise, it is important that the protein contains aromatic amino acids, especially tryptophan. The mgCel6AΔCBM protein contains six tryptophans, eight tyrosines, and six phenylalanines (see Appendix D for mgCel6AΔCBM amino acid sequence).

Note that in this study, the path length b was always equal to 1 cm.

Materials:

- Concentrated target protein (Section 3.16)
- dH_2O
- Disposable cuvette

Method:

Diluted samples to be measured in the spectrophotometer were prepared in 0.2 mL PCR tubes. Since the concentration was variable amongst the different recombinant proteins produced during this study, protein samples were diluted between 50X and 200X using 20 mM Tris-HCl buffer, pH 8.0 to obtain an A_{280} value between 0.1 and 1.0. 100 µL of the diluted sample was transferred to a disposable cuvette which was subsequently placed in the D30 BioPhotometer® instrument. The A_{280} value of each target protein was then measured in five parallels.

To determine the target protein's concentration in mol/L, the protein's average A_{280} value was divided by its extinction coefficient ε, determined using ProtParam. This value was then multiplied by the dilution factor and divided by the protein's molecular weight in Dalton (Da), giving the final protein concentration in mol/L.

Note that in this study, the ProtParam extinction coefficient value for "assuming all pairs of Cys residues form cysteines" was used.

Following protein concentration determination, proteins were stored at 4°C.

3.18. Enzyme Activity Assays

Throughout this study, the wild-type and mutant mgCel6AΔCBM variants were tested in a variety of enzyme activity assays, and while numerous different assay conditions were used, two main categories of activity assays were performed: pH optimum assays (pH-activity

profiles) and pH stability assays. In addition, progress curve assays were also performed to investigate trends in product formation over time.

3.18.1. Substrates

During the study, enzyme activity was assessed on three different substrates (***Table 2.8.1***). The principal substrate was the Borregaard AS-generated BALI™ cellulose (Section 1.2.4), towards which the study was aimed. Activity was also assessed on the model substrates carboxymethyl cellulose (CMC) and phosphoric acid-swollen cellulose (PASC). Substrate solutions and/or suspensions were prepared in dH_2O. The final substrate concentrations were 1% (w/v) in reactions with BALI™ cellulose and CMC, and 0.2% (w/v) in reactions with PASC. A lower concentration of PASC was used due to a shortage of this substrate in the lab.

3.18.2. Reaction Conditions

Enzyme assays were performed either in 96-well Microtiter® plates with plastic sealing tape for shorter 15-minute reactions, or in 2 mL screw cap micro tubes for 24-hour reactions. 150 mM phosphate-citrate buffers of varying pHs were used in all activity assays (Section 2.11.2; ***Table 2.11.1***).

Activity assays were performed in a Thermomixer™ C with shaking: 750 rpm was used in reactions with BALI™ cellulose, 600 rpm in reactions with PASC, and 200 rpm in reactions with CMC. Pre-incubations for pH stability assays were performed in a Mastercycler Gradient PCR machine.

mgCel6A optimum temperatures for varying substrates had previously been determined (Section 1.7.2.2.2) (Jensen et al. 2017). These were found to be 75°C for 15-minute reactions with BALI™ cellulose, 60°C for 24-hour reactions with BALI™ cellulose, and 85°C for 15-minute reactions with CMC. pH optimum assays were therefore carried out at these temperatures when using the respective substrates. The optimum temperature for 15-minute reactions with PASC was not determined; these experiments were therefore performed at 60°C.

Reactions with CMC were stopped by the addition of a 2X total reaction volume of 3,5-dinitrosalicylic acid (DNS) reagent (Section 2.11.5).

24-hour reactions with the BALI™ cellulose performed in screw cap micro tubes were stopped by boiling samples for 10 minutes, whereas15-minute reactions with BALI™ cellulose performed in Microtiter® plates were stopped by adding a 1X total reaction volume of 0.1 M NaOH. After the reactions had been stopped, samples were vacuum filtered using Multiscreen®$_{HTS}$ 96-well plates to isolate the soluble reaction products from the remaining insoluble substrate. Finally, in the case of the 24-hour reaction samples, DNS reagent was added in the same volumetric ratio as described above.

Reactions with PASC were first filtered to isolate soluble reaction products (in the same manner as described above), before the reactions were stopped by the addition of a 2X total reaction volume of 3,5-dinitrosalicylic acid (DNS) reagent.

Subsequent quantification of product formation was performed using the DNS method for reactions stopped with the DNS reagent (Section 3.19.1), or via high-performance anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) for reactions stopped using 0.1 M NaOH (Section 3.19.2).

Reactions were performed in triplicates, unless stated otherwise. Enzyme blank samples were also assayed in triplicates in parallel to the reaction samples, consisting of 75 mM phosphate-citrate buffers of the relevant pHs used in the reactions, as well as 1% (w/v) CMC or BALI™ cellulose, or 0.2% (w/v) PASC, depending on which substrate was used in the enzymatic reactions. The total volumes of the blank samples were equal to the volumes of the enzymatic reactions.

3.18.3. pH Optimum Assays

The wild-type and mutant mgCel6AΔCBM variants were subjected to several pH optimum assays to determine their pH-activity profiles on various substrates. This assay determines the pH value at which the enzyme has the highest activity under the defined conditions (substrate, temperature, and time) of the assay. In this study, pH optimum assays were performed to

determine whether the mutant mgCel6AΔCBM variants had altered pH-activity profiles compared to the wild-type enzyme.

pH optimum assays were performed using all substrates mentioned above. When using CMC and PASC, the reaction time was 15 minutes due to the rapid degradation of these easily accessible substrates. When using the industrial BALI™ cellulose substrate, pH optimum assays were conducted with reaction times of 15 minutes or 24 hours.

To set up enzyme activity assays to generate pH-activity profiles, phosphate-citrate buffer of the appropriate pH was added to the wells of a Microtiter® plate or to micro tubes. To obtain a final buffer concentration of approximately 75 mM, the volume of buffer added was equal to half the total reaction volume. The relevant substrate (in water) was then added to the wells or micro tubes, giving either a 1% (w/v) final substrate concentration in the case of reactions with CMC or BALI™ cellulose, or a 0.2% (w/v) final substrate concentration in the case of reactions with PASC. Finally, the relevant enzyme was added to the reactions. Varying enzyme concentrations from 1 μM to 32 μM were used in the reactions performed throughout this study. Reactions were incubated for either 15 minutes or 24 hours, and stopped either by adding DNS reagent or 0.1 M NaOH, as described above. Product quantification was performed using the DNS method (Section 3.19.1) for reactions stopped with DNS reagent, or using HPAEC-PAD (Section 3.19.2) for reactions stopped with 0.1 M NaOH.

3.18.4. pH Stability Assays

In a pH stability assay, the target enzyme is subjected to pre-incubation in a buffer of relevant pH over varying lengths of time. Pre-incubated enzymes are then tested in a residual activity assay to determine how pH affects the enzyme's stability over time. In this study, pH stability assays were performed to determine whether the engineered mgCel6AΔCBM variants had increased stabilities at more acidic pH values as compared to the wild-type enzyme.

To set up pre-incubations for pH stability assays, phosphate-citrate buffer of varying pHs and enzyme were added to 0.2 mL PCR tubes. The total pre-incubation volume was 60 μL, with a final enzyme concentration of 2 μM and a buffer concentration of 150 mM. The PCR tubes were then incubated in a Mastercycler Gradient PCR machine at 60°C or 65°C. Samples were

removed at particular time points for up to 24 hours and subsequently stored at 4°C. Activity assays with 1% (w/v) CMC were then performed on the pre-incubated samples to assess the enzyme's residual activity. These reactions were prepared by adding 50 µL pre-incubated enzyme to wells in a Microtiter® plate containing 50 µL 2% (w/v) CMC, giving a final substrate concentration of 1% (w/v), a final enzyme concentration of 1 µM, and a final buffer concentration of 75 mM in the reactions. Note that the pHs of the reactions were not adjusted following pre-incubations. The reactions were incubated for 15 minutes and stopped by the addition of 200 µL DNS reagent. Product quantification was performed using the DNS method (Section 3.19.1).

Enzyme half-life was determined by identifying the pre-incubation time required to obtain 50% residual activity.

3.18.5. Progress Curve Assays

A progress curve of an enzymatic reaction shows the amount of product formed over time as a reaction proceeds. During the initial stages of the reaction, the enzyme is operating at constant rate, and is therefore generating product molecules at a linear rate. This rate is known as the initial rate of reaction, v_0. As the substrate is consumed by the enzyme, the rate gradually slows until all the substrate has been consumed. As long as the substrate is in excess, the rate of reaction should theoretically remain equal to the initial rate.

When performing activity assays to determine enzyme optima, product formation over the reaction time period must be linear. It is therefore necessary to verify that the enzyme is fully functional over the entire length of the reaction period, meaning that it does not become unstable or inactive during the assay and that substrate consumption is low relative to the substrate concentration used in the assay. A linear progress curve throughout the reaction enables the variation in enzyme activity to be attributed solely to the factor that is being varied in the assay, allowing the effect of this parameter on enzyme activity to be studied.

In pH optimum assays performed during this study, the effect of pH on the enzyme's activity was investigated (Section 3.18.3). Enzyme progress curves were generated to ensure that

product formation was linear during the reaction time period over which the pH optimum assays were conducted.

Progress curve assays were prepared in the same manner as the 15-minute pH optimum activity assays, however with samples prepared for removal after specific incubation times. Reactions were stopped by the addition of a 2X total reaction volume DNS reagent or a 1X total reaction volume 0.1 M NaOH, as described above. Reaction products were quantified using the DNS method (Section 3.19.1) for reactions stopped with DNS reagent, or using HPAEC-PAD (Section 3.19.2) for reactions stopped with 0.1 M NaOH. Varying temperatures and pHs were used in the progress curve assays to investigate the effects of these conditions on the enzyme's stability during the reaction time.

3.19. Quantification of Reaction Products

As mentioned above, analysis of the reaction products from the various enzyme assays was performed using two main methods: quantifying the amount of free reducing sugars in solution through a colorimetric assay, the DNS method (Section 3.19.1), or quantifying the concentration of specific reaction products using HPAEC with PAD (Section 3.19.2).

3.19.1. 3,5-Dinitrosalicylic Acid (DNS) Method

The 3,5-dinitrosalicylic acid (DNS) method can be used to quantify sugars with reducing ends in a solution. In a basic environment, DNS oxidizes the aldehyde and ketone groups in reducing sugars to carboxylic acid groups. During the oxidation, DNS is reduced to 3-amino-5-nitrosalicylic acid (ANS). This compound has a distinct orange color which is reflected in its strong absorption at 540 nm.

The main reaction product of the mgCel6A enzyme is cellobiose (Glc2) (Jensen et al. 2017). By using cellobiose (Glc2) standards of known concentrations (0.1, 0.2, 0.5, 1.0, 1.5, and 2.0 mg/mL) to create a standard curve, the A_{540} of the reaction samples containing various amounts of reduced ANS can be converted to relative amounts of cellobiose, thus enabling the quantification of reaction products.

Materials:

- DNS Reagent (Section 2.11.5)
- Reaction products to be analyzed
- Cellobiose (Glc2) Standards (0.1, 0.2, 0.5, 1.0, 1.5, and 2.0 mg/mL)

Method:

Stopping reactions with DNS reagent:

Reactions were stopped by adding a 2X volume of DNS reagent. DNS reagent is highly basic, and the addition of this solution to the reactions causes a sharp rise in pH, stopping enzymatic catalysis.

When using the DNS method to quantify products from reactions with the soluble CMC substrate, 200 μL DNS reagent was added directly to 100 μL of the reaction samples.

When using DNS reagent with the insoluble BALI™ cellulose substrate for the longer reaction times, the reactions were stopped by boiling for 10 minutes to denature the enzyme. The samples were then subjected to vacuum filtration to isolate the soluble reaction products, and 200 μL DNS reagent was added to 100 μL soluble filtered product.

When using DNS reagent with the insoluble PASC substrate in shorter 15-minute reactions, the samples were subjected to vacuum filtration to isolate the soluble reaction products from the remaining insoluble substrate. 200 μL DNS was then added to 100 μL soluble filtered product.

200 μL DNS reagent was also added in the same manner as described above to 100 μL of the enzyme blank samples containing only buffer and substrate (boiled and subsequently filtered prior to addition of DNS reagent when using BALI™ cellulose; filtered prior to addition of DNS reagent when using PASC).

Standard Glc2 samples were prepared by adding 200 µL DNS reagent to 100 µL of cellobiose standard solutions of varying concentrations from 0.1 mg/mL to 2.0 mg/mL.

Following the addition of DNS reagent, samples were incubated at 100°C for 15 minutes with 300 rpm. The high reaction temperature allows the reducing sugars to be oxidized by DNS, which is concomitantly reduced to ANS.

Measuring A_{540}:

After cooling the samples to room temperature, the absorbance of all samples at 540 nm was measured. The average A_{540} measurement of the blank samples was subtracted from A_{540} measurements of the reaction samples, while the average A_{540} of samples containing only dH_2O was subtracted from the cellobiose standard samples.

A standard curve was then made by plotting the A_{540} values of the cellobiose standard samples against their cellobiose concentration in mg/mL. This standard curve was used to estimate the cellobiose concentration of the reaction samples, thus enabling quantification of product formation.

3.19.2. High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)

High-performance anion exchange chromatography (HPAEC) is an analytical chromatographic technique used to separate and quantify oligosaccharides of varying length within a mixture. The separation is based on anion exchange using a column matrix containing positively charged groups. When a sample containing negatively charged compounds is applied to the column, these will be retained by electrostatic interactions with the column matrix. Elution is subsequently achieved by applying an eluent containing a high concentration of another negatively charged group, which will outcompete the sample compounds, causing these to be eluted from the column.

Pulsed amperometric detection (PAD) can then be applied to detect the concentration of ions in the column flow-through. Amperometry uses an electric current to detect changes in the ionic concentration of a solution.

A sample containing glucose and various forms of oligosaccharides can be separated based on the number of hydroxyl groups. Longer oligosaccharides carry more hydroxyl groups, and, upon deprotonation in the highly alkaline mobile phase, will thus contain more negative charges than shorter oligosaccharides. This enables the longer oligosaccharide chains to form stronger electrostatic interactions with the column matrix. When an eluent is applied to the column using an increasing gradient, the longer oligosaccharides will remain bound to the column for longer. This ensures that the mono- and oligosaccharides are separated according to chain length, allowing for the analysis of the carbohydrate composition of a sample via PAD.

By also analyzing standard samples of varying concentrations of glucose, cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose (Glc1-Glc6), standard curves can be generated for each of these mono- and oligosaccharides, thus enabling quantification of each of these potential reaction products from a reaction sample by comparison with the known concentrations of the samples used to make the standard curves.

Chromatography was performed using an ICS-3000 System with PAD and the Chromeleon® 7 Chromatography Data System software.

Materials:

- Reaction Products to be analyzed
- HPAEC Eluent A (Section 2.11.6)
- HPAEC Eluent B (Section 2.11.6)
- HPAEC Eluent C (Section 2.11.6)
- Dionex™ CarboPac™ PA1 Column
- Dionex™ ICS-3000 system with PAD detection
- Glc1-Glc6 Standard Solution (0.005, 0.01, 0.025, 0.05, and 0.1 mg/mL)

Method:

Stopping reactions with 0.1 M NaOH:

In this study, a 1X volume of 0.1 M NaOH (equivalent to the volume of the reaction mixture) was used to stop 15-minute enzymatic reactions with BALI™ cellulose substrate. Samples were subsequently subjected to vacuum filtration to separate the soluble reaction products from the insoluble substrate.

A 1X volume of 0.1 M NaOH was also added in the same manner as described above to the enzyme blank samples containing only buffer and BALI™ cellulose substrate, which were then subjected to vacuum filtration.

The highly alkaline 0.1 M NaOH solution stops the reaction by inactivating the enzyme, but also serves to deprotonate soluble reducing sugars and ionize the hydroxyl groups. These are weak acids with relatively high pK_a values, and are therefore deprotonated by reaction with a strongly basic solution. When the ionized carbohydrates are applied to the CarboPac™ PA1 column containing positively charged quaternary ammonium ions, electrostatic interactions will lead to retention of the sugars within the column.

To prepare the HPAEC-PAD standards, a 1 mg/mL stock solution of glucose, cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose (Glc1-Glc6) was used. Standards containing 0.005, 0.01, 0.025, 0.05, and 0.1 mg/mL of Glc1-Glc6 were made by diluting this stock solution with dH_2O.

Samples were transferred to 0.3 mL Snap Ring Micro-Vials before subsequent HPAEC-PAD analysis.

HPAEC-PAD Analysis:

The CarboPac™ PA1 column was first washed with Eluent C, and then calibrated using only Eluent A. An automatic sampler was used to apply the samples onto the column. Using a flow rate of 0.25 mL/minute, the carbohydrate products in each sample were then separated using a linear gradient in several steps, with a gradually increasing concentration of Eluent B. The

eluents were applied in the following concentrations: 0.1 M NaOH to 0.1 M NaOH and 0.1 M CH_3COONa over the course of 10 minutes, increased to 0.1 M NaOH and 0.14 M CH_3COONa over the course of 5 minutes, increased to 0.1 M NaOH and 0.3 M CH_3COONa over the course of 1 minute, followed by an increase to 0.1 M NaOH and 1.0 M CH_3COONa over the course of 2 minutes. The column was then recalibrated using only Eluent A for 11 minutes before the next sample was applied.

The standards containing various concentrations of Glc1-Glc6 were used to generate standard curves, which enabled the subsequent quantification of each mono- and oligosaccharide product in the reaction samples.

While Glc2 is the main reaction product of mgCel6A reactions, minor amounts of Glc1 and Glc3 are also generated (Jensen et al. 2017). The average amounts of Glc1, Glc2, and Glc3 detected in the blank samples were therefore subtracted from the corresponding Glc1-Glc3 product concentrations calculated for the reaction samples based on the standard curves generated.

4. Results

4.1. Bioinformatics and Mutant Design

The domain architecture of mgCel6A was assessed using Pfam, InterPro, and dbCAN, to gain a full understanding of which domains were present in the enzyme (Finn et al. 2016; Finn et al. 2017; Yin et al. 2012). Since the three-dimensional structure of the enzyme had not (and at the present time, has not) yet been solved, a BLASTp search (Camacho et al. 2009; Zhang & Madden 1997) was performed to locate the enzyme's closest homolog with a solved structure. The structure of this protein, Cel6A from *Thermobifida fusca* (*Tf*Cel6A; PDB ID 1TML) (Berman et al. 2000), was used to construct a model of mgCel6AΔCBM using the homology modeling server SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008). The model of the enzyme was examined using the molecular visualization tool PyMOL (pymol.org). A multiple sequence alignment (MSA) was created to compare the amino acid sequences of mgCel6AΔCBM and other family 6 glycoside hydrolases (GH6s). The mgCel6AΔCBM model, the MSA, and literature on pH-engineering of enzymes were used as a basis for mutant design, as outlined below (for an overview of mutants, see ***Table 4.1.1***).

4.1.1. mgCel6A Domain Architecture

The domain architecture of the full-length mgCel6A enzyme was analyzed using the web servers Pfam, InterPro, and dbCAN (Finn et al. 2016; Finn et al. 2017; Yin et al. 2012), and is shown in ***Figure 4.1.1***. mgCel6A consists of a glycoside hydrolase family 6 (GH6) catalytic domain and a CBM2-domain, connected by a flexible proline- and threonine-rich linker region. The catalytic domain is preceded by a signal peptide consisting of 35 amino acids. The complete amino acid sequences of both mgCel6A and mgCel6AΔCBM can be found in Appendix D.

Figure 4.1.1. mgCel6A Domain Architecture. Full-length mgCel6A consists of a signal peptide of 35 amino acids (not shown), followed by a glycoside hydrolase family 6 domain comprising 288 amino acids (GH6, shown in blue), and a 104-amino acid long family 2 carbohydrate-binding module domain (CBM2, shown in green). The two domains are connected by a linker region (shown as a black line), consisting of 40 amino acids. The full amino acid sequence of the enzyme can be found in Appendix D. In this study and the work preceding it, mgCel6A, as well as mgCel6AΔCBM and all variants thereof, were produced without the signal peptide, and with a C-terminal His-tag.

4.1.2. Homology Modeling

To locate the enzyme's closest homolog, a BLASTp search using the catalytic domain of mgCel6A was performed by Marianne S. Jensen (Camacho et al. 2009; Zhang & Madden 1997). The search gave an 84% sequence identity with an unpublished cellulase from *Micromonospora rosaria*. The homolog with the highest sequence identity for which the structure had been solved was Cel6A from *Thermobifida fusca* (*Tf*Cel6A; PDB ID 1TML) (Berman et al. 2000). This 1,4-endoglucanase also consists of a GH6 catalytic domain and a CBM2 domain, however, the solved structure contains only the catalytic domain. The catalytic domain of *Tf*Cel6A shares a 70% sequence identity with mgCel6AΔCBM (Jensen et al. 2017).

*Tf*Cel6A was used to generate a structural model of mgCel6AΔCBM using the homology modeling server SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008). The alignment of the two catalytic domains created by SWISS-MODEL is shown in ***Figure 4.1.2***.

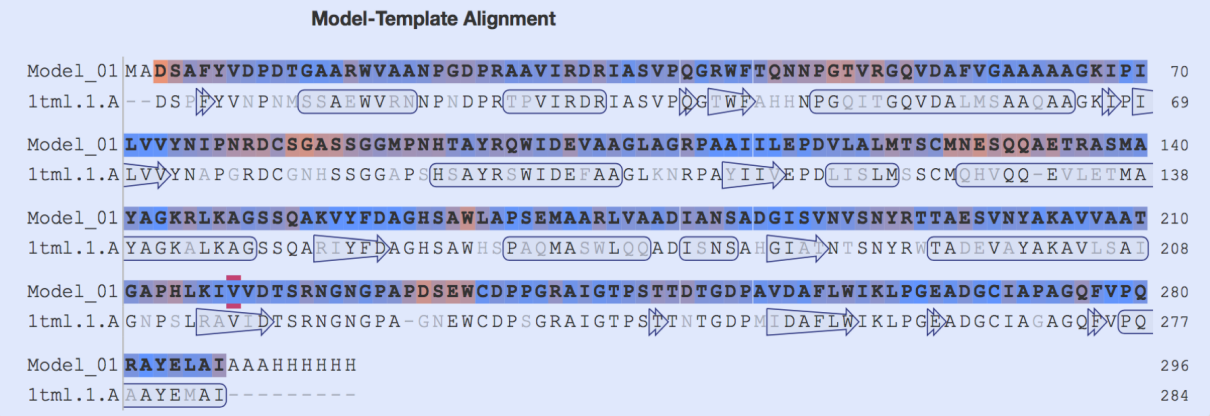


Figure 4.1.2. Alignment of the catalytic domains of *Tf*Cel6A and mgCel6A. mgCel6AΔCBM is designated as Model_01, while the template protein, the catalytic domain of *Tf*Cel6A, is designated as 1tml.1.A. The alignment shows two insertions in the mgCel6AΔCBM amino acid sequence, Ala133 and Pro230. The alignment was generated using the homology modeling server SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008). The alignment was used by the sever to build a model of the mgCel6AΔCBM enzyme, which can be seen in **Figure 4.1.3**. The blue regions of the alignment indicate higher quality area of the model, whereas red regions indicate lower quality regions of the model. Relevant model scores calculated by SWISS-MODEL are presented in Appendix F.

The alignment showed two residues in mgCel6AΔCBM (Ala133 and Pro230) not present in the template sequence. No other insertions or deletions were found in the alignment, and the resulting model was of good quality (QMEAN Score 0.54; GMQE Score 0.85; see Appendix F). The model was subsequently visualized using PyMOL (pymol.org), and is shown in ***Figure 4.1.3***.

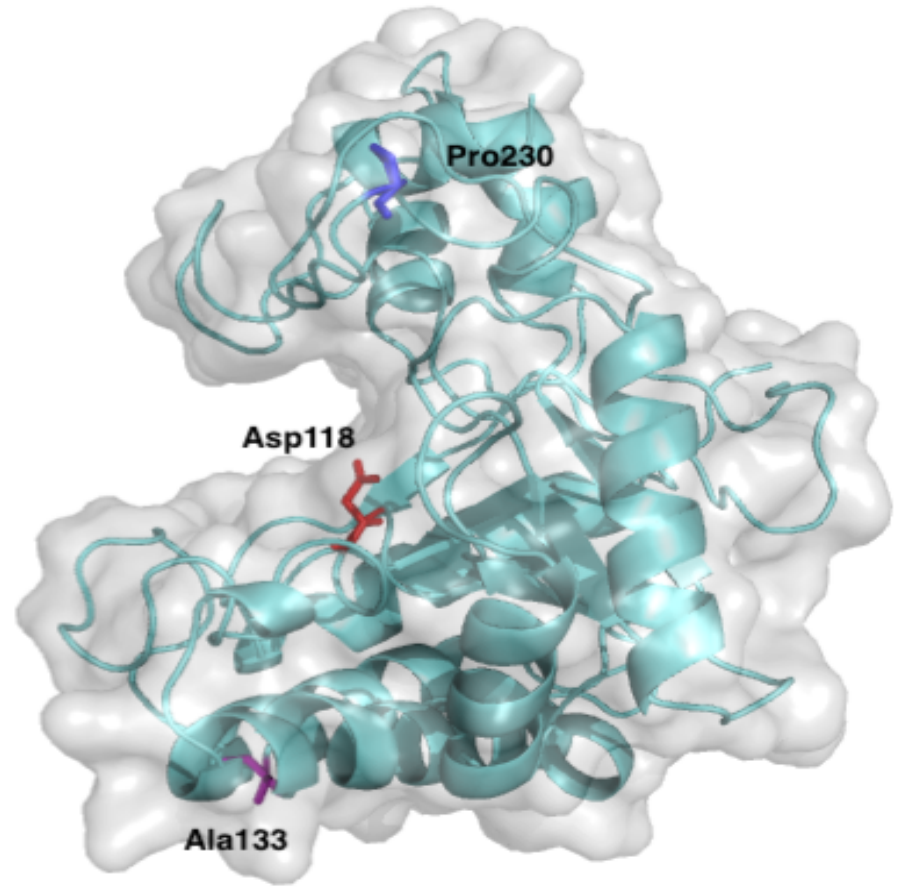


Figure 4.1.3. Structural Model of mgCel6AΔCBM. The model was built based on the structure of the *Tf*Cel6A ortholog. The model shows the enzyme from the side to demonstrate the deep catalytic cleft of the enzyme, and is displayed with a transparent surface and with visualization of secondary structure elements. The catalytic acid (Asp118) is labeled and its side chain is shown as a red stick. The two residues not present in the *Tf*Cel6A template protein, Ala133 and Pro230, are labeled and their side chains are shown as purple and blue sticks, respectively. The figure was made using PyMOL (pymol.org).

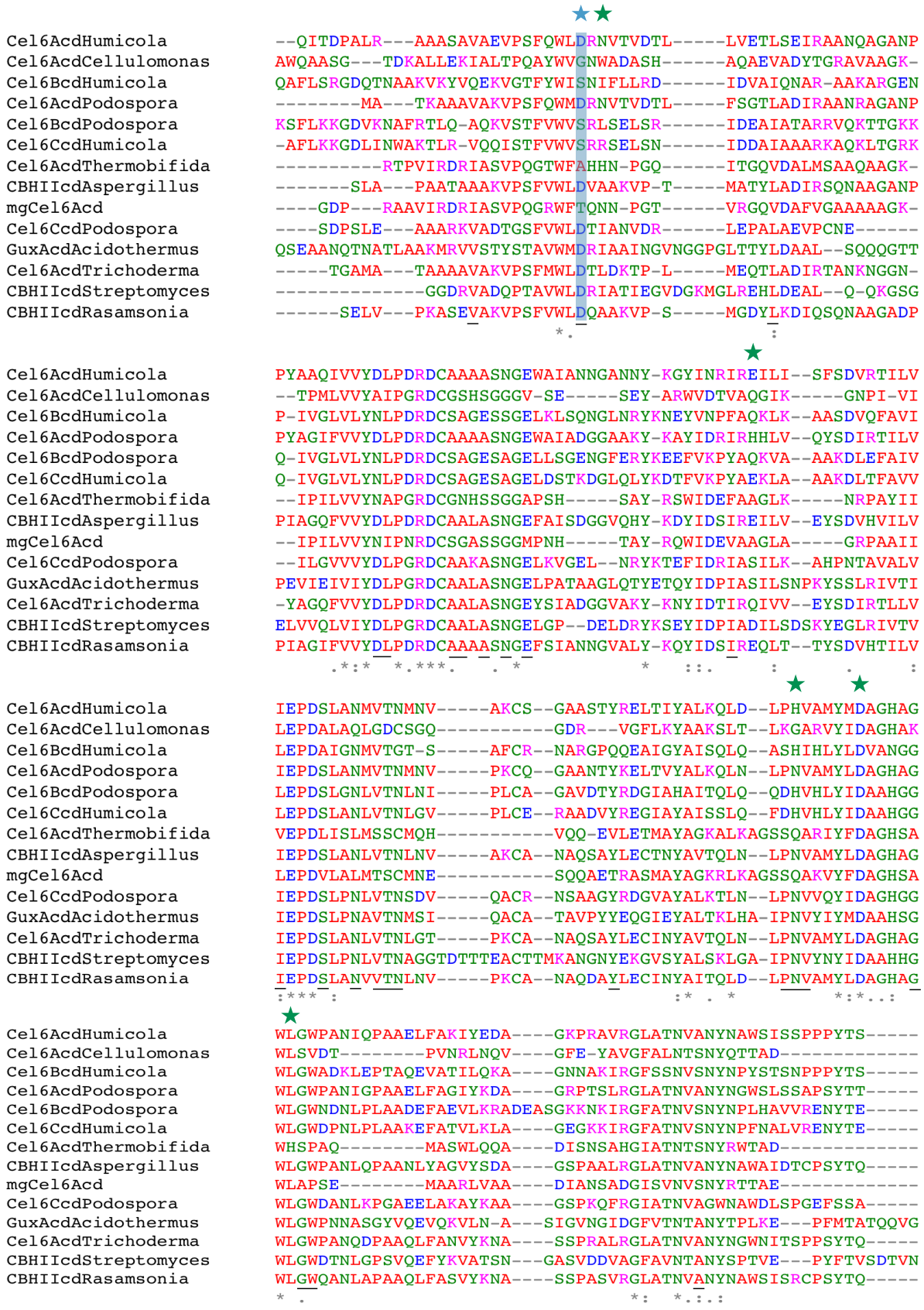
The model shows a classical open active site cleft typical of endoglucanases (Section 1.5.1), and showed that mgCel6AΔCBM contains eight distinct α-helices and seven short β-strands, conforming to the "modified α/β barrel fold" described in the Glycoside Hydrolase Family 6 CAZypedia page (Piens & Davies 2015). The Ala133 and Pro230 were both found to be approximately 18Å from the catalytic acid Asp118. Of the 54 charged residues (including histidines) in the amino acid sequence, the model revealed 47 of these to have their ionizable groups exposed to the solvent to some degree.

4.1.3. Multiple Sequence Alignment (MSA)

An MSA consisting of the catalytic domains of 14 biochemically characterized family 6 glycoside hydrolases (GH6s) (including mgCel6AΔCBM) was made using Clustal Omega (Sievers et al. 2011) to locate potential candidate residues for pH optimum mutants of mgCel6AΔCBM. This approach was largely based on a method utilized in a 2011 pH engineering study by D. W. Cockburn and A. J. Clarke, which aimed to alter the pH-activity profile of *Cf*Cel6A (previously known as CenA), a GH6 cellulase found in *Cellulomonas fimi*. By creating an MSA of GH6 enzymes ordered by descending pH optima, the study identified several amino acids that were conserved amongst the enzymes with a lower pH optimum than the target enzyme, *Cf*Cel6A. The corresponding residues in *Cf*Cel6A were mutated to the conserved amino acids found in the enzymes with lower pH optima, and the mutants were subsequently characterized in a pH optimum assay on 0.4% (w/v) CMC (Cockburn & Clarke 2011).

The 2011 study was based on an MSA consisting of 12 enzymes, including the target enzyme *Cf*Cel6A. The alignment enabled the selection of six mutation candidates, as well as the discovery of an eight-amino acid-long sequence that was conserved in all enzymes with lower pH optima than the target enzyme.

In the present study, the MSA created included eight of the GH6 enzymes used in the 2011 study, as well as an additional five GH6s from the cellulolytic fungi *Humicola insolens* and *Podospora anserina* (***Figure 4.1.4***).


Cel6AcdHumicola
Cel6AcdCellulomonas
Cel6BcdHumicola
Cel6AcdPodospora
Cel6BcdPodospora
Cel6CcdHumicola
Cel6AcdThermobifida
CBHIIcdAspergillus
mgCel6Acd
Cel6CcdPodospora
GuxAcdAcidothermus
Cel6AcdTrichoderma
CBHIIcdStreptomyces
CBHIIcdRasamsonia

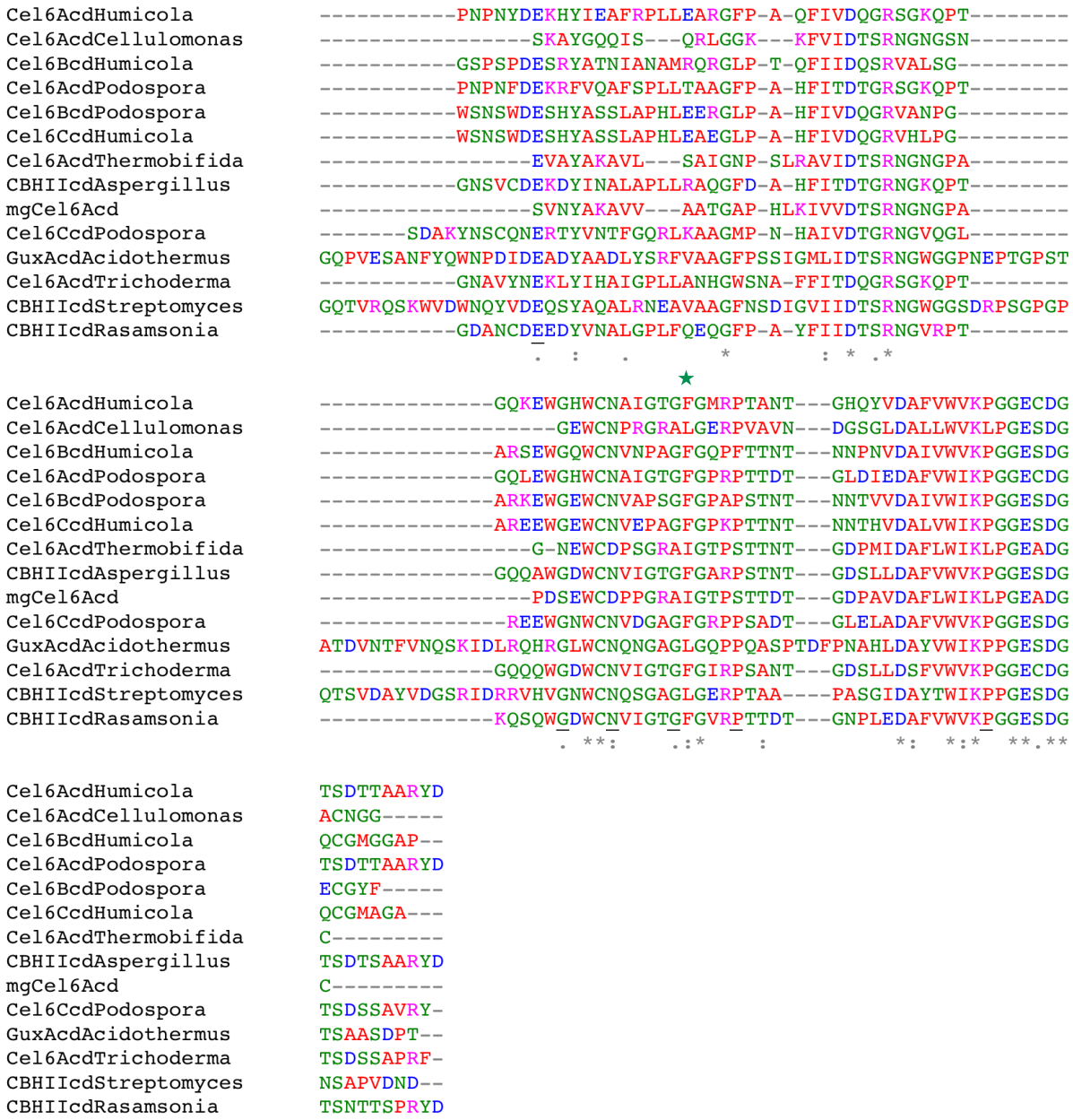


Figure 4.1.4. Multiple Sequence Alignment of mgCel6AΔCBM and Family 6 Glycoside Hydrolases with known pH Optima. The enzymes are listed according to descending pH optimum (see below). The position containing the Thr44 residue, which was selected for mutation based on this MSA, is denoted with a blue star, and the corresponding position of all the enzymes is highlighted in blue. Positions where the enzymes with lower pH optima than mgCel6AΔCBM have fully conserved residues are underlined. Other positions that were mutated in this study are indicated by green stars (note that the Asn21, Ile271, and Gln276 residues, mutated in the M1, M8, and M4 mutants, respectively, are not shown, since the alignment does not cover these positions in the sequence). The MSA was created using Clustal Omega (Sievers et al. 2011), and the colors in the MSA denote various physiochemical properties of the amino acids: red represents small, hydrophobic amino acids, and includes all aromatic amino acids except tyrosine; blue denotes the acidic amino acids aspartate and glutamate; pink represents the basic amino acids lysine and arginine (note: not histidine); green represents the amino acids carrying hydroxyl, sulfhydryl, and amide groups, histidine, and glycine, the only achiral amino acid. Fully conserved residues are marked with an asterisk (*); residues with strongly associated properties are denoted with a colon (:); residues with weakly associated properties are marked with a period (.).

The enzymes included in the MSA and their pH optimum in descending order, as shown in ***Figure 4.1.4***, are: *Humicola insolens* Cel6A, pH optimum 9.0 (on 0.1 mM reduced cellohexaose) (Schülein 1997), *Cellulomonas fimi* Cel6A, pH optimum 7.5 (on 0.4% (w/v) CMC) (Cockburn et al. 2010), *Humicola insolens* Cel6B, pH optimum 7.5 (on 7.5 g/L CMC) (Schülein 1997), *Podospora anserina* Cel6A, pH optimum 7.0 (on 1% (w/v) Avicel®) (Poidevin et al. 2013), *Podospora anserina* Cel6B, pH optimum 7.0 (on 1% (w/v) Avicel®) (Poidevin et al. 2013), *Humicola insolens* Cel6C, pH optimum 6.5 (on 1% (w/v) barley β-glucan) (Xu et al. 2015), *Thermobifida fusca* Cel6A, pH optimum 6.0 (on 0.2% (w/v) PASC) (Wolfgang & Wilson 1999), *Aspergillus nidulans* CBHII (AN5282.2), pH optimum 5.5 (on unknown concentration of CMC) (Bauer et al. 2006), mgCel6AΔCBM, pH optimum 5.0 (on 1% (w/v) CMC) (***Figure 4.3.1***), *Podospora anserina* Cel6C, pH optimum 5.0 (on 1% (w/v) Avicel®) (Poidevin et al. 2013), *Acidothermus cellulolyticus* GuxA, pH optimum not specified, but enzyme described as "active to a pH well below 5.0" (on unknown substrate; unknown reaction time) (Ding et al. 2003), *Hypocrea jecorina* Cel6A, pH optimum 5.0 (on unknown concentration of cellotetraose) (Koivula et al. 2002), *Streptomyces* sp. 43 CBHII, pH optimum 4.0 (on 0.5% (w/v) cellopentaose) (Park et al. 2005), and *Rasamsonia emersonii* CBHII, pH optimum 3.8 (on unknown concentration of 4-Methylumbelliferyl-β-cellotriopyranoside ($MeUmbG_3$)) (Tuohy et al. 2002).

Several residues were found to be universally conserved in the GH6 cellulases with lower pH optima than mgCel6AΔCBM. Of these, conserved charged residues close to the active site were considered most interesting, since they may potentially affect the local charge network of the active site, perhaps by altering pK_a values of critical residues, or by changing the active site's substrate-binding properties. In conjunction with structural studies of the target enzyme model, all but one of the conserved charged amino acids in the corresponding positions in mgCel6AΔCBM were deemed to be too distant from the active site to have any impact on the enzyme catalytic mechanism or pH optimum (see Section 5.5.3 for discussion). The residue that was considered a potential candidate due to its relative proximity to the substrate binding site was Thr44, which was a conserved aspartate in the sequences of the enzymes with lower pH optima than mgCel6AΔCBM. This amino acid was therefore selected as a candidate for an active site mutation (Section 4.1.4.2).

4.1.4. Mutation Criteria and Selecting Mutation Sites

The rational mutation design was based on studies of pH stability and optimum engineering literature, inspections of the structural model, and the MSA of GH6s with known pH optima. Several rational strategies were employed to design mutants of mgCel6AΔCBM, with the aim of reducing the enzyme's pH optimum and increasing its pH stability. The precise goal was to create a mutant mgCel6AΔCBM enzyme capable of operating at pH 5.0 or lower and 60°C or higher for more than 24 hours. ***Table 4.1.1*** shows an overview of all mutants designed in this study.

Table 4.1.1. mgCel6AΔCBM mutants. The table shows the mutant name, the mutation(s), and whether the mutation was inserted on the surface of the enzyme or close to its active site. See Appendix E for the amino acid sequences of all mutants mgCel6AΔCBM variants.

Mutant	Mutation	Surface (S) or Active Site (A) Mutant
M1	N21D	S
M2	N46D	S
M3	Q152E	S
M4	Q276E	S
M5	A104E	S
M6	L165E	S
M7	I242E	S
M8	I271D	S
M9	N21D/N46D/A104E/Q152E/ L165E/I242E/I271D/Q276E	S
M10	T44D	A
M11	D158A	A
M12	D158N	A

Mutants 1-4 were designed to introduce negative charge on the surface of the enzyme. Mutants 5-8 were designed to form surface salt bridges (ion pairs). Mutant M9 contained all eight single point mutations from mutants 1-8. Finally, mutants 10-12 carried mutations introduced near the catalytic site of the enzyme (***Figure 4.1.3***).

4.1.4.1. Design of Surface Mutants M1-M9

4.1.4.1.1. Mutants M1-M4

An enzyme's solubility and stability are reliant, amongst many other properties, upon its net charge, i.e. the sum of all its charged residues. The final proportion of basic to acidic amino acids on the surface of an enzyme will dictate numerous physiochemical properties, including its solubility and stability in solvents of various temperatures, pHs, salt concentrations, etc. While the roles of various intramolecular interactions in maintaining protein stability have been studied for many years (Lindman et al. 2006; Sanchez-Ruiz & Makhatadze 2001; Sippel & Quiocho 2015; van den Burg & Eijsink 2002), notably, the contribution towards stability of isolated surface charges that are not participants in ionic bonds is not very clear.

The theoretical pI of mgCel6AΔCBM was calculated to be 5.96 by ProtParam (***Table 4.1.2***) (Gasteiger et al. 2005). At this pH, the net charge of the protein is neutral. As the pH of the environment decreases, the negatively charged residues in the enzyme will gradually become protonated, altering the net charge of the protein. This will eventually lead to an excess of positive charges on the enzyme surface, which may encounter significant repulsion both amongst themselves through long-range interactions, and with positive dipoles in the surrounding local environment. It is conceivable that this increase in positive charge on the enzyme surface may lead to instability.

In an attempt to counteract this accumulation of positive charge, mutants M1-M4 were designed to introduce negative charges on the enzyme surface. The idea was that an increase in surface negative charges would reduce the pH value required to destabilize the enzyme based on accumulation of positive surface charges (V.G.H. Eijsink; personal communication).

The residues selected for mutation using this strategy were asparagines or glutamines on the surface of mgCel6AΔCBM, which were mutated to aspartates or glutamates, respectively.

The residues targeted for mutation were not presumed to be involved in any ionic bonds, and were located a minimum of 20Å from the enzyme's catalytic acid residue (Asp118). Asparagines and glutamines were selected with the idea of minimizing side effects, since these amino acids are identical in structure to the negatively charged residues aspartate and glutamate, with the exception of the terminal carboxylic acid group. ***Figure 4.1.5*** shows the mutations selected using this approach.

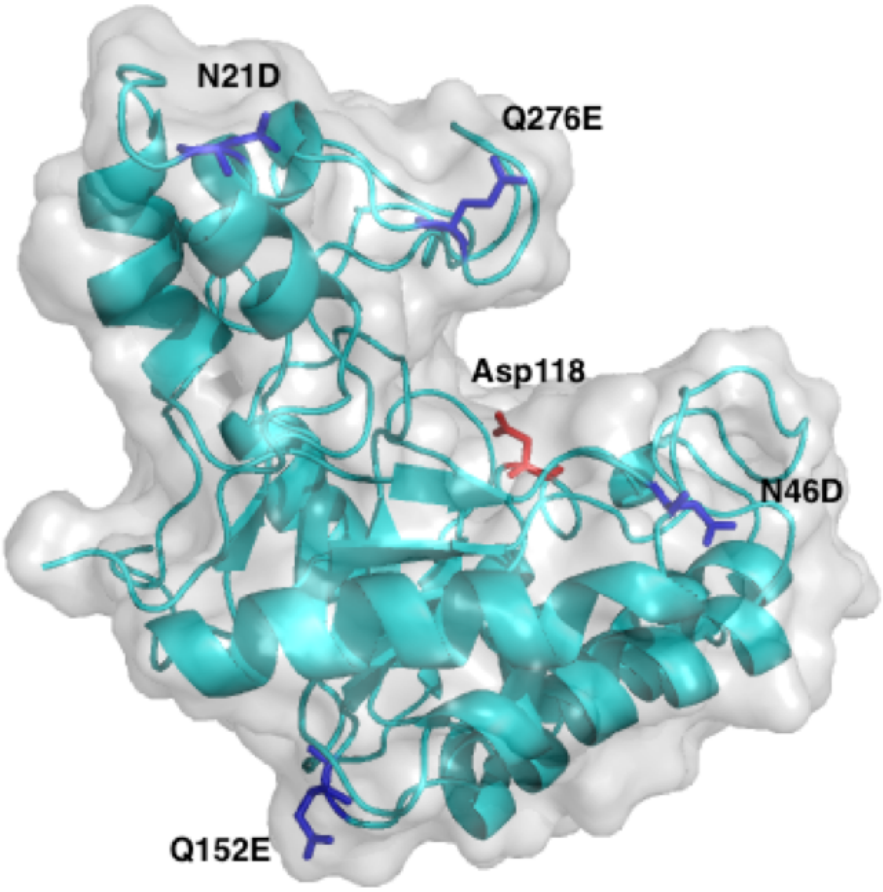


Figure 4.1.5. Model of mgCel6AΔCBM with Mutations 1-4. The model shows the side chains of the four mutated residues N21D (M1), N46D (M2), Q152E (M3), and Q276E (M4) as dark blue sticks. The mutations, which were intended to increase pH stability by introducing negative charge on the enzyme surface, are labeled. The side chain of the catalytic acid (Asp118) is shown as a red stick. The figure was made using PyMOL (pymol.org).

4.1.4.1.2. Mutants M5-M8

As described in Section 1.6.2, the stability of a protein is dictated by its amino acid interaction network, which determines its native three-dimensional structure. One such local interaction is the ionic bond, also referred to as a salt bridge. This involves the interaction between two

oppositely charged residues, typically with their terminal charged groups located within 4Å of one another (Barlow & Thornton 1983).

As mentioned in Section 1.6.4, studies of enzymes produced by thermophilic organisms have found that these often contain large amounts of salt bridges relative to enzymes produced by mesophilic organisms, presumably conferring increased stability at higher temperatures (Kumar et al. 2000; Yip et al. 1998). In addition, while some studies have found salt bridges on the surface of proteins to be stabilizing, many of these studies also highlight the importance of the spatial positioning and the local environments of the residues involved in the ionic bond (Kumar & Nussinov 1999; Makhatadze et al. 2003).

Mutants 5-8 were designed to introduce salt bridges between residues on the surface of mgCel6AΔCBM. In all cases, this was done by mutating surface aliphatic residues found to be within hydrogen-bonding distance to positively charged residues to aspartate or glutamate. This was presumed to result in the formation of salt bridges, which could increase enzyme stability, including stability at low pH values. ***Figure 4.1.6*** shows the mutations selected using this approach.

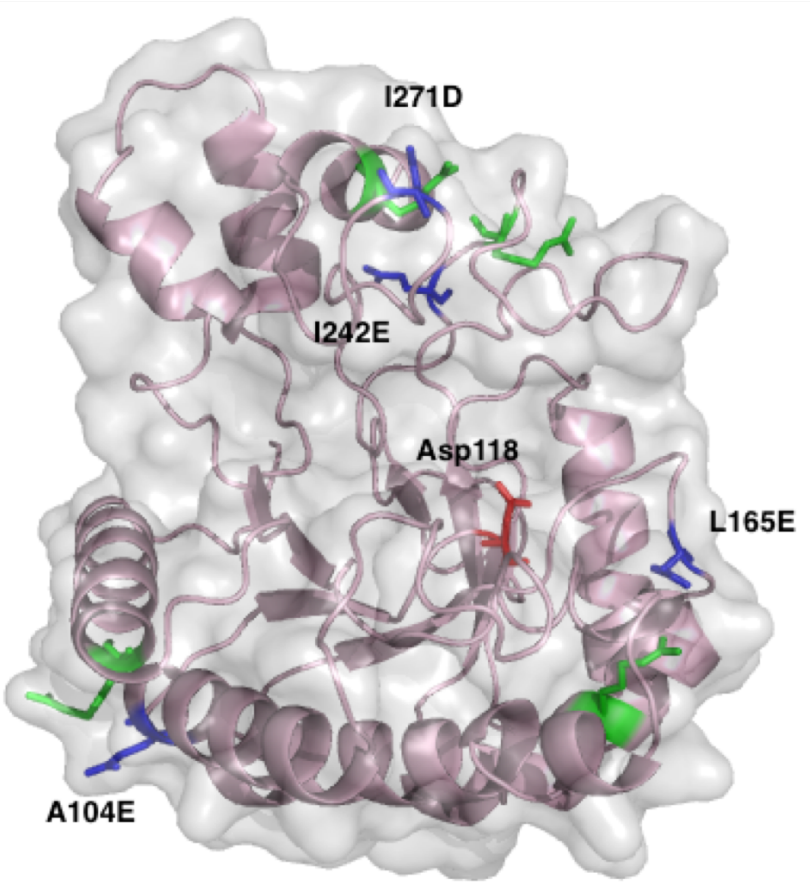


Figure 4.1.6. Model of mgCel6AΔCBM with Mutations 5-8. The model shows the side chains of the four mutated residues A104E (M5), L165E (M6), I242E (M7), and I271D (M8) as blue sticks. The mutations are labeled. These mutations were intended to increase pH stability by forming salt bridges with nearby basic residues. The side chains of the basic residues (all arginines) are shown as green sticks. The catalytic acid, Asp118, is labeled, and its side chain is shown as a red stick. Note that the enzyme's active site is facing towards the viewer. The figure was made using PyMOL (pymol.org).

4.1.4.1.3. Mutant M9

The final stability mutant was designed in the hope of combining the putatively desirable effects of mutants M1-M8, and/or to obtain a more general impression of the mutability of the surface and its effects on enzymatic properties. Theoretically, a combination of all eight point mutations from M1-M8 could result in an additive effect.

The mutations introduced in mutant M9 can be seen in ***Figure 4.1.5*** and ***Figure 4.1.6*** above.

4.1.4.2. Design of Active Site Mutants M10-M12

The mutants designed specifically to reduce the pH optimum of mgCel6AΔCBM were designed based on the multiple sequence alignment (MSA) (***Figure 4.1.4***), as well as data from a previously published pH optimum engineering study of *Thermobifida fusca* Cel6A.

A total of three pH optimum mutants were designed to create an acidic shift in the pH-activity profile of the enzyme. Mutant M10 (T44D) was selected based on the MSA (Section 4.1.3), while mutants M11 and M12 (D158A and D158N) were selected based on a study of the active site of *Tf*Cel6A, where similar mutations had resulted in a shifted pH optimum (Wolfgang & Wilson 1999).

4.1.4.2.1. Mutant M10

The analysis of the MSA (***Figure 4.1.4***) identified a single candidate for mutation: a threonine in position 44, which appeared to be a conserved aspartate in sequences of GH6 enzymes with lower pH optima than mgCel6AΔCBM.

A closer study of Thr44 in the enzyme model (***Figure 4.1.7***) showed that the side chain of this residue is located approximately 5Å from a substrate molecule bound in the active site (carried over from the template structure). When mutated to Asp, this distance decreases to approximately 4Å. Thr44 is located approximately 14Å from the catalytic acid Asp118.

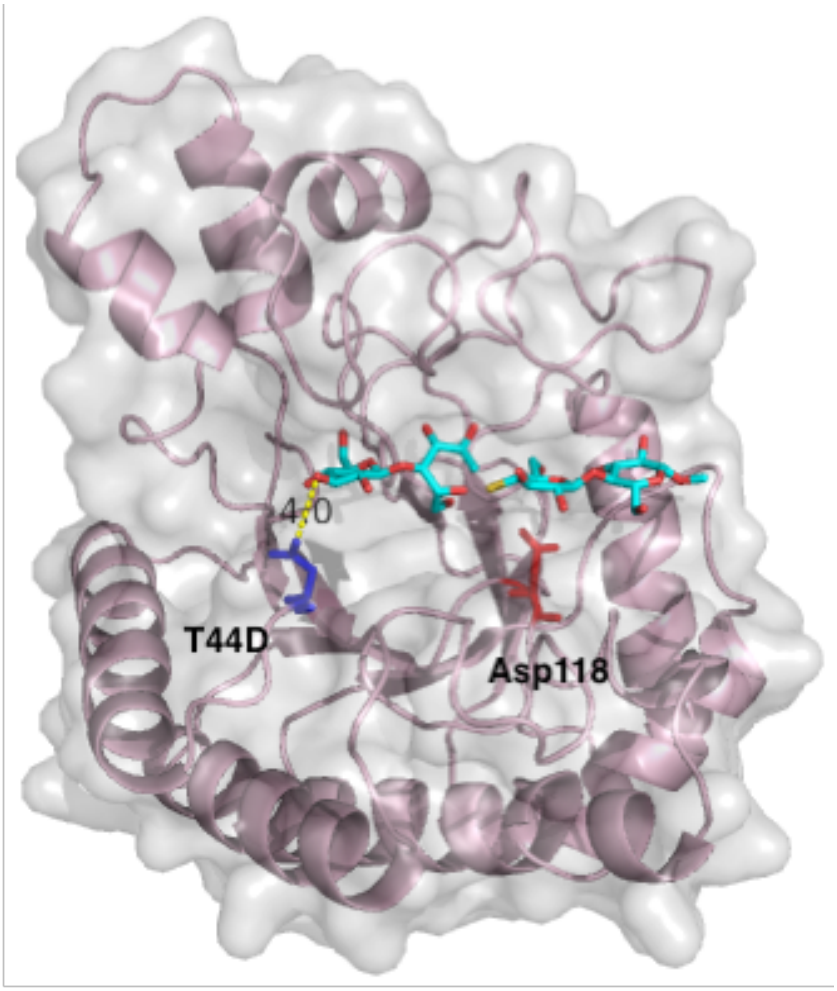


Figure 4.1.7. Structural Context of Mutant M10 (T44D). The active site of the enzyme is facing towards the viewer, and contains a bound methyl cellobiosyl-4-thio-β-cellobioside, as found in the template structure used for modelling mgCel6AΔCBM (in complex with this molecule; PDB ID 2BOD) (Berman et al. 2000). The T44D mutation is labeled, and the side chain of the mutated residue is shown as a dark blue stick. The side chain of the catalytic acid (Asp118) is shown as a red stick. The figure was made using PyMOL (pymol.org).

4.1.4.2.2. Mutants M11-M12

Mutants M11 and M12 were selected based on a study of the active site of *T. fusca* Cel6A. As mentioned, the catalytic domain of *Tf*Cel6A has a 70% sequence identity with mgCel6AΔCBM, and was used as the template for homology modeling of this enzyme.

As described in Section 1.5.3.1, the identity and/or existence of the catalytic base residue in inverting family 6 glycoside hydrolases has yet to be established. The precise mechanism of catalysis in this enzyme family is still being debated, however, several studies have

investigated various amino acids in the active site vicinity, and have thus suggested roles that these residues may play in the catalytic mechanism.

One such analysis is a 1999 study on *Tf*Cel6A by D. E. Wolfgang and D. B. Wilson of residues proposed to be involved in the catalytic mechanism of the enzyme (previously known as *Thermomonospora fusca* E2), as discussed in Section 1.5.3.1. As part of this study, numerous active site mutants were created in an attempt to determine precise roles for each of these residues in the hydrolysis mechanism.

One such residue, Asp156, had previously been proposed in *Cellulomonas fimi* Cel6A to increase the pK_a value of the catalytic acid, Asp117 (Damude et al. 1995). An increased pK_a value of an acidic residue leads to its protonation at higher pH values, and thus to a higher pH optimum if this residue is the catalytic acid. In the 1999 study of *Tf*Cel6A, the Asp156 residue was mutated to both alanine and asparagine to evaluate the effect of these mutations on the pH optimum of the enzyme. According to the study, alanine was chosen to mimic the complete removal of the side chain, whereas asparagine was chosen to simulate removal of only the carboxylic acid group, since this residue is otherwise identical to aspartate in size and orientation.

The results indicated that removing the carboxylic acid group from Asp156 did indeed alter the properties of the catalytic acid. The D156A and D156N mutants both acquired lowered pH optima compared to the wild-type enzyme, although the enzymatic activity was simultaneously considerably reduced. The authors of this study concluded that Asp156 (as well as a nearby aspartate, Asp79, that has previously been speculated to be the catalytic base; see Section 1.5.3.1) raises the pK_a of the catalytic acid.

In mgCel6AΔCBM, residue Asp156 corresponds to Asp158. Based on the results of the study of *Tf*Cel6A, Asp158 was selected for mutation to alanine and asparagine in the hope that these single mutants would reduce the pH optimum of mgCel6AΔCBM. The position of the D158A and D158N mutations is illustrated in ***Figure 4.1.8***.

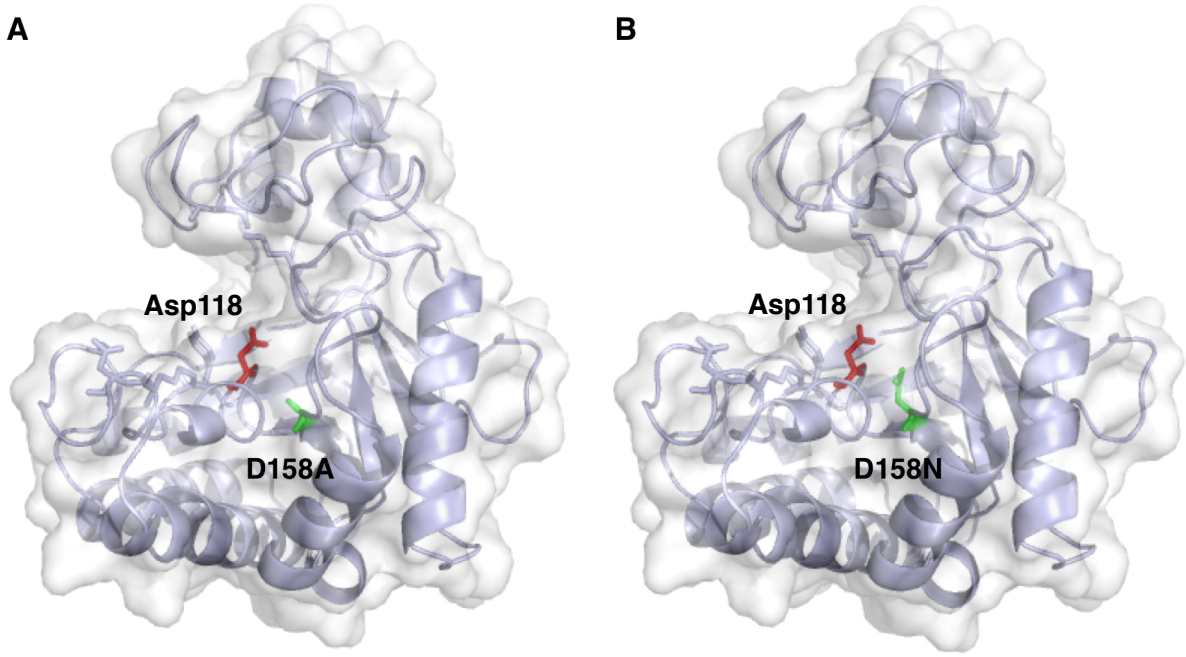


Figure 4.1.8. Models of mgCel6AΔCBM with Mutations 11 and 12. The figure illustrates the D158A (**Panel A**) and D158N (**Panel B**) mutations. These mutations were designed to reduce the pH optimum of mgCel6AΔCBM by altering the residue that increases the pK_a of the catalytic acid. The side chains of the mutated residues are shown as green sticks, while the side chain of the catalytic acid, Asp118, is shown as a red stick. Note that the enzyme is here shown from the opposite side to the view demonstrated in ***Figure 4.1.3***. The figures were made using PyMOL (pymol.org).

4.1.5. Overview of Rationally Designed Mutants

Table 4.1.2 gives an overview of the physicohemical properties of mgCel6AΔCBM and all mgCel6AΔCBM mutants designed during this study. These properties were calculated using the ProtParam tool from ExPASy (Gasteiger et al. 2005).

Table 4.1.2. Physiochemical Properties of mgCel6AΔCBM and mgCel6AΔCBM Mutants. The table shows the name, the relevant mutation, the molecular weight of the protein (MW; in Dalton), the theoretical isoelectric point (pI), and the extinction coefficient (ε) of each mgCel6AΔCBM mutant designed in this study. All properties were calculated using the ProtParam tool from ExPASy (Gasteiger et al. 2005).

Protein	Mutation	MW (Da)	pI	ε
mgCel6AΔCBM	-	30891.55	5.96	1.462
M1	N21D	30892.53	5.81	1.462
M2	N46D	30892.53	5.81	1.462
M3	Q152E	30892.53	5.82	1.462
M4	Q276E	30892.53	5.82	1.462
M5	A104E	30907.51	5.82	1.461
M6	L165E	30907.51	5.82	1.461
M7	I242E	30907.51	5.82	1.461
M8	I271D	30893.48	5.81	1.462
M9	N21D/N46D/A104E/ Q152E/L165E/I242E/ I271D/Q276E	30987.37	5.04	1.458
M10	T44D	30905.53	5.81	1.462
M11	D158A	30847.54	6.12	1.464
M12	D158N	30890.56	6.12	1.462

4.2. Mutant Production

Following the design of mgCel6AΔCBM variants, the mutant genes and the wild-type *mgcel6aΔcbm* gene were created using PCR-based methods. The pNIC-CH plasmid (Appendix A) containing the wild-type *mgcel6aΔcbm* gene was used as the template DNA for mutant gene production.

Site-directed mutagenesis (SDM; Section 3.3.2) was used in an attempt to create the genes coding for mutants 1-8 (*M1*-*M8*). This method only resulted in the construction of two out of eight mutant genes. When attempting to create the mutant *M10*-*M12* genes, splicing by overlap extension (SOE; Section 3.3.3) was used, followed by ligation-independent cloning (LIC; Section 3.7). The use of SOE resulted in the successful production of all three mutant genes created using this method. ***Table 4.2.1*** shows an overview of the outcome of the attempts to produce the various mutant genes. Further details are provided below.

Table 4.2.1. Production Status of *mgcel6aΔcbm* Mutant Genes. The table shows the mutant gene name, the relevant mutation(s), the method used to generate the gene, and the outcome of the attempt to produce the gene.

Gene	Mutation	Production Method	Successfully Produced at Gene Level (+/-)
M1	N21D	SDM Kit	-
M2	N46D	SDM Kit	-
M3	Q152E	SDM Kit	+
M4	Q276E	SDM Kit	-
M5	A104E	SDM Kit	-
M6	L165E	SDM Kit	+
M7	I242E	SDM Kit	-
M8	I271D	SDM Kit	-
M9	N21D/ N46D/ Q152E/ Q276E/ A104E/ L165E/ I242E/ I271D	Gene ordered from GenScript®	+
M10	T44D	SOE	+
M11	D158A	SOE	+
M12	D158N	SOE	+

4.2.1. Gene Production and Cloning

4.2.1.1. Generation of *mgcel6a*, *mgcel6aΔcbm*, and *Mutant 9* (*M9*) Genes

This work was performed by Marianne S. Jensen and Dr. Lasse Fredriksen.

The genes for the two wild-type enzymes and for mutant M9 (*M9*) were ordered ready-made from GenScript®. The genes were amplified through PCRs (Section 3.3.1) and the sizes of the amplified genes were verified using agarose gel electrophoresis.

The genes were subjected to LIC digestion and cloned into the pNIC-CH vector (Section 3.7). The plasmids were then transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Section 3.8), which were plated on LB agar plates supplemented with kanamycin and sucrose (Section 2.10.1). Resulting colonies were examined using colony PCR (Section 3.3.4) to ensure the target gene had been correctly inserted into the vector.

The colonies that grew on the LB agar plates supplemented with kanamycin and sucrose were also used to make overnight cultures (Section 3.1), from which the mutated plasmids were isolated and sequenced (Sections 3.9 and 3.10). Sequencing results indicated correct gene sequences for *mgcel6a*, *mgcel6aΔcbm*, and for *M9*.

4.2.1.2. Generation of *mutant 1 – mutant 8* (*M1-M8*) Genes by Site-Directed Mutagenesis

The genes coding for mutants 1-8 (M1-M8) were produced using site-directed mutagenesis (SDM). The QuikChange II Site-Directed Mutagenesis Kit was used to create these mutant genes. Mutant pNIC-CH plasmids resulting from the mutagenesis protocol (Section 3.3.2) were sequenced in order to verify the presence of the desired mutation. Unfortunately, the sequencing results indicated that only two of the eight mutant genes, *M3* and *M6*, had been correctly generated.

4.2.1.3. Generation of *mutant 10 – mutant 12* (*M10-M12*) Genes using Splicing by Overlap Extension (SOE)

As a result of significant difficulties producing *mgcel6aΔcbm* mutants using the QuikChange II Site-Directed Mutagenesis Kit, splicing by overlap extension (SOE) was used to produce the mutant genes *M10-M12*.

Mutant M10 was designed to contain a single point mutation at amino acid position 44 of the mgCel6A protein, while mutants 11 and 12 were both designed to contain a single point mutation at amino acid position 158.

Since the mutations were located at different positions in the sequence, the left (upstream) and right (downstream) fragments generated during the first PCR of SOE (Section 3.3.3.1) were expected to be of different sizes. The left fragment of *M10* was expected to be 172 bp, while its right fragment was expected to be 760 bp. Since both *M11* and *M12* contained mutations at the same position, the left and right fragments generated for these mutant genes were expected to be of the same size: 514 bp for the left fragment, and 418 bp for the right fragment.

Following generation of the two fragments for each mutant gene, the PCR products were run on a 1% agarose gel (Section 3.4) to estimate the size of each fragment. The resulting gel electrophoresis images are displayed in ***Figure 4.2.1***. The expected sizes of the fragments were verified by comparison to a GeneRuler 1 kb DNA ladder.

Note that the size of the fragments in bp added together does not equal 864 (the length of the *mgcel6aΔcbm* gene). This is due to the presence of a 21 or 22 bp long LIC overhang in the left and right fragments, respectively. In addition, each left fragment contains a 25-bp region of complementarity, allowing it to anneal to the right fragment. This region is eliminated when the two fragments are spliced during the second SOE PCR (Section 3.3.3.2).

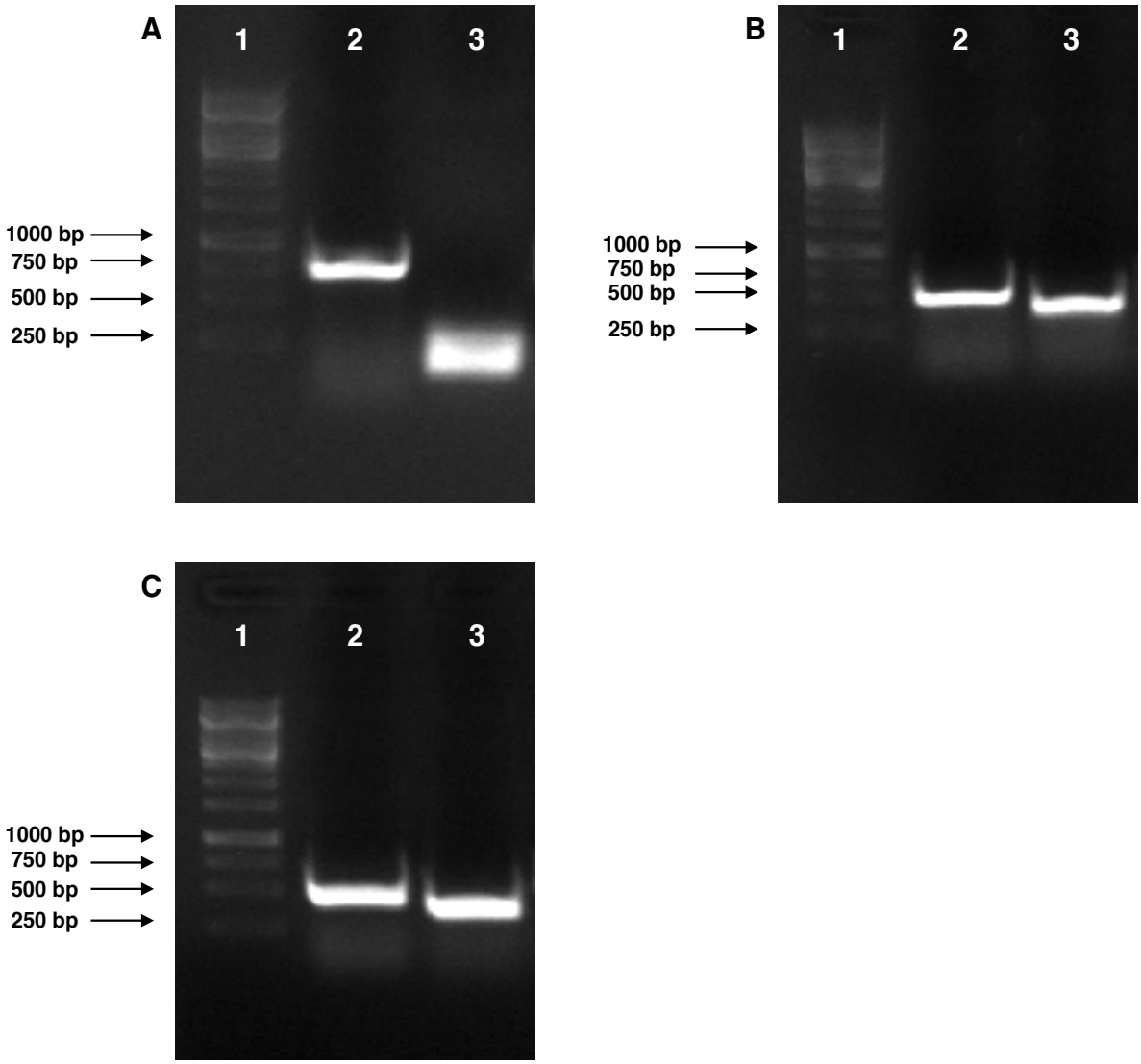


Figure 4.2.1. 1% agarose gel electrophoresis of left and right fragments for *M10-M12* generated during the first PCR of SOE. The GeneRuler 1 kb ladder is shown in lane 1 in all panels. **Panel A:** Gel electrophoresis of PCR products generated for *M10*. Lane 2 shows the right fragment (760 bp), and lane 3 shows the left fragment (172 bp). **Panel B:** Gel electrophoresis of PCR products generated for *M11*. Lane 2 shows the left fragment (514 bp), and lane 3 shows the right fragment (418 bp). **Panel C:** Gel electrophoresis of PCR products generated for *M12*. Lane 2 shows the left fragment (514 bp), and lane 3 shows the right fragment (418 bp).

The DNA fragments (***Figure 4.2.1***) were excised from the gel and purified (Section 3.5). The second SOE PCR was then performed to generate the full-length mutant genes, and the PCR products were again visualized using agarose gel electrophoresis (***Figure 4.2.2***). The correct sizes of the mutant genes were verified by comparison to a GeneRuler 1 kb DNA ladder.

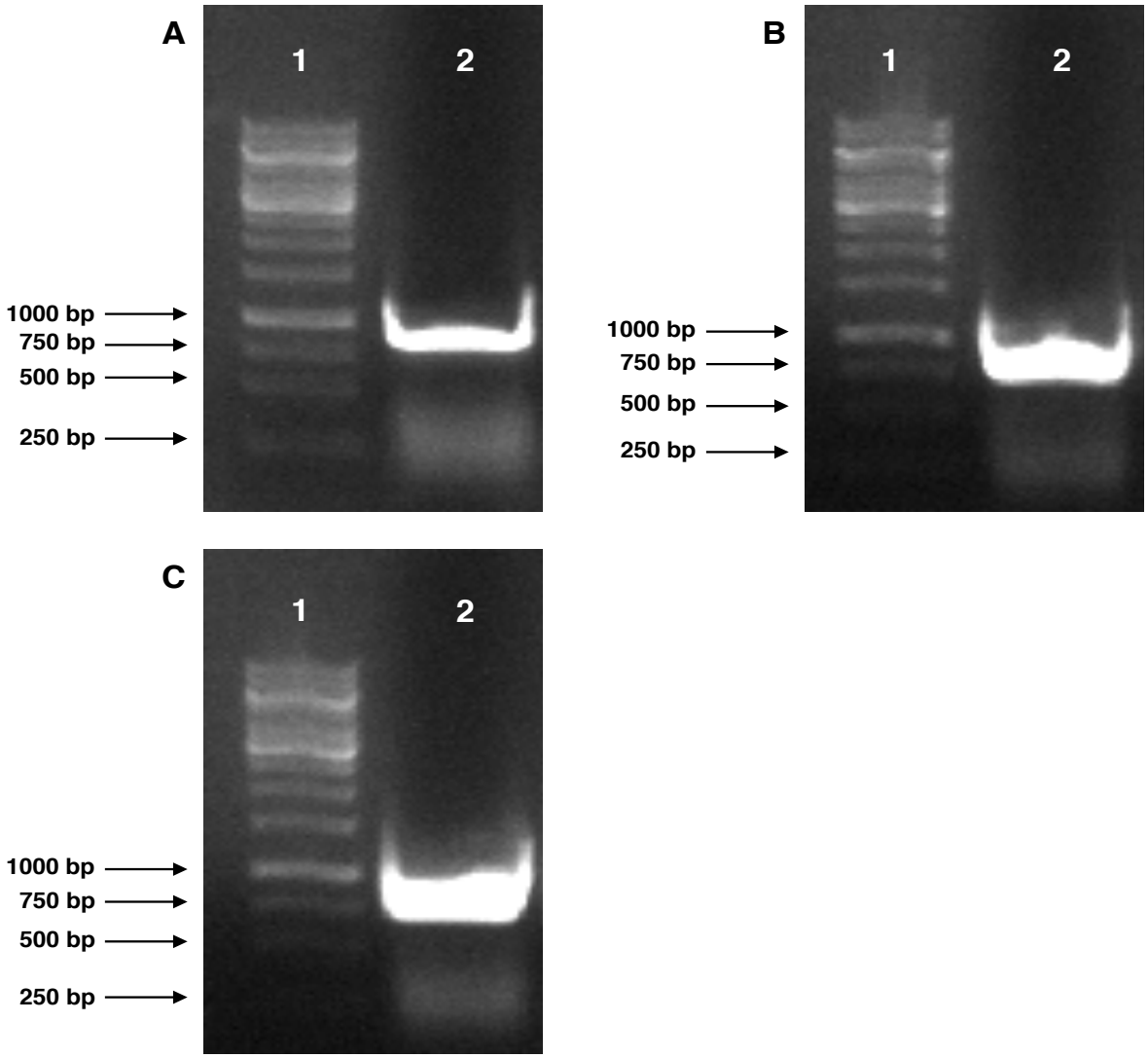


Figure 4.2.2. 1% agarose gel electrophoresis of mutant M10-M12 genes generated during the second PCR of SOE. The figure shows gel electrophoresis of the PCR product generated for *M10* (**Panel A**), *M11* (**Panel B**), and *M12* (**Panel C**). In all panels, lane 1 shows the GeneRuler 1 kb ladder. Lane 2 shows the generated PCR fragment (907 bp).

The DNA fragments containing the desired mutations were subsequently inserted into the pre-digested pNIC-CH vector (Section 3.7) and the recombinant plasmids were transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Section 3.8). Successful transformation of the plasmids containing the correct inserts was verified by colony PCR (results not shown). Sequencing of the isolated plasmids (Section 3.10) verified that the all three mutant genes produced using SOE had been successfully generated and that no changes, other than the desired mutations, had occurred.

This method led to the production of mutant plasmids containing the *M10*, *M11*, and *M12* genes, respectively.

4.2.2. Expression and Purification of Wild-Type and Mutant Enzymes

Following confirmation of the correct gene sequences, the pNIC-CH plasmids carrying the *mgcel6aΔcbm* and mutant genes were transformed into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli* cells (Section 3.11). Expression of target genes was initiated by the addition of the inducer molecule isopropyl β-D-1-thiogalactopyranoside (IPTG) (Section 3.12), cell-free extracts were prepared by sonication (Section 3.13), and the resulting proteins were purified using immobilized metal ion affinity chromatography (IMAC) (Section 3.14).

Nearly all chromatograms for the various target proteins showed sharp, distinct elution peaks once the Ni-NTA Elution Buffer gradient reached approximately 20-25%, corresponding to an imidazole concentration of 100-130 mM (***Figure 4.2.3***). The only exception was the chromatogram for mutant M9, which contained eight point mutations. While this protein also eluted at a similar imidazole concentration, the chromatogram resulting from the purification of this mutant showed a much smaller elution peak, indicating a lower yield of purified protein. All other mutants showed elution peaks more comparable to those observed in the purification of mgCel6AΔCBM and M10 (results not shown).

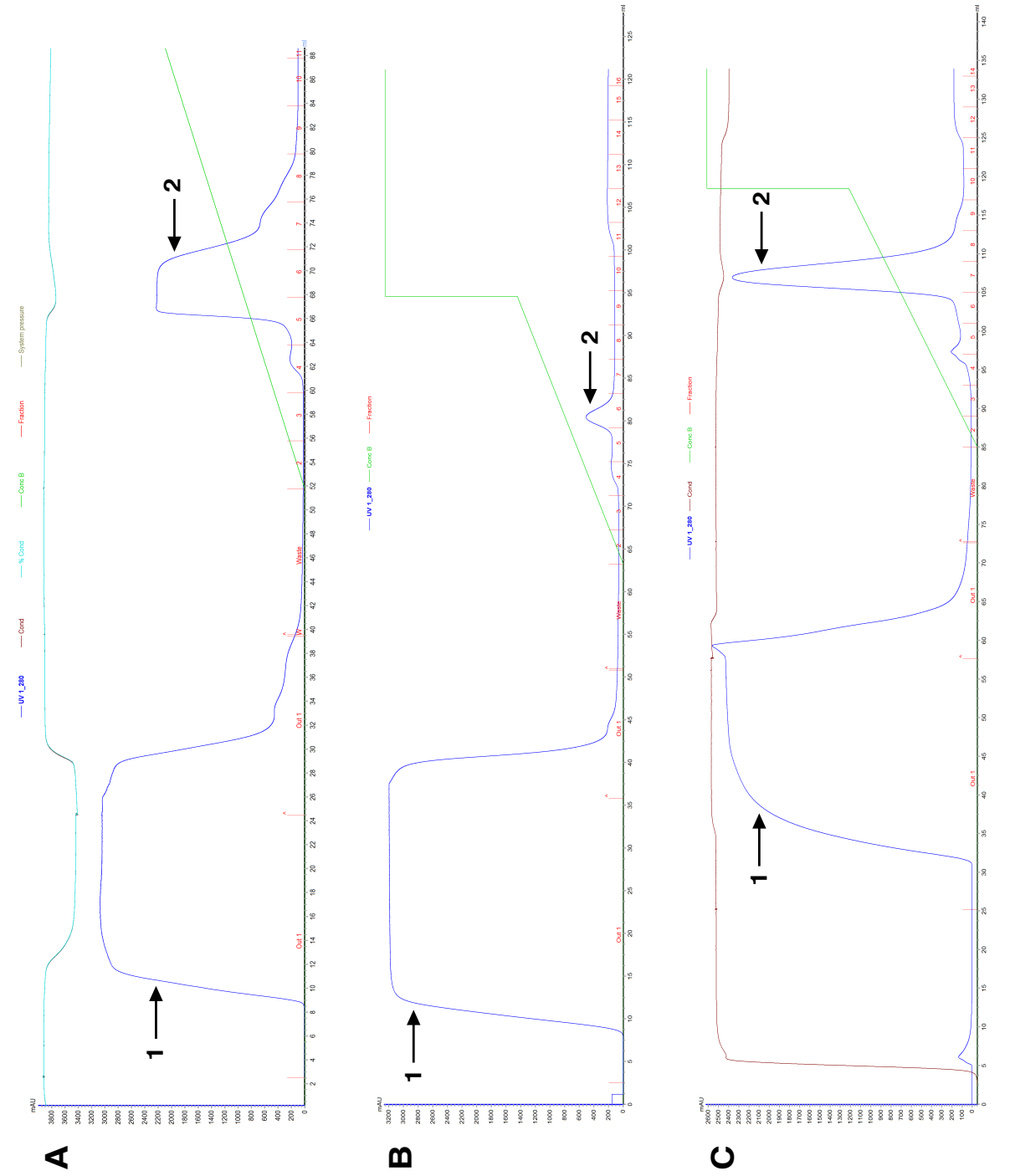


Figure 4.2.3. IMAC Purification of mgCel6AΔCBM, M9, and M10. The chromatograms represent the purification of mgCel6AΔCBM (**Panel A**), the M9 mutant (**Panel B**), and the M10 mutant (**Panel C**). Protein elution from the column was detected by monitoring A_{280} (represented by the dark blue curve) of the column flow-through. The imidazole gradient applied to elute proteins bound to the HisTrap™ column is represented by the light green curve. The light blue and brown curves represent the conductivity of the solution within the column. Fraction numbers are shown in red along the *x*-axes. All chromatograms show two major peaks: the flow-through (non-bound protein) peak is indicated with the number 1, while the peak representing the target protein is indicated with the number 2. Chromatograms for M3 and M6 were nearly indistinguishable from that of mgCel6AΔCBM, whereas chromatograms for M11 and M12 were nearly indistinguishable from that of M10. These chromatograms are therefore not shown. All chromatograms were generated using the Unicorn™ 6.4.1 Software.

Figure 4.2.3 shows a clear difference in the shape of the elution peaks obtained for mgCel6AΔCBM and M10. The elution peaks produced during IMAC purification of M10, M11, and M12 had significantly smaller surface areas than the corresponding elution peaks of mgCel6AΔCBM, M3, and M6, indicating lower yields of purified proteins for these three mutants.

In all chromatograms, minor elution peaks can be seen at the start of the imidazole gradient, just before the elution of the target protein. These most likely correspond to contaminant *E. coli* host proteins (see Section 5.3).

Following IMAC purification, cell pellets resulting from the final centrifugation step after cell lysis, the IMAC flow-throughs, and the purified IMAC fractions corresponding to the target protein elution peaks were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Section 3.15). The resulting gel images are shown in ***Figure 4.2.4***, ***Figure 4.2.5***, and ***Figure 4.2.6***. The mgCel6AΔCBM protein and mutant variants thereof were expected to be approximately 31 kDa (predicted by ProtParam, ***Table 4.1.2***) (Gasteiger et al. 2005), but appeared at positions located slightly below the 30 kDa band in the SDS-PAGE analyses.

More target protein appeared to be located in the pellet remaining after cell lysis of M3-producing cells than in the corresponding samples for mgCel6AΔCBM and M6 (***Figure 4.2.4***). SDS-PAGE analyses also showed that the fractions containing purified M6 contained more contaminations than fractions containing purified mgCel6AΔCBM and M3. The most "contaminated" fraction containing the M6 mutant was therefore not used in the subsequent ultrafiltration of this protein.

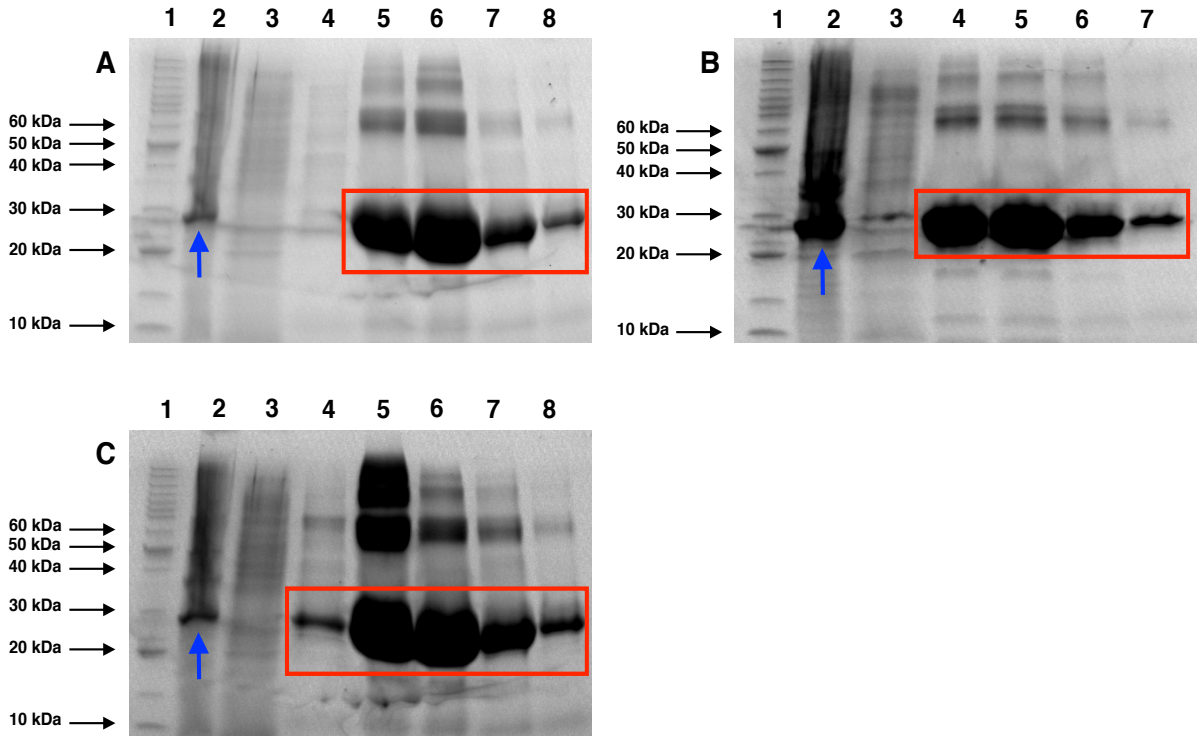


Figure 4.2.4. SDS-PAGE analysis of purified mgCel6AΔCBM, M3, and M6. In all panels, lane 1 shows the BenchMark™ protein ladder, lane 2 shows the pellet resulting from the final centrifugation following sonication of the harvested cells, and lane 3 shows the IMAC flow-through. The gels represent samples from production and purification of mgCel6AΔCBM (**Panel A**), the M3 mutant (**Panel B**), and the M6 mutant (**Panel C**). The target proteins in fractions eluted from the IMAC column are indicated by red squares. Target proteins in the pellets are indicated with blue arrows. Subsequent ultrafiltration was performed on combined fractions shown in lanes 5-8 for mgCel6AΔCBM, lanes 4-7 for M3, and lanes 4, 6, 7, and 8 for M6.

Figure 4.2.5 shows SDS-PAGE analysis of fractions obtained during production and purification of M9. A substantial amount of target protein was located in the pellet resulting from the final centrifugation step following sonication of the cells harboring M9, indicating the presence of insoluble target protein (see Section 5.3).

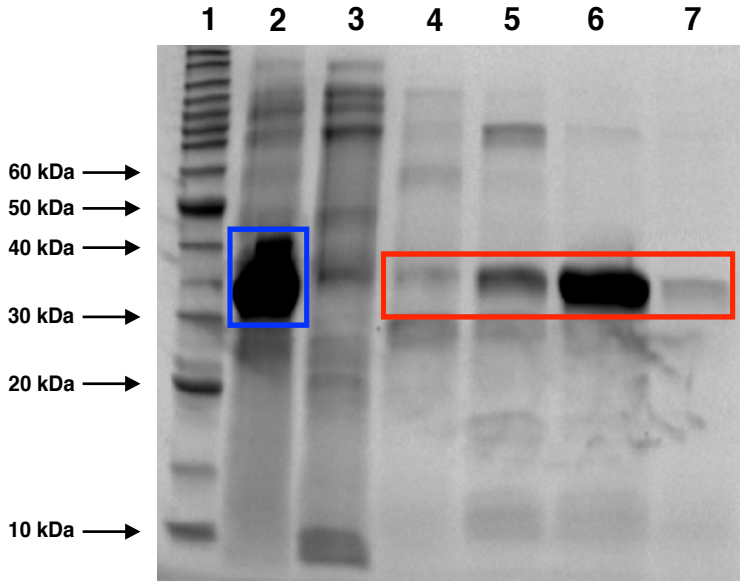


Figure 4.2.5. SDS-PAGE analysis of purified M9. Lane 1 shows the BenchMark™ protein ladder, lane 2 shows the pellet resulting from the final centrifugation following sonication of the harvested cells, and lane 3 shows the IMAC flow-through. Lanes 4-7 show the fractions corresponding to the elution peak of the protein seen in **Figure 4.2.3**. The target protein in elution fractions is indicated by the red square. The target protein in the pellet is indicated by the blue square. All fractions shown (lanes 4-7) were combined and subjected to subsequent ultrafiltration.

SDS-PAGE analysis of fractions generated during the purification of M10, M11, and M12 is shown in ***Figure 4.2.6***. Nearly no target protein was located in the IMAC flow-throughs, but a significant amount of protein was found in fraction 7 for all three mutants. Due to the smaller surface area of the elution peaks of M10, M11, and M12 (***Figure 4.2.3***), only fractions 6, 7, and 8 were run on an SDS-gel. For all three mutants, fractions 6 and 8 showed very weak bands corresponding to approximately 30 kDa, indicating very slight concentrations of the target protein (results not shown). No significant contaminations of any other proteins were detected in these fractions. The pellet samples of these mutants contained minor amounts of target protein (results not shown), similar to the band observed in the SDS-PAGE analysis of the wild-type pellet sample (***Figure 4.2.4***). For each of these mutants, fractions 6, 7, and 8 were combined and subjected to ultrafiltration for M10, M11, and M12.

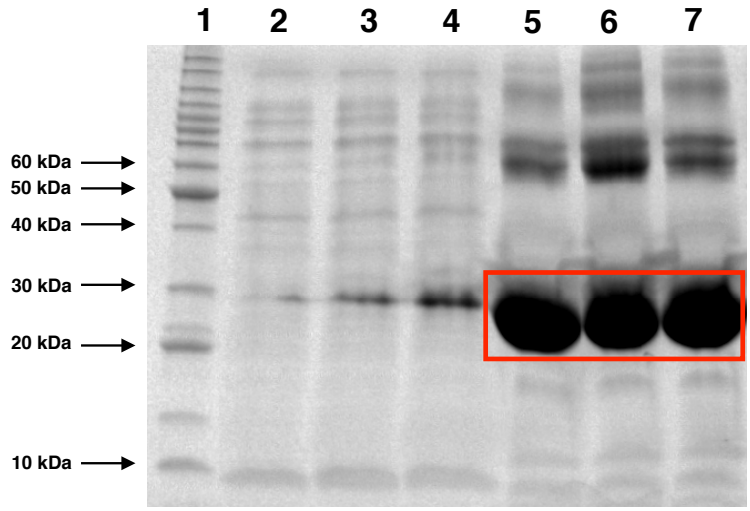


Figure 4.2.6. SDS-PAGE analysis to assess the purification of M10, M11, and M12. Lane 1 shows the BenchMark™ protein ladder. Lanes 2, 3, and 4 show the IMAC flow-through collected during purification of M10, M11, and M12, respectively. Lanes 5, 6, and 7 show the target protein in fraction 7 (as seen in the chromatogram for M10 shown in **Figure 4.2.3**) for M10, M11, and M12, respectively. For each target protein, the fraction shown was combined with fractions 6 and 8 (not shown) and subjected to ultrafiltration.

Following purification and subsequent analysis of purified proteins via SDS-PAGE, all proteins were subjected to ultrafiltration with simultaneous buffer exchange (Section 3.16). The A_{280} of all resulting protein solutions were subsequently measured (Section 3.17), and the final protein concentrations were determined using the extinction coefficients predicted by ProtParam (***Table 4.1.2***) (Gasteiger et al. 2005). These concentrations, along with the total protein yield per liter culture, are shown in ***Table 4.2.2***.

Table 4.2.2. Measured protein concentrations and total protein yields of mgCel6AΔCBM and all mutants. The protein concentrations were calculated from A_{280} measurements using extinction coefficients predicted by ProtParam (Gasteiger et al. 2005).

Protein	Final Concentration (μM)	Protein Yield (mg/L culture)
mgCel6AΔCBM	1943	360
M3	2599	466
M6	1603	288
M9	198	36
M10	693	128
M11	717	132
M12	923	172

4.3. Enzyme Characterization

Several properties of both the full-length mgCel6A enzyme and its catalytic domain mgCel6AΔCBM had been investigated prior to beginning this study, including thermostability and optimal conditions for catalysis on various substrates (Section 1.7.2.2) (Jensen et al. 2017). Initial steps in the present study were therefore limited to pH optimum and stability assays to gain a more comprehensive understanding of how pH affects mgCel6A (purified full-length mgCel6A was provided by Marianne S. Jensen). Potential effects of the CBM2 domain on the physiochemical properties of the enzyme were also investigated.

Successfully expressed and purified mutant enzymes M3, M6, M9, M10, M11, and M12 (see ***Table 4.1.1*** for an overview of mutants) were subsequently characterized to determine the effects of the rationally designed mutations on the optimum pH and on pH stability.

4.3.1. Characterization of the wild-type enzymes, mgCel6A and mgCel6AΔCBM

4.3.1.1. pH Optimum

To generate pH-activity profiles for mgCel6A and mgCel6AΔCBM on CMC and BALI™ cellulose, the enzyme variants were tested at their temperature optima in 15-minute pH optimum assays on both substrates, as well as in a 24-hour assay on the BALI™ cellulose substrate. In addition, a 15-minute pH-activity assay was performed with mgCel6AΔCBM at 60°C using 0.2% (w/v) PASC. All pH optimum assays were conducted as described in Section 3.18.3.

For reactions with CMC and BALI™ cellulose, only the results from the assays using mgCel6A are shown here, since both enzyme variants showed identical optimum pHs in all reactions tested, indicating that the CBM domain did not impact this property in any way. The pH-activity profiles generated are depicted in ***Figure 4.3.1***.

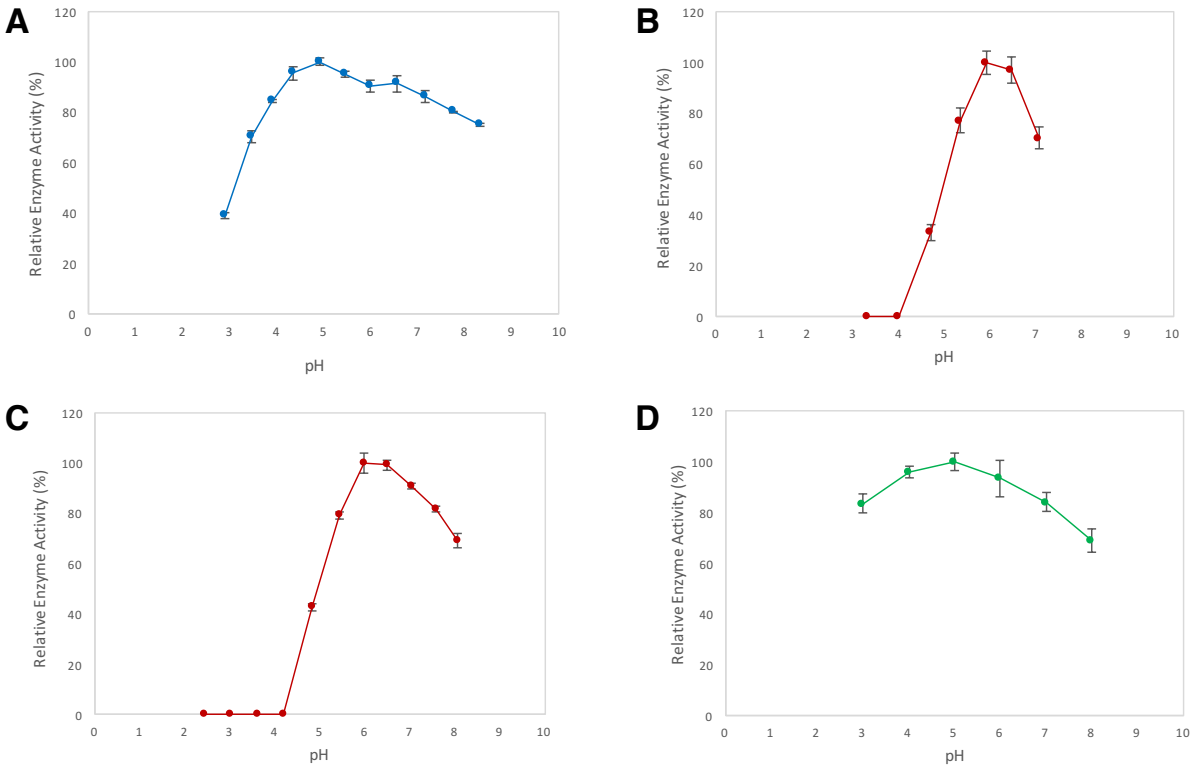


Figure 4.3.1. pH-Activity Profiles of mgCel6A. The figure shows pH-activity profiles for a 15-minute reaction at 85°C with 1% (w/v) CMC (**Panel A**), a 15-minute reaction at 85°C with 1% (w/v) BALI™ cellulose (**Panel B**), and a 24-hour reaction at 60°C with 1% (w/v) BALI™ cellulose (**Panel C**). **Panel D** shows the pH-activity profile of mgCel6AΔCBM for a 15-minute reaction at 60°C with 0.2% (w/v) PASC. The enzyme concentration was 1 µM in all reactions. Reactions were carried out in 75 mM phosphate-citrate buffer using pHs indicated in the *x*-axes of each panel. The buffers had been adjusted to have the correct pH at the relevant temperature before use. Product quantification was performed using the DNS method for the reactions depicted in panels A, C, and D, and using HPAEC-PAD for the reactions shown in panel B. All reactions were performed in triplicates and the standard deviations are shown as error bars. The pH-activity profiles of the mgCel6AΔCBM variant were found to be nearly identical to those of the full-length enzyme on CMC and on BALI™ cellulose, and are therefore not shown. The pH-activity profile of the full-length variant on PASC was unfortunately not investigated due to time constraints.

Figure 4.3.1 shows that the pH optimum of mgCel6A is approximately 5.0 for CMC and PASC, and approximately 6.0 for BALI™ cellulose. In addition, the results show that the large variation in the reaction times used in the two experiments with BALI™ cellulose did not affect the enzyme's optimum pH, suggesting that stability issues do not come into play during the longer reaction. The pH-activity profiles also demonstrate significant differences in terms of the range of pHs tolerated by the enzyme. The pH optimum curves for BALI™ cellulose are much narrower than the corresponding curves for hydrolysis of CMC and PASC, again highlighting the significant influence of the substrate properties on the enzyme's optimal conditions.

4.3.1.2. pH Stability

To further characterize the wild-type enzyme variants, mgCel6A and mgCel6AΔCBM were subjected to pH stability experiments. This was accomplished by pre-incubating the enzyme for up to 24 hours at 65°C and pHs 4.0, 5.0, and 6.0, followed by a 15-minute activity assay on 1% (w/v) CMC as described in Section 3.18.4. Product quantification was performed using the DNS method. Data for mgCel6A are shown in ***Figure 4.3.2***. The figure illustrates that after 8 hours, the residual activity at pH 5.0 was reduced to 75%, while the residual activity of the samples incubated at pH 6.0 was only reduced to 94%. After 24 hours, the residual activity was found to be 91% for samples incubated at pH 6.0, and 60% for samples incubated at pH 5.0. Enzyme samples pre-incubated at pH 4.0 were completely inactive after only 4 hours of pre-incubation at 65°C.

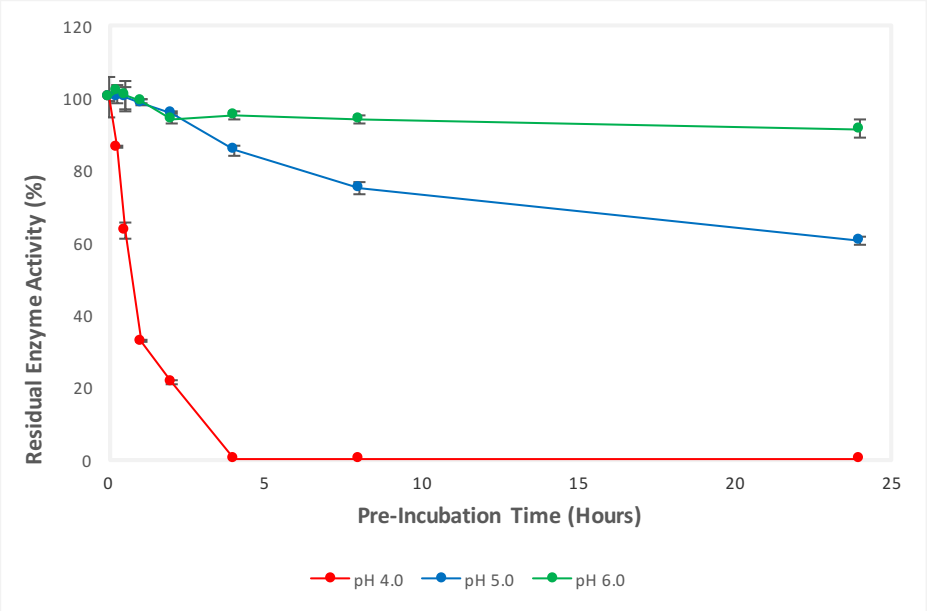


Figure 4.3.2. pH Stability of mgCel6A at 65°C. The figure shows the residual enzymatic activity of mgCel6A following pre-incubation at 65°C. Pre-incubations were performed in 75 mM phosphate-citrate buffer at pHs 4.0, 5.0, and 6.0. The buffers had been adjusted to have the correct pH at the relevant temperature before use. The pre-incubated enzyme samples were then tested for residual activity in a 15-minute assay on 1% (w/v) CMC. The pHs of the reactions were not adjusted following pre-incubation. Product quantification was performed using the DNS method. All pre-incubations and subsequent assays were performed in triplicates and the standard deviations are shown as error bars.

A similar pH stability assay was conducted to compare the stability of mgCel6A and mgCel6AΔCBM at pH 5.0 (***Figure 4.3.3***). In this pre-incubation, the buffer had not been

adjusted to have the correct pH at the relevant pre-incubation temperature. The actual pH of the pre-incubation was therefore calculated to be approximately 4.7. The data displayed in ***Figure 4.3.3*** indicates that there is minimal difference between the pH stability of the two enzyme variants. Based on the decline in activity observed, it seems that the half-life of both mgCel6A and mgCel6AΔCBM at pH 4.7 and 65°C is somewhere between 5 and 6 hours.

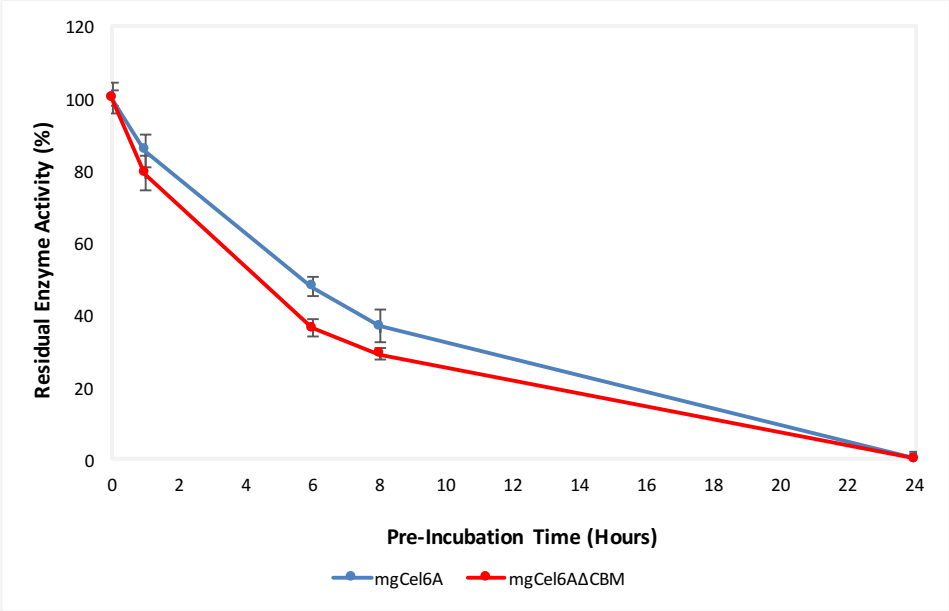


Figure 4.3.3. pH Stability of mgCel6A and mgCel6AΔCBM at 65°C and pH 4.7. The figure shows the residual enzyme activity of mgCel6A and mgCel6AΔCBM following pre-incubations at 65°C. The pre-incubations were performed in approximately 75 mM phosphate-citrate buffer at pH 4.7. The pre-incubated enzyme samples were then tested for residual activity in a 15-minute assay on 1% (w/v) CMC. The pHs of the reactions were not adjusted following pre-incubation. Product yield quantification was performed using the DNS method. All pre-incubations and subsequent assays were performed in triplicates and the standard deviations are indicated with error bars.

The results of these assays, in combination with the data resulting from previous characterizations of mgCel6A and its catalytic domain (Jensen et al. 2017), enabled the selection of the mgCel6AΔCBM variant as the template enzyme into which mutations were introduced.

4.3.2. Characterization of Mutant mgCel6AΔCBM Variants

To evaluate the effects of the mutations, the mutants were subjected to a variety of assays to compare their stability and pH optima to those of the wild-type mgCel6AΔCBM enzyme.

4.3.2.1. Characterization of M3 and M6

The two mutants containing single amino acid mutations on the surface of the enzyme were subjected to 24-hour end point assays at 60°C with 1% (w/v) BALI™ cellulose to determine their activity at pHs 4.0, 5.0, and 6.0 as compared to the activity of mgCel6AΔCBM (***Figure 4.3.4***). The assays were performed as described in Section 3.18.3. Since this reaction proceeds over 24 hours, this assay was also deemed to give a general indication of potential differences in activity at low pH between the mutant enzymes and the wild-type for a longer, more industrially-relevant reaction time.

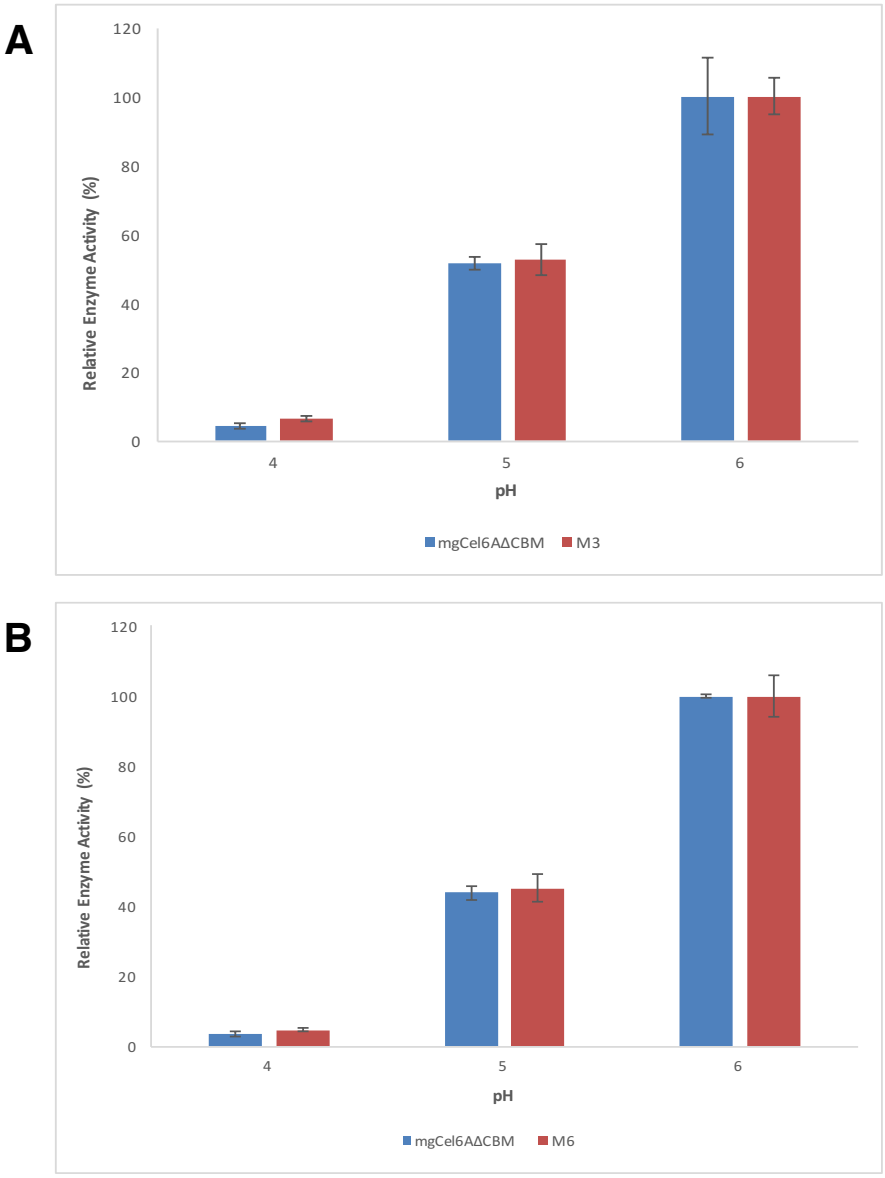


Figure 4.3.4. Activity Profiles of Mutants M3 and M6. Product formation by 1 µM M3 (**Panel A**) or M6 (**Panel B**) was compared to product formation by 1 µM mgCel6AΔCBM in 24-hour reactions at 60°C with 1% (w/v) BALI™ cellulose. Each reaction was performed in 75 mM phosphate-citrate buffer. Product quantification was performed using the DNS method. The buffers used were adjusted to have the correct pH at the relevant temperature. All reactions were performed in triplicates and the standard deviations are shown as error bars. The relative activity has been calculated in terms of A_{540} values for all samples.

Figure 4.3.4 shows that the activity profiles for BALI™ cellulose of both mutants are identical to the pH-profile of the wild-type enzyme. The results indicated that M3 and M6 showed similar amounts of product formation to that of the wild-type enzyme, both at the pH optimum of the enzyme for this reaction type (pH 6.0), and at more acidic pH.

In light of the identical activity profiles of M3, M6, and mgCel6AΔCBM, the characterization of these two mutants was discontinued in favor of producing and characterizing additional mgCel6AΔCBM variants.

4.3.2.2. Characterization of M9

The multiple-mutant M9 was created with the aim of potentially obtaining additive desirable stability effects from all eight single mutations (M1-M8; ***Table 4.1.1***), while simultaneously giving an impression of the general mutability of the enzyme surface. Following production and purification, the M9 mutant and mgCel6AΔCBM were subjected to a stability assay, performed as described in Section 3.18.4. The enzymes were pre-incubated at 65°C and pH 5.0 for 0-24 hours. The pre-incubations were followed by a 15-minute activity assay with 1% (w/v) CMC. The residual activity of the two enzymes is shown in ***Figure 4.3.5***. The residual activity of M9 is shown as relative activity to that of the wild-type.

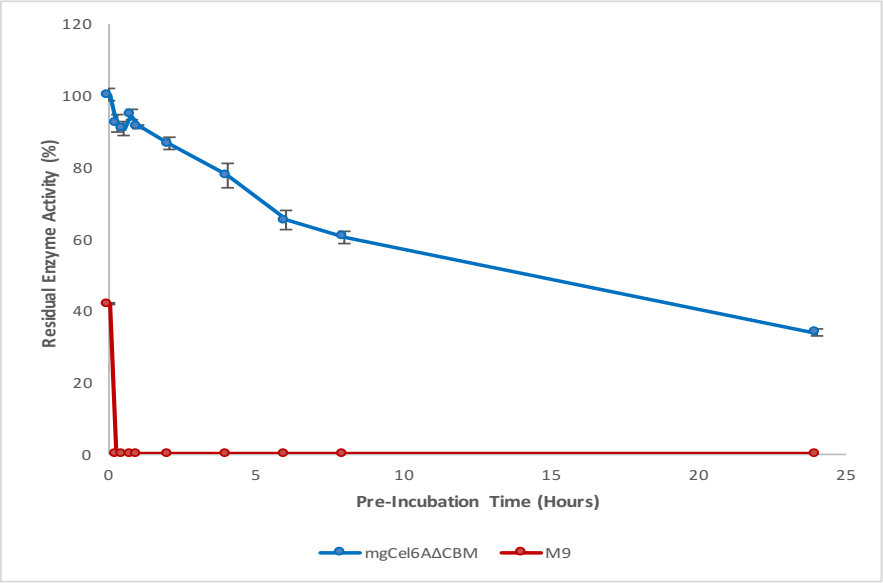


Figure 4.3.5. Stability of Mutant M9 and mgCel6A at pH 5.0 and 65°C. The figure shows the relative residual enzymatic activity of M9 as compared to the residual activity of the wild-type, following pre-incubation at 65°C in 75 mM phosphate-citrate buffer, pH 5.0. The buffer used had been adjusted to have the correct pH at the pre-incubation temperature. The pre-incubated enzyme samples were then tested for residual activity in a standard 15-minute assay at 85°C with 1% (w/v) CMC. The pHs of the reactions were not adjusted following pre-incubation. Product quantification was performed using the DNS method. All pre-incubations and subsequent reactions were performed in triplicates and the standard deviations are shown as error bars.

The data shows a drastic reduction in the thermal stability of M9 as compared to that of the wild-type. In fact, after only a 15-minute pre-incubation at 65°C and pH 5.0, the mutant showed 0% residual activity relative to the wild-type enzyme.

To further investigate the altered thermostability of M9, the mutant and mgCel6AΔCBM were subjected to progress curve assays at temperatures ranging from 30°C-60°C to monitor the stability of the enzymes over a shorter reaction time.

The progress curves were generated as described in Section 3.18.5. At 30°C, 40°C, and 50°C, the mutant enzyme showed identical product formation over time as compared to that of mgCel6AΔCBM (results not shown). However, at 60°C, a clear difference was observed in the stability of the two enzymes (***Figure 4.3.6***).

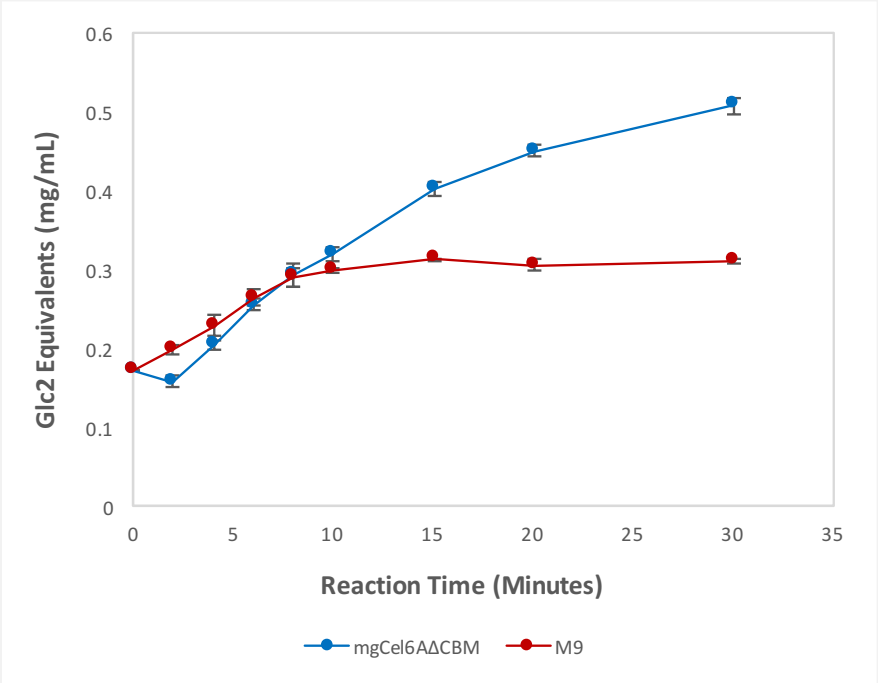


Figure 4.3.6. Progress Curve Experiment with mgCel6AΔCBM and M9. 1 µM enzyme was incubated at 60°C in 75 mM phosphate-citrate buffer, pH 5.0, with 1% (w/v) CMC for up to 30 minutes. The buffer was adjusted to have the correct pH at the relevant temperature. Product quantification was performed using the DNS method. All reactions were performed in triplicates and the standard deviations are shown as error bars.

Figure 4.3.6 shows that the product formation by M9 began to decrease after approximately 8 minutes, while the wild-type enzyme produced a more linear-type progress curve throughout the course of the 30-minute assay.

The clear decrease in thermostability observed during the characterization of M9 indicated that this enzyme was not a potential candidate for use at industrially desirable temperatures of 60°C or higher for more than 24 hours. The characterization of M9 was therefore not further pursued.

4.3.2.3. Characterization of M10, M11, and M12

The final mgCel6AΔCBM mutants, M10, M11, and M12, each contained a single mutation in or near the active site of the enzyme. The M10 mutant was designed based on the MSA of other GH6 enzymes with lower pH optima than the wild-type enzyme (see Section 4.1.4.2.1), and contained the T44D mutation. M11 and M12 each contained a mutation intended to reduce the pK_a of the catalytic acid, D158A and D158N. These two mutants were designed based on literature data (see Section 4.1.4.2.2).

Initial characterization of these mutants in a 24-hour assay on 1% (w/v) BALI™ cellulose at 60°C, using an enzyme concentration of 1 μM, demonstrated that the M10 mutant had a nearly comparable pH-activity profile to that of the wild-type, while mutants M11 and M12 had obtained significantly shifted pH optima. However, the activity of the latter two enzymes was significantly reduced as compared to that of the wild-type, to the extent that product formation could not be determined accurately using the DNS method (results not shown). In order to more accurately assess the pH-activity profiles of these mutants, a new pH optimum assay was performed with increased enzyme concentrations (***Figure 4.3.7***). The assay time was also shortened to 15 minutes to reduce the possibility of potential stability issues conferred by a longer reaction time. The assays were performed as described in Section 3.18.3 and product quantification was performed using HPAEC-PAD.

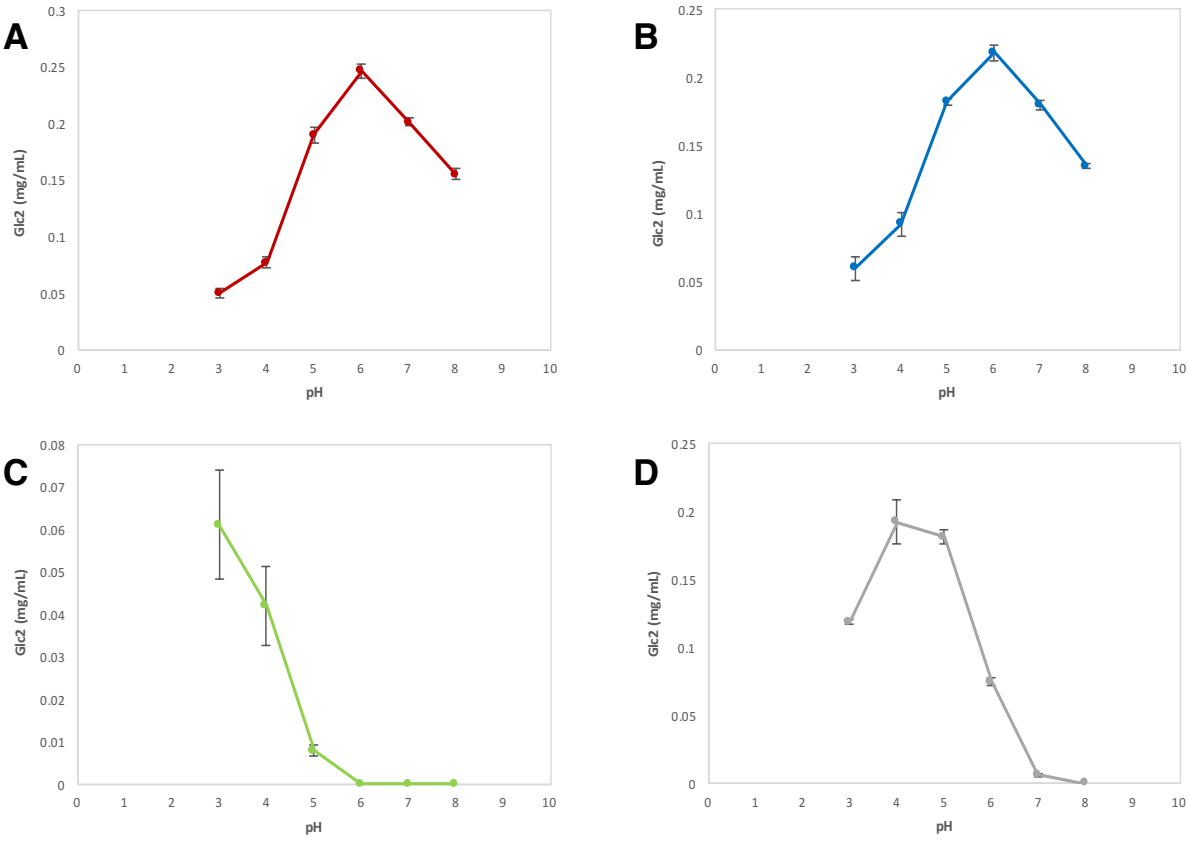


Figure 4.3.7. pH-Activity Profiles of mgCel6AΔCBM, M10, M11, and M12. The figure shows the pH optima of mgCel6AΔCBM (**Panel A**), M10 (**Panel B**), M11 (**Panel C**), and M12 (**Panel D**) derived from product formation in a 15-minute reaction at 60°C with 1% (w/v) BALI™ cellulose. Each reaction was performed in 75 mM phosphate-citrate buffer adjusted to have the correct pH at the relevant temperature before use. The enzyme concentrations used were 8 µM (Panels A and B) or 32 µM (Panels C and D). Product quantification was performed with HPAEC-PAD for all reactions. The activity is shown in terms of amount of product, expressed as cellobiose (Glc2) equivalents. All reactions were performed in triplicates and the standard deviations are indicated with error bars.

The data shown in ***Figure 4.3.7*** demonstrate that the pH optimum of M10 remained similar to that of the wild-type enzyme, namely at approximately 6.0. However, the pH optima of M11 and M12 were shifted towards more acidic pH values. The pH optimum of M12 was found to be approximately 4.0, while the pH optimum of M11 was found to be 3.0 or lower (the exact pH optimum could not be confirmed in this experiment, since buffers with pH values lower than 3.0 were not used).

To assess the stabilities of the enzymes over time, progress curves were generated for mgCel6AΔCBM and M10, M11, and M12 (***Figure 4.3.8***). The assays were performed as detailed in Section 3.18.5.

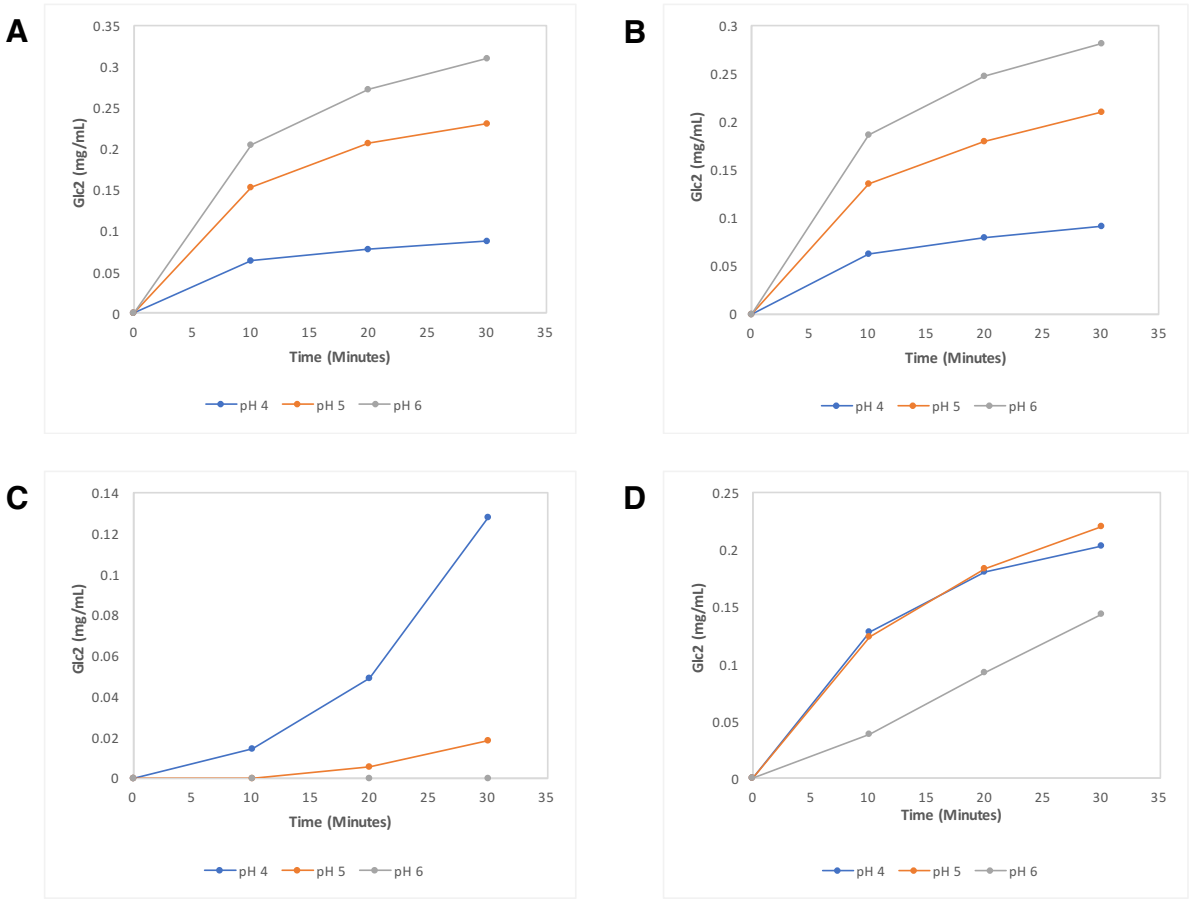


Figure 4.3.8. Progress Curves of mgCel6AΔCBM, M10, M11, and M12. Product formation over time was investigated for mgCel6AΔCBM (**Panel A**), M10 (**Panel B**), M11 (**Panel C**), and M12 (**Panel D**). The enzymes were incubated at 60°C and 75 mM phosphate-citrate buffer pHs 4.0, 5.0, and 6.0 with 1% (w/v) BALI™ cellulose. The enzyme concentrations used were 8 µM (Panels A and B) or 32 µM (Panels C and D). The buffers were adjusted to have the correct pH at the relevant temperature before use. Product quantification was performed using HPAEC-PAD for all reactions. The activity is shown in terms of amount of product, expressed as cellobiose (Glc2) equivalents. In these experiments, single samples were used instead of triplicates.

Figure 4.3.8 shows that product formation was not linear over time. Progress curves of mgCel6AΔCBM and M10 indicate that the activity of the enzyme gradually decreases over time, and that the two enzymes have similar stabilities and pH optima. The progress curve for M11 showed a remarkable increase in activity over time, and points towards an acidic pH optimum of 4.0 or lower. Regrettably, time course experiments to investigate product formation at pH 3.0 for this mutant were not performed. The progress curves for M12 are similar in shape to the curves for the wild-type, but show a pH optimum in the pH 4.0 to 5.0 range.

It is difficult to interpret the non-linearity of the rate of product formation, however, the overall increase in activity between each time point suggests that enzyme instability is most likely not a factor under the conditions used in these assays.

While the pH optima of M11 and M12 were considerably shifted towards more acidic values, their activities were also drastically reduced. To further compare the activities of the wild-type enzyme and M10, M11, and M12, and to evaluate the effects of more industrially relevant conditions on the pH-activity profiles of these enzymes, a 24-hour pH optimum assay was performed with 1% (w/v) BALI™ cellulose using an identical enzyme concentration in all reactions (***Figure 4.3.9***). The assay was performed as described in Section 3.18.3 (the assay was performed by Marianne S. Jensen.

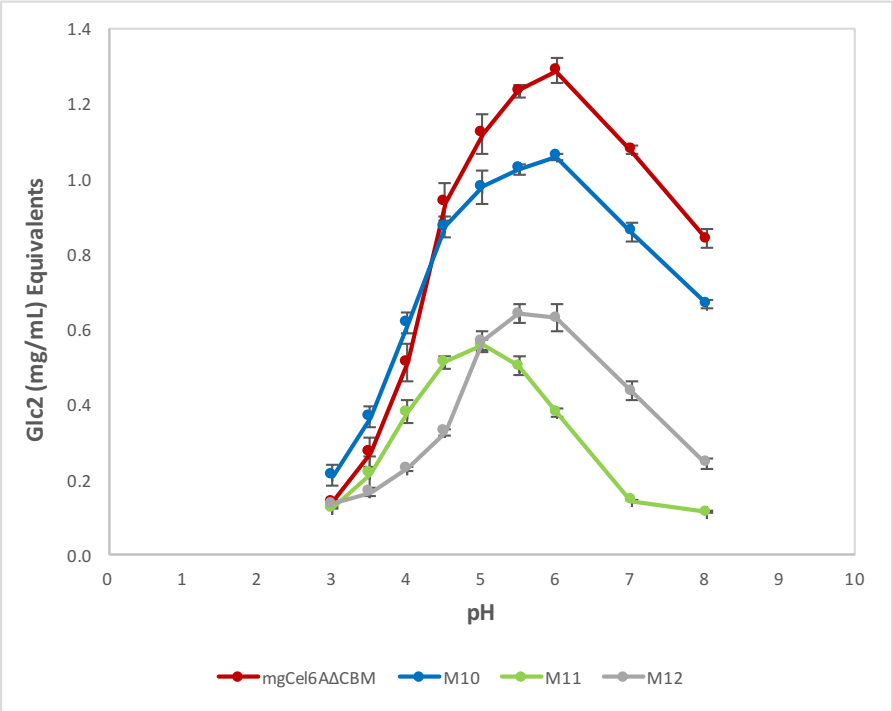


Figure 4.3.9. pH-Activity Profiles of mgCel6AΔCBM, M10, M11, and M12. The pH-activities of the enzymes were assessed in a 24-hour reaction at 60°C with 1% (w/v) BALI™ cellulose. Each reaction was performed in 75 mM phosphate-citrate buffer using an enzyme concentration of 16 µM. The buffers were adjusted to have the correct pH at the relevant temperature before use. Product quantification was performed using the DNS method. All reactions were performed in triplicates and the standard deviations are shown as error bars. Figure Source: Marianne S. Jensen.

In accordance with the results from 15-minute reactions with BALI™ cellulose (***Figure 4.3.7***), the pH optima of mgCel6AΔCBM and M10 were found to be approximately 6.0. However, relative to the 15-minute assays, the pH optima of M11 and M12 in these long-term

assays were higher, having increased from 3.0 and 4.0 to 5.0 and 5.5, respectively. These results may indicate that while the M11 and M12 mutants were most active at pH 3.0 and 4.0, respectively, during the short 15-minute reaction, they were perhaps more stable at higher pH values, resulting in increased product formation at pH values of 5.0 and 5.5, respectively, for the 24-hour reaction.

In a final characterization experiment, the three mutants and the wild-type enzyme were subjected to pH optimum assays with 0.2% (w/v) PASC, as described in Section 3.18.3. As the physiochemical properties of the substrate had previously been shown to significantly affect the pH optimum of mgCel6AΔCBM (***Figure 4.3.1***), it was considered interesting to investigate whether a shift in pH optimum would also be observed in the mutant enzymes with a different substrate. The results of these assays are shown in ***Figure 4.3.10***.

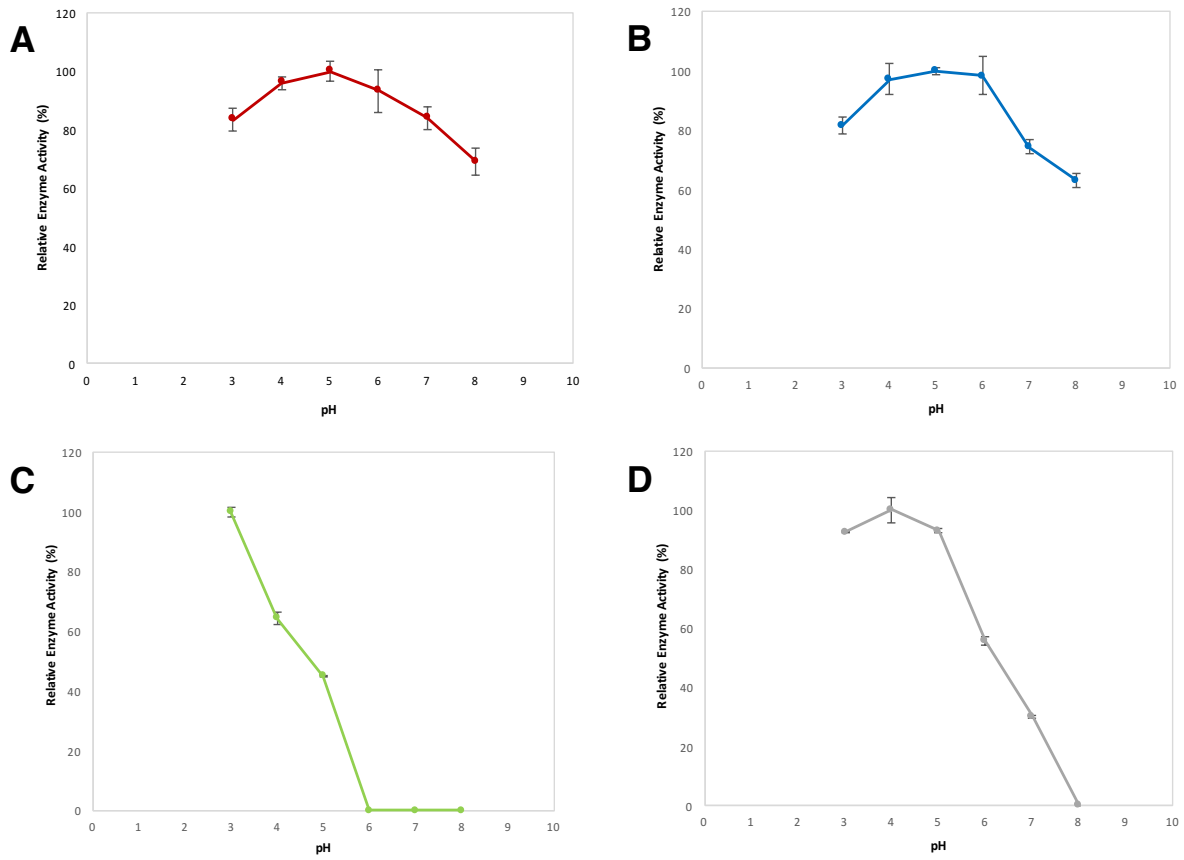


Figure 4.3.10. pH-Activity Profiles of mgCel6AΔCBM, M10, M11, and M12. pH-activity profiles were generated for mgCel6AΔCBM (**Panel A**), M10 (**Panel B**), M11 (**Panel C**), and M12 (**Panel D**) for a 15-minute reaction at 60°C with 0.2% (w/v) PASC. The enzyme concentrations used were 1 µM (Panel A), 8 µM (Panel B), or 32 µM (Panels C and D). Each reaction was performed in 75 mM phosphate-citrate buffer adjusted to have the correct pHs at the relevant temperature. Product quantification was performed using the DNS method. All reactions were performed in triplicates and the standard deviations are shown as error bars.

The results indicated that the pH-activity profile of M10 on PASC was again similar to that of mgCel6AΔCBM: both enzymes were found to have broad pH optima of around 5.0 on this substrate. M11 and M12, however, showed reduced pH optima as compared to the wild-type, similar to the results observed in the 15-minute pH optimum assay performed with BALI™ cellulose (***Figure 4.3.7***). M11 showed a pH optimum of 3.0 on PASC, however, buffers of pH values lower than 3.0 were not used, and the actual pH optimum of M11 under these conditions may therefore in fact be lower than 3.0. M12 was again found to have a pH optimum of 4.0, but the enzyme appeared to have increased enzymatic activity at pH 3.0 on this substrate as compared to its activity at this pH on the BALI™ cellulose substrate (***Figure 4.3.7***).

5. Discussion

The main objective of this study was to characterize and rationally engineer a thermostable GH6 for potential use in the industrial degradation of cellulose-rich biomass. The industrial conditions used in these degradations are often characterized by relatively harsh or extreme conditions, such as high temperatures and low pHs (Cockburn & Clarke 2011; Tynan-Connolly & Nielsen 2007). The mgCel6A enzyme was therefore targeted for mutagenesis with the aim of lowering the protein's pH optimum and increasing its stability at acidic pH.

5.1. Mutant Design

Throughout the course of this study, structural information and sequence comparisons were used to design mutant mgCel6AΔCBM variants. Mutations designed to target the pH-activity and stability of the enzyme were introduced both on the enzyme surface and near its active site.

The model of mgCel6AΔCBM generated using SWISS-MODEL (***Figure 4.1.3***) was deemed to be fairly reliable, due to the high sequence identity (70%) and the fact that no major insertions or deletions were observed in the alignment of the target protein sequence and the sequence of the catalytic domain of *Tf*Cel6A (PDB ID 1TML) (***Figure 4.1.2***) (Schwede et al. 2003). While two single insertions (Ala133 and Pro230) were found in the mgCel6AΔCBM sequence, an examination of the model showed that these residues are located far from the active site (approximately 18Å from the catalytic acid Asp118). The modeled active site and the remainder of the protein were thus considered to be representative of the structure of mgCel6AΔCBM, which indicated that rational mutant design based on the model structure would be reliable.

The amino acid sequence of the GH6 domain of mgCel6A was utilized to create a multiple sequence alignment (MSA) (***Figure 4.1.4***) of family 6 glycoside hydrolases with known pH optima to identify potential candidate residues for mutagenesis. As described in Section 4.1.3, this strategy was adapted from a 2011 engineering study of CenA from *Cellulomonas fimi* (Cockburn & Clarke 2011). The authors of this study cited a number of articles describing the pH optima of the enzymes used in their MSA. However, a closer investigation of these studies

revealed certain discrepancies. Some of the articles did not describe crucial assay parameters such as the substrate identity, substrate concentration, or reaction time, while others interestingly made no mention of the enzyme pH optimum. Therefore, only a selection of the GH6 sequences previously used by Cockburn and Clarke was included in the MSA generated in this study. Five additional enzymes described in studies published after the Cockburn and Clarke study were also included in this MSA.

The lack of concrete published data surrounding the determination of these optima confers a certain degree of unreliability to the interpretation of the MSA. As demonstrated in the present study, the pH optimum of an enzyme for a specific reaction can be affected by the physiochemical characteristics of the substrate and the assay conditions utilized. In addition, stability issues come into play during longer reactions, and can affect thus affect the determination of the pH optimum, as seen in the present study. Therefore, the descriptions of assay parameters such as substrate identity, substrate concentration, and reaction time are essential when comparing pH optima of enzymes.

In the mgCel6A assay performed during this study which resulted in a pH optimum of 5.0, 1% (w/v) CMC was used as the substrate in a 15-minute reaction. When creating the MSA, precautions were therefore taken to select enzymes whose pH optima were determined on similar substrates using short reaction times. However, increasing the amount of amino acid sequences used improves the probability of discovering significant information in the alignment. For this reason, some enzymes for which the pH optimum was determined on unknown or more crystalline substrates, or for which the reaction time was unknown, were also included in the MSA. As a result of the uncertainties surrounding the determination of the pH optima of some of the enzymes used, we cannot be certain that the alignment order provides a correct representation of descending pH optimum. This may have led to incorrect assumptions about which residues are conserved amongst enzymes with lower pH optima than mgCel6AΔCBM. This should be kept in mind when interpreting the results of the MSA and of the characterization of the mutant generated based on this MSA. In order to obtain a relative indication of the pH optima of some of the less well-characterized enzymes, it may be interesting to investigate the environments inhabited by the organisms from which these enzymes that produce these enzymes. Acidic environments would then be of special interest.

In spite of the ambiguities described, the MSA was used to identify 31 residues that were conserved amongst the enzymes presumed to have lower pH optima than mgCel6A. These residues were studied in the mgCel6AΔCBM model to determine their distance from the enzyme active site. The aim of the MSA was to identify candidate residues in mgCel6AΔCBM in or near the active site that could be mutated to a charged residue, with the aim of altering the local electrostatic network within the substrate-binding cleft. Residues found to be further than 5Å from the substrate binding site were therefore excluded as potential candidates. This left a single candidate residue in mgCel6AΔCBM: Thr44, which was replaced by an aspartate in the enzymes presumed to have a lower pH optimum than the target enzyme. This residue was therefore targeted for mutation to an aspartate in mutant M10 (T44D).

In addition, the modeled structure of mgCel6AΔCBM was used in combination with literature studies to select candidate residues for mutation intended to reduce the pH optimum of the enzyme and/or increase its stability at low pH. A total of 11 single mutations (including T44D) were selected for introduction into the enzyme, both on its surface and in its active site, as described in Section 4.1.4. The first 8 single mutations were also combined to create an additional mutant variant, M9, containing all eight single mutations (see ***Table 4.1.1*** for an overview of mutants). Following the rational design of these 12 mutant variants of mgCel6AΔCBM, attempts were made at producing all mutant genes, as described in Section 4.2.1.

5.2. Mutant Gene Production

In an attempt to generate the *M1*-*M8* genes, site-directed mutagenesis (SDM) was employed using the QuikChange II Site-Directed Mutagenesis Kit (Section 3.3.2). The efficacy of this kit proved to be low, as only two out of eight mutant genes were successfully created using this method (*M3* and *M6*). While transformed bacterial colonies were obtained for the remaining mutants, sequencing results showed that the colonies contained either the wild-type *mgcel6aΔcbm* gene, or genes with an incorrect DNA sequence. In the latter case, these genes were found to contain regions corresponding to areas of the pNIC-CH plasmid. The primer sequences used to generate these recombinant genes were therefore analyzed to locate

potential regions of complementarity to the vector. Several short complementary regions were discovered in these primers. While it is difficult to speculate as to why the QuikChange II Site-Directed Mutagenesis method was not very successful, supercoiling of the pNIC-CH plasmid may have enabled the primers to anneal at multiple sites, thus generating a partially incorrect gene (Dr. Lasse Fredriksen; personal communication). Due to time constraints, subsequent efforts to produce these mutant genes were discontinued.

The splicing by overlap extension method (SOE) (Section 3.3.3), however, resulted in the successful production of all three mutant genes made using this strategy (*M10*, *M11*, and *M12*) (***Figure 4.2.2***). In retrospect, the high success rate obtained using this method indicates that SOE should have been utilized to produce all mutant genes. Alternatively, new primer sequences should have been designed, with better control of potential multiple-annealing sites, and the QuikChange II Site-Directed Mutagenesis method should have been reattempted.

5.3. Production and Purification of mgCel6AΔCBM and Mutant Enzymes

All proteins, including the wild-type mgCel6AΔCBM, were expressed in One Shot ® BL21 Star™ (DE3) Chemically Competent *E. coli* cells following induction by isopropyl β-D-1-thiogalactopyranoside (IPTG), as described in Section 3.12.

The presence of distinct elution peaks during IMAC purification (***Figure 4.2.3***) signified that all proteins produced during this study had high affinities for the column. This was also confirmed by the lack of any significant amount of target protein in the IMAC flow-throughs, as verified by SDS-PAGE (***Figure 4.2.4***, ***Figure 4.2.5***, and ***Figure 4.2.6***).

The calculated final protein yields (***Table 4.2.2***) showed a clear variation in the amount of protein obtained for each mutant, however, the lowest yield obtained was for the M9 protein, indicating that this mutant was perhaps more difficult to express than the others. Indeed, a large amount of M9 protein was seen in the pellet of M9-producing cells following SDS-PAGE analysis. The excessive presence of target protein in the pellet may be due to the

formation of inclusion bodies during recombinant protein expression. These aggregates of protein are usually considered signs of incorrect folding and/or reduced stability (Marisch et al. 2013; Rosano & Ceccarelli 2014), and may explain the low overall yield observed for this mutant. While this mutant was designed with the hope of obtaining an additive effect of the putatively stability-conferring mutations from mutants M1-M8, these results seem to suggest that the introduction of the eight point mutations on the surface of this enzyme led to a decrease in stability. This assumption was confirmed in subsequent characterization assays (discussed in Section 5.5.2).

Minor peaks in the chromatograms of all proteins (***Figure 4.2.3***) can be seen at the beginning of the imidazole gradient, immediately prior to the elution of the target protein. These most likely correspond to contaminant *E. coli* host proteins, which can contain surface-exposed histidines and/or metal binding motifs, enabling them to bind to Ni^{2+} columns (Robichon et al. 2011). SDS-PAGE analyses also indicated that none of the IMAC-purified proteins were completely pure (***Figure 4.2.4***, ***Figure 4.2.5***, and ***Figure 4.2.6***). In retrospect, additional SDS-PAGE analyses should have been performed using 10-50X diluted samples of all fractions. This would have made it possible to determine the relative concentrations of contaminants, giving a more precise indication of protein purity, and thus the efficacy of the IMAC purification.

The protein fractions deemed to contain the highest amounts of target protein were pooled and subjected to subsequent ultrafiltration and buffer exchange (Section 3.16). Notably, the method employed during the buffer exchange may not have been optimal, since an identical volume of 20 mM Tris-HCl buffer was added to all protein samples, regardless of their varying imidazole concentrations resulting from IMAC purification.

The use of somewhat contaminated enzyme samples may potentially have affected subsequent activity assays. One possibility is that the presence of alternate proteins and/or imidazole has affected the photometric measurement of protein concentration (Section 3.17), thus leading to an underestimation of the actual concentration of target enzyme. Naturally, such an underestimation would influence the actual amount of enzyme used in all activity assays. In addition, contaminant proteins may have affected the hydrolysis yield in activity assays if these harbor cellulose-degrading properties. However, *E. coli* BL21 Star™ (DE3) is not known to express any such proteins, and the probability of this is therefore low (Gao et al.

2015). While the final concentration of imidazole in the purified proteins was estimated to be relatively low (approximately 0.2-0.3 mM in all final protein samples), the presence of imidazole may also potentially have affected the enzymatic binding of the substrate through potential stacking interactions with aromatic residues in the active site (Liao et al. 2013).

It would therefore have been wise to further purify the target enzymes using size-exclusion chromatography (SEC) following the initial IMAC purification. In addition to further reducing the amount of contaminant proteins, the use of SEC would have allowed for a more complete buffer exchange, ensuring that no residual imidazole was present in the final preparations of purified protein (G. Vaaje-Kolstad; personal communication).

5.4. Characterization of mgCel6A and mgCel6AΔCBM

At the start of this study, the wild-type full-length enzyme and its truncated variant, mgCel6AΔCBM, were characterized in terms of pH optimum and stability. This was done to determine how far the pH optimum would need to be shifted to fulfill the industrial goals of developing a cellulase capable of operating at pH 5.0 or lower and 60°C or higher for more than 24 hours. By comparing the two enzyme variants, the effects of the CBM2 domain, which had previously been examined in terms of thermal stability (Jensen et al. 2017), were further investigated in terms of pH optimum and pH stability.

5.4.1. pH Optimum

As described in Section 4.3.1.1, the full-length and truncated wild-type enzyme variants showed identical pH-activity profiles in all reactions tested, indicating that the CBM domain had no effect on the optimum pH of the enzyme. Initial pH optimum assays on mgCel6A indicated very different pH-activity profiles for the enzyme on different substrates (***Figure 4.3.1***). The enzyme was found to have a pH optimum of approximately 5.0 on CMC and PASC, and 6.0 on BALI™ cellulose, for 15-minute reactions.

The optimum temperature for the 15-minute assay on PASC was unfortunately not investigated due to time constraints. These reactions were therefore performed at 60°C, and

this may not have been the optimal temperature to generate the maximum amount of product from this substrate. Similarly, the 15-minute assay on BALI™ cellulose was performed at 85°C instead of at the enzyme's optimum temperature for this reaction time, 75°C, which may have had an effect on the amount of product generated.

mgCel6A was tested in 15-minute and 24-hour reactions with BALI™ cellulose. No difference was observed in pH optimum between the short and the long reactions, demonstrating that the enzyme is both most stable and most active at this pH on this substrate. However, the different pH optima of mgCel6A in reactions with different substrates demonstrates that the physiochemical properties of the substrate play an important role in determining the optimum conditions of the enzyme during hydrolysis.

In terms of pretreated lignocellulosic substrates such as BALI™ (see Section 1.2.4), residual lignin has been shown to affect both the rate and efficacy of enzymatic hydrolysis by making cellulose chains less accessible to cellulases, or by effecting non-productive enzymatic binding. While less is known about the potential effects of residual lignin on optimal enzyme conditions for hydrolysis, studies have shown that industrial pre-treatment methods can affect the structure of lignin in various ways (Chandra et al. 2007). One such study showed that the use of a liquid hot water pre-treatment method resulted in lignin with an increased proportion of negative charges. The authors postulated that an increase in negative charges in the pre-treated lignin may result in the formation of electrostatic interactions with the active site of cellulases, thus potentially impacting their rate of hydrolysis (Lu et al. 2016). Since the BALI™ cellulose substrate results from a sulfite cooking step (Sjöde et al. 2010), it is conceivable that residual pre-treated lignin can result in non-productive enzymatic binding and/or affect active site electrostatics in a pH-dependent manner, thus influencing the optimum pH of the enzyme for catalysis.

5.4.2. pH Stability

The full-length and truncated variants of mgCel6A were also investigated to assess the potential effects of the CBM2 domain on the enzyme's ability to withstand pre-incubations at lower pH values.

At 65°C, a substantial difference was observed in the stability of the full-length enzyme when pre-incubated in phosphate-citrate buffer of varying pH values (***Figure 4.3.2***). The enzyme was found to be considerably more stable at pH 6.0 for a 24-hour pre-incubation than at pH 5.0 or 4.0.

An identical pH stability assay using 60°C for the pre-incubation indicated that the full-length enzyme was equally stable at pH 5.0 and 6.0 at this temperature for the 24-hour pre-incubation period (results not shown). Notably, the 24-hour pH optimum assay on BALI™ cellulose (***Figure 4.3.1***) was also performed at 60°C, and showed that activity at pH 6.0 was high compared to pH 5.0. These results again highlight the considerable effects of substrate properties on the optimal conditions of the enzyme. The data obtained from these wild-type characterization assays, in combination with the industrial conditions of high temperature and low pH, suggested that designing a mutant variant of mgCel6A should aim to lower its pH optimum to 5.0 or lower, and concomitantly increase its stability at this reduced pH.

The pH stability assays described above were only performed on the full-length enzyme. The results from these assays were considered to be representative of both enzyme variants, however, due to their identical melting points (***Figure 1.7.1***), and nearly indistinguishable half-life curves (***Figure 4.3.3***) (discussed below), indicative of similar stability.

To analyze the potential effects of the CBM domain on the stability of mgCel6A at pH 5.0, the full-length and truncated enzyme variants were subjected to pre-incubations for up to 24 hours at 65°C, followed by a residual activity assay (***Figure 4.3.3***). In these pre-incubations, the phosphate-citrate buffer used had not been adjusted to have the correct pH at 65°C: the actual pH of the pre-incubation solutions was calculated to be approximately 4.7. This slight difference in pH may explain the obvious difference seen in the residual activity of the enzyme at pH 5.0 (and 4.7) in ***Figure 4.3.2*** and ***Figure 4.3.3***, and highlights the importance of adjusting the buffers to have the correct pH at the relevant experiment temperature. This was therefore done in all further experiments in this study.

The wild-type full-length and truncated enzyme variants showed nearly identical levels of residual activity after similar pre-incubation periods at pH 4.7, signifying that the CBM does not affect the stability of the enzyme at acidic pH. It would have been interesting to

investigate this assumption further by pre-incubating the full-length and truncated variants of mgCel6A at lower and higher pH values to observe whether the presence of the CBM may influence the protein's pH stability in more acidic or basic surroundings, and to obtain a true view of the possible effects of the CBM on pH stability.

As discussed in Section 1.6.2, an enzyme's stability is thought to result from the ways in which it folds to obtain its native structure. The burial of hydrophobic residues in the center of the protein, as well as the formation of various types of intramolecular interactions during folding, contribute to the stabilization of the protein's three-dimensional structure. CBM domains are not structurally dependent on catalytic domains, and are known to fold independently of these (Guillén et al. 2010; Venditto et al. 2015). The CBM is thus unlikely to affect the potential denaturation of the catalytic domain at extreme pH values.

5.4.3. Progress Curves

Progress curves investigating mgCel6AΔCBM activity over time on CMC (***Figure 4.3.6***) and BALI™ (***Figure 4.3.8***) were also performed during this study to compare the activity of mutant variants with the wild-type. The enzymatic activity of mgCel6AΔCBM on CMC increased steadily during the initial phase, but appeared to slowly taper as the reaction proceeded. Nonlinear, hyperbolic curves are often observed in the degradation of substrates such as CMC at specific enzyme and substrate concentrations (Zhang et al. 1999). Since dose response assays were not performed, it is difficult to discuss the implications of the gradual reduction in product formation observed for mgCel6AΔCBM.

The reduced product formation over time was also visible on the BALI™ cellulose substrate. The results of this progress curve should not be overinterpreted, since only a single experiment was performed, using single samples instead of triplicates due to time constraints. However, similar hyperbolic trends were observed in mgCel6A progress curves generated on BALI™ cellulose for much longer reaction times (72 hours) (Jensen et al. 2017). The decrease in activity observed over time may thus indicate that the enzyme hydrolyzes the more accessible areas of the BALI™ cellulose substrate in the initial period of the reaction, and that the rate of product formation gradually slows as the number of regions of substrate accessible to the enzyme decreases.

5.5. Characterization of Mutant mgCel6AΔCBM Variants

A total of six mutant variants of mgCel6AΔCBM were designed, produced, and characterized during this study: M3, M6, M9, M10, M11, and M12 (see ***Table 4.1.1*** for an overview of all mutants; see Appendix E for all mutant amino acid sequences). Mutants M3, M6, and M9 were designed with the aim of obtaining increased pH stability, whereas mutants M10, M11, and M12 were designed to increase the enzymatic activity at low pH.

5.5.1. Characterization of Mutants M3 (Q152E) and M6 (L165E)

As discussed in Section 5.2, of the eight point mutations designed to be introduced individually into mgCel6AΔCBM in M1-M8, only the *M3* and *M6* genes were successfully produced. The M3 and M6 variants each contained a single mutation on the surface of the enzyme. Mutant M3 was intended to carry an additional isolated negative charge on its surface (***Figure 4.1.5***). Additional negative surface charge was hoped to counteract the accumulation of positive charge on the enzyme surface at low pH, thus conferring increased stability to the protein at acidic pH values. Mutant M6 was designed to contain an additional surface salt bridge (***Figure 4.1.6***): the formation of salt bridges was hoped to introduce increased stability into the enzyme.

The M3 mutant contained the Q152E mutation, which is located in a loop on the opposite side of the enzyme seen from the active site. The M6 mutant contained the L165E mutation, which is located closer to the active site, in the same loop that harbors the Asp158 residue (mutated to Ala and Asn in mutants M11 and M12; Section 4.1.4.2.2).

The M3 and M6 mutants were characterized in 24-hour pH optimum assays with 1% (w/v) BALI™ cellulose alongside the wild-type mgCel6AΔCBM (***Figure 4.3.4***). As compared to the wild-type, both mutants M3 and M6 showed identical product formation at pHs 4.0, 5.0, and 6.0, and identical pH optima for this 24-hour reaction with BALI™ cellulose. While progress curves and pre-incubation-based pH stability assays were not performed, no increased product formation was observed for the mutants at pH 4.0 or 5.0 for a 24-hour assay. These results indicated that mutants M3 and M6 were not more stable than the wild-type enzyme at acidic pH for an industrially-relevant reaction time and temperature.

While it is difficult to understand why neither mutant showed any difference in stability as compared to that of the wild-type, it is tempting to assume that single point mutations on the surface have too small an effect to cause significant changes in overall stability. Indeed, some authors have claimed that mutations of surface residues have minor effects on protein stability and structure (Dill et al. 2008; Getov et al. 2016; Matthews 1995). However, other studies have shown that the stability of a protein can be significantly improved by altering surface charges (Grimsley et al. 1999; Loladze et al. 1999; Spector et al. 2000; Xiao et al. 2013). Similarly, studies of protein stability have also found that some individual salt bridges can have considerable effects on overall stability (Kumar & Nussinov 1999; Makhatadze et al. 2003; Sun et al. 1991). These examples are more often than not based on continuum electrostatics and model calculations, which allow individual residues' contributions towards overall protein stability to be assessed, thus enabling residues shown to negatively contribute to the protein stability to be selected for mutation, and residues or interactions that contribute favorably to overall stability to be identified. These strategies were not employed in this study, and one could conclude from the (only two) surface point mutations studied here that more advanced mutant design is appropriate.

5.5.2. Characterization of Mutant M9 (N21D/N46D/A104E/Q152E/L165E/I242E/I271D/Q276E)

The multiple mutant M9 contained all eight point mutations introduced in mutants M1-M8 (see Section 4.1.4.1.3). The aim was to reduce the potential accumulation of positive charge on the surface of mgCel6AΔCBM at low pH by introducing negative charges. The mutant also contained four additional putative salt bridges hoped to contribute to increased stability.

The multiple mutant M9 was produced less efficiently than the wild-type (***Table 4.2.2***), and protein precipitation was observed in the pellet after sonication (***Figure 4.2.5***), suggesting that the excess of surface charges may have caused the mutant to aggregate faster due to potential electrostatic interactions with other proteins, or that this mutant had reduced stability. The stability assay performed at 65°C and pH 5.0 (***Figure 4.3.5***) showed a decrease in thermostability as compared to that of the wild-type.

It is important to note, however, that the residual activity of mgCel6AΔCBM at pH 5.0 shown in ***Figure 4.3.5*** was significantly reduced from the residual activity shown in ***Figure 4.3.2***. While these two experiments were performed under identical conditions, using buffers adjusted to have the correct pH at 65°C, the residual activity of mgCel6AΔCBM deviates significantly over time between these two experiments. It is therefore difficult to draw any conclusions from this experiment, since mgCel6AΔCBM did not show the operational stability that has been observed earlier under these conditions. Ideally, the experiment should have been repeated to ensure a similar stability profile of mgCel6AΔCBM before drawing any conclusions about the stability of the M9 mutant.

In spite of these inconsistencies, the results indicated that this mutant was less stable than the wild-type. To assess the reduced thermal stability of mutant M9, progress curves at 30°C, 40°C, 50°C, and 60°C were generated for the mutant and for mgCel6AΔCBM. A clear difference in enzymatic activity was seen only at 60°C (***Figure 4.3.6***), indicating that the thermal stability had only been somewhat reduced in this mutant.

The reduction of thermostability observed during the characterization of mutant M9 indicated that the pH stability assay should have been repeated at a lower temperature (for example 50°C) using a range of pHs, to more correctly assess the effects of the eight point mutations in mutant M9 on its pH stability. Due to the reduction in thermostability, the mutant enzyme was no longer considered a potential candidate for use in industrial conditions, and its characterization was discontinued due to time constraints.

While the pH stability of the M9 mutant was not sufficiently assessed, a decrease in thermostability suggests that one or more of the mutations introduced destabilized mgCel6AΔCBM. It would therefore be highly interesting to perform further mutagenesis studies on the M9 mutant to determine which mutation(s) contributed so significantly to the reduced stability. By “deconstructing” the mutant, removing mutations one by one, it may be possible to gain a deeper understanding of the effects of each mutation on the stability of the mutant. Such studies may perhaps provide clues for future stability engineering ventures.

As mentioned in Section 4.1.4.1.1, the contributions of various intramolecular interactions towards protein stability have been the subject of much interest (Lindman et al. 2006;

Sanchez-Ruiz & Makhatadze 2001; Sippel & Quiocho 2015; van den Burg & Eijsink 2002). However, the contribution of isolated surface charges that are not participants in ionic bonds towards protein stabilization remains uncertain.

In addition, while there is a significant amount of published literature available concerning the engineering of enzymes for improved thermal stability, fewer studies have focused solely on engineering towards increased pH stability. One such study, however, did in fact show that the introduction of several basic residues on the surface of a peroxidase enzyme led to its increased stability at low pH (Sáez-Jiménez et al. 2015). The introduction of positive charges on the enzyme surface may therefore also be an interesting mutational strategy in the continued development of mgCel6AΔCBM. It is important to note, however, that this study was based on comparative engineering with a characterized, highly pH-stable peroxidase, so, while a mutant with a significantly increased pH stability was obtained, the molecular basis for the increased stability is perhaps more difficult to explain.

5.5.3. Characterization of Mutants M10 (T44D), M11 (D158A), and M12 (D158N)

Mutant M10 (T44D) was designed based on the MSA generated of 14 biochemically characterized family 6 glycoside hydrolases, including mgCel6AΔCBM (***Figure 4.1.4***). As described, the Thr44 residue was chosen based on the observation that family 6 GHs with low pH optima have a conserved aspartate in this position. Thr44 was found to be located 4Å from a modeled substrate molecule, and approximately 14Å away from the catalytic acid (Asp118), on the other side of the active site cleft (***Figure 4.1.7***). Despite its considerable distance from the catalytic acid, the mutation was considered to have potential for reducing the pH optimum of the enzyme, based on the observed sequence conservation.

The two final mutants, M11(D158A) and M12 (D158N), were designed based on a strategy employed in a study of the catalytic mechanism of *Tf*Cel6A (previously known as *Thermomonospora fusca* E2) (Section 4.1.4.2.2) (Wolfgang & Wilson 1999). Through mutagenesis studies, the authors demonstrated that an aspartate (Asp156) located in close proximity to the catalytic acid was responsible for raising the pK_a of this residue. In mgCel6AΔCBM, this aspartate corresponds to Asp158. The model of mgCel6AΔCBM

showed that Asp158 and the catalytic acid (Asp118) are separated by approximately 4Å (based on the homology model; predicted distances are therefore not exact) (***Figure 4.1.8***). Asp158 was thus assumed to play the same role in mgCel6AΔCBM. The mutation of the Asp158 residue to alanine (Ala) and asparagine (Asn), respectively, was therefore presumed to reduce the pH optimum of mgCel6AΔCBM by lowering the pK_a of the catalytic acid.

Mutants M10, M11, and M12 were successfully produced and purified via IMAC, but showed lower protein yields compared to the wild-type, mutant M3, and mutant M6 (***Table 4.2.2***). Protein precipitation was not observed in SDS-PAGE, indicating that these proteins were not misfolded or significantly unstable.

A 15-minute pH optimum assay of the M10, M11, and M12 mutants and mgCel6AΔCBM using BALI™ cellulose (***Figure 4.3.7***) indicated that the pH optimum of mgCel6AΔCBM had not been altered by the introduction of the T44D mutation in mutant M10. The pH-activity profiles of M11 and M12, however, showed a significant shift towards more acidic pH values. M12 had a pH optimum of 4.0 for this reaction, as compared to the pH optimum of 6.0 for mgCel6AΔCBM. M11 was found to have a pH optimum of 3.0 or lower. Regrettably, assays at pHs lower than 3.0 were not performed, so it is difficult to conclude whether the pH optimum may, in fact, have been even more reduced.

It should however be noted that this 15-minute assay was performed at 60°C instead of at the optimum temperature for the wild-type enzyme for this reaction time (75°C), and this may have eliminated the possibility of observing potential reductions in thermostability of the M10, M11, and M12 mutants as compared to mgCel6AΔCBM. However, due to time constraints, it was considered more interesting to characterize the pH optimum mutants at an industrially-relevant temperature.

A pH optimum assay performed on 0.2% PASC (***Figure 4.3.10***) also indicated that mutant M10 had an identical pH optimum (5.0) and a very similar pH-activity profile to that of mgCel6AΔCBM on this substrate. Mutants M11 and M12, however, showed pH optima of 3.0 (or lower) and 4.0, respectively. Notably, the M12 mutant appeared to have higher activity at pH 3.0 on the PASC substrate than on the BALI™ cellulose substrate, perhaps reflecting the effects of the substrate characteristics on the pH- activity profile of the enzyme (see Section 5.4.1).

Progress curves at 60°C were subsequently generated to investigate enzymatic activity over time on the BALI™ cellulose substrate (***Figure 4.3.8***). None of the enzymes showed linear product formation over time at their optimum pH (determined in the BALI™ assay, ***Figure 4.3.7***) (note however that a progress curve for M11 at pH 3.0 was not determined).

As discussed in Section 5.4.3, the wild-type enzyme showed a hyperbolic trend at pHs 5.0 and 6.0, with a rapid initial product formation followed by a gradual decrease in enzyme activity. Similar results were observed in the progress curves of mutant M10, indicating that this mutant had a similar stability and pH optimum to the wild-type. Interestingly, mutant M11 showed a nearly exponential increase in product formation at pH 4.0 over the 30-minute reaction time. Mutant M12 showed a near-linear rate of product formation at pH 6.0 throughout the assay, but activity was seen to gradually decrease over time at pHs 4.0 and 5.0, similar to the trend observed in the wild-type and M10 progress curves at pHs 5.0 and 6.0.

While the rate of product formation was not linear for the wild-type enzyme or for any of the three mutants, all the enzymes were active over a 30-minute period at their pH optima (and pH 4.0 for M11), indicating that instability of the enzymes was most likely not a factor under the conditions used in the pH optimum assay on BALI™ (***Figure 4.3.7***). As previously mentioned, however, these results should not be overinterpreted, since the experiment was only performed once, and single samples were used instead of triplicates.

In order to more accurately compare the efficacy of mutants M10, M11, and M12 with that of mgCel6AΔCBM, and simultaneously evaluate the effects of more industrially-relevant conditions on the pH activity profiles of these mutants, a 24-hour reaction with BALI™ cellulose was performed (***Figure 4.3.9***) using identical enzyme concentrations for the wild-type and all three mutants (experiment performed by Marianne S. Jensen). Results from this experiment indicated that the M10 mutant showed an identical pH-activity profile to that of the wild-type, as expected based on the identical pH optimum and stability seen in the 15-minute reaction with BALI™ and the progress curve assay (***Figure 4.3.7*** and ***Figure 4.3.8***, respectively). The pH optima of mutants M11 and M12, however, were increased from what was observed in the 15-minute BALI™ reactions: the pH optimum of M11 increased from 3.0 (or lower) to 5.0, and the pH optimum of M12 increased from 4.0 to 5.5.

The identical pH optimum of mutant M10 (T44D) to that of the wild-type was interesting, since an aspartate residue at this position was found to be conserved in the five GH6s with presumed lower pH optima than mgCel6AΔCBM. Model inspection showed that the Thr44 residue was approximately 14Å from the catalytic acid Asp118, so any potential change in pH optimum in this mutant from that of the wild-type could only have been explained by a significant structural change in the active site during catalysis, enabling the Thr44 residue to move closer to the catalytic acid. This type of structural change seems unlikely, however, as the Thr44 residue is located in a tight loop between a β-strand and an α-helix (***Figure 4.1.7***).

The similarity in both pH optimum and enzymatic activity of mutant M10 to the wild-type seems to indicate that the replacement of the Thr44 residue with an acidic amino acid did not significantly affect productive interactions with the substrate, despite the theoretically reduced distance between the T44D mutated residue and the modeled substrate molecule. These types of interactions are extremely difficult to predict, however, since the use of a model structure confers certain unreliabilities concerning the spatial orientation of the substrate molecule during binding. In addition to these challenges, the detailed catalytic mechanism of the GH6 family is not yet fully understood (Section 1.5.3.1), implying that there may potentially be unpredictable structural transformations occurring in the active site during GH6 substrate hydrolysis. It is also conceivable that the low pH optima of the enzymes with an Asp in this position stems from multiple conserved residues in or near the active site (perhaps including this Asp) that may affect substrate binding and/or active site electrostatics. A more acidic pH optimum may thus be dependent on the presence of several specific residues in certain positions, and the mutation of only one of these positions to an Asp in mgCel6AΔCBM was not sufficient to observe any shift in pH optimum.

The pH-activity profile shifts of mutants M11 (D158A) and M12 (D158N), as well as their substantially reduced activity, were not surprising, since similar results had been obtained in the *Tf*Cel6A study mentioned above (Wolfgang & Wilson 1999). *Tf*Cel6A mutants D156A and D156N showed acidic shifts in their pH-activity profiles (from 6.0 to 4.0 and 5.0, respectively), however the activity of these mutants was greatly reduced: mutants D156A and D156N showed 2.2% and 10% respectively of the wild-type activity on CMC, and 67% and 76% respectively of the wild-type activity on filter paper. The shifted pH optima of the *Tf*Cel6A mutants observed corresponded well with the proposed role of Asp156 in raising the pK_a of the catalytic acid Asp117, as discussed above. Increasing the pK_a value of an acidic

residue leads to its protonation at higher pH values. By removing the carboxylic acid group of the side chain of Asp156, its effect on the pK_a of the catalytic acid is eliminated, and this catalytic residue will thus only be protonated at a lower pH.

In the *Tf*Cel6A study, a lower pH optimum was observed for the D156A mutant than in the D156N mutant. The same results were observed in the present study (mutant M11, D158A, had a lower pH optimum than mutant M12, D158N). As discussed by the authors, this is most likely due to the presence of a terminal oxygen atom in the side chain of the Asn residue of the M12 mutant, which may carry a partial negative charge and thus slightly raise the pK_a of the catalytic acid (Wolfgang & Wilson 1999). The side chain of Ala, which was introduced in the D158A mutation in M11, contains only a methyl group which is unlikely to form any interactions with the catalytic acid, and therefore represents the complete removal of the Asp158 interaction with the catalytic acid.

The reduced pH optima of the present study's M11 and M12 mutants indicate that Asp158 performs the same role in mgCel6A as Asp156 in *Tf*Cel6A: raising the pK_a of the catalytic acid. However, the decrease in activity of the M11 and M12 mutants (which had approximately 40% and 45% of the wild-type activity on BALI™ cellulose, respectively) is more difficult to understand. While it is challenging to predict the effects of any rationally designed mutations, this is especially difficult in the active sites of enzymes where the binding and release of substrate molecules is constantly affecting local conditions in the already dynamic protein structure.

One potential explanation for the reduced activity observed may be that the pK_a values of other charged residues in the active site (ex: D80, R79, K262, and D268), which have been postulated to perform various critical roles during the enzymatic hydrolysis reaction in other GH6s (Section 1.5.3.1) (André et al. 2003; Damude et al. 1995; Wolfgang & Wilson 1999), have not been altered, and are therefore still optimized for catalysis at a higher pH. The shifted pH optima of mutants M11 and M12 indicate that the active site electrostatics have been altered, and many more changes may be needed to optimize for catalysis at low pH. A potential future area of interest may be to attempt to design novel mutations that may lower the pK_a values of other residues shown to be important for catalysis, for example, D80 and D268. However, such studies would be facilitated with a solved crystal structure of the enzyme.

It may also be of interest to attempt to introduce additional mutations that increase activity on cellulose-based substrates. In one study, several such successful mutations were introduced into *Tf*Cel6A by mutating conserved, non-catalytic residues in the active site. One such mutation, K259H, resulted in a 273% increase in activity on CMC, as compared to that of the wild-type (Zhang et al. 2000). This type of engineering of an mgCel6AΔCBM mutant with an already reduced pH optimum may therefore also represent a potential future avenue for further development of a candidate enzyme optimized for use in industrial conditions.

5.6. Conclusion and Future Work

The focus of this study was the rational engineering of the catalytic domain of the thermostable mgCel6A enzyme. The aim was to obtain a mutated mgCel6AΔCBM variant with potential use in industrial conditions, typically characterized by high temperatures and low pH. A total of six mutants of mgCel6AΔCBM were successfully produced and characterized to determine the effects of the mutations on the enzyme's pH optimum and stability.

Two of the six mutants (M11, D158A, and M12, D158N) were found to have reduced pH optima as compared to that of the wild-type enzyme. The potential of these mutants for future industrial use, however, remains limited, due to a concomitant reduction in enzymatic activity. Future mutagenesis studies of these mutants may perhaps aid in understanding the drastically reduced activity, as well as in designing modified versions of these mutants with reduced pH optima and increased activity. Such studies are more likely to be successful with a solved three-dimensional structure.

Protein engineering is an ever-evolving field, and while the generation of novel proteins based solely on rational design and structural studies is an intriguing concept, it is widely known that there are also many associated challenges. More recently, in conjunction with high-throughput screening methods, numerous studies have shown that using a combination of directed evolution and rational engineering is more likely to yield enzymes with specific desirable characteristics (Bommarius et al. 2011; Chica et al. 2005; Eijsink et al. 2005). One such technique, iterative saturation mutagenesis (ISM) (Reetz & Carballeira 2007), is based on the

saturation mutagenesis (SM) of a small number of rationally selected residues, generating a few libraries of mutant proteins from which candidates with promising features can be selected and used as templates in new rounds of SM. In terms of further engineering of mgCel6AΔCBM, the use of methods such as ISM may provide potentially critical information concerning the roles of specific residues surrounding the catalytic acid in determining the pH optimum of the enzyme.

The global abundance of cellulose as an organic renewable resource makes it an interesting candidate for the long-term replacement of fossil resources. Cellulase enzymes represent a promising avenue for efficient conversion of cellulose to its constituent monosaccharides, which can be fermented to produce alcohols that can be further utilized in the production of biofuels, or used to produce other value-added products. There is thus a growing need for the development of cost-effective and efficient cellulases capable of operating under the relatively harsh conditions employed in the industrial degradation of lignocellulosic biomass. The M11 and M12 mutant variants of mgCel6AΔCBM produced in this study may be considered promising future candidates for use in industrial enzymatic cocktails, however, additional studies are required to further understand the significant reduction in activity observed following mutation of the Asp158 residue. The hybrid approach of directed evolution in combination with rational engineering may potentially be a highly relevant and useful strategy in the continued development and engineering of mgCel6AΔCBM towards industrial conditions.

6. References

Abdel-Hamid, A. M., Solbiati, J. O. & Cann, I. K. (2013). Insights into lignin degradation and its potential industrial applications. *Adv Appl Microbiol*, 82: 1-28.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2014). *Molecular Biology of the Cell*. 6th ed. New York: Garland Science.

André, G., Kanchanawong, P., Palma, R., Cho, H., Deng, X., Irwin, D., Himmel, M. E., Wilson, D. B. & Brady, J. W. (2003). Computational and experimental studies of the catalytic mechanism of *Thermobifida fusca* cellulase Cel6A (E2). *Protein Eng*, 16 (2): 125-34.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22 (2): 195-201.

Barlow, D. J. & Thornton, J. M. (1983). Ion-pairs in proteins. *J Mol Biol*, 168 (4): 867-85.

Bauer, S., Vasu, P., Persson, S., Mort, A. J. & Somerville, C. R. (2006). Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc Natl Acad Sci U S A*, 103 (30): 11417-22.

Benkert, P., Biasini, M. & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27 (3): 343-50.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28 (1): 235-242.

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

Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences*, 103 (15): 5869-5874.

Bommarius, A. S., Blum, J. K. & Abrahamson, M. J. (2011). Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr Opin Chem Biol*, 15 (2): 194-200.

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

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2008). Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protocols*, 4 (1): 1-13.

Bornscheuer, U. T. & Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Curr Opin Chem Biol*, 5 (2): 137-43.

Brannigan, J. A. & Wilkinson, A. J. (2002). Protein engineering 20 years on. *Nat Rev Mol Cell Biol*, 3 (12): 964-70.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421-421.

Carter, P. (1986). Site-directed mutagenesis. *Biochem J*, 237 (1): 1-7.

Chami Khazraji, A. & Robert, S. (2013). Interaction Effects between Cellulose and Water in Nanocrystalline and Amorphous Regions: A Novel Approach Using Molecular Modeling. *Journal of Nanomaterials*, 2013: 10.

Chandra, R. P., Bura, R., Mabee, W. E., Berlin, A., Pan, X. & Saddler, J. N. (2007). Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Adv Biochem Eng Biotechnol*, 108: 67-93.

Cherubini, F. (2010). The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management*, 51 (7): 1412-1421.

Chica, R. A., Doucet, N. & Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol*, 16 (4): 378-84.

Cockburn, D. W., Vandenende, C. & Clarke, A. J. (2010). Modulating the pH-activity profile of cellulase by substitution: replacing the general base catalyst aspartate with cysteinesulfinate in cellulase A from *Cellulomonas fimi*. *Biochemistry*, 49 (9): 2042-50.

Cockburn, D. W. & Clarke, A. J. (2011). Modulating the pH-activity profile of cellulase A from *Cellulomonas fimi* by replacement of surface residues. *Protein Eng Des Sel*, 24 (5): 429-37.

Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. (1995). Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. *Biochemistry*, 34 (7): 2220-4.

Daniel, R. M., Dines, M. & Petach, H. H. (1996). The denaturation and degradation of stable enzymes at high temperatures. *Biochemical Journal*, 317 (Pt 1): 1-11.

Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. (1991). Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry*, 30 (49): 11521-9.

Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3 (9): 853-9.

de Gonzalo, G., Colpa, D. I., Habib, M. H. M. & Fraaije, M. W. (2016). Bacterial enzymes involved in lignin degradation. *Journal of Biotechnology*, 236: 110-119.

Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. (2008). The Protein Folding Problem. *Annual review of biophysics*, 37: 289-316.

Ding, S. Y., Adney, W. S., Vinzant, T. B., Himmel, M. E. & Decker, S. R. (2003). *Thermal tolerant cellulase from Acidothermus cellulolyticus*. International Patent Number WO/2003/012109.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, 426 (6968): 884-90.

Eijsink, V. G., Gaseidnes, S., Borchert, T. V. & van den Burg, B. (2005). Directed evolution of enzyme stability. *Biomol Eng*, 22 (1-3): 21-30.

Ferreira-Leitão, V., Gottschalk, L. M. F., Ferrara, M. A., Nepomuceno, A. L., Molinari, H. B. C. & Bon, E. P. S. (2010). Biomass Residues in Brazil: Availability and Potential Uses. *Waste and Biomass Valorization*, 1 (1): 65-76.

Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol*, 224 (3): 771-82.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44 (D1): D279-85.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research*, 45 (Database issue): D190-D199.

Gao, D., Luan, Y., Wang, Q., Liang, Q. & Qi, Q. (2015). Construction of cellulose-utilizing *Escherichia coli* based on a secretable cellulase. *Microbial Cell Factories*, 14: 159.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In Walker, J. M. (ed.) *The Proteomics Protocols Handbook*, pp. 571-607: Humana Press.

Getov, I., Petukh, M. & Alexov, E. (2016). SAAFEC: Predicting the Effect of Single Point Mutations on Protein Folding Free Energy Using a Knowledge-Modified MM/PBSA Approach. *International Journal of Molecular Sciences*, 17 (4): 512.

Gírio, F. M., Fonseca, C., Carvalheiro, F., Duarte, L. C., Marques, S. & Bogel-Łukasik, R. (2010). Hemicelluloses for fuel ethanol: A review. *Bioresour Technol*, 101 (13): 4775-800.

Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci*, 8 (9): 1843-9.

Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology*, 85 (5): 1241-1249.

Habibi, Y., Lucia, L. A. & Rojas, O. J. (2010). Cellulose nanocrystals: chemistry, self-assembly, and applications. *Chem Rev*, 110 (6): 3479-500.

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. & Foust, T. D. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315 (5813): 804-7.

Hirata, A., Adachi, M., Utsumi, S. & Mikami, B. (2004). Engineering of the pH optimum of *Bacillus cereus* beta-amylase: conversion of the pH optimum from a bacterial type to a higher-plant type. *Biochemistry*, 43 (39): 12523-31.

Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels*, 5 (1): 45.

Jaenicke, R. (1991). Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem*, 202 (3): 715-28.

Jensen, M. S., Fredriksen, L., Mackenzie, A. K., Pope, P. B., Chylenski, P., Williamson, A. K., Østby, H., Vaaje-Kolstad, G. & Eijsink, V. G. (2017). *Discovery and characterization of a thermostable GH6 cellulase from a compost metagenome*. Submitted, Under Revision. Unpublished manuscript.

Kaushik, M., Sinha, P., Jaiswal, P., Mahendru, S., Roy, K. & Kukreti, S. (2016). Protein engineering and de novo designing of a biocatalyst. *J Mol Recognit*, 29 (10): 499-503.

Kim, T., Mullaney, E. J., Porres, J. M., Roneker, K. R., Crowe, S., Rice, S., Ko, T., Ullah, A. H. J., Daly, C. B., Welch, R., et al. (2006). Shifting the pH Profile of *Aspergillus niger* PhyA Phytase To Match the Stomach pH Enhances Its Effectiveness as an Animal Feed Additive. *Applied and Environmental Microbiology*, 72 (6): 4397-4403.

Klemm, D., Heublein, B., Fink, H. P. & Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angew Chem Int Ed Engl*, 44 (22): 3358-93.

Koivula, A., Reinikainen, T., Ruohonen, L., Valkeajarvi, A., Claeyssens, M., Teleman, O., Kleywegt, G. J., Szardenings, M., Rouvinen, J., Jones, T. A., et al. (1996). The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. *Protein Eng*, 9 (8): 691-9.

Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., et al. (2002). The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. *J Am Chem Soc*, 124 (34): 10015-24.

Krause, D. O., Denman, S. E., Mackie, R. I., Morrison, M., Rae, A. L., Attwood, G. T. & McSweeney, C. S. (2003). Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiol Rev*, 27 (5): 663-93.
Krieger, F., Moglich, A. & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J Am Chem Soc*, 127 (10): 3346-52.
Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J Mol Biol*, 293 (5): 1241-55.
Kumar, S., Tsai, C. J., Ma, B. & Nussinov, R. (2000). Contribution of salt bridges toward protein thermostability. *J Biomol Struct Dyn*, 17 Suppl 1: 79-85.
Lee, H. V., Hamid, S. B. & Zain, S. K. (2014). Conversion of lignocellulosic biomass to nanocellulose: structure and chemical process. *ScientificWorldJournal*, 2014: 631013.
Lesk, A. M. (2010). *Introduction to Protein Science*. 2nd ed. New York: Oxford University Press.
Liao, S.-M., Du, Q.-S., Meng, J.-Z., Pang, Z.-W. & Huang, R.-B. (2013). The multiple roles of histidine in protein interactions. *Chemistry Central Journal*, 7 (1): 44.
Lindman, S., Xue, W.-F., Szczepankiewicz, O., Bauer, M. C., Nilsson, H. & Linse, S. (2006). Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophysical Journal*, 90 (8): 2911-2921.
Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a Thermostable Protein via Optimization of Charge−Charge Interactions on the Protein Surface. *Biochemistry*, 38 (50): 16419-16423.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, 42 (Database issue): D490-5.
Lu, X., Zheng, X., Li, X. & Zhao, J. (2016). Adsorption and mechanism of cellulase enzymes onto lignin isolated from corn stover pretreated with liquid hot water. *Biotechnology for Biofuels*, 9 (1): 118.
Lutz, S. (2010). Beyond directed evolution--semi-rational protein engineering and design. *Curr Opin Biotechnol*, 21 (6): 734-43.
Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J Mol Biol*, 327 (5): 1135-48.
Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M. & Striedner, G. (2013). Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. *Microbial Cell Factories*, 12: 58-58.
Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu Rev Biochem*, 62: 139-60.
Matthews, B. W. (1995). Studies on protein stability with T4 lysozyme. *Adv Protein Chem*, 46: 249-78.
Mayes, H. B., Knott, B. C., Crowley, M. F., Broadbelt, L. J., Stahlberg, J. & Beckham, G. T. (2016). Who's on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolases. *Chemical Science*, 7 (9): 5955-5968.
Naas, A. E., Mackenzie, A. K., Mravec, J., Schuckel, J., Willats, W. G., Eijsink, V. G. & Pope, P. B. (2014). Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio*, 5 (4): e01401-14.
Naumoff, D. G. (2011). Hierarchical classification of glycoside hydrolases. *Biochemistry (Mosc)*, 76 (6): 622-35.
Nielsen, J. E., Borchert, T. V. & Vriend, G. (2001). The determinants of alpha-amylase pH-activity profiles. *Protein Eng*, 14 (7): 505-12.

Oreskes, N. (2004). Beyond the ivory tower. The scientific consensus on climate change. *Science*, 306 (5702): 1686.
Park, C.-s., Kawaguchi, T., Sumitani, J.-i., Takada, G., Izumori, K. & Arai, M. (2005). Cloning and sequencing of an exoglucanase gene from *Streptomyces* sp. M23, and its expression in *Streptomyces lividans* TK-24. *Journal of Bioscience and Bioengineering*, 99 (4): 434-436.
Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Stahlberg, J. & Beckham, G. T. (2015). Fungal cellulases. *Chem Rev*, 115 (3): 1308-448.
Piens, K. & Davies, G. (2015). *Glycoside Hydrolase Family 6*. In Davies, G. (ed.): CAZypedia. Available at: http://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_6.
Poidevin, L., Feliu, J., Doan, A., Berrin, J. G., Bey, M., Coutinho, P. M., Henrissat, B., Record, E. & Heiss-Blanquet, S. (2013). Insights into exo- and endoglucanase activities of family 6 glycoside hydrolases from *Podospora anserina*. *Appl Environ Microbiol*, 79 (14): 4220-9.
Reddy, A. P., Simmons, C. W., D'Haeseleer, P., Khudyakov, J., Burd, H., Hadi, M., Simmons, B. A., Singer, S. W., Thelen, M. P. & Vandergheynst, J. S. (2013). Discovery of microorganisms and enzymes involved in high-solids decomposition of rice straw using metagenomic analyses. *PLoS One*, 8 (10): e77985.
Reetz, M. T. & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc*, 2 (4): 891-903.
Robichon, C., Luo, J., Causey, T. B., Benner, J. S. & Samuelson, J. C. (2011). Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol*, 77 (13): 4634-46.
Rødsrud, G., Lersch, M. & Sjöde, A. (2012). History and future of world's most advanced biorefinery in operation. *Biomass and Bioenergy*, 46: 46-59.
Rosano, G. L. & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5: 172.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. & Jones, T. A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*, 249 (4967): 380.
Rye, C. S. & Withers, S. G. (2000). Glycosidase mechanisms. *Curr Opin Chem Biol*, 4 (5): 573-80.
Sáez-Jiménez, V., Fernández-Fueyo, E., Medrano, F. J., Romero, A., Martínez, A. T. & Ruiz-Dueñas, F. J. (2015). Improving the pH-stability of Versatile Peroxidase by Comparative Structural Analysis with a Naturally-Stable Manganese Peroxidase. *PLoS One*, 10 (10): e0140984.
Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol*, 19 (4): 132-5.
Scheller, H. V. & Ulvskov, P. (2010). Hemicelluloses. *Annu Rev Plant Biol*, 61: 263-89.
Schülein, M. (1997). Enzymatic properties of cellulases from *Humicola insolens*. *J Biotechnol*, 57 (1-3): 71-81.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, 31 (13): 3381-5.
Sharma, S. & Yazdani, S. S. (2016). Chapter 6 - Diversity of Microbial Cellulase System. In Gupta, V. K. (ed.) *New and Future Developments in Microbial Biotechnology and Bioengineering*, pp. 49-64. Amsterdam: Elsevier.

Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (2): 452-456.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7: 539-539.
Sippel, K. H. & Quiocho, F. A. (2015). Ion–dipole interactions and their functions in proteins. *Protein Science : A Publication of the Protein Society*, 24 (7): 1040-1046.
Sjöde, A., Frölander, A., Lersch, M. & Rødsrud, G. (2010). *Lignocellulosic biomass conversion by sulfite pretreatment*. International Patent Number WO/2010/078930.
Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry*, 39 (5): 872-9.
Sukharnikov, L. O., Cantwell, B. J., Podar, M. & Zhulin, I. B. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol*, 29 (10): 473-9.
Sun, D. P., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry*, 30 (29): 7142-53.
Tuohy, M. G., Walsh, D. J., Murray, P. G., Claeyssens, M., Cuffe, M. M., Savage, A. V. & Coughlan, M. P. (2002). Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1596 (2): 366-380.
Tynan-Connolly, B. M. & Nielsen, J. E. (2007). Redesigning protein p*K*a values. *Protein Sci*, 16 (2): 239-49.
Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M. & Eijsink, V. G. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, 330 (6001): 219-22.
van den Burg, B. & Eijsink, V. G. (2002). Selection of mutations for increased protein stability. *Curr Opin Biotechnol*, 13 (4): 333-7.
Várnai, A., Siika-Aho, M. & Viikari, L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnol Biofuels*, 6 (1): 30.
Venditto, I., Najmudin, S., Luis, A. S., Ferreira, L. M., Sakka, K., Knox, J. P., Gilbert, H. J. & Fontes, C. M. (2015). Family 46 Carbohydrate-binding Modules Contribute to the Enzymatic Hydrolysis of Xyloglucan and beta-1,3-1,4-Glucans through Distinct Mechanisms. *J Biol Chem*, 290 (17): 10572-86.
Vuong, T. V. & Wilson, D. B. (2010). Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnol Bioeng*, 107 (2): 195-205.
Withers, S. G. & Williams, S. (2013). *Glycoside Hydrolases*. In Williams, S. (ed.): CAZypedia. Available at: http://www.cazypedia.org/index.php/Glycoside_hydrolases.
Wolfenden, R., Lu, X. & Young, G. (1998). Spontaneous Hydrolysis of Glycosides. *Journal of the American Chemical Society*, 120 (27): 6814-6815.
Wolfgang, D. E. & Wilson, D. B. (1999). Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2. *Biochemistry*, 38 (30): 9746-51.
Xiao, S., Patsalo, V., Shan, B., Bi, Y., Green, D. F. & Raleigh, D. P. (2013). Rational modification of protein stability by targeting surface sites leads to complicated results. *Proc Natl Acad Sci U S A*, 110 (28): 11337-42.

Xu, X., Li, J., Zhang, W., Huang, H., Shi, P., Luo, H., Liu, B., Zhang, Y., Zhang, Z., Fan, Y., et al. (2015). A Neutral Thermostable β-1,4-Glucanase from *Humicola insolens* Y1 with Potential for Applications in Various Industries. *PLoS ONE*, 10 (4): e0124925.

Yang, H., Li, J., Shin, H. D., Du, G., Liu, L. & Chen, J. (2014). Molecular engineering of industrial enzymes: recent advances and future prospects. *Appl Microbiol Biotechnol*, 98 (1): 23-9.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*, 40 (Web Server issue): W445-51.

Yip, K. S., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur J Biochem*, 255 (2): 336-46.

Zhang, J. & Madden, T. L. (1997). PowerBLAST: A New Network BLAST Application for Interactive or Automated Sequence Analysis and Annotation. *Genome Research*, 7 (6): 649-656.

Zhang, S., Wolfgang, D. E. & Wilson, D. B. (1999). Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol Bioeng*, 66 (1): 35-41.

Zhang, S., Barr, B. K. & Wilson, D. B. (2000). Effects of noncatalytic residue mutations on substrate specificity and ligand binding of *Thermobifida fusca* endocellulase Cel6A. *Eur J Biochem*, 267 (1): 244-52.

7. Appendices

7.1. Appendix A

The pNIC-CH cloning and expression vector was used to produce all proteins generated throughout the course of this study. For more information about the pNIC-CH plasmid, see Section 2.3.2. ***Figure 7.1.1*** and ***Figure 7.1.2*** show the vector map of pNIC-CH and its cloning region, respectively.

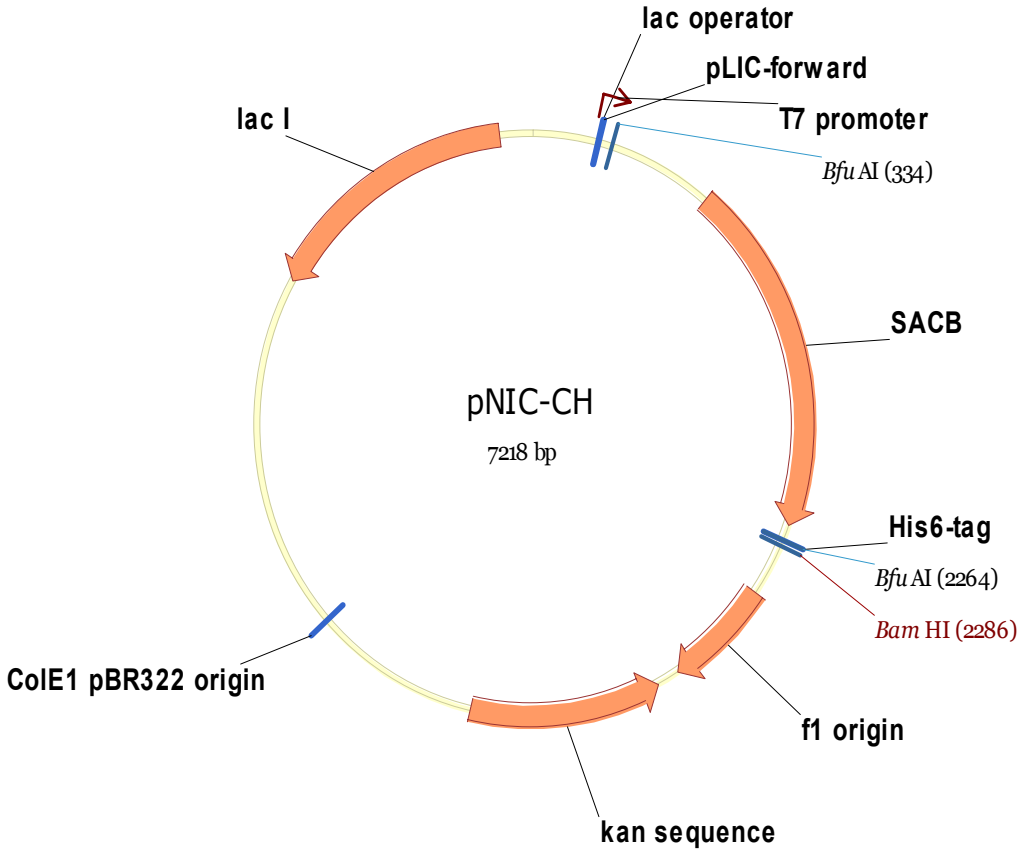


Figure 7.1.1. Vector map of pNIC-CH cloning and expression plasmid. The figure shows an overview of the components of the pNIC-CH vector. The T7 promoter is utilized to transcribe the plasmid genes using the bacterial host T7 RNA Polymerase. The *sacb* gene (SACB) allows for negative selection using sucrose, as described in Section 3.7. Restriction sites flanking the *sacb* gene (shown as *Bfu AI*) facilitate its removal from the vector using BfuAI restriction enzymes. The *aph* gene encoding resistance to kanamycin (Section 2.13) is indicated as kan sequence. The C-terminal 6xHis-tag is shown immediately downstream of the *sacb* gene. Figure obtained from Opher Gileadi Lab via Addgene (Addgene plasmid 26117).

Cloning region in the vector:

5' end:

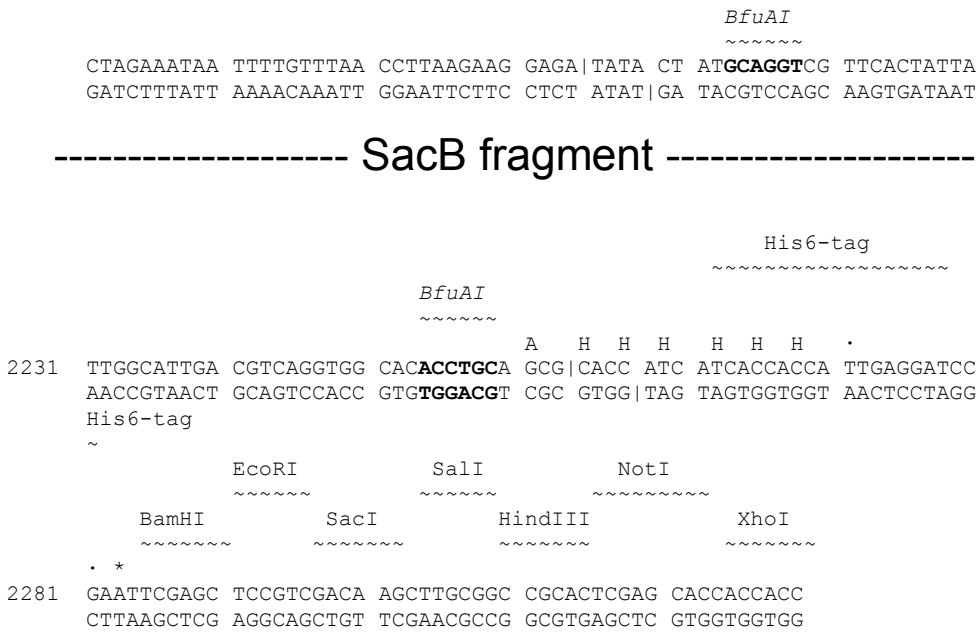


Figure 7.1.2. Sequence of the Cloning Sites of pNIC-CH. The figure shows the cloning region in the pNIC-CH plasmid, and the location of the C-terminal His-tag relative to the *sacb* gene. During digestion of the pNIC-CH plasmid with BfuAI restriction enzymes, the *sacb* gene is removed and overhangs are generated, enabling insertion of a target gene of interest with complementary overhangs. Figure obtained from Opher Gileadi Lab via Addgene (Addgene plasmid 26117).

7.2. Appendix B

Table 7.2.1 shows the sequences of primers used during this study. For a description of the primers, see ***Table 2.4.1***.

Table 7.2.1. Primers by name and sequence. For a description of primers, see **Table 2.4.1**.

Primer	Sequence (5'→3')
mgCel6A_F LIC	TAAGAAGGAGATATACTATGGCAGATAGCG CATTTTATGTTGAT
mgCel6A_R LIC	AATGGTGGTGATGATGGTGCGCGCTGGTAC ATGCACTACCATTCAG
mgCel6AΔCBM_R LIC	AATGGTGGTGATGATGGTGCGCTGCTGCAA TTGCCAGTTCATAT
M1_F	CAACCCAACGTCGGCTAGGCCCACTAGGC
M1_R	GTTGGGTTGCAGCCGATCCGGGTGATCCG
M2_F	CAGCAACCAAATGGGTCCTATTAGGCCCAT GTCAAG
M2_R	GTCGTTGGTTTACCCAGGATAATCCGGGTA CAGTTC
M3_F	CAGACTTTCGTCCATCGTCGCTTCGTTTTC AAATAAAACTACGTC
M3_R	GTCTGAAAGCAGGTAGCAGCGAAGCAAAAG TTTATTTTGATGCAG
M4_F	CATAACGTGGCCGTCCACTTAAACAAGGCG TCGCAC
M4_R	GTATTGCACCGGCAGGTGAATTTGTTCCGC AGCGTG
M5_F	CTAACTACTTCAACGTCTTCCAGACCGTCC AGCAGG
M5_R	GATTGATGAAGTTGCAGAAGGTCTGGCAGG TCGTCC

M6_F	GTCCAGTATCGCGTACCCTTCGTGGCTCGC TTTACCG
M6_R	CAGGTCATAGCGCATGGGAAGCACCGAGCG AAATGGC
M7_F	CACTAGGCGGACCAGCACGTCTTCCGTGGG GCTC
M7_R	GTGATCCGCCTGGTCGTGCAGAAGGCACCC CGAG
M8_F	GGACCACTTCGTCTACCAACACTACGTGGC CGTC
M8_R	CCTGGTGAAGCAGATGGTTGTGATGCACCG GCAG
M10_F	GACCAGAATAATCCGGGTACAGTTCGT
M10_R	GAACTGTACCCGGATTATTCTGGTCAAACC AACGACCCTGCGGAAC
M11_F	GCTGCAGGTCATAGCGCAT
M11_R	CCAGCCATGCGCTATGACCTGCAGCAAAAT AAACTTTTGCCTGGC
M12_F	AATGCAGGTCATAGCGCATGGC
M12_R	CCAGCCATGCGCTATGACCTGCATTAAAAT AAACTTTTGCCTGGCT
pLIC-for	TGTGAGCGGATAACAATTCC
pLIC-rev	AGCAGCCAACTCAGCTTCC

7.3. Appendix C

Figure 7.3.1 and ***Figure 7.3.2*** show the DNA sequences of the full-length *mgcel6a* and truncated *mgcel6a*Δ*cbm* genes, codon-optimized for expression by *E. coli*.

ATGCGTCTGCGTAGCGCACTGGCAGCAGCCGGTGCAGTTGCAGCAACCGCATTTGCAGCCACAGCGGGTGCCGTTGG
TCTGCTGCATCCGACCGCAGCACGTGCAGCAGATAGCGCATTTTATGTTGATCCGGATACCGGTGCAGCACGTTGGG
TTGCAGCCAATCCGGGTGATCCGCGTGCAGCCGTTATTCGTGATCGTATTGCAAGCGTTCCGCAGGGTCGTTGGTTT
ACCCAGAATAATCCGGGTACAGTTCGTGGTCAGGTTGATGCATTTGTTGGTGCAGCCGCAGCAGCAGGTAAAATTCC
GATTCTGGTTGTTTATAACATTCCGAATCGTGATTGTAGCGGTGCAAGCAGCGGTGGTATGCCGAATCATACCGCAT
ATCGTCAGTGGATTGATGAAGTTGCAGCCGGTCTGGCAGGTCGTCCGGCAGCAATTATCCTGGAACCGGATGTTCTG
GCACTGATGACCAGCTGTATGAATGAAAGCCAGCAGGCAGAAACCCGTGCAAGCATGGCATATGCAGGTAAACGTCT
GAAAGCAGGTAGCAGCCAGGCAAAAGTTTATTTTGATGCAGGTCATAGCGCATGGCTGGCACCGAGCGAAATGGCAG
CCCGTCTGGTTGCCGCAGATATTGCAAATAGCGCAGATGGTATTAGCGTGAATGTGAGTAATTATCGTACCACCGCA
GAAAGCGTTAATTATGCAAAAGCAGTTGTTGCCGCAACAGGTGCACCGCATCTGAAAATTGTTGTTGATACCAGCCG
TAATGGTAATGGTCCGGCACCGGATAGCGAATGGTGTGATCCGCCTGGTCGTGCAATTGGCACCCCGAGCACCACCG
ATACAGGTGATCCGGCAGTTGATGCCTTTCTGTGGATTAAACTGCCTGGTGAAGCAGATGGTTGTATTGCACCGGCA
GGTCAGTTTGTTCCGCAGCGTGCATATGAACTGGCAATTGCAGCACCGACCCCTCCGACCACACCGCCTACCACCCC
ACCGACAACGCCACCGACTACTCCTCCGACAACCCCTCCTACGACACCTCCAACAACTCCGCCAACCACACCACCAA
CTACCCCTCCAAATGGTGCATGTCAGGTTAGCTATCAGACCAATCAGTGGCCTGGTGGTTTTACAGCAGATATTCGT
ATTACCAATAATGGCCCTGCACTGAATGGTTGGACCCTGAGCTTTACCGTTGGTAGCGGTGTTGCACTGAGCAATGG
TTGGTCAGGTATTTGGAGCCAGAGCGGCACCCGTATTACCGTTAGCAATGAAGCATGGAATGGTAGCCTGCCGACCG
GTGGCACCGTTACCATTGGTTATCAGGGCACCTATAGCGGTAGTCTGCCTGCACCGACAGATTTTACCCTGAATGGT
AGTGCATGTACCAGC

Figure 7.3.1. DNA Sequence of *mgcel6a*. The figure shows the DNA sequence of the full-length *mgcel6a* gene, codon-optimized for expression by *E. coli*. The signal peptide (105 bp) is shown in red, the catalytic GH6 domain (864 bp) is shown in blue; the linker region (120 bp) is shown in black, and the CBM2 domain (312 bp) is shown in green.

ATGCGTCTGCGTAGCGCACTGGCAGCAGCCGGTGCAGTTGCAGCAACCGCATTTGCAGCCACAGCGGGTGCCGTT
GGTCTGCTGCATCCGACCGCAGCACGTGCAGCAGATAGCGCATTTTATGTTGATCCGGATACCGGTGCAGCACGT
TGGGTTGCAGCCAATCCGGGTGATCCGCGTGCAGCCGTTATTCGTGATCGTATTGCAAGCGTTCCGCAGGGTCGT
TGGTTTACCCAGAATAATCCGGGTACAGTTCGTGGTCAGGTTGATGCATTTGTTGGTGCAGCCGCAGCAGCAGGT
AAAATTCCGATTCTGGTTGTTTATAACATTCCGAATCGTGATTGTAGCGGTGCAAGCAGCGGTGGTATGCCGAAT
CATACCGCATATCGTCAGTGGATTGATGAAGTTGCAGCCGGTCTGGCAGGTCGTCCGGCAGCAATTATCCTGGAA
CCGGATGTTCTGGCACTGATGACCAGCTGTATGAATGAAAGCCAGCAGGCAGAAACCCGTGCAAGCATGGCATAT
GCAGGTAAACGTCTGAAAGCAGGTAGCAGCCAGGCAAAAGTTTATTTTGATGCAGGTCATAGCGCATGGCTGGCA
CCGAGCGAAATGGCAGCCCGTCTGGTTGCCGCAGATATTGCAAATAGCGCAGATGGTATTAGCGTGAATGTGAGT
AATTATCGTACCACCGCAGAAAGCGTTAATTATGCAAAAGCAGTTGTTGCCGCAACAGGTGCACCGCATCTGAAA
ATTGTTGTTGATACCAGCCGTAATGGTAATGGTCCGGCACCGGATAGCGAATGGTGTGATCCGCCTGGTCGTGCA
ATTGGCACCCCGAGCACCACCGATACAGGTGATCCGGCAGTTGATGCCTTTCTGTGGATTAAACTGCCTGGTGAA
GCAGATGGTTGTATTGCACCGGCAGGTCAGTTTGTTCCGCAGCGTGCATATGAACTGGCAATTGCAGCA

Figure 7.3.2. DNA Sequence of *mgcel6a*Δ*cbm*. The figure shows the DNA sequence of the *mgcel6a*Δ*cbm* gene, codon-optimized for expression by *E. coli*. The signal peptide (105 bp) is shown in red, and the catalytic GH6 domain (864 bp) is shown in blue.

7.4. Appendix D

The amino acid sequences of the full-length mgCel6A enzyme and its catalytic domain, mgCel6AΔCBM, are shown in ***Figure 7.4.1*** and ***Figure 7.4.2***, respectively.

MRLRSALAAAGAVAATAFAATAGAVGLLHPTAARAADSAFYVDPDTGAARWVAANPGDP
RAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVGAAAAAGKIPILVVYNIPNRDCSGA
SSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVLALMTSCMNESQQAETRASMAYAG
KRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIANSADGISVNVSNYRTTAESVNY
AKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGRAIGTPSTTDTGDPAVDAFLWI
KLPGEADGCIAPAGQFVPQRAYELAIAAPTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP
PTTPPTTPPNGACQVSYQTNQWPGGFTADIRITNNGPALNGWTLSFTVGSGVALSNGWS
GIWSQSGTRITVSNEAWNGSLPTGGTVTIGYQGTYSGSLPAPTDFTLNGSACTS

Figure 7.4.1. Amino Acid Sequence of mgCel6A. The figure shows the amino acid sequence of the full-length mgCel6A enzyme. The signal peptide (35 amino acids) is shown in red, followed by the catalytic GH6 domain comprising 288 amino acids, which is shown in blue. The linker region (40 amino acids) which connects the catalytic domain to the CBM2 domain is shown in black, and the 104-amino-acid-long CBM2 domain is shown in green.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.4.2. Amino Acid Sequence of mgCel6AΔCBM. The figure shows the amino acid sequence of the truncated mgCel6A enzyme. The amino acid coding for the start codon (Met) is shown in red. The catalytic domain of the mgCel6A enzyme is shown in black (288 amino acids), and the C-terminal 6xHis-tag, as well as the single residue that separates the His-tag from the catalytic domain (Ala), are shown in blue.

7.5. Appendix E

The amino acid sequences of all mutants designed during the course of this study can be seen in the figures below. The amino acid coding for the start codon (Met) is shown at the beginning of each sequence, and the C-terminal 6xHis-tag, as well as the single residue that separates the catalytic domain from the His-tag (Ala), are shown at the end of each sequence.

MADSAFYVDPDTGAARWVAA**D**PGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.1. Amino Acid Sequence of Mutant M1. The inserted mutation N21D is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQ**D**NPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.2. Amino Acid Sequence of Mutant M2. The inserted mutation N46D is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSS**E**AKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.3. Amino Acid Sequence of Mutant M3. The inserted mutation Q152E is shown in red. This mutant was successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAG**E**FVPQRAYELAIAAAHHHHHH

Figure 7.5.4 Amino Acid Sequence of Mutant M4. The inserted mutation Q276E is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAEGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.5. Amino Acid Sequence of Mutant M5. The inserted mutation A104E is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWEAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.6. Amino Acid Sequence of Mutant M6. The inserted mutation L165E is shown in red. This mutant was successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AEGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.7. Amino Acid Sequence of Mutant M7. The inserted mutation I242E is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCDAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.8. Amino Acid Sequence of Mutant M8. The inserted mutation I271D is shown in red. This mutant was not successfully produced.

MADSAFYVDPDTGAARWVAADPGDPRAAVIRDRIASVPQGRWFTQDNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAEGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSEAKVYFDAGHSAWEAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AEGTPSTTDTGDPAVDAFLWIKLPGEADGCDAPAGEFVPQRAYELAIAAAHHHHHH

Figure 7.5.9. Amino Acid Sequence of Mutant M9. The inserted mutations N21D, N46D, A104E, Q152E, L165E, I242E, I271D, and Q276E are shown in red. This mutant was successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFDQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFDAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.10. Amino Acid Sequence of Mutant M10. The inserted mutation T44D is shown in red. This mutant was successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFAAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.11. Amino Acid Sequence of Mutant M11. The inserted mutation D158A is shown in red. This mutant was successfully produced.

MADSAFYVDPDTGAARWVAANPGDPRAAVIRDRIASVPQGRWFTQNNPGTVRGQVDAFVG
AAAAAGKIPILVVYNIPNRDCSGASSGGMPNHTAYRQWIDEVAAGLAGRPAAIILEPDVL
ALMTSCMNESQQAETRASMAYAGKRLKAGSSQAKVYFNAGHSAWLAPSEMAARLVAADIA
NSADGISVNVSNYRTTAESVNYAKAVVAATGAPHLKIVVDTSRNGNGPAPDSEWCDPPGR
AIGTPSTTDTGDPAVDAFLWIKLPGEADGCIAPAGQFVPQRAYELAIAAAHHHHHH

Figure 7.5.12. Amino Acid Sequence of Mutant M12. The inserted mutation D158N is shown in red. This mutant was successfully produced.

7.6. Appendix F

Figure 7.6.1 and ***Figure 7.6.2*** show relevant model data resulting from the generation of the mgCel6AΔCBM model by SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008). The model was built using the catalytic domain of *Tf*Cel6A (PDB ID 1TML) (Berman et al. 2000) as a template. The alignment of the two sequences resulted in 70% sequence identity and is shown in ***Figure 4.1.2***.

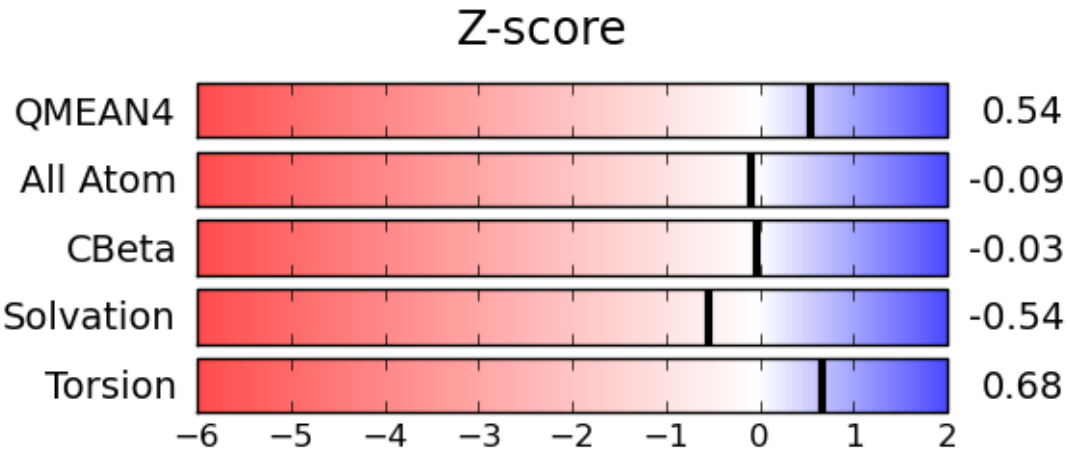


Figure 7.6.1. Normalized QMEAN Score (Z-score) of the model of mgCel6AΔCBM generated using SWISS-MODEL. The QMEAN score is calculated based on the four types of analyses shown: all-atom and C-beta interactions, the solvation potential, which indicates the burial status of all residues, and the torsion angle potential, which analyzes the local geometry of the model over three consecutive amino acids (Benkert et al. 2011). Figure obtained from SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008).

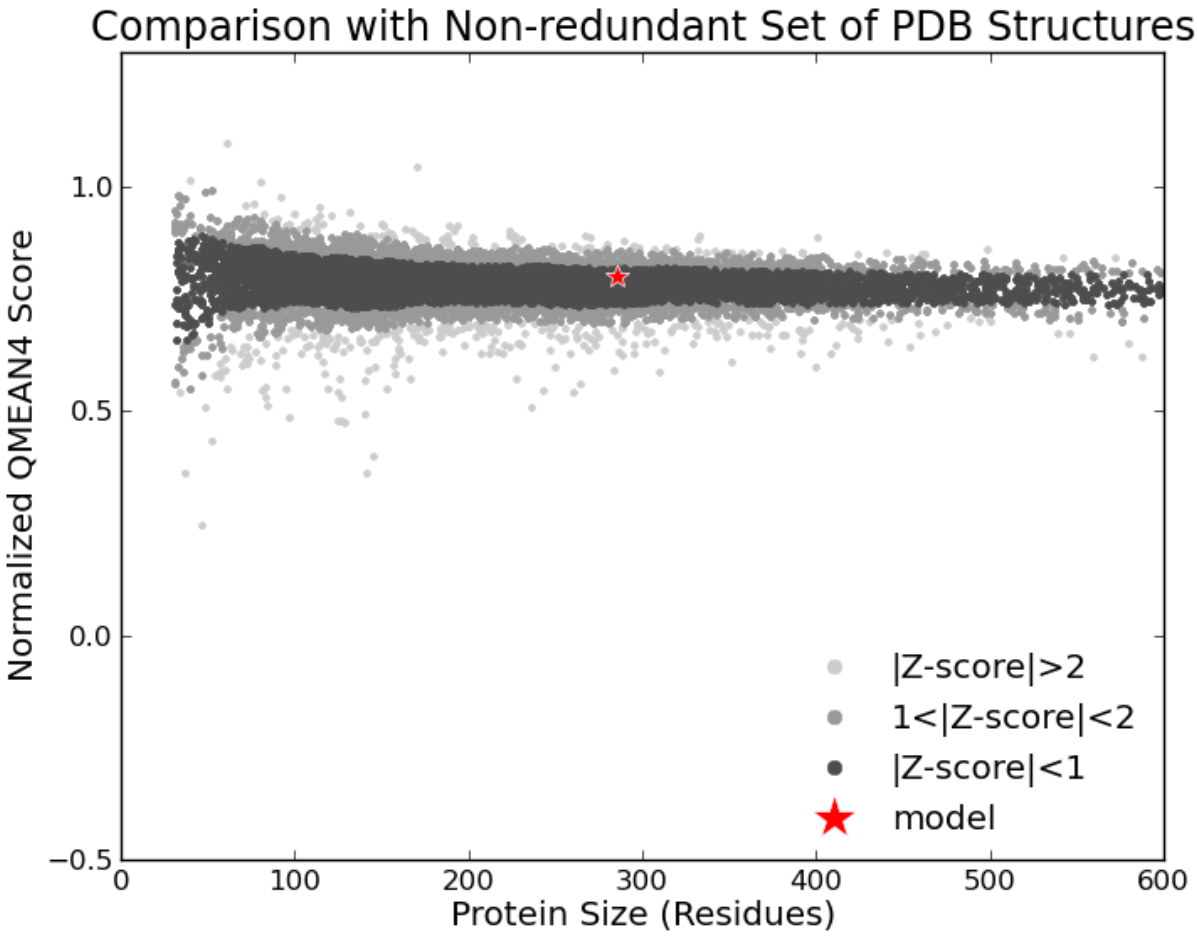
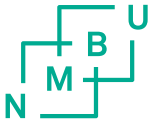


Figure 7.6.2. Comparison of the model of mgCel6AΔCBM generated using SWISS-MODEL with non-redundant PDB structures. The figure shows the normalized QMEAN score (also called Z-score) of the model as a red star. The QMEAN score describes the quality of the model, and is a composite score calculated from analyses of the local geometry, all-atom and C-beta interactions, and the burial statues of all residues in the model (as seen in **Figure 7.6.1**). When this score is related to those expected from proteins of similar sizes with high-resolution solved structures, it is termed a Z-score (Benkert et al. 2011). Figure obtained from SWISS-MODEL (Arnold et al. 2006; Biasini et al. 2014; Bordoli et al. 2008).



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway