

Recombinant expression and stability engineering of lytic polysaccharide monooxygenases

ir. Magali Tanghe

Members of the jury:

Prof. dr. Tom Desmet (Ghent University) (promotor)
Dr. Ingeborg Stals (Ghent University) (promotor)
Prof. dr. ir. Stefaan De Smet (Ghent University) (chairman)
Prof. dr. ir. Yves Briers (Ghent University) (secretary)
Prof. dr. ir. Tina Kyndt (Ghent University)
Prof. dr. Dietmar Haltrich (BOKU University of Natural Resources and Life Sciences, Vienna (Austria))
Prof. dr. Nico Callewaert (VIB - Ghent University)

Promotors:

Prof. Dr. Tom Desmet (Ghent University)
Center for Synthetic Biology
Department of Biochemical and Microbial Technology
Faculty of Bioscience Engineering, Ghent University
Coupure links 653, 9000 Ghent, Belgium

Dr. Ingeborg Stals (Ghent University)
Industrial Catalysis and Adsorption Technology (INCAT)
Faculty of Engineering and Architecture, Ghent University
Valentin Vaerwyckweg 1, 9000 Ghent, Belgium

Dean:

Prof. dr. ir. Marc Van Meirvenne

Rector:

Prof. dr. Anne De Paepe

Magali Tanghe was supported by the IWT-Vlaanderen (Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen, grantagreement nr. 121629, 2013-2016).

Recombinant expression and stability engineering of lytic polysaccharide monooxygenases

Ir. Magali Tanghe

Thesis submitted in fulfillment of the requirements for
the degree of Doctor (PhD) in Applied Biological Sciences:
Cell and Gene Biotechnology
Academic year 2016-2017

Dutch translation of the title:

Recombinante expressie en stabiliteitsengineering van lytische polysaccharide monooxygenasen.

To refer to this thesis:

Tanghe, M. (2017) Recombinant expression and stability engineering of lytic polysaccharide monooxygenases. PhD thesis, Faculty of Bioscience Engineering, Ghent University, Ghent.

Cover illustration:

Design: Magali Tanghe

Trichoderma reesei Cel61A on cellulose substrate.

Printed by University Press, Zelzate.

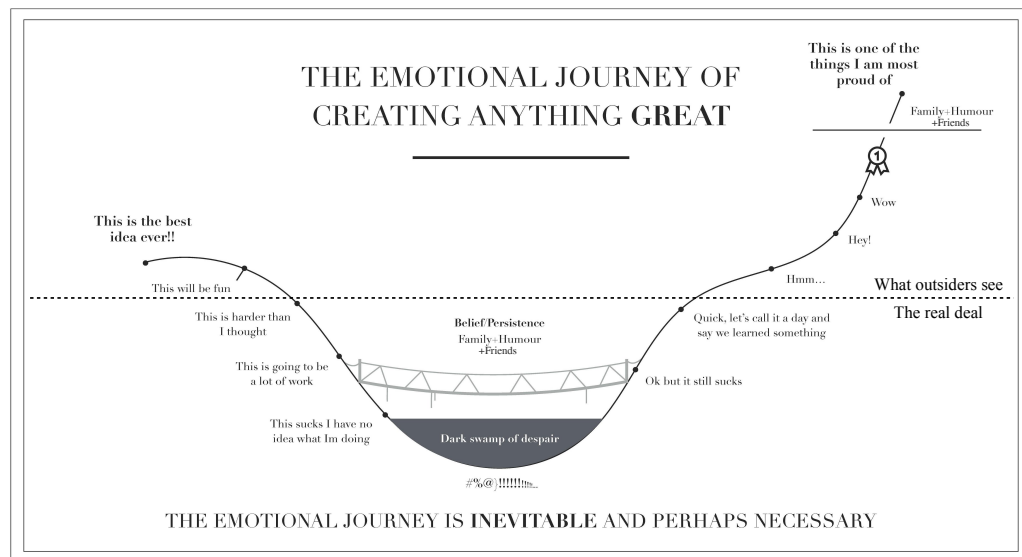
ISBN number: 978-90-5989-977-3

© 2017 by Magali Tanghe. All rights reserved.

The author and the promoters give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

Dankwoord

Een goeie 4 jaar geleden wilde ik heel graag een doctoraat beginnen. Wetenschap vind ik geweldig interessant en ik wilde dus ook mijn steentje bijdragen. Intussen zijn die 4 jaren werkelijk voorbij gevlogen en alle emoties zijn de revue gepasseerd, van pure wanhoop tot opperste extase in een klein eureka-moment. Onlangs vond ik onderstaande figuur 'The emotional journey of creating something great'. Zeer goede vergelijking voor de vele uitdagingen die een mens in zijn leven aangaat, en zonder meer een perfecte weerspiegeling van *Project PhD*. In een doctoraat is het doel om iets nieuws te creëren of te onderzoeken. Dus, per definitie doe je dingen die niet eerder beschreven zijn. Het moeilijke hieraan is dat er zeer weinig of zelfs geen referentiepunten zijn om af te toetsen of je goed bezig bent. Dat brengt heel wat lastige momenten met zich mee en volharding is de enige uitweg. Iedereen maakt de 'Ik zou liever aan de kassa in de Delhaize zitten'-periode door. Maar, het goede nieuws is dat je gaandeweg ook door het bos de bomen weer begint te zien en de voldoening van een afgewerkt project: Ge-wel-dig! Het ware nieuws is, collega's, vrienden en familie vormen de entourage die je nodig hebt en ik ben bijzonder dankbaar voor alle begrip en steun door deze eenzame periode! De grote meerderheid heeft enkel 'peis en vree' gezien, maar jullie weten wel beter. Daarom wil ik nu ook even tijd nemen om een aantal mensen in het bijzonder in de bloemetjes te zetten.



Figuur 1: Realistische voorstelling van een doctoraat.

Eerst en vooral wil ik de leden van de jury bedanken om mijn werk grondig te lezen en nuttige vragen en kritische opmerkingen te geven. Ik ben ervan overtuigd dat dit de laatste *push* was, nodig om dit werk tot een goed einde te brengen. Bedankt professor Stefaan De Smet, professor Yves Briers, professor Tina Kyndt en professor Nico Callewaert. Special thanks to professor Dietmar Haltrich, to make the trip from Austria.

Zonder promotoren, geen doctoraat, zoveel is duidelijk. Daarom wil ik zeker Tom en Ingeborg bedanken. Ik stel het ten zeerste op prijs dat ik de kans heb gekregen onderzoek te doen in een uitstekend uitgerust labo en deel mocht zijn van de Glycodirect groep. Daarnaast mocht ik ook enkele maanden verblijven aan de Universiteit in Graz en heb ik verschillende internationale congressen kunnen bijwonen. Tom, ik weet nog goed hoe ik kwam aankloppen om te 'solliciteren' voor een positie als doctoraatsstudent. Daarna volgde een intensieve voorbereiding voor de ontmoeting met hét IWT, die meteen zijn vruchten heeft afgeworpen. Op 1 januari 2013 kreeg ik dan ook echt mijn plaatsje in de Glycodirect bureau. Toen kon ik nog kiezen en zelfs meerdere plaatsen tegelijk bezetten, intussen puilt dat lokaaltje uit. Glycodirect is een toffe, dynamische, groeiende onderzoeksgroep, iets om trots op te zijn!

Glycodirect is voor mij altijd een hechte *familie* geweest. Er kan zeer hard gewerkt worden, maar af en toe kunnen alle remmen ook eens los. Blijf dat zeker doen! Allereerst wil ik mijn grote voorbeelden bedanken. Mareike, Tommeke and Karel, aka the guidance committee during the deepest point of the *dark swamp of despair*. Without your help, trust and encouragement, I don't know if this little booklet would be here today. I cannot describe how lucky I am to have known you guys, not only for science but also and even more as a person. Mareike, I don't think you always realize, but we still miss you (and your sugar addiction). Tommeke, jij krijgt iedereen aan het lachen met je (flauwe) mopjes of gewoon door je aanstekelijke lach! Karel, jij bent zonder enige twijfel de meest gedreven persoon die ik ken en jij staat altijd voor iedereen klaar. Niet alleen met raad en daad in het labo, maar ook je overweldigende vrolijkheid, en spontane organisatie van De Winter-BBQ hebben glycodirect een paar hoogtepunten bezorgd.

Barbara, lang leve het LPMO-team! Jij weet als geen ander hoe frustrerend het LPMO-onderzoek kan zijn. Je hebt enorm veel volharding en motivatie, ik weet zeker dat je het tot een goed einde zal brengen. Ik wens je echt bijzonder veel succes toe! Renfei, where the f*** are you? Hope life is treating you well... Margo en Griet, jullie zijn 2 *very precious ladies* voor mij. Jullie zijn voor mij al lang meer vriendinnen dan (ex-) collega's. Ik hoop dat we elkaar nog veel gaan zien in de toekomst. Trouwens Margo, dankzij jou wok ik nu in arachide-olie. En Griet, ookal heb je een ander pad ingeslagen: ROCK ON! Jorick, gij aggressive zot! Ceasar'ke? Stevie-ie-ie, er is een probleem met de dionex... ;-) Koen, hoe was dat bij u in Brno? Zorica, we didn't have much time to meet, but for the time we had, you are a lovely lady! En Ophelia, de toekomst van de groep! Het belangrijkste is volharding en ... de weg naar de koepuur...

Mijn thesisstudenten, Lotje, Marijn en Matthias, verdienen ook een speciaal woordje van dank. Na een heel jaar intens met jullie samen te werken, kan ik zeggen dat jullie stuk voor stuk erg gemotiveerde en aangename mensen zijn. Bedankt voor de inzet en het mooie werk. Ik heb graag met jullie samengewerkt.

Glycodirect is uiteraard maar een klein deeltje van het CSB/Inbio team. De sfeer op de tweede verdieping, rechterkant was altijd dik in orde! Het jaarlijkse kerstfeestje (bedankt

Feestcomité) zat steeds steengoed in elkaar, de BBQ was altijd even gezellig, massale deelname aan de Sportdag, de jaarlijkse teambuidling en de spontane uitjes doen je soms even alle zorgen vergeten. Vanaf het eerste moment hebben jullie mij met open armen ontvangen en ik voelde mij meteen thuis in de groep. Herinneringen als Eric die jodelend door de gangen loopt, Frederik 'Klinken is drinken' of DJ Wouter aan de jukebox zullen mij nog lang bijblijven. Pieter, houdt Gert je weer op in de Sportsbar of was het omgekeerd? Robin en Gert, IWT 2013 is een TOP-jaar! Molleke, is dat niet koud aan je benen? Double-D!! Bob, Ik ben zooo blij dat je dan toch de Atlantische oceaan hebt overgestoken. Thomas, ik kom supporteren voor je eerste Triathlon! Delforche, het onoverwinnelijke sporticoon. Soph/fie x3, het ga je goed met practica begeleiden, de algemene vrede bewaren en gisten temmen. Trouwens, bedankt voor de kick-start in gistexpertise! Lisa, heb je nog van die lekkere soep?! Marilyn, immer vrolijke collega! Isabelle VDV, ik mis je... Sylwia, keep up the good work, you will tame that damn *Pichia*! Dries Van Herpe, succes met je welverdiende Baekeland-mandaat! Van Brempt, heb je al een huis gekocht? Tom Delmulle, even vlot in 3-pointers scoren als *Saccharomyces* manipuleren. Nico, goed bezig man! David, ik kom naar de CD-release! Brecht, nog vele jaren vrolijk jammen met Gilles. Gilles, de vrolijke labmanager, maakt al je bestellingen binnen de kortste keren waar. Dominique, Barbara, Anneleen en Isabelle Maryns, bedankt voor de fijne koffieruimte lunches. Stijn en Lien, het ga je goed op jullie nieuwe jobs! En uiteraard de nieuwe lichting, keep calm and keep up the good work/atmosphere!!

Een gezonde geest in een gezond lichaam, of hoe zeggen ze dat ook weer. Bedankt Fietsteam om mij te introduceren in deze toch wel bijzonder leuke hobby. Dankzij jullie weet ik wat echt fietsen is. Soms net iets te uitdagend voor mij, maar ik heb toch steeds geprobeerd om mijn mannetje te staan. Nog eens naar Brugge fietsen? :)

Ik mag ook uiteraard mijn geweldige familie en vrienden niet vergeten. Mama en papa, bedankt om mij de kans te geven om te studeren en steeds achter mij te staan. Bedankt broer en zussen om zo goed voor mij te zorgen tijdens mijn werk-verblijven de laatste maanden. Een bijzondere dank aan Johannes om bij elke ontmoeting uitvoerig te vragen naar de status. Ook bedankt aan Francine, bomma, Jurgen en Ischa! Bedankt voor het begrip, medeleven en ondersteuning. Vrienden, we kennen elkaar al vele jaren en ik hoop dat er nog vele jaren bijkomen. (En, sorry als ik jullie vaak met hetzelfde excuus moest afwijzen...). Bijzondere dank aan Team Gent: Timo, Marjo en Robin, jullie hebben het goede voorbeeld gegeven. Caroline, veel succes! Yes you can!

En dan heb je nog Gert... Ik mocht hem niet vernoemen, maar kijk ik doe het toch! Jij bent het mooiste resultaat van mijn PhD! Dankjewel, om werkelijk elke avond al mijn verhalen en af en toe frustraties te aanhoren, om hier en daar iets na te lezen, om niet (heel) boos te worden als ik zei 'over een halfuurtje ben ik ZEKER thuis' en 2 uur later nog experimenten stond op te zetten in het labo. Bedankt om er te zijn. Bedankt om voor afleiding te zorgen. De laatste maanden waren op zijn minst *challenging*, maar ik ben ook zeer trots op wat jij allemaal bereikt hebt de laatste jaren. Ik kijk ernaar uit om samen met jou de rest van mijn leven te delen!

Gent, Maart 2017
Magali Tanghe

Contents

Outline	1
1 Literature review	5
1.1 The promise of lignocellulosic biomass as renewable resource	6
1.1.1 Renewable resources	6
1.1.2 Structure and challenges of lignocellulose	8
1.1.3 Biorefinery concept	10
1.1.4 Lignocellulosic ethanol: Biochemical versus thermochemical conversion	11
1.2 Enzymes for cellulose hydrolysis	13
1.2.1 Cellulase classification	13
1.2.2 Major actors in the synergistic machinery	14
1.2.3 Microbial sources of cellulase systems	16
1.2.4 Non-hydrolytic factors aiding cellulose hydrolysis	17
1.3 Lytic Polysaccharide MonoOxygenases	19
1.3.1 Discovery and classification of LPMOs	19
1.3.2 Synergy with classical cellulases	21
1.3.3 Protein Structure	23
1.3.4 The oxidative reaction	26
1.3.5 Diverse functions of LPMOs	28
1.4 Biochemical lignocellulose biorefining and LPMO contribution	30
1.4.1 Fermentation strategies for production of biofuels	30
1.4.2 Rethinking the process for LPMO integration	31
1.5 Protein Engineering	34
1.5.1 General principles	34
1.5.2 Directed evolution	36
1.5.3 Rational design	37
1.5.4 Semi-rational design	37
1.6 Protein thermostability	40
1.6.1 The importance of protein stability	40
1.6.2 Thermodynamic versus kinetic stability	40
1.6.3 Increasing protein stability	41
1.7 Lytic polysaccharide monooxygenases and stability	44
1.8 Conclusion	46
2 Expression of TrCel61A in <i>Pichia pastoris</i>	47
Abstract	48
2.1 Introduction	49
2.2 Results and discussion	50
2.2.1 Optimizing the yield of secreted protein	50

2.2.2	Optimizing N-terminal processing	51
2.2.3	Scale-up to fermentation and demonstration of LPMO activity	53
2.2.4	Proof of concept: Expression of <i>Phanerochaete chrysosporium</i> GH61D with native secretion signal	55
2.3	Conclusion	56
2.4	Materials and methods	57
2.4.1	Biological materials	57
2.4.2	Molecular work for <i>TrCel61A</i>	57
2.4.3	Molecular work for <i>PcGH61D</i>	58
2.4.4	Media and growth methods	58
2.4.5	Fermentation and purification	60
2.4.6	Protein technology	60
2.4.7	Protein deglycosylation	61
2.4.8	Activity testing	61
2.4.9	Protein Sequencing	62
2.4.10	Statistical analysis	62
	Supplementary material	63
3	Evaluation of measurement methods	69
	Abstract	70
3.1	Introduction	71
3.2	Results and discussion	73
3.2.1	Differential scanning fluorimetry for measuring the apparent melting temperature	73
3.2.2	HPAEC for analysis of reaction products formed from PASC	75
3.2.3	Proof of concept: stability and activity of a destabilizing mutant	80
3.3	Conclusion	82
3.4	Materials and methods	83
3.4.1	Molecular techniques for creating destabilizing variants	83
3.4.2	Differential scanning fluorimetry	83
3.4.3	Production of <i>MfCDHIIB</i>	84
3.4.4	Activity analysis	84
	Supplementary material	85
4	Expression of LPMO10C from <i>Streptomyces coelicolor</i> and its measurement methods	87
	Abstract	88
4.1	Introduction	89
4.2	Results and discussion	91
4.2.1	Evaluation of a suitable expression system: N-terminal processing	91
4.2.2	Optimization of the purification protocol	95
4.2.3	Validation of assays	96
4.3	Conclusion	99
4.4	Materials and methods	100
4.4.1	Cloning of <i>ScLPMO10C</i>	100
4.4.2	Expression and enzyme extraction	101
4.4.3	Ni-NTA chromatography	102
4.4.4	Enzyme assays	102
	Supplementary material	104

5	Thermostability engineering of <i>Trichoderma reesei</i> Cel61A	107
	Abstract	108
5.1	Introduction	109
5.2	Results and discussion	111
5.2.1	General stability exploration of <i>Tr</i> Cel61A	111
5.2.2	Removal of the carbohydrate binding module	117
5.2.3	Stabilization via N-glycosylation	122
5.2.4	Stabilization via disulfide engineering	124
5.2.5	Stabilization via consensus engineering	128
5.3	Conclusion	136
5.4	Materials and methods	137
5.4.1	Molecular work	137
5.4.2	Enzyme thermostability	141
5.4.3	Enzyme activity	141
5.4.4	Molecular modeling	141
5.4.5	Sequence analysis	141
5.4.6	Statistical analysis	142
	Supplementary material	143
6	Thermostability engineering of <i>Streptomyces coelicolor</i> LPMO10C	153
	Abstract	154
6.1	Introduction	155
6.2	Results and discussion	156
6.2.1	General exploration of stability of the WT ScLPMO10C	156
6.2.2	Stabilization through rational inspection of ScLPMO10C's protein surface	160
6.2.3	Stabilization via disulfide engineering	163
6.3	Conclusion	171
6.4	Materials and methods	172
6.4.1	Site-directed mutagenesis of ScLPMO10C	172
6.4.2	Protein expression and purification	172
6.4.3	Activity testing and HPAEC-PAD	174
6.4.4	Heat treatment	174
	Supplementary material	175
7	General discussion and outlook	183
7.1	Overview	184
7.2	Discussion	186
7.2.1	Key to successful expression: control N-terminal processing	186
7.2.2	LPMO (stability) engineering: Challenge accepted!	188
7.2.3	Stability engineering: is there a perfect route?	190
7.2.4	There are the LPMOs! Problem solved?	192
7.3	Outlook	195
	References	199
	Summary	227
	Samenvatting	231

Curriculum vitae

235

Abbreviations

A

Å	Ångström (= 0.1 nm)
AA	Auxiliary activity family

B

B-FIT	B-factor iterative test
BG	β -glucosidase

C

CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CBP	Consolidated bioprocessing
CD	Catalytic domain
CDH	Cellobiose dehydrogenase
CPEC	Circular polymerase extension cloning

D

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSF	Differential scanning fluorimetry
DbD	Disulfide by Design

E

ABBREVIATIONS

EG Endocellulase

G

GH Glycoside hydrolase (=glycosidase)
GHG Greenhouse gas
Glc_n Glucose chain (β -1,4-bond) of n units
Glc_nGlcA Glucose chain (β -1,4-bond) of n+1 units, whereof the last unit is a gluconic acid (oxidized at position C1)

H

HPAEC-PAD High performance anion exchange chromatography with pulsed amperometric detection

I

IMAC Immobilized metal ion affinity chromatography

K

kDa kilo Dalton

L

LB Lysogeny broth
LCA Life cycle assessment
LPMO Lytic Polysaccharide MonoOxygenase

M

MD Molecular dynamics
MeOH Methanol
MS Mass spectrometry
MODIP MOdeling of Disulfide bridges in Proteins
MSA Multiple sequence alignment

N

Ni-NTA Nickel-nitrilotriacetic acid

O

OD Optical density

P

PcGH61D *Phanerochaete chrysosporium* Glycoside Hydrolase 61D
PASC Phosphoric acid swollen cellulose
PCR Polymerase chain reaction

R

RFU Relative fluorescent units
RMSD Root mean square deviation
RMSF Root mean square fluctuation
RPM Revolutions per minute

S

ScLPMO10C *Streptomyces coelicolor* LPMO10C
SDS-PAGE Sodium dodecyl sulphate - Polyacrylamide gel electrophoresis
SHF Separate hydrolysis and fermentation
SSF Simultaneous saccharification and fermentation

T

TCEP tris(2-carboxyethyl)phosphine
T_m Melting temperature
TrCel61A *Trichoderma reesei* Cel61A

W

ABBREVIATIONS

WT Wild type

Y

YNB Yeast nitrogen base
YPD Yeast extract peptone dextrose

Outline

Of all organic carbon on Earth, the majority is captured as structural components in the plant cell wall, while continuously replenished through photosynthesis. These reserves serve not only as nutrition for different forms of life, but also have an enormous potential as feedstock in so-called biorefineries, wherein practically every molecule that is made from fossil resources can be produced in a sustainable way¹. Such biorefineries will be inevitable in the near future because energy demand is continuously increasing while the fossil resources are gradually depleting and environmental concerns are rising^{2;3}. Lignocellulosic biomass is a broad term for wood, agricultural residues, municipal solid waste such as paper and packaging industry wastes, and dedicated energy crops that contain a significant fraction of structural polysaccharides⁴. These non-edible feedstocks are actually the basis for second generation renewable fuels, whereas the first generation, including sugar/starch-based material, is under debate, mainly because of the *food-versus-fuels* controversy⁵.

Depolymerization of lignocellulosic biomass' dominant component, cellulose, yields D-glucose molecules, that can subsequently form the basis for an endless possibility in fermentations. Well-known end-products are bio-ethanol⁶ and lactic acid, the precursor of polylactic acid (PLA, 'bio-plastics')⁷. Enzymatic degradation of cellulose is however impaired by the stable bond (β -1,4-linkage), crystalline structure and wrapping by hemicellulose and lignin. Nature has evolved many enzymatic strategies to deal with these problems. However, the biggest challenge for using lignocellulosic biomass as renewable resource is to make it a cost-efficient process^{8;9}. Until very recently, the enzymatic degradation of cellulose was thought to be brought about by hydrolytic enzymes only, *i.e.* a mixture of endoglucanases and cellobiohydrolases⁸. However, it was always hard to imagine how an individual cellulose chain could get separated from a microfibril and brought into the active site of these cellulases. Therefore, the presence of a 'missing link' in the process was already suggested in 1950, although it remained undetected for many years¹⁰. With the recent discovery of the lytic polysaccharide monooxygenases (LPMOs), a major step towards better understanding of biomass degradation has been taken. This novel class of metalloproteins, degrades cellulose in a manner completely aberrant to that of cellulases and is able to process crystalline substrates very efficiently^{11;12}. Hence, LPMOs render cellulose more susceptible to be attacked by the canonical cellulases. Exploitation of LPMO properties should result in the development of more efficient enzyme cocktails for low-cost lignocellulosic biomass conversion^{11;13}.

LPMOs were first discovered in the process of chitin degradation with Chitin Binding Protein 21 (CBP21) from *Serratia marcescens*¹⁴ and a little later, the analogy was made towards cellulose degradation with the enzymes GH61A from *Thermoascus aurantiacus* and GH61B and E from *Thielavia terrestris*¹⁵. Four classes of LPMOs exist to date, classified in Auxiliary Activity (AA) families 9, 10, 11 and 13¹⁶, whereas the members of AA9

were initially (wrongly) grouped as glycoside hydrolase family 61 (GH61) and AA10 as cellulose binding module family 33 (CBM33). All four of them perform an oxidative cleavage of a polysaccharide chain, being either (hemi)cellulose (AA9, AA10), chitin (AA10, AA11) or starch (AA13). Since their first description as oxidative enzymes and elucidation of their role as cellulase boosters in 2010¹⁴, these enzymes have revived the interest in enzymatic cellulose degradation, which was earlier thought to be inscrutable and inefficient by some groups. The putative power of these LPMOs is reflected in the number of articles that are published since then (Figure 2). During the period January-October 2016 alone, 31 articles have been published on LPMOs. The most recently discovered group, AA13¹⁷, is somewhat exceptional since it boosts amylases in degradation of starch, what is a more amorphous and readily degradable polymer. The obvious advantageous function of LPMOs in making crystalline substrates (cellulose and chitin) more amenable is therefore questioned as being its core function. This underpins the limited knowledge we have obtained in the past 6 years about this interesting class of enzymes. More diverse functions are gradually emerging such as for example a role in virulence¹⁸, keratinolysis¹⁹ and plant infection²⁰. More research is clearly needed.

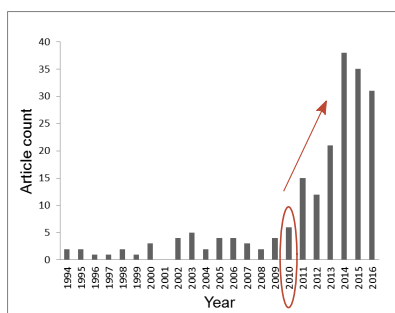


Figure 2: Articles found in the Web Of Science about LPMOs, starting from 1994 until October 2016. Search terms were "polysaccharide monooxygenase OR GH61 OR CBM33 OR LPMO".

Because of the recent discovery of the LPMOs, many aspects of these enzymes still have to be unveiled. Most studies aim to increase the basic understanding of these LPMOs in studies dealing on substrate and electron donor characterization, reaction mechanism or exploration and determination of crystal structure. Very little is known about stability of these enzymes yet, even though this parameter is crucial in industrial applications. Moreover, the canonical cellulases clearly benefit from stability increasing strategies²¹. Advantages attributed to use of higher temperatures include an increased reaction rate and solubility²², a decreased risk on microbial contamination²³ and cell-wall disorganization²⁴.

The aim of this doctoral research is therefore to expand the well-needed knowledge on stability of LPMOs. To that end, two different LPMOs were expressed, assays were evaluated, stability was characterized and an improvement was aimed by protein engineering strategies (Figure 3).

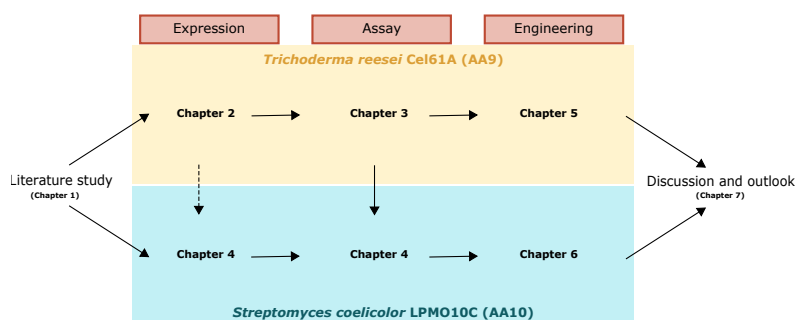


Figure 3: Schematic overview of the chapters in this PhD dissertation.

First, an overview of the required literature is provided in chapter 1. In the framework of biomass-degrading enzymes, only families AA9 and AA10 are considered to be important¹⁹. Therefore, a representative of each family was elected as case study. For family AA9, *Trichoderma reesei* Cel61A was chosen because the fungus *T. reesei* (teleomorph *Hypocrea jecorina*) is the industrial standard when it comes to cellulase production⁸. Its enzymes are exploited among other applications in food/feed processing, textile manufacturing and paper processing²⁵. In this thesis, the LPMO was heterologous expressed in the yeast *Pichia pastoris*, as described in chapter 2. Next, a stability and activity assay were evaluated in chapter 3. As second case study, the AA10 representative *Streptomyces coelicolor* LPMO10C (ScLPMO10C, formerly ScCels2) was expressed in *Escherichia coli*, as described in chapter 4. *Streptomyces coelicolor* is one of the major soil degrading bacteria, while the enzyme ScLPMO10C was one of the first discovered LPMOs. Moreover, family AA10 was first thought to be chitin-active, whereas ScLPMO10C was the first AA10 member that was found to be active on cellulose²⁶. Meanwhile this enzyme is examined in various studies and has become one of the model LPMOs. Different concerns, faced while expressing *TrCel61A*, were also considered while expressing ScLPMO10C in this prokaryotic host. The same assays evaluated for *TrCel61A*, were also tested for ScLPMO10C in chapter 4.

Because stability is typically crucial for industrial applications, chapters 5 and 6 describe various approaches for stability improvement through protein engineering. The first handles on *TrCel61A*, while the second targets ScLPMO10C. To conclude the work, chapter 7 summarizes and discusses the most important findings during this thesis and some future ideas are described.

1

Literature review

Part of this chapter was published as:

Tanghe M., Danneels B., Stals I., Desmet T. (2014) **Enhanced Biomass Degradation by Polysaccharide Monooxygenases**. In: Lin C.S.K. and Luque R. **Renewable Resources for Biorefineries**. 1st ed. Cambridge, UK: The Royal Society of Chemistry 2014. Chapter 3: 64-78.

1.1 The promise of lignocellulosic biomass as renewable resource

Lignocellulosic biomass is a broad term for wood, agricultural residues, municipal solid waste (including paper and packaging industry wastes) and dedicated energy crops that contain a significant fraction of structural polysaccharides⁴. This biomass has enormous potential to serve as renewable feedstock, replacing fossil resources. Since the natural function of a plant cell wall is to provide rigidity however, degradation is highly recalcitrant. The biggest challenge for using biomass as renewable resource is therefore developing a cost-efficient breakdown process^{8,9}.

1.1.1 Renewable resources

About 80 % of the worldwide energy supply in 2014 came from fossil fuels, including oil, coal and gas²⁷. These resources have been the basis for a large variety of fuels and chemicals over the past 100 years, refined in chemical industry. The biggest part hereof goes to energy supply. Meanwhile, the energy demand is continuously increasing and the reserves are becoming gradually depleted. Furthermore, the burning of fossil fuels is one of the big contributors to the increasing amount of carbon dioxide in the atmosphere. Since this greenhouse gas (GHG) is related to global warming, it causes serious environmental concerns. Moreover, most fossil carbon is located in the Middle-East, which is an unstable region and suffers from uncertain availability and varying prices. Taken together, there is an increasing demand for renewable energy resources^{2,3}.

In this respect, current worldwide energy policies are integrating renewable resource goals in their government programs. The renewable energy projections of the 27 EU countries is depicted in figure 1.1. Biomass plays an important role herein with its share of 56 % in 2020²⁸. To meet this goal, the estimated required biomass will increase from 3.8 EJ in 2005 to 10 EJ in 2020 in the EU. Therefore, it is expected that agricultural land will be allocated to grow dedicated energy crops. Because the demand will keep increasing for energy, but also for food/feed and biomaterial production, very efficient agricultural land use is required as well as exploration of alternative resources²⁹.

Other types of renewable energies depend on the availability of resources such as sunlight, wind and water. In this respect, parts of Africa and India have the highest solar capacity factor, while wind capacity factors are highest in Australia and New Zealand and hydroelectric capacity factor is the highest in Canada and South America. The fastest growing type of renewable energy is solar energy, increasing by 8.3 % a year³⁰. However, a barrier to large-scale integration is the variability in supply. Several approaches have been developed to overcome these fluctuations: mainly storage of the energy on moments with

high supply and switchable loads³¹. One interesting solution for integration of renewable energy in the existing system, that supports those needs, is the use of electric vehicles with electric grid connection. The vehicle contains a battery and is plugged in to the electric network. When needed, the battery could put electricity back to the grid, while the battery is loaded when the supply is large³¹. While this technology is further developed, electricity remains a minor fuel based on the world's transportation energy use, with a predicted share of only 1 % in 2040³⁰.

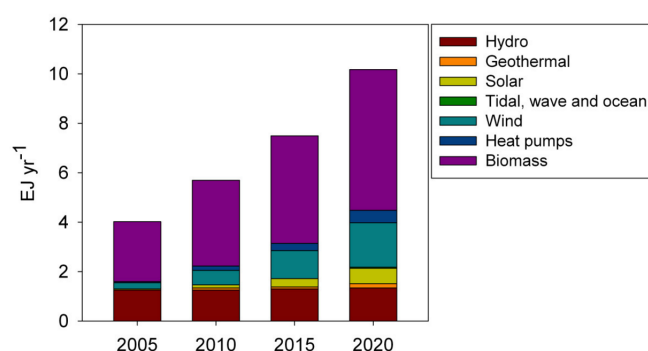


Figure 1.1: Production of energy from renewable resources in the 27 EU countries, as stipulated in the national renewable energy action plans²⁸ (Figure from²⁹)

The quest for alternative carbon sources

In replacement for fossil fuels, alternative carbon sources are searched for. The advantage in comparison to other renewable energy resources such as solar power, is that not only energy but also biofuels and a variety of chemicals and materials can be produced in a sustainable way. Three generations of biofuels are described to date and are discussed below.

The first generation of biofuels comprises direct sugar-containing crops like sugarcane and sugar beet, starch-based crops such as corn and grain, and lipids derived from oilseed such as soybean oil and rapeseed oil. These resources require limited pretreatment and were therefore the first resources used. Nonetheless, the long-term usage was questioned quickly. First of all, they induced the so-called *food-versus-fuel* debate, since the feedstocks used are also major food substances. Secondly, there is a poor energy balance between the energy required for production and the output of the biofuel³². Also, a theoretical maximum of only 10 % of the ethanol required in the US / year was calculated² and a negative impact on the biodiversity, soil quality and even net increase in greenhouse gas (GHG) emission were discussed⁵. Nonetheless, this first generation is the most produced

biofuel with mainly corn (USA) and sugarcane (Brazil) as primary raw material and is used as supplement in gasoline.

An alternative, the second generation of biofuels consumes lignocellulosic biomass. Because cellulose is the main constituent of the plant cell wall, it is the most abundant biological material on Earth³³, and thus a low cost resource. In addition, cellulose is constantly replenished by carbon fixation via photosynthesis, using carbon dioxide. The land use is also more efficient in comparison to the first generation, because here not only the seeds, but the entire plant can be utilized. Moreover, a 90 % reduction in greenhouse gas (GHG) emission was concluded from a life-cycle analysis (LCA) from 2013 comparing cellulosic ethanol with gasoline³⁴. These characteristics make cellulose a highly promising feedstock in biobased industries for production of energy and a variety of chemicals^{1;8;35}. The high cost of pretreatment and enzymes for hydrolysis combined with a low saccharification yield are challenges to face in making it an economically viable process.

Biomass is renewable but also limited to the available land and additional resources (water and nutrients). There are also some concerns rising considering increased use of forest and agricultural waste streams. These include lower soil fertility, soil productivity, decreased organic matter in the soil and lower biodiversity³⁶. Furthermore, there will be pressure on land use because the trade-off between land-use for food supply or non-food biomass production. This not only affects food prices, but also stimulates farmers to increase the productivity by reducing crop rotations and using higher amounts of fertilizer, resulting in acidification and eutrophication^{37;38}. Some studies even argue that there is a net increase in GHG emission in comparison to gasoline when the carbon cost (sacrificed carbon storage and sequestration) for land-use change is included in the life-cycle assessment⁵.

Algae are described for production of third-generation biofuels. They have some outstanding advantages compared to the first two such as fast growth rate and high carbohydrate and lipid content and a lower requirement of arable land. Moreover, they not only fixate carbon dioxide from the atmosphere, but are able to grow on saline and wastewater, whereof they remove pollutants^{39;40}. However, the biggest bottlenecks here are the design of a cost-effective photobioreactor, an efficient harvesting/separation method and metabolic engineering techniques for increasing the productivity^{41–43}.

1.1.2 Structure and challenges of lignocellulose

Lignocellulose is mainly composed out of 3 constituents: cellulose (40-50 %), lignin (20-30 %) and hemicellulose (20-40 %)⁶. The exact proportion depends on the nature of the biomass. Cellulose is a linear homopolymer of D-glucose molecules, linked in a β -1,4-bond. Native cellulose can link up to 10.000 units⁴⁴. These long chains are tightly packed by

hydrogen bonds into crystalline microfibrils, buried in a complex matrix of lignin and hemicellulose (Figure 1.2). Hemicellulose is a heteropolymeric molecule, consisting of pentoses (C5 sugars), hexoses (C6 sugars) and acetylated sugars. Because of its more amorphous structure, it is easier to break down. Lignin is considered the glue of the complex and exploits the strength of cellulose while conferring flexibility. It is an aromatic polymer consisting of mainly *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol. This structure is recalcitrant for microbial and enzymatic attack and therefore, it usually not fermented but rather used for energy generation or chemical extraction⁴⁵.

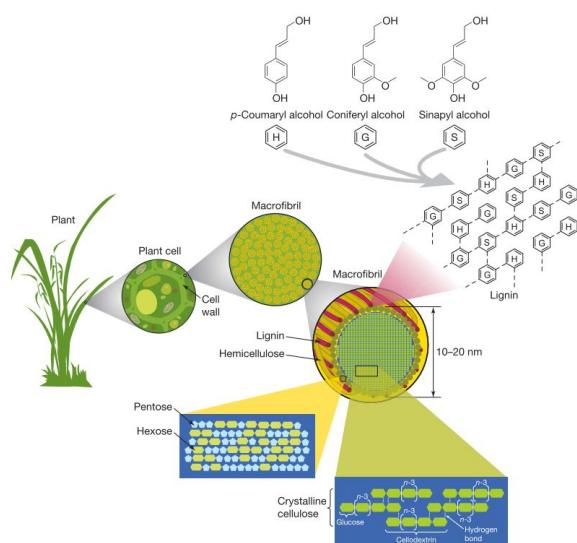


Figure 1.2: Structure of lignocellulosic biomass with cellulose, hemicellulose and lignin as three major building blocks. (Figure from⁶)

Even though starch and cellulose both consist out of D-glucose units as elementary unit, starch is easily degradable, while cellulose is highly recalcitrant towards enzymatic degradation. This has been a serious hurdle for the development of cost-effective biorefineries⁴⁶. Three major reasons form the basis of this problem. First, starch is a storage molecule and therefore its sugars should obviously be easily accessible, while the function of cellulose is to enclose the cell and survive environmental stress such as drought and microbial infection. This is also reflected in the build-up: the stability of the β -1-4-linkage found in cellulose far exceeds the stability of the α -1,4- and α -1,6-bound glucoses, that form amylose and amylopectine respectively in starch. This results in an uncatalyzed half-life time of several million years for cellulose^{47;48}. Second, in the plant cell wall, cellulose is embedded in a matrix of hemicellulose and lignin⁴⁹. This physical barrier limits accessibility makes pretreatment steps crucial⁴. Finally, cellulose chains are associated in so-called microfibrils, which have a high degree of hydrogen bonding and are insoluble in water and

most other solvents. This crystalline structure offers very difficult access for the required enzymes and even water can barely penetrate the structure. To illustrate the difficulty, cellulose requires 40 to 100 times more enzymes to break down in glucose units than starch does³⁵.

1.1.3 Biorefinery concept

Biorefining is defined by the International Energy Agency (IEA) Bioenergy as

*"sustainable processing of biomass into a spectrum of biobased products (food, feed, chemicals, and materials) and bioenergy (biofuels, power and/or heat)"*⁵⁰.

The basis of this definition is a sustainable process, whereas all different generations of biofuels described above are qualified. Concerning the second generation, the heterogeneous structure provides the advantage of different components and putative production of a large variety of chemicals. However, distinct conversion technologies are therefore required. Many processes demand another 10-15 years before commercial stage⁴⁵. A variety of technical processes can be applied in biorefineries: going from fermentation and anaerobic digestion over gasification to chemical processes such as hydrolysis and transesterification. Nonetheless, the majority of biobased products available today, are not produced in a biorefinery, but rather in a single-chain plant.

Transportation fuels obtain a lot of attention concerning biofuels, which is understood by the large fraction of the fossil fuels they consume ($\pm 60\%$ ⁴⁵). Their impact on the environment is important and seems beneficial. For example, blending of biodiesel or bioethanol with fossil diesel reduces particulate matter⁵¹ and bioethanol has a higher octane number than gasoline⁵². Continuous progress in the field is made, yielding engineered microorganisms for the production of a variety of biofuel molecules such as biosynthetic biobutanol, terminal alkenes and fatty alcohols⁵³. In fact, the endeavors done in transportation fuels drive the development of biorefineries. Indeed, despite its high demand and high volume, fuels are low-value products. An integrated biorefinery, that also produces high-value chemicals would give a better return on investment.

Every molecule made from a petroleum refinery can be produced via biorefining, albeit with higher costs⁵⁴. In fact, cellulose and hemicellulose can be hydrolyzed to monosaccharides, that can be isomerized to sweeteners or can be substrate for single cell production or fermentation of a tremendous amount of chemicals such as ethanol, citric acid, L-glutamic acid and penicillin⁴⁴. Generally, the production of a few platform molecules has to be completed. Since the biorefinery is still in its infancy, choosing the most viable molecules is a cumbersome task. The US Department of energy proposed a list of 12 molecules (Fig-

ure 1.3), that was later updated to 10^{55} , whereof other bulk or added-value chemicals can be derived⁵³. A typical example of such a platform molecule is lactic acid as reviewed by Dusselier and coworkers⁷. This is the precursor of polylactic acid (PLA) polymers, the main constituent of biodegradable plastic. Moreover, other fine chemicals, green solvents and fuel precursors are in the range. A second example is succinic acid, easily converted to 1,4-butanediol, tetrahydrofuran and different pyrrolidinone derivatives. However, further strain improvement is indispensable to compete with the chemically produced form⁵⁶.

In addition, lignin can be used as energy source (burning), or can lead to a variety of (aromatic) compounds⁵⁷. One example of a high-potential platform molecule is vanillin. This molecule has not only an obvious application as flavoring component, but is also an intermediate component for production of a wide variety of chemicals such as herbicides, air-fresheners and anti-foaming agents⁵⁸. The hemicellulose component xylose on the other hand is interesting for production of mainly xylitol and furfural⁵⁹.

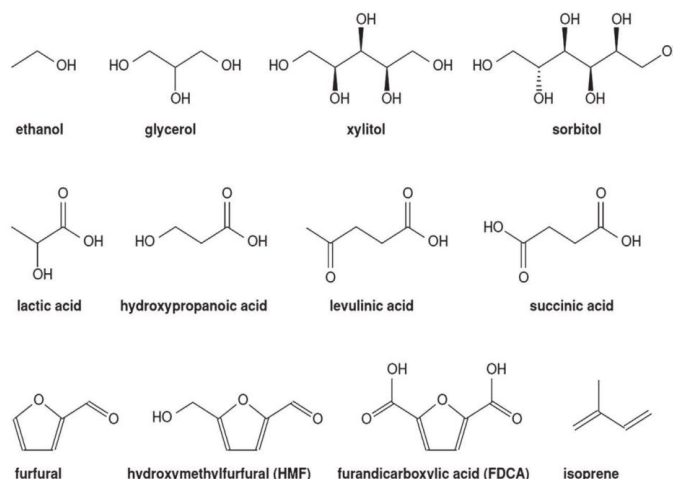


Figure 1.3: Proposed platform molecules from a biorefinery, listed by the US Department of Energy. (Figure from⁶⁰)

1.1.4 Lignocellulosic ethanol: Biochemical versus thermochemical conversion

In the second generation of ethanol production, lignocellulosic biomass is converted into ethanol, either through a biochemical or thermochemical conversion route. The first relies on the use of enzymes and is therefore considered as a sustainable and thus green biotechnological application. Indeed, enzymes are very attractive because of the applicability in mild reaction conditions and their high selectivity and catalytic turnover⁶¹. In biochemical conversion, biomass is mainly converted to polysaccharides, that can be fermented to

ethanol in a subsequent step⁶². The lignin component is mainly burned to generate heat and electricity. The thermochemical route on the other hand involves catalytic synthesis and gasification. During this procedure, raw biomass is heated under oxygen deficient conditions, what results in formation of synthesis gas (syngas, primarily hydrogen and carbon monoxide). This syngas can be burned or further processed into ethanol in the presence of catalysts⁶³.

Although the energy efficiency at plant level is identical for both routes and comparable amounts of ethanol are produced, there are differences from a life cycle perspective. Indeed, a life cycle assessment (LCA) evaluates the true environmental impact by considering an entire life cycle of a process or product. For biochemical conversion routes, several LCA studies exist and confirm that biochemical conversion performs much better than starch-based ethanol production in terms of fossil fuel consumption and greenhouse gas emission (> 50 % GHG reduction compared to gasoline^{64;65}). On the contrary, a negative evolution has been described by indirect impacts such as the use of sulfuric acid for pretreatment purposes followed by lime for neutralization and nutrients consumed during fermentation. For thermochemical conversion methods, published LCA analyses are rare. A comparative study of both routes has been conducted by Mu and coworkers⁶³. The results of ethanol yield, energy efficiency and carbon conversion efficiency are listed in table 1.1. In the current situation, thermochemical conversion performs best for all feedstocks. However, a bigger improvement is expected in the long term for the biochemical conversion. Indeed, the low efficiencies are largely caused by relatively low conversion of (hemi)cellulose (70 %), while this is expected to improve to at least 90 % by intensive research. Intensive research on thermochemical routes can lead to higher alcohol selectivity but cannot exclude co-product formation such as higher molecular weight alcohols. Furthermore, this study also describes a lower GHG emission and fossil fuel consumption for the biochemical route, whereas the thermochemical route performs better in terms of direct and indirect water use⁶³.

Table 1.1: Technical performance of Thermochemical (Thermo) and Biochemical (Bio) plants, based on different feedstocks. Table from⁶³.

	Wood chips		Corn stover		Waste paper		Wheat straw	
	Thermo	Bio	Thermo	Bio	Thermo	Bio	Thermo	Bio
Current								
Ethanol yield (l/dry ton)	283	262	272	264	325	284	270	225
Energy efficiency (as ethanol) %	41	38	36	35	36	31	36	30
Energy efficiency (as ethanol and co-products)	49	44	43	41	43	37	43	36
Carbon conversion efficiency (ethanol/all products)	23/27	21	24/30	24	32/39	28	25/30	21
Near term								
Ethanol yield (l/dry ton)	340	368	326	371	390	399	324	316
Energy efficiency (as ethanol) %	49	53	43	50	43	44	43	42
Energy efficiency (as ethanol and co-products)	58	59	52	54	51	49	51	47
Carbon conversion efficiency (ethanol/all products)	27/33	30	29/35	34	38/46	40	30/36	30
Long term								
Ethanol yield (l/dry ton)	359	411	344	415	411	445	342	353
Energy efficiency (as ethanol) %	51	60	45	55	45	49	45	47
Energy efficiency (as ethanol and co-products)	62	65	54	60	54	54	54	52
Carbon conversion efficiency (ethanol/all products)	29/35	33	31/37	38	41/49	44	32/38	33

1.2 Enzymes for cellulose hydrolysis

Although chemical methods for cellulose decomposition, such as acid hydrolysis and treatment with ionic liquids exist⁶⁶, enzymatic hydrolysis is still the most common and environmentally friendly method. Moreover, the involved cellulases are one of the largest groups of industrial enzymes, used in a wide variety of applications including paper recycling, detergents, juice extraction and cotton processing²¹. A brief overview of the distinct synergistic enzymes working together is given in this section.

1.2.1 Cellulase classification

The Enzyme Commission (EC) of the International Union of Biochemistry and Molecular Biology (IUBMB) has set an enzyme classification system, based on substrate specificity and the type of reaction the enzyme catalyzes⁶⁷. Each enzyme group is defined by a unique 4-digit code, whereof the first digit represents one of the six main classes: (1) oxidoreductases, (2) transferases, (3) hydrolases, (4) lyases, (5) isomerases and (6) ligases. The second and third digit describe the type of substrate, while the last part unambiguously assigns the actual reaction and enzyme specificity. Cellulases are defined as glycoside hydrolases that break the β -1,4-glycosidic linkage between glucose molecules through a mechanism of acid-hydrolysis. They thus belong together with many hemicellulases to the glycoside hydrolases that break a O-glycosyl linkage, assigned with EC-number 3.2.1.x.

Because several thousand glycoside hydrolases exist and no structural information was included in the EC-system, it became insufficient and a different ordering, based on amino

acid sequence, was proposed by Henrissat and coworkers⁶⁸. In this classification, the members originate from public databases and at least one member per family is experimentally characterized. Furthermore, biochemical information is continuously curated and the families largely reflect enzymatic mechanisms and conserved structural folds⁶⁹. The grouping involves catalytic domains and carbohydrate binding domains of all enzymes that build, modify or break linkages in complex carbohydrates or glucoconjugates. This large group of enzymes was called Carbohydrate Active Enzymes or CAZymes, all ordered in the associated database CAZy⁶⁹. Currently, 5 enzyme classes are covered: (1) glycoside hydrolases (GH), (2) glycosyltransferases (GT), (3) polysaccharide lyases (GL), (4) carbohydrate esterases (CE) and recently added (5) auxiliary activities (AA)¹⁶. Carbohydrate-binding modules (CBM) are classified in a separate, non-catalytic class.

A cellulase mixture consists typically of three main components, *i.e.* endoglucanases, exoglucanases and β -glucosidases (Table 1.2). Endoglucanases cut at internal positions of the glucose chain, while exoglucanases work processively from the glucose chain end and β -glucosidases hydrolyze cellobiose to glucose.

Table 1.2: Three main types of cellulase activities, with their EC number and function description.

Activity type	Subtype	EC	Alternative names		Short function description
Endoglucanase		3.2.1.4	1,4- β -D-glucan-4-glucanohydrolase		creation of new chain ends by cutting at internal sites of an amorphous structure
Exoglucanase	cellodextrinase	3.2.1.74	1,4- β -D-glucan hydrolase	glucano-	processive liberation of glucose molecules starting from a chain end
	cellobiohydrolase	3.2.1.91	1,4- β -D-glucan hydrolase	cellobio-	processive liberation of cellobiose molecules starting from a chain end
β -glucosidase		3.2.1.21	β -glucoside glucosidase		hydrolyzation of soluble dextrans and cellobiose to glucose

1.2.2 Major actors in the synergistic machinery

A well-designed enzyme consortium works together to break down the recalcitrant cellulose fibers. The enzymes work in a synergistic way, meaning that the hydrolysis rate of a combination of cellulases exceeds the sum of individual rates. Different synergies have been reported between endo- and exoglucanases, but also between different exoglucanases and between exoglucanases and β -glucosidases. In multimodular systems, even a synergy between the catalytic domain and cellulose-binding domain was described⁸.

Exoglucanases are active on microcrystalline cellulose and work in a processive manner (consecutive action without releasing the substrate), and always start from a chain end. Two types of cellobiohydrolases (CBH) exist, CBH I and CBH II, with a preference of acting from

the reducing and or non-reducing end, respectively. In the catalytic domain, different surface loops come together to form a tunnel region (Fig. 1.4), through which the glucose chain is pulled with release of mainly cellobiose (occasionally glucose and cellotriose)⁷⁰. Mostly, a cellulose-binding module is present (CBM), which is shown to be important for binding and processivity of the enzymes as the CBM brings the catalytic domain and substrate in close proximity⁷¹. CBHs make out an important and substantial part of the cellulase mixture. Indeed, the natural *Trichoderma reesei* cellulase load contains 60 % CBH I and 20 % CBH II. Likewise, optimizing mixtures of *Humicola insolens* resulted in a composition of 70 % CBH I and 30 % CBH II for maximum efficiency⁷².

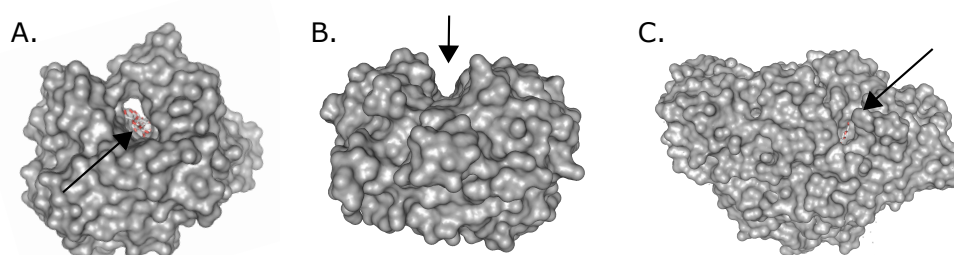


Figure 1.4: Three types of active sites, as found in cellulases from *Trichoderma reesei*. (A) tunnel in CBH I (Cel6A, PDB-code 1QK2) (B) groove or cleft in endoglucanase 2 (Cel5A), with cellotetraose in bound in the tunnel region (PDB-code 3QR5) and (C) pocket or crater in β -glucosidase (Cel3A) with glucose situated in the active site (PDB-code 4I8D). Figure made with Pymol⁷³.

Endoglucanases (EG) create free chain ends for CBHs to work on, by cleaving on internal positions in relative amorphous regions in the cellulose substrate⁷⁴. EGs contain shorter surface loops so that a groove is created rather than of a tunnel region (Fig. 1.4 B)^{75;76}. A CBM is not essential for these proteins, although sometimes present. The optimal *H. insolens* mixture contained only 1-2 % endoglucanase (EG)⁷². Despite this small contribution, many different types of EG exist, probably reflecting the wide variety of cellulose substrates that exist in nature.

β -glucosidases (BG) are not active on cellulose, although they fulfill a very important task. They significantly improve the mixture by removing cellobiose and other short oligosaccharides⁷⁷. This is crucial because CBH for example suffers from end-product inhibition from cellobiose. The *Trichoderma*-mixtures contains a low BG content that is sensitive for glucose inhibition. Therefore, *Aspergillus* BG is usually added to improve the mixture^{77;78}.

1.2.3 Microbial sources of cellulase systems

Free enzyme systems (Fig. 1.5 A) are the most studied and industrially most applied system. They are isolated from cellulolytic fungi and actinomycete bacteria, that release their cellulases in enclosed cavities after penetrating the substrate with their hyphae⁸. The industrial standard today is *Trichoderma reesei* (teleomorph *Hypocrea jecorina*), named after its principal investigator Elwyn T. Reese. During World War II, the US Army was suffering from cotton rot and thus deteriorated tents in the South Pacific. Reese isolated *T. viride* strain QM6a, that was later renamed after him. This actual strain forms the basis of all industrial *Trichoderma* cellulase-producers today⁷⁹. *Trichoderma* secretes an highly efficient cellulase system, although it does not produce a large variety of enzymes⁸⁰: two exocellulases, 5 endocellulases and 2 β -glucosidases. Alternative widely studied systems are *Humicola insolens*^{72;81} or bacteria from genera *Cellulomonas*⁸² and *Thermobifida*, one of the major soil degrading organisms⁸³.

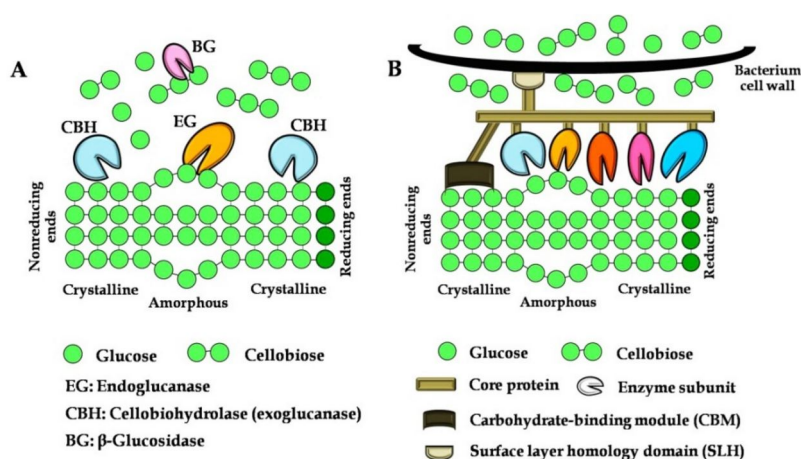


Figure 1.5: Schematic representation of (A) a free cellulase system versus (B) an associated cellulosome (Figure from⁸⁴).

Secondly, cellulases can also exist in large multi-enzyme complexes, called cellulosomes (Fig. 1.5 B). These are mainly found in anaerobic bacteria, that cannot penetrate the substrate and thus compete with other surrounding organisms for the hydrolyzed sugars. Moreover, they only have a limited ATP available for cellulase production. As a solution, nature invented the enzyme-complex, that is associated with the cell wall of the bacterium and is only formed when cellulosic substrate is available. Often a glycocalyx and/or contact corridor connects the cell with the cellulosome to avoid product diffusion and facilitates easy product uptake⁸⁵. The general architecture is common, with a non-enzymatic scaffoldin that binds the various catalytic domains through so-called cohesins. The catalytic domains are connected to a dockerin, that matches to the cohesins⁸⁶. Mostly, also a CBM is attached

to tightly bind the substrate. The exact composition however, can vary depending on the species. The complex is usually 2-16 MDa, but can become up to 100 MDa, called a poly-cellulosome. The first discovered and best characterized cellulosome is from the anaerobic bacterium *Clostridium thermocellum*⁸⁷. It is thought to contain at least 22 catalytic modules including endocellulases, exocellulases, hemicellulases and 1 lichenase⁸⁸. Alternatively, *Rhuminococcus* species, found in rumen of large herbivorous animals, are also well-known cellulosome producers^{85;89}.

1.2.4 Non-hydrolytic factors aiding cellulose hydrolysis

In recent years, increasing attention goes to the importance of non-hydrolytic and even non-enzymatic factors that can help in amorphogenesis, *i.e.* to loosen the difficult crystalline structure of cellulose in order to make it more amenable for cellulase attack⁹⁰. Several factors indeed hamper an efficient degradation. These can be either related to the substrate such as cellulose crystallinity and lignin/hemicellulose wrapping (see above), or related to limiting ability of the cellulases itself such as end-product inhibition and accessibility restricted to the surface of the crystals⁴⁹. In 1950 already, Reese proposed the concept of amorphogenesis with the C₁-C_x theorem for enzymatic breakdown of cellulose¹⁰. Herein the C₁ component activates the degradation by loosening or deaggregation of the structure, while the C_x component subsequently performs the depolymerization. While all cellulases are contained in the C_x component, elucidation of suitable C₁ components seems more cumbersome.

Expansins have been proposed as one of these C₁ factors. They are plant-derived proteins without enzymatic activity. These proteins are thought to enlarge cell walls by disrupting non-covalent interactions in the cell wall (e.g. cellulose-hemicellulose)⁹¹. In this way, they stimulate growing cells⁹². However, when expansins are removed, the cell wall is found restored in its original state⁹³. A large variety in expansins has been described and their function is under investigation. One explanation would be dedication to specific cell types as well as related processes like fruit ripening⁹⁴ and pollination⁹⁵. Adding them to cellulase mixtures yields a synergistic effect⁹⁶.

Similarly, it was shown that some bacteria and fungi also produce expansins with comparable structural and functional characteristics. One example is the EXLX1 protein, from the soil bacterium *Bacillus subtilis*, that enhances root colonization⁹⁷. Furthermore, *T. reesei* produces a protein called swollenin⁹⁸, able to swell cotton fibers, but with a different structure than the expansins. Recent literature also suggests a hydrolytic activity for these proteins⁹⁹. Besides, a third expansin-like protein, loosenin, was found from the basidiomycete *Bjerkandera adusta*¹⁰⁰. It is distantly related to expansins, although can bind polysaccharides, while enhancing hydrolytic enzymes.

A completely different group of non-hydrolytic enzymes that could fill the gap in our understanding of biomass degradation, comprises the recently discovered lytic polysaccharide monooxygenases (LPMOs). These enzymes degrade cellulose in a manner completely different to that of cellulases and is able to process crystalline substrates very efficiently^{11;12}. Hence, LPMOs render cellulose more susceptible to attack by the canonical cellulases and exploitation of their properties should result in the development of more efficient enzyme cocktails for low-cost lignocellulosic biomass¹⁰¹. More information on this interesting class of enzymes is given in the next topic.

1.3 Lytic Polysaccharide MonoOxygenases

Lytic polysaccharide monooxygenases (LPMOs) are a recently discovered class of enzymes that aid in breakdown of the polysaccharide substrates (hemi)cellulose, chitin and starch by a direct oxidative reaction. They come in especially helpful in biomass degradation, which is a difficult and costly process. LPMOs can directly bind the inaccessible and crystalline substrates by their flat binding surface and randomly create new chain ends. Hereby they render the substrate more amenable for the classical cellulases and they boost the entire process.

1.3.1 Discovery and classification of LPMOs

Despite the fact that Reese already indicated some missing factors in the mechanism of cellulose breakdown in 1950¹⁰, Eriksson and coworkers described the beneficial effect of aerobic conditions on natural cellulase mixtures in 1974¹⁰² and it was generally known that redox enzymes were involved in lignin degradation¹⁰³, it took till 2010 before the involvement of lytic polysaccharide monooxygenases (LPMOs) in direct, oxidative depolymerization reaction was elucidated¹⁴.

Some LPMOs were known earlier, although wrongly classified as glycoside hydrolase family 61 (GH61) or carbohydrate binding module 33 (CBM33)¹⁰⁴. Family GH61 contained mainly fungal enzymes, such as for example *T. reesei* Cel61A¹⁰⁵, although a clear function was not assigned yet. These enzymes were co-induced with other cellulases and some harbored a CBM1, suggesting a role in cellulase breakdown. Moreover, they could bind cellulose and a very low endoglucanase activity was measured. The first crystal structure of *T. reesei* Cel61B in 2008¹⁰⁶, revealed an aberrant build-up as compared to classical cellulases. Indeed, no tunnel, cleft or pocket was present, but a flat binding surface with a metal-binding site. Therefore, more questions about an actual glycoside hydrolase character were raised. A similar observation was done by Harris and coworkers in 2010, revealing a crystal structure of *Thielavia terrestris* GH61E¹⁵. They were also the first to show a boosting effect on cellulose hydrolysis when adding the GH61 enzymes to a classical cellulase mixture. In addition, mutating the metal-coordinating residues resulted in loss of the boosting effect.

Besides, non-hydrolytic CBMs were described, such as CBHI from *Streptomyces olivaceoviridis*, that could bind α -chitin but without shown activity¹⁰⁷. An ortholog of this enzyme, *Serratia marcescens* CBP21 (chitin binding protein 21), would be the first LPMO described. The enzyme was classified in family CBM33 and a crystal structure was revealed in 2005¹⁰⁸, showing a similar fold as the GH61 enzymes. To further draw the analogy, mutagenesis experiments of the metal-binding residues impaired the chitinase C synergy. In 2010, finally, Vaaje-Kolstad and coworkers were the first to prove an oxidative mechanism for CBP21 in

chitin degradation and suggested highly similar reactions for GH61 in the structural analogue cellulose¹⁴. Since then, more accurate names have been proposed, such as copper metalloenzyme, oxidohydrolases and (lytic) polysaccharide monooxygenase. In 2013, these enzymes were reclassified as auxiliary activity family 9 (AA9, formerly GH61) and auxiliary activity family 10 (AA10, formerly CBM33) in the CAZy database¹⁶. The name lytic polysaccharide monooxygenases is generally adopted to point to this diverse group of enzymes.

Nowadays, LPMOs include the auxiliary activity families 9, 10, 11¹⁰⁹ and 13¹⁷, as listed in table 1.3. Even though the grouping is based on sequence similarity, families AA9, 11 and 13 contain predominantly enzymes from eukaryotic origin (99 % fungal), while family AA10 includes mostly bacterial enzymes (although also contains some viral representatives). Three different substrate groups are comprised: (hemi)cellulose, chitin and starch. (Hemi)cellulose forms the main substrate for members of family AA9 and part of AA10, while chitin is the preferred substrate for representatives of family AA11 and part of AA10. Families AA9 and AA10 are thought to have a common ancestor. The ancestor of AA10 is suggested to be chitin-active and thus part of the group diverged towards cellulose oxidation¹¹⁰. Surprisingly, starch is the target of AA13 although this is not a crystalline substrate, thereby losing the benefit of the flat binding surface.

Table 1.3: The 4 auxiliary activity classes, representing the LPMOs. The number of enzymes from the different source organisms is derived from CAZy (October 2016).

	Source organism				Main substrate	Oxidation position
	archaea	bacteria	eukaryota	viruses		
AA9 (formerly GH61)	0	0	399	0	cellulosic substrate	C1, C4, C1 / C4
AA10 (formerly CBM33)	1	2036	6	133	chitin, cellulosic substrate	C1, C1 / C4
AA11	0	1	65	0	chitin	C1
AA13	0	0	14	0	starch	C1

For family AA9, a further sequence-based grouping was proposed that subdivides the family in 3 types, varying in their site of oxidation (Fig. 1.6). LPMO type-1 includes the LPMOs that introduce a C1-oxidation, LPMO type-2 includes members that perform a C4-oxidation and LPMO type-3 can oxidize at both sides^{101;111}. One exception, LPMO-3*, only hydroxylates the C1-carbon of the β -1,4-linkage¹¹². In the early days, some researchers also indicated C6-oxidation, which is not performing an actual chain cleavage¹¹³. Other researchers suggested later that this C6-oxidation is not taking place. The subdivision is not absolute however, *e.g.* *Podospira anserina* AA9H was predicted to be an LPMO-2, while it is shown to produce both C1- and C4-oxidized products¹⁷. For cellulose-active AA10 members, C1-oxidation and C1/C4-oxidation pattern exist¹¹⁴, although no pure C4-

oxidizers have been described so far. AA11 and AA13 members are currently only known to oxidize at C1.

A distinct system was suggested by Busk and Lange, who divided the AA9 family based on short conserved sequences using an algorithm called Peptide Pattern Recognition. The algorithm led to 16 subfamilies wherein the already characterized LPMOs belong to only six of these subclasses. The 16 subfamilies thus probably are an overestimation or can indicate much more variability, which is not elucidated yet^{115;116}. Nevertheless, this system does not find general application in further literature.

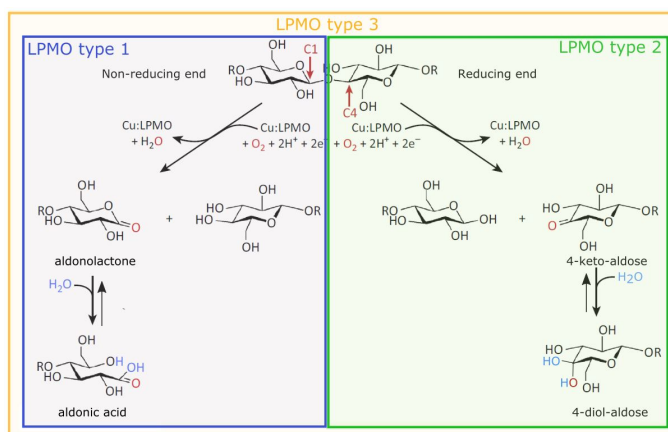


Figure 1.6: LPMOs from family AA9 are further divided into three types depending on the type of oxidation they perform. Type 1 performs oxidation at the reducing end (C1) of the cellulose chain, while type 2 oxidizes the non-reducing end (C4) and type 3 can perform both reactions. (Figure adapted from:¹¹⁷)

1.3.2 Synergy with classical cellulases

Breaking up cellulose fibrils into soluble chains requires a substantial amount of energy, which means that the pretreatment of biomass is a crucial (yet expensive) step. However, LPMOs are able to bind directly to crystalline cellulose thanks to a flat active site that is located at the protein's surface¹⁰⁶. By cutting the chains and introducing oxygen atoms, LPMOs can disrupt the substrate's crystalline structure and thus facilitate the action of cellulases. Furthermore, the electrons needed for this oxidative cleavage can be delivered amongst others by redox active compounds like gallic acid or ascorbic acid (see further for full list of known electron donors). Consequently, a high degree of synergy is to be expected between the redox enzymes on the one hand, and the hydrolytic enzymes on the other hand.

Harris and coworkers described the LPMOs for the first time as 'cellulase-enhancing' factors¹⁵. They described an almost twofold reduction in enzyme loading by addition of *T. aurantiacus* GH61A to Celluclast 1.5L (Novozymes) to reach 91 % cellulose conversion. One year later, Langston *et al.* described a doubled conversion of microcrystalline cellulose

after 72 hours incubation when *TaGH61A* and *HiCDH* (as electron donor) were added to a combination of *HjCBHI* and *HjEGII*¹². The individual cellulase specificities that are present in the mixture were also evaluated in combination with LPMO, whereof it could be concluded that LPMOs enhance the work of cellulases in a generic way rather than improving the action of one specific type of cellulase^{12;15}. Furthermore, the enhancement of LPMO addition to classical cellulase cocktails was clearly proportional to the amount of LPMO added¹¹⁸ and also different LPMOs can have a synergistic effect.

In order to mimic industrial relevant scenarios for enzymatic cellulose breakdown, one should apply real cellulosic substrates such as corn stover, wheat straw and sugarcane bagasse and operate the process at high dry matter content (> 20 %) ^{119;120}. Under these conditions, Cannella and coworkers evaluated the effect of oxidative enzymes in a commercial cellulase cocktail, Cellic CTec 2 (Novozymes), as compared to earlier cocktails, Celluclast 1.5L and Novozym188 (Novozymes), that lack LPMOs¹²¹. As substrate, they used hydrothermally pretreated wheat straw from a pilot scale facility (30 % water insoluble solids) and only used the present lignin as suitable electron donor. The study showed that Cellic CTec 2 had a 25 % higher conversion yield after 144 hours and that only 4 % of the released glucose appeared as the oxidized gluconic acid. Besides, also the pretreatment method appears to have an effect on LPMO activity. Organosolv, hydrothermal and alkaline pretreatment were evaluated considering the chemical composition of the biomass and the subsequent enzymatic hydrolysis and oxidation by the same commercial cellulase-mixture Cellic CTec 2. The highest LPMO activity was observed for the pretreatment that had left the highest lignin content, namely the hydrothermally pretreated biomass (observed for all three substrates). Furthermore, wheat straw resulted in higher cellulose conversion after 96 hours (80 % glucose yield and 0.4 % oxidized cellulose) than sugarcane bagasse (64 % and 0.46 %, respectively) and corn stover (50 % and 0.25 %, respectively)¹²². Surprisingly, the amount of LPMO added to a cellulase mixture for optimal yield appears to be relatively low (about 2 mg/g cellulose) independent of the substrate concentration for both steam pretreated corn stover and popular¹²³.

The known specificities of LPMOs have been broadened in the last years. Indeed, Isaksen and coworkers described activity on short soluble oligosaccharides in 2014¹²⁴, while also different hemicelluloses have been described as substrates such as xyloglycan and lichenan in addition to or exclusive from cellulose specificity^{125–127}. Moreover, synergistic action of LPMOs with hemicellulases (xylanase) has recently been described¹²⁸. This collaboration is found on pure xylose or pure cellulose substrate as well as on pretreated lignocellulosic biomass. Because there is also a large diversity in substrate composition, that demand different compositions of the various cellulases and accessory enzymes (like LPMOs, that are shown to act different on various substrate compositions¹²⁹), the idea of having tailor-made enzyme cocktails is gaining more attention than making one generic

cocktail for efficient saccharification of different sources of biomass^{123;130}.

LPMOs are exclusively found in aerobic organisms, meaning that enzyme-complexes (cellulosomes), produced by anaerobic bacteria, lack these promising type of enzyme. Therefore Arfi and coworkers included *Thermobifida fusca* LPMOs in their designer cellulosome and obtained indeed increased saccharification. The cellulosome-principle also yielded a 1.7 fold improvement in comparison to the free enzymes and even 2.6 fold improvement in comparison to cellulases without LPMO¹³¹.

In complete analogy, chitin-active LPMOs boost the activity of chitinases, also working in a synergy with endo- and exochitinases, while starch active LPMOs (AA13) push activity of β -amylases. A recent study even describes a 30- and 20-fold improvement in alpha- and beta-chitin, respectively, by addition of *Streptomyces griseus* LPMO to chitinases¹³².

1.3.3 Protein Structure

Overall Fold

The location of the active site in LPMOs is very different from that in canonical cellulases. Indeed, the latter contain a binding cleft or tunnel (See figure 1.4 on page 15), whereas the former have a wedge-like structure with a flat binding surface that can adhere to crystalline substrates. Their fold comprises two anti-parallel and twisted β -sheets that form an immunoglobulin- or fibronectin-III like structure (Figure 1.7)¹⁵. Most LPMOs are composed of a single domain, although approximately 30 % of the AA9 members and 30 % of the AA10 members contain an additional carbohydrate module (CBM) to facilitate substrate binding. The first carries mostly a CBM1 (cellulose-binding), while the latter concerns a CBM2 or CBM3 (cellulose-binding) or CBM5 or CBM12 (chitin-binding)¹¹⁰. For family AA13, no less than 60 % carries a CBM20¹³³.

A complete characterization of the substrate-protein interactions is required to get a full understanding of the LPMO action. Obtaining this information is impeded by the insoluble nature of the crystalline substrate. However, to date, some clues are elucidated. Binding cellulosic substrate is usually performed by stacking interactions with aromatic residues in the binding surface, in a comparable way as found for CBM1¹⁰¹. Furthermore, a flexible loop, L2, is suggested to account for cellulose specificity since it is also found in cellulose-specific AA10 members and is reduced in AA11 and AA13 members¹³⁴. Moreover, it enlarges the substrate binding plane and contains aromatic residues, ideally spaced for cellulose binding. MD-simulations revealed two additional flexible loops, called LC and LS, that could be involved in substrate binding¹³⁵, or alternatively, could allow association with its proposed physiological redox partner cellobiose dehydrogenase (CDH). In that way, that

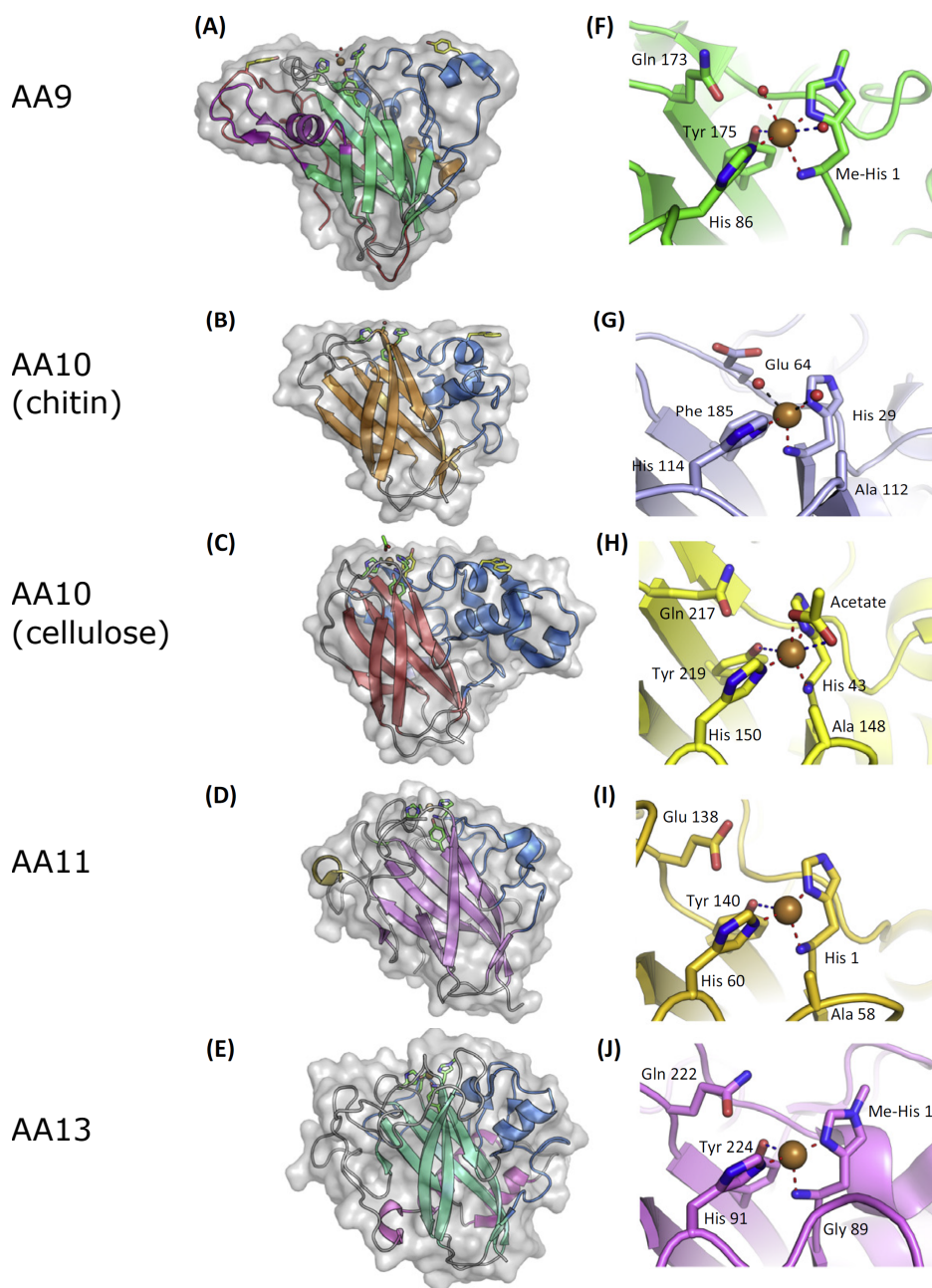


Figure 1.7: Overall and detailed active-site structure of the different auxiliary activity families representing LPMOs, with the substrate binding surface located on top. (A,F) *Thermoascus aurantiacus* AA9 (PDB: 2YET), (B,G) *Enterococcus faecalis* AA10 (PDB 4 ALC), (C,H) *Streptomyces coelicolor* LPMO10B (PDB 4OY6), (D,I) *Aspergillus oryzae* AA11 (PDB 4MAI) and (E,J) *Aspergillus oryzae* AA13 (PDB 4OPB). The L2 loop is colored in blue, LC loops red and LS magenta. The aromatic residues in the surface, responsible for substrate binding are indicated in yellow. The dashed lines in the copper-active site are red for interaction with copper, blue in the axial positions and black for specific chitin binding. (Figure adapted from:¹¹⁷)

enzyme can provide the electrons that are required for the reaction via a conserved hydrogen network that leads from CDH to the active site of the LPMO¹³⁶. Chitin-specific LPMOs on the other hand rather bind their substrate via hydrophilic interactions¹³⁷. These enzymes were first thought to harbor an extra cavity in their surface for the protruding N-acetyl group from chitin¹¹⁴, although a later study demonstrated this is not generally true¹³⁸. An interesting finding by Forsberg and coworkers is that the ability to bind a substrate does not always imply activity on this substrate. Indeed, *Streptomyces coelicolor* LPMO10C can bind chitin, nonetheless lacks activity on this substrate¹³⁹. Therefore, they proposed the geometry of the copper-center would be of major importance.

Furthermore, AA11 and AA13 deviate slightly from the typical flat surface. The first appears to have a more convex substrate binding plane¹⁰⁹, while the latter contains a shallow groove and lacks aromatic residues in the surface. The size of the groove would be large enough to bind an amylose chain, although this is not experimentally confirmed (yet)¹⁷.

Since LPMOs are secreted, post-translational modifications can occur such as disulfide bridges and glycosylation. The N-glycosylation site in LPMO-3 from *Neurospora crassa* is found close to the active site, forming an extension of the flat binding surface. The appearance of an alpha 3₁₀-helix in the L2 loop seems to be correlated to this glycosylation and seems to direct the active site towards the substrate. An identical combination was found in *T. reesei* Cel61B, despite being absent in other family members¹⁰¹. Furthermore, various fungal members have been identified with a methylation of the N-terminal histidine residue. The exact function is still unknown because recombinant expression hosts like *P. pastoris* or *E. coli* lack the machinery to perform this methylation, although the produced heterologous LPMOs are still active^{109;113;135}. Achmann and coworkers proposed the methylation aids copper binding¹⁴⁰.

Active site: copper coordination center

The flat binding surface of an LPMO carries a type 2 copper center, which is the protein's active site. Initial studies were not conclusive on the nature of the metal whereas a number of studies in 2011 revealed copper was the most suitable in AA9 proteins^{11;111;141}, whereafter the same conclusion followed quickly for AA10 members¹⁴². Meanwhile it has been shown that copper is bound very strongly, in nanoMolar scale, in the active center^{140;143;144}. Two histidine residues form the core of this center in a T-shaped geometry, the so-called *His-brace*¹¹, and it is the most conserved region in the protein (Figure 1.8). Copper is bound via the actual N-terminus, the side chain of this required amino-terminal histidine and a second histidine side chain of a distinct residue. Even though most other oxygenases contain 2 or 3 copper sites, LPMOs only contain a single copper. A highly similar structure has also been found in methane monooxygenases^{11;111}.

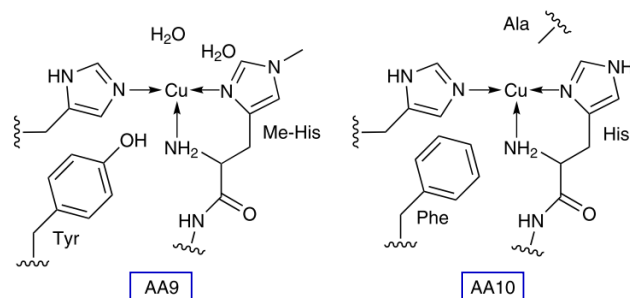


Figure 1.8: Schematic representation of the copper active site in LPMOs from family AA9 and AA10. The N-terminal histidine is methylated for AA9 in this schema. (Figure adapted from:¹⁰⁹)

There is some ambiguity in the remainder of the copper site (apart from the His-brace) including the exact coordinating ligands and the oxidation state of copper. The reason behind this is that X-ray beam causes photoreduction of copper and NMR is impeded by the paramagnetism of Cu(II)¹⁴⁵. Mostly a trigonal bipyramidal structure is found with an oxygen, water or substrate molecule at the axial position out of the protein. Sometimes an additional water molecule is found as equatorial ligand in the His-brace plane. A tyrosine is found on the axial position in the protein for families AA9, 11 and 13 (Fig. 1.7 (F,I,J)). For family AA10, either a tyrosine (Tyr) or a phenylalanine (Phe) can occur, and the last option is most conserved (90 %). All AA10 also contain an alanine (Ala 112 and Ala 148 in Fig 1.7 G and H, respectively), which results in a blockage of the axial position of copper and thus a forced different coordination geometry of exogenous ligands. Despite the conservation of this alanine, it does not imply an identical spatial position. The absence of this alanine is suggested as C4-oxidizing determinant in family AA9^{114;146}. Additionally, family AA11 (so far) also contains this alanine, while a loop region containing a glycine is found in AA13 at this position.

1.3.4 The oxidative reaction

Independent of the substrate preference and independent of the regioselectivity (C1- and/or C4-oxidation in glucose linkages), every LPMO intervention is dependent on an exogenous electron donor source to provide 2 electrons and molecular oxygen to provide one oxygen atom (mono-oxygenase). The oxygen indeed originates from molecular oxygen as shown via $^{18}\text{O}_2$ isotope labeling first for AA10 member *SmCBP21* in chitin degradation¹⁴ and later for AA9 member *NcLPMO9E* (NCU08760) in cellulose degradation¹⁴⁷. However, in a very recent groundbreaking perspective, Bissaro and coworkers claim that hydrogen peroxide and not oxygen would be the actual cosubstrate of the LPMOs¹⁴⁸. Regarding the electron donor, LPMOs show a promiscuity and most recent papers suggest even an LPMO-dependent reducing agent preference¹⁴⁹.

Different exogenous electron donor sources have been described to date (Figure 1.9). The first and most used reducing agent is ascorbic acid^{14;125;147;150}. Alternative small molecules that have been reported are gallic acid, catechin and low molecular-weight lignin products (phenols)^{11;124;151;152}. These are very useful in laboratory experiments, although they are questioned for their physiological relevance regarding their natural availability in the system. For example ascorbic acid is only present in non-lignified plant tissue. As second electron donor, another redox enzyme, cellobiose dehydrogenase (CDH), has been described¹². This enzyme generates electrons at its flavin-binding dehydrogenase domain via the oxidation of cellobiose, and this electron is subsequently shuttled to the heme-binding cytochrome domain of the enzyme¹⁵³. Only this last cytochrome domain is responsible for the direct electron transfer to the LPMO¹³⁶. The speed of electron transfer is dependent on pH and the specific LPMO-CDH combination. This system is suggested to be physiologically more important because almost 60 % of the fungi contain *cdh* and *lpmo* genes, that are often co-expressed¹⁵³. In addition, the LPMO-CDH system is thought to have a faster electron-transfer rate than the small-molecules. To date, CDH is not found in bacteria, although co-expression of LPMO and other redox-enzymes have been described for *Cel-lvibrio japonicus*, suggesting the existence of alternative systems¹⁵⁴. Third, a light-driven oxidation is described through photosynthetic pigments¹⁵⁵. Both soluble chlorophyllin and membrane-bound thylakoids can herein provide the electrons after light-induced excitation of the pigment. The cycle is completed by a pigment reduction via lignin or ascorbic acid. Notably, not only the catalytic activity of LPMOs was significantly increased ($\sim \times 100$), but also the specificity of the investigated enzyme was broadened. As last group of electron donors plant-derived and fungal phenols, especially diphenols, are described. Moreover, a continuous supply can be ensured by the coupled recycling via glucose-methanol-choline oxidoreductases (GMC oxidoreductase)^{152;156}. Some examples are glucose dehydrogenase, glucose oxidase and pyranose dehydrogenase. Interestingly, also the above mentioned CDH belongs to this enzyme class. Care should be taken in whole cellulase systems because phenolic compounds might also inactivate or inhibit cellulases^{157–160}. The biological relevance of this reducing agent promiscuity is not completely understood. One hypothesis is the flexibility to degrade cellulose substrate in various stages of the breakdown while adapting to the changing substrate.

Just like the disagreement in the exact copper-center configuration, there is still discussion on the exact reaction mechanism. First of all, some researches suggest a direct electron transfer at the copper site^{136;153}. Alternatively, long-range electron transfer pathways through the enzyme have been described^{101;135}. The last option merits from the benefit of the LPMO being absorbed to the substrate while electrons can be continuously provided. It can not be excluded that distinct parallel systems might exist. Second, different

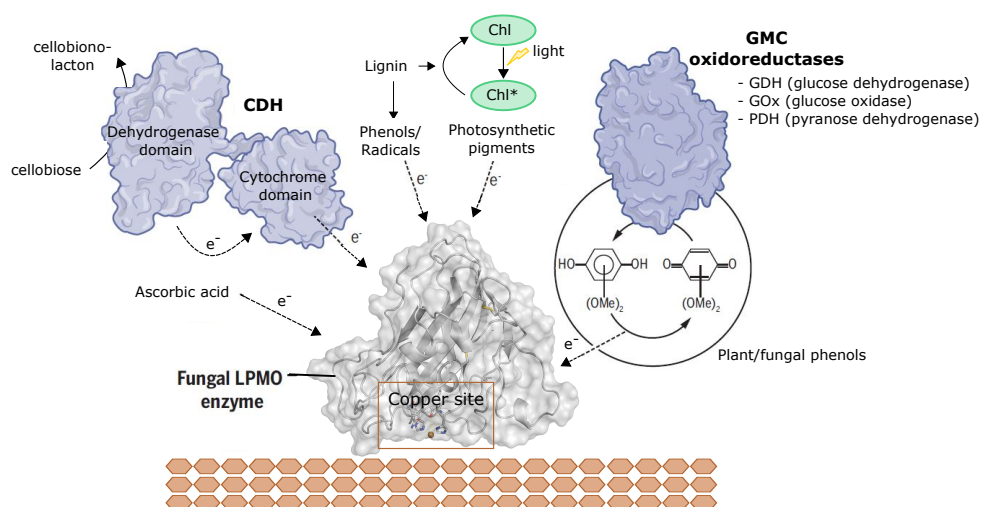


Figure 1.9: Different electron donor sources for a fungal LPMO. (Figure based on ¹⁶¹, ¹⁵² and ¹⁵⁵)

LPMO mechanisms for bond cleavage have been proposed. A consensus exist about the first part though, an electron reduces Cu(II) to Cu (I), where after O₂ is activated and a Cu(II)-superoxo complex is formed. What happens next is to be unveiled. It is unsure if this superoxo species is the principal catalytic form¹⁴⁷ or if a more reactive oxyl-radical is formed afterwards¹⁶².

1.3.5 Diverse functions of LPMOs

An intriguing observation is that several organisms produce multiple LPMO enzymes that display high sequence variability. For example, the soil fungus *Haetomium globosum* contains 44 unique LPMO genes¹⁶³, whereas a cell-wall degrading fungus is estimated to code for an average number of AA9 members in their genome of no less than 10¹⁹. Even more remarkable, some fungi have more LPMOs than cellulases. *Neurospora crassa* for example, codes for 10 LPMOs with an average pairwise sequence identity of 33%. A suggested hypothesis is that this multiplicity covers the diversity in cellulosic substrates, electron donor preference or variability in environmental circumstances (such as pH, salt concentration and temperature), rather than the efficiency of a cellulase mixture^{20;116}. Furthermore, it has been suggested by some authors that different LPMOs are required to attack cellulose concomitantly from different angles¹⁶³ and on different substrate types¹⁵⁰. This also implies that a cellulase mixture has to be optimized for each different type of biomass. It can be concluded that LPMOs (families AA9 and AA10) play an underestimated role in biomass degradation, but a detailed understanding of their mode of action is still lacking^{19;101}.

This role is not attributed to families AA11 and AA13, which lack activity on cellulose. For family AA11, a role in degradation of keratin or another skin-macromolecule was sug-

gested¹⁹, deduced from their prevalence in dermatophytes. Moreover, the starch active AA13 LPMOs are also highly represented in for example *Aspergillus nidulans* and are co-regulated with starch-degrading enzymes¹⁶⁴. An important role in starch-degradation is therefore assumed as well.

Their importance and role in biomass degradation industry is obvious. Nonetheless, a putative role in pathogenesis has come up in recent years, mostly for AA10 members. In that respect, virulence factors of the human pathogens *Listeria monocytogenes*¹⁶⁵ and *Vibrio cholerae*¹⁶⁶ and an insect pathogen¹⁶⁷ were found to contain an LPMO-domain with activity on chitin. The LPMO helps to degrade the peritrophic matrix in the gut of invertebrates by the LPMO as key virulence factor¹⁸. Furthermore, also plant infection has been related to LPMOs (AA9), contained by a variety of necrotrophic fungi. For example, *Colletotrichum graminicola* (maize pathogen) expresses 22 out of 28 LPMOs during its necrotic phase¹⁶⁸. Alternatively, the pathogen *Blumeria graminis* only contains 1 LPMO, associated with infection.

Apart from the oxidative polysaccharide cleavage, LPMOs also produce reactive oxygen species (ROS) in absence of substrate (but presence of electron donor). This ROS-generation can come in handy during infection of a living plant. This plant will induce a sensitive response, that leads to localized cell death in the plant, a benefit for necrotic organisms. Indeed, several AA9-LPMOs are overexpressed during plant infection. At the same time, LPMO expression is sometimes found concomitant with ROS-neutralizing enzymes such as catalase to protect the LPMO producer itself²⁰ and reduce cellulase inactivation¹⁶⁹.

Lastly, a symbiosis of *Sirex noctilio* (woodwasp) and AA10 LPMOs has been described. The wasp leaves eggs in the bark of a pine tree, covered with mucus that include fungal spores and bacteria. These produce cellulose degrading enzymes, such as LPMOs, to provide nutrients for the larvae^{170;171}.

1.4 Biochemical lignocellulose biorefining and LPMO contribution

The development of a more efficient biorefinery process is of very high interest. Indeed, oil can theoretically be replaced by plant biomass to produce a range of chemicals, and of course biofuels. The following text will be mainly focused on biofuels, although the basic concepts of degrading biomass are valid for all processes.

1.4.1 Fermentation strategies for production of biofuels

The conversion of lignocellulose biomass into valuable products, such as a biofuels, takes generally 4 steps: (1) Pretreatment is mostly required to increase cellulose digestibility, (2) Biomass is hydrolyzed to hexose and pentose sugars, (3) The sugar is fermented into a valuable product such as bioethanol (4) Bioethanol is recovered.

A pretreatment method of the biomass is generally performed to make the structure more accessible by (partially) liberating the cellulose from the lignin and hemicellulose components. The pretreatment can be biological, chemical or mechanical or any combination^{1,2}, although the biological treatment is not suitable for industrial purposes because it takes too long. The choice for the ideal method strongly depends on the structure and composition of the biomass used¹⁷². Even though it is an expensive step in the process, pretreatment is inevitable. A study using Ammonia Fibre Explosion (AFEX) as pretreatment, described a four-to five-fold higher glucose yield thanks to this preceding step¹⁷³.

Afterwards, this biomass can be fermented in different configurations, whereof the most prevalent are separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), or consolidated bioprocessing (CBP) (Figure 1.10).

Separate hydrolysis and fermentation (SHF, green strategy) is a two-step process wherein biomass is first hydrolyzed by adding an externally produced enzyme mixture and in a next vessel fermented to ethanol. Additionally, the pentoses in the liquid fraction of pretreatment can be separately hydrolyzed and fermented, albeit not always included¹⁷⁵. The hexoses and pentoses can also be fermented simultaneously in a variant process called separate hydrolysis and co-fermentation (SHcF). This is usually performed by engineered strains, including the desired pentose pathways¹⁷⁶. The main advantage of SHF/SHcF is that both steps can be performed at optimal conditions. Simultaneous saccharification and fermentation (SSF, brown strategy) on the other hand, combines both steps in one vessel. The main benefit of this set-up is the enhanced hydrolysis because the hydrolysis end-products, that can inhibit cellulases, are immediately removed. However, the compatibility of hydrolysis and fermentation conditions is of major importance. Thermotolerant organisms can bring

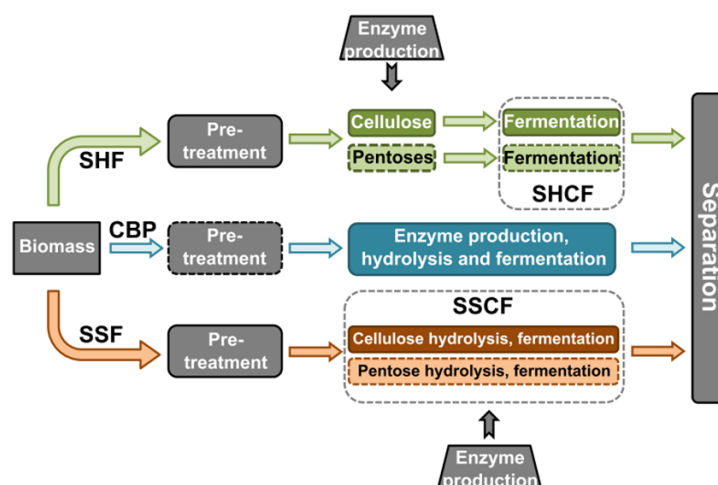


Figure 1.10: Different processing strategies for production of bioethanol (2nd generation) from plant material. Three main routes exist: Separate hydrolysis and (co-)fermentation (SH(C)F), Simultaneous saccharification and (co-)fermentation (SS(C)F) and Consolidated bioprocessing (CBP). The dashed lines around the pentose route, implies this is not always performed and furthermore, a pretreatment in CBP would ideally not be necessary by using organisms that can reduce the recalcitrance during hydrolysis. (Figure from:¹⁷⁴)

a considerable advantage to close the gap between the optimal growth temperature of the organism and the optimal catalytic temperature of the enzyme¹⁷⁴. Likewise, hexoses and pentoses can again be co-fermented in a process called SS_{Co}F¹⁷⁷.

From an economical point of view, integrating different steps is thought to result in lower costs¹⁷⁸. Therefore, consolidated bioprocessing (CBP, blue strategy) encompasses a one-vessel process without addition of externally produced enzymes^{179;180}. Two possible routes exist to create the perfect organism for CBP. On one hand, the native cellulolytic strategy makes use of naturally cellulolytic organisms that are engineered towards higher titers and higher yields. On the other hand, the recombinant cellulolytic strategy involves organisms that also have high production yields and high titers, but need engineering to include expression of cellulolytic enzymes. For this strategy, mostly organisms are selected that are well-known considering genetic background, information and genetic tools. This usually turns out to industrially used organisms such as *Escherichia coli*, *Saccharomyes cerevisiae*, *Bacillus subtilis*¹⁸¹ and *Corynebacterium glutamicum*¹⁸².

1.4.2 Rethinking the process for LPMO integration

It is clear that LPMOs have potential in lowering the enzyme loadings and/or process time by adding them to commercial cellulase mixtures. The importance therefrom lies in the fact that enzyme cost is still underestimated and is one of the main factors to overcome in making the process economically interesting⁴⁶. Most studies are performed at labscale,

using pure cellulose, whereas this differs from real process conditions such as using real biomass, at high dry matter content^{119;120}. Furthermore, the process configuration needs serious rethinking since LPMOs, which are oxidative enzymes, have different needs than the classical cellulases. The most important concerns to tackle are (1) the need for oxygen (2) the need for an electron donor¹⁸³ and (3) the production of non-fermentable oxidized products¹⁸⁴.

As early as 1974, Eriksson and coworkers described experiments with cell-free extract of 4 different fungi, that revealed a beneficial effect on glucose yields in aerobic conditions¹⁰². At the time, the exact reasoning was unclear, whereas we now know that LPMOs are the key answer. Indeed, the improvement of cellulase cocktails by LPMO addition only appeared when oxygen and a suitable electron donor were provided¹⁸³. This premise has implications on the entire process. Before LPMO discovery, SSF seemed to be the most suitable process. However, now that LPMOs are added, SHF appeared to have a 20 % higher glucose yields (or a comparable yield, while 30% reduction in enzyme loading)¹⁸⁵. This difference is caused by the competition for oxygen by the yeast and the LPMO in the SSF set-up. Alternatively, in the anaerobic process of lactic acid bacteria producing lactic acid from biomass, SHF also seems to yield up to 30% more lactic acid despite the incapability of respiration of these bacteria. LPMOs can be activated by oxygen in SSF, while this oxygen supply shifts the metabolic pathways towards acetic acid production. An inverse result was described by Zhang and coworkers, presumably because they treated the substrate different after the initial hydrolysis step¹⁸⁶.

As a cellulose wrapping component, lignin is long believed to be a disturbing factor during biomass degradation. It not only impairs swelling of the cellulose fibers during pretreatment, resulting in reduced accessibility for cellulases to work on^{90;187}, but also causes unproductive binding¹⁸⁸. This view was recently revised, since Dimarogona and coworkers showed that lignin could serve as electron donor for LPMO activity¹¹⁸, after earlier work that showed involvement in redox cycles¹⁰³. This was confirmed in later industrial-like configurations, adding no other electron donor than lignin^{183–185}. Interestingly, the pretreatment method applied, can not only influence the amount of lignin left, but also the extent of LPMO activation¹²².

Every LPMO intervention creates a neutral and an oxidized chain end. The exact ratio depends on the process conditions such as substrate, pretreatment and cellulase cocktail composition, although a relative low prevalence for the oxidized sugars is observed in comparison to glucose. For example, ~ 4 % of the glucose release appeared as gluconic acid¹⁸⁴ and <1 % as Glc4gemGlc (C4-oxidized cellobiose¹⁸³). Why the LPMOs seem to be so slow, is still to be elucidated. The yield of these by-products has to be taken into

consideration for designing an ideal process. These oxidized sugars might cause problems because gluconic acid for example inhibits β -glucosidases more than glucose does¹⁸⁴ and the enzyme works 10x slower on cellobionic acid as on cellobiose. Furthermore, the industrial important yeast *S. cerevisiae* can not metabolize gluconic acid and thus this results in a loss of fermentable sugar. Therefore, a recent study describes the catabolic pathway transfer for cellobionic acid utilization from *N. crassa* to *S. cerevisiae* (Figure 1.11)¹⁸⁹. Earlier, a pathway for cellodextrin¹⁹⁰ and hemicellulose (xylodextrin)¹⁹¹, also from *N. crassa*, were already successfully transferred. The pathways for utilization C4-oxidized products are still unknown, but can be expected in cellulolytic fungi. Additionally, Cannella and coworkers found that less gluconic acid was formed during SHF, where hydrolysis is performed at 50 °C, than during SSF, performed at 30 °C.

Furthermore, the oxidative reaction results in formation of toxic by-products such as hydroxyl radicals and reactive oxygen species, that inactivate cellulases. Therefore, addition of catalase to the mixture would increase the proteins inactivation time¹⁶⁹.

These 3 issues describe the problems that are to be handled when using LPMOs, while they underpin at the same time the importance of not only optimizing one step, but integrating the complete process in order to find the optimal process configuration¹⁹².

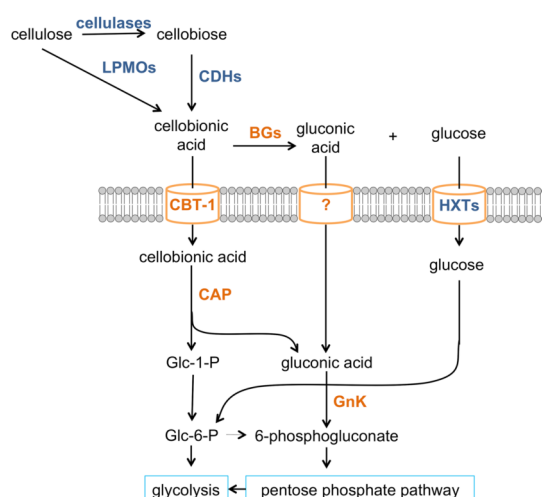


Figure 1.11: Catabolic pathway for cellobionic and gluconic acid consumption in *N. crassa*, and transferred in *S. cerevisiae*. BGs: β -glucosidases, CBT-1: cellobionic transporter, HXTs: hexose transporters, CAP: cellobionic acid phosphorylase, GnK: gluconokinase (Figure from¹⁸⁹)

1.5 Protein Engineering

Enzymes are increasingly used as biocatalyst in industrial applications because they offer a green alternative for conventional chemical catalysts. Indeed, enzymes are sustainable and produce less waste and (toxic) side products while they require less energy and raw material. Moreover, they offer some advantages such as high catalytic turnover (>500 / min) and high selectivity (chemo-, stereo and regioselectivity)⁶¹. However, natural enzymes usually don't meet the requirements for an efficient industrial processes, *i.e.* combine a high stability, activity and selectivity. To overcome these issues, often enzyme engineering techniques are applied to push the biocatalysts beyond their natural limits.

1.5.1 General principles

Iterative rounds of mutagenesis

Irrespective of what characteristic is aimed to improve and what engineering technique is applied, a general principle of iterative rounds of mutagenesis is followed (Figure 1.12). First, changes are made at DNA-level to generate diversity. Next, these genes have to be expressed in a suitable host where after the so-called library is screened or selected for the desired characteristic to identify improved variants. These steps are usually repeated in an iterative way until the optimized enzyme is reached or until no ameliorations are obtained anymore.

The last step, identification of improved mutants is the most challenging. The method should be specific for the characteristic one wants to improve and sensitive to identify variants with only small improvements. Two different approaches exist: *in vivo* selection and *in vitro* screening^{61;193}. The first, selection, applies the Darwinian principle of 'survival of the fittest'. This methodology is typically used when the improved variant could give an advantage to the expression organism, such as for example use of a specific carbon source. After transformation of the library into the expression host, the mixture is grown on minimal medium, containing that specific carbon source. Strains with improved variants will overgrow the ones with less functional variants. The improved variants can subsequently be enriched by repetitive inoculation in fresh medium. Another well-known example of selection is the use of an antibiotic, where only the strains that contain a resistance gene can survive. Screening on the contrary, tests each variant separately for the desired reaction. After transformation of the library in the expression host, single colonies are individually transferred to liquid medium and tested. The disadvantage is thus that also inactive variants (>50 %) can not be excluded on beforehand and are assessed. Typically colorimetric methods, based on chemical or enzymatic reactions, are used as high-throughput screening method (HTS). In recent years, ultra-high throughput screening methods (uHTS) using

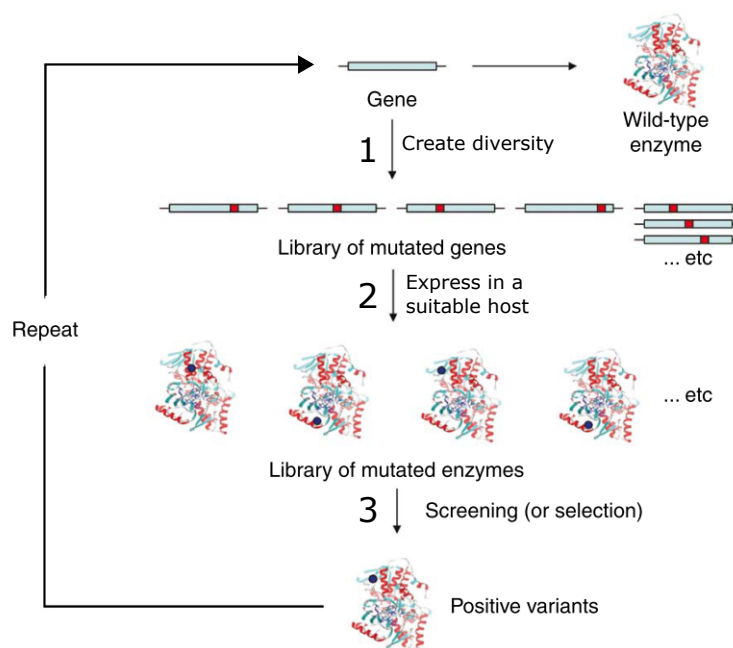


Figure 1.12: General principle of iterative rounds of mutagenesis, expression and screening/selection to create an optimized enzyme. (Adapted from:⁶¹)

microfluidics are gaining attention¹⁹⁴.

The fitness landscape

Even though the number of possible amino acid combinations (sequence space) is astronomical high, only limited combinations lead to folded proteins, let alone to functional entities¹⁹⁵. Nature counts 20 (most abundant) different amino acids, so that for a protein of ~ 300 residues (such as LPMOs) 20^{300} sequences can be built. Taken into account that the entire universe 'only' counts between 10^{70} and 10^{80} molecules, only a very limited number of combinations are indeed prevalent in nature⁶¹. Regarding that screening efforts are mostly limited to 10^6 - 10^8 by practical issues, only a negligible fraction of sequence space can be screened/selected at each round¹⁹⁶. Nature solves this problem by taking only limited searches, which radiate from a parental enzyme and only retains the fittest enzymes. The fitness of a certain enzyme variant is measured under a set of selection conditions and is typically a function of several properties such as substrate affinity, active site accessibility, protein stability and catalytic mechanism¹⁹⁶. The fitness landscape then represents the entire sequence space and its according fitness. This usually looks like a rough mountain chain with peaks and valleys. Enzyme engineering aims to climb uphill,

either in small steps (single mutations) or large steps (recombinations or multiple simultaneous mutations), which can not only reach local maxima but also distinct neighboring maxima.

While walking the fitness landscape, trade-offs between different properties are known to occur. Enzyme stability is for example often compromised when mutations for activity improvement are accumulated¹⁹⁷. Therefore, an enzyme is often first stabilized before the activity is altered¹⁹⁸ and/or rounds of stability improvements are introduced in between activity alteration rounds^{199;200}. Also, the variants can be evaluated on both activity and stability in order to retain only those mutations that solely improve activity without stability trade-off.

Numerous methods exist today to introduce sequence diversity. The type of mutagenesis corresponds largely to the type of protein engineering, as further explained below. In general, random methods correspond to directed evolution, whereas targeted methods include rational design.

1.5.2 Directed evolution

Directed evolution bundles all molecular biology techniques that mimic natural, Darwinian evolution in a test tube, although for one specific gene rather than for a whole genome²⁰¹. The 'fittest' variants are in this approach subjected to a subsequent round of mutagenesis¹⁹⁶. This system is highly comparable to the way Nature takes its 'evolutionary walks'. Since random variability is introduced, no knowledge on the mechanism, nor a crystal structure are required.

Two main types of library generation can be distinguished, *i.e.* asexual or sexual. In asexual evolution, one parental gene (template) is subjected to random mutagenesis such as for example UV-radiation or error-prone PCR (epPCR). The error rate of a polymerase can be increased by addition of $MnCl_2$ or the use of unbalanced dNTPs. However, the diversity suffers from bias caused by the difference in mutations it takes to change the codon for another residue (*e.g.* AGT and AGC both code for serine). Sexual evolution on the contrary, takes homologous enzymes or enzyme variants as starting point that are randomly recombined. The most used method is called DNA shuffling²⁰². An advantage of this technique is that a more complete coverage of the sequence space is reached. More advanced methods are being developed to enhance the random mutagenesis methods such as for example proSAR. It stands for PROtein Sequence Activity Relationships and sorts single mutations in three categories (beneficial, neutral or deleterious) as a training set. Only the beneficial mutations are kept for subsequent rounds²⁰³.

Although the ease of library generation, the biggest bottleneck is the large libraries to screen, associated with limited success rates. The screening method is extremely impor-

tant here and needs to be specific for the desired characteristic. The first law of directed evolution therefore states 'You get what you screen for'^{193;201}. Moreover, when altering the specificity, one can not introduce a completely new functionality. This means to be successful, the starting point should already perform, albeit at a very low level, the desired reaction²⁰⁴. In that respect, the second rule of directed evolution is stated as 'you should select for what is already there'.

Directed evolution has led to many successes, for example significant increases in thermostability and pH stability of a number of xylanases^{205;206} an increased thermostability of an esterase have been obtained²⁰⁷.

1.5.3 Rational design

In contrast to directed evolution, a profound insight into the structure-function relationships is required, which includes the availability of a 3D-structure, mechanistic knowledge and/or knowledge on the determinants for substrate specificity. In particular for stability amelioration, the mechanism of deactivation is a considerable asset to target specifically the most labile regions. Guided by this understanding, whether or not aided by computational techniques, specific amino acid substitutions are proposed and introduced via site-directed mutagenesis (SDM). This technique is also regularly used to prove the inverse: residues that are believed to be crucial are mutated in order to prove their importance and specific role²⁰⁸.

1.5.4 Semi-rational design

Any combination of structural information with random elements is known as semi-rational design. This methodology involves first the targeting of a *hot-spot* via structural insights or any random approach, often followed by site saturation mutagenesis (SSM). Different rounds of mutagenesis are usually performed, known as iterative saturation mutagenesis (ISM), which reduces the molecular work and screening effort significantly as compared to directed evolution²⁰⁹. Libraries are created by use of mutagenic primers with degenerate codons such as NNS or NNK (32 codons for all 20 aminoacids) or NDT (12 codons for 12 different residues)²¹⁰. Libraries can be made for one position or for different positions at the same time to allow synergistic mutations.

Three different approaches can be discriminated, namely based on structure, sequence or computational design.

Structure-based approach

The first, structure based approach, links protein structure with protein functionality. A well-known example hereof is combinatorial active-site saturation testing (CASTing)²¹¹. Because typically 10-15 residues are in actual contact with the substrate (first coordinating sphere) and 20-30 residues are in indirect contact (second coordinating sphere), alteration of the enantioselectivity and/or substrate specificity often benefits from targeting these residues only⁶¹. Therefore, CASTing arranges active site residues in groups of 2 to 3 that are saturated simultaneously²¹¹. A close variant of this technique for stability improvement, B-factor iterative test (B-FIT), targets residues with high crystallographic B-factor. Since a B-factor reflects the thermal displacement, a high B-factor is correlated with high flexibility²¹².

Sequence-based approach

Second, sequence based methods use the diversity that billions of years of natural selection have provided us. The diversity of functional homologous sequences can be studied using a multiple sequence alignment (MSA) and phylogenetic analysis. Even though these sequences hold highly valuable information, the success of actually extracting this information strongly depends on the quality of the alignment, which is especially challenging in variable loop regions. One example of a technique hereof is consensus engineering, where in each amino acid is replaced by the most frequently occurring amino acid in an MSA of homologous sequences²¹³. This approach has been successfully used for increasing the stability. The underlying assumption is that a more frequent residues would contribute to protein stability^{214;215}. The concept can be either used to create new full-length proteins^{216;217} or to create libraries of a protein by targeting specific positions^{218;219}. Other approaches are the construction of ancestral sequences for altering substrate specificity and increasing thermostability. The first is based on the assumption that the parent enzyme would have a higher promiscuity (broad substrate specificity) and the second on the assumption that ancient life was hotter²²⁰. Moreover, correlation networks of residues that tend to change together, can be extracted from the MSA²²¹. Those networks are believed to be directly or indirectly related to protein function or stability.

Computational design

Last, the possibilities of computational design are ever increasing with the fast development of powerful computers. They allow to numerically solve chemical problems with increasing size and complexity²²². Homology modeling and ligand docking are now considered as standard protein engineering tools, although a good scientist is still required to interpret and adjust the results. Some prominent examples of rational protein design are ROSETTA and SCHEMA. The software package ROSETTA is developed for structure prediction and

functional design²²³. One of the most striking possibilities is to model the active site and transition state of the reaction so that a novel biocatalyst can be created from an existing template structure. SCHEMA on the other hand aids in sequence recombination for the creation of chimeric proteins by identification of blocks in the sequence with minimal structural disruption²²⁴.

1.6 Protein thermostability

1.6.1 The importance of protein stability

As stated in the former section, there is an increasing interest in the use of biocatalysts in industrial applications. Of course, the costs herein has to be minimized, resulting in the need for stable, selective and productive catalysts. Stability is often a crucial determinant for economical success and actually limits the widespread use²²⁵. Indeed, an enzyme with increased thermostability has a longer lifetime and can thus perform more turnovers during its life, described as an increased total turnover number (TTN)¹⁹⁸. Furthermore, often higher temperatures are preferred to (1) improve the reaction rate, (2) improve solubility of the substrate and/or product and (3) decrease the risk on microbial contamination. In addition, stable proteins are of high interest as template for mutagenesis strategies (*e.g.* to design a better catalyst). Stable proteins have indeed a higher mutational robustness and evolvability¹⁹⁹.

The stability of a biocatalyst is influenced by many factors such as temperature, pH, oxidative stress, binding of metal ions or co-factors and the presence of organic solvents or surfactants²²⁵. In applications such as production of fine chemicals and pharmaceuticals, tolerance to organic solvents is important and gains increasing attention²²⁶. The reason hereof is that this increased tolerance can allow for higher concentration of substrate and products and thus wider operating conditions. Furthermore, since the detergent industry is the largest application area of biocatalysts, the effect that surfactants have on stability is extremely important. Moreover, stability is sometimes only desired under specific circumstances. For example, in some specific cases therapeutic proteins should be stable under storage conditions (long shelf-life), whereas they should be relatively low stable *in vivo*, because the proteins can otherwise lead to toxicity and side-effects²²⁷. Of all forms, thermostability is the most studied one and discussed in this section. Nonetheless, different forms of stability seem to be interrelated, albeit not through an absolute correlation. For example, thermostability is often closely related to organic solvent stability²²⁸.

1.6.2 Thermodynamic versus kinetic stability

Thermal denaturation of proteins is thought to be at least a 2-step process, as described in the Lumry-Eyring model²²⁹.

$$N \leftrightarrow U \rightarrow F \quad (1.1)$$

wherein N stands for native state, U for unfolded state (inactive) and F for final permanently inactivated enzyme. The first step, $N \leftrightarrow U$, is assumed to proceed in both directions and

is therefore reversible. The second step, $U \rightarrow F$, is irreversible, meaning that it cannot return to the native, neither to the unfolded state. Different phenomena may cause this irreversibility such as aggregation, autolysis and chemical or covalent alteration of residues. This Lumry-Eyring model is the simplest model, although more complex situations can occur *e.g.* the unfolding step might involve a significant amount of intermediate states²³⁰. A more accurate model is thus to change the unfolded form (U) for 'non-native ensemble', which includes unfolded partially unfolded and misfolded proteins. In this view, there is at least one kinetic barrier, which separates the native from the non-native ensemble. Indeed, when denaturing conditions are applied, enzymes from the native state tend to unfold much slower than all intermediate states. Two reasons for this energy barrier are described. The first includes a 'safety mechanism' that guarantees substantial stability in an aggressive environment because the irreversible process cannot go faster than the reversible. The second reason is that the states that form the actual barrier have high free energy and are thus not significantly populated. Therefore, alteration from these states are disfavoured.

Based on this Lumry-Eyring model, two types of stability can be distinguished, *i.e.* thermodynamic and kinetic stability, that clearly refer to different processes²³¹. First, thermodynamic stability of an enzyme measures the tendency of a protein to reversibly unfold. This means that a denaturing influence alters the polypeptide chain so that it can no longer form the active site in order to perform the reaction, although, when the denaturing influence is removed, the unfolding is reversed and the activity restored²³². In practical terms, thermodynamic stability can be represented by the free energy of unfolding (ΔG_u), an unfolding equilibrium constant (K_u) or the melting temperature (T_m). Methods to obtain these parameters include differential scanning calorimetry (DSC), circular dichroism (CD), tryptophan fluorescence or changes in tyrosine absorbance²³¹.

Secondly, kinetic stability measures the tendency to irreversibly unfold and thus the length of time a protein remains active before undergoing irreversible denaturation (matches the second step from the Lumry-Eyring model). In practice, parameters describing this type of stability include the overall observed deactivation constant ($k_{d,obs}$), temperature at which the activity of the enzyme is reduced by half after a specified and fixed amount of time (T_{50}), and the most commonly used half-life of denaturation ($t_{1/2}$), which equals the time it takes to reach half of its activity at a specified and fixed temperature. Measurement of these parameters is mostly done via an enzyme activity test²³¹. Since kinetic stability relates directly to interesting applied parameters, such as shelf-life and half-life time for degradation, it is often more informative than thermodynamic stability.

1.6.3 Increasing protein stability

Numerous methods exist to increase protein stability. One important concern herein is the preservation of the catalytic center, since it is generally considered the most fragile part

of the enzyme, while crucial for activity²³³. Different methods exist such as addition of compounds to stabilize the protein, protein immobilization and enzyme engineering. These three strategies will be described in more detail below.

Additives

The oldest and most commonly applied type of stabilization is the use of additives, usually in quite high concentrations of 1 molar or higher²³⁴. It is a good approach to increase shelf-life or storage stability²³². The largest disadvantage of additives however is that additives have to be separated again from the final product²³¹.

Two categories have been described, namely charged and uncharged additives. The charged group, also referred to as ionic stabilizers, include salts and work through shielding of surface charges. The second group, also called osmolytes, include polyols, amino acids and polysaccharides. They mainly affect solvent viscosity and surface tension. Sugars and polyols in an aqueous solution lead to protein rigidification via strengthening of the hydrophobic interactions with non-polar residues. For example, the addition of glycol chitosan could increase the T_{50} of trypsin from 49 °C to 93 °C²³⁵.

Immobilization

One of the more recent non-biological ways to confer added stability to proteins is to link them covalently to solid or gel-like matrices, including agarose beads, porous glass, zeolite and epoxy resins. Immobilization provides structural protection against intermolecular process and external influences (proteolysis, aggregation, inadmissible organic solvents and oxygen)²³⁶. Positive effects from random immobilization is nonetheless not guaranteed, because even negative effects have been documented²³⁷. Several different methods have been tested. Most promising techniques proved to be adsorption, entrapment, membrane confinement and single or multiple covalent bonding²³². The most applied method is however the use of multipoint covalent attachment, which is based on fixation of critical positions of the enzyme to a stable surrounding. This is obtained through covalent binding of residues of certain amino acids to reactive groups on spacer arms on the carrier substance. Through locking these amino residues in to place, they retain their relative positions during denaturing conditions, resulting in a higher stability (e.g. 5-fold and 18-fold for chymotrypsin and penicillin G acylase respectively as reported by Mateo et al.). Furthermore, this allows the fixation of different subunits in multimeric enzymes, preventing inactivation through inter-subunit dissociation or unconstructive intra-subunit crosslinking²³⁷.

The number of enzyme-support bonds can be correlated with the degree of stabilization

obtained through immobilization. A successfully executed multi-covalent attachment has even led to stabilization factors up to even 30000-fold. The use of covalent bond enzyme also offers the additional advantages of an increased operational control and enzyme recovery through filtration²³². Activity studies however, have shown slight decreases after immobilization, with activity recoveries around 60% on average²³⁶. For increasing the thermodynamic stability of enzymes, the second most common technique is the entrapment of enzymes. When compared to the soluble enzyme of Horseradish peroxidase, an increase in activity of 100-fold and an increase in turn-over after 80 minutes of 230-fold was measured after entrapment in a copolymer²³¹.

Enzyme engineering

As described above, enzyme engineering is an increasingly used method to obtain biocatalysts with desired properties, such as high stability. Apart from directed evolution, (semi-) rational approaches have been very productive. Despite the numerous studies published, there is no general technique to improve thermostability of a protein. Stability not only differs for different proteins, it is also suggested to be fold specific²³⁸. Some underlying principles that form the base of these stabilizing strategies are modification of the protein scaffold to improve hydrophobic interactions or electrostatic-surface interactions^{239;240}, amelioration of the internal packaging²⁴¹, reinforcement of oligomerization state, inclusion of rigidifying mutations (*e.g. increase proline content*)^{242;243} and shortening of surface loops²³⁹.

1.7 Lytic polysaccharide monooxygenases and stability

After reviewing the literature on lytic polysaccharide monooxygenases (LPMOs, section 1.3 starting on page 19) and some basics on protein stability (section 1.6, starting on page 40), an overview is given of the rather limited current literature on LPMOs and their stability.

In 2013, Hemsworth and coworkers resolved an X-ray structure of *Bacillus amyloliquefaciens* AA10A²⁴⁴. They also found that the configuration of the His-brace, with binding of a suitable metal ion is crucial for LPMO stability. Indeed, they studied the effect of different metal ions (copper(II), nickel (II), manganese (II), zinc(II)) in the active site and found that the binding of copper to the LPMO is extremely strong with a low nanomolar dissociation constant. Furthermore, only Zinc(II) showed some measurable binding via isothermal titration calorimetry (ITC). To confirm these data, thermostability experiments were also performed: *BaAA10A* in presence of Cu(II) resulted in an apparent melting temperature (T_m) of 65 °C, while much lower temperatures of 53 °C, 52 °C, and 46 °C were obtained in the presence of Ni(II), Zn(II) and for the apo-enzyme respectively²⁴⁴. A subsequent study concludes that the stability of the LPMO greatly increases when adding substrate¹⁴⁴. Indeed, an increase of 8 °C in T_m is measured by adding chitin (its substrate) to the apo-enzyme, whereas no increase could be measured by adding cellulose and the T_m of a copper-containing *BaAA10A* increased with 3.5 °C. Meanwhile some more apparent melting temperatures of LPMOs of different families were determined (Table 1.4), ranging from 63 to 72 °C.

Table 1.4: Currently known apparent melting temperatures of LPMOs. DSC: Differential scanning calorimetry, DSF: Differential scanning fluorimetry, Intrinsic fluorescence measured at 295 and 345nm.

Family	Enzym	T_m (°C)	Method	Source
AA9	<i>Neurospora crassa</i> PMO-01867 (<i>NcLPMO9J</i>)	66.9	DSC	245
	<i>Neurospora crassa</i> PMO-02916 (<i>NcLPMO9C</i>)	63	DSC	245
	<i>Neurospora crassa</i> PMO-03328 (<i>NcLPMO9F</i>)	68.9	DSC	245
	<i>Neurospora crassa</i> PMO-08760 (<i>NcLPMO9E</i>)	67.9	DSC	245
AA10	<i>Bacillus amyloliquefaciens</i> CBM33 (<i>BaAA10A</i>)	65	DSF	246
	<i>Enterococcus faecalis</i> CBM33A (<i>EfAA10A</i>)	72	intrinsic fluorescence	142
	<i>Serratia marcescens</i> CBP21 (<i>SmAA10A</i>)	70.3		142
	<i>Aspergillus terreus</i> LPMO13A (<i>AtLPMO13A</i>)	70	DSC	133
AA13	<i>Magnaporthe oryzae</i> LPMO13A (<i>MoLPMO13A</i>)	70.9	DSC	133

In the search for more sustainable production of biofuels, alternative systems were explored such as the use of seawater during saccharification in biorefineries. Therefore, two LPMOs from the mangrove-associated fungus *Pestalotiopsis* sp. NCi6 were produced in *Pichia pastoris* and characterized. *PsLPMOA* and *PsLPMOB* were found to have a high stability and activity in the presence of up to 6 % of salt, presumably as a result of the high abundance of negative charges at the protein's surface²⁴⁷. The oxygen-reducing capacity,

measured via H₂O₂-production²⁴⁵, was around 3 U/g, which is comparable to results obtained for other terrestrial LPMOs^{125;245}. The authors were also the first (and so far only) to publish pH-dependent activities and temperature profiles for LPMOs. The salt-responsive LPMOs retained more than 50 % residual activity between pH 5.0 and 6.5, with an optimum at pH 5.5²⁴⁷. In addition, both *PsLPMOA* and *B* retained 100 % activity after 150 minutes at 30 and 40 °C. *PsLPMOA* showed a fast degradation at 60 °C, with an estimated half-life of 2 minutes, while *PsLPMOB* is still fairly stable at this temperature²⁴⁷.

Finally, as an alternative for new economically and environmentally friendly chemicals during biomass pretreatment, ionic liquids are proposed. A recent study that contained only *in silico* molecular dynamics simulations concludes that *ScLPMO10C* and *ScLPMO10B* show a very similar stability behavior in water as in ionic liquids so that ionic liquids would be a viable alternative²⁴⁸.

1.8 Conclusion

There is an increasing demand for renewable resources for the sustainable production of biochemicals and commodity chemicals in biorefineries. Lignocellulosic biomass is thought to be the most promising feedstock to fulfill these needs. Traditional models of enzyme-catalysed polysaccharide degradation involve the concerted action of synergistic endo- and exo-acting cellulases and β -glucosidases. Despite decades of research on these enzymes, the process has long seemed inscrutable due to the highly recalcitrant nature of the substrate. With the discovery of lytic polysaccharide monooxygenases (LPMO), classified as auxiliary activity families 9, 10, 11 and 13, the interest in lignocellulosic biomass has revived. LPMOs are believed to facilitate the action of canonical cellulases by loosening the crystalline packaging of the carbohydrate substrate (cellulose, chitin or starch) by a novel, oxidative reaction and thus render the process much more efficient. Despite the discovery of this new enzymatic function only 6 years ago, the research field is moving in a rapid pace. Our view on these enzymes is continuously broadened and it has become clear that our understanding is far from complete.

In biomass degradation industry, only families AA9 and AA10 are important because only these show activity on cellulosic substrate. LPMOs are considered an inevitable addition to the existing enzyme mixtures to make the degradation process economically interesting. They indeed have the potential to lower the costly enzyme loadings, shorten saccharification time or reduce the severity of pretreatment. Integrating LPMOs in an industrial process however, demands a serious redesign since LPMOs require oxygen and an electron donor, while they produce oxidized products that cannot be metabolized by the traditional fermentation organisms.

In industrial settings, stability is a crucial parameter that is still largely unexplored for LPMOs, yet required for fruitful application of LPMOs. In the field of the traditional cellulases, especially enzyme engineering strategies have been very productive to improve stability, while LPMOs still lack this experimental phase. In the early days, mimicking nature in directed evolution approaches was most often used while rational and semi-rational techniques are gradually arising. These have the advantage of lowering the cost, time and effort needed for screening or selection, whereas the availability of structure-function relationships are essential as starting material. In case of LPMOs, some site-directed mutagenesis studies have been performed to elucidate important residues for the catalytic reaction or for binding the substrate or electron donor. So far, no mutagenesis studies have been described for LPMO stability purposes. Therefore, this thesis will focus on this research area.

2

Expression of *Trichoderma reesei* Cel61A in *Pichia pastoris*

Part of this chapter was published as:

Tanghe M., Danneels B., Camattari A., Glieder A., Vandenberghe I., Devreese B., Stals I., Desmet T. (2015) **Recombinant Expression of *Trichoderma reesei* Cel61A in *Pichia pastoris*: Optimizing Yield and N-terminal Processing.** *Molecular Biotechnology*. 57(11-12): pp.1010-1017.

Abstract

The auxiliary activity family 9 (AA9, formerly GH61) harbors a recently discovered group of oxidative enzymes that boost cellulose degradation. Indeed, these lytic polysaccharide monooxygenases (LPMOs) are able to disrupt the crystalline structure of cellulose, thereby facilitating the work of hydrolytic enzymes involved in biomass degradation. Since these enzymes require an N-terminal histidine residue for activity, their recombinant production as secreted protein is not straightforward. Therefore, in this first chapter, we optimized the expression of *Trichoderma reesei* Cel61A (*TrCel61A*) in the host *Pichia pastoris* so that this basic system can be used in chapter 5 for stability engineering of the protein. The use of the native *TrCel61A* secretion signal instead of the α - mating factor from *Saccharomyces cerevisiae* was found to be crucial, not only to obtain high protein yields (> 400 mg/ L during fermentation) but also to enable the correct processing of the N-terminus. Furthermore, the LPMO activity of the enzyme is demonstrated here for the first time, based on its degradation profile of a cellulosic substrate. Finally, the same system was shown to be successful also for the expression of *Phanerochaete chrysosporium* GH61D.

2.1 Introduction

Lytic Polysaccharide MonoOxygenases (LPMOs) are a recently identified group of enzymes that assist in the breakdown of carbohydrate polymers like cellulose, chitin and starch by oxidative cleavage¹⁷. Although the first representatives were assigned to glycoside hydrolase (GH) families, they have recently been reclassified in auxiliary activity (AA) families^{16;163;249}. Family AA9 (formerly GH61) only contains eukaryotic LPMOs with activity towards cellulose as substrate^{13;118;147}. By disrupting the crystalline structure of cellulose, they facilitate the enzymatic cleavage of classical cellulases, which is expected to significantly boost the production efficiency of second-generation ethanol^{250;251}. The production of several AA9 members has already been described, either by their native hosts^{13;106} or using *Pichia pastoris* for heterologous expression^{113;118;141;245}. Nevertheless, every protein needs optimization of its own expression conditions²⁵².

All LPMOs have a similar active site architecture, where an N-terminal histidine is involved in binding of a copper ion. The correct processing of the protein's N-terminus is therefore crucial to obtain an active enzyme^{11;13}. In particular, the choice of the secretion signal may have a big influence, as this leader sequence needs to be cleaved off after the pre-protein has been targeted to the secretory pathway. In *Pichia pastoris*, the most commonly used secretion signal is the alpha-mating factor (α -MF) originating from *Saccharomyces cerevisiae*. However, many alternatives exist and whether or not the protein can be correctly processed depends on the compatibility of the secretion signal used and machinery of the host cell²⁵³.

The filamentous fungus *Trichoderma reesei* (*Hypocrea jecorina*) is one of the best-studied organisms in the field of biomass valorization^{21;35;254}. Besides a whole array of hydrolytic enzymes, it also produces two LPMOs of family AA9, namely *TrCel61A* and *TrCel61B*. The former consists of a catalytic domain connected to a cellulose-binding domain (CBM1) by a flexible linker, whereas the latter only harbors a catalytic domain, which is 49% identical to that of *Cel61A*¹⁰⁶. Before the discovery of LPMO as a new enzyme class, *Cel61A* was thought to have a very weak endoglucanase (EG) activity and was therefore referred to as EGIV. The enzyme has already been produced in both a heterologous host (*S. cerevisiae*, 1997²⁵⁵) and overexpressed from the native host (*T. reesei*, 2001¹⁰⁵). However, its LPMO activity has not yet been demonstrated.

In this chapter, the recombinant expression of *TrCel61A* in the yeast *Pichia pastoris* is described for the first time, resulting in the highest yields reported so far. This allowed us to confirm that *TrCel61A* indeed is a polysaccharide monooxygenase that generates oxidized oligosaccharides from cellulose substrate.

2.2 Results and discussion

2.2.1 Optimizing the yield of secreted protein

Many factors can influence the yield of secreted recombinant proteins²⁵². In this work, the evaluation was focused on (1) the type of promoter, (2) the codon usage and (3) the secretion signal on the expression yield of *TrCel61A* (Uniprot O14405). To that end, 6 different constructs were prepared and their expression yields were compared by SDS-PAGE analysis (Table 2.1). In the compared analysis of at least three biologic replicates, the *TrCel61A* protein was detected at a molecular mass of about 60 kDa (Figure S2.3), which is higher than its calculated mass of 35 kDa. Furthermore, the bands appear to be slightly diffuse, both due to glycosylation. Indeed, the protein contains 2 putative N-glycosylation positions and a Ser/Thr rich linker, prone to O-glycans. When adding endo-N-acetyl- β -D-glucosaminidase from *Trichoderma reesei* (EndoT) to remove N-glycans, the apparent molecular mass was only decreased to 50 kDa. This suggests an important contribution of O-glycans too, since the difference between the theoretical and observed mass is not completely explained by this experiment.

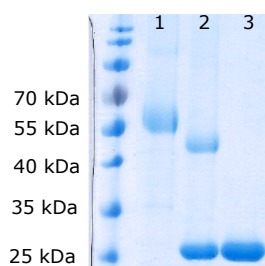


Figure 2.1: SDS-PAGE of *TrCel61A* and deglycosylation experiment. (1) *TrCel61A* (purified via Ni-NTA chromatography) (2) the same enzyme treated with EndoT for deglycosylation (3) EndoT only.

Although the methanol-regulated AOX1 promoter and the constitutive GAP promoter are supposed to be equally strong²⁵⁶, only a very low expression level could be detected for the GAP constructs in our case. This could be an indication that either the secretion pathway was not followed correctly or that the proteins ended in the ER-associated protein degradation (ERAD) pathway as described earlier in literature²⁵⁷. An increase in expression yield ($p=0.002$, Student t-test) was achieved by modifying the codon usage towards *Pichia pastoris* (sequences in supplementary material). Analysis from the rare codon cluster already indicates a large difference for the expression in the native host *Trichoderma reesei* or the recombinant host *Pichia pastoris* when using native codons and a clearly improved pattern for the codon optimized sequence (Fig. S2.1). Improvement in yields by codon optimization have already been reported previously for the recombinant expression of cellobiohydrolase and mannanase²⁵⁸. Interestingly, the codon-optimized native secre-

tion signal of *Trichoderma reesei* Cel61A in combination with the codon optimized gene was found to significant give better results ($p=0.001$, Student t-test) than the more commonly used α -mating factor from *Saccharomyces cerevisiae* (α -MF). At the end of the optimization process, the optimized construct yielded 70 ± 2 mg / L of secreted protein on microscale, which is about 4 times higher than the starting construct (Table 2.1).

Table 2.1: Relative expression levels of secreted *TrCel61A* by *P. pastoris*. Yields were relatively compared by setting the highest value for protein concentrations in the culture supernatant to 100 %, as quantified by digital imaging. The effect of promoter (column A), codon optimization (column B), and secretion signal (column C) was compared. ND not determined, since the obtained concentration was below the quantifiable range. (n=3, error is standard deviation)

Construct	Promoter	Codon optimization	Secretion Signal	Relative yield (%)
Inducible expression				
pPpT4-aMF-TrCel61A-nat	AOX1	No	a-MF (EKR-EA-EA)	26 ± 3
pPpT4-aMF-TrCel61A-CO	AOX1	Yes	a-MF (EKR-EA-EA)	41 ± 5
pPpT4-NSS-TrCel61A-CO	AOX1	Yes	Native secretion signal TrCel61A	100 ± 5
Constitutive expression				
pPpT4-GAP-aMF-TrCel61A-nat	GAP	No	a-MF (EKR-EA-EA)	ND
pPpT4-GAP-aMF-TrCel61A-CO	GAP	Yes	a-MF (EKR-EA-EA)	ND
pPpT4-GAP-NSS-TrCel61A-CO	GAP	Yes	Native secretion signal TrCel61A	ND

2.2.2 Optimizing N-terminal processing

For most proteins the exact amino-terminus is of little importance. However, some exceptions exist where the N-terminus is important for its biological function²⁵⁹ such as for example virulence²⁶⁰. This is also the case for LPMOs, where the correct removal of the secretion signal is absolutely essential, as the enzyme requires an N-terminal histidine for activity. So far, several LPMOs have been produced using the α -MF for secretion purposes in *Pichia pastoris*, but little attention has been paid to its processing^{118;136}. Therefore, N-terminal sequence analysis has now been performed on several constructs (Table 2.2). The α -MF sequence ends with the amino acid sequence EKR, which is recognized by the protease Kex2 in the Golgi apparatus. To increase the cleavage efficiency of Kex2, the signal peptide can be extended with 2 EA repeats, which are subsequently cleaved off by another Golgi-protease Ste13²⁶¹. However, neither of these constructs resulted in a correct N-terminus for *TrCel61A* (Fig. 2.2 and table S2.1 for exact position). The former was not recognized at all by Kex2, resulting in the entire pro-sequence still present in the protein. Although the addition of the EA repeats indeed increased the activity of Kex2, their removal by Ste13 was highly inefficient. As an alternative to the EA repeats, an enterokinase cleavable sequence (DDDDR¹⁴¹) was tested, to check whether an artificially introduced cleavage site could help to circumvent an inefficient biologic post-translational processing. Although this complicates the process by the introduction of an extra step, the result was satisfactory (Figure 2.2), as was also reported for the LMPO from *Phanerochaete chrysosporium*

GH61D by Westereng and coworkers¹⁴¹.

Table 2.2: Effect of secretion signal sequence on N-terminal processing. The different secretion signals that were applied are listed with their amino acid sequence and percentage correctly processed (HIS1) enzyme. The last column indicates prior use of the secretion signal.

Secretion signal name	Secretion signal amino acid sequence	Correctly processed (%)	Ref
<i>S. cerevisiae</i> Alpha-Mating factor	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYS DL EGDFDVAVL PFSNSTNNGLLFINTTASIAAKEEGVSLEKR	0	262
<i>S. cerevisiae</i> Alpha-Mating factor -EA-EA	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGY SDLEGDFDVAVL PFSNSTNNGLLFINTTASIAAKEEGVSL EKREAEA	0	262
<i>S. cerevisiae</i> Alpha-Mating factor -DDDDR	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYS DLEGDFDVAVL PFSNSTNNGLLFINTTASIAAKEEGVSLEKR DDDDR	100	141
<i>T. reesei</i> Cel61A	MIQKLSNLLVTALAVATGVVG	100	-
<i>P. chrysosporium</i> GH61D	MKAFFAVLAVVSAPFVLG	82	-
<i>P. pastoris</i> DDDK-protein	MFNLKTILISTLASIAVA	0	263

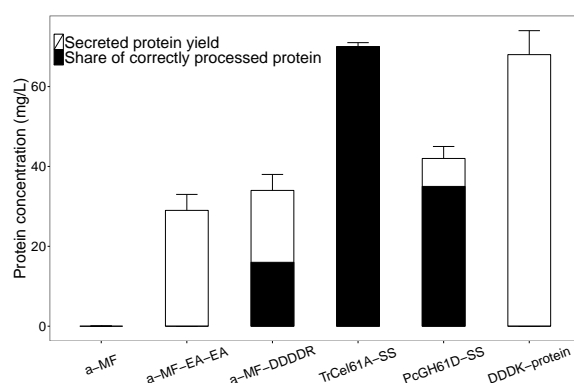


Figure 2.2: Effect of secretion signal on yield and processing. The expression level of the secreted protein (white) and corresponding share of correctly processed form (black) as determined by Edman degradation is given. (aMF is indicated with its different amino acids in the end, DDDK protein = secretion signal of DDDK protein, PcGH61A-SS = *Phanerochaete chrysosporium* GH61D native secretion signal, TrCel61A-SS = *Trichoderma reesei* Cel61A secretion signal) (n=3, error bars show standard deviation)

As a comparison, two related native LPMO secretion signals have also been evaluated, *i.e.* that of *TrCel61A* and that of *Phanerochaete chrysosporium* GH61D (*PcGH61D*). Interestingly, and despite both secretion signals are equally foreign for *Pichia pastoris*, the former is completely and correctly removed by the host, and is thus the preferred option in terms of both protein yield and N-terminal processing. In contrast, the latter only had a removal efficiency of 85% and also resulted in a product with a 2-fold lower expression yield. Finally, inserting an endogenous secretion signal from *P. pastoris* has also been

tried, as one would expect that the cellular machinery is optimized to recognize this signal. For example, the DDDK-protein signal sequence has been proposed as alternative for the α -MF²⁶³. Unfortunately, the N-terminus was found to be processed incorrectly in our case, and, more specifically by cleaving at position His3 in the protein.

2.2.3 Scale-up to fermentation and demonstration of LPMO activity

In order to get a better insight into the expression capabilities for our optimized construct (AOX1 promoter, codon optimized *TrCel61A* secretion signal and gene and native *TrCel61A* secretion signal), the production process was scaled-up to a 2L fed-batch fermentation. Since there is no direct, fast and quantitative activity measurement (yet) available for LPMO, the protein concentration was monitored during the fermentation. To that end, four different methods were compared: (1) Abs280 (Nanodrop 2000c), (2) Bradford protein assay, (3) Bicinchoninic acid (BCA) protein assay and (4) quantitative SDS-PAGE. The first three suffered from a high background, high variation and/or a low response to increases in protein content, most likely due to interference by the culture medium or glycosylation of the protein²⁴⁵. In contrast, quantification of the protein band on SDS-PAGE by digital imaging was found to be a more reliable method. After an induction phase of 96 hours, a protein concentration of 447 mg/L was obtained (Figure 2.3), which outperforms production in the natural host *Trichoderma reesei* by almost a factor four¹⁰⁵ and represents the highest expression level for *TrCel61A* reported so far. Although the enzyme was already expressed in *S. cerevisiae*, comparison was not possible since no yields were reported²⁵⁵. As a comparison, expression of 4 *Neurospora crassa* LPMOs in *Pichia pastoris* yielded 340 mg per liter medium.

More importantly, *TrCel61A* was shown to be active as LPMO by incubating the enzyme with phosphoric acid swollen cellulose (PASC) as substrate and ascorbic acid as electron donor. HPAEC (high performance anion-exchange chromatography, specialized HPLC) analysis showed the formation of cello-oligosaccharides in their neutral and oxidized form (Figure 2.4). Interestingly, the even-numbered aldonic acids (cellobionic acid and cellotetraonic acid; have a higher prevalence than the odd-numbered counterparts (celotronic acid and cellopentaonic acid, $p=0.000025$, Student t-test based on 10 biological replicates). This might suggest a preference in the cleavage mechanism for this enzyme. Furthermore, products appear in the region for C1-oxidized, C4-oxidized and double (C1- and C4-) oxidized products, which classifies *TrCel61A* as a type-3 LPMO¹⁷.

When optimizing the N-terminal processing, one notable characteristic has not been taken into account. Indeed, it is expected that all fungal LPMOs in nature possess an N- ϵ -methylation of the N-terminal histidine residue. In its native host, *Trichoderma reesei*, Cel61A is methylated, while this is not the case in the recombinant host *Pichia pastoris*.

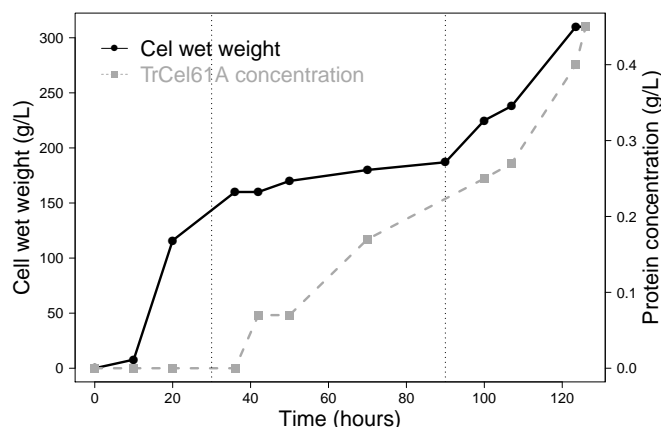


Figure 2.3: Fermentation parameters. Protein concentration (g/L) and cell wet weight (g/L) monitored and given for different time points. The methanol feed was initiated at 30 h (first vertical line). A second increase in cell wet weight was found after 90 h (second vertical line) due to pressure increase caused by foaming that resulted in fouling of the outlet filter.

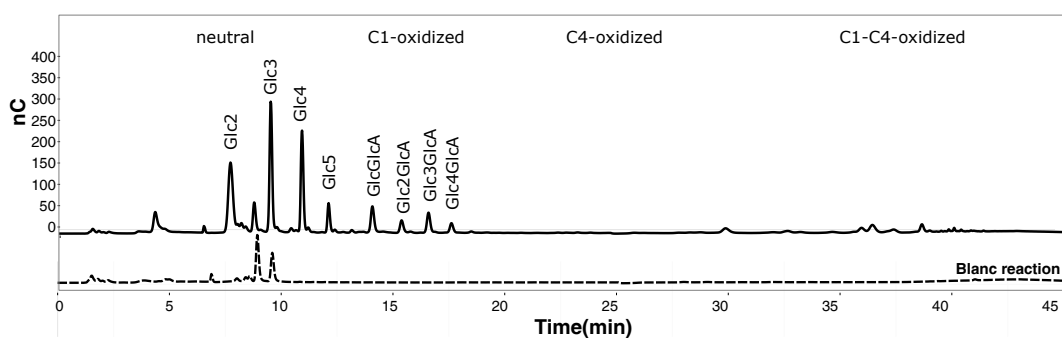


Figure 2.4: Enzymatic activity of recombinant *TrCel61A*. HPAEC-PAD profile of *TrCel61A* activity on PASC at 50 °C. The neutral oligosaccharides, cellobionic acid (GlcGlcA) and cellotrionic acid (Glc₂GlcA) were used as standard (Figure S2.5) whereas the nature of the other oxidized products was inferred from the literature²⁶⁴. The dotted line represent the blanc reaction (buffer was added instead of *TrCel61A*.)

Nevertheless, the enzyme clearly shows activity as an LPMO. This is in accordance to other recombinantly expressed LPMOs (*P. pastoris* or *E. coli*) found in literature^{109;125;141;146}. The function of this unusual post-translational modification is still unknown. The alkyl substituent may alter the electronic properties around the active site²⁴⁶, or explain the stronger Cu(II) affinity¹⁴⁰ and thus impact the enzyme-substrate electrostatic interaction²⁶⁵. As an alternative, it has been proposed that the methylation protects the LPMO from oxidative self-destruction¹⁴⁸. A recent study with *TrCel61A*, produced in its native host, describes also activity on PASC, but the C4-oxidized products are much more abundant (comparable to C1-oxidized products) than in our case. The authors suggest a possible explanation by the methylation²⁶⁶. A more elaborate comparison between the two enzymes would be highly

interesting.

2.2.4 Proof of concept: Expression of *Phanerochaete chrysosporium* GH61D with native secretion signal

Phanerochaete chrysosporium GH61D (Uniprot code H1AE14) was first shown to be a type-1 LPMO by Westereng and coworkers¹⁴¹. They expressed the protein in *Pichia pastoris*, using an enterokinase cleavable part to ensure the HIS1-condition (recognition sequence DDDDR, as used above). Given this article, the outcome in terms of activity is known. As a proof-of-concept, it was therefore evaluated when using the same set-up as *TrCel61A*: its own codon optimized native secretion signal, AOX1-promotor and codon optimized gene-sequence.

PcGH61D is in contrary to *TrCel61A* a single domain protein, *i.e.* only having a catalytic part and no glycosylated linker nor a cellulose-binding-domain. The protein has a theoretical size of 26 kDa, but appears at higher size because of glycosylation (2 putative N-glycosylation positions N-191 and N-221¹³⁵). The obtained yield was 36 ± 16 mg/L for growth on microscale. Westereng and coworkers described a yield of 50 mg/L¹⁴¹. This is within the error margin of our obtained result.

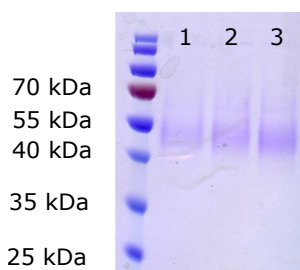


Figure 2.5: SDS-PAGE of *PcGH61D*. Three replicates are shown, grown on microscale.

Secondly, the activity on PASC was measured, resulting in a profile characteristic for a type-1 LPMO, as earlier described¹⁴¹. The detectable range of neutral glucose chains has a degree of polymerization (DP) up to 6 and aldonic acids display a DP up to 7 (Figure 2.6). Here cellotetraonic acid (DP4) is the most prevalent product, while cellosextaonic acid (DP 6) was the most formed product in the former study, despite using the same conditions.

Nonetheless, the use of the expression system was shown to result in an active protein, wherein the use of the native signal sequence lead to fruitful expression of the protein.

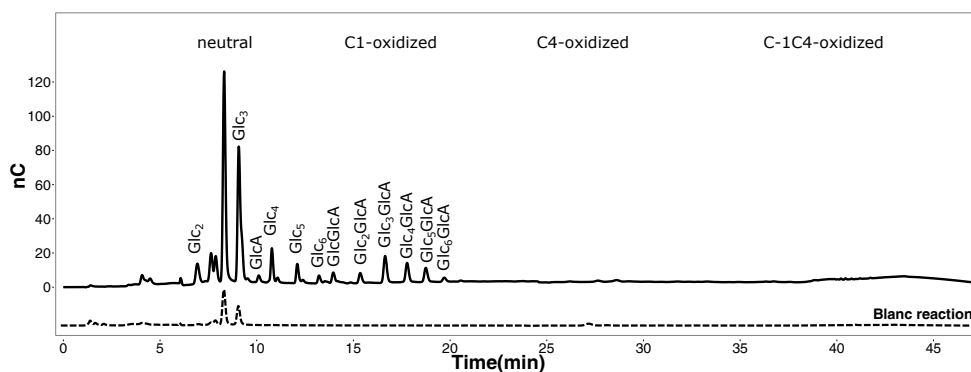


Figure 2.6: HPAEC-trace for *PcGH61D*, incubated on PASC. The dotted line represents the blanc reaction, wherein buffer was added instead of *PcGH61D*.

2.3 Conclusion

The yeast *Pichia pastoris* was found to be a suitable expression host for the LPMO Cel61A from *Trichoderma reesei*, with a protein yield during fermentation of > 400 mg/L when using the AOX1 promoter, a codon optimized gene and the protein's native secretion signal. Considering the importance of an N-terminal histidine residue, we here report that the processing of the native secretion signal is much more accurate than that of the more commonly used alpha-mating factor from *Saccharomyces cerevisiae*. Furthermore, our results demonstrate for the first time that *TrCel61A* is an active LPMO that generates both neutral and oxidized cello-oligo-saccharides from PASC as substrate. Finally, *PcGH61D* was also expressed in *Pichia pastoris* with its native signal sequence and appeared as an active type-1 LPMO. This suggests the native secretion signal can be a valuable option in different cases.

2.4 Materials and methods

2.4.1 Biological materials

The cDNA coding for *Trichoderma reesei* Cel61A (GenBank Y1113) was isolated from *T. reesei* QM6A (MUCL 44908) via total RNA extraction (RNeasy Mini kit, Qiagen) followed by cDNA production (Superscript III first strand synthesis kit, Invitrogen). This gene was cloned in pJET1.2 plasmid (Life Technologies) for further use. A codon optimized secretion signal and gene-sequence were ordered from GeneArt Gene Synthesis (Thermo Fisher Scientific). *Pichia pastoris* strain CBS7435 and all plasmids (pPpT4 plasmid variants, described by Näätsaari²⁶⁷) were kindly provided by the institute of Molecular Biotechnology at TU Graz, Austria.

2.4.2 Molecular work for *TrCel61A*

The molecular constructs were completed in *E. coli* cells (Transformax™ EC100™ Electrocompetent *E. coli* cells from Epicentre). The required primers were ordered from Integrated DNA Technologies (IDT). In all cases the constructs were integrated in the pPpT4 vectors downstream of the promoter (GAP or AOX1) and included an N-terminal secretion signal followed by *Trcel61A* gene with a His6-tag directly attached to its C-terminus (Figure 2.7). The constructs described in the section 'Optimizing the yield of secreted protein' were cloned by restriction-ligation using *XhoI*, *EcoRI* and *NotI*, using Fast digest enzymes (Life Technologies). All other variations of the secretion signal were constructed by various molecular techniques such as the method developed by Sanchis to eliminate a few base-pairs²⁶⁸ and Gibson assembly²⁶⁹ to insert *TrCel61A* preceded by another secretion signal in the pPpT4 backbone. More detailed information on the molecular constructs can be found in Supplementary material table S2.2 .

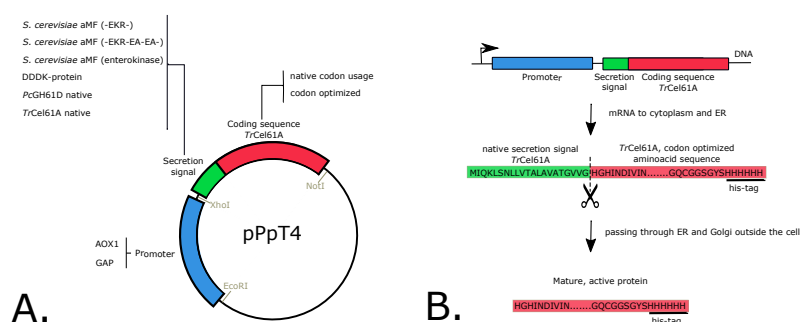


Figure 2.7: Molecular background of cloning strategies. (A.) Representation of the vector with the relative positions of promoter, secretion signal, and coding sequence of *TrCel61A*. All evaluated options are indicated in a list. (B.) Short schematic overview of the different steps going from DNA sequence till mature active protein.

After confirming its correct sequence (LGC Genomics) in *E. coli*, the resulting plasmid was linearized and transferred into freshly prepared electrocompetent *Pichia pastoris* CBS7435 cells²⁷⁰. Positive transformants were selected by incubating the transformation mixture at 30 °C for 48 hours on YPD-agar plates containing 100 µg/mL Zeocin. At least 5 colonies per transformation were grown on microscale (96-deep well plate, see further) and the supernatant was analyzed via SDS-PAGE.

2.4.3 Molecular work for *PcGH61D*

Phanerochaete chrysosporium GH61D (Uniprot code H1AE14) was codon optimized for *Pichia pastoris* and synthetically ordered from GeneArt (LifeTechnologies, cloned on pMA-T vector). The gene sequence was cloned in the pPpT4-vector containing *PcGH61D* native secretion signal already from the experiments with *TrCel61A* via Gibson assembly²⁶⁹, using the primers from table 2.3.

Table 2.3: Primers for cloning *Pcgh61d* in pPpT4-vector after its own native secretion signal. Bold: *Pcgh61d* sequence, underline: 6 x his-tag.

Template	Primer sequence (5' → 3')
pPpT4- <i>TrCel61A</i> - <i>PcGH61D</i> -SS	FWD: ACCAGCTGTTTGGCAAGGTCATCACCATCACCATCACTAGG REV: GAAGTCTGGGAAAGTGTAGTGTC CCACAACACCAGTAGC
pMA-T- <i>PcGH61D</i>	FWD: GCTACTGGTGTGTTGTTGGGAC ACTACACTTTCCAGACTTC REV: CCTAGTGATGGTGATGGTGATG ACCTTGCCAAACAGCTGGT

2.4.4 Media and growth methods

LB-medium was used for growing *E. coli* cultures containing 2% (w/v) agar, if required, and 25 µg / mL Zeocin for selection. The cultures were grown overnight at 37 °C while shaking at 200 rpm.

The standard medium for *P. pastoris* was YPD medium (1% (w/v) yeast extract, 2% (w/v) tryptone and 2% (w/v) glucose) containing 2% (w/v) agar, if required, and 100 µg/mL Zeocin for selection).

Selection of the best producing strain

After *P. pastoris* transformation with the destined vector containing the coding sequence of the desired enzyme variant, 5 colonies per construct were selected and grown on microscale to obtain the colony with best expression, as analyzed via SDS-PAGE on undiluted culture supernatant. If no expression was obtained, the experiment was repeated with 10 colonies each, followed by analysis on SDS-PAGE and western blot. Only the best producing strain for each variant was produced on larger scale and purified for stability and activity

analysis.

For growth experiments, a buffered minimal medium was used (composition in table 2.4). Microscale cultivation²⁷¹ was performed in 96-deep well plates (8x8x40mm with round bottom, Enzygscreen) that were sealed with a low-evaporation sandwich-cover (Enzygscreen). Plates were incubated at 28°C tilted under an angle of 25° shaking at 300 rpm. The plates were inoculated from a single colony on YPD-agar plate in 250 µL BMD1 medium. After 48-60 hours, induction was initiated by adding 250 µL BMM2 medium in the morning to each well and 50 µL BMM10 medium in the evening (at least 8 hours in between). Next, the induction was kept at 2 times 50 µL BMM10 medium a day. After 3-5 days induction, the cultures were harvested by centrifuging 5 minutes, 2000xg at 4 °C.

Table 2.4: Medium composition of buffered minimal medium for growing *Pichia pastoris*. BMD1: Buffered Medium with 1 % Dextrose, BMM2: Buffered Medium with 2 % Methanol, BMM10: Buffered Medium with 10 % Methanol.

Compound	BMD1	BMM2	BMM10
Sodium phosphate buffer, pH 6 (M)	0.200	0.200	0.200
Biotin (%)	$4 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	$4 \cdot 10^{-5}$
Yeast nitrogen base (YNB) (%)	1.34	1.34	1.34
Glucose (%)	1	-	-
Methanol (%)	-	1	5

Enzyme production and purification

The best producing strain for each enzyme variant was grown in 250 mL (unbaffled) shake flasks at 30°C by 200 rpm shaking. Initially the culture was started in 45 mL BMD medium, followed by induction after 48-60 hours by adding 5 mL BMM10 medium. Subsequently methanol supply was kept at 2 shots of 0.5 mL MeOH a day. After 3 to 5 days induction, cultures were harvested by centrifuging at 1500 rpm for 15 minutes (4°C).

The culture supernatant was subjected to ultrafiltration using a stirred bioseparation cell (Amicon) and 10 kDA ultrafiltration disks (Merck Millipore). The 50 mL culture supernatant was filtered and subsequently washed 3 times with 50 mL 50 mM Sodium acetate buffer pH5 and harvested in 2-3 mL final volume. If found necessary, a subsequent his-tag purification was performed as described in chapter 3.

Secondly, His₆-tagged proteins (C-terminal attachment) can be purified in an optional second Ni-NTA Chromatography step. Therefore, 1.5 mL of the nickel NTA-agarose slurry (MC-lab) was added to 10 mL purification columns. After washing with 6 mL of sterile distilled water, the resin was equilibrated with 3 x 8 mL native binding buffer (PBS buffer = 50 mM sodium phosphate and 300 mM sodium chloride at pH 7.4), supplemented with 10 mM imidazole. The sample was adjusted to 10 mM imidazole concentration and it was allowed

to bind to the resin by gentle rotation of the sealed columns while incubating at 4 °C for 30-60 minutes. The resin is settled by gravity and the suspension is aspirated. Next, the resin was washed with 3 x 8 mL of BPS buffer (pH 7.4) containing 20 mM imidazole in PBS. Elution was performed with 10 mL of PBS buffer (pH 7.4) supplemented with 250 mM imidazole. Hereafter, the buffer was exchanged to 50 mM Sodium-acetate buffer pH 5 in Vivaspin filter columns (10000 MWCO PES, Vivaspin, Sartorius).

2.4.5 Fermentation and purification

TrCel61A was fermented in a 2l fermentation vessel (Biostat B, B. Braun Biotech) starting with one liter BSM culture medium, as described by De Winter et al.²⁷². Methanol feed was initiated at a cell wet weight volume of 160 g/L (30 hours batch phase) and was gradually increased to 2g/L/h during 6 hours and kept at this rate for 90 hours. After 60 hours of induction, the pO₂ feed was increased to maintain a dissolved oxygen percentage of 50%. The protein was obtained by centrifuging the fermentation broth for 10 minutes at 10.000 rpm. The resulting supernatant was filtered using depth filtration, followed by cross-flow filtration and buffer exchange for 50 mM sodium acetate buffer pH 5.2 (Vivaflow 200, Sartorius).

2.4.6 Protein technology

SDS-PAGE

Sodium dodecyl sulphate - Polyacrylamide gel electrophoresis or shortly stated SDS-PAGE, was used to check the presence of our desired protein as well as the purity of the samples (culture supernatant or purified proteins)²⁷³. The required gels were made in-house (12 % resolving and 5 % concentrating gel, composition in table 2.5).

Table 2.5: Composition of stacking and resolving gel for preparing 2 gels for SDS-PAGE.

Component	5 % stacking gel	12 % resolving gel
dH ₂ O (mL)	2.85	3.4
Acrylamide (30 %) (mL)	0.85	4
Tris-HCl (0.5 M, pH 6.8) (mL)	1.25	0
Tris-HCl (1.5 M, pH 8.8) (mL)	0	2.5
SDS (10%) (μL)	50	100
APS (10%) (μL)	50	50
TEMED (μL)	5	5
Total volume (mL)	5	10

To perform SDS-PAGE, the gels were placed in the gel electrophoresis unit (Mini- Protean Tetra Cell, Biorad) and submerged in running buffer (0.303 % Tris base, 1.44 % glycine and 0.1 % SDS). Before samples could be loaded, sample preparation required the production of Laemmli buffer (10 % β-Mercaptoethanol, in 62.50 mM Tris-HCl pH 6.8, 26.3 % glycerol,

2 % (w/v) SDS and 0.01 % (w/v) bromophenol blue²⁷³). Ten μL sample buffer was mixed with 20 μL sample and was boiled at 95 °C for 3-5 minutes. For a sample, 12 μL of this mixture or 4 μL of ladder (PAGERuler Prestained protein ladder, ThermoScientific) was then pipetted into the wells and the electrophoresis was performed for 1 hour at 200 V.

The gel was removed from the spacer plate and was stained with the QC Colloidal coomassie stain from Biorad following the manufacturer's instructions.

Western blot

During Western Blot, the proteins are transferred to a Hybond ECL nitrocellulose membrane (Amersham). The blotting sandwich is covered with ice-cold CAPS buffer (10 mM CAPS, pH 11 and 10% methanol.) in a 1D-electrophoresis chamber (Biorad) and the transfer is performed during 60 minutes at 100 V. The blocking was done with 1 % milk powder in PBS buffer. The detection was performed using anti-polyhistidine antibody (produced in mouse, H1029, Sigma-Aldrich), subsequent anti-mouse antibody (produced in goat, A3562, Sigma-Aldrich) and finally NBT/BCIP.

Protein concentration

The protein concentration of the samples was estimated via SDS-PAGE. Together with a reference standard protein ladder, every gel was also provided with a reference protein, namely 0.1 mg/mL bovine serum albumin (BSA), to estimate the protein concentration. The intensity of the desired bands was measured and compared by digital imaging²⁷⁴, using ImageJ (Image Processing and Analysis in Java, available at <http://imagej.nih.gov/ij/>). Samples were diluted to fit in the linear range (0.050 - 0.200 mg/mL protein, Figure S2.4).

2.4.7 Protein deglycosylation

TrCel61A was treated with recombinant Endo-N-acetyl- β -D-glucosaminidase from *Trichoderma reesei* (EndoT) (kind gift from dr. Ingeborg Stals²⁷⁵). An amount of 2 μg EndoT was added to 100 μg LPMO in 50 mM sodium acetate buffer pH 5 and a total volume of 100 μL . This mixture was incubated overnight at room temperature.

2.4.8 Activity testing

To measure activity, ~ 40 μg of *TrCel61A* or *PcGH61D* was mixed with 1 mM ascorbic acid and 1.2% Phosphoric Acid Swollen Cellulose (PASC), prepared from Avicel PH-101 (Sigma-Aldrich) following instructions from Wood²⁷⁶. The volume was adjusted to 1 mL with 10 mM sodium acetate buffer pH 5.2 in Eppendorf tubes. The tubes were incubated overnight at 50 °C while shaking at 1400 rpm in an Eppendorf Thermomixer. The enzyme

was heat-inactivated by incubating the tubes at 95 °C for 10 minutes and the samples analyzed by HPAEC-PAD (ICS3000 CarboPac PA20, Dionex) by the method described by Forsberg et al.²⁶. Identification of some products was done by running an in house 'ladder' containing neutral glucose chains, gluconic acid, cellobionic acid and cellotronic acid (Fig S2.5).

2.4.9 Protein Sequencing

The N-terminus of the proteins was determined by Edman degradation. The proteins were blotted on a PVDF membrane and the desired bands were excised from the blot. Subsequently, the samples were analyzed using a 494 Procise protein sequencer (Applied Biosystems).

2.4.10 Statistical analysis

In order to evaluate if the average yields of two expression conditions were significantly different, a two-sample Student t-test was performed. The zero-hypothesis (H₀) states that there is no difference between the compared conditions. A probability (p-value) is obtained, reflecting the probability that H₀ is true. If $p < 0.05$, a difference is considered statistically significant.

Supplementary material

Codon optimized sequences

- *TrCel61A* native secretion signal, codon optimized for *Pichia pastoris*

ATGATTCAAAAATTGTCTAACTTACTTGTACTGCTTTGGCAGTTGCTACTGGTGTGTGGGA

- *TrCel61A* with C-terminal his-tag, codon optimized

CACGGTCACATTAAACGACATCGTTATCAACGGTGTGGTATCAGGCTTACGACCCAACTACTTTCC
CATACGAATCCAACCCACCAATCGTTGTTGGTTGGACTGCTGCTGATTTGGACAACGGATTGCTTTC
TCCAGACGCTTACCAGAACCCAGACATCATCTGTCACAAGAACGCTACAAACGCTAAGGGTCACGCT
TCTGTTAAGGCTGGTGACACTATTTTGTTCAGTGGGTTCCAGTTCCATGGCCACATCCAGGTCCTA
TCGTTGATTACTTGGCTAACTGTAACGGTGACTGTGAGACTGTTGACAAGACTACTTTGGAGTCTTC
AAGATCGACGGTGTGGTTTGTGTCTGGTGGTGATCCAGGTAAGTGGGCTTCCGACGTTTTGATCT
CCAACAACAACACTTGGGTTGTTAAGATCCCAGACAACCTGGCTCCAGGTAAGTACGTTTTGAGACA
CGAGATTATCGCTTTCGACTCCGCTGGTCAAGCTAACGGTGCTCAAACTACCCACAGTGTTC AAC
ATTGCTGTTTCCGGTCTGGTTCCTTGCAACCATCTGGTGTGGTGGTACTGACTGTACACGCTAC
TGACCCAGGTGTTTTGATCAACATCTACACTTCCCCATTGAACTACATCATCCAGGTCCAAGTGTG
TTTCCGGATTGCCAACTTCTGTTGCTCAAGGTTCTTCTGCTGCTACTGCTACAGCTTCTGCTACTGTT
CCAGGTGGTGGTCTGGTCCAACCTCCAGAACTACTACAAGTCTAGAACTACTCAGGCTTCTCTA
GACCATCTTCTACTCCTCCAGCTACTACTTCTGCTCCTGCTGGTGGTCTACTCAAACCTTGTACGGT
CAATGTGGTGGATCCGGTACTCTGGACCAACTAGATGTGCTCCACCAGCTACATGTTCCACTTTGA
ACCCTTACTACGCTCAGTGTGGTGAACCATCACCATCACCATCACTAG

- *TrCel61A* with C-terminal his-tag, native codon usage

CATGGACATATTAATGACATTGTCATCAACGGGGTGTGGTATCAGGCCTATGATCCTACAACGTTTC
CATACGAGTCAAACCCCCCATAGTAGTGGGCTGGACGGCTGCCGACCTTGACAACGGCTTCGTT
TCACCCGACGCATACCAAAACCCTGACATCATCTGCCACAAGAATGCTACGAATGCCAAGGGGCAC
GCGTCTGTCAAGGCCGGAGACACTATTCTCTCCAGTGGGTGCCAGTTCCATGGCCGCACCCTGG
TCCATTGTGCGACTACCTGGCCAACTGCAATGGTGACTGCGAGACCGTTGACAAGACGACGCTTG
AGTTCTTCAAGATCGATGGCGTTGGTCTCCTCAGCGGCGGGGATCCGGGCACCTGGGCCTCAGAC
GTGCTGATCTCCAACAACAACACCTGGGTCGTCAAGATCCCCGACAATCTTGCGCCAGGCAATTAC
GTGCTCCGCCACGAGATCATCGCGTTACACAGCGCCGGGCAGGCAAACGGCGCTCAGAACTACC
CCCAGTGCTTAAACATTGCCGTCTCAGGCTCGGGTCTCTGCAGCCCAGCGGCTTCTAGGGACC
GACCTCTATCACGCGACGGACCCTGGTGTCTCATCAACATCTACACCAGCCCGCTCAACTACATC
ATCCCTGGACCTACCGTGGTATCAGGCCTGCCAACGAGTGTGCCCAGGGGAGCTCCGCCGCGA
CGGCCACCGCCAGCGCCACTGTTCTGGAGGCGGTAGCGGCCCGACCAGCAGAACACGACAAC
GGCGAGGACGACGAGGCCTCAAGCAGGCCAGCTCTACGCCTCCCGCAACCACGTGGGCACCT
GCTGGCGGCCCAACCCAGACTCTGTACGGCCAGTGTGGTGGCAGCGGTTACAGCGGGCCTACTC
GATGCGCGCCGCGCCAGCCACTTGTCTACCTTGAACCCCTACTACGCCAGTGCCTTAACGACGAC
CATCATCATCATCATATTGA

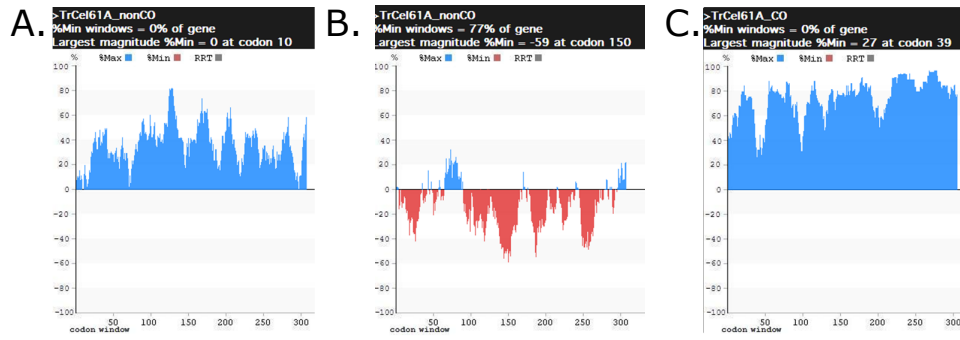


Figure S2.1: Codon clustering analysis for recombinant *TrCel61A*, calculated from the rare codon calculator²⁷⁷. (A) *TrCel61A* with native codon usage in its native host *Trichoderma reesei* (B) native codon usage in the host *Pichia pastoris* (C) codon optimized gene in the host *Pichia pastoris*.

- *PcGH61D* native secretion signal, codon optimized

ATGAAGGCTTTCTTTGCCGTTTGGCAGTTGTCTCTGCTCCATTGTCTTGGGT

- *PcGH61D*, codon optimized

CACTACACTTTCCCAGACTTCATTGAGCCATCCGGTACTGTTACTGGTGACTGGGTTTACGTTAGAG
AGACTCAGAACCACTACTCCAACGGTCCAGTTACTGACGTTACTCCCCAGAGTTCAGATGTTACGA
GTTGGACTTGCAAGAACTGCTGGTCAAACCTCAGACTGCTACTGTTTCTGCTGGTGACACTGTTGGT
TTCAAGGCTAACTCTGCTATCTACCACCCAGGTTACTTGGACGTTATGATGTCTCCAGCTTCTCCTGC
TGCTAACTCTCCAGAAGCTGGTACTGGACAGACATGGTTCAAGATCTACGAAGAGAAGCCACAGTTC
GAGAACGGTCAGTTGGTTTTGACACTACTCAGCAAGAGGTTACTTTCACTATCCCAAAGTCCTTGC
CTTCCGGTCAGTACTTGTGAGAATTGAGCAGATCGCTTGGACGTTGCTTCATCTTACGGTGGTGC
TCAGTTCTACATTGGTTGTGCTCAATTGAACGTTGAGAACGGTGGTAACGGTACTCCAGGTCCATTG
GTTTCTATCCCAGGTGTTTACACTGGTTACGAGCCAGGTATCTTGATCAACATCTACAAGTTGCCAAA
GAACTTCACTGGATACCCAGCTCCAGGACCAGCTGTTTGGCAAGGT

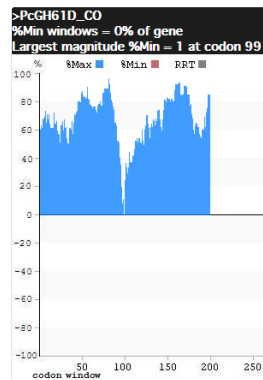


Figure S2.2: Codon clustering analysis for recombinant *PcGH61D*, calculated from the rare codon calculator²⁷⁷, studying the codon optimized gene in the host *Pichia pastoris*.

SDS-PAGE

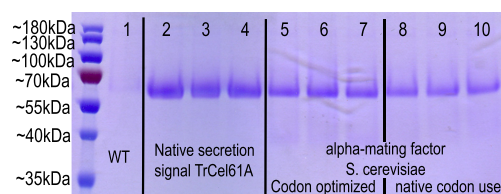


Figure S2.3: SDS-PAGE analysis of 3 different colonies, producing *Trichoderma reesei* Cel61A preceded by different secretion signals. Lane 1: wild-type *Pichia pastoris*, lanes 2-4: codon optimized gene sequence preceded by its native secretion signal, lanes 5-7: codon optimized gene sequence preceded by α -mating factor, lanes 8-10: native codon usage of the gene preceded by α -mating factor.

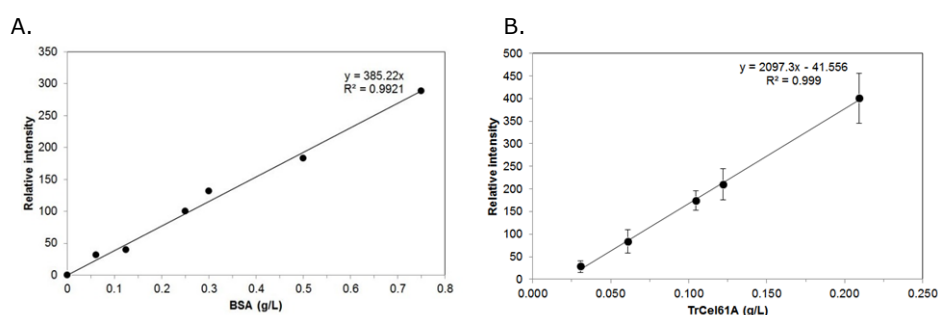


Figure S2.4: Standard curve for different concentrations of (A) bovine serum albumin (BSA) and (B) *TrCel61A* (performed in triplicate) and the relative intensity on SDS-PAGE.

Edman-degradation

Table S2.1: Exact N-terminal processing position following Edman-degradation analysis. Remark: the DDDDR-sequence was not treated with enterokinase in this experiment.

Secretion signal name	Obtained N-terminus	Remark	Percentage (%)
<i>S. cerevisiae</i> Alpha-Mating factor	APVXTTT	α -MF: A20	100
<i>S. cerevisiae</i> Alpha-Mating factor -EA-EA	EAEAHGH APVNTTT	α -MF: EAEA-signal α -MF: A20	57 34
<i>S. cerevisiae</i> Alpha-Mating factor -DDDDR	DDDDRHG APVNTTT PFXNXTN GYDXXX	enterokinase: DDDDRHG α -MF: A20 α -MF: P54 α -MF: G40	56 34 7 3
<i>T. reesei</i> Cel61A	HGHINDI	Correct (H22 = H1 from <i>TrCel61A</i>)	100
<i>P. chrysosporium</i> GH61D	HGHINDI AYDXTTX	Correct (H22 = H1 from <i>TrCel61A</i>) <i>TrCel61A</i> A37	82 18
<i>P. pastoris</i> DDDK-protein	HINDIVI	H24 = H3 from <i>TrCel61A</i>	100

HPAEC-PAD standards

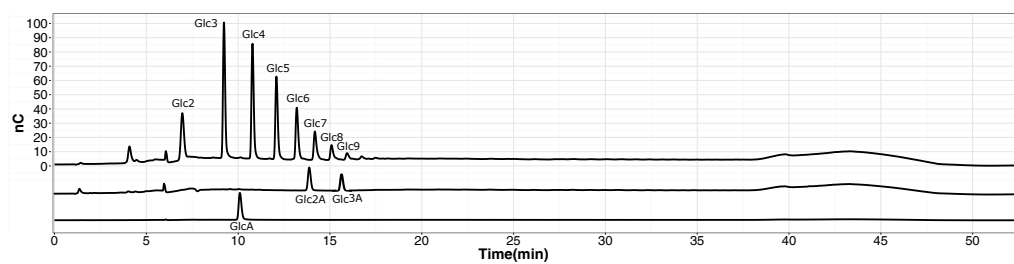


Figure S2.5: HPAEC-PAD standards that were included in every run to identify products. GlcA: gluconic acid, Glc2A: cellobionic acid, Glc3A: celotronic acid and Glcx:glucose chain in β -1,4-bond of x units.

Molecular strategies

Table S2.2: Detailed overview of strategies used to assemble the different constructs. Overview of all constructions that were compared. The molecular strategy and applied primer sequences are summed in the table. In the primer sequences, uppercase *italic* letters are part of the secretion signal sequence, while uppercase bold letters are used for *TrCel6A*+His-tag DNA-sequence.

Construct description	Strategy	Primer sequences
Codon optimized <i>TrCel6A</i> preceded by α -mating factor (EKR-EA-EA)	<ol style="list-style-type: none"> 1. Include XhoI restriction site at 5' and NotI at 3' end of the sequence, starting from a pMAT vector including <i>TrCel6A</i> (primers FW1+REV1, restriction sites underlined) 2. Restriction digest from the PCR product and pPpT4 backbone with XhoI-NotI followed by ligation 	<p>FW1: 5'- <i>AAACTCGAGAAGAGAGAGGCCGAAGCT</i> CACGG TCACATTAACGACATCGTTATC-3'</p> <p>REV1: 5'- tagcgccgcCTAGTGATGGTGATGGTGATGGT TCAAACACTGAGCGTAGTAAGGG-3'</p>
<i>TrCel6A</i> with native codon usage + α -mating factor (EKR-EA-EA)	<ol style="list-style-type: none"> 1. Include XhoI restriction site at 5' and NotI at 3' end of the sequence, starting from a pJET vector including <i>TrCel6A</i> (primers FW1+REV1) 2. Restriction digest from the PCR product and pPpT4 backbone with XhoI-NotI followed by ligation 	<p>FW1: 5'- <i>AAACTCGAGAAAAGAGAGGCTGAAGCT</i> CATGGACATATTAA -3'</p> <p>REV1: 5'- aatcgccgcTCAATGATGATGATGATGATGGTCGTCGTT -3'</p>
<i>TrCel6A</i> preceded by α -Mating factor ending in Enterokinase cleavage site (EKR-DDDDR-)	One-piece Gibson assembly ²⁶⁹ removing E-A-E-A from the α -mating factor and changing it for D-D-D-D-R amino acid sequence (underlined) (Primers FW1 + REV2)	<p>FW1: 5'- <i>GGTGTCTCTCTCGAGAAGAGAGATGACGATGACAGA</i> CACGGTCACATTAACGACATCG -3'</p> <p>REV1: 5'- CGATGTCGTTAATGTGACCGTG <i>TCTGTCATCGTCATCTCTCTCTCGAGAGAGACACC</i> -3'</p>
<i>TrCel6A</i> with native codon usage + α - mating factor (EKR)	Method as described by Sanchis and coworkers ²⁶⁸ to remove amino-acids E-A-E-A at the 3' end of the α -mating factor starting from the complete vector with alpha-mating factor ending in EKR-E-A-E-A (Primers FW1 +REV1)	<p>FW1: 5'- <i>GGTGTCTCTCTCGAGAAGAGA</i> CACGGTCACATTAACGACATCG -3'</p> <p>REV1: 5'- gaagaggagtgggaaatacc -3'</p>

Construct description	Strategy	Primer sequences
DDDK protein native secretion signal	<p>1. Amplify <i>TrCel61A</i> with the DDDK-protein secretion signal in front via PCR amplification and including the sequence into the primer (primers FW1+REV1)</p> <p>2. Gibson assembly (1) to insert gene and secretion signal into pPpT4 vector (insert amplification primers: FW2 +REV1, backbone primers: FW3 + REV2)</p>	<p>FW1: 5'- <i>CTACTTTGGCTTCCATTGCTGTTGCT</i> CACGGTCACATTAACGACATCGTTATC -3'</p> <p>REV1: 5'- <i>cttgagcggccgcct</i>AGTGATGGTGAT GGTGATGGTTCAAACACTGAGCGTAGTAAGG -3'</p> <p>FW2: 5'- <i>GTTTAACTTGAAGACTATTTTGATTCTACTTTGG</i> <i>CTTCCATTGCTGTTGCT</i> -3'</p> <p>FW3: 5'- CCTTACTACGCTCAGTGTTTGAACCATCAC CATCACCATCACTAG<i>gcggccgctcaag</i> -3'</p> <p>REV2: 5'- <i>CAAAATAGTCTTCAAGTTAAACAT</i> <i>cgttcgggaattcttcaataattag</i> -3'</p> <p>FW1: 5'- <i>GTTGTCTCTGCTCCATTGTCTTGGGT</i> CACGGTCACATTAACGACATCG -3'</p>
PcGH61D	Gibson assembly ²⁶⁹ to exchange <i>TrCel61A</i> sequence to pPpT4 vector containing already <i>Phanerochaete chrysosporium</i> GH61D secretion signal (insert amplification primers: FW1 + REV1, backbone primers: FW2 + REV2)	<p>REV1: 5'- <i>ccgctc</i>AATGATGATGATGATGATGGTT CAAACACTGAGCGTAGTAAGG -3'</p> <p>FW2: 5'- CCTTACTACGCTCAGTGTTTGAACCAT CATCATCATCATCATTGA<i>gcgg</i> -3'</p> <p>REV2: 5'- CGATGTCGTTAATGTGACCGTG <i>ACCCAAGACAAATGGAGCAGAGACAAC</i> -3'</p>
<i>TrCel61A</i> preceded by its native secretion signal sequence	<p>1.Precede <i>TrCel61A</i> by its native secretion signal sequence via PCR primers and include EcoRI restriction site at the 5' and NotI site at 3' end of the sequence (primers FW1 + REV1, restriction sites underlined)</p> <p>2.Restriction digest from the PCR product and pPpT4 backbone with EcoRI-NotI followed by ligation</p>	<p>FW1: 5'- <i>aaagaattccgaaacg</i><u><i>ATGATTCAAAAATTGTCT</i></u> <u><i>AACTTACTTGTTACTGCTTTGGCAGTTGCTACTGGTGTGTGGA</i></u> CACGGTCACATTAACGACATCGTTATCAAC - 3'</p> <p>REV1: 5'-<i>tagcggccgc</i>CTAGTGATGGTGATGGTGATGGTTC AAACACTGAGCGTAGTAAGGG -3'</p>

3

Evaluation of measurement methods

Abstract

In this chapter, the measurement methods that will be used to identify stabilized variants of *Trichoderma reesei* Cel61A are evaluated. These will become important when engineering the enzyme in chapter 5.

First, the thermodynamic stability of *Tr*Cel61A was measured via differential scanning fluorimetry (DSF). The corresponding parameter that is obtained and compared between variants is the apparent melting temperature (T_m). DSF appeared to be sensitive and reproducible method with a low technical and biological variability (1 °C). In most cases, purifying the culture supernatant is sufficient to a sufficient purity was obtained after ultrafiltration of the culture supernatant of the recombinantly produced

purification of the recombinantly produced *Tr*Cel61A from the culture medium via ultrafiltration will be sufficient, although a subsequent Ni-NTA chromatography step can increase the quality of the resulting unfolding curve.

Secondly, high performance anion-exchange chromatography (HPAEC, specialized HPLC) was performed for activity analysis on the cellulosic substrate PASC. The peak area of different reaction products was monitored through time, whereby cellobionic acid can serve as a representative for the overall reaction. Neutral and C1-oxidized chains with a degree of polymerization (DP) starting from 5 and up decrease over time, suggesting a specificity of *Tr*Cel61A for short oligosaccharides in addition to the more crystalline cellulosic substrate PASC. Even though cellobiose dehydrogenase (CDH) is a suitable electron donor, ascorbic acid is preferred because of overlap in product formation of both enzymes.

Finally, the methods were shown to be valuable for the evaluation of destabilizing mutations. Disturbing ion pair network (E123-K85) in the surface decreases the stability, as could be measured with the optimized DSF method.

3.1 Introduction

In chapter 2, the expression optimization of *Trichoderma reesei* Cel61A was described, while in a later phase (Chapter 5), some engineering methods will be applied to increase the stability of this LPMO. In that respect, a sensitive and selective measurement method is crucial for assessing the stability and activity of the LPMO variants.

Despite the fact that the enzymatic function of LPMOs has been discovered now for six years, measuring the activity on quantitative and even qualitative level is still the largest bottleneck in studying these interesting enzymes. There are several reasons for the lack of such basic biochemical methods. The actual nature of LPMOs, creating reaction products of different degrees of polymerization combined with oxidation at different positions (C1- or C4-carbon in the β -1,4-linkage), makes the analysis extremely challenging. Additionally, the lack of commercially available standards, use of insoluble substrates, auto-oxidation of the electron donor by dissolved Cu(II)²⁷⁸, and spontaneous decomposition of the C4-oxidized products in aqueous solutions²⁷⁹ and at high pH (\sim HPAEC)²⁸⁰ further hamper the development process of a suitable method.

The most powerful and most popular method till now is still the use of high performance anion-exchange chromatography (HPAEC, a specialized form of HPLC) to analyze the soluble reaction products. The drawbacks of this method though are the low throughput due to lengthy incubation periods and a long analysis program to separate the reaction products, costly standards for identification and underestimation of the LPMO activity because only soluble products are measured. Different studies used the peak area of certain peaks as measure, while a recent study used the peak height of C4-oxidized cellobiose (Glc4gemGlc) to quantify activity of NcLPMO9C¹⁸³. Some alternative approaches, aiming at more high-throughput screening have been proposed. These include the use of chromogenic substrates²⁸¹, the monitoring of ascorbate (electron donor) consumption²⁸² and an unproductive side reaction of the LPMO: reduction of O₂ to H₂O₂²⁴⁵. However, these assays all have drawbacks resulting in a low practical application. Indeed, the use of hydrogels and chromogenic substrates is not developed as a fully quantitative assay since the chlorotriazine groups may hinder the LPMO to access the substrate. And secondly, the use of ascorbic acid is diminishing because of its questionable stability and biological relevance as electron donor¹⁵². Furthermore, the method is still labor-intensive and time-consuming since the reaction mixtures have to be centrifuged, filtered and subsequently analyzed via HPLC. Only the last assay (measuring H₂O₂ production) is used in other studies^{124;125;128;133}, albeit often to measure substrate preference rather than activity. Hence the side reaction only occurs in absence of a suitable substrate.

Engineering approaches require a fast and robust method to measure specifically the desired characteristic. In the case of stability, two types exist: kinetic and thermodynamic stability (See literature study, section 'Protein thermostability' starting on page 40). Since the first is mostly measured via an activity assay, we will omit this by mainly focusing on thermodynamic stability, measured without the catalytic reaction. Various methods exist to follow the reversible thermal denaturation of a protein such as the conventional methods of differential scanning calorimetry (DSC)²⁸³ and circular dichroism (CD)²⁸⁴. However, these methods have the huge disadvantage of low throughput (~ 1 hour / sample) and the need for large amounts of pure protein. Moreover, CD only works in limited buffer conditions. A more recent technique for monitoring protein unfolding is differential scanning fluorimetry (DSF)²⁸⁵, which has the advantages of high-throughput performance (96-well plate), requirement of low enzyme amounts and application in a simple real-time PCR device, available in many labs²⁸⁶. However, the need for a dye can hamper the measurement and makes subsequent cycles of heating/cooling impossible. A very recent method is nanoDSF, also called microscale thermophoresis. This device uses a capillary system for high-throughput and requires only low protein amounts and measures the intrinsic fluorescence of the protein^{287;288}.

This chapter will therefore describe the use of HPAEC-traces and differential scanning fluorimetry for evaluation of activity and thermodynamic stability, respectively. As a proof of concept, the activity and stability of a destabilized mutant will be assessed in the last section.

3.2 Results and discussion

3.2.1 Differential scanning fluorimetry for measuring the apparent melting temperature

Differential scanning fluorimetry (DSF) measures thermal unfolding of a protein by monitoring the fluorescent signal that is caused from dye-binding to the hydrophobic interior of a protein such as exposed in molten globules or thermal unfolding intermediates²⁸⁵. Dyes are selected on this ability combined with the premise of being quenched in polar environments. Some examples of such dyes are 1-anilinonaphtalene-8-sulphonic acid (1,8-ANS) and variants²⁸⁹, Nile Red and Sypro Orange. The assay is also known as thermal shift assay, ThermoFluor assay or high-throughput thermal scanning (HTTS) and was originally developed as a screening method on micro-scale for drug discovery, based on the concept of ligand-induced conformational stability²⁹⁰. Later on, the applicability was extended to identifying enzyme-stabilizing buffers and ligands²⁸⁶ and measuring stability of protein variants²⁸⁹. Differential scanning fluorimetry results in a sharp protein unfolding curve, whereof the inflection point corresponds to the apparent melting temperature (T_m). This can be deduced from the negative first derivative, as calculated by the software of the qPCR device (here Biorad CFX Connect software).

Since DSF will be the main evaluation tool in the stability engineering part, a profound evaluation was done to find the variability of the test. A set-up with 4 parallel Erlenmeyer shake flasks was performed, containing 1 flask with the wild-type *P. pastoris* strain and 3 flasks with the strain producing heterologous TrCel61A, as developed in chapter 2 (TrCel61A-NSS). They were all 4 simultaneously and equally grown and induced for protein expression, followed by purification via ultrafiltration and subsequent Ni-NTA chromatography. The apparent melting temperature was measured in duplicate after each purification step. Obviously, the wild-type *P. pastoris* did not show any protein on SDS-PAGE (S3.1), nor a curved signal during DSF, confirming no background should be taken into account. On the other hand, from the strains producing TrCel61A, a melting curve and derivative (Fig. 3.1) were obtained after ultrafiltration only. In 2 out of the 3 cases, a decent melting temperature could be deduced (Table 3.1), while in the third case (Erlenmeyer 3, yellow curve) a double melting value was obtained. After subsequent Ni-NTA chromatography however, the curves were more smooth and showed an identical 2-step transition, with a T_m equal to the value obtained after ultrafiltration only. Concerning variability, the technical variability was found to be 1 °C which is the measurement error of the device, while the biological variability also equals 1 °C. In conclusion, DSF is a robust assay with low variability. Ultra-filtration provides in most cases an enzyme sample, pure enough to find the apparent melting temperature, while subsequent immobilized metal ion affinity chromatography (IMAC) can provide a solution for difficult samples. Moreover, a difference of

at least 2 °C as compared to the WT enzyme will be necessary to conclude on a significant increase/decrease in thermodynamic stability. Since stability increments are usually very small, out-stepping this difference will be a challenge.

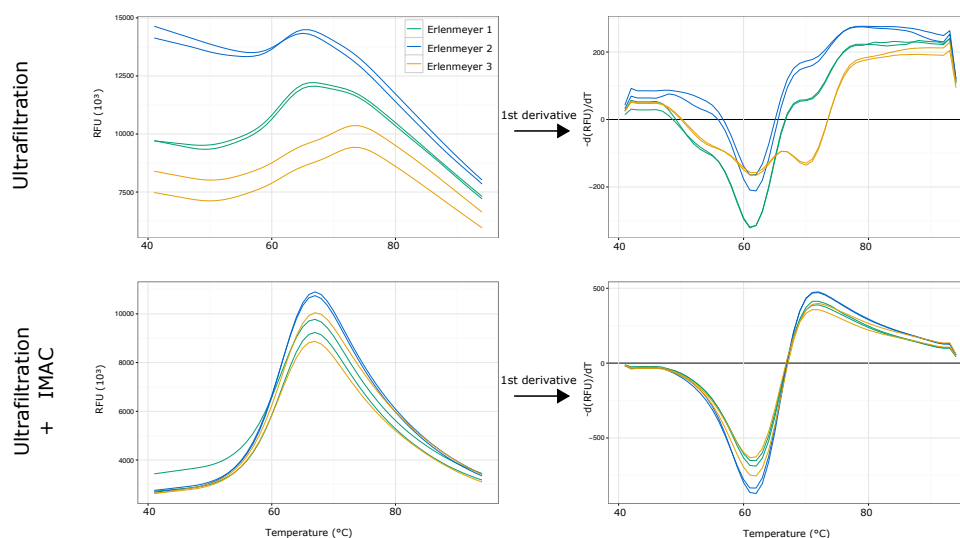


Figure 3.1: Melting curve and negative derivative for proteins purified by ultrafiltration and immobilized metal ion affinity chromatography (IMAC).

Table 3.1: Set-up for measuring variability and the effect of purity on DSF.

Erlenmeyer	Concentration (mg/mL)	T_m (°C)		Average (°C)	Stdev (°C)
		1	2		
Ultrafiltration					
1	0.255	61	61	61	0
2	0.243	62	61	62	1
3	0.237	61 en 70	62 and 70	62 (and 70)	1
				62	1
Ultrafiltration and IMAC					
1	0.174	62	62	62	0
2	0.150	62	61	62	1
3	0.221	61	62	62	1
				62	0

Secondly, an optimal ratio of enzyme and Sypro Orange was assessed. An undiluted sample after purification was taken as starting point ($\sim 4 \mu\text{g}$ LPMO), whereas different dilutions of the enzyme and dye were tested (Figure S3.2). An optimal protocol was found to be in a total of 25 μL sample volume, including 10 μL 1/400 diluted Sypro Orange and about 4 μg LPMO.

Despite *TrCel61A* being a multidomain protein, containing a catalytic domain and celu-

lose binding domain, connected by a linker, the protein only shows one unfolding midpoint temperature. Even though this is contra-intuitive, such findings were observed earlier and were explained as cooperative unfolding of the protein²⁹⁰. As alternative explanation, preferred and strong binding of the dye to one of the domains during the unfolding process was suggested²⁸⁶.

It can also be observed from figure 3.1 that the melting curve is nicely clock shaped for IMAC purified *TrCel61A*, while the fluorescence starts higher for an ultrafiltrated culture supernatant. This observation is also found in further work in this thesis. The high starting signal is most probably due to impurities that interact with the dye such as for example DNA/RNA that is released by the host during cell lysis, while removed during his-tag purification. The impurities should be bigger than the 10 kDa filter cut-off since they are retained, although not found on SDS-PAGE (S3.1).

On a more general note, it should be noted that the obtained fluorescent signal is rather low compared to other enzymes available in the lab and used to check the method. This can be due to the size and shape of the protein, determining the ratio of surface residues compared to buried (hydrophobic) residues that cause the signal. Furthermore, disulfide bridges in the protein might also keep the fold tight, even in denatured form so that the interior is not fully exposed to dye-binding.

3.2.2 HPAEC for analysis of reaction products formed from PASC

Activity tests will be mainly used to check if enzyme variants did not alter the activity dramatically, whereas stability is the main concern during this work. The activity method described here, is the use of high performance anion exchange chromatography (HPAEC) analysis to evaluate the reaction products formed from the cellulosic substrate PASC. A simple end-point measurement was already shown in chapter 2 and demonstrated that *TrCel61A* is a type-3 LPMO, with products appearing in the region for neutral, C1-, C4-oxidized and double (C1- and C4-) oxidized products.

Cellobiose dehydrogenase as electron donor

LPMOs require an electron donor for their oxidative reaction as explained in the literature review (page 26 and following). Often ascorbate is used, although the natural system has been proposed to be relying on a second enzyme, namely cellobiose dehydrogenase (CDH), and this LPMO/CDH system has been applied in different studies^{11;12;111;118;245}. Frommhagen et al. even proposed that LPMOs have, apart from their regioselectivity, also an electron donor preference¹⁴⁹. In our lab, cellobiose dehydrogenase from *Myceliophthora fergusii* was available (*MfCDHIIB*, Uniprot E7D6B9) and therefore it was compared

to the use of ascorbate, as depicted in figure 3.2.

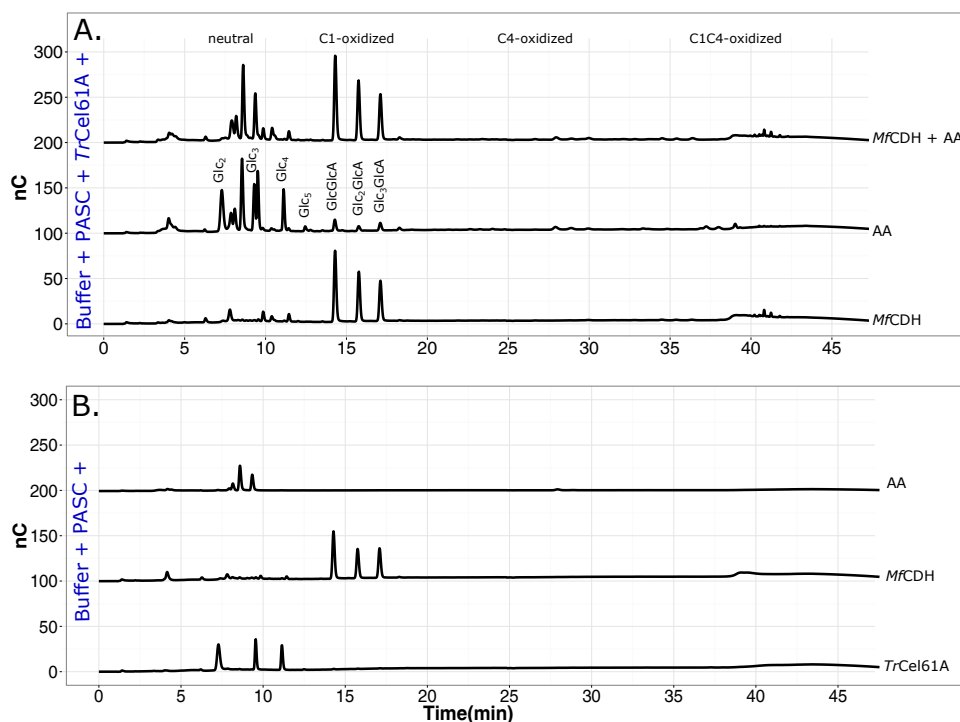


Figure 3.2: Activity of *TrCel61A* on PASC in combination with different electron donors: ascorbic acid (AA) or *Mycelophthora fergusii* cellobiose dehydrogenase (*MfCDH*). Panel A shows the reaction of combining a mixture of buffer, PASC, *TrCel61A* and the indicated electron donor, while panel B shows the result of an activity test when only mixing buffer and PASC and the indicated extra component.

From panel A, a clear difference between the use of ascorbic acid (AA) and CDH that can be observed is the product shift from neutral oligos to aldonic acids (C1-oxidized), respectively. Hence the natural reaction of CDH is the conversion of cellobiose to cellobiono- δ -lactone, which spontaneously hydrolyzes to cellobionic acid²⁹¹. The specificity for cellobioses up to cellopentaose (DP 5) has been demonstrated for CDH from *Sporotrichum pulverulentum*²⁹². *MfCDH* is shown here to also utilize cellotriose and cellotetraose as substrates, reducing the load of neutral glucan chains. Furthermore, the complex products that are eluted around 40 minutes are different in both reactions (precise identification was not done here or in any other study because very high complexity of instable products^{279;280}). However these products did not appear from a control reaction lacking LPMO. When both electron donors are added simultaneously to the LPMO-reaction, a combined pattern of the former two reactions appears, wherein the neutral oligos are all converted to aldonic acid forms and a complex pattern in the C1/C4-regio is observed. It can also be observed in panel B of figure 3.2 that there is already formation of some neutral oligosaccharides when only *TrCel61A* is mixed with substrate, lacking an electron donor. This is

the background activity we observe from an endogenous endoglucanase (GH45, GenBank id CCA40496.1) that appears to be present in our host *P. pastoris*. Similarly, when adding just *MfCDH*, some peaks already appear in the C1-oxidized region. This enzyme was also produced in *P. pastoris* and therefore also suffers from this background endoglucanase. The neutral oligosaccharides are directly oxidized by CDH. To evaluate the influence, experiments with his-tag purified enzymes can be performed (not done in this work).

In conclusion, *TrCel61A* can use either ascorbic acid or CDH as electron donor. However, the use of ascorbic acid is preferred for analysis since the reaction products from CDH (especially C1-oxidized) partially overlap with the reaction products from the LPMO. This might be particularly cumbersome for analysis of the reaction products of type-1 LPMOs, that strictly form C1-oxidized oligosaccharides along with the neutral ones.

Time-course of the reaction on cellulose

In order to evaluate how fast products are formed and what products are formed first by the LPMO, the release of soluble products was monitored on different time points via HPAEC-PAD. Even though the release rate of soluble cello-oligos is considered relatively low for LPMOs, so that lengthy incubations are usually performed, a time-course of *TrCel61A* already reveals a detectable amount of neutral and C1-oxidized products after only 2 hours of incubation with PASC. In the 44 hour-incubation shown (Fig. 3.3), the amount of most soluble cello-oligos kept increasing. Hence, Langston et al. even showed increasing product formation up to 5 days of incubation, albeit at a strongly reduced rate when incubation time increased¹². Notably, the C4-oxidized products remain very small and barely detectable in comparison to the neutral oligos and aldonic acids. The reason behind this is mostly underestimation because of instability of ketoaldoses, especially in alkaline environments such as applied for HPAEC²⁸⁰. For C1-oxidized products on the other hand, no literature was found on instability of the compounds. It is known that C1-oxidized products in solution exist in a chemical equilibrium between their lactone form and aldonic acid form, depending on factors like pH, temperature and concentration²⁹³.

When observing the time-course, it can be noted that the neutral oligos appear quicker than the C1-oxidized products although one would expect to see a neutral product for every charged product following the LPMO reaction. Different reasons can be found for this observation. First, as mentioned above, *P. pastoris* also expresses an endogenous endoglucanase (GH45, GenBank id CCA40496.1). This endoglucanase background is responsible for formation of neutral oligo formation apart from the LPMO formation. Second, detection of C1-oxidized products is less sensitive than neutral ones because products are oxidized at the gold electrode and C1 has the highest reducing tendency. Furthermore, the later the product elutes, the higher the sodium acetate gradient and thus the lower the sig-

nal²⁹⁴. And third, it is known for C4-oxidized products that the 4-ketoglucose (non-reducing end) is lost at HPAEC-conditions²⁸⁰. Thus, some of the non-charged products also result from degradation of the corresponding C4-oxidized compound. Observing the aldonic acid region, *TrCel61A* seems to have a preference for even-numbered products, whereof cellobionic acid is the main product formed. A very strong even-numbered preference has been shown for different bacterial LPMOs, active on chitin^{14;140;295}, although not yet observed for fungal LPMOs^{11;113;135}.

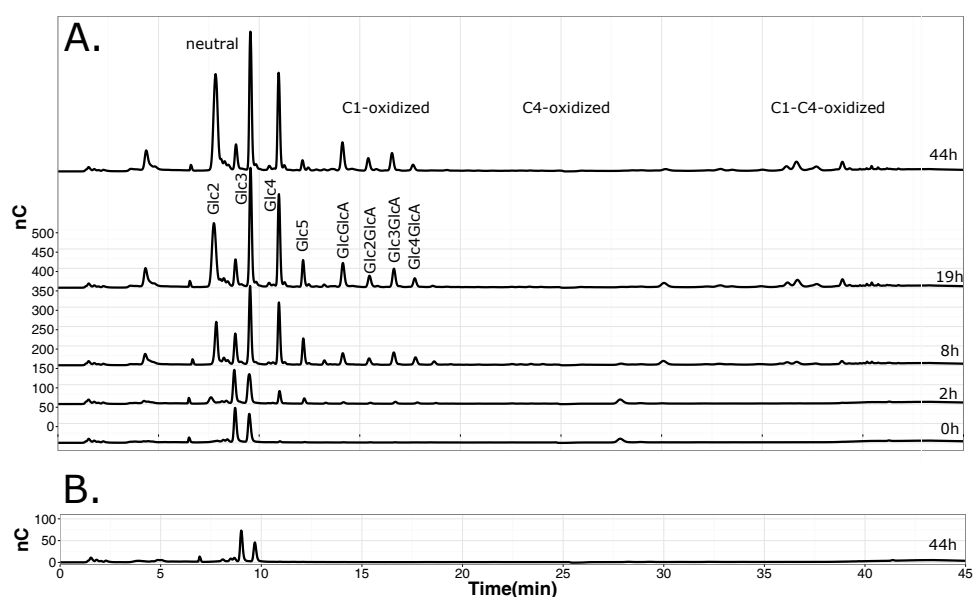


Figure 3.3: Time-course of a reaction of *TrCel61A* on PASC at 50 °C. Samples were analyzed after 0, 2, 8, 19, and 44 h of reaction. (A) Enzyme reaction with the addition of *TrCel61A*. (B) Control reaction performed without the addition of enzyme. Cellobionic acid (GlcGlcA) and cellotrionic acid (Glc2GlcA) were used as standard whereas the nature of the other oxidized products was inferred from the literature²⁶⁴.

Furthermore, the release of cellobiose (DP2) and cellotriose (DP3) and their oxidized counterparts cellobionic acid and cellotrionic acid keep increasing over time, while cellotetraose and cellotetraonic acid (DP4) is no longer linear after 20 hours. This can indicate the substrate is depleted or cellotetraose/cellotetraonic acid is used as substrate. The same trend is observed for products of DP 6 (no more linear after 10 hours) and products of DP5 even start decreasing over time (Fig. 3.4). This suggests that *TrCel61A* is also active on short soluble oligos as described earlier for *Neurospora crassa* LPMO9C¹²⁴. *TrCel61A* would recycle its own reaction product again as substrate when the concentration becomes high enough. A substrate of minimally 4 glucose units would be required. Alternatively, these products can also be the result the background endoglucanase in *P. pastoris*.

Since the HPAEC-trace involves many different reaction products, monitoring the peak area of one product as representative for the reaction would be interesting. Cellobionic acid and cellotronic acid seem to be excellent candidates since they only increase in time (thus, do not serve as substrate) and oxidized products are specific for the LPMO reaction with ascorbic acid as electron donor. Furthermore, they are all nicely separated in the chromatogram and show a decent peak area. Since cellobionic acid is the most prevalent peak, this was chosen for further analyses. An endpoint after overnight reaction (16h) is still in the linear phase, based on figure 3.4

The linear cause between enzyme concentration and product formation was further examined by Ir Barbara Danneels, as written in a recently submitted paper and will be discussed in her PhD thesis.

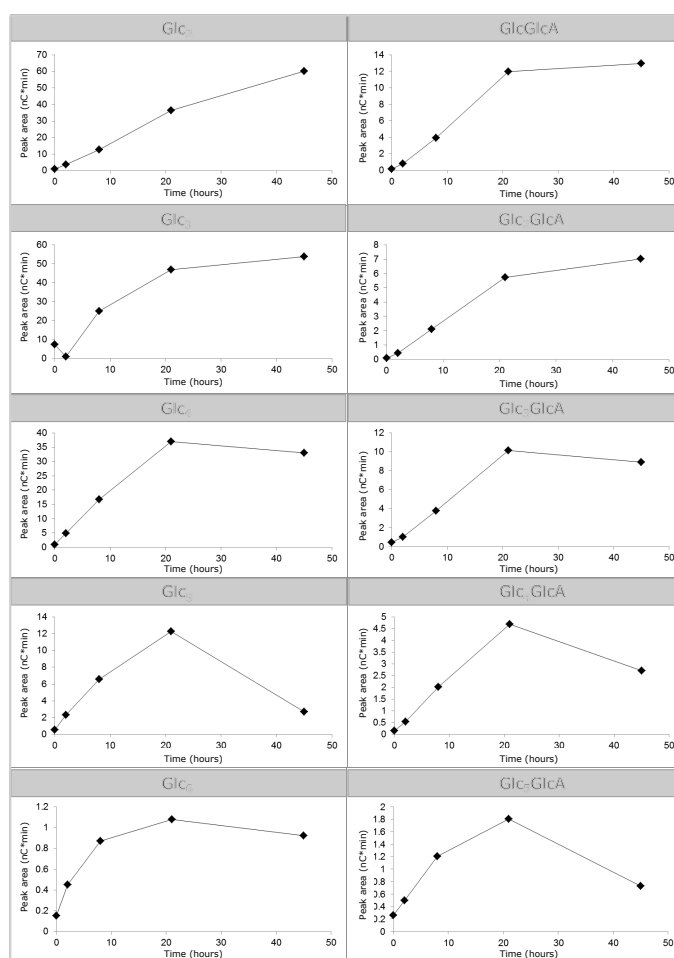


Figure 3.4: Peak area of different products in HPAEC-analysis.

3.2.3 Proof of concept: stability and activity of a destabilizing mutant

In order to evaluate the methods described above, potentially destabilizing mutants were designed and created. Hence, destabilizing a protein is often a less cumbersome task and can equally validate the protocols.

Design of destabilizing variants

Two ion-pair networks were disturbed to destabilize *TrCel61A* by rationally inspecting the protein using Pymol as visualization tool⁷³. The first network is situated at the protein surface and is not conserved in the AA9 family. Exchanging glutamate 123 for an alanine (E123A) will disturb the electrostatic interaction with lysine 85 (Fig.3.5 A). The effect is considered only on stability of the protein, while no effect on activity is expected. The second proposed mutation is a conserved ion-pair in the twisted β -sandwich, defining the core region of the LPMO (Fig.3.5 B). Mutating glutamate 179 into alanine (E179A) might have an influence on the entire protein fold and be much more of a problem for the protein's stability.

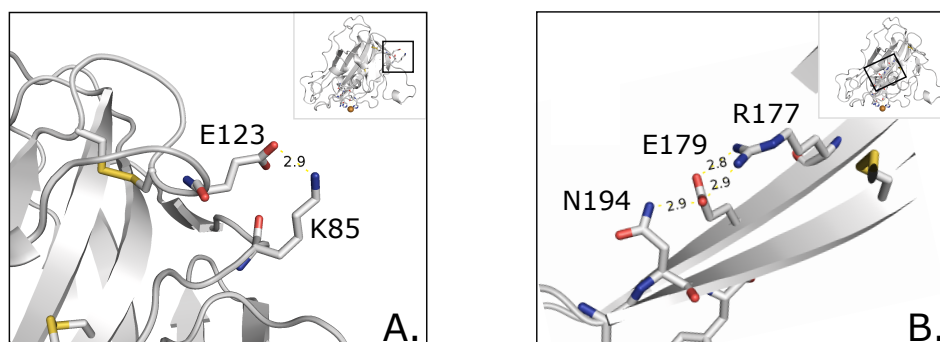


Figure 3.5: Design of destabilizing variants via destroying ion pair networks (A) mutation E123A will destroy network E123–K85 (B) mutation E179A will disturb E179A–R177.

Effect on thermodynamic stability and activity

The enzyme variant carrying mutation E179A, could not be produced. The mutation was most likely too dramatic and resulted in misfolding and the degradation pathway. The other enzyme variant was produced in shake flasks and purified via ultrafiltration. The yield was approximately 60 % from the wild-type *TrCel61A*, which corresponds to 42 mg protein /L culture volume. (SDS-gel in supplementary figure S3.3). A subsequent Ni-NTA chromatography step was not performed since the variant already showed a 2-state unfolding with distinct and unambiguous apparent melting temperature (curves in supplementary fig S3.4). This melting temperature was significantly ($p=0.0004$, Student t-test) decreased with 4 °C to 56 ± 1 °C, as compared to the wild-type enzyme (60 ± 1 °C).

Secondly, the activity of variant *TrCel61A-E123A* has slightly increased (x 1.5) on PASC in comparison to the wild-type LPMO. More elaborate experiments should confirm if the increase of the first is a real increase. However, they were not performed in the course of this thesis work.

Nevertheless, this proof-of-concept shows differences in apparent melting temperature and activity can be observed and the DSF method is rigid enough to screen the stabilizing mutants in chapter 5.

3.3 Conclusion

In this chapter, the use of differential scanning fluorimetry (DSF) for stability evaluation was proven to be a sensible and reproducible measurement (technical and biological variance 1 °C). Mostly purification via ultrafiltration will be sufficient, although Ni-NTA chromatography can improve the quality of the resulting curve, when any doubt. Furthermore, high performance anion-exchange chromatography will be performed for analysis of the activity, that can be monitored following the peak area of cellobionic acid. Even though *TrCel61A* can use cellobiose dehydrogenase (CDH) as electron donor, ascorbic acid is preferred. The underlying reason is the overlap in C1-oxidized products formed by both the LPMO and CDH. Finally, the methods were shown to be valuable by use of destabilizing mutations. When an ion pair network (E123-K85) in the surface of the protein is destroyed, a decrease in thermodynamic stability is observed.

3.4 Materials and methods

The experimental procedure for *Pichia pastoris* transformation, growth, protein production and purification can be found in 'Materials and methods' section of chapter 2.

3.4.1 Molecular techniques for creating destabilizing variants

Destabilizing variants of *TrCel61A*-NSS were created in the pPp-T4-vector, as described in chapter 2. The site-directed mutants were created by use of the method of Sanchis et al.²⁶⁸, using mutagenic forward primers for mutation E123A (5'-AACGGTGACTGT**GCT**ACTGT TGACAAG-3') and E179A (5'-CGTTTTGAGACACGCTATTATC**GCT**TTTGCCTCCGC-3'). The same reverse primer was used in both cases (5'-CGAATGCGAAGGTTAGTAGG-3'), located about 500 bp downstream the stopcodon of the gene sequence of interest. Sanchis et al describe in their method a single PCR reaction, combining the creation of a mega-primer in 5 cycles and a subsequent whole plasmid PCR. Nonetheless, this procedure was split in separate PCR reactions in order to obtain higher yields. The first reaction was performed with Primestart GXL DNA polymerase (Takara Bio Inc) and the template was 1/1000 diluted from a 100-200 ng/ μ purified plasmid concentrate. Subsequently, 2 μ L of the first reaction is added to a second tube to perform the whole-plasmid PCR reaction, using Pfu Ultra AD polymerase (Agilent Technologies), where to 1 μ L template (1/10 dilution) is again added as well as 2% (v/v) DMSO and primers were avoided. After *dpnI* restriction digest, 20 μ L in-house prepared²⁹⁶ electrocompetent BL21 (DE3) cells was subjected to electroporation with 2 μ L of the mixture. Cells were selected on LB plates containing 100 μ g / mL zeocin. The correct sequence of the mutated versions of *TrCel61A* was confirmed by nucleotide sequencing (Macrogen Europe) before transforming *Pichia pastoris* with the corresponding plasmid, followed by enzyme production.

3.4.2 Differential scanning fluorimetry

The apparent melting temperature of purified enzymes was determined by differential scanning fluorimetry (DSF) in a CFX Connect96 cycler (Biorad). Excitation and fluorescent detection were performed using the FRET channel at wavelengths of 450-490 nm and 560-580 nm, respectively. Samples were diluted to equal enzyme concentrations (minimal 0.5 mg/mL) and all variants were measured in triplicate. Each well consisted of 15 μ L purified enzyme and 10 μ L 1/400 diluted Sypro Orange (Sigma-Aldrich). The temperature was linearly increased (1 $^{\circ}$ C / min) from 25 to 95 $^{\circ}$ C. The inflection point of the melting curve, corresponding with the apparent melting temperature, is calculated as the minimum value of the negative first derivative of the melting curve. This value is determined by the CFX Manager software (Biorad).

3.4.3 Production of *Mf*CDHIIB

As alternative electron donor, *Myceliophthora fergusii* cellobiose dehydrogenase IIB (*Mf*-CDHIIB), was available in the lab. This enzyme was produced by dr. Paul Ameloot in *Pichia pastoris* KM71H during fermentation, identical to the one explained in chapter 2. Although the commercial system from Invitrogen with pPicZ-vector was applied, the coding sequence of *Mf*CDHIIB was preceded by its native secretion signal while a 6xHis-tag was C-terminally attached.

3.4.4 Activity analysis

Activity evaluation was performed in 1.5 mL Eppendorf tubes in a 1000 μ L reaction volume. This reaction contained 40-100 μ g *Tr*Cel61A, 1 mM ascorbic acid and 1.2 % (w/v) PASC (prepared from Avicel PH-101²⁷⁶) in 10 mM sodium acetate buffer pH 5. When CDH was used as alternative electron donor, ascorbic acid was replaced by 0.6 μ M *Mf*CDHIIB and 200 μ M lactose. The tubes were incubated in a thermomixer (Eppendorf) at 50 °C, while 1400 rpm shaking. To end the reaction, samples of 75 μ L were boiled at 95 °C for 15 minutes, subsequently incubated on ice for 10 minutes and centrifuged for 15 minutes at maximum speed. The supernatant was diluted tenfold in milliQ water and analyzed via HPAEC-PAD, using the method of Forsberg et al²⁶.

Supplementary material

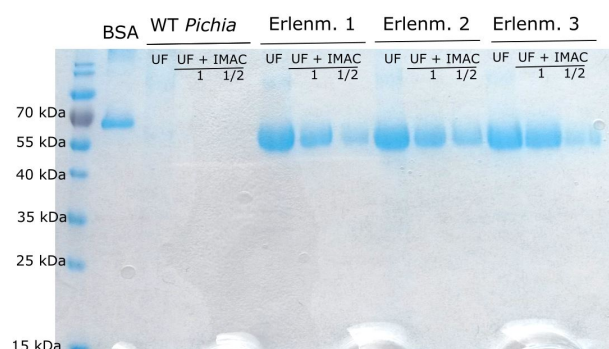


Figure S3.1: SDS-PAGE gel presenting different Erlenmeyer flasks and purification steps in DSF optimization. BSA: 0.1 g/l as a reference for quantification with ImageJ, UF: ultrafiltration, UF+IMAC: sample purified via ultrafiltration and subsequent Ni-NTA chromatography, 1/2: dilution of final fraction, WT *Pichia*: wild-type *Pichia pastoris* CBS7435 strain without heterologous gene insert.

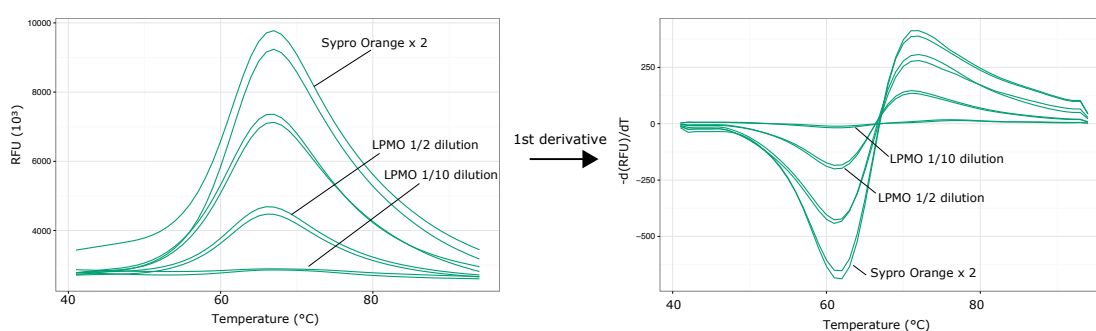


Figure S3.2: Effect of protein and dye concentration on DSF.

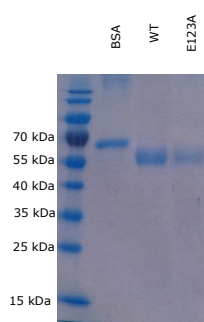


Figure S3.3: SDS-PAGE of destabilizing variant (*TrCel61A-E123A*) to test the measurement methods. BSA: 0.1 g/L as reference for quantification with ImageJ.

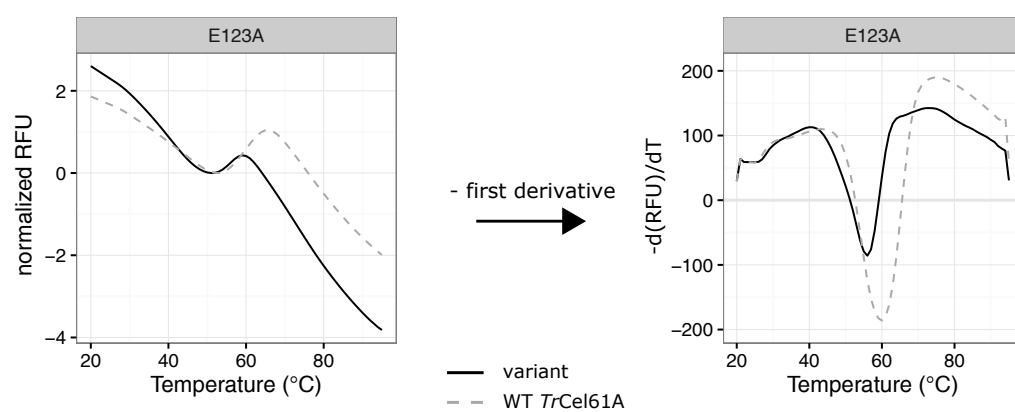


Figure S3.4: Meltcurve and negative derivative of destabilizing variant to evaluate the DSF protocol.

4

Expression of LPMO10C from
Streptomyces coelicolor and its
measurement methods

Abstract

Because lytic polysaccharide monooxygenases (LPMOs) have only recently been discovered, many aspects of these interesting enzymes are still to be unveiled. This work focuses on stability engineering in the next chapters, which requires a solid expression system. The expression of a fungal LPMO, *TrCel61A* from *Trichoderma reesei*, in the yeast *Pichia pastoris* is described in the previous chapters. Here, a second parallel expression system in the prokaryotic workhorse *E. coli* was described for LPMO10C from *Streptomyces coelicolor*. This second system has the huge advantage of a shorter turnover cycle in a bacterial host, which is advantageous for engineering purposes.

First, the expression of ScLPMO10C was studied with special attention to the preservation of the histidine residue at the protein's N-terminus, indispensable to maintain activity. The protein's native secretion signal and isolating the LPMO from the periplasm seemed the most suitable method. Next, purification was optimized and measurement methods validated. The use of differential scanning fluorimetry (DSF) as key method in stability engineering was shown to have very low variability and therefore an excellent assay. Activity was evaluated on the cellulosic substrate PASC and subsequently analyzed by HPAEC-PAD.

4.1 Introduction

Recombinant protein expression by yeasts, as described in chapter 2 is often used because yeasts combine the benefits of being unicellular organisms with eukaryotic characters²⁹⁷. This enables post-translational modifications that are required for the correct folding, stability and activity of many proteins. *Saccharomyces cerevisiae* was the first characterized yeast because of its use in brewing and baking. However, the applicability is limited because often low expression and hyperglycosylation appear^{298;299}. As an alternative, *Pichia pastoris* has become widely applied in science and industry^{261;262;300}. The yeast has several profitable features such as the capability of growing to very high cell density (200 g / L cell dry weight)³⁰¹, accompanied with possible high protein yields of > 5 g / L^{302;303}. Along with that, low endogenous protein secretion facilitates the down-stream processing and can lower the production cost²⁶¹.

Despite all these advantages, when using *P. pastoris* for protein engineering purposes, a disadvantage is the relatively long time required for 1 cycle of mutagenesis. Indeed, enzyme variants are mostly created through a bacterial intermediate (some exceptions exist³⁰⁴), linearization of the vector and integration in the genome of the yeast. Additionally, growing the yeast and producing a protein quickly takes one week (see chapter 2), while this only takes 2 days for an *Escherichia coli* culture. Therefore, a parallel expression system in the bacterial workhorse *E. coli*³⁰⁵ was established that strongly decreases the protein engineering cycle time.

LPMOs are classified following the CAZy database in auxiliary activity families 9, 10, 11 and 13¹⁶. They all share a characteristic flat binding surface to accommodate the substrate, a central twisted β -sandwich and an essential copper ion in the active site, coordinated by a so-called histidine brace. Family AA10, formerly known as CBM33, counts at the moment about 2000 entries, of which about 1800 are from bacterial origin. Most of them are active on chitin although some have shown activity on cellulose. One of these examples is *Streptomyces coelicolor* LPMO10C, formerly known as ScCelS2. The enzyme is active on cellulose and oxidizes the glucose chain to form neutral cellodextrins and gluconic acids. It is one of the first LPMOs that has been discovered and characterized²⁶. Additionally, the availability of its crystal structure has made the enzyme an interesting research subject^{114;248;280;306}. For all those reasons, ScLPMO10C seems a suitable second LPMO test case.

In later stages of this thesis work, the LPMO will be subjected to stability engineering strategies. The main goal of this chapter is therefore to evaluate an expression system in *E. coli*, that produces an active protein and of which the apparent melting temperature can be measured by differential scanning fluorimetry (DSF). A prerequisite is therefore a

pure enzyme of high quality. The first part of this chapter focuses mainly on the correct N-terminus required for activity, then the purification protocol is optimized and finally all assays are tested on the most suitable expression system.

4.2 Results and discussion

4.2.1 Evaluation of a suitable expression system: N-terminal processing

The most challenging part in expressing an active LPMO is to get the amino-terminal processing precisely to obtain a histidine at the first position (His-1). The reason behind this premise is that the histidine and the actual amino-terminus are involved in the coordination of a copper atom in the active center (see Literature review in chapter 1). To ensure this required N-terminus, different systems were compared in chapter 2 for the heterologous expression in the yeast *Pichia pastoris*. Likewise, different set-ups were considered here for heterologous expression in the bacterium *Escherichia coli*. Their expression level and purity, necessary for measuring activities, and especially apparent melting temperatures, were evaluated by SDS-PAGE. Their N-terminal processing on the other hand was investigated by activity on phosphoric acid swollen cellulose (PASC).

Different constructs

Similar to the strategies described for *Pichia pastoris*, the use of the natural cleavage system via a secretion signal and the use of a synthetic cleavable N-terminal extension were both considered to fulfill the His-1 premise (Fig. 4.1).

The first approach, the use of a leader sequence, has been applied for expression of AA10 members in *E. coli* with subsequent isolation of the mature protein from the periplasmic space^{138;139;246;307}. In chapter 2, the native secretion signal of *TrCel61A* outperformed all other options in terms of expression level and accurate His-1 processing. Therefore, the native secretion signal (NSS) of *Streptomyces coelicolor* LPMO10C was cloned upstream the mature protein. As an alternative, the secretion signal of the homologue AA10A (formerly Chitin Binding Protein 21, CBP21) from *Serratia marcescens* was selected, inspired by the successful expression of different AA10 members by Forsberg et al.¹³⁹. This second option was included because the origin of our target protein, *S. coelicolor*, is a gram-positive bacterium, lacking a periplasmic space. The required machinery to recognize the signal peptide, cleave it correctly and translocate the protein might substantially differ in *E. coli*. In order to investigate this, an *in silico* webserver (Signal P³⁰⁸) was used to predict the presence of a signal peptide and the expected position of cleavage. Following this output (Supplementary fig S4.1), the leader sequence of *SmAA10A* is recognized at the desired position in gram-negative and gram-positive bacteria, while ScLPMO10C signal peptide is correctly recognized in a gram-positive bacteria (its parent organism), but below the threshold of a gram-negative bacteria. However, increased values were displayed at the desired cleavage position.

Secondly, His-1 can also be obtained post-expression by protease cleavage^{26;139;144;309}.

Such a system and expression strain (in *E. coli* Origami 2 (DE3) cells) was provided by Ir. Barbara Danneels. The short peptide sequence MIDGR preceded the His-1 and can be cleaved by the protease factor Xa.

The protein was in both expression systems followed by a C-terminal 6xHis tag for purification purposes.

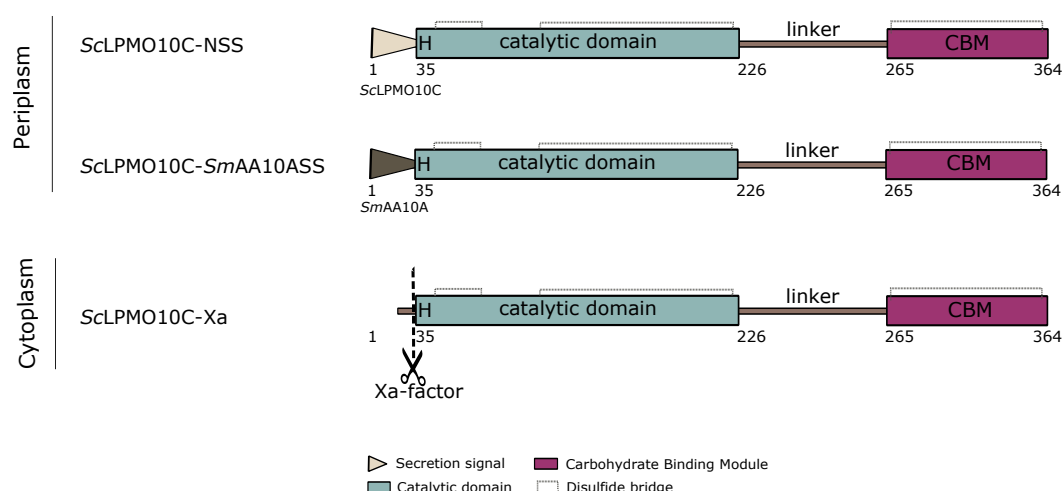


Figure 4.1: Schematic representation of different expression systems for ScLPMO10C.

Since it is difficult to predict where the leader sequence will direct the protein, the constructions including a secretion signal were attempted to isolate from the periplasm, cytoplasm and medium fraction. On the other hand, the Xa-cleavable construct will solely be isolated from the cytoplasm, where proteins are most commonly expressed in *E. coli*.

After finishing the molecular work, the construction with the *Serratia marcescens* secretion signal failed after repeated and different assembly methods such as the method from Gibson²⁶⁹, CPEC³¹⁰ and CLIVA³¹¹. Therefore, only the construct with the native secretion signal, ScLPMO10C-NSS, was compared to the construct with factor Xa recognition site, ScLPMO10C-Xa.

Expression level and purity

Because purity and expression level are important parameters for differential scanning fluorimetry, the proteins are his-tag purified and evaluated by SDS-PAGE.

ScLPMO10C-NSS is found in both cytoplasm and periplasm and not (abundantly enough) in the medium fraction (not shown on SDS-PAGE). A slightly larger amount of protein is isolated from the cytoplasm. ScLPMO10C-Xa, on the other hand, is found in the cytoplasm, in

a much larger amount. This is expected since the cytoplasmic space is the final destination of the protein and the protein amount builds up, while the protein is only on its way out in case of a secretion signal.

Protein isolation from the periplasmic space has several advantages over cytoplasmic isolation such as an exact N-terminal processing, lower protease abundance and easier purification because of lower protein content³¹². The latter is clearly displayed in the SDS-gels (Fig 4.2), where the periplasmic fraction contains very few other proteins at the start of the purification while the cytoplasmic fraction is overloaded.

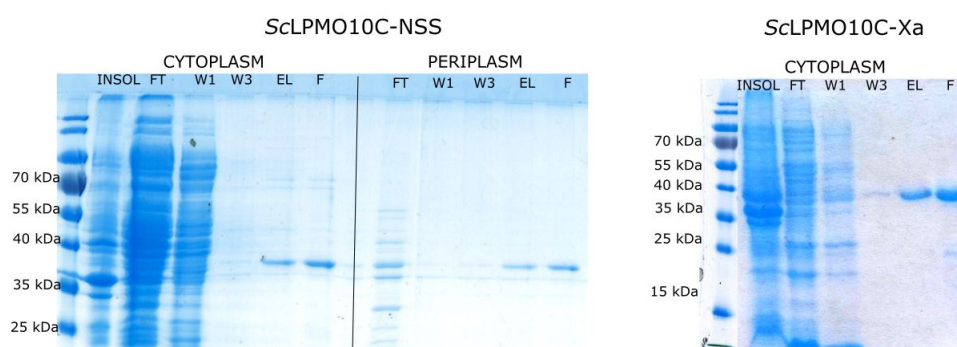


Figure 4.2: SDS-PAGE of purification for ScLPMO10C-NSS and ScLPMO10C-Xa. INSOL: insoluble fraction, FT: flow through from the soluble fraction loaded on the Ni-NTA column, W1: first wash fraction, W3: third wash fraction, EL: eluted fraction, F: final enzyme, buffer exchanged and concentrated via spin column.

An interesting observation in all SDS-gels is the size of the LPMO. The theoretical mass of the mature protein is 35.4 kDa, while this is 38 kDa for ScLPMO10C-NSS with leader sequence attached and 36 kDa for the uncleaved ScLPMO10C-Xa. The apparent weight of the enzyme on SDS-PAGE seems slightly higher and rather 40 than 35 kDa. To investigate this further, MS-analysis was performed at the lab of Prof. dr. Bart Devreese. This MS-analysis did not only confirm the correct cleavage of the secretion signal, but also the presence of 3 disulfide bridges. Indeed, molecular masses of 35384 and 35957 ± 1 Dalton were measured for the cytoplasmic fraction of ScLPMO10C-NSS and ScLPMO10C-Xa, respectively. These numbers are both 6 Dalton lower than the theoretically calculated masses, but can easily be explained by the formation of 3 disulfide bridges. Formation of disulfide bridges in the cytoplasm is highly exceptional³¹³, because an oxidizing environment and the help of proteins Dsb A-C is required. In the case of ScLPMO10C-NSS, produced in *E. coli* BL21 (DE3), the result can be most likely attributed to incomplete extraction of the periplasmic space and thus presence of those proteins in the cytoplasmic fraction that is isolated subsequently. For ScLPMO10C-Xa on the other hand, this is less unusual since it was produced in *E. coli* Origami 2 (DE3), a special strain to allow disulfide formation.

Activity evaluation

As second parameter, the activity of the LPMO, isolated from the different fractions was measured. A protocol from Forsberg and coworkers was used, since they first expressed the enzyme²⁶ and had proven the character to be a lytic polysaccharide monooxygenase, active on PASC.

In accordance with the results from the MS-analysis, confirming that the enzyme ScLPMO10C-NSS was correctly processed in both the cytoplasmic and periplasmic fraction, both fractions showed indeed activity on PASC (Fig. 4.3). For ScLPMO10C on the other hand, the LPMO only showed activity when the protein was cleaved with factor Xa. This result confirms the inevitability of this His-1 for activity and ScLPMO10C as type-1 LPMO (only oxidizes glucose chain at C1 carbon).

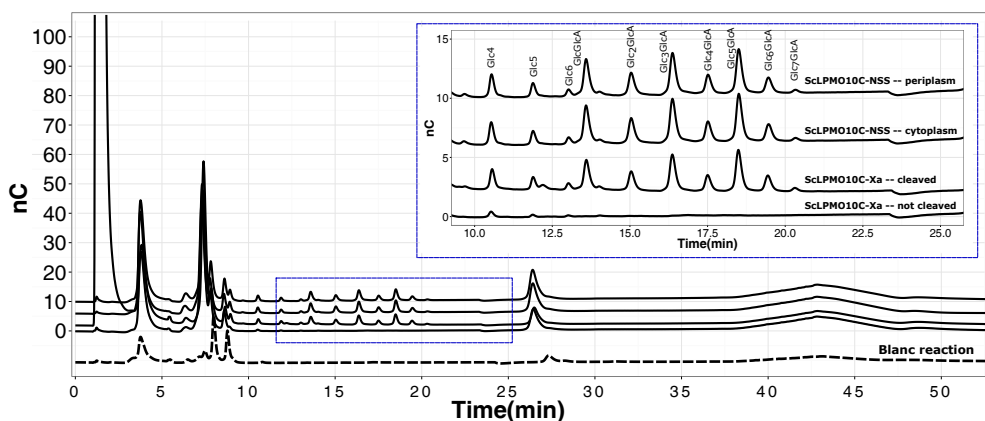


Figure 4.3: HPAEC-PAD trace of ScLPMO10C-NSS and ScLPMO10C-Xa. The dotted line represents the blank reaction (buffer instead of enzyme).

As a general conclusion, all 3 possibilities tested can be used for enzyme expression. Therefore a choice was made, based on practical considerations. Despite the higher enzyme yield, the construction ScLPMO10C-Xa was excluded because of extra process cost, time and purification required for the additional protease cleavage step. Even though ScLPMO10C-NSS isolated from periplasmic and cytoplasmic fraction showed the correct N-terminal processing and subsequent activity, the periplasmic fraction was preferred. The cytoplasmic fraction was less pure, hampering the buffer exchange and concentration procedure by quickly clogging of the filter columns.

4.2.2 Optimization of the purification protocol

The construction ScLPMO10C-NSS was selected as best suitable expression system. However, the protein yield was still quite low. The use of terrific broth (TB), developed by Tartof and Hobbs, provides an extended growth phase by use of a buffered, enriched medium³¹⁴. Obviously, the resulting higher cell densities can also improve recombinant protein yields in *E. coli*^{315;316}. In our experiments TB multiplied the protein yield by a factor 3 compared to the use of classical lysogeny broth (LB). Furthermore, a standard of 250 mL culture volume was purified to ensure enough protein.

However, the immobilized metal ion affinity chromatography (IMAC) purification protocol was already used in the first section of this chapter, it can be further optimized to obtain the highest possible purity and yield. Because an elution based on differences in pH could not improve the process, the imidazole concentration was varied. A fresh periplasm fraction was therefore washed with subsequent increasing imidazole concentrations going from 10 mM to 60 mM (10 mM steps), followed by 100 and 250 mM imidazole concentration. It is clear that most of the impurities are already washed with 20 mM imidazole. Increasing the imidazole concentration improves the purity, while at the same time more of the desired protein is lost. At a concentration of 60 mM, almost all ScLPMO10C is already eluted. Therefore a washing concentration of at least 20 mM and maximum 40 mM imidazole is suggested, while eluting with 100 mM imidazole is sufficient. The actual wash condition best to be used in this thesis is decided upon the applicability in the DSF-method below.

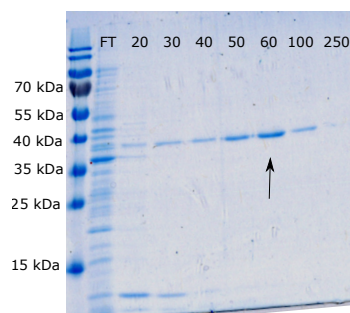


Figure 4.4: ScLPMO10C-NSS (periplasmic fraction) purification with increasing imidazole concentration (in mM). FT: flow through from the soluble fraction loaded on the Ni-NTA column. The arrow points to the imidazole concentration, that elutes most of the enzyme.

4.2.3 Validation of assays

In chapter 3, a measurement method was assessed for evaluation of stability and activity. Here, a validation of these methods with the LPMO ScLPMO10C is performed.

Differential scanning fluorimetry for measuring the apparent melting temperature

Thermodynamic stability is measured by differential scanning fluorimetry (DSF). This results in an apparent melting temperature (T_m), which is a parameter that describes a protein's tendency to unfold²³¹. This will be the main screening assay in chapter 6, working on stability engineering of ScLPMO10C. Because an exact and reproducible assay is the absolute premise for engineering purposes, the use of DSF for ScLPMO10C expressed with their native secretion signal will be evaluated here.

To assess the variability of the assay and the effect of purity and enzyme concentration on DSF, a set of wild-type enzyme cultures was grown and purified with different wash conditions. Three of these cultures were purified using a wash buffer containing 40 mM imidazole, while two more were purified using a wash buffer containing 20 mM imidazole. The last resulted in a higher enzyme concentration (~ 0.3 mg/mL) but as a drawback also a lower purity, while the first resulted in a lower enzyme concentration (~ 0.1 mg/mL), but a higher purity as explained in the previous section. Furthermore, one of 20 mM imidazole samples was diluted to the same concentration (about 3 times) as the 40 mM imidazole samples to see the effect of concentration. The apparent melting temperature of each enzyme was measured in triplicate, listed in table 4.1.

The biological and technical variability both equal 0 °C, corresponding to the identical values between erlenmeyers within the same wash condition and the identical values for every measurement of a certain erlenmeyer, respectively. This is an extremely low variability, making DSF an excellent method as screening assay for stability engineering. The largest error that can be estimated in this case, is the measurement error of the qPCR device, which is 1 °C, and will be taken as error further in this work, unless a larger variability was measured elsewhere.

Table 4.1: Set-up for measuring variability, effect of concentration and purity on DSF.

Erlenmeyer	Imidazole (wash) (mM)	Concentration (mg/mL)	T _m (°C)			Average (°C)	Stdev (°C)
			1	2	3		
A1	40	0.089	51	51	51	51	0
A2	40	0.073	51	51	51	51	0
A3	40	0.086	51	51	51	51	0
B1	20	0.298	50	50	50	50	0
B2	20	0.263	50	50	50	50	0
B2 diluted	20	0.088	51	51	51	51	0

A difference of 1 °C was observed for all enzymes with lower concentration, independent of the purity level. Because this is only a small difference that is equal to the measurement error of the qPCR device, this only suggests a dependence on protein concentration. However, the protein variants that will be compared to each other will be diluted to similar concentrations.

To decide upon what wash condition to use further, the quality of the actual melting curves was also evaluated (Fig. 4.5). The quality of the curve for both treatments is similar and as can be expected, a higher enzyme concentration yielded a higher relative fluorescence. Since the signal is only very low for the 40 mM imidazole washed samples, the final protocol will use a 3 x 8 mL wash buffer containing 20 mM imidazole.

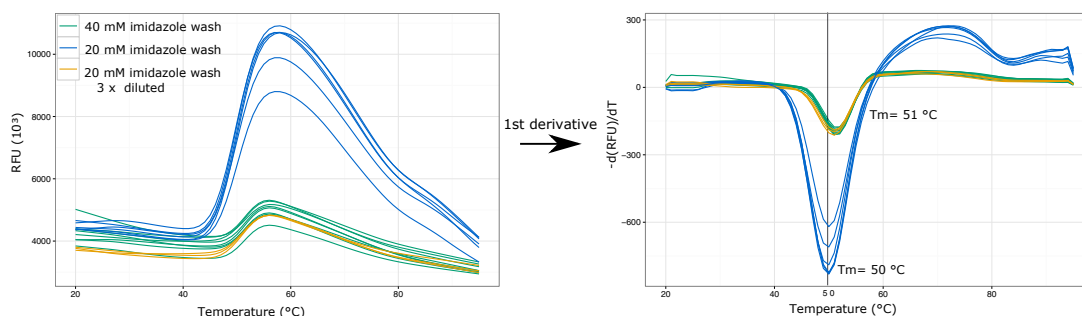


Figure 4.5: Melt curves and corresponding negative derivatives for different wash conditions. The inflection point of the left curve, corresponding to the minimum in the negative first derivative, is the apparent melting temperature of the protein.

HPAEC for analysis of reaction products formed from PASC

The activity of ScLPMO10C on the cellulosic substrate PASC was evaluated by incubation at 37 °C and analysis of the reaction products by high-performance anion-exchange chromatography (HPAEC). A time-course of the reaction is given in figure 4.6. This confirms that ScLPMO10C is a type-1 LPMO (as described in¹¹⁴), strictly producing neutral and C1-oxidized oligosaccharides. The degree of polymerization (DP) of maximum 6 for the neutral oligosaccharides and aldonic acids until DP 8 could be observed. Furthermore, the enzyme seems to have a preference to produce cellotetraonic acid (Glc₃GlcA) and cellosextaonic acid (Glc₅GlcA), which both count an even number of units, as seen for many chitin-active LPMOs from family AA10^{14;140;295} and was thought to be caused in those cases by the stereochemistry of chitin with alternating positioning of the acetyl-amine group. Moreover, the products with higher DP do not decrease over time, suggesting that contrary to *TrCel61A*, ScLPMO10C can not oxidize short oligosaccharides.

LPMO activity is known to be a slow reaction to be measured, as reflected in the slow product release in the time-course. Indeed, it takes 24 hours in this case to measure peak areas of > 0.250 nC*min, much slower than *TrCel61A* in comparable reaction conditions

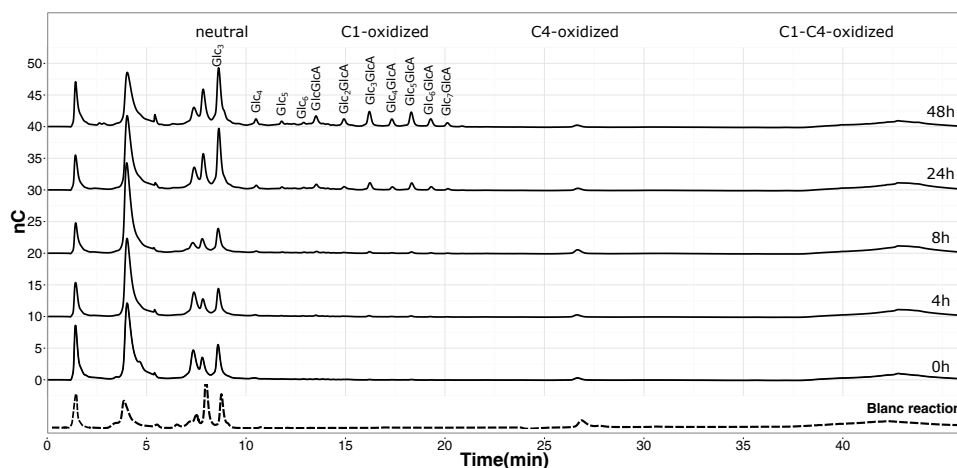


Figure 4.6: Time-course of a reaction of ScLPMO10C on PASC at 37°C. Samples were analyzed after 0, 4, 8, 24, and 48 h of reaction. Cellobionic acid (GlcGlcA) and cellotrionic acid (Glc2GlcA) were used as standard whereas the nature of the other oxidized products was inferred from the literature²⁶⁴. The blanc reaction is represented as a dotted line.

(see figure 3.3 on page 78). The underlying cause might be a preference for different substrates or electron donor (different specificity). The observed HPAEC-trace is practically identical to the one described by Forsberg et al.²⁶ for activity of ScLPMO10C on Avicel, which is the basic substance where PASC was prepared from. However, the result is somewhat surprising since PASC is much less crystalline than Avicel is.

4.3 Conclusion

Comparable to the expression of *TrCel61A* in chapter 2, using the native secretion signal of ScLPMO10C was shown to be the most suitable expression system here, despite the fact that an *in vitro* prediction was not convincing about the cellular localization. ScLPMO10C was isolated from the periplasmic space facilitating the purification procedure. However, the use of terrific broth medium and at least 250 mL culture volume were necessary to obtain a considerable protein yield. Interestingly, isolating the ScLPMO10C from the cytoplasm and/or using an N-terminal cleavable part proved to be useful alternatives. The formation of disulfide bridges and correct processing were confirmed by MS-analysis.

Considering purification and assay evaluation, a higher enzyme concentration prevailed even with slightly lower purity. The quality of unfolding curve was still high and a melting temperature could easily be deduced with a biological and technical variance of 0 °C if the same enzyme concentration was used. For different concentrations, the variability is 1 °C. For activity purposes, ScLPMO10C was incubated with PASC and the products were analyzed via HPAEC-PAD.

4.4 Materials and methods

4.4.1 Cloning of ScLPMO10C

The coding sequence for *Streptomyces coelicolor* LPMO10C (Uniprot Q9RJY2) preceded by its native secretion signal was synthetically ordered from GeneArt Gene Synthesis (Thermo Fisher Scientific) and codon optimized for *E. coli* expression by the same company (sequence in supplementary). The secretion signal of *Serratia marcescens* CBP21 was built in via primer sequence instead of the native secretion signal. The coding DNA sequence was cloned into the pCXP34 vector for constitutive expression³¹⁷ and a His₆-tag was immediately attached downstream the sequence via a three-piece Gibson assembly²⁶⁹. The three pieces were first amplified with Pfu Ultra AD high fidelity polymerase (Agilent technologies). The first piece amplified via polymerase chain reaction (PCR) included the coding sequence of the desired protein (bold bases) as picked up from the p-MAT vector provided by GeneArt. The his₆-tag (underlined) was built in the primer sequence (Primers in table 4.2). The second and third piece were amplified from the destination vector, pCXP34. The second piece started at the His6 built-in and ended in the middle of the beta-lactamase sequence while the third piece completed the beta-lactamase gene and ended right downstream the p34 promoter sequence. After *dpnI* digest and purification of the fragments (MinElute PCR Purification Kit, Qiagen), a Gibson assembly was performed during 1 hour at 50 °C. Next, 2 µL of the mixture was mixed with 20 µL electrocompetent BL21 (DE3) cells and subjected to electroporation (1.8 kV, 200 Ω, 25 µF).

Piece	Template	Primer sequence (5' → 3')
ScLPMO10C-Native secretion signal		
1	pMA-T-ScLPMO10C	FWD: CTCGAATTCGGAGGAAACAAGATGGTTCTGTCGTACCCGTC REV: CTCTCCCATATGGTCGACCTTAATGGTGATGGTGATGGTGCGGTGCAACACAACCAAT
2	pCXP34	FWD: ATTGGTTGTGTTGCACCGCACCATCACCATCACCATTAAAGGTCGACCATATGGGAGAG REV: CCCAACGATCAAGGCGAGTTACATG
3	pCXP34	FWD: TAACTCGCCTTGATCGTTGGGAACC REV: GACGGGTACGACGAACCATCTTTGTTTCCTCCGAATTCGAG
ScLPMO10C-SmCBP21 secretion signal		
1	pMA-T-ScLPMO10C	FWD: [TACCCTGCTCTCTCTGGGCCTGCTGAGCGCGGCCATGTTCTGGCGTTTCGCAAC- AGGCGAATGCC]CATGGTGTTGCAATGATGCC REV: CTCTCCCATATGGTCGACCTTAATGGTGATGGTGATGGTGCGGTGCAACACAACCAAT
2	pCXP34	FWD: ATTGGTTGTGTTGCACCGCACCATCACCACACCATTAGGTCGACCATATGGGAGAG REV: CCCAACGATCAAGGCGAGTTACATG
3	pCXP34	FWD: TAACTCGCCTTGATCGTTGGGAACC REV: GCCGAACATG[GCCGCGCTCAGCAGGCCAGAGAGAGCAGGGTACGGGAAGTT- TTGTTTCA]CTTTGTTTCCTCCGAATTCGAG

Table 4.2: Primers for cloning of ScLPMO10C in pCXP34-vector; Bold face: synthetic gene sequence, Underlined: his₆-tag, FWD: forward primer, REV: reverse primer, (XXX): secretion signal built in.

4.4.2 Expression and enzyme extraction

For ScLPMO10C-NSS, *E. coli* BL21 (DE3) cells were grown at 30 °C, 200 rpm for 18-20 hours in 250 mL terrific broth (TB) medium supplemented with 100 µg/mL ampicillin, if required. Terrific broth (TB) contains 1.2 % tryptone, 2.4 % yeast extract, 0.4 % glycerol, 0.017 M KH₂PO₄ and 0.072 M K₂HPO₄. The culture was inoculated with 1 % (v/v) of an overnight grown 5 mL preculture.

In the optimization phase, growth at 20 °C and 37 °C were also evaluated and the commonly used LB medium (1 % trypton, 0.5 % yeast extract, 0.5 % NaCl) was compared to TB. The protein was isolated from 2 different fractions: the periplasm or the cytoplasm. Both protocols are described below.

ScLPMO10C-Xa was cloned in a pET22b vector and expression was optimized in Origami 2 (DE3) cells by ir. Barbara Danneels. In this protocol, the cells were grown at 37 °C in LB medium until the start of the exponential phase (OD₆₀₀=0.6), when 1 mM IPTG was added for induction. The culture was further incubated at 20 °C for 60 hours, while shaking at 200 rpm.

Periplasmic protein isolation

The biomass was harvested by centrifugation at 3200 x g for 10 minutes (4 °C) and the pellet was resuspended by vortexing in 12.5 mL of the periplasmic fraction buffer (20 % (w/v) sucrose, 100 mM Tris-HCl (pH8) and 1 mM EDTA) and subsequently incubated on ice for 30 minutes. This cell suspension was centrifuged at 4500 x g for a duration of 20 minutes and the supernatant was stored as the periplasmic fraction. Next, the pellet was resuspended in a second solution containing 5 mM MgCl₂, to isolate the osmotic shock fraction. After 20 minutes incubation on ice, the suspension was centrifuged at 4500 x g for 20 minutes and supernatant was kept. These periplasmic and osmotic shock fractions were pooled and formed 1 periplasmic extract to be used. The pellet can either be used for a subsequent cytoplasmic extraction or can be discarded.

Cytoplasmic protein isolation

Subsequent to periplasmic protein extraction, cytosolic protein isolation was performed. The produced biomass was harvested by centrifugation for 20 min at 5000 x g and 4 °C. The cell pellet was stored for at least one night at -20 °C. After thawing the pellet, the cells were resuspended in 5 mL lysis buffer (PBS buffer containing 50 mM NaH₂PO₄ and 300 mM NaCl at pH 7.4, 1 mg/mL lysozyme, 10 µg/mL PMSF and 10 mM imidazole). This cell suspension was incubated on ice during 30 minutes and afterwards sonicated 3 times for 2.5 minutes (Branson sonifier 250, output 3, duty cycle 50 %). To remove the cell debris, the suspension was centrifuged at 8000 rpm for 1 h. The supernatant, containing his₆-tagged proteins was subjected to Ni-NTA chromatography.

4.4.3 Ni-NTA chromatography

All proteins were C-terminally His6-tagged for purification by Ni-NTA chromatography and this technique is applied after either cytoplasmic or periplasmic protein isolation. For these purifications, 1.5 mL of the nickel NTA-agarose slurry (MC-lab) was added to 10 mL columns. This was washed with 6 mL of sterile distilled water and consecutively equilibrated with 3 volumes of native binding buffer (PBS buffer = 50 mM sodium phosphate and 300 mM sodium chloride at pH 7.4). The sample was bound to the resin by adding 8 mL to the slurry and incubating at 4 °C for 30 to 60 minutes while slowly rotating. The resin is settled by gravity and the suspension is aspirated. More sample can be added to bind to the slurry if necessary and this can be repeated until all sample is loaded. Next, the resin was washed three times with 8 mL of BPS buffer (pH 7.4) containing 20 mM imidazole in PBS. Elution was performed with 10 mL of PBS buffer (pH 7.4) supplemented with 100 mM imidazole. After this, the buffer was exchanged to 50 mM Sodium-acetate buffer pH 5 in Vivaspin filter columns (10000 MWCO PES, Vivaspin, Sartorius).

The protein content was analysed by measuring the absorbance at 280 nm. The extinction coefficients for the desired proteins were calculated using the ProtParam tool on the ExPASy server (<http://web.expasy.org/protparam/>), as well as the theoretical molecular weight. Approximately 0.5-1 mg of protein was obtained from 250 ml of culture medium. Purity and molecular weight were verified by SDS-PAGE (12% gel).

For optimization purposes, the concentration of imidazole in the washing buffer and elution buffer was varied. Furthermore, another type of elution was tested, based on the change of protonation at different pH. The elution was hereby induced with 4 different buffers with decreasing pH, using 0.1 M citrate buffers with a pH 6 to 3. However, this method was discarded, as it was more extensive and proved no more efficient.

Protease treatment for ScLPMO10C-Xa was performed after purification and prior to activity test, following manufacturer's instructions (Factor Xa protease, New England Biolabs).

4.4.4 Enzyme assays

Differential scanning fluorimetry

The melting temperature of the purified enzymes was measured using differential scanning fluorimetry (DSF)²⁸⁵ in a CFX Connect96 cycler (Biorad) using the FRET channel (excitation / emission: 450 to 490 nm / 560 to 580 nm). Twenty µL of sample was then mixed with 10 µL of Sypro Orange (Sigma-Aldrich) in a 1/400 dilution in Sodium-acetate buffer pH 5.

All measurements were performed in triplicate. The temperature was increased from 25 to 95 °C with 1 °C / min increment. As the apparent melting temperature (T_m) is defined as the inflection point of the melting curve, the minimum value of the negative first derivative equals the T_m . This value is determined by the CFX Manager software (Biorad).

Activity on PASC and HPAEC-analysis

A 1.5 mL Eppendorf Tube containing 20-50 µg LPMO, 2 mM ascorbic acid as electron donor and 1,2 % PASC in a total of 1 mL of 10 mM ammonium acetate buffer pH 5 was incubated overnight at 37 °C in a thermoshaker while shaking at 1400 rpm. Samples (75 µL) were heated at 95 °C for 10 minutes for deactivation. Subsequently, they were cooled on ice and centrifuged for 10 minutes at max speed. The supernatant was diluted 10 times in milliQ water prior to analysis on HPAEC, using the method developed by Forsberg et al.²⁶. Identification of some products was done by running an in house 'ladder' containing neutral glucose chains, gluconic acid, cellobionic acid and cellotronic acid (Fig S2.5).

Supplementary material

Codon optimized sequences

- Secretion signal *Serratia marcescens* AA10A
ATGAACAAAACCTCCCGTACCCTGCTCTCTCTGGGCCTGCTGAGCGCGGCCATGTTCCGGCGTTTCG
CAACAGGCGAATGCC
- Secretion signal *Streptomyces coelicolor* LPMO10C
ATGGTTCGTCGTACCCGTCTGCTGACCCTGGCAGCAGTTCTGGCAACCCTGCTGGGTAGCCTGGGT
GTTACCCTGCTGCTGGGTCAGGGTCGTGCCGAAGCA
- Coding Sequence *Streptomyces coelicolor* LPMO10C (his₆-tag underlined)
CATGGTGTTCGAATGATGCCTGGTAGCCGTACCTATCTGTGTCAGCTGGATGCAAAAACCGGTACA
GGTGCACTGGATCCGACCAATCCGGCATGTCAGGCAGCCCTGGATCAGAGCGGTGCAACCGCACT
GTATAATTGGTTTGCCGTTCTGGATAGCAATGCCGGTGGTCGTGGTGCAGGTTATGTTCCGGATGG
CACCCCTGTGTAGTGCCGGTGATCGTAGCCCGTATGATTTTAGCGCATATAATGCAGCACGTAGCGA
TTGGCCTCGTACCCATCTGACCAGCGGTGCCACCATTCGGTTGAATATAGCAATTGGGCAGCACA
TCCGGGTGATTTTCGTGTTTATCTGACCAAACCGGGTTGGAGCCCGACCAGCGAACTGGGTTGGGA
TGATCTGGAAGTGAATTCAGACCGTTACCAATCCGCCTCAGCAGGGCAGTCCGGGTACAGATGGTGG
TCACTATTATTGGGATCTGGCACTGCCGAGCGGTCTAGCGGTGATGCACTGATCTTTATGCAGTG
GGTTCGTAGCGATAGCCAAGAAAACCTCTTTAGCTGCAGTGATGTGGTGTGTTGATGGTGGTAATGGT
GAAGTTACCGGTATTCGTGGTAGCGGTAGCACCCCTGATCCTGATCCGACCCCGACACCGACCGA
TCCGACAACCCCTCCGACCCATACCGGTAGCTGTATGGCAGTTTATAGCGTTGAAAATAGCTGGTC
AGGTGGTTTTTCAGGGTAGCGTGGAAGTTATGAATCATGGCACCGAACCGCTGAATGGTTGGGCAGT
TCAGTGGCAGCCTGGTGGTGGTACAACCCTGGGTGGTGTGTGGAATGGTAGCCTGACCAGTGGTA
GTGATGGCACCGTTACCGTTCGTAATGTTGATCATAATCGTGTTGTTCCGCCTGATGGTAGCGTTAC
CTTTGGTTTTACCGCAACCAGCACCGGTAATGATTTCCGGTGGATAGCATTGGTTGTGTTGCACCG
CACCATCACCATCACCATTAA

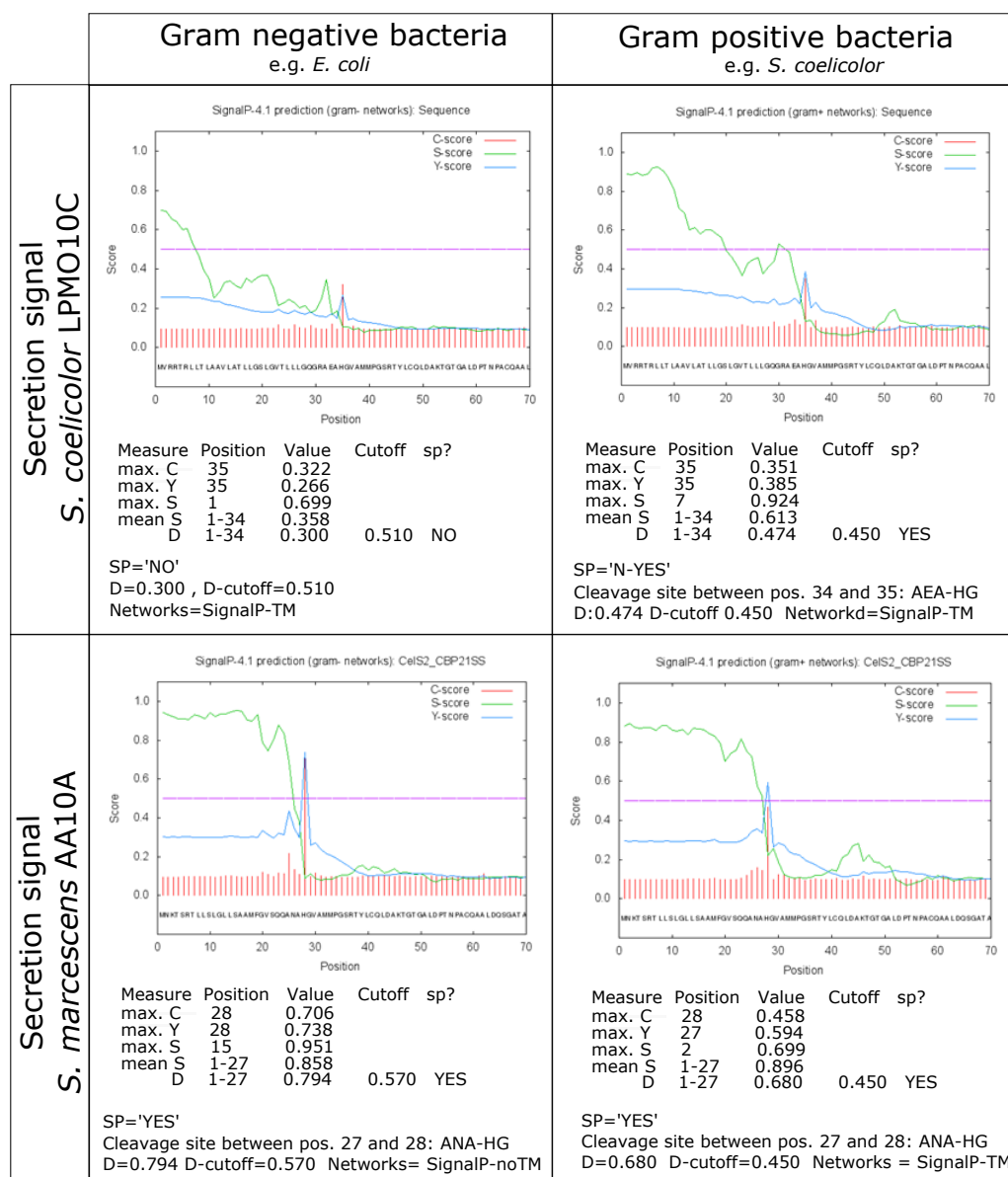


Figure S4.1: Prediction of presence and position of signal sequence with SignalP 4.1 server. The secretion signal of ScLPMO10C and and SmAA10A were both evaluated in gram-negative and gram-positive bacteria. C-score: recognizes signal cleavage position by an increased value, S-score: recognizes positions in signal peptide (high score) and in mature protein (low score), Y-score: geometric average of the C-score and slope of the S-score, mean S: average S-score of the putative signal peptide, D-score: discrimination value (weighted average of mean S-score and max Y-score).³⁰⁸

5

Thermostability engineering of *Trichoderma reesei* Cel61A

Abstract

Since the discovery of Lytic Polysaccharide MonoOxygenases (LPMOs) and their stimulating effect on cellulose degradation, numerous studies have already been published. Most of these focused on the structure and mechanism of LPMOs, but only limited research has been done on the stability of these interesting enzymes. Nonetheless, stability is a very important factor in industrial processes. Therefore, this chapter discusses some stability engineering strategies explored via the LPMO Cel61A from *Trichoderma reesei*.

Since *TrCel61A* is a multidomain protein, it was investigated whether removing the carbohydrate binding module (CBM) would increase stability and could serve as template for the engineering strategies. This analysis revealed that only the variant that preserved the glycosylated linker could be expressed well but a slight decrease in thermodynamic stability was found for this variant. Therefore, the most stable form, the full-length *TrCel61A* was taken as template for engineering strategies.

An initial study of the wild-type *TrCel61A* uncovered that both N-glycosylation at position Asn-158 and the native disulfide bridges contributed to the protein's stability. The first is responsible for a 2 °C increase in apparent melting temperature, while the latter is responsible for an unusual and unaltered activity after a 2 hour incubation at 80 °C. Inspired by these interesting findings, a reverse strategy of adding extra positions for such post-translational modifications was implemented. Even though all variants could be expressed in our host *Pichia pastoris*, they did not show an improvement in thermodynamic stability.

After building a phylogenetic tree, 2 consensus genes were also generated by using the statistical distribution of the different amino acids at each position of the catalytic domain. The resulting sequences were also fused to the linker and CBM domain of *TrCel61A*. Nonetheless, these *de novo* enzymes showed a significant decrease in thermodynamic stability. The MSA and the choice of the sequences therein is of utmost importance. A larger set, thus more variability, only decreased the T_m with 6 °C, while a small set of sequences with the same loop length has led to a decrease with > 20 °C. On the other hand, exchanging 2 small flexible clusters for their consensus counterpart did not notably alter the enzyme's stability.

Despite the negative outcome for the stability engineering approaches for *TrCel61A*, some interesting characteristics of the LPMO were unveiled.

5.1 Introduction

Cellulose, the main component of plant biomass, is the most abundant biopolymer on Earth and has therefore a huge potential as second generation renewable resource³³. However, the natural need for robustness in the plant cell wall results in recalcitrant behavior towards degradation. Bacteria and fungi, that can grow on cellulosic materials are therefore equipped with an entire cellulose degradation machinery, including a range of cellulases such as endo- and exocellulases, beta-glucosidases and (usually) a variety of the newly discovered lytic polysaccharide monooxygenases (LPMOs). This last group of enzymes has gained an enormous interest since its first mentioning^{117;163;318}. The promise of these LPMOs lies in the boosting effect that they exert on the classical cellulases by oxidatively cleaving the glycosidic bond in cellulose^{14;15;147}. In this way, they disrupt the crystalline structure and provide new chain ends for the cellulases to work on. As a consequence, LPMOs have become a compelling addition to the existing industrial cellulose degradation cocktails^{147;319}.

Although LPMOs are intensively studied these days, the majority of the studies deals with elucidation of the reaction mechanism, expression and characterization of new family members. Only very few studies include the stability of LPMOs, whereas this has been a very productive research area in the case of classical cellulases²¹. Increased temperatures can not only increase the reaction rate, solubility and availability of the substrate²² but also offer other advantages such as decreased viscosity and decreased risk of microbial contamination²³. Moreover, altering the stability (and solubility) of a protein can also improve the heterologous production of a protein³²⁰. Furthermore increased temperatures might promote cell-wall disorganization and thus better penetration of the enzymes²⁴. In that view, temperatures above the classical 50-55 °C for industrial lignocellulose breakdown are recommended^{321;322}. Besides, biomass is often pretreated in harsh conditions such as extreme pH, high temperature and high pressure. The classical (hemi)cellulases in the enzyme mixtures have therefore already undergone intensive thermostability engineering while LPMOs still need this experimental phase.

Little information about stability of LPMOs has been reported. Vaaje-Kolstad and coworkers reported that the apparent melting temperature (T_m) is 4 °C higher for *Serratia marcescens* CBP21 with disulfide bridges, compared to a variant without these linkages¹⁴². Furthermore, the metal ion in the active site seems to have a huge effect on the thermodynamic stability. Hemsworth and coworkers described for *Bacillus amyloliquefaciens* CBM33 a difference in melting temperature (T_m) of 20 °C for the apo-enzyme in comparison to a Cu(II)-bound form²⁴⁴. Also, Sprenger et al showed through *in silico* molecular dynamics simulations that ScLPMO10C and ScLPMO10B show a very similar stability behavior in

water as in ionic liquids²⁴⁸. For family AA9, some melting temperatures of *Neurospora crassa* enzymes were already measured²⁴⁵ and a very recent article describes the stability and halotolerance of two LPMOs isolated from a mangrove²⁴⁷.

Despite the discovery of 4 auxiliary activity families (AA9, 10, 11 and 13), only family AA9 and AA10 are considered important in the field of cellulose breakdown¹⁹. Therefore, the focus of this work is limited to stability engineering of a single representative from each family. In this chapter, stability of Cel61A (AA9) from the industrial important *Trichoderma reesei* is studied. The expression was optimized in chapter 2 and measurement methods in chapter 3. The next chapter focuses on an AA10 member from the soil degrading *Streptomyces coelicolor*, namely ScLPMO10C (AA10), whereof expression and measurement methods are optimized in chapter 4.

To date, predicting the effect of a mutation on the protein stability is still one of the most difficult problems, yet unresolved in protein science³²³. On top of that, no generic engineering method exist to increase stability, since mutations are always context dependent²³⁴ and often the concerted action of multiple mutations only leads to stability improvement³²⁴. To circumvent these issues, directed evolution approaches are often applied, followed by screening huge amounts of mutant enzymes in so-called libraries^{207;324–327}. The premise of this strategy however, is the availability of a fast and efficient screening method for activity, which is unfortunately lacking for LPMOs. Therefore, rational design will be applied, limiting the screening effort.

In this chapter, an increase in thermostability of TrCel61A was evaluated on the base of 3 approaches: (1) introducing additional N-glycosylation positions, (2) introducing additional disulfide bridges and (3) designing enzymes using the statistically most prevalent residues (consensus design). The first 2 are facilitated thanks to the eukaryotic nature of *Pichia pastoris*, that serves as expression host, while the latter is a sequence-based manner because of the lack of a crystal structure for TrCel61A. However, the chapter will start with some characteristics of the wild-type enzyme that provide the basis for the engineering strategies and evaluating the effect of removing the carbohydrate binding module of the LPMO.

5.2 Results and discussion

5.2.1 General stability exploration of *TrCel61A*

The first section gives a short overview of some stability-related characteristics that were observed in the wild-type enzyme *TrCel61A* and that form the reasoning behind the thermostability engineering strategies. First, the general build-up of the sequence is analyzed in detail, next the flexible regions in the protein are mapped as possible target residues. Finally, the effect of removal of N-glycosylation recognition sites and putative disulfide bridges on the stability of the protein is evaluated.

General build-up of the protein

For different LPMOs a crystal structure is available, although not for *Trichoderma reesei* Cel61A (UniProt O14405). Therefore, the amino acid sequence is the main source of information. Following the predictions on UniProtKB³²⁸, the LPMO consists of a catalytic domain (CD, His-22–Gly-256), glycosylated linker region (Ser-257–Pro-307) and carbohydrate-binding module 1 (CBM 1, Thr-308–Leu-343) to bind the cellulosic substrate. A schematic overview is given in figure 5.1.

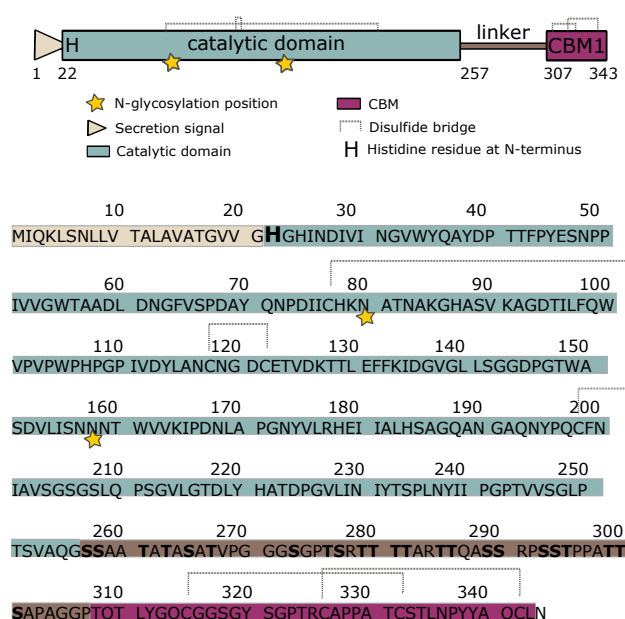


Figure 5.1: Schematic build-up and translated coding sequence of *TrCel61A*.

Serine and threonine residues are highly abundant in the linker region (bold face letters) as putative O-linked glycosylation positions and the region is also rich in proline and glycine

residues. The length of 50 residues is considerably long for a linker region. Additionally, 2 N-linked glycosylation positions can be found, namely Asn-80 and Asn-158 (signal sequence -N-X-S/T-). Furthermore, 2 disulfide bridges can be found in the catalytic domain (Cys-118–Cys-122, Cys-77–Cys-198) and two more in the CBM region (Cys-315–Cys-332, Cys-326–Cys-342). An N-terminal secretion signal (gray colored) leads the protein out of the cell and is cleaved off in order to obtain a functional protein with a histidine residue at the actual N-terminus (see chapter 2).

Mapping flexible regions in *TrCel61A* via B-factor analysis

If a crystal structure is available, this not only reveals the spatial organization of the amino acids in the protein, but also reports the error on the average position of each atom, the so-called B-factor, B-value or thermal displacement factor. Since the size of the electron smear around the average position they represent is linked to local flexibility³²⁹, residues with high B-factors form often targets for stability engineering strategies^{330;331}. This can be explained by the observation that protein unfolding often starts at regions with large thermal fluctuations, as studied via molecular dynamics simulations³³².

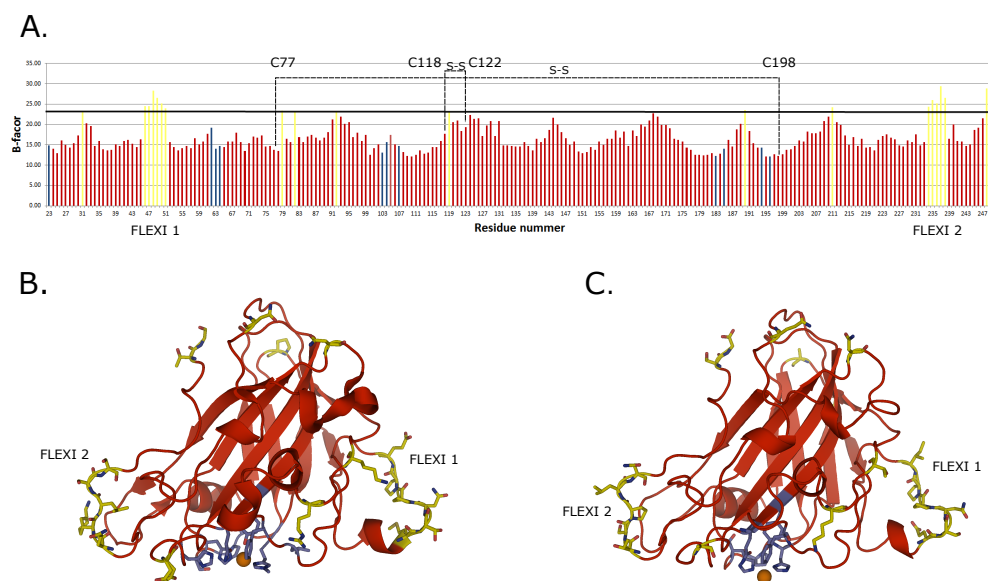


Figure 5.2: Flexible regions in *TaGH61A* and *TrCel61A* according to B-factor analysis. (A) *TaGH61A*: The yellow bars are the top 20 flexible residue (here, B-factor higher than 23), red bars are fairly stable and blue bars represent the residues within 6 Å from the copper ion. Disulfide bridges are also indicated with a dotted line. (B) *TaGH61A*: The flexible regions indicated in a cartoon representation wherein the cellulose binding surface is at the bottom. (C) *TrCel61A*: homology model with top 20 most flexible residues transferred from *TaGH61A*. Figures were prepared with PyMol⁷³.

Despite the benefits it could deliver, no crystal structure has yet been determined for *TrCel61A*. To circumvent this problem, a homology model was built, based on the closest homologue, *Thermoascus aurantiacus* GH61A (Uniprot G3XAP7, PDB 3ZUD-chain A). Because backbone flexibility is usually conserved within a family^{333;334}, the B-factors for *TaGH61A* were analyzed and the top 20 most flexible residues were transferred to *TrCel61A* via a structural overlay and manual comparison (Table S5.1). In figure 5.2, the B-factor of *TaGH61A* is given in function of the residue number. Yellow and blue bars represent the most flexible residues and residues within a distance of 6 Å from the copper ion, respectively. The most labile residues are mostly scattered throughout the sequence with two small clusters (FLEXI 1 and FLEXI 2), close to the protein's substrate binding surface. Panel (B) shows the same residues in cartoon representation for *TaGH61A*, while panel (C) depicts the homology model of *TrCel61A*.

Effect of removing native N-glycosylation positions

N-glycosylation takes place at asparagine residues in the motif Asn - Xxx - Ser/Thr (Xxx is everything except Pro), which is also called the N-glycosylation 'sequon'. Glycosylation takes place co-translationally when the protein enters the endoplasmic reticulum (ER) in eukaryotic organisms^{335;336}.

Different functions have been attributed to these N-glycans such as accelerating the folding (fourfold has been described)³³⁷, facilitating disulfide bond formation^{338;339}, protecting sites for protease attack³⁴⁰, preventing aggregation of (partially) unfolded proteins³⁴¹ and increasing the thermodynamic stability (about 3,1 kCal/mol)³³⁷. Furthermore, it can be required for the protein's functionality, due to an otherwise incorrect fold of the glycosylation-naïve protein³⁴² or it can increase the catalytic efficiency³⁴³.

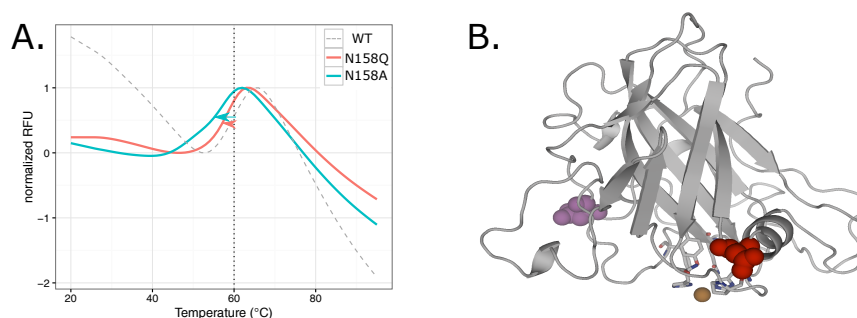


Figure 5.3: Effect of N-glycosylation N158 on thermostability of *TrCel61A*. (A) Melting curves of WT enzyme and variants N158Q, N158A. (B) Cartoon representation of the N-glycosylation sites: red, Asn-158; purple: Asn-80. Figures were prepared with PyMol⁷³.

To explore the function of N-glycosylation recognition site asparagine-158 (N158) in *TrCel61A*, two constructs were created by ir. Barbara Danneels that interchanged as-

paragine 158 for a glutamine residue (N158Q) or an alanine residue (N158A). The first mutation preserves the general properties but omits glycosylation, while the latter changes into a very small and flexible residue. The protein also harbors a second glycosylation position, but that was left untouched (N80, purple in cartoon of figure 5.3 B). The effect on activity and specificity is thoroughly investigated by ir. B. Danneels, yet not reported in this work. Only the effect on thermodynamic stability was evaluated here via measurement of the apparent melting temperature (T_m). A decrease in T_m (compared to wild-type) of 3 °C and 2 °C, respectively, was measured (Figure 5.3 A). Therefore, a reverse strategy of extra N-glycosylation positions in order to stabilize the protein is described in section 5.2.3 on page 122.

Effect of removing native disulfide bridges

As described above, two disulfide bridges are formed in the catalytic domain (CD) of *TrCel61A*. In order to evaluate the effect that they exert on the thermal stability, one of the cysteine partners was mutated into an alanine residue (C122A and C77A). However, none of these variants could be properly produced. Both SDS-PAGE and Western blot (against his₆-tag) failed to reveal any protein. Drastic decrease in protein expression has earlier been observed. For example, Yin and coworkers removed three disulfide bridges of a feruloyl esterase (cysteine to threonine) and observed for each single removal expression levels that were more than 10-25 fold lower in *P. pastoris*³⁴⁴. Similarly, alkaline phosphatase is shown to fail proper folding into its final and active conformation without disulfide bonds resulting in high sensitivity for protease attacks³⁴⁵. Additionally, free sulfhydryl groups are described to cause expression hurdles^{346;347} so that mutating both cysteines of the couple might have resulted in better yields.

In consequence, the disulfide linkages were chemically broken by addition of tris(2-carboxyethyl)phosphine (TCEP). Figure 5.4 A represents the melting curves of the wild-type enzyme and with addition of 0.1 and 1 mM TCEP, which are all identical, indicating that the disulfide bridges did not alter the melting temperature in this case. The location of the disulfide bridges is depicted in panel B.

The effect of the native disulfide bridges on the activity was also analyzed via HPAEC (Figure 5.5 A). Surprisingly, the activity after a 2 hour treatment at 80 °C did not differ significantly ($p=0.1$, Student t-test) from the activity of an enzyme incubated on ice for the same period of time. The enzyme thus retained or regained all its activity after treatment at this temperature, far exceeding the melting temperature. This highly remarkable activity is. Nonetheless, it was only observed in the presence of native disulfide bridges.

The activity tests for *TrCel61A* are preformed at 50 °C. To further explore this residual activity, tests were also performed at higher temperatures (50-80 °C) as depicted in panel

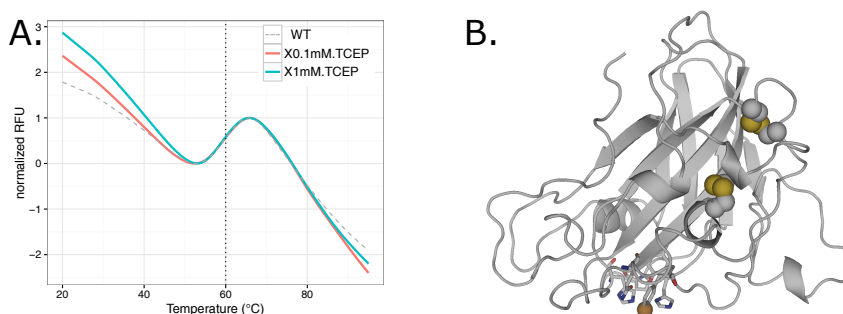


Figure 5.4: Effect of native disulfide bridges on thermostability of *TrCel61A*. (A) Melting curves of WT enzyme and with addition of TCEP. (B) Cartoon representation wherein the disulfide linkages are indicated in yellow spheres. The residues within a distance of the copper ion (orange sphere) are also indicated as sticks.

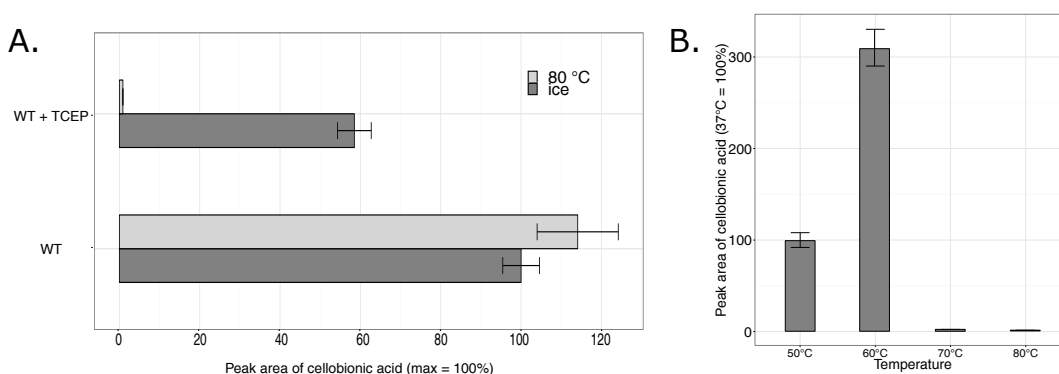


Figure 5.5: Activity of *TrCel61A* on PASC (A) after a 2 hour heat treatment at 80 °C and subsequently incubated for 2 hours on ice compared to identical enzyme aliquots incubated for 4 hours ice. The actual activity test is performed at 50 °C. The effect of the native disulfide bridges is studied by chemically breaking them with TCEP. (B) Activity measured at different temperatures (and without TCEP). The relative cellobionic acid production is given whereas the production at 50 °C after incubation on ice is 100 %.

B. This experiment reveals that no activity is left at 70 and 80 °C, suggesting that the protein loses its active conformation at a temperature of 70 °C and thus regains it again after cooling on ice, thanks to disulfide bridges. On the other hand, the LPMO retaining activity at this high temperature can not be excluded. The cofactor added here, ascorbic acid, is very unstable at this high temperature³⁴⁸ and also oxygen transfer is limited. These factors might be responsible for the lack or strong decrease in activity. Furthermore, figure 5.5 B also shows that activity tests performed at 60 °C yield higher product formation. Since this is very close to the apparent melting temperature, which is always determined without substrate, a stabilizing effect of the substrate^{349–351} cannot be excluded but needs further investigation. A recent paper describes indeed such a stabilizing effect for *B. amyloliquefaciens* AA10 in presence of its substrate chitin¹⁴⁴.

A very unusual property of remaining active at higher temperatures or regaining activity

after treatment at higher temperature is described here. The effect on the thermal stability of introducing more disulfide bridges will therefore be investigated in section 5.2.4 on page 124.

5.2.2 Removal of the carbohydrate binding module

TrCel61A contains a catalytic domain (CD), linker region and carbohydrate binding module (CBM). In general, about 40 % of the fungal (hemi)cellulases contains a CBM³⁵², while this number is similar for the LPMOs of family AA9. Indeed, approximately 30 % of the AA9 members contain a CBM1 and 4 % a CBM0 (unclassified CBM)¹¹⁰.

The precise role of this CBM is still poorly understood, although the most widely accepted theorem is that a CBM helps in substrate binding so that interaction time is increased and the biocatalytic reaction is promoted^{353;354}. Nonetheless, this seems not necessarily true in industrial set-ups, where very high substrate concentrations are added and very little free water is present. In that condition, a cellulase with and without CBM performs equally well³⁵⁵. Furthermore, some state that the function of a CBM not only relies on binding, but also decreases recalcitrance^{356;357}.

In addition, Payne and coworkers found that the linker region, that connects the CD with CBM, is not simply 'linking' both domains, but would aid cellulose binding in an unspecific way and thus contribute to activity of the protein³⁵⁸.

Hence, we wanted to investigate whether a truncated form of *TrCel61A* could be the starting point for the stability engineering effort and if so, whether the single domain protein has advantages in stability. To that end, various constructs were designed and created, devoid of linker region and/or CBM.

Construct design

Three main approaches were taken in designing truncated variants. A schematic overview of all constructs is given in figure 5.7. The starting point is the enzyme produced in chapter 2, *i.e.* native amino acid sequence of *TrCel61A* with 6xHis C-terminally attached for purification purposes.

- **Truncation based on domain prediction**

In the first construct, all residues after the predicted catalytic domain, as stated by UniProtKB (UniProtKB³²⁸, protein ref. O14405), were removed. This means, removing the complete linker and CBM (brown and purple part of figure 5.1, respectively) and attaching a his₆-tag directly to the C-terminal glycine 256. This construct was called *TrCel61A-CD-prediction*.

- **Truncation based on MSA and structural overlay**

A structural overlay (Pymol⁷³) and a multiple sequence alignment (MSA by ClustalW2³⁵⁹) were made to compare all AA9 members with known structure at that time. This analysis includes 6 other AA9 members (Table 5.12, except for *Neurospora crassa* LPMO9 F and C). The analysis revealed that the predicted CD from UniProtKB yielded

0



S

1 0 0

, , , : , , ,

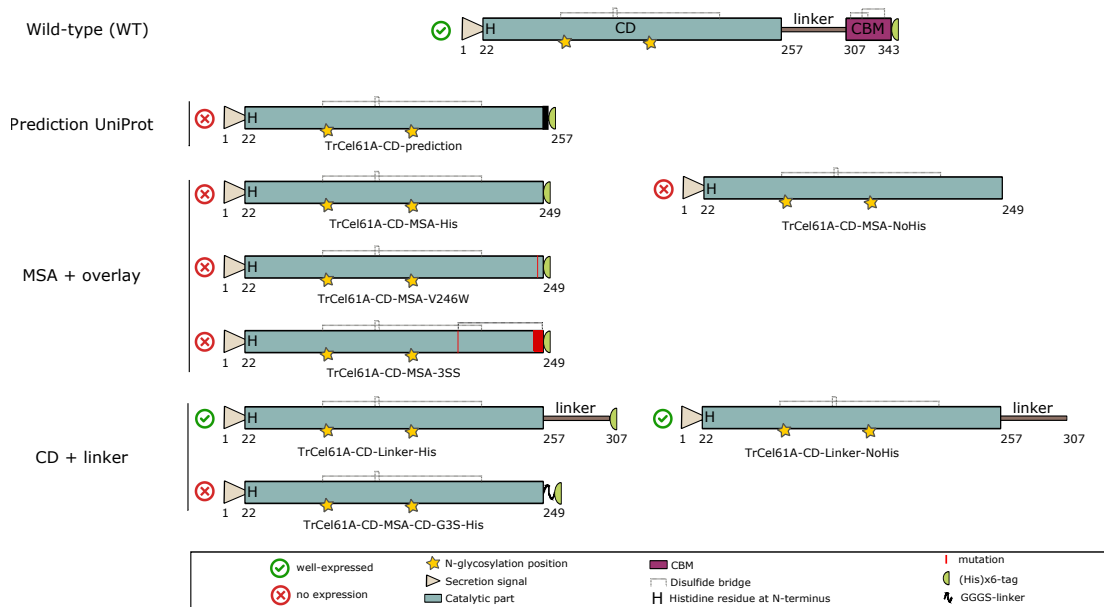


Figure 5.7: Schematic overview of the different constructs made for CBM removal.

Expression of the constructs

All constructs were successfully cloned in the pPpT4-vector downstream the native secretion signal, because this was found to be the most favorable in chapter 2. The first construct prepared was *TrCel61A-CD-prediction*. Considerable effort was done to express this enzyme. Nonetheless, no detectable amount of soluble protein was obtained. Therefore the 7 other constructs were created. Hereof, only the ones that preserved the glycosylated linker (*TrCel61A-CD-Linker-His* and *TrCel61A-CD-Linker-NoHis*) resulted in active protein expression with a yield about 30 % higher than the full-length protein. All other variants could not be detected in the medium fraction with analysis via SDS-PAGE, nor Western Blot for over 10 colonies tested per enzyme variant (indicated with a red cross in figure 5.7).

The theoretical size of the truncated proteins is about 30 kDa, while observed at about 50 kDa due to O- and N-glycosylation (SDS in Fig. S5.2). A possible explanation for the expression of only these proteins might lie in the presence of this glycosylated linker that improves solubility of the protein and can improve stability. Nonetheless, literature does not describe similar problems and only presents simple removal of linker and/or CBM. Linker regions are generally unstructured entities and often expected to be highly glycosylated. From the single signal obtained in differential scanning fluorimetry and the difficulty of removal of this CBM, some interaction between the catalytic domain and CD can be suggested. This interaction might be solely for stability purposes or also functional. A removal of O-glycans (via Jack bean mannosidase) could gain some more insight in the function of

the linker region.

For two truncations, a variant with and without His₆-tag is prepared. No difference was observed between the variants. Therefrom, we conclude the presence of the His₆-tag is not the cause of the expression deficit. Furthermore, separating the CD from the tag could not yield expression: neither the short flexible linker Gly-Gly-Gly-Ser, nor the 8 residue extension for the construct *TrCel61A-CD-prediction* showed improvement. It is known however, that the size and degree of flexibility of a linker region can be important and might influence the success in expression³⁶¹. Anyhow, the addition of glycosylation to the linker seems to improve the expression. An interesting experiment in that respect would be to remove the O-glycosylation from the linker region, for example with Jack bean mannosidase, and look at the effect on protein expression of the wild-type enzyme.

Surprisingly, adapting the C-terminus to existing LPMOs without CBM and introducing an extra disulfide bridge did not yield protein expression. Nevertheless, reducing the mobility of the extremities is known to improve stability. Examples exist for either fixation of N-³⁶² and/or C-terminus²⁴.

Effect on stability and activity

The thermodynamic stability (Fig. S5.6) and activity on PASC were also evaluated, as summarized in table 5.1. The variants without CBM showed a slightly reduced stability (2-3 °C) in comparison to the wild-type enzyme. It is generally thought that the CD and CBM are 2 separate entities that (un)fold independently from each other since they are spatially separated by a long, unstructured linker region³⁶³. However it was observed already that the full-length protein only displays one unfolding midpoint temperature, suggesting cooperative unfolding. That CD and CBM can influence one another is further proposed by Lu and coworkers, describing a β -mannanase, also containing a CBM1 and glycosylated linker³⁶⁴. They describe differences in secondary structure for proteins missing a CBM and/or linker, whereas the glycosylated linker was found to be conformationally stabilizing, reflected in a higher residual activity. Even though this is the opposite effect of what we have observed, both the stabilizing^{365–367} and destabilizing^{368–370} effect of having a CBM have been described before. Nonetheless, the stability decrease is very small. Also, Pellegrini and coworkers studied the stability of *Trichoderma harzianum* Cel7B and obtained very similar results for CD only compared to full-length enzymes although small differences can be observed, depending on pH³⁷¹.

The activity, represented as cellobionic acid production, had remained unchanged. This seems at first sight a surprising result since mostly activity strongly decreases when removing the CBM from a cellulosic enzyme^{354,372}. Hence, the CBM aids in attaching to the substrate and allows a longer interaction time with the substrate. Bennati-Granier seems to confirm this theorem by showing that different LPMOs from *Podospora anserina* containing a CBM1 in nature result in higher soluble product release than the ones naturally devoid of

Table 5.1: Thermodynamic stability and activity of catalytic domain variants. (Error = standard deviation)

Mutation	T_m (°C)				ΔT_m (°C)	Relative activity (%)
	1	2	3	Average		
WT	60	60	59	60 ± 1	-	100
TrCel61A-CD-Linker-NoHis	57	58	57	58 ± 1	-2	91
TrCel61A-CD-Linker-His	56	58	58	57 ± 1	-3	110

CBM¹²⁵. However, the story seems more complex for LPMOs. Indeed, the effect of CBM deletion in different studies appears to be rather modest albeit LPMO-, CBM- and substrate specific³⁷³. For example, Borisova and coworkers found that the CBM1 of *Neurospora crassa* LPMO9C contributed to xyloglucan degradation of the full-length enzyme, but had no effect on PASC digestion¹⁴⁶. Furthermore removal of the CBM in *Thermobifida fusca* E7 and E8 only caused a limited reduction in activity on PASC¹³¹. A second remarkable result is that despite *Streptomyces coelicolor* LPMO10C being only active on cellulose, a strong binding of the CD to α - and β -chitin was observed and only very weak binding to cellulose³⁰⁶. Thanks to the CBM attached, the LPMO can better bind cellulose. In addition, different ratios of neutral and oxidized products were observed for truncated LPMOs^{373;374}. Thus, CBMs have rather an effect on activity or substrate specificity, even though binding to a specific substrate does not imply activity.

In conclusion, these results and discussion clearly show that more research is needed for further elucidation of the role of CBMs in stability and activity of cell-wall degrading enzymes and in particular LPMOs. Even though the expression of the truncated variants was slightly increased and the activity remained unchanged, the stability showed a modest decrease in stability. Therefore, the template for stability engineering was decided to be the full-length enzyme. In that prospect, the experiments started with the most stable variant for more improvement.

5.2.3 Stabilization via N-glycosylation

It is known that N-glycosylation can improve protein stability³⁴⁰. Many studies on the removal of these N-linked sugars and their effect on stability can be found either by site-directed mutagenesis of the N-glycosylation recognition site^{375;376} or by enzymatic removal of the glycan in the final protein^{341;377}. For example, removal of putative N-linked glycosylation positions in glucoamylase from *Aspergillus niger* results in exposure of hydrophobic part of the protein and concomitant increase in aggregation and reduction in thermostability³⁷⁸, although no significant effect on secretion or thermostability was observed for α -amylase from *Aspergillus oryzae*³⁷⁹. On the contrary, much less studies can be found that introduce extra glycosylation positions. Although some examples exist. For example Clark and coworkers added two glycosylation positions from a homologue in α -glucosidase from *Horeum vulgare* and found a stabilizing effect for one mutation and no effect for the other³⁸⁰.

Selecting targets for N-glycosylation

Since N-glycosylation takes place at the motif Asn- Xxx (No Pro)- Ser/Thr, all possible N-glycosylation recognition sites that could be obtained by 1 single mutation can easily be selected searching the amino acid sequence (via python script). Next, all positions were considered on (1) spatial orientation in the homology model and (2) sequence conservation. For the first restriction, the cellulose binding surface, residues within 6 Å from the copper ion and residues close to a native glycosylation³⁷⁶ were avoided as well as buried side chains. For the latter constraint, the complete alignment of LPMO9-type 3 from the consensus strategy (see further) was inspected to avoid mutation of conserved residues, since mutating these can largely decrease stability and more importantly activity³⁸¹. A complete list of all possible mutational positions and their considerations can be found in table S5.2, whereas a compacted list with only the retained options are listed in table 5.2 and shown in figure 5.8.

Effect on thermal stability

All selected site-directed mutants for N-glycosylation were successfully expressed in *Pichia pastoris*. Nonetheless, a considerable variation in expression yield can be observed (reflected on SDS-PAGE gels in supplemental figure S5.3, wherein none of the samples were diluted to be easily comparable). All LPMO variants, except S206N, reveal a slightly larger estimated size after SDS-PAGE than the wild-type enzyme, confirming that at least 6 out of the 7 glycosylation sites were occupied. Clear melting curves were observed for all variants (figure S5.7) and their corresponding apparent melting temperatures, derived from their inflection points, are listed in table 5.2. Hereof, we can conclude that no significant differences ($p=0.17$, one-way ANOVA) were induced by the extra glycosylation positions.

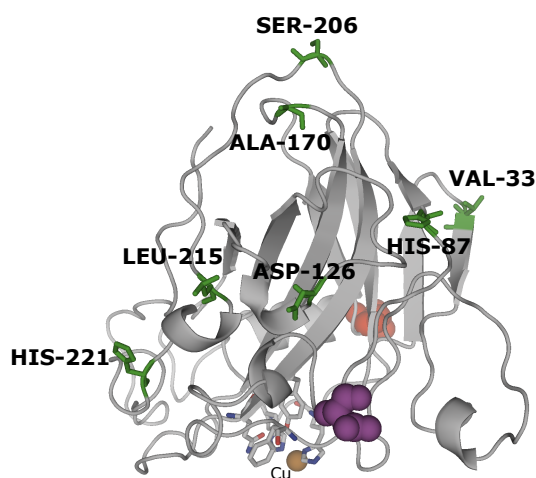


Figure 5.8: Overview of site-directed mutations for additional N-glycosylation positions, indicated as green sticks. The native N-glycosylations are indicated in spheres (red, Asn-158; purple: Asn-80). Residues within 6 Å from the active site are indicated as gray sticks.

Wang and coworkers showed that the impact of the deglycosylation effect on stability was dependent on the degree of glycosylation³⁴¹. Since *Tr*Cel61A already holds an O-glycosylated linker and two N-glycosylation positions, the additional glycosylation might not drastically induce a change in stability. Furthermore, it is known that not all glycosylation positions affect stability^{379;380;382}.

Table 5.2: Overview of stability mutants, aiming for extra N-glycosylation position. (n=3, error = standard deviation)

Nr	Start residue	Original sequon	Mutation	T_m (°C)	ΔT_m (°C)
WT	80	NAT	-	57 ± 1	-
	158	NNT	-		
1	87	HAS	H87N	56 ± 1	-1
2	126	DKT	D126N	57 ± 1	0
3	206	SGS	S206N	56 ± 1	-1
4	215	LGT	L215N	57 ± 1	0
5	221	HAT	H221N	55 ± 1	-2
6	31	NGV	V33S	56 ± 1	-1
7	168	NLA	A170S	56 ± 1	-1

In order to make sure the mutants did not dramatically change the activity, a quick test to determine the cellobionic acid production rate on PASC was determined (Figure 5.9). This graph only reveals little differences in activity so that no notable difference in stability, nor activity were reported here.

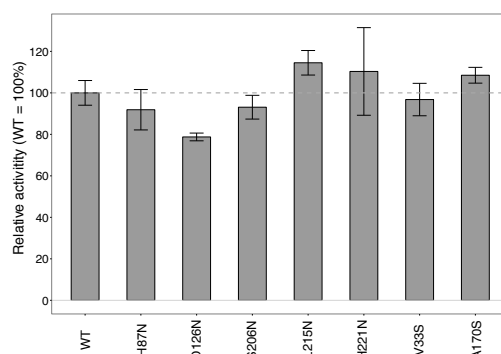


Figure 5.9: Activity of site-directed variants with additional N-glycosylation positions. Activity is measured on PASC at 50 °C, pH5 and followed during 8 hours. The analysis was performed with HPAEC and the slope of cellobionic acid production was relatively compared to the wild-type enzyme. (error = standard error on slope, calculated from 8 data points)

5.2.4 Stabilization via disulfide engineering

Disulfide engineering is a promising strategy to increase protein stability. For example, a rise in melting temperature of 10 °C was described by introduction of three disulfide bonds in Cel7A from *Talaromyces emersonii*³⁷⁷. Despite many positive reports, the outcome can hardly be predicted^{383–385}. Apart from the geometric requirements, such as suitable angles and distances, some guidelines for higher success ratios were reported: beneficial disulfide bridges might be found when (1) rigidifying flexible regions (target high B-factor)³⁸⁶, (2) covering a large loop length (25-75 residues ideally³⁸⁷) or (3) choosing residues close to the protein surface³⁸⁷. To keep up with all the restrictions and requirements, several tools exist to predict at what positions to introduce new stabilizing mutations. Some examples are DISULFIND³⁸⁸, Disulfide by design³⁸⁶, MODelling of Disulfide bridges in Proteins (MODIP)^{387;389} and MAESTRO³⁹⁰.

Selecting target residues for disulfide engineering

Putative positions for additional disulfide introductions were selected by using a combination of two computational tools, *i.e.* Disulfide by Design (DbD)^{386;391} and MODelling of Disulfide bridges in Proteins (MODIP)^{387;389}, a combination that has earlier been examined^{344;392}. The homology model was uploaded as input file, while usually the crystal structure is used. The latter tool, MODIP, ranks its disulfide predictions depending on the likelihood to be formed, with a grade A to D. In this way, six A-graded and 23 B-graded options were listed (figure 5.10). Hereof, Disulfide by Design predicted five of the A-graded options and 15 of the B-graded options. Despite the fact that only A-list members were explored earlier, also B-list options were included here. However, to increase the success ratio, an additional constraint was implied: at least one of the involved residues must be in

the top 20 most flexible positions (see table S5.1). Furthermore, all putative linkages that included a residue that is situated within a distance of six Å from the copper ion were excluded. A final selection of seven putative linkages was retained for *in vitro* analysis (table 5.3).

Noteworthy, the disulfide bridge that was built in earlier to remove the CBM (Page 117 and further), while obtaining good expression for our truncated protein, was only predicted in Disulfide By Design and not by MODIP.

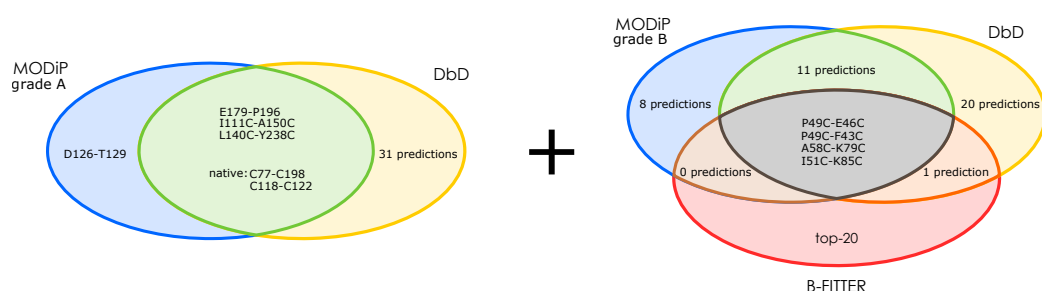


Figure 5.10: Distribution of predicted disulfide bridges. Left: combination of predictions made from DbD and MODiP grade A. The combinations from the green consensus were evaluated. Right: Combination of DbD and MODiP grade B and the most flexible residues. The combinations in the gray overlay were assessed.

Effect on thermodynamic stability

The LPMO variants were all expressed in *P. pastoris* with variable yields (SDS in figure S5.4). Noteworthy, the MODIP A-list predictions all showed a very low yield in comparison to the B-list and the wild-type enzyme. Hence, the assignment into MODIP categories is based on the steric hindrance the newly introduced cysteine side chain might cause to its neighboring residues. Since only a homology model was used as base for the calculations, the prediction has presumably a reduced accuracy.

Surprisingly, just like the N-glycosylation strategy, no improvement in apparent melting temperature, nor destabilizing effect could be observed (table 5.3). In addition, the computational services only predict the possible formation of a disulfide bond, without mentioning the effect on stability. Furthermore, the actual formation of the linkages is not confirmed here. The combination P49C-E46C, having only 2 residues separation, might have been supposed to be too short to obtain stability improvement in advance.

However, some differences in melting curve were observed. For an ideal unfolding, the fluorescent level starts at 0 because the dye, Sypro Orange, is quenched in aqueous environment. While heating the sample, the protein unfolds and at a certain point the fluo-

Table 5.3: Site-directed mutants with additional disulfide bridges and their apparent melting temperatures. The bold face and asterisk indicate the residues from the top 20 most flexible positions. (n=3, error = standard deviation)

Nr	Residue 1	Residue 2	MODIP-grade	DbD	T_m (°C)	ΔT_m (°C)
WT	C77 C118	C198 C122	A A	X X	57 ± 1	-
1	E179C	P196C	A	X	57 ± 1	0
2	I111C	A150C	A	X	56 ± 1	-1
3	L140C	Y238C	A	X	56 ± 1	-1
4	P49C*	E46C*	B	X	56 ± 1	-1
5	P49C*	F43C	B	X	56 ± 1	-1
6	A58C	K79C*	B	X	56 ± 1	-1
7	I51C*	K85C	B	X	57 ± 1	0

rescence quickly rises exponentially till a maximum. Hereafter, the signal decreases again because of aggregation. Now, it stands out that different curves have an aberrant form in comparison to the wild-type (Figure S5.8), indicating that a more complex phenomenon has been conducted. For example, variant P49C-E46C has a small plateau phase at its maximum. This might indicate a second maximum is followed by the first. Also, a small peak in fluorescent signal precedes the main peak in variant I51C-K85C. Furthermore, variant E179C-P196C starts with an extraordinary high initial fluorescence, higher than the maximum at transition temperature. In general, high fluorescence might suggest that the protein does not well behave at starting conditions, indicating (partially) unfolded or molten globule forms³⁹³. All three variants most probably will not unfold in a 2-state-pattern. A possible explanation might be the formation of wrong disulfide bridges or lack of cysteine pairing in at least part of the total protein amount. This suggestion is proposed since the T_m of the WT enzyme is in all cases also found. A more thorough investigation is required to understand the complete process but was not evaluated here since no stability improvement was obtained.

Secondly, we have investigated whether the activity was not completely destroyed by the introduction of the cysteine residues (Fig. 5.11). Based on HPAEC profiles, most variants show a comparable activity to the wild-type enzyme. One exception is found for variant E179C-P189C. This variant was already mentioned above: it has not only a low expression, weak unfolding transition difference combined with high initial fluorescence but also an impaired activity on PASC. The location of the involved residues might explain difficulties in the formation of the disulfide crosslink. E179 and P189 are located on 2 adjacent strands of an anti-parallel β -sheet in the inner core of the twisted β -sandwich. These so-called cross-strand cysteine pairs are not very common in nature, although few examples exist. Some researches claim that they would not impair the conformation of the sheet, although some preferences are found: one of the cysteines is best located at the edge of a strand and the bridge is most likely formed between non-hydrogen bonded amino acid pairs^{394;395}. The first requirement is violated, whereas only the latter is complied. A

second difficulty is most probably the very close proximity of the native Cys-198.

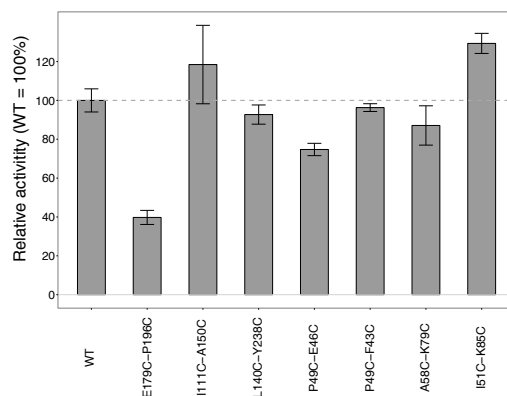


Figure 5.11: Activity of disulfide variants. The height of the bar indicates the relative cellobionic acid production rate as measured on PASC at 50 °C, pH5. (error = standard error on slope, calculated from 8 data points)

5.2.5 Stabilization via consensus engineering

Since no crystal structure is available for *TrCel61A*, the third strategy of consensus engineering is solely based on protein sequence. Although science does not (yet) understand all information encoded, this primary amino acid sequence is a very rich source of information. Indeed, it holds irrefutably and uniquely all instructions required for the protein's fold and biochemical function³⁹⁶

In consensus engineering an amino acid is replaced by the most frequent representative in a set of homologous sequences²¹³, based on the assumption that residues that are more prevalent at a certain position have a positive effect on protein stability²¹⁴. Among stability engineering strategies, the consensus concept has been proposed as a promising alternative for the frequently used directed evolution approach³⁹⁷. Indeed, positive results are faster reached because much larger steps in sequence space can be taken³⁹⁸. Moreover, the success ratio is remarkably high, considering 50 % of the residues exchanged for its consensus counterpart yield stability improvement^{214;227;381;399;400}. Likewise, Lehmann and coworkers propose a direct link for the fungal phytases, between residue conservation and thermostability²¹⁵.

The concept of consensus residues can be applied either by creating a full-length *de novo* protein, using the most frequent residue at each position^{216;217;401–404}, or by mutating only targeted positions^{213;218;219;403;405;406}. Both approaches will be utilized here after a short phylogenetic exploration.

Phylogenetic analysis

The first and most difficult task of consensus engineering is to decide on the set of sequences to build the consensus from. Indeed, Aerts and coworkers showed that the choice of sequences involved can have an influence on the final result²¹⁹. Therefore, a phylogenetic tree was built for the entire auxiliary activity family 9 in order to get some insight in the evolutionary diversity.

The AA9 family was downloaded from CAZy, yielding 263 sequences (September 2015). A multiple sequence alignment (MSA) was made with these sequences and subsequently a phylogenetic tree of the catalytic domains of all sequences was built using maximum likelihood (in MEGA6⁴⁰⁷). To link this information to the AA9 members with available crystal structure (table 5.4), a limited phylogenetic tree was built simultaneously containing all enzymes with available crystal structure (build with endsript 2.0⁴⁰⁸).

When comparing both trees and tracing back all members from the left tree in the right one, a high degree of similarity was found (Figure 5.12). Both trees nicely branched in 3 distinctive groups, corresponding to the 3 types of LPMOs that are described to date¹⁰¹. A similar partitioning was earlier observed in a phylogenetic analysis by Book and cowork-

ers¹¹⁰. Furthermore, the protein of interest, TrCel61A, can be found in the LPMO type 3 group, which is in accordance with the HPAEC traces in chapter 2. The sequence identities in table 5.4 suggest a quite low degree of identity within the family, and lower between the different oxidation types.

Organism	LPMO	Uniprot	PDB	% id	Active site	Type oxidizer	Source
<i>Trichoderma reesei</i>	TrCel61A	O14405	-	100	H22-H107-Y195	PMO-3	409
<i>Thermoascus aurantiacus</i>	TaGH61A	G3XAP7	3ZUD	56.83	H1-H86-Y175	? PMO-3	11
<i>Trichoderma reesei</i>	TrCel61B	Q7Z9M7	2VTC	49.78	H1-H89-Y176	? PMO-3	106
<i>Neurospora crassa</i>	NcLPMO9M	Q7SA19	4EIS	36.82	H1-H82-Y171	PMO-3	13
<i>Neurospora crassa</i>	NcLPMO9C	Q7SHI8	4D7U	35.86	H1-H86-Y166	? PMO-2	146
<i>Neurospora crassa</i>	NcLPMO9D	Q1K8B6	4EIR	30.50	H1-H84-Y168	PMO-2	13
<i>Thielavia terrestris</i>	TtGH61E	D0VWZ9	3EII	30.26	H1-H68-Y153	? PMO-1	15
<i>Neurospora crassa</i>	NcLPMO9F	Q1K4Q1	4QI8	29.59	H1-H72-Y157	? PMO-1	153
<i>Phanerochaete chrysosporium</i>	PcGH61D	H1AE14	4B5Q	27.92	H1-H76-Y160	? PMO-1	135

Table 5.4: Overview of AA9 member with revealed crystal structure. The representatives are ranked from high to low sequence identity (% id) to TrCel61A. The uniprot code, PDB code, active site residues that coordinate the copper and (presumable) type of oxidizer are also listed.

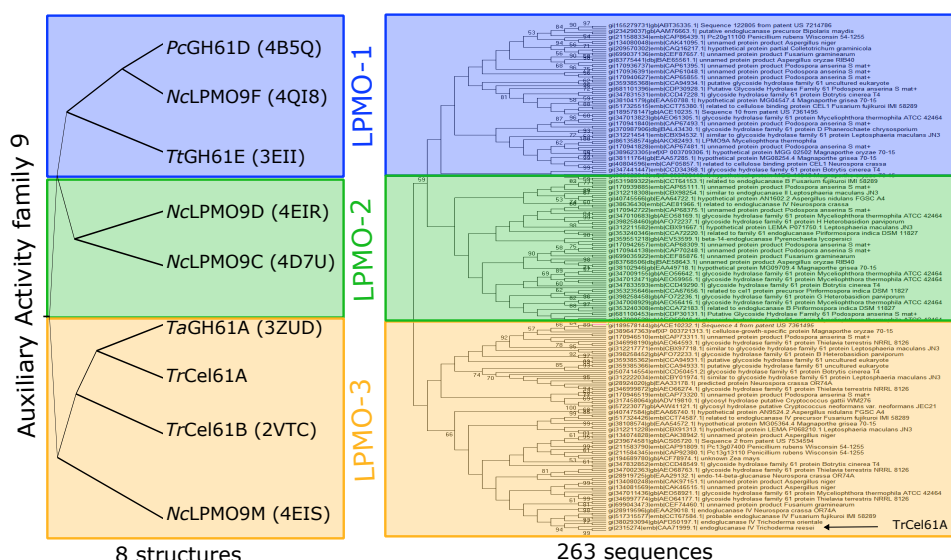


Figure 5.12: Phylogenetic tree of AA9 from all known crystal structures (left) compared to all CAZy members (right). A nice separation in 3 types of oxidizers can be found in both trees. TrCel61A is situated in LPMO-3 region, as expected from chapter 2. The tree was bootstrapped 1000 times and the frequency values > 50% are added.

Design of full-length consensus sequences

It is known that single mutations usually contribute to very limited stability increments, although combination of multiple mutations can make a big difference⁴¹⁰. Therefore, the first approach here is to create 2 full-length consensus genes, based on 2 different sets of input

sequences.

In order to avoid differences caused by specificity features, only the LPMO-type 3 subset that was adopted from the phylogenetic tree was kept. This list of 106 sequences was reduced with another 9 sequences by a 95 % sequence identity filter to compensate for phylogenetic bias⁴¹¹. Indeed, the key concept of consensus-based approach demands a set of independent sequences with identical fold. From the remaining set of 97 sequences, a full-length consensus gene was built, called *Consensus-AA9-LPMO3-95*, wherein the most prevalent amino acid was taken at each position, unless a gap was more frequent than any other residue. In that case, a gap was built in.

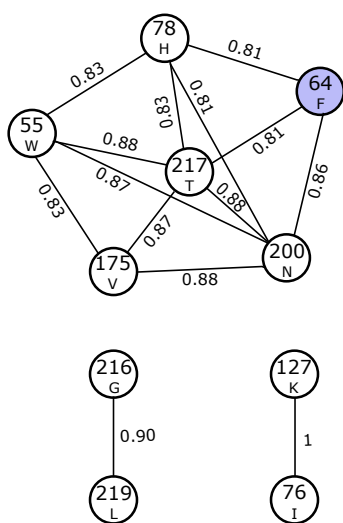
An MSA, however, holds much more information than sequence conservation only⁴⁰³ and positions in an MSA are not completely independent⁴¹². This is exemplified by the network of correlated positions²²¹, that can ameliorate the chances for a correctly folded protein when including them in the analysis^{403;413}. Combinations of residues with a correlation factor of 0.8 - 1 were manually inspected (Figure 5.13), but allowed combinations were preserved at all involved positions in the consensus enzyme. Secondly, the N-glycosylation position 80-NAT, was only present in 16/97 sequences. Although the motif GAT, present in 30/97, was more prevalent, it was adapted to 80-NAT because we have proven earlier that this glycosylation contributes to thermostability. After codon optimization for *Pichia pastoris*, this final consensus enzyme, *Consensus-AA9-LPMO3-95*, was ordered synthetically and cloned in the pPpT4 vector with and without the C-terminal linker and CBM region from *TrCel61A*. The catalytic domain (CD) shared 68 % identity and 79 % similarity with *TrCel61A*.

As a second consensus sequence, the former alignment was further reduced by discarding the sequences with different loop lengths because the quality of the MSA is a crucial factor in the consensus concept. Likewise, an MSA becomes more challenging in loop regions, especially when they span different lengths and in weakly conserved regions. Moreover, it has been shown that longer loops usually result in lower stability⁴¹⁴. The sequences were manually and stepwise removed, while the alignment was rebuilt for every 5 sequences removed from the former alignment. In order to preserve a decent amount of sequences, 'same loop length' was defined as not more than 2 subsequent gaps in comparison to the target enzyme *TrCel61A*. This approach resulted in a reduction to 40 sequences only. The size of the sequence collection varies among different studies with various outcome. As little as about 20 close homologous sequences have been successfully served in an MSA^{215;397;410}, while other successful stories used 31⁴⁰² to 100²²⁷ and even up to > 600 sequences⁴⁰³. Again to ameliorate the quality of the final enzyme, correlated positions were taken into account. Remarkably, different correlation networks were generated (figure 5.13), containing all different residues. More surprising, two couples in this consensus en-

zyme were new to nature, namely P71-K189 and P20-A41. These were manually adapted to existing couples, namely E71-K189 (17,5 % appearance) and P20-D41 (15 % appearance). Furthermore the combination D50-T151 had a very low prevalence of only 2.5 % so that it was exchanged for to S50-T151, which has a tenfold higher prevalence (22.5 %). Moreover, the first N-glycosylation recognition site, *i.e.* 80-NAT, was also violated in this consensus enzyme (DAT in 11/40 sequences, while NAT in 10/40 sequences) so that it was manually adapted. The final resulting enzyme, *Consensus-AA9-LPMO3-95-loops*, shared 73 % sequence identity and 84 % sequence similarity to *TrCel61A*. The protein was also expressed as CD only and CD linked to CBM of *TrCel61A*.

Both consensus enzymes (alignment in figure 5.14) shared a high identity and similarity of 85 % and 90 %, respectively. Furthermore, following the correlation networks, 3 correlations are violated in *TrCel61A*. The residue pair E202-K89, from correlations of the same loop length collection, shows a clear ionic pair interaction, also spatial close to each other in the enzyme model, that could contribute to enzyme stability. Simple introduction of this network in *TrCel61A* is a suggestion for stability evaluation, but was not evaluated here.

consensus-AA9-LPMO3-95



consensus-AA9-LPMO3-95-loops

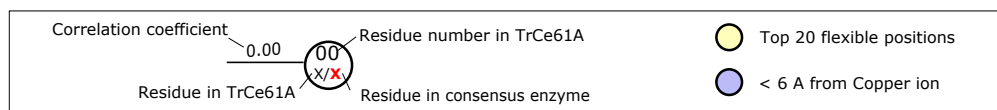
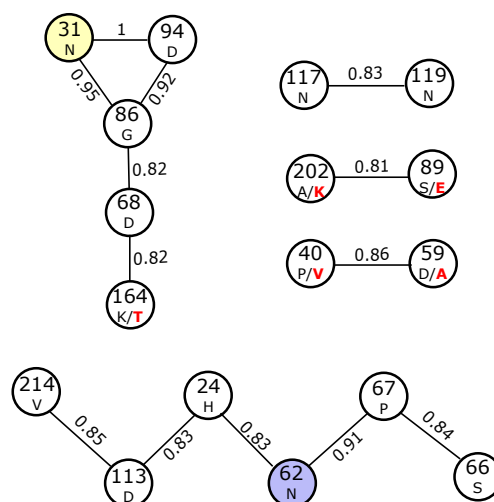


Figure 5.13: Correlation networks extracted from the 2 different MSA's for consensus engineering. Only correlations with a correlation coefficients between 0.8 and 1 are shown.

Design of flexible loop exchange

Besides creating a *de novo* full-length enzyme, B-factor analysis was used to identify so-called *hotspots* for mutagenesis. Only the two most flexible regions in TrCel61A, FLEXI 1 and FLEXI 2 (figure 5.2 on page 112), were exchanged for their consensus counterpart described above (figure 5.14). This resulted in three enzyme variants of TrCel61A: (1) FLEXI 1 mutated into 46-MSNPPE-51 (2) FLEXI 1 mutated into 46-MSNPPT-51 and (3) FLEXI 2 mutated into 233-QSLSS-237. Since both consensus enzymes only differed in position 236 and serine shares the same side chain properties with threonine, only 1 variant was prepared to represent both options.

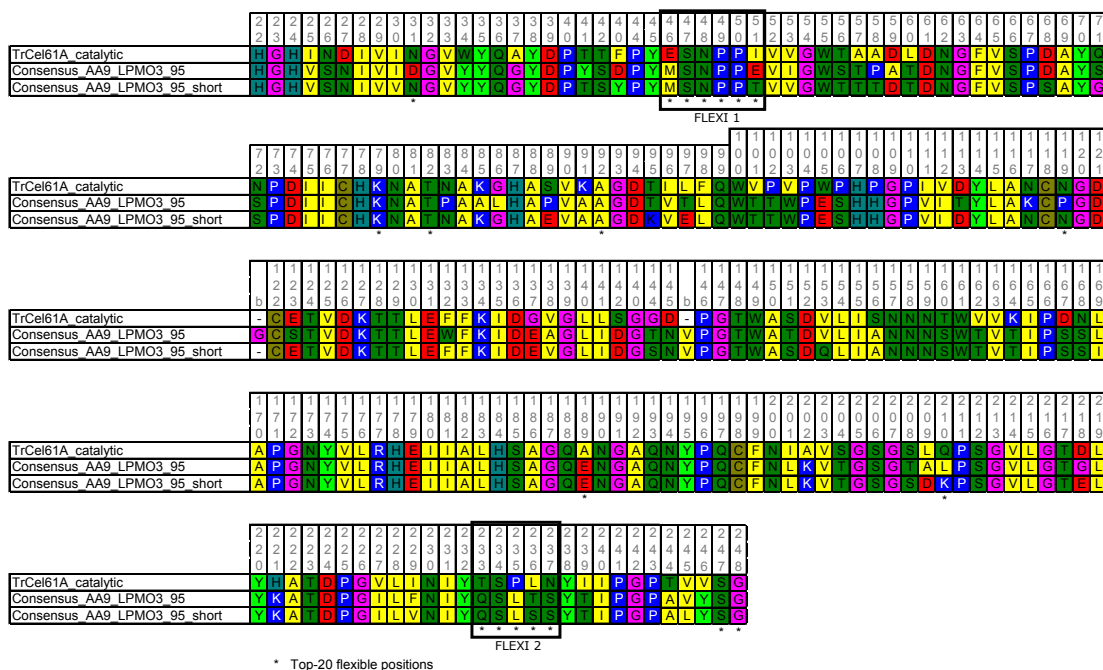


Figure 5.14: Multiple sequence alignment of TrCel61A and the 2 consensus enzymes. The top 20 most flexible residues are indicated with an asterisk as well as the two most flexible loops from the B-factor analysis (figure 5.2 on page 112) are indicated. The residues are colored according to their side chain properties.

Effect on thermodynamic stability

After completing the molecular work, the host *Pichia pastoris* was transformed with the final plasmids containing the coding sequences of the enzyme variants for enzyme production. The yield and purity was compared via SDS-PAGE (SDS-gel in figure S5.5). The enzymes from the flexible loop exchange show a very similar yield to the wild-type enzyme, while the full-length enzymes show a larger variation. Although the expression of the CD only seems more difficult because less intense bands were observed. A complete loss of activity was

observed earlier in this chapter when trying to remove the CBM (see page 117 and further).

In order to evaluate the thermostability, the apparent melting temperature was measured and listed in table 5.5 (melting curves can be found in supplementary figure S5.9). The thermal unfolding curves for the full-length consensus enzymes appeared distorted and therefore inconclusive. As described in chapter 3, a consecutive immobilized metal ion affinity chromatography (IMAC) purification step after the standard ultrafiltration can ameliorate the results and indeed nice melting curves were obtained afterwards.

In accordance to the yields obtained for the loop-swap variants that were similar to the wild-type enzyme, the melting temperatures are also highly comparable. The changes made in the sequence are also rather small so that little effects can be expected. The *de novo* designed variants, on the other hand, show a distinct and clear decrease in thermodynamic stability with a reduction of 6 °C and > 20 °C in T_m for a consensus enzyme calculated from a set of AA9 sequences with LPMO-3 specificity or the same set reduced to sequences with identical loop lengths, respectively. The size of the effect on thermostability is comparable for both the catalytic domain only and the variant linked to the linker and CBM from TrCel61A. This suggests that the main unfolding action is initiated in the catalytic domain. A surprising non-two-state unfolding pattern is found for the catalytic domain only of this last consensus enzyme with a transition temperature at 40 °C, like the variant complemented with at CBM, and 62 °C, like the WT enzyme. Such a phenomenon can result from a different unfolding event for distinct domains in the protein^{415;416} or can be pointed to the presence of different forms of the enzyme⁴¹⁷.

Strategy	Name	configuration	T_m (°C)	ΔT_m (°C)
WT	-	-	62	-
Full-length consensus	Consensus-AA9-LPMO3-95	CD - linker - CBM	56 ± 1	-6
		CD	56 ± 1	-6
	Consensus-AA9-LPMO3-95-loops	CD - linker - CBM	39 ± 1	-23
		CD	40 ± 1 (61 %)	-22
			62 ± 1 (39 %)	0
Loop exchange	46-MSNPPE-51	CD - linker - CBM	62 ± 1	0
	46-MSNPTT-51	CD - linker - CBM	59 ± 1	-3
	233-QSLSS-237	CD - linker - CBM	61 ± 1	-1

Table 5.5: Apparent melting temperatures (T_m) of consensus based enzymes. When multiple peaks appear, the relative percentage of appearance, deduced from peak height, is given between brackets. CD: catalytic domain, CBM: carbohydrate binding module (n=3, error = standard deviation)

Protein folding is a very complex process that is not completely understood yet⁴¹⁸. Obtaining a correctly folded and thus well-expressed *de novo* enzyme is admittedly already an achievement, even for a consensus enzyme. The decrease in stability can be pointed to the nature of the sequences. Cellulase-producing fungi mostly live in mesophile environments⁸⁶ and some in thermophile environments^{419;420}, wherein an apparent melting tem-

perature of ~ 60 °C for *TrCel61A* is already considerable high. Preparing the consensus sequence of the entire family most likely 'averages' all positions so that many compensatory mutations will be introduced too⁴⁰³. The underlying cause is that positions with low sequence conservation contain little information so that taking the consensus residue is most likely not stabilizing⁴⁰⁶. The effect is larger for the subset with same loop-lengths, so that this selection most likely also contained less stable enzymes. Often a vote weight is assigned for different sequences from the same organism to compensate for phylogenetic bias^{215;219}. Such an approach might have improved our results.

The combination of B-factor analysis and consensus approach seems logical and showed already positive outcome³³¹. However, infinite possible strategies and variations on strategies are possible. One alternative to our approach here would be to look only at thermostable enzymes. For example, one could swap the flexible region for the respective region of a thermostable enzyme. Such a design was shown successful for a *Bacillus subtilis* esterase⁴²¹. Differently, the entire L2-loop might also be exchanged for the shortest representative in the family. Also, the other, non-clustered top 20 residues could be mutated for their statistically most prevalent residue²²⁷ or for multiple 'allowed' residues at that specific position. When inspecting the alignment of the consensus enzymes in figure 5.14, both consensus enzymes show some similarities, distinct from *TrCel61A*. Combining this information to the protein model can reveal some interesting mutagenic positions. One example would be to mutate the couple A189-H121 into E189-K221. This last salt bridge was present in both consensus enzymes but lacked in *TrCel61A*. Spatially, they would be close enough to each other, whereas the C- α carbons are separated at only 10 Å in the model. A more thorough study can certainly reveal much more interesting options.

As for the other strategies, the activity was also checked, depicted in figure 5.15. The loop swap variants are in accordance to the melting temperature and expression level, also highly similar in activity to the wild-type *TrCel61A*. On the contrary, the full-length enzymes show a distinct effect. This activity is measured at 50 °C, what is still in the reach of the LPMO-AA9 based enzyme, but surprisingly above the T_m of the enzymes from the same-loop collection. Furthermore, the LPMO-AA9 based enzyme shows a 10 times higher activity due to addition of the CBM. Hence, this could be explained by arguing that the CBM accommodates the substrate. However, the inverse is observed for the variant created using only sequences with same loop length. The variant without CBM showed a 3-fold higher activity. However, its thermal denaturation curve contained a second unfolding event at 62 °C. The presence of this more stable form can be responsible for the higher activity, while completely compensating for the benefits of the CBM. In addition, this variant with CBM still shows a comparable activity to the WT enzyme, even though its T_m is 40 °C. This suggests, as earlier mentioned for the wild-type enzyme, a stabilizing effect after adding the substrate.

These results need further investigation to find the basis of this interesting observation. A full temperature profile of the consensus enzymes in both activity tests would provide very valuable information.

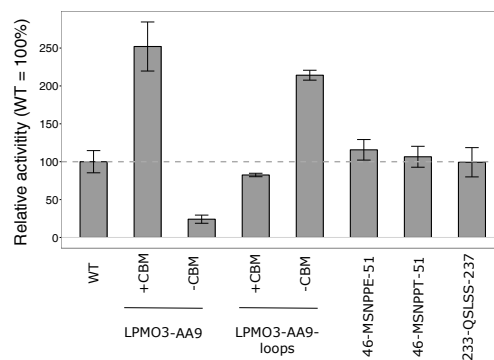


Figure 5.15: Activity of consensus-based variants. (error = standard error on slope, calculated from 8 data points)

5.3 Conclusion

In this chapter, the full length *TrCel61A* was taken as template for stability engineering since removing the CBM appeared to be a cumbersome task, slightly lowering the thermodynamic stability. Three distinct approaches for stability engineering were explored, whereas none of them could improve the thermodynamic stability. The variants of the consensus design even showed a very large decrease (~ 20 °C) in apparent melting temperature. However, they showed interesting increases in activity that need further investigation. Despite this negative outcome, some very interesting stability characteristics of the LPMO were unveiled such as a very unusual and intriguing characteristic of the native disulfide bridges. Indeed, although the wild-type LPMO shows an apparent melting temperature around 60 °C, the protein still shows an unchanged activity after a 2 hour treatment at 80 °C most likely due to the presence of native disulfide bridges. Furthermore a possible stabilizing effect of the substrate was suggested. These very interesting results still need further investigation.

5.4 Materials and methods

The experimental procedure for SDS-PAGE, Western blot analysis, his-tag purification can be found in chapter 3. A more elaborate procedure on *Pichia pastoris* transformation, selecting the best producing strain, enzyme production and purification can be found in chapter 2.

5.4.1 Molecular work

Cloning of full consensus genes

The full consensus genes were ordered from GeneArt (Invitrogen) after codon optimization for *Pichia pastoris*. They were delivered in a pMA-T vector, a simple vector containing an ampicillin resistance gene (β -lactamase) and a Col E1 origin of replication. The 2 consensus genes were cloned via circular polymerase extension cloning (CPEC³¹⁰) in the pPpT4-alpha-vectors downstream the secretion signal of *TrCel61A* and a His₆-tag was c-terminal attached to the gene (figure 5.16). The genes were expressed as modular enzyme, completed with linker and CBM region (panel A) originating from *TrCel61A* and as catalytic part only (panel B).

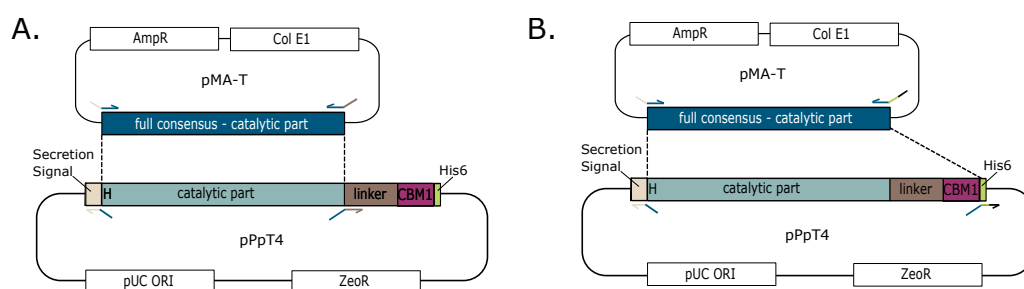


Figure 5.16: Schematic overview of cloning strategy for full consensus genes. The location of the primers is also indicated whereas the color of the arrow indicates the region of the sequence they represent. The upper vector is the pMA-T vector, obtained from GeneArt with the synthetic genes. The pPpT4-vector contains the wild-type *TrCel61A* as used for expression in this work. (A) Cloning strategy to obtain a modular protein, including linker region and CBM. (B) Cloning strategy to obtain a catalytic part only.

Prior to the CPEC assembly, the insert (full consensus gene, picked up from the pMA-T vector) and backbone (pPpT4-vector) were first amplified using Primestar GXL polymerase (Takara Bio, inc) with the primers indicated in table 5.6. Subsequently, the fragments were purified with innuPREP PCRpure kit (Analytik Jena) and eluted in 12 μ L milliQ water. The fragments were assembled in a CPEC reaction containing 1 μ L Q5 high fidelity DNA polymerase (New England Biolabs), 10 μ L 5x Q5 high fidelity buffer (New England Biolabs) 10 mM dNTP mixture, 3% (v/v) DMSO and an equimolar amount of both fragments (100-250 ng). The volume was adjusted to 50 μ L with sterile milliQ water. The total mixture was

subjected to an initial denaturation of 30 seconds at 98 °C, followed by 15 cycles of denaturation (10 sec, 98 °C), annealing (30 sec, 55 °C) and elongation (75 sec, 72 °C). A final elongation of 10 minutes at 72 °C completed the program. Without purification, 2 µL of the mixture was mixed with 20 µL electrocompetent BL21 (DE3) cells and subjected to electroporation. If the sequence was checked to be correct (Macrogen Europe), *Pichia pastoris* was transformed with the corresponding plasmid.

CBM?	Template	Primer sequence (5' → 3')
pPpT4-TrCel61A_AA9_PMO3_95		
YES	pPpT4-alpha-TrCel61A-NSS	FWD: GGTCCAGCTGTTTACTCTGGTTTGCCAACTTCTGTTGCTCAAGG REV: CGATGTTGGAAACGTGACCATGTCCCACAACACCAGTAGC
	pMA-T-TrCel61A-AA9-PMO3	FWD: GCTACTGGTGTGTTGGGACATGGTCACGTTTCCAACATCG REV: CCTTGAGCAACAGAAGTTGGCAAACCAGAGTAAACAGCTGGACC
NO	pPpT4-alpha-TrCel61A-NSS	FWD: CCAGGTCCAGCTGTTTACTCTGGTCATCACCATCACCATCACTAGG REV: CGATGTTGGAAACGTGACCATGTCCCACAACACCAGTAGC
	pMA-T-TrCel61A-AA9-PMO3	FWD: GCTACTGGTGTGTTGGGACATGGTCACGTTTCCAACATCG REV: CCTAGTGATGGTGATGGTGATGACCAGAGTAAACAGCTGGACCTGG
pPpT4-TrCel61A_AA9_PMO3_95_short		
YES	pPpT4-alpha-TrCel61A-NSS	FWD: GGTCCAGCTTTGTACTCTGGTTTGCCAACTTCTGTTGCTCAAGG REV: CGATGTTGGAAACGTGACCATGTCCCACAACACCAGTAGC
	pMA-T-TrCel61A-AA9-PMO3_short	FWD: GCTACTGGTGTGTTGGGACATGGTCACGTTTCCAACATCG REV: CCTTGAGCAACAGAAGTTGGCAAACCAGAGTACAAAGCTGGACC
NO	pPpT4-alpha-TrCel61A-NSS	FWD: CCAGGTCCAGCTTTGTACTCTGGTCATCACCATCACCATCACTAGG REV: CGATGTTGGAAACGTGACCATGTCCCACAACACCAGTAGC
	pMA-T-TrCel61A-AA9-PMO3_short	FWD: GCTACTGGTGTGTTGGGACATGGTCACGTTTCCAACATCG REV: GCCTAGTGATGGTGATGGTGATGACCAGAGTACAAAGCTGGACCTGG

Table 5.6: Primers for cloning full consensus genes in pPpT4-vector. Bold face: synthetic gene sequence, underlined: his₆-tag FWD: forward primer, REV: reverse primer.

Molecular cloning for CBM removal

The constructs were made using circular polymerase extension cloning (CPEC³¹⁰), starting from the pPpT4-vector containing the wild-type full length enzyme from chapter 2. A schematic overview for the construction of a variant lacking linker and CBM is shown in figure 5.17. All other variants were made in analogy.

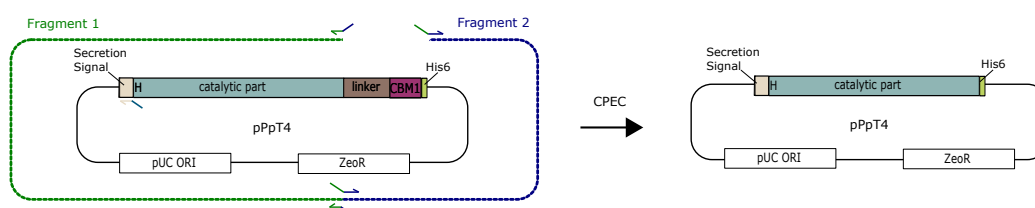


Figure 5.17: Schematic overview of molecular strategy for CBM removal. This is an example for removal of glycosylated linker and CBM, while keeping the 6 x His-tag. Other truncations were made in analogy.

First 2 different fragments were amplified via PCR, whereas the first fragment started at the His₆-tag or terminator region in case of no tag and ended in the zeocine resistance gene. The second fragment completed the zeocine resistance gene and ended in the se-

quence at the position that corresponds to the carboxy terminus of the desired protein. Fragment assembly and CPEC were performed as explained above. The forward and reverse primer in the zeocine resistance region were identical in all cases (fwd: 5'-CTGAG TGGTCTGAGGTCGTGTCTAC-3', rev: 5'-GTAGACACGACCTCAGACCACTCAG-3'). All other required primers can be found in table 5.7.

For construct pPpT4-*TrCel61A*-CD-MSA-3SS, an additional mutation at position 168 was made (N168C) to complete the disulfide bridge with C248. Therefore, the vector with truncated gene was first created followed by site-directed mutagenesis using the method of Sanchis et al.²⁶⁸ and explained in the following alinea.

Fragment	Primer sequence (5' → 3')
pPpT4- <i>TrCel61A</i> -CD-prediction	
1	FWD: ATTGCCAACTTCTGTTGCTCAAGGTCATCACCATCACCATCACTAGGCGGCCGCTCAAG
2	REV: CCGCCTAGTGATGGTGATGGTGATG ACCTTGAGCAACAGAAGTTGGCAATCCGGAAC
pPpT4- <i>TrCel61A</i> -CD-MSA-His	
1	FWD: CCAGGTCCAACCTGTTGTTCCGGACATCACCATCACCATCACTAGG
2	REV: CTAGTGATGGTGATGGTGATG TCCGGAACAACAGTTGGACCTGG
pPpT4- <i>TrCel61A</i> -CD-MSA-NoHis	
1	FWD: CTGTTGTTTCCGGATAGGCGGCCGCTCAAG
2	REV: CTTGAGCGGCCGCTAT CCGGAACAACAG
pPpT4- <i>TrCel61A</i> -CD-MSA-V246W	
1	FWD: CCAGGTCCAACCTGTTGGTCCGGACATCACCATCACCATCACTAGG
2	REV: CTAGTGATGGTGATGGTGATG TCCGGACCAACAGTTGGACCTGG
pPpT4- <i>TrCel61A</i> -CD-MSA-3SS	
1	FWD: CCAGGTCCAGCCGTTTTCACGTGTCATCACCATCACCATCACTAGG
2	REV: CTAGTGATGGTGATGGTGATG ACACGTGAAAACGGCTGGACCTGGGATGATGTAGTTC
Sanchis	FWD: CCCAGACAACTTGGCTCCAGGTGTACGTTTTGAGACACGAG
	REV: GTAGACACGACCTCAGACCACTCAG
pPpT4- <i>TrCel61A</i> -CD-Linker-His	
1	FWD: CCTGCTGGTGGTCCTCATCACCATCACCATCACTAGGCG
2	REV: CGCCTAGTGATGGTGATGGTGATG AGGACCACCAGCAGG
pPpT4- <i>TrCel61A</i> -CD-Linker-NoHis	
1	FWD: CTGCTGGTGGTCCTTAGGCGGCCGCTCAAG
2	REV: CTTGAGCGGCCGCTA AGGACCACCAGCAG
pPpT4- <i>TrCel61A</i> -CD-G3S-His	
1	FWD: CTGTTGTTTCCGGAGGTGGAGGATCTCATCACCATCACCATCACTAG
2	REV: CTAGTGATGGTGATGGTGATGAGATCCTCCACCT CCGGAACAACAG

Table 5.7: Primers for CBM removal of *TrCel61A*. Bold face: *trcel61a* gene sequence, underlined: his₆-tag FWD: forward primer, REV: reverse primer.

Site-directed mutagenesis of *TrCel61A*

Variants of *TrCel61A* were created in the pPp-T4-vector with native secretion signal and C-terminal His-tag and controlled by the AOX1-promotor, as described in chapter 2. The mutations were created by use of the method of Sanchis et al.²⁶⁸, using the forward primers from table 5.8 and reverse primer 5'-CGAATGCGAAGGTTAGTAGG-3', located about 500

bp downstream the stopcodon of our gene sequence. A detailed description of the method can be found in Materials and Methods of chapter 3 on page 83.

Mutation	Primer sequence (5' → 3')
Disulfide removal	
C77A	CCAGACATCATC GCT CACAAGAACGCTAC
C122A	TGTACGGTGAC GCT GAGACTGTTGAC
N-glycosylation introduction	
V33S	CGTTATCAACGGTT CCT GGTATCAGGCTTACG
H87N	CGCTAAGGGT AAT GCTTCTGTTAAGGCTGG
D126N	CGGTGACTGTGAGACTGTT AAT AAGACTACTTTGG
A170S	CCCAGACAACCTT GCT CCAGGTAACACTACG
S206N	GCTGTTTCCGGT AAC GGTTCCTTGC
L215N	CCATCTGGTGTT AAC GGTACTGACTTG
H221N	GGTACTGACTTGTA AC GCTACTGACCC
Disulfide introduction	
F43C	CGACCCAACTACT TGT CCATACGAATCCAAC
P49C	CCATACGAATCCAAC TGT CCAATCGTTGTTGG
E46C + P49C	CTACTTTCCATAC TGT TCCAAC TGCC CAATCGTTGTTGG
I51C	ATCCAACCCACCA TGC GTTGTTGGTTGG
A58C	GGTTGGACTGCT TGT GATTTGGACAACGG
K79C	CATCATCTGTCACT TGC AACGCTACAAACGC
K85C	GAACGCTACAAACGCT TGC GGTCACGCTTCTGTTAAGG
I111C	CCACATCCAGGTCCT TGT GTTGATTACTTGGC
L140C	TCGACGGTGTTGGTTGTT TGT CTGGTGGTGATCC
A150C	GATCCAGGTACTTGGT TGT CCGACGTTTGTATCTCC
E179C	CGTTTTGAGACACT TGT ATTATCGCTTTGCACTCCGC
P196C	CGGTGCTCAAACTACT TGCC AGTGTTCACATTGC
Y238C	CTTCCCCATTGAAC TGT ATCATCCCAGGTC
Consensus: flexible loops	
E46M	CTTTCCCATAC ATGT CCAACCCACC
I51E	CAACCCACCA GAG GTTGTTGGTTG
I51T	CAACCCACCA ACT GTTGTTGGTTG

Table 5.8: Mutagenic primers for stabilizing *TrCel61A*. Bold face: synthetic gene sequence, underlined: his₆-tag.

One variant of the flexible loop exchange (consensus design 233-QSLSSY-238) could not be obtained by the method of Sanchis et al. since the regions to change are too extended (> 2-3 subsequent amino acids). Therefore a 2-piece CPEC assembly was performed, starting from the original pPpT4-plasmid, with WT *Trcel61a* sequence. The procedure was identical to the procedure described above, with overlapping regions in the flexible loops to mutate and the zeocin resistance gene. The first piece started at the region to mutate, ending in the middle of the zeocin resistance gene (fwd: 5'-GGTGTTTTGATCAACATCTAC**CAATCCTTGTCTTCT**ACATCATCCCAGGTCCAAC-3', rev: 5'-GTAGACACGACCTCAGACCACTCAG-3'), while the second piece completed the zeocin resistance and ended in the to mutate region (fwd: 5'-CTGAGTGGTCTGAGGTCGTGTCTAC-3', rev: 5'-GTTGGACCTGGGATGATGTAG**GAAGACAAGGATTGG**TAGATGTTGATCAAAACAC-3').

5.4.2 Enzyme thermostability

The apparent melting temperature of all enzyme variants was determined using differential scanning fluorimetry (DSF) in a CFX Connect96 cycler (Biorad), as described in chapter 3. Samples were diluted to equal enzyme concentrations (minimal 0.5 mg/mL) and all variants were measured in triplicate. Each well consisted of 15 μ L purified enzyme and 10 μ L 1/400 diluted Sypro Orange.

When different unfolding actions appeared, the height of the unfolding peak (amplitude) was calculated to determine the relative amount in the respective unfolding event.

5.4.3 Enzyme activity

To determine the activity of the enzyme variants on PASC, not an endpoint measurement, but a kinetic follow-up of cellobionic acid production was applied, while taking regular samples during 8 hours and 1 sample after 20 hours. The experimental data were fitted in a linear curve wherein the slope was estimated via the least squares principle with accompanying standard error on this coefficient.

5.4.4 Molecular modeling

Because a crystal structure was lacking, a homology model for *Trichoderma reesei* Cel61A was generated with the YASARA software (YASARA bioscience, Vienna)^{422;423}. The software searches for homologues and builds a model. Five template structures were first selected by Yasara and 5 different initial models were built using these structures. The crystal structure from the closely related LPMO from *Thermoascus aurantiacus* isozyme A (56% identity, 68% similarity) was found to be the best template (PDB entry 3ZUD-chain A) to build the final model and could not be improved by combination of different models. Yasara considered the model as satisfactory with a z-score of -3.369. Only the catalytic part was modeled since the homologue did not harbor a CBM, nor a linker region.

5.4.5 Sequence analysis

The B-FITTER tool was used to determine the most flexible residues within the protein. This software, freely available from the developers, ranks all residues from high to low B-factor by calculating the mean B-factor per amino acid using all atoms except hydrogen²¹².

TrCel61A is part of the auxiliary activity family 9 on CAZy (Cazy.org)¹⁶. The complete set of proteins classified as AA9 family was downloaded from CAZY via the Research tools developed by Alexander Holm Viborg (research.AHV.dk). A classical multiple sequence alignments (MSA) was performed on amino acid level using the MUSCLE algorithm⁴²⁴ with standard gap penalties in MEGA6.0⁴⁰⁷. The sequences were trimmed to the catalytic

domain only, meaning to remove anything before histidine-22 (=N-terminus because of preceding secretion signal) and removing the possible present linker and CBM. Furthermore a reduction in the set of sequences was done by removing all sequences that resembled more than 95 % in a pairwise comparison by CD-HIT^{425;426}. Next, a phylogenetic tree was built (MEGA6.0 software) by using the maximum likelihood method (JTT Matrix-based substitution model⁴²⁷) and 1000 bootstraps were performed to test the phylogeny. An unrooted tree was built after clustering the data by the nearest neighbor-joining algorithm. All positions containing gaps were eliminated, whereas the tree was built on a total of 66 positions in the final dataset of 263 sequences.

The consensus residues were determined by taking the highest prevalent residue at a certain position after MSA, unless a GAP was most prevalent. Correlation networks were studied using the correlation mutation analysis tool (Comulator) from 3DM^{428;429}.

5.4.6 Statistical analysis

Significant differences between one specific enzyme variant and the wild-type enzyme were determined with the Student t-test, while ANOVA was used to determine if there are significant differences when comparing all variants and wild-type enzyme at once. If probabilities (p-values) lower than 0.05 were obtained, the result is considered statistically significant.

Supplementary material

Rank	Residue <i>Ta</i> GH61A	B-value (Å ²)	Residue <i>Tr</i> CEL61A
1	GLY-228	31.59	GLY-227
2	SER-216	29.34	LEU-215
3	THR-227	28.86	SER-226
4	ASN-27	28.37	ASN-27
5	PRO-28	26.52	PRO-28
6	SER-217	26.51	ASN-216
7	LYS-214	25.94	SER-213
8	PRO-29	25.22	PRO-29
9	LEU-215	25.15	PRO-214
10	SER-26	24.49	SER-26
11	MET-25	24.48	GLU-25
12	GLN-213	24.42	THR-212
13	ASN-190	24.29	GLN-189
14	GLU-30	23.90	ILE-30
15	GLN-169	23.49	ALA-168
16	ARG-58	23.42	LYS-58
17	LYS-61	23.15	THR-61
18	ASP-10	23.13	ASN-10
19	ASN-98	23.09	ASN-98
20	PRO-71	23.07	ALA-71

Table S5.1: Top 20 most flexible residues in *Ta*GH61A and transfer to *Tr*CEL61A.

Residue start	motif	remark
1. adapted motif X - X (No PRO) - SER or THR		
40	PTT	close to cellulose binding surface, P not directly at surface
45	YES	in cellulose binding surface
54	GWT	not at protein's surface
64	FVS	F64 at < 6 Å from Cu ion
80	NAT	native glycosylation
87	HAS	possibility
93	GDT	G93 conserved residue
122	CET	disulfide bridge involved
126	DKT	Possibility
140	LLS	L140 not completely surface exposed
146	PGT	P146 in cellulose binding surface
149	WAS	not at protein's surface
154	LIS	oriented inwards on alpha-helix and close to other glycosylation position
158	NNT	native glycosylation
183	LHS	H184 at < 6 Å from Cu ion
202	AVS	A202 not completely surface exposed
206	SGS	Possibility
215	LGT	Possibility
221	HAT	Possibility
231	IYT	Y232 semi-conserved
245	VVS	semi-conserved regio
2. adapted motif ASN - X (No PRO) - X		
26	NDI	N26 not completely surface exposed
31	NGV	Possibility
62	NGF	all at distance < 6 Å from Cu ion
80	NAT	native glycosylation
83	NAK	close to native glycosylation
117	NCN	disulfide bridge involved
157	NNN	next to N158 = native glycosylation position
158	NNT	native glycosylation
168	NLA	Possibility
173	NYV	V175 semi-conserved
190	NGA	A192 semi-conserved
194	NYP	Y195 involved in copper-binding
200	NIA	N200 not surface exposed
230	NIY	Y232 semi conserved
237	NYI	Y238 conserved
3. adapted motif ASN - PRO - SER or THR		
NONE		

Table S5.2: Overview of all possible N-glycosylation sequons by single mutation effort. Three different adaptations of the conserved motif are possible and each yield different options. The putative positions that were evaluated in the lab are colored gray.

Multiple sequence alignment

ESPRIT figuur - combi met CBM verwijdering mogelijk?, ZEKER 3ZUD moet erin staan

Codon optimized gene sequences• *TrCel61A-AA9-LPMO3-95*

CATGGTCACGTTTCCAACATCGTTATCGACGGTGTTTACTACCAGGGTTACGACCCATACTCTGACC
 CTTACATGTCTAACCCACCAGAGGTTATTGGTTGGTCCACTCCAGCTACTGACAACGGTTTTGTTTCT
 CCAGACGCTTACTCCTCCCCAGACATTATCTGTCAAGAAGCTACTCCAGCTGCTTTGCATGCTC
 CAGTTGCTGCTGGTGATACTGTTACATTGCAGTGGACTACTTGGCCAGAATCTCATCACGGTCCAGT
 TATCACTTACTTGGCTAAGTGTCCAGGTGACGGTTGTTCCACTGTTGACAAGACTACTTTGGAGTGG
 TTCAAGATCGACGAGGCTGGTTTGATTGACGGAACAACTTCCAGGTACTTGGGCTACTGACGTTT
 TGATCGCTAACAACTCCTGGACTGTTACTATCCCATCCTCTTTGGCTCCAGGTAACACTACGTTTTG
 AGACACGAGATTATCGCTTTGCACTCCGCTGGTCAAGAAAACGGTGCTCAAACTACCCACAGTGTT
 TCAACTTGAAGGTTACTGGTTCCGGTACTGCTTTGCCATCTGGTGTTTTGGGTACTGGTTTGTACAA
 GGCTACTGATCCAGGTATCTTGTTCAACATCTACCAGTCCTTGACTTCCTACACTATCCAGGTCCAG
 CTGTTTACTCTGGT

• *TrCel61A-AA9-LPMO3-95-short*

CATGGTCACGTTTCCAACATCGTTGTTAACGGTGTTTACTACCAGGGTTACGACCCAACTTCCTACC
 CTTACATGTCTAACCCACCAACTGTTGTTGGTTGGACTACTACTGACACTGACAACGGTTTCGTTTCC
 CCATCTGCTTACGGTTCCCCAGACATTATCTGTCAAGAAGCTACAAACGCTAAGGGTCACGCTG
 AAGTTGCTGCTGGTGATAAGGTTGAGTTGCAGTGGACTACTTGGCCAGAATCTCATCACGGTCCAGT
 TATCGATTACTTGGCTAACTGTAACGGTGACTGTGAGACTGTTGACAAGACTACTTTGGAGTTCTTCA
 AGATCGACGAGGTTGGTTTGATCGACGGTTCTAACGTTCCAGGTACTTGGGCTTCCGACCAGTTGAT
 TGCTAACAACTCCTGGACTGTTACTATCCCATCCTCCATTGCTCCAGGTAACACTACGTTTTGAGAC
 ACGAGATTATCGCTTTGCACTCCGCTGGTCAAGAAAACGGTGCTCAAACTACCCACAGTGTTTCAA
 CTTGAAGGTTACTGGTTCCGGTCCGACAAGCCATCTGGTGTTTTGGGTACTGAGTTGTACAAGGCT
 ACTGACCCAGGTATCTTGTTAACATCTACCAGTCCTTGCTCCTACACTATTCCAGGTCCAGCTTT
 GTACTCTGGT

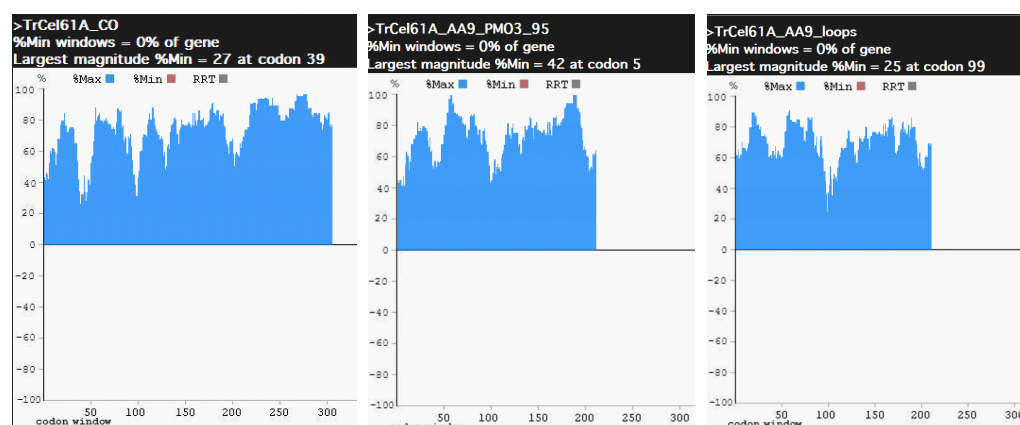


Figure S5.1: Codon clustering analysis for recombinant expression of the WT *TrCel61A* and the full consensus genes in the host *Pichia pastoris* as calculated from the rare codon calculator²⁷⁷. The codon usage table was used from (<http://www.kazusa.or.jp/codon/>).

SDS-PAGE analysis

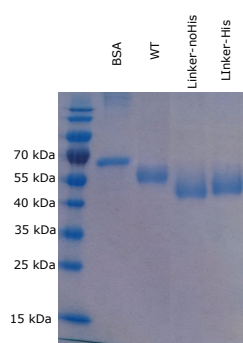


Figure S5.2: SDS-PAGE of truncated variants. BSA: 0.1 g/L as reference for quantification with ImageJ, Linker-noHis and Linker-His: truncated variants of *TrCel61A* lacking CBM.

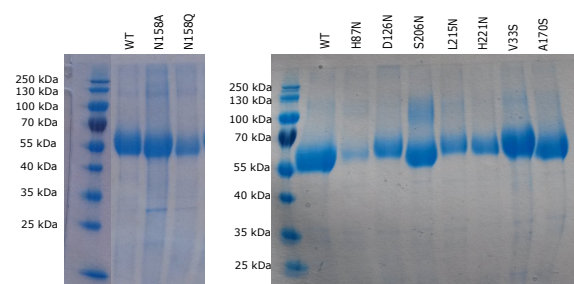


Figure S5.3: SDS-PAGE analysis of all site-directed mutants for removal of N158 glycan (LEFT) or introduction of additional N-glycosylation positions (RIGHT) in *TrCel61A*. All enzymes were purified via ultrafiltration.

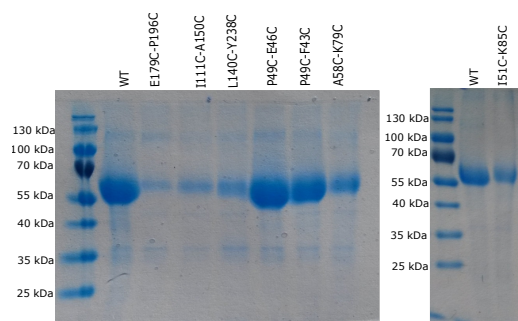


Figure S5.4: SDS-PAGE analysis of all site-directed mutants for additional disulfide bridges in *Tr*Cel61A. All enzymes were purified via ultrafiltration.

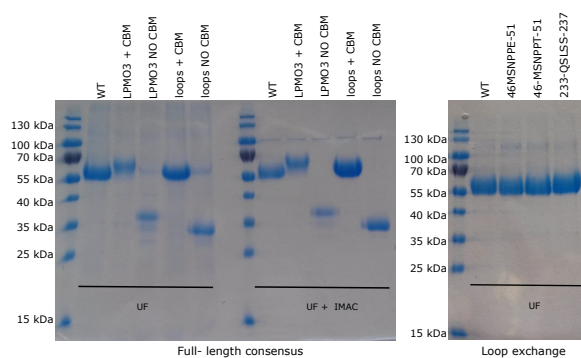


Figure S5.5: SDS-PAGE analysis of consensus enzymes and flexible loop exchange variants for *Tr*Cel61A. The full-length consensus enzymes were purified with ultrafiltration (UF) and subsequent immobilized metal ion affinity chromatography (IMAC) against the C-terminal his-tag.

Melting curves

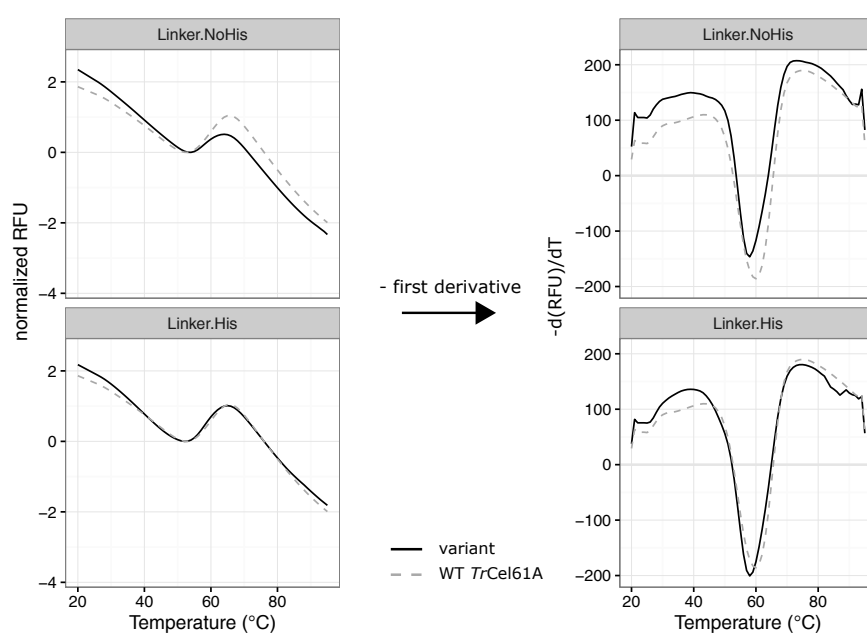


Figure S5.6: Meltcurve and negative derivative of truncated variant of *TrCel61A*.

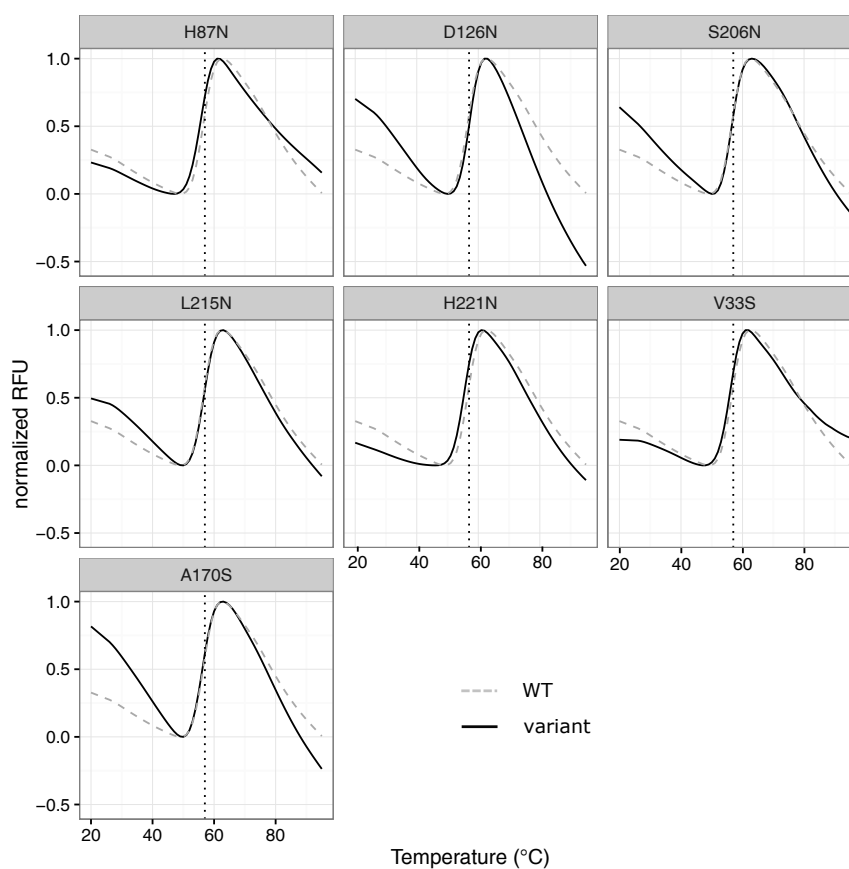


Figure S5.7: Overview of DSF curves for all variants introducing additional putative N-glycosylation positions. The inflection point of the curve corresponds to the apparent melting temperature. This is 57 °C for the wild-type *Tr*Cel61A, as indicated with a black dotted line. Experiment was performed in triplicate, but only one curve is shown here for clarity. The relative fluorescent units (RFU) were normalized to be better comparable.

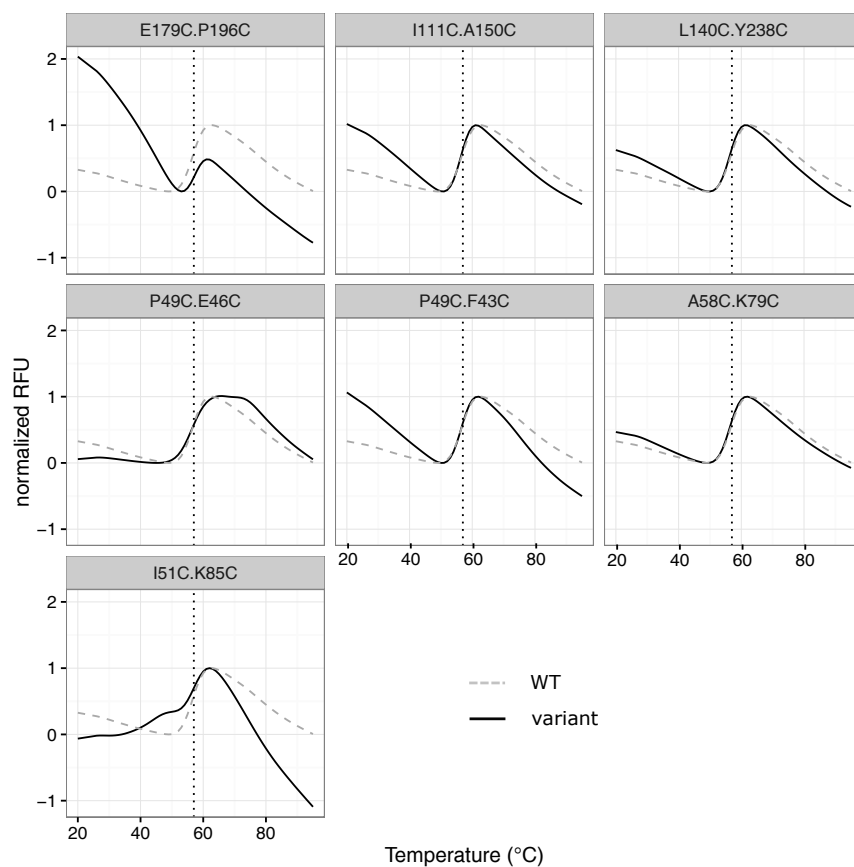


Figure S5.8: Overview of DSF curves for all variants introducing additional putative disulfide bridges.

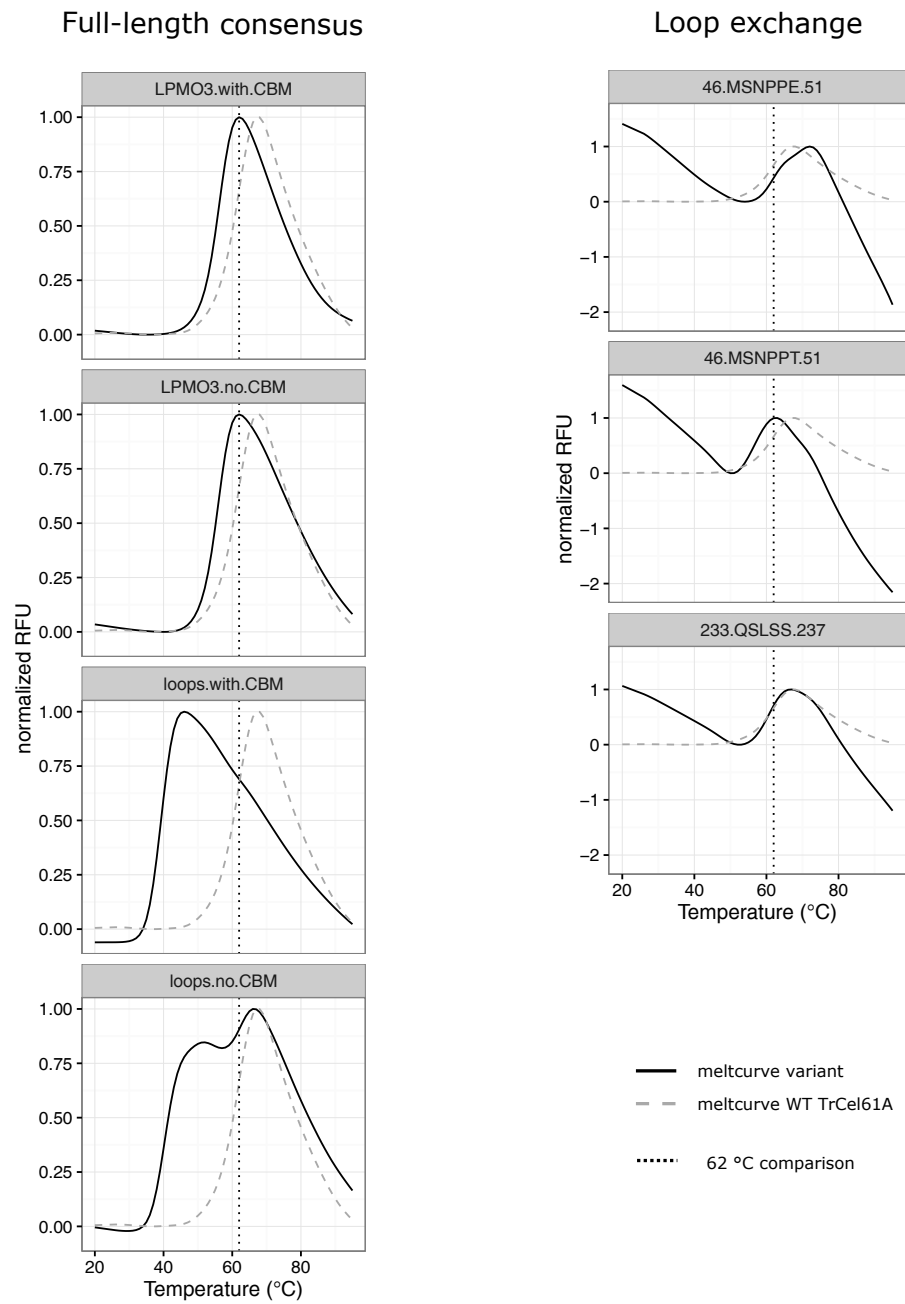


Figure S5.9: Overview of DSF curves for the consensus-based variants. The inflection point of the curve corresponds to the apparent melting temperature.

6

Thermostability engineering of *Streptomyces coelicolor* LPMO10C

Part of this chapter was published as:

Tanghe M., Danneels B., Last M., Beerens K., Stals I., Desmet T. (2017) **Disulfide bridges as essential elements for the thermostability of lytic polysaccharide monooxygenase LPMO10C from *Streptomyces coelicolor*.** *Protein Engineering, Design and selection*. doi:10.1093/protein/gzx014. pp. 1-8

Abstract

Lytic polysaccharide monooxygenases are known to boost the cellulose degradation process. Because very little is known about the stability of these interesting enzymes, this chapter will focus on enhancing the stability of a bacterial LPMO, namely *Streptomyces coelicolor* LPMO10C, formerly known as ScCelS2.

In order to stabilize *Streptomyces coelicolor* LPMO10C, some rational mutants were designed based on improving interactions in the protein's surface such as creating ion pair networks and increasing hydrophilic contact with the solvent. Nonetheless, these mutations could not improve the apparent melting temperature.

As second strategy, extra disulfide bridges were introduced in the wild-type ScLPMO10C. The reason behind this was not only to increase the thermodynamic stability, but also to assess the role of disulfide bridges in this LPMO. Although *Streptomyces coelicolor* LPMO10C has an apparent melting temperature of only 51 °C, the enzyme was found to retain 34 % of its activity after 2 hour incubation at 80 °C. However, this was no longer true when its disulfide bridges were chemically broken prior to the incubation. To further assess this interesting finding, additional disulfide bridges were then introduced. Sixteen putative new linkages were designed and subsequently evaluated based on apparent melting temperature (T_m). Four of the resulting variants showed an improvement in T_m , ranging from 2 to 9 °C. Combining the positive disulfide introductions yielded additional improvements (up to 19 °C) but aberrant unfolding patterns became apparent in some cases, resulting in a diminished capacity for heat resistance. Nonetheless, the best variant, a combination of A143C-P183C and S73C-A115C, displayed a 12 °C increase in T_m and was able to retain no less than 60 % of its activity after heat treatment (compared to only 34 % for the wild-type).

Disulfide engineering is thus a valuable strategy for LPMO stability improvement. Extra disulfide bridges not only contribute to a higher stability but can also increase resistance to heat treatment.

6.1 Introduction

As stated already in the introduction of the former chapter, LPMOs are considered a key factor in cellulose breakdown because of their boosting effect on the entire process^{14;15;147}. In this context, LPMOs are an indispensable addition to the current industrial cellulose degradation cocktails^{147;319}. Even though stability is a main concern in industrial processes^{23;231}, the current literature lacks information about this important parameter for LPMOs. A broadened substrate specificity for different LPMO families has been unveiled including the closely related chitin and even the structurally different starch⁴³⁰. The actual biomass (lignocellulose) targeting families seem to be limited to only auxiliary activity families 9 and 10¹⁹. The previous chapter dealt with the stability improvement of an AA9 representative. Therefore, this chapter will focus on an AA10 representative, namely *Streptomyces coelicolor* LPMO10C, for which expression and measurement methods have been optimized in chapter 4 (starting from page 87).

Despite the exploration of various engineering strategies in chapter 5, the outcome was limited. For this reason, a different approach was taken here. Also, the availability of a crystal structure of ScLPMO10C, enlarges the possibilities. To that end, ScLPMO10C was firstly explored in terms of sequence conservation, flexibility and heat resistance to obtain some insight in the structure - function relationships. Because the role of surface residues in stability is highly underestimated⁴³¹, these were examined in the first stability engineering strategy. The 3D structure of the protein was examined and residue substitutions were suggested that might form interactions such as ion pairs^{432;433} or increase hydrophilic contact with the solvent. Stability through ion pair networks on the surface has been widely used as a strategy for stability engineering^{434–436} because of the high prevalence of extended ion pair networks on the surface of (hyper)thermostable enzymes^{433;437;438}.

In a last approach disulfide engineering was applied even though it was not successful for *TrCel61A* thermostability engineering. The reason behind this is not only to (possibly) increase the stability, but also to explore the interesting capabilities of these linkages. Disulfide bridges seem very important, as shown for *TrCel61A* in chapter 5 (page 114) and its high heat tolerance.

In this chapter, after exploration of the WT enzyme, the 2 strategies for stability engineering are explained. These start with an explanation of the target residues for mutagenesis and proceed with determination of the apparent melting temperatures as first screening method. Only positive hits were retrieved for more tests considering kinetic activity. The activity evaluation was solely based on HPAEC-PAD analysis since the H₂O₂-method appeared not suitable for his-tag purified enzymes (see chapter 4).

6.2 Results and discussion

6.2.1 General exploration of stability of the WT ScLPMO10C

Some general observations about stability of the wild-type *Streptomyces coelicolor* LPMO10C enzyme will be reported here. This includes the importance of the amino-terminus on the thermal stability, an overview of flexible regions as target residues for stability engineering and a remarkable residual activity after heat treatment.

Overall structure

Despite AA9 and AA10 being both LPMO families sharing the same overall structure, they only show a low sequence identity and similarity apart from the active site residues¹¹⁰. Moreover the largest part of the conserved regions are located at the inside of the protein^{19;115}. As illustration, all AA10 members with known crystal structure are listed in table S6.1. Only 30% of the AA10 members harbor a cellulose-binding domain, that can serve for cellulose binding (CBM2 or CBM3) or chitin binding (CBM5 or CBM12)¹¹⁰. ScLPMO10C harbors a catalytic domain (CD), formerly classified as CBM33 and known as a chitin-binding domain. This part is connected through a linker to a CBM2 domain. The protein harbors 3 disulfide bonds, of which two in the CD (C48-C66 and C103-C222, figure 6.3) and 1 in the CBM (C265-C361).

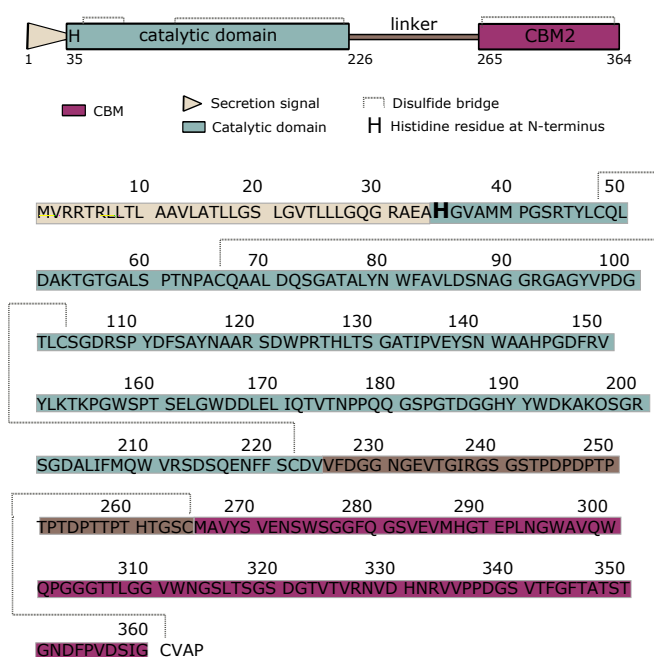


Figure 6.1: Schematic build-up and translated coding sequence of ScLPMO10C.

Importance of the N-terminus

A characteristic trait of LPMOs is the need for an N-terminal histidine residue. We have shown earlier in this thesis that a His-1 residue is inevitable to obtain an active LPMO (see also chapter 4). Additionally, the effect of a free His-1 on the thermodynamic stability was also assessed here. A variant of ScLPMO10C was created in that chapter, called LPMO10C-Xa, that has an additional N-terminal extension that can be cleaved off with protease factor Xa to yield an active protein with His-1 (see also page 91 and following). The melting temperature of this LPMO10C-Xa was determined before and after the cleavage with Xa-factor. The obtained melting temperature of 44 ± 1 °C for the construct with extra addition at its N-terminus was significantly lower ($p=0.0002$, Student t-test) than the 51 ± 1 °C obtained for a correct N-terminus (Figure 6.2). This result shows that the impaired copper coordination center not only results in a loss of activity, but also introduces a new unfolding region, *i.e.* the most labile region in a protein where the denaturation will initiate under stress conditions such as heat increase in this case³³⁰. In accordance with these results, a very recent report of Chaplin et al. concludes that AA10 member *Streptomyces lividans* LPMO10E with N-terminal methionine-glycine extension can still bind copper but has a weaker affinity to bind this copper than the His-1 variant⁴³⁹.

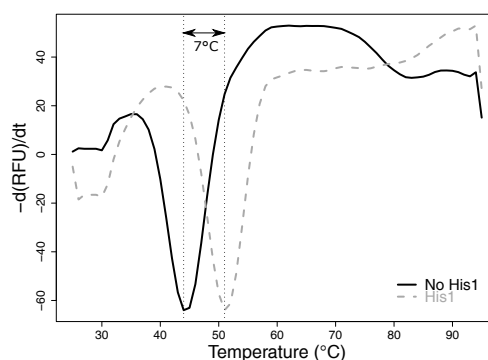


Figure 6.2: Influence of His-1 on thermostability. The apparent melting temperature of ScLPMO10C with a correct N-terminus is 7 °C higher than with an additional N-terminal part attached. The graph shows the negative derivative of the melting temperature so that the minimum value represents the apparent melting temperature (T_m).

The copper coordination center is indeed really prone to instability as was already shown by Hemsworth et al. They described for *Bacillus amyloquelaciens* CBM33 that the thermal unfolding temperature decreased with 19 °C when all copper is removed. The same decrease was reported when Manganese (II) was added instead, while only a 11 °C decrease was shown for Zinc (II) or Nickel (II)^{143;144}. Thus, having an N-terminal extension shows significant decrease in melting temperature, but seems to have less impact on the overall fold than having no or an improper metal ion.

Mapping flexible regions in ScLPMO10C via B-factor analysis

Several (semi-) rational engineering strategies for stabilization include rigidifying flexible regions²⁴² because these are thought to be initiators for denaturation. Hence, fixing these targets can highly increase the stability of your protein³³⁰. Therefore, possible target regions were mapped by identification of the flexible regions using B-factor iterative test (B-FITTER program available from the developers using PDB 4OY7-chain A)²¹². This software assigns every residue with a value for local flexibility that is the average B-factor of all atoms (except hydrogen) in that specific residue and ranks them from high to low value. Herein, the B-factor corresponds to the size of the atomic electron density smearing around the average position of the atom and thus correlates with local flexibility²³¹.

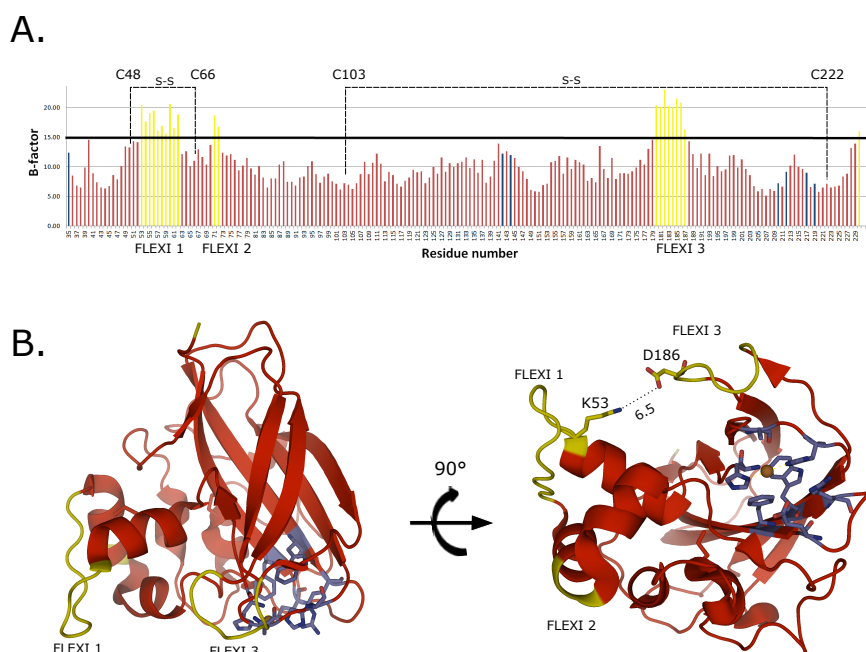


Figure 6.3: Flexible regions in ScLPMO10C according to B-factor analysis. (A) The yellow bars are the top 20 flexible residues (here, B-factor higher than 15), red bars are fairly stable and blue bars represent the residues within 6 Å from the copper ion (H35, A142, H144, W210, R212, E217, F219). Disulfide bridges are also indicated with a dotted line. (B) The flexible regions indicated in a cartoon representation where the left side is a front view with cellulose binding surface at the bottom and the right side is a view from the binding surface onwards. Three flexible regions are indicated on the graph and cartoon with FLEXI 1 to 3. The shortest distance between flexible loop 1 and 3 is also indicated with the distance in Angstrom. Figures were prepared with PyMol⁷³.

The result for ScLPMO10C is shown in figure 6.3, where the top 20 (or 10 %) most flexible residues are colored yellow, while the presumed active site residues are presented in blue. The graph shows that the most flexible positions are mostly focused in 3 clusters, that are all loop regions. There is a higher clustering as was the case for *TrCel61A*, as shown in figure 5.2 on page 112. The first flexible group of ScLPMO10C (FLEXI 1,

residues Lys-53–Thr-62) corresponds to the region that Sprenger and coworkers obtained by molecular dynamics simulations to be the most flexible²⁴⁸. This group is also part of the so-called L2-loop (Pro-41–Arg-125) that is highly variable in LPMOs and that is responsible in ScLPMO10C for at least 50% of the cellulose binding capacity¹¹⁴. The second (short) group (FLEXI 2, Asp-71–Gln-72) is still located in the L2-loop but a little distant from the actual cellulose binding surface. The last group (FLEXI 3, Gln-180–Gly-187) corresponds to the short loop (LS-loop) in family AA9, also known for substrate binding but was not further referred to in AA10. Interestingly, the first group and third group are both positioned at the cellulose binding surface. This might suggest a required flexibility in binding the substrate. Finally, G230 is also highly flexible, but this can be expected from a terminal residue.

Improved heat resistance due to native disulfide bridges

The thermal stability of the wild-type enzyme was first examined by Differential Scanning Fluorimetry (DSF)²⁸⁵, which yielded an apparent melting temperature (T_m) of 51 °C, a value that is quite typical for a mesophilic enzyme⁴⁴⁰ (chapter 4).

In contrast to the enzyme's rather normal thermodynamic stability, a surprisingly high kinetic stability was observed at elevated temperatures. Indeed, after two hours of incubation at 80 °C, a residual activity of 34 % could still be detected on phosphoric acid swollen cellulose (PASC) (figure 6.4). However, this was only true when the native disulfide bridges were kept intact, since prior treatment with tris(2-carboxyethyl)phosphine (TCEP) to break these bonds^{441;442} resulted in complete loss of activity under similar incubation conditions. The chemical treatment did by itself also lower the enzyme's activity (to 82%) but not enough to explain the dramatic decrease in stability. To further assess the role of disulfide bridges in the enzyme's heat resistance, the effect of introducing additional disulfide bridges was investigated next.

This is a surprising and very intriguing result since 80 °C far exceeds the melting temperature of 51 °C so one would expect a denatured enzyme, yet not active. Indeed, when measuring activity at higher temperatures (Figure 6.4 B), there is no activity at 80 °C. Similar reports of increased resistance for thermal denaturation have been described^{443;444}. They assigned the effect to a lower energy requirement for reclaiming the original conformation after (partial) heat denaturation. The reason behind this would be that disulfide bridges would lower the entropy in the denatured state.

With respect to this interesting result, in one of the thermostability engineering strategies further in this chapter (section 6.2.3 on page and 163 further), the effect of more disulfide bridges is explored. The hypothesis to investigate is whether more disulfide bridges can even further increase this heat resistance.

In addition, figure 6.4 B also reveals a higher activity at 50 °C, thus a probably more

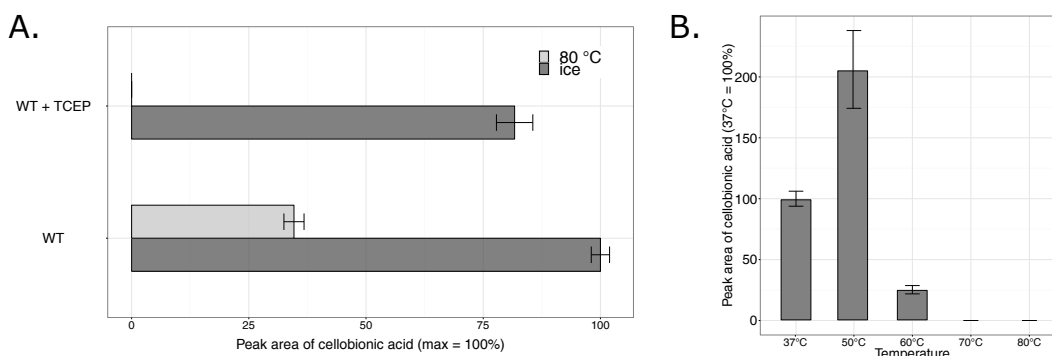


Figure 6.4: (A) Effect of a 2-hour 80 °C treatment on the WT ScLPMO10C. The dark gray bars represent the activity when the enzyme is incubated on ice for 4 hours, while the light gray bars show the activity when the enzyme was pre-incubated for 2 hours at 80 °C and subsequently 2 hours on ice. A concentration of 10 mM TCEP was added to break the disulfide bridges chemically. The activity is calculated as the relative peak area on HPEAC-PAD analysis of the cellobionic acid. (B) Relative activity at different temperatures (without TCEP), while the activity at 37 °C is taken as 100.

suitable temperature than the 37 °C applied in this work. The optimum temperature is presumably a little under 50 °C. A possible stabilizing effect of the substrate³⁵¹ cannot be excluded since there is a high activity at the protein's melting temperature (measured without substrate) and even at 60 °C a little activity can still be measured. Similar results were observed for *TrCel61A* in the previous chapter.

6.2.2 Stabilization through rational inspection of ScLPMO10C's protein surface

Selecting residues for mutagenesis

From the analysis of flexible regions earlier (Figure 6.3 on page 158), two flexible protrusions were found in the protein surface (FLEXI 1 and FLEXI 3). Their closest connection is lysine 53 (K53) and aspartate 186 (D186) at a distance of 6.5 Å. While these residues are oppositely charged, this distance is too large to indeed form a salt bridge (<4 Å)⁴⁴⁵. Therefore, in a first approach, an attempt was made to facilitate the formation of a salt bridge by decreasing this distance and at the same time fixating both loops in order to stabilize the enzyme. This can be obtained by the use of amino acids with a larger side chain, while maintaining the opposite charges. Accordingly, 2 mutations are proposed: change lysine 53 for arginine (K53R) and/or aspartate 183 for glutamate (D183E) resulting in formation of 3 putative salt bridges (new shortest distances for rotamers without steric clashes, calculated with PyMol, are found between brackets): K53–D183E (4 Å), K53R–D183 (5.5 Å) or K53R–D183E (3.6 Å) in the 3 different enzyme variants.

As a second approach, the crystal structure of ScLPMO10C (PDB 4OY7-chain A) was

visualized in PyMol⁷³ and the entire protein surface was inspected for some additional interactions, mainly at the surface and focused around the flexible regions. Hereby, the copper-coordinating residues were left untouched as well as the amino acids within a distance of 6 Å from the copper ion. Furthermore, the alignment in figure S6.1 (page 175) was used as a guideline not to alter conserved residues (red box/white letters). This resulted in introduction of 5 ion pair interactions: 4 apolar residues at the surface were replaced by a polar alternative for hydrophilic interaction with the solvent and in 1 case an ion pair interaction was suggested to be facilitated. Indeed, the methyl group of threonine 175 was found as a possible physical barrier between the negatively charged aspartate 147 and the positively charged arginine 149 and was therefore substituted by a serine (T175S). An overview of all suggested mutations are given in the table below (table 6.1).

Strategy	Proposed mutation
Flexible loop connection	K53R D183E K53R + D183E
Ion pair network	L50D (Connecting to K53) A58K (connecting to D60) T62K (Connecting to D59) T173D (Connecting to R149) T173R (Connecting to D149)
Avoid apolar residues at protein surface	A58S A195S L170T A94S
Avoid -CH ₃ group to interrupt an ion pair	T175S

Table 6.1: Overview of rational, site-directed mutations in ScLPMO10C.

Effect on the stability

ScLPMO10C and its enzyme variants were expressed in the periplasmic space of *E. coli*. An SDS-PAGE containing all purified variants is shown in supplementary figure S6.3, page 178. As screening method for improved (thermo)stability, the protein's melting temperature (T_m) of all variants was measured using differential scanning fluorimetry (DSF), as shown in S6.5, page 179. This is a parameter that represents a protein's tendency to unfold as a function of temperature²³¹.

The comparison in T_m is shown in figure 6.5. The loop connecting variants showed 1 increase of 1 °C for mutation K53R. Since this difference is not bigger than the measurement error, it is not considered as a positive hit. The other substitutions did not differ from the wild type, suggesting that formation of an ion pair could not be obtained or did not alter the thermodynamic stability of the protein. Secondly, introducing additional ion pairs at the surface appeared to be a more tedious task because it results in higher differences in T_m , all decreasing in comparison to the wild-type. Finally, increasing polar interactions with the

solvent yielded equally stable enzymes as well as removing the possibly interfering methyl group.

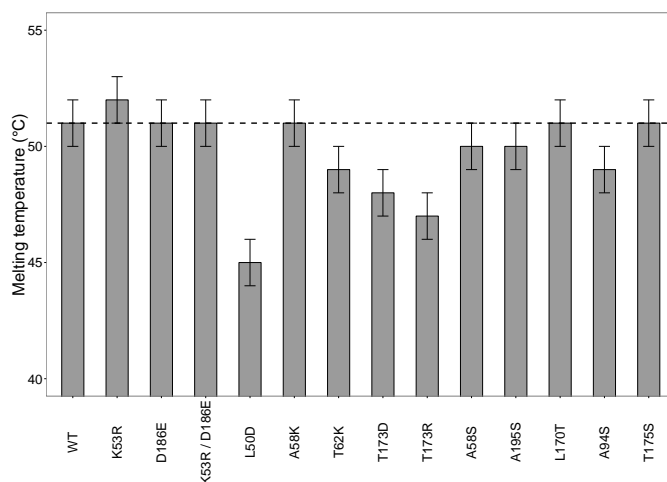


Figure 6.5: Effect of rational, site-directed mutations on the protein's melting temperature. (n=3, error = standard deviation)

However, the stability gain frequently comes at the expense of functional efficiency, presumably by over-rigidifying catalytic motions.

Despite the fact that decrease or loss of activity was not expected from changing surface residues far from the active site, over-rigidifying is frequently described to result in loss of functional efficiency of the protein^{197;446}. Additionally, the flexible regions at the binding surface suggest the need for flexibility. Therefore, a single overnight activity test on PASC was done to verify activity was not lost. The activity here is the amount of cellobionic acid formed, measured as the peak area obtained after HPAEC-PAD analysis. Other products showed similar patterns. The graph indeed reveals similar activities for all variants (figure 6.6).

The mutations suggested here are based on a clear rational mutagenesis strategy that requires a thorough insight into the structure - function relationships, but anyhow difficult to predict because of complex networks that are formed. However, targeting regions with high B factor has already proved its success^{213;436}. What further complicates the procedure is that very low increases in T_m can be expected from such an approach, so that multiple concerted mutations will only lead to significant increases^{324;410}. A smarter method could probably increase the success ratio. One option is to take the strength of the charge-charge interaction into account (depends on spatial orientation) and the local context of the salt bridge⁴⁴⁷. An even more elaborate alternative, that is more universally applicable but also requires a lot of technical knowledge, is to use computational models to calculate the energy and optimize the surface charges. Such an approach has been described successful by Gribenko and coworkers for 2 distinct enzymes⁴⁴⁸.

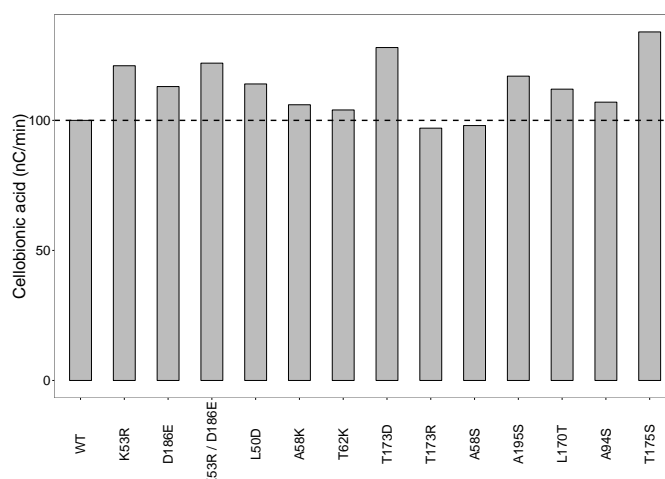


Figure 6.6: Effect of site-directed mutations on the protein's activity. Measured as the cellobionic acid produced after overnight incubation at 37 °C and products measured with HPAEC-PAD. (n=1)

6.2.3 Stabilization via disulfide engineering

Selecting targets for disulfide introduction

In this second approach, additional disulfide bridges will be introduced in ScLPMO10C for increased thermostability. In addition, disulfide bridges seem very important for the LPMO: indeed, they are unusually formed in the cytoplasm of the prokaryotic host and they are responsible for an increased heat tolerance. Therefore, the disulfide linkages are further studied in this section.

To introduce additional disulfide bonds, only the catalytic domain (CD) of the enzyme was considered since not all AA10 members harbor a CBM. Additionally, the computational tools that were used for the mutational design require a crystal structure as input, which is only available for the catalytic domain of ScLPMO10C (residues 35-230).

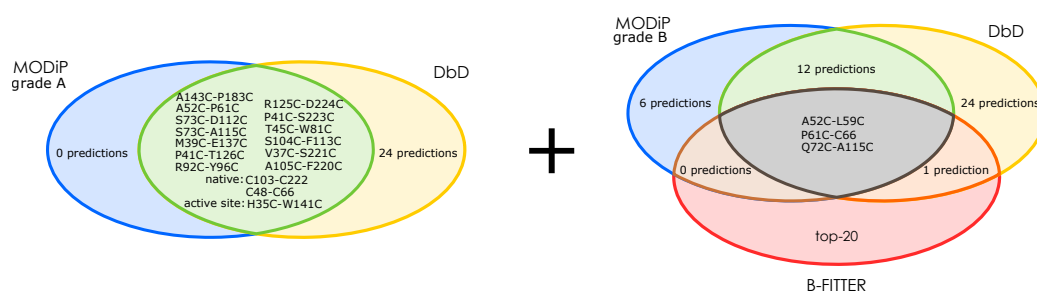


Figure 6.7: Distribution of all predicted disulfide bridges. Left: combination of MODiP-grade A and DbD, wherein the suggestions in the green overlap were evaluated except for the native crosslinks and options involving the active site. Right: combination of MODiP-grade B, DbD and flexible residues, wherein only the options in the black overlap were evaluated.

Target positions were selected based on the combination of 2 computational tools. The first tool, called MOdelling of DIsulfide bridges in Proteins (MODIP)^{387;389}, resulted in a list of 57 possible linkages, graded A to D reflecting a decreasing geometric and steric feasibility. Sixteen of them were graded A and 21 graded B. The options with grade C and D were not further considered since they are not very likely to be formed. The second tool, Disulfide by Design (DbD, v 2.12)³⁸⁶ yielded a list of 40 possibilities. All 16 A-graded options of MODIP were predicted by DbD too, including the 2 native disulfide bonds. Since these show the highest possibility to be formed, they will be considered for mutation experiments. From the 21 B-graded MODIP predictions, 15 were predicted by DbD too (overview figure 6.7). Because a B-graded MODIP score means the disulfide bridges are less likely to be formed, an additional restriction was added. This B-list of 15 options was thus reduced to only 3 options by solely considering bonds that include a residue in the top 20 (or 10 %) of most flexible positions as determined by their B-factor²¹². Furthermore, to avoid activity loss, the targeted cysteine couples that included one of the copper coordinating residues (His-35, His-144 and Phe-219) were eliminated from the entire A and B-list. In total, a list of 16 disulfide introductions was preserved in this way and can be found in table 6.2. The residues part of the top 20 most flexible positions are indicated with an asterisk and also the sum of B factors from both possible cysteine partners is listed. Interestingly a rearrangement of a disulfide bridge was also included: the native bond C48-C66 being substituted by P61C-C66. Therefore, cysteine at position 48 was mutated into an alanine to avoid expression hurdles or aberrant disulfide bonds due to the free -SH group^{346;347;449}.

Effect on thermal stability

All of the predicted disulfide bridges were then introduced individually by site-directed mutagenesis (SDS-PAGE gel with purified enzymes in supplementary figure S6.4). The stability of the corresponding enzyme variants was screened by DSF (Table 6.2 and Fig. S6.6). Unfortunately, the effect of three of the suggested crosslinks could not be measured because the protein expression was much too low for practical purposes. The other linkages showed a variable effect on T_m , *i.e.* nine resulted in a significant decrease ($p < 0.05$, Student t-test) and four in an increase ($p < 0.05$, Student t-test). Interestingly, mutation S73C was involved in two beneficial combinations, but only the one with the highest increase (S73C-A115C) will be used in further combinations. Despite the fact that A115C was involved in the highest thermostability improvement, this mutation was also part of a slightly destabilizing disulfide introduction (Q72C-A115C).

It should be noted that few variants failed to express, even after various optimization attempts (Table 6.2). This positional effect has earlier been described for disulfide introductions³⁴⁶ but is very hard to predict or rationalize. Furthermore, the effect of a single mutation can also be remarkably variable, as several residues could form the basis of more

than one bond but with a different outcome, *i.e.* from 2 beneficial combinations (S73C), over a beneficial and neutral introduction (A115C), a beneficial and disadvantageous effect (P61C), to non-expressed proteins (P41C). This suggests that the stabilizing effect is caused by a specific residue couple and cannot be predicted by one residue only.

Nr	Residue 1	Residue 2	MODIP-grade	DbD	Sum of B-factors	T_m (°C)	ΔT_m (°C)
WT	C48	C66	A	X	20.96	51 ± 1	-
	C103	C222	A	X	13.89		
1	A143C	P183C *	A	X	32.78	53 ± 1	+2
2	A52C	P61C *	A	X	31.40	56 ± 1	+5
3	A52C	L59C *	B	X	30.75	45 ± 1	-6
4	P61C *	C66 (C48A)	B	X	27.78	N.d.	/
5	Q72C *	A115C	B	X	24.48	49 ± 1	-2
6	S73C	D112C	A	X	22.66	54 ± 1	+3
7	S73C	A115C	A	X	21.04	60 ± 1	+9
8	M39C	E137C	A	X	18.52	47 ± 1	-4
9	P41C	T126C	A	X	17.3	N.d.	/
10	R92C	Y96C	A	X	16.48	44 ± 1	-7
11	R125C	D224C	A	X	15.49	40 ± 1	-11
12	P41C	S223C	A	X	15.03	N.d.	/
13	T45C	W81C	A	X	15.01	44 ± 1	-7
14	S104C	F113C	A	X	14.72	40 ± 1	-11
15	V37C	S221C	A	X	13.45	44 ± 1	-7
16	A105C	F220C	A	X	12.07	46 ± 1	-5

Table 6.2: Overview of all extra disulfide bonds evaluated in this study with their resulting melting temperature (T_m). The melting temperature (T_m) of 3 combinations could not be determined (N.d.). The bold face and asterisk (*) after the residue means the WT residue is in the top 20 of most flexible residues. The sum of B factors of both residues is also given. The stabilizing linkages are highlighted in gray.

Every disulfide bond that increased the thermal stability was also evaluated in combination with another stabilizing linkage via a full factorial design (table 6.3). An SDS-PAGE gel with all purified enzymes can be found in figure S6.4 and DSF curves in figure S6.7. Since serine 73 was involved in 2 beneficial combinations, only the one with the highest increase (S73C-A115C) was used in combinations. The combined variants revealed that their effect on T_m is more or less additive. However, a second peak at a lower temperature in the melting curve started to emerge in all cases (Figure 6.8). The combination of A143C-P183C with either A52C-P61C or S73C-D115C (Double 1 (D1) and Double 2 (D2)) resulted in both cases in an additional increase leading to a T_m of 59 °C and 63 °C respectively, and a second signal that is relatively small. However for the combination A52C-P61C and S73C-D115C (Double 3 (D3)), the left peak became more prominent than the right peak. With the variant that contained all three of the additional disulfide bridges (Triple (T)), the unfolding curve even appeared to be the result of multiple overlapping peaks, from which T_m -values of 40, 51 and 70 °C could be deduced.

To summarize the results, the thermal stability was increased based on the melting temperature from 51 °C for the wild-type to 60 °C for the best single disulfide introduction (S73C-A115C) and 63 °C for the best combined disulfide introduction (A143C-P183C and S73C-A115C, Double 2 in table 6.3). Since biomass degradation is usually performed at

Combinations	A143C-P183C ($\Delta T_m = +2$ °C)	A52C-P61C ($\Delta T_m = +5$ °C)	S73C-A115C ($\Delta T_m = +9$ °C)	T_m (°C) *	ΔT_m (°C)
Double 1 (D1)	X	X		33 ± 1 (8 %) 59 ± 1 (92%)	-18 +8
Double 2 (D2)	X		X	35 ± 1 (22 %) 63 ± 1 (78 %)	-16 +12
Double 3 (D3)		X	X	46 ± 1 (69 %) 67 ± 1 (31 %)	-5 +16
Triple (T)	X	X	X	40 ± 1 (42 %) 51 ± 1 (47 %) 70 ± 1 (11 %)	-11 +0 +19

Table 6.3: Overview of all combined variants and the apparent melting temperature. *: Values derived from the different peaks in the spectrum, with the relative amplitude (peak heights) in between brackets. (n=3, error = standard deviation)

50-60 °C, increasing the T_m to values above that temperature is a very useful outcome. Indeed, the classical cellulases have already been extensively engineered for exactly that purpose. Lantz and coworkers present an overview of industrially relevant *Trichoderma reesei* (now *Hypocrea jecorina*) cellulases with an apparent melting temperature ranging from 61 to 77 °C. They even improved the mixture by engineering *HjCel6A* to obtain a T_m increase from 67 to 74 °C²⁵. So even the T_m of our best variant is still at the lower end of the spectrum and further engineering techniques might be required to avoid limiting circumstances due to the LPMO performance.

Additionally, the activity was measured after overnight incubation with PASC to make sure activity was not lost due to the disulfide introductions. The resulting amount of cellobionic acid formation is shown in figure S6.8. However, this must be interpreted carefully because the test was only performed once. The activity might be decreased for 2 beneficial mutants S73C-D112C and S73C-A115C. In contrary, these decreases were not observed in combinations, suggesting the observed decrease included a measurement error. Furthermore, 2 disulfide introductions might increase the activity at 37 °C. More thorough investigation is therefore required to make any conclusion.

Improved kinetic stability due to additional disulfide bridges

In order to investigate the practical effects of the ambiguous melting pattern, the combined variants were also evaluated with respect to their resistance against inactivation (kinetic stability) at higher temperatures (2h at 80 °C). All enzymes showed a comparable activity before incubation, meaning that the mutations did not affect their catalytic machinery (Figure 6.9). Differences could be observed. However, when comparing the samples that were pre-incubated for two hours at 80 °C. In particular, variant D2 offered a clearly improved kinetic stability, with 58% of its activity remaining compared to only 34% for the wild-type enzyme (p=0.02, Student t-test). Interestingly, the variants for which the T_m curves showed

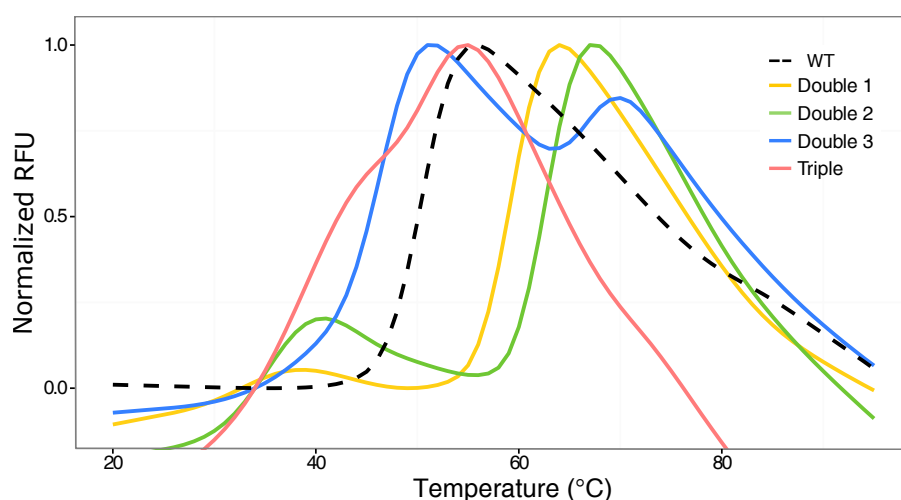


Figure 6.8: Melting curves of combined disulfide introductions. The melting curve of the wild-type enzyme is shown in black (dotted curve), while the combined disulfide introductions are colored as follows: Double 1 (A143C-P183C and A52C-P61C) in yellow, Double 2 (A143C-P183C and S73C-A115C) in green, Double 3 (A52C-P61C and S73C-A115C) in blue, Triple (A143C-P183C, A52C-P61C and S73C-A115C) in red. The experiments were performed in triplicate, but only one curve is shown here for clarity.

pronounced multiple peaks (*i.e.* D3 and T) had a very low residual activity, indicating that the mixed denaturation pattern has a negative effect on heat resistance.

When combining the individual disulfide bridges, an unexpected phenomenon was observed, *i.e.* the appearance of additional unfolding peaks in the melting curve (Fig. 6.8). Although one peak always seems to originate from the additional effect of the beneficial bonds, the other signals correspond to T_m -values that are considerably lower, even lower than that of the wild-type enzyme (Table 6.3). Furthermore, the variants for which the additional peaks became very prominent (D3 and T) also showed a diminished residual activity after heat treatment (Figure 6.9). One possible explanation could be the formation of unintended bonds between the available cysteine residues, resulting in isomers that are less stable than the target structure. Such behavior is known as disulfide scrambling⁴⁵⁰ and has been reported in literature for leech carboxypeptidase inhibitor⁴⁵¹, *Thermobifida fusca* Cel6A³⁴⁶ and IgG2 monoclonal antibodies⁴⁵². For example, residues Ala-52 and Ser-73 are both relatively close to the native cross-linked Cys-66 and could become part of a mismatched pair after their mutagenesis to cysteine. Remarkably however, the introduction of either A52C-P61C or S73C-A112C seems to have a beneficial effect in combination with the native C66-C48 bond, while the combination of all three crosslinks (in D3 or T) results in a deteriorated stability. When all parameters are taken into account, the best performing variant seems to be D2 (combination of A143C-P183C and S73C-A115C) as this variant shows only a very small aberrant peak in its melting curve and offers an almost doubled residual activity after heat treatment.

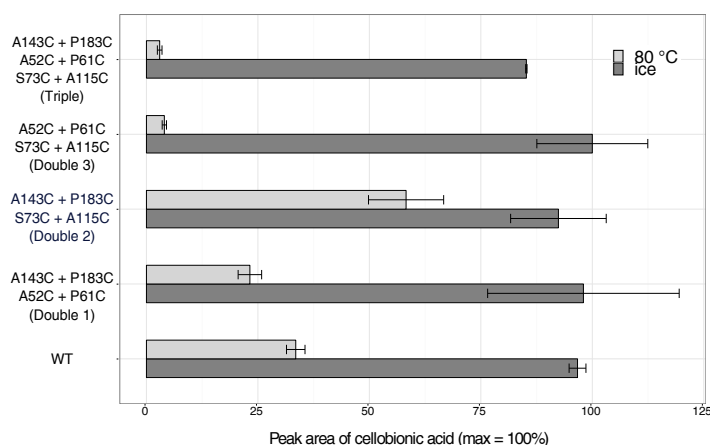


Figure 6.9: Effect of heat treatment on variants with combined crosslinks. The peak area of cellobionic acid was taken as comparison after HPAEC-PAD analysis. The dark gray shows the activity when the enzymes were incubated for 4 hours on ice. The light gray bars indicate the activity left when pretreated for 2 hours at 80 °C and subsequently 2 hours on ice. (n=3, error = standard deviation)

Structural implications

Disulfide engineering has been applied in the field of cellulases with varying outcome in terms of stability improvement^{377;382;453}. Moreover, the location of the crosslink has to be carefully chosen not to perturb the activity. For example, adding extra disulfide bridges across the tunnel region of the exocellulase *Thermobifida fusca* Cel6B did not yield an increase in melting temperature and resulted in a decrease in ligand binding so that lower activity and processivity were measured⁴⁵⁴. In the case of ScLPMO10C, the actual copper-coordination residues (His-35, His-144 and Phe-219) were left untouched. Interestingly, alanine 143 was involved in a putative crosslink with stabilizing outcome even though its neighboring residues histidine 144 (binds copper) and alanine 142 (shields copper^{114;246}) are crucial and strictly conserved. Additionally, the introduced crosslink **A143C**-P183C is situated in the cellulose binding surface. Nonetheless, similar activities (after incubation on ice) were measured for all of the combined variants (figure 6.9), what indicates the activity is left untouched. This could probably be expected because alanine 143 itself is not a conserved residue within the AA10 family (Figure S6.1) and the disulfide bond did not alter any of the aromatic residues in the surface, still allowing the protein to bind its substrate. Noteworthy, all other beneficial disulfide bonds are situated in the L2-loop (indicated in green on figure 6.10). This loop is suggested to be responsible for substrate specificity¹³, because it differs significantly in length and conformation between the different LPMOs in a family. In LPMO10C, it would be responsible for at least half of the substrate binding surface¹¹⁴. Nonetheless, all enzyme variants showed a comparable product profile that contains none of the keto-aldehydic acids that are characteristic for the other LPMO classes (Figure S6.9).

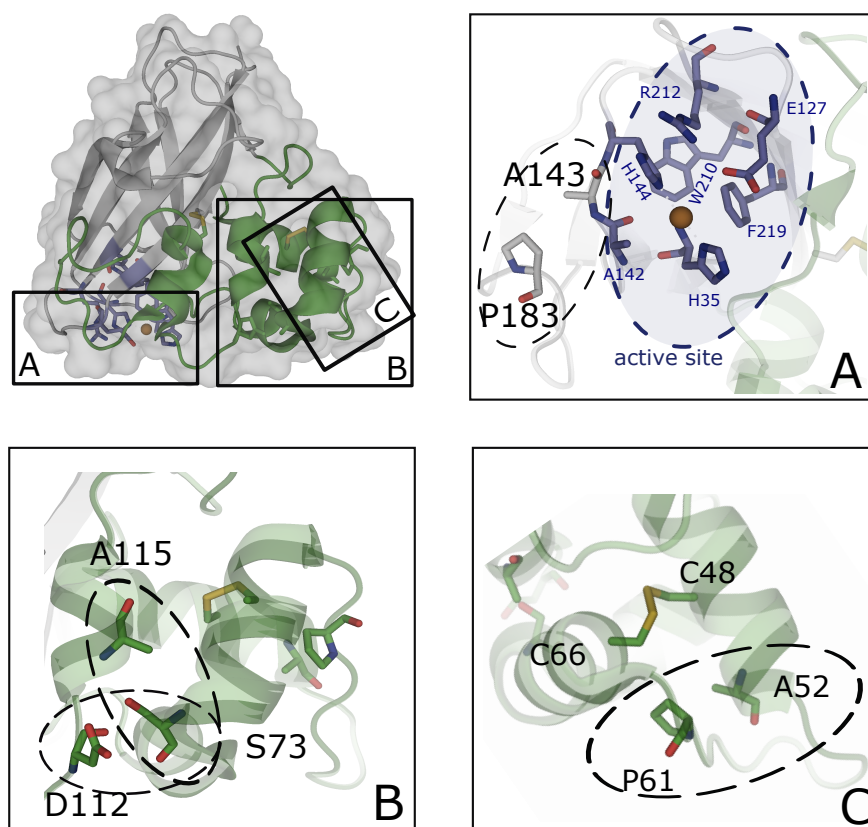


Figure 6.10: Detailed view of the location of the beneficial disulfide bridges. The variable L2 loop is colored in green and the active site is colored blue, while the copper ion is represented by an orange sphere and the native S-S bonds are colored yellow. An overview pictures is given where the cellulose binding surface is at the bottom of the protein. (A) Location of residues A143-P183 from a bottom point of view. They are located next to the active site, indicated in blue, and situated in the cellulose binding surface. (B) P61 and A52 can form a crosslink that is about parallel to the native C66-C48 bond. (C) S73 can form a disulfide bridge with either D112 or A115, both resulting in a higher melting temperature. Figures were made with PyMol⁷³

Figure 6.10 B and C shows the location of the 3 other beneficial disulfide introductions. Cysteine pair A52C-P61C forms a connection in the most flexible region of the protein (FLEXI 1 from figure 6.3 on page 158) and is almost parallel to the native bond C48-C66. Although longer loop lengths (>25 residues) are usually preferred for a stabilizing effect⁴⁵⁵, this shows an exceptional case for the native as well as the engineered linkage. The last two possibilities (Panel B) connect serine 73 to either alanine 115 or aspartate 112 (table 6.2). Serine 73 is not part of the 10 % most flexible residues, but residue 71 and 72 are, which might contribute to the stabilizing effect (for overview with flexible regions, see also additional figure S6.2). The connecting residues have a medium to low flexibility (B factor of D112 is 10.6 and A115 8.6 while the average is 10.5). All 3 residues are found at the protein surface what is known to have a beneficial effect³⁸⁷. The crosslink S73C-A115C shows the biggest increase in thermostability, probably because it connects 2 (almost) parallel alpha-

helices, which could highly increase the overall packing.

Reflections on the strategy of disulfide introductions

The results presented in this topic, describe the introduction of 16 additional crosslinks, a rather big amount that can teach some lessons for future disulfide engineering, especially in the light of set reduction. First of all, the last 6 couples in table 6.2 all had a negative effect on the melting temperature. This can be explained by the close proximity of a native disulfide bridge. Indeed they are all at 3 or less residues apart from a native cysteine. However, alanine 52 is situated at only 4 residues of native cysteine 48 and did yield a 5 °C increase in T_m in one case and a 6 °C decrease in another. Secondly, the combination of MODIP (grade A) and disulfide by design has been used before^{344;392;453} and shown to be valuable in this study. Despite the use of MODIP grade B was shown here to be omitted. Furthermore, B factors (as used before³⁹²) seems a reliable parameter to reduce the list of possible linkages. Taking MODIP grade A and the 5 highest sum of B factors would have reduced the tested set to 5 disulfide introductions only, and at the same time would have resulted in an enormous increase in success ratio. One must take care nonetheless, that not the result with the highest sum of B factors yields the highest T_m increase. In this case, the highest increase corresponded to only the fourth one on the list.

As a general remark, although stability engineering through disulfide introduction does not seem the method of choice for an enzyme produced in *E. coli*, we have shown here the potency the method holds in this expression host too. Most proteins are expressed in the cytosol while this environment does not provide the oxidative needs for disulfide formation. In LPMO research, correct processing of the protein is very important to obtain a free histidine at the N-terminus for an intact copper-coordination center. Cytoplasmic expression combined with the use of a cleavable part has shown its usefulness²⁶, but also expression in the periplasmic space is an option often considered^{142;246;306;456}. Periplasmic expression also holds other advantages than an intact N-terminus, namely the possibility to form disulfide bridges (presence of Dsb machinery and oxidative environment) and the lower protease content³¹². Our results show a positive outcome of this expression and engineering strategy combination. However, a drawback is the relatively low protein yields. This issue can be addressed by the use of knock-out strains that provide a sufficiently oxidizing environment with formation of disulfide bridges in the cytoplasm and provide the correct machinery^{457;458}. Furthermore, the protein can be obtained from the supernatant with successful disulfide introduction⁴⁵⁴. This can contribute to making the method more valuable in *E. coli*.

6.3 Conclusion

Rational engineering of surface residues via introduction of extra ion pair networks is a complex operation, that did not yield stabilized variants in this work. However, this method has proven valuable in former work. The use of computer algorithms and combination of multiple mutant residues could improve the success ratios.

Disulfide engineering was shown to be a powerful tool for increasing the thermodynamic stability of ScLPMO10C. Indeed, the enzyme's apparent melting temperature could be increased from 51 °C to 63 °C by the introduction of two additional disulfide bridges (A143C-P183C and S73C-A115C). Furthermore, this variant displayed a drastic increase in residual activity after incubation at 80 °C. Remarkably, however, some mutations resulted in a melting pattern with multiple peaks and concomitant deterioration of thermostability. This means that caution is advised when several cysteine residues are introduced in close proximity, since these can become part of unwanted bonds. For the design of disulfide bridges, the combination of MODIP (grade A) and DbD was confirmed to be especially valuable when targeting the most flexible positions (high B-factor) in the protein structure.

6.4 Materials and methods

The experimental procedure for differential scanning fluorimetry (DSF) can be found in chapter 4. For comparable results, the protein samples were diluted to the concentration of the sample with the lowest yield, with a minimum of 0,200 mg / mL.

6.4.1 Site-directed mutagenesis of ScLPMO10C

Streptomyces coelicolor LPMO10C preceded by its native secretion signal (ScLPMO10C-NSS) served as template for mutagenesis. The protein sequence is cloned in the constitutive expression plasmid pCXP34³¹⁷, explained in chapter 4.

All mutants were created using the method of Sanchis and coworkers²⁶⁸. However, instead of using a single PCR-reaction, the creation of the mega-primer and whole plasmid PCR were separately performed to obtain higher yields. For these reactions Primestar GXL DNA polymerase (Takara Bio Inc) and Pfu Ultra AD polymerase (Agilent Technologies) were used respectively following manufactures' instructions. The required forward primer sequences are listed in table 6.4, while the same reverse primer (5'-CATCCGTCAGGATGGCCTTC-3') was used for all mutants. After *dpnI* restriction digest, 20 μ L in-house prepared²⁹⁶ electrocompetent BL21 (DE3) cells was subjected to electroporation with 2 μ L of the mixture. Cells were selected on LB plates containing 100 μ g / mL ampicillin. The correct sequence of the mutated versions of ScLPMO10C was confirmed (Macrogen Europe) before enzyme production.

6.4.2 Protein expression and purification

E. coli cells were grown in 250 mL terrific broth medium (1.2 % tryptone, 2.4 % yeast extract, 0.4 % glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4) containing 100 μ g / mL ampicillin. The cultures were inoculated with 2.5 mL of an overnight-grown preculture and they were incubated at 30 °C for 18-20 hours, while shaking at 200 rpm.

The cells were pelleted by centrifuging 10 minutes at 3200 x g at 4 °C and the periplasmic fraction was isolated in two steps. The first fraction was obtained by re-suspending the pellet in 12.5 mL (1/20 of the culture volume) of a 20 % (w/v) sucrose, 100 mM Tris-HCl (pH 8) and 1 mM EDTA solution and subsequently incubating on ice for 30 minutes. After centrifugation (20 minutes 4500 x g), the supernatant was kept. Next, the cell pellet was resuspended in 12.5 mL 5 mM MgCl_2 solution and incubated for 20 minutes on ice. The suspension was centrifuged for 20 minutes at 4500 x g again and the resulting supernatant was pooled with the former one and stored at -20°C upon use, while the cell pellet was discarded.

The protein was purified by immobilized metal ion affinity chromatography (IMAC) using 1.5 mL Ni-NTA agarose slurry (MC-lab) in 10 mL purification columns. The resin was first

Mutation	Primer sequence (5' → 3')
Rational mutants (see section 6.2.2)	
L50D	CCTATCTGTGTCAGG AT GATGCAAAAACCGG
K53R	TGTCAGCTGGATGCA CGT ACCGGTACAGGTGC
A58K	ACCGGTACAGGT AA ACTGGATCCGACC
A58S	AAACCGGTACAGGT AG CTGGATCCGACC
T62K	GCACTGGATCCG AAAA ATCCGGCATGTC
A94S	CCGGTGGTCGTGGT AG CGGTTATGTTCC
L170T	GGGATGATCTGGAA ACC ATTAGACCGTTACC
T173D	GGAACTGATTCAG GAT GTACCAATCCGCC
T173R	GGAACTGATTCAG CG CGTTACCAATCCGCCTC
T175S	CTGATTCAGACCGTT AG CAATCCGCCTCAGC
D186E	CAGTCCGGGTACAG AA GGTGGTCACTATTATTGG
A195S	TATTATTGGGATCTG AG CTGCCGAGCGGTC
Disulfide introductions (see section 6.2.3)	
V37C	GAAGCACATGGT TGT GCAATGATGCCTGG
M39C	ATGGTGTTGCAT GC ATGCCTGGTAGC
P41C	TTGCAATGATG TGC GGTAGCCGTACCTATC
T45C	GCCTGGTAGCCGT TGT TATCTGTGTCAGC
C48A	AGCCGTACCTATCTG GCG CAGCTGGATGCAAAAACC
A52C	CTGTGTCAGCTGGAT TGT AAAACCGGTACAGG
L59C	ACCGGTACAGGTGCAT TGT GATCCGACCAATCC
P61C	ACAGGTGCACTGGAT TGT ACCAATCCGGCATG
Q72C	CAGCCCTGGAT TGT AGCGGTGCAAC
S73C	CTGGATCAG TGT GGTGCAACCGCAC
W81C	CCGCACTGTATAAT TG TTTTGCCGTTCTGGATAGC
R92C+Y96C	AATGCCGGTGGT TGT GGTGCAAGTT GCG TTCCGGATGG
S104C	CACCCTGTGT TGC GCCGGTGATCGTAG
A105C	CACCCTGTGTAGT TGT GGTGATCGTAGC
D112C	CGATCGTAGCCGTAT TGT TTTAGCGCATATAATGC
F113C	GCCCGTATGAT TGT AGCGCATATAATGCAGC
A115C	CCGTATGATTTAGCT TGT TATAATGCAGCACGTAGCG
R125C	AGCGATTGGCCT TGT ACCCATCTGACC
T126C	ATTGGCCTCGTT GCC ATCTGACCAGCGG
T129C	CCTCGTACCCATCTG TGT AGCGGTGCCACATTCC
E137C	CACCATTCCGGTT TGC TATAGCAATTGGG
A143C	GCAATTGGGCAT TGT CATCCGGGTGATTTTCG
P183C	CCTCAGCAGGGCAGT TGT GGTACAGATGGTGG
F220C	GCCAAGAAAACCTC TGT AGCTGCAGTGATGTGG
S221C	AACTTCTTT TGT TGCAGTGATGTGGTG
S223C	CTTAGCTGCT TGT GATGTGGTGTGGATGG
D224C	TAGCTGCAGT TGT GTGGTGTGGATGG
G230C	GTGGTGTGGATG TGT AATGGTGAAGTTACCGG

Table 6.4: Overview of the mutagenic forward primers used for site-directed mutations in ScLPMO10C.

washed with 10 mL distilled water and equilibrated with 2 x 10 mL PBS buffer (50 mM sodium phosphate and 300 mM sodium chloride at pH 7.4) containing 10 mM imidazole. The same concentration imidazole was added to the samples and the first 10 mL was applied to the column. The columns were incubated at 4 °C for 30-60 minutes while gently rotating to allow binding to the resin. The suspension is settled down by gravity. More sample is applied and incubated at 4 °C until all sample is loaded to the column. Next, the column is washed with 3 x 8 mL 20 mM imidazole in PBS-buffer and the protein is finally eluted in 10 mL 100 mM imidazole solution. The buffer is exchanged for 50 mM Sodium-acetate buffer pH 5 and 25 times concentrated using Vivaspin 6 columns with 10 K PES membrane (Sartorius). The protein content was measured using a Nanodrop (absorbance at 280nm). The extinction coefficient of 75775 M⁻¹ cm⁻¹ was calculated using the Protparam tool on the ExPASy server (<http://web.expasy.org/protparam/>). The purity was evaluated

using SDS-PAGE (12% gels) and subsequent coomassie staining (QC Colloidal coomassie, Biorad).

6.4.3 Activity testing and HPAEC-PAD

The activity on phosphoric acid swollen cellulose (PASC) was measured to ensure all mutants had remained active. For that purpose, 5-10 μ M of ScLPMO10C was mixed with 1 mM ascorbic acid and 1.2 % (PASC), prepared from Avicel PH-101 (Sigma- Aldrich) following instructions from Wood²⁷⁶. A 10 mM sodium acetate buffer pH 5 was used to adjust the volume to the final volume (usually 150 μ L or 750 μ L). The Eppendorf tubes were incubated overnight at 37 °C while shaking at 1400 rpm in an Eppendorf Thermomixer. The reaction was stopped by heat-inactivated the enzyme at 95 °C for 10 min. The reaction products were analyzed by HPAEC-PAD using the method described by Forsberg and coworkers²⁶. The peak area of cellobionic acid was used as a measure for activity since it is a characteristic product that gives a strong signal and is nicely separated from the others.

6.4.4 Heat treatment

To evaluate the heat resistance, the best mutants were subjected to a treatment at 80 °C (hot water bath) for 2 hours and subsequent incubation for 2 hours on ice. Identical enzyme aliquots were incubated for 4 hours on ice at the same time. A mixture containing ascorbic acid, PASC and buffer was added to the enzyme after incubation to start an activity test at 37 °C. The peak area of the products after HPAEC-PAD analysis was used to compare the amount of activity with or without heat treatment. If necessary, disulfide bridges were chemically broken by addition of 10 mM tris(2-carboxyethyl)phosphine (TCEP)^{441;442}.

Supplementary material

Multiple sequence alignment

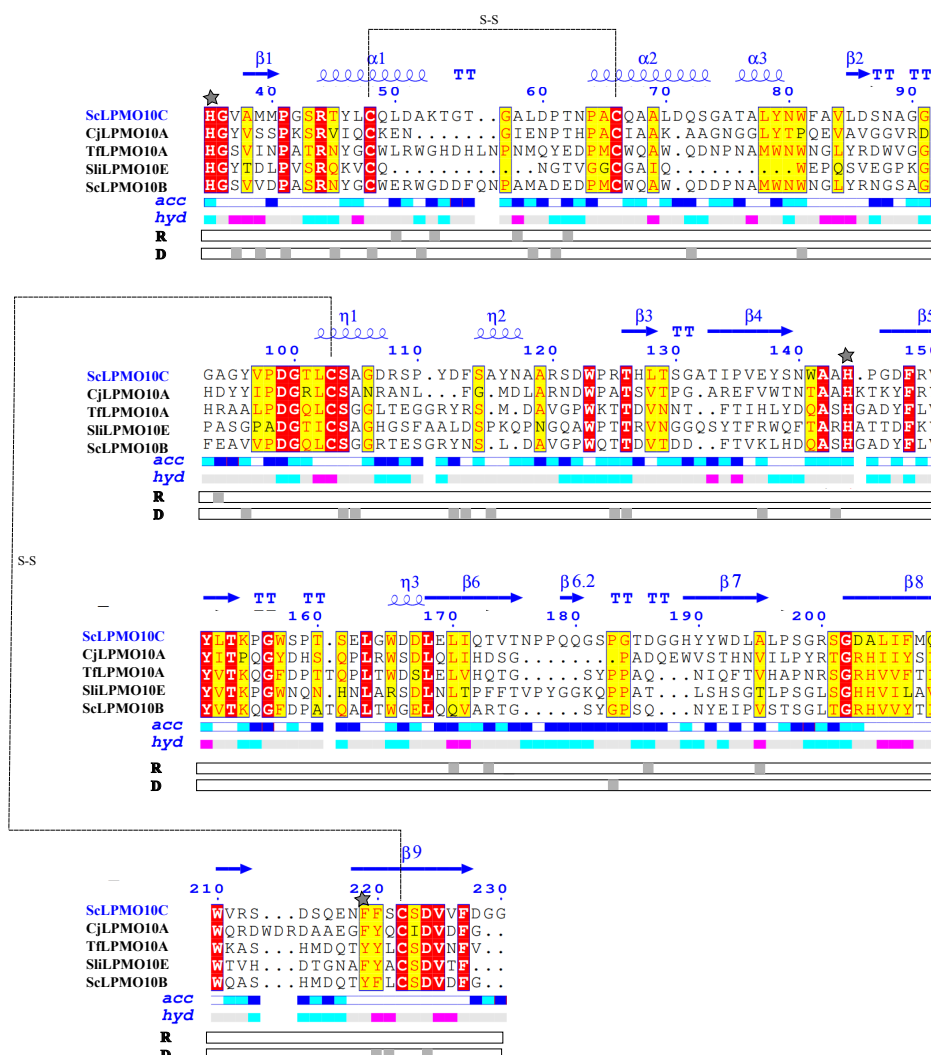


Figure S6.1: Multiple sequence alignment (MSA) of ScLPMO10C and the bacterial AA10 members with known crystal structure and highest sequence identity (Table S6.1). The secondary structure is indicated on top, as well as the disulfide bridges (S-S). (red box/white character: strictly conserved residue; yellow box/red character: similarity within group (black letter differs); acc: accessibility score (white is buried, cyan is intermediate, blue is accessible and blue with red border is highly exposed), hyd: hydrophobicity score (pink is hydrophobic, gray is intermediate and cyan is hydrophilic, R: residues involved in rational design approach (section 6.2.2), D: residues involved in disulfide engineering strategy (section 6.2.3). A gray star indicates the copper coordinating residues. Figure prepared with (and adapted from) Endscript (<http://endscript.ibcp.fr/ESPrpt/ENDscript/>)⁴⁰⁸.

Organism	LPMO	UniProt ID	PDB	% identity	Active site residues	Substrate	Source
<i>Streptomyces coelicolor</i>	ScLPMO10C	Q9RJY2	4OY7	100	H35-H144-F219	cellulose (C1)	114
<i>Cellvibrio japonicus</i>	CjLPMO10A	B3PJ79	5FJQ	35.80	H37-H136-F207	chitin	138
<i>Thermobifida fusca</i>	TfLPMO10A (E7)	Q47QG3	4GBO	32.40	H37-H144-Y213	cellulose (C1/C4), chitin	114
<i>Streptomyces lividans</i>	SlLPMO10E	-	5FTZA	32.35	H30-H120-F193	chitin (C1)	439
<i>Streptomyces coelicolor</i>	ScLPMO10B	Q9RJC1	4OY8	31.28	H43-H150-Y219	cellulose (C1/C4), chitin (C1)	114
<i>Anomala entropoxvirus</i>	fusolin (ACV34)	-	4YN1	27.37	H1-H144-F225	N.d.	167
unidentified entropoxvirus	fusolin	-	4YN2	27.13	H1-H142-F222	N.d.	167
<i>Vibrio cholerae</i>	VcAA10b	Q9KLD5	2XWX	25.86	H24-H121-F193	chitin (C1)	307
<i>Melolontha melolontha</i>	fusolin	Q83389	4X29	25.79	H1-H144-F225	N.d.	167
<i>Bacillus amyloquefaciens</i>	BaAA10A (ChbB)	E1UUUV3	2YOX	25.14	H8-H125-F196	α and β chitin	144;246
<i>Enterococcus faecalis</i>	EfAA10A (EfCBM33A)	Q838S1	4ALT	24.53	H29-H114-F185	chitin(C1)	459
<i>Jonesia denitrificans</i>	JdLPMO10A	C7R4IO	5AA7	22.70	H32-H109-F164	chitin (C1)	460
<i>Serratia marcescens</i>	SmAA10A (CBP21)	O83009	2BEM	22.62	H28-H114-F187	chitin (C1)	461
<i>Burkholderia pseudomallei</i>	BpAA10A	Q3JY22	3UAM	19.34	H19-H122-F205	N.d.	-

Table S6.1: Overview of all AA10 members with revealed crystal structure. The PDB code, active site residues and substrate specificity are given as well. The percentage of identity is compared to ScLPMO10C.

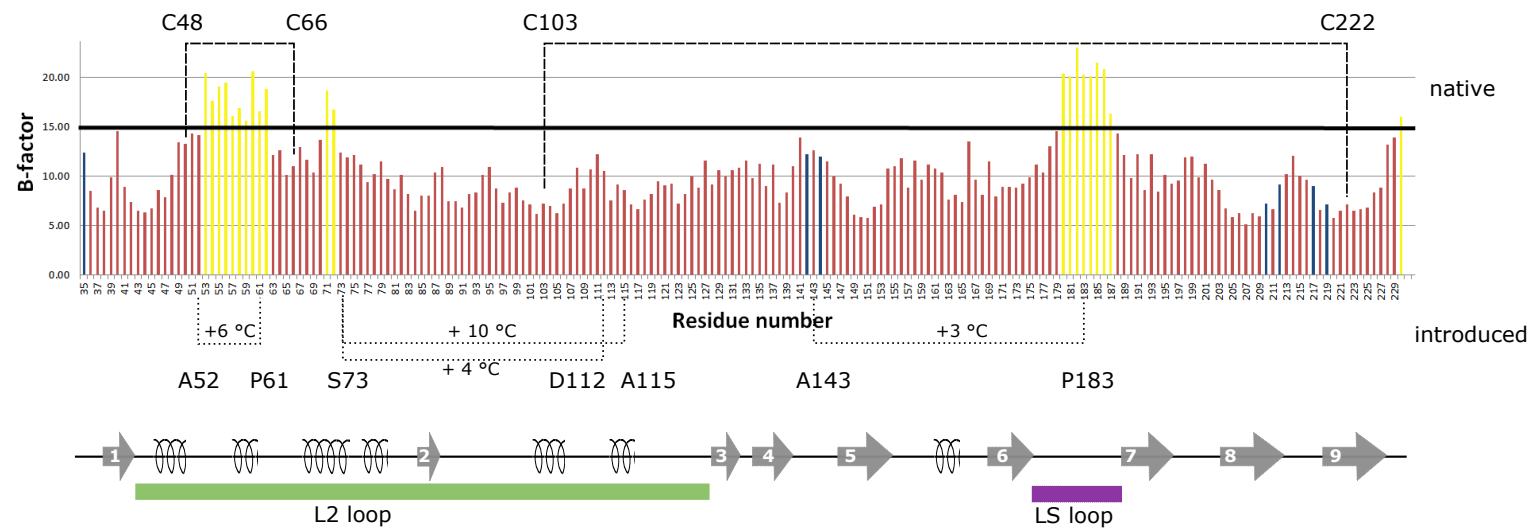


Figure S6.2: Beneficial disulfide introductions linked to B factor. The upper part shows again the average B-factor per residue in function of the residue number. Yellow bars are the top 20 most flexible amino acids, blue bars represent the residues within a distance of 6 Å from the copper ion. The native disulfide bridges are indicated on top, while the introduced and beneficial ones are drawn below the diagram. The bottom figure adds the secondary structure of the protein, including an indication of the flexible L2 and LS loops.

SDS-PAGE analysis

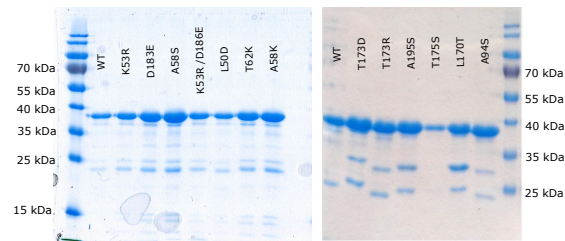


Figure S6.3: SDS-PAGE analysis of all rational mutants of ScLPMO10C.

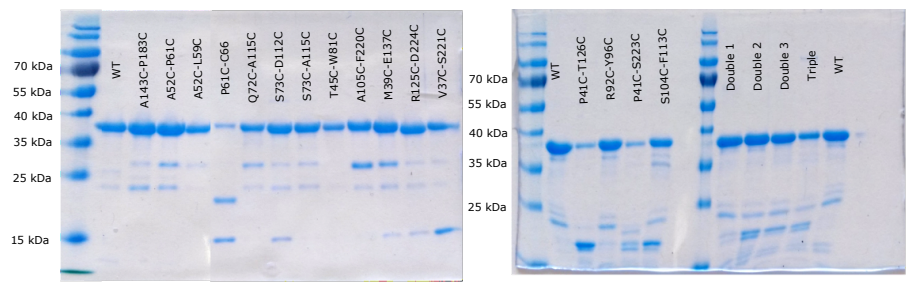


Figure S6.4: SDS-PAGE analysis of all variants of ScLPMO10C with introduced disulfide linkages.

Melting curves

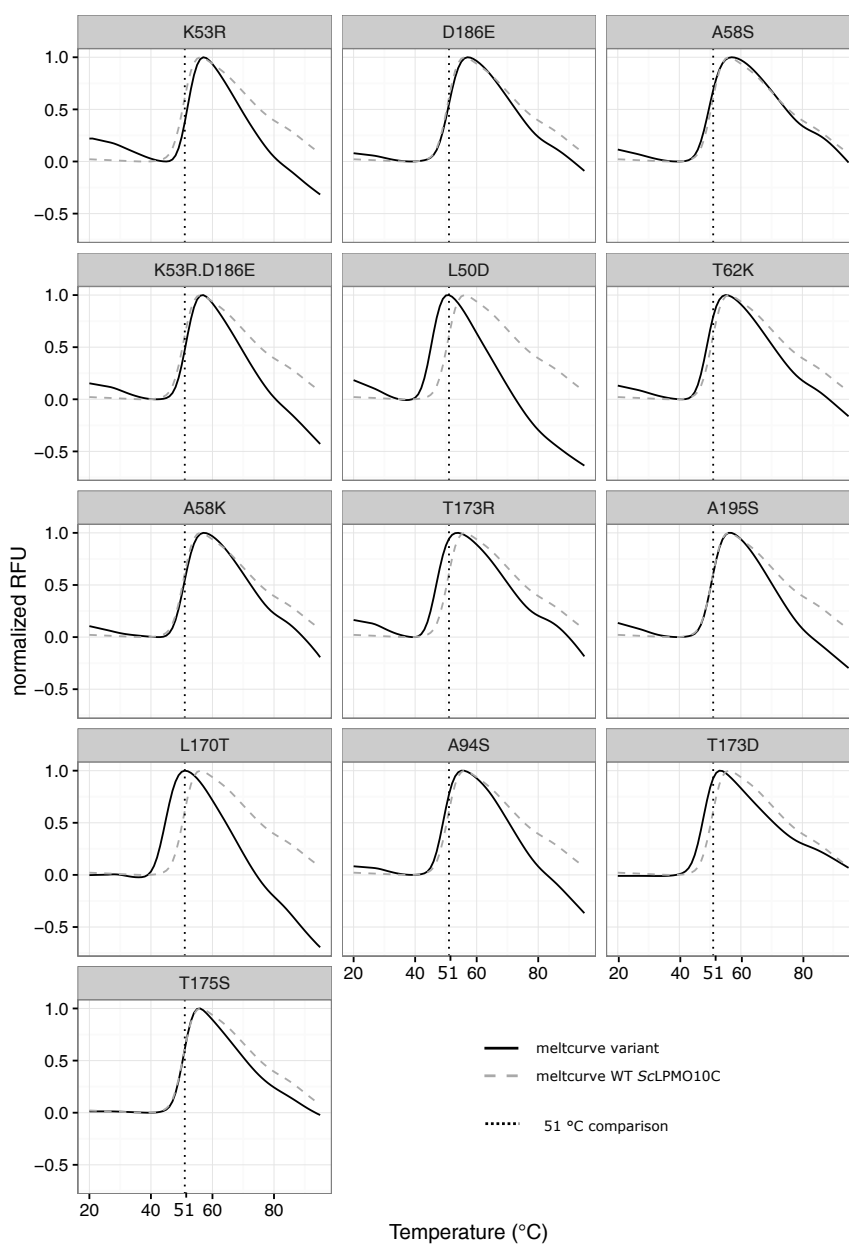


Figure S6.5: Overview of DSF curves for all mutants for the rational engineering strategy. The inflection point of the curve corresponds to the apparent melting temperature. This is 51 °C for the wild-type ScLPMO10C, as indicated with a black dotted line. Experiment was performed in triplicate, but only one curve is shown here for clarity. The relative fluorescent units (RFU) were normalized for better comparison.

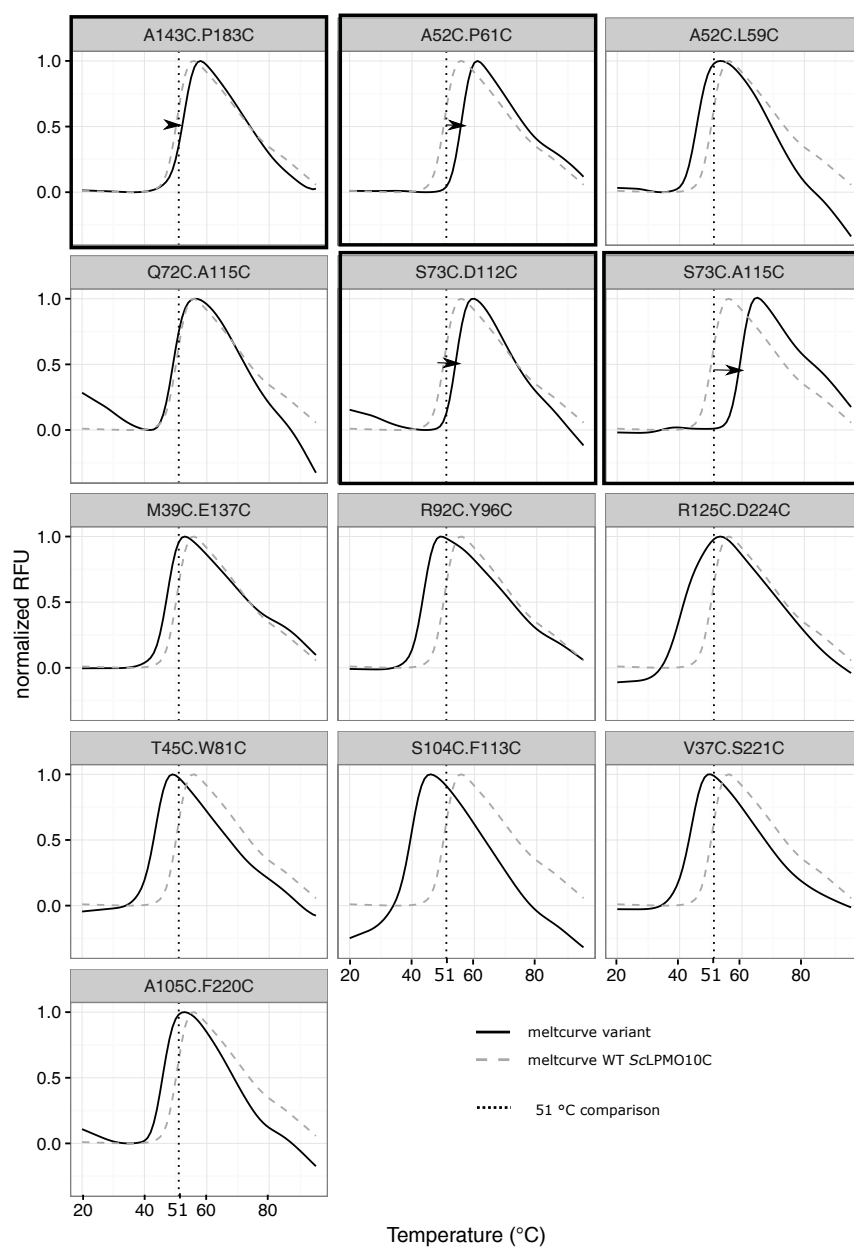


Figure S6.6: Overview of melting curves for all mutants for the disulfide engineering strategy. The melting curve shows the increase in fluorescence while heating the sample from 20 to 95 °C (1 °C / min increment). The inflection point of the curve corresponds to the apparent melting temperature. This is 51 °C for the wild-type ScLPMO10C, as indicated with a gray dotted line. Experiment was performed in triplicate, but only one curve is shown here for clarity. The relative fluorescent units (RFU) were normalized for better comparison. The black frames indicate the beneficial variants.

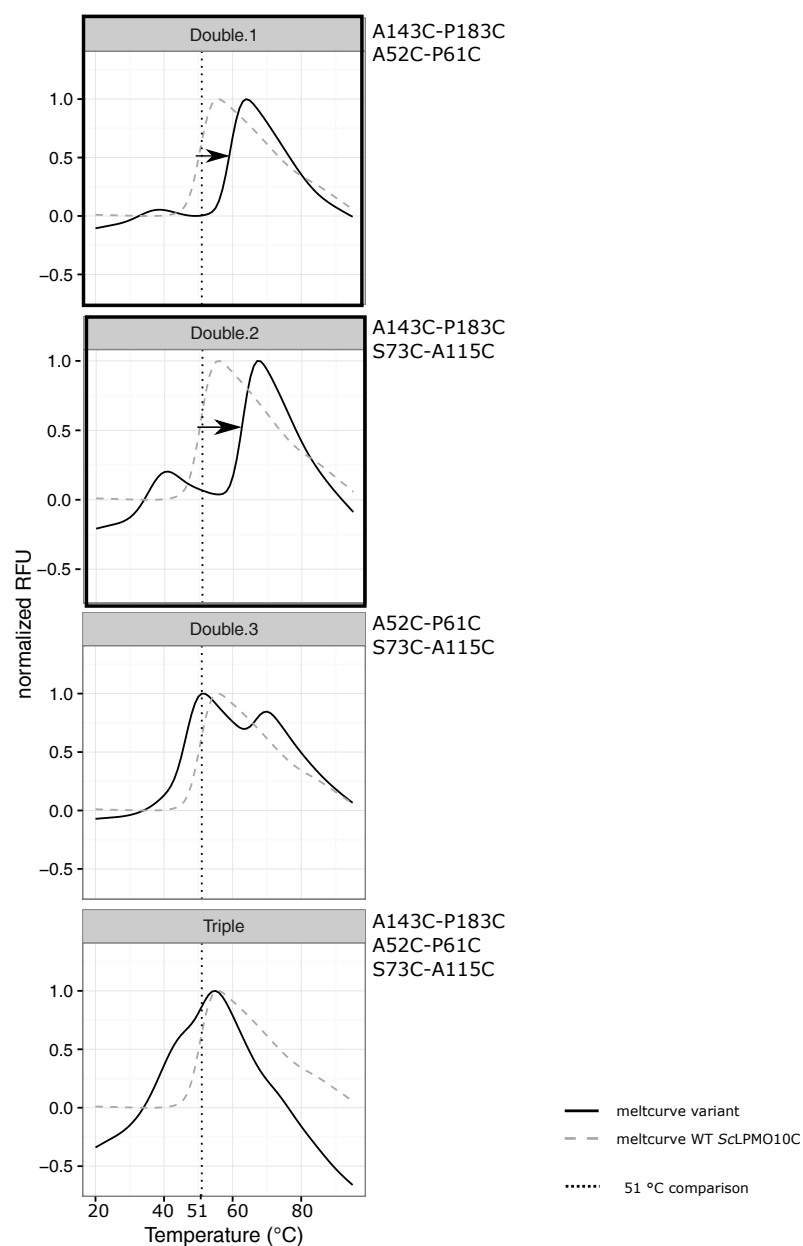


Figure S6.7: Overview of melting curves for combined mutants in the disulfide engineering strategy. The melting curve shows the increase in fluorescence while heating the sample from 20 to 95 °C (1 °C / min increment). The inflection point of the curve corresponds to the apparent melting temperature. This is 51 °C for the wild-type ScLPMO10C, as indicated with a gray dotted line. Experiment was performed in triplicate, but only one curve is shown here for clarity. The relative fluorescent units (RFU) were normalized for better comparison. The black frames indicate the beneficial variants.

Others

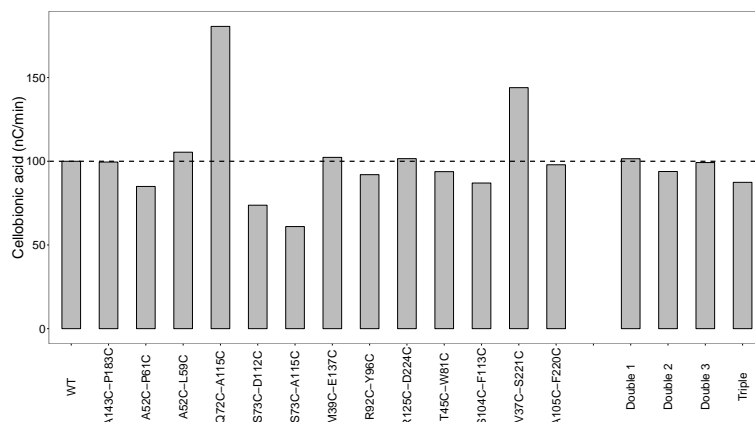


Figure S6.8: Effect of additional disulfide bridges on the protein's activity. Measured as the cellobionic acid produced after overnight incubation at 37 °C and products measured with HPAEC-PAD. (Double 1: A143C-P183C and A52C-P61C, Double 2: A143C-P183C and S73C-A115C, Double 3: A52C-P61C and S73C-A115C, Triple: A143C-P183C and A52C-P61C and S73C-A115C) (n=1)

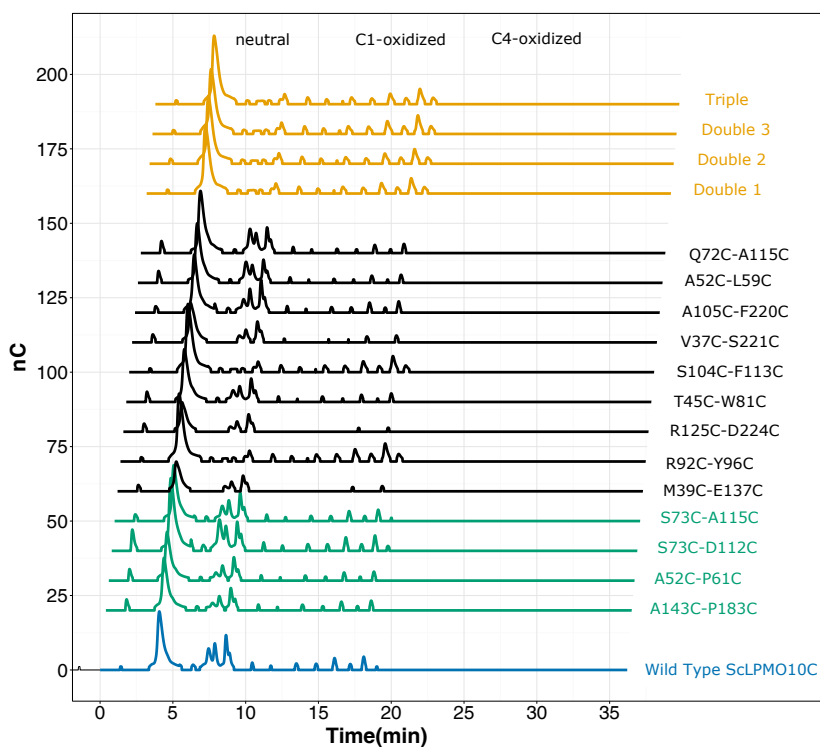


Figure S6.9: HPAEC-PAD profile of all disulfide variants. The proteins were incubated overnight with PASC at 37 °C, pH5. After inactivation (95 °C), the product profile was analyzed using HPAEC-PAD. Neutral cello-oligosaccharides and aldonic acids (C1-oxidized) are formed, while keto-aldonic acids (C4-oxidized) are lacking. Beneficial single and combined disulfide introductions are indicated in green and gold, respectively.

7

General discussion and outlook

7.1 Overview

The discovery of lytic polysaccharide monooxygenases (LPMOs) has revolutionized our view on cellulose degradation^{14;163}. Indeed, these copper-dependent enzymes boost the activity of the classical endo- and exocellulases by making the substrate more amenable after oxidative cleavage of crystalline cellulose¹⁵. The biggest bottleneck in making lignocellulosic biomass a suitable feedstock for second generation biorefineries is the cost, mainly attributed to pretreatment and required high enzyme loadings. Therefore, these LPMOs have the potential to reduce the costs⁴⁶². As a matter of fact, they can potentially decrease the severity of the pretreatment, reduce enzyme loadings or shorten saccharification time.

In this PhD thesis, the final goal was to study and improve the stability of LPMOs through various protein engineering strategies. Indeed, stability is an industrially important parameter, yet unexplored for LPMOs. To that end, two different enzymes from the two main LPMO classes for biomass degradation were selected as case study¹⁹, namely *Trichoderma reesei* Cel61A as representative of auxiliary activity family 9 (AA9) and *Streptomyces coelicolor* LPMO10C of auxiliary activity family 10 (AA10). For each enzyme, three subsequent stages were run through (Figure 7.1): first, the enzymes were heterologously expressed. Next, an evaluation method for stability and activity was tested, where after different protein engineering techniques were applied, as listed in the scheme.

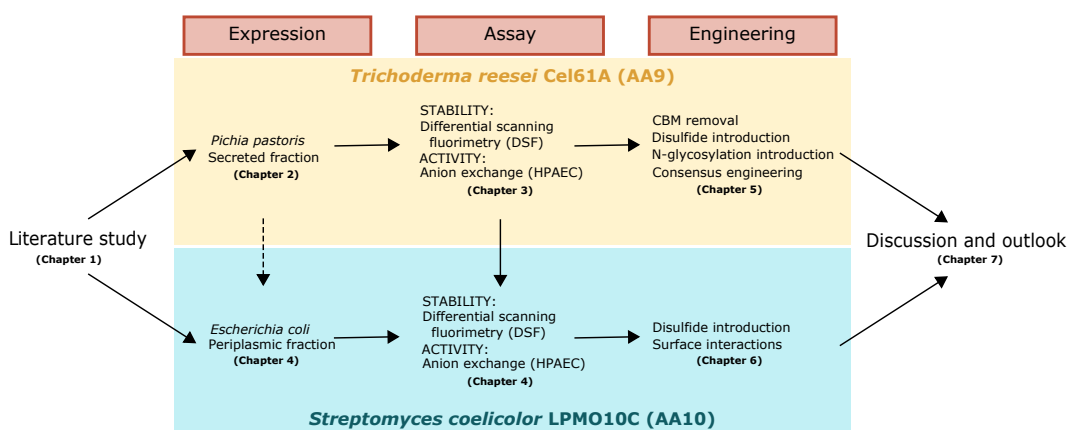


Figure 7.1: Schematic overview of the PhD thesis. (vertical arrow point at transfer of knowledge from one part to succeed in another)

In summary, the ten major achievements of this PhD thesis are:

1. **Construction of an inducible expression system for *TrCel61A* in *P. pastoris*** and proof that the native secretion signal is superior in comparison to the generally used

α -mating factor from *S. cerevisiae* considering both yield and N-terminal processing (Chapter 2).

2. **Construction of a constitutive expression system for ScLPMO10C in *E. coli*** with both a cleavable N-terminal part, when the LPMO is isolated from the cytoplasm or a mature enzyme isolated from the periplasmic fraction by use of the native secretion signal (Chapter 4).
3. **Evaluation of DSF for measuring the thermodynamic stability of an LPMO** for both *TrCel61A* (Chapter 3) and *ScLPMO10C* (Chapter 4). The method is reproducible, accurate and sensitive to measure differences as small as 1 °C.
4. **Characterization of *TrCel61A*** as a type-3 LPMO with suggested additional activity on short oligos (DP 5 and higher) and an apparent melting temperature of 62 ± 1 °C (Chapter 3).
5. **Characterization of *ScLPMO10C*** as a type-1 LPMO (confirmation of earlier literature) with an apparent melting temperature of 51 ± 1 °C (Chapter 4).
6. **Expression of an active *TrCel61A*, lacking CBM**, when glycosylated linker is still attached to the protein. A slight decrease in stability was observed for the resulting enzyme (Chapter 5).
7. **Successful introduction of additional disulfide bridges and N-glycosylation positions in *TrCel61A***, while still obtaining active enzyme variants with a comparable thermodynamic stability. This mutagenesis strategy was inspired by the stabilizing effect of native glycans and disulfide bridges (Chapter 5).
8. **Construction of 2 *de novo* enzymes, based on a consensus strategy of family AA9 (LPMO-3)**, where both enzymes were expressed with and without linker and CBM (Chapter 5).
9. **Disulfide bridges were found to be important in LPMOs**: Both wild-type enzymes, *i.e.* *ScLPMO10C* and *TrCel61A*, show an unusual activity after 80 °C treatment due to disulfide bridges. Introduction of extra disulfide bridges yields a highly variable effect on stability of the protein (Chapter 6)
10. **A +12 °C increase in thermodynamic stability was obtained by disulfide engineering**. The variant also exhibited and improved residual activity (\sim x2 compared to wild-type) after a 2 hour treatment at 80 °C (Chapter 6).

7.2 Discussion

7.2.1 Key to successful expression: control N-terminal processing

Despite the fact that the exact amino-terminal processing does not matter for most proteins, it is known that LPMOs have the unusual requirement of strict processing to obtain a precise amino-terminus. Indeed, an N-terminal histidine side chain is involved in copper coordination together with the actual NH₂-terminus and a second distant histidine side chain, forming all together the so-called "histidine brace"¹¹. Focusing on this special characteristic has led to successful heterologous expression in this PhD thesis in both *P. pastoris* and *E. coli* as expression host. In Chapter 2, different secretion signals were compared in that respect, with very promising results: higher yields than earlier described in *P. pastoris* for LPMOs and certainty of a 100 % correctly N-processed enzyme population by use of the protein's native secretion signal. Moreover, the comparison also showed that the most commonly used α -mating factor from *S. cerevisiae* is not always the best option. Many alternatives can be evaluated, whereas the native secretion signal in particular should be taken into account. Furthermore, for expressing ScLPMO10C in *E. coli*, the sequence was preceded by its native secretion signal as well. This resulted in successful expression of an active LPMO. In addition, it was found that an N-terminal extension to ScLPMO10C resulted not only in complete activity loss but also in a significant decrease in thermodynamic stability (-7 °C). This clearly indicates the enzyme fold has been disturbed and a new weak point for denaturation initiation has been introduced (Chapter 4). Complete removal or replacement of the copper ion by other divalent cations was earlier shown in literature to have an even larger effect, with decreases up to -20 °C in thermodynamic stability²⁴⁶. Although the N-terminal histidine residue is closely evaluated, the N- ϵ -methylation that is present in all fungal LPMOs has not been taken into account²⁶⁵. In case of TrCel61A, the native host does perform this modification, while *P. pastoris* is unable to do this. The function of the alkyl substituent is unknown so far. A recent study with TrCel61A, produced in the native host, suggest a difference in regiospecificity (relatively more C4-oxidation)²⁶⁶.

Even though the heterologous expression systems of the two LPMOs seem very different at first sight, especially because the eukaryotic and prokaryotic character of the hosts, they comprise a highly parallel strategy (Table 7.1). In both cases, the use of a secretion signal was the preferred option to secure an authentic N-terminus combined with a relatively simple purification method from the medium fraction or periplasmic fraction, respectively. Using an N-terminal cleavable parts was in both hosts established, although concluded to be more labor-intensive. First, in *P. pastoris*, isolation of a secreted protein from the medium fraction was evaluated because this secretory pathway also benefits from formation of disulfide bridges and N-glycosylation, which both were shown to contribute to stability (Chapter 5). Moreover, following this strategy also merits the well-known advan-

tage of low endogenous protein content and low protease activity which results in simple protein purification²⁶¹. Indeed, a simple ultrafiltration was in most cases sufficient. Second, in *E. coli*, periplasmic expression was evaluated because secretion in the medium fraction is not widely used since it was found to be rather complicated⁴⁶³. Periplasmic expression in *E. coli* has comparable advantages to the secreted fraction from *P. pastoris* such as the ability to form disulfide bridges and low protease content (compared to cytoplasm)³⁰⁵. Nonetheless, techniques for cytoplasmic expression are much better developed, whereas periplasmic expression is mostly used if problems occur for the cytoplasm such as formation of inclusion bodies, wrong protein formation and protein toxicity³⁰⁵.

Next to these similarities mentioned above, a clear difference between the two systems can be found in the way expression is switched on: in *P. pastoris* an inducible expression was used²⁶¹, while a constitutive expression for *E. coli*³¹⁷. The huge advantage of an inducible system is the strict regulation, while the advantage of constitutive expression is the ease and lower labor-input, especially when thinking about large mutant libraries. One inducible promoter (Glyceraldehyde-3-phosphate dehydrogenase (GAP)) was also evaluated for *P. pastoris*, but did not yield good expression of the protein (Chapter 2). No further effort was put into varying the promoter strength since large libraries would anyhow not be convenient in an integrative expression system and in addition, as long as no fast and convenient LPMO assay exists, library screening is out of reach (see next topic).

<i>TrCel61A</i> in <i>P. pastoris</i>	<i>ScLPMO10C</i> in <i>E. coli</i>
1. Native secretion signal	1. Native secretion signal
<ul style="list-style-type: none"> • Authentic N-terminus ensured • More convenient than cleavable part (enterokinase cleavage) • Performs better than other secretion signals in terms of yield and N-terminal processing 	<ul style="list-style-type: none"> • Authentic N-terminus ensured • More convenient than cleavable part (factor Xa)
2. Isolation from medium fraction	2. Isolation from periplasmic fraction
<ul style="list-style-type: none"> • Disulfide bridge formation • Glycosylation • Low protease activity • Low endogenous protein secretion • Easy down-stream processing (Ultra-Filtration) 	<ul style="list-style-type: none"> • Disulfide bridge formation • Low protease activity • Low endogeneous protein load in periplasm • Easy down-stream processing (e.g. osmotic shock)
3. Inducible expression (pAOX1)	3. Constitutive expression (pCX34)
<ul style="list-style-type: none"> • Strict regulation possible • More labor-intensive 	<ul style="list-style-type: none"> • No regulation possible • Less labor intensive, preferred for library screenings

Table 7.1: Comparison of two different expression systems for the two LPMO families.

Just like other cellulases, LPMOs are always secreted enzymes in nature. This is easily explained by its function to degrade an extracellular substrate, so that the resulting simple sugars are transported in the cell as nutrients. An additional benefit of using this strategy is that active LPMOs will not end up in the cytosol. Assuming they would, in absence of

cellulose substrate, start generating reactive oxygen species (ROS) and provide stress in the cell, thereby damaging DNA, RNA, proteins and lipids.

7.2.2 LPMO (stability) engineering: Challenge accepted!

Regardless of what characteristic of a protein one wants to engineer and how the genetic diversity is created, a crucial prerequisite for finding a variant with the desired characteristic is a good screening or selection method. This is the main reason why so little studies published today involve engineering of LPMOs, despite the high interest in these proteins.

Analysis of LPMO activity is very challenging on quantitative and even qualitative level¹⁶⁶. The lack of a quick and convenient biochemical method has several reasons. First of all, the actual nature of the LPMO reaction results in generation of products with different degree of polymerization combined with oxidation at different positions (C1- and/or C4-oxidation have been described). Furthermore, the shortage of commercially available standards, use of insoluble substrates, auto-oxidation of the electron donor by dissolved Cu(II) and spontaneous decomposition of the C4-oxidized products in aqueous solutions and at high pH (~ HPAEC) further hampers the development process of a suitable method²⁸⁰. In addition, the resulting product of one LPMO intervention can form the substrate for its next action.

The most powerful and thus most popular method to date is the use of high performance anion-exchange chromatography (HPAEC, a specialized HPLC), also used for activity measurement in this PhD thesis. The analysis method separates neutral from C1- and C4- oxidized cellulose products and also internally on degree of polymerization. However, the method is still lengthy and labor-intensive (typically >30 minutes per HPAEC-run is required). Moreover, the LPMO does much more than one can observe since only soluble products are analyzed and the interpretation is complicated by subsequent use of reaction products (as for example shown in Chapter 3 for *TrCel61A*). In addition, for type-3 LPMOs, the double oxidized products are very difficult to separate, let alone identify. Instead of monitoring all products of the HPAEC-trace, often the reaction rate of one single product is followed as representative for the reaction. One should choose a short oligo to make sure it is an actual end-product of the reaction. Furthermore, it should be preferentially one of the predominant products and be nicely separated in the chromatogram. In that respect, cellobionic acid was selected in this PhD thesis. While we used peak area of the resulting product for quantification, peak height can also be used as shown in a recent publication, in which the authors monitored C4-oxidized cellobiose (Glc4gemGlc)¹⁸³. However, the pitfall is that you screen for production of that specific product only and not the total amount of oxidation reactions performed. In other words, if the LPMO variant would make less of the monitored product, but 100 times more of another product, one will never find it.

Some alternative approaches that endeavor more high-throughput screening have been proposed. These include the use of chromogenic substrates²⁸¹, the monitoring of ascorbate (electron donor) consumption²⁸² and an unproductive side reaction of the LPMO: reduction of O_2 to H_2O_2 ²⁴⁵. However, these assays have all drawbacks resulting in a low practical application. For example the chromogenic substrates are not designed for enzyme kinetics because enzyme accessibility can be hindered in the hydrogels by the chlorotriazine groups^{281;464}. The use of ascorbate on the other hand is being diminished because the biological relevance is questioned as well as its stability during long incubation periods¹⁵². Only the last assay (reduction of O_2 to H_2O_2) is highly promising and has been applied in various studies^{124;125;128;133}. Nonetheless, the assay is often used to measure substrate preference rather than activity, hence the side reaction only occurs in absence of a suitable substrate. In addition, the side reaction can have different optima (such as pH optimum as described by Yu et al²⁸²) than the actual polysaccharide oxidation reaction. Design of a high-throughput method for LPMOs is highly cumbersome, considering the large variety in substrates and products and additionally in electron donor specificity¹⁴⁹, while a specific assay is required. Presumably focusing on the common factor, the oxidation, could offer a solution. Such an approach would also present a universal assay for cellulose as well as hemicellulose substrates. Often high-throughput methods are colorimetric. In that respect one can think on the use of redox indicators.

Now, in this PhD research in particular, the aim was to increase the stability of LPMOs. Enzymes are prone to denaturation by different factors such as temperature, but also pH and various chemicals²²⁵. It is often assumed that they are all related in a way so that enzymes with a higher thermal stability for example would also offer more resistance to denaturing chemicals. However, thermal stability is the best studied form and main subject in this thesis. Two fundamental types can be distinguished, namely thermodynamic and kinetic stability²³¹. The first refers to the protein's tendency to reversibly unfold, while the last refers to the time it takes for a protein to irreversibly unfold and thus become inactive. Although both terms clearly refer to different processes, they are often related. Indeed, irreversible loss of activity is initiated by partial unfolding, but in contrary to the thermodynamic equilibrium, the kinetic process is followed by permanent change such as aggregation or proteolytic cleavage. Measurement of kinetic stability usually relies on a solid activity test to determine parameters like half-life of denaturation ($t_{1/2}$), optimum operating temperature (T_{opt}) and temperature at which only half of the enzyme activity is retained after a specified amount of time (T_{50}). Because no fast and convenient activity test exists, measuring those parameters for kinetic activity would be a cumbersome task. Therefore, we focused on thermodynamic stability, which can be measured without the catalytic reaction. Parameters reflecting this type of stability are unfolding equilibrium constant (K_u), melting temperature (T_m) and free energy of unfolding (ΔG_u). These mostly require a pure protein, so that high-

throughput possibilities are limited. In this PhD thesis, apparent melting temperature was measured by use of differential scanning fluorimetry (DSF). This is a good initial screen to find promising enzyme variants and the result contains very interesting information from scientific point of view. However, to find the real merit of a variant in industrial applications, *i.e.* its actual lifetime in a process, kinetic stability parameters are more interesting. The promising variants should therefore be further characterized on kinetic stability. To that end, the positive hits from disulfide engineering in Chapter 6, were additionally evaluated concerning their residual activity after 80 °C treatment before deciding on the *best* variant. Although this is a limited kinetic stability test, it already indicated important differences between the variants. In the light of more practical applications, it would be very useful to add more experiments such as incubation at industrial relevant incubation temperatures and times to validate the LPMO variants. For example, one could incubate the LPMO and its variants for a few days at different temperatures (for example 50, 60 and 65 °C) and measure the residual activity afterwards. When incubating the enzymes together with substrate, a suitable electron donor and a beta-glucosidase, the glucose content can be measured (by HPLC for example) to find the best performing enzyme variant. Similar experiments are performed in companies like Novozymes (See patent application 'Variants of polypeptides having cellulolytic enhancing activity and polynucleotides encoding same', number US 20130219568 A1) to find improved enzyme variants (in this case even for higher stability). Instead of adding a beta-glucosidase, a very small amount of an industrial cellulase mixture (such as Accelerase Trio (Dupont) or Novozym 188 (Novozymes)) can be added and one could also monitor the gluconic acid for example to quantify the activity performed by the LPMO.

In conclusion, activity testing of LPMOs is still in its infancy and measuring thermodynamic stability requires a rather pure protein. This means that protein engineering endeavors are limited to rational methods, resulting in only small amounts of enzyme variants. This poses a second major challenge in engineering LPMOs. Indeed, these enzymes are still very recently discovered and very little structure-function information is available. However, this makes the work also extremely interesting and of high-value for the scientific community.

7.2.3 Stability engineering: is there a perfect route?

Stability remains a very important enzyme parameter and often determinant for economic success²²⁵. First, industrial applications require stable enzymes because higher temperatures offer advantages such as reduced risk on contamination, decreased viscosity and increased reaction rate. Second, the most stable variants are often taken as template for mutagenesis¹⁹⁹. As a matter of fact, introducing mutations that improve function are often compromised on stability and/or lead to expression loss. Third, stability engineering is

highly interesting from a scientific point of view: the molecular basis of the protein folding problem, which is related to many diseases, is still unresolved. One valid option is always to look for a thermostable variant in nature, although these are not always available. For LPMOs, 4 years ago, nothing was known on stability so that industrial (*T. reesei*) and/or scientifically (*S. coelicolor*) relevant representatives were chosen to study in this thesis. In addition, the availability of a crystal structure for ScLPMO10C was a convincing factor.

Despite the numerous studies on protein stability, the effect of a mutation is still hard to predict. It is even considered one of the most difficult problems, unresolved in protein science³²³. In addition, no generic engineering method exist to increase stability, since mutations are always context dependent²³⁴ and one single mutation often results in a very humble improvement, if even detectable. It is usually only the concerted action of multiple mutations that leads to a significant stability improvement³²⁴.

It should be clear by now that there is no such thing as the shortest road to success in stability engineering rather than trial and (unfortunately) error. In this PhD thesis, rational engineering methods were applied including disulfide engineering for both *TrCel61A* and ScLPMO10C, introduction of N-glycosylation positions and consensus engineering for *TrCel61A* and rational inspection of the protein surface for ScLPMO10C. The highest success rates could most probably be expected from disulfide engineering and rational surface inspection, while rather big differences can result from the first and only subtle differences from the last. Consensus engineering and N-glycosylation are both somewhat *long-shot* strategies because they have only known limited preceding examples and no clear rules exist to improve the chances on success. For example consensus engineering is known to strongly depend on the input sequences²¹⁹, although this can be rationalized in many ways. For *de novo* full consensus genes, the actual expression of a fully functional LPMO is already an appreciable achievement and the difference in stability found is highly interesting. Since the apparent melting temperature of *TrCel61A* is already reasonable high, a more advanced technique would be to only withdraw the LPMO sequences from thermophilic organisms for example and start building the consensus therefrom. In our experiments, the mutations were presumably neutralized by combining sequences with higher as well as lower thermostability.

The engineering approach with the highest outcome in this work is disulfide engineering. Disulfide bridges are known to have the ability of lowering the conformational entropy as compared to the unfolded state. Although this results in an increased stability, the success ratio of disulfide introduction is limited. This is also illustrated in this research: 0/7 stabilizing introductions for *TrCel61A* (Chapter 5) and 4/16 for ScLPMO10C (Chapter 6). The underlying cause is that we only have limited knowledge on where to add a disulfide

bridge in the protein. Some generic rules are set such as the understanding that larger loop lengths have higher chances on success. And also, computational tools evaluate ideal geometric requirements to find suitable positions for cysteine pairs. Nonetheless, these methods are not aiming at the core of the problem. As a matter of fact, the information one actually needs is: how does the protein unfold and what parts initiate the process? Indeed, fixating regions that are involved in early unfolding events can only stabilize the protein⁴⁶⁵. So, on a more general note, one should first study the unfolding through for example a molecular dynamics (MD) simulation and subsequently determine how to stabilize the unfolding initiators. Despite the promising character of such an approach, there are some impediments. First, this strategy requires coordinates of a 3D structure of the protein and second the approach results in complex calculations. Indeed, one not only needs special hardware (computer clusters) to perform the calculations on, but also an advanced knowledge to choose the most suitable forcefield and its parameters. Furthermore, the calculation always leads to an outcome, although it is difficult to estimate the reliability thereof. It is clear that these techniques hold an enormous potential to learn a lot about the dynamics of the protein, and to save time on extensive lab work. Notwithstanding, a further development and user-friendliness are inevitable to open the door of this fascinating area to a more wide research community. As a proxy for this techniques, B-factor analysis is a very useful tool to find target regions. Indeed, regions with high B-factor represent highly flexible regions, that are likely to unfold early in the denaturation process. In addition, more user-friendly *in silico* calculation tools are gradually becoming available such as for example the FoldX web server. Empirical terms are used so that also non-specialist can easily use the tool to predict stabilizing mutations of a protein, by calculation of the change in free energy between the wild-type enzyme and its variants⁴⁶⁶.

7.2.4 There are the LPMOs! Problem solved?

There is no doubt about the potential that LPMOs hold in biomass industry. However there is one question remaining: Will it be enough? From its natural function as constituent of the plant cell wall, lignocellulosic biomass is meant to be onerous in degradation. Therefore, large enzyme loadings are required, what increases the cost significantly. LPMOs facilitate the process, but how helpful can they really be? At the same time, (bio)fuels are low-value products that will certainly not be economical viable without their integration in a complete biorefinery. These questions are highly interesting, still not possible to fully answer yet.

First, LPMOs are still in *early development phase*, with the discovery of their enzymatic function only six years ago¹⁴. The one paper on LPMOs is followed by the other, while our view is continuously broadened. For example, the story started with activity on chitin, while we know now that LPMOs are not only active on cellulose but also hemicellulose^{125–127},

soluble oligosaccharides¹²⁴ and even starch¹⁷. In addition, the range of possible electron donors is continuously expanding. Lignin, first thought to be a disturbing factor, is now found to serve as electron donor^{118;183–185} and moreover, would be much more interesting than ascorbic acid in a biomass degradation process¹⁶¹. Although, the knowledge on LPMOs is developing in a rapid pace, the full potential of these enzymes is clearly not revealed yet.

Second, the entire process needs a serious make-over to fulfill the needs of these oxidative enzymes. Indeed, in contrary to glycoside hydrolases, LPMOs require oxygen for their reaction^{14;147}. This is more complex than one would assume at first sight. Indeed, simultaneous saccharification and fermentation (SSF) was long thought to be the most efficient process, although when aerobic organisms are used, they go in competition for oxygen with the LPMOs¹⁸³. Alternatively, when anaerobic fermentation is applied, the addition of oxygen alters the metabolism, resulting in undesired side-products⁴⁶⁷. In this respect, separate fermentation and hydrolysis (SFH) seems again more convenient. Furthermore, LPMOs produce different products than glycoside hydrolases: namely oxidized oligosaccharides. Even though, these oxidized forms are much less prevalent than glucose in the end of the process, they could significantly disturb the process. Indeed, they inhibit classical cellulases, *e.g.* gluconic acid inhibits β -glucosidases more than glucose does and these β -glucosidases have a much higher preference (10x) for cellobiose than cellobionic acid as substrate¹⁸⁴. In addition, the classical fermentation organisms cannot metabolize these oxidized products (unless the required pathways are introduced). From an industrial point of view, this is very important because it means that part of the glucose is lost and cannot be fermented into ethanol. These concerns are gradually arising, as some recent articles demonstrate (also described in literature study "Rethinking the process for LPMO integration").

On a more general note, considering the current knowledge, LPMOs are an inevitable addition to cellulase cocktails. Regarding the large variety in LPMOs, some seem more appropriate than others. An ideal process would probably contain multiple LPMOs because LPMOs themselves can also work in synergy. One hypothesis therefore is that different LPMOs can attack the cellulose fibers from different angles. In addition to cellulose-active LPMOs, also hemicellulose active LPMOs could be added to biomass. Specificity for short oligosaccharides is not a main requirement since the other cellulases can degrade them as well. Lignocellulose material is highly variable and also different pretreatment methods result in variable carbohydrate composition of the substrate¹²². Different behaviors of AA9 enzymes have been observed in these various situations¹²⁹. Therefore it is becoming more generally believed that the enzyme composition has to be adjusted to the biomass type¹³⁰. Because high dry matter content is applied in this context, a CBM domain seems not to be required. However, it would not harm the process. Furthermore, lignin can serve as

electron donor¹²¹, while cellobiose dehydrogenase should be omitted because the enzyme also produces undesired, non-fermentable, oxidized oligosaccharides. As alternative, the most recently found light-driven electron transfer (via photosynthetic machinery)¹⁵⁵ would be highly interesting because it would be possible to only activate the LPMOs when needed by a simple light pulse. In nature, not all cellulose niches have access to light such as for example cellulases in the bark of a tree. Therefore, there must exist LPMOs, with preference for a different electron donors. And finally, of course, a stable enzyme is required which performs well at 50-60 °C in a SFH set-up.

7.3 Outlook

In this PhD thesis, many enzyme variants were created and screened on the characteristic of thermodynamic stability. As the first law of directed evolution states 'You get what you screen for', we might have missed other, yet interesting changes in for example enzyme specificity. Even though this was not the initial aim of the mutagenesis strategies, it would be worth characterizing them in more depth. Especially the mutants created during consensus engineering (Chapter 5) are considered interesting. Indeed, type-3 LPMOs are all able to oxidize at both the C1- and C4-carbon of the β -1,4 linkage. However, there is a range in the ratio of the different oxidized products that can vary. Other than that, there is a wide diversity in possible substrates going from PASC (used here) to more crystalline forms of cellulose over all types of hemicelluloses to soluble oligosaccharides. These substrate specificity differences are moreover not taken into account in the 3 distinct LPMO types. Furthermore, the variant of *TrCel61A* lacking a CBM would also be a high-valued result since the exact function of a CBM is still not completely unraveled (see Chapter 5). A supplementary comparison with the other LPMO from *T. reesei*, i.e. *TrCel61B*, which naturally does not carry a CBM would be interesting to find out the rather humble diversity of this fungus in LPMO production. Other variants (disulfide bridges and N-glycosylation) are thought to be less interesting in this respect because the copper-coordination center and substrate binding plane were both left untouched in the design.

The outcome of the rational engineering techniques, applied in this PhD thesis, was rather limited, reflecting the fact that science does not understand protein stability well. During the last years, advances in the direction of high-throughput methods are gradually becoming available so that one can start thinking in the direction of more random or semi-rational engineering approaches. First, effective library creation has become possible: the efficiency of *P. pastoris* transformation techniques has been increased last years and also in *E. coli*, periplasmic expression systems with higher yields as described here are developed⁴⁶⁸. Second, more high-throughput systems for screening on microtiterplate-scale are being studied. One example is the fluorescent labeling of insoluble C1-oxidized products, with the advantage of not underestimating the LPMO performance compared to other techniques that only measure soluble products (e.g. HPAEC)⁴⁶⁴. Alternatively, MS-analysis of the product-mixture can also be done in high-throughput format (e.g. 384-target plate MALDI at the lab of Prof. B. Devreese). Besides, thermostability measurement via DSF can be performed in 96-well format. These techniques open the door to more engineering techniques. One example is iterative saturation mutagenesis (ISM)²⁰⁹, where amino acids can be randomized at the most flexible position(s) (e.g. via B-FIT analyses in Chapters 5 and 6) and in iterative cycles. Such a strategy can be narrowed by only using 'allowed' residues, for example by looking to a multiple sequence alignment of LPMOs from ther-

mophile organisms. Also, the packaging of the protein can be improved by filling cavities in the protein. In a similar way, hotspots can be determined in a rational way and later fully saturated by the different amino acids or only by larger ones. Furthermore, all random techniques can be applied going from DNA-shuffling to error-prone PCR and use of mutator strains in order to create variability in the wild type sequence.

One of the biggest questions in LPMO research today is still: Why is there such a large diversity? Sometimes, even more than 30 LPMO variants are present in one single organism. Different reasons have been proposed such as variety in substrate specificity, regioselectivity and electron donor preference. This would permit LPMO action in different stages of biomass degradation and various environmental circumstances. From the point of view of the classical cellulases, CBH I works from the reducing end and CBH II from the non-reducing end. Maybe, it is not a coincidence that LPMOs can also introduce an oxygen at the reducing and/or non-reducing end? There might be an activating mechanism and subsequent synergy that we don't understand yet and that clearly needs more investigation.

In order to find flexible regions in a protein, molecular dynamics (MD) simulations can be a valuable addition to B-FIT analysis for proteins whereof a crystal structure is available. Furthermore, the finding that some LPMOs are active on short soluble oligos makes the outcome of molecular docking studies with these substrates more valuable. Indeed, docking studies with short oligos are more easily attainable than for cellulose substrate. Cellulose is an extended substrate with many possibilities in rotability. Therefore, it demands extremely difficult calculations. Moreover, this only includes cellulose and not even hemicellulose or lignin. Even more trustworthy than a simulation is of course elucidation of the first LPMO crystal structure, bound to a short soluble glucose chain. In the near future, such a composition can definitely be expected, while being of major importance for the LPMO community.

The highest goal of numerous studies involving LPMOs is to propose an amelioration to the existing cellulase degradation process. The route can be either through a milder pre-treatment or an improved cellulase mixture (faster saccharification with lower enzyme loading). These ultimately aim of course in lowering costs for making lignocellulosic biomass a suitable renewable feedstock. In this respect, to know the true benefit of a single optimized parameter in the degradation process (e.g. stabilized LPMOs), its contribution to the entire synergistic process has to be evaluated. Many calculations can offer interesting insights, although they always have to be experimentally validated to find the true benefit. Moreover, making lignocellulosic biomass a cost-efficient feedstock is a challenge that puzzles scientists all over the world. Finding the ultimate process in cellulose degradation demands really a multi-disciplinary endeavor, tackling the problem from different angles (process-optimization, enzyme engineering, metabolic engineering) and in distinct bottlenecks of the

entire process. An inspiring, completely different and additive approach to this work is for example the design of genetically modified energy plants with lower lignin content, which require lower pretreatment⁴⁶⁹. Another recent, yet interesting suggestion is that the use of seawater in biorefineries, by applying halotolerant LPMOs (and cellulolytic system)²⁴⁷, could offer even more environmental advantages. The solution has to be found in bringing such ideas all together.

It is clear that the last paper on LPMO research is far from near. Thanks to their applicability in biomass industry, the research on LPMOs has grown very quickly. Their underlying natural function in degradation of organic matter, is not yet fully understood. The newest advances describe 100x activity improvement by the use of photosynthetic pigments (excited via sunlight) as electron donor¹⁵⁵. This literally sheds, again, new light on the LPMO reaction. Nonetheless, LPMOs hold much potential in other research areas too. This is emphasized by the observation of different functions last years such as a role in virulence¹⁸, keratinolysis¹⁹ and plant infection²⁰. Furthermore, considering the large variety on LPMOs, for each application, a suitable LPMO or LPMO mixture can be found or designed. That is really a task of matching all parts. For example a study demonstrated that mixing a catalytic domain of one LPMO and CBM form another LPMO could alter and/or increase its activity³⁷³. Many questions still have to be resolved such as: what are the so-called 'second sphere' residues, important for activity? Is there an equivalent for CDH in family AA10? What is the rate determining step in the reaction? Have we reached the full range of possible substrates and electron donors?

The work bundled in this PhD thesis is certainly only the beginning...

References

- [1] V. Menon and M. Rao, "Trends in bioconversion of lignocellulose: Biofuels, platform chemicals and biorefinery concept," *Progress in Energy and Combustion Science*, vol. 38, no. 4, pp. 522–550, 2012.
- [2] V. B. Agbor, N. Cicek, R. Sparling, A. Berlin, and D. B. Levin, "Biomass pretreatment: fundamentals toward application.," *Biotechnology advances*, vol. 29, no. 6, pp. 675–85, 2011.
- [3] J. Goldemberg and J. Goldemberg, "Ethanol for a sustainable energy future," *Science*, vol. 315, no. 5813, pp. 808–810, 2007.
- [4] K. A. Gray, L. Zhao, and M. Emptage, "Bioethanol," *Current Opinion in Chemical Biology*, vol. 10, no. 2, pp. 141–146, 2006.
- [5] T. Searchinger, R. Heimlich, R. A. Houghton, F. Dong, A. Elobeid, J. Fabiosa, S. Tokgoz, D. Hayes, and T.-h. Yu, "Use of U.S. croplands for biofuels increase greenhouse gases through emissions from land-use change," *Science*, vol. 319, no. February, pp. 1238–1241, 2008.
- [6] E. M. Rubin, "Genomics of cellulosic biofuels.," *Nature*, vol. 454, no. 7206, pp. 841–5, 2008.
- [7] M. Dusselier, P. Van Wouwe, A. Dewaele, E. Makshina, and B. F. Sels, "Lactic acid as a platform chemical in the biobased economy: the role of chemocatalysis," *Energy & Environmental Science*, vol. 6, no. 5, p. 1415, 2013.
- [8] L. R. Lynd, P. J. Weimer, W. H. van Zyl, and I. S. Pretorius, "Microbial Cellulose Utilization: Fundamentals and Biotechnology," *Microbiology and Molecular Biology Reviews*, vol. 66, no. 3, pp. 506–577, 2002.
- [9] M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust, "Biomass recalcitrance: engineering plants and enzymes for biofuels production.," *Science*, vol. 315, no. 5813, pp. 804–7, 2007.
- [10] E. T. Reese, R. G. H. Siu, and H. S. Levinson, "The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis," *Journal of Bacteriology*, vol. 59, pp. 485–497, 1950.
- [11] R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J.-C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, and P. H. Walton, "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 37, pp. 15079–84, 2011.
- [12] J. a. Langston, T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko, and M. D. Sweeney, "Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61.," *Applied and environmental microbiology*, vol. 77, pp. 7007–15, oct 2011.
- [13] X. Li, W. T. B. Iv, C. M. Phillips, M. A. Marletta, W. T. Beeson, and J. H. D. Cate, "Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases.," *Structure*, vol. 20, no. 6, pp. 1051–1061, 2012.
- [14] G. Vaaje-Kolstad, B. Westereng, S. J. Horn, Z. Liu, H. Zhai, M. Sørlie, and V. G. H. Eijsink, "An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides.," *Science*, vol. 330, no. 6001, pp. 219–222, 2010.
- [15] P. V. Harris, D. Welner, K. C. McFarland, E. Re, J.-C. Navarro Poulsen, K. Brown, R. Salbo, H. Ding, E. Vlasenko, S. Merino, F. Xu, J. Cherry, S. Larsen, and L. Lo Leggio, "Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family.," *Biochemistry*, vol. 49, no. 15, pp. 3305–16, 2010.

REFERENCES

- [16] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, and B. Henrissat, "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes," *Biotechnology for biofuels*, vol. 6, p. 41, 2013.
- [17] V. V. Vu, W. T. Beeson, E. a. Span, E. R. Farquhar, and M. a. Marletta, "A family of starch-active polysaccharide monooxygenases.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 13822–7, sep 2014.
- [18] J. Ebeling, H. Knispel, G. Hertlein, A. Fünfhaus, and E. Genersch, "Biology of *Paenibacillus* larvae, a deadly pathogen of honey bee larvae," *Applied Microbiology and Biotechnology*, vol. 100, no. 17, pp. 7387–7395, 2016.
- [19] P. K. Busk and L. Lange, "Classification of fungal and bacterial lytic polysaccharide monooxygenases," *BMC Genomics*, vol. 16, no. 1, pp. 1–13, 2015.
- [20] K. S. Johansen, "Lytic polysaccharide monooxygenases: The microbial power tool for lignocellulose degradation," *Trends in Plant Science*, vol. 21, no. 11, pp. 926–936, 2016.
- [21] D. B. Wilson, "Cellulases and biofuels," *Current opinion in biotechnology*, vol. 20, no. 3, pp. 295–9, 2009.
- [22] K. Egorova and G. Antranikian, "Industrial relevance of thermophilic Archaea," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 649–655, 2005.
- [23] P. Turner, G. Mamo, and E. N. Karlsson, "Potential and utilization of thermophiles and thermostable enzymes in biorefining.," *Microbial cell factories*, vol. 6, p. 9, 2007.
- [24] G. Paës and M. J. O'Donohue, "Engineering increased thermostability in the thermostable GH-11 xylanase from *Thermobacillus xylanilyticus*," *Journal of Biotechnology*, vol. 125, no. 3, pp. 338–350, 2006.
- [25] S. E. Lantz, F. Goedegebuur, R. Hommes, T. Kaper, B. R. Kelemen, C. Mitchinson, L. Wallace, J. Ståhlberg, and E. A. Larenas, "Hypocrea jecorina CEL6A protein engineering.," *Biotechnology for biofuels*, vol. 3, p. 20, 2010.
- [26] Z. Forsberg, G. Vaaje-Kolstad, B. Westereng, A. C. Bunæs, Y. Stenstrøm, A. MacKenzie, M. Sørli, S. J. Horn, and V. G. H. Eijsink, "Cleavage of cellulose by a CBM33 protein," *Protein science*, vol. 20, no. 9, pp. 1479–83, 2011.
- [27] International Energy Agency, "Key world energy statistics 2016," tech. rep., 2016.
- [28] L. Beurskens and M. Hekkenberg, "Renewable energy projections as published in the National Renewable Energy Action Plans of the European member states covering all 27 EU member states," tech. rep., 2011.
- [29] N. S. Bentsen and C. Felby, "Biomass for energy in the European Union - a review of bioenergy resource assessments.," *Biotechnology for biofuels*, vol. 5, no. 1, p. 25, 2012.
- [30] U.S. Energy Information Administration, "International Energy Outlook 2016," Tech. Rep. May 2016, 2016.
- [31] D. B. Richardson, "Electric vehicles and the electric grid: A review of modeling approaches, Impacts, and renewable energy integration," *Renewable and Sustainable Energy Reviews*, vol. 19, pp. 247–254, 2013.
- [32] M. J. Groom, E. M. Gray, and P. A. Townsend, "Biofuels and biodiversity: Principles for creating better policies for biofuel production," *Conservation Biology*, vol. 22, no. 3, pp. 602–609, 2008.
- [33] A. O'sullivan, "Cellulose: the structure slowly unravels," *Cellulose*, vol. 4, no. 3, pp. 173–207, 1997.
- [34] L. Caspeta and J. Nielsen, "Economic and environmental impacts of microbial biodiesel.," *Nature biotechnology*, vol. 31, no. 9, pp. 789–93, 2013.
- [35] S. T. Merino and J. Cherry, "Progress and challenges in enzyme development for biomass utilization," *Advances in Biochemical Engineering/Biotechnology*, vol. 108, no. June, pp. 95–120, 2007.
- [36] N. Scarlat, J. F. Dallemand, F. Monforti-Ferrario, and V. Nita, "The role of biomass and bioenergy in a future bioeconomy: Policies and facts," *Environmental Development*, vol. 15, pp. 3–34, 2015.
- [37] S. Kim and B. E. Dale, "Life cycle assessment of various cropping systems utilized for producing biofuels: Bioethanol and biodiesel," *Biomass and Bioenergy*, vol. 29, no. 6, pp. 426–439, 2005.

- [38] F. Cherubini and A. H. Strømman, "Life cycle assessment of bioenergy systems: State of the art and future challenges," *Bioresource Technology*, vol. 102, no. 2, pp. 437–451, 2011.
- [39] J. P. Maity, J. Bundschuh, C.-Y. Chen, and P. Bhattacharya, "Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives - A mini review," *Energy*, vol. 78, pp. 104–113, 2014.
- [40] F. Alam, S. Mobin, and H. Chowdhury, "Third generation biofuel from Algae," *Procedia Engineering*, vol. 105, no. 1, pp. 763–768, 2015.
- [41] J. Lü, C. Sheahan, and P. Fu, "Metabolic engineering of algae for fourth generation biofuels production," *Energy & Environmental Science*, vol. 4, pp. 2451–2466, 2011.
- [42] D. R. Georgianna and S. P. Mayfield, "Exploiting diversity and synthetic biology for the production of algal biofuels," *Nature*, vol. 488, no. 7411, pp. 329–335, 2012.
- [43] S.-H. Ho, X. Ye, T. Hasunuma, J.-S. Chang, and A. Kondo, "Perspectives on engineering strategies for improving biofuel production from microalgae - A critical review," *Biotechnology Advances*, vol. 32, no. 8, pp. 1448–1459, 2014.
- [44] Y. H. Fan, L. T.; Gharpuray, M. M.; Lee, *Cellulose hydrolysis*, vol. 3. Berlin: Springer-Verlag, 1987.
- [45] F. Cherubini, "The biorefinery concept: Using biomass instead of oil for producing energy and chemicals," *Energy Conversion and Management*, vol. 51, no. 7, pp. 1412–1421, 2010.
- [46] D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. a. Simmons, and H. W. Blanch, "The challenge of enzyme cost in the production of lignocellulosic biofuels," *Biotechnology and bioengineering*, vol. 109, no. 4, pp. 1083–1087, 2012.
- [47] R. Wolfenden, X. Lu, and G. Young, "Spontaneous hydrolysis of glycosides," *Journal of American Chemical Society*, vol. 120, pp. 6814–6815, 1998.
- [48] D. B. Wilson, "Microbial diversity of cellulose hydrolysis," *Current opinion in microbiology*, vol. 14, no. 3, pp. 259–63, 2011.
- [49] S. D. Mansfield, C. Mooney, and J. N. Saddler, "Substrate and enzyme characteristics that limit cellulose hydrolysis," *Biotechnology Progress*, vol. 15, no. 5, pp. 804–816, 1999.
- [50] G. Jungmeier, R. Van Ree, E. de Jong, H. Stichnothe, I. de Bari, H. Jorgensen, M. Wellisch, K. Torr, K. Habu, G. Garnier, and J. Spaeth, "The possible role of biorefineries in a bioeconomy- activities of IEA bioenergy task 42 "Biorefining"," in *Conference proceedings of the 21st European biomass conference*, (Copenhagen), 2013.
- [51] A. C. Hansen, Q. Zhang, and P. W. L. Lyne, "Ethanol-diesel fuel blends - A review," *Bioresource Technology*, vol. 96, no. 3, pp. 277–285, 2005.
- [52] H. Kim and B. Choi, "The effect of biodiesel and bioethanol blended diesel fuel on nanoparticles and exhaust emissions from CRDI diesel engine," *Renewable Energy*, vol. 35, no. 1, pp. 157–163, 2010.
- [53] F. Zhang, S. Rodriguez, and J. D. Keasling, "Metabolic engineering of microbial pathways for advanced biofuels production," *Current Opinion in Biotechnology*, vol. 22, no. 6, pp. 775–783, 2011.
- [54] J. van Haveren, E. L. Scott, and J. Sanders, "Bulk chemicals from biomass," *Biofuels, Bioproducts and Biorefining*, vol. 2, pp. 41–57, 2008.
- [55] J. J. Bozell and G. R. Petersen, "Technology development for the production of biobased products from biorefinery carbohydrates - the US Department of Energy's "Top 10" revisited," *Green Chemistry*, vol. 12, no. 4, p. 539, 2010.
- [56] I. Bechthold, K. Bretz, S. Kabasci, R. Kopitzky, and A. Springer, "Succinic acid: A new platform chemical for biobased polymers from renewable resources," *Chemical Engineering and Technology*, vol. 31, no. 5, pp. 647–654, 2008.

REFERENCES

- [57] A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. a. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. a. Tuskan, and C. E. Wyman, "Lignin valorization: improving lignin processing in the biorefinery," *Science*, vol. 344, no. 6185, p. 1246843, 2014.
- [58] N. J. Walton, M. J. Mayer, and A. Narbad, "Vanillin," *Phytochemistry*, vol. 63, no. 5, pp. 505–515, 2003.
- [59] R. Kumar, S. Singh, and O. V. Singh, "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives.," *Journal of industrial microbiology & biotechnology*, vol. 35, no. 5, pp. 377–91, 2008.
- [60] E. D. Jong and G. Jungmeier, *Biorefinery concepts in comparison to petrochemical refineries*. 2015.
- [61] N. J. Turner, "Directed evolution drives the next generation of biocatalysts.," *Nature chemical biology*, vol. 5, no. 8, pp. 567–73, 2009.
- [62] S. Naik, V. V. Goud, P. K. Rout, and A. K. Dalai, "Production of first and second generation biofuels: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 14, pp. 578–597, feb 2010.
- [63] D. Mu, T. Seager, P. S. Rao, and F. Zhao, "Comparative life cycle assessment of lignocellulosic ethanol production: Biochemical versus thermochemical conversion," *Environmental Management*, vol. 46, no. 4, pp. 565–578, 2010.
- [64] Y. Kalogo, S. Habibi, H. L. Maclean, and S. V. Joshi, "Environmental implications of municipal solid waste-derived ethanol policy analysis," vol. 41, no. December 2006, pp. 35–41, 2007.
- [65] S. Spatari, Y. Zhang, and H. L. Maclean, "Life cycle assessment of switchgrass- and corn automobiles," *Environmental Science & Technology*, vol. 39, no. 24, pp. 9750–9758, 2005.
- [66] J. B. Binder and R. T. Raines, "Fermentable sugars by chemical hydrolysis of biomass.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 10, pp. 4516–21, 2010.
- [67] E. C. Webb, *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. San Diego, California: Academic Press, 1992.
- [68] B. Henrissat, M. Vegetales, and F. Grenoble, "A classification of glycosyl hydrolases based sequence similarities amino acid," *Biochemical Journal*, vol. 280, pp. 309–316, 1991.
- [69] B. I. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics," *Nucleic Acids Research*, vol. 37, no. SUPPL. 1, pp. 233–238, 2009.
- [70] C. Divne, J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Petterson, J. K. C. Knowles, T. T. Teeri, and T. A. Jones, "The three-dimensional crystal structure of the catalytic cor," *Science*, vol. 265, no. 5171, pp. 524–528, 1994.
- [71] H. Palonen, M. Tenkanen, and M. Linder, "Dynamic interaction of *Trichoderma reesei* cellobiohydrolases Cel6A and Cel7A and cellulose at equilibrium and during hydrolysis," *Applied and Environmental Microbiology*, vol. 65, no. 12, pp. 5229–5233, 1999.
- [72] C. Boisset, C. Pétrequin, H. Chanzy, B. Henrissat, and M. Schlein, "Optimized mixtures of recombinant *Humicola insolens* cellulases for the biodegradation of crystalline cellulose," *Biotechnology and Bioengineering*, vol. 72, no. 3, pp. 339–345, 2001.
- [73] W. L. DeLano and J. W. Lam, "PyMOL: a communications tool for computational models," *Abstracts of Papers of the American Chemical Society*, vol. 230, pp. U1371–U1372, 2005.
- [74] T. T. Teeri, A. Koivula, M. Linder, G. Wohlfahrt, C. Divne, and T. a. Jones, "Trichoderma reesei cellobiohydrolases: why so efficient on crystalline cellulose?," *Biochemical Society transactions*, vol. 26, no. 2, pp. 173–178, 1998.
- [75] G. J. Kleywegt, J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Ståhlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri, and T. a. Jones, "The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes.," *Journal of molecular biology*, vol. 272, no. 3, pp. 383–397, 1997.

- [76] M. Sandgren, A. Shaw, T. H. Ropp, S. Wu, R. Bott, A. D. Cameron, J. Ståhlberg, C. Mitchinson, and T. A. Jones, "The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 Å resolution," *Journal of molecular biology*, vol. 308, no. 2, pp. 295–310, 2001.
- [77] D. Sternberg, P. Vuayakumar, and E. T. Reese, " β -Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose," *Canadian Journal of Microbiology*, vol. 23, no. 2, pp. 139–147, 1977.
- [78] K. Réczey, A. Brumbauer, M. Bollók, Z. Szengyel, and G. Zacchi, "Use of hemicellulose hydrolysate for beta-glucosidase fermentation," *Applied biochemistry and biotechnology*, vol. 70–72, pp. 225–35, 1998.
- [79] B. Seiboth, C. Ivanova, and V. Seidl-seiboth, "*Trichoderma reesei* : A fungal enzyme producer for cellulosic biofuels," in *Biofuel Production - Recent Developments and Prospects* (M. A. Dos Santos Bernardes, ed.), ch. 13, pp. 309–340, InTech, 2011.
- [80] D. Martinez, R. M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. E. Baker, J. Chapman, O. Chertkov, P. M. Coutinho, D. Cullen, E. G. J. Danchin, I. V. Grigoriev, P. Harris, M. Jackson, C. P. Kubicek, C. S. Han, I. Ho, L. F. Larrondo, A. L. de Leon, J. K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A. A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. L. Schoch, J. Yao, R. Barabote, R. Barbote, M. A. Nelson, C. Detter, D. Bruce, C. R. Kuske, G. Xie, P. Richardson, D. S. Rokhsar, S. M. Lucas, E. M. Rubin, N. Dunn-Coleman, M. Ward, and T. S. Brettin, "Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)," *Nature biotechnology*, vol. 26, no. 5, pp. 553–560, 2008.
- [81] M. Schülein, "Enzymatic properties of cellulases from *Humicola insolens*," *Journal of Biotechnology*, vol. 57, no. 1–3, pp. 71–81, 1997.
- [82] D. W. Thayer, S. V. Lowther, and J. G. Phillips, "Cellulolytic activities of strains of the genus *cellulomonas*," *International Journal of Systematic Bacteriology*, vol. 34, no. 4, pp. 432–438, 1984.
- [83] A. Lykidis, K. Mavromatis, N. Ivanova, I. Anderson, M. Land, G. DiBartolo, M. Martinez, A. Lapidus, S. Lucas, A. Copeland, P. Richardson, D. B. Wilson, and N. Kyrpides, "Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX," *Journal of Bacteriology*, vol. 189, no. 6, pp. 2477–2486, 2007.
- [84] K. Ratanakhanokchai, R. Waeonukul, P. Pason, C. Tachaapaikoon, K. L. Kyu, K. Sakka, A. Kosugi, and Y. Mori, "Paenibacillus curdlanolyticus Strain B-6 multienzyme complex : A novel system for biomass utilization," in *Biomass Now - Cultivation and utilization* (M. D. Matovic, ed.), ch. 16, pp. 369–394, InTech, 1st ed., 2013.
- [85] P. J. Weimer, N. P. J. Price, O. Kroukamp, L. M. Joubert, G. M. Wolfaardt, and W. H. Van Zyl, "Studies of the extracellular glycocalyx of the anaerobic cellulolytic bacterium *Ruminococcus albus* 7," *Applied and Environmental Microbiology*, vol. 72, no. 12, pp. 7559–7566, 2006.
- [86] R. H. Doi, "Cellulases of mesophilic microorganisms: Cellulosome and noncellulosome producers," *Annals of the New York Academy of Sciences*, vol. 1125, no. 530, pp. 267–279, 2008.
- [87] C. Boisset, H. Chanzy, B. Henrissat, R. Lamed, Y. Shoham, and E. A. Bayer, "Digestion of crystalline cellulose substrates by the clostridium thermocellum cellulosome: structural and morphological aspects," *The Biochemical journal*, vol. 835, pp. 829–35, 1999.
- [88] E. A. Bayer, H. Chanzy, R. Lamed, and Y. Shoham, "Cellulose, cellulases and cellulosomes," *Current Opinion in Structural Biology*, vol. 8, no. 5, pp. 548–557, 1998.
- [89] H. Ohara, S. Karita, T. Kimura, K. Sakka, and K. Ohmiya, "Characterization of the cellulolytic complex (cellulosome) from *Ruminococcus albus*," *Bioscience, biotechnology, and biochemistry*, vol. 64, no. 2, pp. 254–260, 2000.
- [90] V. Arantes and J. N. Saddler, "Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis," *Biotechnology for biofuels*, vol. 3, no. 4, pp. 1–11, 2010.
- [91] S. McQueen-Mason, D. M. Durachko, and D. J. Cosgrove, "Two endogenous proteins that induce cell wall extension in plants," *The Plant cell*, vol. 4, no. November, pp. 1425–1433, 1992.
- [92] D. J. Cosgrove, "Growth of the plant cell wall," *Nature reviews. Molecular cell biology*, vol. 6, no. 11, pp. 850–861, 2005.

REFERENCES

- [93] S. J. McQueen-Mason and D. J. Cosgrove, "Expansin mode of action on cell walls. Analysis of wall hydrolysis, stress relaxation, and binding.," *Plant physiology*, vol. 107, no. 1, pp. 87–100, 1995.
- [94] J. K. C. Rose and A. B. Bennett, "Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: Parallels between cell expansion and fruit ripening," *Trends in Plant Science*, vol. 4, no. 5, pp. 176–183, 1999.
- [95] D. J. Cosgrove, P. Bedinger, and D. M. Durachko, "Group I allergens of grass pollen as cell wall-loosening agents," *Proceedings of the National Academy of Sciences*, vol. 94, no. 12, pp. 6559–6564, 1997.
- [96] E. S. Kim, H. J. Lee, W.-g. Bang, I.-g. Choi, and K. H. Kim, "Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose.," *Biotechnology and bioengineering*, vol. 102, no. 5, pp. 1342–1353, 2009.
- [97] F. Kerff, A. Amoroso, R. Herman, E. Sauvage, S. Petrella, P. Filée, P. Charlier, B. Joris, A. Tabuchi, N. Nikolaidis, and D. J. Cosgrove, "Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 44, pp. 16876–16881, 2008.
- [98] M. Eibinger, K. Sigl, J. Sattelmow, T. Ganner, J. Ramoni, B. Seiboth, H. Plank, and B. Nidetzky, "Functional characterization of the native swollenin from *Trichoderma reesei*: study of its possible role as C1 factor of enzymatic lignocellulose conversion," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 178, 2016.
- [99] M. Andberg, M. Penttilä, and M. Saloheimo, "Swollenin from *Trichoderma reesei* exhibits hydrolytic activity against cellulosic substrates with features of both endoglucanases and cellobiohydrolases," *Bioresource Technology*, vol. 181, pp. 105–113, 2015.
- [100] R. E. Quiroz-Castañeda, C. Martínez-Anaya, L. I. Cuervo-Soto, L. Segovia, and J. L. Folch-Mallol, "Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*," *Microbial cell factories*, vol. 10, p. 8, 2011.
- [101] X. Li, W. T. Beeson, C. M. Phillips, M. a. Marletta, and J. H. D. Cate, "Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases," *Structure*, vol. 20, no. 6, pp. 1051–1061, 2012.
- [102] K.-e. Eriksson, B. Pettersson, and U. Westermark, "Oxidation: an important enzyme reaction in fungal degradation of cellulose," *FEBS Letters*, vol. 49, no. 2, pp. 282–285, 1974.
- [103] C. Felby, B. R. Nielsen, P. O. Olesen, and L. H. Skibsted, "Identification and quantification of radical reaction intermediates by electron spin resonance spectrometry of laccase-catalyzed oxidation of wood fibers from beech (*Fagus sylvatica*)," *Applied Microbiology and Biotechnology*, vol. 48, no. 4, pp. 459–464, 1997.
- [104] L. L. Leggio, D. Welner, and L. De Maria, "A structural overview of GH61 proteins - fungal cellulose degrading polysaccharide monooxygenases," *Computational and Structural Biotechnology Journal*, vol. 2, no. 3, pp. 1–8, 2012.
- [105] J. Karlsson, M. Saloheimo, M. Siika-aho, M. Tenkanen, M. Penttilä, and F. Tjerneld, "Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*," *European Journal of Biochemistry*, vol. 268, pp. 6498–6507, dec 2001.
- [106] S. Karkehabadi, H. Hansson, S. Kim, K. Piens, C. Mitchinson, and M. Sandgren, "The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 Å resolution," *Journal of molecular biology*, vol. 383, no. 1, pp. 144–54, 2008.
- [107] J. Schnellmann, A. Zeltins, H. Blaak, and H. Schrempf, "The novel lectin-like protein CHB1 is encoded by a chitin-inducible *Streptomyces olivaceoviridis* gene and binds specifically to crystalline β -chitin of fungi and other organisms," *Molecular Microbiology*, vol. 13, pp. 807–819, sep 1994.
- [108] G. Vaaje-Kolstad, D. R. Houston, A. H. K. Riemen, V. G. H. Eijsink, and D. M. F. Van Aalten, "Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21," *Journal of Biological Chemistry*, vol. 280, no. 12, pp. 11313–11319, 2005.
- [109] G. R. Hemsworth, B. Henrissat, G. J. Davies, and P. H. Walton, "Discovery and characterization of a new family of lytic polysaccharide monooxygenases," *Nature Chemical Biology*, vol. 10, pp. 122–126, 2014.

- [110] A. J. Book, R. M. Yennamalli, T. E. Takasuka, C. R. Currie, G. N. Phillips, and B. G. Fox, "Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases," *Biotechnology for Biofuels*, vol. 7, no. 1, p. 109, 2014.
- [111] C. M. Phillips, W. T. Beeson, J. H. Cate, and M. A. Marletta, "Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*," *ACS chemical biology*, vol. 6, no. 12, pp. 1399–406, 2011.
- [112] V. V. Vu, W. T. Beeson, C. M. Phillips, J. H. D. Cate, and M. A. Marletta, "Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases," *Journal of the American Chemical Society*, vol. 136, no. 2, pp. 562–565, 2014.
- [113] M. Bey, S. Zhou, L. Poidevin, B. Henrissat, P. M. Coutinho, J.-G. Berrin, and J.-C. Sigoillot, "Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from *Podospora anserina*," *Applied and environmental microbiology*, vol. 79, no. 2, pp. 488–96, 2013.
- [114] Z. Forsberg, a. K. Mackenzie, M. Sorlie, a. K. Rohr, R. Helland, a. S. Arvai, G. Vaaje-Kolstad, and V. G. H. Eijsink, "Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8446–8451, 2014.
- [115] P. K. Busk and L. Lange, "Function-based classification of carbohydrate-active enzymes by recognition of short, conserved peptide motifs," *Applied and Environmental Microbiology*, vol. 79, no. 11, pp. 3380–3391, 2013.
- [116] P. K. Busk, M. Lange, B. Pilgaard, and L. Lange, "Several genes encoding enzymes with the same activity are necessary for aerobic fungal degradation of cellulose in nature.," *PloS one*, vol. 9, no. 12, p. e114138, 2014.
- [117] G. R. Hemsworth, E. M. Johnston, and G. J. Davies, "Lytic polysaccharide monooxygenases in biomass conversion," *Trends in Biotechnology*, vol. 33, no. 12, pp. 747–761, 2015.
- [118] M. Dimarogona, E. Topakas, L. Olsson, and P. Christakopoulos, "Lignin boosts the cellulase performance of a GH-61 enzyme from *Sporotrichum thermophile*," *Bioresource technology*, vol. 110, pp. 480–7, apr 2012.
- [119] J. Larsen, M. Øtergaard Petersen, L. Thirup, H. W. Li, and F. K. Iversen, "The IBUS process - Lignocellulosic bioethanol close to a commercial reality," *Chemical Engineering and Technology*, vol. 31, no. 5, pp. 765–772, 2008.
- [120] S. Di Risio, C. S. Hu, B. A. Saville, D. Liao, and J. Lorie, "Large-scale, high-solids enzymatic hydrolysis of steam-exploded poplar," *Biofuels, Bioproducts and Biorefining*, vol. 6, no. 3, pp. 246–256, 2012.
- [121] D. Cannella, C.-W. C. Hsieh, C. Felby, and H. Jørgensen, "Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content.," *Biotechnology for biofuels*, vol. 5, no. 1, p. 26, 2012.
- [122] U. F. Rodríguez-Zúñiga, D. Cannella, R. d. C. Giordano, R. d. L. C. Giordano, H. Jørgensen, and C. Felby, "Lignocellulose pretreatment technologies affect the level of enzymatic cellulose oxidation by LPMO," *Green Chemistry*, vol. 17, no. 5, pp. 2896–2903, 2015.
- [123] J. Hu, R. Chandra, V. Arantes, K. Gourlay, J. Susan van Dyk, and J. N. Saddler, "The addition of accessory enzymes enhances the hydrolytic performance of cellulase enzymes at high solid loadings," *Bioresource Technology*, vol. 186, pp. 149–53, 2015.
- [124] T. Isaksen, B. Westereng, F. L. Aachmann, J. W. Agger, D. Kracher, R. Kittl, R. Ludwig, D. Haltrich, V. G. H. Eijsink, and S. J. Horn, "A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides," *The Journal of biological chemistry*, vol. 289, no. 5, pp. 2632–42, 2014.
- [125] C. Bennati-Granier, S. Garajova, C. Champion, S. Grisel, M. Haon, S. Zhou, M. Fanuel, D. Ropartz, H. Rogniaux, I. Gimbert, E. Record, and J.-G. Berrin, "Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*," *Biotechnology for Biofuels*, vol. 8, no. 1, p. 90, 2015.

REFERENCES

- [126] Y. Kojima, A. Várnai, T. Ishida, N. Sunagawa, D. M. Petrovic, K. Igarashi, J. Jellison, B. Goodell, G. Alfredsen, B. Westereng, V. G. Eijsink, and M. Yoshida, "Characterization of an LPMO from the brown-rot fungus *Gloeophyllum trabeum* with broad xyloglucan specificity, and its action on cellulose-xyloglucan complexes," *Applied and Environmental Microbiology*, vol. 82, no. 22, pp. 6557–6572, 2016.
- [127] L. Nekiunaite, D. M. Petrović, B. Westereng, G. Vaaje-Kolstad, M. Abou Hachem, A. Várnai, and V. G. Eijsink, "Fg LPMO9A from *Fusarium graminearum* cleaves xyloglucan independently of the backbone substitution pattern," *FEBS Letters*, vol. 590, no. 19, pp. 3346–3356, 2016.
- [128] I. J. Kim, H. J. Youn, and K. H. Kim, "Synergism of an auxiliary activity 9 (AA9) from *Chaetomium globosum* with xylanase on the hydrolysis of xylan and lignocellulose," *Process Biochemistry*, vol. 51, no. 10, pp. 1445–1451, 2016.
- [129] I. J. Kim, N. Seo, H. J. An, J.-H. Kim, P. V. Harris, and K. H. Kim, "Type-dependent action modes of TtAA9E and TaAA9A acting on cellulose and differently pretreated lignocellulosic substrates," *Biotechnology for Biofuels*, vol. 10, no. 1, p. 46, 2017.
- [130] A. Karnaouri, L. Matsakas, E. Topakas, U. Rova, and P. Christakopoulos, "Development of thermophilic tailor-made enzyme mixtures for the bioconversion of agricultural and forest residues," *Frontiers in Microbiology*, vol. 7, no. FEB, pp. 1–14, 2016.
- [131] Y. Arfi, M. Shamsoum, I. Rogachev, Y. Peleg, and E. a. Bayer, "Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 25, pp. 9109–14, 2014.
- [132] Y. S. Nakagawa, M. Kudo, J. S. M. Loose, T. Ishikawa, K. Totani, V. G. H. Eijsink, and G. Vaaje-Kolstad, "A small lytic polysaccharide monooxygenase from *Streptomyces griseus* targeting alpha- and beta-chitin," *FEBS Journal*, vol. 282, no. 6, pp. 1065–1079, 2015.
- [133] L. Nekiunaite, T. Isaksen, G. Vaaje-Kolstad, and M. Abou Hachem, "Fungal lytic polysaccharide monooxygenases bind starch and β -cyclodextrin similarly to amylolytic hydrolases," *FEBS Letters*, 2016.
- [134] A. J. Book, R. M. Yennamalli, T. E. Takasuka, C. R. Currie, G. N. Phillips, and B. G. Fox, "Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases," *Biotechnology for Biofuels*, vol. 7, no. 1, p. 109, 2014.
- [135] M. Wu, G. T. Beckham, A. M. Larsson, T. Ishida, S. Kim, C. M. Payne, M. E. Himmel, M. F. Crowley, S. J. Horn, B. Westereng, K. Igarashi, M. Samejima, J. Ståhlberg, V. G. H. Eijsink, and M. Sandgren, "Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus *Phanerochaete chrysosporium*," *The Journal of biological chemistry*, pp. 1–21, mar 2013.
- [136] C. Sygmund, D. Kracher, S. Scheiblbrandner, K. Zahma, A. K. G. Felice, W. Harreither, R. Kittl, and R. Ludwig, "Characterization of the two *Neurospora crassa* cellobiose dehydrogenases and their connection to oxidative cellulose degradation," *Applied and environmental microbiology*, vol. 78, pp. 6161–71, sep 2012.
- [137] K. E. H. Frandsen and L. Lo Leggio, "Lytic polysaccharide monooxygenases: a crystallographer's view on a new class of biomass-degrading enzymes," *IUCrJ*, vol. 3, no. 6, pp. 18779–18784, 2016.
- [138] Z. Forsberg, C. E. Nelson, B. Dalhus, S. Mekasha, J. S. M. Loose, L. I. Crouch, Å. K. Røhr, J. G. Gardner, V. G. H. Eijsink, and G. Vaaje-Kolstad, "Structural and functional analysis of a lytic polysaccharide monooxygenase important for efficient utilization of chitin in *Cellvibrio japonicus*," *Journal of Biological Chemistry*, vol. 291, no. 14, pp. 7300–7312, 2016.
- [139] Z. Forsberg, A. K. Røhr, S. Mekasha, K. K. Andersson, V. G. H. Eijsink, G. Vaaje-Kolstad, and M. Sørlie, "Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases," *Biochemistry*, vol. 53, no. 10, pp. 1647–56, 2014.
- [140] F. L. Achmann, M. Sørlie, G. Skjåk-Bræk, V. G. H. Eijsink, and G. Vaaje-Kolstad, "NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 46, pp. 18779–84, 2012.

- [141] B. Westereng, T. Ishida, G. Vaaje-Kolstad, M. Wu, V. G. H. Eijsink, K. Igarashi, M. Samejima, J. Ståhlberg, S. J. Horn, and M. Sandgren, "The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose.," *PLoS one*, vol. 6, p. e27807, jan 2011.
- [142] G. Vaaje-Kolstad, L. A. Bøhle, S. Gåseidnes, B. Dalhus, M. Bjørås, G. Mathiesen, and V. G. H. Eijsink, "Characterization of the chitinolytic machinery of *Enterococcus faecalis* V583 and high-resolution structure of its oxidative CBM33 enzyme.," *Journal of molecular biology*, vol. 416, no. 2, pp. 239–54, 2012.
- [143] G. R. Hemsworth, E. J. Taylor, R. Q. Kim, R. C. Gregory, S. J. Lewis, J. P. Turkenburg, A. Parkin, G. J. Davies, and P. H. Walton, "The copper active site of CBM33 polysaccharide oxygenases," *Journal of the American Chemical Society*, vol. 135, no. 16, pp. 6069–6077, 2013.
- [144] R. C. Gregory, G. R. Hemsworth, J. P. Turkenburg, S. Hart, P. Walton, and G. Davies, "Activity, stability and 3-D structure of the Cu(II) form of a chitin-active lytic polysaccharide monooxygenase from *Bacillus amyloliquefaciens*," *Dalton Trans.*, vol. 45, pp. 16904–16912, 2016.
- [145] W. T. Beeson, V. V. Vu, E. A. Span, C. M. Phillips, and M. A. Marletta, "Cellulose degradation by polysaccharide monooxygenases," *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 923–946, 2015.
- [146] A. S. Borisova, T. Isaksen, M. Dimarogona, A. a. Kognole, G. Mathiesen, A. Várnai, Å. K. Røhr, C. M. Payne, M. Sørli, M. Sandgren, and V. G. H. Eijsink, "Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity," *Journal of Biological Chemistry*, vol. 290, no. 38, pp. 22955–22969, 2015.
- [147] W. T. Beeson, C. M. Phillips, J. H. D. Cate, and M. a. Marletta, "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases," *Journal of the American Chemical Society*, vol. 134, no. 2, pp. 890–2, 2012.
- [148] B. Bissaro, A. K. Rohr, M. Skaugen, Z. Forsberg, S. J. Horn, G. Vaaje-Kolstad, and V. Eijsink, "Fenton-type chemistry by a copper enzyme: molecular mechanism of polysaccharide oxidative cleavage," *bioRxiv*, p. doi: <http://dx.doi.org/10.1101/097022>, 2016.
- [149] M. Frommhagen, M. J. Koetsier, A. H. Westphal, J. Visser, S. W. A. Hinz, J.-P. Vincken, W. J. H. van Berkel, M. A. Kabel, and H. Gruppen, "Lytic polysaccharide monooxygenases from *Myceliophthora thermophila* C1 differ in substrate preference and reducing agent specificity," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 186, 2016.
- [150] J. W. Agger, T. Isaksen, A. Várnai, S. Vidal-Melgosa, W. G. T. Willats, R. Ludwig, S. J. Horn, V. G. H. Eijsink, and B. Westereng, "Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 17, pp. 6287–92, 2014.
- [151] B. Westereng, D. Cannella, J. Wittrup Agger, H. Jørgensen, M. Larsen Andersen, V. G. H. Eijsink, and C. Felby, "Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer.," *Scientific reports*, vol. 5, no. November, p. 18561, 2015.
- [152] D. Kracher, S. Scheiblbrandner, A. K. G. Felice, E. Breslmayr, M. Preims, D. Haltrich, V. G. H. Eijsink, and R. Ludwig, "Extracellular electron transfer systems fuel cellulose oxidative degradation," *Science*, vol. 352, no. 6289, pp. 1098–1101, 2016.
- [153] T.-C. Tan, D. Kracher, R. Gandini, C. Sygmund, R. Kittl, D. Haltrich, B. M. Hällberg, R. Ludwig, and C. Divine, "Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation," *Nature Communications*, vol. 6, p. 7542, 2015.
- [154] J. G. Gardner, L. Crouch, A. Labourel, Z. Forsberg, Y. V. Bukhman, G. Vaaje-Kolstad, H. J. Gilbert, and D. H. Keating, "Systems biology defines the biological significance of redox-active proteins during cellulose degradation in an aerobic bacterium.," *Molecular microbiology*, vol. 94, pp. 1121–1133, oct 2014.
- [155] D. Cannella, K. Möllers, N.-U. Frigaard, P. Jensen, M. Bjerrum, K. Johansen, and C. Felby, "Light-driven oxidation of polysaccharides by photosynthetic pigments and a metalloenzyme," *Nat. Comm.*, vol. 7, no. 11134, pp. 1–8, 2016.
- [156] S. Garajova, Y. Mathieu, M. R. Beccia, C. Bennati-Granier, F. Biaso, M. Fanuel, D. Ropartz, B. Guigliarelli, E. Record, H. Rogniaux, B. Henrissat, and J.-G. Berrin, "Single-domain flavoenzymes trigger lytic polysaccharide monooxygenases for oxidative degradation of cellulose," *Scientific Reports*, vol. 6, p. 28276, 2016.

REFERENCES

- [157] E. Ximenes, Y. Kim, N. Mosier, B. Dien, and M. Ladisch, "Deactivation of cellulases by phenols," *Enzyme and Microbial Technology*, vol. 48, no. 1, pp. 54–60, 2011.
- [158] A. Tejirian and F. Xu, "Inhibition of enzymatic cellulolysis by phenolic compounds," *Enzyme and Microbial Technology*, vol. 48, no. 3, pp. 239–247, 2011.
- [159] Y. Kim, E. Ximenes, N. S. Mosier, and M. R. Ladisch, "Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass," *Enzyme and Microbial Technology*, vol. 48, no. 4-5, pp. 408–415, 2011.
- [160] M. Michelin, E. Ximenes, M. de Lourdes Teixeira de Moraes Polizeli, and M. R. Ladisch, "Effect of phenolic compounds from pretreated sugarcane bagasse on cellulolytic and hemicellulolytic activities," *Bioresource Technology*, vol. 199, pp. 275–278, 2016.
- [161] A. T. Martinez, "How to break down crystalline cellulose," *Science*, vol. 352, no. 6289, pp. 1050–1051, 2016.
- [162] S. Kim, J. Ståhlberg, M. Sandgren, R. S. Paton, and G. T. Beckham, *Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxy, oxygen-rebound mechanism.*, vol. 111, 2014.
- [163] S. J. Horn, G. Vaaje-Kolstad, B. Westereng, and V. G. Eijsink, "Novel enzymes for the degradation of cellulose," *Biotechnology for biofuels*, vol. 5, p. 45, jan 2012.
- [164] L. Nekiunaite, M. Ø. Arntzen, B. Svensson, G. Vaaje-Kolstad, and M. Abou Hachem, "Lytic polysaccharide monooxygenases and other oxidative enzymes are abundantly secreted by *Aspergillus nidulans* grown on different starches," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 187, 2016.
- [165] D. K. Paspaliari, J. S. M. Loose, M. H. Larsen, and G. Vaaje-Kolstad, "Listeria monocytogenes has a functional chitinolytic system and an active lytic polysaccharide monooxygenase," *The FEBS Journal*, jan 2015.
- [166] J. S. M. Loose, Z. Forsberg, M. W. Fraaije, V. G. H. Eijsink, and G. Vaaje-Kolstad, "A rapid quantitative activity assay shows that the *Vibrio cholerae* colonization factor GbpA is an active lytic polysaccharide monooxygenase," *FEBS letters*, vol. 588, no. 18, pp. 3435–40, 2014.
- [167] E. Chiu, M. Hijnen, R. D. Bunker, M. Boudes, C. Rajendran, K. Aizel, V. Oliéric, C. Schulze-Briese, W. Mitsuhashi, V. Young, V. K. Ward, M. Bergoin, P. Metcalf, and F. Coulibaly, "Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization," *Proceedings of the National Academy of Sciences*, vol. 112, no. 13, pp. 3973–3978, 2015.
- [168] R. J. O'Connell, M. R. Thon, S. Hacquard, S. G. Amyotte, J. Kleemann, M. F. Torres, U. Damm, E. a. Buiate, L. Epstein, N. Alkan, J. Altmüller, L. Alvarado-Balderrama, C. a. Bauser, C. Becker, B. W. Birren, Z. Chen, J. Choi, J. A. Crouch, J. P. Duwick, M. a. Farman, P. Gan, D. Heiman, B. Henrissat, R. J. Howard, M. Kabbage, C. Koch, B. Kracher, Y. Kubo, A. D. Law, M.-H. Lebrun, Y.-H. Lee, I. Miyara, N. Moore, U. Neumann, K. Nordström, D. G. Panaccione, R. Panstruga, M. Place, R. H. Proctor, D. Prusky, G. Rech, R. Reinhardt, J. a. Rollins, S. Rounsley, C. L. Schardl, D. C. Schwartz, N. Shenoy, K. Shirasu, U. R. Sikhakolli, K. Stüber, S. a. Sukno, J. a. Sweigard, Y. Takano, H. Takahara, F. Trail, H. C. van der Does, L. M. Voll, I. Will, S. Young, Q. Zeng, J. Zhang, S. Zhou, M. B. Dickman, P. Schulze-Lefert, E. Ver Loren van Themaat, L.-J. Ma, and L. J. Vaillancourt, "Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses," *Nature genetics*, vol. 44, no. 9, pp. 1060–5, 2012.
- [169] B. R. Scott, H. Z. Huang, J. Frickman, R. Halvorsen, and K. S. Johansen, "Catalase improves saccharification of lignocellulose by reducing lytic polysaccharide monooxygenase-associated enzyme inactivation," *Biotechnology Letters*, vol. 38, no. 3, pp. 425–434, 2016.
- [170] A. S. Adams, M. S. Jordan, S. M. Adams, G. Suen, L. a. Goodwin, K. W. Davenport, C. R. Currie, and K. F. Raffa, "Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*," *The ISME journal*, vol. 5, no. 8, pp. 1323–1331, 2011.
- [171] T. E. Takasuka, A. J. Book, G. R. Lewin, C. R. Currie, and B. G. Fox, "Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*," *Scientific Reports*, vol. 3, pp. 1–10, jan 2013.
- [172] S. Sun, S. Sun, X. Cao, and R. Sun, "The role of pretreatment in improving the enzymatic hydrolysis of lignocellulosic materials," *Bioresource Technology*, vol. 199, pp. 49–58, 2016.
- [173] B. E. Dale, C. K. Leong, T. K. Pham, V. M. Esquivel, I. Rios, and V. M. Latimer, "Hydrolysis of lignocellulosics at low enzyme levels: Application of the AFEX process," *Bioresource Technology*, vol. 56, no. 1, pp. 111–116, 1996.

-
- [174] S. E. Blumer-Schuetz, S. D. Brown, K. B. Sander, E. A. Bayer, I. Kataeva, J. V. Zurawski, J. M. Conway, M. W. W. Adams, and R. M. Kelly, "Thermophilic lignocellulose deconstruction," *FEMS Microbiology Reviews*, vol. 38, no. 3, pp. 393–448, 2014.
 - [175] C. C. Geddes, I. U. Nieves, and L. O. Ingram, "Advances in ethanol production," *Current opinion in biotechnology*, vol. 22, no. 3, pp. 312–9, 2011.
 - [176] M. Bettiga, O. Bengtsson, B. Hahn-Hägerdal, and M. F. Gorwa-Grauslund, "Arabinose and xylose fermentation by recombinant *Saccharomyces cerevisiae* expressing a fungal pentose utilization pathway," *Microbial cell factories*, vol. 8, p. 40, 2009.
 - [177] A. Margeot, B. Hahn-Hägerdal, M. Edlund, R. Slade, and F. Monot, "New improvements for lignocellulosic ethanol," *Current Opinion in Biotechnology*, vol. 20, no. 3, pp. 372–380, 2009.
 - [178] B. Hahn-Hägerdal, M. Galbe, M. F. Gorwa-Grauslund, G. Lidén, and G. Zacchi, "Bio-ethanol - the fuel of tomorrow from the residues of today," *Trends in Biotechnology*, vol. 24, no. 12, pp. 549–556, 2006.
 - [179] L. R. Lynd, W. H. Van Zyl, J. E. McBride, and M. Laser, "Consolidated bioprocessing of cellulosic biomass: An update," *Current Opinion in Biotechnology*, vol. 16, no. 5, pp. 577–583, 2005.
 - [180] L. R. Lynd, M. S. Laser, D. Bransby, B. E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J. D. McMillan, J. Sheehan, and C. E. Wyman, "How biotech can transform biofuels," *Nature biotechnology*, vol. 26, no. 2, pp. 169–172, 2008.
 - [181] D. G. Olson, J. E. McBride, A. Joe Shaw, and L. R. Lynd, "Recent progress in consolidated bioprocessing," *Current Opinion in Biotechnology*, vol. 23, no. 3, pp. 396–405, 2012.
 - [182] S. S. Yim, J. W. Choi, S. H. Lee, and K. J. Jeong, "Modular optimization of hemicellulose-utilizing pathway in *Corynebacterium glutamicum* for consolidated bioprocessing of hemicellulosic biomass," *ACS Synthetic Biology*, vol. 5, no. 4, pp. 334–343, 2016.
 - [183] G. Müller, A. Várnai, K. S. Johansen, V. G. H. Eijsink, and S. J. Horn, "Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions," *Biotechnology for biofuels*, vol. 8, no. 1, p. 187, 2015.
 - [184] D. Cannella, C.-W. C. Hsieh, C. Felby, and H. Jørgensen, "Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content," *Biotechnology for biofuels*, vol. 5, p. 26, jan 2012.
 - [185] D. Cannella and H. Jørgensen, "Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production?," *Biotechnology and Bioengineering*, vol. 111, no. 1, pp. 59–68, 2014.
 - [186] Y. Zhang, A. Kumar, P. R. Hardwidge, T. Tanaka, A. Kondo, and P. V. Vadlani, "D-lactic acid production from renewable lignocellulosic biomass via genetically modified *Lactobacillus plantarum*," *Biotechnology Progress*, vol. 32, no. 2, pp. 271–278, 2016.
 - [187] Y.-H. P. Zhang, J. Ciu, L. R. Lynd, and L. R. Kuang, "A Transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure," *Biomacromolecules*, vol. 7, pp. 644–648, 2006.
 - [188] L. Kumar, V. Arantes, R. Chandra, and J. Saddler, "The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility," *Bioresource Technology*, vol. 103, no. 1, pp. 201–208, 2012.
 - [189] X. Li, K. Chomvong, V. Y. Yu, J. M. Liang, Y. Lin, and J. H. D. Cate, "Cellobionic acid utilization: from *Neurospora crassa* to *Saccharomyces cerevisiae*," *Biotechnology for Biofuels*, vol. 8, no. 1, p. 120, 2015.
 - [190] J. M. Galazka, C. Tian, W. T. Beeson, B. Martinez, N. L. Glass, and J. H. D. Cate, "Cellodextrin transport in yeast for improved biofuel production," *Science*, vol. 330, no. October, pp. 84–86, 2010.
 - [191] X. Li, V. Y. Yu, Y. Lin, K. Chomvong, R. Estrela, A. Park, J. M. Liang, E. A. Znameroski, J. Feehan, S. R. Kim, Y. S. Jin, N. Louise Glass, and J. H. D. Cate, "Expanding xylose metabolism in yeast for plant cell wall conversion to biofuels," *eLife*, vol. 2015, no. 4, pp. 1–55, 2015.

REFERENCES

- [192] M. Á. B. Alcántara, J. Dobruchowska, P. Azadi, B. D. García, F. P. Molina-Heredia, and F. M. Reyes-Sosa, "Recalcitrant carbohydrates after enzymatic hydrolysis of pretreated lignocellulosic biomass," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 207, 2016.
- [193] A. Aharoni, A. D. Griffiths, and D. S. Tawfik, "High-throughput screens and selections of enzyme-encoding genes," *Current Opinion in Chemical Biology*, vol. 9, no. 2, pp. 210–216, 2005.
- [194] U. T. Bornscheuer, "Protein engineering: beating the odds," *Nature chemical biology*, vol. 12, pp. 54–55, 2016.
- [195] a. D. Keefe and J. W. Szostak, "Functional proteins from a random-sequence library.," *Nature*, vol. 410, no. April, pp. 715–718, 2001.
- [196] P. a. Dalby, "Strategy and success for the directed evolution of enzymes.," *Current opinion in structural biology*, vol. 21, pp. 473–80, aug 2011.
- [197] K. Sohail Siddiqui, "Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability," *Critical Reviews in Biotechnology*, vol. 8551, pp. 1–14, 2016.
- [198] A. S. Bommarius, J. K. Blum, and M. J. Abrahamson, "Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst.," *Current opinion in chemical biology*, vol. 15, no. 2, pp. 194–200, 2011.
- [199] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, "Protein stability promotes evolvability.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 15, pp. 5869–74, 2006.
- [200] N. G. Brown, J. M. Pennington, W. Huang, T. Ayvaz, and T. Palzkill, "Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM beta-lactamases," *Journal of Molecular Biology*, vol. 404, no. 5, pp. 832–846, 2010.
- [201] C. Schmidt-Dannert and F. H. Arnold, "Directed evolution of industrial enzymes," *Trends in Biotechnology*, vol. 17, no. 4, pp. 135–136, 1999.
- [202] P. A. Patten, R. . I. Howardt, and W. P. Stemmer, "Applications of DNA shuffling to pharmaceuticals and vaccines," *Current Opinion in Biotechnology*, vol. 8, pp. 724–733, 1997.
- [203] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. a. Sheldon, and G. W. Huisman, "Improving catalytic function by ProSAR-driven enzyme evolution.," *Nature biotechnology*, vol. 25, no. 3, pp. 338–44, 2007.
- [204] S. G. Peisajovich and D. S. Tawfik, "Protein engineers turned evolutionists.," *Nature methods*, vol. 4, no. 12, pp. 991–994, 2007.
- [205] Q. Wang and T. Xia, "Enhancement of the activity and alkaline pH stability of *Thermobifida fusca* xylanase A by directed evolution," *Biotechnology Letters*, vol. 30, no. 5, pp. 937–944, 2008.
- [206] D. E. Stephens, S. Singh, and K. Permaul, "Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability," *FEMS Microbiology Letters*, vol. 293, no. 1, pp. 42–47, 2009.
- [207] J. H. Kim, G. S. Choi, S. B. Kim, W. H. Kim, J. Y. Lee, Y. W. Ryu, and G. J. Kim, "Enhanced thermostability and tolerance of high substrate concentration of an esterase by directed evolution," *Journal of Molecular Catalysis B: Enzymatic*, vol. 27, no. 4-6, pp. 169–175, 2004.
- [208] T. Verhaeghe, M. Diricks, D. Aerts, W. Soetaert, and T. Desmet, "Mapping the acceptor site of sucrose phosphorylase from *Bifidobacterium adolescentis* by alanine scanning," *Journal of Molecular Catalysis B: Enzymatic*, vol. 96, pp. 81–88, 2013.
- [209] M. T. Reetz and J. D. Carballeira, "Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes.," *Nature protocols*, vol. 2, no. 4, pp. 891–903, 2007.
- [210] S. Kille, C. G. Acevedo-rocha, L. P. Parra, Z.-g. Zhang, D. J. Opperman, M. T. Reetz, and J. P. Acevedo, "Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis," *ACS chemical biology*, vol. 2, no. 2, pp. 83–92, 2013.

- [211] M. T. Reetz, L.-W. Wang, and M. Bocola, "Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space," *Angewandte Chemie*, vol. 118, no. 8, pp. 1258–1263, 2006.
- [212] M. T. Reetz, J. D. Carballeira, and A. Vogel, "Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability.," *Angewandte Chemie (International ed. in English)*, vol. 45, no. 46, pp. 7745–51, 2006.
- [213] H. Jochens, D. Aerts, and U. T. Bornscheuer, "Thermostabilization of an esterase by alignment-guided focussed directed evolution.," *Protein engineering, design & selection : PEDS*, vol. 23, no. 12, pp. 903–9, 2010.
- [214] B. Steipe, B. Schiller, A. Plückthun, and S. Steinbacher, "Sequence statistics reliably predict stabilizing mutations in a protein domain.," *Journal of molecular biology*, vol. 240, no. 3, pp. 188–192, 1994.
- [215] M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes, and A. P. van Loon, "From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase," *Protein Engineering Design and Selection*, vol. 13, no. 1, pp. 49–57, 2000.
- [216] E. R. G. Main, Y. Xiong, M. J. Cocco, L. D'Andrea, and L. Regan, "Design of stable alpha-helical arrays from an idealized TPR motif," *Structure*, vol. 11, no. 5, pp. 497–508, 2003.
- [217] H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, and A. Plückthun, "Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins," *Journal of Molecular Biology*, vol. 332, no. 2, pp. 489–503, 2003.
- [218] N. Amin, A. D. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti, and V. Schellenberger, "Construction of stabilized proteins by combinatorial consensus mutagenesis," *Protein Engineering, Design and Selection*, vol. 17, no. 11, pp. 787–793, 2004.
- [219] D. Aerts, T. Verhaeghe, H.-j. Joosten, G. Vriend, and W. Soetaert, "Consensus engineering of sucrose phosphorylase : The outcome reflects the sequence input," vol. 110, no. 10, pp. 2563–2572, 2013.
- [220] D. Suplatov, V. Voevodin, and V. Švedas, "Robust enzyme design: Bioinformatic tools for improved protein stability.," *Biotechnology journal*, vol. 10, pp. 1–12, dec 2014.
- [221] J. Franceus, T. Verhaeghe, and T. Desmet, "Correlated positions in protein evolution and engineering," *Journal of Industrial Microbiology & Biotechnology*, pp. doi:10.1007/s10295-016-1811-1, 2016.
- [222] G. Seddon, V. Lounnas, R. McGuire, T. van den Bergh, R. P. Bywater, L. Oliveira, and G. Vriend, "Drug design for ever, from hype to hope.," *Journal of computer-aided molecular design*, vol. 26, no. 1, pp. 137–50, 2012.
- [223] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler, "Practically useful: What the Rosetta protein modeling suite can do for you," *Biochemistry*, vol. 49, no. 14, pp. 2987–2998, 2010.
- [224] P. Heinzelman, C. D. Snow, M. A. Smith, X. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull, and F. H. Arnold, "SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability," *Journal of Biological Chemistry*, vol. 284, no. 39, pp. 26229–26233, 2009.
- [225] V. G. H. Eijssink, S. Gaseidnes, T. V. Borchert, and B. Van Den Burg, "Directed evolution of enzyme stability," *Biomolecular Engineering*, vol. 22, no. 1-3, pp. 21–30, 2005.
- [226] A. L. Serdakowski and J. S. Dordick, "Enzyme activation for organic solvents made easy," *Trends in Biotechnology*, vol. 26, no. 1, pp. 48–54, 2008.
- [227] A. L. Pey, D. Rodríguez-Larrea, S. Bomke, S. Dammers, R. Godoy-Ruiz, M. M. Garcia-Mira, and J. M. Sanchez-Ruiz, "Engineering proteins with tunable thermodynamic and kinetic stabilities," *Proteins: Structure, Function and Genetics*, vol. 71, no. 1, pp. 165–174, 2008.
- [228] E. Vazquez-Figueroa, V. Yeh, J. M. Broering, J. F. Chaparro-Riggers, and A. S. Bommaris, "Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media," *Protein Engineering, Design and Selection*, vol. 21, no. 11, pp. 673–680, 2008.

REFERENCES

- [229] R. Lumry and H. Eyring, "Conformation changes of proteins," *J. Phys. Chem*, vol. 58, no. 2, pp. 110–120, 1954.
- [230] I. M. Plaza Del Pino, B. Ibarra-Molero, and J. M. Sanchez-Ruiz, "Lower kinetic limit to protein thermal stability: A proposal regarding protein stability in vivo and its relation with misfolding diseases," *Proteins: Structure, Function and Genetics*, vol. 40, no. 1, pp. 58–70, 2000.
- [231] K. M. Polizzi, A. S. Bommarius, J. M. Broering, and J. F. Chaparro-Riggers, "Stability of biocatalysts," *Current opinion in chemical biology*, vol. 11, no. 2, pp. 220–5, 2007.
- [232] P. V. Iyer and L. Ananthanarayan, "Enzyme stability and stabilization - Aqueous and non-aqueous environment," *Process Biochemistry*, vol. 43, no. 10, pp. 1019–1032, 2008.
- [233] Y. Xie, J. An, G. Yang, G. Wu, Y. Zhang, L. Cui, and Y. Feng, "Enhanced enzyme kinetic stability by increasing rigidity within the active site," *Journal of Biological Chemistry*, vol. 289, no. 11, pp. 7994–8006, 2014.
- [234] C. O. Fagain, "Understanding and increasing protein stability," *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, vol. 1252, no. 1, pp. 1–14, 1995.
- [235] L. Fernández, L. Gómez, H. L. Ramírez, M. L. Villalonga, and R. Villalonga, "Thermal stabilization of trypsin with glycol chitosan," *Journal of Molecular Catalysis B: Enzymatic*, vol. 34, no. 1-6, pp. 14–17, 2005.
- [236] C. Mateo, J. M. Palomo, G. Fernandez-Lorente, J. M. Guisan, and R. Fernandez-Lafuente, "Improvement of enzyme activity, stability and selectivity via immobilization techniques," *Enzyme and Microbial Technology*, vol. 40, no. 6, pp. 1451–1463, 2007.
- [237] C. Mateo, O. Abian, R. Fernandez-Lafuente, and J. M. Guisan, "Increase in conformational stability of enzymes immobilized on epoxy-activated supports by favoring additional multipoint covalent attachment," *Enzyme and Microbial Technology*, vol. 26, no. 7, pp. 509–515, 2000.
- [238] R. M. Yennamalli, A. J. Rader, J. D. Wolt, and T. Z. Sen, "Thermostability in endoglucanases is fold-specific," *BMC Structural Biology*, vol. 11, no. 1, p. 10, 2011.
- [239] A. Goldman, "How to make my blood boil Two recent papers comparing the structure of a hyperthermophilic protein with," *Current Biology*, pp. 1277–1279, 1995.
- [240] P. Strop and S. L. Mayo, "Contribution of surface salt bridges to protein stability," *Biochemistry*, vol. 39, no. 6, pp. 1251–1255, 2000.
- [241] A. Mitrovic, K. Flicker, G. Steinkellner, K. Gruber, C. Reisinger, G. Schirrmacher, A. Camattari, and A. Glieder, "Thermostability Improvement of Endoglucanase Cel7B from *Hypocrea pseudokoningii*," *Journal of Molecular Catalysis B: Enzymatic*, vol. 103, pp. 16–23, 2013.
- [242] H. Yu and H. Huang, "Engineering proteins for thermostability through rigidifying flexible sites," *Biotechnology Advances*, vol. 32, no. 2, pp. 308–315, 2014.
- [243] A. Karshikoff, L. Nilsson, and R. Ladenstein, "Rigidity versus flexibility: the dilemma of understanding protein thermal stability," *FEBS Journal*, vol. 282, no. 20, pp. 3899–3917, 2015.
- [244] G. R. Hemsworth, E. J. Taylor, R. Q. Kim, R. C. Gregory, S. J. Lewis, J. P. Turkenburg, A. Parkin, G. J. Davies, and P. H. Walton, "The copper active site of CBM33 polysaccharide oxygenases," *Journal of the American Chemical Society*, vol. 135, no. 16, pp. 6069–77, 2013.
- [245] R. Kittl, D. Kracher, D. Burgstaller, D. Haltrich, and R. Ludwig, "Production of four *Neurospora crassa* lytic polysaccharide monooxygenases in *Pichia pastoris* monitored by a fluorimetric assay," *Biotechnology for biofuels*, vol. 5, p. 79, oct 2012.
- [246] G. R. Hemsworth, G. J. Davies, and P. H. Walton, "Recent insights into copper-containing lytic polysaccharide mono-oxygenases," *Current opinion in structural biology*, vol. 23, no. 5, pp. 660–8, 2013.
- [247] I. Patel, D. Kracher, S. Ma, S. Garajova, M. Haon, C. B. Faulds, J.-G. Berrin, R. Ludwig, and E. Record, "Salt-responsive lytic polysaccharide monooxygenases from the mangrove fungus *Pestalotiopsis* sp. NCi6," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 108, 2016.

-
- [248] K. G. Sprenger, A. Choudhury, J. L. Kaar, and J. Pfaendtner, "The lytic polysaccharide monooxygenases ScLPMO10B and ScLPMO10C are stable in ionic liquids as determined by molecular simulations," *The Journal of Physical Chemistry B*, vol. 120, no. 16, pp. 3863–3872, 2016.
- [249] L. L. Leggio, D. Welner, and L. D. Maria, "A structural overview of GH61 proteins – fungal cellulose degrading polysaccharide monooxygenases," *Computational and Structural Biotechnology Journal*, vol. 2, sep 2012.
- [250] E. Ekwe, I. Morgenstern, A. Tsang, R. Storms, and J. Powlowski, "Non-hydrolytic cellulose active proteins: Research progress and potential application in biorefineries," *Industrial Biotechnology*, vol. 9, pp. 123–131, jun 2013.
- [251] M. Dimarogona, E. Topakas, and P. Christakopoulos, "Recalcitrant polysaccharide degradation by novel oxidative biocatalysts," *Applied microbiology and biotechnology*, vol. 97, no. 19, pp. 8455–65, 2013.
- [252] R. Daly and M. T. W. Hearn, "Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production," *Journal of molecular recognition*, vol. 18, no. 2, pp. 119–38, 2005.
- [253] R. S. Hegde and H. D. Bernstein, "The surprising complexity of signal sequences," *Trends in Biochemical Sciences*, vol. 31, no. 10, pp. 563–571, 2006.
- [254] C. P. Kubicek, M. Mikus, A. Schuster, M. Schmoll, and B. Seiboth, "Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*," *Biotechnology for biofuels*, vol. 2, p. 19, jan 2009.
- [255] M. Saloheimo, T. Nakari-Setälä, M. Tenkanen, and M. Penttilä, "cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast," *European journal of biochemistry*, vol. 249, no. 2, pp. 584–91, 1997.
- [256] J. M. Cregg, I. Tolstorukov, A. Kusari, J. Sunga, K. Madden, and T. Chappell, *Expression in the yeast Pichia pastoris*, vol. 463. Elsevier Inc., 1 ed., jan 2009.
- [257] H. Hohenblum, B. Gasser, M. Maurer, N. Borth, and D. Mattanovich, "Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*," *Biotechnology and Bioengineering*, vol. 85, no. 4, pp. 367–375, 2004.
- [258] A. Mellitzer, R. Weis, A. Glieder, and K. Flicker, "Expression of lignocellulolytic enzymes in *Pichia pastoris*," *Microbial cell factories*, vol. 11, no. 1, p. 61, 2012.
- [259] E. Nemeth, G. C. Preza, C. L. Jung, J. Kaplan, A. J. Waring, and T. Ganz, "The N-terminus of hepcidin is essential for its interaction with ferroportin: Structure-function study," *Blood*, vol. 107, no. 1, pp. 328–333, 2006.
- [260] M. Rigard, J. E. Bröms, A. Mosnier, M. Hologne, A. Martin, L. Lindgren, C. Punginelli, C. Lays, O. Walker, A. Charbit, P. Telouk, W. Conlan, L. Terradot, A. Sjöstedt, and T. Henry, "Francisella tularensis IgG belongs to a novel family of PAAR-like T6SS proteins and harbors a unique N-terminal extension required for virulence," *PloS one*, vol. 12, no. 9, p. e1005821, 2016.
- [261] J. L. Cereghino and J. M. Cregg, "Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*," *FEMS microbiology reviews*, vol. 24, no. 1, pp. 45–66, 2000.
- [262] M. Ahmad, M. Hirz, H. Pichler, and H. Schwab, "Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production," *Applied Microbiology and Biotechnology*, vol. 98, pp. 5301–5317, 2014.
- [263] N. Govindappa, M. Hanumanthappa, K. Venkatarangaiah, S. Periyasamy, S. Sreenivas, R. Soni, and K. Sastry, "A new signal sequence for recombinant protein secretion in *Pichia pastoris*," *Journal of microbiology and biotechnology*, vol. 24, no. 3, pp. 337–45, 2014.
- [264] Z. Forsberg, A. K. Mackenzie, M. Sørlie, Å. K. Røhr, R. Helland, A. S. Arvai, G. Vaaje-Kolstad, and V. G. H. Eijsink, "Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 8446–51, jun 2014.

REFERENCES

- [265] K. E. H. Frandsen, T. J. Simmons, P. Dupree, J.-C. N. Poulsen, G. R. Hemsworth, L. Ciano, E. M. Johnston, M. Tovborg, K. S. Johansen, P. von Freiesleben, L. Marmuse, S. Fort, S. Cottaz, H. Driguez, B. Henrissat, N. Lenfant, F. Tuna, A. Baldansuren, G. J. Davies, L. Lo Leggio, and P. H. Walton, "The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases," *Nature Chemical Biology*, no. February, 2016.
- [266] B. C. Pierce, J. Wittrup, J. Wichmann, and A. S. Meyer, "Oxidative cleavage and hydrolytic boosting of cellulose in soybean spent flakes by *Trichoderma reesei* Cel61A lytic polysaccharide monooxygenase," *Enzyme and Microbial Technology*, vol. 98, pp. 58–66, 2017.
- [267] L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner, and A. Glieder, "Deletion of the *pichia pastoris* ku70 homologue facilitates platform strain generation for gene expression and synthetic biology," *PLoS ONE*, vol. 7, no. 6, p. e39720, 2012.
- [268] J. Sanchis, L. Fernández, J. D. Carballeira, J. Drone, Y. Gumulya, H. Höbenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J.-P. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly, and M. T. Reetz, "Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates," *Applied microbiology and biotechnology*, vol. 81, no. 2, pp. 387–97, 2008.
- [269] D. G. Gibson, L. Young, R.-y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature methods*, vol. 6, no. 5, pp. 343–345, 2009.
- [270] J. Lin-Cereghino, W. W. Wong, S. Xiong, W. Giang, L. T. Luong, J. Vu, S. D. Johnson, and G. P. Lin-Cereghino, "Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*," *Biotechniques*, vol. 38, no. 1, pp. 44–48, 2005.
- [271] R. Weis, R. Luiten, W. Skranc, H. Schwab, M. Wubbolts, and A. Glieder, "Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena," *FEMS yeast research*, vol. 5, no. 2, pp. 179–89, 2004.
- [272] K. De Winter, D. Šimčíková, B. Schalck, L. Weignerová, H. Pelantova, W. Soetaert, T. Desmet, and V. Kren, "Chemoenzymatic synthesis of α -l-rhamnosides using recombinant α -l-rhamnosidase from *Aspergillus terreus*," *Bioresource Technology*, vol. 147, pp. 640–644, 2013.
- [273] U. K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, 1970.
- [274] L. W. Sun, Y. Zhao, L. P. Niu, R. Jiang, Y. Song, H. Feng, K. Feng, and C. Qi, "A rapid method for determining the concentration of recombinant protein secreted from *Pichia pastoris*," *Journal of Physics: Conference Series*, vol. 276, p. 012144, 2011.
- [275] I. Stals, S. Karkehabadi, S. Kim, M. Ward, A. van Landschoot, B. Devreese, and M. Sandgren, "High resolution crystal structure of the Endo-N-Acetyl-beta-D-glucosaminidase responsible for the Deglycosylation of *Hypocrea jecorina* cellulases," *PLoS ONE*, vol. 7, no. 7, 2012.
- [276] T. M. Wood, "Preparation of crystalline, amorphous, and dyed cellulase substrates," *Methods in Enzymology*, vol. 160, pp. 19–25, 1988.
- [277] T. F. Clarke and P. L. Clark, "Rare codons cluster," *PLoS one*, vol. 3, no. 10, p. e3412, 2008.
- [278] G. R. Buettner and P. G. Czapski, "Ascorbate autoxidation in the presence of iron and copper chelates," *Free Radical Research*, vol. 1, no. 6, pp. 349–353, 1986.
- [279] S. Freimund and S. Köpper, "The composition of 2-keto aldoses in organic solvents as determined by NMR spectroscopy," *Carbohydrate Research*, vol. 339, no. 2, pp. 217–220, 2004.
- [280] B. Westereng, M. Ø. Arntzen, F. L. Aachmann, A. Várnai, V. G. Eijsink, and J. W. Agger, "Simultaneous analysis of C1 and C4 oxidized oligosaccharides, the products of lytic polysaccharide monooxygenases (LPMOs) acting on cellulose," *Journal of Chromatography A*, vol. 1445, pp. 46–54, 2016.
- [281] S. K. Kračun, J. Schückel, B. Westereng, L. G. Thygesen, R. N. Monrad, V. G. H. Eijsink, and W. G. T. Willats, "A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes," *Biotechnology for Biofuels*, vol. 8, no. 70, 2015.

- [282] M.-J. Yu, S.-H. Yoon, and Y.-W. Kim, "Overproduction and characterization of a lytic polysaccharide monoxygenase in *Bacillus subtilis* using an assay based on ascorbate consumption," *Enzyme and Microbial Technology*, vol. 93-94, pp. 150–165, 2016.
- [283] C. M. Johnson, "Differential scanning calorimetry as a tool for protein folding and stability.," *Archives of biochemistry and biophysics*, vol. 531, no. 1-2, pp. 100–9, 2013.
- [284] N. J. Greenfield, "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions," *Nature protocols*, vol. 1, no. 6, pp. 2527–2535, 2006.
- [285] F. H. Niesen, H. Berglund, and M. Vedadi, "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.," *Nature protocols*, vol. 2, pp. 2212–21, jan 2007.
- [286] U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, and P. Nordlund, "Thermofluor-based high-throughput stability optimization of proteins for structural studies," *Analytical Biochemistry*, vol. 357, no. 2, pp. 289–298, 2006.
- [287] M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, and D. Breitsprecher, "MicroScale Thermophoresis: Interaction analysis and beyond," *Journal of Molecular Structure*, vol. 1077, pp. 101–113, 2014.
- [288] C. G. Alexander, R. Wanner, C. M. Johnson, D. Breitsprecher, G. Winter, S. Duhr, P. Baaske, and N. Ferguson, "Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1844, no. 12, pp. 2241–2250, 2014.
- [289] J. J. Lavinder, S. B. Hari, B. J. Sullivan, and T. J. Magliery, "High-throughput thermal scanning: A general, rapid dye-binding thermal shift screen for protein engineering," *Journal of the American Chemical Society*, vol. 131, no. 11, pp. 3794–3795, 2009.
- [290] M. W. Pantoliano, E. C. Petrella, J. D. Kwansnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, and F. Salemme, "High-density miniaturized thermal shift assays as a general strategy for drug discovery," *Journal of biomolecular screening*, vol. 6, no. 6, pp. 429–440, 2001.
- [291] G. Henriksson, G. Johansson, and G. Pettersson, "A critical review of cellobiose dehydrogenases," *Journal of Biotechnology*, vol. 78, no. 2, pp. 93–113, 2000.
- [292] U. Westermark and K.-E. Eriksson, "Purification and properties of cellobiose: Quinone oxidoreductase from *Spootrichum pulverulentum*," *Acta Chemica Scandinavica B*, vol. 29, pp. 419–424, 1975.
- [293] S. Ramachandran, P. Fontanille, A. Pandey, and C. Larroche, "Gluconic Acid : properties , applications and microbial production," *Food Technol. Biotechnol.*, vol. 44, no. 2, pp. 185–195, 2006.
- [294] C. Corradini, A. Cavazza, and C. Bignardi, "High-performance anion-exchange chromatography coupled with pulsed electrochemical detection as a powerful tool to evaluate carbohydrates of food interest: Principles and applications," *International Journal of Carbohydrate Chemistry*, vol. 2012, pp. 1–13, 2012.
- [295] G. Vaaje-Kolstad, S. J. Horn, M. Sørli, and V. G. H. Eijsink, "The chitinolytic machinery of *Serratia marcescens* - A model system for enzymatic degradation of recalcitrant polysaccharides," *FEBS Journal*, vol. 280, no. 13, pp. 3028–3049, 2013.
- [296] D. J. Warren, "Preparation of highly efficient electrocompetent *Escherichia coli* using glycerol/mannitol density step centrifugation," *Analytical Biochemistry*, vol. 413, no. 2, pp. 206–207, 2011.
- [297] E. Celik and P. Calik, "Production of recombinant proteins by yeast cells.," *Biotechnology advances*, vol. 30, no. 5, pp. 1108–1118, 2012.
- [298] G. Gellissen and C. P. Hollenberg, "Application of yeasts in gene expression studies : a comparison of *Saccharomyces cerevisiae*, *Hansenula polymorpha* and *Kluyveromyces lactis* - a review," *Gene*, vol. 190, pp. 87–97, 1997.
- [299] T. Bulter, M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer, and F. H. Arnold, "Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution," *Applied and Environmental Microbiology*, vol. 69, no. 2, pp. 987–995, 2003.

REFERENCES

- [300] D. Weinacker, C. Rabert, A. B. Zepeda, C. A. Figueroa, A. Pessoa, and J. G. Fariás, "Applications of recombinant *Pichia pastoris* in the healthcare industry," *Brazilian Journal of Microbiology*, vol. 44, no. 4, pp. 1043–1048, 2013.
- [301] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, "Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production.," *Biotechnology and bio-engineering*, vol. 107, no. 2, pp. 357–68, 2010.
- [302] M. Hasslacher, M. Schall, M. Hayn, R. Bona, K. Rumbold, J. Lückl, H. Griengl, S. D. Kohlwein, and H. Schwab, "High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts," *Protein Expression and Purification*, vol. 11, no. 1, pp. 61–71, 1997.
- [303] A. S. Xiong, Q. H. Yao, R. H. Peng, P. L. Han, Z. M. Cheng, and Y. Li, "High level expression of a recombinant acid phytase gene in *Pichia pastoris*," *Journal of Applied Microbiology*, vol. 98, no. 2, pp. 418–428, 2005.
- [304] L. Fernández, N. Jiao, P. Soni, Y. Gumulya, L. G. D. E. Oliveira, M. T. Reetz, and L. G. de Oliveira, "An efficient method for mutant library creation in *Pichia pastoris* useful in directed evolution," *Biocatalysis and Biotransformation*, vol. 28, no. 2, pp. 122–129, 2010.
- [305] J. H. Choi and S. Y. Lee, "Secretory and extracellular production of recombinant proteins using *Escherichia coli*," *Applied microbiology and biotechnology*, vol. 64, no. 5, pp. 625–35, 2004.
- [306] Z. Forsberg, A. K. Röhr, S. Mekasha, K. K. Andersson, V. G. H. Eijsink, G. Vaaje-Kolstad, and M. Sør-lie, "Comparative Study of Two Chitin-Active and Two Cellulose-Active AA10-Type Lytic Polysaccharide Monooxygenases.," *Biochemistry*, vol. 53, pp. 1647–56, mar 2014.
- [307] E. Wong, G. Vaaje-Kolstad, A. Ghosh, R. Hurtado-Guerrero, P. V. Konarev, A. F. M. Ibrahim, D. I. Svergun, V. G. H. Eijsink, N. S. Chatterjee, and D. M. F. van Aalten, "The *Vibrio cholerae* colonization factor GbpA possesses a modular structure that governs binding to different host surfaces," *PLoS Pathogens*, vol. 8, no. 1, pp. 1–12, 2012.
- [308] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.
- [309] S. S. Ghatge, A. a. Telke, T. R. Waghmode, Y. Lee, K.-W. Lee, D.-B. Oh, H.-D. Shin, and S.-W. Kim, "Multifunctional cellulolytic auxiliary activity protein HcAA10-2 from *Hahella chejuensis* enhances enzymatic hydrolysis of crystalline cellulose.," *Applied microbiology and biotechnology*, oct 2014.
- [310] J. Quan and J. Tian, "Circular polymerase extension cloning of complex gene libraries and pathways," *PLoS ONE*, vol. 4, no. 7, p. e6441, 2009.
- [311] R. Zou, K. Zhou, G. Stephanopoulos, and H. P. Too, "Combinatorial engineering of 1-deoxy-D-xylulose 5-phosphate pathway using cross-lapping in vitro assembly (CLIVA) method," *PLoS ONE*, vol. 8, no. 11, pp. 1–12, 2013.
- [312] F. Baneyx and M. Mujacic, "Recombinant protein folding and misfolding in *Escherichia coli*," *Nature biotechnology*, vol. 22, no. 11, pp. 1399–408, 2004.
- [313] J. K. Locker and G. Griffiths, "An unconventional role for cytoplasmic disulfide bonds in vaccinia virus proteins," *Journal of Cell Biology*, vol. 144, no. 2, pp. 267–279, 1999.
- [314] K. Tartof and C. Hobbs, "Improved media for growing plasmid and cosmid clones," *Focus*, vol. 9, no. 2, p. 10, 1987.
- [315] R. S. Islam, D. Tisi, M. S. Levy, and G. J. Lye, "Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design," *Biotechnology progress*, vol. 23, pp. 785–793, 2007.
- [316] M. Osadska, H. Bonkova, J. Krahulec, S. Stuchlik, and J. Turna, "Optimization of expression of untagged and histidine-tagged human recombinant thrombin precursors in *Escherichia coli*," *Applied Microbiology and Biotechnology*, vol. 98, no. 22, pp. 9259–9270, 2014.
- [317] D. Aerts, T. Verhaeghe, M. De Mey, T. Desmet, and W. Soetaert, "A constitutive expression system for high-throughput screening," *Engineering in Life Sciences*, vol. 11, no. 1, pp. 10–19, 2011.

-
- [318] E. A. Span and M. A. Marletta, "The framework of polysaccharide monooxygenase structure and chemistry," *Current Opinion in Structural Biology*, vol. 35, pp. 93–99, 2015.
- [319] S. Jung, Y. Song, H. M. Kim, and H.-J. Bae, "Enhanced lignocellulosic biomass hydrolysis by oxidative lytic polysaccharide monooxygenases (LPMOs) GH61 from *Gloeophyllum trabeum*," *Enzyme and Microbial Technology*, vol. 77, pp. 38–45, 2015.
- [320] C. Roodveldt, A. Aharoni, and D. S. Tawfik, "Directed evolution of proteins for heterologous expression and stability," *Current Opinion in Structural Biology*, vol. 15, no. 1 SPEC. ISS., pp. 50–56, 2005.
- [321] C. J. Yeoman, Y. Han, D. Dodd, C. M. Schroeder, R. I. Mackie, and I. K. O. Cann, *Thermostable enzymes as biocatalysts in the biofuel industry*, vol. 70. Elsevier Inc., 1 ed., 2010.
- [322] A. Bhalla, N. Bansal, S. Kumar, K. M. Bischoff, and R. K. Sani, "Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes," *Bioresource Technology*, vol. 128, pp. 751–759, 2013.
- [323] T. J. Magliery, J. J. Lavinder, and B. J. Sullivan, "Protein stability by number: High-throughput and statistical approaches to one of protein science's most difficult problems," *Current Opinion in Chemical Biology*, vol. 15, no. 3, pp. 443–451, 2011.
- [324] T. W. Johannes, R. D. Woodyer, and H. Zhao, "Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration," *Society*, vol. 71, no. 10, pp. 5728–5734, 2005.
- [325] H. Zhao and F. H. Arnold, "Directed evolution converts subtilisin E into a functional equivalent of thermitase," *Protein engineering*, vol. 12, no. 1, pp. 47–53, 1999.
- [326] H. Uchiyama, T. Inaoka, T. Ohkuma-Soyejima, H. Togame, Y. Shibanaka, T. Yoshimoto, and T. Kokubo, "Directed evolution to improve the thermostability of prolyl endopeptidase," *J Biochem*, vol. 128, no. 3, pp. 441–447, 2000.
- [327] H. Komeda, N. Ishikawa, and Y. Asano, "Enhancement of the thermostability and catalytic activity of D-stereospecific amino-acid amidase from *Ochrobactrum anthropi* SV3 by directed evolution," *Journal of Molecular Catalysis B: Enzymatic*, vol. 21, no. 4-6, pp. 283–290, 2003.
- [328] T. U. Consortium, "UniProt: A hub for protein information," *Nucleic Acids Research*, vol. 43, no. D1, pp. D204–D212, 2015.
- [329] S. Parthasarathy and M. R. Murthy, "Protein thermal stability: insights from atomic displacement parameters (B values)," *Protein engineering*, vol. 13, no. 1, pp. 9–13, 2000.
- [330] R. Ulbrich-Hofmann, U. Arnold, and J. Mansfeld, "The concept of the unfolding region for approaching the mechanisms of enzyme stabilization," *Journal of Molecular Catalysis B: Enzymatic*, vol. 7, no. 1-4, pp. 125–131, 1999.
- [331] J. K. Blum, M. D. Ricketts, and A. S. Bommaris, "Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method," *Journal of Biotechnology*, vol. 160, no. 3-4, pp. 214–221, 2012.
- [332] T. Lazaridis, I. Lee, and M. Karplus, "Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin," *Protein science : a publication of the Protein Society*, vol. 6, pp. 2589–2605, 1997.
- [333] R. Jaenicke and G. Böhm, "The stability of proteins in extreme environments," *Current Opinion in Structural Biology*, vol. 8, no. 6, pp. 738–748, 1998.
- [334] M. C. Brown, D. Verma, C. Russell, D. J. Jacobs, and D. R. Livesay, "A case study comparing quantitative stability/flexibility relationships across five metallo- β -lactamases highlighting differences within NDM-1," *Methods in Molecular Biology*, vol. 1084, pp. 227–238, 2014.
- [335] D. J. Kelleher and R. Gilmore, "An evolving view of the eukaryotic oligosaccharyltransferase," *Glycobiology*, vol. 16, no. 4, pp. 47–62, 2006.
- [336] J. L. Price, E. K. Culyba, W. Chen, A. N. Murray, S. R. Hanson, C. H. Wong, E. T. Powers, and J. W. Kelly, "N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability," *Biopolymers*, vol. 98, no. 3, pp. 195–211, 2012.

REFERENCES

- [337] S. R. Hanson, E. K. Culyba, T.-L. Hsu, C.-H. Wong, J. W. Kelly, and E. T. Powers, "The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 9, pp. 3131–6, 2009.
- [338] S. E. O'Connor, J. Pohlmann, B. Imperiali, I. Saskiawan, and K. Yamamoto, "Probing the effect of the outer saccharide residues of N-linked glycans on peptide conformation," *Journal of the American Chemical Society*, vol. 123, no. 25, pp. 6187–6188, 2001.
- [339] C. J. Bosques and B. Imperiali, "The interplay of glycosylation and disulfide formation influences fibrillization in a prion protein fragment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 13, pp. 7593–8, 2003.
- [340] P. M. Rudd, T. Elliott, P. Cresswell, I. A. Wilson, and R. A. Dwek, "Glycosylation and the immune system," *Science*, vol. 291, no. 5512, pp. 2370–2376, 2001.
- [341] C. Wang, M. Eufemi, C. Turano, and A. Giartosio, "Influence of the carbohydrate moiety on the stability of glycoproteins," *Biochemistry*, vol. 35, no. 23, pp. 7299–7307, 1996.
- [342] M. A. Recny, M. A. Luther, M. H. Knoppers, E. A. Neidhardt, S. S. Khandekar, M. F. Concino, P. A. Schimke, M. A. Francis, U. Moebius, B. B. Reinhold, V. N. Reinhold, and E. L. Reinherz, "N-glycosylation is required for human CD2 immunoadhesion functions," *Journal of Biological Chemistry*, vol. 267, no. 31, pp. 22428–22432, 1992.
- [343] E. Rodriguez, Z. a. Wood, P. a. Karplus, and X. G. Lei, "Site-directed mutagenesis improves catalytic efficiency and thermostability of Escherichia coli pH 2.5 acid phosphatase/phytase expressed in Pichia pastoris," *Archives of biochemistry and biophysics*, vol. 382, no. 1, pp. 105–112, 2000.
- [344] X. Yin, D. Hu, J. F. Li, Y. He, T. D. Zhu, and M. C. Wu, "Contribution of disulfide bridges to the thermostability of a type a feruloyl esterase from Aspergillus usamii," *PLoS ONE*, vol. 10, no. 5, p. e0126864, 2015.
- [345] J. C. Bardwell, K. McGovern, and J. Beckwith, "Identification of a protein required for disulfide bond formation in vivo," *Cell*, vol. 67, no. 3, pp. 581–589, 1991.
- [346] Y. C. Ai, S. Zhang, and D. B. Wilson, "Positional expression effects of cysteine mutations in the Thermobifida fusca cellulase Cel6A and Cel6B catalytic domains," *Enzyme and Microbial Technology*, vol. 32, no. 2, pp. 331–336, 2003.
- [347] I. Fremaux, S. Mazères, A. Brisson-Lougarre, M. Arnaud, C. Ladurantie, and D. Fournier, "Improvement of Drosophila acetylcholinesterase stability by elimination of a free cysteine," *BMC biochemistry*, vol. 3, p. 21, 2002.
- [348] O. Oyetade, G. Oyeleke, B. Adegoke, and A. Akintunde, "Stability studies on ascorbic acid (Vitamin C) from different sources," *IOSR Journal of Applied Chemistry*, vol. 2, no. 4, pp. 20–24, 2012.
- [349] R. Griessler, S. D'Auria, F. Tanfani, and B. Nidetzky, "Thermal denaturation pathway of starch phosphorylase from Corynebacterium callunae: oxyanion binding provides the glue that efficiently stabilizes the dimer structure of the protein," *Protein science : a publication of the Protein Society*, vol. 9, no. 6, pp. 1149–61, 2000.
- [350] A. Lejeune, M. Vanhove, J. Lamotte-Brasseur, R. H. Pain, J. M. Frère, and A. Matagne, "Quantitative analysis of the stabilization by substrate of Staphylococcus aureus PC1 β -lactamase," *Chemistry and Biology*, vol. 8, no. 8, pp. 831–842, 2001.
- [351] O. Turunen, M. Vuorio, F. Fenel, and M. Leisola, "Engineering of multiple arginines into the Ser/Thr surface of Trichoderma reesei endo-1,4-beta-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH," *Protein engineering*, vol. 15, no. 2, pp. 141–145, 2002.
- [352] A. Várnai, M. R. Mäkelä, D. T. Djajadi, J. Rahikainen, A. Hatakka, and L. Viikari, *Carbohydrate-binding modules of fungal cellulases. Occurrence in nature, function, and relevance in industrial biomass conversion*, vol. 88, 2014.
- [353] O. Shoseyov, Z. Shani, and I. Levy, "Carbohydrate binding modules: Biochemical properties and novel applications," *Microbiology and Molecular Biology Reviews*, vol. 70, no. 2, pp. 283–295, 2006.

-
- [354] C. Hervé, A. Rogowski, A. W. Blake, S. E. Marcus, H. J. Gilbert, and J. P. Knox, "Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 34, pp. 15293–15298, 2010.
- [355] A. Várnai, M. Siika-Aho, and L. Viikari, "Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs," *Biotechnology for biofuels*, vol. 6, no. 1, p. 30, 2013.
- [356] N. Din, N. R. Gilkes, B. Tekant, R. C. Miller, A. J. Warren, and D. G. Kilburn, "Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase," *Nature Biotechnology*, vol. 9, no. 1, pp. 1096–1099, 1991.
- [357] B. L. Mello and I. Polikarpov, "Family 1 carbohydrate binding-modules enhance saccharification rates," *AMB Express*, vol. 4, no. 1, p. 36, 2014.
- [358] C. M. Payne, M. G. Resch, L. Chen, M. F. Crowley, M. E. Himmel, L. E. Taylor, M. Sandgren, J. Ståhlberg, I. Stals, Z. Tan, and G. T. Beckham, "Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 36, pp. 14646–51, 2013.
- [359] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [360] F. Sterpone and S. Melchionna, "Thermophilic proteins: insight and perspective from in silico experiments," *Chemical Society reviews*, vol. 41, no. 5, pp. 1665–76, 2012.
- [361] J. S. Klein, S. Jiang, R. P. Galimidi, J. R. Keeffe, P. J. Bjorkman, and L. Regan, "Design and characterization of structured protein linkers with differing flexibilities," *Protein Engineering, Design and Selection*, vol. 27, no. 10, pp. 325–330, 2014.
- [362] F. Fenel, M. Leisola, J. Jänis, and O. Turunen, "A de novo designed N-terminal disulphide bridge stabilizes the *Trichoderma reesei* endo-1,4- β -xylanase II," *Journal of Biotechnology*, vol. 108, no. 2, pp. 137–143, 2004.
- [363] A. B. Boraston, D. N. Bolam, H. J. Gilbert, and G. J. Davies, "Carbohydrate-binding modules: fine-tuning polysaccharide recognition," *The Biochemical journal*, vol. 382, no. Pt 3, pp. 769–781, 2004.
- [364] H. Lu, H. Luo, P. Shi, H. Huang, K. Meng, P. Yang, and B. Yao, "A novel thermophilic endo-beta-1,4-mannanase from *Aspergillus nidulans* XZ3: Functional roles of carbohydrate-binding module and Thr/Ser-rich linker region," *Applied Microbiology and Biotechnology*, vol. 98, no. 5, pp. 2155–2163, 2014.
- [365] T. A. Pham, J. G. Berrin, E. Record, K. A. To, and J. C. Sigoillot, "Hydrolysis of softwood by *Aspergillus* mannanase: Role of a carbohydrate-binding module," *Journal of Biotechnology*, vol. 148, no. 4, pp. 163–170, 2010.
- [366] P. Verjans, E. Dornez, M. Segers, S. Van Campenhout, K. Bernaerts, T. Beliën, J. A. Delcour, and C. M. Courtin, "Truncated derivatives of a multidomain thermophilic glycosyl hydrolase family 10 xylanase from *Thermotoga maritima* reveal structure related activity profiles and substrate hydrolysis patterns," *Journal of Biotechnology*, vol. 145, no. 2, pp. 160–167, 2010.
- [367] M. Couturier, J. Feliu, M. Haon, D. Navarro, L. Lesage-Meessen, P. M. Coutinho, and J.-G. Berrin, "A thermostable GH45 endoglucanase from yeast: impact of its atypical multimodularity on activity," *Microbial cell factories*, vol. 10, no. December, p. 103, 2011.
- [368] Y. Wang, H. Yuan, J. Wang, and Z. Yu, "Truncation of the cellulose binding domain improved thermal stability of endo-beta-1,4-glucanase from *Bacillus subtilis* JA18," *Bioresource Technology*, vol. 100, no. 1, pp. 345–349, 2009.
- [369] H. Ravalason, I. Herpoël-Gimbert, E. Record, F. Bertaud, S. Grisel, S. de Weert, C. A. M. J. J. van den Hondel, M. Asther, M. Petit-Conil, and J. C. Sigoillot, "Fusion of a family 1 carbohydrate binding module of *Aspergillus niger* to the *Pycnoporus cinnabarinus* laccase for efficient softwood kraft pulp biobleaching," *Journal of Biotechnology*, vol. 142, no. 3–4, pp. 220–226, 2009.

REFERENCES

- [370] R. Cheng, J. Chen, X. Yu, Y. Wang, S. Wang, and J. Zhang, "Recombinant production and characterization of full-length and truncated β -1,3-glucanase PglA from *Paenibacillus* sp. S09.," *BMC biotechnology*, vol. 13, p. 105, 2013.
- [371] V. O. A. Pellegrini, V. I. Serpa, A. S. Godoy, C. M. Camilo, A. Bernardes, C. A. Rezende, N. P. Junior, J. P. L. Franco Cairo, F. M. Squina, and I. Polikarpov, "Recombinant *Trichoderma harzianum* endoglucanase I (Cel7B) is a highly acidic and promiscuous carbohydrate-active enzyme," *Applied Microbiology and Biotechnology*, vol. 99, no. 22, pp. 9591–9604, 2015.
- [372] J. Hall, G. W. Black, L. M. Ferreira, S. J. Millward-Sadler, B. R. Ali, G. P. Hazlewood, and H. J. Gilbert, "The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel.," *The Biochemical journal*, vol. 309, pp. 749–756, 1995.
- [373] L. I. Crouch, A. Labourel, P. H. Walton, G. J. Davies, and H. J. Gilbert, "The contribution of non-catalytic carbohydrate binding modules to the activity lytic polysaccharide monooxygenases," *Journal of Biological Chemistry*, vol. 291, no. 14, pp. 7439–7449, 2016.
- [374] Z. Forsberg, G. Vaaje-Kolstad, B. Westereng, A. C. Bunæs, Y. Stenstrøm, A. MacKenzie, M. Sørli, S. J. Horn, and V. G. H. Eijsink, "Cleavage of cellulose by a CBM33 protein.," *Protein science : a publication of the Protein Society*, vol. 20, pp. 1479–1483, sep 2011.
- [375] Y. Yasuda, S. Ikeda, H. Sakai, T. Tsukuba, K. Okamoto, K. Nishishita, A. Akamine, Y. Kato, and K. Yamamoto, "Role of N-glycosylation in cathepsin E. A comparative study of cathepsin E with distinct N-linked oligosaccharides and its nonglycosylated mutant," *European Journal of Biochemistry*, vol. 266, no. 2, pp. 383–391, 1999.
- [376] G. Li, Q. Yan, A. Nita-Lazar, R. S. Haltiwanger, and W. J. Lennarz, "Studies on the N-glycosylation of the subunits of oligosaccharyl transferase in *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 280, no. 3, pp. 1864–1871, 2005.
- [377] S. P. Voutilainen, P. G. Murray, M. G. Tuohy, and A. Koivula, "Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity," *Protein Engineering Design and Selection*, vol. 23, no. 2, pp. 69–79, 2010.
- [378] J. Jafari-Aghdam, K. Khajeh, B. Ranjbar, and M. Nemat-Gorgani, "Deglycosylation of glucoamylase from *Aspergillus niger*: effects on structure, activity and stability.," *Biochimica et biophysica acta*, vol. 1750, no. 1, pp. 61–68, 2005.
- [379] S. H. Eriksen, B. Jensen, and J. Olsen, "Effect of N-linked glycosylation on secretion, activity, and stability of alpha-amylase from *Aspergillus oryzae*," *Current microbiology*, vol. 37, no. 2, pp. 117–22, 1998.
- [380] S. E. Clark, E. H. Muslin, and C. A. Henson, "Effect of adding and removing N-glycosylation recognition sites on the thermostability of barley alpha-glucosidase," *Protein Engineering, Design and Selection*, vol. 17, no. 3, pp. 245–249, 2004.
- [381] R. Godoy-Ruiz, R. Perez-Jimenez, B. Ibarra-Molero, and J. M. Sanchez-Ruiz, "A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization.," *Biophysical journal*, vol. 89, no. 5, pp. 3320–31, 2005.
- [382] A. Mitrovic, K. Flicker, G. Steinkellner, K. Gruber, C. Reisinger, G. Schirrmacher, A. Camattari, and A. Glieder, "Thermostability improvement of endoglucanase Cel7B from *Hypocrea pseudokoningii*," *Journal of Molecular Catalysis B: Enzymatic*, vol. 103, pp. 16–23, 2014.
- [383] C. R. Robinson and R. T. Sauer, "Striking stabilization of Arc repressor by an engineered disulfide bond," *Biochemistry*, vol. 39, no. 40, pp. 12494–12502, 2000.
- [384] R. Guzzi, L. Andolfi, S. Cannistraro, M. P. Verbeet, G. W. Canters, and L. Sportelli, "Thermal stability of wild type and disulfide bridge containing mutant of poplar plastocyanin," *Biophysical Chemistry*, vol. 112, no. 1, pp. 35–43, 2004.
- [385] O. R. Siadat, A. Lougarre, L. Lamouroux, C. Ladurantie, and D. Fournier, "The effect of engineered disulfide bonds on the stability of *Drosophila melanogaster* acetylcholinesterase.," *BMC biochemistry*, vol. 7, no. 1, p. 12, 2006.

-
- [386] D. B. Craig and A. A. Dombkowski, "Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins.," *BMC bioinformatics*, vol. 14, no. 1, p. 346, 2013.
- [387] V. S. Dani, C. Ramakrishnan, and R. Varadarajan, "MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins.," *Protein engineering*, vol. 16, no. 3, pp. 187–193, 2003.
- [388] A. Ceroni, A. Passerini, A. Vullo, and P. Frasconi, "Disulfind: A disulfide bonding state and cysteine connectivity prediction server," *Nucleic Acids Research*, vol. 34, pp. W177–W181, 2006.
- [389] R. Sowdhamini, N. Srinivasn, B. Shoichet, V. Santi, Daniel, C. Ramakrishnan, and P. Balam, "Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis," *Protein engineering*, vol. 3, no. 2, pp. 95–103, 1989.
- [390] J. Laimer, H. Hofer, M. Fritz, S. Wegenkittl, and P. Lackner, "MAESTRO - multi agent stability prediction upon point mutations," *BMC Bioinformatics*, vol. 16, no. 1, p. 116, 2015.
- [391] A. a. Dombkowski, "Disulfide by Design: A computational method for the rational design of disulfide bonds in proteins," *Bioinformatics*, vol. 19, no. 14, pp. 1852–1853, 2003.
- [392] Q. A. T. Le, J. C. Joo, Y. J. Yoo, and Y. H. Kim, "Development of thermostable *Candida antarctica* lipase B through novel in silico design of disulfide bridge," *Biotechnology and Bioengineering*, vol. 109, no. 4, pp. 867–876, 2012.
- [393] O. B. Ptitsyn, R. H. Pain, G. V. Semisotnov, E. Zerovnik, and O. L. Razgulyaer, "Evidence for a molten globule state as a general intermediates in protein folding," *FEBS Letters*, vol. 262, no. 1, pp. 20–24, 1990.
- [394] K. Chakraborty, S. Thakurela, R. S. Prajapati, S. Indu, P. S. S. Ali, C. Ramakrishnan, and R. Varadarajan, "Protein stabilization by introduction of cross-strand disulfides," *Biochemistry*, vol. 44, no. 44, pp. 14638–14646, 2005.
- [395] O. Khakshoor and J. S. Nowick, "Use of Disulfide "Staples" to Stabilize beta-Sheet Quaternary Structure," *Organic Letters (ACS Publications)*, vol. 11, no. 14, pp. 3000–3003, 2009.
- [396] C. B. Anfinsen, "Principles that Govern the Folding of Protein Chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.
- [397] M. Lehmann, L. Pasamontes, S. F. Lassen, and M. Wyss, "The consensus concept for thermostability engineering of proteins," *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, vol. 1543, no. 2, pp. 408–415, 2000.
- [398] M. Lehmann and M. Wyss, "Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution," *Current Opinion in Biotechnology*, vol. 12, no. 4, pp. 371–375, 2001.
- [399] E. Ohage and B. Steipe, "Intrabody construction and expression. I. The critical role of VL domain stability," *Journal of molecular biology*, vol. 291, pp. 1119–1128, 1999.
- [400] a. Knappik, L. Ge, a. Honegger, P. Pack, M. Fischer, G. Wellnhofer, a. Hoess, J. Wölle, a. Plückthun, and B. Virnekäs, "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides.," *Journal of molecular biology*, vol. 296, no. 1, pp. 57–86, 2000.
- [401] P. Forrer, H. K. Binz, M. T. Stumpp, and A. Pluckthun, "Consensus design of repeat proteins," *Chem-BioChem*, vol. 5, no. 2, pp. 183–189, 2004.
- [402] M. Dai, H. E. Fisher, J. Temirov, C. Kiss, M. E. Phipps, P. Pavlik, J. H. Werner, and A. R. M. Bradbury, "The creation of a novel fluorescent protein by guided consensus engineering," *Protein Engineering, Design and Selection*, vol. 20, no. 2, pp. 69–79, 2007.
- [403] B. J. Sullivan, V. Durani, and T. J. Magliery, "Trioephosphate isomerase by consensus design: Dramatic differences in physical properties and activity of related variants," *Journal of Molecular Biology*, vol. 413, no. 1, pp. 195–208, 2011.

REFERENCES

- [404] S. A. Jacobs, M. D. Diem, J. Luo, A. Teplyakov, G. Obmolova, T. Malia, G. L. Gilliland, and K. T. Oneil, "Design of novel FN3 domains with high stability by a consensus sequence approach," *Protein Engineering, Design and Selection*, vol. 25, no. 3, pp. 107–117, 2012.
- [405] A. M. Loening, T. D. Fenn, A. M. Wu, and S. S. Gambhir, "Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output," *Protein Engineering, Design and Selection*, vol. 19, no. 9, pp. 391–400, 2006.
- [406] B. J. Sullivan, T. Nguyen, V. Durani, D. Mathur, S. Rojas, M. Thomas, T. Syu, and T. J. Magliery, "Stabilizing proteins from sequence statistics: The interplay of conservation and correlation in triosephosphate isomerase stability," *Journal of Molecular Biology*, vol. 420, no. 4-5, pp. 384–399, 2012.
- [407] K. Tamura, G. Stecher, D. Peterson, A. Filipowski, and S. Kumar, "MEGA6: Molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [408] X. Robert and P. Gouet, "Deciphering key features in protein structures with the new ENDscript server," *Nucleic Acids Research*, vol. 42, no. W1, pp. 320–324, 2014.
- [409] M. Tanghe, B. Danneels, A. Camattari, A. Glieder, I. Vandenberghe, B. Devreese, I. Stals, and T. Desmet, "Recombinant Expression of Trichoderma reesei Cel61A in Pichia pastoris: Optimizing Yield and N-terminal Processing," *Molecular Biotechnology*, vol. 57, no. 11-12, pp. 1010–1017, 2015.
- [410] M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. G. M. van Loon, and M. Wyss, "The consensus concept for thermostability engineering of proteins: further proof of concept," *Protein engineering*, vol. 15, no. 5, pp. 403–411, 2002.
- [411] C. Jäckel, J. D. Bloom, P. Kast, F. H. Arnold, and D. Hilvert, "Consensus protein design without phylogenetic bias," *Journal of Molecular Biology*, vol. 399, no. 4, pp. 541–546, 2010.
- [412] M. Socolich, S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner, and R. Ranganathan, "Evolutionary information for specifying a protein fold," *Nature*, vol. 437, no. 7058, pp. 512–518, 2005.
- [413] W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe, and R. Ranganathan, "Natural-like function in artificial WW domains," *Nature*, vol. 437, no. 7058, pp. 579–583, 2005.
- [414] a. D. Nagi and L. Regan, "An inverse correlation between loop length and stability in a four-helix-bundle protein," *Folding & design*, vol. 2, no. 1, pp. 67–75, 1997.
- [415] C. Budiman, Y. Koga, K. Takano, and S. Kanaya, "FK506-binding protein 22 from a psychrophilic bacterium, a cold shock-inducible peptidyl prolyl isomerase with the ability to assist in protein folding," *International Journal of Molecular Sciences*, vol. 12, no. 8, pp. 5261–5284, 2011.
- [416] A. Pey, T. Majtan, J. Sanchez-Ruiz, and J. Kraus, "Human cystathionine beta-synthase (CBS) contains two classes of binding sites for S-adenosyl-L-methionine (SAM): complex regulation of CBS activity and stability by SAM," *Biochemical Journal*, vol. 121, pp. 109–121, 2012.
- [417] F. Abate, E. Malito, R. Cozzi, P. Lo Surdo, D. Maione, and M. J. Bottomley, "Apo, Zn²⁺-bound and Mn²⁺-bound structures reveal ligand-binding properties of SitA from the pathogen Staphylococcus pseudintermedius," *Bioscience Reports*, vol. 34, no. 6, pp. 743–758, 2014.
- [418] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, "The protein folding problem," *Annual Review of Biophysics*, vol. 37, pp. 289–316, 2008.
- [419] H. Durand, P. Soucaille, and G. Tiraby, "Comparative study of cellulases and hemicellulases from four fungi: mesophiles Trichoderma reesei and Penicillium sp. and thermophiles Thielavia terrestris and Sporotrichum thermophilum," *Enzyme Microbial Technology*, vol. 6, pp. 175–180, 1984.
- [420] R. Maheshwari, G. Bharadwaj, and M. K. Bhat, "Thermophilic Fungi : Their Physiology and Enzymes," *Microbiology and Molecular Biology Reviews*, vol. 64, no. 3, pp. 461–488, 2000.
- [421] M. G. Gall, A. Nobili, I. V. Pavlidis, and U. T. Bornscheuer, "Improved thermostability of a Bacillus subtilis esterase by domain exchange," *Applied Microbiology and Biotechnology*, vol. 98, no. 4, pp. 1719–1726, 2014.
- [422] E. Krieger and G. Vriend, "Models@Home: distributed computing in bioinformatics using a screensaver based approach," *Bioinformatics (Oxford, England)*, vol. 18, pp. 315–8, feb 2002.

-
- [423] E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, and K. Karplus, "Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8.," *Proteins*, vol. 77 Suppl 9, pp. 114–22, jan 2009.
- [424] R. C. Edgar, "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [425] W. Li and A. Godzik, "Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics*, vol. 22, no. 13, pp. 1658–1659, 2006.
- [426] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: Accelerated for clustering the next-generation sequencing data," *Bioinformatics*, vol. 28, no. 23, pp. 3150–3152, 2012.
- [427] D. Jones, W. Taylor, and J. Thornton, "The rapid generation of mutation data matrices from protein sequences," *Computer Applications in the Biosciences*, vol. 8, no. 3, pp. 275–282, 1992.
- [428] R. K. P. Kuipers, H. J. Joosten, E. Verwiël, S. Paans, J. Akerboom, J. Van Der Oost, N. G. H. Leferink, W. J. H. Van Berkel, G. Vriend, and P. J. Schaap, "Correlated mutation analyses on super-family alignments reveal functionally important residues," *Proteins: Structure, Function and Bioinformatics*, vol. 76, no. 3, pp. 608–616, 2009.
- [429] R. K. Kuipers, H.-J. Joosten, W. J. H. van Berkel, N. G. H. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. a. P. M. dos Santos, and P. J. Schaap, "3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities.," *Proteins*, vol. 78, no. 9, pp. 2101–13, 2010.
- [430] L. Lo Leggio, T. J. Simmons, J.-C. N. Poulsen, K. E. H. Frandsen, G. R. Hemsworth, M. a. Stringer, P. von Freiesleben, M. Tovborg, K. S. Johansen, L. De Maria, P. V. Harris, C.-L. Soong, P. Dupree, T. Tryfona, N. Lenfant, B. Henrissat, G. J. Davies, and P. H. Walton, "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase," *Nature Communications*, vol. 6, p. 5961, 2015.
- [431] S. S. Strickler, A. V. Gribenko, A. V. Gribenko, T. R. Keiffer, J. Tomlinson, T. Reihle, V. V. Loladze, and G. I. Makhataдзе, "Protein stability and surface electrostatics: A charged relationship," *Biochemistry*, vol. 45, no. 9, pp. 2761–2766, 2006.
- [432] A. Horovitz, L. Serrano, B. Avron, M. Bycroft, and A. R. Fersht, "Strength and Co-operativity of contributions of surface salt bridges to protein stability," *Journal of Molecular Biology*, vol. 216, pp. 1031–1044, 1990.
- [433] K. S. Yip, K. L. Britton, T. J. Stillman, J. Lebbink, W. M. de Vos, F. T. Robb, C. Vetriani, D. Maeder, and D. W. Rice, "Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family.," *European journal of biochemistry / FEBS*, vol. 255, no. 2, pp. 336–46, 1998.
- [434] E. J. Spek, a. H. Bui, M. Lu, and N. R. Kallenbach, "Surface salt bridges stabilize the GCN4 leucine zipper," *Protein science : a publication of the Protein Society*, vol. 7, no. 11, pp. 2431–2437, 1998.
- [435] H. Leemhuis, H. J. Rozeboom, B. W. Dijkstra, and L. Dijkhuizen, "Improved thermostability of *Bacillus circulans* cyclodextrin glycosyltransferase by the introduction of a salt bridge," *Proteins: Structure, Function and Genetics*, vol. 54, no. 1, pp. 128–134, 2004.
- [436] A. Cerdobbel, K. De Winter, D. Aerts, R. Kuipers, H.-J. Joosten, W. Soetaert, and T. Desmet, "Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis.," *Protein engineering, design & selection : PEDS*, vol. 24, no. 11, pp. 829–34, 2011.
- [437] S. Kumar, C.-J. Tsai, and R. Nussinov, "Factors enhancing protein thermostability," *Protein Engineering*, vol. 13, no. 3, pp. 179–191, 2000.
- [438] A. Karshikoff and R. Ladenstein, "Ion pairs and the thermotolerance of proteins from hyperthermophiles: A 'traffic rule' for hot roads," *Trends in Biochemical Sciences*, vol. 26, no. 9, pp. 550–556, 2001.
- [439] A. K. Chaplin, M. T. Wilson, M. A. Hough, D. A. Svistunenko, G. R. Hemsworth, P. H. Walton, E. Vijgenboom, and J. A. R. Worrall, "Heterogeneity in the histidine-brace copper coordination sphere in AA10 lytic polysaccharide monooxygenases," *Journal of Biological Chemistry*, vol. 291, no. 24, pp. 12838–12850, 2016.

REFERENCES

- [440] G. Feller, "Protein stability and enzyme activity at extreme biological temperatures.," *Journal of physics. Condensed matter : an Institute of Physics journal*, vol. 22, no. 32, pp. 323101–321018, 2010.
- [441] E. B. Getz, M. Xiao, T. Chakrabarty, R. Cooke, and P. R. Selvin, "A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry.," *Analytical biochemistry*, vol. 273, no. 1, pp. 73–80, 1999.
- [442] J. R. Auclair, K. J. Boggio, G. A. Petsko, D. Ringe, and J. N. Agar, "Strategies for stabilizing superoxide dismutase (SOD1), the protein destabilized in the most common form of familial amyotrophic lateral sclerosis.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 50, pp. 21394–9, 2010.
- [443] R. Wetzel, L. J. Perry, W. a. Baase, and W. J. Becktel, "Disulfide bonds and thermal stability in T4 lysozyme.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 2, pp. 401–5, 1988.
- [444] T. Florczak, M. Daroch, M. C. Wilkinson, A. Bialkowska, A. D. Bates, M. Turkiewicz, and L. A. Iwanejko, "Purification, characterisation and expression in *Saccharomyces cerevisiae* of LipG7 an enantioselective, cold-adapted lipase from the Antarctic filamentous fungus *Geomyces* sp. P7 with unusual thermostability characteristics," *Enzyme and Microbial Technology*, vol. 53, no. 1, pp. 18–24, 2013.
- [445] D. J. Barlow and J. M. Thornton, "Ion-pairs in proteins," *Journal of Molecular Biology*, vol. 168, no. 4, pp. 867–885, 1983.
- [446] R. Sterner and W. Liebl, "Thermophilic adaptation of proteins.," *Critical reviews in biochemistry and molecular biology*, vol. 36, no. 1, pp. 39–106, 2001.
- [447] G. I. Makhatadze, V. V. Loladze, D. N. Ermolenko, X. Chen, and S. T. Thomas, "Contribution of surface salt bridges to protein stability: Guidelines for protein engineering," *Journal of Molecular Biology*, vol. 327, no. 5, pp. 1135–1148, 2003.
- [448] A. V. Gribenko, M. M. Patel, J. Liu, S. a. McCallum, C. Wang, and G. I. Makhatadze, "Rational stabilization of enzymes by computational redesign of surface charge-charge interactions.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 8, pp. 2601–2606, 2009.
- [449] W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga, "Disulfide bonds and protein folding," *Biochemistry*, vol. 39, no. 15, pp. 4207–4216, 2000.
- [450] L. Zhang, C. P. Chou, and M. Moo-Young, "Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by *Escherichia coli* expression system," *Biotechnology Advances*, vol. 29, no. 6, pp. 923–929, 2011.
- [451] J. L. Arolas, V. Castillo, S. Bronsoms, F. X. Aviles, and S. Ventura, "Designing out disulfide bonds of leech carboxypeptidase inhibitor: Implications for its folding, stability and function," *Journal of Molecular Biology*, vol. 392, no. 2, pp. 529–546, 2009.
- [452] X. Wang, S. Kumar, and S. K. Singh, "Disulfide scrambling in IgG2 monoclonal antibodies: Insights from molecular dynamics simulations," *Pharmaceutical Research*, vol. 28, no. 12, pp. 3128–3144, 2011.
- [453] S. Badiyan, D. R. Bevan, and C. Zhang, "Study and design of stability in GH5 cellulases.," *Biotechnology and bioengineering*, vol. 109, no. 1, pp. 31–44, 2012.
- [454] S. Zhang, D. C. Irwin, and D. B. Wilson, "Site-directed mutation of noncatalytic residues of *Thermobifida fusca* exocellulase Cel6B," *European Journal of Biochemistry*, vol. 267, no. 11, pp. 3101–3115, 2000.
- [455] V. Abkevich and E. Shakhnovich, "What can disulfide bonds tell us about protein energetics, function and folding: Simulations and bioninformatics analysis," *Journal of Molecular Biology*, vol. 300, no. 4, pp. 975–985, 2000.
- [456] Y. S. Nakagawa, V. G. H. Eijssink, K. Totani, and G. Vaaje-Kolstad, "Conversion of α -chitin substrates with varying particle size and crystallinity reveals substrate preferences of the chitinases and lytic polysaccharide monooxygenase of *Serratia marcescens*," *Journal of agricultural and food chemistry*, vol. 61, no. 46, pp. 11061–6, 2013.

- [457] P. H. Bessette, F. Aslund, J. Beckwith, and G. Georgiou, "Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 24, pp. 13703–13708, 1999.
- [458] J. Lobstein, C. A. Emrich, C. Jeans, M. Faulkner, P. Riggs, and M. Berkmen, "SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm," *Microbial Cell Factories*, vol. 11, no. 1, p. 1, 2012.
- [459] M. Gudmundsson, S. Kim, M. Wu, T. Ishida, M. H. Momeni, G. Vaaje-Kolstad, D. Lundberg, A. Royant, J. Ståhlberg, V. G. H. Eijsink, G. T. Beckham, and M. Sandgren, "Structural and Electronic Snapshots during the Transition from a Cu(II) to Cu(I) Metal Center of a Lytic Polysaccharide Monooxygenase by X-ray Photoreduction.," *The Journal of biological chemistry*, vol. 289, pp. 18782–92, jul 2014.
- [460] S. Mekasha, Z. Forsberg, B. Dalhus, J.-p. Bacik, S. Choudhary, C. Schmidt-dannert, G. Vaaje-kolstad, and V. G. H. Eijsink, "Structural and functional characterization of a small chitin-active lytic polysaccharide monooxygenase domain of a multi-modular chitinase from Jonesia denitrificans," *FEBS letters*, vol. 590, no. 1, pp. 34–42, 2016.
- [461] G. Vaaje-Kolstad, S. J. Horn, D. M. F. van Aalten, B. Synstad, and V. G. H. Eijsink, "The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation," *Journal of Biological Chemistry*, vol. 280, no. 31, pp. 28492–28497, 2005.
- [462] K. S. Johansen, "Discovery and industrial applications of lytic polysaccharide mono-oxygenases," *Biochemical Society Transactions*, vol. 44, no. 1, pp. 143–149, 2016.
- [463] F. J. M. Mergulhão, D. K. Summers, and G. A. Monteiro, "Recombinant protein secretion in Escherichia coli," *Biotechnology Advances*, vol. 23, no. 3, pp. 177–202, 2005.
- [464] T. V. Vuong, B. Liu, M. Sandgren, and E. R. Master, "Microplate-based detection of lytic polysaccharide monooxygenase activity by fluorescence-labeling of insoluble oxidized products," *Biomacromolecules*, vol. 18, no. 2, pp. 610–616, 2017.
- [465] A. A. Dombkowski, K. Z. Sultana, and D. B. Craig, "Protein disulfide engineering," *FEBS Letters*, vol. 588, no. 2, pp. 206–212, 2014.
- [466] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The FoldX web server: an online force field," *Nucleic Acids Research*, vol. 33, no. Web Server, pp. W382–W388, 2005.
- [467] G. Müller, D. C. Kalyani, and S. J. Horn, "LPMOs in cellulase mixtures affect fermentation strategies for lactic acid production from lignocellulosic biomass," *Biotechnology and Bioengineering*, pp. in press, DOI: 10.1002/bit.26091, 2016.
- [468] G. Courtade, S. B. Le, G. I. Sætrom, T. Brautaset, and F. L. Aachmann, "A novel expression system for lytic polysaccharide monooxygenases," *Carbohydrate Research*, p. doi: 10.1016/j.carres.2017.02.003, 2017.
- [469] R. Vanholme, B. Demedts, K. Morreel, J. Ralph, and W. Boerjan, "Lignin biosynthesis and structure.," *Plant physiology*, vol. 153, no. 3, pp. 895–905, 2010.

REFERENCES

Summary

Lignocellulosic biomass is believed to be the most promising renewable feedstock for sustainable production of biofuels and commodity chemicals in biorefineries. While the need for sustainable systems is continuously increasing, the depolymerization of lignocellulose is impeded by its high recalcitrance. This makes the development of a cost-effective process extremely challenging. In enzymatic degradation, the classical view of synergistic endo- and exocellulases working together has recently been revolutionized by the discovery of the lytic polysaccharide monooxygenases (LPMOs). These metalloproteins oxidize the glycosidic linkage, thereby rendering the crystalline substrate more amorphous and thus more amenable for traditional cellulases. In that way, LPMOs boost the efficiency of the degradation process.

It is clear that LPMOs have a huge potential for cost reduction in biomass-converting industry and the knowledge on these enzymes is quickly expanding. Nonetheless, many aspects of this interesting group of enzymes still have to be elucidated. One of the concerns with LPMOs is protein stability, which is an industrially important parameter. Processes at higher temperatures indeed offer advantages such as a reduced risk of contamination and an increased reaction rate. Therefore, this PhD thesis aims to study and increase the stability of 2 LPMOs via various protein engineering techniques. Since only two classes of LPMOs show activity on (hemi)cellulose, only these are important for the biomass industry and they are therefore the focus of this PhD thesis. More specifically, one member of auxiliary activity family 9 (AA9) and one from AA10 were selected to study. For family AA9, on the one hand, an LPMO of the well known cellulase-producing fungus *Trichoderma reesei* was examined, namely *TrCel61A*. For family AA10 on the other hand, the first described cellulose-active bacterial LPMO was studied, *i.e.* LPMO10C (formerly CelS2) from the soil bacterium *Streptomyces coelicolor*.

The first enzyme, *TrCel61A*, was heterologously expressed in the eukaryotic host *Pichia pastoris* to enable post-translational modifications. A peculiar concern in expressing an LPMO is the special characteristic that LPMOs require a histidine residue at their N-terminus (His-1) in order to remain active. To that end, different secretion signals were compared. The protein's native secretion signal outperformed the widely used α -mating factor of *S. cerevisiae* among others. The highest LPMO yield in *P. pastoris* (> 400 mg/L during fermentation) was described and a 100 % correct processing was guaranteed.

Next, the activity and stability were measured of the previously expressed protein. Phosphoric acid swollen cellulose (PASC) served as substrate in the activity test, while the analysis of the reaction products was performed by high performance anion-exchange chromatography (HPAEC, specialized HPLC). For the first time LPMO activity of *TrCel61A* on PASC was demonstrated and furthermore, the protein was found to be a type-3 LPMO,

generating neutral, C1- and C4-oxidized products.

Furthermore, the enzyme's thermostability was measured via differential scanning fluorimetry (DSF), which yields a thermal denaturation curve. The resulting parameter that can be compared between enzymes is the apparent melting temperature (T_m), which matches the inflection point. The method was found to be sensitive and reproducible with a low technical and biological variability. Moreover, for TrCel61A, an apparent melting temperature of 62 ± 1 °C was measured.

Finally, some protein engineering strategies were applied to TrCel61A. First, the catalytic domain (CD) only of the protein was expressed, but did not show a higher stability. Therefore, the full-length protein (including linker and carbohydrate binding module) was taken as starting point for engineering. Second, even though N-glycosylation and native disulfide bridges were found to contribute to the wild-type protein stability, a reverse strategy of adding extra positions for such post-translational modifications did not yield more stable variants. Finally, the principle of consensus engineering was applied after building a phylogenetic tree. However, these *de novo* designed enzymes showed a significantly decreased thermodynamic stability.

The second enzyme, ScLPMO10C, was heterologously expressed in the prokaryotic host *Escherichia coli*. Taking the His-1 requirement into account, the protein was expressed in the periplasmic space by preceding the coding sequence by its native secretion signal. ScLPMO10C was also active on PASC as a type-1 LPMO, producing only neutral and C1-oxidized products and had an apparent melting temperature of 51 ± 1 °C. Despite this rather moderate melting temperature, an unusual 34 % residual activity was measured after 2 hours incubation at 80 °C, which could be attributed to the native disulfide bridges.

Finally, ScLPMO was also subjected to protein engineering strategies in view of increasing its stability. Even though some rational mutations at the protein's surface did not have the desired effect, disulfide engineering yielded different variants with improved apparent melting temperature, ranging from +2 to +9 °C. By combining the positive disulfide introductions, the best variant obtained, displayed a +12 °C increase in T_m and was able to retain no less than 60 % of its activity after heat treatment (compared to only 34 % for the wild-type). This improvement brings the enzyme's apparent melting point to an industrially relevant temperature.

In conclusion, even though not all engineering strategies yielded more stable LPMO variants, some interesting stability characteristics were discovered for LPMOs. Furthermore, disulfide bridges are thought to be essential elements in LPMO stability.

Samenvatting

Van lignocellulose biomassa wordt algemeen gedacht dat het de meest belovende hernieuwbare grondstof is voor duurzame productie van biobrandstoffen en alledaagse chemicaliën in zogenaamde bioraffinaderijen. Terwijl de nood voor zulke hernieuwbare systemen voortdurend toeneemt, blijkt het gebruik van lignocellulose biomassa erg moeilijk door zijn weerstand tegen depolymerisatie. Hierdoor verloopt de ontwikkeling van een kostenefficiënt proces erg moeizaam. In de klassieke modellen voor enzymatische afbraak van cellulose, werken verschillende endo- en exocellulasen samen voor hydrolytische splitsing van de β -1,4-binding tussen de glucose moleculen. Recent werd dit beeld volledig herzien door de ontdekking van de lytische polysaccharide monooxygenasen (LPMO's). Deze metalloproteïnen oxideren de glycosidische binding, waardoor de kristalliene structuur verbroken wordt en dus meer toegankelijk voor de klassieke cellulasen. Op die manier maken ze het depolymerisatieproces efficiënter.

Er bestaat weinig twijfel over het enorme potentieel in kostenvermindering dat deze enzymen teweeg kunnen brengen. Hierdoor krijgen de LPMO's zeer veel aandacht en groeit de kennis over deze interessante klasse van enzymen erg snel. Vele eigenschappen en aspecten zijn totnogtoe niet bekend. Eén van deze eigenschappen is de enzym stabiliteit. Nochtans is dit een industrieel zeer interessante parameter omdat processen op hogere temperatuur een verlaagd kans op contaminatie en een verhoogde reactiesnelheid tot gevolg hebben. Daarom werd in deze doctoraatsthesis een studie gemaakt van de stabiliteit van 2 LPMO's en werden er verschillende *enzyme engineering* strategieën uitgetest om de stabiliteit te verhogen. Slechts 2 LPMO klassen zijn actief op (hemi)cellulose substraat, waardoor ze interessant zijn voor gebruik in bioraffinaderijen. Daarom werd het onderzoek enkel op deze klassen toegespitst. Het gaat over de *auxiliary activity* families 9 (AA9) en 10 (AA10). Van elke klasse werd 1 enzym bestudeerd. Voor familie AA9, werd een LPMO van *Trichoderma reesei* uitgekozen, namelijk TrCel61A. Deze natuurlijke schimmel is ervoor bekend een efficiënt cellulase systeem te produceren. Voor familie AA10, werd het eerste bacteriële LPMO uitgepikt dat actief bevonden werd op cellulose substraat. Het gaat over ScLPMO10C (vroeger ScCelS2) uit de bodembacterie *Streptomyces coelicolor*.

Het eerste enzym, TrCel61A, werd heteroloog tot expressie gebracht in de eukaryote gist *Pichia pastoris* om post-translationele modificaties mogelijk te maken. Een bijzondere eigenschap van LPMOs, waarmee men rekening moet houden tijdens productie, is dat een N-terminaal histidine residue (His-1) noodzakelijk is voor de activiteit. Om dit te bekomen werd het eiwit extracellulair tot expressie gebracht en werden er verschillende secretiesignalen vergeleken. Hieruit bleek dat het natieve secretiesignaal van TrCel61A veel beter presteerde dan onder andere het algemeen gebruikte α -secretiesignaal uit *S. cereviae*. De hoogste LPMO opbrengst in *P. pastoris* kon bekomen worden (> 400 mg/ L tijdens een fermentatie), waarbij tevens ook een correct N-terminus (His-1) gegarandeerd werd.

Vervolgens werden de activiteit en stabiliteit gemeten van *TrCel61A*. *Phosphoric acid swollen cellulose* (PASC) werd hierbij als substraat gebruikt en de resulterende reactieproducten werden geanalyseerd via *high performance anion-exchange chromatography* (HPAEC, een gespecialiseerde soort HPLC). Voor het eerst kon LPMO activiteit voor *TrCel61A* beschreven worden en bovendien kon aangetoond worden dat het om een type-3 LPMO gaat, omwille van de generatie van neutrale zowel als C1- en C4-geoxideerde producten.

Daarnaast werd de stabiliteit gemeten via *differential scanning fluorimetry* (DSF), waarbij een thermische denaturatie curve bekomen wordt. Als parameter wordt hierbij de schijnbare smelttemperatuur (T_m) vergeleken tussen enzymen. Deze parameter komt overeen met het buigpunt van de smeltcurve. De meetmethode werd als gevoelig en herhaalbaar beschouwd met een lage technische en biologische variabiliteit. Daarnaast werd voor *TrCel61A*, een schijnbare smelttemperatuur van 62 ± 1 °C opgemeten.

Ten slotte werden een aantal *enzyme engineering* strategieën toegepast op *TrCel61A*. Allereerst werd hierbij enkel het eigenlijke katalytische deel tot expressie gebracht. Aangezien dit geen verbetering in stabiliteit opbracht, werd het volledige enzym (met linker regio en koolhydraat bindende module) als startpunt voor de *engineering* genomen. Ten tweede bleek introductie van extra N-gycosylatie posities en disulfidebruggen geen verbetering in stabiliteit teweeg te brengen. Als laatste werd het principe van *consensus engineering* toegepast nadat een fylogenetische boom was opgesteld. Hoewel deze enzym varianten goed tot expressie kwamen, brachten ze een significante daling in thermodynamische stabiliteit met zich mee.

Het tweede enzym, *ScLPMO10C*, werd geproduceerd in de prokaryote gastheer *Escherichia coli*. Rekening houdend met de His-1 voorwaarde, werd het enzym naar de periplasmatische ruimte geleid door gebruik te maken van een secretiesignaal. *ScLPMO10C* vertoonde ook activiteit op PASC als type-1 LPMO, waarbij neutrale en C1-geoxideerde producten bekomen worden. Hoewel een bescheiden schijnbare smelttemperatuur van 51 ± 1 °C gemeten werd, bleek het enzym een verrassend hoge residuele activiteit (34 %) te vertonen nadat het 2 uur op 80 °C geïncubeerd was. Deze speciale eigenschap werd toegewezen aan de disulfide bruggen.

Ten slotte werden ook een aantal *engineering* strategieën toegepast met het oog op verbetering van de stabiliteit. Hoewel enkele rationele veranderingen in het enzym oppervlak niet tot een verbeterde stabiliteit leidden, kon de introductie van disulfide bruggen dat wel. Bij verschillende varianten werd een verhoging in T_m gemeten, gaande van +2 tot +9 °C. Door combinatie van deze positieve disulfide bruggen kon een sterk verbeterde variant

bekomen worden. Hiervoor werd een verbetering van +12 °C op vlak van T_m gemeten en deze variant vertoonde na 2 uur incubatie op 80 °C wel 60 % residuele activiteit (ten opzichte van 34 % voor het oorspronkelijk enzym). Deze verbetering is zeer belangrijk en brengt de smelttemperatuur naar een industrieel relevante temperatuur.

Om af te sluiten, ookal leidden niet alle *engineering* strategieën tot verhoging in stabiliteit, een aantal interessante aspecten in verband met LPMO stabiliteit konden worden blootgelegd. Daarnaast kon ook besloten worden dat disulfide bruggen essentiële elementen zijn in de stabiliteit van LPMOs.

Curriculum vitae

Personalialia

Name	Magali Tanghe
Professional address	Centre for Industrial Biotechnology and Biocatalysis Faculty of Bioscience Engineering Ghent University Coupure Links 653 9000 Ghent - Belgium
E-mail	magali.tanghe@ugent.be magali.tanghe@gmail.com
Tel.	+32 9 264 99 21
Web	www.biocatalysis.ugent.be, www.csb.ugent.be
Nationality	Belgian
Place of birth	Wilrijk (Antwerp)
Date of birth	April 12th, 1989

Education

- 2013–2017 **PhD in Applied Biological Sciences**, Faculty of Bioscience Engineering, Ghent University, Ghent
PhD thesis: Characterization and engineering of the stability of lytic polysaccharide monooxygenases.
Promoters: Prof. dr. Tom Desmet and dr. Ingeborg Stals
- 2010–2012 **Msc in Bioscience Engineering - Cell and Gene Biotechnology**, Faculty of Bioscience Engineering, Ghent University, Ghent
Msc thesis: Interaction partners of a nucleoplasmic lectin.
Promoters: Prof. dr. Els J.M. Van Damme
- Atlantis Program: 6-month exchange at Kansas State University (USA), August – December 2010
- 2007–2010 **Bsc. in Bioscience Engineering - Cell and Gene Biotechnology**, Faculty of Bioscience Engineering, University of Antwerp, Antwerp

Conferences and courses

LPMO symposium November 15th – November 18th, 2016, Copenhagen (Denmark)

Poster presentation: M. Tanghe, B. Danneels, M. Last, K. Beerens, I. Stals and T. Desmet (2016) Disulfide bridges as essential elements for the thermodynamic stability of LPMO10C from *Streptomyces coelicolor*.

11th international conference on Protein Stability (Protstab2016), May 9th – May 11th, 2016, Istanbul (Turkey)

Poster presentation: M. Tanghe, B. Danneels and T. Desmet (2016) Effect of amino-terminus and disulphide bridges on stability of a lytic polysaccharide monooxygenase.

4th international conference on Novel Enzymes October 14th – October 17th, 2014, Ghent (Belgium)

Ghent Bio-economy Summer School August 19th – August 22nd, 2013, Ghent (Belgium)

9th international Conference on Renewable Resources and Biorefineries (RRB9) June 5th – June 7th, 2013, Antwerp (Belgium)

Poster presentation: M. Tanghe, B. Danneels, I. Stals, S. De Maeseneire and T. Desmet (2013) Library construction in *Pichia pastoris* for PMO engineering.

Yeasterday 2013 May 24th, 2013, Ghent (Belgium)

Poster presentation: M. Tanghe, B. Danneels, I. Stals, S. De Maeseneire and T. Desmet (2013) Library construction in *Pichia pastoris* for PMO engineering.

3DM/YASARA/Utopia course (Bio-Product), January 21st – January 23rd, 2013, Nijmegen (The Netherlands)

International mobility

Institute of Molecular Biotechnology (Graz University of Technology) January 6th – March 21st, 2014, Graz (Austria)

CWO Grant, Ghent University (Faculty of Bioscience Engineering)

Project: Expression of a polysaccharide monooxygenases in *Pichia pastoris*

Supervisors: dr. Andrea Camattari, Prof. dr. Anton Glieder

Peer-reviewed publications

Published article

Tanghe M., Danneels B., Camattari A., Glieder A., Vandenberghe I., Devreese B., Stals I., Desmet T. (2015) Recombinant Expression of *Trichoderma reesei* Cel61A in *Pichia pastoris*: Optimizing Yield and N-terminal Processing. *Molecular Biotechnology*. 57(11-12): pp.1010-1017.

Tanghe M., Danneels B., Last M., Beerens K., Stals I., Desmet T. (2017) Disulfide bridges as essential elements for the thermostability of lytic polysaccharide monooxygenase LPMO10C from *Streptomyces coelicolor*. *Protein Engineering Design and Selection*. doi: 10.1093/protein/gzx014. pp. 1-8

Submitted articles

Danneels B., Tanghe M., Joosten H. J., Stals I., Desmet T. (2017) A Quantitative regio-selectivity indicator diagram for lytic polysaccharide monooxygenases active on cellulose. *Analytical Biochemistry*. **Submitted**

Published book chapter

Tanghe M., Danneels B., Stals I., Desmet T. (2014) Chapter 3: Enhanced biomass degradation by polysaccharide monooxygenases. *Renewable resources for biorefineries*. (RSC Green Chemistry) pp.64-78.

Student guidance

Practical courses

Microbiology (2012-2013)

Biocatalysis – structural analysis of enzymes and computer exercises (2013-2014)

Biocatalysis – bioethanol (2014-2015)

Msc thesis

Smolders L. (2013-2014): Expression of LPMOs in *P. pastoris*.

Last M. (2015-2016): Stability engineering of a bacterial LPMO.

De Munter M. (2015-2016): Stability engineering of a fungal LPMO.

